TOOL CONDITION MONITORING
– AN INTELLIGENT INTEGRATED SENSOR APPROACH

WANG WENHUI
(B. Eng., M. Eng., Beijing Institute of Technology)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2005

Acknowledgements

First and foremost, I want to express my most sincere gratitude to my supervisors, Associate Professor G. S. Hong and Associate Professor Y. S. Wong. They provided me valuable supervision, constructive guidance, incisive insight and enthusiastic encouragement throughout my project.

I also would like to thank National University of Singapore for offering me research scholarship and research facilities. Without these supports, my graduate study will be impossible.

Special thanks go to Mr. K. S. Neo, Mr. C. S. Lee, Mr. S. C. Lim, Mr. C. L. Wong, Mr. Simon Tan and all the technicians at Workshop 2 for their technical assistance, and to Mdm. W. H. Liaw and Mdm. T. L. Wang in Control and Mechatronics Lab 2.

Many thanks are given to Experimental Mechanics Lab for allowing the use of the experimental equipment on phase-shifting. Generous help from Mr. Chen Lujie on the experiment is greatly appreciated.

My gratitude is also extended to the colleagues and friends in our lab and NUS, Mr. Du Tiehua, Mr. Wang Zhigang, Mr. Ong Wee Liat, Mr. Dong Jianfei, Ms. Sun Jie, Mr. Zhu Kunpeng and many others, for their enlightening discussion and suggestions.

Finally, I owe my deepest thanks to my parents and brothers for their unconditional and selfless encouragement, love and support.

Table of Contents

Summary

Sensor integration has shown much potential to enable a tool condition monitoring (TCM) system to be more accurate, robust and effective as the sensors can complement and reinforce each other. The main objective of this thesis is to incorporate one direct sensor (vision) and one indirect sensor (force) to create an intelligent integrated TCM system for on-line monitoring of flank wear and breakage in milling. To achieve this objective, two subsystems including a vision-based subsystem and a force-based subsystem have been developed to work in-cycle and in-process respectively. Experiments on both the subsystems and the integrated system were conducted to verify the integration scheme. To measure crater wear, a full-field optical method based on phase-shifting was also proposed and demonstrated.

For the vision-based subsystem, images were first captured with the spindle stands stationary. These were then processed with the individual image processing method, giving sub-pixel accuracy. A rough-to-fine strategy was employed to locate the point on the boundary of the wear land in two steps. The binary edge image was firstly used to locate the boundary point roughly. The gray-level image was then used to locate the boundary point more precisely using a moment-invariance based edge detection method in the vicinity of the rough point. Based on the individual image processing method, the successive image analysis method was developed to capture and process moving images captured while the spindle was rotating. A trigger-capture mechanism was introduced in response to the spindle rotation and successive images were processed on the basis of their correlation.

For the force-based subsystem, two force features in time domain based on average force and standard deviation were extracted from the cutting force signal and included to train a Self-organizing map (SOM) network. The SOM network was used locally in

the sense that the feature vectors used to train and apply the network were derived from two neighboring machining passes. The wear measured in-cycle by vision and the force features extracted in-process in the previous pass were used to train the SOM network. After the training, the SOM network was applied immediately to the next machining pass to estimate flank wear.

Apart from flank wear estimation, breakage and crater wear were also studied. To detect breakage, two other force features, which are residual error and peak rate, were used. This preliminary detection result was verified by vision. To measure crater wear, the phase-shifting method was employed. Four images of the rake face on which different fringes were projected were analyzed to give the phase map, which was converted to a 3-D map of crater wear after calibration.

Experimental results showed that the breakage was detected and verified successfully, and the flank wear was estimated well, especially at the linear wear stage. The crater wear was accurately and robustly measured by phase-shifting method. This study has demonstrated that it is possible to use this sensor integration scheme to monitor breakage and flank wear on-line in milling process quite accurately, robustly and effectively over a wide range of machining conditions.

List of Tables

List of Figures

List of Symbols

N_T^E	Threshold to ascertain $P_E(i)$
$\overline{G}_w(i)$	Average gray level in the wear land
$\overline{G}_{\bar{w}}(i)$	Average gray level in the unworn area
$\Delta\overline{G}(i)$	Average gray level difference between $\overline{G}_w(i)$ and $\overline{G}_{\bar{w}}(i)$
$G_{w-\bar{w}}$	Minimum gray-level difference between $\overline{G}_w(i)$ and $\overline{G}_{\bar{w}}(i)$
$\overline{m}_i$	Moment in one image
$\overline{w}$	Average wear width of previous pass
ε	Residual error of average force
α	Angle of a line's normal
θ	An angle
σ	Standard deviation
Φ	Parameters in *AR*1 model
β	Parameters in *AR*1 model
λ	Forgetting factor in *AR*1 model
$\mu(k)$	Average gray level of gray class *k*
$\omega(k)$	Sum of gray level of gray class *k*
$\phi(x,y)$	Phase function
α_0	Angle of the normal to the reference line
ϕ_1, ϕ_2	Phase maps in location 1 and 2 in calibration
τ_1, τ_2	Time constants in *SOM*
σ_B	Standard of gray class classification
$\triangle D$	Interval between two neighboring scan lines
φ_{en}	Entry angle of cut
φ_{ex}	Exit angle of cut
Δh	Depth difference in calibration for phase-shifting
α_i	Discretized α
φ_i	Cutting edge rotation angle of the *i*th tooth
$\Re^n$	*n*-dimensional real-value space
μ_T	Average gray level of one image

1-*D*	One dimension
2-*D*	Two dimension
3-*D*	Three dimension
a	Slope for a line
A(*R*)	Average gray level of image *R* in window *W*
A(*U*)	Average gray level of image *U* in window *W*
A(*x*,*y*)	Original gray-level image
AE	Acoustic emission
AI	Artificial intelligence
AR	Autoregressive
ART	Adaptive resonance theory
A_w	Area of the wear land
b	Intercept for a line
B	A binary image in general
B(*x*,*y*)	Binary image by binarizing *S*(*x*,*y*)
$B_E^+(x,y)$	Binary edge image of $E^+(x,y)$
BK	Breakage
BP	Back propagation algorithm
c	Winner in *SOM*
C	Local contrast
C'	Wanted contrast
CC	Cross-correlation coefficient
CCD	Charge coupled device
CC_W	Cross-correlation coefficient in window *W*
CDF	Cumulative distribution function
CID	Charge injection device
CNC	Computer numerical control
CNNN	Condensed nearest-neighbor network
CP	Chipping
CW	Crater wear
d	Distance from origin to a line
D	Diameter of cutter
d_0	Distance of the origin to the reference line

DAQ	Data acquisition
d_i	Discretized d
E	Local edge gray value
$E(x,y)$	Edge image by Sobel operator
$E^+(x,y)$	Enhanced edge image
F	Cutting force
$f(i,j)$	Instantaneous force
$f(x,y)$	An input image in general
f_a	First order differencing of average force
F_a	Average force
F_e	Force features in wear estimation
FFT	Fast Fourier transform
F_m	Peak value of cutting force in one rotation
f_{pt}	Feed per tooth per revolution
f_r	Rotation frequency
F_r	Instantaneous radial force
f_t	Tooth frequency
F_t	Instantaneous tangential force
FW	Flank wear
F_x	Cutting force in X direction
F_y	Cutting force in Y direction
$g(x,y)$	An edge image in general
$G(x,y)$	Background intensity
G_{max}	The maximum level in a gray-level image
G_{min}	The minimum level in a gray-level image
$G_{PDF=max}$	The gray level with maximum density in histogram
G_x	Sobel operator for vertical edges
G_y	Sobel operator for horizontal edges
H	Histogram of one gray-level image
$H(x,y)/G(x,y)$	Fringe contrast
h_1	Gray level 1
h_2	Gray level 2
$h_{ci}(t)$	Neighboring function in SOM

HT	Hough transform
h_{tr}	True undeformed chip thickness
$K(t)$	Estimation gain in $AR1$ model
K_{pr}	Peak rate of cutting force
k_r	Ratio of the tangential force and radial force
k_s	Specific cutting force coefficient
$L(i)$	Scan line i
$l(t)$	Learning rate function in SOM
LCD	Liquid crystal display
LDF	Linear discriminant function
L_{max}	A border line in parallel scanning
LP	Low pass
$LWDM$	Long working distance microscope
M	Margin set for the critical area
$M_{\Delta\phi}$	Mean of the phase difference
$M(x,y)$	Binary edge image after morphology
$MARMS$	Moving average of the root mean square
m_i	Weight vector in SOM
MLP	Multi-layer perceptron
m_t	Number of teeth or inserts
n	Data length in general
N	Number of sampling points in one rotation
N_1	Number of white pixels in the window with length w_1
NB	No breakage
N_L	Number of scan lines with wear
N_{max}	The maximum number of white pixels on some parallel scan line
NN	Neural network
N_{Vi}	Number of input samples in the Voronoi set of unit i
$O(x,y)$	Median filtered gray-level image
P	Parameters in $AR1$ model
$P_A(i)$	Start point of scan line i
$P_B(i)$	End point of scan line i
$P_E(i)$	Rough point on the boundary of the wear land on scan line i

PMP	Phase measuring profilometry
$P_{RB}(i)$	Fine point on the boundary of the wear land on scan line *i*
r	Contrast enhancement rate
R	Reference image
$R(x,y)$	Object reflectivity function characterizing the surface nature
RBF	Radial basis function
R_c	Radius of the cutter
RCE	Restricted coulomb energy
r_i ,r_c	Location vectors of unit *i* or *c* in *SOM*
ROI	Region of interest
rpm	Rotation per minute
r_w	Ratio upon which white pixels can be removed in parallel scanning
s	Spindle speed
S	Structuring element in morphology
$S(x,y)$	Histogram-stretched gray-level image
s_i	Sum of the vectors in the Voronoi set of unit *i*
SMC	Spindle motor current
SOM	Self-organizing map
SVM	Support vector machine
S_x	Nucleus of the structuring element in morphology
T	Temperature
TBD	Tool breakage detection
TC	Time constant
TCM	Tool condition monitoring
T_E	Threshold to binarize $E^+(x,y)$
T_I	Integration time of the *CCD* camera
T_{pr}	Threshold for peak rate in breakage detection
T_{re}	Threshold for residual error in breakage detection
T_S	Threshold to binarize $S(x,y)$
TS	Transducer sensitivity
TWD	Tool wear detection
TWE	Tool wear estimation
u	Number of units in *SOM*

U	Worn insert image
V	Possible maximum width of the wear land
V_0	Flank wear measured by microscope
V_1	Flank wear measured by *CCD* camera
VB	Width of flank wear
VB_{ave}	Average width of flank wear
VB_{max}	Maximum width of flank wear
Vt	Vibration
w	Width of the critical area
W	Window to calculate CC_W
$W(i)$	Flank wear value on scan line i
w_1	Window length in locating $P_E(i)$
w_2	Lower bound of the number of pixels to locate $P_{RB}(i)$
w_3	Upper bound of the number of pixels to locate $P_{RB}(i)$
w_c	Chip width
Z_0	Crater wear depth measured by microscope
Z_1	Crater wear depth measured by phase-shifting

Chapter 1
Introduction

1.1 Problem statement

In manufacturing, it is desirable to reduce labor cost, minimize operator's errors, and enhance the productivity and quality of products (Huang et al., 1999). To achieve this goal, on-line tool condition monitoring (TCM) is one of the most important techniques (Lin and Lin, 1996). It helps to operate the machine tool at its maximum efficiency by detecting and measuring the tool conditions such as flank wear, crater wear, chipping, breakage and so on. A successful TCM system can increase productivity, and hence competitiveness, by maximizing tool life, minimizing machine down time, reducing scrap and preventing workpiece damage (Donnell et al., 2001).

A significant amount of TCM research has been dedicated to monitor tool conditions on-line. However, most of the TCM techniques developed are for single-point cutting processes, such as turning. The results may not be directly suitable for the multi-tooth milling process (Lin and Lin, 1996). Although milling is a very important machining process in manufacturing, much less effort has been made to monitor it (Byrne et al., 1995). The systems developed for milling still need to be more reliable, robust and responsive for truly automated manufacturing (Prickett and Johns, 1999). Obviously, there is still much to understand and do before on-line TCM systems in milling can be used in industry.

For decades, researchers have proposed numerous methods based on sensors to monitor tool conditions in milling on-line. In the early years, only a single sensor was used but it was found to be inadequate. Recently, one trend is to combine two or more

sensors in one system to achieve better performance. Therefore, in this research, vision-based sensor and force-based sensor are integrated to implement an on-line TCM system that can monitor progressive flank wear and detect breakage in milling.

1.2 Motivation

A TCM system is basically an information flow and processing system (Niu et al., 1998) that integrates the following three functional blocks: the information source selection and acquisition (sensor and data collection); information processing and refinement (feature extraction); and decision making based on the refined information (condition identification). It is essentially a sensor-based system. Consequently, according to the sensor type, TCM techniques can be generally classified into direct and indirect methodologies (Kurada and Bradley, 1997a). The direct methods rely on sensors that measure tool condition *in situ*, such as vision, mechanical probes and proximity sensors. Indirect methods, by contrast, measure signals that indirectly indicate the tool conditions with sensors such as force, acoustic emission (AE), vibration, current/torque, and power sensors.

Early TCM research focuses on one single sensor in the TCM systems. However, use of a sole sensor, either by direct or indirect methods, to monitor the tool condition is not satisfactory. Although accurate, the direct method can only monitor the conditions between cuts or tool changeovers, and thus continual monitoring is not achieved. By contrast, the indirect method, which deploys force or AE sensors, can monitor conditions continually and on-line. But in most cases, it is not sufficient for the sole sensor to provide condition-sensitive features. Accordingly, the performance of TCM systems using a single sensor is not satisfactory and as a result, few successful applications in industrial environment have been reported (Byrne et al., 1995).

To replace the manual monitoring of the tool condition is one of the goals of TCM research. Less downtime, higher productivity, higher surface finish quality and more powerful, but cheaper unmanned tool change decisions are necessities for industrial application (Donnell et al., 2001). More research, with the aim of developing a TCM system with higher reliability, robustness, and response is needed (Byrne et al., 1995; Kurada and Bradley, 1997a; Prickett and Johns, 1999). With this goal, sensor fusion, integration of two or more sensors in one TCM system, has been recently researched. It shows great potential for empowering the system with these abilities (Byrne et al., 1995).

Available sensor fusion methods include multiple indirect sensor fusion and direct plus indirect sensor fusion. Artificial intelligence (AI), especially neural networks (NNs), is the predominant technique in the former method. Even though these methods provide a systematic approach for sensor fusion, the need for extensive training of the neural networks is still a major drawback (Park and Ulsoy, 1993a). More importantly, either supervised or unsupervised neural networks cannot adapt to various cutting conditions. Further research is needed to address this problem partially, if not completely.

By contrast, direct plus indirect sensor fusion seems more attractive due to its valuable advantage that the two different types of sensors can counteract drawbacks of each other and reinforce each other. However, few papers on this scheme have been published. Accordingly, this fusion strategy is used and its implementation of each subsystem is presented in this thesis.

1.3 Objectives and scope of work

The aim is to develop an on-line TCM system which can monitor the flank wear and breakage in milling by integrating vision and force sensors. The specific objectives are to:

1. Build a vision subsystem that can monitor the flank wear with good accuracy and robustness while the spindle rotates.
2. Develop a vision subsystem that can measure the crater wear efficiently.
3. Extract relevant features from the force signal which are sensitive to flank wear and breakage.
4. Implement a force subsystem that can monitor breakage and flank wear based on the extracted features.
5. Integrate the vision and force subsystems into an on-line TCM system.

With these objectives achieved, the developed techniques, subsystems and system can provide:

1. An advanced vision system to monitor flank wear dynamically whereby the cutting operation is minimally interrupted.
2. An efficient method to monitor crater wear with the insert in the milling cutter.
3. An integrated approach for monitoring flank wear and breakage on-line in milling, which is adaptive to various cutting conditions.

To achieve the objectives, the scope of work includes:

1. Integration of one direct (vision) and one indirect (force) sensors.
2. The flank wear along the major cutting edge is studied as generally this wear is the most important aspect to monitor.

3. Experimental setup design for capturing images of tool inserts in the milling cutter rotating with low speed. Moving and still images are to be processed with appropriate techniques.
4. Investigation of a non-contact method of crater wear measurement is proposed, which is based on phase-shifting and fringe projection. But crater wear is not considered in on-line monitoring since flank wear is more often considered in research.
5. Identification and application of suitable neural networks as the estimator to predict the flank wears and tool breakage in milling.

Tool conditions such as wear and chipping/breakage and wear mechanisms in milling are reviewed and the sensors used to monitor these conditions are discussed, especially vision and force sensors. These two sensors are separately reviewed as single-sensor methods, which lay the foundation for sensor integration. After single-sensor methods, multiple-sensor methods are reviewed. By surveying the literature, research gaps in vision and force domain are identified, and hence research orientation is highlighted.

1.4 Organization of the thesis

This thesis is organized as follows:

Chapter 2 reviews the current TCM systems with focus on use of sensors. Basic techniques and systems of TCM are presented and various sensors and their corresponding signal processing methods are reviewed, including direct and indirect sensors, and sensor fusion methods.

Chapter 3 outlines the overall on-line monitoring framework of the proposed TCM system, which integrates in-cycle image processing module and in-process force analysis module.

Chapter 4 presents the individual image processing methods to measure flank wear and detect breakage. Unlike the traditional thresholding-based methods, a rough-to-fine strategy is considered and a threshold-independent edge detection method based on moment invariance is employed for more robust determination of the wear edge with sub-pixel accuracy. The chipped-away part of the insert is quantified to detect breakage.

Chapter 5 extends the work of Chapter 4 that utilizes successive images to analyze the in-cycle processing. The system uses close correlation between successive images to measure flank wear during in-cycle process, whereby the images are captured while the spindle rotates.

Chapter 6 describes a phase-shifting method using fringe patterns to measure crater wear by constructing a 3-D map of the tool insert. By solving and then unwrapping the phase map obtained from four images with different fringe patterns, the 3-D profile of the tool insert is obtained, which provides the overall size of the crater wear land.

Chapter 7 proposes a self-organizing map (SOM) network used to estimate the flank wear in-process based on features extracted from cutting force. The SOM network is trained in a batch mode after each pass using the two features and interpolated wear values. The trained SOM network is applied to the next cutting pass to estimate the flank wear. Breakage detection based on force features is also investigated.

Chapter 8 shows the on-line experimental results under various cutting conditions.

Chapter 9 concludes the thesis and recommends work for future research.

Chapter 2
Literature review

2.1 Tool condition monitoring (TCM) and sensors

2.1.1 TCM

Generally, tool wear, breakage/chipping, chatter, chip breakage, and built-up edge are the tool conditions to be monitored in TCM. So far, most of research work has focused on tool wear and breakage monitoring. One of the main reasons is that these two phenomena are crucial in TCM and difficult to handle; thus bringing about several research challenges to this field. Since these two conditions are to be monitored in this thesis, the review concentrates on them.

Tool wear is defined as change in shape of the cutting edges and their neighboring regions of a tool from its original shape, resulting in progressive loss of tool material during cutting (ISO8688-1, 1989). It has two categories: flank wear and crater wear, as shown in Figure 2.1.

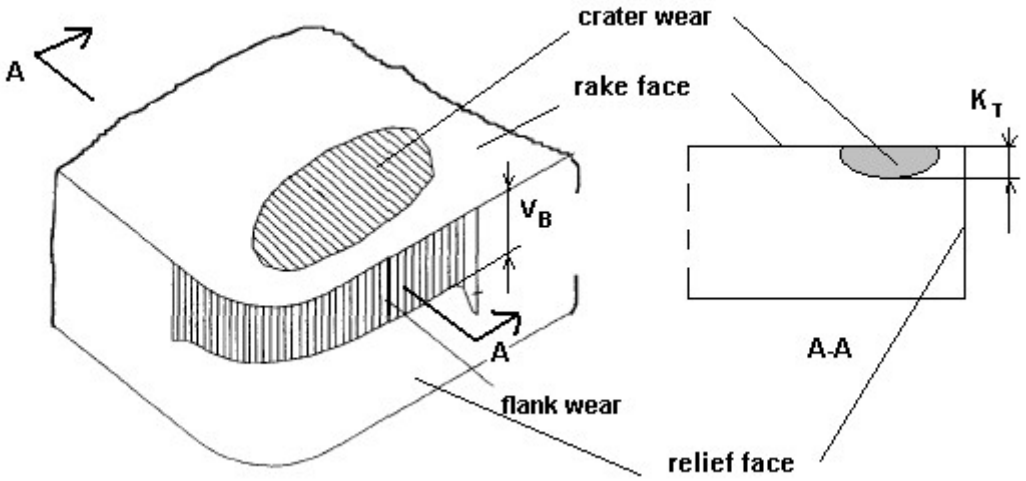


Figure 2.1 Sketch of flank wear and crater wear

Flank wear. Flank wear occurs on the tool relief face (or flank face). It develops under almost any cutting condition, and usually in three stages (Ber and Friedman, 1976), as shown in Figure 2.2. The first stage is a rapid initial wear stage in which the wear develops rapidly to a certain level, within a relatively short time. In the second stage, the wear progresses linearly for a comparatively longer period of time. Much of the useful tool life is within this stage, and therefore this stage is of most concern. The last stage is a rapid accelerated wearing period. In this stage, the wear rate increases rapidly and it is usually recommended that the tool be replaced before this stage. Flank wear predominates under low cutting speed (low cutting temperature).

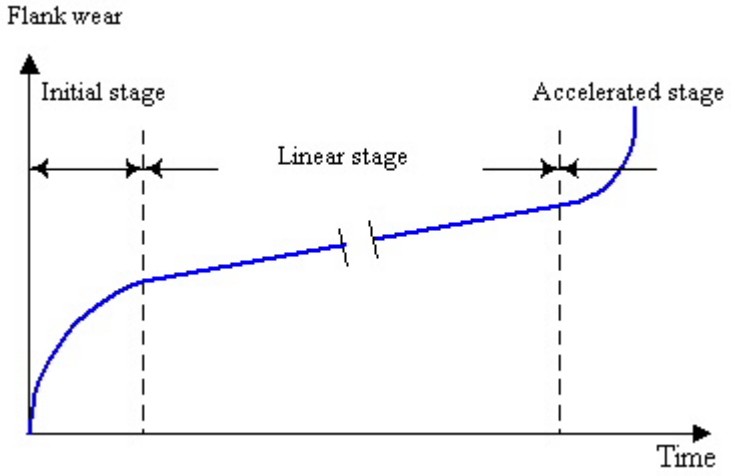


Figure 2.2 Three stages of flank wear

Nose wear forms at the nose radius and near the end relief face of the tool. The wear is partially a continuation of the flank wear around the nose radius and partially a series of grooves that often develops at the front of the tool. It is similar to and is often considered as part of the flank wear. Accordingly, in this thesis, it is also not considered separately.

Crater wear. Crater wear occurs on the rake face. It develops under high cutting speeds or high feeds. The development of crater wear is closely related to the cutting temperature and pressure (Cook, 1973). The crater depth is generally a maximum at a

substantial distance from the major cutting edge, where the cutting temperature and pressure are high. Under certain circumstances the crater may break off from the tool face to intersect the tool major flanks. The general tool geometry correspondingly can vary considerably. Sometimes, fracture or catastrophic failure of the tool results from serious crater wear.

Since flank wear appears in all cutting operations, and it directly affects the quality of machined part, its monitoring is usually considered to be more essential than crater wear.

Chipping/breakage. Chipping happens when the edge line breaks away from the tool cutting edge, rather than wears. A sudden load in intermittent cutting (as in milling) or thermal gradients are two main reasons of chipping. Gross inconsistencies in the workpiece composition or its structure also contribute to chipping. Figure 2.3 depicts the chipping of a tool. When the tool has chipping with length of more than 1mm, this chipping is called breakage. Table 2.1 shows the three kinds of chipping in terms of the size of the chipped pieces (ISO 8688-1, 1989).

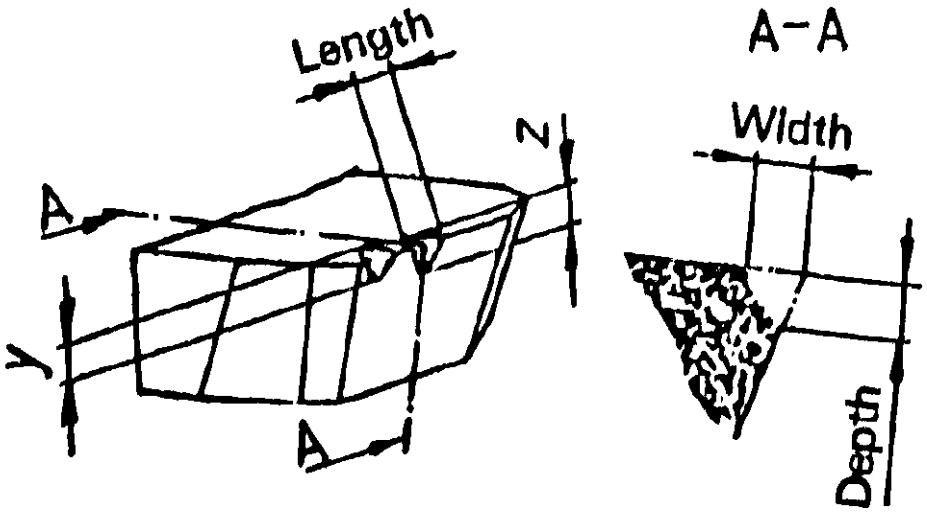


Figure 2.3 Chipping illustration

Table 2.1 Three types of chipping

Length (mm)		*y* or *z* (mm)		
		Small	Normal	Large
Mirco-chipping	< 0.3	0.2	0.25	0.3
Macro-chipping	0.3 to 1	0.25	0.4	0.5
Breakage	> 1	N/A	N/A	N/A

There are many kinds of wear mechanisms responsible for tool wear. Depending on circumstances, the following mechanisms have been outlined (Teo, 1992):

Adhesive wear. Adhesive wear arises from molecular adhesion occurring between tool and workpiece. When the chip slides, it tears away minute particles of the tool material and causes tool wear. This kind of wear can occur at any cutting speed.

Abrasive wear. Abrasive wear involves the removal of the tool material by the scoring action of the inherently hard particles in the machined workpiece, such as inclusions and carbides, causing continuous wear on the surface of the tool. This kind of wear can also occur at any cutting speed.

Oxidative wear. An oxidation process generally occurs at high cutting temperature, particularly at the outer-edge of contact zones where there is free atmospheric contact. As a result, there is a general weakening of the tool matrix which facilitates the tool wear.

Diffusive wear. In diffusive wear, mutual dissolution of the materials occurs between the workpiece and tool and weakens the tool material. It often occurs at a high cutting temperature.

Superficial plastic deformation. This process has a major influence on the crater wear rate when machining with high speed steel tools near the limit of their endurance. The chip is deformed at a very high strain rate and can exert sufficient shear stress onto the surface layer of the tool to deform the latter at a low strain rate. This effect removes tool material from the crater region.

Plastic deformation of the cutting edge. The thermal weakening of the cutting edge region with plastic deformation occurs under normal applied load. It occurs at high cutting temperature and depends on the temperature rise of the tool. Once this deformation occurs, the blunt edge causes an additional heat source as it rubs on the workpiece, further weakening the material and leading to plastic collapse.

2.1.2 Sensors

As mentioned in Chapter 1, a TCM system is essentially a sensor-based system. Hence, sensors are crucial to TCM. To be successful in a machining environment, sensors should meet the following requirements (Kurada and Bradley, 1997a, Niu et al., 1998, Pedersen, 1990, Byrne et al., 1995):

- Good correlation between the sensor signal and the tool condition;
- The response should be fast enough for feedback control;
- Simple in design and rugged in construction and easily integrated into system together with other control and measurement equipment;
- Non-contact, accurate, low-cost and reliable;
- No interference with the machining process;
- Resistant to dirt, chips and mechanical, electromagnetic and thermal influences;
- Function independent of tool or workpiece, and signal transmission reliable.

Numerous sensor types are available for monitoring aspects of the machining environments (Moriwaki, 1993). Generally, the sensors fall into two categories (Cook, 1973): direct and indirect sensors, among which the commonly used are shown in Table 2.2.

Table 2.2 Sensor types in TCM

Direct sensors	Proximity sensors	Takeyama et al., 1967, Stoferle and Bellmann, 1975
	Radioactive sensors	Uehara, 1973, Cook and Subramanian, 1978
	Vision sensors	Yang M. Y. and Kwon O. D., 1996, Kurada and Bradley, 1997b, Karthik et al., 1997, Wong et al., 1997, Pfeifer and Wiegers , 2000, Xu and Luxmoore, 1997, Prasad and Ramamoorthy, 2001, Lanzetta, 2001, Mannan et al., 2000, Park and Ulsoy, 1993b, Maeda et al., 1987a, 1987b, Giusti et al., 1987, Jeon and Kim, 1988, Oguamanam et al., 1994, Lee et al., 1994, Teshima et al., 1993, Giardini et al., 1996
Indirect sensors	Force sensors	Lan and Naerheim, 1986, Altintas, 1988, Altintas and Yellowley, 1989, Lin and Lin, 1996, Tarn and Tomizuka, 1989, Elbestawi et al., 1989, Elbestawi et al., 1991, Tarng, 1990, Tansel et al., 1992a, 1995, Tansel and Mclaughlin, 1993a, 1993b, Leem and Dornfeld, 1995, Zhang et al., 1995, Elanayar and Shin, 1995, Santanu et al., 1996; Xue et al., 1997; Lee and Tarng, 1999; Rene de Jesus et al., 2004
	Vibration sensors	Lee et al., 1987; Tlusty and Tarng, 1988; Reif, and Cahine, 1988; Coker and Shin, 1996; Li et al., 2000a
	AE sensors	Sampath and Vajpayee, 1987; Diei and Dornfeld, 1987a, 1987b; Liu and Liang, 1991; Wilcox et al., 1997; Jemielniak and Otman, 1998a, 1998b
	Power sensors	Matsushima et al., 1982; Bertok et al., 1983; Yellowley, 1985; Stein and Wang, 1990; Li et al., 2000b

Direct sensors

Proximity sensors. These estimate tool wear by measuring the change in the distance between the cutting edge and the workpiece. This distance can be measured by electrical feeler micrometers and pneumatic touch probes. The measurement is affected by the thermal expansion of the tool, deflection or vibration of the workpiece and the deflection of the cutting tool due to the cutting force.

Radioactive sensors. These assess tool wear by monitoring the amount of radioactive material deposited on the chips from the flank face of the cutting tool where the radioactive material is implanted. The need for collecting chips on-line and the hazardous nature of radioactive material limits this technique for laboratory environment.

Vision sensors. These measure tool wear by extracting various morphological parameters with image processing techniques. Because of the availability of CCD camera, vision sensors have been widely used as a direct TCM method. Due to the hostility of the cutting environment (presence of lubricant, built-up-edge or metal deposits on the cutting tool), current vision sensors can only be used between cutting cycles.

Indirect sensors

Unlike direct sensors, indirect sensors measure one or more parameters that can be correlated with tool conditions. The most commonly used indirect sensors are (Kurada and Bradley, 1997a, Bahr et al., 1997, Byrne et al., 1995):

***Cutting force sensors*.** There are typically dynamometers mounted on a tool holder or under the workpiece to monitor the cutting force. Cutting force has been proven to be closely correlated to both the flank wear and breakage or chipping and thus has been extensively used.

AE sensors. These record the elastic stress waves, known as acoustic emission (AE), which is generated by different sources such as friction on the rake face and the flank, plastic deformation in the shear zone, crack formation and propagation, impact of the chip at the workpiece and chip breakage. AE has been very successful in its application to TCM during turning operations (Sampath and Vajpayee, 1987). Its application to milling has been less straightforward. It is difficult to distinguish pulse shock loading occurring during the entry and exit of each individual tooth to the workpiece from that generated during tooth fracture.

***Vibration sensors*.** These sense the level of vibration caused by the friction between the flank face of the cutting tool and the workpiece/the internal fractures of the tool. This type of sensor has the advantage of simplicity and low cost. However, it is widely

understood that vibration monitoring for indirect tool wear detection may not be as accurate or reliable as the methods based on force and AE (Verma and Kline, 1990).

***Current/power sensors*.** These measure current or effective power of feed drives or main spindle. It has some disadvantages. Tool breakages are not detected directly, but only after the consequential damage has occurred. Furthermore, the spindle power is proportional to the resultant cutting force in the direction of primary motion: the least wear-sensitive parameter. This makes wear monitoring very difficult (Byrne et al., 1995).

These direct and indirect sensors, however, should not be necessarily mutually exclusive. On the contrary, integration of several sensors or sensor fusion technique has attracted much attention recently to better and more robustly characterize the cutting conditions.

To summarize, Table 2.3 shows the tool conditions and their corresponding sensing techniques.

Table 2.3 Tool conditions and sensing signals

Tool condition	Causes	Happening frequency	Detectable physical effects	Signals
Wear: Crater wear Flank wear	-Friction -Abrasion -Adhesion -Diffusion	Predominates at high cutting speed or high feed rate (Zhou et al., 1995) Predominates at low cutting speed 3-5 (Lanzetta, 2001)	-Change of force in flank face -Rubbing between tool flank face and workpiece -Change in the effective rake angle	Force AE Vision
Breakage: Fracture Chipping Cracking	-Severe flank wear or crater wear -Impact on tool during entry, exit or by action of swarf -Plastic deformation -The tearing off of materials from tool faces -Non-uniform thermal expansion and contraction -Mechanical stress cycles on tool	Dominant mode for more than a quarter of all the advanced tooling material (Kurada and Bradley, 1997a) 3-5 (Lanzetta, 2001)	-Change in cutting force -Fracture of tool -Change in shear deformation during chip formation -crack formation and propagation	Force AE Vision
Chatter	-The regenerative effects -The mode coupling effect		Vibration of tool	Vibration
Chip breakage	-Bending moment -Local fracture stress	1* (Lanzetta, 2001)	-Change in shear deformation during chip formation and chip/tool interface	AE Force
Built-up edge	-Adhesion -Cratering	1/50-1/100 (Lanzetta, 2001)	Change in effective rake angle	Force

*Assume the occurrence frequency of chip breakage is 1

2.2 Single sensor

In general, in terms of the number of sensors involved, the monitoring methods fall into two categories: single-sensor methods and multi-sensor or sensor fusion methods. In this subsection and the one that follows, these two methods are reviewed.

As previously mentioned, any one of the direct or indirect sensors can be used in monitoring tool conditions. For conciseness, only vision and force sensors are reviewed herein since they are deployed in this thesis.

2.2.1 Vision

Illuminated by appropriate lighting, the wear land (both flank wear and crater wear) reflects the incident light and is sensed by a camera. The image is then processed to get the size of the wear.

Vision method has some valuable advantages (Pfeifer and Wiegers, 2000; Lanzetta, 2001; Pedersen, 1990; Giusti et al., 1987):

- The obtained measurement results are independent from the actual cutting process and its parameters;
- High accuracy, as CCD cameras can get images of the tool area 5×5 mm^2 with high resolution of about 10 μm/pixel;
- No effects on the machine stiffness;
- Universal. Systems using computer vision techniques can measure many different kinds of tools without requiring physical adjustment.
- Tool wear appears in a large variety of forms, which can be classified into several typical groups of wear. As a result, prediction of the tool life can be far more reliable.

In constructing TCM systems with vision sensors, some key issues should be taken into consideration:

- Optimized illumination. Generally, two light sources are used, including directional lighting and structured lighting (Lanzetta, 2001). Only with an optimized illumination, the contours of the worn area on the cutting edge can be extracted from the image with a high degree of reliability. Thus far, hardly there are any systematic approaches found in literature to handle the adaptive adjustment of illumination parameters. Few techniques were reported for application in industrial environment. The main problem is either the lack of measurement robustness or the required complexity in the lighting devices (Park and Ulsoy, 1993b).
- Vision sensors are sensitive to outside environment, such as disturbances, dirt, chips, fluids, and mechanical influences (Lanzetta, 2001).
- The measurement system should be calibrated to provide absolute units of measurement (Kurada and Bradley, 1997a).
- Appropriate segmentation techniques (which can be categorized into 4 classes: thresholding, edge detection, region growing, and split and merge (Bahr et al., 1997)) determine the accuracy of measurement results.

A typical TCM system based on machine vision is shown in Figure 2.4.

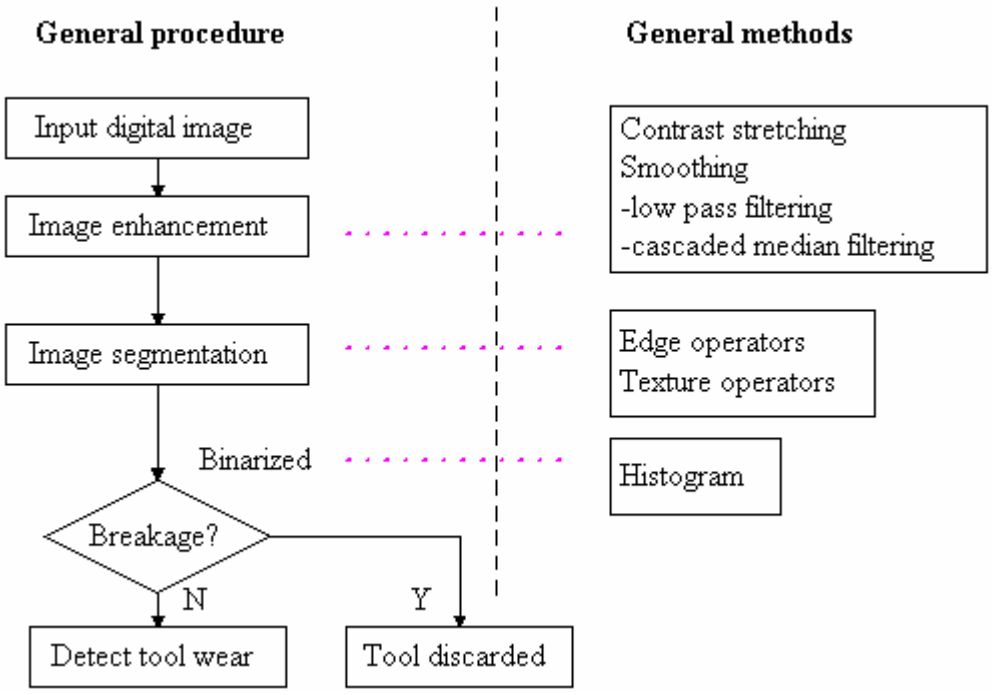


Figure 2.4 General framework of image analysis for TCM

Flank wear by vision

To measure flank wear, three approaches based on rake face image, neural networks and flank face image have been proposed.

Rake face image. In this approach, the image of the rake face rather than that of the flank face was measured (Maeda et al., 1987b). The coordinate transformation relationship between the images of the rake face and flank face was parameterized by rake angle, side angle and cutting edge angle. In this way, by measuring the change of the rake face image of the worn tool with respect to that of the new tool, an entire contour of the flank face could be constructed. However, the accuracy depends on how precisely the orientation of the tool can be achieved with respect to the camera. Another problem is that it may be difficult to obtain the rake face image when the tool is mounted on the mill holder. This approach is, therefore, not considered in the proposed system.

Neural networks. In this approach, the image data were input to a neural network to give the flank wear. Teshima et al. (1993) trained a BP neural network to estimate the cutting tool life. The entire image was partitioned into many small regions, each of which was assigned a value by their gray-level. The image data, together with the cutting conditions, was input to a BP net to estimate the tool information. Giardini et al. (1996) proposed a similar approach but did not consider the cutting conditions in training the neural network. This approach has two disadvantages: lots of trial cutting should be carried out in order to collect training data, and the generalization of the neural network would be low for various cutting conditions. Consequently, this approach is not considered in the proposed system either.

Flank face image. This approach directly uses the image of the flank face to measure flank wear. Most of the researchers used this approach, which is reviewed chronologically below.

One of the earliest approaches was that by Giusti et al. (1987) who employed a special lighting system. They deployed optical fibers to illuminate the flank face and laser beam to illuminate the rake face. The two lighting subsystems worked independently and were successfully integrated into the machining system as a prototype. The image processing technique was very basic and not too robust.

Besides lighting, the image processing technique is also very important. Jeon and Kim (1988) designed an optical system for flank wear measurement using two parameters: VB and VB_{max} with a measurement resolution of 0.1 mm. The image processing began with binarizing the gray-level image into non-wear area and wear land. The binary image was then projected along the cutting edge of the tool. By checking the distribution of the projection, some pixels outside the wear land were

rejected as noise. The wear land contour was determined by neighboring pixels. This rather simple method to measure the flank wear had quite low accuracy.

To improve accuracy, Pederson (1989) developed a prototype experimental computer vision system for flank wear measurement. The gray-level image was binarized with a threshold determined from a smoothed histogram. To remove noise and close unconnected areas, a sequence of shrink and expand operations were applied. Although the resolution of the image was improved to 0.01 mm, the method did not consider actual machining process, in which small chips or dirt could introduce much noise on the flank face, affecting system's robustness.

In an attempt to enhance robustness, Ogumanam et al. (1994) built up a system that extracted five features and classified the tool as good, worn and broken from the captured tool image. The image was segmented firstly by multi-threshold learnt from trial tests. Then neighborhood and connectivity were used to label the unlabelled pixels after the first pass. The wear land was thus detected by searching the pixels with the same specified labels. The cutting edge was detected by Hough transform and was used to detect breakage. This method provided a systematic solution to the measurement of the degree of flank wear and breakage. But it required a series of tests to determine thresholds.

The first two aforementioned approaches based on rake face image and neural networks are not the research trend because their practical use is limited. In contrast, the third method based on flank face image is of research interest and focus as it is more practical. However, to facilitate applications in the industry, there is the need to improve the method for the following aspects:

1. The above systems using flank face image provide pixel accuracy, which may not be adequate for today's precision manufacturing. Actually, sub-pixel accuracy is of greater importance. An optical system with sub-pixel accuracy is therefore needed.

2. Typically, a single image is captured at the instant when the tool is parked in a stationary position at a specified location. In practice, it is more efficient to capture the image when the tool is still rotating. Thus far, there is no reported approach to capturing images from a rotating tool. Accordingly, a system that can capture the image when the tool rotates and process the captured moving image is to be developed.

3. Generally the image is processed with a simple procedure. Noise due to dirt and small adhered chips may not be completely removed and the resulting noise can significantly affect the robustness of the system. Consequently, a robust processing procedure is needed.

Crater wear by vision

To measure crater wear, it is necessary to first construct a 3-D image of the crater wear. Crater wear is not used as commonly as flank wear to indicate the degree of wear. In literature, only a few researchers have reported study in this topic and their work is discussed below.

Giusti et al. (1987) proposed a method using laser fringe to measure crater wear. Laser fringe, obtained after the laser beam goes through a diffraction grating, is projected onto the rake face. Fringe patterns differ between the crater wear area and other areas. With the fringe projection angle, and by tracking the boundary of the fringe in the wear area, the crater wear can be measured. Maeda et al. (1987a) proposed a similar method to measure crater wear. In both methods, the projection angle must be known, and only one fringe pattern is used, which can be sensitive to the fringe spacing.

Yang and Kwon (1996) applied an automatic focusing technique to measure a crater wear depth with a 1-D search algorithm for finding the best focus. The intensity of the illumination is important for both contour detection and best focus search. Thus, light can affect the accuracy easily. The use of a servo control system in finding the best focus increases the complexity of the system.

Karthik et al. (1997), Prasad and Ramamoorthy (2001) constructed a 3-D map of the crater wear using a stereo imaging technique. Two images captured with the camera at different locations relative to the object are used to reconstruct its 3-D image, which is used to measure the crater wear. Because it is necessary to search the corresponding pair of points in the two images, the computation time is long. The use of two cameras is costly for practical application and light can also affect the measurement.

It can be seen that the crater wear measurement systems are subject to ambient light or their processing is not full-field. Because the ambient light will affect the accuracy of the result and the non-full-field processing will make the system complicated and difficult to handle, there is a need to develop a system that is robust to ambient light and can process the image in full-field sense.

