Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Intrinsic feature extraction using discriminant diffusion mapping analysis for automated tool wear evaluation*

Yi-xiang HUANG†1, Xiao LIU[2], Cheng-liang LIU[1], Yan-ming LI[1]

[1]*State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China*

[2]*Shanghai Aerospace Equipment Manufacturer, Shanghai 200245, China*

†E-mail: huang.yixiang@sjtu.edu.cn

Received Aug. 30, 2016; Revision accepted Jan. 23, 2017; Crosschecked Nov. 27, 2018

Abstract: We present a method of discriminant diffusion maps analysis (DDMA) for evaluating tool wear during milling processes. As a dimensionality reduction technique, the DDMA method is used to fuse and reduce the original features extracted from both the time and frequency domains, by preserving the diffusion distances within the intrinsic feature space and coupling the features to a discriminant kernel to refine the information from the high-dimensional feature space. The proposed DDMA method consists of three main steps: (1) signal processing and feature extraction; (2) intrinsic dimensionality estimation; (3) feature fusion implementation through feature space mapping with diffusion distance preservation. DDMA has been applied to current signals measured from the spindle in a machine center during a milling experiment to evaluate the tool wear status. Compared with the popular principle component analysis method, DDMA can better preserve the useful intrinsic information related to tool wear status. Thus, two important aspects are highlighted in this study: the benefits of the significantly lower dimension of the intrinsic features that are sensitive to tool wear, and the convenient availability of current signals in most industrial machine centers.

Key words: Tool condition monitoring; Manifold learning; Dimensionality reduction; Diffusion mapping analysis; Intrinsic feature extraction

https://doi.org/10.1631/FITEE.1601512 **CLC number:** TP277

1 Introduction

Tool condition monitoring (TCM) is a key technology in modern manufacturing processes. In the manufacturing process, tool wear caused by abrasion can severely harm the processing precision and productivity. Thus, it is necessary to develop effective TCM systems to reduce costs and improve productivity. There are two main challenges in TCM development: finding signals that are suitable to describe the tool status, and determining a feature set extracted from the signals that can effectively capture the performance degradation of the tools.

The conventional way to measure tool wear is to use some devices to measure the wear directly. A drawback of this type of direct measurement is that the tool must be taken out of circulation to be checked for tool wear. Some researchers have tried another way that is more direct to measure tool wear, which involves applying non-contact machine vision techniques (Sortino, 2003; Su et al., 2006; D'Addona and Teti, 2013; D'Addona et al., 2015, 2017). However, in vision analysis techniques, special care must be taken to avoid the interference of reflections from the cutting edge. Moreover, disturbances from cutting fluids or lubricants may severely hinder vision-based wear detection. Thus, considerable research has been devoted to finding indirect ways to conduct automated evaluation of tool wear status. Various models have

* Project supported by the National Natural Science Foundation of China (No. 51305258), the National Science and Technology Major Project, China (No. 2014ZX04015021), and the Shanghai Science Project, China (No. 1411104600)

ORCID: Yi-xiang HUANG, http://orcid.org/0000-0001-8384-1566

© Zhejiang University and Springer-Verlag GmbH Germany, part of Springer Nature 2018

been established to describe the relationships between tool wear and indirect signals such as vibration (Scheffer and Heyns, 2001; Abu-Mahfouz, 2003), acoustic emission (Lee et al., 2006; Yao and Chien, 2014), force (Huang et al., 2007; Zhou et al., 2009), temperature (Young, 1996), spindle motor current (Li and Tso, 1999; Oh et al., 2004; Franco-Gasca et al., 2006), and sensor fusion (Karam et al., 2016). To find suitable signals to describe the tool status in industrial applications, the cost, availability, and robustness of the TCM system should be considered. As mentioned, vision-based techniques may be easily interfered with and they are expensive. Vibration sensors are usually costly, and when they are installed closely or attached to the workpiece, sensor wires may interfere with the movements of the spindle and the holder during the machining process, which means an additional redesign cost and technical difficulties to reduce the interference. The cost of the acoustic emission technique is also high and it is ineffective in detecting gradual failures (Dimla, 2000). For temperature signals, the infrared thermograph camera is perhaps the most expensive sensor and can be impaired easily by lubricants. Force signals suffer from a similar problem because of the vibration affecting the sensor wires, which requires a lot of extra consideration due to possible dynamic interference. Thus, the current signals from servomotors may be the most practical variable because they are closely related to forces that are similar to torque measurements, and the current signals are usually available in most modern manufacturing systems. Furthermore, there is no problem with the dynamic interference of lubricants or sensor wires. Thus, we choose to analyze current signals in this study.

