

Recognition of Porosity in Wood Microscopic Anatomical Images

Shen Pan^{1,2} and Mineichi Kudo¹

¹ Graduate School of Information Science and Technology, Hokkaido University,
Sapporo, 060-0814, Japan

{panshen,mine}@main.ist.hokudai.ac.jp

² Department of Information Management and Information Systems,
Hefei University of Technology, Hefei, 230009, China

Abstract. The size and configuration of pores are key features for wood identification. In this paper, these features are extracted and then used for construction of a decision tree to recognize three different kinds of pore distributions in wood microscopic images. The contribution of this paper lies in three aspects. Firstly, two different sets of features about pores were proposed and extracted; Secondly, two decision trees were built with those two sets by C4.5 algorithm; Finally, the acceptable recognition results of up to 75.6% were obtained and the possibility to improve was discussed.

Keywords: wood identification, porosity of wood, wood microscopic image, C4.5.

1 Introduction

Intelligent systems for recognition of wood species have been developed to identify woods according to some features, particularly wood anatomy features such as vessels, perforation plates, parenchyma and so on. It is expected that such a process can be done automatically by a computer without any manual intervention. Some of the latest intelligent recognition systems are based on macroscopic features such as color and texture in macroscopic images. About 30 different kinds of woods have been recognized by using these systems [1] [2]. The advantage of these systems is due to the simple process. Neither special equipment such as a microscope nor wood slicing is required. Nevertheless, information obtainable from macroscopic images is limited and is not sufficient for identifying a wide range of woods. Therefore, information from microscopic features is necessary for accurate classification of species in a wide range of woods [3]. Indeed, the International Association of Wood Anatomists (IAWA) published a list of microscopic features for hardwood identification [4]. From the list published by IAWA, we can find over 100 features that are used to identify hard wood. On the other hand, for human inspectors, much training time is necessary for gaining sufficient ability to use such complicated features. The same thing happens even

to a computer if all the features are given. It is also known that too many features degrade the classification performance in general, so that feature selection has been discussed in a long history of pattern recognition [5]. Feature selection also helps to reduce time and labor for measuring the values of the features. We therefore focus first on the most important features according to the domain knowledge. One of these features is vessels.

The vessels of hard wood appear as pores in a cross section of wood slide. The size, distribution, combination and arrangement of pores are important features to recognize the species of hard wood [6], and the pore distribution in particular contributes most to recognition. Pores have three kinds of different distribution shapes which are also known as porosity according to their early wood/late wood transition as depicted in Fig. 1: *ring*, *semi-ring* and *diffuse*. In ring porous wood, each region surrounded by two growth rings has large pores in the early wood zone and small vessels in the late wood zone. The large pores can be observed with naked eyes, but the small vessels can only be observed by a microscope. In semi-ring porous wood, the pores in the early wood zone have a large diameter and gradually decrease in size toward the late wood zone. In some cases, semi-ring porous woods also have pores of the same sizes in early wood and late wood, but the frequency of pores in early wood is higher than that in late wood. In diffuse porous wood, pores of almost the same sizes are distributed uniformly across the entire zone [7].

Pattern recognition technique and digital image analysis technology have been successfully integrated into a strong tool for dealing with many aspects of agriculture, such as inspection and grading of agriculture and food products [8][9][10], tracking animal movements [11], machine vision based guidance systems [12][13], analysis of vertical vegetation structure [14], green vegetation detection [15][16], and weed identification [17][18]. However, in the case of wood species recognition by microscopic information, there hardly exists any work on discussing how to recognize these features by computers, although IAWA published the list of microscopic features for hardwood identification about 20 years ago. This paper

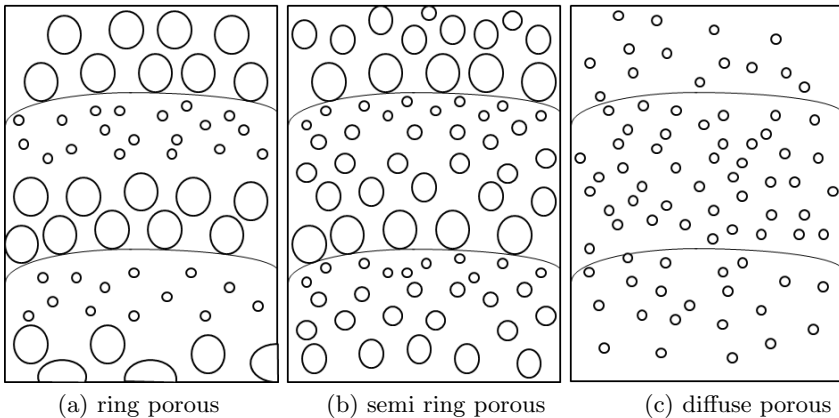


Fig. 1. Typical configurations of three different kinds of wood porosity

gives an algorithm to recognize the three kinds of porosity in hard woods. Section 2 presents the samples and the methods used in the paper. Section 3 gives the results and discusses them. Finally, the conclusion is given in Section 4.

