CLASSIFYING FEATURES IN CT IMAGERY: ACCURACY FOR SOME SINGLE AND MULTIPLESPECIES CLASSIFIERS

Daniel L. Schmoldt  
USDA Forest Service  
Brooks Forest Products Center  
Blacksburg VA 24061-0503 USA

Jing He  
Communication Technology Group  
COMSAT LAB  
22300 COMSAT Drive  
Clarksburg MD 20871 USA

A. Lynn Abbott  
Bradley Department of Electrical Engineering  
Virginia Tech  
Blacksburg VA 24061-0111 USA

Abstract
Our current approach to automatically label features in CT images of hardwood logs classifies each pixel of an image individually. These feature classifiers use a back-propagation artificial neural network (ANN) and feature vectors that include a small, local neighborhood of pixels and the distance of the target pixel to the center of the log. Initially, this type of ANN was able to classify clear wood, bark, decay, knots, and voids in CT images of two species of oak (*Quercus rubra*, L., *Quercus nigra*, L.) with 95% pixel-wise accuracy. Recently we have investigated other ANN classifiers, comparing 2-D versus 3-D neighborhoods and species-dependent (single species) versus species-independent (multiple species) classifiers using oak, yellow poplar (*Liriodendron tulipifera*, L.), and black cherry (*Prunus serotina*, L.) CT images. When considered individually, the resulting species-dependent classifiers yield similar levels of accuracy (96-98%); however, all classifiers achieve greater than 91% accuracy. 3-D neighborhoods work better for multiple-species classifiers and 2-D is better for single-species. Multiple-species classifiers, whose training included both cherry and yellow poplar examples, exhibit the lowest accuracy. Nevertheless, when this combination of species is avoided, there is no statistical difference in accuracy between single- and multiple-species classifiers, suggesting that a multiple-species classifier can be applied broadly with high accuracy. Because all reported accuracy values are prior to postprocessing operations (which visually improve classification accuracy), we are confident that even the least accurate classifiers would be adequate for industrial implementation.
1 Introduction

Most of the high- and medium-grade hardwood lumber sawn in the US goes into the production of appearance-related goods, e.g., furniture, cabinets, moulding. Because the value of hardwood lumber is directly related to the percentage of clear-wood cuttings that it contains, each log must be sawn to minimize the defects in the resulting boards. Traditionally, the sawyer chooses a sawing strategy by visually examining the exterior of a log and dynamically adjusting the cutting face as sawing exposes the log interior. This type of sawing is “information limited” in the sense that the sawyer only has knowledge of external indicators of internal features (i.e. defects). Developing nondestructive sensing and analysis methods that can accurately detect and characterize interior defects is critical to future efficiency improvements for sawmills (Occena 1991).

A tacit assumption for eventual application of internal scanning to log sawing is that knowledge of internal defects can lead to choosing the best sawing position and sawing method. Making the correct choice will allow mills to realize increased value gains. Log breakdown in this scenario is “fully informed”, where the sawyer has knowledge about internal feature size, type, and location. CT scanning has been investigated (e.g., Aune 1995, Benson-Cooper et al. 1982, Birkeland and Holoyen 1987, Burgess 1985, Cown and Clement 1983, Davis and Wells 1992, Grönlund 1992, Grundberg and Grönlund 1992, Harless et al. 1991, Hodges et al. 1990, Hopkins et al. 1982, Lindgren 1991, Onoe et al. 1984, Roder 1989, Schmoldt 1996, Taylor et al. 1984) as providing processing-critical internal defect information. Between the steps of (1) scanning and (2) applying scan information to log sawing, however, lies a third step in which defects and the log surface are automatically delineated.