When analyzing signals, researchers have developed numerous techniques to extract the correct feature set. Abu-Mahfouz (2003) analyzed vibration signals by transforming the signals into 28 features including averaged harmonic wavelet coefficients, local peaks of the Burg power spectral density, and some statistical measures, and then used a multi-layer feed-forward neural network to detect and classify the drill wear. Scheffer and Heyns (2001) selected 13 features to identify tool status including mean, root mean square (RMS), crest factor, variance, skewness and kurtosis, spectral energies, Shannon entropy from wavelet analysis, and the coefficients for auto-regressive, moving average, and auto-regressive moving average models. Handling and processing such a large number of features would not be easy, and could lead to the problem of "the curse of dimensionality" (Korn et al., 2001). The high-dimensionality of features has always been an obstacle to efficient data processing and fusion. Traditional techniques to tackle this high-dimensionality problem include principle component analysis (PCA) (Zhou et al., 2009), independent component analysis (ICA) (Bingham and Hyvärinen, 2000), and linear discriminant analysis (LDA) (Harmouche et al., 2014). However, most of these methods are suitable for linear analysis, and cannot sufficiently capture the nonlinear characters from manufacturing systems. Recently, a new trend in dimensionality reduction techniques, the manifold learning method, has emerged and been successfully applied to condition-based monitoring (CBM) of gear or bearing systems (Jiang et al., 2009; Sipola et al., 2014). In this study, we introduce a manifold learning method to analyze current signals and fuse the features for tool wear evaluation. Discriminant diffusion maps analysis (DDMA) (Huang et al., 2013) is used for current signal analysis. This method consists of mainly three steps: (1) extraction of original features from raw signals; (2) parameter estimation of DDMA, including dimensionality estimation of the original feature space, estimation of the discriminant information, and the establishment of the affinity matrix; (3) feature fusion implementation through feature space mapping with diffusion distance preservation. Then the low-dimensional features can be input to classifiers or estimators for tool wear detection.

2 Discriminant diffusion maps analysis methodology

DDMA is developed based on the diffusion maps (DM) analysis originally proposed by Coifman and Lafon (2006). The main difference between DDMA and DM is that DDMA uses a pre-classification step to obtain the discriminant information before the calculation of dimensionality reduction, while DM neglects such information and proceeds directly to dimensionality reduction. DDMA shares a similar framework for density invariant embedding with DM.

The original idea is to define a graph of the data set. Each data point can be regarded as a vertex on the graph, and all the vertices are connected to other vertices by weighted edges.

Let the triple $(X, \mathcal{L}, \mu)$ be a measure space, where X is the data set ($X\in\mathbb{R}^D$, where D is the original dimensionality), $\mathcal{L}$ is a σ-algebra of its subsets, and μ: $\mathcal{L}\to[0, +\infty]$ is a measure, representing the distribution of the data points on the graph. The edge weight is quantified by calculating the kernel function $k(x, y)$, $\forall x, y\in X$, which satisfies:

(1) symmetry condition: $k(x, y)=k(y, x)$;

(2) positivity preservation: $k(x, y)\geq 0$.

In the standard DM analysis, the weight function $k(\cdot, \cdot)$ usually adopts some isotropic kernel, such as the Gaussian kernel. To better use the discriminant information for dimensionality reduction, DDMA uses a discriminant Gaussian kernel instead. The mathematical expression is defined as

$$k_\rho(x,y)=\exp\left(-\frac{\|x-y\|^2}{2\rho(x,y)}\right),\quad \forall x,y\in X, \tag{1}$$

where ρ is the discriminant kernel width:

$$\rho(x,y)=\begin{cases}\omega\rho_{\mathrm{W}(L)}(x,y),\ l(x)=l(y)=L,\ \forall L\in\Omega,\\ \omega^{-1}\min\{\rho_{\mathrm{W}(x)},\ \rho_{\mathrm{W}(y)},\ \rho_{\mathrm{B}}(x,y)\},\ l(x)\neq l(y),\end{cases} \tag{2}$$

where x and y are the data points, $l(x)$ and $l(y)$ are the labels of x and y, respectively, ω is a constant which may enhance the discriminant result, ρ_{W} is the kernel width for data points within the same cluster, that is, of the same label L, Ω is the whole label set of the data, and ρ_{B} is the kernel width for data points of different labels.

$$\rho_{\mathrm{W}(L)}(x,y)=\frac{1}{N_{\mathrm{A}}}\sum_{y\in\delta x}\min\|x-y\|^2_{l(x)=l(y)=L},\quad x\in X, \tag{3}$$

$$\rho_{\mathrm{B}}(x,y)=\frac{1}{A}\sum_{y\in\delta x}\min\|x-y\|^2_{l(x)\neq l(y)},\quad x\in X, \tag{4}$$

$$A=\min\{N_{\mathrm{A}},\ N_{L=l(x)},\ N_{L=l(y)}\}, \tag{5}$$

where y is within the neighborhood of x based on the affinity matrix of the data set, N_{A} is the pre-set size of the neighborhood, N_L is the number of points with the same cluster label L, and A is a number taking the minimum value from N_{A}, $N_{L=l(x)}$, and $N_{L=l(y)}$.

As mentioned previously, any data point x can be regarded as a vertex on a weighted graph, and the construction of a Markov chain on the graph of the data can be used to find relevant structures in complex geometrics, such as clusters. If the degree of any vertex is described as

$$\deg(x)=\sum_{x,y\in X}k_\rho(x,y) \tag{6}$$

and the one-step transition probability from x to y is

$$q(x,y)=\frac{k_\rho(x,y)}{\deg(x)}, \tag{7}$$

then the Markov chain can be presented by the transition matrix $\boldsymbol{Q}$ that contains all the entries of $q(x, y)$, $\forall x, y\in X$. The diffusion distance $\mathcal{D}$ between vertices is defined as

$$\mathcal{D}^2(x,y)=\|q(x,\cdot)-q(y,\cdot)\|^2_{\frac{1}{\upsilon}}=\sum_{z\in X}\frac{(q(x,z)-q(y,z))^2}{\upsilon(z)}, \tag{8}$$

where $\upsilon(z)$ is the stationary distribution of the Markov chain, given by

$$\upsilon(z)=\frac{\deg(z)}{\sum_{z'\in X}\deg(z')}. \tag{9}$$

Note that if X is a finite set and assuming that the graph is all connected, the stationary distribution $\upsilon(z)$ is unique. As mentioned in Coifman and Lafon (2006), the diffusion distance may capture some local geometric features from the data set by the kernel. The spectral properties of the Markov chain can be related to the geometry of the vertex, such as the local density distribution of the data points. Thus, the Markov chain defines fast and slow directions of propagation based on the values taken by the kernel. Based on Lafon's theory of diffusion map analysis, the discriminant kernel in DDMA further enhances the utilization of data connectivity by assigning the discriminant kernel weights along the propagation paths.

