Machine Vision and Applications (2003) 13: 275–285
Digital Object Identifier (DOI) 10.1007/s00138-002-0084-z

Machine Vision and Applications

Wood inspection with non-supervised clustering

Olli Silvén, Matti Niskanen, Hannu Kauppinen

Department of Electrical Engineering, University of Oulu, P.O.B. 4500, 90015 Oulu, Finland; e-mail: Olli.Silven@ee.oulu.fi

Received: 16 December 2000 / Accepted: 8 December 2001

Abstract. The appearance of sawn timber has huge natural variations that the human inspector easily compensates for mentally when determining the types of defects and the grade of each board. However, for automatic wood inspection systems these variations are a major source for complication. This makes it difficult to use textbook methodologies for visual inspection. These methodologies generally aim at systems that are trained in a supervised manner with samples of defects and good material, but selecting and labeling the samples is an error-prone process that limits the accuracy that can be achieved. We present a non-supervised clustering-based approach for detecting and recognizing defects in lumber boards. A key idea is to employ a self-organizing map (SOM) for discriminating between sound wood and defects. Human involvement needed for training is minimal. The approach has been tested with color images of lumber boards, and the achieved false detection and error escape rates are low. The approach also provides a self-intuitive visual user interface.

Key words: Self-organizing map – user interface – color – texture

1 Introduction

The visual grading of wood surfaces is a problem that combines very formal grading rules, large natural variations in the target material, and highly subjective appearance criteria. These factors make it a very challenging task, which is not ideal for humans, who have difficulty in following the strict and detailed grading standards. These standards consider every single defect type, size, and position, while human classification seems to be based on an overall impression, augmented with some specific rules, such as checking for the presence of knots at the edges of the board. Nonetheless, a human inspector seems to painlessly compensate for the variations in the material mentally, while this part of the task seems to be particularly difficult for machine vision systems used for automatic grading.

Correspondence to: O. Silvén

Human visual inspection rarely achieves better than 70% performance in grading lumber. Huber et al. (1985) found that in grading red oak lumber, human inspectors achieved 68% of the perfect result when recognizing, locating, and identifying defects. Grönlund (1995) found that when grading boards into four grades, only 60% of the boards were assigned the same grade by two different expert inspectors.

In automatic visual inspection, it is not straightforward to determine the actual accuracy achieved, as the results need to be compared to the ones of humans. Lampinen et al. (1994), in their comparative study, found that knots in pine boards could be classified at 80% accuracy. The test material consisted of about 400 high-quality color images of knots classified by experts. Kauppinen (1999) has determined that this result corresponds to 80% accuracy in grading boards when Nordic Timber (1994) rules are enforced. We believe this is the best that can be achieved with state-of-the-art automatic inspection systems applying this grading standard.

The performance of inspection systems suffers from the variations in the material; e.g., normal wood grain and knots may be difficult to discriminate in all circumstances. A big part of the blame falls upon the popular thresholding-based detection methods in use. In addition, the appearances of defects vary, so one board's sound knot is another board's dark one, requiring contextual information for proper classification. Figure 1 shows six knots classified as sound and another six classified as dry by a human expert. The range of variation within classes is substantial, causing difficulties for selecting features for classification.

The textbook methods proposed for visual inspection generally aim at systems that are trained by showing samples of defects and sound material as presented by Fig. 2. However, we believe this supervised training approach is fundamentally flawed in wood inspection, and in many comparable applications. Selecting and labeling the samples is an error-prone process that limits the accuracy that can be achieved, and this principle is not easily made adaptive to cope with the variations of material. The result is more likely to be a frustrating, everlasting training task rather than an ideal, accurate automatic visual-inspection device.

Not surprisingly, current state-of-the-art systems seem to employ rule-based classification, and that in itself is a rather

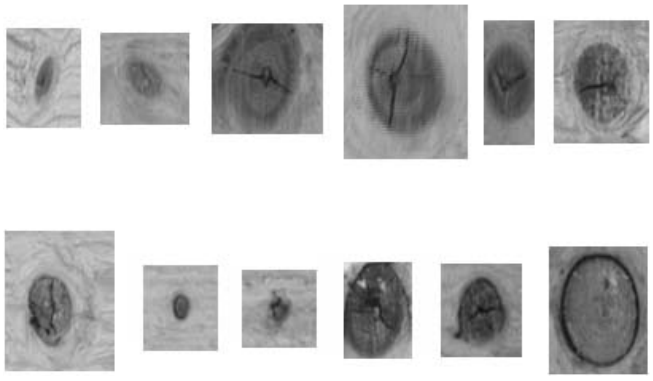


Fig. 1a,b. Examples of knots: the *top row* shows sound knots; the *bottom rows* shows dry knots. Within classes the variation is large; e.g., the color of some knots can be lighter than the background of others

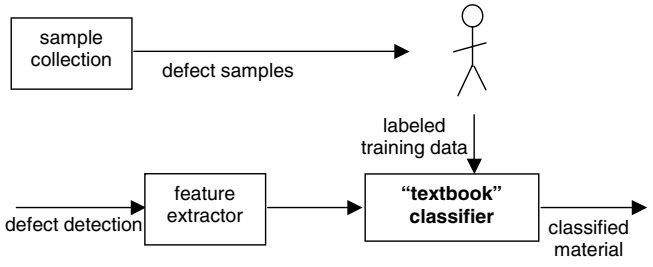


Fig. 2. The block diagram of visual surface inspection using supervised textbook classification

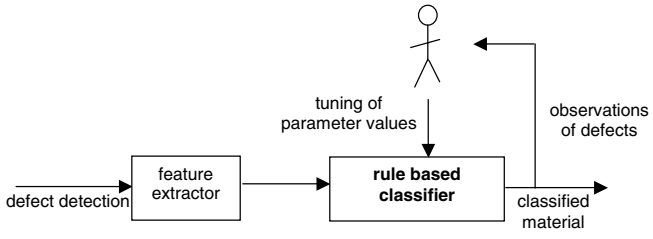


Fig. 3. Training the rule-based classifier: human expert tunes the parameters of the system based on observations on the output of the system

rigid approach. The tuning of the parameters requires the understanding of the classifier, which may take time to acquire. Rule-based classifier training is illustrated in Fig. 3.

