Pattern Anal Applic (2004) 7: 333–342
DOI 10.1007/s10044-004-0232-3

THEORETICAL ADVANCES

Franz Pernkopf

Detection of surface defects on raw steel blocks using Bayesian network classifiers

Received: 3 October 2003 / Accepted: 18 September 2004 / Published online: 19 November 2004

Abstract This paper proposes an approach that detects surface defects with three-dimensional characteristics on scale-covered steel blocks. The surface reflection properties of the flawless surface changes strongly. Light sectioning is used to acquire the surface range data of the steel block. These sections are arbitrarily located within a range of a few millimeters due to vibrations of the steel block on the conveyor. After the recovery of the depth map, segments of the surface are classified according to a set of extracted features by means of Bayesian network classifiers. For establishing the structure of the Bayesian network, a floating search algorithm is applied, which achieves a good tradeoff between classification performance and computational efficiency for structure learning. This search algorithm enables conditional exclusions of previously added attributes and/or arcs from the network. The experiments show that the selective unrestricted Bayesian network classifier outperforms the naïve Bayes and the tree-augmented naïve Bayes decision rules concerning the classification rate. More than 98% of the surface segments have been classified correctly.

Keywords Surface inspection · Range imaging · Bayesian network classifier · Feature selection · Light sectioning

1 Introduction

There is an increasing demand in industry for automatic inspection systems to control the quality of products. The ever more stringent customer demands are well-founded on the high costs incurred for poor quality, due to the resulting costs for correction. Newman and Jain [1] defined the task of inspection as "inspection is the process of determining if a product deviates from a given set of specifications."

Basically, two different approaches for acquiring the surface image are considered in the literature: intensity imaging techniques and range imaging methods [1, 2].

Many surface inspection approaches for steel products based on intensity imaging have been proposed. In this introduction, we briefly present the most prominent systems. Dupont et al. [3] proposed a system for flat steel products that is optimized using a cost matrix approach. They achieve a recognition rate of 84.5% with a multilayer neural network. Pernkopf and O'Leary [4] presented an inspection system for machined metallic high-precision surfaces, such as bearing rolls. An optimal illumination setup in terms of maximizing the contrast between surface defects and a flawless surface is derived by considering the machined surface as a composition of microfacets using the Torrance-Sparrow reflection model [5, 6]. They achieve a classification rate of 85.4% using a k-NN classifier based on a selected subset of features. Platero et al. [7] developed a system for the detection of eight types of defects on continuous aluminum strips with a recognition performance of 95.7%. Many more systems for the inspection of metallic surfaces have been suggested in various proceedings of the International Society for Optical Engineering (Proceedings of SPIE) [8]. All these approaches, unlike our inspection application, have in common that the reflection property and, accordingly, the optical appearance of the flawless surface is homogeneous. Inspection systems that use range images have mostly been applied for complex objects, solder joints, and printed circuit boards. A comprehensive overview is provided in [1, 8].

For many inspection applications of metallic surfaces, an acceptable intensity image cannot be produced neither with bright field or dark field lighting, nor with diffuse illumination. This is the case if the reflection

F. Pernkopf (✉)
Department of Electrical Engineering,
University of Washington, M254 EE/CSE Building,
Box 352500, WA Seattle, 98195-2500, USA

Present address: F. Pernkopf
Institute of Signal Processing and Speech Communication,
Graz University of Technology,
Inffeldgasse 12, 8010 Graz, Austria
E-mail: pernkopf@tugraz.at

property across the intact surface changes and the defects may not be emphasized with respect to their background using intensity imaging. Surface defects with three-dimensional characteristics, e.g., cavities and cracks, are visualized with a higher contrast by means of range imaging. Depending on the range imaging approach [9], the data might be less affected by a change of the reflection property across the flawless surface.

In this paper, we deal with the surface inspection of milled steel blocks, which are partially covered with scale. The cross-section of the block is approximately quadratic, with a size varying from 130×130 mm to 160×160 mm, whereby the edges are round. The length is about 10 m. Defects with a minimal length of 3 mm, a width of 0.5 mm, and a depth of 1 mm have to be identified on blocks moving at a speed of 1.85 m/s.

The reflection properties across the flawless surface changes strongly due to scale on the surface. Since intensity imaging results in a poor performance, we use range imaging based on light sectioning techniques to acquire the surface data of the steel block with its embedded flaws. Of course, the variations of the surface reflectance affects the range accuracy [10]. However, we are interested in a qualitative measure of the surface finish. The exact dimensions of the product are irrelevant in this application. The range image is less sensitive to the inhomogeneous reflectance of the flawless surface. When we use light sectioning to acquire the surface image, the data are affected by vibrations caused by the movement of the steel block on the conveyor. After the recovery of the depth map from the surface range data, segments of the surface are classified according to a set of extracted features by means of Bayesian network classifiers. The segments of the surface should be classified into three classes:

1. Surface flaws: basically, there are different types of flaws that might be classified according to their genesis during production. We treat them as non-specific surface errors, since they have the degradation of the surface quality in common. Defects occur at the round edge of the steel block. Fig. 1 shows different flaws and their dimensions. Additionally, the inhomogeneity of the surface caused by scale is observable.
2. Pseudo errors: pseudo errors are caused by an extremely inhomogeneous surface reflection property, which results in a spurious specular reflection [6] of the light in the direction of the imaging sensor. This entails an overmodulation of the sensor and distorted range data. In particular, we are confronted with these errors when the scale of the surface is removed mechanically through the carrier apparatus during the production. Thus, small surface areas appear shiny. These pseudo errors cannot be avoided by changing the setup of the light sectioning technique.
3. Intact surface segment: pseudo errors and intact surface segments do not degrade the quality of the steel blocks.

