J Intell Manuf (2019) 30:657–669
https://doi.org/10.1007/s10845-016-1272-4

Tool wear monitoring in ultrasonic welding using high-order decomposition

Yaser Zerehsaz[1] · **Chenhui Shao**[2] · **Jionghua Jin**[1]

Received: 25 February 2016 / Accepted: 22 October 2016 / Published online: 1 November 2016

Abstract Ultrasonic welding has been used for joining lithium-ion battery cells in electric vehicle manufacturing. The geometric profile change of tool shape significantly affects the weld quality and should be monitored during production. In this paper, a high-order decomposition method is suggested for tool wear monitoring. In the proposed monitoring scheme, a low dimensional set of monitoring features is extracted from the high dimensional tool profile measurement data for detecting tool wear at an early stage. Furthermore, the proposed method can be effectively used to analyze the data cross-correlation structure in order to help identify the unusual wear pattern and find the associated assignable cause. The effectiveness of the proposed monitoring method was demonstrated using a simulation and a real-world case study.

Keywords Ultrasonic metal welding · Tool wear monitoring · High-order representation · Principal component analysis (PCA) · High-order singular value decomposition (HOSVD)

Introduction

Ultrasonic welding has been well used for joining lithium-ion battery cells in electric vehicle manufacturing. Ultrasonic welding tool wear has a significant impact on the weld quality of lithium-ion batteries (Shao et al. 2016), because the failure of a single weld may lead to the malfunction of the entire battery pack. To ensure the weld quality, a commonly used preventive maintenance practice in most plants is to replace tools when the number of welding operations reaches a preset limit. Such limit is often set very conservatively, resulting in waste of useful tool life. Therefore, an accurate tool wear monitoring algorithm is critically needed to reduce the unnecessary maintenance cost induced by inappropriate tool replacements.

Tool wear monitoring has received tremendous attention over the last several decades. Most of papers have focused on machining, where some sensing signals, such as vibration, force, acoustic emission, and electric current are often collected and analyzed. A data transform, like fast Fourier transformation (FFT), wavelet decomposition, principal component analysis (Jolliffe 2005), and dominant feature identification (DFI) can be applied to extract relevant features for tool life prediction, classification or tool wear monitoring. For instance, Zhou et al. (2011) used the DFI method to extract the features from an acoustic emission signal. An autoregressive moving average (ARMA) model was then utilized for predicting tool wear in a ball-nose cutter of a high-speed milling machine. Shi and Gindy (2007) employed the PCA method on multiple sensor signals represented by a data matrix for the purpose of extracting important features related to tool wear. A least squares support vector machine (LS-SVM) method was further used to build a tool wear prediction model based on the extracted features. The methodology was applied to predict the wear in the teeth of a high-speed steel broaching tool. Li et al. (1999) applied a discrete wavelet transform to extract the features from acoustic emission and electric current signals. The extracted features were used for detecting the breakage of the tool in a drilling machine. A thorough review on different tool wear

✉ Yaser Zerehsaz
yzereh@umich.edu

[1] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA

[2] Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

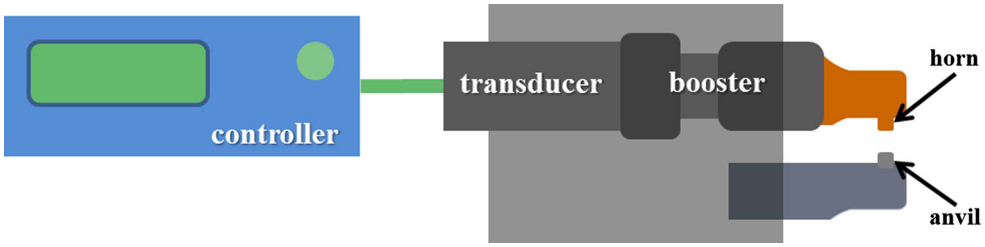

Fig. 1 Ultrasonic welding system (Shao et al. 2013)

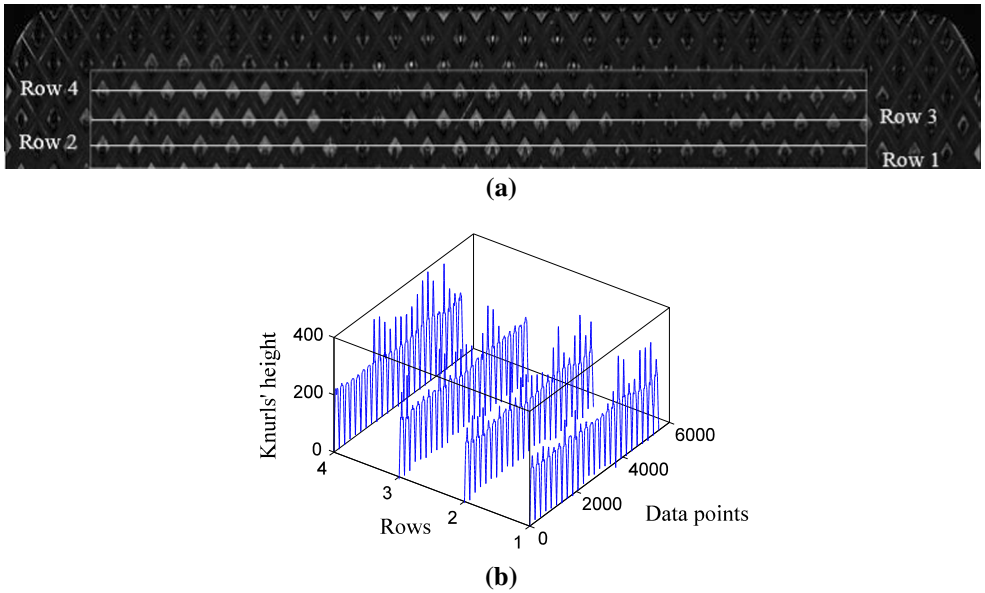

Fig. 2 Optical image of an anvil and knurl profiles. **a** Optical image of an anvil, **b** cross-sectional view of knurls profiles at the highlighted 4 rows

monitoring techniques can be found in Abellan and Subiron 2010.

Tool wear monitoring of ultrasonic welding is more challenging than that of machining processes, as the former tool wear mechanism is more complicated and has not been thoroughly understood. In ultrasonic welding, high-frequency energy is used to produce acoustic sound waves. The generated vibration produces oscillating shears between thin metal sheets held together under a clamping force perpendicular to the interface between the workpieces. The clamping force and the oscillation lead to deforming the material and forming the final joining. Because of the high frequency oscillation and workpiece material deformation, a small relative movement of the contacting surfaces between the tool (anvil in Fig. 1) and the workpiece is inevitable. This interaction will lead to anvil wear during the repeated welding operations.

Figure 1 shows major components in an ultrasonic welding machine consisting of a controller, a transducer, a booster, a horn and an anvil. Both horn and anvil contain pyramid-shaped teeth called knurls. The tool wear in ultrasonic welding is characterized based on the profile change of knurls' shape. Figure 2a shows an optical image of knurls in an anvil. The highlighted area (4 rows) shows the major welding area that is underneath the horn pads. Figure 2b gives the cross-sectional view of the knurl profiles at the highlighted area.

The first purpose of this paper is to develop an effective monitoring strategy for early detection of tool wear based on the knurls profile measurements (e.g. Fig. 2b) in an ultrasonic welding process. The second aim of the paper is to analyze tool wear variation pattern and to identify unusual patterns caused by the misalignment of the anvil with respect to the horn. This would help improve the tool setup to increase the tool life. In the remainder of this section, existing research papers related to tool wear monitoring for ultrasonic metal welding will be reviewed; moreover, the research challenges in using multidimensional data to early detect tool wear and unusual wear pattern will be summarized.

