

# Visual Inspection of Machined Metallic High-Precision Surfaces

**Franz Pernkopf**

*Institute of Automation, University of Leoben, Peter-Tunner-Strasse 27, A 8700 Leoben, Austria  
Email: franz.pernkopf@unileoben.ac.at*

**Paul O'Leary**

*Institute of Automation, University of Leoben, Peter-Tunner-Strasse 27, A 8700 Leoben, Austria  
Email: automation@unileoben.ac.at*

*Received 31 July 2001 and in revised form 12 January 2002*

This paper presents a surface inspection prototype of an automatic system for precision ground metallic surfaces, in this case bearing rolls. The surface reflectance properties are modeled and verified with optical experiments. The aim being to determine the optical arrangement for illumination and observation, where the contrast between errors and intact surface is maximized. A new adaptive threshold selection algorithm for segmentation is presented. Additionally, is included an evaluation of a large number of published sequential search algorithms for selection of the best subset of features for the classification with a comparison of their computational requirements. Finally, the results of classification for 540 flaw images are presented.

**Keywords and phrases:** visual inspection, surface reflection, flaw detection, segmentation, statistical classification, feature selection.

## 1. INTRODUCTION

In industry, there is an increasing demand for automatic surface inspection systems for quality control of final products. Thereby, image processing techniques play a crucial role in this growing field. The customer demands are well founded on the high costs of poor quality, with the resulting cost for correction. The human assessment of the product quality in repetitive tasks, for example, inspection of textiles, achieves a maximum reliability of only 80% [1]. Thus, replacing the human visual inspection with a fully automatic system will normally increase the accuracy of flaw detection and reduce costs. A survey of automated visual inspection can be found in [2]. Many papers of surface inspection on steel products have recently been published [3, 4], most of which concentrate on flat objects.

In this paper, an automatic visual surface inspection system for bearing rolls, that is, precision surfaces, is presented. A coarse concept of the automated solution for real-time visual inspection of bearing rolls is shown in Figure 1.

It consists of five main parts:

- (1) an optical arrangement for the acquisition of the surface image data, including the mechanical hardware, camera, lighting, and frame-grabber (see Section 3),
- (2) flaw detection (see Section 4),
- (3) image segmentation (see Section 5.1),

- (4) feature extraction (see Section 5.2),
- (5) feature selection (see Section 5.3), and
- (6) classification (see Section 5.4).

The problem at hand is presented in Section 2, whereby, the demands on the inspection system and the main characteristics of the flaws which must be detected are described. The design of the optical arrangement (see Section 3) is based on a modeling of the surface reflectance properties [5, 6, 7]. Exploratory measurements were performed under laboratory conditions to verify the model and to determine the optimal position of the light sources. The inspection of a single roll requires the processing of approximately 7 MBytes of data. Consequently, in the first step a statistical procedure, which is numerically efficient (see Section 4), is used to locate irregularities in the acquired data. These portions of the data are extracted and submitted as subimages for further processing. In this manner, a data reduction of approximately 1 : 100 is achieved.

The classification of the flaws in the subimages is presented in Section 5. These images are segmented so as to separate the flaw from the surroundings, so enabling the calculation of characteristic features. During online processing these features are used for classification. The identification and selection of the feature set with the best classification performance is a complex task and is performed a priori offline. To

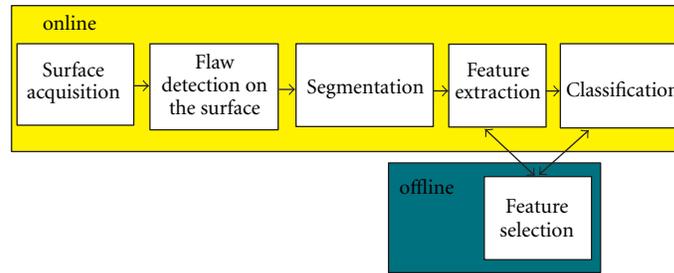


FIGURE 1: System concept.



FIGURE 2: Different bearing rolls.

this end, a comparative study of sequential feature selection algorithms with experimental results from this application is presented. The paper closes with a summary and gives a perspective on the future work.

## 2. PROBLEM STATEMENT

In the production of high quality bearings, the surfaces of the bearing rolls have to be free of flaws and defects which can arise during the production process. To guarantee a long period of operation and low friction, bearing rolls which have flaws with a diameter larger than 0.1 mm should be removed from the production lot. Whereby, the small defects tend to have a better contrast than the larger defects. The dimensions of the bearing rolls differ widely, the length is between 45 and 100 mm and the diameter varies from 30 to 45 mm. The different rolls are shown in Figure 2.

The cycle time for detection and classification of all flaws on one small bearing roll is specified as 1.2 seconds. Consequently, the data throughput for 100% inspection corresponds to approximately 6 MB per second.

Visual inspection is performed straight after the grinding process, so the reflection properties of the rough ground surface are highly specular and not yet influenced by inhomogeneities of the corrosion protection liquid.

### 2.1. Characteristic of flaws

The characteristics of the flaws on the metallic surface depend strongly on the final treatment. There are three main causes of surface irregularities in this application:

- (i) material defects, which are present in the incoming material prior to grinding;

- (ii) flaws produced during grinding;
- (iii) chafe marks due to material handling.

According to the human inspectors five different classes of irregularities have been established during the production. Thereby, the first three classes degrade the roughness measure of the surface.

(1) Material flaws (class 1): they exhibit a very low and homogeneous grey value which lightens at the edge of the defect. They show no particular shape or size. These errors already exist in the blank and they appear as a cavity in the ground surface of the bearing roll.

(2) Grinding flaws (class 2): these flaws arise at the end of the grinding process when the bearing roll exits the grinding machine. They commonly exhibit a triangular shape with a predominant orientation. The grey value of the error is not as low and homogeneous as in class 1. The visual appearance of these flaws depend strongly on the illumination direction (see Section 3). The roughness of the surface is elevated.

(3) Scratches (class 3): scratches develop from the abrasive grain of the grinding wheel. Their appearance is very thin and long with a low homogeneous grey value, and the orientation is mostly between  $60^\circ$  and  $120^\circ$ . They also degrade the roughness measure of the surface.

(4) Chafe marks (class 4): chafe marks are caused by the manipulators. They show no particular shape, but the object itself is normally very inhomogeneous. These flaws are in fact a polishing of the surface.

(5) Spots (class 5): the genesis of spots is unsolved. They cannot be removed by the cleaning process before the visual inspection. Dark spots show a small size and a medium grey value. They do not alter the roughness measure of the surface.

A representative example of each class is shown in Figure 3.

An additional complication in this task is that the flaw types 4 and 5 should not lead to a part being removed from the production lot. This places additional requirements on the classification algorithms.