2.2.2 Force

Unlike vision sensors, cutting force can be used on-line to detect tool breakage/failure, classify wear level, estimate wear value, or predict tool life, etc. Here, wear usually refers to flank wear as it serves as the wear indicator more often than crater wear (ISO 8688-1, 1989). As an indirect monitoring signal, cutting force has to be processed and analyzed to extract a set of features that are close to tool conditions. The widely used signal processing approaches for feature extraction include:

- Time domain analysis (such as differencing of force (Altintas and Yellowley, 1989), maximum force level and total amplitude (Tarn and Tomizuka, 1989),

variable force (Tarng et al., 1994), statistical value of force (Leem and Dornfeld, 1995));

- Time series analysis (such as AR model (Lan and Naerheim, 1986; Altintas, 1988);
- Power spectrum analysis (such as FFT (Tarng, 1990; Elbestawi et al., 1991; Oraby, 1995));
- Wavelet analysis (Kasashima et al., 1995; Wu and Du, 1996; Xue et al., 1997; Lee and Tarng, 1999; Rene de Jesus et al., 2004).

Artificial intelligence (AI) techniques are widely used to map the extracted features to tool conditions. They are:

- Neural networks (such as MLP (Rangwala and Dornfeld, 1987; Monostori, 1993; Ko and Cho, 1994; Santanu et al., 1996), SOM (Leem and Dornfeld, 1995), ART and ART2 (Burke and Rangwala, 1991; Tansel and Mclaughlin, 1993b), RBF (Elanayar and Shin, 1995), RCE (Tansel et al., 1992b));
- Fuzzy logic (Li and Elbastawi, 1996);
- Pattern recognition (Elbestawi et al., 1989).

Among the ongoing discussed AI techniques, neural networks are the most popular due to the following favorable advantages (Dimla et al., 1997):

- Fault tolerance and adaptability;
- Data-driven nature;
- Noise suppression capabilities;
- Parallel processing capabilities.

For force, the predominant research uses supervised neural networks such as MLP, RBF, RCE or SVM. To train these networks, there is a need to collect a lot of data through experiments under various cutting conditions and perform off-line training.

Additionally, the generalization of the trained networks is not good enough for various cutting conditions. Therefore, there is a need to develop an unsupervised network to use less data in training. Furthermore, the network should be adaptive to various cutting conditions. The features for flank wear estimation and breakage should be sensitive to the two conditions and their extraction should be easy to implement on-line.

There are also model-based methods of relating tool wear to cutting force (Koren, 1978; Marques and Mesquita, 1991; Park and Ulsoy, 1993a; Lin and Lin, 1996) or other measurable signals (Abu-Zahra and Nayfeh, 2000). To implement such a model, a large database must be established through numerous experiments to furnish the constants in the models. This disadvantage renders it not so amenable to practical applications.

Table 2.4 sums up the cutting force features used in literature and the methods to make decision based on them.

Table 2.4 Force features and decision-making: review

Conditions	Features	Decision making	Reference
TBD[1]	Residual error	Thresholding	Lan and Naerheim, 1986; Altintas, 1988
	Components in tool breakage zone	Thresholding	Tarng, 1990
	Sum of the squares of residual errors	Thresholding	Tansel and Mclaughlin, 1993a
	Peak rate	Thresholding	Zhang et al., 1995
	Ten normalized averages in one tool revolution	RCE, ART2	Tansel et al., 1992a, 1992b
	Average force and variable force	MLP	Tarng et al., 1994
	Shape characteristic vectors from wavelet coefficients	ART2	Xue et al., 1997
	Wavelet detail coefficients	Thresholding	Kasashima et al., 1995; Lee and Tarng, 1999
TWD[2]	1st and 2nd order differencing	Thresholding	Altintas and Yellowley, 1989
	Maximum level, Total amplitude, combined incremental changes, amplitude ratio	Thresholding	Tarn and Tomizuka, 1989
	Power spectral density	LDF-classifier	Elbestawi et al., 1989
	Ratio between harmonics	Thresholding	Elbestawi et al., 1991
	Power spectral density and mean, standard deviation, skew, kurtosis	SOM	Leem and Dornfeld, 1995
	Wavelet transform coefficient	ART2	Tansel et al., 1995
	Ratio between the instantaneous increase in the cumulative distribution function (CDF)	Thresholding	Oraby, 1995
TWE[3]	Average force	MLP	Santanu et al., 1996; Lin and Lin, 1996
	Clustering of spectrum and wavelet coefficients	Recurrent NN	Kamarthi and Pittner, 1997

1. TBD: Tool breakage detection.
2. TWD: Tool wear detection.
3. TWE: Tool wear estimation.

2.3 Multiple sensors: sensor fusion and sensor integration

To provide reliable, robust and accurate result in tough industrial environments, researchers have combined two or more sensors, which is termed sensor fusion, in one monitoring system. Thus far, sensor fusion has attracted much attention and research highlight in TCM. Generally speaking, sensor fusion can be divided into two types: fusion by multiple indirect sensors and fusion by indirect plus direct sensors.

2.3.1 Multiple indirect sensors

When using multiple indirect sensors, majority of the proposed methods use neural networks for decision making, partially because of the good ability of neural networks to map the non-linear relationship between input and output. Functionally, multiple-indirect-sensor fusion method is analogous to single-sensor method in the monitoring flow in TCM, involving data acquisition, signal preprocessing, feature extraction, and decision making. The difference is that the feature vector is derived from multiple sensors instead of a single sensor. In principle, in sensor fusion, features from the multiple sensors are able to complement and reinforce each other to result in a better TCM performance.

Therefore, as in the single-sensor method, in multiple indirect sensor fusion method, the features of each sensor signal are extracted using the same or different processing methods. These features and the observed corresponding tool conditions are used to train a neural network. The trained network subsequently provides appropriate output for given input, in a form of either a condition class or an estimation of the tool wear. Table 2.5 shows some indirect sensor fusion systems.

Table 2.5 Multiple indirect sensor fusion systems

Condition	Signals	Decision -making	Success (%)	Reference
TWE	F, AE, SMC	MLP	--	Dornfeld, 1990
TWD	Vt, F	MLP	97	Ko and Cho, 1994
TWE	T, F, AE	MLP	--	Masory, 1991
TWE	T, AE, SMC	MLP	>90	Ruiz et al., 1993
TBD	Vt, F	MLP	100	Markos et al., 1993
TWE	Vt, F	SOM	~95	Kamarthi et al., 1991
TWD	Vt, F	MLP	96.5-98	Monostori, 1993
TWD	AE, F	MLP	96.9	Moriwaki and Mori, 1993
TWD	F, T, AE	MLP	--	Chryssolouris and Domroese, 1988

F = cutting force, AE = acoustic emission, SMC = spindle motor current, Vt = vibration, T = temperature, CNNN = condensed nearest-neighbor network, TWE = tool wear estimation, TWD = tool wear detection, TBD = tool breakage detection, SOM = self-organizing map, MLP = multiple layer perceptron

Even though these methods provide a systematic approach for sensor fusion, the need for extensive training of the neural networks remains a major drawback.

To address this problem, Leem and Dornfeld (1995) designed an unsupervised customized neural network to map the features of force and AE, together with cutting conditions at different wear levels. The neural network is trained by unsupervised Kohonen's feature map procedure followed by an input feature scaling algorithm. The network shows several practical and reliable properties for sensor-based tool-wear monitoring systems at work: off-line feature selection not required, fast learning, fewer teaching samples, and the avoidance of serious types of error (i.e., the reduction of cases where tool wear is misclassified into something other than the level adjacent to the correct one). Burke and Rangwala (1991) also proposed an unsupervised network, ART, to overcome this problem. However, even the unsupervised network still uses the cutting conditions in learning, which means the network is not adaptive to a wide range of cutting conditions.

2.3.2 Direct plus indirect sensors

For the foregoing indirect methods, even sensor fusion is not satisfactory, perhaps because indirect methods can only correlate with actual tool wear conditions throughout its entire monitoring process. In other words, indirect methods are blind to any direct condition information to correct its error, if any. Hence, researchers attempt to combine direct and indirect methods, in an attempt to reduce the drawbacks of one to the other and result in more accurate on-line TCM. A few papers have been published on this sensor fusion scheme, and they are reviewed chronologically below.

Park and Ulsoy (1993a, 1993b) proposed a tool wear monitoring system integrating an adaptive observer and computer vision. A simplified linear model was formulated on the relationship between the cutting force and flank wear. By observing the measured cutting force, some of the unknown parameters could be estimated by the constructed adaptive observer. Other parameters could be calculated or evaluated by empirical equations or data book. The less accurate indirect flank wear measuring method was intermittently calibrated by a more accurate direct measurement method. However, the force-wear model was developed based on turning and not be directly applicable in milling.

Bahr et al. (1997) proposed a similar system, which featured a direct vision measurement and indirect vibration-based on-line monitoring. The vision measurement provided geometrical properties of the tool, such as length, width, perimeter, and area of the wear land to decide the extent of wear. Vibration sensor, on the other hand, was used to monitor the tool condition on-line and could detect tool breakage. A neural network (SOM) was used in image segmentation, which required teaching sample pixels to be provided in training so that the image processing module was not

automatic, since. Another problem was that the vibration signal was not as reliable or accurate as force signal.

Mannan et al. (2000) also studied the combination of direct and indirect sensors. The difference from Bahr et al. was that Mannan et al. monitored the tool wear by inspecting the workpiece surface instead of the tool, and the indirect sensor was a microphone (sensing sound). In the analysis of the image, a neural network was deployed. This may be a drawback, since it is necessary to collect much data to train the neural network. Even though the correspondence of the tool wear and the workpiece surface is very close, it is doubtful that the generalization of the neural network can guarantee an accurate measurement of the tool wear, given a different cutting condition.

From the above survey, there is still much room for improvement to the performance of the reported sensor fusion schemes because the processing methods for one or more of the sensors in the system have disadvantages. It is therefore meaningful to try the sensor integration by vision and force, which are processed with more effective methods, to hopefully achieve a better performance. For the general concept, the overall scheme of the sensor integration is described and discussed in Chapter 3.

Chapter 3
Framework for on-line TCM by multi-sensor integration

3.1 Overview

As mentioned earlier, in this thesis, the focus is on vision and force sensor integration. Before explaining the details of each sensor technology, the general framework for on-line TCM by multiple sensor integration is briefly introduced. Therefore, this chapter describes how the two sensors are combined to work together and how the two sensors play their individual role in monitoring tool conditions.

For clarity, two terms are defined in this thesis: in-cycle and in-process. “In-cycle” refers to “periodic”, such as, between machining cycles or during part changeovers. “In-process,” on the contrary, refers to that during machining. The overall scheme of the proposed on-line TCM system is shown in Figure 3.1.

For flank wear, accurate vision measurement is obtained by processing the images of the tool captured in-cycle. Two features of the cutting force extracted in-process and the flank wear increment in this cutting block are used to train the unsupervised Kohonen’s SOM neural network, which in turn serves to estimate the wear increment in the next cutting block. The two individual sensors play their respective roles in alternative intervals.

To handle exceptions, i.e., breakage event, two force features extracted in-process are firstly checked with preset thresholds. If the thresholds are reached, the image of the tool is then captured and processed to verify if breakage has really occurred.

Note that in this thesis, for convenience of experiment, one pass of machining a cube is taken as one block.

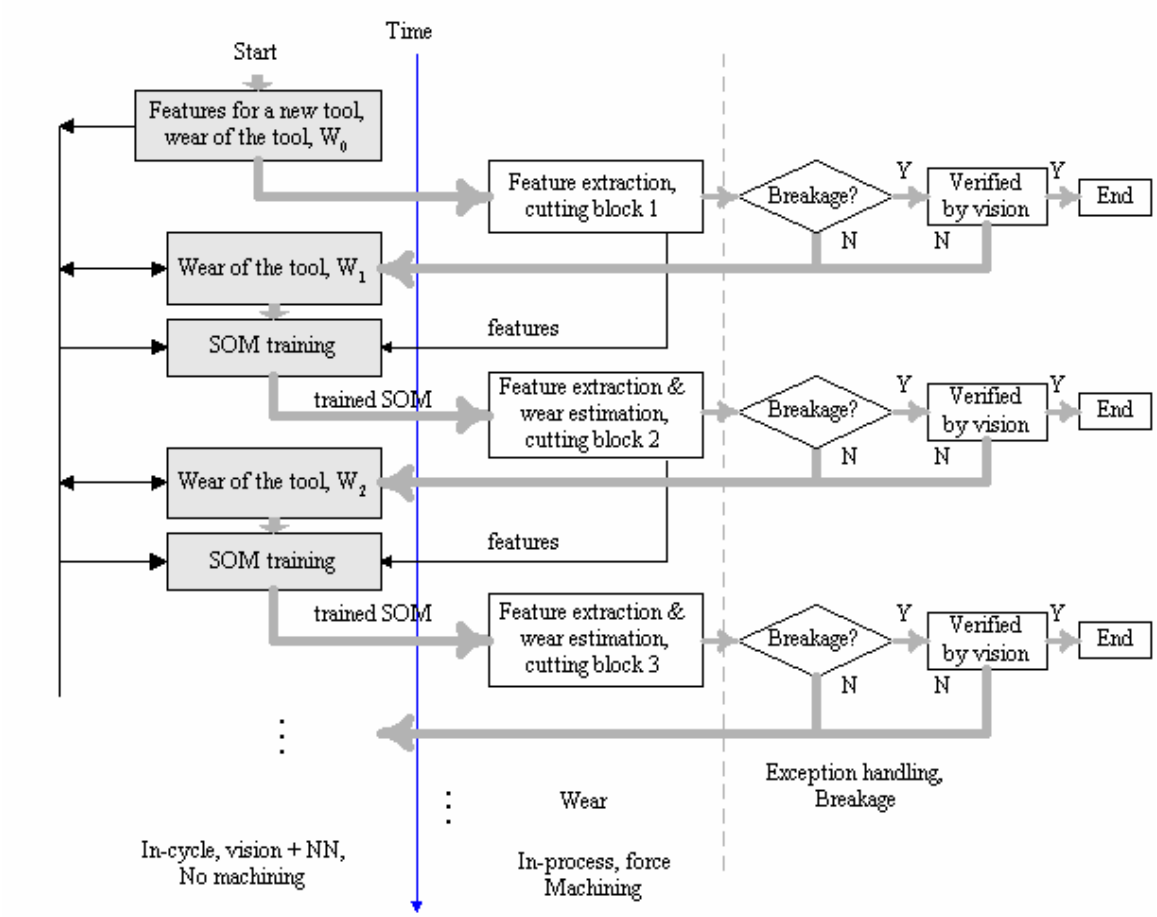


Figure 3.1 Overall scheme of the proposed on-line TCM system

3.2 In-cycle tool wear measurement by vision

From Figure 3.1, it is necessary to know the wear values at the beginning and end of one cutting block to train the SOM net. To perform this task, a CCD camera has been chosen to work in-cycle. For this vision subsystem, there are some key issues:

- It should interrupt the machining procedure as little as possible.
- It should provide accurate measurement results as fast as possible.
- It should be more or less robust to the ambient environment, particularly lighting.
- It can be generalized to a wide range of tool types.

In the light of the above, techniques used to process the images are previewed herein:

For flank wear

1. Not only binary image, but also gray-level image, is involved to measure the flank wear area, compared to traditional thresholding-based methods that use binary image only. Hence, more accurate result can be achieved by avoiding the selection of wrong threshold.
2. Rough-to-fine strategy is employed. The wear edge is located step by step by gradually narrowing the search scope. In this way, the flank wear area can be more reliably detected.
3. Noise is reduced by various means, i.e., morphological operation, windowing technique, median filter, horizontal and vertical scanning.
4. Processing area is reduced from the entire image to the region of interest (wear land) to speed up the processing.
5. Threshold-independent method based on moment invariance is introduced to detect the wear edge with sub-pixel accuracy.
6. Trigger-capture mechanism is proposed and implemented to capture moving image series as the spindle rotates.
7. Successive moving image analysis based on individual image processing is deployed to process the consecutive image series, using the correlation within the image series to make the measurement more accurate and robust.

Techniques 1-5 will be discussed in detail in Chapter 4, which focuses on the individual image processing and lays the foundation for successive image analysis. Then techniques 6 and 7 mentioned above will be described in Chapter 5, which focuses on successive image analysis.

3.3 In-process wear estimation by force

The force subsystem works in-process while cutting is in-progress. The challenges of wear estimation through force signature are:

1. Features suitable for on-line use and close to wear progress should be extracted and properly processed, say, scaled.
2. The dimension and the extraction procedure of the features should be as simple as possible due to the on-line application.
3. A good unsupervised neural network should be selected according to the features. Its training should converge fast and end before the next machining cycle.
4. Cutting-condition independent estimation method is preferred to make the system adaptive to various cutting conditions.

The solution to these challenges is presented in Chapter 7.

3.4 Breakage detection and verification

Generally, breakage detection is based on thresholding. Once one or more features of a sensed signal (AE, or force, etc.) have a value exceeding a preset value (threshold), the tool is viewed as failed. However, to specify such a threshold is a big challenge. Accordingly, a wrong decision on the breakage may be made due to a deviated threshold. To address this problem, there is an available tool – vision. So once again, to verify this breakage, the flank face is inspected by vision. Part of Chapter 4 and 7 touch on the issues on breakage.

Chapter 4
Individual image processing

4.1 System configuration

Figure 4.1 shows the camera system mounted on a CNC machining center. The camera system consists of the following:

- EG&G Reticon MD4256C CCD camera with a 40 mm-extender to zoom on to the insert. The image size is 256 pixel × 256 pixel. The field of view of the camera is about 3 mm × 3 mm. The horizontal and vertical pixel resolution is 12 μm/pixel. A telecentric lens Computar TEC-55 is used. The vision system was calibrated with the method proposed by Blahusch et al. (1999).
- Frame grabber EG&G Reticon SB4001.
- Adjustable fibre-optic lighting.
- Microcomputer Pentium III 550.

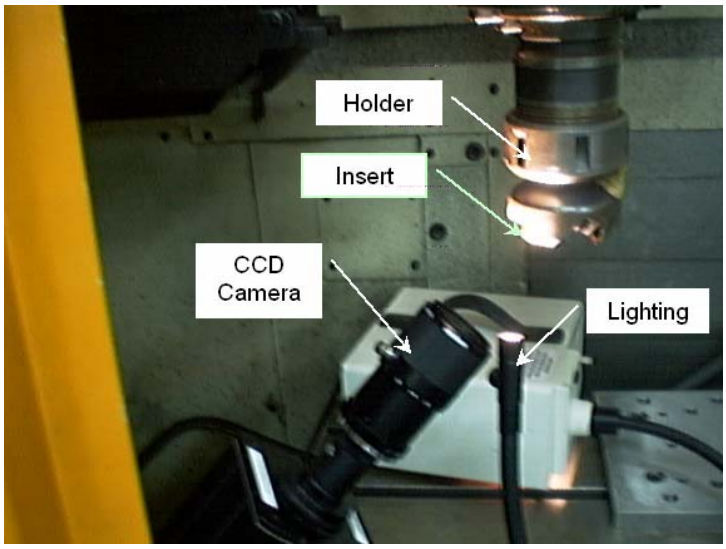


Figure 4.1 Experimental setup for individual image processing

During the experiment, the milling cutter was programmed to park at a preset

position with respect to the CCD camera so that a sharp image could be captured. For different inserts, the spindle was manually rotated to an appropriate position where the flank face could be clearly inspected by the camera. In this way, the pixel resolution was consistent for all images of different inserts. The insert remained still in the slot of the cutter when its image was captured.

4.2 Definition of terms

To proceed, some key terms used throughout the thesis are defined.

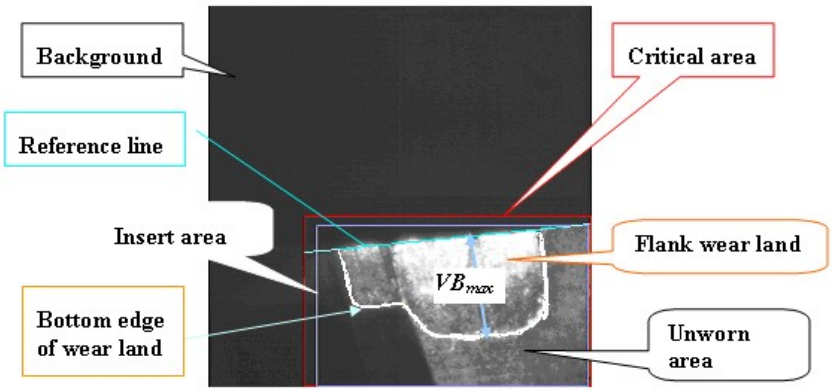


Figure 4.2 Definition of key terms

As shown in Figure 4.2, a typical CCD-captured insert image consists of three different areas:

a). The ***background*** (relative to the insert), where each pixel has a low gray level (dark area).

b). The ***unworn area*** of the insert, where each pixel has an intermediate gray level.

c). The ***flank wear land*** (of the insert), where each pixel has a high gray level (bright area).

The ***insert area*** is defined as the minimal rectangular area that circumvents the whole insert appearing in the captured image. The ***critical area*** is defined as the rectangular area obtained by extending the insert area by some appropriate amount to accommodate some errors. The ***reference line***, refers to the cutting edge of the insert.

The ***bottom edge of wear land*** is a set of points, which are on the boundary between the worn and unworn areas.

The aim of this chapter is to find and quantify the flank wear land.

4.3 Identification of the critical area

To reduce the processing time and memory requirement, the processing area of the captured image is confined to the critical area. The background, unworn area and flank wear land, which are at different gray levels, provide the means of identifying the critical area, based on the schematic steps shown in Figure 4.3.

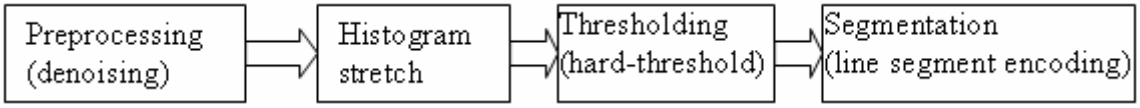


Figure 4.3 Schematic steps for identification of the critical area

4.3.1 Preprocessing

In preprocessing, noise is reduced using median filtering (Gonzalez and Woods, 2002), which has the advantage of keeping the detailed information while minimizing the noise so that edges can be clearly identified. This is important as in the following steps, edges in the image are used to determine the critical area. The original image and the median-filtered image are denoted as $A(x,y)$ and $O(x,y)$, respectively. Here, the value of a pixel in $O(x,y)$ is replaced by the median of the gray levels in the neighborhood of that pixel in $A(x,y)$. As can be seen from the median filtered image (Figure 4.4 (b)), the pixels within each area appear more uniform and smooth.

In digital image processing, the median filter is carried out in a window, which can also be called as a mask. The side value (size) of the window is usually odd. A large size of the window will result in much computation time and may bring about the

unexpectedly distortion of the resultant image. A small size window may not get the noise removed. In this case, 5 × 5 window is selected.

4.3.2 Histogram stretch

The histogram stretch is employed for two purposes. Firstly, it is used to improve the contrast of the median-filtered image. Secondly, it is used to differentiate the average gray level of the wear land and that of the unworn area pixels, within a relatively constant interval, by mapping the original gray-level interval into a fixed interval of [0,255] with a linear transform (Gonzalez and Woods, 2002).

Suppose the linear transform is:

$$y = ax + b \tag{4.1}$$

where x is any gray level of the pre-transformed image, y is the corresponding desired gray level after transform; a and b are unknown variables, which can be calculated by solving two equations.

Consider that the gray level range of the pre-transformed image lies over [G_{max}, G_{min}], the desired gray level interval is generally [0, 255]. Two equations are obtained:

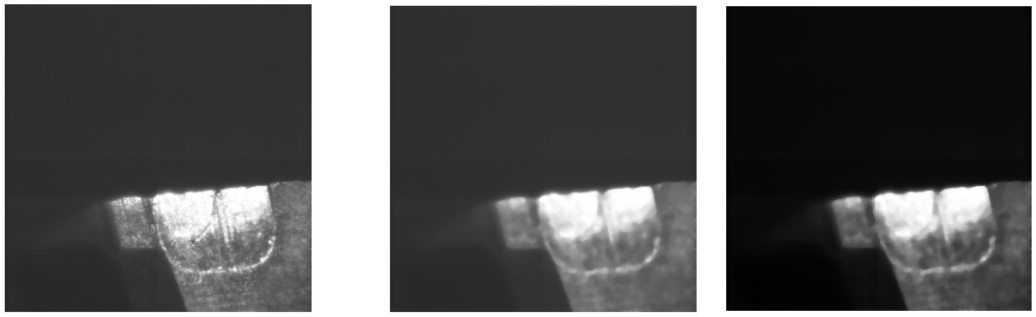
$$0 = kG_{min} + b \tag{4.2}$$

$$255 = kG_{max} + b \tag{4.3}$$

which give:

$$a = \frac{255}{G_{\max} - G_{\min}},\ b = -\frac{255}{G_{\max} - G_{\min}} G_{\min} \tag{4.4}$$

The image after histogram stretch has greater contrast in each gray-level area (Figure 4.4 (c)).



(a) (b) (c)

(a) Original image *A(x,y)*, (b) Median-filtered image *O(x,y)*, (c) Histogram-stretched image *S(x,y)*.

Figure 4.4 Gray-level images after preprocessing and histogram stretch

4.3.3 Thresholding

The background pixels are assigned a value of '0' and other pixels '1', with a suitable threshold T_S. The resultant binary image is denoted as $B(x,y)$. It is obtained through this formula:

$$B(x,y) = \begin{cases} 1 & ,if \quad S(x,y) \geq T_S \\ 0 & ,else \end{cases} \tag{4.5}$$

This threshold is easily obtained by searching the gray-level histogram, with the knowledge that the background has the lowest gray level.

As shown in Figure 4.5, a typical histogram of the image $S(x,y)$ has three visible clusters, which correspond background, unworn area and wear land respectively as the gray level increases. Additionally, the background has the maximum density value in the histogram distribution.

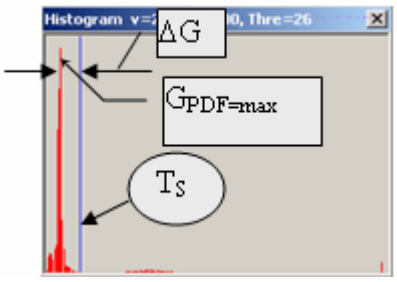
Since the unworn area and wear land as a whole in this context is the object to be labeled with '1', the iterative threshold selection algorithm (Sonka et al., 1999) is not

suitable. There are two main reasons. One is that having too many iterations requires time. The other is that the histogram is sparse, i.e., the background has very compact distribution, while the non-background has very wide distribution. If the threshold is selected with this algorithm, some non-background pixels will be wrongly classified into background in the resultant binary image. Accordingly, the obtained critical area may be shrunk compared to its minimal size.

From the histogram pattern, a simple way to get a suitable threshold is designed.

a). Search the histogram, finding the gray level with the maximum density value, i.e., $G_{PDF=max}$;

b). Let $T_S = 2G_{PDF=max}$.



(a) Histogram of image $S(x,y)$

(b) Resultant binary image

Figure 4.5 Image thresholding

In this method, the underlying hypothesis is that the background has a Gaussian distribution. Consequently, $G_{PDF=max}$ is not only the mean of the Gaussian but also the value of several times of standard deviation, say 5σ, such that outside of range [0, $2G_{PDF=max}$] are there nearly no background pixels. Figure 4.5 (b) shows the resultant binary image.

4.3.4 Extraction of the critical area

To extract the critical area, the "line segment encoding" segmentation method (Castleman, 1996) is used.

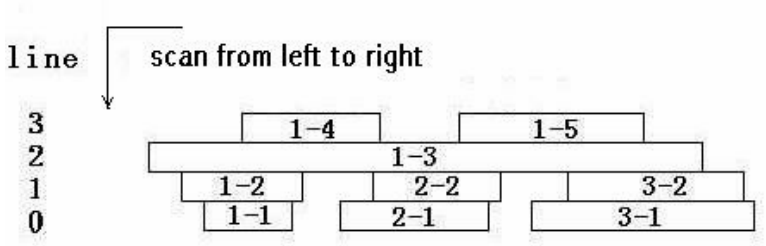


Figure 4.6 Line coding method sketch map

Suppose there is a region in an image as shown in Figure 4.6, the algorithm goes as follows.

Scanned from line 0, all the connected pixels whose values are 1 will be viewed as a section. According to this rule, in line 0, three sections are obtained. They belong to three different regions which are encoded as region 1, region 2 and region 3. The three sections are encoded as 1-1, 2-1 and 3-1, which mean they are the first sections of three regions.

In line 1, the algorithm finds other 3 sections, which are neighbors of the three sections found previously in line 2. The three new sections belong to region 1, region 2 and region 3 respectively, so the algorithm encodes them 1-2, 2-2 and 3-2, which mean they are the second sections of three regions.

In line 2, only one section is ascertained. It is neighbor of all the three sections of 2-1, 2-2 and 3-2, so region 1, region 2 and region 3 actually are only one region. They will be merged into region 1 as a result. At the same time, this new section will be encoded as 1-3, which means it is the third section of region 1.

Perform the scanning of each line for the entire image in the same way, thus all isolated regions are obtained.

The insert area is selected out from all candidate areas after segmentation with the criterion that the insert area is the greatest one in terms of area. In Figure 4.5 (b), the insert area is outlined. After the insert area is found, the critical area is thus obtained

by magnifying the insert area by some amount, in order to accommodate possible errors in processing. The image processing to identify and measure the flank wear land is subsequently performed within the critical area.

4.4 Identification of flank wear land

The determination of the flank wear land requires the extraction of the edge image to obtain the reference line.

4.4.1 Edge detection and enhancement

The fundamental principle of edge detection is based on the abrupt or gradual gray values' change of several continuous pixels. For an input image $f(x,y)$, its edge image can be obtained by calculating its gradient magnitude via any one of the following equations (Gonzalez and Woods, 2002):

$$g(x,y) = \max\{|G_x|,|G_y|\} = \max\{|\frac{\partial f}{\partial x}|,|\frac{\partial f}{\partial y}|\} \tag{4.6}$$

$$g(x,y) = |G_x| + |G_y| \tag{4.7}$$

$$g(x,y) = \sqrt{G_x^2 + G_y^2} \tag{4.8}$$

These three equations give slightly different results. But the complexity of Eqs. (4.6) and (4.7) is less than Eq. (4.8) and they are usually used in practice.

In computing the edge image in digital image processing, a kernel is often used that is a square matrix to convolve the image, which is also essentially a matrix. For the edge detection operation, the Sobel operator, Eq. (4.8) is applied as it incurs almost no extra noise to the image.

Sobel operator (Sonka et al., 1999; Gonzalez and Woods, 2002) has two kernels, G_x and G_y, weighting the central row and column respectively, which are shown as follows:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{4.9}$$

As the horizontal edges are more critical in the proposed method, only the horizontal edge operator G_y is calculated with the image data. Figure 4.7 (a) shows the Sobel-operated image, which is denoted as $E(x,y)$.

In order to improve the edge contrast of $E(x,y)$, a local window function, called adaptive contrast enhancement at edges (Klette and Zamperoni, 1996), is employed.

The local edge gray value E is defined as a weighted and normalized average gray value inside the window whose center is the pixel at (x,y).

$$E = \frac{\sum_{i,j=0}^{n-1} d_{i,j} f(x-I+i, y-J+j)}{\sum_{i,j=0}^{n-1} d_{i,j}} \tag{4.10}$$

where n is the size of the window and given an odd value according to the general rule. The weight $d_{i,j}$ is edge value of the pixel at $(x\text{-}I+i, y\text{-}J+j)$. $I=J=n/2$.

The local contract C between gray value $f(x,y)$ and E is defined by using the Canberra distance measure:

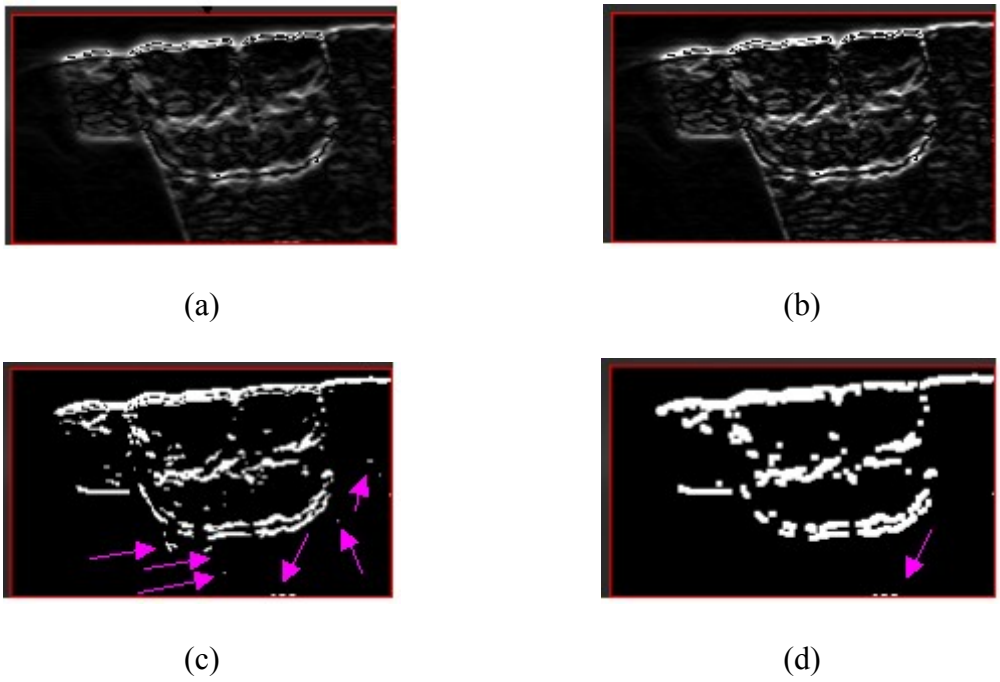
$$C = \frac{|f(x,y)-E|}{f(x,y)+E}, \text{ with } 0 \le C \le 1 \tag{4.11}$$

Let the wanted contrast $C' = C^r$, the resulting $h(x,y)$ can be obtained by the following equation:

$$h(x,y)=\begin{cases} E\dfrac{1-C'}{1+C'} & \text{,if } f(x,y)\le E \\ E\dfrac{1+C'}{1-C'} & \text{,else} \end{cases} \tag{4.12}$$

When $r<1$, the contrast can be enhanced.

Figure 4.7 (b) shows the enhanced image $E^+(x,y)$ which shows a higher-contrast edge.



(a) (b)

(c) (d)

(a) Sobel-operated gray-level edge image $E(x,y)$, (b)Enhanced gray-level edge image $E^+(x,y)$, (c) Binary edge image $B_E{}^+(x,y)$ after Otsu method, (d) Binary edge image $M(x,y)$ after morphology.

Figure 4.7 Edge and binary edge images confined to the critical area outlined by the red rectangle (Arrows indicate noise patches)

4.4.2 Thresholding the edge image

As the pattern of the histogram of the image $E^+(x,y)$ is quite different from that of the image $S(x,y)$, an adaptive determination of an optimal threshold based on the Otsu method (Otsu, 1979) is employed:

a). Calculate the histogram $H(i)$ of the edge image $E^+(x,y)$;

b). Calculate the average gray level of the image;

$$\mu_T = \sum_{i=0}^{255} i \cdot H(i) \tag{4.13}$$

c). Calculate the average of the gray class k and the sum of the class histogram:

$$\mu(k) = \sum_{i=0}^{k} i \cdot H(i), \ \omega(k) = \sum_{i=0}^{k} H(i); \tag{4.14}$$

d). Calculate the standard of class classification:

$$\sigma_B = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1-\omega(k)]}; \tag{4.15}$$

e). Finally, calculate k so that σ_B reaches its maximum, the optimum threshold is then:

$$T_E = k. \tag{4.16}$$

Figure 4.7 (c) shows the resultant binary image, $B_E^+(x,y)$, binarized with the threshold via Otsu method.

4.4.3 Reference line parameterization by Hough transform (HT)

As mentioned earlier, the reference line is actually the crater face when seen from the flank face. This line acts as the reference line in detection of flank wear and should

be located precisely. For line detection, Hough transform (HT) (Klette and Zamperoni, 1996; Castleman, 1996) is an effective method often used.

Principle of HT

A straight line in x-y coordinate system can be described in the spatial domain by a set of parameters. For example, as shown in Figure 4.8, line $y = ax + b$ is parameterized by the slope a and intercept b. In another coordinate system, polar system, this line can be written in a parameter set (d,α), where d is its distance from the origin and α the angle of its normal. The d-α plane is parameter space.

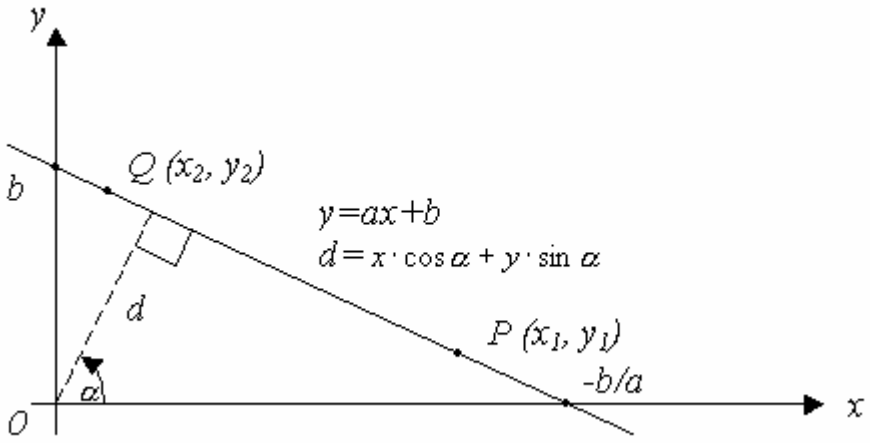


Figure 4.8 Principle of Hough transform

In light of the transform, any point (x_i,y_i) on the line satisfies the following condition:

$$d = x_i \cos \alpha + y_i \sin \alpha \tag{4.17}$$

For example, the coordinates of collinear points P and Q on the line L, satisfy the conditions:

$$d = x_1 \cos \alpha + y_1 \sin \alpha \tag{4.18}$$

$$d = x_2 \cos \alpha + y_2 \sin \alpha \tag{4.19}$$

Therefore, P and Q will "vote" for the specified set (d,α). Considering all pixels in an image, calculate their votes for each discretized pair (d_i,α_i); obviously, the line that has the most collinear points has a peak "vote" value in the parameter space. By searching the peak value, the line is identified.

Hence, the HT reduces the problem of searching locally for lines in image space to searching globally for peaks in parameter space

Data preparation

The data for HT are the parts from the morphology resultant image $M(x,y)$ in the processing. If all pixels are individually used as the input, too much storage space is required. Considering there are some connected points which have same value, '0' or '1', only the two end points of the data segments, which are composed of a series of connected points with the same value of '1', are enough to be input as the spatial points for HT. Figure 4.9 exemplifies this idea.

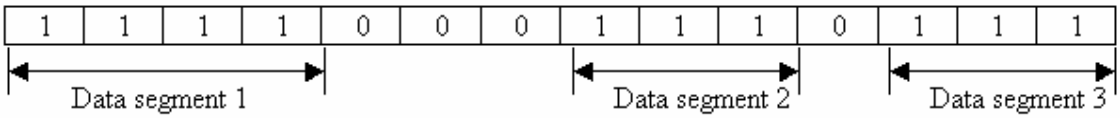


Figure 4.9 Data structure for Hough transform

For one line scanned from left to right, there are three data segments found, each of which has a start point and an end point. These two end points' coordinates and the distance between the two points are recorded as a data structure. Obviously, this data structure is more efficient than every point's coordinates themselves as the input to HT. In this example, 6 points for three segments are recorded instead of all the 10 points. Even plus the storage space for the distance, the storage saving here is remarkable (about 40%).

The same operation has been manipulated along each column in order to be able to detect the line whose slope's absolute value is less than one.

Reduction of the transform area

This has been achieved after identification of the critical area. Now that the HT is performed just in the critical area, which is smaller than the entire image area. As a result, the storage space decreases and computation speed increases.

Fast calculation of cosine and sine

Basically, in calculating cosine and sine or even the right-hand value of Eq. (4.17), look-up tables are a good choice. But in some dynamic calculation of cosine and sine because of their adaptive resolution, although a very detailed look-up table works, it consumes too much memory. In what follows, a simple numerical calculation method is presented.

In HT, calculation of cosine and sine of each discretizated angle value over the entire search interval, which is always $[0,180^0)$ (Immerkær, 1998). It is time-consuming to directly calculate the cosine and sine values of each angle increment by using the functions provided by the compiler, i.e., cos(x) and sin(x) (Obviously, as processors get faster, this numerical method may not be necessary). According to the triangular relationship, the following is performed (with reference to Figure 4.10):

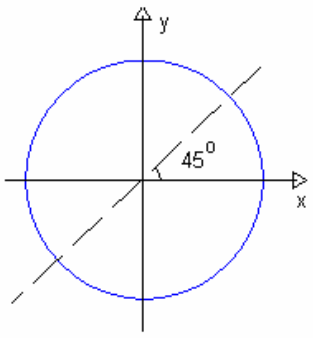


Figure 4.10 Triangular symmetry relationship regarding 45^0, 90^0, 180^0

a). For $\alpha \in [0^0, 45^0]$, calculate $\cos\alpha$ and $\sin\alpha$ directly by functions cos(x) and sin(x).

b). For $\alpha \in (45^0, 90^0]$, according to $\cos\alpha = \sin(90^0-\alpha)$ and $\sin\alpha = \cos(90^0-\alpha)$, $\cos\alpha$ and $\sin\alpha$ can be obtained directly by copying the values of the corresponding $\sin\alpha$ and $\cos\alpha$ obtained in a).

c). For $\alpha \in (90^0, 180^0]$, according to $\cos\alpha = -\cos(180^0-\alpha)$ and $\sin\alpha = \sin(180^0-\alpha)$, $\cos\alpha$ and $\sin\alpha$ can be obtained directly by copying the values of the corresponding $\cos\alpha$ and $\sin\alpha$ obtained in a) and b).

Constraint of the parameter space

In hardware configuration, the reference line's approximate orientation can be easily controlled within a range, say, $(0,90^0)$, so that an angle searching area can be limited in a narrow interval rather than a 180-degree search for the HT. Based on the same principle, the polar distance of the reference line can also be easily estimated within a range, which can serve as the polar distance search interval.

This strategy brings about three remarkable advantages:

1). Faster computation can be achieved as a result of narrowing both the distance and angle search intervals;

2). Less run-time memory is needed because of the reduced scale of parameter sets in parameter space;

3). More accurate in locating the reference line results by avoiding the risk of transforming points on other line patterns, which might create unexpected results.

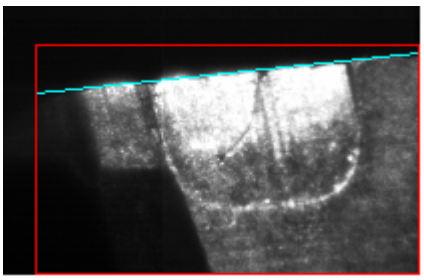


Figure 4.11 The identified reference line

The reference line

The reference line produces the most noticeable line pattern among all the candidate line patterns available in the critical area. Based upon this fact, by seeking through the resulting HT parameter space with Quick Sort algorithm (Klette and Zamperoni, 1996), the set (d,α) with the greatest accumulation value is taken as the parameters of the reference line. Figure 4.11 shows the reference line superimposed on the gray-level image.

4.4.4 Morphology

As indicated by the arrows in Figure 4.7 (c), $B_E^+(x,y)$ consists of small areas of white patches that can be regarded as noise. To remove these, two morphology operators, namely, erosion and dilation, are performed on the binary image $B_E^+(x,y)$ sequentially. The main purpose of the erosion operation is to remove small areas regarded as noise that will introduce errors when locating the wear land. The purpose of dilation is to connect broken segments that originally are on one boundary of a region.