In this paper, we present a solution for shortcomings in the detection and recognition of the defects of lumber boards, describe experiments giving quantitative data on performance, and discuss the consequences for inspection system design and the principles of grading. The proposed approach relies on non-supervised training and does not require labeling of individual samples, but uses a clustering method to discover whether the samples fall into a finite set of categories based on the similarity of their features. Thus, it is not affected by human errors in training. The approach also provides a self-intuitive visual user interface that is very easy to understand even for novices in the field. The block diagram of a SOM-based classifier is presented in Fig. 4.

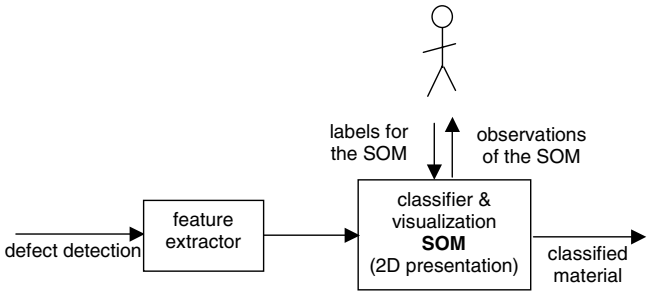


Fig. 4. The block diagram of inspection using the SOM-based classifier

2 Performance criteria

In order to rank and tune machine visual-inspection methods, performance criteria in good correlation with those achieved by human judgment are needed. That is, we have to compare the human results to those given by a machine; however, this is not always straightforward to do.

There are two different approaches to compare the performance of wood inspection methods. First, the grades of the classified boards can be compared to their "true" grades. In practice, these results depend highly on the applied grading rules and would be difficult to generalize for other grading standards. The inspection method that gives, say, a perfect result for one set of criteria may fail for another. The second approach, and the method we use in this paper, is to compare the results for individual defects. For this purpose, we need material in which the defects have been labeled; that is, the defects have been detected and recognized by some reasonably reliable means.

In our approach, wood inspection is divided into two phases: detection and recognition. Performances of both phases are evaluated separately, although the number and quality of detections affect recognition accuracy.

Defect-detection accuracy is often characterized by *defect-detection* (D), *error-escape* (E) and *false-alarm* (F) rates. The necessary information for calculating these are:

- the number of detections produced by the inspection system $= N_{\mathrm{det}}$,
- the number of the real labeled defects $= N_{\mathrm{lab}}$, and
- the number of detected labeled defects $= N_{\mathrm{det\_lab}}$.

The defect-detection rate indicates the proportion of detected labeled defects to real defects, and is determined by

$$D = N_{\mathrm{det\_lab}}/N_{\mathrm{lab}} \; . \tag{1}$$

The error-escape rate indicates the ratio of missed defects to real defects, and is related to the defect-detection rate as follows:

$$E = 1 - D \; . \tag{2}$$

The false-alarm rate is defined as the ratio of detections that are not in the area of labeled defects to the total number of detections

$$F = (N_{\mathrm{det}} - N_{\mathrm{det\_lab}})/N_{\mathrm{det}} \; . \tag{3}$$

Error-escape and defect-detection rates seem to include all the information needed to evaluate defect-detection accuracy.

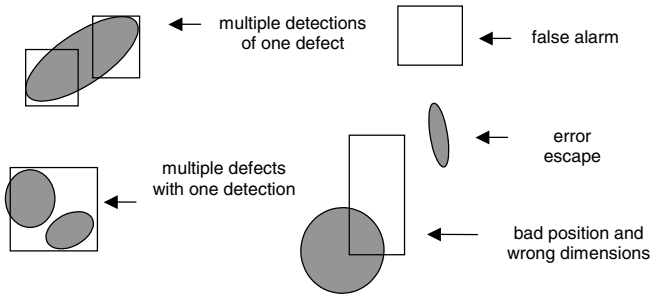


Fig. 5. Examples of detections. The *ellipses* denote real labeled defects and the *rectangles* denote detections made by the inspection system

In practice, the determination of the rates is not as clear-cut as described above. There may be several detections from a single defect, one detection may cover several defects, or there may be detections which do not match very well with the real defects or have wrong dimensions. Figure 5 illustrates these problematic, but very common situations. The number of detected and missed defects depends on the interpretation of results.

The error-escape and false-alarm rates do not contain any information about the quality of the detection. It is difficult to determine the class, exact position and size of the defect if the detection overlaps only a little with the real defect. A measure describing the quality of the match of the detection is, for example, the overlap of the detection and the corresponding defect. However, the boundaries of the defects are not always well defined and the "detection overlap" may be extremely difficult to determine. Evaluating the number of multiple detections may also be desired. In this presentation we evaluate the performance of the detection method using plain error-escape and false-alarm rates, but acknowledge a need for a more descriptive measure.

Some real detections presented in Fig. 6 illustrate the problems in evaluating the detection performance. There are no real defects in Fig. 6a, but there is a false alarm. There is one defect in Fig. 6b and two in Fig. 6c, but only one region is detected in Fig. 6c. In Fig. 6d there is an error escape, and in Fig. 6e, two detections from a single defect. One may say that there are two defect detections in Fig. 6f, even though the dimensions and position of detections do not match well with the real defect. Finally, the detection in Fig. 6g is open to various interpretations: in it, only the bark area of an encased knot has been detected while the knot itself is interpreted as sound wood.

The recognition accuracy is evaluated using a confusion matrix that shows the level of disagreement between automatic recognition and the human labelings for each class.

3 Self-organizing maps

The self-organizing map (SOM) is the key component in the detection and recognition method of this work. It is an algorithm used to visualize and interpret large high-dimensional data sets by projecting them to a low-dimensional space. The SOM theory and algorithm is presented thoroughly by Kohonen (1997).

Typically, a SOM consists of a two-dimensional regular grid of nodes, also known as neurons. A model of multidimensional observation m_i, eventually a vector consisting of features, is associated with each node. The training of the SOM is usually carried out by a sequential process, where for each data vector x in the training set, the closest node from the map is chosen. The choice of this node c, also called the best matching unit (BMU), satisfies

$$c = \arg\min_i \{\|x - m_i\|\} \ . \tag{4}$$

The BMU and its neighborhood kernel, which consists of a few surrounding nodes in the map, are then tuned to the direction of the training vector as

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \ , \tag{5}$$

where t is the discrete-time coordinate, and h_{ci} is a neighborhood function, a decreasing function of the distance between the node c and i on the map grid. One commonly used neighborhood function is Gaussian

$$h_{ci} = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \ , \tag{6}$$

where $\alpha(t)$ is the learning rate, and $\sigma(t)$ defines the width of the kernel.