Early work on automatically labeling internal log defects established the feasibility of utilizing CT images for this purpose (e.g., Funt and Bryant 1987, Taylor et al. 1984). These researchers, and others since, have employed a variety of methods to segment different regions of a CT image and then to interpret, or label, those segmented regions. While those initial efforts demonstrated feasibility, they had some limitations, e.g., anecdotal, or limited statistical, defect labeling accuracy reported, no effort to assess or to achieve real-time operability, and little use of texture information. Recent work by us (Li et al. 1996, Schmoldt et al. 1997) has demonstrated highly accurate labeling of log defects in CT imagery. In contrast to the previous global approaches that separate the tasks of segmentation and region labeling, this approach operates using local, pixel neighborhoods primarily, and effectively combines segmentation and labeling into a single classification step. A feed-forward artificial neural network (ANN) can be trained to accept CT values from a small 2-dimensional (2-D) or 3-dimensional (3-D) neighborhood about the target pixel, and then assigns a particular class label to each pixel. In order to accommodate different types of hardwoods, a histogram-based preprocessing step normalizes CT density values prior to ANN classification. Morphological postprocessing is used to refine the shapes of detected image regions. This approach mitigates the limitations of previous approaches, that is, accuracy can be evaluated quantitatively, defect labeling can be accomplished in real time, and texture information is utilized in the segmentation-classification step.
Accuracy achieved by this classification approach is very high (95%) at the pixel level (Schmoldt et al. 1997). This previous work, however, used two species of oak only, and processed 3-D neighborhoods almost exclusively. The current study extends that work to look at the interaction of neighborhood dimensionally (2-D vs. 3-D) and single- vs. multiple-species classifiers, with respect to their impact on classifier accuracy. The issue that we sought to resolve here is whether we could develop species-independent classifiers of high accuracy using our ANN, local-neighborhood approach.

2 Neural Net Classifiers

We have developed several species-dependent classifiers and several species-independent classifiers for different shaped neighborhoods in CT images. Both 2-D and 3-D neighborhoods have been considered. All of these classifiers contain the same modules, which are: (1) a preprocessing module, (2) an ANN-based classifier, and (3) a post-processing module. The preprocessing module separates wood from background and internal voids, and normalizes the CT density values. The ANN classifier labels each pixel of the image. The post-processing step removes some of the spurious misclassifications. The major difference between the various classifiers we have developed is that they are trained with different types of input features and have different sets of ANN weights.

2.1 The Preprocessing Module

2.1.1 Background Segmentation

Background segmentation, which separates the wood region (foreground) from background and internal voids, is the first step of the preprocessing module. This step eliminates portions of the image from further analysis which, in turn, simplifies the classification procedure and decreases classification time. Background thresholding can be accomplished either statically or dynamically. This research applies Otsu’s dynamic thresholding method (Otsu 1979). It has demonstrated effectiveness for segmenting CT images of hardwood logs previously (Li et al. 1996).

Otsu’s method assumes a bimodal distribution, in which a threshold \( t \) is selected for the histogram \( h(i) \) to minimizes the weighted within-mode variances (or, alternatively, maximizes the weighted between-mode variances) When using this thresholding method directly, however, decay in CT images was found to be categorized with background because the presence of decay in an image creates a trimodal histogram. To avoid this problem, a weighting function \( w(i) \) is applied to the original image histogram before applying Otsu’s method. This weighting function is given by

\[
w(i) = 1 - \exp\left(-\left(\frac{i - \frac{x_{ce}}{2}}{b}\right)^2\right),
\]

(1)
where $i$ is the CT number, $b$ is set to 2047 (the largest possible CT value in 11-bit data), and $x_{cw}$ is the value of the clear wood peak in the histogram. Then the best threshold is determined by applying Otsu’s method to the new histogram function, $h'(i) = h(i) \cdot w(i)$. After thresholding the original CT image, the background region is set to zero, and these pixels are ignored in subsequent processing steps. The original CT values are not modified in this step.

2.1.2 Normalization
Normalizing CT image values is the second step of the preprocessing module. The values in CT images are directly related to the density of the object. Because different species and different logs vary in density, somewhat different ranges of CT values can result. Histogram normalization translates the original CT image values into new values without disturbing the invariant associations that internal log features have with particular regions of the CT histogram. These associations seem to be, in our experience, consistent across many different species of logs in the green state (i.e., freshly cut).

