Transactions on Machine Learning
and Data Mining
Vol. 5, No. 1 (2012) 45-62

ISSN:1865-6781,
ISBN: 978-3-942952-11-8

ibai Publishing
www.ibai-publishing.org

Recognition of Wood Porosity Based on Direction Insensitive Feature Sets

Shen Pan[1,2] and Mineichi Kudo[1]

[1] Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060-0814, Japan
{panshen,mine}@main.ist.hokudai.ac.jp
[2] Department of Electronic Business, Hefei University of Technology, Hefei, 230009, China

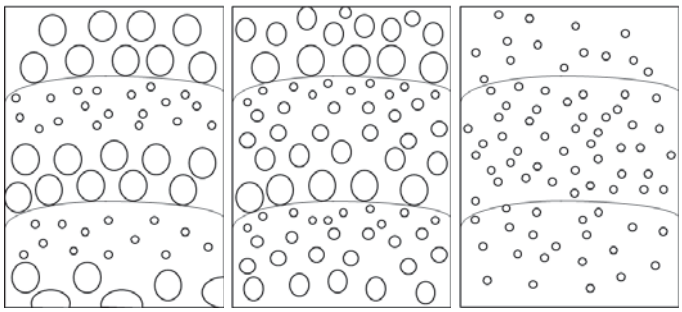
Abstract. The size and configuration of pores are key features for automatic wood identification. In this paper, two feature sets insensitive to rotation and transition change are extracted and then used for construction of decision trees for recognizing three different kinds of pore distributions in wood microscopic images. The contribution of this paper lies in three aspects. Firstly, two direction insensitive sets of features for porosity classification are designed and extracted, Secondly, for turning the found classification rule into human-readable knowledge, decision trees are built with the feature sets by C4.5 algorithm; Finally, rules extracted from the decision trees are explained according to domain knowledge of wood science.

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. Much training time is necessary for human inspectors to gain sufficient capability to handle such complicated features. It implies also that we need a very large number of training samples even if we rely on a computer to learn a classification rule, and a non-ignorable amount of cost time is necessary for measuring these features. In addition, it is known that too many features deteriorate the classification ability of a classifier learned from examples, so that the effectiveness of feature selection has been repeatedly demonstrated 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 on a small number of features, such as vessels, that are known to be useful for classification from domain knowledge. One of these features is vessels.



(a) ring porous (b) semi ring porous (c) diffuse porous

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

The vessels of hard wood appear as pores in a cross section of wood slice. 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, except for ours[19][20], on discussing how to recognize these features automatically by computers, although IAWA published the list of microscopic features for hardwood identification about 20 years ago. This paper gives an algorithm to recognize the three kinds of porosity in hard woods. Section 2 presents the samples used in the paper. Section 3 gives algorithms used to extract direction insensitive feature sets. The decision trees generated by C4.5 algorithm and other classifiers are shown in Section 4. Section 5 gives the results and discussion on them. Finally, the conclusion is given in Section 6. The samples used in this paper is the same as those used in ICDM2011[20], but there are some extended contents: (1)Both of the feature sets used in ICDM2011 are redesigned and improved into direction insensitive feature sets that means the rotation of original images will not impact the final results. (2)More reasonable and readable decision trees are generated and then explained by domain knowledge. (3)Some misclassified samples are analyzed and the problems in our recognition algorithm are explained based on these samples.

2 Images and Segmentation Algorithm

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. As space is limited, we cannot give all these information that is obtained from TWTNo, readers can find their interested information in the database according to the paper published in ICDM2011[20] because the same samples were used.

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

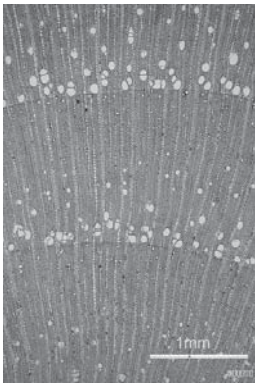


Fig. 2. One image of ring porous wood: *Araliaceae Kalopanax pictus*. It can be classified to semi ring porous wood at the same time.

image of *Araliaceae Kalopanax pictus* which was taken by a Nikon D100 camera in 2002.

