

Evaluation of a model using local features and a codebook for wood identification

SW Hwang^{1*}, K Kobayashi², and J Sugiyama^{1,3}

¹ Research Institute Sustainable Humanosphere, Kyoto University, Kyoto, 611-0011, Japan

² Graduate School of Agricultural and Life Sciences, Department of Biomaterial Science, The University of Tokyo, Tokyo, 133-8657, Japan

³ College of Materials Science and Engineering, Nanjing Forestry University, Nanjing 210037, China

*email: sungwook_hwang@rish.kyoto-u.ac.jp

Abstract. We designed a model for wood identification based on scale-invariant feature transform (SIFT) descriptors and a codebook. A dataset consisting of cross-sectional optical micrographs of the Lauraceae family including 39 species was used for identification. The bag-of-features (BOF) model was superior to the model combined SIFT descriptors with a classifier. Among the four classifiers applied to both models, the support vector machine (SVM) achieved the best identification performance with 99.4% accuracy. From the feature importance calculated by the random forests and the inverse document frequency (IDF) score, it was also confirmed that cell corner-based features are more informative for the identification of Lauraceae. In particular, cell corners in vessels are not only important for species identification, but also reveal that they are species-specific features. The computer vision-based model was suitable for Lauraceae identification and enabled the quantification of anatomical structures that are not possible with conventional visual inspection for wood identification.

1. Introduction

Studies related to wood such as botany, cultural property science, and archeology, as well as wood science start with the accurate identification of species. This is because wood not only presents a variety of characteristics by species, but also contains species-specific information. The conventional wood identification process is performed by visually inspecting the morphological structure of the three orthogonal planes. This standard visual inspection is the most reliable method to date.

The vast variety of trees distributed throughout the world, or a country, and even in a city have different anatomical characteristics by species. The International Association of Wood Anatomists' (IAWA) list of microscopic features for hardwood identification is an international standard for wood identification that consists of 221 feature codes [1]. There are 163 feature codes to identify Japanese hardwoods alone [2]. Therefore, long-term training and experience are required to identify wood by visual inspection. Nevertheless, there are certain limitations to this method. Wood identification by visual inspection is generally classifiable to the genus level, and furthermore, species that have the identical anatomical characteristics from the feature code are not classifiable. There are many species with the same anatomical characteristics among trees classified as different species in plant taxa. Such difficult-to-identify species often cause social problems.

To overcome the limitations of the conventional method, many researchers have conducted alternative methodological studies, such as deoxyribonucleic acid (DNA) analysis [3-5] and



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

chemometrics [6-8], and they reported successful results. However, these approaches require a great deal of effort to identify a species and are not suitable for processing large databases with various species.

We consider the computer vision-based image classification technique to be a promising alternative to overcoming the limitations of conventional methods. Various machine learning algorithms have been developed for classifying images from acquiring information on images. These techniques have also been applied to wood identification and include texture features, local or global features, and neural network-based models. The gray-level co-occurrence matrix (GLCM) was a popular method for extracting texture features from wood images [9-11]. Tou et al. [9] performed wood recognition with a combination of GLCM texture and the Gabor filter. Kobayashi et al. [10, 11] reported successful results in wood recognition using textures extracted from low-resolution computed tomography data and stereographs. This research suggested the applicability of the non-destructive species classification method for wood cultural properties via computer vision technology. Wood classification by local feature extraction and a codebook-based method has also been performed [12, 13]. However, in many studies, it is quite difficult to understand the model's achievements in relation to the morphological structure of the wood.

In this study, the model was designed using local features and a codebook for wood identification. We evaluated the performance of the model using different identification schemes based on the identification accuracy. The feature importance was calculated in order to determine which anatomical features are important for species identification.

2. Dataset

The Lauraceae family was used for wood identification. Lauraceae is known to be a difficult family to identify owing to their very large and complex species composition and morphological similarity. The wood blocks of 11 genera including 39 species were received from the RISH Xylarium, Kyoto University. To make microscopy slides (Figure 1), all the blocks were cut to 15 μm thickness by a sliding microtome. The wood sections were then stained with safranin and embedded on the slide. These preparation processes are the same as those of the conventional method by visual inspection. However, we prepared only the cross-sectional slides and not the three orthogonal sections. The cross section images were acquired with OlympusTM 2 \times (0.08 NA) PlanApo objective lens, using a BX51 optical microscope equipped with a DP73 charge-coupled device (CCD) camera. The original image was in red, green, and blue (RGB) color and had a size of 4800 \times 3600 pixels with a pixel resolution of 0.74 μm .