The transformation we developed is:

$$x_{\text{norm}} = \frac{1}{x_0} \left[ x_0 \cdot \frac{x_{cw} - x_c}{1 + \exp(\alpha \left( \frac{x_{cw}}{2} - x_0 \right))} \right],$$

where $x_0$ is the original CT value, $x_{\text{norm}}$ is the normalized value, $x_{cw}$ is the original CT value of the clear wood peak, $x_s$ is an arbitrarily selected anchor value that is greater than the CT value of the clear wood peak. The quantity $\alpha$ is a constant that determines the steepness of the curve and has been set to $10/x_{cw}$. After histogram normalization, the new value of the clear wood peak in an image histogram is approximately 1.0. This translation also stretches the histogram so that mid-histogram features, such as decay, are not compressed into the clear wood portion of the histogram. Normalized CT values for each pixel are used directly by the ANN classifiers.

2.2 ANN Classifiers
The ANN classifier is the seminal part of this classification system. Back-propagation neural networks were chosen because of their documented effectiveness for pattern-matching problems, and their relative ease of use. Using an ANN, each non-background pixel is labeled. This section describes the feature vectors and classifier topologies used.

2.2.1 Feature Vectors
Selecting useful features for an ANN is extremely important because they determine how well the classifier learns and consequently how it will perform on unseen data. In this work, the features of each pixel that are extracted from a CT image are the histogram-normalized values of the pixels. These pixels belong to the neighborhood of the pixel under consideration (the target pixel). For 2-D analysis, a pixel’s neighborhood contains the pixels within a 5×5 window; for 3-D analysis, its neighborhood contains the pixels within a 3×3×3 window, i.e.
including 3x3 windows from adjacent CT images. Additionally, because some defects, such as splits, are near to the center, and some of them, such as bark and sapwood, are close to the outside edge of the log, the distance from the center of the log to the target pixel is also used as a feature. This distance measure contains contextual (or global) information that can improve classification. The neighborhood of a pixel under consideration for 2-D and 3-D analysis is shown in Figure 1.

Figure 1: ANNs containing a 2-D window (a) and a 3-D window (b) illustrate the network topology and relationship of input images to output classifications. The top left pixel in (a) is the input of the first node in the 2-D ANN, while the top left pixel in the previous slice (b) is mapped to the input for the first input node of the 3-D ANN. The distance $r$ is the last input to the ANN in both cases.
2.2.2 Topology

The topology of a neural network has an effect on the speed of convergence during training, and on the accuracy of the classification. Based on prior results (Li et al. 1996), the number of hidden nodes was chosen to be 12. The number of output nodes for the ANNs differed, however. In different families of species-dependent and species-independent classifiers, there are different defects to be labeled. For example, oak classifiers detect five classes: clear wood, knots, bark splits, and decay. Yellow poplar (*Liriodendron tulipifera*, L.) and oak combined classifiers identify six classes: heartwood, knots, bark, splits, decay, and yellow poplar sapwood. In 2-D classifiers, the topology is 26-12-5 or 26-12-6, which means that the structure of the neural network has 26 input nodes, 12 hidden nodes, and 5 or 6 output nodes. In 3-dimensional classifiers, the topology is 28-12-5 or 28-12-6, which has a similar interpretation.

2.3 Post-processing

Because classification features are based primarily on local neighborhoods, spurious misclassifications tend to occur at isolated points. A post-processing module is used to remove these small regions, and therefore improve overall system performance. The module includes two mathematical morphological operations: erosion and dilation.