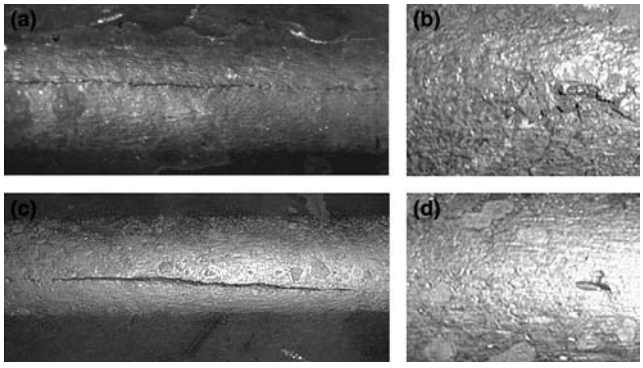

Fig. 1a–d Different surface flaws. **a** 30×0.5 mm. **b** 10×7 mm. **c** 60×1 mm. **d** 3×1 mm

The remainder of the paper is organized as follows: Sect. 2 summarizes the data acquisition approach, with particular emphasis on light sectioning techniques. Section 3 treats a method for depth map extraction, whereby the vibration artifacts that exist due to the movement of the steel blocks on the conveyor have to be considered. Section 4 summarizes the extracted features used for classification. Different Bayesian network classifiers are presented in Sect. 5 and experimental results of the classification of the surface segments are shown in Sect. 6. The paper closes with a summary.

2 Data acquisition: light sectioning

The light sectioning method [2, 9–12] is a well-known measurement technique for the optical determination of sections of objects. We use a laser light stripe projected onto the steel block from one direction and viewed from a different direction using a camera. Through the known arrangement of the laser light source and the camera, the height information can be determined. The complete 3D model is gathered by moving the object in one direction while its sections are scanned in a sequential manner. The principle and physical setup for light sectioning is shown in Fig. 2.

One of the main drawbacks is that the sections are determined sequentially, which is time consuming. We use a fast camera that delivers more than 2,000 sections per second [13]. Each section is composed of 512 range values with a range resolution of 512 pixels. For detecting the smallest specified surface flaw, we acquire range data at rates of 1.2×10^6 samples/s. In our application, occlusions [12] may occur only in the case of flaws on the surface. This results in isolated missing data values, which are recovered by interpolation from neighboring regions.

As mentioned above, the data shows vibrations caused through the movement of the steel block on the conveyor. These vibrations result in a varying position of the acquired sections. Figure 3 shows a surface segment with an embedded crack and the vibration artifacts caused by the movement.

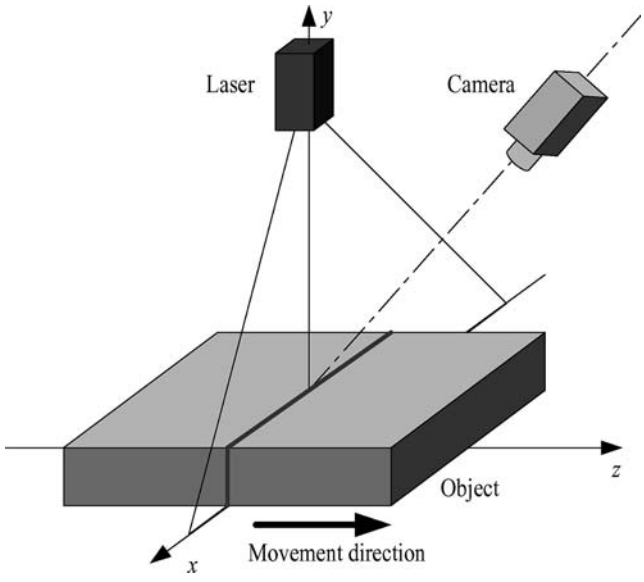

Fig. 2 Principle and physical setup for light sectioning

3 Depth map extraction using singular value decomposition

Before the features are extracted, we have to recover the depth map of the surface range data. Thereby, the vibration artifacts that exist due to the movement of the steel blocks on the conveyor have to be removed. The measuring head for light sectioning is focused on the round edge of the steel block. This means that the edge is in the center of the acquired sections (see Fig. 4a). In order to classify the data, the depth of the flaws has to be recovered. Therefore, a model of the three-dimensional shape of the surface is approximated. This approximation of the surface segment [14] is determined using singular value decomposition (SVD) [15]. Basically, many different algorithms exist for approximating a surface; however, this approach is simple and straightforward. The surface segment of range data $\mathbf{A}$ of size $m \times n$ is decomposed by means of SVD as $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}$, where $\mathbf{U}$ is an $m \times m$ orthogonal matrix, $\mathbf{V}$ is an $n \times n$ orthogonal matrix, and $\mathbf{S}$ is an $m \times n$ diagonal matrix composed of the singular values S_{ii} ordered from the largest to the smallest. It has been observed that the shapes of the surface segments are either planar or curved in one direction. Therefore, the first few singular values capture most of the variance of such simple surface shapes.