2.2 Segmentation Algorithm

In a microscopic cross sectional image, we can find 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.

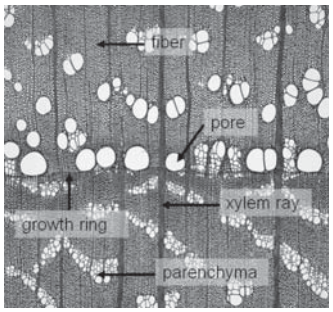


Fig. 3. Many tissues including pores found in an microscopic image

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 μm, while those of small pores are less than 100 μm. 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.

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 ith 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.

3 Feature Extraction

3.1 Previous Works

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. Two kinds of feature set have been already designed and extracted by authors [20]: One of them is the diameter change along to the vertical direction (early to late zones) and the other is the histogram features based on the nearest pore pairs. In general, features are desirable to be invariant to rotation. In our past experiments, scale seemed almost the same because of the microscopic measurement, but rotation and translation were not considered carefully. Therefore, in this paper, we will show two algorithms to extract feature sets, both of them are similar to previous algorithms but some improvements are achieved to ensure the extracted feature sets are invariant to rotation, scale and translation.

3.2 Features Based on the Nearest Pore Pairs

Nearest pore pairs can provide size, distance and angle information of the pairs. Fig. 4 shows an example of the nearest pore pairs. Note that this nearest relationship is not symmetric, that is, the fact that pore q to pore p is closest does not mean that p is the closest to q. In Fig. 4, the nearest pore to p_i is p_j, but the nearest pore to p_j is p_k. Suppose the current pore is p_i and its nearest neighbor pore is p_j, we can normalize the size s_i and distance d_i into $(0, 1]$ according to their maximum values. After normalization, s_i and d_i can be treated as invariant features, but the angle θ_i is not. For example, if the original image is rotated by γ, then the angle between p_i and p_j becomes $\theta_i + \gamma$. Fig. 5(a) and (b) show the histograms before and after $\gamma = \frac{\pi}{4}$ degree rotation respectively. The histogram after rotation γ keeps the shape of histogram, but the bins shift to the right by γ

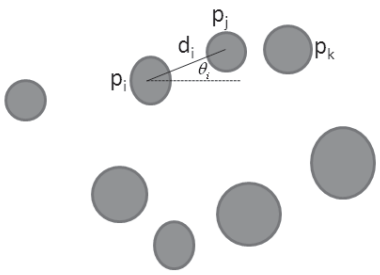
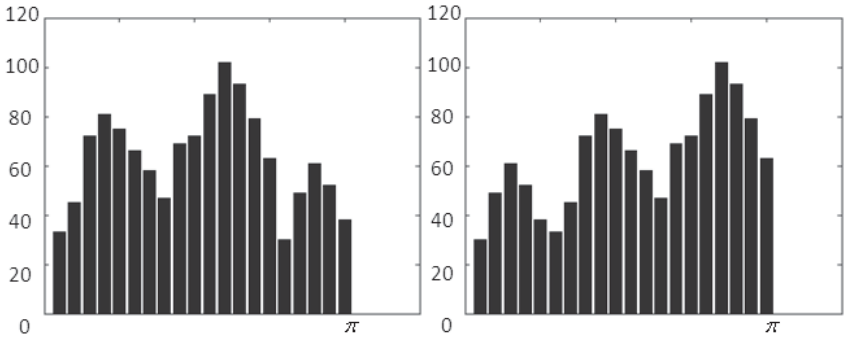


Fig. 4. Nearest pore pairs

and take modulo by π. To solve this problem, therefore, we move the mode (the most frequent angle) to the left most position, that is, at 0°. Such a modified histogram is invariant to rotate. Finally, the steps to extract features based on the nearest pore pairs can be expressed as following:



(a) Before rotation (b) After rotation

Fig. 5. The angle histogram before and after rotation

Step1. For each p_i, quantize the value s_i of area into one of three size labels 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 $(0,\ 1]$. These rules were determined from experiments because the accuracy of final results would be worse with other possible rules.