The definitions of erosion and dilation are the following:

Erosion:

$$B * S = \{x: S_x \subset B\} \tag{4.20}$$

Dilation:

$$B \oplus S = \{x: S_x \cap B \neq \varnothing\} \tag{4.21}$$

where B stands for the function of binary image; S stands for the structuring element; S_x stands for the nucleus of the structuring element.

Figure 4.12 shows the visible effects of erosion and dilation on a binary image.

(a) Erosion (b) Dilation

Figure 4.12 Morphological operation

The resultant image $M(x,y)$ after the morphology operation is shown in Figure 4.7 (d). It can be seen that much noise has been removed, and most pixels on the bottom edge of wear land and reference line have been connected. The remaining noise can be removed further through windowing technique discussed later.

4.4.5 Image rotation

To make the subsequent calculation easier, with the reference line in hand, the critical area is rotated such that the reference line aligns with the x-axis (the horizontal line) using bilinear interpolation.

A rotation through an angle θ about the origin can be expressed in homogeneous coordinates as:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4.22}$$

where (X,Y) is the new coordinate, (x,y) is the old coordinate.

Obviously, the new coordinate has fractional value, which means that one output pixel falls into the space between four input pixels. So interpolation is necessary to determine what gray level corresponds to that position. For binary image, nearest neighbor interpolation is ok, since there are only two levels: 0 or 1. For gray level image, however, the nearest neighbor interpolation will result in sawtooth effect at

some of the edges. In this case, bilinear interpolation method is then used, which produces better result at cost of slight increase in execution time.

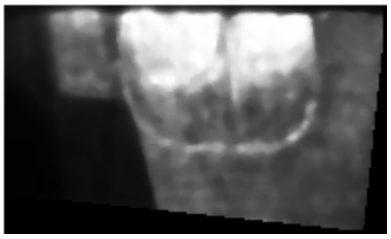
Let $f(x,y)$ be a function of two variables that is known at the four vertices of the unit square. The value of $f(x,y)$ at an arbitrary point (x,y) inside the square can be established by interpolation through:

$$f(x,0) = f(0,0) + x\,[f(1,0) - f(0,0)] \tag{4.23}$$

$$f(x,1) = f(0,1) + x\,[f(1,1) - f(0,1)] \tag{4.24}$$

$$f(x,y) = f(x,0) + y\,[f(x,1) - f(x,0)] \tag{4.25}$$

In real calculation, the unit square is replaced by the four neighboring pixels of the fractional position. Figure 4.13 shows the rotated binary image and gray level image.



(a) Rotated gray-level image

(b) Rotated binary image

Figure 4.13 Image rotation

4.5 Flank wear measurement

The flank wear measurement is based on the resultant histogram stretch image (gray-level image) $S(x,y)$ and the binary image after morphology $M(x,y)$. The processing is confined to the critical area, which is the region of interest (ROI).

As illustrated in Figure 4.14, the measurement procedure, herein referred to as ***orthogonal scanning***, involves repetitive scanning vertically to determine points on

the bottom edge of wear land and further, to get the wear value along each line and the entire wear land.

A ***scan line*** is the vertical line that starts from a point on the reference line and ends at a point on the boundary of the critical area. For simplicity, for the *i*th scan line $L(i)$, the start point is denoted as $P_A(i)$, and the end point $P_B(i)$ with respect to the origin at the upper-left corner of the critical area. The interval, $\triangle D$, between each pair of adjacent scan lines $L(i)$ and $L(i+1)$ is typically '1' pixel apart.

The measurement procedure uses information provided by the preceding processing, as indicated in Figure 4.15. In wear detection module, a rough-to-fine strategy is used, i.e., the edge point, denoted as $P_E(i)$, is firstly located in the binary image $M(x,y)$ in *i*th scan line, using a windowing technique. The detected $P_E(i)$ is used as a reference point to more precisely detect in the gray-level image $S(x,y)$ the point on the bottom edge of wear land, denoted as $P_{RB}(i)$. In this way, the distance between $P_{RB}(i)$ and $P_A(i)$ gives a flank wear value, $W(i)$. Detailed explanation follows.

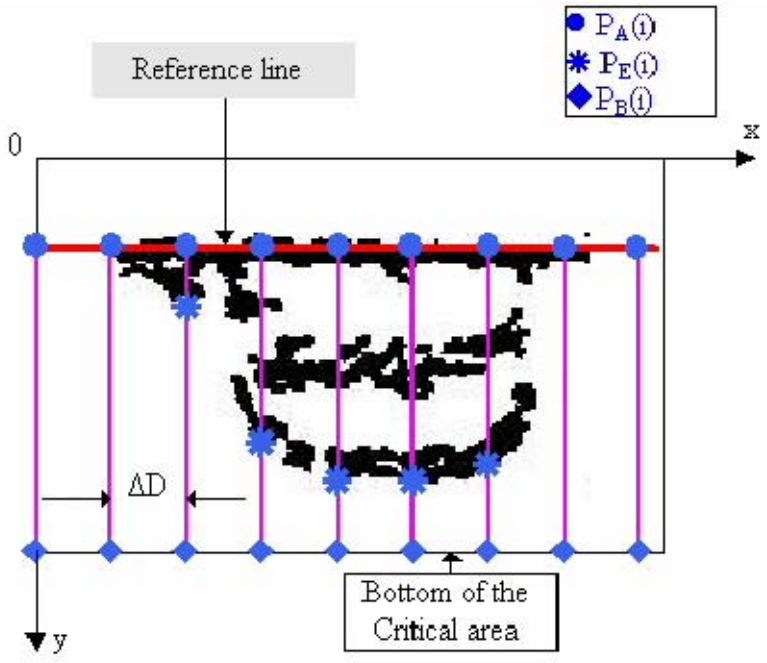


Figure 4.14 Illustration of orthogonal scanning

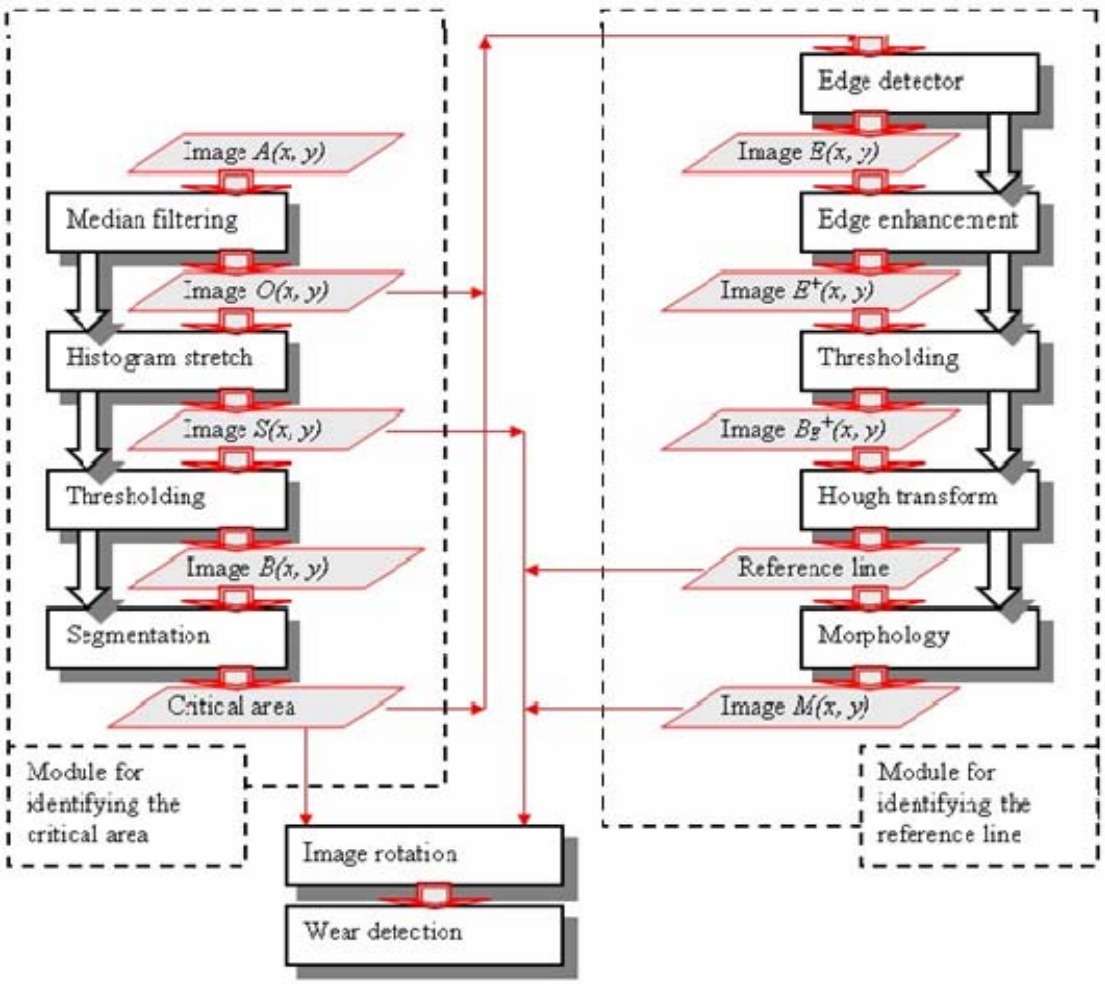


Figure 4.15 Flow chart of procedures for wear detection

4.5.1 Rough bottom edge detection

With reference to Figure 4.14, to locate rough bottom edge point is equivalent to locate point $P_E(i)$. To do this, the binary image $M(x,y)$ is used.

Determination of scan lines

The origin is taken to be at the upper-left corner of the critical area, whose coordinate system is shown in Figure 4.14. A scan line can be determined easily through getting the coordinates of its two end points, $P_A(i)$ and $P_B(i)$ now that the parameters for the critical area and the reference line are known.

Determination of $P_E(i)$

In performing this task, the procedure involves scanning from $P_B(i)$ to $P_A(i)$. The first white ('1') pixel may be basically regarded as $P_E(i)$. However, $P_E(i)$ obtained in this way is easily subject to noise (for example, noise indicated by the arrow in Figure 4.7 (d) will be regarded as $P_E(i)$). Considering that the noise is normally isolated and small, a windowing technique is introduced to improve the performance.

After finding a candidate edge point whose binary value is '1', a given number of adjacent pixels (a w_1-*pixel* length window) is considered. If the number of '1' pixels (N_1) in the window exceeds a preset threshold N_T^E, this candidate edge point is taken as $P_E(i)$; otherwise, find another candidate edge point, and repeat the above operation. In this process, the window is moving, as illustrated in Figure 4.16, where it is supposed that the window length is 4-pixel (w_1=4), N_T^E=3. In practice, selection of w_1 and N_T^E is based on noise size and distribution observed in the binary edge image. Observing that noise is small and isolated, w_1 = 10 and N_T^E = 4 is a good choice after trial.

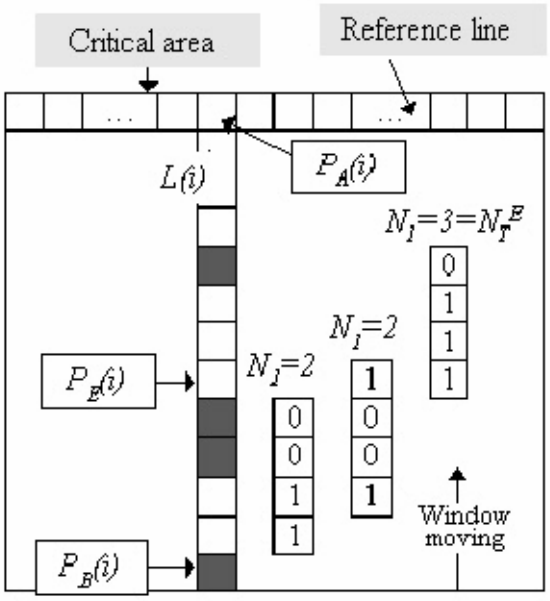


Figure 4.16 Moving window

4.5.2 Fine bottom edge detection

In this section, the wear point is determined with the reference line, $P_E(i)$, $P_B(i)$ and the gray-level image $S(x,y)$ after histogram stretch.

Processing gray-level data

The difference in the gray level between the wear land and the unworn area of the scan line $L(i)$ serves as a criterion in determining $P_{RB}(i)$. The average gray level in the wear land, $\overline{G}_w(i)$, and the average gray level in the unworn area, $\overline{G}_{\overline{w}}(i)$ can be easily computed since $P_A(i)$, $P_E(i)$, and $P_B(i)$ are available for scan line $L(i)$.

The average gray level difference, $\Delta\overline{G}(i)$, between $\overline{G}_w(i)$ and $\overline{G}_{\overline{w}}(i)$ is given by:

$$\Delta\overline{G}(i) = \overline{G}_w(i) - \overline{G}_{\overline{w}}(i) \tag{4.26}$$

A scan line $L(i)$ with wear (that is, $P_{RB}(i)$ exists on this line) has to meet the following constraint:

$$\Delta\overline{G}(i) \geq G_{w-\overline{w}} \tag{4.27}$$

where $G_{w-\overline{w}}$ reflects the minimum gray-level difference between the wear land and unworn area. This can be obtained experimentally.

For scan line $L(i)$ satisfying Eq. (4.27), assume that wear occurs in the vicinity of w_3 pixels of $P_E(i)$.

Moment-invariant edge detection

To locate the wear bottom point, a threshold independent method (Tae et al., 1999), based on moment invariance, is introduced.

A scan line across a step edge in the absence of noise is characterized by a set of numbers x_i's (i=0, 1, 2,…, n-1) that are either monotonically non-decreasing or non-

increasing. An ideal edge is a sequence of one brightness value h_1, followed by a sequence of another brightness value h_2, as shown in Figure 4.17 (a), where k denotes the edge location and n is the number of input data.

The first three moments of the input data can be calculated by:

$$\overline{m}_i = \frac{1}{n}\sum_{j=0}^{n-1}(x_j)^i \,,\; i = 1, 2, 3 \tag{4.28}$$

The solutions of the edge are given by:

$$h_1 = \overline{m}_1 - \sigma\sqrt{\frac{p_2}{p_1}} \tag{4.29}$$

$$h_2 = \overline{m}_1 + \sigma\sqrt{\frac{p_1}{p_2}} \tag{4.30}$$

$$p_1 = \frac{1}{2}\left[1 + s\sqrt{\frac{1}{4+s^2}}\right] \tag{4.31}$$

Where

$$s = \frac{\overline{m}_3 + 2\overline{m}_1^{\,3} - 3\overline{m}_1\overline{m}_2}{\sigma^3}, \quad \sigma = \sqrt{\overline{m}_2 - \overline{m}_1^{\,2}}, \quad p_1 + p_2 = 1 \tag{4.32}$$

Thus, the edge location is given with a sub-pixel precision by:

$$k = p_1 n \tag{4.33}$$

In this case, considering that the real data in either sequence of brightness value are not so uniform, for each scan line, the above edge detection method is repeated on gray-level data with increasing length from w_2 to w_3 (see Figure 4.17 (b)), each time

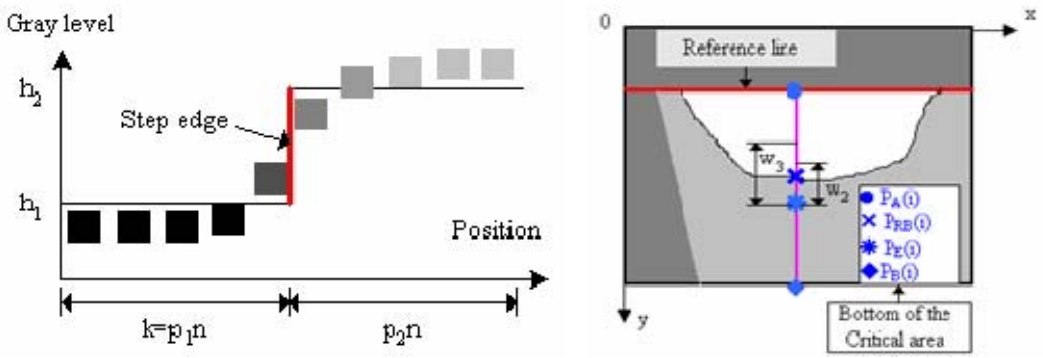
the input data starting from $P_E(i)$ with length increased by '1' pixel. For clarity, postfix form $h_1(n)$ is defined as h_1 calculated with data's length n, and the other parameters follow the same rule.

The distance K from $P_E(i)$ to $P_{RB}(i)$ is then given by:

$$K = k(j) = p_1(j)\,j, \tag{4.34}$$

where $j = \arg\max_n \{h_2(n) - h_1(n)\}$, $n = w_2, w_2+1, w_2+2,\ldots,w_3$.

Correspondingly, $P_{RB}(i)$ can be determined by offsetting $P_E(i)$ with the value of K.



(a) Moment-invariant method (b) Searching scope

Figure 4.17 Searching bottom edge of wear land

4.5.3 Parameters of the wear land

Once $P_{RB}(i)$ is obtained, the wear land on line $L(i)$ can be calculated, i.e., $W(i)$ is given by:

$$W(i) = |P_{RB}(i) - P_A(i)| \tag{4.35}$$

A_w of the wear land is given by:

$$A_w = \sum W(i) \tag{4.36}$$

Maximum wear height VB_{max}:

$$VB_{max} = \max\{W(i) \mid i=0, 1, 2, \ldots, N_L-1\} \tag{4.37}$$

Average wear height VB_{ave}:

$$VB_{ave} = A_w / N_L \tag{4.38}$$

where, N_L is the number of scan lines having wear.

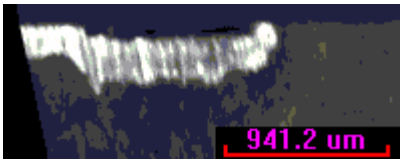
4.6 Breakage detection

Sometimes, when micro-chipping develops in size to breakage, the tool must be replaced. To detect breakage is very significant. In this section, the means by vision to inspect breakage are examined.

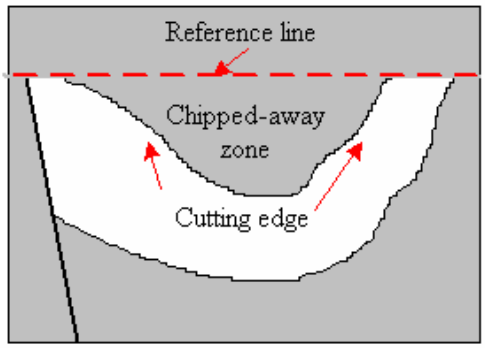
As mentioned in Chapter 2, breakage occurs as a block of material chipped away along the cutting edge. Figure 4.18 (a) shows a sample image of a broken tool. It can be observed that the chipped-away zone actually has same gray level as the background. With this information, the binary image alone can be used to detect this chipped-away zone, as shown in Figure 4.18 (b).

The processing is confined to the critical area. In this region of interest (ROI), the image can be binarized using the optimal threshold calculated with Otsu method. The small regions (isolated small blocks or noise) of the resultant binary image are firstly removed so that within the area circumvented by the reference line and the cutting edge there is no noise. After this, the chipped-away zone can be easily extracted since the location of the pixels on the cutting edge can be figured out in the binary image scanned from one pixel on the reference line downwards, the first white point is the pixel on the cutting edge. Again, orthogonal scanning is used and thus its detail description is omitted. Figure 4.19 shows the binary image of the ROI and the

extracted chipped-away zone for the sample as shown in Figure 4.18 (a). The upper part is the binary image of Figure 4.18 (a), and the lower part is the extracted breakage.



(a) A tool with breakage



(b) Breakage sketch

Figure 4.18 Breakage detection

Figure 4.19 Detected breakage.

4.7 Experimental results

In this chapter, the focus is on the tool wear measurement. Thus, experiments were carried out to measure tool wear rather than breakage. That is, the samples tested were worn-out tools, not broken tools.

Wear characteristics of three types of inserts, namely, A30N (Group A), AC325 (Group B) and ACZ350 (Group C) have been investigated. Table 4.1 shows ten image samples and the measurement results obtained. The results obtained by using the method proposed by Jeon and Kim (1988), which is a typical method based on only measuring the binary image, and the direct manual measurement with a tool-maker's microscope are also listed for comparison. When measuring V_2, the threshold for each

image was set to be the same at 150 (since the lighting condition was the same for all samples), by which the wear land was seen to be white ('1').

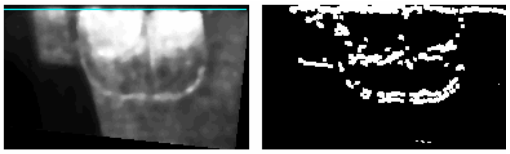
Figure 4.20 shows the ROI of 10 samples, in each of which the left part includes $S(x,y)$ superposed by the reference line (drawn in line). Below $S(x,y)$ is shown the wear measurement parameters. The right image is the binary edge image $M(x,y)$.

According to Table 4.1, among the 10 samples, results of 8 samples (Samples 1, 2, 4, 5, 6, 8, 9, 10) had a relative error within ±5%, assuming that results obtained by manual operation with a microscope are considered the correct or reference values. These results indicate an error of about 1 pixel size. Samples 3 and 7 had a relative error greater than 5%, where the wear was at its initial stage.

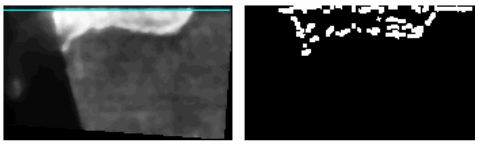
Table 4.1 Comparison of flank wear measurement results

Group	Sample No.	Wear VB_{max} (μm)		Error (μm) $\Delta_1 = V_1 - V_0$ ($\Delta_2 = V_2 - V_0$)	Δ_1 / V_0 (%) (Δ_2 / V_0)	Computing time** (s)	Comments
		V_1 by this method (V_2 *)	V_0 by microscope				
A (A30N)	1	816 (812)	821	-5 (-9)	-0.61 (-1.1)	1.822 (0.01***)	
	2	367 (212)	365	2 (-153)	0.55 (-41.92)	1.382	
	3	78 (24)	85	-7 (-61)	-8.24 (-71.76)	1.191	
	4	288 (259)	282	6 (-23)	2.13 (-8.16)	1.542	
	5	640 (600)	629	11 (-29)	1.75 (-4.61)	1.822	2
B (AC325)	6	441 (435)	426	15 (9)	3.52 (2.11)	1.533	
	7	180 (129)	167	13 (-38)	7.78 (-22.75)	1.472	1
	8	193 (94)	196	-3 (-102)	-1.53 (-52.04)	1.222	
	9	258 (94)	248	10 (-154)	4.03 (-62.1)	1.332	1
C (ACZ350)	10	236 (224)	230	6 (-6)	2.61 (-2.61)	1.462	

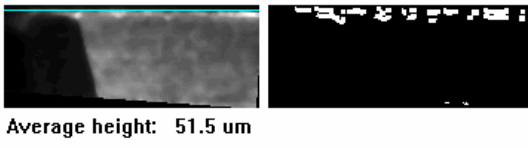
*, V_2 obtained by method of Jeon and Kim (1988), corresponding parameters listed in the brackets.
**, All time based on the same PC.
***, Same time for each image sample.
1, Reference line accuracy expected to be improved.
2, Critical area unexpected.



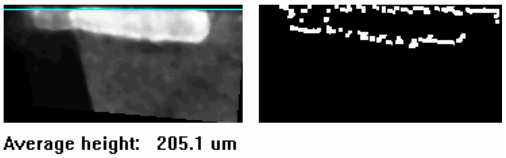
(a) Sample 1



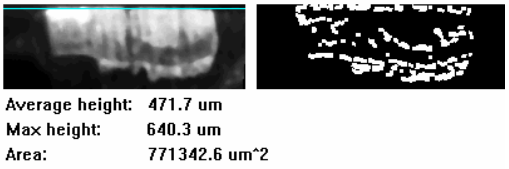
(b) Sample 2



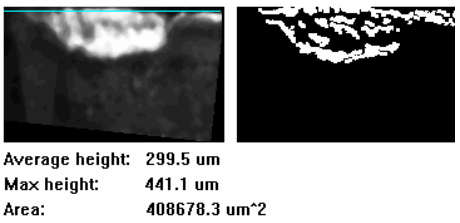
(c) Sample 3



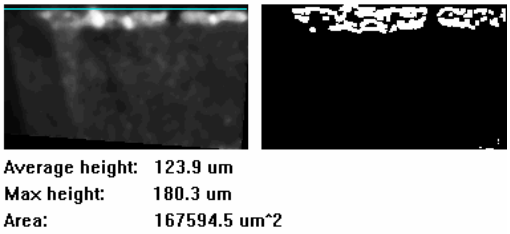
(d) Sample 4



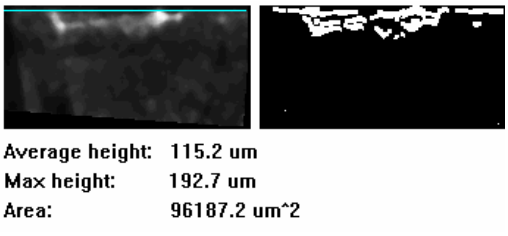
(e) Sample 5



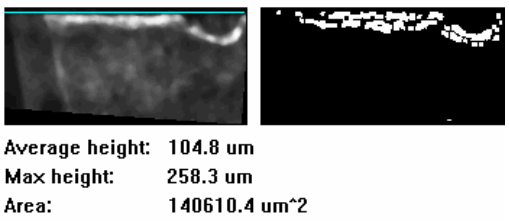
(f) Sample 6



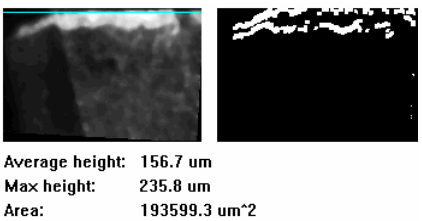
(g) Sample 7



(h) Sample 8



(i) Sample 9



(j) Sample 10

Figure 4.20 Flank wear measurement results

4.8 Discussion

The performance of this method depends partially on the accuracy of the reference line (See Samples 7, 9) when wear is negligible or little. Because the wear land at this stage is small, the edges obtained will affect the HT seriously. As a result, the reference line may not be calculated precisely.

Interestingly, however, a smaller processing area than expected seems to have no serious effect on the final results (See Sample 5). In binarizing image $S(x,y)$ into $B(x,y)$, if a threshold is given such that some pixels in unworn area are set to '1' rather than '0', given the pixels in wear land are set to '1' (which is often the case), the insert area obtained by segmentation will contain a whole set of pixels in the wear land. In other words, all candidate wear pixels are counted into the critical area. In this way, the result will not be affected.

In Table 4.2, some 'error's are big, for example, Sample 6 had an error 15 μm. It seems large. However, considering that when using microscope to measure the flank wear, human eyes make subjective decision on the wear land bottom boundary. Different people give different measured values through different times. It is, therefore, reasonable to take this error as tolerable.

Compared to the traditional thresholding procedure, this method with more procedure involved needs more time to compute as a necessary result of more operations than thresholding. Regardless of this, even using software, the computation time is within 2 s, which can be reduced with a more powerful PC. Given that the procedure is implemented by using a professional image processing hardware, the speed will be higher and hence satisfies the practical real-time need. In addition, this method has a better performance in terms of accuracy since it takes edge image and gray-level image into consideration, which makes use of as much available

information as possible. More importantly, the wear bottom edge is determined through a threshold independent method with sub-pixel accuracy, surpassing the traditional thresholding-based methods in this sense. With the same setting, lighting condition and algorithm parameters for three kinds of inserts, the results are satisfactory, which shows the system has certain robustness. Table 4.2 compares other methods (Kurada and Bradley, 1997a) and the proposed method.

Although the results show that this individual image processing method is successful in measuring flank wear and detecting breakage, the spindle stand is stationary as the image is captured. As mentioned earlier, it is more desirable to keep the spindle rotating when the measurement is made. To realize it, further development is needed. And this is presented in Chapter 5.

Table 4.2 Comparison of vision-based flank wear measurement systems

Methods	Hardware		Software		Results	Application area
	Light	Camera	Algorithm to get wear	Speed	Accuracy	
Lee et al. (1986)	Non-coherent	VIDICON (488 × 380)	Interactive segmentation	Low	+	Lab
Giusti et al. (1987)	Non-coherent (FW) Coherent (CW)	VIDICON	Thresholding by 10-pixel wide stripes	Med.	0.03 mm (FW) 0.03 mm (CW)	In-cycle
Jeon and Kim (1988)	Coherent (dia. 0.8mm)	VIDICON (0.001 mm)	Filter and fill binary image, contour detection	High	0.1 mm	Lab
Pedersen (1990)	Non-coherent	Philips LDH 0600 CCD (0.001 mm)	Threshold value from smoothed histogram	Med.	+	In-cycle*
Park and Ulsoy (1993b)	Non-coherent	TN 2500 CID (0.0015 mm)	Thresholding	Low	~5% (FW)	Lab
Du et al. (1993)	Non-coherent	CCD (0.0043 mm)	Template matching	Med.	+	In-cycle
Oguamanam (1994)	Non-coherent	CCD (RGB)	Thresholding to label three distinct regions	Med.	Confidence Level of 0.002	Lab
Proposed method	Non-coherent	CCD (0.012 mm)	Threshold independent, moment invariance	Low	≤0.015 mm (FW)	In-cycle

FW, Flank wear.
CW, Crater wear.
+, Not available.
*, An in-cycle sensor examines the tool periodically, for example, between machining blocks or during part changeovers.

Chapter 5
Successive image analysis

5.1 Problem statement

Chapter 4 presented a systematic method of measuring flank wear with sub-pixel accuracy for individual images. These still images were captured when the tool insert was mounted on a stationary milling holder. Actually, reported research studied only still images. However, to capture the image, the insert was either removed or retained in the tool holder. For the first case (Maeda et al., 1987b; Jeon and Kim, 1988; Oguamanam et al., 1994), the insert was placed on some measurement table where a still image of the insert was captured and measured. The mounting/dismounting of the insert was time-consuming and might cause run-out problem between the inserts. For the second case (Giusti et al., 1987; Pedersen, 1990; Park and Ulsoy, 1993b; Kurada and Bradley, 1997b; Pfeifer and Wiegers, 2000), although the mounting/dismounting of the insert was avoided, the image was captured when the spindle was stationary, so that a sharp image of the insert could be captured. This obviously does not meet the demand of unmanned manufacturing. A vision system that can capture and process the images while holder is rotating is necessarily required.

To develop such a system, there are some challenges:

- The mechanism to capture the image, i.e., in what manner the camera is controlled to capture the image;
- Since the object is moving, the ability of the camera to capture image with little blurring should be considered;

- There are in most cases several inserts mounted on the holder, and these need to be distinguished separately when processing their images;
- Since a series of images of a same tool insert can be obtained, their correlation can be appropriately used to speed up the processing and enhance the accuracy.

In the remaining text of this chapter, a system based on successive image analysis is presented for periodic measurement of flank wear in milling. The successive images are captured while the spindle is rotating. The trigger-capture hardware mechanism is described. With the image series, a method of processing image series to measure flank wear is presented instead of processing only individual still images, using the relationship between the successive images.

5.2 System configuration

5.2.1 Experimental setup

The experimental setup is shown in Figure 5.1. The same system described in Section 4.1 was used in addition:

- a laser trigger PD45VP6C200 with working range of 250 mm.
- four pieces of reflective tapes mounted on the spindle to serve as markers.

Compared to Figure 4.1 in Chapter 4, where the image was captured when the holder stayed stationary, the above two additional items were involved in order to capture the image as the spindle rotates.

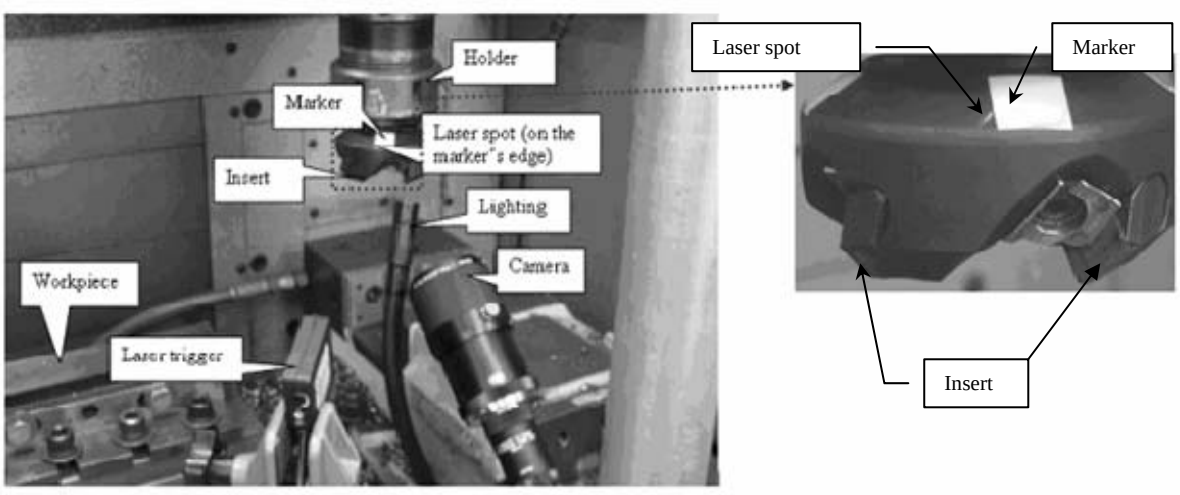


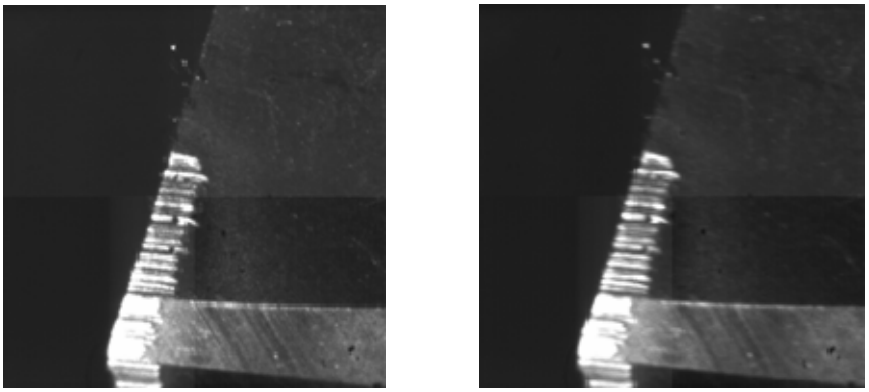
Figure 5.1 Experimental setup for successive image analysis

To get an image as sharp as possible for a moving object, theoretically, the integration time of the camera is expected to be shorter. Therefore, in this case, the integration time, T_I, of the camera working under trigger mode was set to 1.03 ms, which is the shortest time interval that can be set for the camera. The spindle speed, *s*, is also desired to be as low as possible to get the minimum blur, given by Eq. (5.1). Therefore, in this setup, *s* was set to 20 rpm, which is the minimum spindle speed for the Makino CNC machine used. The diameter of the cutter *D* is 50 mm. Accordingly, the blur of the captured moving images can be approximated by:

$$Blur = \pi \cdot D \cdot s \cdot T_I / 60 \ \mu m = 3.14 \times 50 \times 20 \times 1.03 / 60 \cong 54 \ \mu m \qquad (5.1)$$

Considering that the insert is moving in a circle and its movement trajectory with respect to the camera is not parallel to the image plane, the blur is slightly smaller than 54 μm, which is equivalent to 4-5 pixels. Figure 5.2 shows the still and moving image of the same insert. It can be observed that the blur does not seriously affect the overall gray-level distribution of the different areas of the insert, although some non-critical details in the moving image were slightly blurred, such as the undulation in the worn region. In other words, the moving image can be approximately viewed as a translated

still image, with negligible blur.



(a) Still image (b) Moving image

Figure 5.2 Still and moving images of the same insert

5.2.2 Experimental procedure

For four inserts mounted into four different holder slots, moving images of the four inserts can be captured. The experimental procedure is as follows:

(i). Calibration and referencing. A worn insert is mounted into the holder. Adjust the position of the camera (which has been calibrated earlier) and the holder so that a sharp image of the insert can be obtained. The current position of the tool is stored in the CNC machine controller as the *image capture position*.

(ii). Adjust the intensity and position of the lighting so that the image of the insert has good contrast, i.e., the whole image has distinct parts: background, unworn part of the insert, and worn part of the insert.

(iii). Dismount the worn insert, and mount all the four fresh inserts into different slots. For each insert, rotate the spindle slightly, so that the insert's possible worn part can be fully seen by the camera. Put the tapes on the holder almost evenly spaced to mark each insert. When the spindle rotates, by sensing the occurrence of the edge of the tape, a laser trigger sends a signal to the camera, which then captures an image of the insert. Each insert is marked by a tape. Therefore, in one revolution, the camera is

triggered by the laser trigger for four times to capture four moving images corresponding to the four different inserts.

(iv). Before machining commences, capture the images of the fresh inserts as the spindle rotates at a speed of 20 rpm.

(v). Using predetermined machining conditions, the CNC machine is programmed to cut the workpiece for one pass. It then moves the tool to the image capture position at a lower spindle speed of 20 rpm. A new set of images of the four inserts is captured.

Repeat the fifth step to capture a set of four images of the inserts after each machining pass. By processing each image set, the flank wear of each insert is obtained after each pass.

Note that in steps (i) and (ii), a worn insert is used in setting up the positions of the camera and lighting for suitable orientation and intensity of the lighting. This setup is to ensure that a good image of the insert is captured for successful measurement of flank wear. The worn insert serves as a sample to help determine the setup. In the subsequent image capture, the setup is not changed, and steps (i) and (ii) are no longer needed.

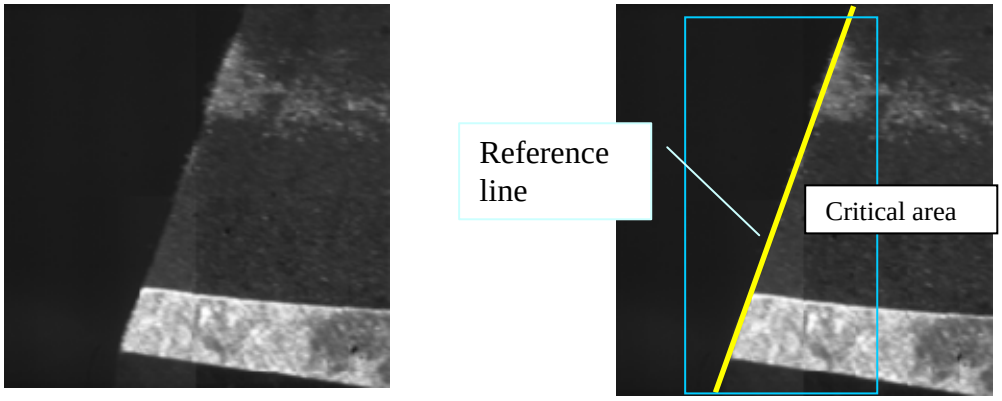
5.3 Reference image processing

In this configuration, the blur (Eq. (5.1)) is minimal and the captured moving image is almost as sharp as the still image as shown earlier (Figure 5.2). So the deblurring operation is omitted.

The ongoing method in this chapter is based on the static image method described in Chapter 4, where the images are not moving but static. However, the method has been modified to take advantage of correlation features among successive images, which provide important clues for measuring the flank wear. For convenience, images of all

inserts after each pass are called as an *image set*, two images of the same insert from two successive passes as an *image pair*. The images related to the fresh inserts are indicated as 'reference' image, while those related to the gradually worn inserts are indicated as 'worn insert' image.

Images in the reference image set are captured when the inserts are fresh. Figure 5.3 (a) shows an image of a fresh tool insert.



(a) Image of a fresh tool insert

(b) Two features extracted

Figure 5.3 Image processing for a reference image

Two features can be extracted from this image set, which provide important information to process the subsequent worn insert image set to obtain the flank wear continually. As shown in figure 5.3 (b), the two features are:

- The critical area (outlined by solid rectangle)
- The reference line (cutting edge)

5.3.1 Critical area redefined dynamically

Examining the moving image in Figure 5.3 (b), it can be observed that the cutter area is a large portion of the entire image. According to the definition of the critical area in Chapter 4, its scope covers this large portion, see the dashed rectangle in Figure 5.3 (b). As a result, the image is processed over this large region that increases computation time. Hence the critical area is redefined.

Because of the constant relative position between the camera and the insert, the insert position in the image is relatively fixed. Given the depth of cut, the contact length between the major cutting edge and the workpiece can be roughly estimated. The flank wear occurs along this length. Consider Figure 5.2 (b) as an example, the upper border of the critical area can be fixed. The bottom border of the critical area can be obtained by tailoring the cutter area if necessary since the wear in the major cutting edge is of interest.

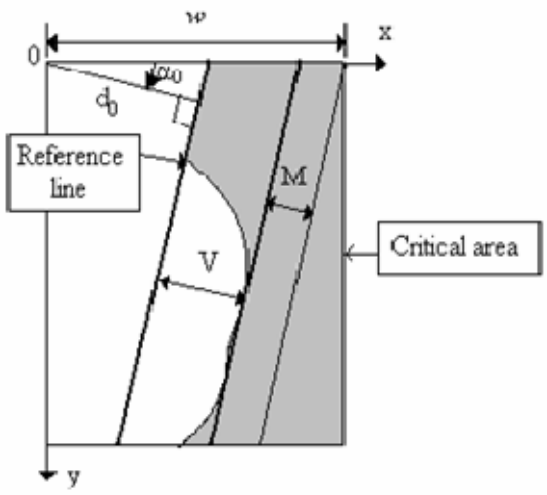


Figure 5.4 Determination of the right border of the critical area

In addition, suppose that the maximum width of the wear land is *V*, the right border of the critical area can be calculated, as shown in Figure 5.4.

$$w = (d_0+V+M)/\cos\alpha_0 \tag{5.2}$$

where,

w - the width of the critical area;

d_0 - the distance from the origin to the reference line;

α_0 - the angle from the x-axis to the normal to the reference line;

V - the maximum width of the wear land;

M - a margin set for the critical area.

Obviously, *V* is a time-varying parameter. In particular, *V*=0 when the insert is fresh. After each pass, *V* is increased by some amount as wear progresses. Hence a dynamic critical area can be used. The new critical area when the tool insert is fresh is shown as the solid rectangle in Figure 5.3 (b).

5.3.2 Reference line

The reference line is obtained with Hough transform in the entire image. The upper and bottom borders are defined before Hough transform, while the right border of the critical area is defined after Hough transform. The procedure to determine the critical area and the reference line is shown in Figure 5.5. Detailed information can be found in Chapter 4 and thus is omitted here.

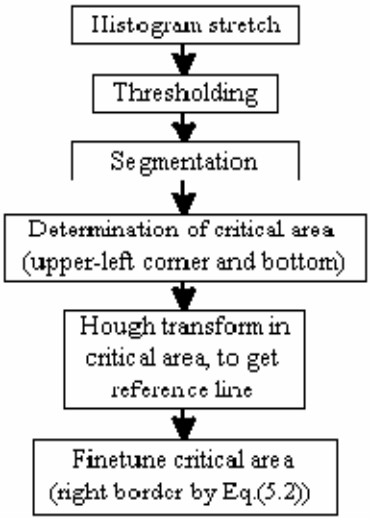


Figure 5.5 Processing blocks for the reference image

5.4 Worn image processing

The procedure is somewhat similar to that static image method described in Chapter 4. For the moving image case, however, unique steps have been developed and described below.

5.4.1 Index and order inserts

The objective of this step is to index and order each insert after each pass with its image so that a correct image pair can be determined, in the case of more than one insert mounted on the holder. This is required since in the worn insert image sets, the capture may not be in the same order as in the reference set. Take four inserts for example. In the reference set, the captured images start with that of insert 1. Thus four ordered images are obtained and their insert indices are 1, 2, 3, and 4, respectively; that is, P0I1 for insert 1, P0I2 for insert 2, P0I3 for insert 3 and P0I4 for insert 4. While in the subsequent worn insert set (e.g., *x*th), the captured images can start from any one of the four insert images as the first image captured can be any one of the rotating inserts, i.e., not necessarily that of insert 1. Thus, the images PxI1, PxI2, PxI3 and PxI4 may have any one of the following insert indices: {1, 2, 3, 4}, {2, 3, 4, 1}, {3, 4, 1, 2}, or {4, 1, 2, 3}. Obviously, the image PxI1 and P0I1 are for different insert in the latter three cases. Accordingly, it is necessary to identify the insert with its captured image.

