Knot Detection in X-ray Images of Wood Planks using Dictionary Learning

Mattias Hansson, Alexandru Enescu, Sami S. Brandt
Image Group, Dept. Computer Science, University of Copenhagen
Sigurdsgade 41, 2100 København Ø, Denmark
nmhansson@di.ku.dk

Abstract

This paper considers a novel application of x-ray imaging of planks, for the purpose of detecting knots in high quality furniture wood. X-ray imaging allows the detection of knots invisible from the surface to conventional cameras. Our approach is based on texture analysis, or more specifically, discriminative dictionary learning. Experiments show that the knot detection and segmentation can be accurately performed by our approach. This is a promising result and can be directly applied in industrial processing of furniture wood.

1 Introduction

Detecting knots in wood planks is a highly relevant task in high-quality furniture production, where the tolerance of knots in the finished product is very low. Knots may divided into a multitude of categories, such as splits, decay, wanes, sound and unsound, but in this paper knot types are not differentiated. The detection is a difficult problem, since knots are non-uniform structures composed of a pattern of bright and dark intensities.

This paper proposes a pattern recognition method for knot detection based on dictionary learning; a block diagram of the method is given in Fig. 1. There are various approaches for wood defect detection: optical cameras, x-ray, microwaves and ultrasound. In the following, we focus on the x-ray modality for detection and segmentation. In our set-up, x-ray was chosen as imaging modality, since this allows detection of knots at all depths.

Many strategies have been proposed for wood defect detection solely from X-ray images. Funt *et al.* [1] divided x-ray images into four categories with a threshold determined by derivatives of intensity histograms. Thereafter, the size and orientation of highest density regions were determined and labeled knots. In [2], the image was first filtered after which a wood-type specific threshold was applied in five concentric surfaces in the log. The positions of the knots were predicted from 11 parameters. Thresholding is however problematic when logs have a high moisture content since knots have similar density to sapwood [3]. In [4], knots were classified by thresholding in each CT slice, after which the convex hull of each segmentation was computed. The resulting contour was discretized by b-splines and used as an input to a Kalman filtering model, which produced a three-dimensional reconstruction of knots. Rojas *et al.* [5] described the minimum distance (MDC) and the Maximum Likelihood classifier (MLC). These classifiers identify 5 categories (sapwood, heartwood, rot, splits, knots and bark). Classification was performed by nearest class mean distance and most probable class, respectively. In [6], a neural classifier was trained from image features consisting of intensity values, distance between the pixel of interest and the center of the log, and seven textural features: homogeneity, contrast, dissimilarity, mean, SD, entropy and angular second moment. In the experiments, black spruce logs were classified into heartwood, sapwood, bark and knots.

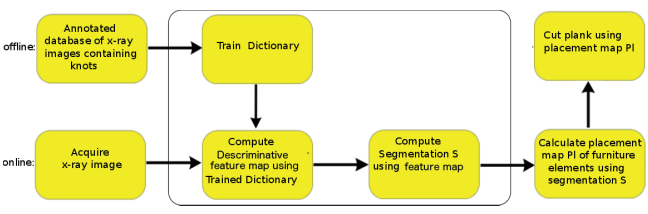


Figure 1: Block diagram of the proposed framework. This paper considers the pattern recognition tasks enclosed by the large rectangle.

The validation measures typically used in defect detection with the x-ray modality are producer's accuracy (PA), overall accuracy (OA), detection rate (DR) and false alarm rate (FA). Producer's accuracy is defined as the total number of correct pixels in the category divided by the total number of ground truth pixels in that category. Overall accuracy is defined as the total number of correctly classified pixels in all the categories divided by the total number pixels. Detection rate is the rate of correctly identified knots in an image. False alarm rate denotes the rate of incorrect detections. Correctly identified implies that at least one pixel of the knot is detected. A proper detection rate should be paired up with accuracy for a useful result. Performance of the works above for x-ray knot segmentation ranges between 76.4% to 97.6% in OA [5], [6] and 97.2% to 100% in DR [2], [4].

All the current methods for knot detection rely on pre-defined features; however, they often fail since they are not effective enough due to the variability of the knot appearance. To date, to the best of our knowledge, non-parametric dictionary learning has not yet been used in knot segmentation, though it has been been widely applied for image representation with state-of-art results in many problems including image segmentation [7]. We hence approach wood defect detection by sparse representation and dictionary learning. In dictionary learning, the image patch basis is simultaneously learnt with their coefficients from the training set. The basis captures the variation of the dataset to the point that it is possible to reconstruct the data using the basis patches. We propose

a model in which discriminative feature maps are obtained through discriminative dictionary learning [8], and the maps are then classified to knot and normal wood areas by a Gaussian mixture model (GMM).