2 Materials and Methods

In the following, the information about images used in this paper is introduced firstly, then the proposed process for distinguishing three kinds of hard wood species is described in three parts in order: segmentation, feature extraction and classification.

2.1 Image Data

Wood microscopic images were collected as basic research materials in this study. We selected 135 microscopic cross sectional images from the database of Japanese woods (<http://f030091.ffpri.affrc.go.jp/index-E1.html>) that include three different kinds of wood pore distribution. These 135 images are divided into 45 diffuse-porous images, 45 ring-porous images and 45 semi ring-porous images. In the database of Japanese woods, every wood image has an identification key named TWTwNo. Besides the original image, we can find more detailed information such as wood species, collection date, collection place and collectors in the database according to the TWTwNo. The TWTwNo information of all the images used in this paper can be found in Table 1.

All of the images have the same size of 1500×997 , which means a height of 1500 pixels and width of 997 pixels, and, they were saved in JPEG image format. A scale bar of 1 mm is marked at the right bottom corner of each image. One image of ring porous wood is shown in Fig. 2. It is a microscopic cross-sectional image of *Araliaceae Kalopanax pictus* which was taken by a Nikon D100 camera in 2002.

Table 1. TWTwNo of every image

Porosity	TWTwNo
diffuse	13908 16201 16016 15927 5691 18155 15129 16944 15098 1307
	18072 4340 14822 4356 4397 15860 16082 17072 6377 18134
	12829 6369 419 14291 13899 12831 12846 14352 13879 14911
	12836 12909 15094 12820 15087 12907 15223 757 521 16063
	43 416 12916 14256 15168
ring	17512 13971 17535 2669 15934 15504 14866 15494 4000 13900
	14334 3385 19817 4334 9308 16954 17525 4337 5774 9323
	423 15897 14174 739 13874 18074 17050 13421 16941 17969
	5775 9321 25 6363 13966 3373 4818 4000 13956 18025
	516 2874 6329 14887 17545
semi ring	14281 14289 14277 18565 16268 15486 16976 15671 16315 14367
	6365 18394 15910 12843 3407 14810 14904 17321 640 4343
	14759 14368 2228 460 14275 16952 2578 14279 18602 18011
	14766 2873 13925 17559 15942 16282 14888 16294 426 2556
	15515 15527 15800 17544 18549

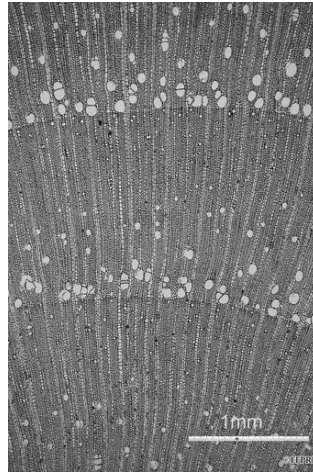


Fig. 2. One image of ring porous wood: *Araliaceae Kalopanax pictus*

2.2 Segmentation Algorithm

In a microscopic cross sectional image, we can recognize many tissues other than pores such as xylem ray, parenchyma, growth rings, fibers and other tissues (Fig. 3). Therefore only pores have to be spotted.

There are two difficulties to be solved in pore segmentation. The first difficulty is due to the variety of sizes and the variety of shapes. For example, large pores have tangential diameters of more than $300 \mu m$, while those of small pores are less than $100 \mu m$. Some pores exist solitarily, while others are multiple or even arranged in a chain, cluster or band. The other difficulty comes from the existence of fibers and longitudinal parenchyma. The parenchyma and pore are similar in color and shape but different in size. Thus, taking such a slight difference into consideration is necessary to improve the accuracy of segmentation.