According to the proof in Nadler et al. (2006), the diffusion distance in Eq. (8) can also be written as

$$\mathcal{D}^2(x,y)=\sum_{n=1}^{D}\lambda_n^2(\boldsymbol{\phi}_n(x)-\boldsymbol{\phi}_n(y))^2, \tag{10}$$

$$\begin{cases}\boldsymbol{\varphi}_n\boldsymbol{Q}=\lambda_n\boldsymbol{\varphi}_n,\\ \boldsymbol{Q}\boldsymbol{\phi}_n=\lambda_n\boldsymbol{\phi}_n,\end{cases} \tag{11}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ are the right and left eigenvectors of transition matrix $\boldsymbol{Q}$ respectively, and λ is the eigenvalue. Eq. (10) indicates that the diffusion distance $\mathcal{D}$ can also be computed by the right eigenvectors. Here, the order of the eigenvalues is $1=\lambda_0>\lambda_1\geq\cdots\geq\lambda_{D-1}\geq 0$. Then the discriminant map can be defined as

$$\mathcal{M}: x\mapsto(\lambda_1\boldsymbol{\phi}_1(x),\lambda_2\boldsymbol{\phi}_2(x),\cdots,\lambda_D\boldsymbol{\phi}_D(x))^{\mathrm{T}}. \tag{12}$$

As mentioned earlier, the main purpose of manifold learning methods is to reduce the dimensionality of the feature space. In a traditional dimensionality reduction process, the number of reduced dimensionality can be determined arbitrarily. For example, one can select a three-dimensional space to conveniently visualize the data points in the feature space. Another popular method is to use the number of dominant eigenvalues to minimize the error function with a pre-set threshold (Coifman and Lafon, 2006). However, the disadvantage of this method is that the target reduced dimensionality is often overestimated (Hein and Audibert, 2005), and thus may not be qualified as the intrinsic dimensionality. Therefore, for manifold learning methods, researchers have tried to use fractal-based methods to estimate the intrinsic dimensionality of the data set, where the underlying dynamics within the data set can be better described by estimating the dimension based on fractals in many cases (Camastra and Vinciarelli, 2002; Kunze et al., 2012). Thus, in this study, a fractal-based dimension is adopted as the target dimensionality. Specifically, the correlation dimensionality estimation is applied because of its relatively easy calculation of active degrees of freedom, thereby providing a good measure of the complexity of the data set (Borovkova et al., 1999). A formal definition of the correlation dimension is expressed as (Camastra, 2003)

$$D_{\mathrm{r}}=\lim_{e\to 0}\frac{\ln f(e)}{\ln e}, \tag{13}$$

$$f(e)=\lim_{N\to\infty}\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}g(e-|x_i-x_j|), \tag{14}$$

$$g(\Delta)=\begin{cases}1, & \Delta>0,\\ 0, & \Delta\leq 0,\end{cases} \tag{15}$$

where e (e>0) is a pre-set threshold of the distances between pairs of data points, $f(\cdot)$ is the correlation integral function, and $g(\Delta)$ is a step function.

3 Experiment with the milling process

3.1 Experiment setup and signal acquisition

To validate DDMA in an application for tool wear detection, the method has been applied to the data set collected from a face milling experiment (Goebel and Yan, 2000), courtesy of the NASA data repository, for prognostics validation (Agogino and Goebel, 2007). In this experiment, the tools consist of six inserts of type KC710 on a 70-mm face mill, mounted on a Matsuura MC-510V CNC vertical machining center (Fig. 1).

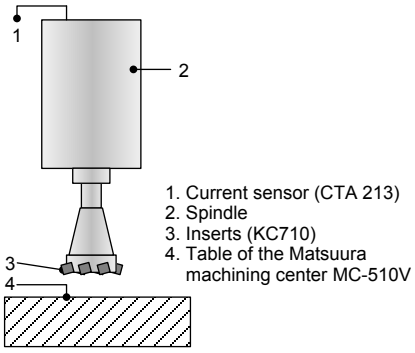

Fig. 1 Test stand for the face milling tool experiment

The processing parameters used in the experiments were set to the settings recommended by the manufacturer. There are 16 run-to-failure cases in total for the tools, with a different number of runs in each case. The depth of cut was either 1.5 mm or 0.75 mm. The cutting speed was 200 m/min. The feed was taken at either 0.5 mm/rev or 0.25 mm/rev. The adopted criterion to determine the failure status was the flank wear, which was measured after each run. The worn area could usually be divided into three zones, and zones C, B, and N are for nose wear, flank wear, and notch wear, respectively (Fig. 2). The flank

wear VB is defined as the mean width of zone *B*, that is, the average distance from the cutting edge to the end of the abrasive wear on the flank face of the tool (ISO, 1989). Although the criterion for a worn tool as recommended in the ISO standard is that the VB value exceeds the threshold of 0.35 mm, to better understand the degradation status of the tools, the milling process was continued for several runs even after the VB threshold had been reached. In the experiment, the VB values were manually measured after each run.