## 3. OPTICAL ARRANGEMENT AND EXPERIMENTAL SETUP

An optimal illumination can enhance the visual appearance of the surface defects. Especially for flaws of class 2 and class 3 the positioning of the light source significantly influences the

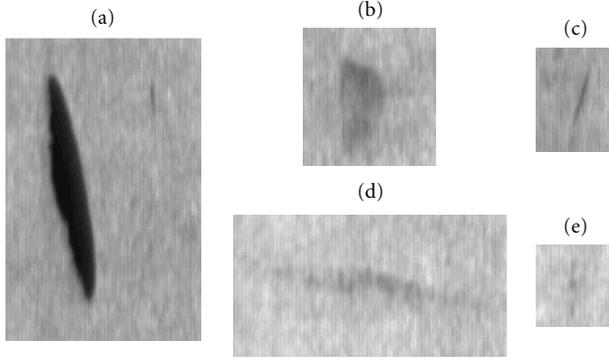


FIGURE 3: Different flaws: (a) material flaw, (b) grinding flaw, (c) scratch, (d) chafe mark, (e) spot.

image contrast. These defects are characterized by an altering microstructure of the surface, which degrades the roughness measure. Therefore, the optimal direction of the incident light is derived from the surface topography. For this reason, a modeling of the surface reflectance properties have been performed and validated by means of optical experiments.

### 3.1. Reflection model

The surface is regarded as a composition of a collection of microfacets which constitute a surface patch having a mean normal vector  $\mathbf{n}$ . In analyzing the surface reflection properties, the roughness of the surface is determined by the slope distribution of the microfacets. Each microfacet is characterized by the slope  $\alpha$ , which is the tilt angle of its normal vector with respect to the mean normal vector  $\mathbf{n}$ . The surface is described by the standard deviation  $\sigma_\alpha$  of the assumed mean-centered normal distribution.

Nayar et al. [7] compared the Beckmann-Spizzochino [5] physical optics model to the Torrance-Sparrow [6] geometrical optics model and proposed a unified reflectance framework for smooth and rough surfaces, which is comprised of three reflection components: the diffuse lobe  $I_{dl}$ , the specular lobe  $I_{sl}$ , and the specular spike  $I_{ss}$ . These three components are shown in Figure 4. The radiance in the sensor direction of the three components is determined through the intersections of the lobes with the line joining the light source and the origin [7].

The light reflected from the surface in the direction of the camera causes an image with the intensity

$$I_{im} = I_{dl} + I_{sl} + I_{ss}. \quad (1)$$

(1) Diffuse lobe: the diffuse component is represented by the Lambertian model,

$$I_{dl} = K_{dl} \cos \theta_i, \quad (2)$$

where  $K_{dl}$  denotes the strength of the diffuse lobe and  $\theta_i$  is the angle of the incident light rays.

(2) Specular lobe: due to its simpler mathematical form,

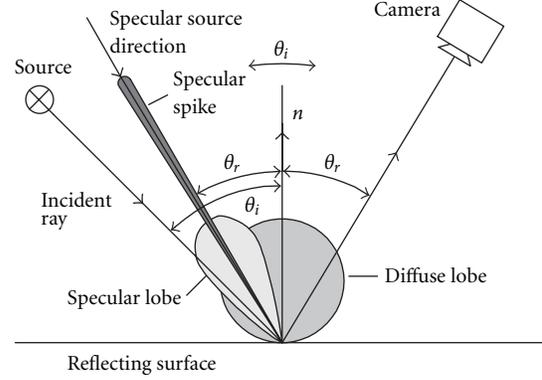


FIGURE 4: Three reflection components for a fixed viewing direction and a varying source angle (from Nayar et al. [7]).

a Torrance-Sparrow model [6] is used to predict the specular lobe,

$$I_{sl} = K_{sl} \exp\left(-\frac{\alpha^2}{2\sigma_\alpha^2}\right) F(\theta'_i, \eta') G(\theta_i, \theta_r, \phi_r), \quad (3)$$

where  $K_{sl}$  is the magnitude of the specular lobe and the exponential term describes the slope distribution of the microfacets assuming a normal distribution.

The Fresnel coefficient is represented by  $F(\theta'_i, \eta')$ , where  $\theta'_i$  is the local angle of incidence and  $\eta'$  is the complex index of refraction. In the literature it is observed that for metals, the Fresnel coefficient is nearly constant until  $\theta'_i$  approaches  $75^\circ$  [7]. Therefore, this coefficient is assumed constant in this application.

The geometric attenuation factor  $G(\theta_i, \theta_r, \phi_r)$  [6] describes the shadowing and masking effects of facets by adjacent facets, where  $\theta_r$  and  $\phi_r$  are the direction of the camera. The factor  $G$  equals to unity over a large range for  $\theta_i$  when the viewing direction coincides with the mean normal vector  $\mathbf{n}$ .

(3) Specular spike: the specular spike component is a very sharp function which is approximated by the delta function,

$$I_{ss} = K_{ss} \delta(\theta_i - \theta_r) \delta(\phi_r), \quad (4)$$

where  $K_{ss}$  is the strength of the specular spike component.

The coordinate system used for the model is shown in Figure 5.

The ratio  $K_{sl}/K_{ss}$  is dependent on the surface roughness [7]. For rough surfaces, the specular lobe is dominant compared to the specular spike which is negligible when the standard deviation of the surface height  $\sigma_h$ , divided by the wavelength  $\lambda$  of the light, exceeds a factor of 1.5 [7]. The surface roughness of the bearing rolls is small, however, the optical measurements (see Section 3.3) show that no specular spike is present. The change of the surface roughness caused by the flaws only influences the specular lobe component in the reflection model. Hence, the reflectance model is restricted to the specular lobe component to describe the difference of the reflectance of an intact and flawed surface.

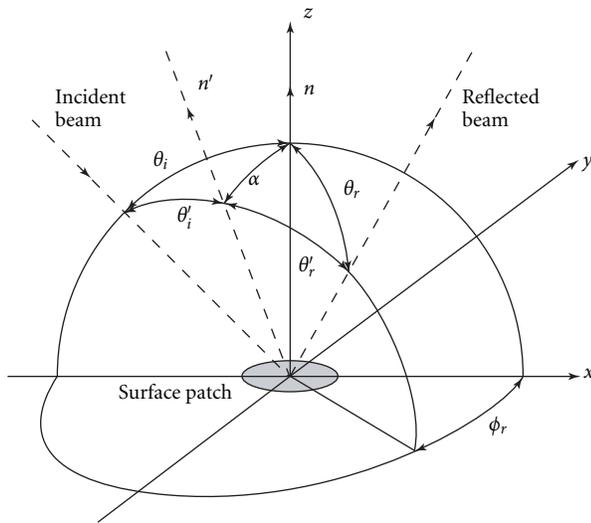


FIGURE 5: Coordinate system for the reflection model (adopted from Torrance and Sparrow [6]).