Some previous research was conducted to use indirect process sensing signals for the tool wear monitoring in the ultrasonic welding. Shao et al. (2014) studied the relationship between the online vibration signals and tool conditions for predicting the remaining lifetime of an anvil. They constructed a prediction model using the FFT method to extract

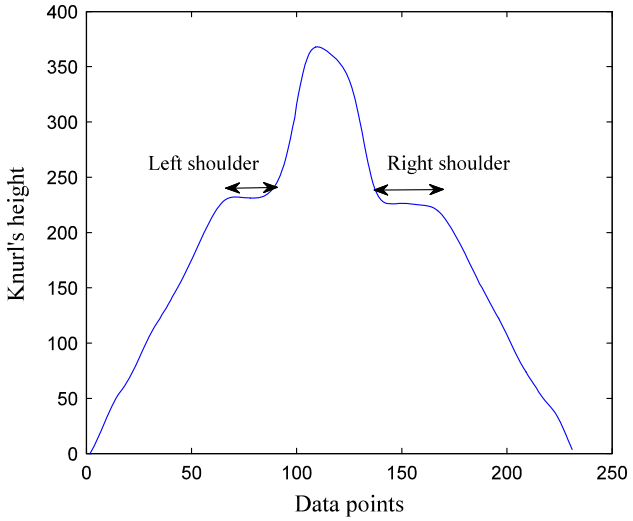

Fig. 3 A knurl with shoulders appearing on both sides

the dominant frequency of vibration signals as the response variable and the number of welds performed by the anvil as the predictor. The remaining tool life was predicted based on the prediction of the dominant frequency feature exceeding a preset threshold. This approach has two major drawbacks: one problem is that the preset threshold is sensitive to the materials and operating conditions. The other drawback is that even if the indirect process sensing signal shows dramatic change when the knurls are worn, it is impossible to analyze the spatial cross-correlation among knurls on the anvil. In practice, the information of wear pattern in the anvil would help find the root causes of unusual tool wear at an early stage to prevent severe production loss.

Shao et al. (2016) used the FFT method to extract the monitoring features from the knurls profiles measurements. They subsequently built various classifiers on the extracted features to classify each anvil to one of four wear statuses predefined based on operating experience and engineering knowledge. Instead of their classification approach, another alternative way is to simply build a monitoring control chart based on the extracted FFT features (Kisic et al. 2015). However, some early slight wear appearing on the knurls' profile shoulders (Fig. 3) or the knurls' peaks may not significantly affect the FFT features. Furthermore, if only one or a few knurls are worn in the same column of knurls, and the rest of knurls are normal, the FFT features would not significantly change. Therefore, the FFT features are not sensitive in detecting the early or local wear of an anvil.

In addition to the FFT features, two other features, which can be directly extracted from the knurl profile, were suggested by Shao et al. (2016) for classification. The first feature was the variability of the knurls' heights. A decreasing trend in the knurls' height variability indicates that the knurls start to wear off. The second feature was the average of the widths of the right and left shoulders as shown in Fig. 3. The larger these shoulders are the more the knurls are worn. As it is shown in "Monitoring performance comparison" section, using these simple features for tool wear monitoring is not always effective. This inefficacy reveals itself, for instance, when the peaks are slightly worn, or small wear occurs at only one side of a knurl (one shoulder). For the aforementioned reasons, an alternative method with the ability of detecting slight wear or asymmetric wear in the shape of knurls is needed (please see section "Monitoring performance comparison" for performance comparison).

There are two key challenges in establishing an appropriate methodology for monitoring and analysis of tool wear in an ultrasonic metal welding process: the monitoring method must be able to promptly detect slight wear in the shape of knurls. The other challenge is to have the ability to systematically analyze spatial cross-correlation among different rows of knurls profiles. Wear pattern analysis is very useful in identifying the associated root causes. For example, as shown in Fig. 2, if the first two rows have more wear than that of Row 3 and Row 4, this may indicate a misalignment of the anvil with respect to the horn. Early correction of such a tool misalignment problem will help save production loss.

To understand the shape difference between normal and worn knurls, a set of normal knurls profiles and worn knurls profiles are shown in Fig. 4a, b, respectively. It can be clearly seen that normal knurls have a consistent profile shape change that is quite different from that of worn knurls. Therefore, the monitoring method can be set based on the change pattern (variation pattern) of the knurls profile shapes.

PCA using singular value decomposition is one possible solution to analyze the multivariate variation pattern. In order to use PCA, the dataset must be represented by a matrix. For the tool wear analysis, the knurls' profiles measured at one specific sampling time on an anvil should be sequentially stacked up and represented by a row vector. Hence, different rows of the matrix represent different anvil samples. As discussed before, the spatial correlation among the knurls at different rows of an anvil should be considered in the analysis. However, stacking up all knurls' profiles will destroy such a spatial cross-correlation structure. To overcome this shortcoming, a high-order array called tensor is suggested in the paper to represent the multidimensional structure of tool wear data. Specifically, the ultrasonic tool wear dataset is represented by a three-way tensor including three dimensions defined as follows: the first dimension is the positional row index of knurls, the second dimension is the positional column index of knurls, and third dimension is the data point index of each knurl profile. Using such a multidimensional representation, the spatial correlation structure among different rows is clearly preserved. Analysis of this spatial correlation pattern can facilitate process fault diagnosis. For instance, if the knurls profiles wear in the first and

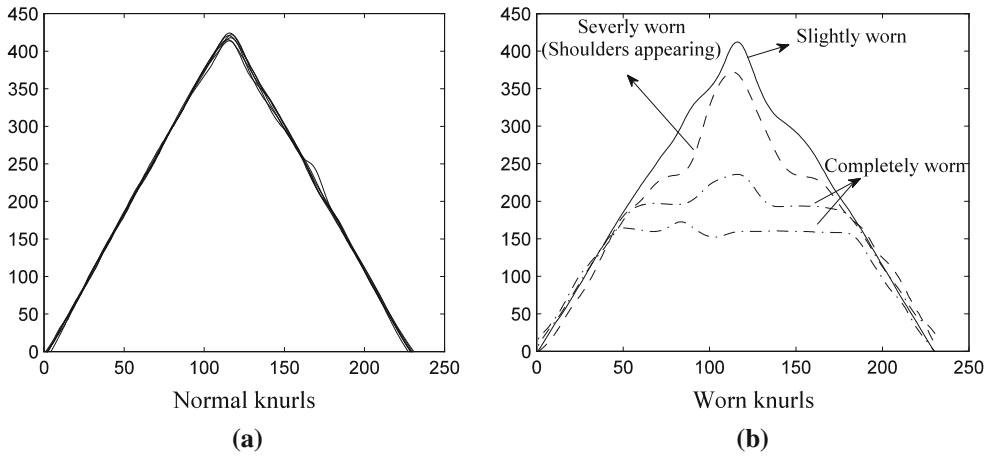

Fig. 4 Illustration of normal and worn knurl profiles. **a** Overlap of four healthy knurls on an anvil, **b** Overlap of four worn out knurls on an anvil

second rows are positively correlated, it means that the knurls in these two rows are getting worn with a similar pattern. On the other hand, a negative correlation between two rows indicates that one row is getting worn faster than the other row. This unusual wear pattern is an indicator for the existence of a tool installation problem.

Several authors have favored the use of a multidimensional array (tensor) representation when the dataset has more than two dimensions (He et al. 2005; Paynabar et al. 2013; Yan et al. 2015). In this paper, a high-order decomposition method called high-order singular value decomposition (HOSVD) is used to factorize the tensor representing the data and extract effective monitoring features afterwards. A multivariate monitoring chart will be constructed based on those extracted features.

To sum up, the method suggested in this paper aims at systematically analyzing the variation pattern of knurls profiles wear and extracting effective monitoring features for tool wear monitoring and root cause inference. In order to capture the spatial cross-correlation structure of the knurls among different rows, a high-order array is employed to represent the knurls' profile measurements. For variation pattern analysis of such a tensorial dataset, the HOSVD method is employed. The HOSVD method can systematically check whether spatial correlation exists among rows of knurls in the anvil (please see section "Comparison of HOSVD and PCA in capturing the correlation structure"). Furthermore, a T^2 control chart is constructed to monitor the extracted features based on the HOSVD results. This will be discussed in detail in section "HOSVD method and HOSVD-based T^2 control chart construction".

The reminder of this paper is organized as follows: section "A brief review of tensor data representation and the HOSVD-based T^2 control chart" provides a brief review of different strategies including a high-order array for representing the data; additionally, the HOSVD method and the multivariate T^2 control chart constructed based on the extracted monitoring features are discussed. A simulation study is conducted in section "Performance comparison of the HOSVD, PCA and FFT methods" to show the superiority of the HOSVD method to the PCA and FFT methods in two aspects, namely, (1) effectively capturing the true variation pattern of the data, and (2) accurately detecting slight/early wear in the knurls. A case study is also performed in section "Case study" to demonstrate the effectiveness of the proposed method. The concluding remarks are presented in section "Conclusion".