After training, high-dimensional vectors of the nodes in the map have been tuned to represent the training data. In addition, the data vectors of neighboring nodes are located close in high-dimensional feature space, while dissimilar data is projected further apart from each other in the map. In its pure form, the SOM defines an "elastic net" of points that are fitted to the input signal space to approximate its density function in an ordered way (Kohonen 1997).

A trained SOM can be used as a non-supervised classifier. In classification, for a feature vector to be classified, the node from the SOM whose vector is most similar to it is chosen using Eq. (4). The chosen node determines the class for the classified feature vector. Classes can be assigned to nodes beforehand, such as after the training phase, or manually after clustering, based on an overview of the map.

4 The wood surface inspection approach

In our approach, wood inspection is divided into two phases. In the first one, we detect possible defected regions. In the second one, these are examined as individual defects. Classifications in both phases are based on non-supervised clustering, implemented using self-organizing maps (SOMs).

The test material used in our experiments has been prepared by VTT Building Technology. It consists of a large number of pine boards with ground truth classifications for each defect. The results reported here are based on a set of 42 board images totaling around 900 Mb of data and over 1000 labeled defects (Sommardahl and Usenius 1999). The imaging resolution has been 0.5 mm and a color line-scan camera has been used for image acquisition. The typical board size is 5 m by 0.35 m.

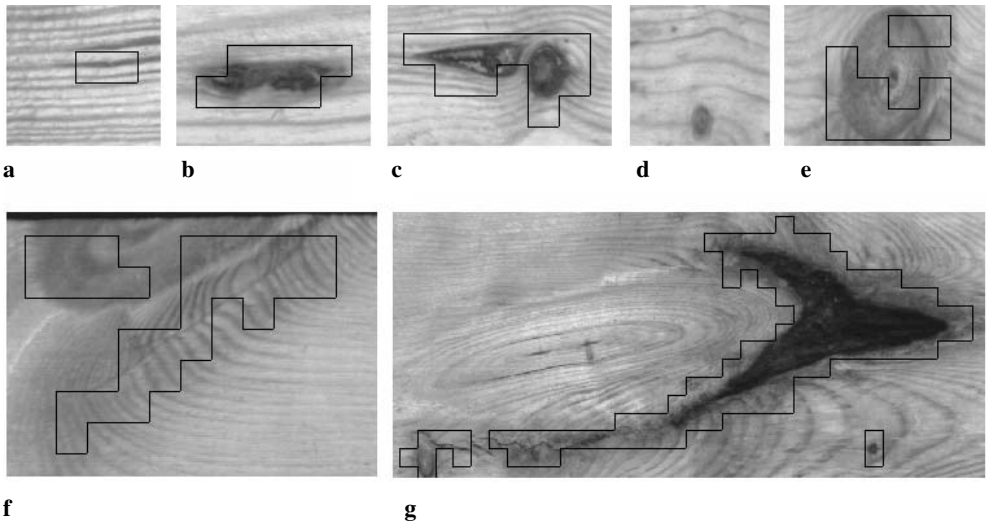


Fig. 6a–g. Real detections. Parts **a**, **b** and **d** are clear. Also the parts **c** and **e** are unambiguous while **f** and **g** are more difficult to interpret

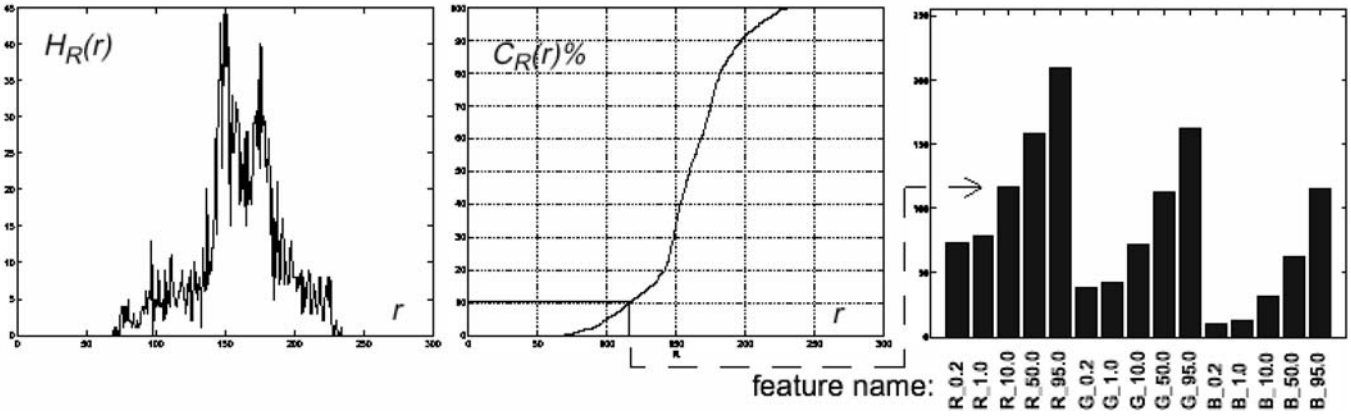


Fig. 7. Finding the 10% centile from a red-channel histogram. Kauppinen (1999)

4.1 Features

Most of the features used in our experiments are centiles based on color histogram features introduced for wood inspection by Silvén and Kauppinen (1994). The centiles are intensity values selected according to cumulative channel histograms. For example, the 10th centile of a channel denotes an intensity value larger than that of 10% of all the pixels in the region in question. The calculation of a 10% red centile is illustrated in Fig. 7. A histogram of the red channel (left) is first collected for the image. Then, centile values (right) are efficiently calculated from the cumulative channel histograms (center) as illustrated.

In addition to plain centile values, differences of two centiles of the same or different color channels can also be used as features. The centile features work well with highly varying lumber images and are only marginally more expensive to calculate than the trivial adaptive thresholding methods used in most automatic lumber grading systems. In practice, both approaches require only one access to image data per pixel.

In addition to centiles, we have experimented with second-order texture features, such as local binary patterns (LBP) (Ojala et al. 1996) and statistical features calculated from co-occurrence matrices (Haralick and Shapiro 1992). The computation of a local binary pattern (LBP) is shown in Fig. 8.

