Precision Engineering 34 (2010) 807–810

Contents lists available at ScienceDirect

Precision Engineering

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

Technical note

A new approach to analysis of surface topography

Zhengkai Zhang*, Youyun Zhang, Yongsheng Zhu

Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an 710049, PR China

ARTICLE INFO

Article history:
Received 27 June 2009
Received in revised form 26 April 2010
Accepted 7 May 2010
Available online 13 May 2010

Keywords:
Profile
Surface topography
Empirical mode decomposition (EMD)
Intrinsic mode functions (IMFs)

ABSTRACT

It is well known that an engineering surface is composed of a large number of wavelengths of roughness that are superimposed on each other. This paper proposes a new method for surface topography analysis based on empirical mode decomposition. The method provides good adaptive separation of surface profile into multiple bands. Applications are conducted by using a milled engineering surfaces to demonstrate the feasibility and applicability of the empirical mode decomposition method in the analysis of engineering surfaces.

1. Introduction

It is well known that engineering surfaces are comprised of a range of spatial wavelengths. Because these multi-scale features are related to different aspects of the processes the surface has undergone and closely related to the friction and wear properties of a surface, the analysis and characterization of these features becomes an important aspect of manufacture.

In order to separate surface profile data into different wavelength components, many kinds of filtering techniques are adopted, such as 2RC filter and Gaussian filter [1]. Liu and Raja [2] and Josso et al. [3] presented a study on the application of wavelet filter for analyzing multi-scale engineering surfaces. Morphological filter was applied by Dietzsch et al. [4]. If the above methods are adopted, we should pre-set a series of parameters such as cutoff and bandwidth, and there is only little correlation between these subjective defined parameters and actual components of profile.

In 1998, Huang first proposed empirical mode decomposition (EMD) [5], which can decompose the non-linear and non-stationary signals, then give a better understanding of the physics behind the signals. The major advantage of the EMD is that the decomposition result is derived from the signal itself. Hence, the analysis is adaptive.

In this paper, EMD method is adopted to investigate the multi-scale properties of surface roughness, a milled surface profile data will be used to validate the capabilities of the EMD method.

2. A brief review of EMD method

2.1. EMD algorithm

EMD method was originally used for the analysis of vibration signal. Because the EMD method is highly efficient in non-stationary and non-linear data analysis. It has been widely applied to many kinds of signal analysis. For example, fault diagnosis of machinery [6], potential field data analysis [7]. EMD method is a self-adaptive analysis method which can decompose a complicated signal into a collection of intrinsic mode functions (IMFs) based on the local characteristic scale of the signal.

The essence of the EMD is to identify the IMF by characteristic time or spatial scales, which can be defined locally by the distance between two extrema of an oscillatory mode. The EMD picks out the highest frequency oscillation that remains in the signal. Thus, locally, each IMF contains lower frequency oscillations than the one extracted just before [8].

EMD method is developed from the simple assumption that any signal consists of different simple intrinsic modes of oscillations. Each linear or non-linear mode will have the same number of extrema and zero-crossings. There is only one extrema between successive zero-crossings. Each mode should be independent of the others [9]. In this way, each signal could be decomposed into a number of intrinsic mode functions (IMFs), each of which must satisfy the following definition [10]:

(1) In the whole data set, the number of extrema and the number of zero-crossings must either equal or differ at most by one.
(2) At any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero.

* Corresponding author. Tel.: +86 029 82669159; fax: +86 029 83237910.
E-mail address: woodncy@yahoo.com (Z. Zhang).

0141-6359/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.precisioneng.2010.05.002

The IMFs indicate the simple oscillation mode involved in the signal. EMD, a "sifting" process, is used to extract the IMFs by the following steps [10]:

(1) Identify all the extrema of the signal, and connect all the local maxima by a cubic spline line as the upper envelope. Repeat the procedure on the local minima to produce the lower envelope.
(2) Designate the mean of the two envelopes as $m_1(t)$, and the difference between the signals $x(t)$ and $m_1(t)$ as the first component, $h_1(t)$, i.e.

$$x(t) - m_1(t) = h_1(t). \tag{1}$$

(3) If $h_1(t)$ is an IMF, take it as the first IMF of $x(t)$ (the criterion of IMF is shown in Section 2.2). If $h_1(t)$ is not an IMF, take it as the original signal and repeat the steps above until $h_{1k}(t)$ is an IMF, and designate $h_{1k}(t)$ as $c_1(t)$:

$$c_1(t) = h_{1k}(t). \tag{2}$$

(4) Separate the first IMF $c_1(t)$ from $x(t)$ by

$$x(t) - c_1(t) = r_1(t). \tag{3}$$

(5) Treat residue $r_1(t)$ as the original signal and subject it to the same process as above, so that we can get other IMFs, $c_2(t)$, $c_3(t), \ldots, c_n(t)$, which satisfy

$$\left.\begin{array}{c} r_1(t) - c_2(t) = r_2(t) \\ \bullet \\ \bullet \\ \bullet \\ r_{n-1}(t) - c_n(t) = r_n(t) \end{array}\right\} \tag{4}$$

(6) By summing up Eqs. (3) and (4), we finally obtain

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t). \tag{5}$$

Thus, one can achieve a decomposition of the signal into n-empirical modes, and a residue $r_n(t)$, which is the mean trend of $x(t)$. The IMFs $c_1(t), c_2(t), \ldots, c_n(t)$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and change with the variation of signal $x(t)$. So, EMD is a self-adaptive signal decomposition method.