A brief review of tensor data representation and the HOSVD-based T^2 Control Chart

In this section, firstly, different data representation strategies are discussed. The basic notation relating to the tensor representation used throughout the paper is introduced in section "Data representation and basic high-order algebraic operations" followed by some useful multilinear algebraic operations. Secondly, the HOSVD method and the T^2 control chart constructed upon the extracted features from HOSVD are elaborated in section "HOSVD method and HOSVD-based T^2 control chart construction".

Data representation and basic high-order algebraic operations

Table 1 lists all the symbols used in this paper. Two fairly different strategies are used to represent the tool wear data. A simple matrix representation called low-order representation is a classical way to represent a dataset. A matrix $\mathbf{X} \in \mathbb{R}^{24\times 924}$ is used to represent the knurls' profiles shown in Fig. 2. In the following analyses, this is called Representation I. The number of rows in this matrix indicates

Table 1 List of symbols

x	Scalers
$\mathbf{x}$	Vectors
$\mathbf{X}$	Matrices
$\mathbf{X}^{(n)}$	The nth matrix in a group of matrices
$\mathcal{X}^{I_1 \times I_2 \times \ldots \times I_N}$	An N-dimensional (Nth-order) tensor
I_n	The number of elements in the nth order
$x_{i_1 i_2 \ldots i_n}$	$i_1 i_2 \ldots i_n$ th element in tensor $\mathcal{X}^{I_1 \times I_2 \times \ldots \times I_N}$

that each anvil has 24 columns of knurls, and the number of columns ($4 \times 231 = 924$) represents the stacked-up rows of knurls' profiles with each knurl profile having 231 data points. As discussed before, the spatial cross-correlation among different rows of knurls on an anvil breaks down in Representation I. Alternatively, a different low-order representation $\mathbf{Y} \in \mathbb{R}^{96 \times 231}$ denoted as Representation II can also be used, where each knurl profile is considered as a row in matrix $\mathbf{Y}$, and 96 ($24 \times 4 = 96$) rows of matrix $\mathbf{Y}$ correspond to stacked-up rows of 24 columns of knurls. In this way, the cross-correlations among different rows and different columns are mixed together and cannot be distinguished.

The tool wear dataset shown in Fig. 1b can be, furthermore, represented by a 3rd-order tensor $\mathcal{X}^{4 \times 231 \times 24}$ with 4 rows as Mode 1, 231 data points of each knurl profile as Mode 2, and 24 columns of knurls on an anvil as Mode 3. Some commonly used multilinear algebraic operations for tensors are introduced as follows:

(1) *Tensor matricization* It is used to transform a tensor into a matrix. Specifically, for mode-n matricization, each row vector of the new matrix is obtained by stacking up the elements of all other modes, and the number of rows is equal to I_n, i.e., the resultant matrix $\mathbf{X}^{(n)}$ has the dimension of $I_n \times I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N$. That is, when matricizing a tensor, the $i_1 i_2 \ldots i_N$th element in the tensor $\mathcal{X}^{I_1 \times I_2 \times \ldots \times I_N}$ is mapped to element (i_n, j) of matrix $\mathbf{X}^{(n)}$, where $j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k$ and $J_k = \sum_{\substack{m=1 \\ m \neq n}}^{k-1} I_m$.

(2) *Tensor vectorization* Tensor vectorization is simply the process of rearranging the tensor $\mathcal{X}^{I_1 \times I_2 \times \ldots \times I_N}$ to produce a vector $\mathbf{v}$ with size $I_1 I_2 \ldots I_N$.

(3) *Tensor-matrix product* It generally yields a new tensor. Specifically, mode-n product of a tensor $\mathcal{X}^{I_1 \times I_2 \times \ldots \times I_N}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ is denoted by $\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N}$. Each element of the resulting new tensor is obtained as

$$\mathcal{Y}_{i_1 i_2 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \ldots i_N} u_{j i_n} \quad (1)$$

where $u_{j i_n}$ is element (j, i_n) in matrix $\mathbf{U}$.

(4) *Tensor-vector product* It is a special case of tensor-matrix product. Specifically, the mode-n product of a tensor $\mathcal{X}^{I_1 \times I_2 \times \ldots \times I_N}$ by a vector $\mathbf{v} \in \mathrm{R}^{I_n}$ is defined as $(\mathcal{X} \times_n \mathbf{v})_{i_1 i_2 \ldots i_{n-1} i_{n+1} \ldots i_N} = \sum_{i_n}^{I_n} x_{i_1 i_2 \ldots i_N} v_{i_n}$, where v_{i_n} is the i_nth element of vector $\mathbf{v}$. This mode-n tensor-vector product changes the tensor order from N into $N-1$. For a comprehensive review regarding the multilinear algebra and high-order decomposition methods, the interested readers can refer to Kolda and Bader (2009).

HOSVD method and HOSVD-based T^2 control chart construction

The general idea in HOSVD known as Tucker decomposition is to factorize a tensor $\mathcal{X}^{I_1 \times I_2 \times \ldots I_N}$ into a core tensor denoted by $\boldsymbol{\Psi}^{k_1 \times k_2 \times \ldots \times k_N}$ using a set of factor matrices $\mathbf{V}^{(n)} \in \mathbb{R}^{I_n \times k_n}$; $n = 1, 2, \ldots, N$, where $k_n < I_n$ is the number of components in mode n. In the context of ultrasonic welding dataset, a tensor $\mathcal{X}^{4 \times 231 \times 24}$ is used to represent the dataset. The factor matrices $\mathbf{V}^{(1)} \in \mathbb{R}^{4 \times k_1}$, $\mathbf{V}^{(2)} \in \mathbb{R}^{231 \times k_2}$ and $\mathbf{V}^{(3)} \in \mathbb{R}^{24 \times k_3}$ are the matrices of singular vectors for Mode 1, Mode 2 and Mode 3, respectively. Note that the HOSVD method enables the practitioners to choose different number of components k_n for each mode n, and the number of components can be determined based on a predefined threshold of the explained variability. To obtain the optimal orthogonal singular vectors for HOSVD, the objective function defined in Eq. (2) is to minimize the L_2 norm of the residuals which are the difference between the original tensor $\mathcal{X}$ and the approximated tensor $\hat{\mathcal{X}} = \boldsymbol{\Psi} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \times_3 \ldots \times_N \mathbf{V}^{(N)}$ (De Lathauwer et al. 2000); i.e.,

$$\begin{aligned} &\text{Minimize } \|\mathcal{X} - \boldsymbol{\Psi} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \times_3 \ldots \\ &\quad \times_N \mathbf{V}^{(N)}\|_2^2, \text{ or equivalently} \\ &\text{Maximize } \boldsymbol{\Psi} = \mathcal{X} \times_1 \mathbf{V}^{(1)^T} \times_2 \mathbf{V}^{(2)^T} \times_3 \ldots \times_N \mathbf{V}^{(N)^T} \\ &\text{s.t: } \mathbf{V}^{(n)^T} \mathbf{V}^{(n)} = \mathbf{I} \text{ for } n = 1, 2, \ldots, N \end{aligned} \quad (2)$$

Unlike the regular low-order SVD, there is no closed-form solution for the high-order decomposition problem in Eq. (2). In practice, the problem is solved iteratively by fixing all factor matrices but one factor matrix in mode n; $n = 1, 2, \ldots, N$, and obtaining the solution for mode-n's factor matrix. This procedure can be repeated for all other factor matrices in different modes. Figure 5 gives the high-order orthogonal iterations (HOOI) algorithm proposed by De Lathauwer et al. (2000).