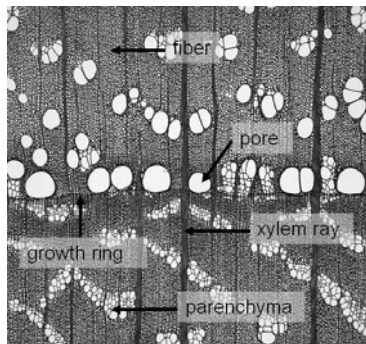


Fig. 3. Many tissues including pores found in an microscopic image

The authors gave already an effective algorithm based on mathematical morphology to solve the above problem [19]. The algorithm uses a disk shape structuring element that can change its radius according to the size of pores. All the pores information are saved as $p_i = (x_i, y_i, s_i)$, after getting an accurate result of segmentation, where p_i means the i th pore, x_i and y_i are the center positions of the pore, s_i represents the area of the pore. The range of positions is $0 \leq x_i \leq 997$, $0 \leq y_i \leq 1500$. The unit of area s_i is pixel and $s_i > 0$. It is worth noting that there are some inaccuracy in $p_i = (x_i, y_i, s_i)$ because mathematical morphology cannot provide very precise edges of pores in segmentation results. For example, the erosion operation will cause reduction of pores area. Fortunately this slight inaccuracy will not cause serious problem to the following process.

2.3 Feature Extraction

The individual pore information $p_i = (x_i, y_i, s_i)$ is not directly useful for recognizing the three kinds of porosity. The configuration also has to be taken into consideration. A promising feature seems to be the diameter change along to the vertical direction (early to late zones) as explained before. Therefore we use such features as the first feature set. In addition, we prepare another feature set. In general, features are desirable to be invariant to rotation, scale and translation of images. In our material, scale seems almost the same because of the microscopic measurement, but rotation and translation should be considered. Therefore, we focus on local features determined by the nearest pairs of pores. For each pore, we find the nearest pore and measure the size difference, the relative direction and the distance between them. After that, we construct a histogram over these values so that the histogram features are invariant to rotation and translation. Strictly speaking, the rotation makes change the values but it is only within the change of histogram bin numbers.

In this section, we will show two feature sets: one is of the features connected to diameter change of pores along to the vertical direction and another is of the features that are invariant to rotation and translation.

The features of vertical direction. It is noted that the difficulty in recognizing the growth rings leads to the difficulty in detecting the local diameter change of pores; however, it is not difficult to consider the global diameter change of all the pores in the whole image along the vertical direction. The following procedure is applied to extract such a global diameter change.

- Step1. Divide the image into 30 equally-sized divisions D_j , $j \in [1, 2, \dots, 30]$, along by the vertical direction. A sub-image D_j is a strip with a size of 50×997 .
- Step2. Normalize all the areas of pores into $[0 \ 1]$ by the maximum value of pore areas.
- Step3. Calculate the average \overline{S}_j of areas of pores in each D_j .

Step4. In order to inspect the global diameter change of pores, we sort D_j by the value \bar{S}_j , so that D_1 can be regarded as the begin of the early zone and D_{30} can be regarded as the end of the late zone.

We sort D_j because there are always more than one growth ring in an image and the internal varies. After sorting, we can treat all the pores as those between two growth rings (Fig. 4).

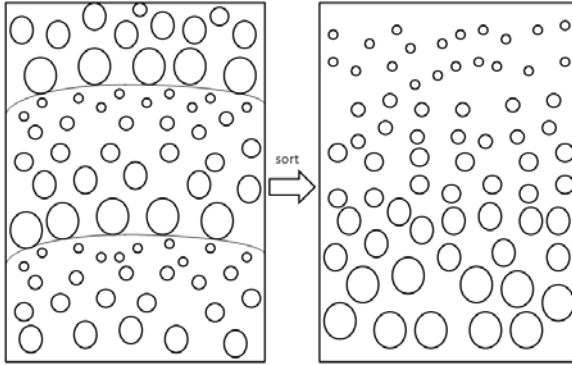


Fig. 4. Sort D_j by the value \bar{S}_j

As a result, we have a graph feature set of 30 values.

The results are shown in Fig. 5 for one example of each class. We can observe that the average areas (thus the radii) of pores decrease gradually from early zones to late zones and some degree of difference between these three classes is detectable.

The variety in graph of diffuse porosity is the least among three kinds of porosity. Indeed, the size of pores in diffuse porous wood is almost the same regardless of position. The ring porosity has the largest variety because there are tremendous changes in pore size between early wood and late wood. The variety of semi ring porosity is between diffuse and ring porosity.