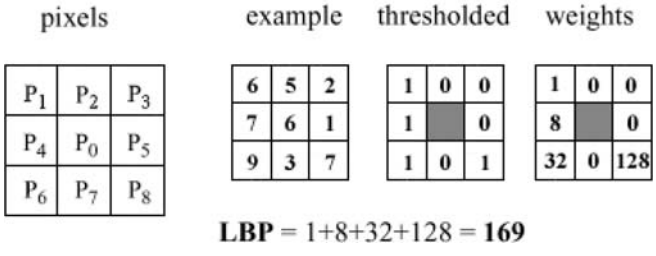


Fig. 8. Computation of LBP

The original 3×3 neighborhood is thresholded by the value of the center pixel. The values of the pixels in the thresholded neighborhood are multiplied by the weights given to the corresponding pixels. Finally, the values of the eight pixels are summed to obtain a number for this neighborhood. The LBP histogram computed over a region is used for texture description. LBP provides us with knowledge about the spatial structure of the local image texture, and it is invariant against any monotonic grayscale transformation; however, it does not address the contrast of texture, which is important in the discrimination of some textures.

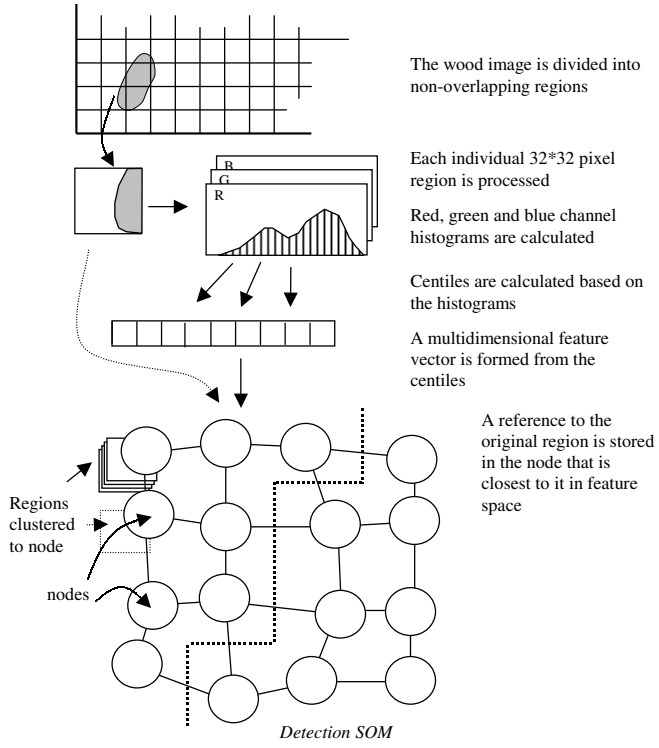


Fig. 9. The defect-detection method. In the SOM at the bottom of the figure, the distance between adjacent nodes represents their relative distance in a high-dimensional feature space. The *dashed line* is a manually drawn border between sound and defective wood

Table 1. Feature sets used at detection test

	Color channels	Number of centiles	Number of LBP values
Feature set 1	red	32	32
Feature set 2	red, green, blue	30	–

4.2 Defect detection

The defect-detection method used is shown in the diagram of Fig. 9. The wood image captured by the imaging system is divided into non-overlapping rectangular regions of a fixed size (e.g. 32×32 pixels). A set of feature values – in our case, centiles of color histograms – is calculated for each region, resulting in a feature vector. Each feature vector is then mapped into a self-organizing map that we call *detection SOM*.

In the SOM, for each feature vector, the node that is closest in the feature space is chosen, and reference to corresponding region is added to the respective node to enable accessing of the original data via the SOM. The result is a 2D map in which regions with similar feature vectors are located close to each other.

The dashed line in the SOM in Fig. 9 is the boundary between sound wood and suspected defects. This boundary is set during the basic training stage of the algorithm and is the only human interaction required. In practice, this defect-detection approach can be interpreted as thresholding in a multidimensional feature space.

An example of a SOM used for defect detection from pine boards is shown in Fig. 10. At each node an image of one randomly selected region is shown from among those clustered to it. In our experimental system, clicking any node with a mouse results in showing all the regions clustered to the node, enabling the user to make observations on the uniformity of the members of the cluster. The gray nodes have no sample items. The similarities between adjacent nodes are apparent. Dry knots and sound wood are at opposite corners of the SOM, while the shakes and other defects are in-between. Two alternatives for the boundary between sound wood and suspected defect regions are shown in the figure.

In our experiments, setting a fixed border between sound and defective wood has given good results, despite material variations, and only minor improvement has been achieved using adaptive, or board-specific, boundaries. The boundary can be determined quite easily by just taking a glance at the SOM. Figure 11 shows what happens with two different selections – optimistic and pessimistic – for the boundary. With the optimistic selection, the defective regions detected are quite small; while the pessimistic selection includes even the transitions from sound wood to defects in the detections. The number of error escapes is small in both cases.

In practice, we favor pessimistic boundaries as they result in regions that include information of the background of the defect. This improves the recognition results as, for example, knots with the same color but with different kinds of backgrounds are clustered into their own nodes in the detection SOM.

4.2.1 Test setup

The defect-detection approach was tested with the two different sets of features presented in Table 1. Due to the very high data rate in real-time inspection, defect detection requires computationally low-cost features. Obtaining the histograms of color channels requires only one access to the image per pixel. The centiles are then fast to calculate and have been found to be powerful features.

The cost of the implementation reduces if detection can be made using only one color channel. The discrimination appears to be best in the red channel ($> 600\,\mathrm{nm}$). However, the red-channel histogram does not alone provide enough information of the wood surface. In addition, some second-order texture features are required. LBPs are efficient and computationally cheap features and thus used here in feature set 1.

Both feature sets were used in the test with region sizes of 16×16, 24×24, 32×32, and 40×40 pixels to find the size providing good error-escape and false-alarm rates. In addition to the detection rates, the region sizes have also some other effects. Smaller regions need more computation time, but give a better evaluation of the size and position of the defects, and better discrimination between adjacent defects.

The size of the detection SOM was 22×18 nodes, which fits reasonably onto a 1280×1024 pixel display. While bigger maps may give slightly better discrimination, the smaller ones are computationally cheaper. The influence of the size of the SOM has been analyzed by Niskanen et al. (2001).