Initialize $\mathbf{V}^{(n)} \in \mathbb{R}^{I_n \times k_n}$, *for* n =1, 2, ..., N

Repeat until convergence or maximum iteration reached

For n = 1, 2, ..., N *do*

$\mathcal{G} = \mathcal{X} \times_1 \mathbf{V}^{(1)^T} \ldots \times_{n-1} \mathbf{V}^{(n-1)^T} \times_{n+1} \mathbf{V}^{(n+1)^T} \ldots \times_N \mathbf{V}^{(N)^T} \in \mathbb{R}^{k_1 \times k_2 \times \ldots k_{n-1} \times I_n \times k_{n+1} \times \ldots k_N}$

matricize tensor $\mathcal{G}$, *apply SVD, and choose* $\mathbf{V}^{(n)}$ *to be the left leading* k_n *singular vectors*

End for

$\boldsymbol{\Psi} = \mathcal{X} \times_1 \mathbf{V}^{(1)^T} \times_2 \mathbf{V}^{(2)^T} \ldots \times_N \mathbf{V}^{(N)^T}$

Convergence criterion: in each iteration k, stop if for all n =1, 2, ..., N, $\left\|\mathbf{V}_k^{(n)} - \mathbf{V}_{k-1}^{(n)}\right\|_2 < \varepsilon$, *where* ε *is a small value*

Return $\boldsymbol{\Psi}, \mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(N)}$

End procedure

Fig. 5 High-order orthogonal iteration algorithm (De Lathauwer et al. 2000)

After decomposing the original tensor of data $\mathcal{X}^{4\times231\times24}$ ($I_1 = 4, I_2 = 231$, and $I_3 = 24$), the monitoring features can be obtained by projecting the centralized tensor $\mathcal{X}^{4\times231\times24}$ using the knurls profiles factor matrix $\mathbf{V}^{(2)} \in \mathbb{R}^{231\times k_2}$; that is, the tensor of scores is obtained as

$$\mathcal{Z} = \mathcal{X} \times_2 \mathbf{V}^{(2)^T} \in \mathbb{R}^{4\times k_2\times 24} \tag{3}$$

where $\mathcal{Z}$ is the tensor of scores. The number of components k_n; $n = 1, 2, 3$, can be determined based on the explained variance defined as $\frac{\|\boldsymbol{\Psi}\|_2^2}{\|\mathcal{X}\|_2^2}$, where $\boldsymbol{\Psi} = \mathcal{X} \times_1 \mathbf{V}^{(1)^T} \times_2 \mathbf{V}^{(2)^T} \ldots \times_N \mathbf{V}^{(N)^T}$ is the core tensor, $\mathcal{X}$ is the original tensor representing the data, $\|\mathcal{X}\|_2^2$ is the squared Frobenious norm of tensor $\mathcal{X}$, and it is defined as $\|\mathcal{X}\|_2^2 = \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2$. Tensor $\mathcal{Z}$ can be transformed into a matrix $\mathbf{Z} \in \mathbb{R}^{24\times 4k_2}$ via tensor matricization over Mode 3. Rows of matrix $\mathbf{Z}$ are considered as samples used to estimate the mean vector and covariance matrix of the extracted features.

As discussed, the construction of the HOSVD-based T^2 control chart involves two steps. (1) Using the algorithm in Fig. 5 to estimate the matrix of singular vectors (factor matrices) based on the data from normal anvils. The features are computed as in Eq. (3), and a $24 \times 4k_2$ matrix of features is obtained. (2) An appropriate control chart is constructed for monitoring the extracted features. Since there is usually more than one feature, a multivariate T^2 control chart is set up to monitor these $p = 4k_2$ extracted features. The details will be shown in the case study provided in Section "Conclusion". Let $\mathbf{z}_i \in \mathbb{R}^{4k_2\times 1}$ denote the ith row of matrix $\mathbf{Z}$, then the T_i^2 statistic can be computed as

$$T_i^2 = (\mathbf{z}_i - \bar{\mathbf{z}})^T \mathbf{S}^{-1} (\mathbf{z}_i - \bar{\mathbf{z}})\,; i = 1, 2, \ldots, I_3 = 24 \tag{4}$$

where the jth element of mean vector $\bar{\mathbf{z}}$ is $\frac{\sum_{i=1}^{24} z_{ij}}{24}$, z_{ij} is the jth element of the vector $\mathbf{z}_i$ for $j = 1, 2, \ldots, p = 4k_2$, and $\mathbf{S} \in \mathbb{R}^{p\times p}$ is the sample covariance matrix of features estimated using in-control 24 samples represented by matrix $\mathbf{Z}$ (Hotelling 1947). Provided that the features follow a normal distribution, the upper control limit is computed as UCL $= \frac{p(I_3+1)(I_3-1)}{I_3(I_3-p)} f_{\alpha,p,I_3-p}$ in Phase II and UCL $= \frac{(I_3-1)^2}{I_3}\beta_{\alpha,p/2,(I_3-p-1)/2}$ for Phase I, where f_{α,ν_1,ν_2} is the $(1-\alpha)100^{\text{th}}$ percentile of an F distribution with ν_1 and ν_2 degrees of freedom, and $\beta_{\alpha,a,b}$ is the $(1-\alpha)100$th percentile of a Beta distribution with parameters a and b (Tracy et al. 1992). In case the features do not follow a normal distribution, the upper control limit can be obtained using the empirical distribution of T_i^2's. When the PCA transform is used for a low-order data representation, a similar procedure is followed for constructing a T^2 chart using the transformed PC features.

Performance comparison of the HOSVD, PCA and FFT methods

In general, there are two key aspects to be considered when analyzing tool wear data in an ultrasonic welding process. The first important issue is studying the correlation structure of the knurls across different rows of an anvil. This is helpful in understanding the knurls' wear pattern on an anvil. For this purpose, section "Comparison of HOSVD and PCA in capturing the correlation structure" shows the superiority of HOSVD to PCA in capturing the true correlation structure of the data. The second favorable aspect is that the monitoring method can effectively detect worn knurls at an early wear level. Section "Monitoring performance comparison" provides a performance comparison among four control charts established based on the HOSVD, PCA, FFT and the knurls' height methods.

Comparison of HOSVD and PCA in capturing the correlation structure

In this subsection, a surrogated simulation study is conducted to compare the performance of HOSVD and PCA in explaining the spatial cross-correlation among the rows of knurls on an anvil. Since FFT does not provide any information regarding the variation pattern of the data, it is excluded from this comparison. A set of tensorial data represented by tensor $\mathcal{X} \in \mathbb{R}^{4\times231\times24}$ is generated to represent an anvil having 4 rows of knurls with each row containing 24 columns of knurls. Each knurl is represented by a profile having 231 data points. The elements of tensor $\mathcal{X}$ are denoted by x_{ijk} corresponding to the jth point of a knurl profile located on Row i and Column k on an anvil. The simulation condition is set such that Row 1 and Row 2 have the same high variance. Row 3 and Row 4 have equal low variances. Furthermore, Row 1 and Row 2 are positively correlated while there is a negative correlation between Row 3 and Row 4. These parameters are specifically represented by

$$\sigma^2_{R_1} = \sigma^2_{R_2} > \sigma^2_{R_3} = \sigma^2_{R_4}, \quad \rho_{R_1R_2} > 0 \text{ and } \rho_{R_3R_4} < 0 \tag{5}$$

where $\sigma^2_{R_i}$ is the variance of Row i, and $\rho_{R_iR_j}$ is the correlation between Row i and Row j. In this simulation, these parameters are specified as $\sigma^2_{R_1} = \sigma^2_{R_2} = 100 > \sigma^2_{R_3} = \sigma^2_{R_4} = 60$, $\rho_{R_1R_2} = \rho$, $\rho_{R_3R_4} = -\rho$, and $\rho = 0.5$. In order to simulate the knurls profiles, a mixed-effect model is used, which is defined as

$$\mathbf{y}^{(i)}_k = \mathbf{B}^{(i)}\left(\boldsymbol{\beta}^{(i)} + \mathbf{r}^{(i)}_k\right) + \boldsymbol{\varepsilon}^{(i)}_k \quad k = 1, 2, \ldots, 24, i = 1, 2, 3, 4 \tag{6}$$

where $\mathbf{y}^{(i)}_k$ is the 231×1 vector of the kth knurl profile in Row i, $\mathbf{B}^{(i)}$ is the $231 \times L$ matrix of B-spline basis values with L knots. The vector of fixed-effect coefficients is denoted by $\boldsymbol{\beta}^{(i)} \in \mathbb{R}^{L\times1}$, and it is computed as follows: a set of normal knurls is selected from the dataset, and each knurl profile is regressed on B-spline basis values stored in matrix $\mathbf{B}^{(i)}$. The coefficients are computed using the least-square method. The mean vector of the computed coefficients is considered as the vector of fixed-effect B-spline coefficients $\boldsymbol{\beta}^{(i)}$. $\mathbf{r}^{(i)}_k \in \mathbb{R}^{L\times1}$ is the vector of random-effect coefficients that are normally distributed with a zero mean vector and the covariance matrix $\mathbf{U} \in \mathrm{R}^{L\times L}$ with elements given in Eq. (5). The vector of random errors $\boldsymbol{\varepsilon}^{(i)}_k \in \mathbb{R}^{231\times1}$ is assumed to follow a normal distribution with a zero mean vector and the diagonal covariance matrix $\sigma^2_\varepsilon\mathbf{I}$. In this paper, the random errors' variability is $\sigma^2_\varepsilon = 0.1$ and $\mathbf{I} \in \mathbb{R}^{231\times231}$ is an identity matrix. An important property of the mixed-effect model in Eq. (6) is that it has a high flexibility in modeling the cross-correlation among different rows of an anvil.