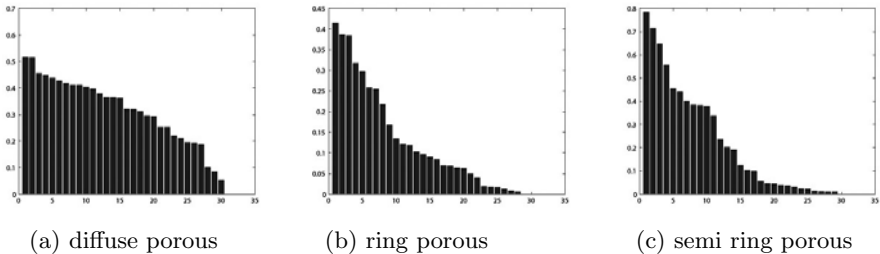


Fig. 5. Area change of pores along to the vertical line (1 to 30 according to the early to late zones)

Invariant features to rotation and translation. Next, a feature set invariant to rotation and transition is extracted as follows:

Step1. For each p_i , quantize the value s_i of area into one of three values according to the rule $\{1 : s_i \in [0, 0.15); 2 : s_i \in [0.15, 0.35); 3 : s_i \in [0.35, 1.0)\}$.

Here, s_i is normalized to $(0, 1]$.

Step2. Find the nearest neighbor p_j of each p_i with the same quantized size.

Step3. Calculate the angle θ_i and distance d_i between $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$:

$$\theta_i(p_i, p_j) = \arctan \frac{y_j - y_i}{x_j - x_i} \tag{1}$$

$$d_i(p_i, p_j) = \sqrt{(y_i - y_j)^2 + (x_i - x_j)^2} \tag{2}$$

Step4. Quantize the normalized distance $d_i \in (0, 1]$ into one of three values according to $\{1 : d_i \in [0, 0.25); 2 : d_i \in [0.25, 0.45); 3 : d_i \in [0.45, 1.0)\}$.

Similarly, the angle θ_i is quantized according to $\{1 : \theta_i \in [0, \frac{\pi}{3}); 2 : \theta_i \in [\frac{\pi}{3}, \frac{\pi}{2}); 3 : \theta_i \in [\frac{\pi}{2}, \frac{2\pi}{3}); 4 : \theta_i \in [\frac{2\pi}{3}, \pi)\}$.

Step5. Construct a histogram H over $36(= 3 \times 4 \times 3)$ bins by quantizing the (p_i, p_j) pair for every p_i , where p_j is the nearest to p_i and both have the same size. Here, $H[a, b, c]$ corresponds to the frequency of pairs (p_i, p_j) producing a th size, b th angle and c th distance.

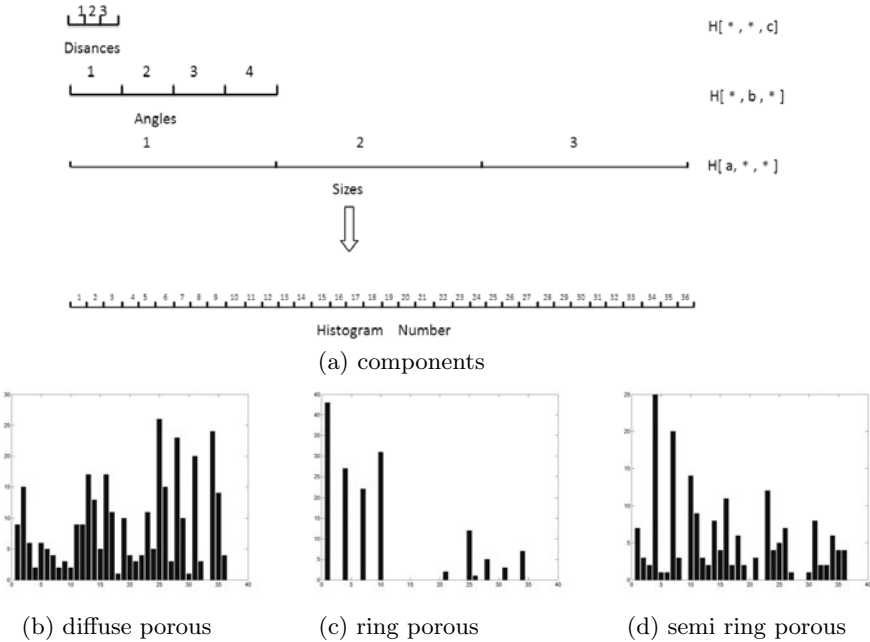


Fig. 6. Histogram of invariant features.

