Bag of Words Representation and SVM Classifier for Timber Knots Detection on Color Images

Mohamad Mazen Hittawe
mohamad-mazen.hittawe
@u-bourgogne.fr

Désiré Sidibé
dro-desire.didibe
@u-bourgogne.fr

Fabrice Mériaudeau
fabrice.meriaudeau
@u-bourgogne.fr

Le2i UMR CNRS 6306, Université de Bourgogne
12 rue de la fonderie, 71200 Le Creusot, France

Abstract

Knots as well as their density have a huge impact on the mechanical properties of wood boards. This paper addresses the issue of their automatic detection. An image processing pipeline which associates low level processing (contrast enhancement, thresholding, mathematical morphology) with bag-of-words approach is developed. We propose a SVM classification based on features obtained by SURF descriptors on RGB images, followed by a dictionary created using the bag-of-words approach. Our method was tested on color images from two different datasets with a total number of 640 knots. The mean recall (true positive) rate achieved was (92%) and (97%) for a single dictionary (built only on samples from the first dataset), for the two datasets respectively, illustrating the robustness of our method.

1 Introduction

Wood used for structural purposes needs to fulfill some requirements and must be sorted by grades to ensure the designer that the product meets some mechanical specification. The wood quality is resolved considering the quantity of defects and their surfaces. Knots, by nature, are the most prominent defects found in wood examination. Knots lessen the mechanical properties of timber and have also been viewed as an undesirable visual feature (for wood furniture). Consequently, wood boards with few knots are exchanged at higher costs in the business sector than boards with numerous knots. There has been a developing concern toward innovative work of wood imperfection detection as well as characterization techniques and frameworks. Several studies have been carried out in order to define the knottiness of a timber. The knottiness of a board is characterized by a value which is used in prediction's models of mechanical properties [4]. Among the very few published methods which are dealing with wood defect detection, attention is only brought to methods dealing with knots detection. In [6], the authors described a method that combines texture features and percentile features using simple spatial operators. They detected the defects in parquet slabs, and showed that co-occurrence features outperformed the signed differences and local binary patterns (LBP), which demonstrates that the intensity information is very descriptive, the dataset was extracted from an application environment developed for detecting the defects in parquet slabs. The results shown that the performance can be improved if additional features were extracted. The main weakness of this method using color percentile features is that the dependencies between neighboring pixels are not described. In [7], the authors used non-supervised clustering for distinguishing between defects and clear wood based on Self-Organizing Map (SOM). Multidimensional feature vectors containing texture and color is used for training. However, the main limitation of this approach is the selection of the most discriminative features. Fuzzy Min-Max neural network for image segmentation was proposed in [8]. It combines many steps, the first of which is a seed selection process followed by a Fuzzy Min-Max neural network used for classification. The system outputs minimum bounding boxes enclosing the defects. In [11], the authors applied image segmentation on digital veneer images through a series of morphological operations to separate the regions which contain knots. They also applied an adaptive thresholding method to the gray-scale and red components of the segmented images to enhance the accuracy of knot detection. However, this approach produces a lot of false positive detection.

In this paper, we propose a method based on classification using the bag of words (BoW) approach. First, a dictionary based on SURF features [1] extracted from knot and clear wood regions in a training step, is created. Then, for a given test image, the potential knot regions are detected in a pre-processing step, then SURF features for each candidate regions are extracted. These features are transformed into a histogram of words occurrences using the learned dictionary in the training step. Finally we use the histogram as input to a linear kernel SVM for classification.

The rest of the paper is organized as follows: on section 2 our proposed method is explained. Experimental results are presented in section 3, and finally section 4 concludes the paper.

2 Proposed Method

The proposed method comprises two steps: i) a preprocessing step for improving the contrast of the input image and for detecting potential knots regions in the image, and ii) a classification step for detecting the timber knots.

