A texture based descriptors used for real time tool condition monitoring

A Antic1, M Milosevic1, R Cep2 and N Ungureanu3

1University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
2VŠB-Technical University of Ostrava, Faculty of Mechanical Engineering, Ostrava, Czech Republic
3Technical University of Cluj-Napoca, Faculty of Engineering, Department of Engineering and Technology Management, Baia Mare, Romania

E-mail: *antica@uns.ac.rs*

**Abstract.** Tool Condition Monitoring Systems (TCMs), in the task of tool wear recognition, are highly dependent on the choice of descriptors. It is very important to extract the appropriate marking that contains information on the tool wear condition from the sensor signal. We consider the portions of the Short-Term Discrete Fourier Transform (STDFT) spectrum obtained by speaking from a vibration signal sensor as a 2D textured image. This is achieved by setting the STDFT timeline as the first dimension and the frequency scales as the second dimension of the resulting textured image. We divide the resulting textured image into special particular patches of texture, focusing on a portion of the frequency range of interest. Applying the set filter bank, 2D textons are allocated for each predefined frequency band. From 2D textons, for each frequency band, we extract probability density function (PDF) data in the form of lower order statistical moments. The applied method gives robust descriptors of the Tool wear state.

1. **Introduction**

Preventing possible damages to the workpiece and machine tool due to the unexpected breakage of the tool in the machining process requires the introduction of constant monitoring of the condition of tool wear, which provide a quick response to the unexpected failure of the tool error. Many academic and practical studies have been conducted to monitor the state of the real-time tool monitoring (TCM) in the turning process, which is a very important contribution to increasing the safety and stability of the production process [1]. The state of the machining process and its control play an important role in predicting tool wear. Different signal processing techniques are applied to obtain crucial information. The development of information technology has made it possible to use multiple sensors to effectively monitor of tool wear condition [2]. Tool fracture prediction can be implemented by effectively monitoring changes occurring on the tool itself. Cho et al., in their work, divided the tool changes, and tool wear states, into the following categories: tool breakage, tool crunching, and tool wear [3]. The main task in using a modern machining system is to develop and deploy reliable and robust TCM systems.

In research studies and practical solutions, different methods are applied to collect tool failure information using the appropriate sensors and thus the corresponding signals used in TCMs. The basic division is in two groups: direct methods (consisting of laser, optical and ultrasonic sensors that allow
direct measurement, for example in [4] and indirect, based on sensors that bring to the state of processing by monitoring cutting forces, vibrations, temperature, instantaneous consumption etc. The application of the direct method is very expensive and difficult to apply in the workspace of the machine, indirect methods much more economical and more frequent in practice. Among indirect methods, vibration signal monitoring (i.e. accelerometers) is highly correlated with surface roughness in turning operations [5], as well as wear of cutting tools [6], therefore, these signals and their processing are highly applicable in TCM tasks.

Modern TCM systems based on artificial intelligence and their precision in determining the state of wear and tear depends primarily on the choice of descriptors, i.e. a feature extracted from different sensor signals. The fact is that unless the descriptors describe the signal appropriately, other techniques, such as feature extraction or feature selection, as well as recognition methodology, are also ineffective. Bahr in [7], and Tsai et. all in [8] were the first to apply descriptors derived from vibrational signals in TCM tasks. In fact, they used RMS and / or the mean value of the vibration sensor signal to detect an increase in the vibration magnitude, which corresponds to an increase in the vibration signal energy generated by the flank wear. Also, in [9], the mean and peak values of vibration sensor signals are used in the TCM task. In [6], Dimila analysed the correlation of vibration signal features to cutting-tool wear, in both the time and the frequency domains, during turning operations. Time domain features were deemed to be more sensitive to the cutting condition than tool wear, whereas certain peak values in the frequency domain correlated well with the measured wear values. In [10], the influence of power spectral density (PSD) for prediction of surface robustness is analysed. In [11], the average, variance, and single harmonic from the vibration sensor signal are used as descriptors in order to efficiently model the tool condition. In [12], authors used wavelet analysis and lower order moments, such as average value, standard deviation, power value, kurtosis value, harmonics frequency, skew value, etc., in order to express work-piece and spindle vibrations in the X, Y and Z directions. In [13], authors analysed five harmonic values of vibration signals in the frequency domain and the signal average in the time domain.