For the matching, a cross-correlation coefficient (*CC*) is used. This is a classical representation of the area-based methods in matching corresponding image pairs in image register technique (Zitová and Flusser, 2003), and is defined by:

$$CC_W = \frac{\sum_W (R - A(R))(U - A(U))}{\sqrt{\sum_W (R - A(R))^2 \sum_W (U - A(U))^2}} \tag{5.3}$$

where *W* refers to the same window for the reference image and worn insert image, *R* is the reference image, *U* is the worn insert image, *A*(*R*) and *A*(*U*) are their average gray levels in *W*.

CC of the first image instead of each in the worn set with the reference set is calculated. Once the first image's index is determined, the others are subsequently

uniquely determined, since the index set has only four unique cases, each of which begins with a unique index.

The configuration has been set such that each insert appears at a different position in the captured image, as shown in Figure 5.6 (Each captured image has only one insert. In Figure 5.6, all four insert images are put together to show they are at different locations). The trigger-capture mechanism ensures the same position of the same insert in successive images. Thus it is reasonable to select a small-size window (in dashed line) to calculate *CC*. The accuracy of image pair is practically 100% guaranteed. Once the correct image pair is obtained, the reference lines in the worn insert image set are made available by taking those obtained in processing the reference image set.

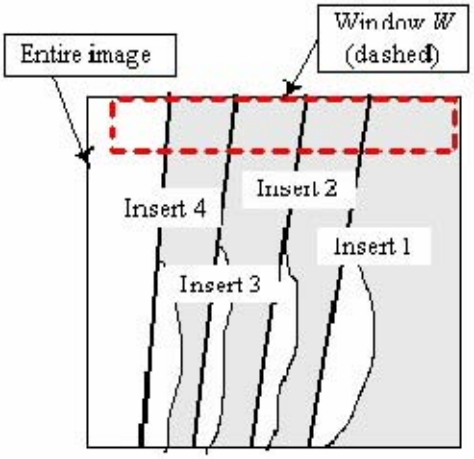


Figure 5.6 Four inserts put together with window to match image pairs

5.4.2 Parallel scanning

In Chapter 4, orthogonal scanning method is used to determine the wear, where the scanning is operated in the direction perpendicular to the reference line. Here, for greater robustness against noise outside the flank wear land, parallel scanning is applied on the binary image $B_E^+(x,y)$, which is the resultant image after binarizing the edge image following contrast enhancement.

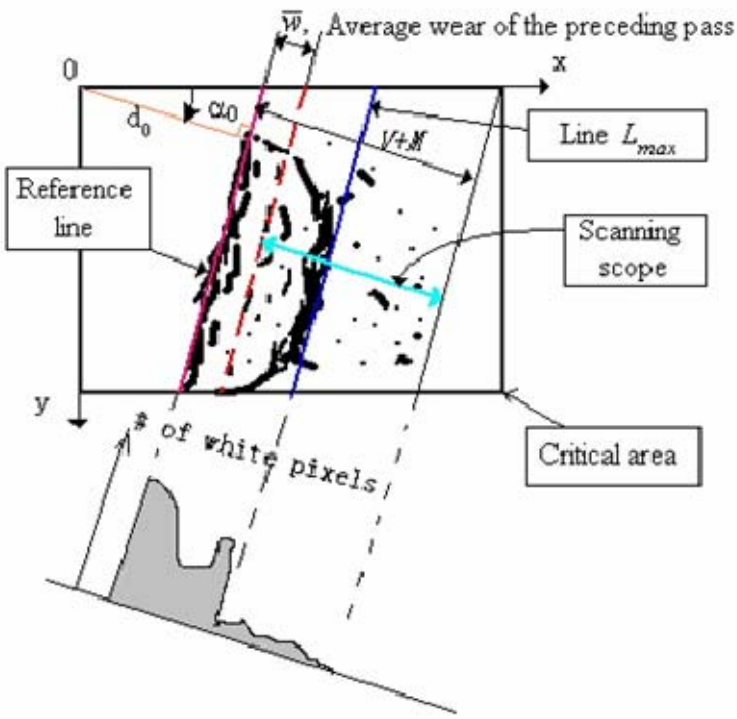


Figure 5.7 Parallel scanning scheme

The parallel scanning is operated parallel to the reference line, as shown in Figure 5.7. It proceeds as follows:

(i). Scan along each parallel line with distance i to the reference line, and get the number $N(i)$ of white pixels. The scanning is confined to the critical area, i.e., i = 0, 1, 2, …, V+M-1.

(ii). Find the maximum value N_{max}, of number set N, and the related index I.

$$N_{max} = \max\{N(i)\},$$

$$I = \{i | N_{max} = N(i)\} \tag{5.4}$$

(iii). Let $j = I, I+1, I+2, \ldots, V+M-1$, if

$$N(j) < r_w \cdot N_{max} \tag{5.5}$$

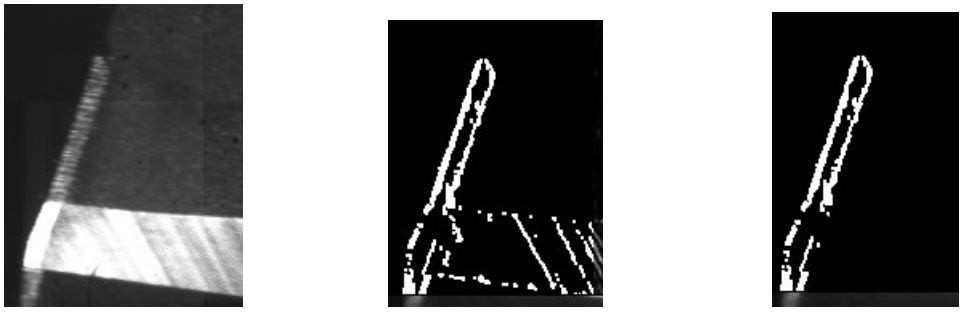
go to (iv).

(iv). Let $D_{max} = j+d_0$. Turn all white points outside line $L_{max} = (D_{max}, \alpha_0)$ black, and terminate.

where r_w is a percentage ratio, indicating the correlation between a line within the flank wear land and L_{max} in terms of the white-pixel number $N(i)$.

Note that since the average wear of the preceding pass is known (denoted as $\overline{w}$), in step (i), it does not have to scan right from $i = 0$. Instead, scanning from $i = \overline{w}$ is reasonable. Because the aim is to remove noise outside the boundary of the wear land, the priority is given to scanning closer to the boundary.

Figure 5.8 shows the original gray image and images before and after parallel scanning. The necessity (from (b)) and effectiveness (from (c)) of parallel scanning can be observed.



(a) (b) (c)

(a) Original gray image, (b) Binary edge image before parallel scanning, (c) Binary edge image after parallel scanning.

Figure 5.8 Parallel scanning in practice

5.4.3 Wear measurement

The whole procedure to get the flank wear is illustrated in Figure 5.9. Detailed information can be found in Chapter 4 and thus is omitted here.

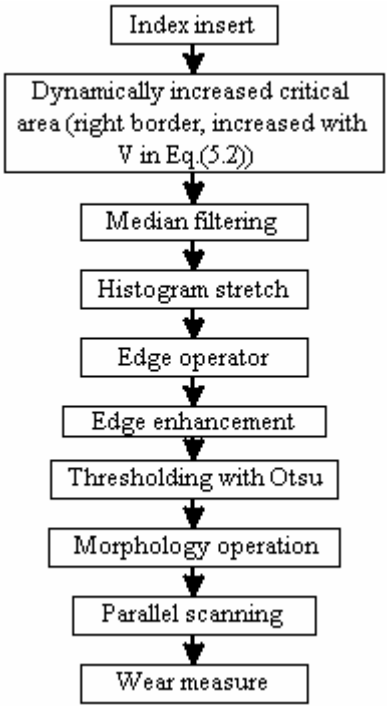


Figure 5.9 Procedure to measure flank wear

5.5 Experimental results

Five tests were done under various cutting conditions as shown in Table 5.1.

After each machining pass, an image set was captured. For the image set, taking Test 1 in which four inserts were mounted on the holder as an example, the maximum flank wear of each insert was measured. The four maximum values were then averaged to get the flank wear for this pass. In this way, the flank wear values of 53 passes were computed and plotted in Figure 5.10 (a). Also in Figure 5.10 (a), the manual measurement is plotted for comparison, which was made after 3 passes, i.e., after pass 3, 6, 9, …, etc. Figure 5.10 (b)-(e) shows the results for Tests 2-5.

Note that 'Automatic, individual images' are based on the individual image processing method. 'Automatic, correlated images' refer to the results obtained using the successive image analysis method. 'Manual measurement' refers to the results obtained by manual measurement of the flank wear land, which is subjectively

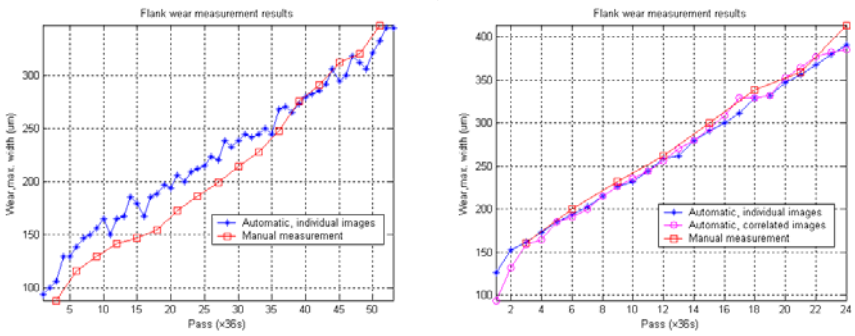
determined by the naked eye from the captured successive images. 'Measurement with microscope' refers to the results obtained by taking out the inserts from the holder and measuring their wear with a tool-maker's microscope.

In Figure 5.10 (a), no results with 'Automatic, correlated images' are shown, because in Test 1, the markers (pieces of clay rather than reflective tape) had position shifts after machining (the markers will be discussed separately later in Section 5.6.2) and the same insert appeared in different positions in its image pair. Figure 5.10 (a) is shown to verify the effectiveness of the individual image processing method. In Figure 5.10 (b)-(d), three results are plotted to further verify the effectiveness of the successive image analysis method. Figure 5.10 (e) shows the results obtained using successive images and a tool-maker's microscope to verify that the successive image analysis method gives relatively good accuracy.

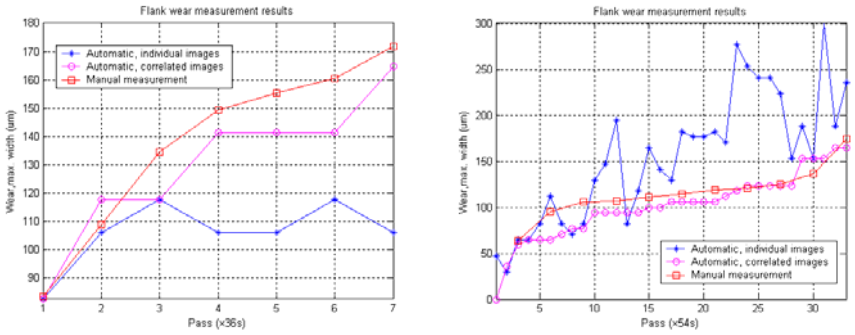
Table 5.1 Parameters in dry machining for successive image analysis

Parameters	Test 1	Test 2	Test 3	Test 4	Test 5
Type of inserts (ISO SDKN42MT)	A30N, uncoated		AC325, coated		A30N, uncoated
Number of inserts mounted	4	4	1	2	2
Workpiece	ASSAB718HH steel				
Length of the workpiece (mm)	205				
Diameter of cutter (mm)	50				
Spindle speed (rpm) (m/min)	800 125.6	1000 157	1000 157	1000 157	1000/800* 157/125.6
Feed rate (mm/min)	300	300	300	200	100
Feed per tooth (mm/tooth)	0.094	0.075	0.3	0.1	0.05
Depth of cut (mm)	1	2	2	1	2
Time/pass (s)	36	36	36	54	108
Immersion rate	Full				
Total passes observed	53	24	10	37	10
Conditions observed	Gradual wear		Breakage from pass 8	Fracture at pass 34	Worn out, fracture at pass 10
Manual measure interval	3	3	1	3	3**

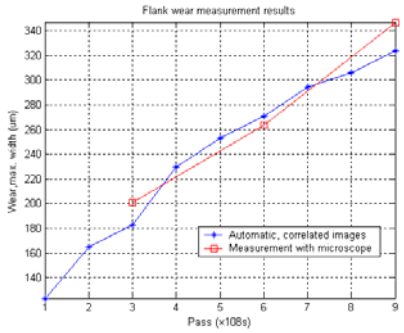
*, The speed was reduced after pass 6 due to the rapid tool wear progress.
**, Only this test has been measured with a tool-maker's microscope.



(a) Test 1 (b) Test 2



(c) Test 3 (d) Test 4



(e) Test 5

Figure 5.10 Flank wear measurement against pass (time)

5.6 Discussion

5.6.1 Results

Figure 5.10 (a) shows that the flank wear obtained by the individual image processing method was quite good in terms of accuracy. The method gave a result that

had small deviation from the manual measurement. In Figure 5.10 (a), even the maximum deviation which occurred after pass 14 was only about 50 μm. Comparing the two results throughout the whole test (53 passes), it was found that although the individual image processing method gave a result greater or smaller than the manual measurement, the overall average deviation was small. This finding again reinforces the conclusion that the individual image processing method is effective and accurate as shown in Chapter 4. Furthermore, it was observed that when the flank wear became severe (with a greater value), the deviation was smaller (see the results from passes 35-53). This result is favorable since it is preferable to be able to monitor the flank wear accurately as the tool is at the worn stage.

The results measured by the individual image processing method tracked the manual measurement well in Figure 5.10 (a) and (b). However, Figure 5.10 (c) and (d) show that the results obtained using the individual image processing method did not match the manual measurement in the same way. An important reason is that Tests 3 and 4 shown in Figure 5.10 (c) and (d) used AC325 inserts, which are coated with small particles to enhance the performance of the insert. For this kind of insert, its surface is not as smooth and uniform as that of the uncoated insert A30N. When the coated insert is used to cut workpiece after a period of time, the dirt will fill up the space between particles. Also some of the coated particles will be removed with the chips from the surface of the insert. Furthermore, due to the temperature of the machining, the color of the particles will fade away to different extents. These factors will collectively make the surface of the insert ‘noisy’ in the sense that the image of the insert has no distinct three areas which are supposed to have separable gray levels. This means that, the gray level of the unworn area of the insert is not evenly distributed as assumed in the intermediate level, which results in more noise in the binary image. Consequently,

some wear value measured by the individual image processing method may be greater than the actual. Similarly, the gray level of the wear land does not exhibit a uniform distribution at the bright level. Some pixels in the wear land, particularly those along the wear land boundary are slightly darker than or even approaching the gray level of the unworn area. Thus, the measured wear value will be smaller than the actual.

Although the results shown for coated inserts were not so accurate, the accurate results for uncoated inserts indicate that for a specific image of a coated insert, it is likely to achieve a better result by deliberately fine-tuning the parameters in the individual image processing method. Although fine-tuning the parameters needs more trials, the accurate result implies that this method can be used as the basis for successive image analysis.

Figure 5.10 (a), (b), (c) and (d) show that the flank wear obtained by the individual image processing method had some fluctuation. It did not always increase. On the contrary, for some two successive passes, say, pass k and pass k+1, the wear after pass k+1 can be less than pass k. One noticeable example is the measured results after pass 12 and 13 as shown in Figure 5.10 (d), where the wear after pass 13 was smaller than pass 14 by more than 100 μm. This, definitely, is not logical since as machining proceeds, the wear should become greater in principle. Although it can remain unchanged in calculation when the wear increases so slightly that the increment is undetectable for the camera with an insufficiently high resolution, the wear should by no means decrease. Two main reasons may account for this observation. One reason is that small chips may or may not adhere to the surface of the insert after each pass (In Figure 5.2, for example, small chips exist in the unworn area). If the small chips stay around the boundary of the wear land, since chips look bright to the camera, the method will regard chips as wear, and the flank wear result will be greater. Another

reason is that, after one pass of machining, the surface of the insert, especially near the wear land boundary, will change the gray level. Sometimes, the boundary after pass k+1 looks shrunken, compared to that after pass k. Thus, the wear after pass k+1 is less than that after pass k. For example, in Figure 5.11, the surface of the insert looks different in successive two images, especially for the surface around the wear land boundary.

Results obtained with the successive image analysis method are shown in Figure 5.10 (b)-(e). Figure 5.10 (b) strongly demonstrates the good accuracy of this method. Throughout the whole test, the wear using successive image analysis was nearly the same as the manual measurement. One significant improvement of the successive image analysis method over the individual image processing method lies in accuracy. This can be seen in Figure 5.10 (d), where the results tracked the manual measurement very well. This can be explained by two factors. One factor is that a more precise binary edge image was obtained thanks to parallel scanning that can remove noise outside the wear land. The other factor is that a more precise critical area was obtained due to the knowledge of the neighboring (previous) wear values. Another notable improvement of this method over the individual image processing method lies in the logical increment of wear values as machining progresses. This can be seen in any one of Figure 5.10 (b)-(e), where the wear values kept increasing or remained unchanged as time went by. This was achieved because the critical area was expanded dynamically pass-by-pass.

(a) Image after pass 3 (b) Image after pass 4

Figure 5.11 Test 3, images after pass 3 and pass 4

Furthermore, Figure 5.10 (e) shows the good agreement between the results obtained using the successive image analysis method and a tool-maker's microscope. It also shows that after the insert was measured with the microscope, the remounting of the insert onto the holder affected the wear measurement. For example, the automatic results 1-3, 4-6, 7-9 were of different patterns (greater or smaller than) around the microscope results. This is a natural consequence of the remounting since there is no way to mount the insert back to the same position where it was removed from the holder for microscope measurement. Accordingly, there will be geometric shifts for the reference line and the critical area in the entire image. Therefore, depending on the shift orientation, the automatic result is greater or smaller than the microscope measurement. This observation indicates that removal and remounting of the insert from and back to the holder for microscope measurement can alter the wear change pattern. In this aspect, the successive image analysis method is better than the microscope measurement method in the sense that the former does not affect the wear rate since the insert always stays on the holder.

It is also worth noting that the successive image analysis method gave a result with a greater deviation at the early wear stage; for example, the deviation reached up to 35 μm after pass 6 in Figure 5.10 (d). This can be attributed to the use of a median filter, which has both positive and negative effects. When wear is small, after median

filtering, the inside and boundary pixels of the wear land will be partially immersed in the unworn area in terms of gray level. In other words, the wear land is shrunken when segmented from the unworn area. The measured wear is more likely to be less than the actual. However, as wear grows, the wear land is more distinguished and hence, the camera can capture a more uniform bright wear land. Thus, the measured result is better. But median filter cannot be simply discarded to reduce the deviation at the early wear stage. If median filtering is not applied, too much noise remains in the gray-level and binary edge images, which may result in a greater deviation. Moreover, greater deviation occurring at the early stage is tolerable in practice since it is more important to get a more precise wear value as the tool tends to be worn out at the later stage rather than the early wear stage. Therefore, it is preferable to employ the median filter as the positive effect surpasses its negative effect.

The wear shown in Figure 5.10 (d) developed in three stages, i.e., initially stage (wear increases sharply) from pass 1 to 3, linear stage (wear increases slowly) from pass 4 to 28, and accelerated stage (wear increases abruptly till the tool is worn out) from pass 29. This observation confirms the theory of three stages for progressive wear, as mentioned in Chapter 2. However, the wear development curves shown in Figure 5.10 (a) and (b) have no evident linear stage. This is probably because the tool has no coated material. Under some cutting conditions, flank wear progresses nearly linearly from a fresh tool to a worn-out or broken tool. The absence of the linear stage was also observed in Figure 5.10 (c) and (e). For these two cases, the tool failed (with breakage or fracture) very soon. Therefore, it may be argued that under some circumstances, the tool does not necessarily wear out in the typical three stages.

5.6.2 Experimental setup

Normal lighting or strobe lighting

In this experiment, normal lighting was projected on the cutter when capturing images. Because the spindle was rotating during capture, to reduce blurring, integration time of the camera was set very short (recall the time is only 1.03 ms) and the spindle speed was set low (only 20 rpm). The normal lighting was thus set to be very intensive to provide adequate exposure of the insert surface for the camera. This configuration is acceptable, since the speed is slowed down only when there is no machining. As the image capture was quite fast, there was little negative effect on the actual machining.

Strobe lighting is a possible alternative. Its remarkable advantage is that the spindle speed may remain unchanged when capturing images, so that images can be captured faster and even on the fly. But the problems are: 1) If the spindle speed is high, say, 800 rpm, which is 40 times the set speed presently. For the experimental setup presented in this thesis, to maintain the acceptable or ignorable blurring, strobe flash duration should be less than 1.03/40 ms, or about 25 μs. Correspondingly, the lighting intensity should be 40 times the normal lighting, which now has been quite intensive (the intensity looks higher than a 150W-bulb). 40 times this lighting will cost the strobe circuit board to be much higher, especially with a quite high strobe flash rate. 2) The timing sequences of strobe fire signal, image capture signal, and laser trigger signal should be controlled properly on order of microsecond. Extra control circuit is required, which again costs more. In other words, even with strobe lighting, the spindle speed when capturing images should be slowed down to make the system cheaper and simpler.

Space for the system

This system needs a little more space. The frame holding the camera should provide enough stabilization and adjustment range for the camera. In addition, since the camera

should be within the machining volume of the CNC machine (because the holder cannot go outside of the machine door), the frame hampers the door from being closed. Meanwhile, the camera's wire connection is the factor which prevents closing the door. As such, all of these: camera, frame, laser trigger and its fixture, plus lighting set need space. If the system is to be used in real industry, these components should be integrated and made more compact for use on the CNC machine.

Markers to index different insert

In this experiment, four pieces of clay were initially attempted as markers. A big problem occurred after some passes of machining: the clay became softer and lost its stickiness. This resulted from the increasing temperature after a period of machining and high-intensity lighting projected on the holder. Another accompanying phenomenon was that the shape of the clay was distorted. Two factors contributed to this occurrence: 1) The centrifugal effect during machining at a spindle speed that cannot be ignored. 2) Relatively high-frequency impact during entry and exit of the inserts through the workpiece.

Consequently, the marker for triggering the camera to capture its corresponding insert may not stay at the same place after each pass, or even be thrown off the holder. As a result, the same insert appears in different positions in the image pair. In this case, the reference line, critical area obtained in the reference set cannot be directly used in the worn set and have to be recalculated. Result for Test 1 is one such case. Fortunately, the experiment shows that the reflective tape works well as a marker. Tests 2-5 were carried on with reflective tapes as the markers. In other words, to be used in real industry, the markers should be at least immune to high temperature, distortion and displacement.

Others

As observed in machining, small chips may remain on the surface of the cutter when capturing images, which will affect the measurement result. Good chip removal will be most helpful in practice. For the lighting and the laser trigger, keeping their positions fixed with respect to the holder is necessary for this method. Because the integration time for the camera is short, high-intensity lighting is needed in order for the insert surface to be exposed sufficiently. Thus, it is necessary to consider cooling, not only the lighting system, but also the camera system (especially working with long continuous time). In addition, it is a necessity to protect the lens of the camera from the chips produced during the machining.

So far, the optical method has been presented to measure flank wear. Actually, crater wear is also an important wear mechanism, especially in high-speed machining. Therefore, in the next chapter, crater wear measurement based on phase-shifting will be resolved.

Chapter 6
Crater wear measurement by phase-shifting method

6.1 Problem statement

Flank wear and crater wear are the two main wear mechanisms in machining. They are primarily used in determining the tool conditions for the study of tool wear characteristics. In recent decades, researchers have been conducting research on TCM using image processing techniques to measure flank and crater wear (Giusti et al., 1987; Jeon and Kim, 1988; Pedersen, 1990; Oguamanam et al., 1994; Yang and Kwon, 1996; Karthik et al., 1997). It has been demonstrated that image processing techniques can be effective in measuring flank wear and crater wear. When the cutting speed is high, it is more likely for crater wear to occur. Recent trends are toward high-speed machining and thus, there is increasing need to be able to measure crater wear, which is hence the focus of the research described in this chapter.

To measure crater wear, a 3-D image of the crater wear is needed. As reviewed in Chapter 2, several researchers (Giusti et al., 1987; Yang and Kwon, 1996; Karthik et al., 1997) have developed and used various methods to obtain the 3-D profile of the crater wear. But those methods are not satisfactory in terms of processing speed, system complexity, robustness and accuracy. Therefore, in this chapter, a new method of non-contact tool wear measurement is presented. It employs a phase-shifting method using fringe patterns to measure crater wear by constructing a 3-D map of the tool insert. Four fringe patterns with various phase shifts are projected onto the rake face of the tool upon which four gray-level images are captured. By solving and then unwrapping the phase map with the four images, the 3-D profile of the tool is obtained

to provide the overall size of the crater wear land. Sample tool inserts under different machining conditions have been tested and the results demonstrate that the system is robust to ambient light and is accurate and applicable for use in real industrial environment to measure crater wear. The full-field method has high precision, permitting interpolation to 1/1000 of a fringe period and low sensitivity to background, contrast variations and noise (Halioua and Liu, 1989). It does not require precise orientation of the incident ray (lighting) and the camera and can construct the 3-D map of the tool accurately and automatically.

6.2 Principle of phase-shifting method

Halioua and Liu (1989) have developed a novel 3-D sensing approach known as phase measuring profilometry (PMP), based on the use of sinusoidal grating projection and digital phase-shifting techniques.

With reference to Figure 6.1, an incident ray *P* strikes the object and the reference plane at points *D* and *A* respectively. When viewed from an offset direction at height *h*, *A* is displaced to a new position *C*. It is the displacement $\Delta x = \overline{AC}$ which carries information about the object height $Z = h(x,y)$ from the reference plane, or the range $L - h(x,y)$ from the view point.

6.2.1 Phase measuring profilometry (PMP) model

When a sinusoidal grating pattern is projected onto a 3-D diffuse object, the mathematical representation of the deformed grating image may be expressed in the general form:

$$I(x,y)=R(x,y)[G(x,y)+H(x,y)\cos\phi(x,y)] \tag{6.1}$$

where, $R(x,y)$ is the object reflectivity characterizing the surface nature, $G(x,y)$ represents the background intensity and $H(x,y)/G(x,y)$ the fringe contrast, and $\phi(x,y)$ is the phase function, which characterizes the fringe deformation and is related to the object shape $Z = h(x,y)$. The surface shape is thus converted into a phase distribution.

Normally, in practice, phase function $\phi(x,y)$ can be obtained easily. For example, four phase-shifted sinusoidal fringe patterns are projected onto an object surface with phase shifts of 0, $\pi/2$, π, and $3\pi/2$ within one period of the fringe. Following Eq. (6.1), one can derive four equations (Eq. 6.2-6.5) with respect to these phase shifts:

$$I_1 = R\,[G + H\cos\phi] \tag{6.2}$$

$$I_2 = R\,[G + H\cos(\phi+\pi/2)] \tag{6.3}$$

$$I_3 = R\,[G + H\cos(\phi+\pi)] \tag{6.4}$$

$$I_4 = R[G + H\cos(\phi+3\pi/2)] \tag{6.5}$$

By solving them, the following phase function can be expressed as:

$$\phi=\arctan[(I_4-I_2)/(I_1-I_3)] \tag{6.6}$$

The phase value computed according to Eq. (6.6) ranges from $-\pi$ to π. Thus the phase distribution is wrapped into this range and, consequently, has discontinuities with 2π-phase jumps. These discontinuities can be corrected easily by adding or subtracting 2π by checking the phase jump value. The resultant phase map is called an unwrapped map.

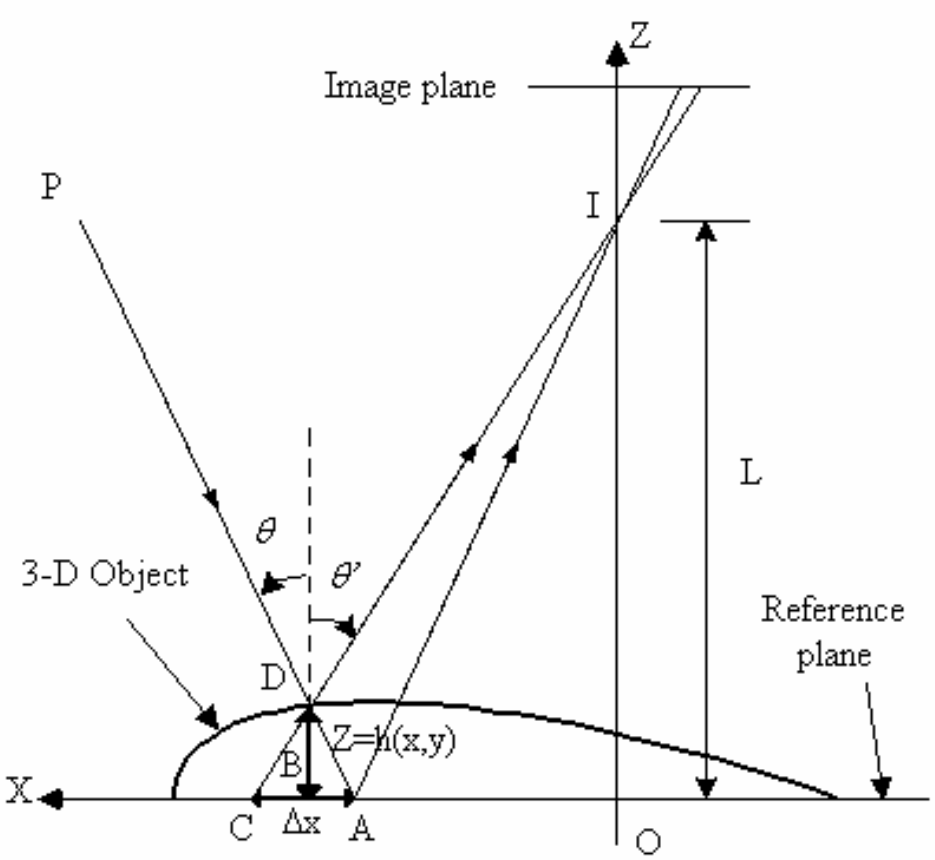


Figure 6.1 Optical geometry for fringe projection

6.2.2 Relation between phase function and shape

The projected sinusoidal pattern as seen on the reference plane has a period p_0 and the intensity produced at a point C is given by:

$$I_C = G + H\cos(2\pi\overline{OC}/p_0) \tag{6.7}$$

Consider r_i as one of the detectors in an array of a CCD camera and is used to measure the intensity at C on the reference plane and at D on the object. The intensity observed at D on the object is the same as that observed at A if projected on the reference plane (without the object), modified by the object reflectivity R; thus:

$$I_D = R[G + H\cos(2\pi\overline{OA}/p_0)] \tag{6.8}$$

Using $I_D=RI_C$, and Eqs. (6.7) and (6.8), the phase difference ϕ_{CD} in phase values at points C and D, observed by the same detector r_i, is related to the geometric distance $\overline{AC}$:

$$\overline{AC} = (p_0/2\pi)\phi_{CD} \tag{6.9}$$

In turn $\overline{AC}$ is related to the object height $h = \overline{BD}$ by:

$$h = \overline{AC}/(\tan\theta + \tan\theta') = (p_0/2\pi)\phi_{CD}/(\tan\theta+\tan\theta') = K\phi_{CD} \tag{6.10}$$

where θ and θ' represent the respective directions of illumination and observation as shown in Figure 6.1.

6.3 Experimental setup

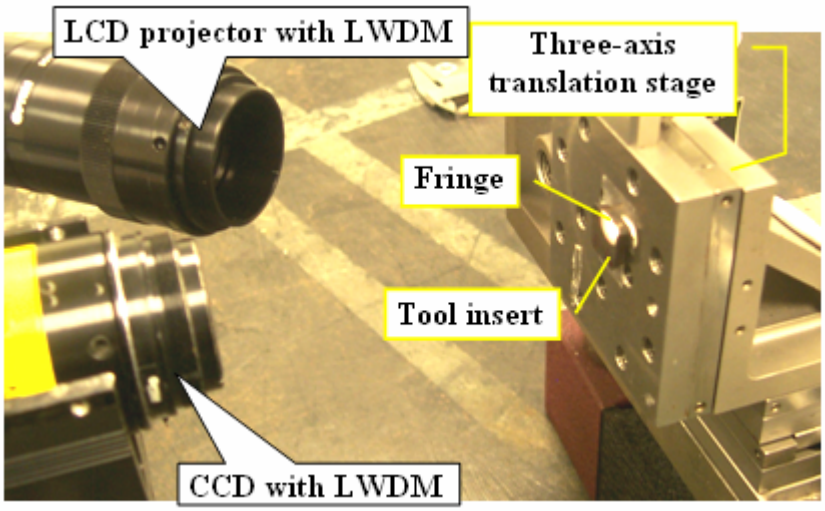


Figure 6.2 Experimental setup for 3-D crater wear measurement

As shown in Figure 6.2, the experimental system consists of an LCD projector mounted with a long working distance microscope (LWDM), an LCD controller, a CCD camera mounted with another LWDM, a three-axis translation stage, and a frame

grabber. The tool insert is put on the stage. Four linear sinusoidal fringe patterns with phase shift values of 0, $\pi/2$, π, and $3\pi/2$ are produced by the LCD controller and projected onto the rake face of the tool insert by the LCD projector through its LWDM. The contouring fringe patterns are then captured by the CCD camera through its LWDM and stored via the frame grabber. By processing the 256-graylevel images of fringe patterns, a 3-D map of the crater wear region can be obtained.

6.4 Experimental results

6.4.1 System calibration

With reference to Eq. (6.10), the system can be calibrated to convert a phase value to a physical dimension, say, in microns by determining a suitable constant factor K using the following simple method:

- When the stage stays at some initial place, capture four phase-shifting images of the rake face, and solve out the phase function according to Eq. (6.6) and unwrapped it to a phase map (ϕ_1);
- Move the stage towards or away from the camera by Δh (50 μm in this case), and capture another set of four phase shifting images of the rake face. Again solve out the phase function and get the phase map (ϕ_2);
- Trim two parts of the two phase functions from the same region, and compute the mean of the phase difference, $M_{\Delta\phi}$. Then, the factor $K=\Delta h / M_{\Delta\phi}$.

After calibration, either ϕ_1 or ϕ_2 is used to convert the phase to its physical dimension with Eq. (6.10).

6.4.2 3-D crater wear of samples

Two types of tool insert, namely, A30N (uncoated) and AC325 (coated), were tested. The fringe patterns were projected onto the rake face of the inserts and the captured images were 768 × 576 pixel at 256-graylevel.

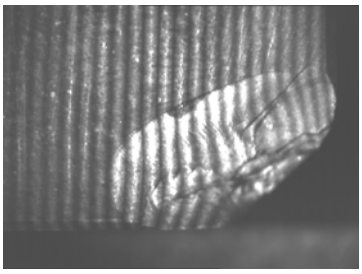
Figures 6.3-6.9 show the results of the seven samples of tool inserts. For each sample, the fringe image of the rake face projected with the phase shift value of '0', and the constructed 3-D map of the tool insert are shown. In the 3-D map, the image was truncated to focus on the crater wear area.

For each sample, a 3-D map constructed using laser scanning is also shown for comparison. The laser scanning was performed with Mahr's multi-sensor OMS400 measuring machine. Each time, only one line can be scanned. Therefore, to construct the 3-D map to sufficient resolution, a large number of scans are necessary.

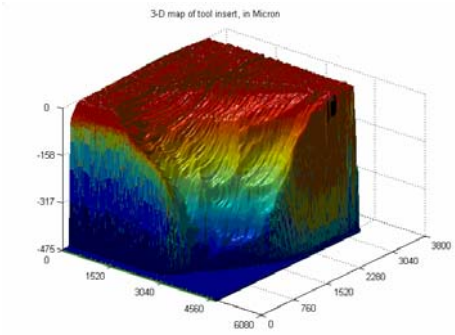
To verify the accuracy of the phase-shifting method, the maximum depths measured with phase-shifting method and with a microscope are compared in Table 6.1. The microscope is Keyence VHX-100 digital microscope. By manual focus on the rake face and the lowest point, the maximum depth can be measured. This measurement can be taken as the benchmark and the results measured by phase-shifting method can be compared accordingly.

Table 6.1 Maximum crater wear depths for seven insert samples

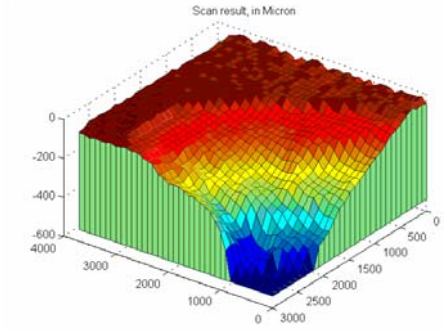
Sample #	1	2	3	4	5	6	7
Insert Type (ISO SDKN42MT)	A30N, uncoated			AC325, coated			
Z_0, measured by microscope (μm)	468	549	433	302	234	221	585
Z_1, measured by phase-shifting (μm)	450	554	448	315	220	238	578
$\Delta Z=Z_1-Z_0$ (μm)	-18	5	15	13	-14	17	-7
$\Delta Z/Z_0$ (%)	-3.8	0.9	3.5	4.3	-6.0	7.7	-1.2



(a) Rake face, fringe phase shift = 0

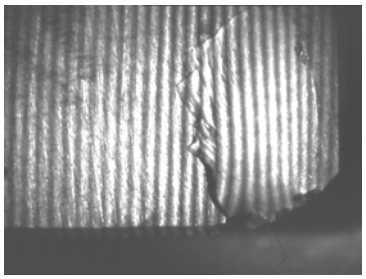


(b) 3-D map using phase-shifting, maximum depth = 450 μm

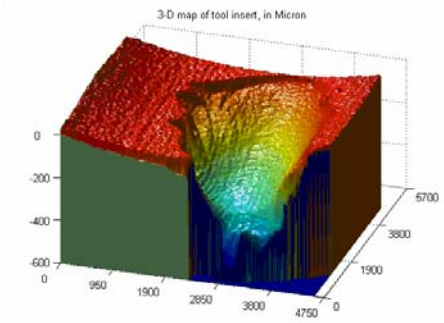


(c) 3-D map using laser scanning, maximum depth = 473 μm

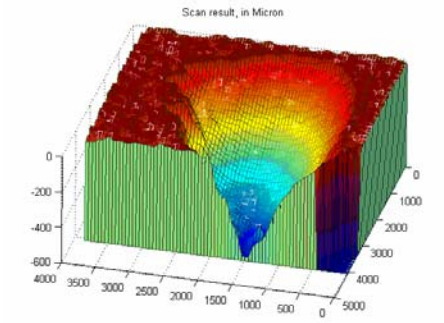
Figure 6.3 Sample 1



(a) Rake face, fringe phase shift = 0

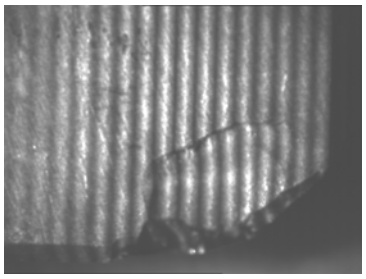


(b) 3-D map using phase-shifting, maximum depth = 554 μm

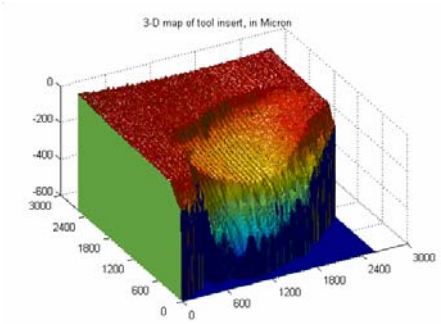


(c) 3-D map using laser scanning, maximum depth = 553 μm

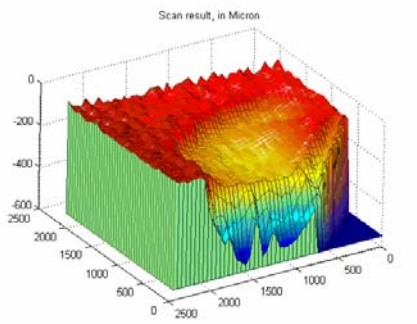
Figure 6.4 Sample 2



(a) Rake face, fringe phase shift = 0

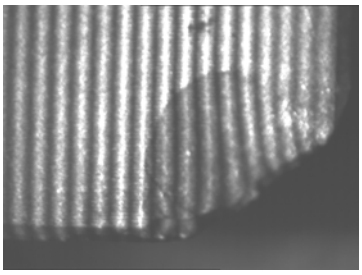


(b) 3-D map using phase-shifting, maximum depth = 448 μm

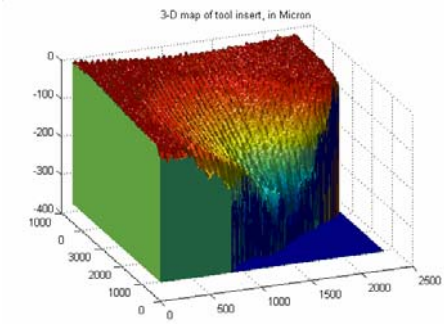


(c) 3-D map using laser scanning, maximum depth = 440 μm

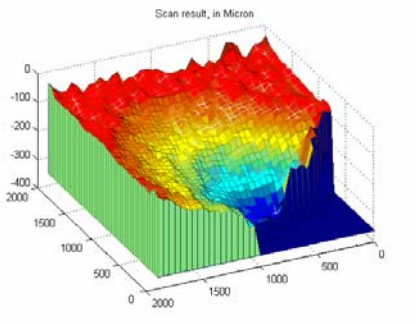
Figure 6.5 Sample 3



(a) Rake face, fringe phase shift = 0

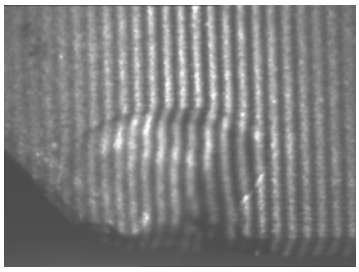


(b) 3-D map using phase-shifting, maximum depth = 315 μm

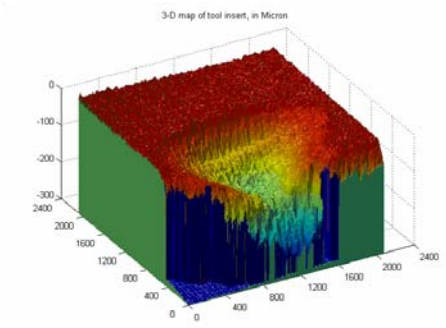


(c) 3-D map using laser scanning, maximum depth = 306 μm

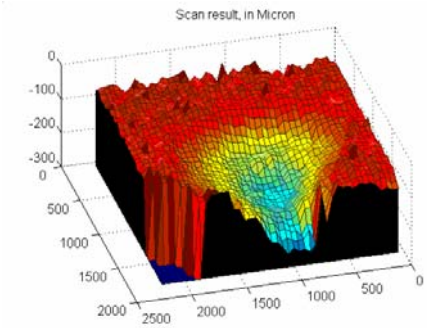
Figure 6.6 Sample 4



(a) Rake face, fringe phase shift = 0

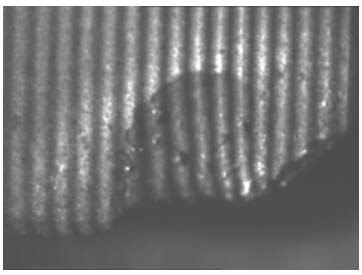


(b) 3-D map using phase-shifting,
maximum depth = 220 μm

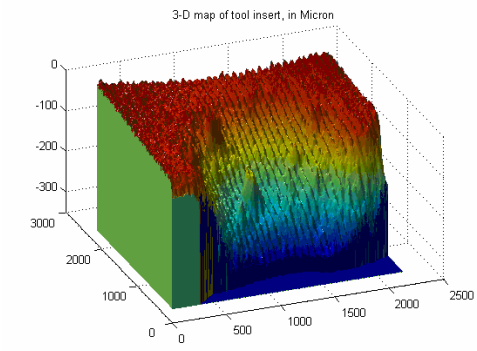


(c) 3-D map using laser scanning,
maximum depth = 237 μm

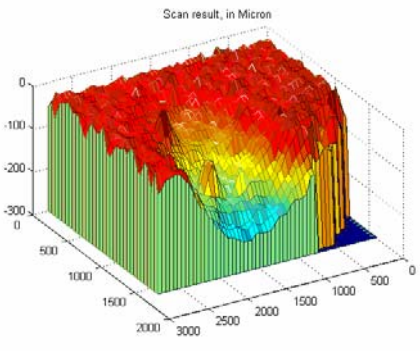
Figure 6.7 Sample 5



(a) Rake face, fringe phase shift = 0

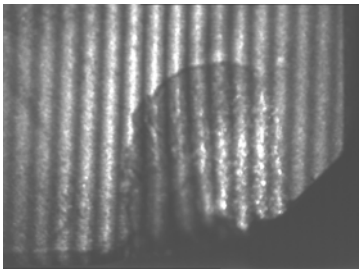


(b) 3-D map using phase-shifting,
maximum depth = 238 μm

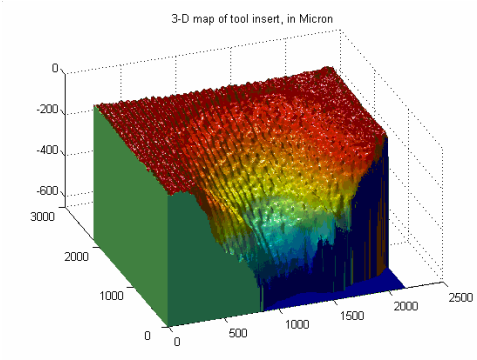


(c) 3-D map using laser scanning,
maximum depth = 227 μm

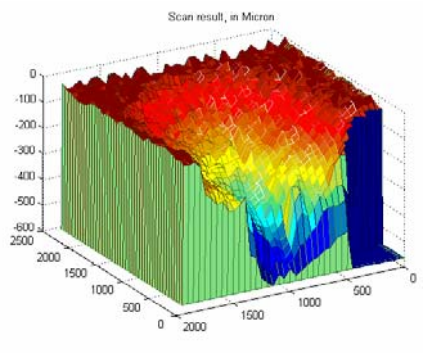
Figure 6.8 Sample 6



(a) Rake face, fringe phase shift = 0



(b) 3-D map using phase-shifting,
maximum depth = 578 μm



(c) 3-D map using laser scanning,
maximum depth = 576 μm

Figure 6.9 Sample 7

From the above figures, one can see that the 3-D map of the tool insert, and the whole size of the crater wear can be easily obtained. From Table 6.1, the deviation of the result measured with phase-shifting method between that measured with the microscope is small (maximum absolute deviation was only -18μm for Sample 1 and maximum percentage error was 7.7% for Sample 6). Additionally, the results given by this method also match those given by a laser measuring machine (particularly, the maximum depth on the edge, which has small deviation). But this method is of full-field nature, which surpasses laser scanning that must scan line-by-line. Accordingly, to construct a 3-D map with higher resolution, the laser scanning resolution has to be higher. This will make the hardware more complex and costly, and it will take more

time to scan the crater wear area. By contrast, the efficiency of the phase-shifting method as a full-field method is higher.