2.1 Pre-processing

Knots appear as regions of dark intensities in the RGB color images of wood timbers. As such, they may be detected through thresholding. However, the

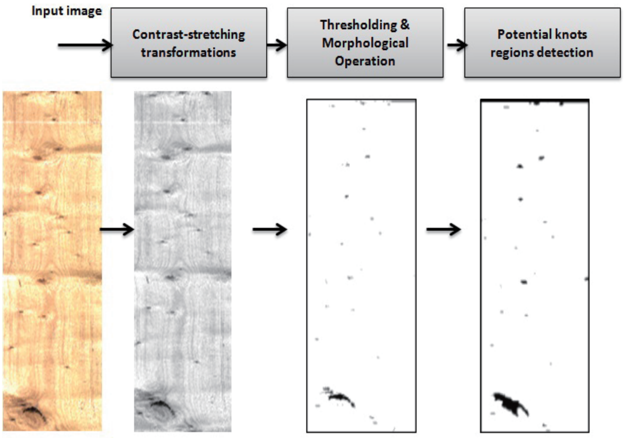

Figure 1. Pre-processing flowchart

threshold value may be different for each image, requiring empirical determination. We therefore use Otsu's method for automatic determination of the threshold [5].

Prior to thresholding, contrast stretching is performed to isolate dark regions and thereby knots. Contrast stretching is the process of transforming the intensity characteristics of the image such that dark regions are further darkened and bright regions are further brightened [3]. We employ the following stretching operator:

$$I_s = \frac{1}{1 + (\frac{\overline{I}}{I+\epsilon})^n}, \tag{1}$$

where I is the input image, $\overline{I}$ its mean intensity value, and I_s is the output of the transformation. ε is a term added to avoid division by 0, and n controls the slope of the applied function.

After contrast stretching, potential knots regions are detected by thresholding the input image using a threshold obtained with Otsu's method. Furthermore, morphological operations (such as *opening* and *closing*) are performed subsequent to thresholding in order to form connected components comprising the knots. *Opening* detaches poorly connected regions, while *closing* forms connected components. Figure. 1 shows an overview of the pre-pocessing steps used to detect knots in color images of timbers.

While the thresholding method finds the knots in the input image, it also produces many false detection in cases where the intensity characteristics of the knots in the image are not distinguishable or in case of noise in the image. In the next subsection, we propose a classification method to reduce the number of false detections.

2.2 Bag-of-words representation and classification

Our method uses the bag-of-words (BoW) representation of images for classification. BoW introduced in [10] is a powerful image representation method that has been used in different applications such as object recognition and image category classification. In our work, we adopt a supervised classification method. We first form a training set by collecting a set of images of knots regions in different images. The training set also contains images of clear wood regions (that is regions not corresponding to knots). Our training set contains 100 knots images and 100 clear wood images respectively. We then extract a set of low-level features from the training images and build a visual vocabulary, or codebook, by quantization of the low-level features. SURF features [1] are extracted from the three channels of the RGB color space. We use SURF because it is fast and invariant to image transformations like illumination changes, rotation and scale changes.

The features extracted from the entire set of training examples are then used to create a codebook using K-means clustering. If we define K clusters in the feature space, then the visual dictionary or codebook will contain K words each one being the center of one cluster. After creating the codebook, each of the training example is represented as a histogram of size K obtained by calculating the frequency of occurrences of each of the K words in the features extracted from the image.

Finally, the obtained histogram representations are used as input features to train a linear SVM classifier [2] for distinguishing between knots areas and other regions in a timber image. The SVM finds a linear hyper-plane which maximizes the margin in this higher dimensional space. The training vectors x are mapped into a higher dimensional space by a kernel function $k(x, y)$, here we use linear kernel defined as following:

$$K(x, y) = x^T y. \tag{2}$$

For detecting knots in a new test image, the preprocessing method explained in section 2.1 is first applied which results in a set of potential knots regions in the image. We then extract SURF features from each potential region and find the closest word in the codebook to each individual feature. The frequency of occurrences of the visual words forms the histogram representation of the region. Finally, the SVM classifier uses this histogram representation as input feature vector to predict the label of this region.

A diagram that summarizes the proposed method is shown in Figure. 2.