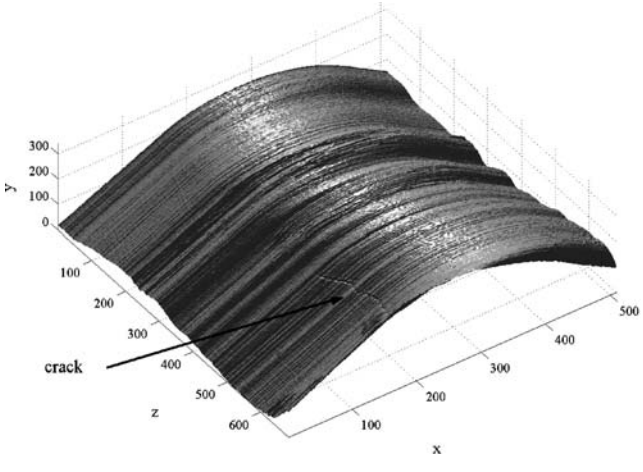

Fig. 3 Acquired surface data with an embedded crack

Hence, the idea is to reassemble a model of the surface segment by dropping all terms from the SVD that correspond to singular values smaller than, e.g., S_{22}. This means that all singular values except the first two are set to zero in $\mathbf{S}'$:

$$\mathbf{S}' = \begin{pmatrix} S_{11} & 0 & 0 & \cdots & 0 \\ 0 & S_{22} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{1}$$

and the smoothed surface approximant (or model) $\mathbf{A}^{\mathrm{a}}$ is obtained according to $\mathbf{A}^{\mathrm{a}} = \mathbf{U}\mathbf{S}'\mathbf{V}^{\mathrm{T}}$. The resulting model of the surface in Fig. 4a is shown in Fig. 4b.

Once we have the model, the orthogonal distance between the original surface segment $\mathbf{A}$ and the surface approximant $\mathbf{A}^{\mathrm{a}}$ can be determined as proposed in [16]. However, a simple subtraction of the model from the range data ($\mathbf{A} - \mathbf{A}^{\mathrm{a}}$) giving the depth map is sufficient (see Fig. 4c), since the differences contributed by the orthogonal distance are negligible due to the noise in the data. Our approach for surface approximation compensates for the vibrations in the data if the size of the surface segments is moderate. An approach for removing the vibrations in the data by means of a geometric transformation is given in [16].

4 Feature extraction

In order to evaluate the quality of the surface, the data are divided into a set of overlapping segments (block size 30×30), whereby each segment is represented by 40 features. According to these features, the segment is assigned to one of the classes specified in Sect. 1. The features should discriminate well between erroneous surface segments, intact surface blocks, and pseudo errors caused by inhomogeneous surface reflection properties. The relevant features in this classification task are not known at the outset. Thus, we derive many features and use a selected subset for classification. With respect to the optical appearance of the flaws, we mainly focus on features that describe the geometric shape and the statistics of the flaws. These features are summarized in Table 1.

Basically, five different methods for feature extraction were used in this application. These approaches may be roughly divided into statistical features, SVD descriptors, wavelet transform features, region-based methods,

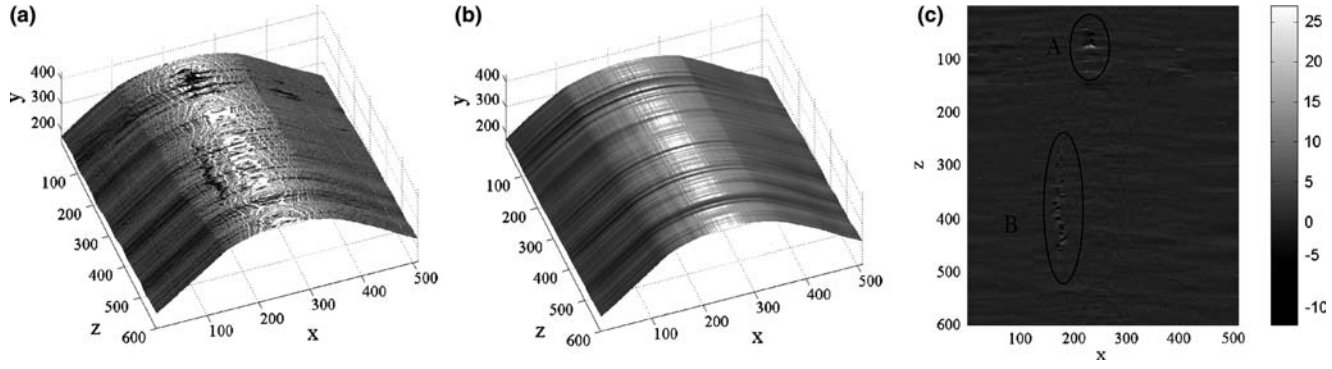

Fig. 4a–c Surface approximation. **a** Acquired surface segment (with vibrations). **b** Smoothed surface approximant. **c** Depth map (A = pseudo error, B = surface error).

and features from the discrete cosine transform (DCT). As shown in Table 1 (see *Basis* column), the descriptors are either derived from segments (30×30) of the range data directly, or from the depth map:

- Singular value decomposition features: the first few singular values derived directly from the range data of surface segments are used as descriptors [15, 17]. For flawless surface segments, the first singular value covers most of the variance.
- Wavelet transform features: the surface sections are transformed by means of a 1-dimensional discrete wavelet transform using Daubechies wavelets of the fourth order. The first-level detail coefficients show the discontinuity in the erroneous sections. Therefore, the maximum, the minimum, and the areas of the thresholded coefficient image serve as features. The mean polar distance is the mean distance from all set pixels of the thresholded coefficient image to the center of gravity [18].
- Discrete cosine transform (DCT) features: the lowest frequencies of the 2-dimensional DCT [19] are used as descriptors of the range data segment. The DCT has the property that most of the significant information about the surface segments is concentrated in a few coefficients of the DCT.
- Statistical features: first-order statistics involve parameters such as mean, variance, skewness, and kurtosis [20]. These features are determined from the distribution of the depth values in the depth map of surface segments. The variance is a measure of the spread of the histogram. The skewness evaluates the skew of a distribution. The distribution belongs to a negative skew in the case where the left tail is larger than the right one, and vice versa. The kurtosis is a measure of how outlier-prone the distribution is. This measure evaluates the deviation of the distribution from the normal distribution and is used for the detection of very small flaws with an extreme depth value.
- Region-based features: simple region-based descriptors are derived from the thresholded depth map. The aim is to describe the geometric shape of the flaws. The threshold is selected to be $T=1$ and $T=-1$. These features are described in [19, 20]. The error likelihood measure is described in detail in [21]. Due to the fact that the flaws on the surface of the steel block are existent in neighboring sections, this measure considers the neighborhood of flaws on subsequent sections. Therefore, a kernel is used for weighting the neighborhood of a defect. The shape of the kernel can be Gaussian. However, a special kernel was developed for emphasizing a preference direction of the flaws on the surface [21].

5 Bayesian network classifier

In classification problems, the relevant features are often unknown a priori. Thus, many features are derived and those which do not contribute or even degrade the classification performance are removed from the set of extracted features during classification. Feature selection has become important for numerous pattern recognition and data analysis approaches [22–24]. The main purpose of feature selection is to reduce the number of extracted features to a set of a few significant ones, whilst maintaining the classification rate. The reduction of the feature set may even improve the classification rate by reducing estimation errors associated with finite sample size effects [25].

Another approach to achieve an improvement of the classification accuracy is to model statistical dependencies between attributes. Therefore, the framework of Bayesian networks [26] is used for building classifiers [27, 28].

A Bayesian network [26, 29–31] $B=\langle G, \Theta\rangle$ is a directed acyclic graph G, which models probabilistic relationships among a set of random variables $\mathbf{U}=\{X_1,..., X_n, \Omega\}=\{U_1,..., U_{n+1}\}$, where each variable in $\mathbf{U}$ has specific states or values denoted by lower case letters $\{x_1,..., x_n, \omega\}$. The symbol n denotes the number of attributes of the classifier. Each vertex (node) of the graph represents a random variable, while the edges capture the dependencies between the variables. The network encodes the conditional independence relationships so that each node is independent of its non-

Table 1 List of features used for the classification experiments. Features are based on data segments of size 30×30. Symbol T denotes the threshold value

No.	Feature name	Method	Basis
1	Second singular value S_{22}	SVD	Features based on range data surface segments
2	Third singular value S_{33}		
3	Fourth singular value S_{44}		
4	Maximum of first-level detail coefficients	Wavelet transform	
5	Minimum of first-level detail coefficients		
6	Area of thresholded ($T=1$) first-level detail coefficients		
7	Polar mean distance of thresholded ($T=1$) first-level detail coefficients		
8	Area of thresholded ($T=-1$) first-level detail coefficients		
9	Polar mean distance of thresholded ($T=-1$) first-level detail coefficients		
10	DCT coefficient $B_{0,0}$	2-dimensional discrete cosine transform (DCT)	
11	DCT coefficient $B_{0,1}$		
12	DCT coefficient $B_{0,2}$		
13	DCT coefficient $B_{0,3}$		
14	DCT coefficient $B_{1,0}$		
15	DCT coefficient $B_{2,0}$		
16	DCT coefficient $B_{3,0}$		
17	Standard deviation	1st order statistics	
18	Skewness		Features based on the depth map of surface data segments
19	Kurtosis		
20	Maximum of depth map	Region-based features	
21	Minimum of depth map		
22	Error likelihood measure		
23	Area of regions		
24	Polar mean distance of regions		
25	Euler number		Features based on thresholded depth map ($T=1$) of surface data segments
26	Eccentricity of the regions		
27	Length of the major axis of the bounding rectangle of the regions		
28	Length of the minor axis of the bounding rectangle of the regions		
29	Orientation of the bounding rectangle of the regions		
30	Diameter of a circle with equivalent area as the regions		
31	Ratio of the region area to the bounding rectangle		
32	Area of regions		
33	Polar mean distance of regions		
34	Euler number		Features based on thresholded depth map ($T=-1$) of surface data segments
35	Eccentricity of the regions		
36	Length of the major axis of the bounding rectangle of the regions		
37	Length of the minor axis of the bounding rectangle of the regions		
38	Orientation of the bounding rectangle of the regions		
39	Diameter of a circle with equivalent area as the regions		
40	Ratio of the region area to the bounding rectangle		

descendants, given its parents. These conditional independence relationships reduce the number of parameters needed to represent the probability distribution. The symbol Θ represents the set of parameters that quantify the network. Each node U_i is represented as a local conditional probability distribution, given its parents Π_{U_i}. The joint probability distribution of the network is determined by these local conditional probability distributions as:

$$P(\mathbf{U}) = \prod_{i=1}^{n+1} P(U_i|\Pi_{U_i}) \; . \tag{2}$$

Basically, two different techniques for parameter learning are available; the maximum likelihood estimation and the Bayesian approach [30]. In this paper, the parameters of the network are estimated by the maximum likelihood method. In the following, three different types of Bayesian network classifiers are presented: the naïve Bayes classifier, the tree-augmented naïve Bayes classifier, and the selective unrestricted Bayesian network classifier.