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

Step3. Calculate the angle θ_i and the Euclidean distance d_i between $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$ such as those shown in Fig. 4.

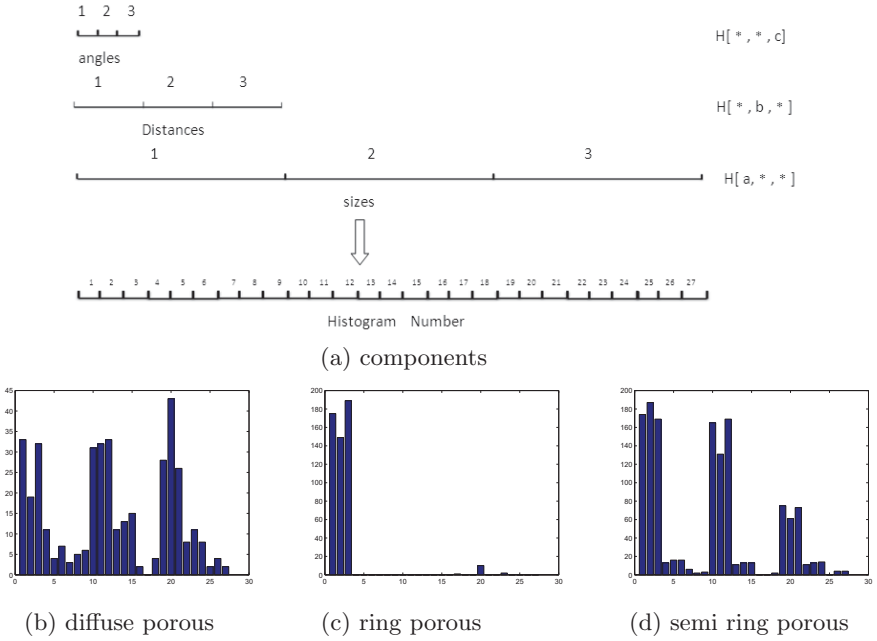
Step4.of For each θ_i, make the histogram to the left most by the following formula:

$$\theta_i' = (\theta_i - \theta_{mode})\%\pi \tag{1}$$

here, θ_{mode} means the mode angle (the most frequent angle) occurred in all pore pairs, % means the modulo operation.

Step5. 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 $\{1 : \theta_i' \in [0, \frac{\pi}{3}); 2 : \theta_i' \in [\frac{\pi}{3}, \frac{2\pi}{3}); 3 : \theta_i' \in [\frac{2\pi}{3}, \pi);\}$. Like the rules used in Step 1, all of these rules were decided by our experiments too.

Step6. Construct a histogram H over $27(= 3 \times 3 \times 3)$ 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 roughly same size label. Here, $H[a, b, c]$ corresponds to the frequency of pairs (p_i, p_j) producing ath size, bth distance and cth angle in their quantized codes.



(a) components

(b) diffuse porous (c) ring porous (d) semi ring porous

Fig. 6. Histogram of features based on the nearest pore pairs.The X-axis is the number of components and the y-axis is the number of pore pairs in each component.

Fig. 6(a) shows the components of histogram, Fig. 6(b),(c),(d) are three different histograms for three different porosities. From Fig. 6(b),(c),(d), we can observe that different porosity types have different histogram shapes. For diffuse porosity, almost every bin in histogram has a non-zero value. This means that we can find different sizes of pore pairs with different angles and distances. For ring porosity, most of pores located in the small size and big size position of the histogram, middle size pores are hardly found. In semi ring porosity, we can

find all the three size pore pairs, most of them have a short distance, the angle between them may be small, middle or large.