Fig. 6 shows three different histograms for three different porosities. From Fig. 6, we can observe that different porosity types have different histograms of invariant features. For diffuse porosity, almost every bin in histogram has a non-zero value. It means that pores in diffuse porosity have different sizes, angles and distances. It is reasonable because pores are distributed uniformly without any order in diffuse porosity. For ring porosity, we can find some near-zero values of components in the histogram, especially in the middle area of the histogram. A possible reason is that the translation of pore size from early wood zone to late wood zone is rapid, so that only large and small pores are observable. We also notice that close pairs of pores ($\#bin=1,4,7,\dots$) are much more found in ring porosity. Semi ring porosity shows an intermediate characteristic between two others.

2.4 Porosity Recognition

The decision tree with features of vertical direction. In order to analyze the discriminative information of those feature sets, we use C4.5 algorithm [20]. C4.5 algorithm generates a set of classification rules as a decision tree.

First, we used the first set of features of vertical direction. The decision tree is shown in Fig. 7. We can see some simple rules from the tree. For example, if the average pore area is larger than 0.05 in the 24th band (of 30 bands), and the average pore area is larger than 0.1553 in the 26th band, then the image will be classified to 'D' (diffuse porosity). In the total 135 images, there are 41 images classified to 'D' by this rule, however, the other 3 images are misclassified to 'D'. For the decision tree, the dominant rules are three. For diffuse porous woods, the pore area in late wood zone such as the 24th and 26th bands should be relatively large. For ring porous woods, the average pore area in late zone (the 24th band) should be very small. The semi-ring porosity should not satisfy either of these two rules. These rules are almost consistent to our knowledge of porosity.

The decision tree with invariant features. Next, in order to know the detailed relationship between the invariant features and the porosity, we use C4.5 again with the second set of features.

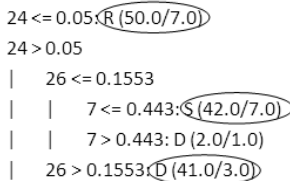


Fig. 7. Decision tree with features of vertical direction. The node such as ' $n \leq v : k(a/b)$ ' means a samples are classified as porosity k , $k \in \{D', R', S'\}$, if the value of the n th feature is less than or equal to v , while b samples are misclassified.

Fig. 8 is the decision tree for porosity recognition given by C4.5 algorithm. Each node corresponds to the decision at one component in the histogram. From this figure, we find that most ring porosity images (circled in the decision tree) satisfy the condition $[2,3,1] \leq 8$, $[2,1,3] \leq 4$, $[2,4,2] \leq 3$ and $[2,1,2] \leq 2$. It means roughly that middle size pores ($[2,*,*]$) should be less regardless of the angle and distance. This is consistent to the observation seen in the rule for ring porous images with the first set of features. For most semi-ring porous images (circled in the decision tree) satisfy the condition $[2,3,1] > 8$, $[1,4,1] > 19$, $[3,2,1] \leq 32$ and $[3,4,1] > 4$. It means that this porosity is decided by pairs with small distance ($[*, *, 1]$). The change of pore size is smooth in semi-ring porosity, and therefore same size neighbors tend to be found in short distances. For the same reason, ' $[3,2,1] \leq 32$ ' says that large and close pairs in vertical direction should be not so many. For most diffuse porous images (circled in the decision tree), the length of condition part of the most dominant rule is shorter than those of two others ($[2,3,1] > 8$ and $[1,4,1] \leq 19$). Such a simpler rule implies that the diffuse porosity is easy to be separated from the other two porosities.

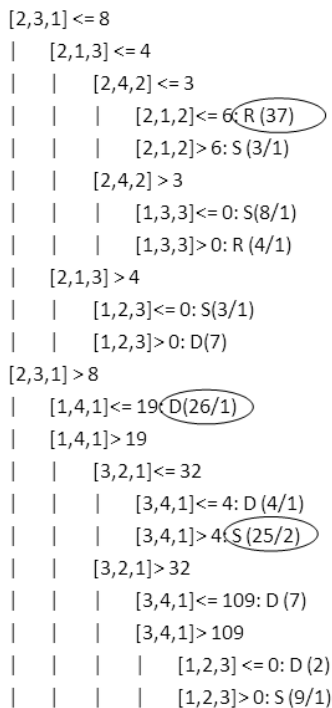


Fig. 8. The decision tree with invariant features. The node such as ' $n \leq v : k(a/b)$ ' means a samples are classified as porosity k , $k \in \{D', R', S'\}$, if the value of the n th feature is less than or equal to v , while b samples are misclassified.