This study proposes tool wear monitoring strategy in which new descriptors based on image texture and feature extraction have been implemented to be applied in TCMs that use vibration sensor signals. In fact, the textons approach to constructing descriptors is completely new in the field of TCM to the authors knowledge. The proposed descriptors obtained from the vibrational sensor signals are based on the spectral domain. They rely on what the module of the Short-Term Discrete Furrier Transform (STDFT) spectrum obtained from a particular sensor signal will view as a 2D textured "image". We identified time and its scale as the first dimension, and frequency and frequency scale as the second dimension of the STDFT sensor signal module of interest. The 2D textured image obtained in this way is then divided into separate discrete narrow 2D texture patches, which cover a portion of the frequency band of interest. Then, by applying the appropriate filter bank for each predefined frequency band, we extract 2D textons (see [14]), i.e., small dimensional vectors in the filter response space. Our goal is to use the aforementioned filter response for the corresponding textons, to encode fine differences in the texture structure of the aforementioned texture patches that correspond to the degree of tool wear.

This approach is based on the assumption, which was later confirmed in the experiments, that the basic physical process of cutting and forming a chip is generated by different states of tool wear directly related to the structure of the aforementioned textons. However, the main problem we face when trying to use a texture-based approach is that most features, i.e. Descriptors used in texture recognition are inappropriate in this particular application. The reason for this is that the discriminability of texture patches belonging to different classes of tool wear states is contained in vertical lines distributed over the entire frequency range, and many of the techniques used in texture analysis could not account for this. Therefore, it is crucial in our approach to select the appropriate filter bank that will be able to efficiently extract the presence information of these lines, for each frequency band. In this paper, we apply a particular filter bank, proposed in [14], which is used in texture recognition problems, which has the aforementioned ability. We apply it to 2D patches of
texture, obtained from the STDFT vibration sensor signal, to efficiently extract information on said lines, in the form of a texton, i.e. characteristic in filter response space.

We additionally use the modal probability distribution for each frequency range, which describe each particular utterance in the filter response space, and use them as low-dimensional, robust descriptors. The marks are obtained by using the first four statistical moments for the final marks to be obtained, which is a special saying that will be used in the training or recognition process. In this way, information on the texton probability distributions corresponding to each frequency band is extracted and placed in small features vector to be used effectively in the training and recognition in TCM tasks.

2. Apparatus used in experiments

The experiment setup and processing were performed on a CNC GU 600 lathe manufactured by INDEX installed in the laboratory of the Faculty of Technical Sciences in Novi Sad. Investigation of the tool wear process involved monitoring the mechanism of dominant wear through the following parameters: wear belt, crater wear, and tool life. During the turning process, a vibrating signal was detected on the tool shank. Generated signals from the machining process were recorded for each tool pass. The placement of the sensors on the tool as well as the workpiece dimensions used in this experiment are shown in Figure 1. On the experiment we used two cutting speeds, 180 to 250m/min, correlated with 0.15 and 0.3mm/rev feed rates. In the experiment was used 20x20mm cross section of the tool shank. The machining process was performed with P25 tool inserts marked TNMM 110408. Used sensors set is Accelerometers Kistler 8002 was fixed onto the tool holder, where the measured acceleration of vibrations signal. All signals was sampled at 625 kHz, using A/D converter NI 625 USB, National Instruments. Workpiece material was 42CrMo4, 310 HB hardness and 950MPa.

The features extraction and selection of the descriptors used in the training as well as in the recognition phase of the classifier is performed from a signal of different lengths obtained from the vibration sensor. The focus in studies is on the problem of recognizing 2D texture from the perspective of tool wear recognition and classification. The most challenging problems present in the problem of recognizing the actual texture with 3D variations come from the variability of those textures, i.e. the fact that classical texture is primarily a function of the following variables: texture surface, its albedo (i.e., the reflection of the corresponding surface factor), lighting, cameras, and its viewing position.
We have noticed that the texture corresponding to the STDFT spectrogram of the vibration sensor signal that we consider to be seen as a 2D texture does not possess this kind of variability [15]. By identifying the presence of vertical lines in the performed resulting STDFT (viewed as texture that is of interest), our goal is to effectively isolate them, and/or to obtain reliable information about their presence. Therefore, imposes the need to extract texture features used by certain specific filter banks able to identify them.