Empirical mode decomposition (EMD) has the feature of not assuming a time or spatial series is linear or stationary (like Fourier analysis). When dealing with surface topography data, where most variables exhibit non-linear and non-stationary behavior, this feature is particularly useful, allowing more meaningful quantification of the proportion of variance in a spatial series due to fluctuations at different spatial scales than other techniques. Furthermore, the EMD does not use any pre-determined filter or wavelet function, which means that the decomposition results are more objective and accurate than other techniques.

2.2. The criterion for stopping the sifting process

The main process of EMD decomposition is the sifting process. Through the sifting process, we can extract the IMFs from original signal. Therefore, we should set up a criterion for stopping the sifting process to ensure the IMF components retain enough physical sense. In order to ensure the orthogonality and stability of IMF, we used the energy difference tracking method as the stopping criterion for sifting process [11].

Suppose that signal $x(t)$ contains a finite number of mutually irrelevant components $x_i(t)$.

$$x(t) = x_1(t) + x_2(t) + \cdots + x_n(t) = \sum_{i=1}^{n} x_i(t). \tag{6}$$

The energy of signal $x(t)$ is calculated as

$$E_x = \int_{-\infty}^{\infty} x^2(t)\ \mathrm{d}t = \int_{-\infty}^{\infty} \left[\sum_{i=1}^{n} x_i(t)\right]^2 \mathrm{d}t. \tag{7}$$

And we define the energy of $x_i(t)$ as

$$E_i = \int_{-\infty}^{\infty} {x_i}^2(t)\ \mathrm{d}t. \tag{8}$$

Due to the irrelevance between $\{x_i(t), i = 1, 2, \ldots, n\}$, in other words, because of the orthogonality, then we can obtain

$$\int_{-\infty}^{\infty} x_i(t) x_j(t)\ \mathrm{d}t \approx 0, \quad i \neq j \tag{9}$$

And then signal $x(t)$ has total energy as

$$E_x = \int_{-\infty}^{\infty} \left[\sum_{i=1}^{n} x_i(t)\right]^2 \mathrm{d}t = \int_{-\infty}^{\infty} {x_1}^2(t)\,\mathrm{d}t + \int_{-\infty}^{\infty} {x_2}^2(t)\,\mathrm{d}t + \cdots + \int_{-\infty}^{\infty} {x_n}^2(t)\,\mathrm{d}t = E_1 + E_2 + \cdots + E_n \tag{10}$$

If EMD is used to decompose signal and it is supposed that the component $c_1(t)$ is exactly the orthogonal component $x_1(t)$ of $x(t)$, after $c_1(t)$ has been separated from $x(t)$, the residual signal energy is calculated as follows:

$$E_{2,\ldots,n} = \int_{-\infty}^{\infty} \left[\sum_{i=2}^{n} x_i(t)\right]^2 \mathrm{d}t = \int_{-\infty}^{\infty} \left[\sum_{i=2}^{n} {x_i}^2(t)\right] \mathrm{d}t \tag{11}$$

and

$$E_x = E_1 + E_{2,\ldots,n} \tag{12}$$

If $c_1(t)$ is not the orthogonal component of $x(t)$, then

$$E_x - (E_1 + E_{2,\ldots,n}) \neq 0 \tag{13}$$

We define E_{error} as follows:

$$E_{\mathrm{error}} = |E_x - (E_1 + E_{2,\ldots,n})| \tag{14}$$

Considering that there are large difference among the energy value of these IMFs, in order to facilitate follow-up analysis and calculation, we make these energy value normalized.

$$\overline{E}_{\mathrm{error}} = \frac{E_{\mathrm{error}}}{E_1} \tag{15}$$

The smaller the $\overline{E}_{\mathrm{error}}$, the more entire integrity and orthogonality the decomposed results will attain. Hence, we can track $\overline{E}_{\mathrm{error}}$ when the signal is decomposed by EMD method. When $\overline{E}_{\mathrm{error}}$ reaches a certain minimum (in this paper, we set the $\overline{E}_{\mathrm{error}} = 0.01$), sifting process is completed. Thus the obtained IMF component is an orthogonal one of the original signal.

3. The decomposition of a milled surface profile by EMD method

The measured profile of a milled surface is shown in Fig. 1 [12]. The data were measured at a 0.25 μm sampling interval for a length of 5.6 mm.