5.1 Naïve Bayes classifier

The naïve Bayes (NB) decision rule [32] assumes that all the attributes are conditionally independent, given the class label. As reported in the literature [27], the performance of the naïve Bayes classifier is surprisingly good, even if the independence assumption between attributes is unrealistic in most of the data sets. Independence between the features ignores any correlation among them. The attribute values of X_i and X_j ($X_i \neq X_j$) are conditionally independent, given the class label of node Ω. Hence, x_i is conditionally independent of x_j,

given class ω, whenever $P(x_i|\omega,\ x_j\ ) = P(x_i|\omega)$ for all $x_i \in X_i$, $x_j \in X_j$, $\omega \in \Omega$, and when $P(x_j, \omega) > 0$. The structure of the naïve Bayes classifier represented as a Bayesian network is illustrated in Fig. 5. Feature selection is introduced to this network by removing irrelevant features by means of a search algorithm (see Sect. 5.4). This extension of the naïve Bayes decision rule is known as the selective naïve Bayes classifier (SNB).

The structure in Fig. 5 shows that each attribute is conditionally independent of the remaining attributes, given the class label ω of the class variable. The class variable Ω is the only parent for each attribute X_i, denoted as $\Pi_{x_i} = \{\Omega\}$ for all $1 \le i \le n$. Hence, the joint probability distribution $P(X_1,..., X_n, \Omega)$ for this network is determined to be $P(X_1, \ldots, X_n, \Omega) = \prod_{i=1}^{n+1} P(U_i|\Pi_{U_i}) = P(\Omega)\prod_{i=1}^{n} P(X_i|\Omega)$, and, from the definition of conditional probability, the probability for the classes in Ω, given the values of the attributes, is $P(\Omega|X_1, \ldots, X_n) = \alpha P(\Omega)\prod_{i=1}^{n} P(X_i|\Omega)$, where α is a normalization constant.

5.2 Tree-augmented naïve Bayes classifier

Since the features may be correlated and the independence assumption of the naïve Bayes classifier is unrealistic, Friedman et al. [27] introduce the *tree-augmented naïve Bayes classifier* (TAN). It is based on the structure of the naïve Bayes network where the class variable is the parent of each attribute. Hence, the posterior probability $P(\Omega|X_1,..., X_n)$ takes all the attributes into account. Additionally, edges (arcs) among the attributes are allowed in order to capture the correlations among them. Each attribute may have at most one other attribute as an additional parent, which means that there is an arc in the graph from feature X_i to feature X_j. This implies that these two attributes X_i and X_j are not independent, given the class label. The influence of X_j on the class probabilities depends also on the value of X_i. An example of a tree-augmented naïve Bayes network is shown in Fig. 6. A tree-augmented naïve Bayes network is initialized as a naïve Bayes network. Additional arcs between attributes are learned by means of a search algorithm (see Sect. 5.4). The maximum number of arcs added to relax the independence assumption between the attributes is $n-1$.

5.3 Selective unrestricted Bayesian network classifier

The selective unrestricted Bayesian network classifier (SUN) [28] (see Fig. 7) can be viewed as a generalization of the tree-augmented naïve Bayes network. The class node is equally treated as an attribute node and may have attribute nodes as parents. The attributes need not be connected directly to the class node, as for the tree-augmented naïve Bayes network. After initialization, the network consists of the nodes without any arcs. A search algorithm (see Sect. 5.4) adds arcs to the network according to an evaluation criterion. If there is no arc between an attribute and the classifier network, then the attribute is not considered during classification. During the determination of the network structure, irrelevant features are not included and the classifier is based on a subset of selected features. This unrestricted network structure maximizes the classification performance by removing irrelevant features and relaxing the independence assumptions between correlated features.

Since this network is unrestricted, the computational demands for determining the network structure is huge, especially if there is a large number of attributes available. Additionally, the size of the conditional probability tables of the nodes increases exponentially with the number of parents. This might result in a more unreliable probability estimate of the nodes that have a large number of parents.

The posterior probability distribution of Ω given the value of all attributes is only sensitive to those attributes that form the Markov blanket of node Ω [26]. The Markov blanket of the class node Ω consists of the direct parents of Ω, the direct successors (children) of Ω, and all the direct parents of the direct successors (children) of the class node Ω. All the features outside the Markov blanket do not have any effect on the classification performance. Introducing this knowledge into the search algorithm reduces the search space and the computational effort for determining the structure of the classifier.

5.4 Search-and-score structure learning

We use the cross-validation classification accuracy estimate as the scoring function J for evaluating the