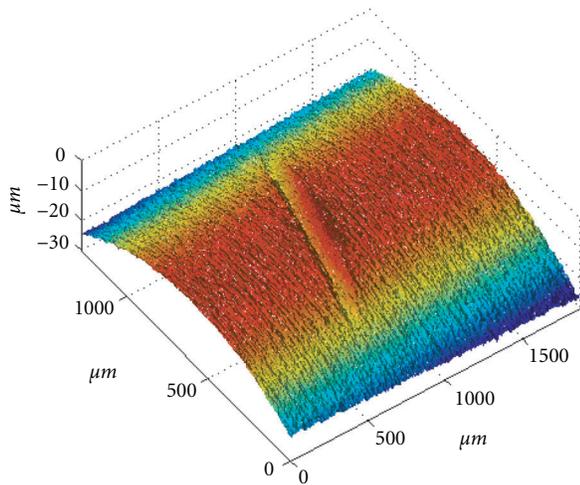


FIGURE 6: 3D profile of a scratch.

### 3.2. Measurements of surface roughness

The proposed reflection model is verified with the measurements of the surface roughness. The surface height measurements of the flaws and the adjacent error-free surface were acquired using a topography measuring apparatus (Wyko NT-1000, Veeco Instruments, Inc., New York, USA). The 3D surface of a scratch is shown in Figure 6.

Basically, the reflection properties of machined surfaces must be divided into a reflection function parallel and perpendicular to the machining direction [8]. In the proposed arrangement (see Figure 11), the camera, the light source, and the axis of the roll lie in a plane. Hence, the reflection function parallel to the machining direction can be neglected and the investigations concentrate on modeling the reflection function in the plane given by the optical setup. The

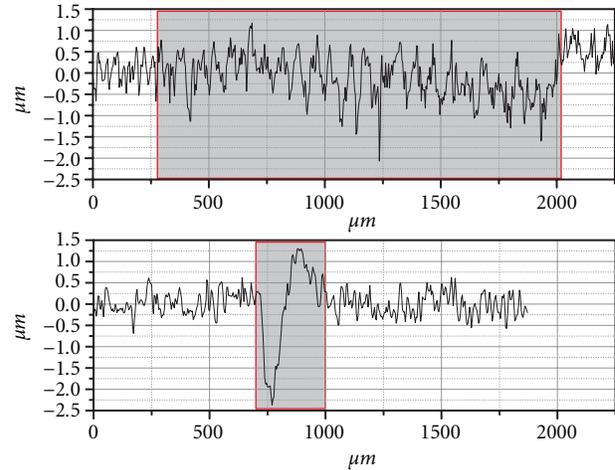


FIGURE 7: Roughness profiles of a grinding flaw and a scratch.

roughness profiles for a scratch and grinding error measured in the direction of the roll axis are visualized in Figure 7. Both flaws degrade the surface smoothness.

The standard deviation of the approximately normal distributed surface slopes  $\sigma_\alpha$  is for the errors lower than for the intact surface. The change of the surface height in case of a flaw is at least twice that of the error-free surface, also the frequency of the main oscillation of the roughness profile is lower and the slopes are smoother, these effects result in a lower  $\sigma_\alpha$ . The measured  $\sigma_\alpha$  for the grinding error is 4.8 and for an intact surface 5.5. This difference causes a shift of the reflection property.

The reflection intensity of the specular lobe component is computed according to the model (see (3)) and plotted in Figure 8. The viewing direction is fixed at  $\theta_r = 0^\circ$  and the source direction  $\theta_i$  is varied.

The difference in the reflected intensities (see Figure 8) for the intact surface and a grinding flaw reflects the difference in surface roughness. The largest difference is given for an angle of illumination  $\theta_i$  in the range from  $10^\circ$  to  $30^\circ$ . Hence, an illumination angle in this range should lead to an optimum contrast.

### 3.3. Optical measurements

The influence of the surface roughness on the reflected light was measured using a camera and a point light source. The source direction  $\theta_i$  is varied from  $0^\circ$  to  $90^\circ$  while the sensor direction coincides with the normal vector  $\mathbf{n}$  of the surface and is constant with  $\theta_r$  of  $0^\circ$ . The intensity measurements are obtained for both an error-free and a flawed area by means of averaging. In Figure 9, the measured intensity distribution is plotted.

A rough surface leads to less light being reflected into the camera. The maximum contrast between a flaw and intact surface is achieved for an illumination angle in the range from  $10^\circ$  to  $40^\circ$ . This experimental result corresponds well with the model prediction (see Figure 8), despite, the simplifications and assumptions which were made in [6, 7].

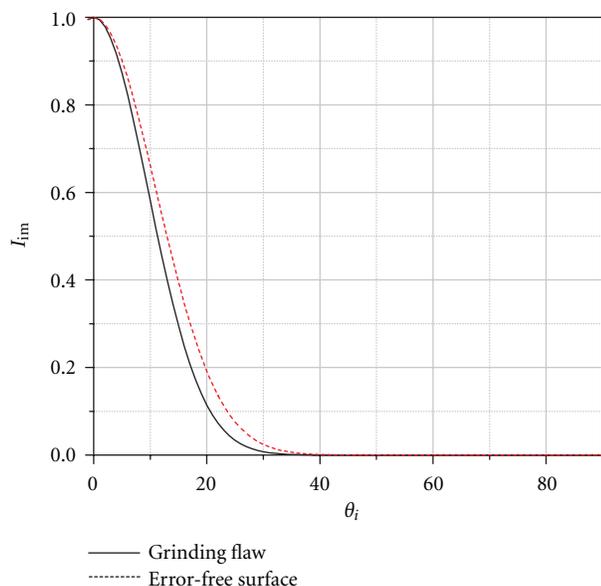


FIGURE 8: The relative reflected intensity for an error-free surface and a grinding error is computed according to the reflection model which is restricted to the specular lobe component.

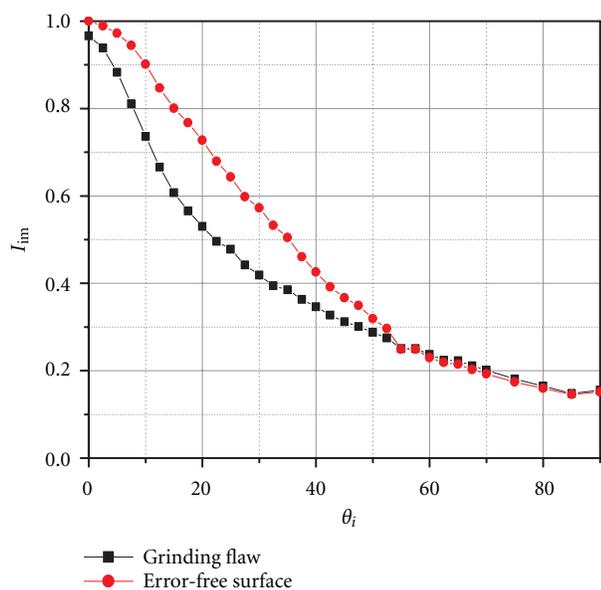


FIGURE 9: Relative intensity measured for an error-free surface and a grinding error for a fixed viewing direction  $\theta_r = 0^\circ$  and a varying source direction.

