Wood Defect Detection using Grayscale Images and an Optimized Feature Set

P. Cavalin[2], L. S. Oliveira[1,2], A. L. Koerich[1,2], A. S. Britto Jr[1,2].

[1]Pontifical Catholic University of Parana
Graduate Program in Computer Science
Rua Imaculada Conceição, 1155, Curitiba, PR, 80215-901, BRAZIL

[2]INVISYS Intelligent Vision Systems
Rua Prof. Algacyr Munhoz Mader, 3775, Curitiba, PR, 80350-010, BRAZIL
{cavalin,soares,alekoe,alceu}@ppgia.pucpr.br

***Abstract*— In this paper we address the issue of detecting defects in wood using features extracted from grayscale images. The feature set proposed here is based on the concept of texture and it is computed from the co-occurrence matrices. The features provide measures of properties such as smoothness, coarseness, and regularity. Comparative experiments using a color image based feature set extracted from percentile histograms are carried to demonstrate the efficiency of the proposed feature set. Two different learning paradigms, neural networks and support vector machines, and a feature selection algorithm based on multi-objective genetic algorithms were considered in our experiments. The experimental results show that after feature selection, the grayscale image based feature set achieves very competitive performance for the problem of wood defect detection relative to the color image based features.**

Fig. 1. Example of a cutting optimization machine. The human operator feeds the machine.

I. Introduction

Natural resources such as wood have become scarce and very expensive. Maximize the usage and reduce the rejection (losses) is a great challenge for the wood industry. The process to maximize the value of wood can be divided into three parts. Initially, the wood is taken to a sawmill and then one needs to decide whether the wood is more valuable as lumber, veneer, or chips. If it is for lumber, them the boards cut from it must be edged and trimmed. This is a process that requires someone to decide how to trim off effective parts and make the board as valuable as possible. Thereafter, someone must examine the board and give it a grade, based on the quality of the wood and presence of defects. Finally, someone cuts the lumber again to produce defect free dimension parts.

In order to facilitate the job of producing defect free dimension parts, some companies have developed machines to perform length cutting optimization (Figure 1). However, a great deal of human interaction is still necessary in most of the cases, since an operator must identify and mark (assign) the defects in the lumber using a fluorescent crayon. It is well known that operators seldom follow strict rules, but it is mostly based on observing the general visual appearance of the lumber. Here, it is worth of remark, that some operators are less efficient than others to detect very small defects such as cracks or spots.

When observing the general visual appearance of the lumber, one takes into consideration color and shape [7]. In light of this, some authors have shown through experimentation that color conveys very discriminant information and should be used in automated wood inspection systems, specially for grading [9], [6], [2].

In this paper we address mainly the problem of defects detection to build a fully automatic lumber optimization machine. Our goal is to build a low-cost robust algorithm to detect defect in wood. Since it has to be cost-effective, we have chosen monochromatic sensors which are considerably less expensive than color ones, particulary when considering line-scan sensors. This lead us to the challenge of building a robust feature set based on grayscale images. To assess the performance of the proposed feature set it is compared with the one proposed by Kauppinen [5], which is based on cumulative histograms extracted from the R, G and B channels of color images. Two different learning paradigms were used in our experiments: Support Vector Machines and Neural Networks. Feature selection is carried out using multi-objective genetic algorithms to avoid any possible influence of irrelevant features in this comparison. Experimental results show that after feature selection, both grayscale and color-based features achieve similar performance for the problem of wood defect detection.

This paper is organized as follows: Section II describes two different set of features used in our experiments. Section III presents the methodology applied to perform feature selection and Section IV describes the database considered in this work. Section V reports the experimental results and Section VI concludes this works.

II. Features

This section describes two different set of features used in our experiments. The color-based feature set takes into account percentile histograms extracted from the R, G, and B color channels [5] while the proposed grayscale-based feature set are extracted from the co-occurrence matrices [1].

A. Color-based Features

The features extracted from R, G and B color channels cumulative histograms [5] are simple, but very discriminative. Some percentile values are selected, and the color value at that percentile is used as primitive.