6.5 Discussion

For the unwrapped phase map, there are two important post-processing steps to remove the background noise and thus, improve the quality of the 3-D map.

Outside the rake face is the background. There is much noise introduced through the unwrapping process. Because there are no fringes on the background, the unwrapping algorithm will accumulate phase error on the background. The range of the background noise usually is wider than that of the points in the crater wear land. In visualization, it will immerse the useful information because the valid range of phase in the wear land just spans a small part of z axis and make the true crater wear land hardly visible. A mask can be produced and applied to this noisy phase map.

It is easy to obtain a binary image by thresholding the fringe image of the rake face since the gray-level difference between the tool surface and the background is distinct. Pixels inside the tool surface are turned to white (‘1’), and those outside (background pixels) are turned to black (‘0’). The binary image can be applied to the phase map to remove the background noise. Figure 6.10 shows the mask image for Sample 1.

Figure 6.10 The mask image for Sample 1

The non-uniform surface of the wear land introduces noise on the tool surface. To reduce it, a median filter (Gonzalez and Woods, 2002) can be applied, whereby the

value of a pixel is replaced by the median of the gray levels in the neighborhood of that pixel. The median filter is used here because the noise is impulse or salt-and-pepper noise, which can be reduced effectively especially by this filter. As a result, this filter introduces little error on the filtered phase map in terms of actual phase values, while the noise is removed. A 5 × 5 filter has been found to be effective by experiment.

It is necessary to pay attention to the intensity of the fringe projected on the rake face. Excessive intensity will make the captured image saturated (gray-level of the pixels perpendicular to the fringe is not sinusoidal). As a result, the accuracy will be affected. To move the LCD projector away from the rake face and/or adjust the position of the CCD camera can easily address this problem.

Generally, given a fixed image resolution, with more fringes projected on the crater wear area, the resolution of this method will be higher. But if the fringe is too narrow, the abrupt change of depth cannot be accurately observed. Hence, there is a compromise between resolution and accuracy when selecting the fringe width.

Figure 6.11 demonstrates the experimental setup with the insert on the milling holder. By adjusting the positions and angles of the projector and the CCD camera, the crater wear land can be seen by the camera clearly. A 3-D map with very similar accuracy of the crater wear can be constructed. Figure 6.12 shows the fringe image captured using the experimental setup in Figure 6.11 with the insert on the milling tool holder. The constructed 3-D map of Sample 1 with this setup can be compared with the constructed 3-D map using the experimental setup shown in Figure 6.2. Compared to Figure 6.3 (b), the details in the crater wear differ slightly. Theoretically, the performance of the system is independent of the location of the tool. Because the orientation and location of the camera and fringe projector can be adjusted in response to the change of the location of the tool, the relative coordinates of these components

can be fixed. That is to say, whether the tool is on the stage or on the holder, the system is able to get nearly the same result. The only requirement is to get a good fringe image such that fringes can reflect depth change.

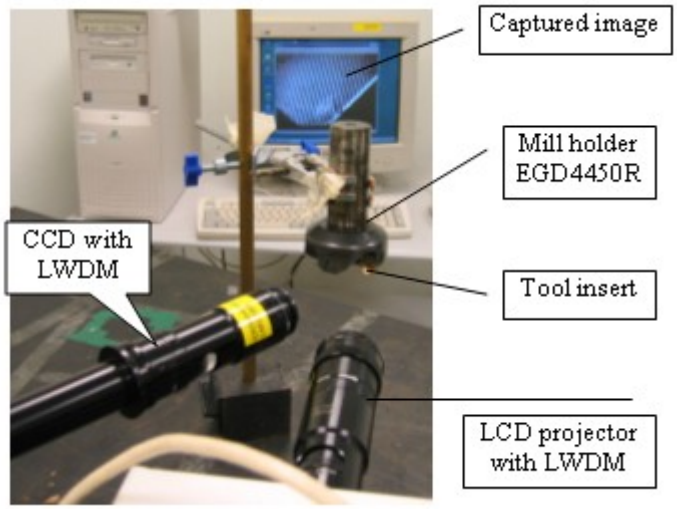
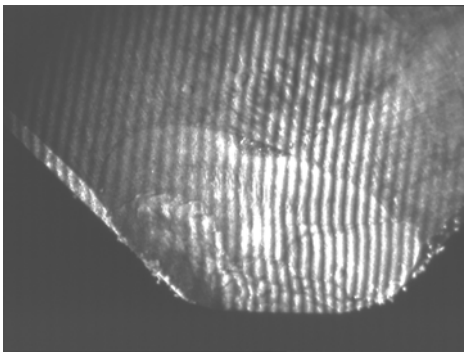
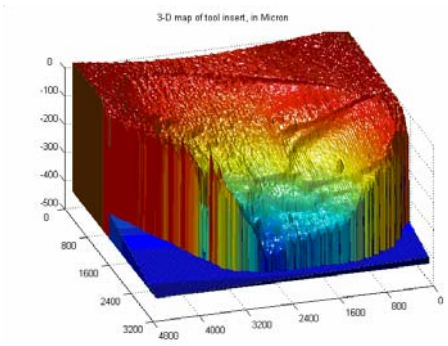


Figure 6.11 Experimental setup tried with a mill holder



(a) Fringe image



(b) 3-D map, maximum depth = 456 μm

Figure 6.12 Sample 1 reprocessed with the setup shown in Figure 6.11

With a suitable setup, the system can be used in an industrial environment to measure and monitor crater wear. For example, in monitoring the crater wear, the tool can be moved away from the cutting zone and coolant to a pre-specified location, where the fringes can be projected and images captured with the equipment already appropriately oriented and placed. With the captured images, the pre-calibrated system can give the crater wear map.

Limited by the availability of the experimental devices, this crater wear measurement system has not been put *in situ* to process the images captured when spindle rotates using trigger-capture mechanism as did in the successive image analysis method for flank wear measurement. Therefore, in the next chapter, crater wear is not considered in wear estimation via the self-organizing map network.

Chapter 7
Flank wear estimation and breakage detection by force

7.1 Problem statement

As mentioned in Chapter 3, cutting force is used to monitor both the breakage and flank wear in-process. Most researchers proposed monitoring strategy based on force features and neural networks, especially supervised learning networks, to make decisions on tool conditions, as reviewed in Chapter 2. Supervised learning needs collection of lots of data and the training is off-line. The generalization of the trained networks is also not good for varied cutting conditions. Although unsupervised networks, such as SOM, have been investigated, the training is still based on the whole dataset collected from a fresh tool to a worn-out tool. It is not adaptive to various cutting conditions. Thus in this chapter, one of the challenges is to estimate flank wear in-process without off-line training and involvement of cutting conditions. Another challenge is to detect breakage in time.

In accordance with these two challenges, in the following sections, the following key issues are presented:

- A self-organizing map (SOM) network is introduced as estimator.
- SOM is locally trained in a batch mode and applied immediately.
- Two features of the cutting force, which closely indicate flank wear, are extracted in-process and used to train the SOM.
- Force features in time domain, frequency domain and wavelet domain are extracted and compared. The sensitive features are identified and used to detect breakage.

7.2 Kohonen's self-organizing map

7.2.1 Why SOM

Self-Organizing Map (SOM) has the ability to form a topographic map in which the spatial locations of the neurons in the lattice are indicative of intrinsic statistical features contained in the input patterns (Kohonen, 1982; Kohonen, 1990; Kohonen, 2001). It transforms high-dimensional input space into a low-dimensional grid where the location of a unit represents the density of the input data. In this transformation, relative distances between data points are preserved.

In this case where the force signal is closely related to wear progress, SOM can map the input signal (force) with same patterns to an output unit and accordingly, the conditions can be classified or the flank wear be estimated. A big advantage is that the majority of the input feature vectors can be used in unsupervised learning, and just a small portion are needed for supervised learning together with their outputs.

Generally, wear progresses extensively at the short initial stage and progresses nearly linearly at the linear stage, within which most of the useful tool life lies. Change of a worn tool is usually recommended before the end of the linear stage (Zhou et al., 1995). The force signal at this stage is also stable in some sense. Consequently, the force features in the previous machining pass can be used to train the SOM network, which is subsequently applied to the current machining pass to estimate flank wear. With this scheme, the traditional off-line training of the neural network, which generally uses overall dataset collected off-line (i.e., tool from fresh to worn-out), is avoided. More importantly, cutting conditions are not involved in this training as traditional neural network methods do.

7.2.2 Principle

There exist many versions of the SOM. As shown in Figure 7.1, the basic SOM defines a mapping from the input data space $\Re^n$ onto a 1-D or 2-D array of units (Kohonen, 1982; Kohonen, 1990; Kohonen, 2001), whereby the latter is more often the case.

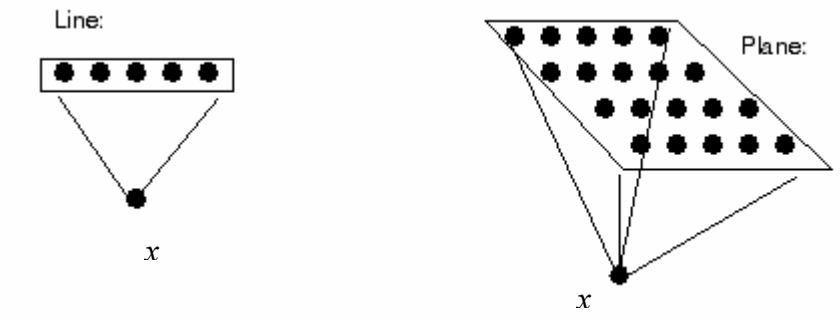


Figure 7.1 Mapping of SOM

With each unit i, a weight vector $m_i \in \Re^n$ is associated. A metric function is used to compare an input vector $x \in \Re^n$ with each m_i. The closest match is defined as response: the input is thus mapped onto this location. A common metric function used is the Euclidean norm $||x - m_i||$ .The best-matching unit or winner, c, is defined as:

$$\|x - m_c\| = \min_i\{\|x - m_i\|\}, \text{ or } c = \arg\min_i\{\|x - m_i\|\} \tag{7.1}$$

For the purpose of topological ordering, during learning, the weight vector of those units that are topographically close to the winner c will also be updated from the same input. The learning rule is given below:

$$m_i(t + 1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \tag{7.2}$$

where t is the discrete-time coordinate, and $h_{ci}(t)$ is a neighborhood kernel defined over the lattice points.

Usually $h_{ci}(t)=h(\|r_c - r_i\|, t)$, where $r_c \in \Re^2$ and $r_i \in \Re^2$ are the location vectors of units c and i, respectively, in the array. With increasing $\|r_c - r_i\|$ and time t, $h_{ci}(t) \to 0$.

The following function is used as the neighborhood kernel.

$$h_{ci}(t) = l(t) \exp(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}) \tag{7.3}$$

where $l(t)$ is some monotonically decreasing function of time ($0<l(t)<1$), $\sigma(t)$ defines the width of the kernel and is also some monotonically decreasing function of time, as shown below:

$$\alpha(t) = \alpha(0) \exp(-t/\tau_1),\ \sigma(t) = \sigma(0) \exp(-t/\tau_2) \tag{7.4}$$

where τ_1 and τ_2 are time constants.

The training is usually performed in two stages. In the first stage, relatively large initial learning rate $l(0)$ and neighborhood radius $\sigma(0)$ are used. In the second stage, both learning rate and neighborhood radius are small right from the beginning. This procedure corresponds to first tuning the SOM approximately to the same space as the input data and then fine-tuning the map.

7.2.3 Batch training algorithm

Instead of using a single input vector at a time, the whole input set is presented to the map before any adjustments are made – hence the name "batch." (Kohonen, 2001) Each unit has a Voronoi set, which collects a list of input vectors whose winners belong to the current neighborhood of the unit. So, in each training step, the data set is

partitioned into Voronoi sets whose number is equal to that of units. The algorithm proceeds as follows:

- Calculate the sum of the vectors in each Voronoi set:

$$s_i(t) = \sum_{j=1}^{N_{Vi}} x_j \tag{7.5}$$

where N_{Vi} is the number of input samples in the Voronoi set of unit i.

- Update the weight vectors:

$$m_i(t+1) = \frac{\sum_{j=1}^{u} h_{ij}(t) s_j(t)}{\sum_{j=1}^{u} N_{Vj} h_{ij}(t)} \tag{7.6}$$

where u is the number of units.

In batch training algorithm, no convergence problems will be encountered. Compared to basic SOM algorithm (sequential training algorithm) aforementioned, this algorithm does not involve learning rate and performs faster. A few iterations will usually suffice.

7.3 SOM as estimator

Asymmetric mapping (Kohonen, 2001) is employed here. Here, the input data consisting of force features and wanted output (wear increment value) are called labeled data; otherwise those consisting of only force features are called unlabeled. Consider then that it is often much easier and less expensive to collect masses of unlabeled input data than a much smaller number of well-classified training samples. The unlabeled data should be utilized first for a preliminary unsupervised learning, which readily approximates the density function of input data. Then the labeled data

are used for a refinement of the density function. So, two-phase learning strategy is more reasonable and applied.

7.3.1 Phase one

Presentation of all the available unlabeled input data to a normal SOM, whereby the input vectors are of the form:

$$x=[in\ \phi]^T$$

where $in=[f_1, f_2, \ldots, f_m]^T$, a feature vector at one instance of time, its corresponding output vector is denoted as $out=[o_1, o_2, \ldots, o_n]^T$ (in this case, n=1, flank wear increment); the symbol ϕ means the "don't care" condition, i.e., when searching for the winner, only the *in* part of x is compared with the corresponding components of the weight vectors, and no output (*out*) part is thereby present in the learning algorithm.

7.3.2 Phase two

After convergence of phase one, the labeled data are applied to the SOM and training is continued. During this second phase, the winner is determined only based on the *in* part of x, while the weight vectors corresponding to the *in* part of x are no longer changed. Instead, the SOM learning algorithm is now applied to change the weight vectors of the *out* part only.

So, the input vector x in winner search and learning are different.

During winner search: $x=[in\ \phi]^T$; during learning: $x=[\phi\ out]^T$. Notice that the *in* and *out* in phase two are labeled, whereas in phase one they are unlabeled.

7.4 Estimation by SOM

7.4.1 Feature extraction

The average force and the standard deviation over one rotation have been found to be correlated to flank wear well (Elbestawi et al., 1989; Leem and Dornfeld, 1995; Lin and Lin, 1996; Santanu et al., 1996).

$$F_a(i) = \frac{1}{N}\sum_{j=1}^{N} f(i,j) \tag{7.7}$$

$$\sigma(i) = \sqrt{\frac{1}{N-1}\cdot\sum_{j=1}^{N}\left[f(i,j) - F_a(i)\right]^2} \tag{7.8}$$

$F_a(i)$, $\sigma(i)$ refer to the average force and standard deviation over rotation i, $f(i,j)$ refers to the raw force at instance of time j in rotation i, N is the number of sample points in one rotation.

The average force and standard deviation in each machining pass instead of in the whole test are processed to extract features. That is to say, the features are extracted pass by pass.

The two features are firstly filtered with median algorithm. The objective is to remove the great fluctuation and to make them smoother. Direct use of the filtered features is still not a good idea, since in SOM, input data should be normalized in order to remove the effect on weight vectors imposed by the greater value of some input data. To do the scaling, take the filtered force in the kth pass as an example. The force at the beginning (force at 1st rotation) of the kth pass is denoted as $L(0)$, then any force at nth rotation in this pass $L(n)$ is scaled by:

$$F'(n)=(L(n)-L(0))/L(0); \tag{7.9}$$

The scaled value is then accumulated to form a stable feature by the following formula:

$$F_e(n) = \sum_{i=0}^{n} F'(i) \tag{7.10}$$

The scaled value $F_e(n)$ is now suited to be the feature. The same scaling method is applied to the filtered standard deviation.

7.4.2 Working with SOM

A 1-D SOM is constructed with 15 units. In training, the majority of the data in the *k*th pass are unlabeled (only the two force features) and used in phase one. A small portion of the data are labeled (the two force features and its corresponding wear increment value) according to the overall wear increment during this whole pass. The labeled data are used in phase two. The trained SOM is then applied on each data in the (k+1)th pass to estimate flank wear increment at different rotations.

The output of the SOM is flank wear increment with respect to the wear value at the beginning of this (k+1)th pass. This wear value has been available by means of interpolation of the measured flank wear which is obtained with the tool-maker's microscope. Thus, the absolute wear at each time is the summation of the wear value plus the wear increment value estimated by SOM.

7.5 Breakage detection

Breakage detection is an important task in TCM. To detect breakage, typically there are three methods based on features or patterns extracted in different domains, i.e., time domain, frequency domain and wavelet domain. In what follows, these three

methods will be compared via case study and identify which method is more sensitive to breakage.

7.5.1 Features in time domain

This method has been the mostly studied since the early stage of research in TCM. Many features have been proposed as shown in Table 2.4. Here, two features are especially focused on, i.e., residual error and force peak rate.

Altintas (1988) has shown that the residual error of the average force rather than instantaneous force was able to embody the breakage. With this conclusion, the calculation of residual error on basis of rotation rather than every sampling point is by any means more time-saving. The residual error is based on AR1 model. Its calculation follows the procedure stated below.

AR1 model for the first differencing of the average force can be expressed as:

$$f_a(t) = \Phi f_a(t\text{-}1) + \varepsilon(t) \tag{7.11}$$

$$f_a(t) = F_a(t) - F_a(t\text{-}1) \tag{7.12}$$

where $F_a(t)$ is the average force over the t-th rotation period. The residual error can therefore be calculated by:

$$\varepsilon(t) = f_a(t) - f_a(t-1)\cdot\hat{\Phi}(t-1) \tag{7.13}$$

where Φ can be evaluated recursively by:

$$\hat{\Phi}(t) = \hat{\Phi}(t-1) + K(t-1)\cdot\varepsilon(t) \tag{7.14}$$

where $K(t)$ is updated by:

$$K(t) = \frac{P(t) \cdot f_a(t)}{\lambda + P(t) \cdot f_a^2(t)} \tag{7.15}$$

And $P(t)$ can be updated by:

$$P(t+1) = \frac{P(t)}{\lambda}[1 - K(t) \cdot f_a(t)] \tag{7.16}$$

The initial conditions are $\Phi(0) = 0$, and $P(0) = \beta$ where β is a large number. The forgetting factor λ is $0.9 \le \lambda \le 1$.

A residual error exceeding a preset threshold (i.e., $\varepsilon > T_{re}$, T_{re} is the threshold) indicates breakage.

To calculate residual error, the average force over one rotation is considered. The detail force, i.e., instantaneous force at each sampling point is not considered. This may lead to mistake when the average force cannot reflect the breakage. Therefore, peak rate of cutting forces (Zhang et al., 1995) is taken into account too. Its definition is given by:

$$K_{pr}(t) = \frac{F_m(t) - F_m(t-1)}{|F_m(t)| + |F_m(t-1)|} \tag{7.17}$$

where $F_m(t)$ is the peak value of the cutting force over the t-th rotation period. The force peak rate K_{pr} is dimensionless and independent of the cutting conditions. A value exceeding a preset threshold (i.e., $K_{pr} > T_{pr}$, T_{pr} is the threshold) indicates tool breakage.

As pointed out by Altintas (1988) and Zhang et al. (1995), it is very difficult to specify T_r and T_{pr}, e.g., many trial tests are required. In this thesis, since conditions are monitored pass-by-pass, dynamical T_r and T_{pr} can be obtained without any trial tests.

In the first pass, it is assumed that there is no breakage, the maximum amplitudes of residual error and peak rate can be obtained. These two values are taken as the thresholds for the beginning of the second pass. From the second pass onwards, as long as no breakage has been detected by vision after the *k*th pass, the maximum amplitudes of the two features in the *k*th pass can be taken as the thresholds for the beginning of the (*k*+1)th pass. Within each pass, if breakage is detected by force features but not verified by vision, i.e. breakage does not really occur, the thresholds are updated by increasing a little, say, 5%.

Figures 7.2 and 7.3 show the average force, instantaneous force, and the two extracted features. Clearly, the breakage can be detected through the two abrupt changes of the residual error and peak rate. In the following two sections, the dataset got from the same test will be used to do FFT and wavelet transform for comparison.

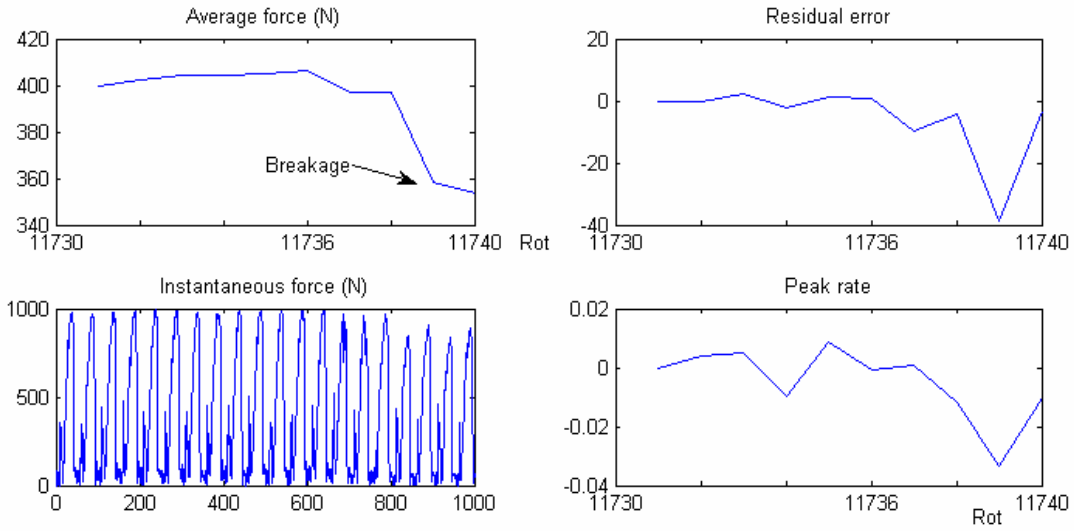


Figure 7.2 Residual error and peak rate for dataset 1, s = 1200 rpm, 2 inserts mounted

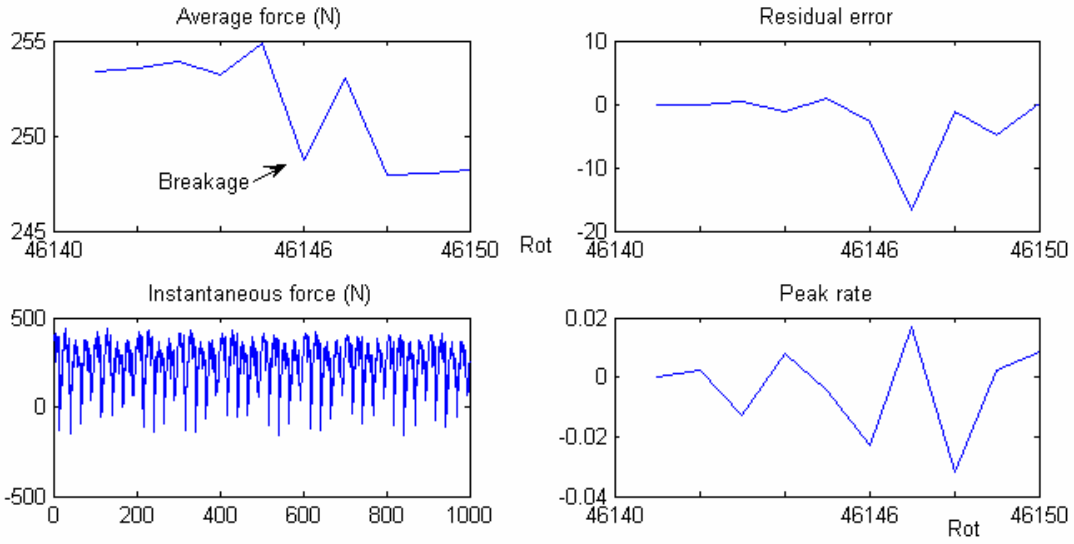


Figure 7.3 Residual error and peak rate for dataset 2, s = 1200 rpm, 4 inserts mounted

7.5.2 Features in frequency domain

It is stated that there exists a tool breakage zone in the spectral density of the cutting force. Between the rotation frequency, f_r and tooth frequency, f_t ($f_t = m_t \times f_r$, m_t is the number of teeth or inserts), the harmonics of f_r are indicators when breakage occurs (Tarng, 1990). Using a different chip thickness determination method from Tarng, the theoretical research based on a milling force model reinforces this finding.

In milling, as shown in Figure 7.4, the instantaneous tangential force acting on the cutting edge is modeled to be proportional to the area of the chip cut (Koenigsberger and Sabberwal, 1961):

$$F_t = k_s \cdot w_c \cdot h_{tr} \tag{7.18}$$

where,

F_t, instantaneous tangential force on the cutting edge;

k_s, specific cutting force coefficient;

w_c, chip width;

h_{tr}, true undeformed chip thickness.

According to Li et al. (2001), the true undeformed chip thickness can be approximated by this formula:

$$h_{tr} = |MC| - |NC|$$
$$\cong R_c \left(1 - \sqrt{1 - \frac{2 f_{pt} \sin\varphi_i}{R_c + \frac{m_t f_{pt}}{2\pi}\cos\varphi_i} - \frac{f_{pt}^2 \cos 2\varphi_i}{\left(R_c + \frac{m_t f_{pt}}{2\pi}\cos\varphi_i \right)^2} + \frac{f_{pt}^3 \sin\varphi_i \cos^2\varphi_i}{\left(R_c + \frac{m_t f_{pt}}{2\pi}\cos\varphi_i \right)^3}} \right) \quad (7.19)$$

R_c, radius of the cutter;

f_{pt}, feed per tooth per revolution;

m_t, number of teeth on the cutter,

φ_i, cutting edge rotation angle of the ith tooth.

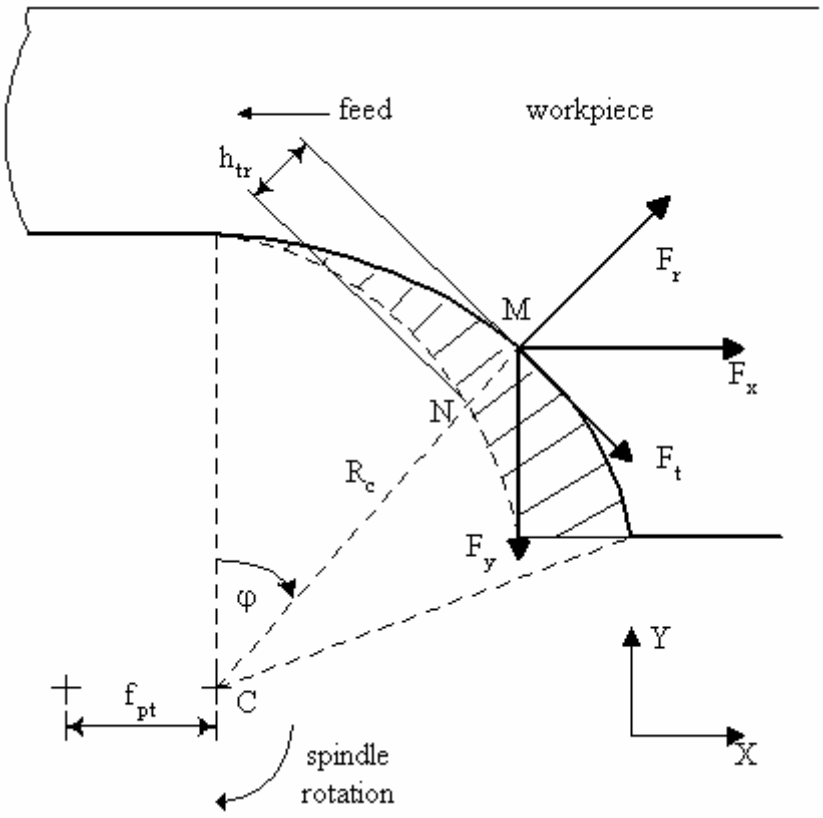


Figure 7.4 Force model in milling

The corresponding instantaneous radial force, F_r, has a linear relationship with F_t, and can be modeled by:

$$F_r = k_r \cdot F_t \tag{7.20}$$

where,

k_r, the ratio of the tangential force and radial force.

The instantaneous cutting forces in the X and Y directions can be obtained by decomposing the cutting forces F_t and F_r into the X and Y directions:

$$F_x = F_t \cdot \cos\varphi + F_r \cdot \sin\varphi \tag{7.21}$$

$$F_y = -F_t \cdot \sin\varphi + F_r \cdot \cos\varphi \tag{7.22}$$

In multi-tooth milling, the instantaneous cutting forces in the X and Y directions can be expressed as:

$$FX = \sum_{i=1}^{m_t} \delta(i) F_x(\varphi_i) \tag{7.23}$$

$$FY = \sum_{i=1}^{m_t} \delta(i) F_y(\varphi_i) \tag{7.24}$$

and

$$\delta(i) = 1, \text{ if } \varphi_{en} \le \varphi_i \le \varphi_{ex},$$
$$0, \text{ otherwise} \tag{7.25}$$

where,

φ_{en}, entry angle of cut;

φ_{ex}, exit angle of cut;

Finally, the instantaneous resultant cutting force, F, can be expressed as:

$$F = (FX^2 + FY^2)^{1/2}. \quad (7.26)$$

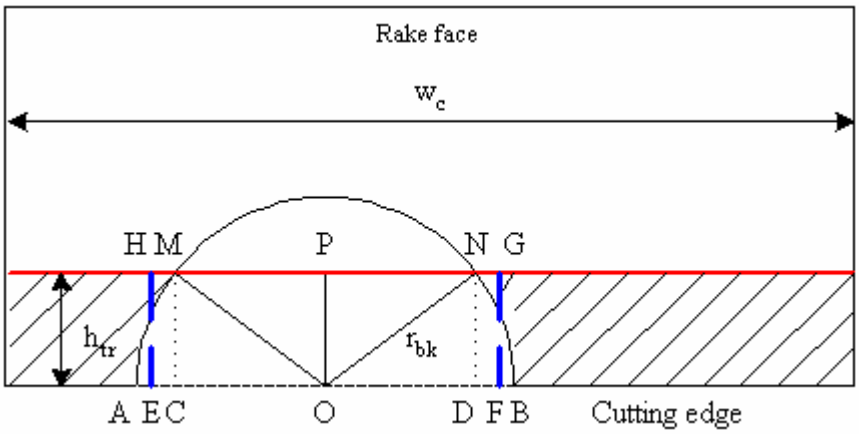


Figure 7.5 Assumed breakage geometry

To simulate the breakage, suppose the breakage forms a circle with a radius of r_{bk}, as shown in Figure 7.5. At some instance, when the chip thickness is h_{tr}, the shadowed area is the chip cut. Thus, for this broken insert, the area outlined by the arc AM, line MN, arc NB, and line BA is uncut. This uncut area can be viewed as a rectangle EFGH. From the viewpoint of the force, for the broken insert this uncut area is equivalent to a decreasing w_c. Conversely, for the insert following this broken insert, it cuts with an increasing w_c. The change of w_c is equivalent to |EF| and can be calculated easily.

Using the following parameters, a simulation was conducted on the force and its spectrum in the cases of with and without breakage: spindle speed 1200 rpm, feedrate 200 mm/min, 4 inserts, sampling rate 2000 Hz, w_c = 2 mm, breakage occurs on the 1st insert, r_{bk} = 1 mm. φ_{en} = 45 degree, φ_{ex} = 135 degree. The following findings have been obtained:

1. FFT over one rotation cannot provide the complete information of frequency components, especially the harmonics of f_r at the breakage zone. This can be

seen in Figure 7.6. In contrast, FFT over two or more rotations can provide harmonics of f_r, as shown in Figure 7.7.

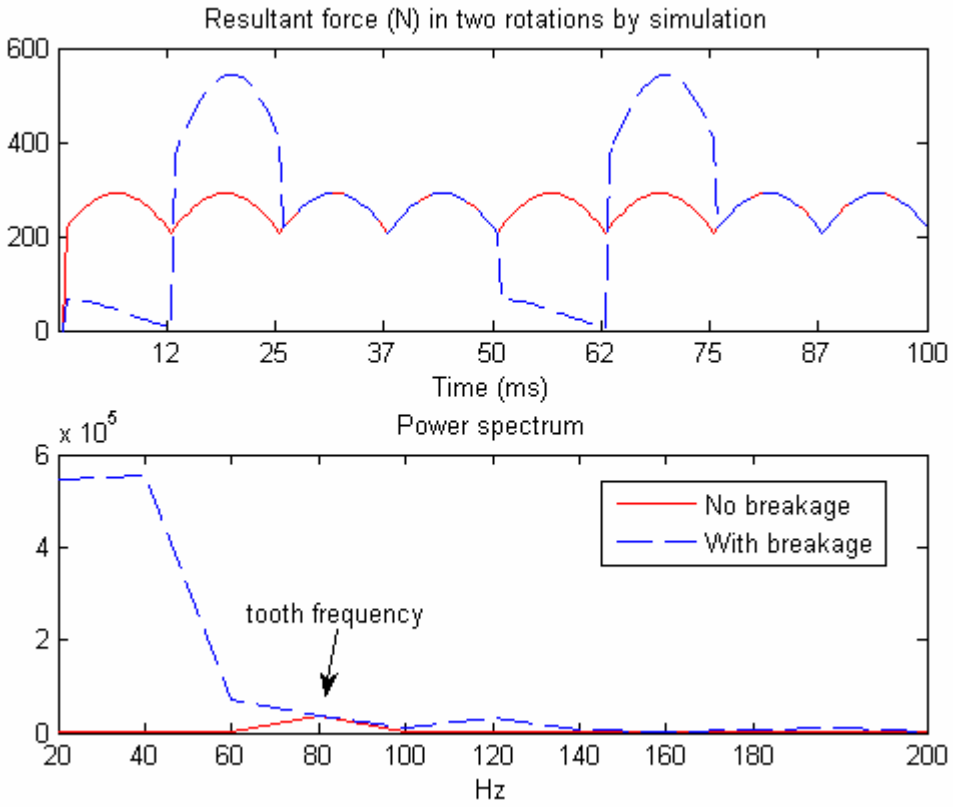


Figure 7.6 Simulated force and its power spectrum (FFT over one rotation)

2. Only with breakage, harmonics of f_r at the breakage zone stand out, indicating the breakage. This was observed in Figure 7.7.

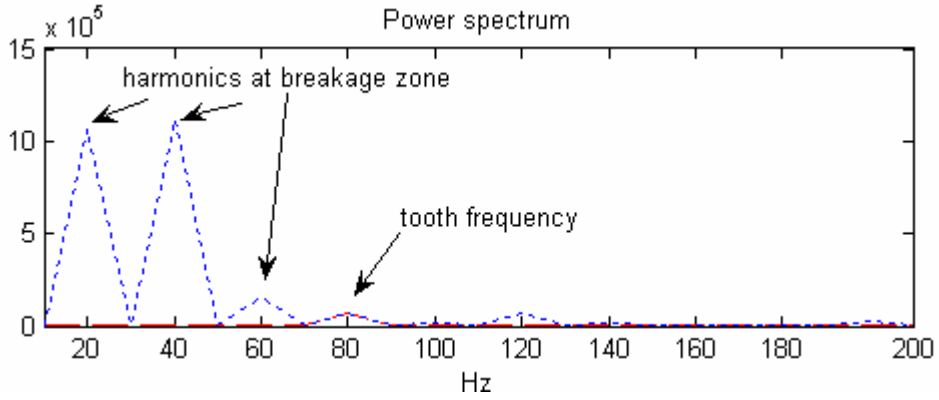


Figure 7.7 Power spectrum of the simulated force (FFT over two rotations)

In the following, FFT is performed on the real force, i.e., sampled signal in machining tests. Note that in the power spectrum plots, 'NB', 'CP' or 'BK' refers to

that one of these three events: no breakage, chipping, or breakage has occurred by some rotation(s). The following have been observed.

1. Like simulated force, for sampled force, FFT over one rotation cannot provide complete frequency components, opposed to FFT over two or more rotations. This is shown in Figures 7.8 - 7.11.

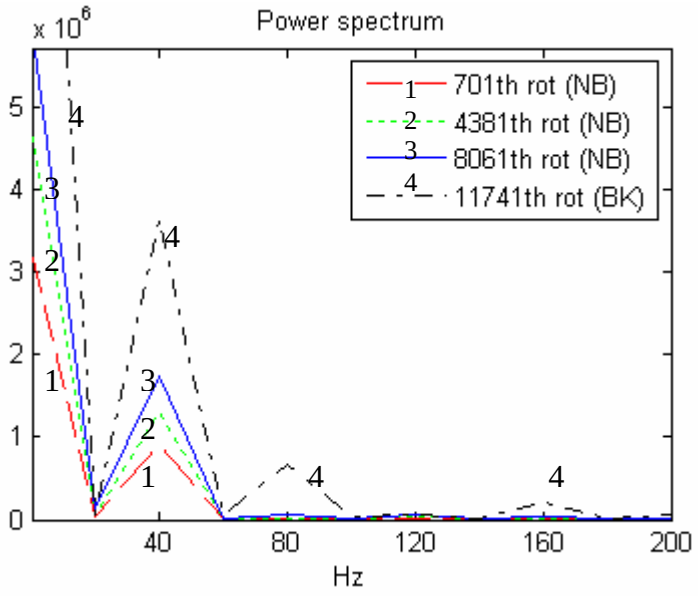


Figure 7.8 Power spectrum before and after breakage (FFT over one rotation), for dataset 1

2. The harmonics at the breakage zone are not reliable for breakage detection as expected from the theoretical simulation.

In Figure 7.10, the harmonic at 20 Hz when the insert was broken did not have a greater value than that of a fresh insert. Likewise, in Figure 7.11, a greater peak value is not observed at 20 Hz, 40 Hz and 60 Hz when the insert was broken. The peaks at 40 Hz in Figure 7.10 and at 80 Hz in Figure 7.11 do have a greater value when the insert was broken. However, this cannot be used as the breakage indicator, because this increasing amplitude in frequency domain is just a mapping of the increasing amplitude of cutting force in time domain. That is, the cutting forces for 701st, 4381st,

8061st and 11741st rotations increase in Figure 7.10, the same is true for 67th, 11587th, 34627th and 46147th rotations in Figure 7.11.