### 3.4. Practical implementation for the production environment

The bearing roll is illuminated axially from two sides using a focused lightline in order to achieve a uniform illumination across the complete surface. Due to the geometry of the light source, the light rays impinging on the surface are distributed over a wide range of incident angles, as shown in Figure 10.

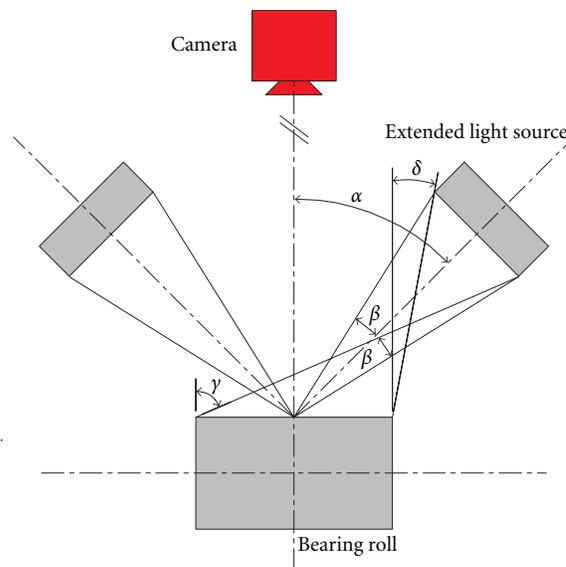


FIGURE 10: Arrangement of the light sources.

Further, the illumination across the roll is nonuniform with respect to each light source. The left side of the roll is illuminated primarily by the left light source and vice versa. The best illumination angle has been determined experimentally to be  $\alpha = 45^\circ$ . This implies, taking the mechanical arrangement and dimensions into account, that  $\beta = 14^\circ$ ,  $\gamma = 65^\circ$ , and  $\delta = 18^\circ$ . Consequently, the angle of incidence  $\theta_i$  for most of the light rays lies between  $\delta = 18^\circ$  and  $\alpha + \beta = 59^\circ$ .

For an illumination angle  $\theta_i$  greater than  $60^\circ$  and a viewing angle  $\theta_r$  of  $0^\circ$ , the shadowing and masking effects of facets by adjacent facets modeled in the Torrance-Sparrow model [6, 7] with the geometrical attenuation factor must be considered. This would be the case for a portion of the surface if the roll was illuminated using only one light source (e.g., for the left portion of the roll when illuminating from the right-hand side).

### 3.5. Experimental setup

A monochrome CCD line-scan camera with 2048 pixels is employed. Approximately 3500 lines for one perimeter of the small rolls are scanned. The sampling interval is chosen to be less than a half of the smallest flaw which should be detected. The pixel size in the horizontal direction is 0.049 mm and 0.027 mm in the vertical direction. The mechanical arrangement is shown in Figure 11.

## 4. FLAW DETECTION

The flaws have to be detected in the acquired surface image, where they are embedded in a more or less homogeneous texture. The roughness of the ground rolls causes an image composed of statistical textures which are characterized by a stochastic spatial distribution of grey values. In comparison, structural textures are composed by texture primitives which occur repeatedly in a deterministic manner.

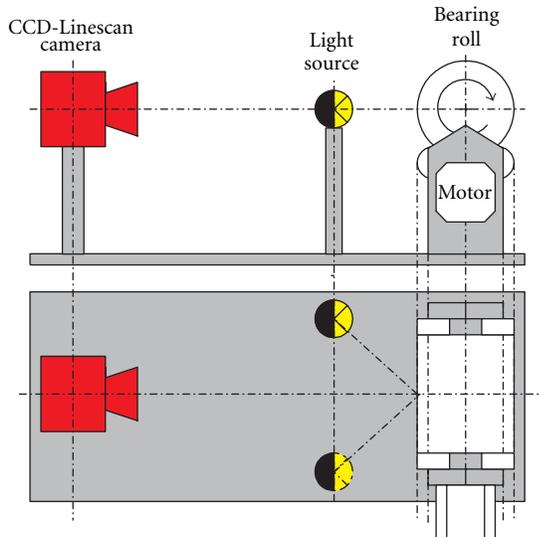


FIGURE 11: Experimental arrangement.

In the first step, regions of interest (i.e., areas with possible errors) are determined by moving a sliding window over the entire surface. In this window a reduced set of features is calculated to determine if a possible flaw is present or not. Regions with feature values exceeding predetermined limits are extracted as subimages for further processing. This step brings a data reduction of approximately 1 : 100. Basically, two main approaches to textural features are proposed in the literature, either features of the spatial domain, or of the spatial-frequency domain. The most difficult task is to extract adequate features, which are most descriptive and selective according to the appearance of the flaws. Due to the real-time demands in this application only first-order statistical features [9] of the spatial domain, for example, mean, variance, skewness, and kurtosis are computed. They deliver a reliable detection of the surface flaws even in the case of weak contrast.

## 5. CLASSIFICATION

The classification of the detected irregularities is achieved by means of characteristic descriptors (features). Thus each subimage is described by a feature vector,

$$\mathbf{x}^T = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad (5)$$

where  $x_i$  represents the  $i$ th descriptor and  $N$  is the number of all features.

To this end, the irregularity within the subimages is segmented and features are derived. The most appropriate features for classification are not known a priori. Hence, the features must be selected in a prior offline step from a large set so as to give optimal classification performance.

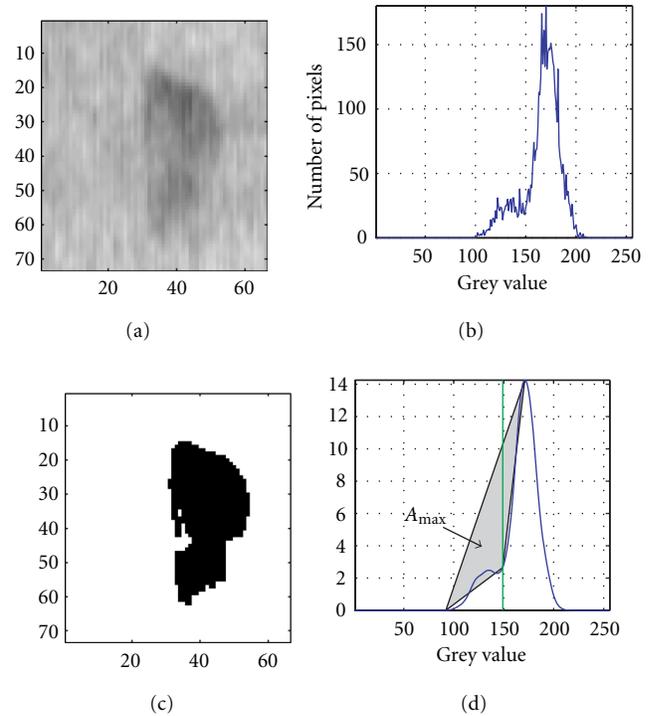


FIGURE 12: Segmentation through threshold based on maximizing the area of the triangle: (a) grinding error, (b) histogram, (c) binary image after applying the threshold and the morphological filtering, (d) smoothed histogram, grey line represents the threshold  $T = 149$ .