3.3 Features Based on Diameter Change

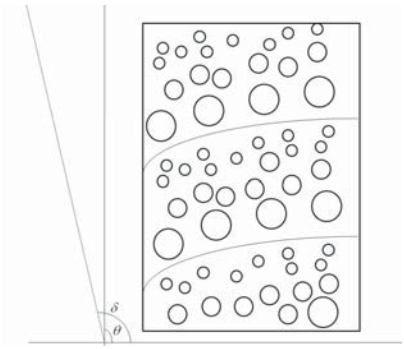


Fig. 7. The proper angle to inspect the global diameter change

It is noted that the difficulty in recognizing the growth rings leads to the difficulty in detecting the local diameter change of pores between two growth rings; however, it is not difficult to consider the global diameter change of all the pores in the whole image along the vertical direction. Before inspecting the global diameter change, the proper angle of original image should be calculated firstly. Fig. 7 shows an example to find the proper angle to inspect the global diameter change. In Fig. 7, $\theta = \frac{\pi}{2}$ is the original angle of input image; however, this angle is not fit for inspecting the diameter change because the ideal direction is perpendicular to the tangent of the centers of growth rings. For Fig. 7, the best choice is to detect the diameter change along the dotted line, that has angle δ with the horizontal direction. In order to find the proper δ, the frequency of large pores is checked because along the angle δ, the large pores and small pores will occur alternately, especially in ring and semi ring porous wood images. This phenomenon can be measured by the entropy [21] of large pores: the value of entropy will be the least with the angle δ because the regular appearance of large pores. As a result, the following procedure is applied to calculate the angle δ.

Step1.Set $\delta = \frac{\pi}{2}$.

Step2.Divide the image into 30 equally-sized divisions $D_j,\ j \in [1, 2, ..., 30]$, along the δ direction.

Step3.Normalize all the areas of pores into [0 1] by the maximum value of pore areas.

Step4. 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 $(0,\ 1]$.

Step5.Calculate the number n_j of large pores (s_i=3) and the frequency $p(h_j)$ in each D_j by

$$p(h_j) = \frac{n_j}{N} \tag{2}$$

Here, N is the total number of pores in the image.

Step6.Calculate the entropy of large pores $H(h_\delta)$ by the following formula:

$$H(h_\delta) = -\sum p(h_j) \log p(h_j) \tag{3}$$

Step7.If $\delta \leq \pi$ set $\delta = \delta + \frac{\pi}{18}$ and go to step 2, else go to Step 8.

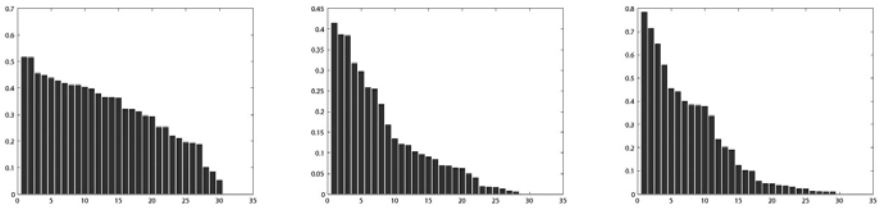
Step8. Choose the angle δ^* to make the entropy minimum as:

$$\delta^* = \arg\min_{\delta} H(h_\delta) \tag{4}$$

After getting the proper angle δ^*, the diameter change along the angle will be inspected. The same steps as those in prior paper [20] are used to extract the features based on diameter change, only the new position coordinate $(x_i' \;\; y_i')$ for each p_i should be calculated as:

$$(x_i' \;\; y_i')^T = \begin{pmatrix} cos(\delta^*) & -sin(\delta^*) \\ sin(\delta^*) & cos(\delta^*) \end{pmatrix} (x_i \;\; y_i)^T \tag{5}$$

Here, $(x_i \;\; y_i)$ is the original position of each pore. As a result, we have a feature



(a) diffuse porous (b) ring porous (c) semi ring porous

Fig. 8. Diameter change of pores (1 to 30 according to the early to late zones). The X-axis is the 30 divisions and the y-axis is the average area of pores in each division.

set of 30 values. The results are shown in Fig. 8 for one example of each class. We can observe that the average areas (thus the diameter) 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.