Table 2 Percentage of explained variance by the components of HOSVD and PCA methods

Number of components	Method		
	HOSVD (%)	PCA I (%)	PCA II (%)
1	30.22	27.33	40.33
2	24.66	23.40	28.3
3	12.05	15.61	21.01
Total explained variability	66.93	66.34	89.64

Both the HOSVD and PCA methods are applied to the simulated data. Table 2 gives the percentage of explained variance by the first three components of each method. PCA I and PCA II denote the PCA method applied to Representation I ($\mathbf{X} \in \mathbb{R}^{24\times924}$) and Representation II ($\mathbf{Y} \in \mathbb{R}^{96\times231}$) defined in section "Data representation and basic high-order algebraic operations", respectively.

Figure 6a shows the first singular vector of HOSVD method for Mode 1, where Row 1 and Row 2 have the largest weights among all 4 rows. This result is expected because the first two rows are set to have the highest variability in the simulation. Moreover, the simulated positive correlation between Row 1 and Row 2 is well captured by this singular vector. The second singular vector computed by HOSVD is plotted in Fig. 6b. It can be observed that the HOSVD method gives almost zero weights to Row 1 and Row 2, while Row 3 and Row 4 have a large weight with an opposite sign reflecting the negative correlation between Row 3 and Row 4.

Figure 7 uses the eigentensors of HOSVD to illustrate the variation pattern in Mode 1 and Mode 2 simultaneously, where the eigentensor for the lth component is calculated as $\mathbf{v}^{(1)} \circ \mathbf{v}^{(2)}$ with $\mathbf{v}^{(1)} \in \mathbb{R}^{4\times1}$ and $\mathbf{v}^{(2)} \in \mathbb{R}^{231\times1}$ corresponding to the lth columns of factor matrices $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$, respectively, and "$\circ$" denoting the outer product between two vectors. As expected, in the first eigentensor (Fig. 7), Row 1 and Row 2 have the highest weights with a positive correlation, while in the second eigentensor shown in Fig. 7b, Row 3 and Row 4 are highlighted with a negative correlation.

For comparison, the PCA method is further applied to the simulated dataset with Representation I ($\mathbf{X} \in \mathbb{R}^{24\times924}$) and Representation II ($\mathbf{Y} \in \mathbb{R}^{96\times231}$). The first and second eigenvectors estimated by PCA I are plotted in Fig. 8a, b, respectively. Different from the HOSVD's results in Figs. 7 and 8 shows that both eigenvectors of PCA I reflect the large variability of Row 1 and Row 2, and the negative correlation between Row 3 and Row 4 is not obviously reflected by the second eigenvector. This shows that PCA cannot fully capture the correlation of Rows 3 and 4.

Figure 9 shows the first and second eigenvectors of PCA II results. As expected, PCA II cannot provide any information regarding the correlation structure of the rows due to the way the data are represented.

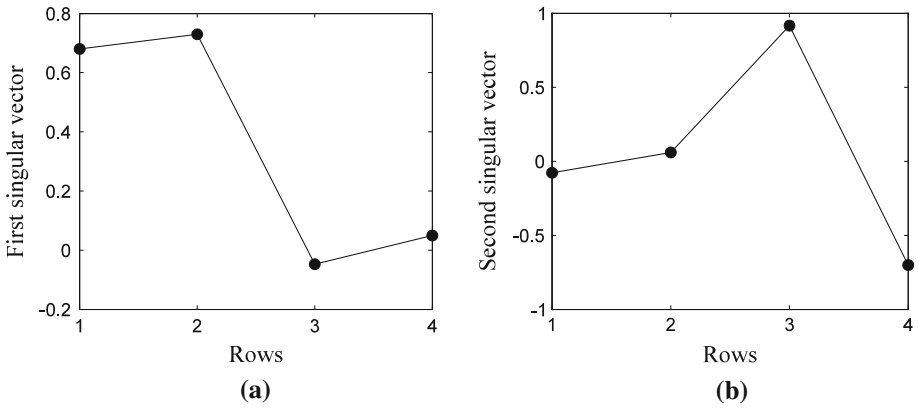

Fig. 6 First and second singular vectors computed by HOSVD method. **a** First Mode-1 singular vector, **b** second Mode-1 singular vector

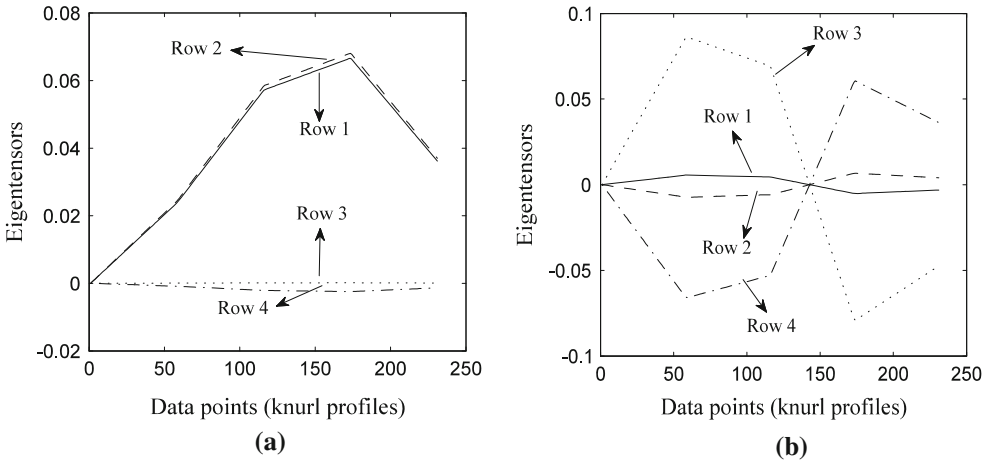

Fig. 7 Eigentesors estimated by the HOSVD method. **a** First eigentensor, **b** second eigentesor

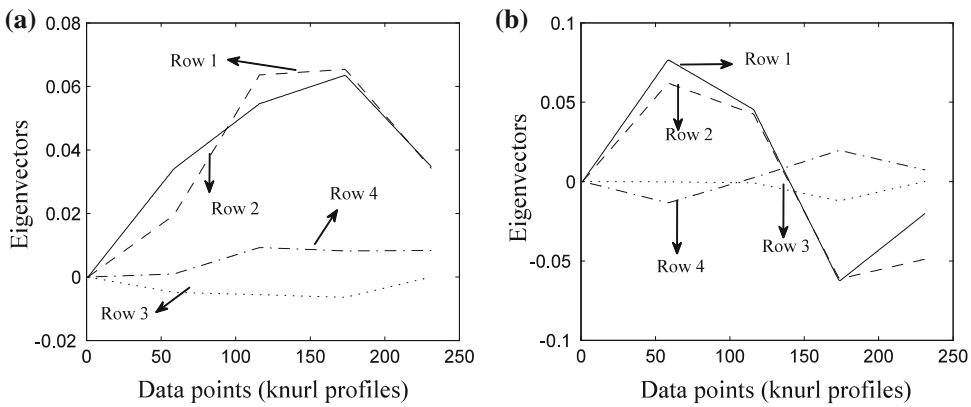

Fig. 8 First and second eigenvectors estimated by PCA I. **a** First eigenvector obtained using PCA, **b** second eigenvector obtained using PCA

The above comparison suggests that the HOSVD method has been able to more effectively capture the true spatial cross-correlation structure among the rows of knurls on an anvil. In practice, this correlation structure is sensitive to the relative position of the anvil to the workpiece/horn in the ultrasonic welding machine, and it can reflect the fact whether the anvil is properly set up or not. Therefore, the HOSVD method is suggested in this paper to monitor the tool wear with a diagnostic capability of identifying the anvil's misalignment problem.