To construct an image dataset, the original images were converted to 8-bit grayscale and cropped to a size of 3600 \times 3600 pixels. The images were then resized to pixel resolutions of 1.47, 2.94, 5.88, and 11.76 μm to determine the optimal resolution for Lauraceae identification. Finally, the dataset was constructed with 1658 cross-sectional images.

3. Identification model

Figure 2 presents the schemes of the identification model. The model was implemented within the

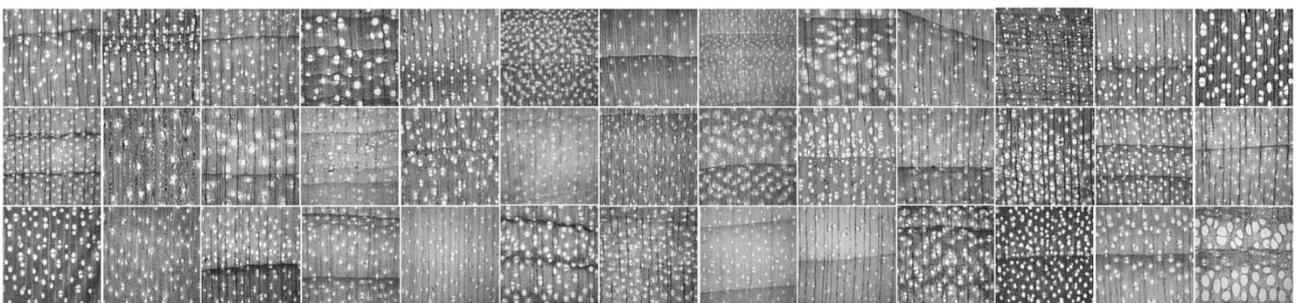


Figure 1. Cross-sectional images of 39 species in the Lauraceae dataset.

bag-of-features (BOF) framework based on local features and had three sub-models.

3.1. Feature extraction

Local features were extracted from the images using the scale-invariant feature transform (SIFT) algorithm [14, 15]. SIFT features are robust to changes in scale, rotation, and illumination, and have demonstrated its superior performance in a variety of image classification problems [16-18]. All images in the dataset were converted to SIFT descriptors after the keypoints were extracted by the algorithm. We implemented the SIFT algorithm by adopting the parameters of the number of layers in each octave of two. The Gaussian filter was applied to the image of each layer of 1.6, the contrast threshold of 0.06, and the edge threshold of 10. When the feature extraction was completed, each image was represented by a SIFT descriptor, which was a 128-dimensional vector.

The algorithm extracted local features such as blobs, corners, and edges from the difference in Gaussian images. These SIFT features effectively catch wood fibers, axial and ray parenchyma cells, and vessels, which are the main anatomical features we observe in cross-sectional optical micrographs of wood in species identification [13]. The blobs could be applied to all cells with lumen and the corners were present on all cell walls between adjacent tissues. In addition, the edges could detect vessels and rays. Thus, the local features that the SIFT algorithm detected from the wood were not very different from the anatomical features observed by wood anatomists.

3.2. The model combined SIFT descriptors with a classifier

The first sub-model is a simple model that combined SIFT descriptors with a classifier (Model 1 in Figure 2). The images with pixel resolutions of 0.74 μm , 1.47 μm , 2.94 μm , 5.88 μm , and 11.76 μm were used to determine optimal pixel resolution for Lauraceae identification. Based on the identification accuracy, the optimal pixel resolution was determined and used in subsequent models.

Three different classifiers, k-nearest neighbor (k-NN), linear discriminant analysis (LDA), and support vector machine (SVM) were used for data learning and species identification.

3.3. Bag-of-features model

The second sub-model (Model 2) is a BOF model [19, 20] that converted SIFT keypoints into codewords via clustering. In this model, the dataset was divided into training and test sets at a ratio of 4 to 1, which were used for learning and identification, respectively. The images in the training set were not represented by the SIFT descriptors, and all the extracted keypoints were used to generate codewords by mini-batch k -means clustering [21] implemented by the k -means++ algorithm [22] with a processing batch size of 100. Various numbers of clusters (k -size) ranging from 100 to 1000 were considered to determine the optimal k -size. Once the codebook was created, images could be quantified by the codewords to which all their keypoints belong. In other words, each image was represented by a feature histogram with k -size bins, and each bin represented the number of keypoints contained in the corresponding codeword in the image. This process is called vector quantization and is the core of the BOF model.