4 Porosity Recognition

4.1 The Decision Tree with Features Based on the Nearest Pore Pairs

In order to analyze the discriminative information of those feature sets, we use C4.5 algorithm [22]. C4.5 algorithm generates a set of classification rules as a decision tree.

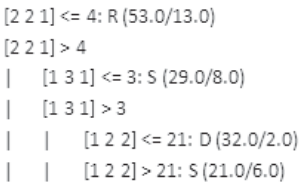


Fig. 9. The decision tree with features based on the nearest pore pairs. The node such as $'[i,j,k] \leq v : k(a/b)'$ means a samples are classified as porosity k, $k \in \{'D','R','S'\}$, if the value of the $[i,j,k]$th feature is less than or equal to v, while b samples are misclassified. $[i,j,k]$ corresponds to the frequency of nearest pore pairs with ith size, jth distance and kth angle.

Fig. 9 is the complete 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 satisfy the rule [2,2,1]≤4. It means roughly that the middle size pore pairs with a middle distance and small angle should be less than 4. This is consistent to the observation that most pores have large or small sizes and few middle size pores can be found in ring porous images. Some semi-ring porous samples satisfy the rule [2,2,1]>4,[1,3,1]≤3, others satisfy the rule [2,2,1]>4,[1,3,1]>3 and [1,2,2]>21. The former rule tells us that there should be a certain number of middle size pore pairs with middle distance and small angle ([2,2,1]), and the number of small size pore pairs with large distance and small angle ([1,3,1]) should be less than 3; the latter rule tells us that if there are a certain number of pore pairs with small size large distance and small angle ([1,3,1]), then the number of small size pores with middle distance and middle angle ([1,3,1]) should be more than 21. For diffuse porous images, most of them satisfy the condition [2,2,1]>4,[1,3,1]>3 and [1,2,2]≤21. It is similar to one condition of semi ring porosity, but we can find small size pore pairs with middle distance and middle angle ([1,2,2]) should not be very much.

In general, the rules given by the decision tree show the different components in the histogram of pore pairs that are used as key points in recognition. However, it is a little difficult in explaining these rules by domain knowledge about porosity, because the definitions of porosity are based on diameter change

instead of the nearest pore pairs. In the next part, we will check the decision tree generated by diameter change and explain it.

4.2 The Decision Tree with Features Based on Diameter Change

In this part, C4.5 is used again to know the detailed relationship between the features based on diameter change the porosity.

```
23 <= 0.051: R(50.0/7.0)
23 > 0.051
|   26 <= 0.1557: S (43.0/8.0)
|   26 > 0.1557: D (42.0/3.0)
```

Fig. 10. Decision tree with features based on diameter change. 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 nth feature is less than or equal to v, while b samples are misclassified.

The complete decision tree is shown in Fig. 10. We can see some simple rules from the tree. For example, if the average pore area is less than 0.051 in the 23th band (of 30 bands), then the sample will be classified to 'R' (ring porosity). The 23th band can be treated as the position in the late wood zone in the total 30 bands. If the pores in this position have very small sizes, it means there is a rapid diameter change from early wood zone, so the porosity will be classified as ring. Otherwise, if the 23th band is not small enough, then the 26th band should be inspected. The 26th band can be treated as a later wood zone than 23th band. If pores in the 26th band are relatively large (larger than 0.1557), the porosity will be classified as diffuse because if most of pores in late wood zone have relatively large sizes, the diameter change will not be obvious in the whole scope; else it means that a gradual size change of pores exists, the porosity will be classified as semi ring.

4.3 Other Classifiers

The support vector machine (SVM) is the most popular in pattern recognition recently. It ensures the generality (the capability of correct recognition for unseen samples) by taking the largest margin between samples of different classes[23]. The Naive Bayes classifier works well when the number of features is relative large compared to the size of training sample set [24]. By these two classifiers, explainable or reasonable results that can be combined with domain knowledge such as decision trees can not be given directly, so we only show the confusion matrix of each classifier.