3 Results and Discussion

3.1 Classification Performance

We obtained an estimate of correct recognition by 10-fold cross-validation [21].

Confusion matrix with vertical direction features. The confusion matrix of 10-fold cross-validation comes as below when the features of vertical direction are used:

D	R	S	<-classified as
38	0	7	D=diffuse
0	37	8	R=ring
10	8	27	S=semi

In the confusion matrix, the numbers in diagonal are the number of samples correctly classified; the others are the numbers of incorrectly classified samples. There are totally 102 samples classified correctly, bringing the accuracy of 75.6%. Especially no misclassification occurred between diffuse porosity and ring porosity. For the other combinations, the accuracy is not so high.

Confusion matrix with invariant features. The confusion matrix when invariant features are used is given as below:

D	R	S	<-classified as
30	3	12	D=diffuse
4	35	6	R=ring
13	6	26	S=semi

In total, 91 samples were classified correctly at recognition rate of 67.4%. The classification between ring porous and semi-ring porous is improved from the previous result, while the classification rate between ring porous and diffuse porous is worse.

Judging from this result and the description of the decision rule, this set of classification rules seems a little too complicated than necessary and cause over-fitting to the training data.

3.2 Discussion

Either 75.6% or 67.4% is not so good when we compare these values with those of many applications of pattern recognition. However, it should be noted that even a well-trained inspector sometimes fails to recognize the porosity of given sample images. It implies that the attainable classification rate might be not so high. For example, the image of Fig. 2 is allowed to assign to both of ring porosity and semi-ring porosity according to Microscopic Identification of Japanese Woods (<http://f030091.ffpri.affrc.go.jp/fmi/xsl/IDB01-E/home.xsl>). In other words, there are some cases in which no one knows the correct answer or more than one correct answer exists.

Ring porosity and diffuse porosity are two opposite porosity types, so that we cannot find any wood species belongs to both porosity types at the same time in practice. Indeed, it is not difficult for human to distinguish them because there is clear difference in their appearance of pore distribution in images. Our decision trees also succeeded to recognize them with high accuracy. Most of misclassified samples are between the other porosity pairs, one is ring and semi ring porosity pair, the other is diffuse and semi ring porosity pair. These porosity pairs are not easy to distinguish even by a well-trained inspector. Fig. 9 demonstrates this.

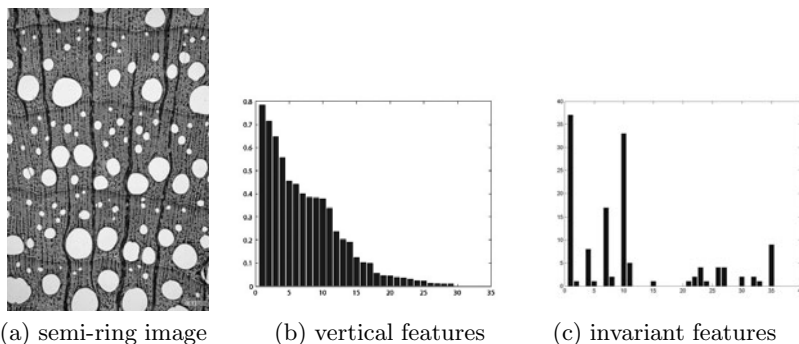


Fig. 9. A case in which semi-ring porous sample is misclassified as ring porosity

From the microscopic image (Fig. 9(a)), it is hard to classify the image correctly from the pore distribution. Indeed, there are some pores whose sizes are between large and small. The correct porosity is 'semi-ring' but we might think it as 'ring' because there are large pores in early zone and small pores in late wood zone. The graph of vertical features also detected a rapid change of pore size from large to small (Fig. 9(b)). As a result, the rule found the fact that the pore size in late wood (the 24th component in the graph) is small enough (≤ 0.05), thus classified it as ring porosity. The decision tree of invariant features also misclassified because the number of middle size pores is very small.