For data learning, k-NN, LDA, and SVM classifiers were used as well as Model 1. The images in the test set were also represented as feature histograms based on the codebook generated in the training phase. They were then identified by the learned classifiers.

3.4. Random forest classifier and feature importance

The last model is a random forests classifier, which was applied to the BOF model (Model 3). The random forest (RF) was used for feature identification as well as species identification. The RF is an ensemble method for classification [23]. It is a powerful classifier created by combining a number of weak classifiers (the decision trees in this model). This classifier avoids overfitting and has a strong generalization characteristic. This is achieved by random sampling of observations and variables, and by combining multiple basic learners during the process of generating the decision trees.

The identical training and test sets as the previous model were applied to the RF classifier. The training process in RF began with bootstrap aggregating (bagging) [24], a sampling technique for creating tree learners. The bagging created trees repeatedly by random sampling with replacement in

the same size as the training set. In this process, because certain images are duplicated in each tree, approximately one-third of the images in the training were not selected for tree learning. These images are called out-of-bag (OOB) samples and they were used for the validation of each tree. We determined the optimal number of trees for Lauraceae identification by tracking the OOB errors, the mean of the validation errors of all tree learners, from RFs with varying numbers of trees.

RF and bagging differ in the way that they select features (codewords) to separate classes in the tree. Unlike the bagging, the RF does not use all features for learning each tree and uses a subset of random features. To determine the optimal number of features for a subset, we calculated the OOB error using three different feature numbers: k features (k -size), root square of k features, and binary log of k features. The RF provided us with information on the importance of features (or variables) in classification, which is what we really wanted to obtain from the RF.

In addition, we calculated the inverse document frequency (IDF) score, which represents the rarity of the features (or codewords), and compared it with the feature importance of RF. The IDF was calculated by the following equation:

$$IDF = \log\left(\frac{N + 1}{df_j + 1}\right) \quad (1)$$

where df_j is the number of images containing feature j , and N is the number of images. Based on the feature importance of RF and the IDF score, we investigated which codewords were used as important criteria for Lauraceae identification. We then visualized the SIFT keypoints belonging to the codeword in order to grasp their corresponding anatomical features.

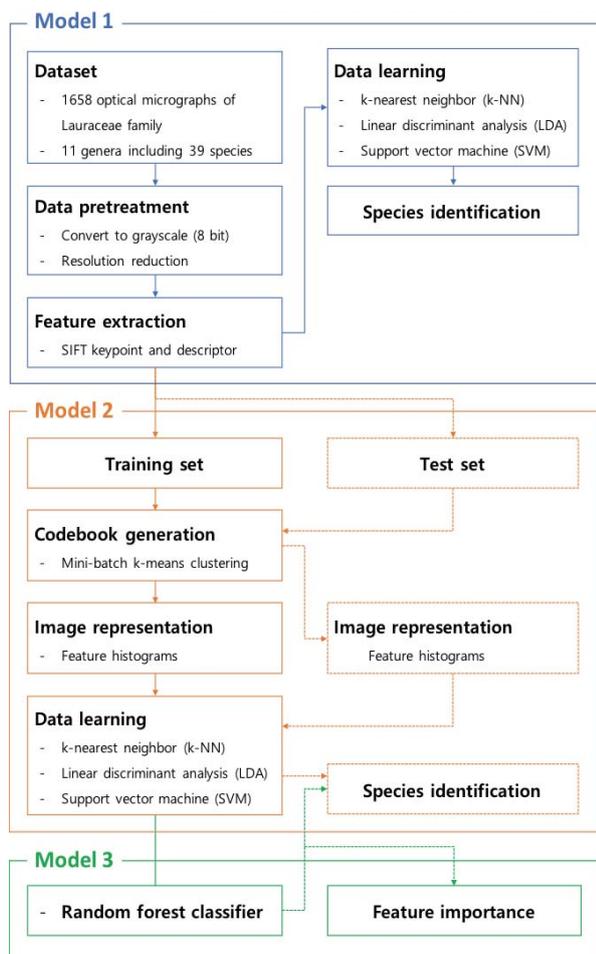


Figure 2. Schemes of the identification model.