Due to the high variation in wood material, it is probably impossible to find features that separate all the defective areas from sound wood. For that reason, some defective and sound wood regions are confused with one another in the feature space and thus also on the SOM. One part of the defect-

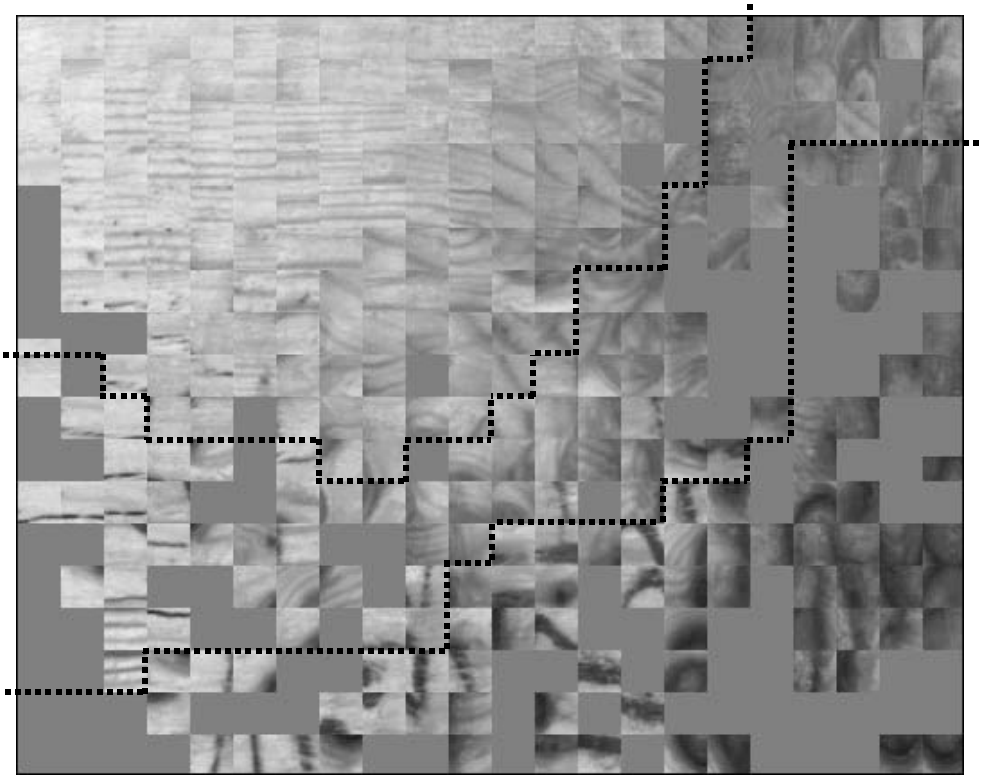
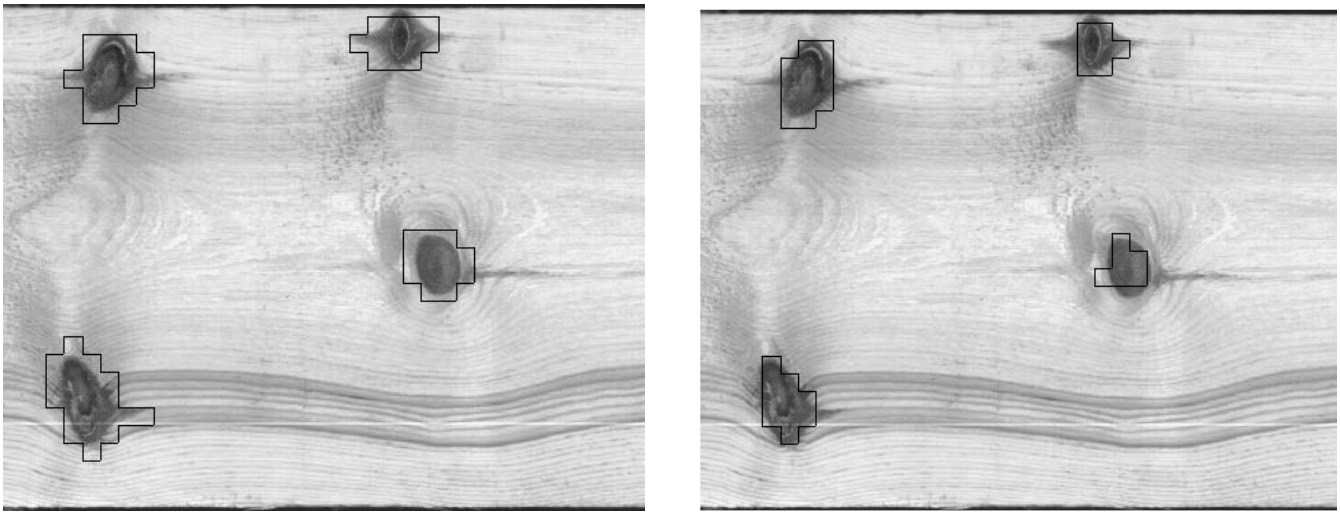


Fig. 10. Detection SOM: The *dashed lines* are manually drawn, pessimistic (*upper*) and optimistic (*lower*) approximations of the boundary between sound wood and defects



a **b**

Fig. 11. a Pessimistic and **b** optimistic selection of the boundary. The boundaries used are sketched in Fig. 10

detection algorithm is to select the boundary between sound and defective areas in the detection SOM. Due to confusion, especially while using a fixed border with highly varying data, there is no "correct" boundary selection, but the selection turns out to be a decision on how much of this border zone in the SOM is interpreted as suspected defects. The boundary selection is thus a compromise between the error-escape and the false-alarm rates. We favor low error-escape rates, since most of the false alarms can be eliminated later.

4.2.2 Detection results

Average false-alarm rates, compared to detections made by humans, with two different fixed boundary selections are presented in Table 2. Both optimistic and pessimistic selections give a low error-escape rate, 5% and 2.5% respectively.

Increasing the false-alarm rate also increases the sizes of detections. In practice, a 40% false-alarm rate results in about a 50% increase in the amount of the board area that requires further analysis; in other words, if the real defects occupy 2% of the board, some 97% of the data is still eliminated at the detection stage.