3 Experimental Results

In this section we present quantitative and qualitative results obtained by the proposed method. The proposed method is evaluated using two different datasets which have different wood characteristics. The first dataset consists of 252 images of Epicea wood with dimensions 650 × 4500 pixels. This dataset is directly extracted from the common production of a sawmill where the ground-truth was provided manually. The second dataset is a subset of the university of Oulu wood dataset [9]. We have randomly selected 50 images containing about 113 knots. The images have resolution 488 × 512.

For creating the visual dictionary, we have extracted 100 knots regions and 100 clear wood regions from the first dataset (Epicea dataset) to form our training set. As explained in section 2.2, we first extract SURF features from the entire training set of images, and create a codebook by quantization of the features into a set

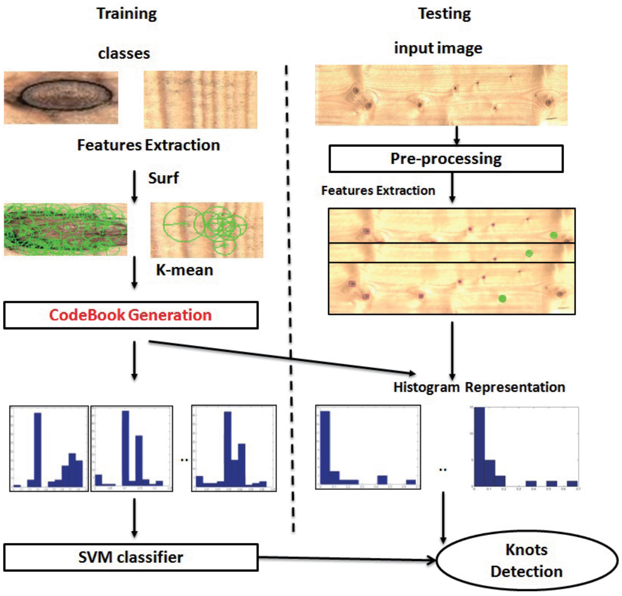

Figure 2. Proposed Algorithm

of words. Then, each training image is coded as a histogram representing the frequency of occurrences of the visual words. Finally, the histogram representation are used as input to train a linear SVM classifier. For testing, a second set of images of wood boards containing 527 knot regions, not used for training, is used.

In order to evaluate the performance of the method, we used three measures which are *recall*, *precision* and *false negative rate* (FN). The *recall* (or true positive rate) represents the proportion of actual knots correctly identified by the method. The *precision* is the proportion of positively identified cases that are actual knots, and FN represents the proportion of actual knots that are not detected by the method. These measures are defined by the following equations:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$FNrate = \frac{FN}{TP + FN} \tag{5}$$

For the first dataset, which contains 527 knots, the proposed method correctly detects 485 knots which corresponds to a recall of 0.92% and a precision of 0.9% when using a dictionary of size 25 words. Figure. 3 shows the evolution of *recall* and *precision* as the size of the dictionary varies. As can be seen, a dictionary of 25 words achieves the highest performance both in terms of *recall* and *precision*.

The overall value of FN rate is 0.08 for Epicea dataset, which corresponds to 42 knots not detected. In Figure. 4 we show the histogram of knots which are not detected in Epicea dataset. The x-axis represents the area of knots in pixels and the y-axis represents the number of knots. The dimension of the timber board in Epicea dataset is $4000mm \times 150mm$ =600000 mm^2 and the image resolution is $645 \times 4038 = 2604510$ pixels.

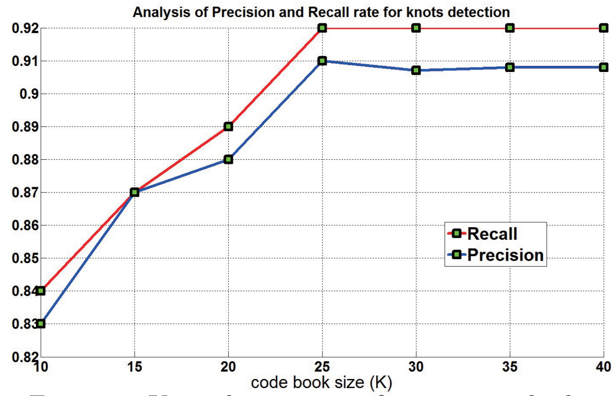

Figure 3. Knot detection performance with the Epicea datatset.