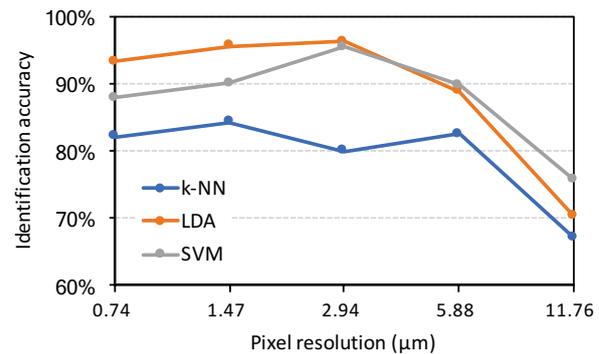


Figure 3. Identification accuracies of the three different classifiers with reducing pixel resolutions.

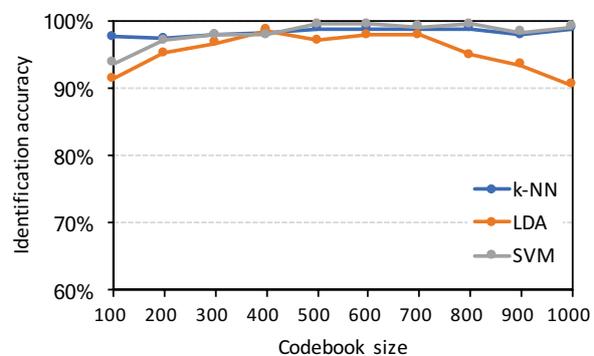


Figure 4. Identification accuracies of the three different classifiers with various codebook sizes.

4. Results and discussion

This section presents the results of the identification performances and the feature importance of the models. We have compared the performance of sub-models and classifiers and discussed which anatomical features are important for the identification of Lauraceae.

4.1. Performance of the model combined SIFT descriptors with a classifier

Figure 3 presents the identification accuracy of Model 1 that combined simple SIFT descriptors with the three different classifiers at various pixel resolutions of images. The LDA and SVM showed a similar pattern in identification accuracy. Their accuracies increased slightly with resolution, decreasing from the original resolution of $0.74 \mu\text{m}/\text{pixel}$, and achieved the best accuracy of 96.3% and 95.4%, respectively, at a pixel resolution of $2.94 \mu\text{m}$, the pixel resolution equivalent to a quarter of the original size. In the images with pixel resolutions of $0.74 \mu\text{m}$ and $1.47 \mu\text{m}$, which are higher than $2.95 \mu\text{m}$, the identification accuracy was somewhat lower, even though more keypoints were detected. This means that many tissues or artifacts that are not informative for identification are detected as features in the high-resolution cross-sectional optical micrographs. In this model, k-NN was inferior to other classifiers.

From the identification results, an optimal image resolution for Lauraceae identification was determined to be $2.94 \mu\text{m}$, corresponding to an image size of 900×900 pixels, and a dataset based on the pixel resolution was used in the subsequent BOF model. The results of the model demonstrate that the SIFT algorithm extracts informative local features for Lauraceae identification from the cross-sectional optical micrographs.

4.2. Performance of the BOF model

The identification performance of the BOF model was higher than that of the previous model combined SIFT descriptors and classifiers (Figure 4). The accuracy of all classifiers improved and larger codebooks yielded better accuracies. The highest accuracy was achieved at a codebook size of 500 and was maintained at similar levels in larger sizes. However, the LDA achieved the best accuracy of 97.9% at codebook sizes of 600 and 700, and then the performance declined at larger sizes. The SVM showed a remarkable identification performance of 99.4% at a codebook size 500. In the BOF framework, the identification accuracy of k-NN was 98.8%, which is much higher than the accuracy of 84.3% in the previous model.

From these results, it is confirmed that the generation of a codebook and the data represented by a feature histogram in the BOF model are more effective in identifying Lauraceae than the data represented by a SIFT descriptor in the previous model.