After passing though an ANN classifier, a CT image is labeled and treated as a gray-level image. Then the image is post-processed by the morphological operations of erosion followed by dilation using a 5-point structuring element. In a CT image, splits appear close to the center of a log image, and their appearance after classification is a narrow line. If a split is post-processed, it is often deleted by the erosion operation. Hence, for all classifiers in our study, an entire image is not post-processed, only the outer regions of the log are post-processed. The range of the post-processed region of an image is currently selected manually. Each pixel whose distance $r$ is greater than 0.75 times the ideal log radius is chosen to be post-processed. This approach deletes misclassified small areas—which occur mostly near the outer edges of the log—but still retains important information (like splits) near the center of the log.

2.4 Training and Testing

An entire training/testing set for one hardwood species consists of approximately 1000 samples. 10-fold cross validation was used to evaluate the accuracy of each classifier. This means that the training set is randomly divided into 10 mutually exclusive test partitions of approximately equal size. For each of the 10 stages of training, one partition is designated as the test set, and the remaining samples in other partitions are used to train the neural network. In successive stages, different partitions are used for testing and the remaining samples are used for training. The average classification accuracy over all 10 stages of training is reported as the cross-validated classification accuracy. Cross validation not only provides a nearly unbiased estimate of the true classifier error rate, but the 10 estimates provided allow statistical analyses to be performed.
3 Experimental Design
As noted above, 1000 samples were taken from each of the species: oak, yellow poplar, and black cherry (*Prunus serotina*, L.). The percentages of these samples for each feature type across the different species appear in Table 1. Both oak and yellow poplar images have pixel resolution 2.5×2.5×2.5 mm$^3$. Whereas, the cherry log images were generated by a different scanner at a different resolution, approx. 0.95×0.95×0.95 mm$^3$. Because image texture differs at these different resolutions, we could not combine data across resolutions for multiple-species classifier development. Consequently, 3×3×3 neighborhoods in the 512×512 cherry images (cherry_512) were combined to produce new images (cherry_170) with approximately 2.84×2.84×2.84 mm$^3$ resolution. We felt that these averaged images would provide comparable texture to our earlier 2.5×2.5×2.5 mm$^3$ images. Having multiple resolutions within the same species also allowed us to compare classifier accuracy for 2 different image resolutions.

Table 1. Distribution of training/testing samples taken from different logs and different species. Decay was not present in the yellow poplar samples, and sapwood was not distinguished cherry and oak.

<table>
<thead>
<tr>
<th>Species</th>
<th>clear</th>
<th>knots</th>
<th>bark</th>
<th>splits</th>
<th>decay</th>
<th>sapwood</th>
</tr>
</thead>
<tbody>
<tr>
<td>cherry_170</td>
<td>47%</td>
<td>16%</td>
<td>15%</td>
<td>11%</td>
<td>11%</td>
<td></td>
</tr>
<tr>
<td>cherry_512</td>
<td>43%</td>
<td>16%</td>
<td>17%</td>
<td>12%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>oak</td>
<td>38%</td>
<td>13%</td>
<td>16%</td>
<td>17%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>yellow poplar</td>
<td>46%</td>
<td>15%</td>
<td>15%</td>
<td>5%</td>
<td>19%</td>
<td></td>
</tr>
</tbody>
</table>

Initial attempts to process yellow poplar images used a 28-12-4 topology for the 3-D ANN, which means that this preliminary classifier had four outputs: clear wood, knots, bark and splits (decay is not present in our yellow poplar images). For yellow poplar logs is which both heartwood and sapwood are present, the classifier performed quite poorly. This occurs because CT image values (density) for heartwood and sapwood are very different in yellow poplar. Therefore, it became necessary to distinguish yellow poplar sapwood from the generalized clear wood class (adding an additional classifier output for this class) in order to develop accurate classifiers that used yellow poplar image data.