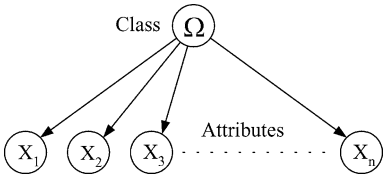

Fig. 5 Structure of a naïve Bayes (NB) network

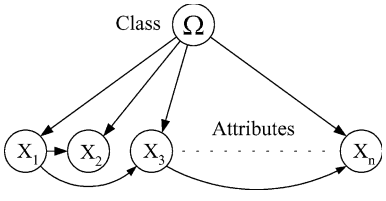

Fig. 6 Structure of a tree-augmented naïve Bayes network

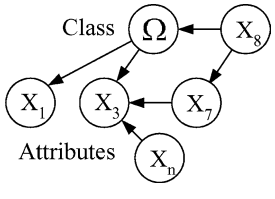

Fig. 7 Structure of a selective unrestricted Bayesian network

performance of the networks. Normally, a hill climbing search is used to learn the structure of the network. An improvement of the hill climbing search is to apply the classical floating search algorithm (CFS), which is used in feature selection applications [33]. We adopted this algorithm for learning the network structure of tree-augmented naïve Bayes classifiers and selective unrestricted Bayesian network classifiers [34]. The main disadvantage of the hill climbing search is that, once an arc has been added to the network structure, the algorithm has no mechanism for removing the arc at a later stage. Hence, this algorithm suffers from the *nesting* effect [35]. To overcome this drawback, the floating search method is used. This algorithm allows conditional exclusions of previously added attributes and/or arcs from the network. Hence, this algorithm is able to correct disadvantageous decisions that have been performed in previous steps. Therefore, it may approximate the optimal solution in a better way than (a) hill climbing search. However, this search strategy uses more evaluations to obtain the network structure and, therefore, it is computationally less efficient than (a) hill climbing search.

An efficient evaluation of the classifier during structure learning may be achieved by ordering the training instances so that the misclassified samples of previous classifications are classified first [36]. The classification algorithm can be terminated as soon as the number of misclassified samples exceeds the error rate of the current best classifier network.

6 Experimental results

Experiments have been performed on a data set $\mathcal{S}$ consisting of 516 surface segments uniformly distributed into three classes. Each sample (surface segment) is represented by 40 features (see Table 1). The data set is divided into six mutually exclusive subsets $\mathcal{S} = \{D_1, D_2, D_3, D_4, D_5, \mathcal{H}\}$. During the structure learning experiments, a five-fold cross-validation classification accuracy estimate is used as the scoring function J for finding the optimal structure of the network (see Fig. 8).

Therefore, the data set parts D_1, D_2, D_3, D_4, and D_5 are used. Each part is comprised of 90 samples. The accuracy estimate of the classifier is given by the successful predictions on the permutation of the data set parts D_1, D_2, D_3, D_4, and D_5. All the structure learning experiments are based on exactly the same cross-validation folds. The established classifiers are validated on a separate hold-out data set $1\mathcal{H}$, which has never been used during the optimization experiments [24].

The attributes in the data sets are continuous-valued. Since the classifiers are constructed for multinomial attributes, the features have been discretized in the manner described in Fayyad and Irani [37], whereby the partition boundaries for discretizing the continuous-valued attributes have been established only through the training data set of the corresponding cross-validation folds. Zero probabilities of the conditional probability tables are replaced with $\varepsilon = 0.00001$.

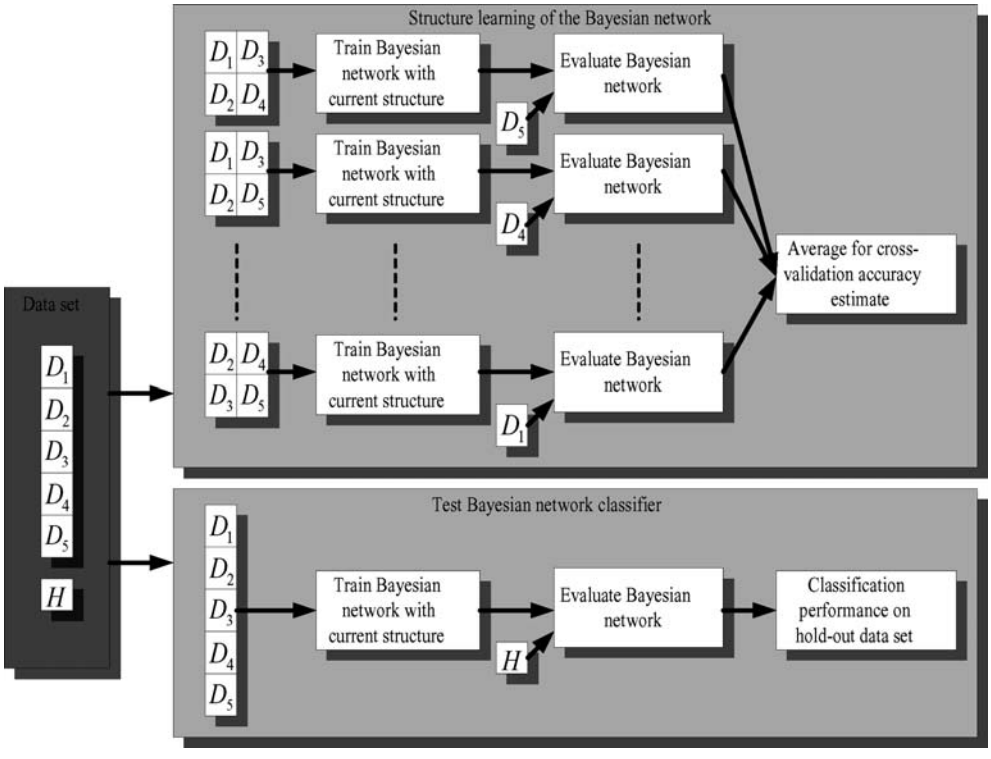

Fig. 8 Cross-validation method for accuracy estimation used for structure learning

Table 2 compares the classification methods discussed in previous sections. It gives the five-fold cross validation classification accuracy estimate (%CV5) and the performance on the hold-out data set (%H). It also shows the number of classifier evaluations (#Evaluations) used for the search, the number of independent probabilities (#Parameters), and the number of features (#Features) and arcs (#Arcs) used to achieve this classification accuracy estimate. The following abbreviations are used for the different classification approaches:

- **NB**: naïve Bayes classifier
- **CFS-SNB**: selective Naïve Bayes classifier using the classical floating search
- **HCS-TAN**: tree-augmented naïve Bayes classifier using the hill-climbing search
- **CFS-TAN**: tree-augmented naïve Bayes classifier using the classical floating search
- **CFS-SUN**: selective unrestricted Bayesian network using the classical floating search

The selective naïve Bayes classifier (CFS-SNB) achieves a better %CV5 classification accuracy estimate than the naïve Bayes (NB) approach based on all the available attributes. However, the performance on the hold-out data set is similar. The computational demands for establishing the CFS-SNB classifier is relatively small. For the tree-augmented naïve Bayes classifier, the same result is achieved either with the hill climbing (HCS) or with the classical floating search (CFS) algorithm. This means that the CFS method does not perform any backward steps during the search. This is also observable in the number of evaluations used. The TAN classifier uses all the extracted features and 12 arcs are added to the naïve Bayes structure (#Arcs = 52). The TAN classifier uses 533 independent probabilities, which have to be estimated from the data set. It is only slightly better than the selective naïve Bayes classifier. However, the CFS-SNB classifier has a much simpler structure and a smaller number of parameters is necessary. The selective unrestricted Bayesian network achieves the best classification accuracy estimate on the five-fold cross-validation and hold-out data set among the Bayesian network classifiers. For achieving this result, 230 probabilties have to be estimated and the structure (see Fig. 9) consists of 12 selected attributes and 14 arcs, whereas the TAN and NB structure do not enable feature selection. Additionally, the number of classifier evaluations used for determining the structure of the TAN network is high compared to learning the structure of the CFS-SUN, since the Markov blanket is used during the search for the SUN network structure.

Table 2 Comparison of classification approaches

	%CV5	%H	#Arcs	#Features	#Evaluations	#Parameters
NB	89.11 ± 2.87	95.45	40	40	1	275
CFS-SNB	96.44 ± 2.11	96.96	20	20	2,098	122
HCS-TAN	97.11 ± 2.02	96.96	52	40	17,893	533
CFS-TAN	97.11 ± 2.02	96.96	52	40	17,958	533
CFS-SUN	98.66 ± 0.81	98.48	14	12	4,097	230

7 Conclusion

In this paper, we present an approach for the detection and classification of small three-dimensional defects on scale-covered steel surfaces. The reflection properties across the flawless surface changes strongly, which results in a poor performance of intensity imaging. Thus, we use light sectioning to acquire the surface range image. Due to vibrations of the steel block on the conveyor, the acquired sections are arbitrarily located within a range of a few millimeters that has to be considered during the recovery of the depth map. According to a set of extracted features, segments of the surface data are classified by means of Bayesian network classifiers. A classification success rate of more than 98% shows the excellent performance of our approach in extremely harsh environments. The system fails if the surface flaws are completely covered with scale. Additionally, the appearance of the scale is dependent on the composition of the steel. This might lead to defect detection problems for different kinds of steel.

For learning the network structure of the tree-augmented naïve Bayes network and the selective unrestricted Bayesian network classifier, the classical sequential floating search algorithm is used. This algorithm is capable of removing previously added arcs/attributes at a later stage of the search if they turn out to be irrelevant.

The experiments on a data set of 516 surface segments show that the selective unrestricted Bayesian network classifier achieves a better five-fold cross-validation classification accuracy estimate during structure learning and a better classification performance on an external hold-out data set compared to tree-augmented naïve Bayes, selective naïve Bayes, and naïve Bayes classifiers. This classifier is based on a selected subset of features.

8 Originality and contributions

This paper proposes an approach for the surface analysis of scale-covered steel blocks. The data used in the

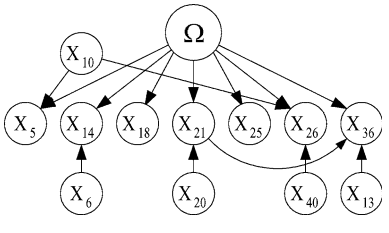

Fig. 9 Structure of the CFS-SUN classifier

paper were gained in a steel plant. The results show the excellent performance of our approach in extremely harsh environments. More than 98% of the surface segments have been classified correctly. We use light sectioning to acquire the range image of the surface of the steel block with its embedded flaws. This reduces the impact of a changing reflection property across the intact steel surface on the image quality. To our knowledge, a similar approach for defect detection has not been published before. This research is highly relevant for practical applications, in particular, for the steel industry.

For learning the network structure of the Bayesian network classifiers, we adopt the sequential forward floating feature selection algorithm [34]. This floating search algorithm facilitates the correction of disadvantageous decisions made in previous steps. Therefore, it may approximate the optimal solution in a better way than a hill climbing search. We show in this paper how well the structure learning algorithm performs on data from a practical application.

9 About the author

Franz Pernkopf received his MSc (Dipl. Ing.) degree in Electrical Engineering at Graz University of Technology, Austria, in the summer of 1999. He earned a PhD degree from the University of Leoben, Austria, in 2002. In 2002, he was awarded the Erwin Schrödinger Fellowship. He is currently a Research Associate in the Department of Electrical Engineering at the University of Washington, Seattle, USA. His research interests include Bayesian networks, feature selection, finite mixture models, and statistical pattern recognition.