In the case of the SIFT descriptor, all the features extracted from the image were represented by a 128-dimensional vector so that the descriptors reflected the characteristics of numerically dominant

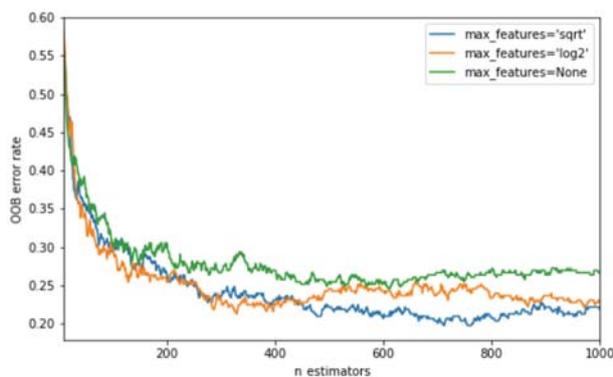


Figure 5. Variation in the out-of-bag errors of three different numbers of features (`max_features`) by increasing the number of trees (`n_estimators`).

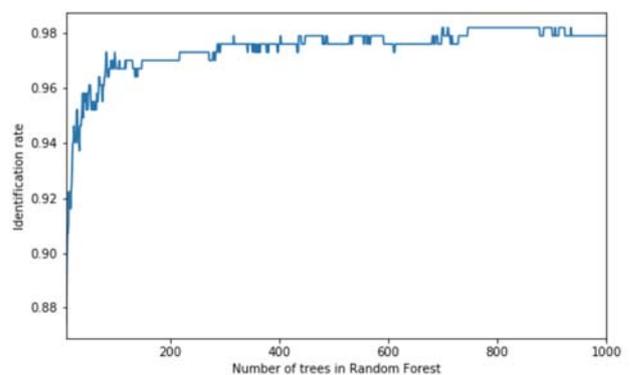


Figure 6. Variation in the identification accuracy with respect to the number of trees.

anatomical features, such as cell lumina and cell corners. In other words, the SIFT descriptor does not fully represent the characteristics of numerically inferior features, including axial parenchyma cells, vessels, and rays. On the other hand, in the feature histogram of the BOF, each codeword represented various anatomical features by the clustering process. Thus, such inferior features were quantified with their own bins in the histogram, so that the feature histogram of the BOF was more informative than the simple SIFT descriptor. It is considered that the difference of the image representation method produces the difference of performance of the models.

4.3. The BOF model using random forest classifier

Based on the codebook size 500 in the BOF model, the OOB error rates of a subset (`max_features`) of three different numbers of features were calculated for various number of trees (`n_estimators`) (Figure 5). The OOB error rates for all schemes decreased with increasing `n_estimator`. ‘`max_features = None`,’ which used all 500 codewords, was not suitable for feature sampling because it produced the highest OOB error rate regardless of the number of `n_estimators`. The binary log of features (`max_features = ‘log2’`) showed a relatively low error rate at the initial stage as compared to the other schemes, but the error rate increased after the `n_estimator` of 400. The optimal feature sampling scheme for tree learners was the square root of features (`max_features = ‘sqrt’`). The ‘`sqrt`’ scheme achieved a minimum OOB error rate in the `n_estimators` ranging from 700 to 750. Therefore, the identification by RF was calculated based on the number of square root of features.

Figure 6 presents the variation in identification accuracy for the number of trees (`n_estimator`). As confirmed in the OOB error rate, more trees produced better performance, with the best accuracy of 98.2% achieved at the `n_estimator` of 750.

4.4. Performance comparison of identification models and classifiers

All models and classifiers considered in this study achieved excellent identification performance (Table 1). The k-NN, which has a relatively low accuracy in the SIFT descriptor-classifier combined model, showed better identification performance than the LDA in the BOF model, indicating that k-NN is more suitable for larger data than LDA. The SVM has proven to be the best classification model for Lauraceae identification by achieving the highest level of accuracy in all sub-models. The performance of the BOF model was surprisingly good. The identification accuracy of the four classifiers that operated within the BOF framework achieved excellent results with a maximum of 99.4% by SVM and the lowest of 97.9% by LDA. From these results, it was confirmed that based on local features extracted by SIFT, the BOF model is a very promising and powerful method for Lauraceae identification.

Table 1. Comparison of the performances of all identification schemes.