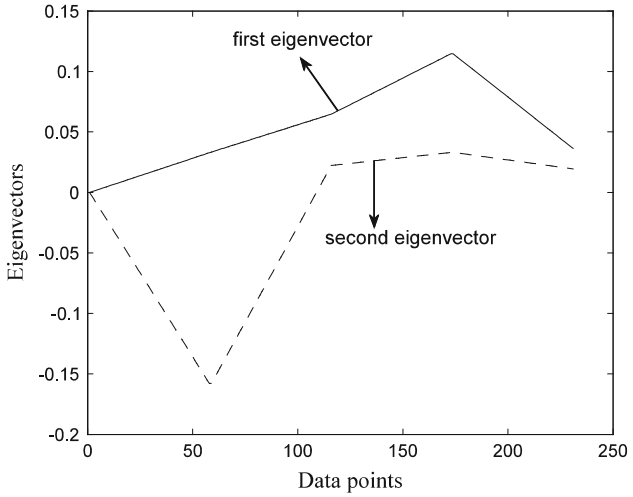

Fig. 9 First and second eigenvectors estimated by PCA II

Monitoring performance comparison

A 3D microscope was utilized to measure the tool surfaces, and the knurl heights were extracted from the surface measurements. In this subsection, four different methods for monitoring the tool wear data are compared: (1) the HOSVD-based T^2 control chart, (2) PCA-based control chart, (3) FFT-based control chart, and (4) a control chart to monitor the knurls' height. The FFT method is applied to each column containing four knurls on an anvil in order to extract the frequency-domain features for representing the repeated pattern. Frequency-domain features are the amplitudes corresponding to the dominant frequencies obtained from FFT method applied to each column of knurls profiles. Figure 10a shows one column of 4 knurls profiles, and the frequency-domain profile is plotted in Fig. 10b with sampling frequency set to 5.5. For the fourth method, to calculate the height of the knurls profiles, firstly, a general baseline is determined

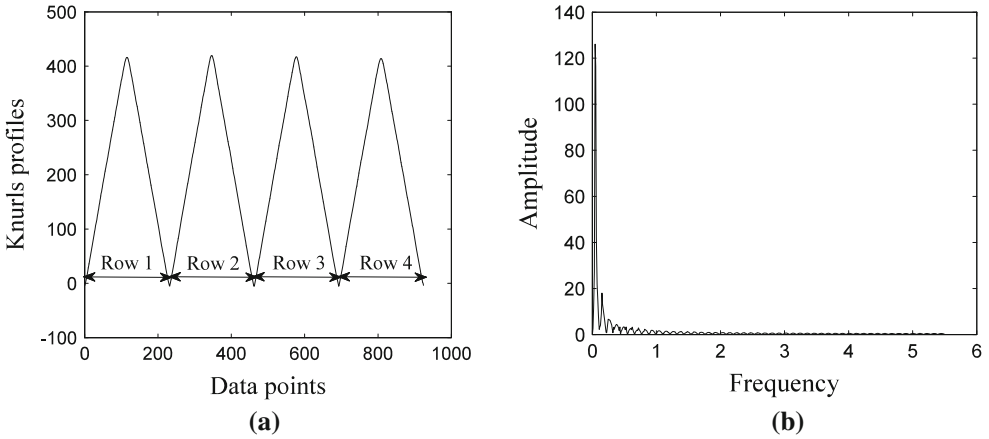

Fig. 10 A column of knurls for FFT and FFT result. **a** Knurls profiles in one column, **b** frequency-domain profile

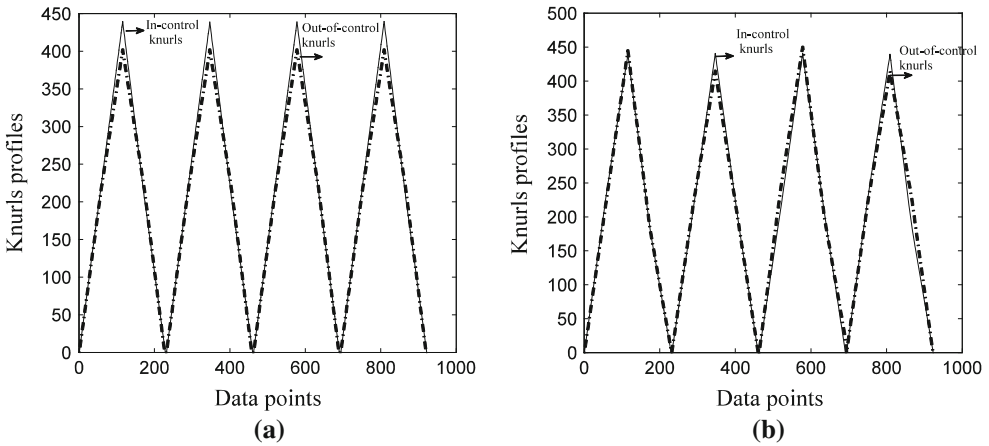

Fig. 11 Simulated in-control and out-of-control knurls in a column of anvil. **a** Scenario I: all knurls are worn, **b** Scenario II: two out of four knurls are worn

as the lowest point of knurls profiles on an anvil. Secondly, using this baseline, for the normal knurls, the height of the knurls is calculated as the distance between the peak (the highest point on the knurl profile) and the baseline point. For a worn knurl with shoulders as shown in Figure 3, the height is the difference between the lower shoulder and the baseline. Finally, for a completely worn knurl, the height is the distance between the flat line on the top and the baseline.

The monitoring performance of all methods is assessed using a simulation study performed in two different scenarios. In the first scenario (Scenario I) as shown in Fig. 11a, all four rows of knurls located in one column are worn, while in Scenario II as shown in Fig. 11b, two rows of knurls are completely healthy and the other two rows of knurls are slightly worn. The in-control knurls are all simulated using the mixed-effect model in Eq. (6). A set of slightly-worn knurls are used to calculate the fixed-effect coefficients in the mixed-effect model to simulate the out-of-control knurls. Four different values of correlation coefficient $\rho = 0.1, 0.5, 0.7$ and 0.95 are considered for the cross-correlation among the rows. The cross-correlation between Row 1 and Row 2 is positive while Row 3 and Row 4 have negative correlation. Either the correlation between Row 1 and Row 3 or the correlation between Row 2 and Row 4 is negligible. A set of 400 anvils is simulated using the above-mentioned procedure. Each anvil contains 4 rows with each row having 24 columns of knurls. Note that in Scenario II as shown in Fig. 11b, two out of four rows in each column are arbitrarily selected. Then, in each simulation run, a set of slightly-worn knurls are simulated. This way of the simulation enables us to generate a completely random wear pattern on the anvils.

As discussed in section "Data representation and basic high-order algebraic operations", the dataset can be represented in three ways. Based on Representation I, a 9600×924 matrix can represent the whole simulated dataset. Following Representation II, the dataset is represented using a 38400×231 matrix. The third way is to use a tensor representation with the dimension of $4 \times 231 \times 9600$. The PCA method is applied to above two matrix representations.