Let $C_k(i) = \sum_{j=1}^{i} H_k(j)$ be the cumulative histogram value of color i at channel k, where $H_k(j)$ is the histogram value for color j at channel k, and N_k the depth of color channel k, the feature value for percentile y is:

$$F_k(y) = i, \qquad \text{where } C_k(i) \simeq y,\ 1 \leq i \leq N_k \tag{1}$$

Due to the intensity changes of pure percentile features, a color calibration scheme is used. Invariant features against the shift and the width of the histogram can be obtained by computing differences of two percentiles and normalizing them by the difference of the maximum and minimum percentile values of the histogram:

$$f = \frac{F_k(y_1) - F_k(y_2)}{F_k(100\%) - F_k(0\%)} \tag{2}$$

One way to select the percentile features is selecting equal portions in a range between the maximum and minimum percentile values. Since wood images can have some noise, it is reasonable to select the maximum and minimum as 95% and 5%, and dividing equally the 90% within the range. In our experiments, the feature vector is selected dividing that range by 10, and having ten percentile features for each color channel. So, the number of features is 30 (10 × 3).

B. Grayscale-based Features

An important approach to describe a region is through the quantification of its texture content. It intuitively provides measures of properties such as smoothness, coarseness, and regularity. The grayscale-based features proposed in this paper to detect defects in wood are based on texture and are computed from co-occurrence matrices. Those properties are statistical measures extracted from a matrix that represents the relationship between pixels within the region.

A co-occurrence matrix (CM) [1] is the joint probability occurrence of graylevel i and j within a defined spatial relation in an image. That spatial relation is defined in terms of a distance d and an angle θ. Given a CM, some statistical information can be extract from it. Assuming that Ng is the graylevel depth, and $p(i,j)$ is the probability of the co-occurrence of graylevel i and graylevel j observing consecutive pixels at distance d and angle θ, to describe wood texture, the following measures were chosen as features, where:

$$f_1 = \text{Contrast} = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng}(i-j)^2\ p(i,j) \tag{3}$$

$$f_2 = \text{Energy} = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng}(p(i,j))^2 \tag{4}$$

$$f_3 = \text{Entropy} = -\sum_{i=1}^{Ng}\sum_{j=1}^{Ng} p(i,j)\log(p(i,j)) \tag{5}$$

$$f_4 = \text{Correlation} = \frac{p(i,j) - \mu_x\mu_y}{\sigma_x^2\sigma_y^2} \tag{6}$$

$$\text{where } \mu_x = \sum_{i=1}^{Ng} i\ p_x(i), \quad p_x(i) = \sum_{j=1}^{Ng} p(i,j)$$

$$\sigma_x^2 = \sum_{i=1}^{Ng}(i-\mu_x)^2 p_x(i)$$

$$\mu_y = \sum_{j=1}^{Ng} j\ p_y(j), \quad p_y(j) = \sum_{i=1}^{Ng} p(i,j)$$

$$\sigma_y^2 = \sum_{j=1}^{Ng}(j-\mu_y)^2 p_y(j)$$

The measures above can be extracted at different distances d and angles θ, where a corresponding CM is computed. For every CM, four features are added to the feature vector $\bar{f}$. An additional measure at the graylevel space is the average graylevel of the image. Therefore, the size of $\bar{f}$ is $4 \times N_{d\theta} + 1$, where $N_{d\theta}$ is the number of combinations between d and θ. In our experiments we have tried different values for d as well as different angles. The best setup we have found is $d = 1$ and $\theta = [0, 90, 180, 270]$. This yields an 17-dimensional feature vector.

III. Feature Selection

An important issue in constructing classifiers is the selection of the best discriminative features. In many applications, it is not unusual to find problems involving hundreds of features. However, it has been observed that beyond a certain point, the inclusion of additional features leads to a worse rather than better performance [11]. Moreover, the choice of features to represent the patterns affects several aspects of the pattern recognition problem such as accuracy, required learning time, and the necessary number of samples.

This apparent paradox presents us with a feature selection problem in automatic design of pattern classifiers. Such a problem refers to the task of identifying and selecting an effective subset of features to represent patterns from a larger set of often mutually redundant or even irrelevant features.

Feature selection is not a trivial problem since features are seldom entirely independent. There may be redundancy, where certain features are correlated so that it is not necessary to include all of them in modeling, and interdependence, where two or more features between them convey important information that is obscure if any of them is included on its own.

In the context of practical applications, feature selection presents a multi-criterion optimization function, e.g., number of features and accuracy of classification. It has been demonstrated that multi-objective genetic algorithms offer a particularly attractive approach to solve this kind of problems since they are generally quite effective in rapid global search of large, non-linear and poorly understood spaces, and can cope with several objective in a very clever way [8]. In light of this, we have used the strategy proposed by Oliveira et al. [8] to perform feature selection. It is based on a powerful multi-objective genetic algorithm (MOGA) called Non-Dominated Sorting Algorithm (NSGA) [4].