To extract the relevant characteristics to be used in the actual task of identifying tool wear states in a TCM system, we first extract the relevant descriptors in the form of textons characteristics obtained in the appropriate filter output space, and then model their probability distribution for each band frequency using the first four moments of those outputs, as final, compressed very robust descriptors.

3. Texture based features – forming the texton

The task is to identify the discrete time frame $k$ of the STDFT spectrogram of the utterance of interest, as the discrete $X - axis$ of the textured image and the discrete frequency $\omega$, as the discrete $Y - axis$ of the corresponding textured images. The hypothesis we confirm in our experiments is that the shape of the STDFT spectrogram shown in 2D space discriminatively for different tool wear states is highly correlated with the characteristics (number, position, structure of the vertical lines) that appear in the corresponding textured images (see Figure 2). In fact, the condition of tool wear is tightly related to the characteristics of those lines. The lines actually represent sudden jumps in the time frequency range of the signal received from certain sensors, at different time and frequency scales, induced by the cutting process and the tool wear basically. Obtaining information with respect to these lines is directly related to the selection of the filter bank to be used for that particular task.

The novel texture based descriptors in the task of TCM and named them the Texture Based Tool Condition Descriptors (TBTD). The TBTCD descriptors are obtained by considering the STDFT spectrogram of the windowed time frames delivered from the previously mentioned utterance, as the 2D textured image.

The feature extraction procedure is as follows: A standard technique is applied for most texture recognition approaches it is divide the STDFT textured image into texture patches spread across in different filter bands. Then, using the appropriate filter bank, we extract texton-based descriptors as the feature vectors in the low dimensional filter response space. As already mentioned, it is crucial to select the appropriate filter bank that is able to efficiently extract the presence of the mentioned lines for each frequency band. The best results are achieved in [16] for the mentioned tasks are those that use the reactions of spatially invariant filter banks, i.e. filter banks based on a set of maximum response filters, such as (BFS, MR8, MR4 and MRS4), which have been reported and discussed in the classification assignment material [14, 17], as well as the Leung and Malik filter bank. A rotationally invariant, multi-scale MR8 filter has been shown to produce better results than any of the previously mentioned filters in texture recognition applications. In this application for the TCM task, we decided to use the MR8 filter, shown in Figure 3, without excluding the last two isotropic components, as they do not extract the corresponding as texton features.

Let $F(x, y)$ be the textured image that corresponds, i.e., is identified with the STDFT spectrogram of the vibration sensor signal of the particular utterance $s$. The STDFT spectrogram is defined as the square module of the spectral density, i.e., $|S(k, \omega)|^2$, where

$$S(k, \omega) = \sum_{n=-\infty}^{\infty} s(n) w(k-n) e^{-i\omega n}$$

(1)

is the STDFT of signal $s$. In (1), by $k$ we denote the discrete time frame, by $\omega$ we denote discrete time frequency, i.e., discrete frequency beam, while $w$ is the windowing sequence that we use (we use Hamming experiments), with the fixed length $K$. We identified $x$ axes of the textured image with the Discrete quadrature mirror frame $k$ of the mentioned STDFT, for $k = 0, \cdots, k_{max}$, so that it holds $x_{max} = k_{max}$. Actually, to every discrete time frame $k$ in STDFT, there is a corresponding discrete index $x$ of the texture patch in the textured image $F$. Also $\omega$, the discrete frequencies $\omega = 0, \cdots, \omega_{max}$ are identified.
with discrete Y axes of the textured image F so that it holds $y_{max} = \omega_{max}$. We now have the following interpretation of STDFT in the form of textured image:

$$F(x, y) = |S(k, \omega)|^2$$

(2)

$x = 0, \ldots, x_{max}$, $y = 0, \ldots, y_{max}$

$k = 0, \ldots, k_{max}$, $\omega = 0, \ldots, \omega_{max}$

We further do all our analysis on textured image $F(x, y)$ obtained as it is explained previously, where for the easier modeling, without loss of generalization, we consider the continuous case $x \in [0, x_{max}]$ and $y \in [0, y_{max}]$.