This means the area of one pixel is equal to $0.23mm^2$. In the worst case, the largest knot not detected by our method has an area of 600 pixels which corresponds to a knot of size 11.76×11.76 mm^2. The mean size, in pixels, of knots not detected is equal to 150 pixels and corresponds to area of 6×6 mm^2, which will not affect the stiffness of the lumber. Therefore, our method misses in the pre-processing step only knots of very small size.

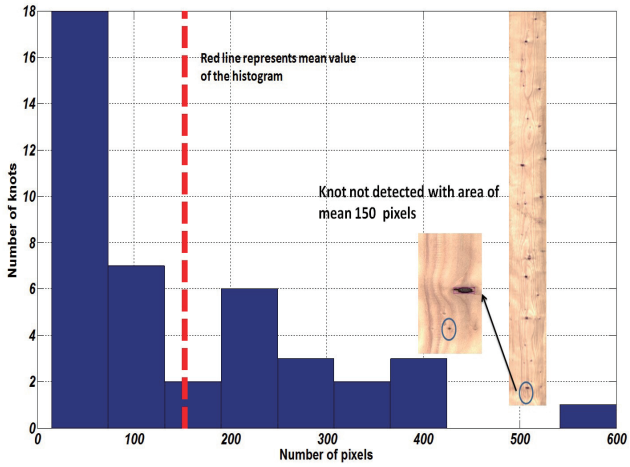

Figure 4. Analysis of knots not detected.

We also evaluate our method with the Oulu [9] dataset which is a complete different dataset from the Epicea dataset used to create the dictionary. The best results for this dataset in terms of *precision* and *recall* were obtained for a dictionary size of 35 words which is illustrated in Figure. 5. The method correctly detected 109 knots out of 113, which corresponds to a *recall* of 0.97% and a *precision* of 0.87%. The overall values of FN rate for the Oulu dataset is 0.03 which corresponds to 4 knots not detected.

The results are summarized in the (table(1)) and Figure. 6 shows some qualitative knots detection results on color images for the two datasets.