### 5.1. Segmentation

The most widely used method for segmentation is thresholding. The threshold is derived from the histogram. The grey value distributions of the flaw and the surrounding statistical texture may overlap so the smoothed histogram has only a single peak. Some of the defects in this application show a very low contrast which results in a unimodal histogram having a tail. It is assumed that the flaw, represented as a more or less homogeneous region in the image, is mapped to the tail of the histogram. Consequently, the threshold has to be located at the point where the curvature of the histogram slope changes into the tail. Before the threshold is determined the noise in the histogram has to be reduced sufficiently by smoothing. This is necessary to guarantee a successful operation of the thresholding algorithm.

The histogram  $H(i)$  is convoluted,

$$\tilde{H}(i) = H(i) * g(i, \sigma) \quad (6)$$

with a one-dimensional Gaussian kernel

$$g(i, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{i^2}{2\sigma^2}\right), \quad (7)$$

where  $i$  is the grey value,  $\sigma$  is the width of the kernel, and  $\tilde{H}(i)$  denotes the smoothed histogram [10]. An appropriate value for  $\sigma$  must be chosen to avoid degradation of the algorithm

performance. A too large value of  $\sigma$  leads to over-smoothing and a loss of characteristics, whereby a too low value results in insufficient smoothing.

Once the histogram has been smoothed, the threshold algorithm based on maximizing the area of a triangle is applied. The point at the maximum of the histogram  $\mathbf{p}_1 = [x_1, y_1]$  and the point corresponding to the end of the left tail  $\mathbf{p}_2 = [x_2, y_2]$  are chosen to form the baseline of the triangle. The point on the histogram  $\mathbf{p}_3 = [x_i, y_i]$  is now sought, which has the maximum normal distance to the baseline. This is equivalent to determining the maximum area enclosed by the three points. However, determining the enclosed area  $A_i$  is numerically more efficient than determining the normal distance. The area  $A_i$  is determined as half of the magnitude of the cross product between the  $x$  and the  $y$  components of the given three points  $\mathbf{p}_1$ ,  $\mathbf{p}_2$ , and  $\mathbf{p}_3$ . This is equivalent to determining the determinant as

$$A_i = \frac{1}{2} \begin{vmatrix} x_i & y_i & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}, \quad (8)$$

or

$$A_i = \frac{1}{2} [x_i(y_1 - y_2) - y_i(x_1 - x_2) + (x_1 y_2 - x_2 y_1)]. \quad (9)$$

This equation can be further simplified, since  $\mathbf{p}_1$  and  $\mathbf{p}_2$  are constant during the search:

$$A_i = x_i c_0 + y_i c_1 + c_2. \quad (10)$$

The threshold value  $T$  is

$$T = x_i \mid x_2 < x_i < x_1 \wedge A_i = \max(A_i). \quad (11)$$

The result of the segmentation is shown in Figure 12.

The threshold generated in this manner leads to a compact segmentation of the flaw. However, the major advantage is that the algorithm is independent of the number of peaks in the histogram. In the case of a bimodal histogram the threshold level is selected in the minimum between the peaks. When the image contrast is low, the selected threshold level lies at the transition of the distinct classes (flaw/surroundings) of the unimodal histogram. If the peak caused by the flaw is larger than the peak of the surroundings this method fails. However, in this application the surrounding area is always dominant.

## 5.2. Feature extraction

The result of the segmentation is a well-defined object in the subimage. This is necessary for the next step where shape describing features are computed as well as grey value dependent descriptors. The algorithms for the computation of the different features are described in [9, 11, 12] and summarized in Table 1. Altogether, 54 features were extracted of each flaw in a preliminary step.

Basically, four different approaches for feature extraction were used that may roughly be divided into statistical features, contour-based descriptors, region-based methods, and gradient features.

(1) Region-based features: simple region-based descriptors are derived from the binary image. These features refer either to all segments obtained after segmentation or to the largest segment extracted through a region labeling algorithm described in Sonka et al. [11]. Most of the features are derived from the region projections and from translation invariant central moments.

(2) Statistical features: these first-order statistic measures are computed either from the image or separately from the flaw and its background. Additionally, the ratio of the mean and the standard deviation of the flaw and the appropriate background serve as features.

(3) Contour-based features: the boundary of the largest segment in the binary image is described with the help of Fourier descriptors [9, 11]. Therefore, the contour is represented as a closed curve  $\vec{c}(i) = (c_x(i), c_y(i))$ , where  $i = 1, \dots, L$ ,  $L$  is the curve length, and  $c_x(i)$  and  $c_y(i)$  are the coordinates in  $x$  and  $y$ . The Fourier descriptors are computed for  $c_x(i)$  and  $c_y(i)$  and therefrom the first few translation, rotation, and magnification invariant descriptors are used as measure for the shape of the flaw.

(4) Gradient features: the gradient of an image describes abrupt changes of the grey value between neighboring pixels. The maximum of the gradient in both directions is used to discriminate between low contrast and clearly defined flaws.

## 5.3. Feature selection

In real-world classification problems, the relevant features for classification are often unknown a priori. Hence, the task of feature selection is to reduce the number of extracted features to a set of a few significant ones, which optimize the classification performance. Therefore, many algorithms exist which typically consist of four basic steps [13]:

- (1) a generation procedure to generate the next subset of features  $X$ ;
- (2) an evaluation criterion  $J$  to evaluate the quality of  $X$ ;
- (3) a stopping criterion for concluding the search. It can be based either on the generation procedure or on the evaluation function;
- (4) a validation procedure for verifying the validity of the selected subset.

Feature selection is performed offline and hence, only the selected features are considered during online classification.

The best subset

$$X = \{x_i \mid i = 1, \dots, d; x_i \in Y\} \quad (12)$$

is selected from the set

$$Y = \{y_i \mid i = 1, \dots, D\}, \quad (13)$$

where  $D$  is the number of extracted features and  $d \leq D$  denotes the size of the feature subset [14, 15]. A feature selection criterion function  $J(X)$  evaluates a chosen subset  $X$ ,

TABLE 1: List of features used for the feature selection experiments.