Model	Identification		Parameters		
	accuracy (%)	Pixel resolution (μm)	Codebook size	Number of features	Number of trees
SIFT-kNN	84.3	1.47	-	-	-
SIFT-LDA	96.3	2.94	-	-	-
SIFT-SVM	95.4	2.94	-	-	-
BOF-kNN	98.8	2.94	500	-	-
BOF-LDA	97.9	2.94	600	-	-
BOF-SVM	99.4	2.94	500	-	-
BOF-RF	98.2	2.94	500	‘sqrt’	750

4.5. Feature importance calculated by the random forest classifier and IDF score

All codewords have their own feature importance value, and certain codewords have a significantly higher value. The codewords with the highest twenty feature importance values calculated by RF and their corresponding anatomical features are presented in Table 2, and Figure 7 shows the morphological appearances and structures of the anatomical features in the cross section.

Table 2. Informative codewords with the highest twenty feature importance values and their corresponding anatomical features.

Rank	Codeword	Feature importance	Anatomical features
1	82	8.31E-03	CC-V&R
2	301	7.57E-03	CC-WF&AP
3	160	7.30E-03	CC-V
4	489	6.68E-03	CC-general
5	377	6.36E-03	CC-general
6	314	6.25E-03	CC-V
7	289	6.08E-03	CC-V
8	493	5.96E-03	CL-general
9	395	5.74E-03	CL-general
10	106	5.72E-03	CC-V
11	25	5.47E-03	CC-R
12	380	5.30E-03	CC-V
13	91	5.23E-03	CL-WF&R
14	157	5.16E-03	CC-V&R
15	296	5.04E-03	CL-WF
16	136	5.01E-03	CL-general
17	161	4.99E-03	CC-R
18	74	4.93E-03	CC-WF
19	298	4.87E-03	CC-V
20	125	4.82E-03	CL-general

Table 3. The codewords with highest twenty IDF values and their corresponding anatomical features.

Rank	Codeword	IDF score	Anatomical features
1	466	1.97E-01	CC-V&AP
2	275	1.24E-01	CC-V
3	177	1.07E-01	CC-V
4	157	1.04E-01	CC-V&R
5	232	9.38E-02	CC-V&AP
6	422	9.05E-02	CC-general
7	136	8.40E-02	CL-general
8	454	8.40E-02	CC-general
9	368	8.08E-02	CL-general
10	319	7.76E-02	CC-R
11	407	7.76E-02	CC-R
12	479	7.76E-02	CC-WF&R
13	333	7.43E-02	CL-general
14	397	7.43E-02	CC-WF
15	283	7.11E-02	CL-R
16	378	6.79E-02	CC-general
17	8	5.84E-02	CC-general
18	101	5.84E-02	CC-WF
19	299	5.52E-02	CL-general
20	223	5.21E-02	CC-general

Note: CC, cell corners; CL, cell lumina; V, vessels; R, ray parenchyma cells; WF, wood fibers; AP, axial parenchyma cells; and general, the features can observe in various anatomical features.

Informative codewords for Lauraceae identification were primarily cell corner-based features. It is also confirmed by Hwang et al. [13] that cell corners are important features for wood classification. Thirteen of the twenty codewords were cell corner-based features. In particular, the importance of vessels, which are numerically inferior anatomical features, was remarkably high. This means that the Lauraceae family has useful information for identification in cell corners that exist between vessels and adjacent anatomical features (Figure 7). This difference in the vessels can be attributed to the presence or absence, and/or the numerical diversity of the anatomical feature by species. According to the codewords with the highest twenty IDF scores (Table 3), the vessels also had higher IDF scores. Therefore, the vessels are not only species-specific features of Lauraceae species, but also exist in different numbers by species. The cell corners of wood fibers in the RF also had a high importance, but their importance in the TFDIF score was somewhat lower. This implies that the cell corners of wood fibers are common features shared by many species, and the number of distributions varies with species.

The feature importance derived from TFDIF scores and RF provided useful information on which anatomical features are species specific or common and the quantitative differences by species. This information has helped us understand what the computer vision observes from the wood and how it identifies the Lauraceae family.