Different signals were collected during the milling procedure, including current signals, vibration signals, and acoustic emission signals. We focused on the current signals, as noted earlier, which can provide relatively good accessibility than other types of signals in most practical manufacturing production lines. Discussions on vibrations or acoustic emission signals are not within the scope of this study. Details can be found in Agogino and Goebel (2007). The current sensors were installed on the wire of the spindle motor. The types of Hall-effect current sensor were Omron K3TB-A1015 and CTA 213 for AC and DC signals, respectively. The data were collected at a sampling rate of 250 Hz.

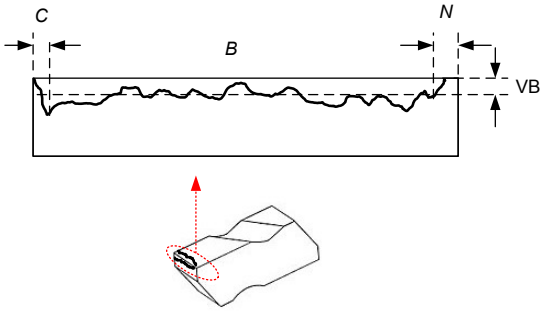

Fig. 2 Illustration of the worn area of a tool
Zones *C*, *B*, and *N* are for nose wear, flank wear, and notch wear, respectively

The application of DDMA to the tool wear data consists of the following steps (Fig. 3): (1) data cleansing; (2) feature extraction; (3) intrinsic dimensionality estimation; (4) discriminant analysis; (5) diffusion distance calculation; (6) construction of low-dimensional embedding; (7) tool wear evaluation using the intrinsic features. Steps (4)–(7) are the main DDMA.

3.2 Data cleansing and preprocessing

To validate the DDMA methodology, all 16 cases were used with 167 milling runs in total. Due to corrupt or missing records in some cases, it is necessary to clean and preprocess the data first. The cleansing and preprocessing procedures for the data are listed as follows:

1. Removing the abnormal data

The current signals were almost zeros during the 95th run. For the 18th run, the current signals were corrupted by incorrect signals; that is, the values of which are much larger or smaller than the range of the sensor. Thus, the data from these two runs were removed before analysis, and the rest of the milling runs were re-indexed from 1 to 165, as well as their case numbers.

2. Handling the missing data

There are several missing VB values, such as in

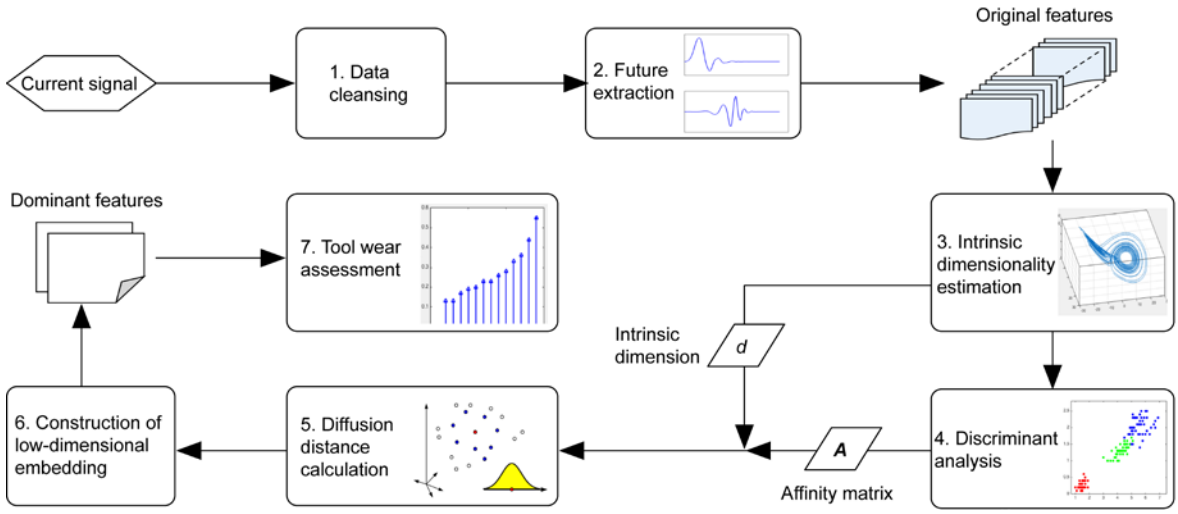

Fig. 3 Flowchart of the discriminant diffusion maps analysis (DDMA) for tool wear evaluation

the 2nd and 3rd runs. The linear interpolation technique was used to fill the unknown gaps. Fig. 4 shows the VB curves for these 16 cases after interpolation. Note that there is only one zero-value VB sample in case 13.

3. Obtaining the discriminant information

The measured VB values were divided into five discrete levels of tool wear (Table 1).

4. Aligning the starting points of the milling processes

The 1.1-fold maximum value of the first 100 DC current signal samples was used as the threshold to determine the starting point of each run, and the signals were discarded before the starting point.