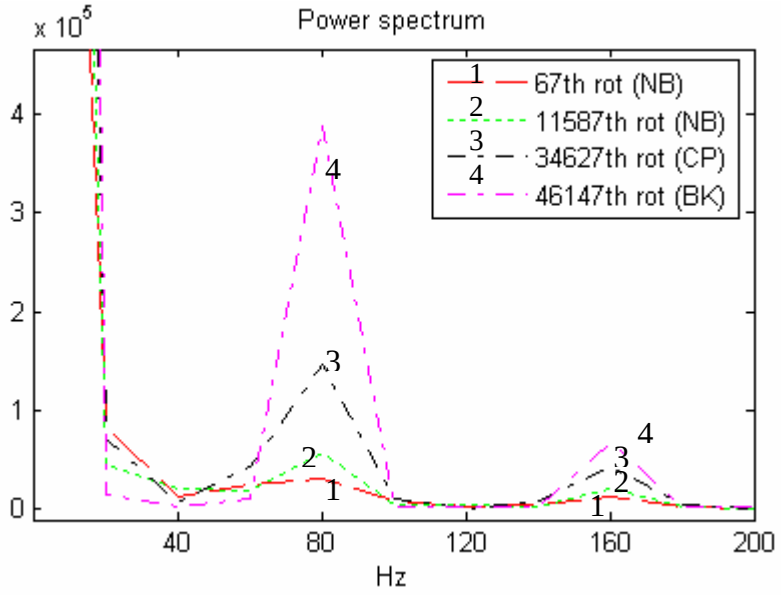


Figure 7.9 Power spectrum before and after breakage (FFT over one rotation), for dataset 2

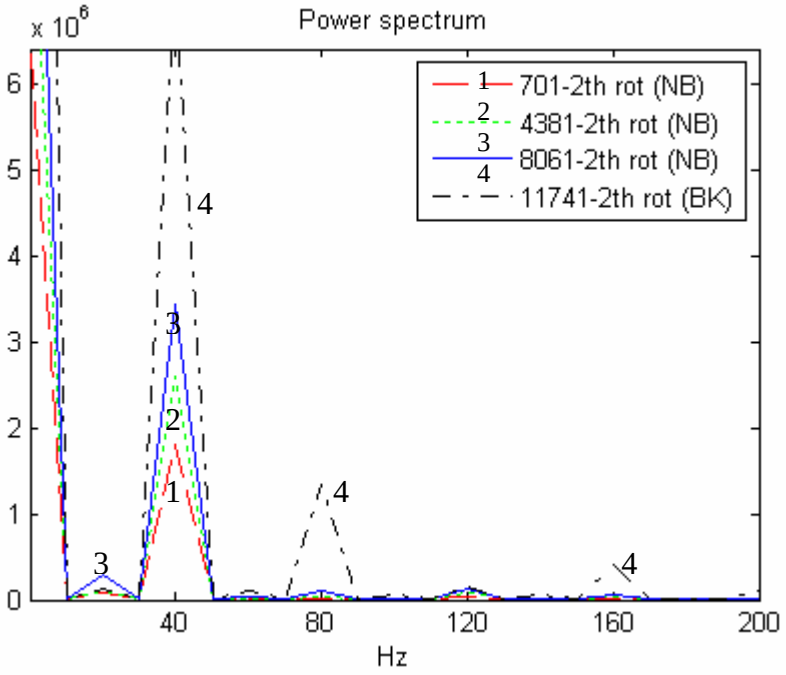


Figure 7.10 Power spectrum before and after breakage (FFT over two rotations), for dataset 1

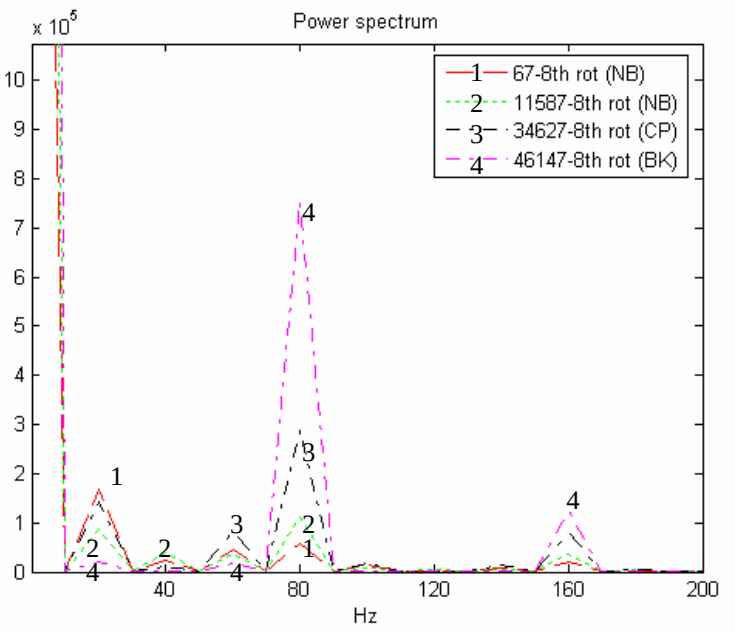


Figure 7.11 Power spectrum before and after breakage (FFT over two rotations), for dataset 2

Therefore, harmonics at the breakage zone are not reliable indicators of breakage in practice. A very important reason is that run-out can cause the same frequency component distribution at the breakage zone as breakage does. Removal of the run-out in machining is a big challenge to make frequency usable for breakage detection. But, it is impossible to eliminate run-out completely. Therefore, in practice, the effect caused by run-out always exists in the cutting force. Frequency pattern is not employed in this thesis for breakage detection.

7.5.3 Features in wavelet domain

Wavelet transform, having good resolution both in time and frequency domains, was adopted in breakage detection (Kasashima et al., 1995; Lee and Tarng, 1999). It is argued that detail components of the wavelet transformation of the force signal can be used to detect breakage by making the transients stand out.

Basically, wavelet transform decomposes a signal into *approximations* and *details* by low-pass and high-pass filters. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components. This decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. Each iteration of decomposition is called a level. Figure 7.12 shows a decomposition of a signal to level 3.

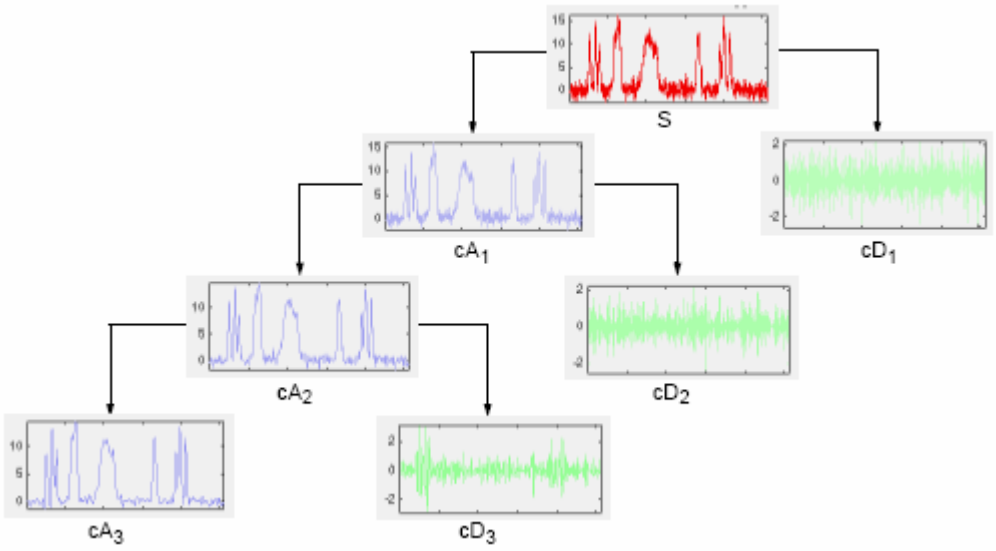


Figure 7.12 An example of wavelet decomposition

The dataset 1 and 2 involved in previous two sections were again used to do wavelet transform to level 5 with wavelet db4. The transformed details and approximation are shown in Figures 7.13 and 7.14, where A5 means approximation at level 5. D5 means details at level 5. D4 means details at level 4, and so on.

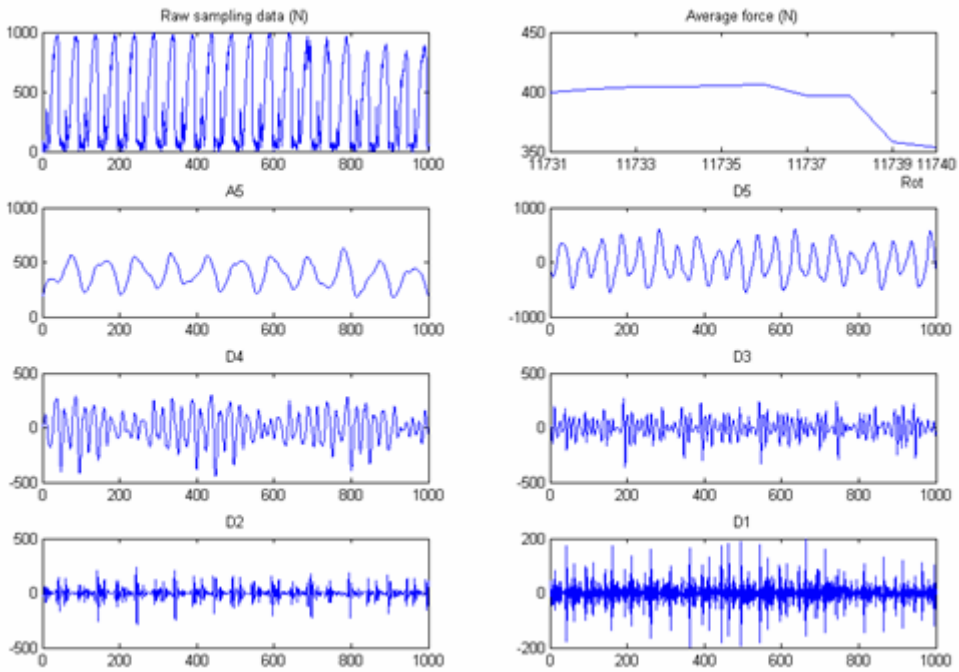


Figure 7.13 Wavelet transform for dataset 1

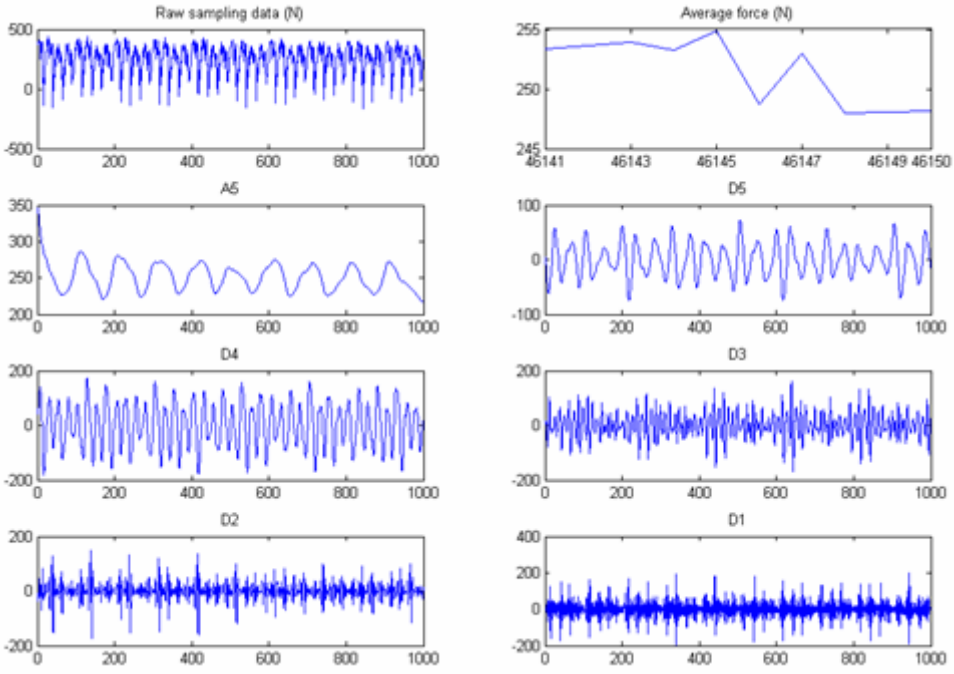


Figure 7.14 Wavelet transform for dataset 2

From Figures 7.13 and 7.14, the approximation and the details do not show any specific patterns when breakage occurs. It indicates that wavelet transform cannot be used to detect breakage.

In summary, collectively considering the effectiveness of the above three methods in time domain, frequency domain and wavelet domain, two features in time domain, i.e., residual error of average force and peak rate are easy and effective in detecting breakage. They are accordingly used in this thesis.

7.6 Experimental results

This section concentrates on wear estimation results rather than breakage detection. The latter will be shown in on-line cases, which are the focus of the next chapter.

7.6.1 Setup for force system

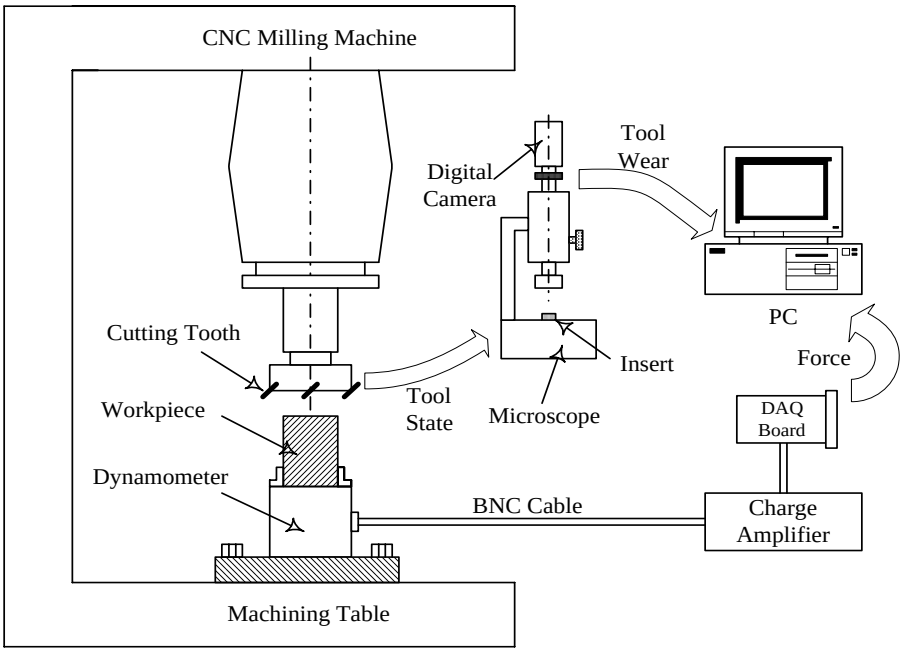


Figure 7.15 Experimental setup for force subsystem

The experimental setup is sketched in Figure 7.15. The devices are listed in Table 7.1.

Table 7.1 Experimental devices for force subsystem

Devices
Makino CNC milling machine with Funuc controller
EGD 4450R cutter
ASSAB718HH workpiece (210 mm × 43 mm × 106 mm)
Kistler 9265B Quartz 3-Component Dynamometer
Kistler 5019A Multi-channel Charge Amplifier
NI-DAQ PCI 1200 Board
Olympus microscope and Panasonic digital camera
PC with P3 300 MHz

The cutting force is captured by the dynamometer in the form of charges, and converted to voltages by the charge amplifier. The voltage signal is then sampled via the data acquisition (DAQ) card PCI-120. The digital voltage level is transferred to force level with the specified mechanical unit. Table 7.2 shows the parameters specified for the charge amplifier and the DAQ card.

Table 7.2 Parameters for charge amplifier and DAQ card

Charge amplifier	Channel Transducer Sensitivity / TS [pC/Mechanical unit] = pC/N Scale / SC [Mechanical units/V] = N/V Low-pass Filter / LP Time constant / TC (High-pass filter)	1 7.85 600 1K Hz Long
DAQ PCI-1200	Input range (V) Input mode Polarity Sampling rate (Hz) Channel Gain	±5 Single-ended Bipolar 2000 0 1

To get stable and useful force, the force signal is sampled only when the tool was fully engaged into workpiece. The effective sampling period, in this case, was therefore not throughout the whole length of workpiece, but part of it. This is shown in Figure 7.16.

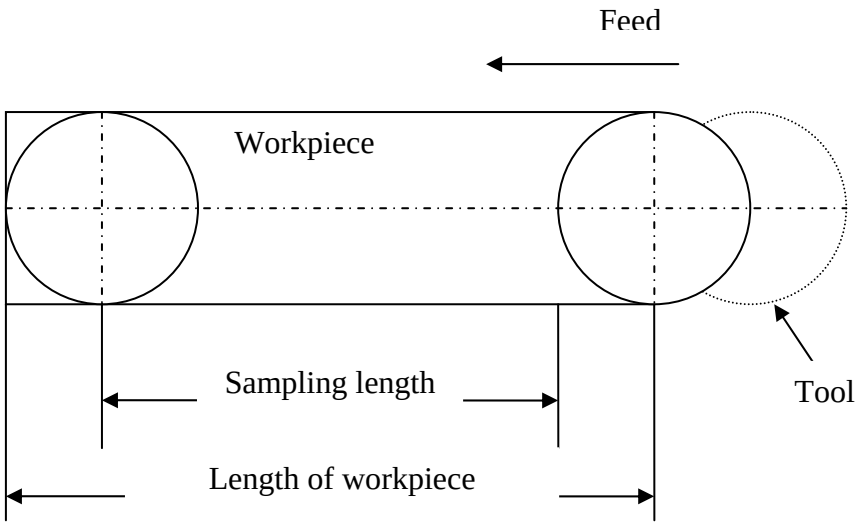


Figure 7.16 Effective force sampling period

The cutting force was recorded and saved on hard disk for off-line processing. After some number of passes, the tool inserts were taken away from the milling holder and inspected by a tool-maker's microscope for flank wear value. The wear values in between the measured passes were calculated by interpolating the measured values. Therefore, for each pass, the force features and wear increment value were extracted throughout this pass. Then SOM was trained by using the wear increment value and the extracted features. The trained SOM used the force features extracted in a new machining pass to estimate the wear increment and thus, the estimated flank wear in this pass. Throughout this period, SOM was trained locally after each pass and applied immediately to the next pass. It was updated pass-by-pass.

7.6.2 Wear estimation results by SOM and comments

Twenty tests with various cutting conditions and two types of inserts were tested and these are shown in Table 7.3. The experimental results are shown in Figures 7.17-7.26. For each test, the estimated wear and actual wear measured with the microscope mentioned in Table 7.1 are plotted for comparison, which is shown in the first row of

the figure. The second row of the figure shows the average force throughout the test. And the third row shows the two features extracted from the cutting force.

Table 7.3 Parameters of cutting tests for off-line wear estimation

Insert type	Test #	Spindle speed (rpm)	Feed rate (mm/min)	Depth of cut (mm)	# of inserts mounted	Time per pass (s)	# of passes
AC325 (coated)	a1	800	150	1	4	72	120
	a2	1000	100	1	2	108	50
	a3	1000	100	1	4	108	82
	a4	1000	200	1	2	54	60
	a5	1000	300	1	4	36	79
	a6	1200	150	1	2	72	40
	a7	1200	200	1	2	54	35
	a8	1200	300	1	4	36	65
	a9	600	100	2	4	108	73
	a10	600	200	2	4	54	101
	a11	800	100	2	2	108	56
	a12	1000	100	1	4	108	72
A30N (uncoated)	b1	800	200	1	4	54	97
	b2	800	300	1	4	36	30
	b3	1000	200	1	2	54	68
	b4	1000	300	1	4	36	20
	b5	1000	300	2	4	36	50
	b6	1200	100	1	2	108	6
	b7	1200	200	1	4	54	35
	b8	800	300	1	4	36	53