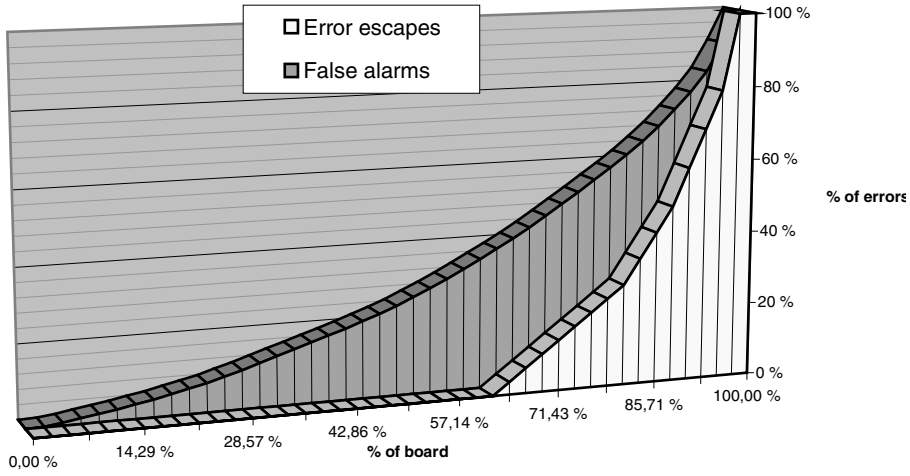


Fig. 12. Cumulative distribution of error escapes and false alarms on boards in the test set. Most of the boards are inspected without any error escapes. The error-escape rate is 2.5% for the whole test material

Fig. 13. False alarms clustered to self-organizing map to provide an overview of false-alarm types

Table 2. False-alarm rates at fixed error-escape rates

		Border	
Feature set	Region size	Optimistic false-alarm rate (error-escape rate ≈5%)	Pessimistic false-alarm rate (error-escape rate ≈2.5%)
1	40	22.9%	33.9%
	32	28.7%	39.8%
	24	37.3%	64.0%
	16	44.1%	67.4%
2	40	39.8%	44.9%
	32	38.7%	50.5%
	24	55.6%	67.0%
	16	69.7%	74.9%

Table 3. Error escapes with two different boundaries

	Optimistic	Pessimistic
Shake (90)	38.9%	13.3%
Pitch pocket (40)	17.5%	12.5%
Sound knot (267)	3.7%	1.5%
Pitch wood (6)	33.3%	33.3%
Dry knot (520)	0.8%	0.6%
Core stripe (8)	0.0%	0.0%
Pin knot (6)	0.0%	0.0%
Bark pocket (4)	0.0%	0.0%
Barkringed knot (5)	0.0%	0.0%
Black knot (28)	0.0%	0.0%

4.2.3 Cumulative distribution of detection errors

The cumulative distribution of error escapes and false alarms on different boards at about a 2.5% error-escape rate is presented in Fig. 12 when detection is made with a pessimistic boundary. The false alarms are distributed quite evenly, while the error escapes clearly came from some difficult boards. The easiest 60% of the boards are inspected without missing any defects, whereas the toughest 10% of boards contain about half of the error escapes. Most of these tough boards have been sawn from the heart side of the log, where small, difficult-to-detect defects are present.

4.2.4 Types of the error escapes

The classes of the error escapes when feature set 1 and a region size of 32×32 pixels are used are presented in Table 3. Using a pessimistic boundary results in few error escapes. Most of them are narrow shakes that are difficult to detect with histogram features. Notice, that there are some seldom-occurring defects.

4.2.5 Types of the false alarms

The reasons for false alarms are harder to assess, since the human-given ground truth grades are only for the defects. An overview of false alarms is shown in Fig. 13, where they are clustered into a self-organizing map. Using the SOM in this phase has nothing to do with the detection algorithm, but by using it we try to clarify the types of the false alarms.

The SOM of the false alarms also provides a means to find better features for discriminating defects from the sound wood. About half of the false alarms appear to be small dark spots, which are not defects, but dirt and strong grain. Among them there are some real defects, approximately 20% of the false alarms, which for some reason had not been labeled by human experts.

4.3 Defect recognition

The defect-recognition method is schematically depicted in Fig. 14. The defect-detection stage results in a set of suspicious regions. These are now subjected to connected-components analysis to obtain regions for the recognition phase. For each connected set of regions, a new set of features, such as centiles, is calculated.

At this stage, only a small fraction of the original data remains enabling us to use more sophisticated and computationally complicated methods for analysis. For example, the category of a knot depends on its aspect ratio, location, and size. However, for presentational clarity, we did not use any of this kind of information in the experiments described here.

Again, the feature vectors are fed into a SOM that clusters similar defects close to each other. Figure 15 shows the result when the detections from several boards have been used to produce the *recognition SOM*. For each node an image of one suspected defect is shown, which is randomly chosen from the region in the respective cluster. The size of the SOM in Fig. 15 is small and does not then provide very good resolution between different types of defects. However, the images in the figure clarify the idea of the method.

It is essential to realize that so far there has been no need to label any defects or collect training material. Nor have we used any special rule-based knowledge about defects.

Evaluating the recognition accuracy of this approach is not straightforward. When we take a look at the clusters at any individual node in the SOM, we see that the defects are strikingly similar and clearly belong to same category, as Fig. 16 demonstrates. However, when we compare them with the corresponding independent manual labeling, prepared during the material acquisition stage, there is 10–20% disagreement. This is in line with the results compiled by Lampinen et al. (1994).

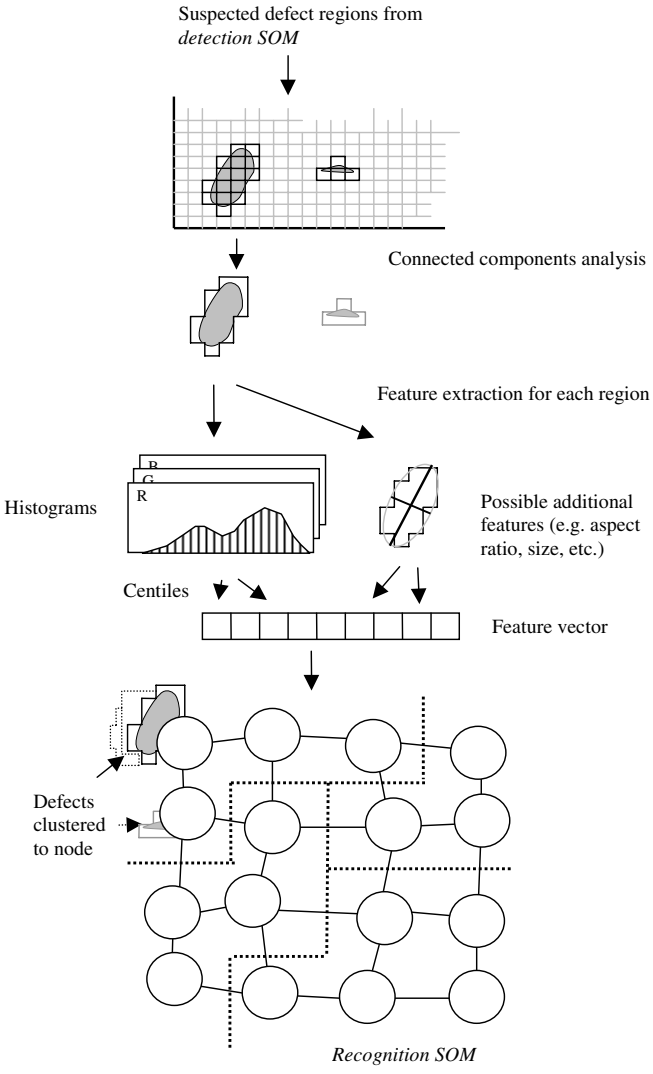


Fig. 14. Recognition method. Defected areas form regions that are clustered to SOM as individual defects