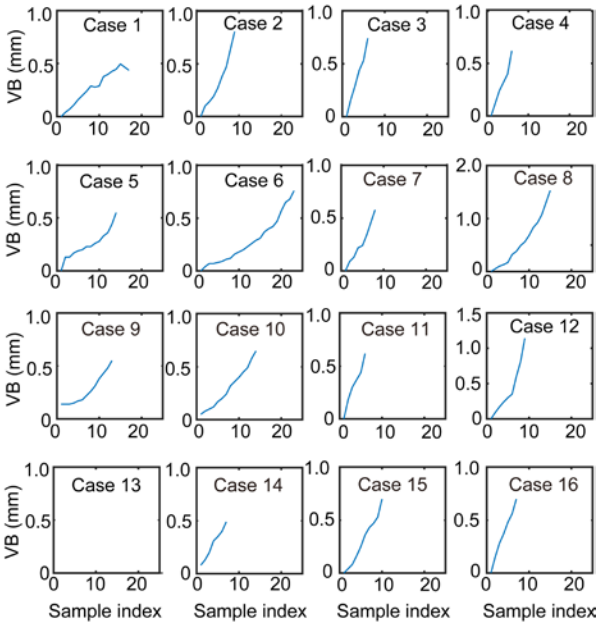

Fig. 4 VB curves of the milling processing cases 1–16

Table 1 Tool wear levels

Wear level	VB range (mm)
1	VB≤0.15
2	0.15 <VB≤0.30
3	0.30 <VB≤0.45
4	0.45 <VB≤0.60
5	0.60 <VB

3.3 Feature extraction

The typical waveforms of the current signals are shown in Fig. 5, corresponding to different levels of measured VB values. The signals were processed by lifting wavelet packet analysis, and several statistics were extracted from the wavelet packet coefficients. The wavelet lifting scheme, proposed by Sweldens (1998), improves the computational efficiency through three steps: splitting, predicting, and updating. The splitting step divides the signals into even or odd series. The predicting and updating steps can be written in a factorization form. A biorthogonal wavelet with orders of vanishing moments, three for

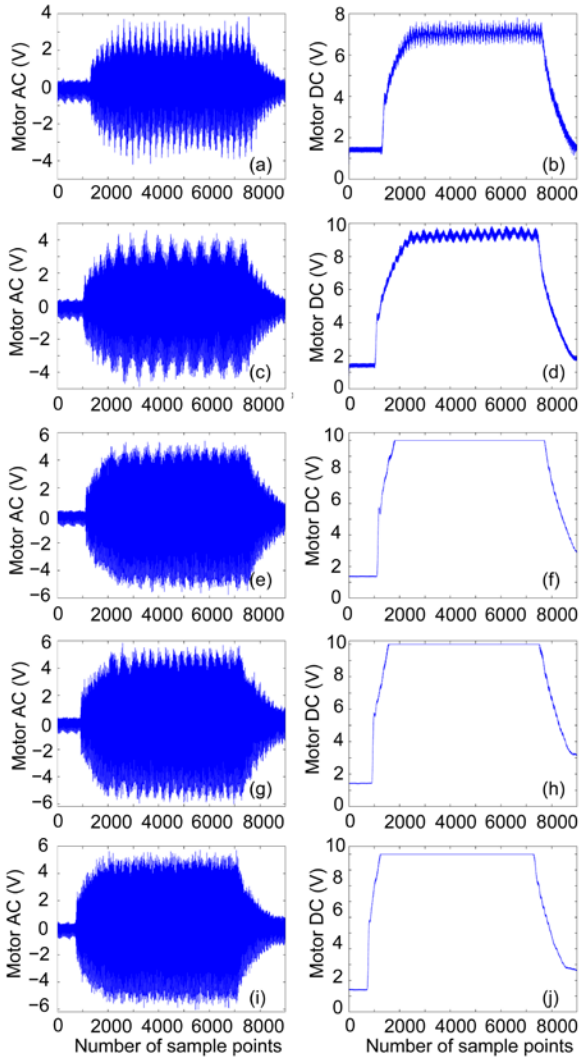

Fig. 5 Typical current signal waveforms for the spindle drive

(a) AC signals when VB≤0.15 mm; (b) DC signals when VB≤0.15 mm; (c) AC signals when 0.15 mm<VB≤0.30 mm; (d) DC signals when 0.15 mm<VB≤0.30 mm; (e) AC signals when 0.30 mm<VB≤0.45 mm; (f) DC signals when 0.30 mm<VB≤0.45 mm; (g) AC signals when 0.45 mm<VB≤0.60 mm; (h) DC signals when 0.45 mm<VB ≤0.60 mm; (i) AC signals when 0.60 mm <VB; (j) DC signals when 0.60 mm<VB

reconstruction and one for decomposition, was chosen for the wavelet packet analysis. Its filters for decomposition were [−0.3536, 1.0607, 1.0607, −0.3536] and [−0.1768, 0.5303, −0.5303, 0.1768]. The corresponding lifting scheme can be represented by its factorization form as follows:

$$H(z)=\begin{bmatrix}0.471 & 0\\ 0 & 2.121\end{bmatrix}\cdot\begin{bmatrix}1 & 0\\ 0.333z & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 1.125+0.375z^{-1}\\ 0 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0\\ -0.444 & 1\end{bmatrix}. \tag{16}$$