The diffuse and semi ring porosity pair is also confusing. Fig. 10 gives an example of this case.

From the above microscopic image (Fig. 10(a)), we can find that the frequency of pores in early wood is higher than that in the late wood. It is a strong evidence for the image to be semi-ring porous. However, the graph of vertical features tells us that the pore size is almost the same, so the rule classified the image as diffuse porosity. The decision tree of invariant features also fails because the 19th component is larger than 8 and the 10th component is smaller than 19.

In case of invariant features, there are also some misclassification between diffuse porosity and ring porosity. In some diffuse porous samples, most of the pores have almost the same size. Fig. 11 shows a diffuse porous sample and its histogram. Most of the pores are large (concentrating on the right of the histogram) and the number of middle size pores ($[2,*,*]$) is very small (less than 10). It caused

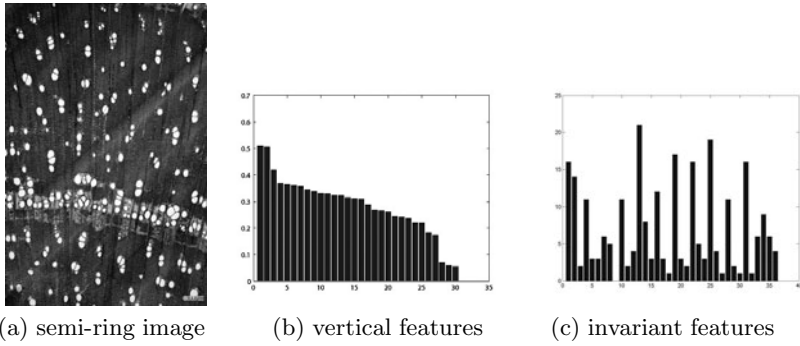


Fig. 10. A case in which semi-ring porous sample is misclassified as diffuse porosity

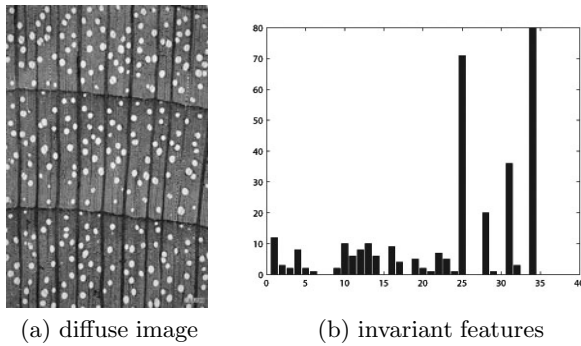


Fig. 11. A case in which diffuse porous sample is misclassified as ring porosity

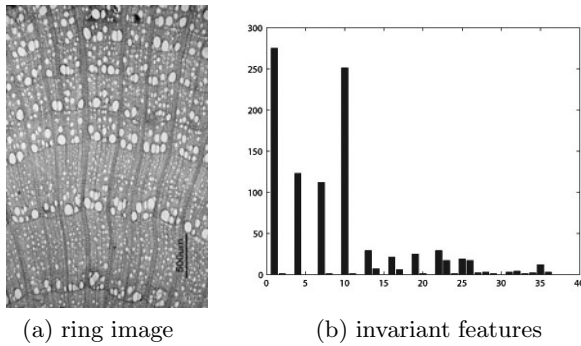


Fig. 12. A case in which ring porous sample is misclassified as diffuse porosity

misclassification from diffuse to ring porous. A reverse case (from ring to diffuse) is shown in Fig. 12. Undoubtedly, Fig. 12(a) looks ring porous, but the corresponding histogram satisfied a rule for diffuse porous with condition $[2, 3, 1] > 8$, $[1, 4, 1] > 19$, $[3, 2, 1] \leq 32$, $[3, 4, 1] \leq 4$.

4 Conclusion

In this paper, a novel procedure has been introduced to recognize three kinds of porosity in wood microscopic images. In the procedure, mathematical morphology is used to segment pores from an image; then two different kinds of feature sets are extracted; those features are used with C4.5 algorithm to generate decision trees. The estimator of 10-fold cross-validation is used to verify the classification performance of the decision trees. As a result, we found that both decision trees distinguish well between diffuse and ring porosity, but do not between semi-ring porosity and the other two kinds of porosity. Although some reasons of failure were investigated from the domain knowledge, it is still necessary to do more work on understanding the decision trees and analyzing the reason of over-fitting. On the other hand, since vertical feature set has a better result in recognizing diffuse porosity and ring porosity, invariant feature set has a better result in recognizing ring porosity and semi-ring porosity. Combining these two feature sets may be an other method to improve the accuracy of the three classes.