Using 10-fold cross-validation we developed individual classifiers for each species—oak, yellow poplar, and cherry—using both 2-D and 3-D feature vectors (6 classifiers). Images used were the nominal (2.5mm)$^3$ resolution. We also developed multiple-species classifiers: pairing 2 species at a time, and also combining all 3 species together. These were also trained using 2-D and 3-D feature vectors for a total of 8 multiple-species classifiers. Finally, the finer resolution cherry images (0.95mm)$^3$ were used to train both a 2-D and 3-D classifier.
4 Results
Classification accuracies for the different classifiers appear in Figure 2. These line plots seem to indicate that 2-D has higher accuracy than 3-D for single-species classifiers, and the reverse performance for multiple-species classifiers. However, it is impossible to determine from these performance estimates whether these apparent differences reflect real accuracy differences. However, because ten-fold cross-validation was used, each trained classifier actually has 10 estimates of classification accuracy, resulting from the accuracy rates from each partition of the data sets. Therefore, these estimates can be used as samples in statistical Analysis of Variance (ANOVA).

Figure 2: 2-D and 3-D classifier accuracies are plotted for each of the ANN classifiers—oak (RO), cherry (CH), yellow poplar (YP), 512×512 cherry (CH_512), cherry/oak (CH_RO), cherry/yellow poplar (CH_YP), oak/yellow poplar (RO_YP), and all 3 species combined (COMB).

In our first statistical test, we separate the full set of classification rates into two groups: dimensionality, which includes two-dimensional and threedimensional classifiers, and cardinality, which includes single (species-dependent) and multiple (species-independent) classifiers. ANOVA treatments, in this case, are single and multiple cardinality, and are blocked on the dimensionality of the classifiers (2-D or 3-D). The f-ratio results for the dimensionality and cardinality are 0.055 ($p = 0.815$) and 27.4 ($p < 0.001$), respectively. It is clear that the F ratio of the dimensionality is much lower than that of cardinality (the former F-ratio is not significant), which indicates (at this point) that differences exist between the mean classification rates for the single- and multiple-species classifiers. The interaction of dimensionality and cardinality is also significant, indicating a combined effect. This can be seen in the average classification rates of Figure 2, where 2-D rates are generally higher for single-species classifiers and 3-D rates are generally higher for multiple-species classifiers.
To understand greater details about the differences between dimensionality and cardinality, we performed ANOVAs for single- and multiple-species classifiers separately. For the single-species classifiers, ANOVA treatments are species (CH, RO, and YP) and dimensionality (2-D and 3-D) is used for blocking. F-ratio values for species and dimensionality are 11.4 \( (p < 0.0005) \) and 9.53 \( (p = 0.003) \), respectively. Probability values associated with post-hoc T-tests are shown in Table 2. They demonstrate that the classification rates for the cherry-specific classifier is significantly different from both those of the oak- and the yellow poplar-specific classifiers. However, there is no significant difference between the oak and yellow poplar single-species classifiers.

### Table 2. A matrix of pair-wise T-test probability values for the classification rates of cherry, oak and yellow poplar single-species classifiers.

<table>
<thead>
<tr>
<th></th>
<th>CH</th>
<th>RO</th>
<th>YP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RO</td>
<td>0.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>YP</td>
<td>0.001</td>
<td>0.995</td>
<td>1.000</td>
</tr>
</tbody>
</table>

For the multiple-species classifiers, ANOVA treatments are species (CH_RO, CH_YP, COMB, RO_YP) and dimensionality (2-D and 3-D) is used for blocking. F-ratio values for species and dimensionality are 39.3 \( (p < 0.005) \) and 4.97 \( (p = 0.032) \), respectively. Probability values associated with post-hoc T-tests are shown in Table 3. Those values indicate that the CH_RO classifier has significantly greater accuracy than the other 3 multiple-species classifiers. In addition, the RO_YP classifier has greater accuracy than the two, lowest accuracy classifiers, COMB and CH_YP. Both of those latter 2 classifiers contain both cherry and yellow poplar samples, which seem to create classification problems. T-tests indicate that COMB and CH_YP are not significantly different from one another.

### Table 3. A matrix of pair-wise T-test probability values for the classification rates of cherry, oak and yellow poplar multiple classifiers.