The decomposition level was five, which can serve a general purpose. We selected the values of the mean F_1, variation F_2, skewness F_3, the kurtosis F_4 of the wavelet coefficients from each node of the wavelet packet tree as the feature vectors (Eqs. (17)–(20)), and applied a z-score method to normalize the features (Eq. (21)). Thus, the dimension of the original feature space extracted from the DC and AC current signals was $2^5\times4\times2=256$ (Fig. 6).

$$F_1=\frac{1}{N}\sum_{n=1}^{N}x_n, \tag{17}$$

$$F_2=\left[\frac{1}{N}\sum_{n=1}^{N}(x_n-F_1)^2\right]^{\frac{1}{2}}, \tag{18}$$

$$F_3=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x_n-F_1}{F_2}\right)^3, \tag{19}$$

$$F_4=\frac{\frac{1}{N}\sum_{n=1}^{N}(x_i-F_1)^4}{\left(\frac{1}{N}\sum_{n=1}^{N}(x_i-F_1)^2\right)^2}, \tag{20}$$

$$y_n=\frac{x_n-F_1}{F_2}. \tag{21}$$

3.4 Intrinsic dimensionality estimation

As shown in Fig. 6, it is not easy to evaluate and differentiate the worn tool status given the original features. Therefore, the proposed DDMA method was used to enhance the useful information within the feature space. As described above, we first estimated the intrinsic dimensionality of the original feature space using its correlation dimension. To estimate the dimension and obtain the discriminant information, we divided the whole data set into a training set of 25% and a testing set of 75% of the data set. The selection process was random. The average result was 2.4336, rounded to the nearest smaller integer 2.

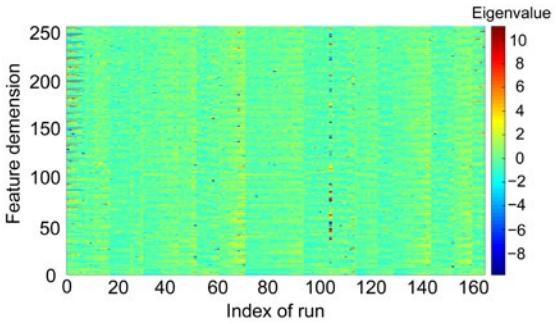

Fig. 6 Original features extracted from 165 runs with AC and DC signals

3.5 Main discriminant diffusion maps analysis

Next, the DDMA method described in Section 2 was applied to the original feature vectors, and their corresponding intrinsic features were calculated in the dimensionality-reduced space. Denote the original feature vectors as $\boldsymbol{X}$:

$$\boldsymbol{X}=(y_{i,j,k})_{i=1,2,\,j=1,2,3,4,\,k\in Z,\,k\in[1,32]}, \tag{22}$$

where the normalized feature columns y are defined in Section 3.3, i indicates that the feature is from the AC (i=1) or DC (i=2) signals, j indicates that the value is of mean (j=1), variation (j=2), skewness (j=3), or kurtosis (j=4) calculated from Eqs. (17)–(21), and k indicates the features from the k^{th} node of the wavelet packet tree after the five-level decomposition. Then discriminant analysis was applied on the feature set by training the multi-class support vector machine model to generate the discriminant labels, with which the parameter ρ in Eq. (1) can be selected according to Eqs. (2)–(5). Then the Markov matrix with all entries of the diffusion distances can be calculated from Eqs. (6)–(9). Thus, the right eigenvectors of the Markov matrix can be obtained and multiplied by the eigenvalues to generate the diffusion coordinates (Eqs. (11) and (12)). Because the intrinsic dimensionality was estimated to be two, we retained the first two coordinates as the intrinsic feature vectors.

Fig. 7a clearly shows that the two-dimensional intrinsic feature vectors are gathered to form several

clusters where most of them are consistent with their true tool wear levels. The results from PCA, which is a popular dimensionality reduction technique, are given in Fig. 7b. It can be observed that the feature vectors of different tool wear levels are tangled and difficult to differentiate.

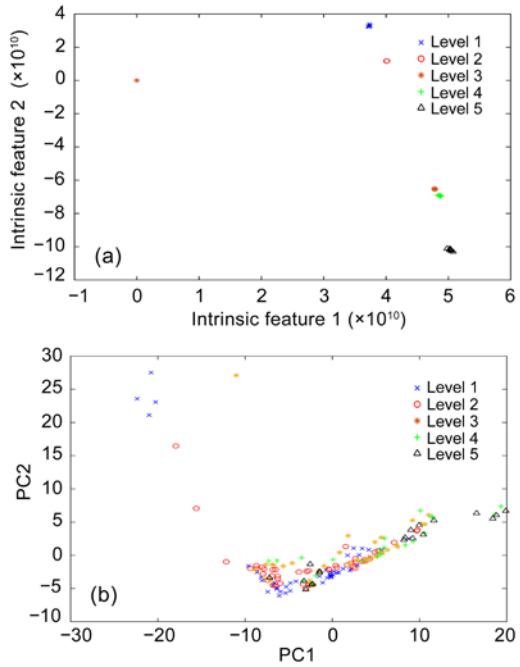

Fig. 7 Two-dimensional intrinsic features marked with true tool wear levels extracted by DDMA (a) and PCA (b)

3.6 Tool wear evaluation using intrinsic features

Furthermore, to precisely evaluate the tool wear status, the VB values were divided into five discrete levels as mentioned. Then a simple nearest neighbor classifier was used to classify the levels of tool wear and evaluate the tool status. Of all 16 cases, 25% were randomly selected as the training set, and the rest of the cases were the test set. The results are shown in Fig. 8. Cases 1, 2, 6, and 11 were the training set, including 55 runs. The accuracy of the tool wear evaluation by DDMA was 100% for the training set, and 99.09% for the test set, which validates that the DDMA gave the preferable property with very little loss of useful information during dimension reduction and intrinsic feature transformation processes. On the other hand, the accuracy of the PCA method was 27.27% for the training set, and 26.36% for the test set, which indicates a poor ability to preserve information within the original features of the current signals.