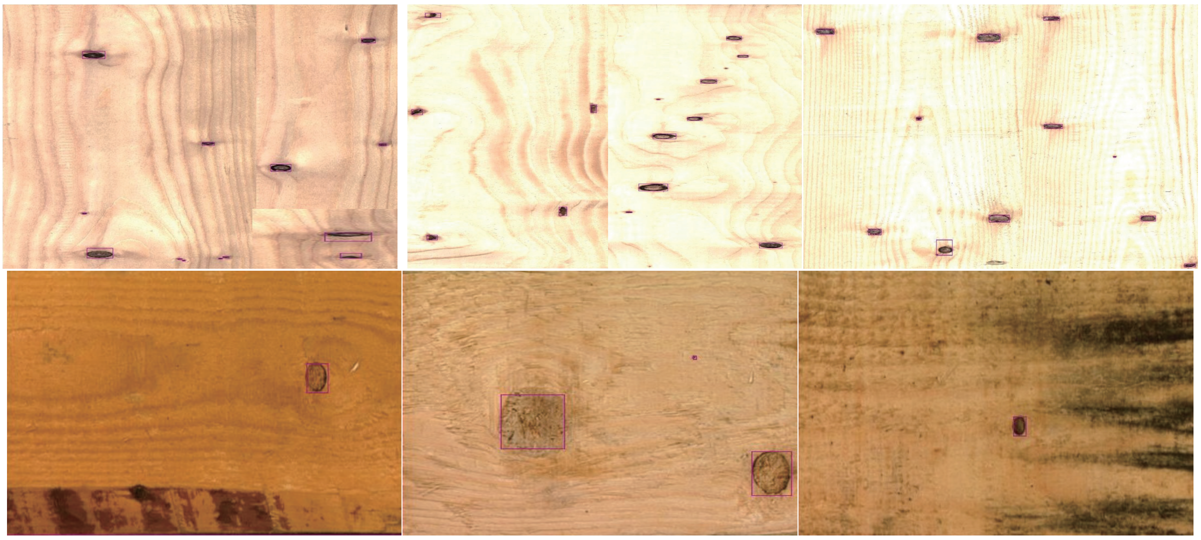

Figure 6. Qualitative results of the proposed method. First row shows results for the Epicea dataset, and Second row shows results for Oulu datset.

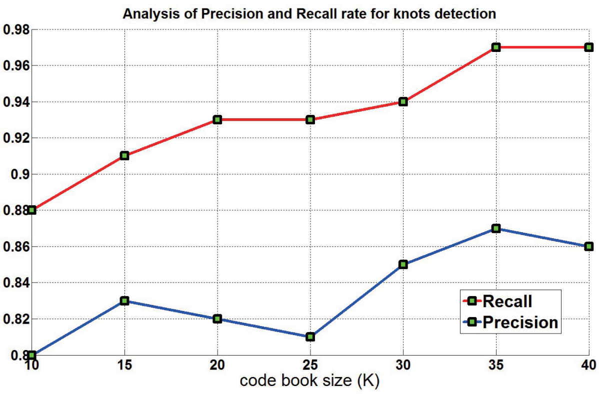

Figure 5. Knot detection performance with the Oulu datatset [9].

Dataset	Total count of knots	Recall	Precision	FNrate
Epicea	527	0.92	0.90	0.08
Oulu [9]	113	0.97	0.87	0.03

Table 1. Knots detection results.

4 Conclusion

In this work, we proposed a method for automatic timber knot detection in color images. The method is based on a pre-processing step which detects potential knot regions using contrast enhancement and thresholding. Then, a classification step based on the bag-of-word approach and SVM classifier is proposed to reduce the false positive cases. Experimental results using two different datasets show the good performance of our method with a recall of 0.92 and 0.97 for both dataset, and precision of 0.90 and 0.87 respectively. The knots that are not detected correspond to very small areas on the board, about an area of 36 mm^2 on a timber board of size $4000mm \times 150mm$. Our future works include the analysis of other wood defects such as cracks and the evaluation with other types of wood.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[2] C-C. Chang and C-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, 2011.

[3] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. Digital image processing using matlab. *Upper Saddle River, NJ Jensen: Prentice Hall,*, 2004.

[4] M. Hittawe, M, D Sidibe, and F. Meridaudeau. A machine vision based approach for timber knots detection. In *International Conference on Quality Control by Artificial Vision*, 2015.

[5] B. Hu and N. Gong. An improved otsu's thresholding segmentation method. *Microelectronics & Computer*, 12:043, 2009.

[6] J. Kyllonen and M. Pietikainen. 5-1 visual inspection of parquet slabs by combining color and texture. 2000.

[7] M. Niskanen, O. Silvén, and H. Kauppinen. Color and texture based wood inspection with non-supervised clustering. In *Proceedings of the scandinavian Conference on image analysis*, pages 336–342, 2001.

[8] GA Ruza and PA Estéveza. Image segmentation using fuzzy min-max neural networks for wood defect detection. In *Intelligent Production Machines and Systems-First I* PROMS Virtual Conference: Proceedings and CD-ROM set*, page 183. Elsevier, 2005.

[9] O. Silvén, M. Niskanen, and H. Kauppinen. Wood inspection with non-supervised clustering. *Machine Vision and Applications*, 13(5-6):275–285, 2003.

[10] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1470–1477, October 2003.

[11] C. Todoroki, E.C. Lowell, and D. Dykstra. Automated knot detection with visual post-processing of douglas-fir veneer images. *Computers and electronics in agriculture*, 70(1):163–171, 2010.