In other words, if manually labeled samples are used in training an inspection system, it is difficult to achieve better than 80% accuracy (Kauppinen 1999). We believe that minimal human involvement in training is a key to high accuracy.

4.3.1 Test setup

In order to evaluate the classification performance of the recognition SOM, we measure the internal variation in the nodes. Each node is named after the most common class represented by sample items clustered into it and the count of detections in each class is reported using a confusion matrix. Thus, all the results are related to human labeling. The size of the recognition SOM is 30×20 nodes, and classification is made with 14 color centile features without using any additional information. The size of the recognition SOM was selected to give a good resolution while not forgetting the demands of the visualization.

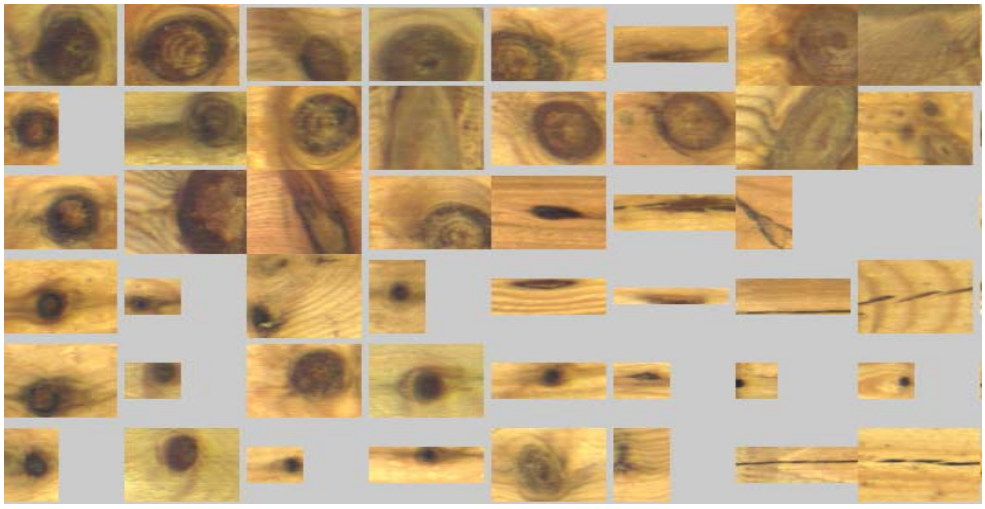
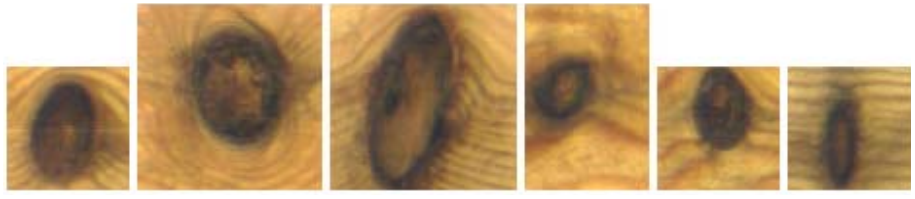


Fig. 15. Classification SOM. Defects with similar appearance are close to each other



a

b

c

Fig. 16a–c. Defects mapped to three different SOM nodes: **a** dry knots; **b** sound knots; and **c** shakes

Results for classifying human-detected defects

Two different recognition tests were made. In the first one all human-labeled defects were classified in order to evaluate the recognition power independently from detection capability. The features were calculated from rectangular areas that human experts had marked to indicate the positions of the defects. In other words, human-defect detection was used rather than the algorithm presented in Figs. 9 and 14. The results of the first test are presented in Table 4.

Results for recognizing automatically detected defects

In the second test, the detected suspected areas obtained from the detection test in Sect. 4.2.2 were classified. This test agrees better with a real problem, where detections do not always match very well with the real defects. Detection was made with feature set 1 and 32×32 pixels regions with a pessimistic boundary. Error escapes and thus non-classified defects were 2.5% of total defects.

The defects were classified based only on the regions that were detected by the detection phase, and each suspected region was named to match as well as possible the human-labeled regions. The confusion matrix is presented in Table 5. The false-alarm rate of 39.8% decreases to less than five percent; however, the 2.5% error-escape rate grows to 12.8 % since some detected real defects are now classified as sound wood. Note that here the false-alarm rate affects the total error percentage as correctly classified sound wood decreases it.

Since the classification rate is evaluated by comparing the obtained classes with the human labeling, all the confusions are not obviously errors. For example, classifying the defect as a dry knot instead of sound may even correct the human

Table 4. Confusion matrix of classifying human-labeled errors

	Shake	Dry knot	Sound knot	Black knot	Pitch pocket	Rest	Error (%)
Shake	81	3	6				10.00%
Dry knot		509	13	5	1		3.60%
Sound knot	2	60	203	2			23.97%
Black knot		15		13			53.57%
Pitch pocket	2	12	5		21		47.50%
Rest		5	1			9	40.00%
Total error							13.64%

Table 5. Confusion matrix of classifying detected areas

	Sound wood	Shake	Dry knot	Sound knot	Black knot	Pitch pocket	Rest	Error (%)
Sound wood	617	32	31	19	4			12.2%
Shake	52	68	2	1				44.7%
Dry knot	23	1	449	10	16			10.0%
Sound knot	14		30	195				18.4%
Black knot			10		18			35.7%
Pitch pocket	7	1	8	5		13		61.8%
Rest	4		9	3	2	1	7	73.1%
Total error	13.9%	33.3%	16.7%	16.3%	55.0%	7.1%	0.0%	17.3%

error, while confusing a shake with a sound knot is a clear error.