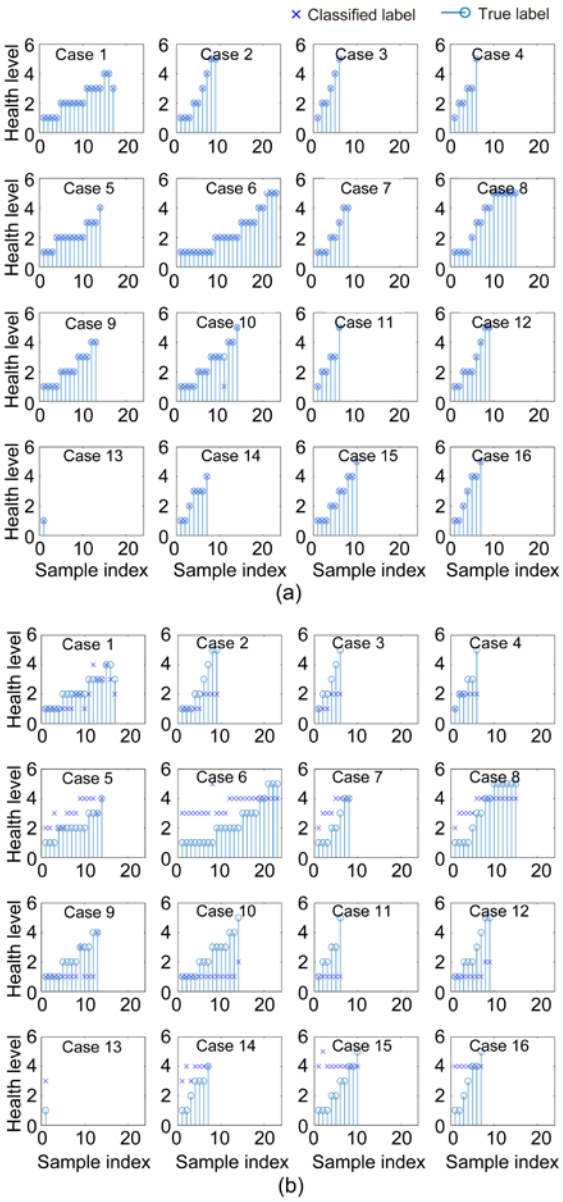

Fig. 8 Results of the tool wear evaluation based on intrinsic features transformed by DDMA (a) and PCA (b)

4 Conclusions

The DDMA technique has been studied for intrinsic feature extraction in evaluating tool wear. In principle, the method applies the Markov random walks theory to a graph and uses a discriminant kernel scheme to refine the high-dimensional feature space. The performance of the proposed DDMA method has been evaluated on AC and DC current signals collected from a milling process experiment, which can

be accessed easily in industrial machine centers. The results of the experiment showed that DDMA can transform the original high-dimensional features to an intrinsic feature space with a significantly lower dimension for the machining processes.

References

Abu-Mahfouz I, 2003. Drilling wear detection and classification using vibration signals and artificial neural network. *Int J Mach Tools Manuf*, 43(7):707-720. https://doi.org/10.1016/S0890-6955(03)00023-3

Agogino A, Goebel K, 2007. Milling Data Set. Available from https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#milling

Bingham E, Hyvärinen A, 2000. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int J Neur Syst*, 10(1):1-8. https://doi.org/10.1142/S0129065700000028

Borovkova S, Burton R, Dehling H, 1999. Consistency of the Takens estimator for the correlation dimension. *Ann Appl Prob*, 9(2):376-390. https://doi.org/10.1214/aoap/1029962747

Camastra F, 2003. Data dimensionality estimation methods: a survey. *Patt Recogn*, 36(12):2945-2954. https://doi.org/10.1016/S0031-3203(03)00176-6

Camastra F, Vinciarelli A, 2002. Estimating the intrinsic dimension of data with a fractal-based method. *IEEE Trans Patt Anal Mach Intell*, 24(10):1404-1407. https://doi.org/10.1109/TPAMI.2002.1039212

Coifman RR, Lafon S, 2006. Diffusion maps. *Appl Comput Harmon Anal*, 21(1):5-30. https://doi.org/10.1016/j.acha.2006.04.006

D'Addona DM, Teti R, 2013. Image data processing via neural networks for tool wear prediction. *Proc CIRP*, 12:252-257. https://doi.org/10.1016/j.procir.2013.09.044

D'Addona DM, Matarazzo D, Ullah AMMS, et al., 2015. Tool wear control through cognitive paradigms. *Proc CIRP*, 33: 221-226. https://doi.org/10.1016/j.procir.2015.06.040

D'Addona DM, Ullah AMMS, Matarazzo D, 2017. Tool-wear prediction and pattern-recognition using artificial neural network and DNA-based computing. *J Intell Manuf*, 28(6):1285-1301. https://doi.org/10.1007/s10845-015-1155-0

Dimla EDS, 2000. Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. *Int J Mach Tools Manuf*, 40(8):1073-1098. https://doi.org/10.1016/S0890-6955(99)00122-4

Franco-Gasca LA, Herrera-Ruiz G, Peniche-Vera R, et al., 2006. Sensorless tool failure monitoring system for drilling machines. *Int J Mach Tools Manuf*, 46(3-4):381-386. https://doi.org/10.1016/j.ijmachtools.2005.05.012

Goebel K, Yan W, 2000. Feature selection for tool wear diagnosis using soft computing techniques. ASME Int Mechanical Engineering Congress and Exposition, p.157-163.

