Research Journal of Applied Sciences 11 (6): 374-379, 2016
ISSN: 1815-932X

Comparison of Feature Extraction Techniques on Segmented Wood Defects

[1]Wan Siti Halimatul Munirah Wan Ahmad, [1]Hau-Lee Tong, [1]Hu Ng
[1]Timothy Tzen-Vun Yap and [2]Mohammad Faizal Ahmad Fauzi
[1]Faculty of Computing and Informatics,
[2]Faculty of Enginnering, Multimedia University, 63100 Cyberjaya, Malaysia

Abstract: This study introduces a analysis on the comparison of feature extraction techniques on the segmented wood defects. The method used is to classify four types of wood defects, namely knot, crack, holes and algae. The wood defect dataset used in this research consisted of 145 images were obtained from various sources. Fuzzy C-Means (FCM) is employed to segment wood defects into four clusters. Six types of feature extraction techniques namely Colour Histogram, Colour Coherence Vector, Local Binary Pattern, Gacfrdfbor Transform, Discrete Wavelet Frame and Gray Level Co-occurrence Matric are employed to describe the image's feature. The performances of Support Vector Machines (SVM), Bayes Networks and tree-based classifiers are comparedon the different defects and the classifiers' performances for each extraction technique are investigated. The experiment shows promising results with the highest classification accuracy of 94.5%, achieved by Random Forest classifier using Colour Histogram features. The proposed framework is useful in the automation of the detection of wood defects and is a superior alternative to manual selection and classification in the wood quality control.

Key words: Segment, wood, defect, crack, SMV

INTRODUCTION

In wood industry, quality control is an important aspect to guarantee that the product remainto a distinct group of qualitystandard. This would attract interest from potential buyer as well as to maintain the company's reputation of supplying high quality woods. The quality of wood is affected by the amount of defects and its dissemination.Currently, human experts that have been trained are needed to perform most lumber board inspections to detect the defects and wood grading. However, wood inspection for lengt hyperiod might in stigateeye fatigue, therefore lesse refficiency and accuracy in the quality examination (Gu *et al*., 2010). The manual work is also very time consuming, prone to human error and subject to each individual's performance based on one's technical skill, familiarity and attentiveness.

Semi or fully automatic wood inspection and classification are very useful to detect defects and categorize them into separate classes. The wood features are used to extract the information in the wood images and a classifier is used to classify the features into predefined defect clusters. The evaluation of classifiers is diverging, based on the extracted wood features used to depict the image and their intrinsic capability to contract with non linearity, deficient information and noise.

A lot of systems have been examinedand produced on automatic wood defects detection and classification. Recently, a study based on Minimum Distance Classifier and gray level histogram (GLH) features were conducted in (Packianather and Kapoor, 2015). There are seventeen features were derived from GLH to represent the proposed features. Promising results were reported with an increase of around 10% in the overall classification accuracy. In (Mahram *et al*., 2012), Gray Level Co-occurrence Matrix (GLCM) was used in the classification of wood defects, with local Binary Pattern (LBP) and statistical moments as the feature extraction techniques and with k-Nearest Neighborhood (k-NN) and SVM as classifiers. The study reported 100% classification accuracy using a combination of GLCM and LBP using either SVM or -NN. (Andersson and Vicen, 2010) had also adopted the SVM classifier in their wood knot classification work with the highest accuracy of 98%. In a review study by Pham and Alcock (1998), they found that tree-based classifier performs better than statistical classifier (Bayesian) and k-NN in wood defect classification.

MATERIALS AND METHODS

The framework of the proposed system is described in Fig. 1.