For PCA I method, 7 components are used for constructing the monitoring chart since they are sufficient to cover 90% of the data variability. For PCA II method, four PC features can explain about 90% of the data variability. When the HOSVD method is applied to the tensor of data directly, in order to represent 90% of the variability of the data, we need to use 4 components for the row's mode, 3 for the knurls profiles mode, and 6 for the columns of knurls mode. For the FFT method, since the first frequency feature is sufficient to explain about 99% of the signal's energy, only this feature is used for constructing the associated control chart. After extracting the monitoring features using each method, a T^2 control chart is used for monitoring the features extracted from the PCA and HOSVD methods. A Shewhart control chart is used for monitoring the magnitude of the dominant

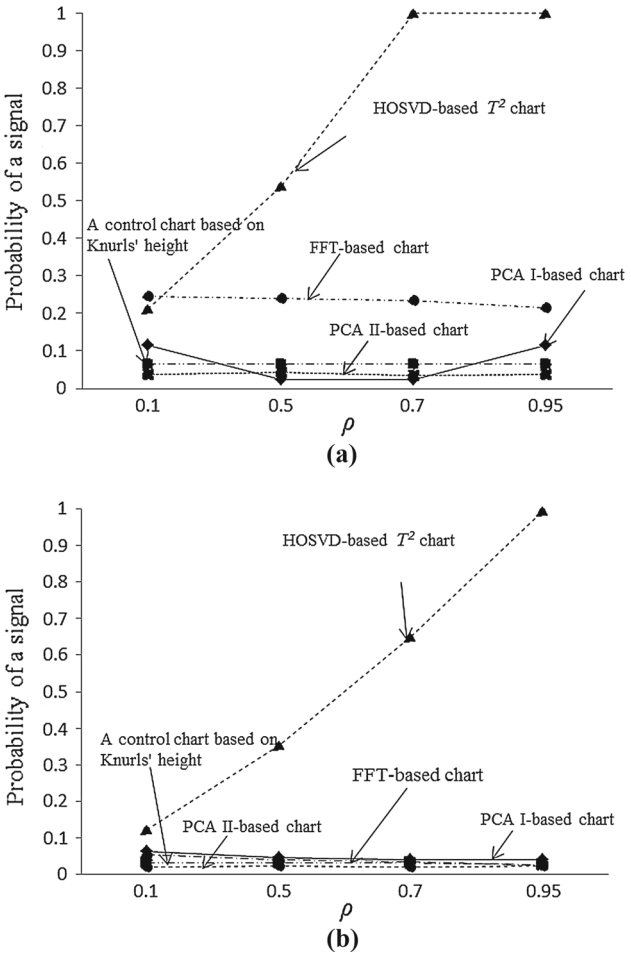

Fig. 12 Probability of signal with regard to different correlation coefficients. **a** Scenario I, **b** Scenario II

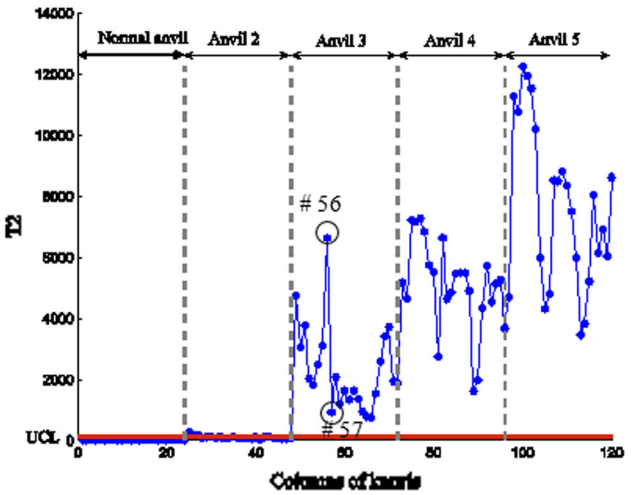

Fig. 13 T^2 control chart to monitor knurls' wear in all five anvils

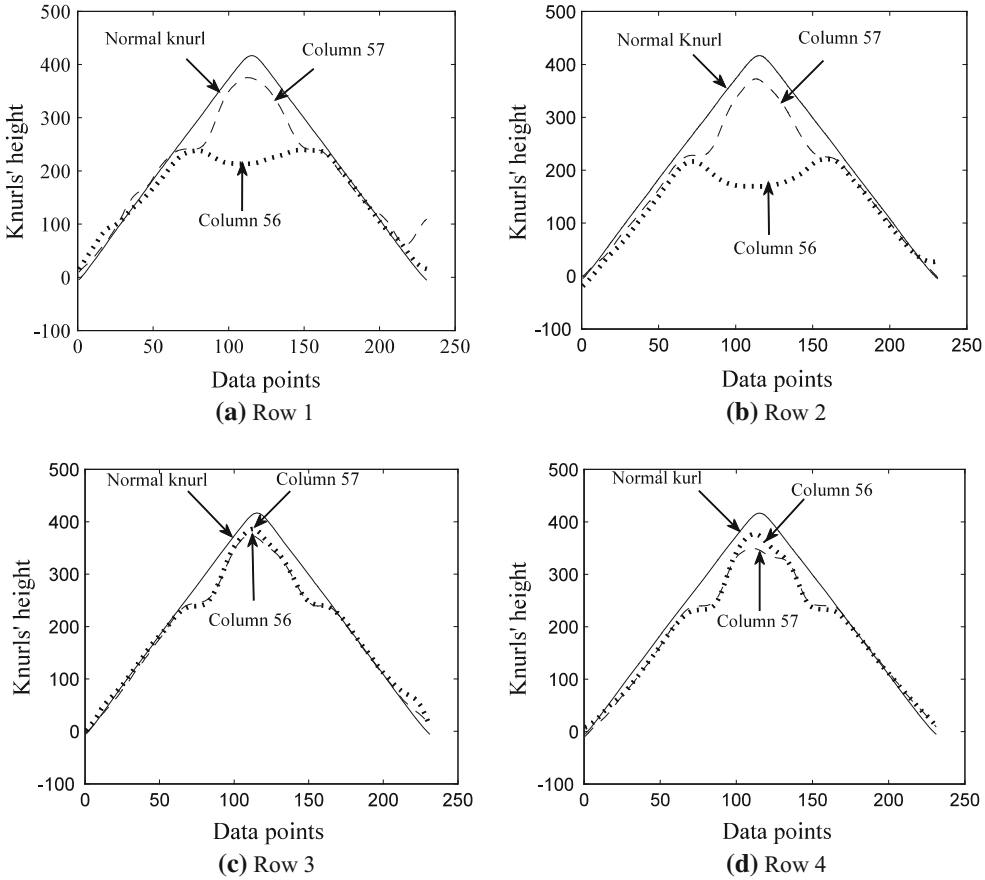

Fig. 14 Visual comparison of Column 56 versus Column 57

frequency in the FFT method. In addition, the knurls' heights are monitored using a Shewhart control chart.

To compare the monitoring performance, in each simulation run, 400 in-control anvils are generated, and the Type-I error rate of each method is computed as the proportion of out-of-control samples to all simulated samples. In this study, 10,000 simulation runs are conducted. The final estimation of the Type-I error rate is computed by averaging the Type-I error rates of the 10,000 runs' results. The upper control limit is then set in such a way resulting in an overall Type-I error rate equal to 0.5%. The probability of triggering an alarm for detecting worn knurls, i.e., detection power, is used as a criterion for comparing the detection performance of aforementioned three methods.

Figure 12a gives the detection power of all four methods under different correlation coefficients ρ in Scenario I. The HOSVD-based method outperforms all other methods except when the cross-correlation among the rows is $\rho = 0.1$. In this case, the FFT method performs slightly better than HOSVD-based T^2 chart. The detection power of the HOSVD-based T^2 chart increases with the increase of the cross-correlation since the tensor data representation can truly preserve the actual data cross-correlation structure among the rows as discussed in section "Comparison of HOSVD and PCA in capturing the correlation structure". The two PCA-based charts perform almost similarly in detecting the out-of-control situation and their performance is also not affected significantly by the amount of correlation among the rows. As expected, the control chart constructed based on the knurls' heights has shown poor performance in detect-

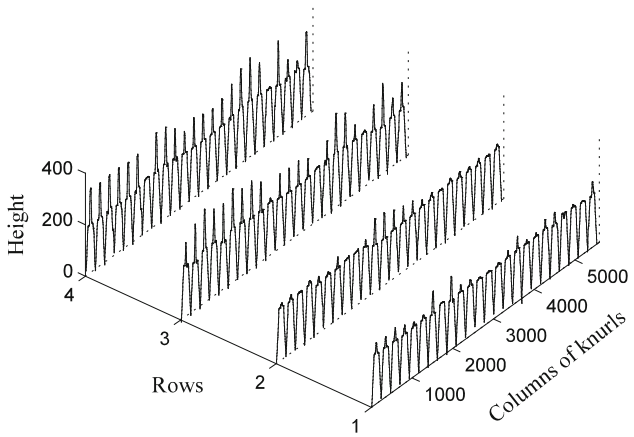

Fig. 15 Visualization of a misaligned anvil

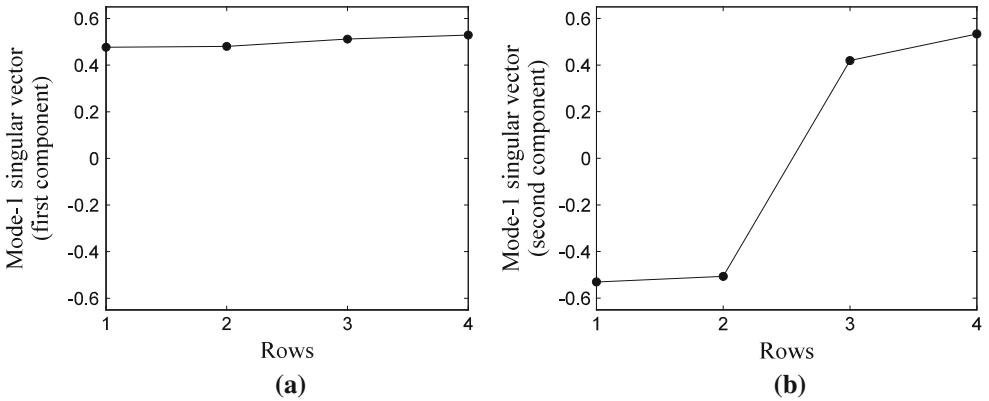

Fig. 16 Mode-1 singular vectors obtained using HOSVD method. **a** First component, **b** second component