<table>
<thead>
<tr>
<th></th>
<th>CH_RO</th>
<th>CH_YP</th>
<th>COMB</th>
<th>RO_YP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH_RO</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH_YP</td>
<td>0.000</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMB</td>
<td>0.000</td>
<td>0.606</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>RO_YP</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Based on the obvious classification problems stemming from combining cherry and yellow poplar samples, we performed our original ANOVA again. This time, treatments were cardinality again, but only CH_RO and RO_YP were included in the multiple-species classifiers (no cherry and yellow poplar combinations). In addition, the fine resolution (0.95mm) cherry classifier (CH_512) was excluded from the single-species classifiers. As before, we blocked the ANOVA on dimensionality (2-D and 3-D). The resulting F-ratio value
for cardinality is 0.050, which indicates that there is no difference between single- and multiple-species classification rates when cherry/yellow poplar combinations are removed.

Finally, we perform an ANOVA to compare the effect of CT resolution on classifier performance. We eliminated the effect of dimensionality by blocking on it. We found that the classifier for finer resolution cherry images does not differ significantly in classification rate from the classifier used for coarser resolution cherry images.

5 Conclusions

Eight single-species classifiers were trained using both 2-D and 3-D image data. The accuracy of all 8 classifiers is above 95%. Six, two-species classifiers were also trained using both 2-D and 3-D image data. Two of them are oak and yellow poplar combined classifiers, two of them are oak and cherry combined classifiers, and two are cherry and yellow poplar combined classifiers. Their accuracy is 90%-97%. Finally, two, three-species classifiers (oak, yellow poplar and cherry) were generated for 2-D and 3-D analysis. These two classifiers identified six kinds of defects: clear wood, knots, bark, splits, decay and yellow poplar sapwood. Their accuracy is about 91%-92%.

For single-species classifiers, the performance of 2-D classifiers is better than that of 3-D classifiers. For multiple-species classifiers, the performance of 3-D classifiers is better than that of 2-D classifiers. We conjecture that in single-species classification multiple image planes contain redundant data that may be unimportant, or even counter-productive, for accurate classification. For multiple-species classification, however, the extra information contained in previous and subsequent CT slices seems to aid feature labeling. Consequently, as we increase the species mix that a classifier must deal with, it appears that 3-D features are important for attaining high accuracy.

Higher resolution images do not seem to have a significant difference on performance. We were able to achieve similar accuracies in cherry using ~1mm resolution and ~3mm resolution. This means that our ANN classification approach is general enough to be applied broadly to CT images of varying resolutions. All that is required is resolution-specific training so that the classifier can learn the local neighborhood patterns.

In comparing single-species classifiers and multiple-species classifiers, the performance of the former is better than that of the latter when cherry and yellow poplar data are used. On the other hand, when that species combination is excluded, there is no significant difference between classification accuracy for single- and multiple-species classifiers. Yellow poplar has traditionally been difficult to deal with because it possesses many intrinsic differences (wood structure, density) to most other fine-grained hardwoods, e.g., cherry. Yellow poplar was included in the study because it is an extreme case, and we desired to delineate a worst-case scenario. Consequently, the difficulty we experienced in combining it with cherry here is neither surprising, nor particularly worrisome. All of these accuracies (90%-98%) should be acceptable for industrial use. Furthermore, it should be noted that all reported accuracies are prior to post-processing. We have visually determined (via classified images) that post-
processing does improve accuracy, but we do not yet have a quantitative estimate for that improvement. Consequently, we expect that the range of accuracies reported above is actually higher, which further enhances its potential for industrial use.

6 References


Proceedings from the 3rd International Seminar/Workshop on Scanning Technology and Image Processing on Wood

SKELLEFTEÅ, SWEDEN
Aug. 17 – 19, 1998

Edited by Ph.D. Owe Lindgren
Ph.D. Anders Grönlund
Ph.D. Olle Hagman

SKELLEFTEÅ CAMPUS
Department of Wood Technology