We briefly describe the filter that we use: Let

$$G(\sigma_x, \sigma_y, \theta, x, y) = \frac{1}{2\pi \sigma_x \sigma_y} e^{(A(\theta)^T [x, y])}$$

(3)

be the directed anisotropic Gaussian kernel. The fixed terms $\sigma_1 > 0$ and $\sigma_2 > 0$ denote scales in $x \in [0, x_{max}]$, $y \in [0, y_{max}]$ direction, respectively, while $\theta$ and $A(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$, for $\theta \in [0, 2\pi]$, denote the orientation of the kernel (3) and its 2D rotational matrix, respectively. The MR8 filter bank that we use in our application consist of 38 Gaussian filters of the form (3), at 6 different orientations $\theta_j \in \{\frac{j\pi}{6} \mid j = 0, \ldots, 5\}$, and 3 different scales $\{\sigma_x, \sigma_y\} \in \{(1,3), (2,6), (4,12)\}$, and also 6 Laplacians of Gaussians, defined as $\Delta G(\sigma_x, \sigma_y, \theta, x, y)$, where $\Delta = \partial^2_x + \partial^2_y$ is the Laplace operator of differentiation, evaluated at the same orientations and scales (for more details see [13, 18]).

The filter outputs are collapsed, by taking only the maximum values, across all orientations, obtaining 6 different filter responses. Additional two responses are isotropic, thus unable to extract characteristic lines in the textured image F, so we do not use them in application.

Further we proceed with the procedure of texton extraction. We divide the interval $[0, y_{max}]$ where texture image F has its values into $M$ sub-intervals $[y_{q-1}, y_q) \mid q = 1, \ldots, M]$, 0, $y_M = y_{max}$ where for every index $x \in [0, \ldots, x_{max}$, the sub-patch of $F(x, y)$, is defined as matrix of pixels bounded by $[y_{q-1}, y_q]$ on Y access and by $[x, x+T]$ on X access. Here, T is predefined width of the patch. We denote such described textured patch as $P_{x,q}$, $q = 1, \ldots, M$. The texton in the form of filter response are then formed for every interval $[y_{q-1}, y_q]$ and discrete coordinate x. We note, that the interval $[y_{q-1}, y_q]$ of the textured image F, corresponds to the frequency band $[\omega_{q-1}, \omega_q]$ of the STDFT spectrogram of particular utterance.

Actually, for each X and each different band q, texton $v_{x,q}$ is composed of $Q = 6$ components of the MR8 filter bank.

Thus, we have:

$$v_{x,q}^l = \left[ v_{x,q}^{l(1)} \ | \ v_{x,q}^{l(2)} \right] \mid l \in L$$

(4)

where

$$v_{x,q}^{l(1)} = \max_{\theta} G(\sigma_x, \sigma_y, \theta, x, y) \ast P_{x,q}$$

$$v_{x,q}^{l(2)} = \max_{\theta} \Delta G(\sigma_x, \sigma_y, \theta, x, y) \ast P_{x,q}$$

$$\theta \in \{\frac{j\pi}{6} \mid j = 0, \ldots, 5\}$$

and

$L = \{(\sigma_x, \sigma_y)(1,3), (2,6), (4,12)\}$.

In (4), for simplicity we enumerate: $l = 1, 2, 3$. 


3.1. Using moments for describing the probability distribution in the filter response space

We model the probability distribution of textons described in the previous section, for every band \( q \) separately. As a final result, we obtain compressed information about the mentioned distributions in the form of robust features.