2 Dictionary learning

Dictionary learning is a sparse modeling method which allows the joint formation of the basis elements together with their coefficients. The method is more flexible than the application of predefined bases, such as Fourier basis or wavelets [9], since there is no need to require the basis elements to be orthogonal. Solving for the optimal sparse representation is an NP-complete problem, to which approximation schemes such as matching pursuit (MP) [10] and orthogonal matching pursuit (OMP) [11] have been introduced. The two main algorithms for dictionary learning are K-SVD [12] and MOD [13], of which K-SVD performs slightly better than MOD in the reconstructive setting [7]. In contrast to optimal reconstruction, the D-KSVD model [8] focuses on the discrimination of a particular part of the signal.

In reconstructive dictionary learning the idea is to minimize the reconstruction error of the sparse representation of data. The reconstructive dictionary is learnt by solving the problem

$$\min_{\alpha,D} \sum_l ||x_l - D\alpha_l||_2^2 \text{ s.t. } ||\alpha_l||_0 \leq L \ \forall l \ , \qquad (1)$$

where α_l is a sparse representation of the data vector x_l, $||\cdot||_0$ is the number of zero elements in the vector, and L is a sparsity factor. The l-th data vector is denoted by $x_l \in \mathbb{R}^n$, corresponding to the l:th square patch of size the c from an input image. The dictionary D is composed of k atoms of the size n, *i.e.* , $D \in \mathbb{R}^{n\times k}$. A standard way of solving (1) is to update α and D alternately. By OMP the optimal sparse representation $\alpha^\star$ is obtained as

$$\alpha^\star(x, D) \equiv \arg\min_{\alpha\in\mathbb{R}^k} \ R(x, D, \alpha) \ , \qquad (2)$$

where $R(x, D, \alpha) \equiv ||x - D\alpha||_2$ is the reconstruction error. A dictionary update algorithm, such as MOD or K-SVD, is then employed, where the K-SVD algorithm updates the non-zero coefficients of $\alpha^\star$ and D concurrently.

In discriminative dictionary learning, the learnt class dictionary is designed to reconstruct the underlying class well and the other classes poorly. In [8], N dictionaries were trained to discriminate N classes of data. Soft-max discriminative cost functions $C_i^\lambda(y_1, y_2, ..., y_N)$ were used to maintain separation between the N dictionaries modelling N distinct classes of objects. These are defined as $C_i^\lambda(y_1, y_2, ..., y_N) \equiv \log\sum_{j=1}^N e^{-\lambda(y_j - y_i)}$, which are close to zero when y_i is the smallest value among y_j and gives asymptotic linear penalty $\lambda(y_i - \min_j y_j)$, $i = 1, ..., N$. The soft-max discriminative cost functions yields the discriminative criterion

$$\min_{D_m} \sum_{\substack{i=1..N \\ l\in S_i}} C_i^\lambda\big(\{R(x_l, D_j, \alpha_{lj})\}_{j=1}^N\big) + \gamma\lambda R(x_l, D_i, \alpha_{li}) \ , \qquad (3)$$

where $D_m \in \{D_1, ..., D_N\}$, and S_i is the classified input data group i. Reconstruction and discrimination are balanced by the two non-negative parameters λ and γ.

3 Gaussian mixture model fitting

Let I be an x-ray image from which T overlapping patches of the size $c \times c$ are selected. The discriminative feature map I_{discr} is defined as the difference of reconstruction errors for the knot and the regular wood dictionary, or

$$I^l_{\text{discr}} = R(x_l, \alpha^\star, D_{\text{knot}}) - R(x_l, \alpha^\star, D_{\text{wood}}) \ , \qquad (4)$$

where x_l is the vectorized patch l.

We define the cross-validation measure for the image I as the discriminative difference

$$\mathcal{D}(I) = \sum_{l\in\text{knot}} I^l_{\text{discr}} - \sum_{l\notin\text{knot}} I^l_{\text{discr}} \ . \qquad (5)$$

The rationale is the heuristics that the best discriminative dictionary for knots should reconstruct the knot region well but the regular wood area poorly; the opposite should hold for the regular wood dictionary. The optimal parameters values are those that minimise (5).