Corresponding Author: Wan Siti Halimatul Munirah Wan Ahmad, Faculty of Computing and Informatics, 63100 Cyberjaya, Malaysia

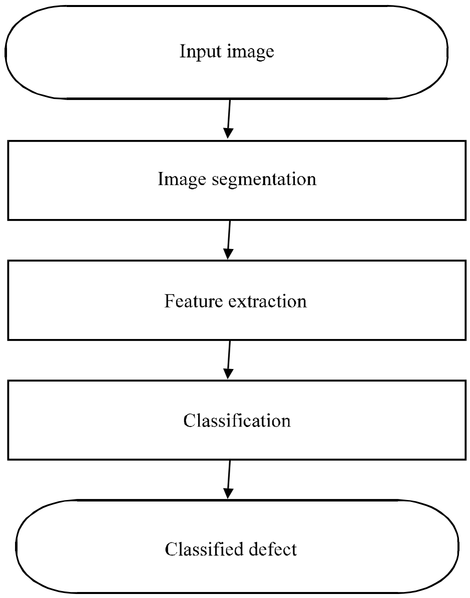

Fig. 1: Flowchart of the proposed system

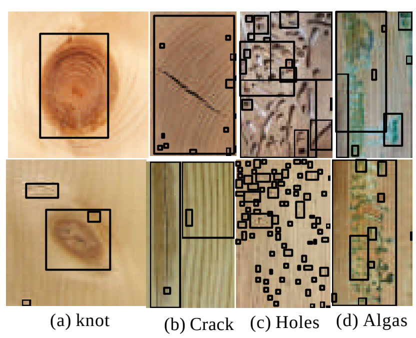

Fig. 2: Examples of the defects from the database

Wood defect dataset: The wood defects dataset used in this work consist of 145 images, comes from various sources. The images are in JPG format, with variousre solutions, extending from 100-450 pixels for the width and height. The images were made up of 4 major wood defects, where 30 of them were knots, 2 were cracks, 101 were holes and 12 were with algae. The example of each defect is shown in Fig. 2 with the bounding box highlighting the detected defect.

Image segmentation: FCM is as well-known segmentation method due to having good ambiguity and has been used for <20 years (Bezdek *et al.*, 1992). In this experiment, FCM is applied to segment the wood images into four desired clusters for the feature extraction process. This method is chosen based on its ability to divide certain pixels into several defined cluster with spatial (neighbourhood) information. Since, the wood defects images mainly consist of the background (wood colour), wood grain, shadows and several shades of colours, FCM with four clusters are tested for the segmentation. After analysing all four clusters, the first cluster (cluster 1) is taken as the defect region as it consists most of the defects region for all images in the database. Figure 3 listed the four clusters images for defects. The first cluster is further enhanced with morphological operations such as opening, dilation and filling to refine the defect regions.

Feature extraction: The features of the defects are then extracted using six feature extraction techniques. The techniques are Colour Histogram, Colour Coherence Vector, Gray Level Co-occurrence Matrix, Local Binary Pattern, Gabor Transform and Discrete Wavelet Frame. These techniques are chosen based on the literature. Colour Histogram is chosen instead of GLH (Packianather and Kapoor, 2015) to be able to analyze the algae and an additional color-based technique using Colour Coherence Vector is also applied to integrate some spatial data to the colour features. Gray Level Co-occurrence Matrix and Local Binary Pattern are employed based on their good performances by Mahram *et al.* (2012). In addition, Gabor Transform and Discrete Wavelet Frame are also tested based on Chacon and Alonso (2006) and Wang *et al.* (2009), respectively.

Colour Histogram. Colour Histogram (CH), introduced by Swain and Ballard (Swain and Ballard, 1991) is the most ordinary method to depict low-level colour contents of images. The CH is represented by aarray of bins where each bin denotes a particular colour. It is acquired by checking the amount of pixels that belongs by each bin based on their original colour. Colour histogram can be generated either as three independent colour distributions (one for each of the Red-Green-Blue (RGB) primary colours) or, more commonly as a joint distribution of all three primary colours (the so-called 3D colour histogram). Usually, the obtained histograms are regularize based on the total number of image pixels. In this research method, we experimented with 3D colour histogram using 64 colour bins, resulting in 64-dimensional feature vectors.