Acknowledgements The author is grateful to Prof. Gernot Kubin for his valuable comments on this paper, and to Ingo Reindl and Voest Alpine Donawitz Stahl for providing the data. Parts of the work were done while the author was with the Institute of Automation at the University of Leoben, Austria.

References

1. Newman TS, Jain AK (1995) A survey of automated visual inspection. Computer Vis Image Und 61(2):231–262
2. Pernkopf F, O'Leary P (2003) Image acquisition techniques for automatic visual inspection of metallic surfaces. NDT&E Int 36(8):609–617
3. Dupont F, Odet C, Carton M (1997) Optimization of the recognition of defects in flat steel products with the cost matrices theory. NDT&E Int 30(1):3–10
4. Pernkopf F, O'Leary P (2002) Visual inspection of machined metallic high-precision surfaces. Eurasip J Appl Signal Process 2002(7):667–678
5. Torrance KE, Sparrow EM (1967) Theory for off-specular reflection from roughened surfaces. J Opt Soc Am 57(9):1105–1114
6. Nayar SK, Ikeuchi K, Kanade T (1991) Surface reflection: physical and geometrical perspectives. IEEE Trans Pattern Anal Machine Intell 13(7):611–634
7. Platero C, Fernandez C, Campoy P, Aracil R (1996) Surface analysis of cast aluminum by means of artificial vision and A.I. based techniques. Mach Vis Appl 2665:36–46
8. Machine vision applications in industrial inspection (various papers). Proceedings of SPIE. Homepage at http://www.spie.org/
9. Indyk D, Velastin SA (1994) Survey of range vision systems. Mechatronics 4(4):417–449
10. Curless B, Levoy M (1995) Better optical triangulation through spacetime analysis. In: Proceedings of the 5th international conference on computer vision (ICCV'95), Boston, Massachusetts, June 1996, pp 987–994
11. Kanade T (1987) Three-dimensional machine vision. Kluwer, Dordrecht, The Netherlands
12. Johannesson M (1995) SIMD architectures for range and radar imaging. PhD thesis, Department of Electrical Engineering, Linköping University, Sweden
13. SICK IVP (Integrated Vision Products). IVP Ranger SAH5 product information. Company homepage at http://www.ivp.se.
14. Long AE, Long CA (2001) Surface approximation and interpolation via matrix SVD. Coll Math J 32(1):20–25
15. Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn. John Hopkins University Press, Baltimore
16. Pernkopf F, Pernkopf F, O'Leary P (2002) Detection of surface defects on raw milled steel blocks using range imaging. In: Proceedings of the 14th IT&S/SPIE annual symposium on electronic imaging, San Jose, California, January 2002, pp 170–181
17. Datta BN (1995) Numerical linear algebra and applications. Brooks/Cole Publishing, Belmont, California
18. Pernkopf F (2002) Automatic visual inspection of metallic surfaces. PhD thesis, University of Leoben, Austria
19. Sonka M, Hlavac V, Boyle R (1999) Image processing, analysis, and machine vision, 2nd edn. International Thomson Publishing, London
20. Gonzalez RC, Woods RE (1992) Digital image processing, 3rd edn. Addison-Wesley, Reading, Massachusetts
21. Pernkopf F, O'Leary P (2003) Shape description and analysis of range data for milled steel blocks. In: Proceedings of the 15th IT&S/SPIE symposium on electronic imaging, Santa Clara, California, January 2003, pp 74–81
22. Jain AK, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. IEEE Trans Pattern Anal Machine Intell 19(2):153–158
23. Dash M, Liu H (1997) Feature selection for classification. Intell Data Anal 1(3):131–156
24. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1–2):273–324
25. Jain AK, Chandrasekaran B (1982) Dimensionality and sample size considerations in pattern recognition in practice, vol 2. In: Handbook of Statistics. North-Holland, Amsterdam
26. Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, San Francisco, California
27. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. Machine Learn 29:131–163
28. Singh M, Provan GM (1996) Efficient learning of selective Bayesian network classifiers. In: Proceedings of the 13th international conference on machine learning (ICML'96), Bari, Italy, July 1996, pp 453–461
29. Jensen FV (1996) An introduction to Bayesian networks. UCL Press, London
30. Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ (1999) Probabilistic networks and expert systems, 1st edn. Springer, Berlin Heidelberg New York
31. Heckerman D (1995) A tutorial on learning Bayesian networks. Technical report MSR-TR-95-06, Microsoft Research, Redmond, Washington
32. Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley, New York

33. Pudil P, Novovičová J, Kittler J (1994) Floating search methods in feature selection. Pattern Recogn Lett 15(11):1119–1125
34. Pernkopf F, O'Leary P (2003) Floating search algorithm for structure learning of Bayesian network classifiers. Pattern Recogn Lett 24(15):2839–2848
35. Kittler J (1978) Feature set search algorithms. In: Chen CH (ed) Pattern recognition and signal processing. Sijtho and Noordho, Alphen aan den Rijn, The Netherlands, pp 41–60
36. Keogh EJ, Pazzani MJ (1999) Learning augmented Bayesian classifiers: a comparison of distribution-based and classification-based approaches. In: Proceedings of the 7th international workshop on artificial intelligence and statistics (Uncertainty'99), Fort Lauderdale, Florida, January 1999, pp 225–230
37. Fayyad UM, Irani KB (1993) Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of the 13th international joint conference on artificial intelligence, Chambéry, France, August/September 1993, pp 1022–1027