Acknowledgement. The core program of C4.5 is realized by Weka [22]. The feature extraction program is coded by Matlab. The wood microscopic images used in this study were obtained from the database of Japanese woods that are copyrighted by the Forestry and Forest Products Research Institute.

References

1. Tou, J.Y., Lau, P.Y., Tay, Y.H.: Computer Vision-based Wood Recognition System. Paper presented at the Proceedings of the International Workshop on Advanced Image Technology, Bangkok, Thailand (2007)
2. Khalid, M., Yusof, R., Liew, E., Nadaraj, M.: Design of an intelligent wood species recognitions system. *International Journal of Simulation System, Science and Technology* 9(3), 9–19 (2008)
3. Xu, F.: *Anatomical Figures for Wood Identification* (in Chinese with English title). Chemical Industry Press, Beijing (2008)
4. Wheeler, E.A., Baas, P., Gasson, P.E.: IAWA list of microscopic features for hardwood identification. *IAWA Bull (N.S.)* 10, 219–332 (1989)
5. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern recognition. *Pattern Recognition* 33(1), 25–41 (2000)
6. Xu, Y.M.: *Wood Science*. China Forestry Publishing House, Beijing (2006) (in Chinese)
7. Bond, B., Hamner, P.: *Wood Identification for Hardwood and Softwood Species Native to Tennessee*. Agricultural Extension Service, Knoxville (2002)
8. Brosnan, T., Sun, D.W.: Inspection and Grading of Agricultural and Food Products by Computer Vision Systems—a Review. *Computers and Electronics in Agriculture* 36(2-3), 193–213 (2002)
9. Bulanon, D.M., Kataoka, Y., Hiroma, T.: A Segmentation Algorithm for Automatic Recognition of Fuji Apples at Harvest. *Biosystems Engineering* 83(4), 405–412 (2002)
10. Nakano, K.: Application of Neural Networks to the color Grading of Apples. *Computers and Electronics in Agriculture* 18(2-3), 105–116 (1997)

11. Tillet, R.D., Onyango, C.M., Marchant, J.A.: Using Model-Based Image Processing to Track Animal Movements. *Computers and Electronics in Agriculture* 17(2), 249–261 (1997)
12. Reid, J., Searcy, S.: Vision-based guidance of an agricultural tractor. *IEEE Control Systems Magazine* 7(2), 39–43 (1987)
13. Sogaard, H.T., Olsen, H.J.: Determination of crop rows by image analysis without segmentation. *Computers and Electronics in Agriculture* 38(2), 141–158 (2003)
14. Zehm, A., Nobis, M., Schwabe, A.: Multiparameter analysis of vertical vegetation structure based on digital image processing. *Flora-Morphology, Distribution, Functional Ecology of Plants* 198(2), 142–160 (2003)
15. Laliberte, A.S., Rango, A., Herrick, J.E., Fredrickson Ed, L., Burkett, L.: An object-based image analysis approach for determining fractional cover of senescent and green vegetation with digital plot photography. *Journal of Arid Environments* 69(1), 1–14 (2007)
16. Zheng, L., Zhang, J., Wang, Q.: Mean-shift-based color segmentation of images containing green vegetation. *Computers and Electronics in Agriculture* 65(1), 93–98 (2009)
17. Tellaeché, A., Burgos-Artizzu, X.P., Pajares, G., Ribeiro, A.: A vision based method for weeds identification through the Bayesian decision theory. *Pattern Recognition* 41(2), 521–530 (2008)
18. Bakker, T., Wouters, H., Asselt van, K., Bontsema, J., Tang, L., Muller, J., Straten van, G.: A vision based row detection system for sugar beet. *Computers and Electronics in Agriculture* 60(1), 87–95 (2008)
19. Pan, S., Kudo, M.: Segmentation of pores in wood microscopic images based on mathematical morphology with a variable structuring element. *Computers and Electronics in Agriculture* 75(2), 250–260 (2011)
20. Quinlan, J.R.: *Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo (1993)
21. Geisser, S.: *Predictive Inference*. Chapman and Hall, New York (1993)
22. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11(1) (2009)