Colour coherence vector: Colour Coherence Vector (CCV) was introduced by (Pass *et al.*, 1996) where the method integrates some spatial datain an image. Every pixel in a specified bin is being classified into two groups; coherent and incoherent. A pixel will be considered as coherent if it is fit intoto a big connected group of relatedpixels; else it is consider as incoherent. Firstly, the procedureis to obtain the 3D colour histograms with n number of bins.

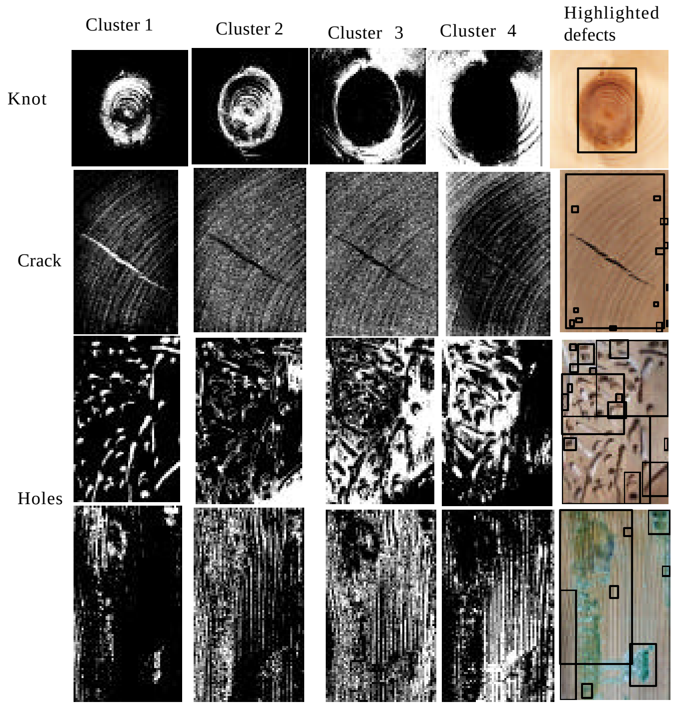

Fig. 3: Example of each defect from the database with the clusters images

Next, every pixel in all the bins are analyzed for their coherency by comparing the size of the region the pixels belong to with a predefined threshold value τ. For this study as in the colour histogram, 64 colour bins are used. A number of values of τ were employed and it is found that the optimal value of τ is 1% of the entire count of pixels in the particular image. Therefore, the feature vector for CCV is calculated as 128-dimensional (64 bins×2 categories).

Local Binary Pattern (LBP): In order to illustrate the local textural patterns of an image, LBP is an intensity-texture based techniquees tabl is hed by Ojala *et al.* (1996). Its different versions in the spatial domain for still images was discussed in detail (Pietikäinen *et al.*, 2011). LBP uses a circular neighbourhood, where the pixel values are altered bi-linearly when the sampling point be located not at the center of a pixel. To acquirea label for the center pixel, the neighbourhood pixels are thresholded by its center pixel value and then will be multiplied by powers of two, followed by summation. In this study, a neighbourhood of 8 pixels with radius 1 is used, resulting a total of 2^8 = 256 different labels. A histogram with 256 bin will be generated from the interpolated image and the pixel count of each bin is taken as the FV, thus the FV length for LBP is 256.

Gabor transform: Gabor transform which was introduced by Manjunath and Ma (1996), extracts texture data from an image. From experiment, it is observed that the best parameters for total number of scales, S and orientations, K are 4 and 6, respectively. Gabor transform will produce S×K output images, thus there will be aaggregate of 24-dimensional features for every defect image.

Discrete Wavelet Frame (DWF): In DWF, four wavelet coefficient images with identical dimension as the input image are produced in order to an over-complete wavelet decomposition (Unser, 1995) where the filtered image is considered noneof the sub-sampled. These channels produced the coefficient images such as low-low (L-L), low-high (L-H), high-low (H-L) and high-high (H-H). After that, decompositions process are carry out on the L-L channels, together as with other wavelet transforms. In this work, three-level decompositions with the Haar wavelet is used to produce a 10-dimensional FV.