No.	Feature name	Method
1	Area of all segments	
2	Boundary length of all segments	
3	Compactness of all segments	
4	Center of gravity in $x$	
5	Center of gravity in $y$	
6	Polar measure minimum	
7	Polar measure maximum	
8	Polar measure mean	
9	Ratio of width to height of all segments	
10	Height of all segments	
11	Width of all segments	
12	Roundness of all segments	
13	Number of discontinuation in the horizontal projection	
14	Number of discontinuation in the vertical projection	
15	Number of free columns within the object	
16	Number of free rows within the object	
17	Number of objects	
18	Area of the largest segment	Region-based features
19	Boundary length of the largest segment	
20	Direction of the largest segment	
21	Length of the bounding rectangle of the largest segment	
22	Width of the bounding rectangle of the largest segment	
23	Rectangularity of the largest segment	
24	Compactness of the largest segment	
25	Axis ratio of the bounding rectangle	
26	$a_0$ of the first order fit of the horizontal projection of the largest segment	
27	$a_1$ of the first order fit of the horizontal projection of the largest segment	
28	Angle of the regression line of the horizontal projection of the largest segment	
29	Absolute error of the regression line to the horizontal projection	
30	$b_0$ of the second-order fit of the horizontal projection of the largest segment	
31	$b_1$ of the second-order fit of the horizontal projection of the largest segment	
32	$b_2$ of the second-order fit of the horizontal projection of the largest segment	
33	Absolute error of the second-order fit to the horizontal projection	
34	Ratio of the areas	
35	Eccentricity	
36	Ratio area/width	
37	Mean grey value of all segments	Statistical features
38	Standard deviation of all segments	
39	Mean grey value of the background	
40	Standard deviation of the background	
41	Ratio of the mean of the segments and the background	
42	Ratio of the standard deviation of the segments and the background	
43	Maximum grey value of the image	
44	Minimum grey value of the image	
45	Moment of 4th order of the image	
46	Skewness of the image	
47	Standard deviation of the image	
48	Mean of the image	
49	Second harmonic coefficient of the Fourier descriptor	Contour-based features
50	Third harmonic coefficient of the Fourier descriptor	
51	Fourth harmonic coefficient of the Fourier descriptor	
52	Fifth harmonic coefficient of the Fourier descriptor	
53	Maximum of the gradient in $x$ direction	Gradient features
54	Maximum of the gradient in $y$ direction	

whereby a higher value of  $J$  indicates a better subset. Formally, the problem of feature selection is to find a subset  $X \subseteq Y$  such that the number of chosen features  $|X|$  is  $d$  and  $J$  reaches the maximum

$$J(X^{\text{opt}}) = \max_{X \subseteq Y, |X|=d} J(X). \quad (14)$$

The evaluation criterion  $J$  is proposed to be the performance of a statistical classifier (classification rate) used as decision rule. Other evaluation measures are available, but Dash and Liu [13] showed that the best accuracy of selecting the subset is achieved by using the same classifier as for the online application. Unfortunately, this is computationally demanding.

Extensive investigations were carried out with suboptimal sequential feature selection algorithms. Further, genetic algorithms were applied to the same data set for feature selection (see [12]). Sequential feature selection algorithms search in a sequential deterministic manner for the suboptimal best feature subset. The forward methods start with an empty set and add features until a stopping criterion concludes the search. The backward algorithms are the counterpart. They begin with all features selected and remove features iteratively. The well-known suboptimal sequential algorithms are listed and a comparative study is given afterwards.

(1) Sequential forward selection (SFS): with each iteration one feature among the remaining features is added to the subset, so that the subset maximizes the evaluation criterion  $J$ .

(2) Sequential backward selection (SBS): in each step one feature is rejected so that the remaining subset gives the best result.

(3) Plus  $l$ -take away  $r$  selection (PTA ( $l, r$ )): this iterative method enlarges the subset by adding  $l$  features with the help of the SFS algorithm in the first step. Afterwards  $r$  features are removed with the SBS algorithm.

(4) Generalized sequential forward selection (GSFS ( $r$ )): at each stage  $r$  features are added simultaneously instead of adding just one feature to the subset at a time like the SFS method.

(5) Generalized sequential backward selection (GSBS ( $r$ )) is the counterpart of the GSFS method.

(6) Generalized plus  $l$ -take away  $r$  selection (GPTA ( $l, r$ )): the difference between the PTA and the GPTA method is that the former approach employs the SFS and the SBS procedures instead of the GSFS and GSBS algorithms.

(7) Sequential forward floating selection (SFFS): the SFFS includes new features with the help of the SFS procedure. Afterwards conditional exclusions of the worst features in the previously updated subset take place. Therefore, the parameters  $l$  and  $r$  are superfluous.

(8) Sequential backward floating selection (SBFS). This method is the counterpart of the SFFS method.

(9) Adaptive sequential forward floating selection (ASFFS ( $r_{\text{max}}, b, d$ )): this algorithm is similar to the SFFS procedure where the SFS and the SBS methods are replaced by their generalized versions GSFS ( $r$ ) and GSBS ( $r$ ). The level of generalization  $r$  is determined dynamically. For a detailed description we refer to [16].

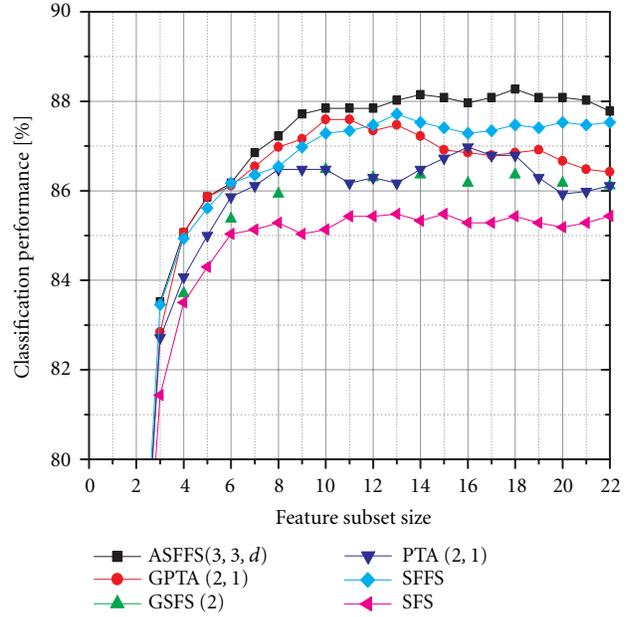


FIGURE 13: Results of feature selection obtained by different sequential optimization methods for a subset size up to  $d = 22$  using the 3NN classification result as evaluation criterion. The performance is achieved by averaging four optimized classification results gained over a rotation scheme of four data set parts.

(10) Adaptive sequential backward floating selection (ASBFS ( $r_{\text{max}}, b, d$ )): this method is the counterpart of the ASFFS algorithm.

The SFFS and SBFS algorithms are described in Pudil et al [17]. The recently published adaptive floating algorithms (ASFFS, ASBFS) are presented in [16]. All other sequential methods are explained in [14, 18]. Some interesting comparative studies of feature selection algorithms are shown in [15, 19].

