A Database for Forest Species Recognition

Pedro Luiz de Paula Filho[†‡], Luiz S. Oliveira[†], Alceu S. Britto Jr.[♯]
[†]Universidade Federal do Paraná, Departamento de Informática, Curitiba, Brazil
[‡]Universidade do Contestado, Canoinhas/Porto Unio, Brazil
[♯]Pontifícia Universidade Católica do Paraná, Curitiba, Brazil
lesoliveira@.inf.ufpr.br, pedro@pu.unc.br, alceu@ppgia.pucpr.br

Abstract

In this work we introduce a database of Brazilian forest species which contains images of 11 different species of the Brazilian flora. It had been built in collaboration with the Laboratory of Wood Anatomy of the Federal University of Parana (UFPR), which assures the correct labeling of the selected species. Such a database permits further research on forest species recognition and serve as basis for comparisons. Preliminary results are also presented to give an insight about the complexity of the data.

1 Introduction

The correct identification of forest species is of vital need for the activity of forest management and safe trade of timber species. The classification of a type of log when this one is not found in the forest where there are flowers, fruit, and leaves is a very complex activity. However, the recognition of forest species is important in several areas, as stated by Khalid et al [1] and Tou et al [2], especially its application in the industry and preventing frauds where a wood trader might mix a noble species with a cheaper one, or even try to export wood whose species is endangered. Generally, the recognition is done by well trained specialists, but few reach good accuracy in classification due to the time it takes for their training, hence they are not enough to meet the industry demands.

Computer vision systems are a very interesting alternative in this case. However, to implement a reliable system, a representative database should be built so that machine learning methods can be trained properly. In spite of the large number of applications, few works in the literature have addressed the recognition of forest species [1, 2]. As pointed out by Khalid et al in [1], the main difficulty lies in obtaining a consistent labeled database.

In this work we introduce a database of Brazilian forest species which contains images of 11 different species of the Brazilian flora. It had been built in collaboration with the Laboratory of Wood Anatomy of the Federal University of Parana (UFPR), which assures the correct labeling of the selected species. Such a database permits further research on forest species recognition and will serve as basis for comparisons. Besides, we present some preliminary results using color and co-occurrence features to give some insight about the complexity of the problem.

2 Database

The database described in this work contains 347 images of 11 different Brazilian forest species which were acquired in the Laboratory of Wood Anatomy of the UFPR (Federal University of Parana). In order to build this dataset two main criteria were adopted: the availability of labeled samples in the laboratory and the level of confusion when the species were classified by human experts. In this first version of the database, we selected those species that normally create doubts in their classification to the naked eye. Figure 1 shows some samples of the database while Table 1 gives some details about each specie.

Figure 1. Database samples.

A protocol was defined for the acquisition of images to ensure standardized distance and lighting. The camera was fixed into a base along with two halogen lamps so that we would have constancy of light. A low-cost digital camera (SONY DSC T20) with both macro mode and auto-focus enabled was used to acquire the images. The distance between the camera lens and the wood samples was 20 millimeters. The images were obtained with a resolution of $3,264 \times 2,448$ pixels. It is important to mention that the only technique applied to the wood sample before acquisition was sanding. This technique, along with a low-cost camera, requires no structure in the field and alike more complex techniques such as wood boiling, microtome use, it also ensure a good visualization of the wood structures.

Table 1. Species of the Database

Label	Popular Name	Scientific Name	Images
0	Amap	B. Parinarioides	34
1	Cumaru	Dipteryx SP	35
2	Eucalipto	Eucalyptus SP	22
3	Ip	Tabebuia SP	38
4	Jacarand	Machaerium SP	29
5	Marup	Simarouba amara	15
6	Maaranduba	Manilkara huberi	25
7	Pau Marfim	B. Riedelianum	17
8	Sucupira	Bowdichia SP	45
9	Tatajuba	Bagassa guianensis	44
10	Tuari	Couratari SP	43

After some analysis we have noticed that some species have a great intra-class variability, which can have a considerable impact on the recognition performance. This variability can be either something inherent to the class, like the different colors observed in Figure 2a or something produced by the cutting apparatus, like the diagonal marks in Figure 2b. One way to cope with this variability is to split the images into smaller images so that more homogeneous images could be generated. In this way, during testing the images would be equally segmented into smaller images and the final decision would be produced by some combination method such as voting, average, etc. After some experiments, we found out that segmenting the original images into 25 pieces of 130×98 pixels produced good results.

The database was divided into 104, 78, and 173 images for training, validation, and testing, respectively. The class distribution reported in Table 1 was respected so that all the classes would be properly represented.

3 Preliminary Experiments

To give an insight about the complexity of the proposed database, in this section we report some experiments using color and co-occurrence features. The feature vector used in this experiments is composed of 18 color features extracted from the RGB, HSV and CIELUV color models and 24 features extracted from the co-occurrence matrix. The machine learning model used was an Multilayer Perceptron (MLP) trained with the gradient descent applied to a sum-of-squares error function. The transfer function employed is the standard sigmoid function. In order to monitor the generalization performance during learning and terminate the algorithm when there is no longer an improvement, we have used the method of validation.

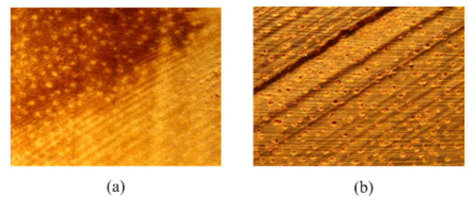

Figure 2. Intra-class variability: (a) intrinsic features and (b) mark produced by the cutting apparatus.

As mentioned before the testing set is composed of 173 images. Two different recognition strategies were tested. In the first one, all the images (training, validation, and testing) are segmented into 25 pieces. After training, the classifier assigns a class for each sub-image and the final decision can be obtained through any combination rule. In our experiments, the voting rule surpassed other combination rules. In the second strategy, the images are used without any segmentation. The first strategy reaches 82% of recognition rate against 65% of the second strategy. In other words, the first strategy outperforms the second one in about 17%.

4 Conclusion

This short communication introduces a database of Brazilian forest species which will be very useful for further investigations on the automatic recognition of Brazilian forest species. The first version of the database is composed of 317 images of 11 different species of the Brazilian flora. Preliminary results show a recognition rate of 82%, which compares to other results found in the literature [2] using different databases. As soon as new species become available and labeled, they will be incorporated into the database.

References

[1] M. Khalid, E. L. Y. Lee, R. Yusof, and M. Nadaraj. Design of an intelligent wood species recognition system. *IJSSST*, 9(3), 2008.

[2] J. Y. Tou, P. Y. Lau, and Y. H. Tay. Computer vision-based wood recognition system. *Proceedings of International Workshop on Advanced Image Technology*, 2007.