Harmouche J, Delpha C, Diallo D, 2014. Linear discriminant analysis for the discrimination of faults in bearing balls by using spectral features. Int Conf on Green Energy, p.182-187. https://doi.org/10.1109/ICGE.2014.6835419

Hein M, Audibert JY, 2005. Intrinsic dimensionality estimation of submanifolds in $\mathbb{R}^d$. Int Conf on Machine Learning, p.289-296. https://doi.org/10.1145/1102351.1102388

Huang SN, Tan KK, Wong YS, et al., 2007. Tool wear detection and fault diagnosis based on cutting force monitoring. *Int J Mach Tools Manuf*, 47(3-4):444-451. https://doi.org/10.1016/j.ijmachtools.2006.06.011

Huang YX, Zha XF, Lee J, et al., 2013. Discriminant diffusion maps analysis: a robust manifold learner for dimensionality reduction and its applications in machine condition monitoring and fault diagnosis. *Mech Syst Signal Process*, 34(1-2):277-297. https://doi.org/10.1016/j.ymssp.2012.04.021

ISO, 1989. Tool life testing in milling—part 1: face milling, ISO 8688-1:1989. International Organization for Standardization, Geneva.

Jiang QS, Jia MP, Hu JZ, et al., 2009. Machinery fault diagnosis using supervised manifold learning. *Mech Syst Signal Process*, 23(7):2301-2311. https://doi.org/10.1016/j.ymssp.2009.02.006

Karam S, Centobelli P, D'Addona DM, et al., 2016. Online prediction of cutting tool life in turning via cognitive decision making. *Proc CIRP*, 41:927-932. https://doi.org/10.1016/j.procir.2016.01.002

Korn F, Pagel BU, Faloutsos C, 2001. On the "dimensionality curse" and the "self-similarity blessing". *IEEE Trans Knowl Data Eng*, 13(1):96-111. https://doi.org/10.1109/69.908983

Kunze H, Torre DL, Mendivil F, et al., 2012. Fractal-Based Methods in Analysis. Springer, New York, USA, p.1-16. https://doi.org/10.1007/978-1-4614-1891-7

Lee DE, Hwang I, Valente CMO, et al., 2006. Precision manufacturing process monitoring with acoustic emission. *Int J Mach Tools Manuf*, 46(2):176-188. https://doi.org/10.1016/j.ijmachtools.2005.04.001

Li XL, Tso SK, 1999. Drill wear monitoring based on current signals. *Wear*, 231(2):172-178. https://doi.org/10.1016/S0043-1648(99)00130-1

Nadler B, Lafon S, Coifman RR, et al., 2006. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Appl Comput Harmon Anal*, 21(1):113-127. https://doi.org/10.1016/j.acha.2005.07.004

Oh YT, Kwon WT, Chu CN, 2004. Drilling torque control using spindle motor current and its effect on tool wear. *Int J Adv Manuf Technol*, 24(5-6):327-334. https://doi.org/10.1007/s00170-002-1490-0

Scheffer C, Heyns PS, 2001. Wear monitoring in turning operations using vibration and strain measurements. *Mech Syst Signal Process*, 15(6):1185-1202. https://doi.org/10.1006/mssp.2000.1364

Sipola T, Ristaniemi T, Averbuch A, 2014. Gear classification and fault detection using a diffusion map framework. *Patt*

Recogn Lett, 53:53-61. https://doi.org/10.1016/j.patrec.2014.10.019

Sortino M, 2003. Application of statistical filtering for optical detection of tool wear. *Int J Mach Tools Manuf*, 43(5): 493-497. https://doi.org/10.1016/S0890-6955(02)00266-3

Su JC, Huang CK, Tarng YS, 2006. An automated flank wear measurement of microdrills using machine vision. *J Mater Process Technol*, 180(1-3):328-335. https://doi.org/10.1016/j.jmatprotec.2006.07.001

Sweldens W, 1998. The lifting scheme: a construction of second generation wavelets. *SIAM J Math Anal*, 29(2): 511-546. https://doi.org/10.1137/S0036141095289051

Yao CW, Chien YX, 2014. A diagnosis method of wear and tool life for an endmill by ultrasonic detection. *J Manuf Syst*, 33(1):129-138. https://doi.org/10.1016/j.jmsy.2013.05.003

Young HT, 1996. Cutting temperature responses to flank wear. *Wear*, 201(1-2):117-120. https://doi.org/10.1016/S0043-1648(96)07227-4

Zhou JH, Pang CK, Lewis FL, et al., 2009. Intelligent diagnosis and prognosis of tool wear using dominant feature identification. *IEEE Trans Ind Inform*, 5(4):454-464. https://doi.org/10.1109/TII.2009.2023318