We model I_{discr} as a 2-component Gaussian mixture model, one component for the background wood and the other knot. In addition, in the histograms of the error images, the pixels with $I_{\text{discr}} > \mu_{\text{wood}}$ of the pixel distribution, should always be classified as background wood, see Fig.2b. We hence model the posterior class probability model for the pixel value z in I_{discr} as

$$P(\text{"}x \text{ represents a knot"}|z) = \begin{cases} \frac{\pi\mathcal{G}(z|\mu_{\text{knot}},\sigma^2_{\text{knot}})}{\pi\mathcal{G}(z|\mu_{\text{knot}},\sigma^2_{\text{knot}})+(1-\pi)\mathcal{G}(z|\mu_{\text{wood}},\sigma^2_{\text{wood}})} & \text{if } z < \mu_{\text{wood}}, \\ 0 & \text{otherwise.} \end{cases} \qquad (6)$$

where $\mathcal{G}(z|\mu_{\text{knot}}, \sigma^2_{\text{knot}})$ is the Gaussian pdf conditioned to the knot and π is the mixing coefficient. The posterior probabilities provided by the GMM can be used to classify discriminative maps into knots and normal wood at the threshold 0.5 that yields the binary classification

$$\mathcal{C}(x) = \begin{cases} 1 & \text{if } P(\text{"}x \text{ represents a knot"}|z) > 0.5, \\ 0 & \text{otherwise.} \end{cases} \qquad (7)$$

These binary, modified Gaussian mixture model (MGMM) classifications are used as seeds to various segmentation methods in Sec. 4.

4 Experiments

4.1 Data

We used oak plank data taken by a single source x-ray scanning machine prototype, where the planks typically were 3 cm thick. The data set consisted of 72 manually annotated images taken from separate planks with the x-ray source positioned in the 90 degree angle of incidence to the plank. A Gaussian mask with $\sigma = 0.5$ was applied to images as pre-processing.

From the images, 20 were randomly selected for cross-validation of model parameters, while the remaining images were only used in testing. The annotation was performed by a trained non-expert with conventional image processing software. The manual segmentation task was challenging as the existence of a knot could not be verified by visual examination of the plank. For cross-validation, we used 280×280 subimages of knots, clipped from 13000×2266 scans. For testing, we used as large as 500×500 images containing knots to guarantee that a large part of the image also represents the background wood. This was done for the purpose of evaluating the occurrence false-positives. The image patches were individually normalised to zero mean and unity variance before further processing.

4.2 Parameters

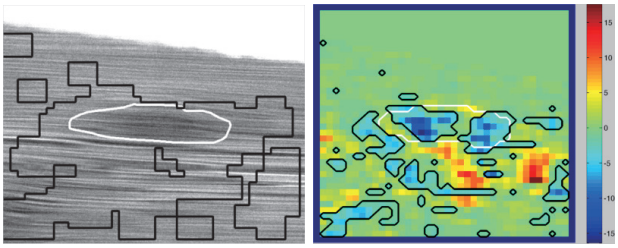
The parameters λ, γ and ϵ in (3) were determined through 5-fold cross-validation over 20 wood samples. We cross-validated over the parameter ranges $\lambda \in [0.1, 0.5]$, $\gamma \in [0.1, 0.5]$, and $\epsilon \in [0.01, 0.03]$, with the corresponding step-sizes of 0.1, 0.1 and 0.01. The remaining parameters were set as in [8]: dictionary size to $k = 128$, sparsity factor to $L = 4$, and patch size to $c \times c = 12 \times 12$. It would have been preferable to cross-validate over all the parameters, but due to the computational load, we focused on the three first parameters of the discriminative model.

4.3 Evaluated segmentation methods

After the cross-validation, posterior knot probability images were created from 52 test images. The posterior maps were then segmented by 6 different graph-based and active contour methods: min-cut graph cut (GC) [14], normalized cuts (NC) [15], grow cut (GrC) [16] and markov random field (MRF) model solved by ICM [17], Hybrid MGMM graph cuts, local Chan-Vese algorithm (CV) [18], and hybrid active contours algorithm (HAC) [19].