Gray Level Co-occurrence Matrix (GLCM): This method was introduced by Haralick and Shanmugam (1973) to calculate the co-occurrence matrix of an image. This algorithm evaluate the frequency of a pixel with a particular intensity appears in relation with another pixel j at a specific orientation θ and distance d. In this research work, the co-occurrence matrix is computed with four directions: θ = 0°, 45°, 90° and 135° with distance d = 1.

Classifiers: In this researcher 10 fold cross validation is employed. The extracted instances were randomly dividedinto 10 equal folds. The accuracy rate is computed by the overall number of correct classifications, divided by the number of instances in the database for this research, 5 classifiers are used: The Support Vector Machine (SVM), the Bayesian classifier (Bayes Network, BN); (Pearl, 1988) and the tree-based classifiers (J48, Naive Bayes Tree (NBTree) and Random Forest (RF). These classifiers were chosen from many available due to their good evaluation in many research work found by other researchers in pattern recognition as mentioned in the first section of this study.

Evaluation of the classification: For the assessment of the performance of classification it is evaluated by referring to the following metrics.

Classification accuracy: Capability of the model to correctly calculate the class label of newly inserted unseen data. The accuracy is computed as:

$$\eta = \frac{N_c}{N_t} \quad (1)$$

Where:

N_c = The number of correctly classified tested cases

N_t = The number of tested examples

Misclassification (Error)

Kappa statistic: A chance-corrected assessment between the classifications and the correctly labelled classes. It is computed by calculating the expected agreement from the perceived agreement and divides it by the highest possible agreement. The amountfound has to be larger than 0, else it is measuredas “random guessing”. The explanation of Kappa values is deliberatein (Viera and Garrett, 2005).

ROC Area: The value of the Receiver Operating Characteristic (ROC) area is estimated close to 1.0 for an supreme classification any value <0.5 is consider as “random guessing”.

RESULTS AND DISCUSSION

The evaluations of the classification and feature extraction techniques are summarized in Table 1 and Fig. 4. Table 1listed the overall performances where the compared measures are the classification accuracy, error and Kappa value whilst. Figure 4 illustrates the performances based on the classes (knot, crack, holes and algae) where the TP and ROC rate are measured.

Table 1:Overall classifiers and feature extraction techniques performances

Variables	Accuracy (Number)	Accuracy (%)	Error (Number)	Error (%)	Kappa
CH					
SVM	130	89.6552	15	10.3450	0.7527
BN	131	90.3448	14	9.6552	0.7921
J48	133	91.7241	12	8.2760	0.8176
NBTree	133	91.7241	12	8.2760	0.8175
RF	137	94.4828	8	5.5170	0.8745
CCV					
SVM	131	90.3448	14	9.6550	0.7773
BN	128	88.2759	17	11.7241	0.7456
J48	132	91.0345	13	8.9660	0.8075
NBTree	128	88.2759	17	11.7240	0.7350
RF	129	88.9655	16	11.0350	0.7330
LBP					
SVM	118	81.3793	27	18.6210	0.5521
BN	101	69.6552	44	30.3448	0.3678
J48	102	70.3448	43	29.6550	0.3859
NBTree	109	75.1724	36	24.8280	0.4675
RF	114	78.6207	31	21.3790	0.4929
Gabor					
SVM	123	84.8276	22	15.1720	0.6486
BN	116	80.0000	29	20.0000	0.5931
J48	123	84.8276	22	15.1720	0.6638
NBTree	121	83.4483	24	16.5520	0.6360
RF	124	85.5172	21	14.4830	0.6600
DWF					
SVM	126	86.8966	19	13.1030	0.6917
BN	122	84.1379	23	15.8620	0.6684
J48	116	80.0000	29	20.0000	0.5582
NBTree	124	85.5172	21	14.4830	0.6865
RF	130	89.6552	15	10.3450	0.7673
GLCM					
SVM	128	88.2759	17	11.7240	0.7151
BN	113	77.9310	32	22.0690	0.5412
J48	124	85.5170	21	14.4830	0.6696
NBTree	115	79.3103	30	20.6900	0.5495
RF	128	88.2759	17	11.7240	0.7236