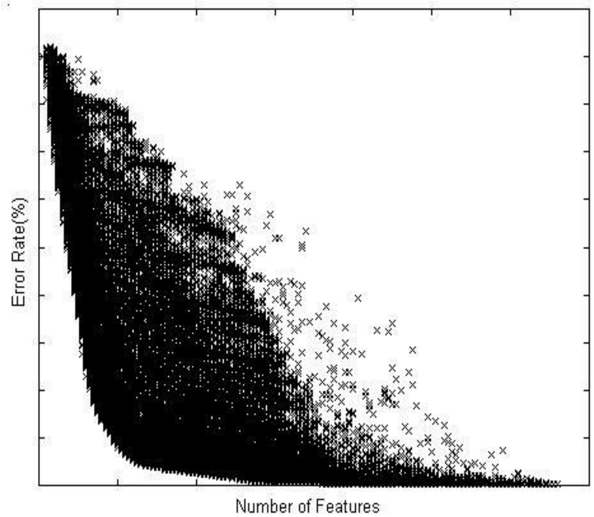
Differently of a single genetic algorithm, NSGA produces a set of potential solutions known as Pareto-optimal solution. This allows the user to try different trade-offs between the objectives being optimized. A very interesting way to pick a solution is to rely on an independent validation set to avoid an overfitted solution. For more details, please refer to [8].

The idea behind the NSGA is that a ranking selection method is used to emphasize good points and a niche method is used to maintain stable subpopulations of good points. It differs from simple genetic algorithm only in the way the selection operator works. The crossover and mutation remain as usual. Before the selection is performed, the population is ranked based on an individual's nondomination. The nondominated individuals present in the population are first identified from the current population. Then, all these individuals are assumed to constitute the first nondominated front in the population and assigned a large dummy fitness value. The same fitness value is assigned to give an equal reproductive potential to all these nondominated individuals.

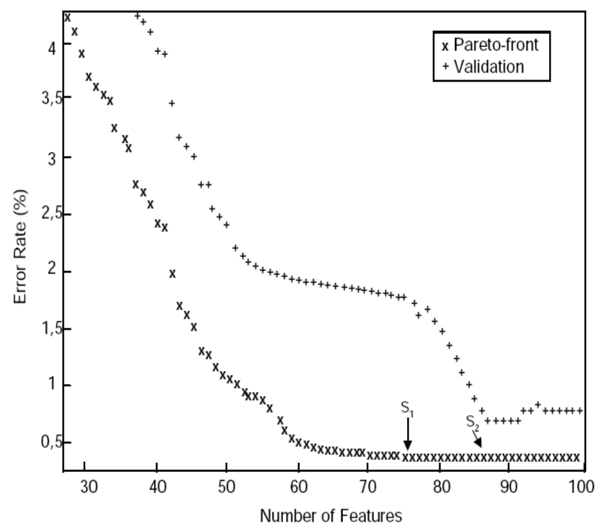
In order to maintain the diversity in the population, these classified individuals are then shared with their dummy fitness values. Sharing is achieved by performing selection operation using degraded fitness values obtained by dividing the original fitness value of an individual by a quantity proportional to the number of individuals around it. Thereafter, the population is reproduced according to the dummy fitness values. Since individuals in the first front have the maximum fitness value, they get more copies than the rest of the population. The efficiency of NSGA lies in the way multiple objectives are reduced to a dummy fitness function using nondominated sorting procedures.

In the end, the algorithm produces a set of potential solutions that can be chosen by a decision maker. In order to support this choice, a good strategy lies in using an independent validation set to avoid an overfitted solution. Figure III depicts a classical Pareto-optimal front for the feature selection problem.

If we analyze only the Pareto-front, the best trade-off between the number of features and the error rate is the solution S_1.However, by analyzing the validation curve, we can observe that such a solution supplies a poor generlization



(a) Example of the evolution in the objective plane



(b) Example of a Pareto-optimal from produced by the MOGA

on an unknown database. We can also observe that the accuracy/complexity trade-off that has the best generalization on the validation set is the solution S_2.

IV. Database

The approach is based on implicit segmentation, which means that the image is partitioned through a pre-defined grid regardless its content. In this way, partitioning is made to non-overlapping rectangular regions of 32×32 pixels [10].