MGMM classification, estimated by the EM algorithm [20], was used as an initialisation for GC, GrC, CV and HAC. In the case of CV and HAC, the convex hull was computed for the points labeled as interior by MGMM. In the case of GC, NC, and GrC, seeds were sampled from the interior and exterior points. Additionally, for NC, the similarity weighting parameters were set to $\sigma_I = 0.1$, $\sigma_X = 150$ and $R = 50$. Hybrid GMM graph cuts, were set to use GMM probabilities (weights) as seeds to the region term in graph cut segmentation.

As evaluation measures, we used area based measures, Hausdorff distance and Jaccard Index, for segmentation validation since the methods do not necessarily produce single closed contours. The measures are defined as follows. Let X and Y be two sets of points. Hausdorff distance $HD(X,Y) = \max\{\sup_{x\in X} \inf_{y\in Y} d(x,y), \sup_{y\in Y} \inf_{x\in X} d(x,y)\}$, where $d(x,y)$ is the Euclidean distance between points x and y. Jaccard Index $JI(X,Y) = |X \cap Y|/|X \cup Y|$. We also used Overall Accuracy, Detection rate and False alarm rate, as described in Sec. 1.



(a) (b)

Figure 2: An example of poor performance. (a) X-ray image, ground truth (white) and MGMM classification (black); (b) the corresponding discriminative feature map.

5 Results

Our knot segmentation results are displayed in Table 1 and in Fig. 3. The cross-validation yielded the estimates $\hat{\lambda} = 0.4$, $\hat{\gamma} = 0.4$ and $\hat{\epsilon} = 0.02$. It can be seen that MGMM classification had the best overall performance, 91.7% OA and the lowest mean HD (9.8 pixels) from the ground truth, although several methods provide satisfactory performance. It also had a high DR (96.9%). The relatively high FA (25.6%) and low values of JI were partly due to the fact that the annotations were approximate, since the boundaries of knots were often unclear. The DR and FA results of MRF and Grow cut were due to oversegmentation, as indicated by HD and JI. In a few cases, 4 of 52, the dictionary was not sufficient to capture the knots, see Fig. 2a. This could have been resolved by increasing the number of training examples.

6 Conclusions and future work

We have proposed a method for detecting knots in wood planks from x-ray images while an automatic defect detection and wood processing have a high cost-saving potential due to reduction in waste. The method is based on discriminative dictionaries, where the discriminative feature maps are classified to knot and normal wood areas by a Gaussian Mixture

Table 1: Segmentation results. HD = Hausdorff distance (mean ± Standard Error of mean (SEM) pixels), JI = Jaccard Index (mean ± SEM), OA=Overall Accuracy (%), DR = Detection rate (%), FA = False Alarm rate (%).

Method \ Measure	HD	JI	OA	DR	FA
MGMM (proposed method)	9.8 ± 0.3	0.45 ± 0.03	91.7	96.9	25.6
Hybrid GMM Graph Cut	10.3 ± 0.4	0.43 ± 0.03	89.9	96.9	26.7
Graph Cut [14]	12.1 ± 0.5	0.35 ± 0.03	79.5	98.4	33.3
Normalized Cuts [15]	10.5 ± 0.4	0.42 ± 0.03	90.3	96.9	33.3
Grow Cut [16]	11.2 ± 0.5	0.37 ± 0.03	83.6	98.4	0
MRF [17]	17.3 ± 0.4	0.14 ± 0.02	34.6	100	48.4
Hybrid Active Contours [19]	9.9 ± 0.4	0.44 ± 0.03	90.7	95.3	15.3
Chan-Vese [18]	9.9 ± 0.4	0.43 ± 0.03	89.7	100	3