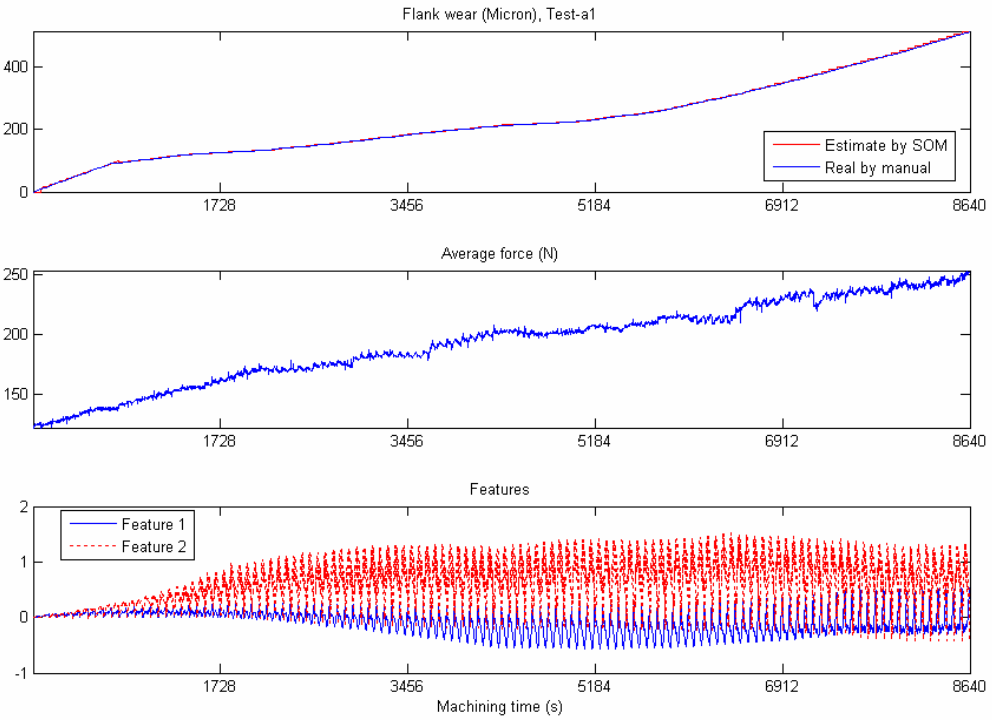


Figure 7.17 Wear estimation result for Test a1

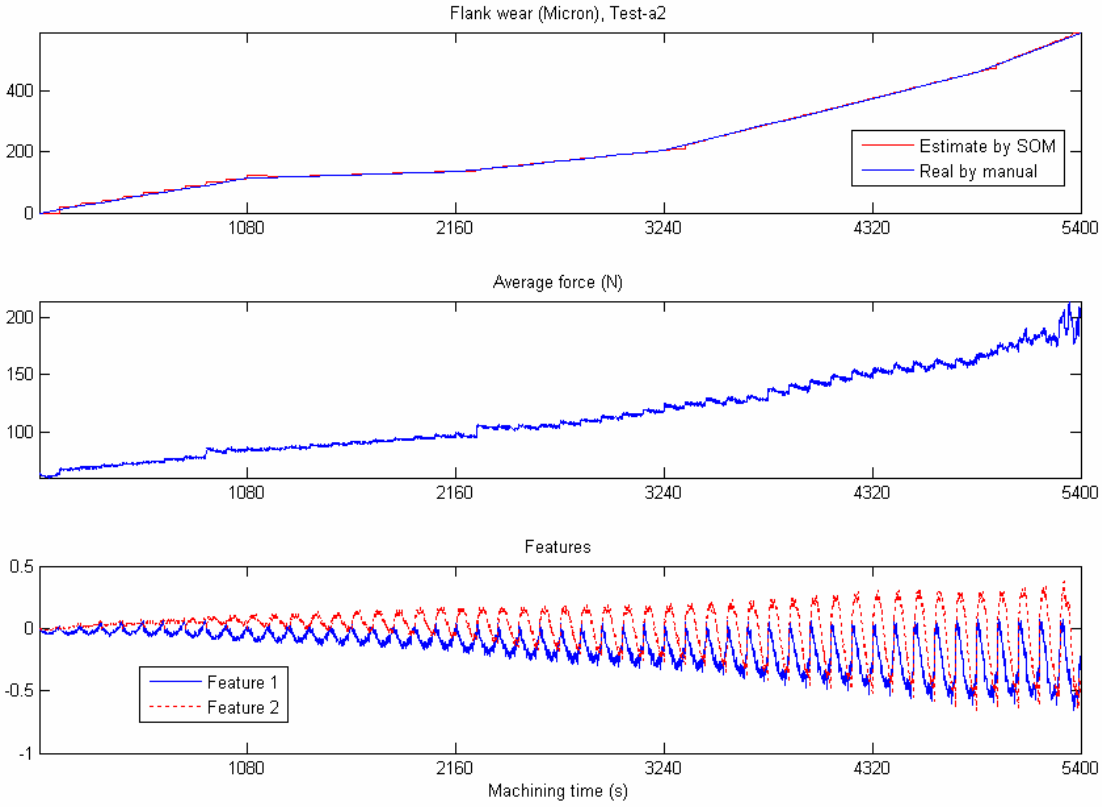


Figure 7.18 Wear estimation result for Test a2

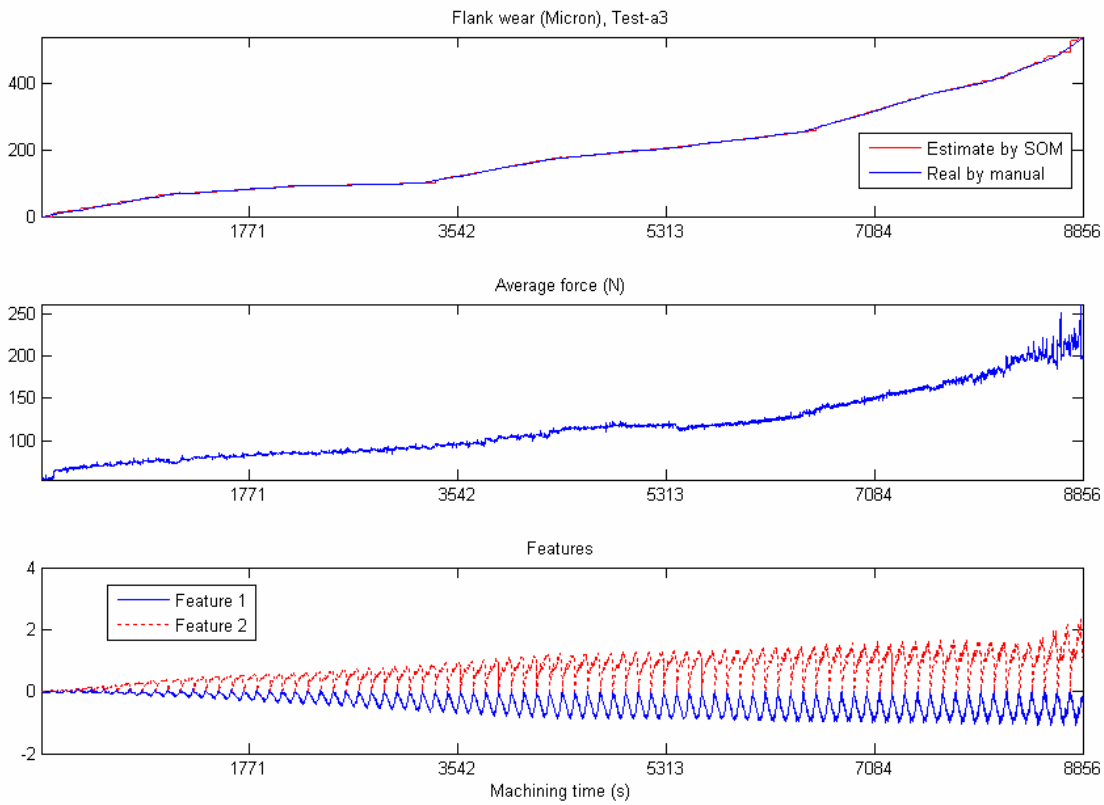


Figure 7.19 Wear estimation result for Test a3

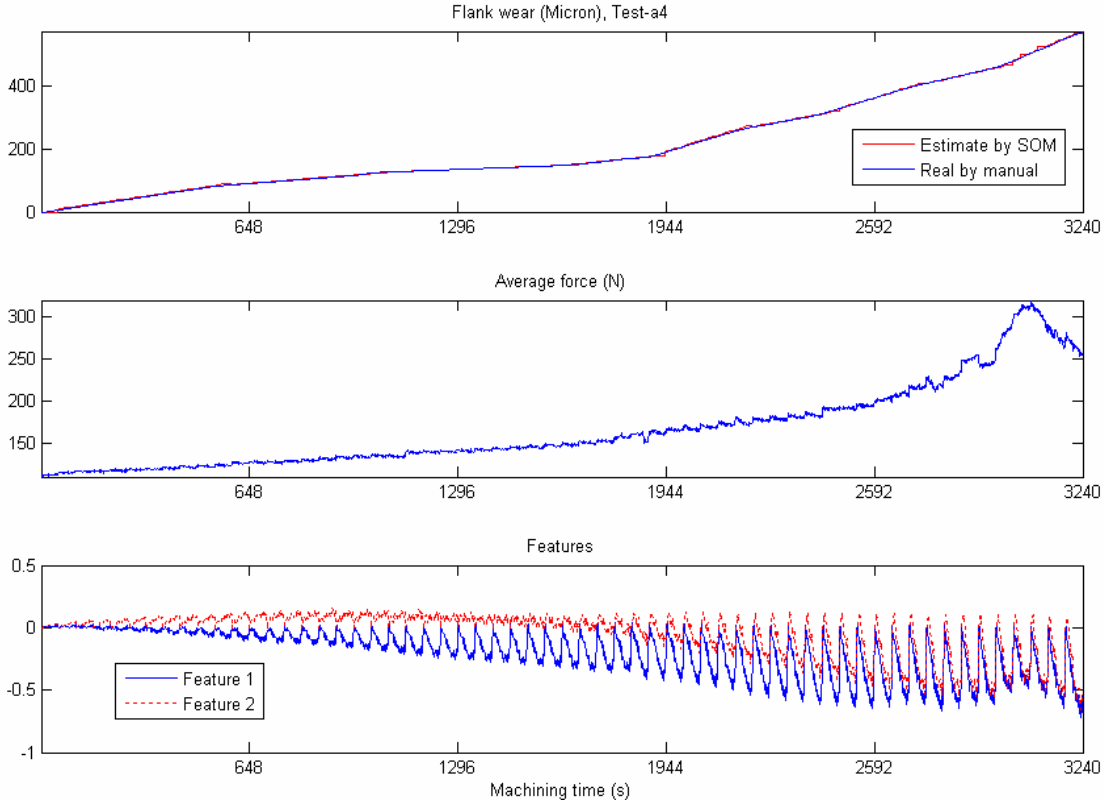


Figure 7.20 Wear estimation result for Test a4

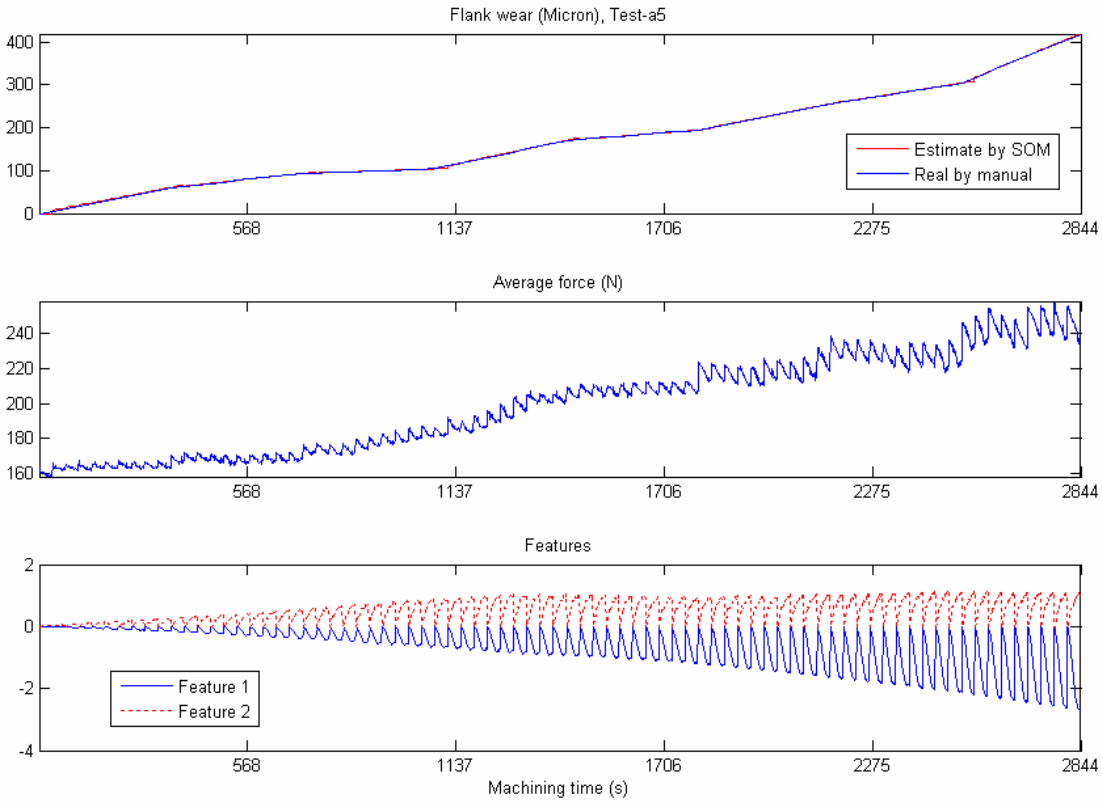


Figure 7.21 Wear estimation result for Test a5

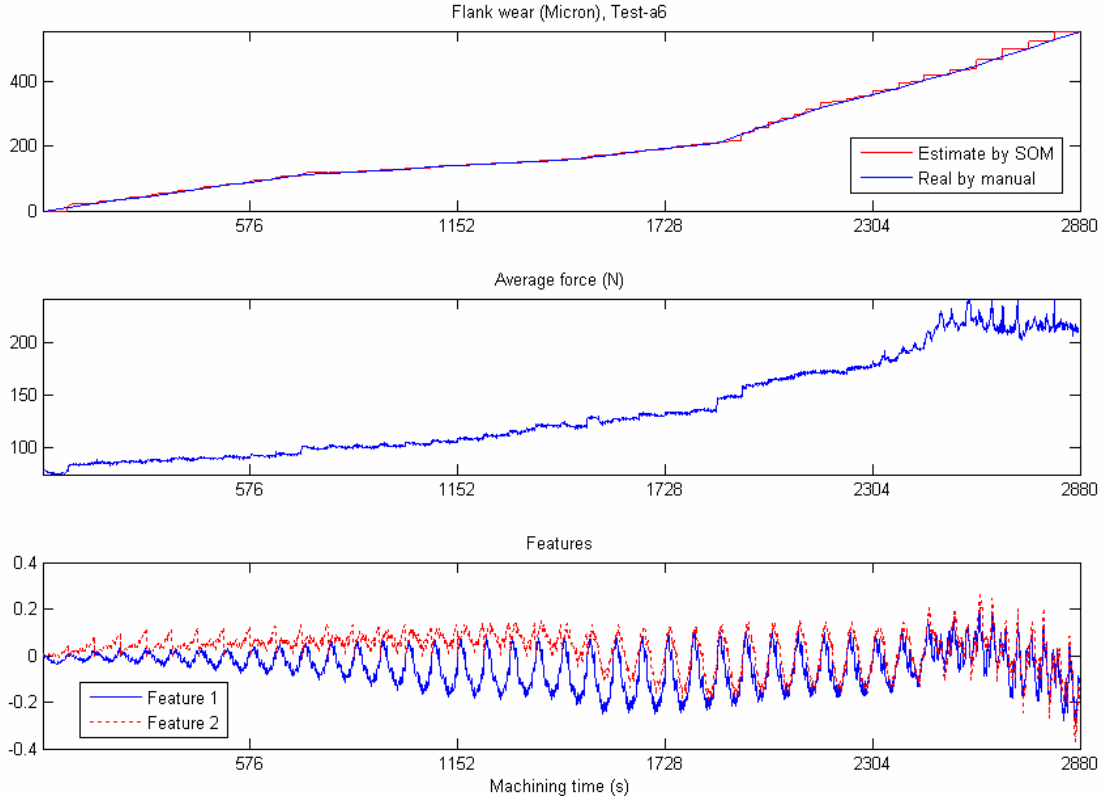


Figure 7.22 Wear estimation result for Test a6

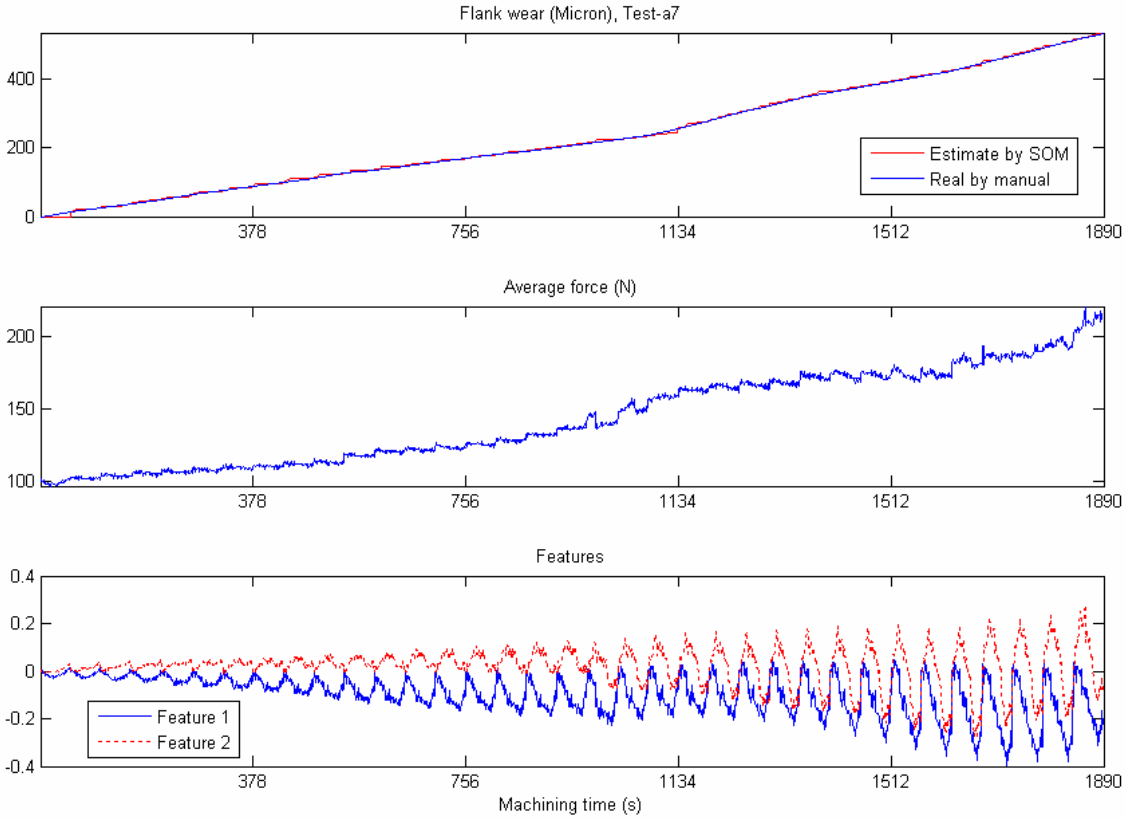


Figure 7.23 Wear estimation result for Test a7

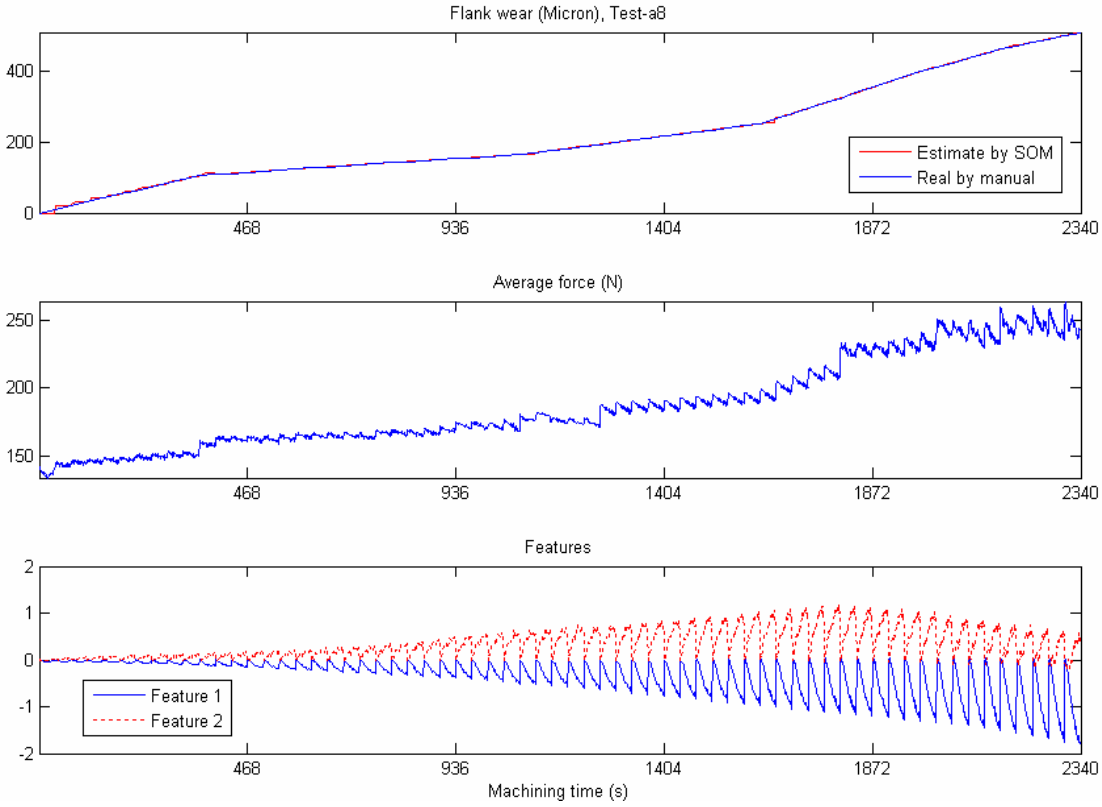


Figure 7.24 Wear estimation result for Test a8

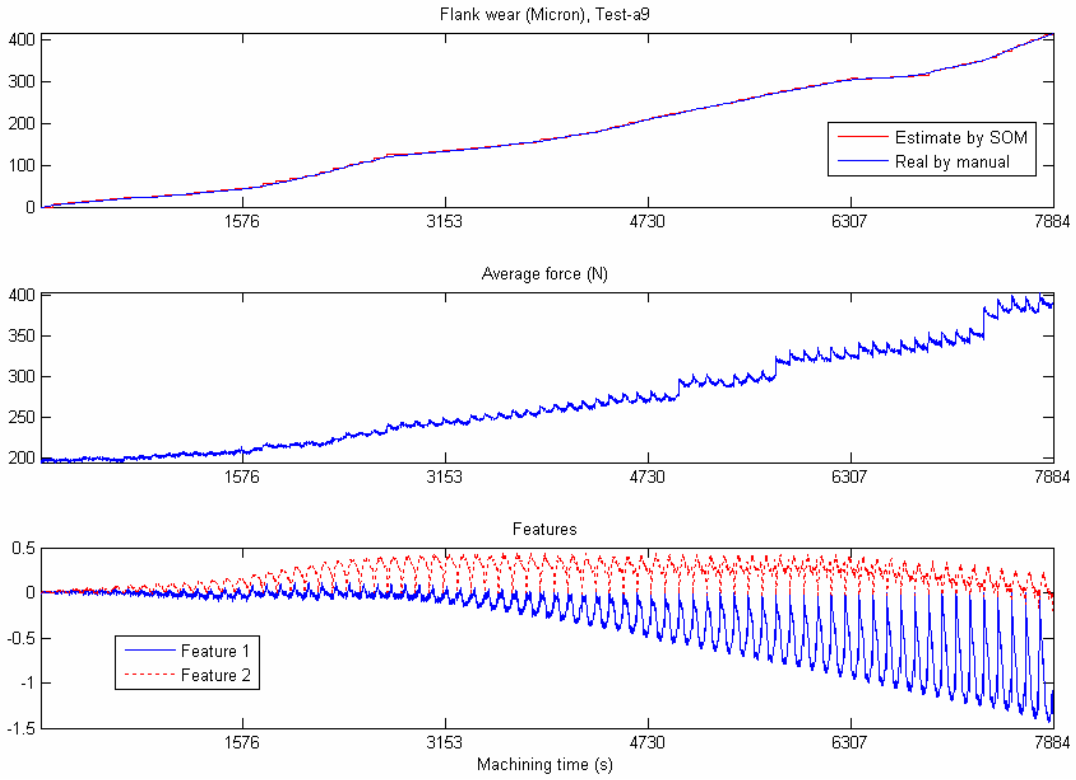


Figure 7.25 Wear estimation result for Test a9

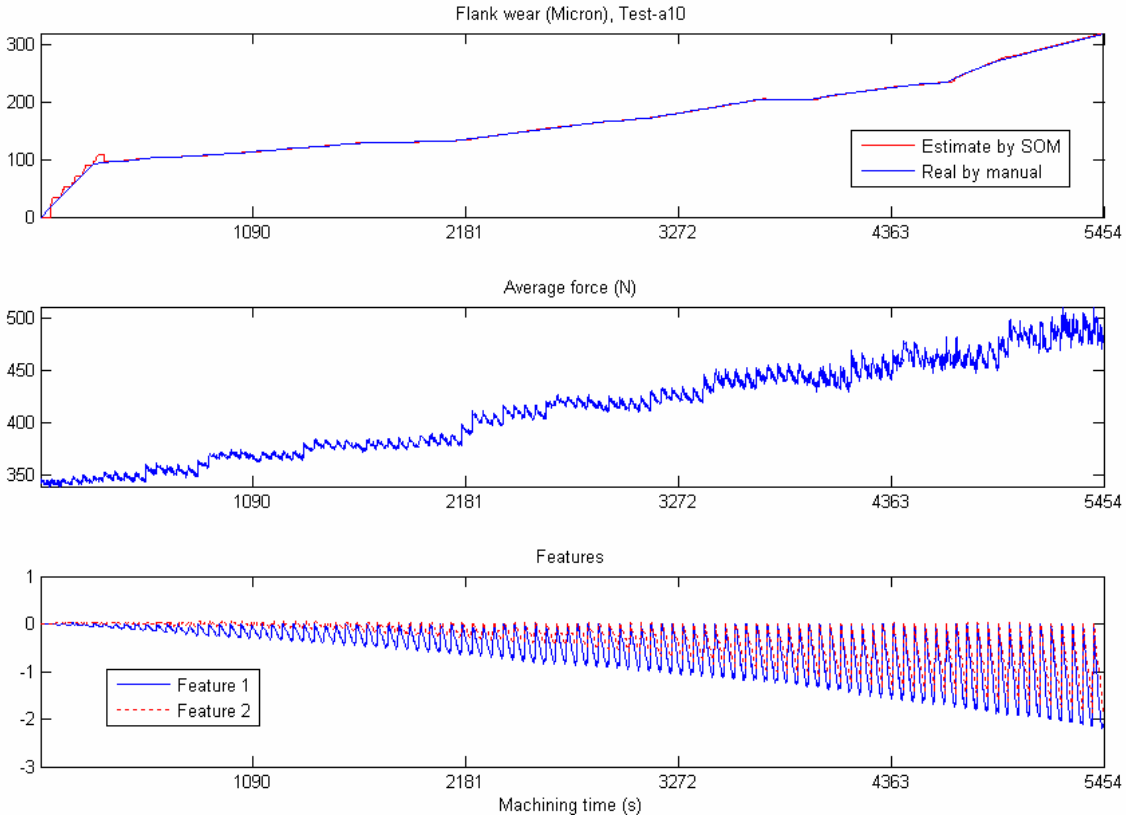


Figure 7.26 Wear estimation result for Test a10

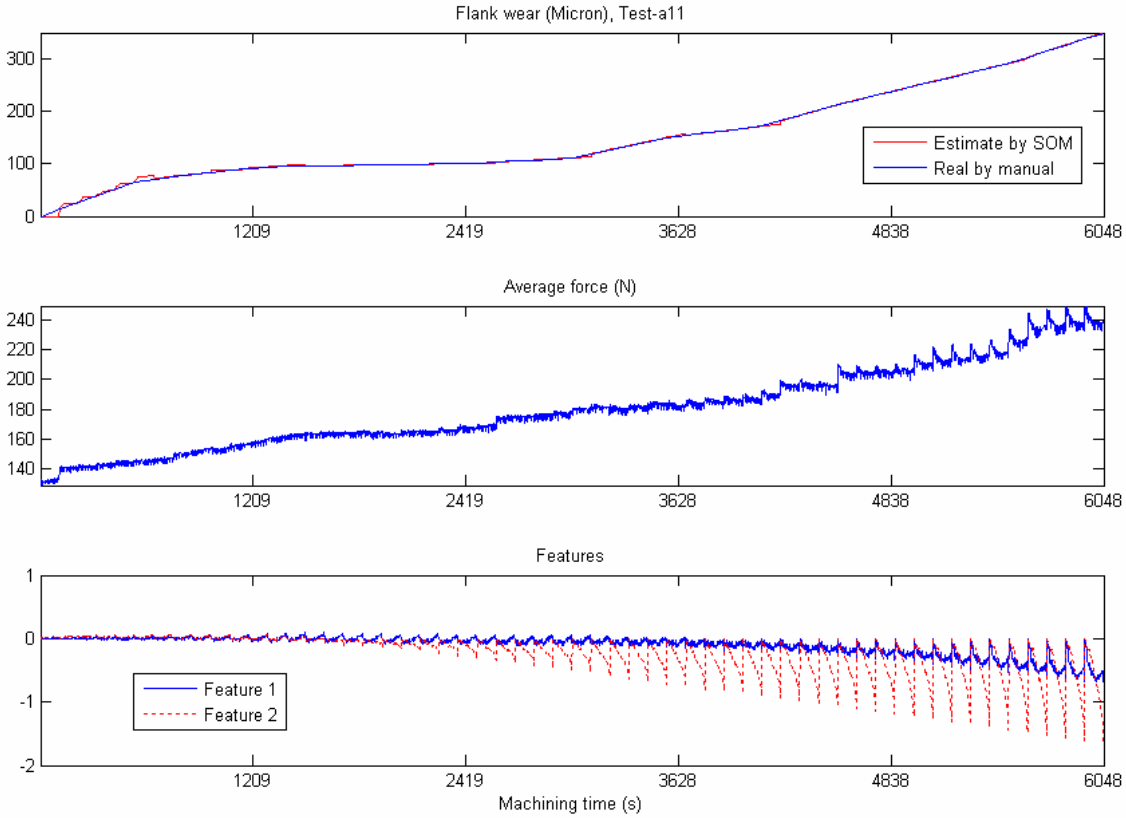


Figure 7.27 Wear estimation result for Test a11

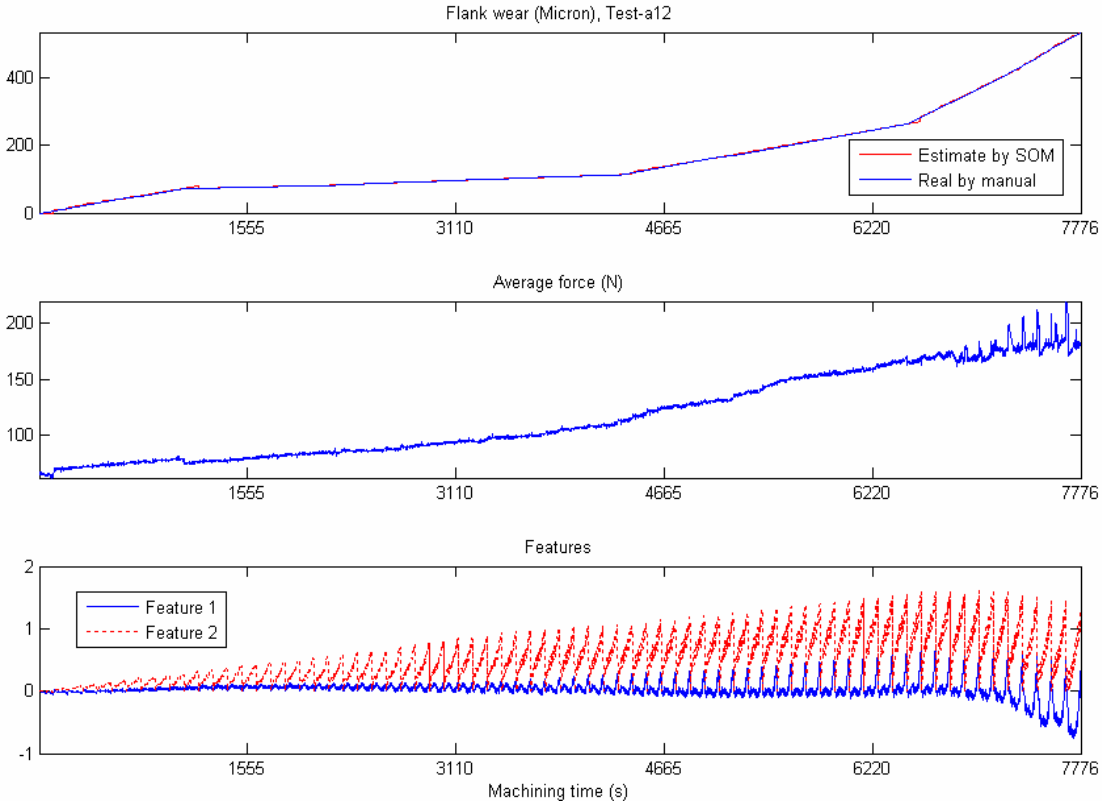


Figure 7.28 Wear estimation result for Test a12

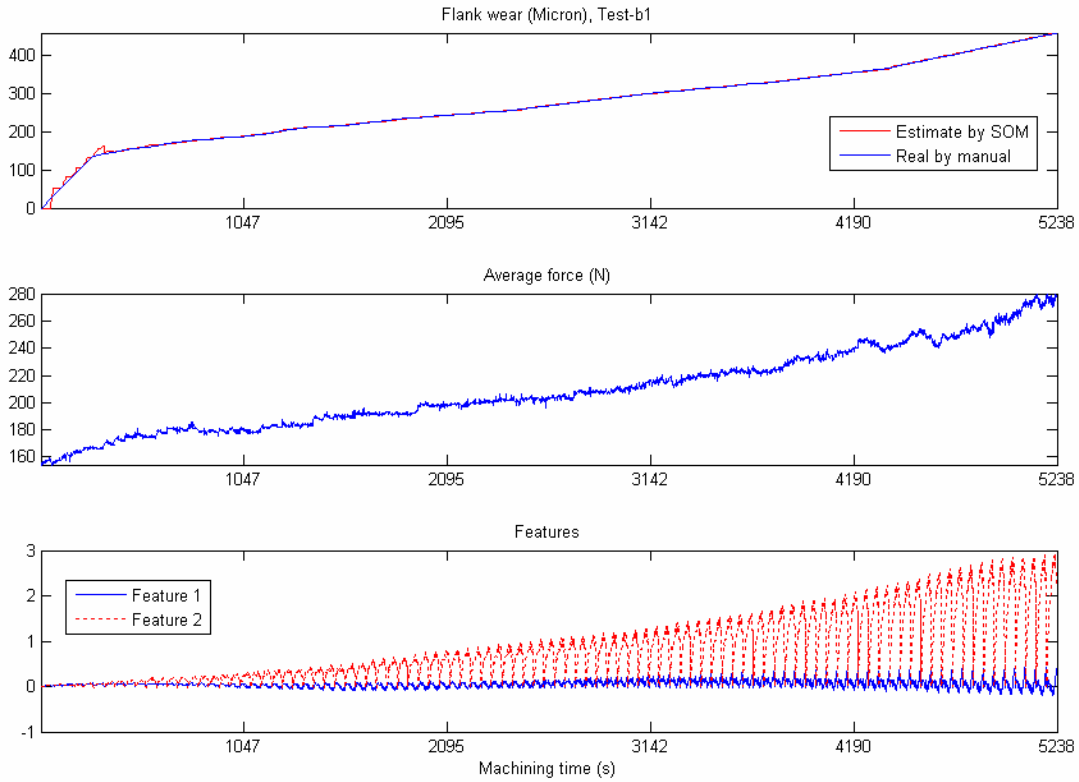


Figure 7.29 Wear estimation result for Test b1

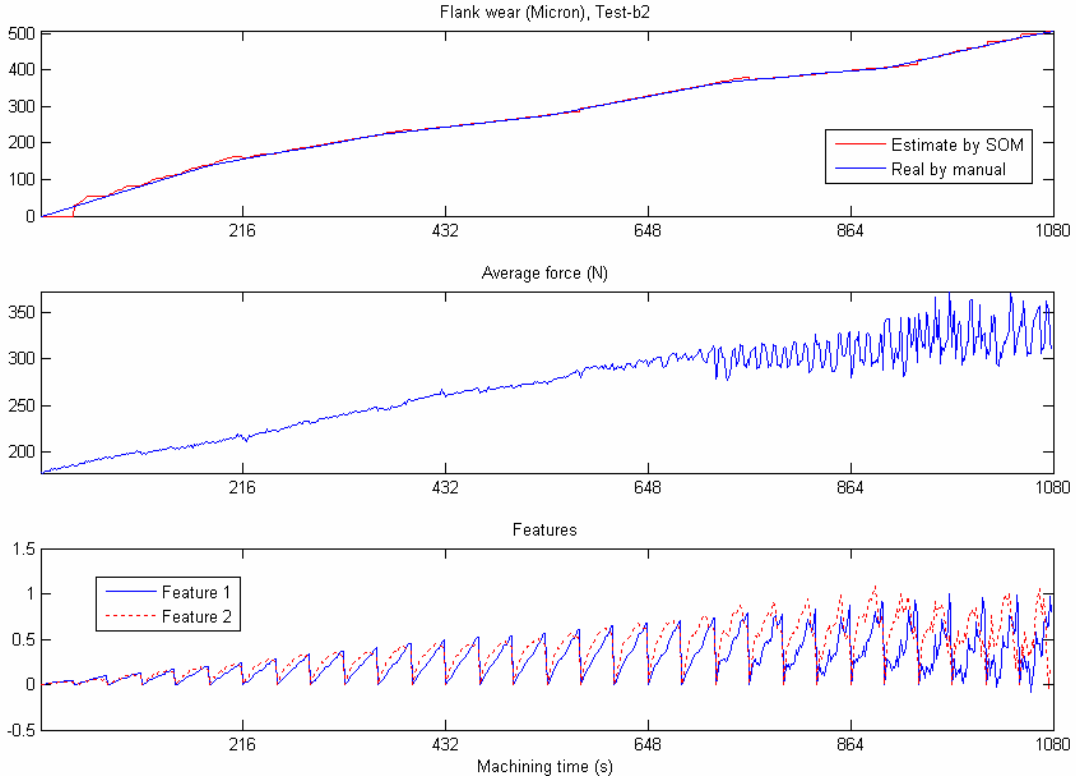


Figure 7.30 Wear estimation result for Test b2

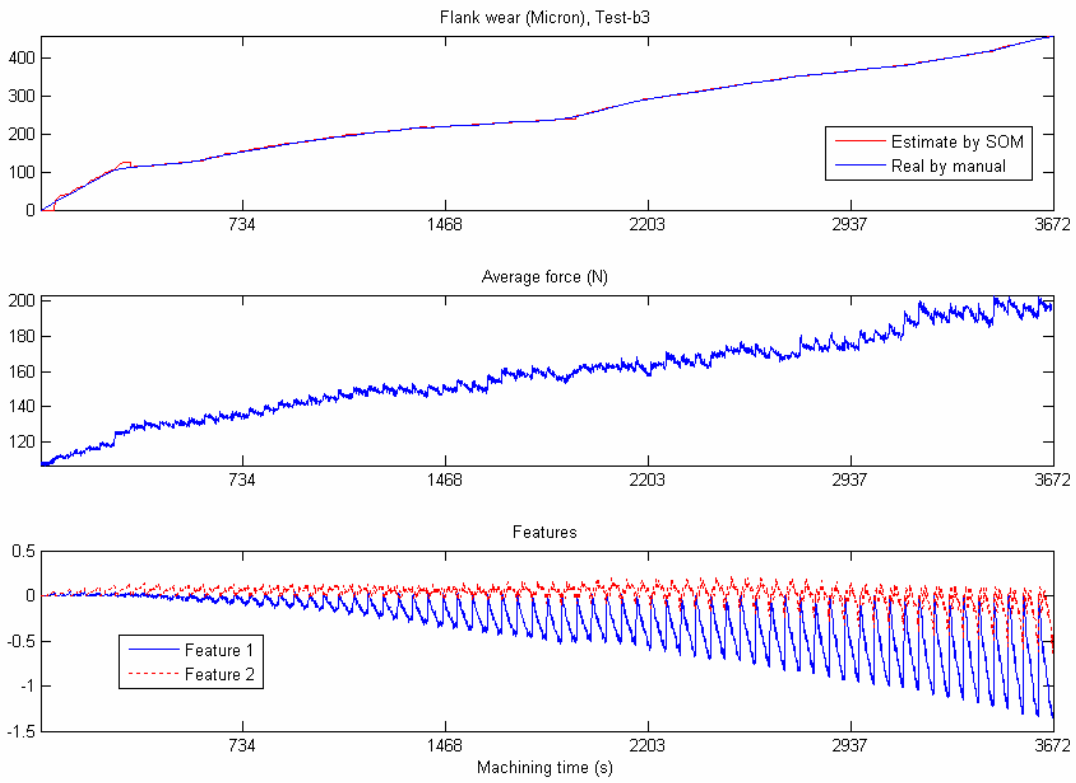


Figure 7.31 Wear estimation result for Test b3

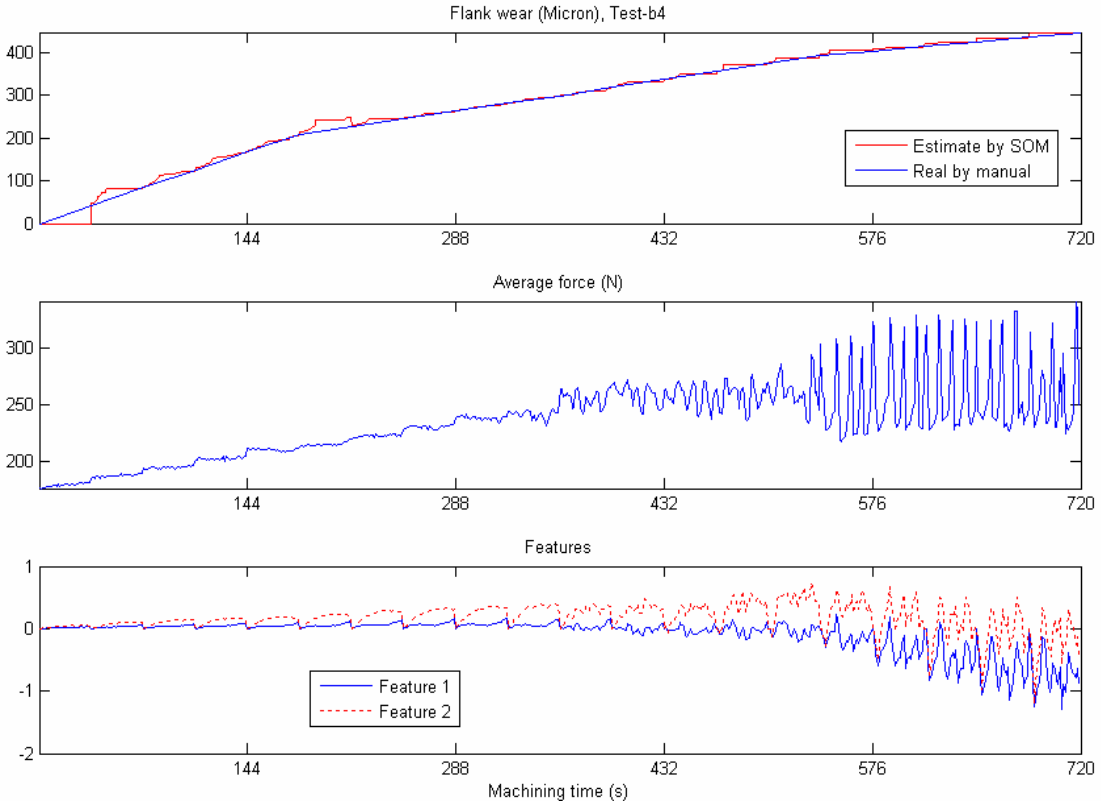


Figure 7.32 Wear estimation result for Test b4

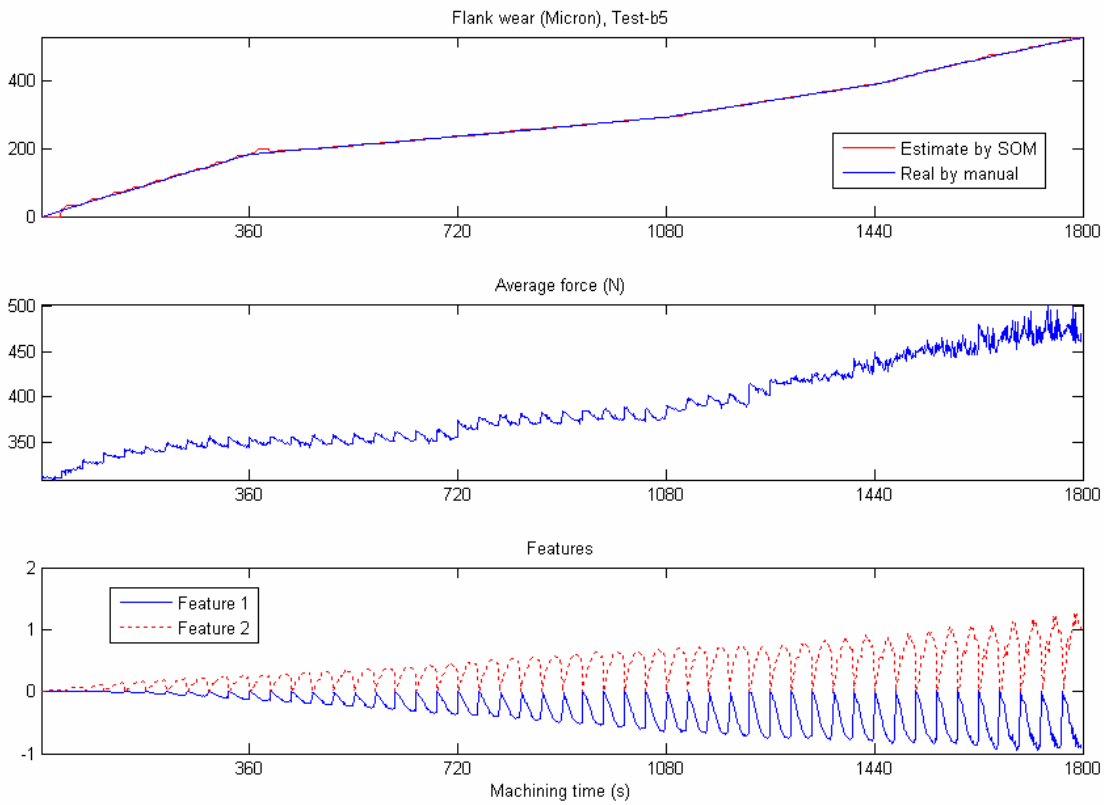


Figure 7.33 Wear estimation result for Test b5

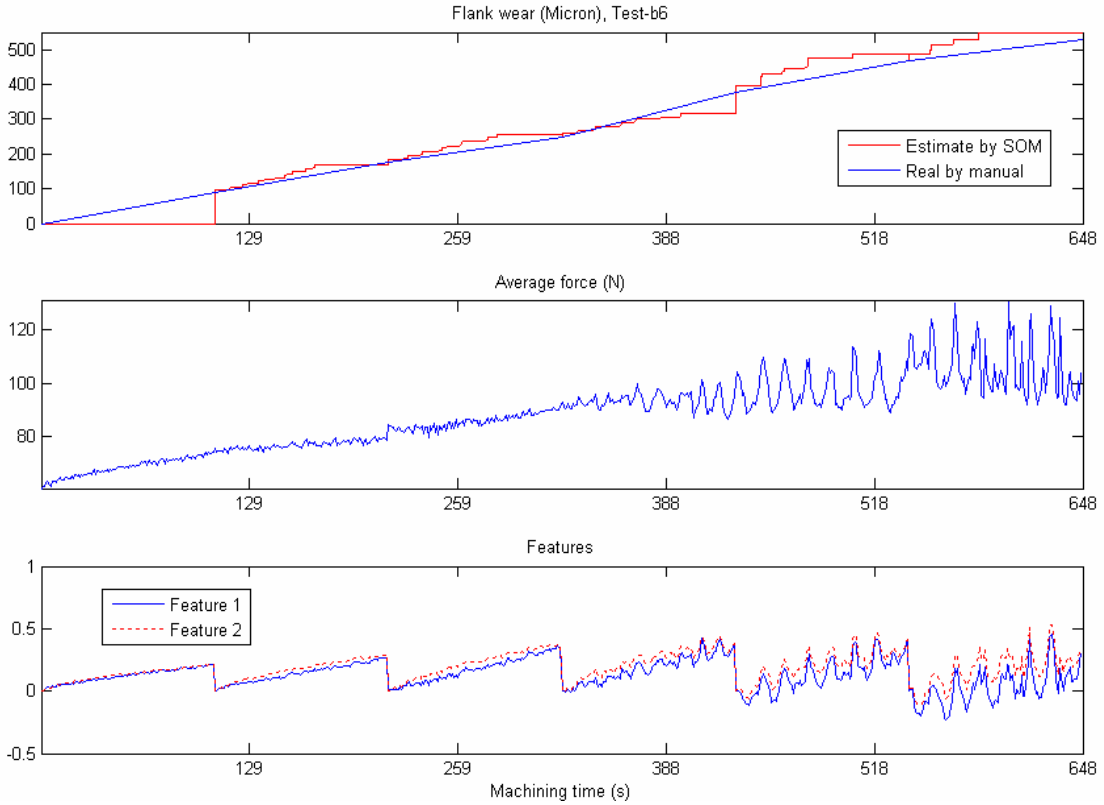


Figure 7.34 Wear estimation result for Test b6

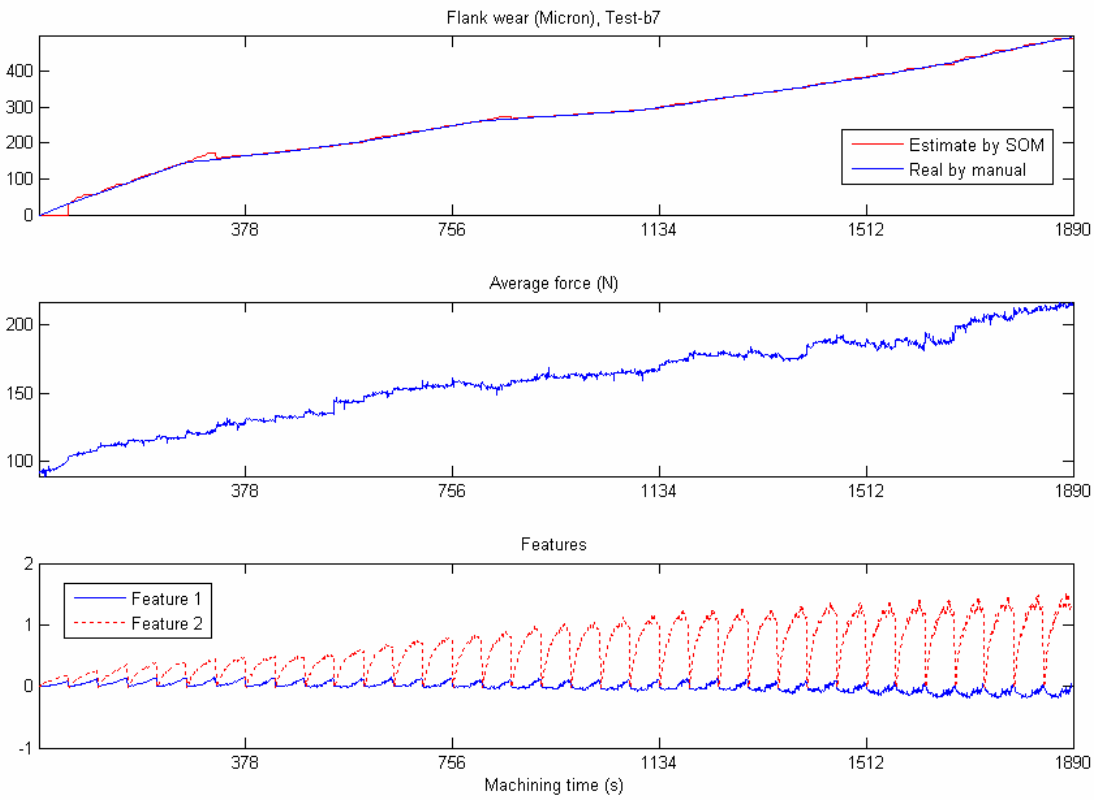


Figure 7.35 Wear estimation result for Test b7

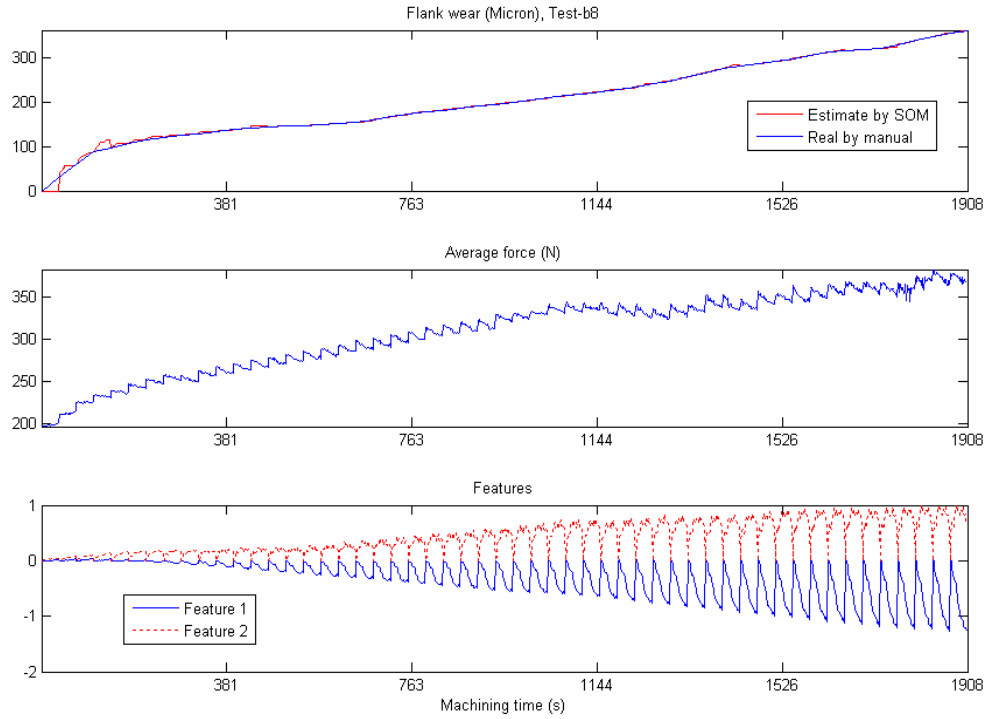


Figure 7.36 Wear estimation result for Test b8

Figures 7.17-7.36 strongly demonstrate that the wear estimation results track the actual values very well, especially for those tests wherein flank wear increases slowly

and stably, i.e., many passes in the test. This can be observed from most of the figures except Figures 7.32 and 7.34, which show the results for Tests b4 and b6. From Table 7.3, it was found that there were only 20 and 6 passes for these two tests, respectively. Compared to other tests, the tools in these two tests were worn out faster. This worn-out rate can be reflected through cutting force. Thus, it was observed that the two features in these two tests for successive passes had different patterns in terms of the amplitude and shape. Because the SOM functions locally, i.e., it is trained to approximate the wear flank increment in the current pass with the increment in the previous pass based on the two force features. When the increments in these two successive passes have really different values, say, the increment in the current pass exceeds that in the previous pass, the SOM still tracks the increment in the previous pass and cannot give a greater increment. In this case, the estimated result is less than the real. Reversely, if the increment in the current pass is less than that in the previous pass, the SOM is likely to give an equal or greater increment compared to the real, depending largely on the amplitude of the features. Nevertheless, the deviation in these tests is tolerable, say, even in the worst case Test b6 in Figure 7.34, the maximum deviation was about 30 μm at the end of pass 4.

It is also worth noting that the two features are not redundant. Most of the figures show that the patterns of the two features are different, including the amplitude and shape. It is therefore not a good idea to use only one of them in order to make the calculation simpler and faster. In addition, looking into the features in two successive passes (see Figure 7.34 for fine resolution in detail), it was found that they were quite similar in patterns. More importantly, even when the tool is approaching worn out, the average force has tremendous fluctuations, the features are stable (Figures 7.19, 7.22,

7.28, 7.30, 7.33 and 7.34 etc.). This property is useful for SOM to track the wear increment locally.

More discussion will be presented in the next chapter in connection with the on-line monitoring.

7.7 Concluding remarks

In this chapter, wear estimation and breakage detection methods based on cutting force have been presented. To estimate wear, SOM is employed, which is trained locally and applied immediately after the training. It is updated with the cutting pass. Features of cutting force in time domain, frequency domain and wavelet domain have been compared to show that the two features in time domain, i.e., residual error of average force and peak rate, can be effective to detect breakage. Experimental results demonstrate the ability of SOM as wear estimator to give accurate estimation.

The accurate estimation has been achieved because of two factors. One is the similarity and stability of the features extracted from cutting force, the other is that the manual wear measurement was used in the training. Recalling that the goal in this research is to replace the manual measurement by the proposed successive image analysis method, it is necessary to investigate if the estimation is still good in terms of accuracy when the replacement takes place. Therefore, in the next chapter, the proposed vision subsystem and force subsystem are put together to form a sensor integration system. The effectiveness of this integrated system in wear estimation and breakage detection is demonstrated.

Chapter 8
Experiment for on-line TCM

8.1 Experimental setup

Referring to Figure 3.1, the integrated on-line TCM system actually combines the vision subsystem and force subsystem into one collaborative system where the former functions in-cycle and the latter functions in-process. Therefore, as shown in Figure 8.1, the complete experimental setup is composed of the two subsystems, setup shown in Figure 5.1 together with that shown in Figure 7.17. The devices of each subsystem have been described in Chapters 5 and 7.

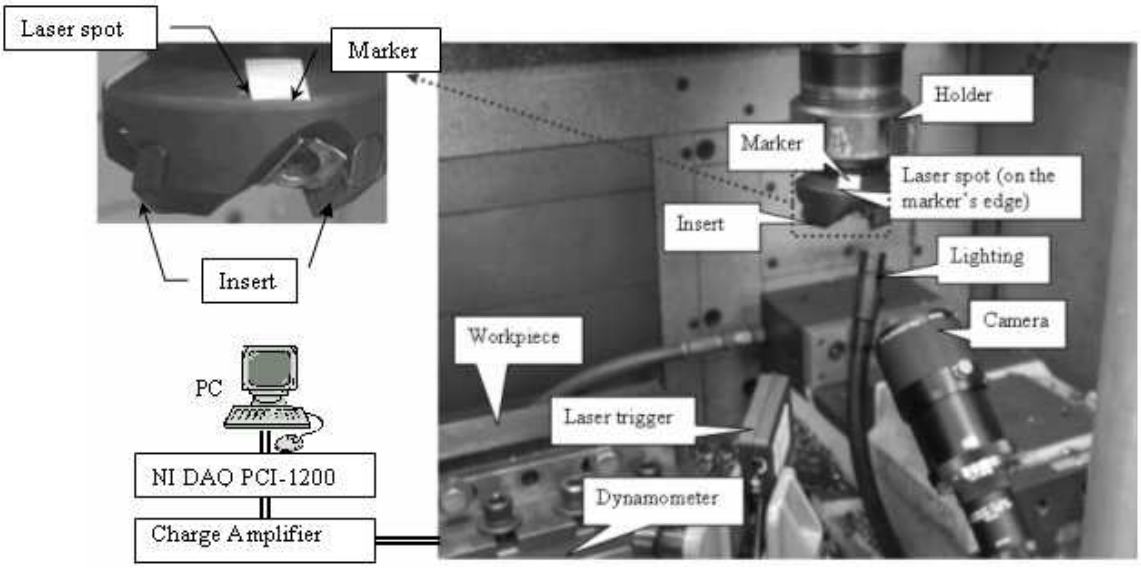


Figure 8.1 Experimental setup for on-line TCM

Before machining commences, images of the fresh inserts were captured while the spindle was rotating at a speed of 20 rpm. The current position of the tool was stored in the CNC machine controller as the *image capture point*. Images were processed to get the reference line and critical area for future use. When machining, the traverse force signal was low-pass filtered (1K Hz low-pass filter) and sampled at 2K Hz via the data

acquisition card PCI-1200. Four features of the cutting force were extracted in-cycle, namely, two features for breakage detection (residual error of average force, peak rate) and two features based on average force and standard deviation. The first two features were checked immediately after extraction with a preset threshold to detect breakage. According to the overall TCM strategy mentioned in Chapter 3, if breakage is detected, the machining is paused. The tool insert is programmed to park at the image capture point to be inspected by the vision subsystem. If breakage is verified by vision, the tool insert is considered to be broken and disposed. If vision does not detect breakage, the machining is resumed, and the thresholds for the two features will be increased some amount, say, 5%. However, in the experiment, in order to observe the force trend after a breakage event, machining was continued with the broken tool insert till the end of the ongoing machining pass.

After one machining pass, the tool was programmed to park at the image capture point and images of the inserts were captured while the spindle was rotating at a speed of 20 rpm. The wear value was thus obtained by successive image analysis. And the chipped-away material along the cutting edge was quantified to verify the breakage. If the breakage occurred, the machining was terminated. Otherwise, the SOM was trained in-cycle by using the wear increment value (measured in-cycle) and the extracted features (extracted in-process). After training, a new machining pass began. During this machining pass, the SOM used the two features extracted in-process to estimate the wear increment and hence, the estimated flank wear in the pass. The experiment was repeated in this manner till the insert was worn out or seriously broken. Note that SOM was trained locally with the latest data collected in-process and kept updating after each machining pass.

8.2 Experimental results

Experiment with different cutting conditions and two types of inserts was carried out. The parameters for the tests are shown in Table 8.1. For Test 1, because the tool was worn very slowly, to save material and time, it was not continued after pass 104. The experimental results for the six tests are shown in Figures 8.2-8.7.

Table 8.1 Parameters in dry milling for on-line TCM

Parameters	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Inserts (ISO SDKN42MT)	AC325, coated			A30N, uncoated		
Number of inserts	2					
Workpiece	ASSAB718HH steel					
Length of the workpiece (mm)	205					
Diameter of cutter (mm)	50					
Spindle speed (rpm) (m/min)	600 94.2	1000 157	1200 188.4	1000 157	800 125.6	1200 188.4
Feed rate (mm/min)	200	200	150	200	300	150
Feed per tooth (mm/tooth)	0.17	0.1	0.063	0.1	0.188	0.063
Depth of cut (mm)	1					
Time/pass (s)	54	54	72	54	36	72
Immersion rate	Full					
Number of machining passes	104	27	30	27	20	13
Conditions observed	Gradual wear	Breakage				

There was no breakage observed or detected in Test 1. So in Figure 8.2, only the wear estimation result is shown. For the other tests, breakage was detected in some pass and the image of the insert was inspected by vision. Therefore, in each of Figures 8.3-8.7, (a) shows the wear estimation result throughout the entire test (till the end of the pass listed in Table 8.1). (b) shows the breakage result inspected by vision, with gray-level image of the insert in left part, binary image in right-upper part and breakage in right-bottom part. (c) shows the force and force features, i.e., residual error

and peak rate against rotations, when breakage was detected in some pass by these two features (except Figure 8.3 for Test 2, no breakage detected).

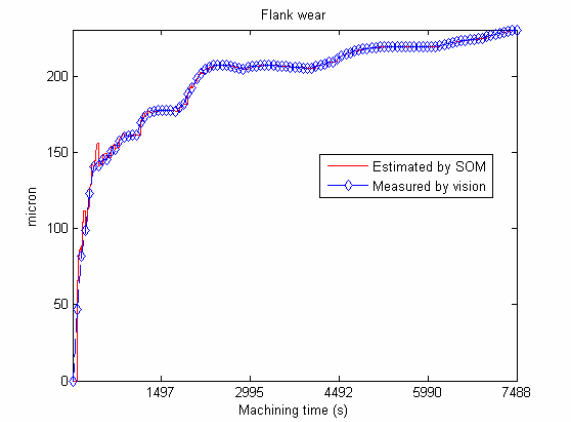
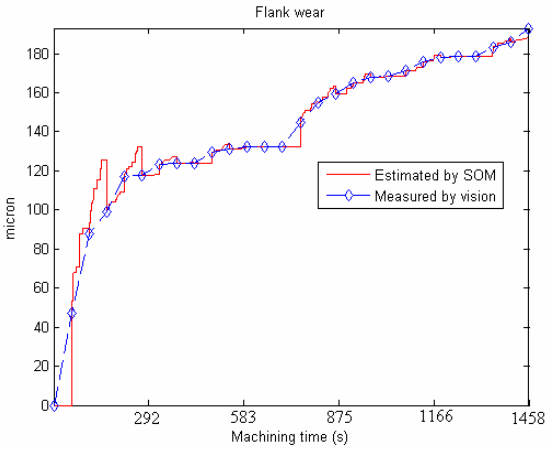
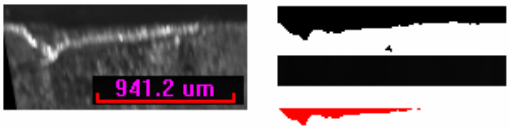


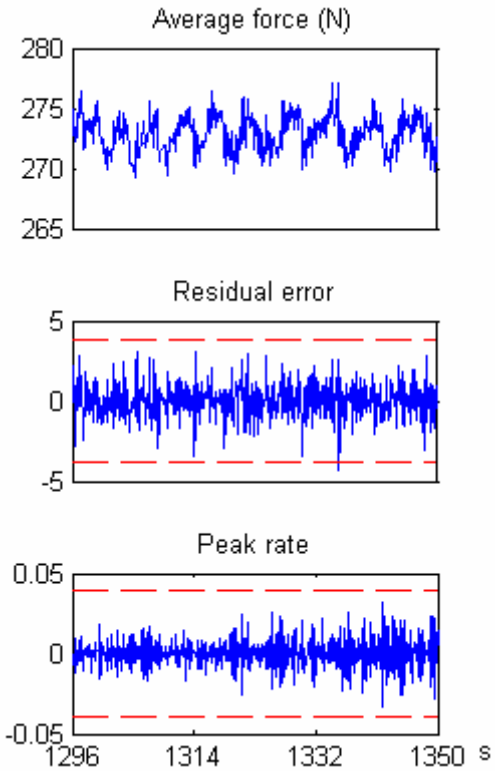
Figure 8.2 On-line TCM result for Test 1



(a) Wear estimation result

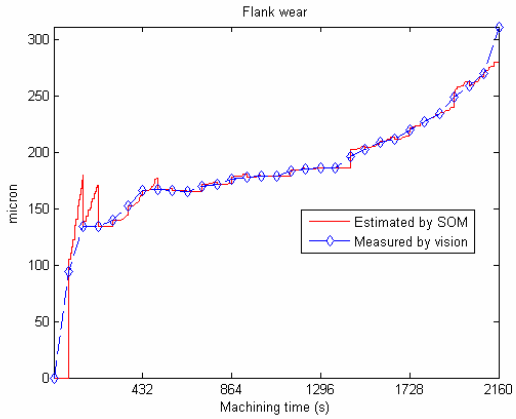


(b) Breakage by vision after pass 27



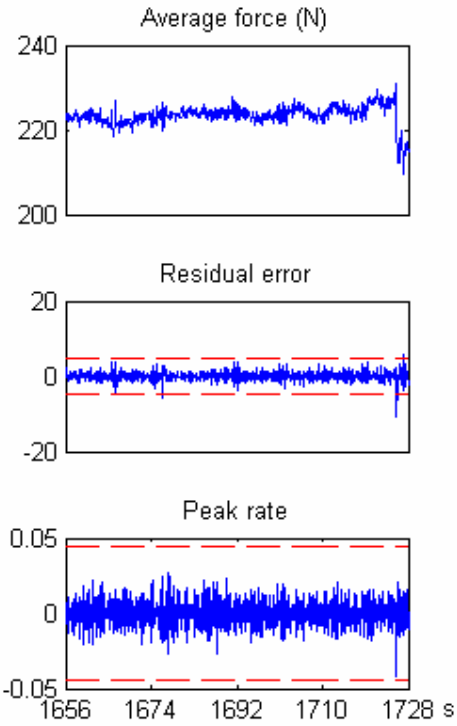
(c) Force features in pass 27

Figure 8.3 On-line TCM result for Test 2



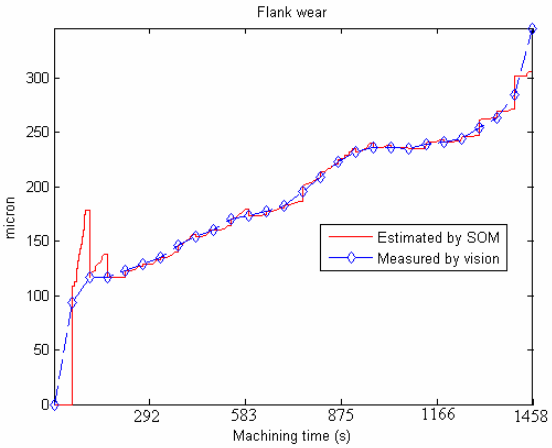
(a) Wear estimation result

(b) Breakage by vision after pass 24

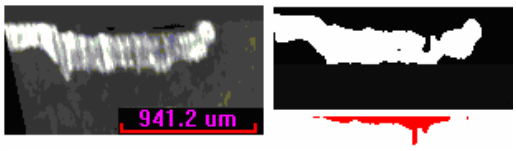


(c) Force features in pass 24

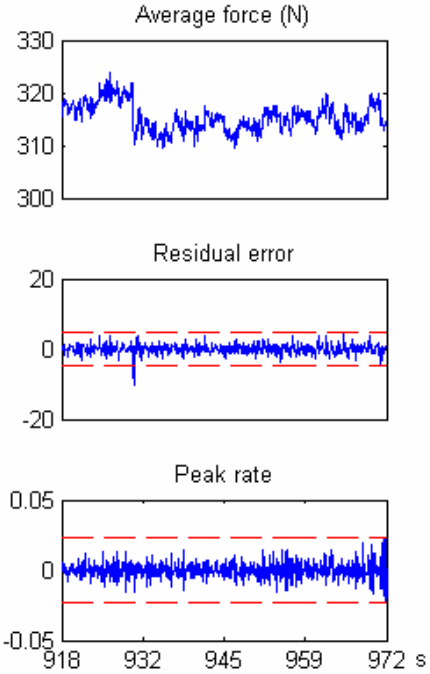
Figure 8.4 On-line TCM result for Test 3



(a) Wear estimation result

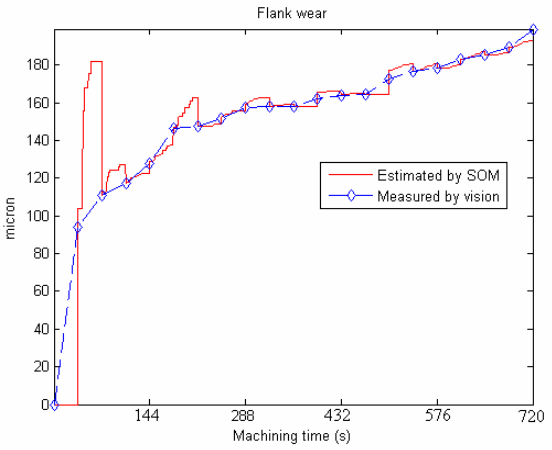


(b) Breakage by vision after pass 18

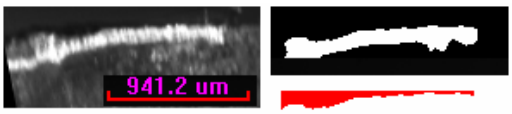


(c) Force features in pass 18

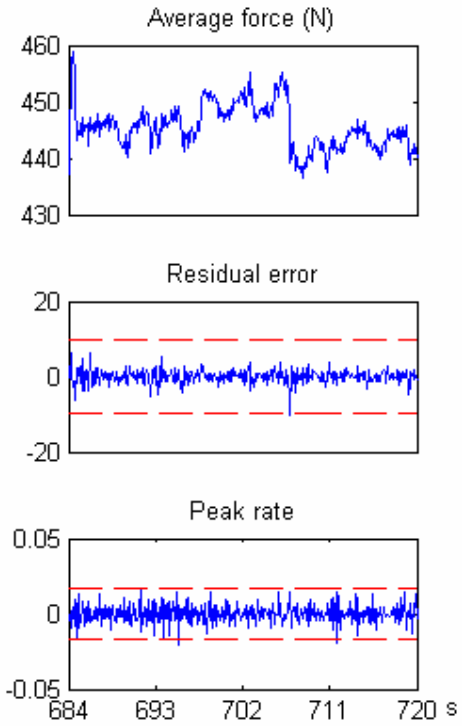
Figure 8.5 On-line TCM result for Test 4



(a) Wear estimation result

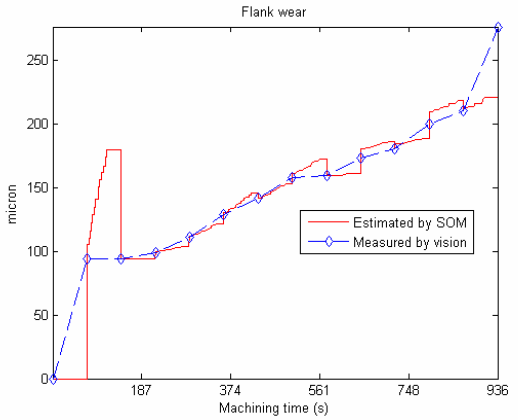


(b) Breakage by vision after pass 20

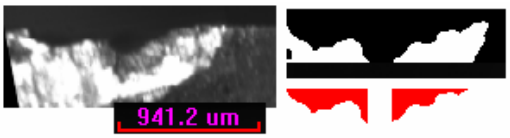


(c) Force features in pass 20

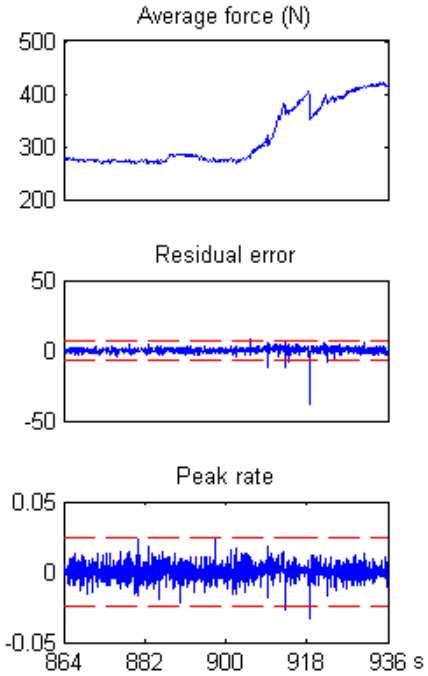
Figure 8.6 On-line TCM result for Test 5



(a) Wear estimation result



(b) Breakage by vision after pass 13



(c) Force features in pass 13

Figure 8.7 On-line TCM result for Test 6

Some important findings were observed from the figures for the six tests:

From all the figures for the six tests, it was observed that the flank wear estimation result is good in association with the in-cycle measurement by vision, especially at the linear wear stage, although there is great deviation at the initial wear stage.