We proceed as follows: For the fixed utterance \( u \), for fixed band \( q = 1, \ldots, M \), we have the set of observation feature, i.e., texton descriptor vectors

\[
v_{x,q} = [v_{x,q}^1, v_{x,q}^2, v_{x,q}^3]
\]

where \( v_{x,q}^j \) given by (4). It is obtained for all patches \( P_{x,q} \). It holds that, where \( x_{\text{max}}^u = k_{\text{max}}^u \) (see (2)) is the number of time samples in the utterance \( u \) which, of course, can vary, dependent of utterance. We consider \( v_{x,q} \) to be the observations, i.e., realizations of the random variable \( V_q^u \) with the probability distribution function (pdf) \( p_{V_q^u} \). As, for fixed band \( q \), there is a unique correspondence between the probability distribution \( p_{V_q^u} \) of \( V_q^u \) and its characteristic function given by

\[
k_{V_q^u}(t) = E[e^{itV_q^u}] = \sum_{j=0}^{\infty} \frac{(itj)^j}{j!} E\left( (V_q^u)^j \right)
\]

where one could observe that (5) is given as an expansion which depend only on moments \( m_{q,j}^u = E\left( (V_q^u)^j \right) \), where the averaging is conducted regarding \( x \). Thus, by taking only first few moments into account we manage to (roughly) approximate pdf \( p_{V_q^u} \), and yet obtain robust features i.e., descriptors which can be used in the recognition part of TCM task

The final feature vector assigned to each utterance \( u \) is obtained either as \( P \) moments per 3 scale, or \( P \) moments per scale case. For the first case, for each band \( q = 1, \ldots, M \), one vector containing \( P \) moment components \( m_{q,j}^u = 1, \ldots, P \) obtained by averaging vectors \( v_{x,q} \) given by (4), corresponding to patch \( P_{x,q} \) where the averaging is conducted regarding \( x \). Then, all the mentioned component \( m_{q,j}^u \)s concatenated into the single vector containing \( MP \) components:

\[
v^u = [m_1^{u,1} \ldots m_1^{u,MP} \ldots m_M^{u,1} \ldots m_M^{u,MP}]
\]

uniquely representing the utterance \( u \). For the second case, a vector containing \( 3MP \) moment components \( m_{q,j}^{u,l} = 1, \ldots, P \); a \( P \) for each scale \( l = 1,2,3 \), is obtained also by averaging in \( x \). Actually, for the second case, we obtain:

\[
v^u = [m_1^{u,1,l} \ldots m_1^{u,MP,l} \ldots m_M^{u,1,l} \ldots m_M^{u,MP,l}],
\]

\[
v^u = [v_1^{u,l}v_2^{u,l}v_3^{u,l}].
\]

4. Experimental setup

The classification of the wear state is made by an indirect method by extracting features from the vibration signal generated on the tool shank during the cutting process. Experimental tests were conducted on more than 150 recorded vibration signals. The signals are corresponding with 3 tool wear states (50 signals for each state), based on the wear band (new plate, i.e. no wear, little wear, i.e. up to 0.25 mm, and high wear, i.e. above 0.5 mm). The tool wear process is a continuous process because the wear condition of the tool is constantly changing and the boundaries between classes cannot be uniquely defined. The analysis of the results was carried out using fivefold cross-validation, i.e. a sample of 120 vibration signals used for training the classifier and another 30 vibration signals were used for the test (while all three categories were presented equally in the test and in the control set). An average value of 5 results were its analysed to obtain a final estimate.

Forming the training set started by computing the vibration signal spectrum, using a Short-Term Discrete Furrier Transform for the frequency range of \( M = 256 \) samples of equal length of Hamming windows, with \( M/4 \) overlap between segments (see Figure 2). The signals are 125000 samples in length, with a sampling frequency range over the 10 kHz. As a result of the applied transformations of signal, is obtained by matrix containing 129 frequencies (\( M/2 + 1 \) for real signals and an equal
number of frequency points, although a limited frequency range for subsequent transformations) and 650 time bins (number of segments) was used. Each segment represents an estimate of the short-term time-localized frequency content of the vibration input signals. Each spectrograph was normalized to exclude the DC component. The convolution is performed between the spectrum and the corresponding filter from the MR8 filter bank described in 3.1 and it is shown in Figure 3. We have selected $P = 3$ moments: variance, skewness and kurtosis, while the first moment, i.e. Mean, excluded from consideration because we only considered normalized spectrograms, i.e. correspondingly normalized textured images, representing a total of 9 marks per signal.

5. Experimental results

In Table 1, the average classification accuracy for 5 folds (F1 to F5) is presented for fuzzy classification conducted per each scale separately (3 moments per scale). The output values are averaged for all 3 scales in order to obtain the final result. These values are classified in the following...
way: less than 1.5 - class 1, 1.5 to 2.5 - class 2, more than 2.5 - class 3. The final results for 5 folds, including the target result, are presented in Fig. 4. The results are presented for 250 clusters (this was around the optimal value).