The feature selection algorithms were applied to a data set of 540 flaw images equally distributed into five classes gained from the surface inspection and a total of  $D = 54$  extracted features of each image were used. The feature selection algorithms were compared in terms of the quality of the classification result. The performance of the 3-nearest-neighbor decision rule serves as optimization criterion. The evaluation of the sequential forward algorithms is shown in Figure 13.

Generally, the floating algorithms perform better than their nonfloating counterparts. The adaptive floating method (ASFFS ( $3, 3, d$ )) performs best.<sup>1</sup> Whereby, there is only a marginal difference to the result obtained by the classical floating algorithm (SFFS). For a small subset size,

<sup>1</sup>For the ASFFS ( $3, 3, d$ ) method, different optimization runs for each subset size  $d$  were performed and the best achieved results are summarized. With the given parameter setting, the classification performance of a subset size of  $d$  within a neighborhood of 3 is optimized more thoroughly, whereby the highest level of generalization is 3.

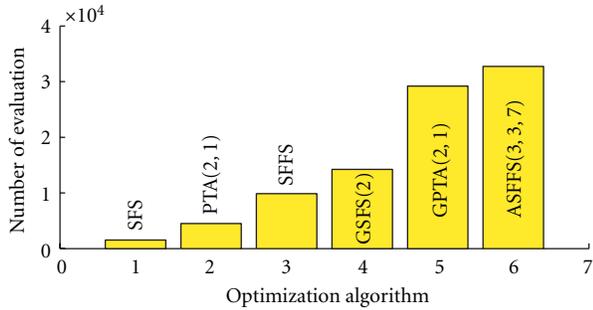


FIGURE 14: Approximate number of evaluations for different sequential selection algorithms.

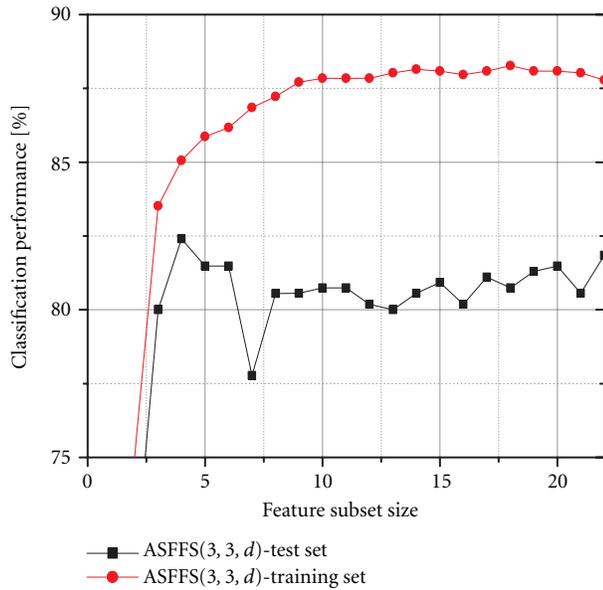


FIGURE 15: Verification of the results of the ASFFS(3,3,d) algorithm obtained by using the 3NN decision rule as evaluation criterion.

the generalized PTA method achieves roughly the same performance as the ASFFS algorithm. The approximate number of evaluations performed during optimization by different algorithms is shown in Figure 14. The number of all nonfloating algorithms is fixed for a given parameter setting. For the floating algorithms the number also depends on the training set due to the floating property. The adaptive floating algorithm with the chosen parameters is approximately three times more costly than the classical floating algorithm.

Due to overfitting of the training set the selected features for an optimal classification of the training data produce inferior accuracy on independent test data. In practice the result of the feature selection optimization must be validated with a test set of objects [13]. Therefore, several methods are suggested in the literature [20, 21].

The four performances obtained by optimizing with four data set parts (rotation scheme) and the classification estimate of the best feature subsets found for the remaining test data are averaged and shown in Figure 15.

The effect that a small subset of features yields a better classification performance on test data is apparent. The best classification performance of the training data is obtained with a feature subset size of 18, whereby the best classification performance of the test data set is achieved for a feature subset size around 4. Therefore, the chosen features are those with numbers 19, 22, 24, and 41 of Table 1.

For the decision of the optimal size of the feature subset some conflicting issues should be taken into account. Discriminatory information may be lost in choosing too few features. Otherwise, a smaller subset results in lower computational effort and the designed classifier yields a better generalization.

#### 5.4. Statistical classification

Prior to classification the descriptors were scaled [22] by their sample means and standard deviations. Each feature  $x'_{mn}$  is normalized using the following equation:

$$x_{mn} = \frac{x'_{mn} - \mu_n}{\sigma_n}, \quad m = 1, 2, \dots, M, \quad n = 1, 2, \dots, N, \quad (15)$$

where

$$\mu_n = \frac{1}{M} \sum_{m=1}^M x'_{mn}, \quad \sigma_n = \sqrt{\frac{1}{M} \sum_{m=1}^M (x'_{mn} - \mu_n)^2}, \quad (16)$$

are the mean value and the standard deviation,  $N$  is the number of features, and  $M$  is the number of samples. This scaling approach achieves an objective normalization of the different features with the statistical properties of  $\mu_n = 0$  and  $\sigma_n = 1$ .

During classification the object is assigned to a certain class according to the selected features. The Bayes decision rule is a fundamental probabilistic approach to statistical pattern recognition [23, 24, 25].

Patterns  $\mathbf{x}$  are assigned to a class  $\omega_j$  as

$$\mathbf{x} \rightarrow \omega_j \quad \text{if } P(\omega_j | \mathbf{x}) = \max_{i=1, \dots, t} P(\omega_i | \mathbf{x}), \quad (17)$$

where  $t$  is the number of classes,  $P(\omega_i | \mathbf{x})$  is the a posteriori probability of the  $i$ th class computed by the probability density function  $p(\mathbf{x} | \omega_i)$  as

$$P(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)P(\omega_i)}{\sum_{j=1}^t p(\mathbf{x} | \omega_j)P(\omega_j)}. \quad (18)$$

Various classification schemes were developed by estimating or approximating the class density function  $p(\mathbf{x} | \omega_i)$  and the a priori class probability  $P(\omega_i)$ . Simple decision rules assume this probability density function as parametric form. Most frequently the multivariate normal density distribution, which is characterized by the mean vector  $\boldsymbol{\mu}_i$  and the covariance matrix  $\boldsymbol{\Sigma}_i$  for each class  $i$ , is assumed. This leads to the well-known minimum distance and Mahalanobis distance decision rules given in [23].

Due to the strong overlap of the classes in the feature space of this classification task parametric decision rules are

TABLE 2: Classification table of a set of 108 observations in each class using a 3NN decision rule.