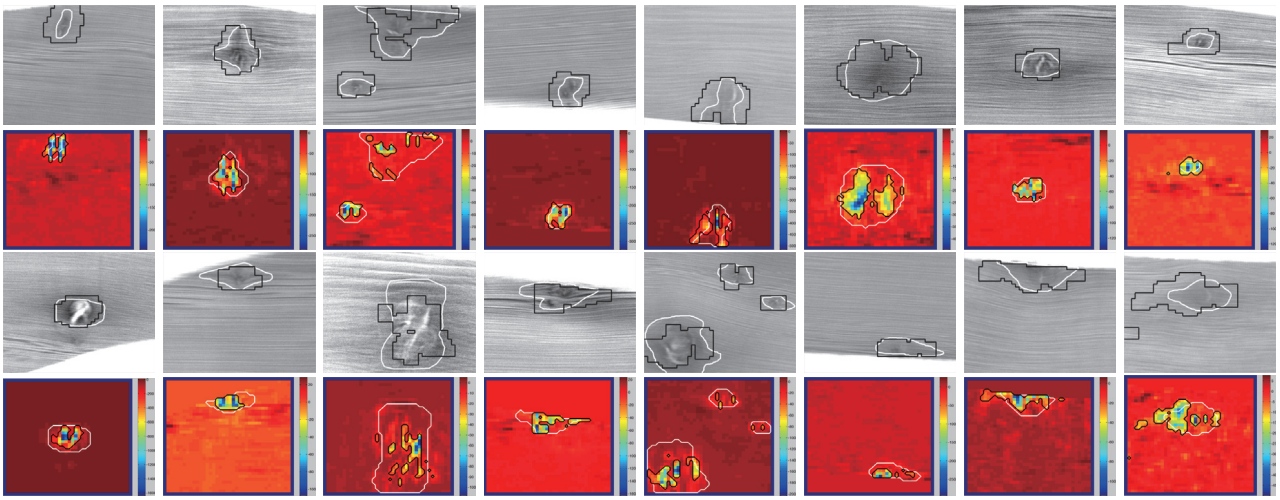


Figure 3: (row 1 and 3) 500×500 X-ray subimages of wood, overlaid with manual knot segmentations (white) and MGMM classifications (black). (row 2 and 4) Discriminative feature maps I_{discr} corresponding to the X-ray subimages.

Model. In our experiments, knots were detected at the mean Hausdorff distance of 9.8 pixels. The results are promising and demonstrate the viability of the approach in the industrial processing of furniture wood.

References

[1] B. V. Funt and E. C. Bryant, "Detection of internal log defects by automatic interpretation of computer tomography images," *Forest. Prod.*, vol. 37, no. 1, pp. 56–62, 1987.

[2] S. Grundberg and A. Grönlund, "Log scanning: extraction of knot geometry in CT-volumes," in *In Proc. Scanning Technology and Image Processing on Wood*, 1992.

[3] Q. Wei, B. Leblon, and A. La Rocque, "On The Use of X-ray Computed Tomography for Determining Wood Properties: A Review," *Can. J. Forest. Res.*, vol. 41, no. 11, pp. 2120–2140, 2011.

[4] S. M. Bhandarkar, X. Luo, R. F. Daniels, and E. W. Tollner, "Automated planning and optimization of lumber production using machine vision and computed tomography," *IEEE Trans. Autom. Sci. Eng.*, vol. 5, no. 4, pp. 677–695, 2008.

[5] G. Rojas, A. Condal, R. Beauregard, D. Verret, and R. E. Hernández, "Identification of internal defect of sugar maple logs from CT images using supervised classification methods," *Holz. Roh. Werkst.*, vol. 64, no. 4, pp. 295–303, 2006.

[6] Q. Wei, Y. H. Chui, B. Leblon, and S. Y. Zhang, "Identification of selected internal wood characteristics in computed tomography images of black spruce: a comparison study," *Wood Sci.*, vol. 55, no. 3, pp. 175–180, 2009.

[7] M. Elad, *Sparse and Redundant Representations:From Theory to Applications in Signal and Image Processing.* Springer Berlin, 2010.

[8] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. CVPR*, 2008, pp. 1–8.

[9] S. Mallat, *A Wavelet Tour of Signal Processing (Third Edition).* Academic Press, 2009.

[10] S. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[11] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[12] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[13] K. Engan, S. O. Aase, and J. H. Husoy, "Frame Based Signal Compression using Method of Optimal Directions (MOD)," in *Proc. ISCAS*, vol. 4, 1999, pp. 1–4 vol.4.

[14] Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation," *Int. J. Comput. Vision*, vol. 70, no. 2, pp. 109–131, Nov. 2006.

[15] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[16] V. Vezhnevets and V. Konouchine, "GrowCut: Interactive Multi-Label ND Image Segmentation by Cellular Automata," in *Proc. Graphicon*, 2005, pp. 150–156.

[17] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Roy. Stat. Soc. B. Met.*, vol. 48, no. 3, pp. 259–302, 1986.

[18] T. Chan and L. Vese, "Active Contours Without Edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.

[19] S. Lankton, D. Nain, A. Yezzi, and A. Tannenbaum, "Hybrid Geodesic Region-Based Curve Evolutions for Image Segmentation," *Proc. SPIE*, vol. 6510, 2007.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Stat. Soc. B. Met.*, vol. 39, pp. 1–38, 1977.