5 Results and Discussion

5.1 Confusion Matrix of C4.5 with Features Based on the Nearest Pore Pairs

We obtained an estimate of correct recognition by 10-fold cross-validation [25]. The confusion matrix of 10-fold cross-validation comes as below when the features based on the nearest pore pairs are used:

D	R	S	<-classified as
31	5	9	D=diffuse
3	37	5	R=ring
10	6	29	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 97 samples classified correctly, bringing the accuracy of 71.9%. Especially low misclassification occurred between diffuse porosity and ring porosity . For the other combinations, the accuracy is not so high. The accuracy between semi ring and diffuse seems the lowest in the matrix.

5.2 Confusion Matrix of C4.5 with Features Based on Diameter Change

The confusion matrix when features based on diameter change are used is given as below:

D	R	S	<-classified as
39	1	5	D=diffuse
0	41	4	R=ring
7	7	31	S=semi

In total, 111 samples were classified correctly at recognition rate of 82.2%. All of the classification between diffuse, ring and semi-ring porosity is improved from the previous result. However, like the results of the nearest pore pairs, the recognition performance between diffuse and ring porosity is the best, while the recognition rate between semi ring and diffuse is the lowest.

5.3 Confusion Matrices of SVM

After comparing the accuracy of different kernels, the the polynomial kernel that can be expressed by $K(X_i, X_j) = (\frac{1}{k}{X_i}^T X_j)^3$ is selected for processing the feature set based on nearest pore pairs; while linear kernel that can be expressed by a linear combination of ${X_i}^T X_j$ is selected for processing the feature set based on diameter change.

The confusion matrix when features based on nearest pore pairs are used is given as below:

D	R	S	<-classified as
33	3	9	D=diffuse
2	33	10	R=ring
5	10	30	S=semi

The confusion matrix when features based on diameter change are used is given as below:

D	R	S	<-classified as
41	1	3	D=diffuse
0	39	6	R=ring
9	5	31	S=semi

From the two confusion matrices, we can get the accuracy about 82.2% for diameter change features and 71.1% for nearest pore pairs features. This proves that both of the feature sets are effective for SVM, but proper kernels should be selected in advance. Like the results given by C4.5 algorithm, both of them have high accuracy in classifying diffuse and ring porosity; however, worse accuracy occurs when semi ring porosity is considered. In general, the results tell us that the feature set based on diameter have a better accuracy.

5.4 Confusion Matrices of Naive Bayes

From the stand point of statistical learning, we choose the Naive Bayes classifier as an other choice.

The confusion matrix when features based on nearest pore pairs are used is given as below:

D	R	S	<-classified as
35	4	6	D=diffuse
2	41	2	R=ring
14	12	19	S=semi

The matrix tells us that 95 samples are classified and the accuracy is about 70.4%.

The confusion matrix when features based on diameter change are used is given as below:

D	R	S	<-classified as
41	1	3	D=diffuse
0	42	3	R=ring
8	7	30	S=semi

In this confusing matrix, we find that totally 113 samples can be classified successfully. It means that the accuracy is about 83.7%. It is the highest accuracy in all classifiers and feature sets used in this paper.

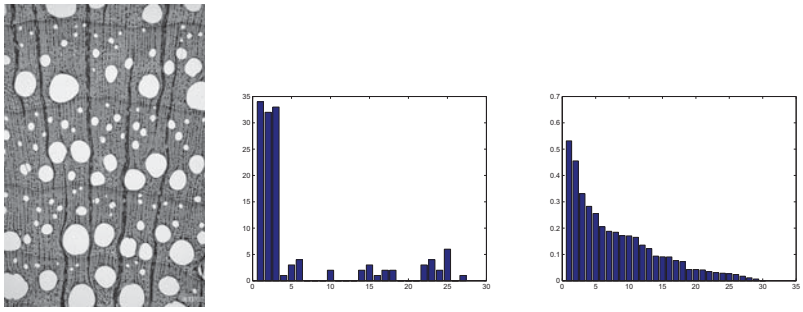
From above results, we can conclude that the two kinds of features is effective not only for C4.5 algorithm, but also for other classifiers such as SVM and Naive Bayes. If the Naive Bayes classifier is used to process the feature set based on diameter change, the highest accuracy can be got; however, neither the Naive Bayes nor SVM can give us rules that can be explained by domain knowledge. So they are only used to prove that our feature sets is effective to different classifiers. This paper will focus on C4.5 algorithm and explanation of decision trees.