ing slightly-worn knurls. In Scenario II, it is observable in Fig. 12b that the HOSVD-based T^2 chart significantly outperforms all other methods. Based on these findings, it can be concluded that the HOSVD-based method is a more reasonable option for monitoring the tool wear data when a strong cross-correlation exists.

Case study

This section presents the results of applying the HOSVD-based method for monitoring the wear of knurls using a real production dataset. The dataset consists of five anvils: Anvil 1 is new, and Anvils 2–5 are ordered with increasing wear levels. The HOSVD method is used to extract the features followed by a T^2 control chart constructed based on the extracted features. To explain 90 % of the data variability, the number of components in Mode 1, Mode 2 and Mode 3 are chosen as $k_1 = 4$, $k_2 = 3$ and $k_3 = 2$, respectively. Hence, the tensor scores $\mathcal{Z}$ given in Eq. (3) is a $4 \times 3 \times 24$ tensor, and the matrix of extracted monitoring features is represented by $\mathbf{Z} \in \mathbb{R}^{24\times 12}$ for constructing the T^2 control chart Fig. 13 shows the T^2 control chart for all five anvils. Each point on the control chart represents the T^2 statistic computed for a column of knurls on an anvil; as a result, a set of 24 points represents one anvil. The control limit and the parameters are estimated using the new Anvil 1 with all the 24 T^2 statistics being within the control limit. Since some of the knurls on Anvil 2 are slightly worn, there are some out-of-control points in the chart. In Anvil 3, most of knurls are severely and completely worn; as a result, all points are out of control, and this is the same for Anvil 4 and Anvil 5. The aforementioned observations show that the proposed control chart can effectively detect the worn knurls at different statuses.

To further verify the accuracy of the proposed method, Column 56's knurls are compared to those of Column 57 marked in Fig. 13. Since the 56th T^2 statistic is much higher than the 57th T^2 statistic, it is expected that the knurls in the 56th Column are more worn than those in the 57th Column. Figure 14 compares the knurls for Column 56, Column 57 and a normal knurl without any wear, in which the 4 rows of knurls are compared separately in Fig. 14a–d, respectively. Figure 14a, b clearly show that the knurls in Row 1 and Row 2 of Column 56 are completely worn. In contrast, Column 57's knurls are only slightly worn. The knurls of Row 3 and Row 4 shown in Fig. 14c, d do not show a significant difference in the wear levels between Column 56 and Column 57. Hence, it is concluded that the severe wear level in Row 1 and Row 2 is responsible for the high value of T^2 statistic in Column 56.

As discussed in section "Comparison of HOSVD and PCA in capturing the correlation structure", one of the strengths of the HOSVD method is its superior capability in the analysis of the variation pattern for the tensor data, or specifically in this case, explicit analysis of the cross-correlation among different rows of knurls in an anvil. This can help provide information regarding the knurls' wear pattern on the anvil and show whether the anvil is misaligned or not. As an example, Fig. 15 shows completely worn knurls in Row 1 and Row 2 while less wear can be seen in Row 3 and Row 4. The proposed HOSVD method is applied to this anvil data using the tensor representation. The resultant Mode-1 singular vectors are plotted in Fig. 16, in which the first singular vector gives an average of all four rows while the second singular vector clearly shows the negative cross-correlation of Rows 1 and 2 with Rows 3 and 4. This negative cross-correlation, as pointed out, reflects the unusual wear pattern on the anvil due to the misalignment in the anvil installation (Fig. 16).

Conclusion

A high-order decomposition based control chart is proposed for monitoring the tool wear in ultrasonic metal welding process. The proposed method has the advantage in detecting slight wear in the knurls at an early stage. Furthermore,

the HOSVD method can help find the unusual knurls' wear patterns among different rows of knurls in an anvil. Studying knurls' wear patterns on the anvil can help further identify and remove potential causes of excessive or unusual wear of the anvil.

References

Abellan-Nebot, J. V., & Subirón, F. R. (2010). A review of machining monitoring systems based on artificial intelligence process models. *The International Journal of Advanced Manufacturing Technology*, *47*(1–4), 237–257.

de Lathauwer, L., de Moor, B., & Vandewalle, J. (2000). On the best rank-1 and rank-(R1, R2,., Rn) approximation of high order tensors. *SIAM Journal of Matrix Analysis and Applications*, *21*(4), 1324–1342.

He, X., Cai, D., & Niyogi, P. (2005). *Tensor subspace analysis, advances in neural information processing systems, 18 (NIPS)*. Cambridge: MIT Press.

Hotelling, H. (1947). Multivariate quality control. *Techniques of Statistical Analysis*, *1947*, 114–184.

Jolliffe, I. (2005). *Principal component analysis*. New York: Wiley Online Library.

Kisić, E., Durović, Z., Kovačević, B., & Petrović, V. (2015). Application of T^2 control charts and hidden Markov models in condition-based maintenance at thermoelectric power plants. Hindawi corporation, Shock and Vibration, 2015, Article ID 960349.

Kolda, T., & Bader, B. (2009). Tensor decompositions and applications. *SIAM Rev*, *51*(3), 455–500.

Li, X., Dong, S., & Yuan, Z. (1999). Discrete wavelet transform for tool breakage monitoring. *International Journal of Machine Tools and Manufacture*, *39*(12), 1935–1944.

Mason, R. L., Tracy, N. D., & Young, J. C. (1995). Decomposition of T2 for multivariate control chart interpretation. *Journal of Quality Technology*, *27*(2), 109–119.

Paynabar, K., Jin, J. J., & Pacella, M. (2013). Monitoring and diagnosis of multichannel nonlinear profile variations using uncorrelated multilinear principal component analysis. *IIE Transactions*, *45*(11), 1235–1247.

Shao, C., Guo, W., Kim, T. H., Jin, J. J., Hu, S. J., Spicer, J. P., & Abell, J. A. (2014). Characterization and monitoring of tool wear in ultrasonic metal welding. In *9th international workshop on microfactories, Honolulu, Hawaii*, October 5–8, (pp. 161–169).

Shao, C., Kim, T. H., Jin, J. J., Hu, S. J., Spicer, J. P., & Abell, J. A. (2016). Tool wear monitoring for ultrasonic metal welding of lithium-ion batteries. *ASME Journal of Manufacturing Science and Engineering*, *138*(5), 051005.

Shao, C., Paynabar, K., Kim, T. H., Jin, J. J., Hu, S. J., Spicer, J. P., et al. (2013). Feature selection for manufacturing process monitoring using cross-validation. *Journal of Manufacturing Systems*, *32*(4), 550–555.

Shi, D., & Gindy, N. N. (2007). Tool wear predictive model based on least squares support vector machines. *Mechanical Systems and Signal Processing*, *21*(4), 1799–1814.

Yan, H., Paynabar, K., & Shi, J. (2015). Image-based process monitoring using low-rank tensor decomposition. *IEEE Transactions on Automation Science and Engineering*, *99*, 1–12.

Zhou, J. H., Pang, C. K., Zhong, Z. W., & Lewis, F. L. (2011). Tool wear monitoring using acoustic emissions by dominant-feature identification. *IEEE Transactions on Instrumentation and Measurement*, *60*(2), 547–559.