The features are calculated for each region and classified to relevant classes. In the case of defect detection, we are dealing with two classes: good wood and defects. In order to build the training set, a graphical interface has been developed. It segments the image into 32×32 rectangular regions, and those containing some kind of defect are marked as class 1 (defect). The unmarked ones are hence labeled as class -1 (good wood).

The database considered in this work consists of 400 images extracted from Pinus lumber for training and other 100 for testing. The database contains several kind of defects all

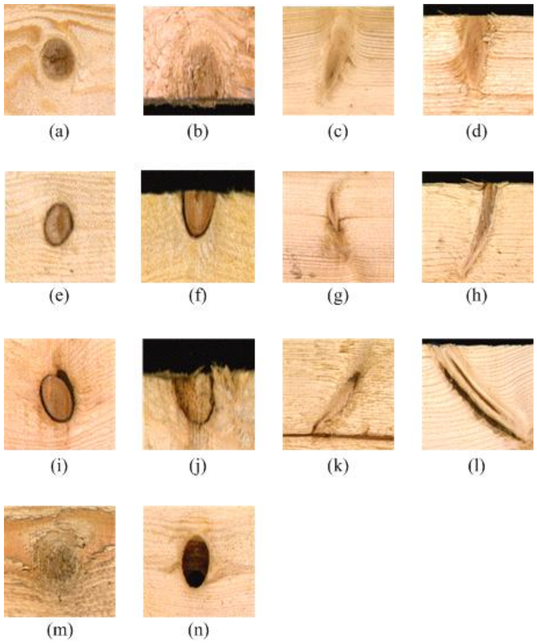


Fig. 2. Different types of knots found in the database: (a-d) Sound knots (normal, edge, leaf, and horn), (e-h) Dry knots (normal, edge, leaf, and horn), (i-l) Encased knots (normal, edge, leaf, and horn), (m) Decayed knots, and (n)knot hole

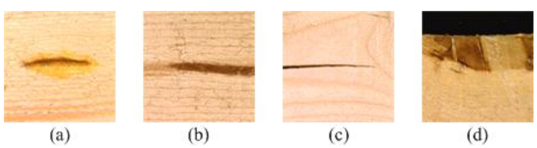


Fig. 3. Other defects: (a) Resin pocket, (b) core stripe, (c) split, and (d) wane

defects depicted in Figures 2 and Figure 3. Figure 4 shows the example of an image segmented into rectangular regions. The blocks which contain defect are highlighted.

The images were collect with a Samsung CCD camera with 8-bits accuracy per color channel. This allows further comparison between color and grayscale-based features. In order to create the gray scale image, we have used the formula for luminance, which is given in Equation 7

$$GRAY = [0.299, 0.587, 0.114] \times [R, G, B]^T \tag{7}$$

V. Experiments

To evaluate the performance of the grayscale-based feature set, we have considered two different machine learning paradigms: Neural Networks and Support Vector Machines. The neural network is an MLP trained with the gradient descent applied to a sum-of-squares error function. The transfer function employed is the standard sigmoid function. The generalization performance is monitored through a validation set. The parameters of the networks were set empirically.

The other machine learning algorithm considered here is the Support Vector Machine [12]. It has gained a lot of attention of machine learning and pattern recognition communities due

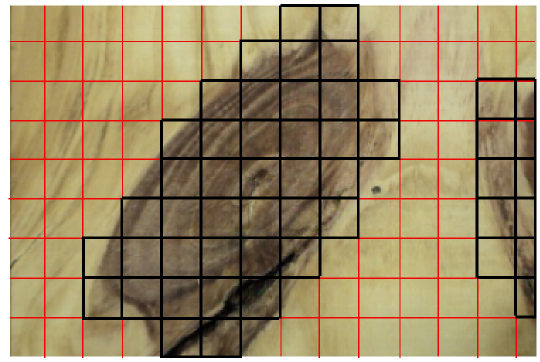


Fig. 4. Pine lumber image segmented into rectangular regions (defects are highlighted).

to its ability of generalizing well even in high dimensional spaces under small training set conditions. In our experiments we have used the LIBSVM package [3]. The parameters of the SVM were set through a grid search tool available in the LIBSVM.

The NSGA used for feature selection is based on bit representation, one-point crossover, bit-flip mutation, and roulette wheel selection (with elitism). The following parameter settings were employed: population size = 128, number of generations = 1000, probability of crossover = 0.8, probability of mutation = 0.007, and niche distance (σ_{share}) = 0.5.