### Table 1 The average fuzzy classification accuracy (separate scales)

| Tool Wear       | new plate (class 1) | 0.25 mm (class 1) | 0.5 mm or more (class 3) |
|-----------------|---------------------|-------------------|--------------------------|
| scale 1         | 96 %                | 100 %             | 98 %                     |
| scale 2         | 98 %                | 100 %             | 100 %                    |
| scale 3         | 96 %                | 98 %              | 100 %                    |
| final classification | 98 %            | 100 %             | 100 %                    |

![Figure 4](image)

**Figure 4.** The final results for fuzzy classification (separate scales) for 5 fold

The classification accuracy obtained by using 9 coefficients (3 moments per 3 scales) for 5 folds is given in Table 2. The optimal number of clusters was smaller in this case, having in mind the fact that we had the same amount of data (number of signals) in higher dimensional space. Better results were obtained in the case of sum of the output values for separate scales pondered by 1/3, than in the case of classification for all 3 scales together.

### Table 2 The average fuzzy classification accuracy (3 scales together)

| Tool Wear       | new plate (class 1) | 0.25 mm (class 1) | 0.5 mm or more (class 3) |
|-----------------|---------------------|-------------------|--------------------------|
| 3 scales        | 98 %                | 100 %             | 90 %                     |
Acknowledgments
This paper presents a segment of the research on the project "Contemporary approaches in the development of special solutions bearing in mechanical engineering and medical prosthetics" project number TR 35025, financed by the Ministry of Education and Science of the Republic of Serbia.

References
[1] Siddhpura A and Paurobally R 2013 International Journal of Advanced Manufacturing Technology, vol. 65, p.371–393,
[2] Abellan-Nebot J V and Subirón F R 2010 International Journal of Advanced Manufacturing Technology, vol. 4,7 p.237–257,
[3] Cho S, Binsaeid S and Asfour S 2010 International Journal of Advanced Manufacturing Technology vol. 46, p.681-694.
[4] Wong Y S, Nee A Y C, Li X Q and Reisdorf C 1997 Journal of Materiel Processing Technology, vol. 63, p.205–210
[5] Jang D Y, Choi Y G, Kim H G and Hsiao A 1996 International Journal Machine Tools and Manufacturing, vol. 36, p.453–464.
[6] Dimla D E 2002 International Journal of Advanced Manufacturing Technology, vol. 19, p.705–713.
[7] Bahr B, Motavalli S and Arfi T 1997 International Journal of Computer Integrated Manufacturing, vol. 10,.314–323
[8] Tsai Y H, Chen J C and Lou S J 1999 International Journal Machine Tools and Manufacturing, vol. 39, p.583–605
[9] Haber R E, Jimenez J E, Peres CR and Alique J R 2004 Sensors and Actuators A: Physical, vol. 116, ISSN: 0924-4247 p.539–545.
[10] Abouelatta O B and Madl J 2001 Journal of Materials Processing Technology, vol. 118, p.269–277
[11] Chen J C and Chen W L 1999 Journal of Intelligent Manufacturing, vol. 10, p.187–197
[12] Al-Habaibeh A and Gindy N 2000 Journal of Materials Processing Technology, vol. 107, p.243–251
[13] Kuo R J and Cohen P H 1999 Neural Networks, vol. 12, p.355–370
[14] Varma M and Zisserman A 2009 IEEE Transactions on Pattern Analysis and Machine Intelligence, 31, 11:2032 - 2047.
[15] Varma M and Zisserman A 2005 International Journal of Computer Vision, vol. 62, (1-2), p.61-81
[16] Lazebnik S, Schmid C and Ponce J 2005 IEEE Trans. Pattern Analysis and Machine Intelligence, 27, 8:1265-1278.
[17] Varma M 2004 Statistical Approaches to Texture Classification, PhD thesis, Univ. of Oxford.
[18] Varma M and Zisserman A 2004 Image and Vision Computing, vol. 22, (14), ISSN: 0262-8856 p.1175-1183