Correct class	Predicted class				
	1	2	3	4	5
1	94 17.41%	1 0.19%	2 0.37%	0 0.00%	11 2.04%
2	0 0.00%	95 17.59%	3 0.56%	10 1.85%	0 0.00%
3	3 0.56%	1 0.19%	94 17.41%	3 0.56%	7 1.30%
4	0 0.00%	10 1.85%	5 0.93%	88 16.30%	5 0.93%
5	8 1.48%	1 0.19%	7 1.30%	2 0.37%	90 16.67%

inferior compared to the nonparametric decision rules such as the  $k$ -nearest-neighbor ( $k$ NN) approach.

The proposed  $k$ NN decision rule aims to estimate directly the a posteriori probability function  $P(\omega_i | \mathbf{x})$ . A reference set of labeled samples is available and  $k$ -nearest-neighbors of the actually unknown object are extracted. The new object is assigned to the most frequently occurring class of its neighbors stated as follows:

$$\mathbf{x} \rightarrow \omega_j \quad \text{if } k_j = \max_{i=1,\dots,t} k_i. \quad (19)$$

For a finite training set the  $k$ NN decision rule will approximate the Bayes rule [25]. To establish the optimal value of  $k$ , Raudys and Jain [21] recommend estimating the classification performance for a different number of  $k$ .

Finally, for the classification experiments, 108 samples for each of the 5 classes of flaws are classified using the established four features. For the classification the actual chosen object is removed from the reference set of the  $k$ NN decision rule. In this manner, 85.37% of the data is classified correctly (see Table 2).

The main characteristics of the classification table are summarized as follows:

- (i) wrong classified objects of class 1 are mostly assigned to class 5 and vice versa;
- (ii) observations which belong to class 3 are, in the case of misclassification, mostly categorized to class 5;
- (iii) objects of class 2 and class 4 also strongly interfere.

In consideration of the fact that flaws of types 4 and 5 on the bearing rolls should not lead to a removal from the production lot, the performed classification result is not satisfactory. Therefore, an extension of the classification with a rejection class is suggested.

## 6. CONCLUSION AND FURTHER WORK

This paper presents a prototype for the detection and classification of flaws on bearing rolls.

The first part concentrates on the optical reflectance properties of the rough surface of the machined rolls. The

change of the microstructure of the surface in case of an error strongly influences the reflectance function. This was shown with a reflectance model using the roughness measurements, and the results were validated with optical experiments. Additionally, considerations about the optimal illumination angle for the lightline source used to achieve a constant illumination across the bearing roll were performed.

The second part deals with the issues of flaw detection, segmentation, feature extraction, feature selection, and classification. A robust adaptive segmentation algorithm, which works on unimodal and bimodal histograms based on maximizing the area of a triangle embedded in the smoothed histogram, was shown. In order to classify the detected flaws, features were derived, and the most significant features which optimize the classification performance were extracted using sequential feature selection algorithms. Therefore, a comparative study of suboptimal feature selection methods has been performed. The effect that a small subset of features yields a better classification performance on an independent test data set was shown.

The future work focuses on algorithms for the classification of the flaws. There exists a huge variety of different classification algorithms which might perform better than the current  $k$ NN decision rule. Special attention will be dedicated to Bayesian network classifiers which seem to be the most powerful decision rules at present. Additionally, a combination of different classifiers could increase the classification performance.

## ACKNOWLEDGMENTS

The authors are grateful to E. G. Badisch for measuring the roughness profiles of the surfaces and to E. Pernkopf for her valuable comments.

## REFERENCES

- [1] B. Smith, "Making war on defects: six-sigma design," *IEEE Spectrum*, vol. 30, no. 9, pp. 43–47, 1993.
- [2] T. S. Newman and A. K. Jain, "A survey of automated visual inspection," *Computer Vision and Image Understanding*, vol. 61, no. 2, pp. 231–262, 1995.
- [3] F. Dupont, C. Odet, and M. Carton, "Optimization of the recognition of defects in flat steel products with the cost matrices theory," *NDT and E International*, vol. 30, no. 1, pp. 3–10, 1997.
- [4] S. A. Stefani, C. R. Nagarajah, and R. Willgross, "A surface inspection technique for continuously extruded cylindrical products," *Mesurement Science and Technology*, vol. 10, no. 2, pp. N21–N25, 1999.
- [5] P. Beckmann and A. Spizzochino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Pergamon, New York, NY, USA, 1963.
- [6] K. E. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *Journal of the Optical Society of America*, vol. 57, no. 9, pp. 1105–1114, 1967.
- [7] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface reflection: Physical and geometrical perspectives," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 611–634, 1991.

- [8] P. Kierkegaard, *Estimating circular shape of machined parts by computer vision*, Ph.D. thesis, University of Linköping, Linköping, Sweden, 1995.
- [9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Mass, USA, 1992.
- [10] D.-M. Tsai, "A fast thresholding selection procedure for multimodal and unimodal histograms," *Pattern Recognition Letters*, vol. 16, no. 1, pp. 653–666, 1995.
- [11] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, International Thomson Publishing, London, UK, 2nd edition, 1999.
- [12] F. Pernkopf, *Automatic visual inspection of metallic surfaces*, Ph.D. thesis, University of Leoben, Leoben, Austria, 2002.
- [13] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.
- [14] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1982.
- [15] A. K. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.
- [16] P. Somol, P. Pudil, J. Novovičová, and P. Paclík, "Adaptive floating search methods in feature selection," *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1157–1163, 1999.
- [17] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [18] J. Kittler, "Feature set search algorithms," in *Pattern Recognition and Signal Processing*, C. H. Chen, Ed., pp. 41–60, 1978.
- [19] F. J. Ferri, P. Pudil, M. Hatef, and J. Kittler, "Comparative study of techniques for large-scale feature selection," in *Pattern Recognition in Practice IV*, E. S. Gelsema and L. N. Kanal, Eds., pp. 403–413, 1994.
- [20] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. of the 14th International Joint Conference on Artificial Intelligence*, pp. 1137–1143, Montreal, Canada, 1995.
- [21] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 252–264, 1991.
- [22] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley and Sons, New York, NY, USA, 1990.
- [23] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley-Interscience Publications, New York, NY, USA, 1973.
- [24] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, NY, USA, 1990.
- [25] J. Kittler, "Statistical classification," *Vistas in Astronomy*, vol. 41, no. 3, pp. 405–410, 1997.

**Paul O'Leary** received a Bachelor in mathematics and a Bachelor in electronic engineering from the Trinity College, Dublin, Ireland, in 1982. Then he joined the Philips International Institute, Eindhoven, Netherlands, where he received the M.S. degree in 1984. He earned a Ph.D. degree in signal processing from the University of Pavia, Italy, 1993. Currently he is Professor at the Institute of Automation, University of Leoben.

**Franz Pernkopf** received his M.S. degree in electrical engineering at Graz University of Technology, Austria, in summer 1999. He earned a Ph.D. degree from the University of Leoben, Austria, 2002. His research interests include surface inspection, feature selection, and statistical pattern recognition.