Breakage is successfully inspected by vision. This can be seen from Figures 8.3 (b)-8.7 (b).

Breakage can be detected by force features using dynamic thresholding. Residual error and peak rate are two complementary features to detect breakage. When one feature fails to detect breakage, the other can be used to detect breakage. For example, in Tests 2 (Figure 8.3 (c)), 3 (Figure 8.4 (c)) and 4 (Figure 8.5 (c)), residual error better indicated breakage than peak rate; while in Test 6 (Figure 8.7 (c)), the opposite was true.

8.3 Discussion and summary

Before the first machining pass, no *a priori* knowledge of the data was assumed in the training of the SOM network. The wear estimation values were set to zero based on the pass. At the initial wear stage, the wear grows rapidly, and then the wear increases at a much slower rate at the linear wear stage. During this wear rate transition, the SOM gave an estimate with greater deviation from that by the vision measurement. In Test 6 (Figure 8.7 (a)), there was an even greater deviation up to 75 μm. The reason is that the SOM applied to a machining pass with a slow wear increment was trained with data from the previous machining pass with a more rapid wear rate. However, this deviation is tolerable since the tool is hardly worn out in this short period of time.

At the linear wear stage, the SOM tracked the wear value well. However, when the wear rate changed between two successive passes, i.e., wear in the current pass was

slower or faster than the previous pass, there was some resultant deviation. This deviation was small, about 15 μm and thus tolerable.

When there is tool breakage in machining, the estimated error becomes large. This can be explained by the rather different patterns in the force features extracted before and after breakage. Figure 8.8 shows the average force and features in pass 12 and 13 for Test 6, respectively, whereby in pass 13 there was breakage, as shown in Figure 8.7 (b).

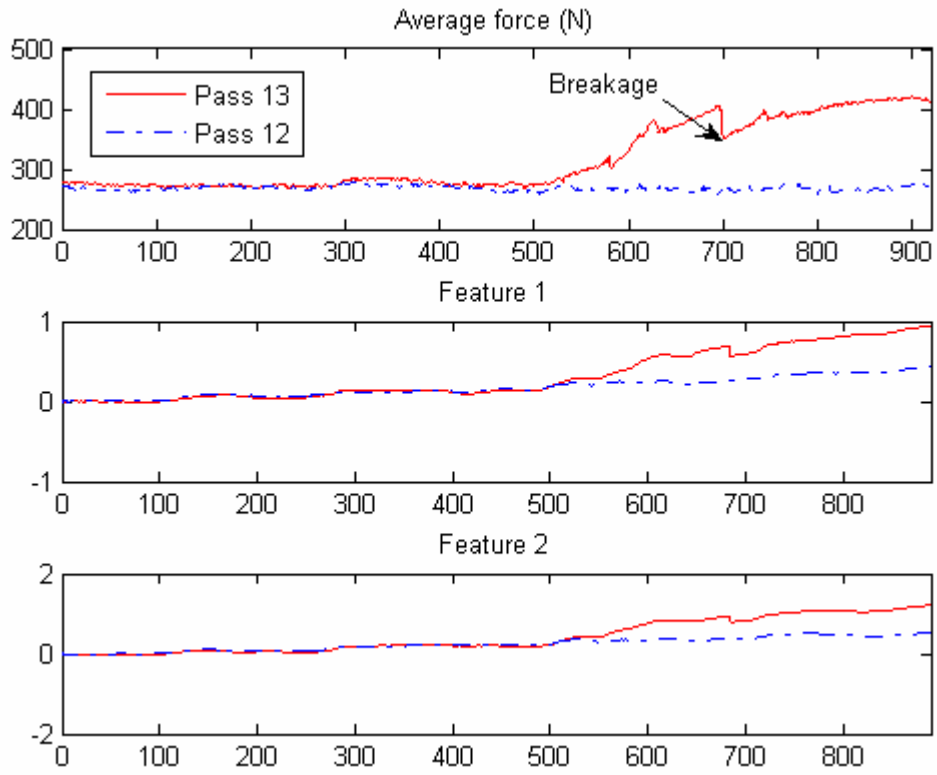


Figure 8.8 Average forces and features in two neighboring passes for Test 6

When the thresholds are too sensitive, it will lead to many cases of false breakage detection, i.e., many iterations of vision verification are required and thus much interruption of machining process. However, the proposed dynamic thresholds, determined by using the correlation between two successive passes are not too sensitive. In the above tests, it was observed that at only several points the thresholds were exceeded. The thresholds are accordingly reasonable.

On the PC Pentium III 300, only 2 s or so are taken to get the flank wear, to quantity the chipped-away zone and to finish training the SOM. The time can be reduced further given that special image processing hardware is designed to process these in-cycle tasks. Additionally, the in-cycle processing can concurrently proceed as the tool holder resets to its starting point for a new machining pass. In this way, the in-cycle processing will not, or slightly, be interrupted during the normal cutting process.

The comparison of the proposed sensor fusion method and the methods reviewed in Chapter 2 (section 2.3.2) is shown in Table 8.2. Method 1 is based on turning, thus the linear force-flank wear model may not be applicable to milling. Furthermore, its image processing module is too simple to be robust in industrial environment. Method 2 does not provide on-line wear estimation with vibration signal. In addition, using SOM to segment the image needs manual selection of teaching sample pixels. Use of the RBF network in Method 3 means extensive experiment over wide range of cutting conditions has to be carried out in order to collect the training data. The proposed method, however, uses updated wear value and force features to estimate the wear on-line. And the vision system can provide robust and fast wear measurement with good accuracy by successive moving image analysis while the spindle does not stop. Breakage detection is also achieved by vision and force features in the method.

In summary, the on-line tests have proven the effectiveness of the proposed sensor fusion strategy in flank wear estimation and breakage detection in milling process. The features of this method are:

- Sensor fusion, i.e., integration of vision and force sensors, is used.
- Force features, namely, residual error and peak rate, are used to detect breakage in-process.

- In-cycle vision measurement based on successive image analysis is accurate, robust, and fast, providing necessary reliable knowledge for the SOM estimator. Breakage can be verified successfully.
- SOM network is used to estimate flank wear in-process with the following merits:

  a). Only two features of the cutting force required, which can thus be implemented easily and quickly, especially favorable for on-line implementation;

  b). Unsupervised batch training, and hence easier data collection and less computation time;

  c). Cutting condition independent. SOM works in a repetitively updating mode: trained locally and applied immediately. It is thus applicable for various cutting conditions in the sense that there is no need to train by collecting lots of force samples under different conditions, as opposed to traditional neural networks.

Table 8.2 Comparison of TCM systems using indirect sensor(s) and vision

Methods	Process/ condition	Indirect signal and processing	Image processing	Fusion method
1 (Park and Ulsoy, 1993a,b)	Turning/ wear	Force, an adaptive observer on wear-force relationship	Segmentation by thresholding	Vision system provides parameters for the observer and calibrates the observer
2 (Bahr et al., 1997)	Turning/ wear and breakage	Vibration, moving average of the root mean square (MARMS)	Segmentation by SOM	Vision system measures wear between cuts, vibration system gives breakage or unforeseen situation. Wear value is not given on-line
3 (Mannan et al., 2000)	Turning/ wear classification	Sound, energy of two wavelet coefficients	Column projection and run length	Both features from sound and vision are fed to a RBF neural network to classify the degree of tool wear
This	Milling/ wear and breakage	Force, processed average force and standard deviation for wear estimation, residual error and peak rate for breakage	Moment-based edge detection, threshold-independent, and spindle rotates	Vision system provides wear value for the SOM network to estimate wear on-line with force features. SOM is updated with the wear value repeatedly. Breakage is detected via force features and verified by vision

Chapter 9
Conclusions and recommendations for future work

9.1 Conclusions

In this research, one direct sensor (vision) and one indirect sensor (force) are used to create an intelligent integrated TCM system for on-line monitoring of flank wear and breakage in milling. Two subsystems, namely, a vision subsystem and a force subsystem have been developed to work in-cycle and in-process respectively. Experiments on both the subsystems and the integrated system have been conducted to verify the integration scheme. A full-field optical method based on phase-shifting has been proposed to measure crater wear,.

The research reported in this thesis has made the following contributions:

1. An effective individual image processing method for measuring flank wear and detecting breakage.

In this proposed method, the image captured when the spindle stands still is processed with sub-pixel accuracy. Rough-to-fine strategy is employed to locate the points on the boundary of the wear land in two steps. The binary edge image is firstly used to roughly locate the boundary point. The gray-level image is then used to locate the boundary point precisely in the vicinity of the rough point with a moment-invariance edge detection method. After trimming the image to a critical area to speed up the calculation, the reference line is detected via Hough transform. To enhance the robustness, noise is reduced by median filter, morphology and moving window techniques. To make the method less sensitive to light, a thresholding-based

segmentation in detecting the boundary point is avoided. The off-line measurement results of sample inserts were accurate and breakage was detected successfully.

However, the performance is still affected by the orientation and intensity of light. The same problem exists in the optical systems proposed by Jeon and Kim (1988), Pederson (1989) and Ogumanam et al. (1994). Generally, the orientation and intensity of light are deliberately adjusted to make the three areas distinguishable to the camera in order for this method to be effective. With the adjustment, the image of the flank wear has three areas with different gray levels: low level for background, intermediate level for the unworn area of the cutter, and high level for the wear land. Deliberate adjustment of the orientation and intensity is rather troublesome and cumbersome. This is, however, an inevitable requirement for any optical method with visible light. It cannot be eliminated completely. In this method threshold selection is avoided in detecting the point on the wear land boundary. Furthermore, although it needs time and effort to make the initial adjustment ready, once this is done, the whole experimental setup can be used for all inserts of the same type since the relative geometric relationship between the camera and the insert is fixed.

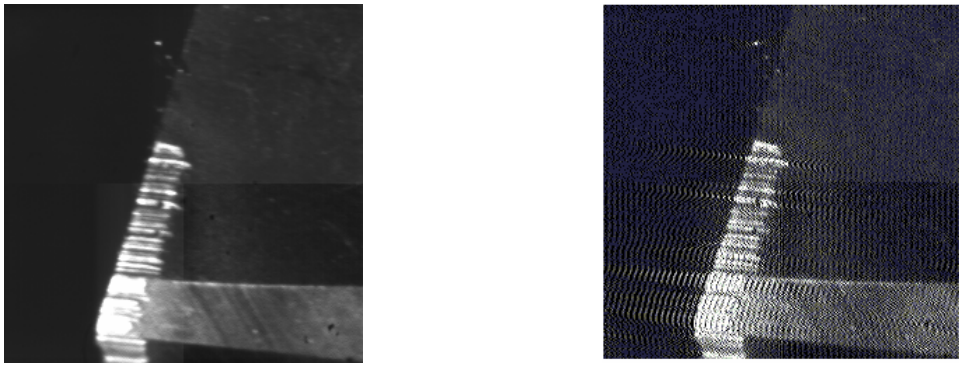
There is another problem with the individual image processing method. Given different parameters, the method may give different measurement results. In some cases, say, for the coated inserts, the measurement results have a large deviation. The main reason is that there is no *a priori* knowledge given to the method regarding the wear since there is only one individual image to be processed. However, in the successive image analysis method, since a series of images with close correlation are involved, a newly captured image can be processed with the knowledge obtained already from its neighboring image captured earlier. Therefore, the effect of parameters on the results, especially on the consistency of the results, is reduced.

2. A more robust and effective successive image analysis method based on processing moving images captured while the spindle rotates.

Based on the aforementioned individual image processing method, the successive image analysis method has been developed to capture and process the moving images when the spindle rotates. A trigger-capture mechanism is introduced in response to the spindle rotation. This mechanism ensures that the same insert appears at the same location in its image series. It also helps to control the integration time of the camera to reduce blur imposed on the image due to the rotation of the spindle. The hardware setup, therefore, leads to close correlation between successive images. In view of this, the critical area can be expanded dynamically as wear progresses. And the reference line and part information of the critical area extracted from the reference image can be reused for all subsequent images. This improves the accuracy, robustness and speed of the method. To reduce noise for these *in situ* images (opposed to images in off-line measurement), parallel scanning is proposed. It has been shown that this method is better than its predecessor (i.e., the earlier individual image processing method) in terms of accuracy, robustness, consistency and speed.

Nevertheless, it is noteworthy that this method with the present setup cannot be directly used in the industry. Firstly, the image is captured at a very low spindle speed (only 20 rpm) because of the limitation of the experimental devices. Despite such a low speed, there exists some blur resulting from the rotation in the image. If the spindle speed is increased, with the same experimental setup in this case, the blur will be severe such that the wear land is very distorted and undetectable. In this case, deblurring is a must. Typically, there are software algorithms to do deblurring, say, by convolving with some deblurring function designed in connection with the rotation

speed and orientation. Figure 9.1 shows a moving image (as shown in Figure 5.2 (b)) and its deblurred counterpart by a Wiener filter.



(a) Moving image (b) Deblurred image

Figure 9.1 Deblurring result

It was found that the deblurred image suffered from ringing effect, which made the deblurred image unacceptable. It is therefore better to employ appropriate hardware to reduce the blur. This can be done in two ways. One way is to use strobe light that can provide strong illumination in a short period of integration time of the camera. The other way is to use a camera with higher speed coupled with a normal light (non-strobe) that can provide more intense lighting.

Secondly, the experimental procedure is a little more complex. But like the adjustment of the orientation and intensity of the light, after the setup for one insert is done, it can be used for all inserts of the same type. More importantly, if the developed system is integrated into the CNC machine, the coordinates of the camera, light and image capture point can be easily obtained, adjusted and stored. Hence, the whole system can be used easily.

Thirdly, the commercial devices used in the experiments are not compact and so need much space. This may hamper them from being integrated into the CNC machine. However, as electronic technologies progress rapidly, it is possible to design more compact cameras and light systems at low cost.

3. An effective and robust 3-D crater wear by phase-shifting method.

Four images of the rake face with different fringes projected on are analyzed to give the phase map, which is converted to a 3-D map of crater wear after calibration. This is a full-field optical non-contact method. Its performance has been found to be independent of light intensity, unlike other methods based on visible light. Therefore, it is effective and robust.

As the devices for phase-shifting are not available for experimental use on the CNC machine, the crater wear is measured off-line rather than on-line. For off-line measurement, the current method is good. However, for on-line measurement, the method has two limitations. One is that the algorithm needs a long time to calculate the 3-D map of the crater wear (in the order of minutes on PIII PC). Special image processing hardware can be used to speed up the calculation. The other limitation is that the devices need much space, similar to that for the flank wear measurement system.

4. An effective flank wear estimation via SOM.

The estimation is carried out on a pass-by-pass basis through an unsupervised SOM network. Two features sensitive to flank wear are extracted in time domain from the cutting force. The features and wear increment in the previous pass are used to train the SOM. After the training, the SOM is applied immediately to the succeeding pass, estimating the flank wear increment in this pass. Both off-line and on-line results show that this method is successful, especially at wear linear stage. Training without cutting conditions makes this method adaptive to various cutting conditions. The use of SOM avoids troublesome data collection and tedious off-line training.

However, flank wear estimation in a pass-by-pass manner limits its feasibility in the industry. In the experiment, throughout the test, the vision subsystems works

periodically, i.e., at the same frequency no matter how serious the wear is. This may be a problem for industrial case. For example, when the wear becomes more serious, the condition is likely to be monitored more frequently by vision. It is easy to change the monitoring frequency in this approach. However, there is no optimum strategy to change the frequency. The design of such an optimum strategy should be explored in future work.

5. Breakage detection via two force features in time domain and verified by vision.

Features in time domain, frequency domain and wavelet domain are compared to identify their effectiveness in detecting breakage. It was found that features in time domain, i.e., residual error and peak rate are sensitive to breakage. Breakage or non-breakage can be verified by vision successfully.

In the detection of breakage, the thresholds are set dynamically with the correlation of two successive passes. This threshold determination method is reasonable and objective.

Collectively speaking, the TCM system based on intelligent sensor integration of vision and force has potential to be used to monitor flank wear and breakage in milling process in the industry.

9.2 Recommendations for future work

In the previous section of Conclusions, some limitations are mentioned. In association with these limitations, recommendations for future work are given below:

1. Hardware setup for capturing images when the spindle rotates at higher speed should be considered as part of the whole CNC machine system. The combination of high-speed camera and light (either a strobe light or a normal light) should be considered carefully in terms of their working temperature,

cost and working space. With a higher spindle speed, it is also necessary to control the laser trigger signal precisely so that the reference line in successive images has tolerable or no shift geometrically. The hardware should be miniaturized in size to be feasible for integration into the CNC machine.

2. Crater wear should be considered in on-line monitoring. Firstly, like the successive image analysis method, it should be studied whether the crater wear could be measured when the spindle rotates via trigger-capture mechanism. Then from this point, the measurement devices for crater wear should be miniaturized to be integrated in the CNC machine.
3. New features from force or other sensing signals should be extracted to reflect the change of wear rate and used as a clue to design an optimum control strategy for the frequency with which the vision subsystem works. From this aspect, besides force, other signals may be considered, for example, vibration or current. The wear rate given by vision can also be used in this regard.
4. The current TCM system is dealing with cases where the machining is very simple (rectangular workpiece milling). For complex shapes to be made, the TCM system may not function properly. It should be further investigated.
5. There is no a quantitative description about the estimation error in using the SOM as in-process flank wear estimator. How to quantify the error and use it to improve the estimation performance should be studied.

References

Abu-Zahra, N. H. and Nayfeh, T. H. Analytical model for tool wear monitoring in turning operations using ultrasound waves, Int. J. Mach. Tools Manufact., 40 (11), pp. 1619-1635. 2000.

Altintas, Y. In-process detection of tool breakage using time series monitoring of cutting forces, Int. J. Mach. Tools Manufact., 28 (2), pp. 157-172. 1988.

Altintas, Y. and Yellowley, I. In process detection of tool failure in milling using cutting force models, ASME J. of Eng. for Industry, 111, pp. 149-157. 1989.

Bahr, B., Motavalli, S. and Arfi, T. Sensor fusion for monitoring machine tool condition, Int. J. Computer Integrated Manufacturing, 10 (5), pp. 314-323. 1997.

Ber, A. and Friedman, M. Y. On the mechanism of flank wear in carbide tools, Annals of the CIRP, 15, pp. 211-216, 1976.

Bertok, P., Takata, S., Matsushima, J., Ootsuka, J. and Sata, T. A system for monitoring the machining operation by referring to a predicted cutting torque pattern, Annals of the CIRP, 32 (1), pp. 439-443. 1983.

Blahusch, G., Eckstein, W., Steger, C. and Lanser, S. Algorithms and evaluation of a high precision tool measurement system, 5th International Conference on Quality Control by Artificial Vision, 1999, pp. 31 - 36.

Burke, L. I. and Rangwala, S. Tool condition monitoring in metal cutting: a neural network approach, J. Intell. Manufact., 2 (5), pp. 269-280. 1991.

Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., Konig, W. and Teti, R. Tool Condition Monitoring (TCM) - The Status of Research and Industrial Application, Annals of the CIRP, 44 (2), pp. 541-567. 1995.

Castleman, K. R. Digital image processing, pp. 116-121, pp. 480-481, New Jersey: Prentice Hall. 1996.

Chryssolouris, G. and Domroese, M. Sensor integration for tool wear estimation in machining, Proceedings of the Winter Annual Meeting of the ASME, Symposium on Sensors and Controls for Manufacturing, 1988, pp. 115-123.

Coker, S. A. and Shin, Y. C. In process control of surface roughness due to tool wear using ultrasonic system, Int. J. Mach. Tools Manufact., 36 (3), pp. 411-422. 1996.

Cook, N. H. Tool wear and tool life, ASME J. of Eng. for Industry, pp. 931-938, 1973.

Cook, N. H. and Subramanian, K. Micro-isotope tool wear sensor, Annals of the CIRP, 27 (1), pp. 73-78. 1978.

Diei, E. N. and Dornfeld, D. A. A model of tool fracture generated acoustic emission during machining, ASME J. of Eng. for Industry, 109, pp. 227-233. 1987a.

Diei, E. N. and Dornfeld, D. A. Acoustic emission sensing of tool wear in face milling, ASME J. of Eng. for Industry, 109, pp. 234-240. 1987b.

Dimla JR, D. E., Lister, P. M. and Leighton, N. J. Neural network solutions to the tool condition monitoring problem in metal cutting - a critical review of methods, Int. J. Mach. Tools Manufact., 37 (9), pp. 1219-1241. 1997.

Donnell, G.O., Young, P., Kelly, K. and Byrne, G. Towards the improvement of tool condition monitoring systems in the manufacturing environment, J. of Materials Processing Tech., 119, pp. 133-139. 2001.

Dornfeld, D. A. Neural network sensor fusion for tool condition monitoring, Annals of the CIRP, 39 (1), pp. 101-105. 1990.

Du, R., Zhang, B., Hungerford, W. and Pryor, T. Tool condition monitoring and compensation in finish turning using optical sensor, *1993 ASME Winter Annual Meeting, Symposium of Mechatronics*. DSC-Vol. 50/PED-Vol., Nov. 1993, 63, Chicago, IL, USA, pp. 245-253.

Elanayar, S. and Shin, Y. C. Robust tool wear estimation with radial basis function neural networks, Journal of Dynamic Systems, Measurement, and Control, 117, pp. 459-467. 1995.

Elbestawi, M. A., Marks, J. and Papazafiriou, T. A. Process monitoring in milling by pattern recognition, Mech. Sys. Signal Process., 3 (3), pp. 305-315. 1989.

Elbestawi, M. A., Papazafiriou, T. A. and Du, R. X. In-process monitoring of tool wear in milling using cutting force signature, Int. J. Mach. Tools Manufact., 31 (1), pp. 55-73. 1991.

Giardini, C., Ceretti, E. and Maccarini, G. A neural network architecture for tool wear detection through digital camera observations. In Advanced Manufacturing Systems and Technology,

CISM Courses and Lectures No. 372, ed by Kuljanic E., pp. 137-144. New York: Springer Verlag, Wien. 1996.

Giusti, F., Santochi, M. and Tantussi, G. On-line sensing of flank and crater wear of cutting tools, Annals of the CIRP, 36 (1), pp. 41-44. 1987.

Gonzalez, R. C. and Woods, R. E. Digital image processing, 2nd edition, pp. 85-94, pp. 234, New Jersey: Prentice Hall. 2002.

Halioua, M. and Liu, H.C. Optical three-dimensional sensing by phase measuring profileometry, Opt. Laser Eng., 11 (3), pp. 185-215. 1989.

Huang, P.T., Chen, J.C. and Chou, C.Y. A statistical approach in detecting tool breakage in end milling operations, J. of Industrial Technology, 15 (3), pp. 1-7. 1999.

Immerkær, J. Some remarks on the straight line Hough transform, Pattern Recognition Letters, 19 (12), pp. 1133-1135. 1998.

International Standard (ISO 8688-1, 1st edition): Tool Life Testing in Milling - part 1: Face milling. 1989.

Jemielniak, K. and Otman, O. Catastrophic tool failure detection based on acoustic emission signal analysis, Annals of the CIRP, 47, pp. 31-34. 1998a.

Jemielniak, K. and Otman, O. Tool failure detection based on analysis of acoustic emission signals, J. of Materials Processing Tech., 76 (1-3), pp. 192-197. 1998b.

Jeon, J. U. and Kim, S. W. Optical flank wear monitoring of cutting tools by image processing, Wear, 127, pp. 207-217. 1988.

Kamarthi, S. V. and Pittner, S. Fourier and wavelet transform for flank wear estimation - a comparison, Mech. Sys. Signal Process., 11 (6), pp. 791-908. 1997.

Kamarthi, S. V., Sankar, G. S., Cohen, P. H. and Kumara, S. R. T. On-line tool wear monitoring using a Kohonen feature map, Proc. of the Artificial Neural Networks in Engineering (ANNIE), 1991, St. Louis, MO, pp. 639-644.

Karthik, A., Chandra, S., Ramamoorthy, B. and Das, S. 3D tool wear measurement and visualization using stereo imaging, Int. J. Mach. Tools Manufact., 37 (11), pp. 1573-1581. 1997.

Kasashima, N., Mori, K., Ruiz, G. H. and Taniguchi, N. Online failure detection in face milling using discrete wavelet transform, Annals of the CIRP, 44 (1), pp. 483-487. 1995.

Klette, R. and Zamperoni, P. Handbook of image processing operators, pp. 195-310, New York: J. Wiley & Sons. 1996.

Ko, T. J. and Cho, D. W. Cutting state monitoring in milling by a neural network, Int. J. Mach. Tools Manufact., 34 (5), pp. 659-676. 1994.

Kohonen, T. Self-organized formation of topologically correct feature maps, Biological cybernetics, 43, pp. 59-69. 1982.

Kohonen, T. The self-organizing map, Proceedings of the IEEE, 78 (9), pp. 1464-1480. 1990.

Kohonen, T. Self-organizing maps, 3rd ed., pp. 105-110, pp. 270-313, New York: Springer. 2001.

Koren, Y. Flank wear model of cutting tools using control theory, Trans. ASME, 100, pp. 103-109. 1978.

Kurada, S. and Bradley, C. A review of machine vision sensors for tool condition monitoring, Computers in Industry, 34, pp. 55-72. 1997a.

Kurada, S. and Bradley, C. A machine vision system for tool wear assessment, Tribology International, 30 (4), pp. 295-304. 1997b.

Lan, M. S. and Naerheim, Y. In-process detection of tool breakage in milling, J. of Eng. Ind., 108, pp. 191-197. 1986.

Lanzetta, M. A new flexible high-resolution vision sensor for tool condition monitoring, J. of Materials Processing Tech., 119, pp. 73-82. 2001.

Lee, Y. H., Sutherland, D. R. and Schneider, J. G. Analysis of light strip images on cutting tools to extract 3D wear information, Proc. SPIE, 2298, pp. 519-530. 1994.

Lee, B. Y. and Tarng, Y. S. Milling cutter breakage detection by the discrete wavelet transform, Mechatronics, 9, pp. 225-234. 1999.

Lee, M. Y., Thomas, C. E. and Wildes, G. Review - prospects for in-process diagnosis of metal cutting by monitoring vibration signals, J. Mat. Sci., 22, pp. 3821-3890. 1987.

Leem, C. S. and Dornfeld, D. A. A customized neural network for sensor fusion in on-line monitoring of cutting tool wear, ASME J. of Eng. for Industry, 117, pp. 152-159. 1995.

Li, S. and Elbestawi, M. A. Tool condition monitoring in machining by fuzzy neural networks, J. of Dynamic Systems, Measurement, and Control, 118, pp. 665-672. 1996.

Li, X. L, Tso, S. K. and Wang, J. Real-time tool condition monitoring using wavelet transforms and fuzzy techniques, IEEE Transactions on systems, man, and cybernetics - part C: applications and reviews, 30 (3), pp. 352-357. 2000a.

Li, X., Dong, S. and Venuvinod, P. K. Hybrid learning for tool wear monitoring, Int. J. Adv. Manuf. Tech., 16, pp. 303-307. 2000b.

Lin, S. C. and Lin, R. J. Tool wear monitoring in face milling using force signals, Wear, 198, pp. 136-142. 1996.

Liu, M. and Liang, S. Y. Analytical modeling of acoustic emission for monitoring of peripheral milling process, Int. J. Mach. Tools Manufact., 31 (4), pp. 589-606. 1991.

Maeda, Y., Uchida, H. and Yamamoto, A. Contour mapping of cutting tool face with the aid of digital image processing technique, JSPE, 21 (2), pp. 136-138. 1987a.

Maeda, Y., Uchida, H. and Yamamoto, A. Estimation of wear land width of cutting tool flank with the aid of digital image processing technique, JSPE, 21 (3), pp. 211-213. 1987b.

Mannan, M. A., Ashraf, A. K. and Ma, J. Application of image and sound analysis techniques to monitor the condition of cutting tools, Pattern Recognition Letters, 21, pp. 969-979. 2000.

Markos, S., Szalay, T., Szollos, G. and Raifu, A. W. Monitoring of milling processes based on artificial intelligence, Mechatronics, 3 (2), pp. 231-240. 1993.

Marques, M. J. M. and Mesquita, R. M. D. Monitoring of the wear of sintered high-speed steel tools, J. of Materials Processing Tech., 25, pp. 195-213. 1991.

Masory, O. Detection of tool wear using multi-sensor readings defused by artificial neural network, Proc. SPIE, 1469 (2), pp. 515-525. 1991.

Matsushima, K., Bertok, P. and Sata, T. In process detection of tool breakage by monitoring the spindle current of a machine tool, ASME J. of Measurement and Control for Batch Manufacturing, pp. 145-154. 1982.

Monostori, L. A. step towards intelligent manufacturing: modeling and monitoring of manufacturing processes through artificial neural networks. Annals of the CIRP, 42 (1), pp. 485-488. 1993.

Moriwaki, T. and Mori, Y. Recognition of cutting state based on neural network sensor fusion. JSPE, 59 (5), pp. 779-784. 1993.

Niu, Y. M., Wong, Y. S. and Hong, G. S. An intelligent sensor system approach for reliable tool flank wear recognition, Int. J. Adv. Manuf. Tech., 14, pp. 77-84. 1998.

Oguamanam, D. C. D., Raafat, H. and Taboun, S. M. A machine vision system for wear monitoring and breakage detection of single-point cutting tools, Computers and Industrial Engineering, 26 (3), pp. 575-598, 1994.

Oraby, S. E. Monitoring of turning operation via force signals, part 1: recognition of different tool failure forms by spectral analysis, Wear, 184, pp. 133-143. 1995.

Otsu, N. A threshold selection method from gray level histogram. IEEE Trans., SMC-9: pp. 62-69. 1979.

Park, J. J. and Ulsoy, A. G. On-line flank wear estimation using an adaptive observer and computer vision, Part 1: Theory, Trans. ASME J. Eng. Ind., 115, pp. 30-36. 1993a.

Park, J. J. and Ulsoy, A. G. On-line flank wear estimation using an adaptive observer and computer vision, Part 2: Experiment, Trans. ASME J. Eng. Ind., 115, pp. 37-43. 1993b.

Pedersen, K. B. Wear measurement of cutting tools by computer vision, Int. J. Mach. Tools Manuf., 30 (1), pp. 131-139. 1990.

Pfeifer, T. and Wiegers, L. Reliable tool wear monitoring by optimized image and illumination control in machine vision, Measurement, 28, pp. 209-218. 2000.

Prasad, K. N. and Ramamoorthy, B. Tool wear evaluation by stereo vision and prediction by artificial neural network, J. of Materials Processing Tech., 112, pp. 43-52. 2001.

Prickett, P.W. and Johns, C. An overview of approaches to end milling tool monitoring, Int. J. Mach. Tools Manufact., 39, pp. 105-122. 1999.

Rangwala, S. and Dornfeld, D. Integration of sensors via neural networks for detection of tool wear states, ASME PED, 25, pp. 109-120. 1987.

Reif, Z. F. and Cahine, J. On line monitoring of tool condition monitoring by means of vibration, in: 9th Eng. Appl. of Mech., 1988, London, Ontario, 533-542.

Rene de Jesus, R. T., Gilberto, H. R., Ivan, T. V. and Juan Carlos, J. C. FPGA based on-line tool breakage detection system for CNC milling machines, Mechatronics, 14, pp. 439-454. 2004.

Ruiz, A., Guinea, D., Barrios, J. and Betancourt, F. An empirical multi-sensor estimation of tool wear, Mech. Sys. Signal Process., 7 (2), pp. 105-119. 1993.

Sampath, A. and Vajpayee, S. Tool health monitoring using acoustic emission, Int. J. Prod. Res. 25 (5), pp. 703-719. 1987.

Santanu, D., Chattopadhyay, A. B. and Murthy, A. S. R. Force parameters for on-line tool wear estimation: a neural network approach, Neural networks, 9 (9), pp. 1639-1645. 1996.

Sonka, M., Hlavac, V. and Boyle, R. Image processing, analysis, and machine vision, 2nd edition, pp. 127-133, CA: PWS Pub, 1999.

Stein, J. L. and Wang, C. H. Analysis of power monitoring on AC induction drive systems, ASME J. of Dynamic Systems Measurement and Control, 112, pp. 239-248. 1990.

Stoferle, T. H. and Bellmann, B. Continuous measuring of flank wear, Proceedings of Sixteenth International Machine Tool Design and Research Conference, 1975, pp. 573-578.

Tae Hyeon Kim, Young Shik Moon and Chang Soo Han, An efficient method of estimating edge locations with subpixel accuracy in noisy images, Proceedings of IEEE Region 10 conference, TENCON 99, pp. 589-592.

Takeyama, H., Doi, Y., Mitsoka, T. and Sekiguchi, H. Sensors of tool life for optimization of machining, Proceedings of Eighth International Machine Tool Design and Research Conference, 1967, pp. 191-208.

Tansel, I. N. and Mclaughlin, C. Detection of tool breakage in milling operations - I. The time series analysis approach, Int. J. Mach. Tools Manufact., 33 (4), pp. 531-544. 1993a.

Tansel, I. N. and Mclaughlin, C. Detection of tool breakage in milling operations - II. The neural network approach, Int. J. Mach. Tools Manufact., 33 (4), pp. 545-558. 1993b.

Tansel, I. N., Mekdeci, C. and Mclaughlin, C. Detection of tool failure in end milling with wavelet transformations and neural networks, Int. J. Mach. Tools Manufact., 35 (8), pp. 1137-1147. 1995.

Tansel, I. N., Mekdeci, C., Rodriguez, O. and Uragun, B. Monitoring drill conditions with wavelet based encoding and neural networks, Int. J. Mach. Tools Manufact., 33 (4), pp. 559-575. 1992a.

Tansel, I. N., Rodriguez, O. and Mekdeci, C. Detection of tool breakage in micro-drilling operation with RCE neural networks, ASME PED, 47 (1), pp. 83-88. 1992b.

Tarn, J. H. and Tomizuka, M. On-line monitoring of tool and cutting conditions in milling, ASME J. of Eng. for Industry, 111, pp. 206-211. 1989.

Tarng, Y. S. Study of milling cutting force pulsation applied to the detection of tool breakage, Int. J. Mach. Tools Manufact., 30 (4), pp. 651-660. 1990.

Tarng, Y. S., Hseih, Y. W. and Hwang, S. T. Sensing tool breakage in face milling with a neural network, Int. J. Mach. Tools Manufact., 34 (3), pp. 341-350. 1994.

Teo, S. C. Development of an on-line tool wear monitoring system using dynamic cutting force, Master thesis, National University of Singapore, 1992.

Teshima, T., Shibasaka, T., Takuma, M. and Yamamoto, A. Estimation of cutting tool life by processing tool image data with neural network, Annals of the CIRP, 42 (1), pp. 59-62. 1993.

Tlusty, J. and Tarng, Y. S. Sensing cutter breakage in milling, Annals of the CIRP, 37, pp. 45-54. 1988.

Uehara, K. New attempts for short time tool-life testing, Annals of the CIRP, 22, pp. 23-24. 1973.

Verma, A. and Kline, W. A. Force - transducer applications on CNC lathe, Transactions of SME Advanced Machining Technology III Conference, Sep. 1990, Chicago, IL, pp. 4-6.

Wilcox, S. J., Reuben, R. L. and Souquet, P. The use of cutting force and acoustic emission signals for the monitoring of tool insert geometry during rough face milling, Int. J. Mach. Tools Manufact., 37 (4), pp. 481-494. 1997.

Wong, Y. S., Nee, A. Y. C., Li, X. Q. and Reisdorf, C. Tool condition monitoring using laser scatter pattern, J. of Materials Processing Tech., 63, pp. 205-210. 1997.

Wu, Y. and Du, R. Feature extraction and assessment using wavelet packets for monitoring of machining processes, Mech. Sys. Signal Process., 10 (1), pp. 29-53. 1996.

Xu, K. and Luxmoore, A. R. An integrated system for automatic wear particle analysis, Wear, 208, pp. 184-193. 1997.

Xue, H. J., Yang, K. C. and Yang, R. The shape characteristic detection of tool breakage in milling operations, Int. J. Mach. Tools Manufact., 37 (11), pp. 1651-1660. 1997.

Yang, M. Y. and Kwon, O. D. Crater wear measurement using computer vision and automatic focusing, J. of Materials Processing Tech., 58, pp. 362-367. 1996.

Yellowley, I. Observations on the mean values of force torque and specific power in the peripheral milling process, J. of Machine Tool Des. and Res., 25 (4), pp. 337-346. 1985.

Zhang, D. Y., Han, Y. T. and Chen, D. C. On-line detection of tool breakages using telemetering of cutting force in milling, Int. J. Mach. Tools Manufact., 35 (1), pp. 19-27. 1995.

Zhou, Q., Hong, G. S. and Rahman, M. A new tool life criterion for tool condition monitoring using a neural network, Engng Applic. Artif. Intell. 8 (5), pp. 579-588. 1995.

Zitová, B. and Flusser, J. Image registration methods: a survey, Image and Vision Computing, 21 (11), pp. 977-1000. 2003.

Published works

Wang, W. H., Hong, G. S. and Wong, Y. S. Flank wear measurement by a threshold independent method with sub-pixel accuracy, Int. J. Mach. Tools Manufact., 2006, 46(2), 199-207.

Wang, W. H., Wong, Y. S. and Hong, G. S. Flank wear measurement by successive image analysis, Computers in Industry, 2005, 56(8-9), 816-830.

Wang, W. H., Wong, Y. S. and Hong, G. S. 3-D measurement of crater wear by phase-shifting method, Wear, In press.

Wang, W. H., Hong, G. S., Wong, Y. S. and Zhu, K. P. Sensor fusion for on-line tool condition monitoring in milling, Int. J. Prod. Res., In press.

Zhu, K. P., Hong, G. S., Wong, Y. S. and Wang W. H. Cutting force denoising in micro-milling tool condition monitoring, Int. J. Prod. Res., In press.

Wang, W. H., Wong, Y. S. and Hong, G. S. Flank wear measurement based on moving images, in ISIST 2004, Vol. 2, Aug. 18~22, 2004, Xi'an, China, pp. 159-164.

Wang, W. H., Wong, Y. S. and Hong, G. S. Sensor fusion for on-line tool condition monitoring in milling, International symposium on collaborative research in applied science (ISOCRIAS), 7-9 October 2005, The University of British Columbia, Vancouver, BC, Canada.