From the overall results obtained, it is observed that the tree-based RF classifier achieved the highest classification accuracy and Kappa value when using CH as feature extraction technique. Its performances are above 85% accuracy for all techniques except LBP (78.6%). CH and CCV are the two best methods where the accuracies for all tested classifiers are close to 90% or above. SVM classifier performs best with CCV and the rest of the classifiers are with CH. Based on the overall results, it is seen that LBP descriptor is not suitable to represent the wood defects features due to the low measures for all the classifiers.

From the analysis of the TP rate values in Fig. 4, it is found that the best feature extraction technique for class ‘Knot’ is CCV for the rest of the classifiers except BN. For case ‘Crack’, only Gabor with BN is able to correctly classify part of the case with TP rate of 0.5. This illustrates that Gabor is a good texture-based method where it could extract the minimum information in the crack images, whilst BN is able to classify minimum case number as case ‘Crack’ is only 2. For class ‘Holes’, the highest rate is

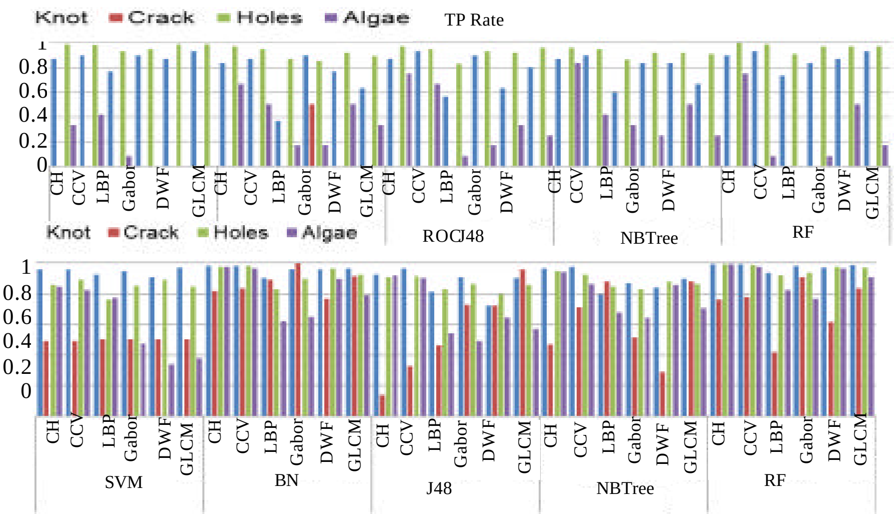

Fig. 4: TP Rate and ROC graphs for each defect class

obtained by CH using RF with TP rate = 1. Lastly, for class ‘Algae’, the best measure (0.833) is still using colour features by CH, with NBTree classifier. From the overall results, it is observed that the best feature extraction technique is CH, whilst the best classifier is NBTree. It is also found that the easiest case to detect is from class ‘Holes’, followed by class ‘Knot’, due to the high number of cases (101 and 30, respectively) as well as the homogeneity of the case appearances.

From the investigation of the ROC values, almost all classifiers obtained >0.9 for class ‘Knot’, except a few in J48 and NBTree. For class ‘Crack’, even though only BN-Gabor gave a TP rate value but more than half of the tested classifier-feature-extraction-technique combination had ROC values >0.7. The scenario suggested that these classifiers are not doing “random guessing”; instead the case cannot be classified correctly due to insufficient data. The BN mainly, being a probabilistic-graphical-model classifier, has 14 out of 24 ROC values above 0.9 and none is below 0.6. For SVM, the classifier is doing “random guessing” for classification of class ‘Crack’ and part of the class ‘Algae’ because the ROC values are 0.5 or lower. While for other classifications the values are above 0.7. For J48 classifier, part of the class ‘Crack’ and ‘Algae’ are randomly-guessed based on the low ROC values whereas other classes obtained high ROC values, especially when using colour-based features. For NBTree and RF classifiers, only three and one features for classification of class ‘Crack’ are doing “random guessing”, respectively with the lowest ROC value (0.288) using NBTree-DWF. The values for classification of other classes are mostly above 0.8; with some cases not <0.64.