At this stage, introducing rules and defect-specific features would improve the accuracy. In particular, a "shake detector" feature would drop the error rate for this particular defect. However, discriminating normal grain from shakes is a demanding texture-analysis task.

5 Training the inspection system

Although the methodology we are proposing for wood inspection is based on non-supervised clustering, it cannot be applied in a blind manner; a sound strategy is needed in preparing the detection and recognition of SOMs. In the next sections we describe the approaches that have been found powerful in building SOMs for this application.

A fundamental problem in visual inspection is the selection of efficient features for defect detection and recognition purposes. In practice, these features need to be searched using test material that has been prepared in a supervised manner. In our case, this very time-consuming process was performed on a small amount of material using a kNN-classifier.

5.1 Building the defect-detection SOM

The SOM is a projection of the probability density function (Kohonen 1995). This causes problems for visualizing rare classes. Namely, only a fraction of the surface area of a wooden board contains defects. If we simply divide the image of the board into, say, 32×32 pixel regions, calculate the features and build the SOM, the result has the defects concentrated into a few nodes at one of the corners while the remaining nodes contain sound wood regions.

This prevents the straightforward application of SOM methodology, because determining the boundary between sound wood and defects is difficult. In practice, it would be beneficial to have wide transition regions in the SOM between different categories of wood as the discrimination result would then be more robust with respect to location of boundary.

Because of this need, the defect-detection SOM is built using selected test material that contains about the same amount of sound wood and defects. This training material originates from several boards to capture the whole range of variation of lumber. However, the selection process does not require any labeling of the samples or any particular expertise in choosing the material.

5.2 Building the recognition SOM

In order to obtain an ideal recognition SOM, we should have around the same number of every defect category in the sample set, because the under-represented categories may not get a node of their own, while the larger classes conquer most of the map. To cope with this phenomenon, we must give some of the samples more weight by using them multiple times when building the SOM. In practice, teaching the SOM is an iterative process in which adding weight to some sample items causes the whole map to change and brings up some other nodes whose weights should then be adjusted. Clearly, an efficient tool is needed for this purpose.

5.3 Training tools

We use a software package, called G-SOM (1999), to create self-organizing maps. It provides visual support for finding which sample items are clustered in each node, so one can easily modify the training data and see the discrimination power of the approach. This tool enables one to label each node and the respective samples manually in a very efficient manner, so it can be used for preparing training material for supervised classification solutions, if desired.

6 Discussion

The weak points in the automatic visual inspection of lumber are human involvement in system training and overly simplistic defect-detection methods. The subjective human judgment appears to be in conflict with precise, formal grading rules and defect criteria. The decision of a human on the category of any individual defect is inherently error-prone and should thus be avoided; however, a human can easily label a group of defects into a single category. This is exactly the kind of approach that the presented solution supports.

The main purpose of this paper was to present a non-supervised inspection approach, not detection and recognition results. However, the results presented here are good or considerably better than results obtained earlier for the same test material with other approaches.

Several problems related to the approach were discussed. Selecting the most suitable features to be used in classification is in practice a supervised process. Also making the "balanced" SOMs, where different categories of wood are separated and get nodes of their own, requires work either in selecting samples or as an iterative training process.

SOM provides an intuitive user interface by giving a concrete way of seeing a classification problem, even for a user not familiar with SOMs. In practice, the concept of SOM can be hidden in the user interface. Among method developers and system integrators it has been received well as a promising tool for method development and testing. We believe that the proposed approach facilitates the training of the inspection system and the use of inspection systems in these kinds of problematic environments.

Acknowledgements. The support of TEKES for this line of research since 1992 is gratefully acknowledged. We also thank Arto Usenius and Olof Sommardahl at VTT Building Technology, and Hannu Rautio and Toni Piirainen at the University of Oulu for their invaluable contributions.

References

1. Grönlund U (1995) Quality improvements in forest products industry. Dissertation, Luleå University of Technology, Sweden
2. G-SOM tool (1999), University of Oulu, Department of Electrical Engineering. http://www.ee.oulu.fi/research/imag/gsom/
3. Haralick R, Shapiro L (1992) Computer and robot vision, vol 1. Addison-Wesley, Reading, Mass.
4. Huber HA, McMillin CW, McKinney JP (1985) Lumber defect detection abilities of furniture rough mill employees. For Prod J 35:79–82
5. Kauppinen H (1999) Development of a color machine vision method for wood surface inspection. Dissertation, Acta Universitatis Ouluensis C141, University of Oulu, Department of Electrical Engineering, Oulu, Finland
6. Kohonen T. (1997) Self-organizing maps. Springer, Berlin Heidelberg New York
7. Lampinen J, Smolander S, Silvén O, Kauppinen H (1994) Wood defect recognition: a comparative study. In: Workshop on machine vision for advanced production, Oulu, Finland, 2–3 June, 1994
8. Niskanen M, Silvén O, Kauppinen H (2001) Experiments with SOM based inspection of wood. In: Proceedings of the international conference on quality control by artificial vision (QCAV 2001), vol 2, Le Creusot, France, 21–23 May, 2001
9. Nordic Timber, grading rules (1994) Pohjoismainen sahatavara, lajitteluohjeet (in Finnish). Suomen Sahateollisuusmiesten Yhdistys
10. Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on feature distributions. Pattern Recogn 29:51–59
11. Silvén O, Kauppinen H (1994) Color vision based methodology for grading lumber. In: Proceedings of the 12th international conference on pattern recognition, vol 1, Jerusalem, 9–13 October, 1994
12. Sommardahl O, Usenius A (1999) Wood image data, VTT Technology. We are trying to release the data via our WWW pages by the time the paper is published

Olli Silvén is the professor of signal processing engineering at the University of Oulu, Finland. He received his Dr. Tech. degree in Electrical Engineering from the same University 1988. He has worked with several visual inpection applications, ranging from defect detection on printed wiring boards to sorting of food materials.

Matti Niskanen received his MSc degree in information engineering from University of Oulu, Finland, in 2000. He is currently in PhD program in the Department of Electrical Engineering and a researcher in the Machine Vision Group at the same university. His main areas of interest are image analysis and pattern recognition.

Hannu Kauppinen received his Dr. Tech. degree in electrical engineering from the University of Oulu, Finland, in 2000. He is currently a senior researcher in the Machine Vision Group, University of Oulu. He has worked with visual inspection applications, especially with wood surface inspection.