5.5 Discussion

Even the best accuracy 83.7% 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.

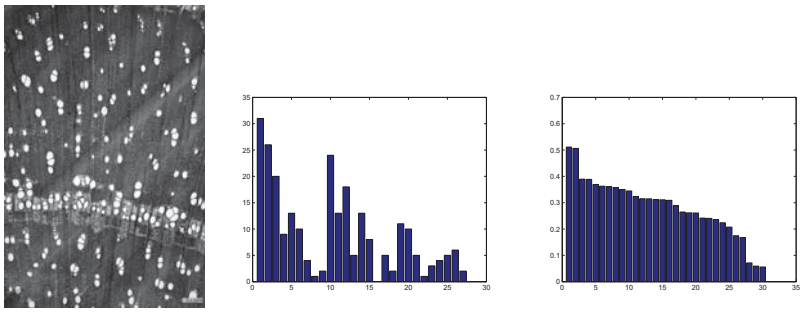


(a) semi-ring image (b) nearest pore pairs (c) diameter change

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

Fig. 11 demonstrates one of these cases. From the microscopic image, 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. From Fig. 11(b), although there are some pores with middle sizes, the number of them is significantly less, most of pores have small and large sizes. So the value of the 13th component in the histogram ([2 2 1] in the decision tree) is zero, it is classified as ring porosity according to the decision tree of the nearest pore pairs. The features based on diameter change also detected a rapid change of pore size from large to small (Fig. 11(c)). As a result, the rule found the fact that the pore size in late wood (the 23th component in the graph) is small enough ($\leq$0.051), thus classified it as ring porosity.

The diffuse and semi ring porosity pair is also confusing. Fig. 12 gives an example of this case. From the microscopic image (Fig. 12(a)), we can find that

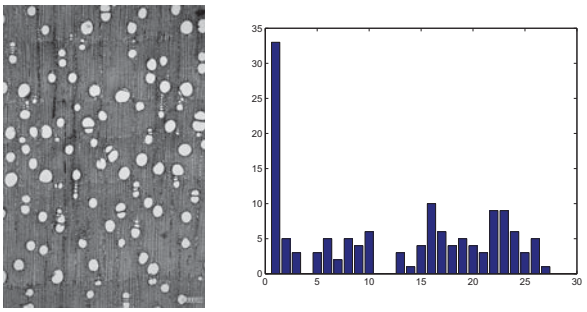


(a) semi-ring image (b) nearest pore pairs (c) diameter change

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

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 features based on diameter change (Fig. 12(c)) tells us that the change of pore size is very little, so the rule classified the image as diffuse porosity. The histogram of features based on the nearest pore pairs (Fig. 12(b)) also fails because almost every bin has a non-zero value.

In case of features based on the nearest pore pairs, 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. 13 shows a diffuse porous sample and its histogram. Most of the pores are small (concentrating on the right of the histogram) and the number of middle size pores ([2,2,1]) is very small (less than 4). It caused misclassification from diffuse to ring porous.



(a) diffuse image (b) nearest pore pairs

Fig. 13. A case in which diffuse porous sample is misclassified as ring porosity

In general, 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. After analyzing the mistaken samples, we believe the main reasons of misclassification in semi ring porous samples are following:

(1)The number of middle size pores is far less than the number of large and small sizes, so they are "omitted" by the decision trees.

(2)Some semi ring porous samples have almost the same pore size but different frequency, our decision trees are in effective to these cases.