5. Conclusion

From the cross-sectional optical micrographs of the Lauraceae family, the SIFT algorithm detected informative local features for identification, and the codebook effectively represented various anatomical features. The BOF framework demonstrated that it is a more robust model for identification than the simple model that combined SIFT descriptors with a classifier. The cell corner-based features were more influential on Lauraceae identification than the cell lumen-based features. In particular, cell corners that connect vessels with adjacent anatomical features were informative. Wood

is composed of various tissues and each tissue has its own morphological characteristics. Computer vision has proven that there are unique and informative features in the contact regions of various adjacent tissues, and not just one independent tissue. This study is not a computer-aid identification based on anatomical keys. Local features are detected from wood images by computer vision, and classification models learn the features and classify categories. Thus, our findings may differ from the currently established wood anatomy. In further studies, we are focusing on reducing the gap between informatics and anatomy, and such studies are underway.

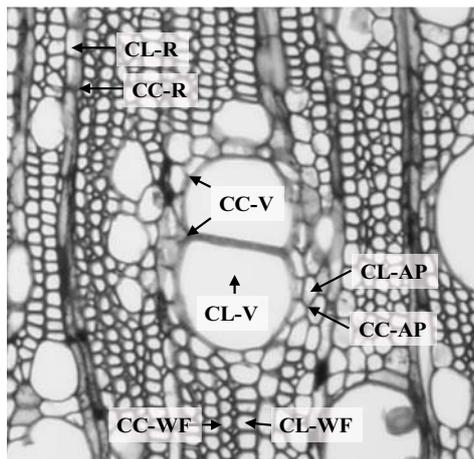


Figure 7. Various wood cells forming deciduous trees observed in a cross section of *Cinnamomum camphora*. CC, cell corners; CL, cell lumina; V, vessels; R, ray parenchyma cells; WF, wood fibers; and AP, axial parenchyma cells.

References

- [1] Wheeler E A, Bass P, and Gasson P E 1989 *IAWA Bulletin, New ser.* **10** 219–332
- [2] Forestry and Forest Products Research Institute. <http://db.ffpri.affrc.go.jp/WoodDB/IDBK-E/home.php>
- [3] Ohyama M, Baba K, and Itoh T 2001 *J Wood Sci.* **47** 81–86
- [4] Watanabe U, Abe H, Yoshida K, and Sugiyama J 2015 *J. Wood Sci.* **61** 1–9
- [5] Liu Z F, Ci X Q, Li L, Li H W, Conran J G, and Li J 2017 *PLOS ONE* **12** e0175788
- [6] Hwang S W, Horikawa Y, Lee W H, and Sugiyama J 2017 *J. Wood Sci.* **62** 156–167
- [7] Horikawa Y, Mizuno-Tazuru S, and Sugiyama J 2015 *J. Wood Sci.* **61** 251–261
- [8] Tsuchikawa S, Inoue K, Noma J, and Hayashi K 2003 *J. Wood Sci.* **49** 29–35
- [9] Tou J Y, Tay T H, and Lau P Y 2009 *ICNC '09, IEEE* **5** 8–12
- [10] Kobayashi K, Akada M, Torigoe T, Imazu S, and Sugiyama J 2015 *J. Wood Sci.* **61** 630–640
- [11] Kobayashi K, Hwang S W, Lee W H, and Sugiyama J 2017 *J. Wood Sci.* **63** 322–330
- [12] Hu S, Li K, and Bao X 2015 *CISP 2015, IEEE* 702–706
- [13] Hwang S W, Kobayashi K, Zhai S, and Sugiyama J 2018 *J. Wood Sci.* **64** 69–77
- [14] Lowe DG 2004 *J. Comput. Vis.* **60** 91–110
- [15] Stanford University CS131. http://vision.stanford.edu/teaching/cs131_fall1516/
- [16] Nurhaida I, Noviyamto A, Manurung R, and Arymurthy AM 2015 *Proced. Comput. Sci.* **59** 567–576
- [17] Huang S, Cai C, and Zhang Y 2009 *CiSE 2009* 1–4
- [18] Mikolajczyk K, and Schmid C 2005 *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1615–1630
- [19] Sivic J, and Zisserman A 2003 *ICCV '03, IEEE* 1470–1477
- [20] UCF Computer Vision Video Lectures 2012. (Video file) Retrieved from <https://www.youtube.com/watch?v=iGZpJZhqEME>
- [21] Sculley D 2010 *WWW '10* 1177–1178
- [22] Arthur D, and Vassilvitskii S 2007 *ACM-SIAM Symp. Discret. Algorithm* 1027–1035
- [23] Breiman L 2001 *Mach. Learn.* **45** 5–32
- [24] Breiman L 1996 *Mach. Learn.* **24** 123–140