In the first experiment we have considered the original feature vectors, i.e., without feature selection. The two first columns of Table I report these results. It can be observed from this table that the color-based features achieve better results than grayscale-based features for both learning models.

TABLE I

Performance of the classifiers on the test set before and after feature selection.

Feature	Before FS		After FS	
Set	MLP	SVM	MLP	SVM
Color	98.0	98.7	98.1	98.7
Grayscale	95.5	95.8	97.5	98.0

On the other hand, it can also be noticed from this table that after feature selection, both SVM and MLP trained with the grayscale-based feature set achieves similar performance to those trained with color-based features. In terms of number of features, in both cases it was reduced considerably. The color-based and grayscale-based feature sets had 21 and 8 features removed, respectively. This corroborates to the importance of feature selection in any kind of pattern recognition system and makes it clear that in this case both feature sets contain mutually redundant or even irrelevant features.

After analysing the results, we could notice that the system is able to detect small defects such as cracks and spots. Figure 5 shows an example of Pine lumber classification. We can see that the defect was detect but some regions of good wood were misclassified. This kind of error is also due to the contrast information inside of the region being classified. However, such a kind of confusion can be eliminated through a postprocessing stage based on binarization [5].

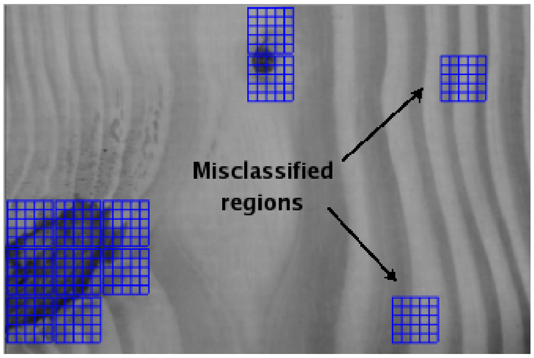


Fig. 5. Example of misclassified regions of Pine lumber.

VI. Conclusion

In this paper we have addressed the problem of defect detection in wood. Our goal is to build a robust and low-cost algorithm so that it can be applied to an automatic lumber optimization machine. As stated before, in order to reduce the overall costs we have chosen grayscale-based features over the color-based ones.

We have demonstrated through experimentation that it is possible to achieve similar performance of color-based systems using just grayscale information. It is worth to remark, though, the importance of the feature selection step in all this process. In both cases, mutually redundant or even irrelevant features are removed. In the case of the grayscale-based feature set, we have noticed an improvement of performance in addition to the reduction of the feature set. For future work we plan to assess the suitability of this feature set for grading.

References

[1] E. H. Aria, M. R. Saradjian, J. Amini, and C. Lucas. Generalized cooccurrence matrix to classify irs-1d images using neural network. In *Proceedings of XXth ISPRS Congress*, Instanbul, Turkey, July 2004.

[2] C. C. Brunner, A. G. Maristany, D. A. Butler, D. Vanleeuwen, and J. W. Funck. An evaluation of color spaces for detecting defects in Douglas-fir veener. *Industrial Metrology*, pages 169–184, 1992.

[3] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/c̃jlin/libsvm.

[4] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley and Sons Ltd, 2^{nd} edition, April 2002.

[5] H. Kauppinen. *Development of A Color Machine Vision Method for Wood Surface Inspection*. PhD thesis, Oulu University Library, 1999.

[6] H. Kauppinen. A two stage defect recognition method for parquet slab grading. In *Procs of the International Conference on Pattern Recognition*, pages 803–806, 2000.

[7] H. Kauppinen and O. Silven. The effect of illumination variations on color-based wood defect classification. In *Procs of the International Conference on Pattern Recognition*, pages 828–832, 1996.

[8] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen. A methodology for feature selection using multi-objective genetic algorithms for handwritten digit string recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(6):903–930, 2003.

[9] S. Radovan, P. George, M. Panagiotis, G. Manos, A. Robert, and D. Igor. An approach for automated inspection of wood boards. In *Procs of the International Conference on Image Processing*, pages 798–801, 2001.

[10] P. J. Sobey and E. C. Semple. Detection and sizing visual features in wood using tonal measures and a classification algorithm. *Pattern Recogntion*, 22:367–380, 1989.

[11] G. V. Trunk. A problem of dimensionality: A simple example. *IEEE Trans.on Pattern Analysis and Machine Intelligence*, 1(3):306–307, 1979.

[12] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, 1998.