For the first reason, it can be partly explained by the conflict between obscure definitions of porosity and the precise measure of decision trees in our algorithm. As the definition of semi ring porosity, the large, middle and small sizes pores can be found in this porous images, and there is a gradual diameter change in these pores; however, we can not find the detailed information about how many pores of each size are or how extent the size change in the definition. A precise value in decision trees may be reasonable for most cases, but sometimes it is invalid. For the second reason, it can be explained by the lost information about position of each pores in our algorithm. Without these information, the frequency of pores in different position is hard to be considered. It is a problem to be solved in the future research.

6 Conclusion

In this paper, two kinds of feature sets that are insensitive to direction are extracted to recognize three kinds of porosity in wood microscopic images. Then those features are used with C4.5 algorithm to generate decision trees. The efficiency of the two feature sets can also be proved by other classifiers. After trying to explain the rules given by decision trees according to domain knowledge, we find that the rules based on diameter change are easy to be read and understood. The estimator of 10-fold cross-validation is used to verify the classification performance of the decision trees. As a result, we find 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 are investigated, it is still necessary to do more work on understanding the decision trees and analyzing the reason of over-fitting. On the other hand, both of the feature sets lack the information about the positions of every pore that is helpful to recognize the semi ring porosity sometimes, so combining position information to these two feature sets may be an other method to improve the accuracy of the three kinds of porosity. Moreover, other features [26] will be investigated and combined with those presented in this paper to improve the accuracy in the future.

Acknowledgments

The core program of C4.5 is realized by Weka. 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. We are grateful to the two anonymous reviewers who gave us much good advice.

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, China (2008)
4. Wheeler, EA., Baas, P., Gasson, PE.: IAWA list of microscopic features for hardwood identification. IAWA Bull. (N.S.) 10, 219-332 (1989)
5. Kudo,M., Sklansky, J.:Comparision of algorithms that select features for pattern recognition. Pattern Recognition 33(1),25-41 (2000)
6. Xu,YM.: Wood Science. China Forestry Publishing House (in Chinese). Beijing, China (2006)
7. Bond, B., Hamner, P.: Wood Identification for Hardwood and Softwood Species Native to Tennessee. Agricultural Extension Service. Knoxville, TN, USA (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. Tellaeche, 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. Pan, S., Kudo, M.: Recognition of porosity in wood microscopic anatomical images. In: Perner, P. (Ed.), ICDM 2011, LNAI, vol. 6870, pp. 147-160, Springer Verlag, Heidelberg (2011)
21. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423, 623-656 (1948)
22. Quinlan, J.R.: Programs for Machine Learning. Morgan Kaufmann Publishers. San Mateo, CA, USA (1993)
23. Kim, H.C., et al.: Pattern Classification Using Support Vector Machine Ensemble. Paper presented at the Proceedings of the 16th International Conference on Pattern Recognition. Quebec City, Canada (2002)
24. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian Network Classifiers. Machine Learning 29, 131-163 (1997)
25. Geisser, S.: Predictive Inference. Chapman and Hall. New York, USA (1993)
26. Serra, J.: Image Analysis and mathematical Morphology. Academic Press. Academic Press. London (1982)

Vitae

Shen Pan, received his bachelor and masters degree of Eng. from Hefei University of Technology and then worked as an assistant in the same university. In 2009, he began to study PhD course in Hokkaido University and will receive Dr. Eng. degree in March, 2012. His current research interests are application systems of image processing, pattern recognition, and data mining.

Prof. Mineichi Kudo, received his Dr. Eng. degree in Information Engineering from the Hokkaido University in 1988.At Hokkaido University, he was an instructor (1988-1994), an associate professor (1994-2001) and is a professor (2001-). In 2001 he received with professor Jack Sklansky "the twenty-seventh annual pattern recognition society award for the most original manuscript from all 2000 Pattern Recognition issues." His current research interests include design of pattern recognition systems, image processing, data mining and computational learning theory. He is a member of the Pattern Recognition Society and the IEEE.