CONCLUSION

In this research, we have investigated the performances of six well-known feature extraction technique techniques with five different classifiers in classifying four types of wood defects, namely knot, crack, holes and algae. From a total of 145 defect images, it is found that in overall, Colour Histogram features with Random Forest classifier gave the best classification accuracy (94.48%) with the least error (5.517%) and highest Kappa value (0.8745). This study is a preliminary experiment to our next work on selecting the features based on the evaluation of these results and will be used in a fully automated wood defect detection and classification system.

ACKNOWLEDGEMENTS

The researcher wish to extend heartfelt acknowledgement to the Ministry of Higher Education, Malaysia for supporting this project through the grant approved MMU/RMC-PL/SL/FRGS/2013/002.

REFERENCES

Bezdek, J.C., L.O. Hall and L. Clarke, 1992. Review of MR image segmentation techniques using pattern recognition. Med. Phys., 20: 1033-1048.

Chacon, M. and G. Alonso, 2006. Wood Defects Classification Using a SOM/FFP Approach with Minimum Dimension Feature Vector. In: Advances in Neural Networks, Wang, J., Z. Yi, J. Zurada, B.L. Lu and H. Yin (Eds.). Vol. 3973, Springer, New York, pp: 1105-1110.

Gu, I.Y.H., H. Andersson and R. Vicen, 2010. Wood defect classification based on image analysis and support vector machines. Wood Sci. Technol., 44: 693-704.

Haralick, R.M. and K. Shanmugam, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernetics, 3: 610-621.

Mahram, A., M.G. Shayesteh and S. Jafarpour, 2012. Classification of wood surface defects with hybrid usage of statistical and textural features. Proceedings of the 35th International Conference on Telecommunications and Signal Processing, July 3-4, 2012, Prague, pp: 749-752.

Manjunath, B.S. and W.Y. Ma, 1996. Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell., 18: 837-842.

Ojala, T., M. Pietikainen and D. Harwood, 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29: 51-59.

Packianather, M.S. and B. Kapoor, 2015. A wrapper-based feature selection approach using Bees Algorithm for a wood defect classification system. Proceedings of the 10th System of Systems Engineering Conference, May 17-20, 2015, San Antonio, TX., pp: 498-503.

Pass, G., R. Zabih and J. Miller, 1996. Comparing images using color coherence vectors. Proceedings of the 4th ACM International Conference on Multimedia, November 18-22, 1996, Boston, Massachusetts, USA., pp: 65-73.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco, CA., USA., ISBN: 9781558604797, Pages: 552.

Phan, D.T. and R.J. Alcock, 1998. Automated grading and defect detection: A review. Forest Prod. J., 48: 34-42.

Pietikainen, M., A. Hadid, G. Zhao and T. Ahonen, 2011. Local Binary Patterns for Still Images. In: Computer Vision Using Local Binary Patterns, Pietikainen, M., A. Hadid, G. Zhao and T. Ahonen (Eds.). Vol. 40, Springer, London, pp: 13-47.

Swain, M.J. and D.H. Ballard, 1991. Color indexing. Int. J. Comput. Vision, 7: 11-32.

Unser, M., 1995. Texture classification and segmentation using wavelet frames. IEEE Trans. Image Process., 4: 1549-1560.

Viera, A.J. and J.M. Garrett, 2005. Understanding interobserver agreement: The kappa statistic. Fam Med., 37: 360-363.

Wang, L., L. Li, W. Qi and H. Yang, 2009. Pattern recognition and size determination of internal wood defects based on wavelet neural networks. Comput. Electr. Agric., 69: 142-148.