

Wood defect detection method with PCA feature fusion and compressed sensing

Yizhuo Zhang¹ · Chao Xu¹ · Chao Li¹ · Huiling Yu¹ · Jun Cao¹

Received: 10 September 2014 / Accepted: 4 December 2014 / Published online: 30 April 2015
© Northeast Forestry University and Springer-Verlag Berlin Heidelberg 2015

Abstract We used principal component analysis (PCA) and compressed sensing to detect wood defects from wood plate images. PCA makes it possible to reduce data redundancy and feature dimensions and compressed sensing, used as a classifier, improves identification accuracy. We extracted 25 features, including geometry and regional features, gray-scale texture features, and invariant moment features, from wood board images and then integrated them using PCA, and selected eight principal components to express defects. After the fusion process, we used the features to construct a data dictionary, and realized the classification of defects by computing the optimal solution of the data dictionary in l_1 norm using the least square method. We tested 50 *Xylosma* samples of live knots, dead knots, and cracks. The average detection time with PCA feature fusion and without were 0.2015 and 0.7125 ms, respectively. The original detection accuracy by SOM neural network was 87 %, but after compressed sensing, it was 92 %.

Keywords Principal component analysis · Compressed sensing · Wood board classification · Defect detection

Introduction

Wood defect detection is an important process in wood board manufacture, and its results directly influence the quality of wood products. Wood-defect detection includes

image acquisition, image segmentation, feature extraction, and defect classification (Ruz et al. 2009). In the process of image acquisition, surface information about the wood board (Estévez et al. 2003) was collected by an industrial camera. Pham and Alcock (1998) summarized 32 feature vectors of four types, including windows, shapes, statistical value and gray-scale. Ruz et al. (2009) proposed three methods for feature selection: statistical method, “leave one out” method, and genetic algorithms. The results showed that the genetic algorithm has the best effect. Our previous experiments showed that the following features, including gray-scale, texture, invariant moment, and geometry region, could give a complete representation of the defects (Zhang et al. 2013). However, as various features lead to complex computation and affect detection speed, the application of feature fusion becomes necessary.

Classification is a critical process in defect detection using the MLP neuron network classifier, Pham and Alcock (1999) analyzed the precision of the classification and found that the number of neurons in hidden layers had no obvious influences on the results, yet the final results were greatly affected by the learning rate. Castellani and Rowlands (2009) experimented with decoration board classification, using a neuron network together with a genetic algorithm, but it was only effective in recognizing wood board with a single defect on its surface. When there were two or more types of defects in the same image, this method was even less effective.

Gu et al. (2008) proposed a support-vector machine to classify four kinds of defect, used B-spline to identify the boundary and area of the defects, and chose internal color, edge color, and the external color as defect features in classifying. As the accuracy of the B-spline boundary is questionable, the speed and accuracy of recognition were affected. Zhang et al. (2013) proposed defect detection

The online version is available at <http://www.springerlink.com>

Corresponding editor: Yu Lei

✉ Jun Cao
zdhcj@163.com

¹ Northeast Forestry University, Harbin 150040, China

based on a SOM (self-organizing map) neural network that requires fewer training samples.

For improving the accuracy of wood-board defect detection and overcoming the disadvantages of multi-dimension and computation complexity, we focused on feature fusion and classifier design. Through linear transformation, the PCA method could discover data variety from multi-dimension and reduce the data dimension by preserving features with the biggest contribution in variance. Compressed sensing is a signal processing method proposed by Donoho (2006) and Candes (2006). Signals are compressed or made sparse through proper transformation. By calculating the optimized feature matrix, defect samples are detected. Because complicated training processes are unnecessary for the compressed sensing method, takes less computation time and produces better classification results.

Materials and methods

Materials

We focused on three wood board defects: dead knots, live knots, and wood cracks. The size of the board in our experiment was $40 \times 20 \times 2$ cm. The wood species was *Xylosma*. The experiment was conducted in MatlabR2012, a platform with a 64-bit PC (Corei3, Double-core, 2.25 GHz), and an Oscar F810C IRF camera was used to obtain experiment images. To make the images clearer, two parallel LEDs were used for illumination. In addition, 50 8-bit gray levels images of 128×128 pixels were used for training (20 live knots, 20 dead knots and 10 cracks).

Feature extraction and fusion of wood defects

Feature extraction and fusion is the first step in wood defect identification. The features should include as much defect information as possible and require less calculation work at the same time. First, we extracted 25 features of three types including geometry features, regional features, texture features, and invariant moment. Then, we conducted features normalization, used principal component analysis (PCA) for feature fusion, and selected features with greater contribution for defect identification. The extraction and fusion process is shown in Fig. 1.

We got a complete representation of the defects from 25 features of three types in the wood board images. By extracting and observing these features, the same type of defects had similar feature values and different kinds of defects had different feature values. However, 25 features contained a large number of repeat information with a duplicate expression, when feature numbers increased and

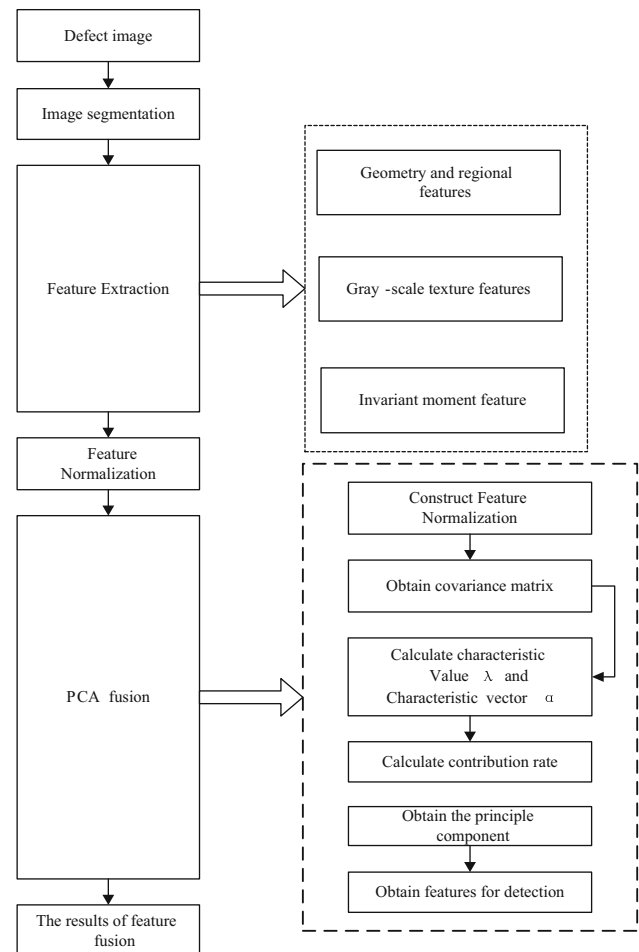


Fig. 1 Flow diagram of feature extraction and fusion

the amount of computation work increased. PCA was implemented to exchange and fuse these features and reduce the number of feature dimensions. Each new feature was obtained by linear combination and transformation of the original features which ensured information preserving of the image. The steps of PCA feature fusion are as follows:

Build the standardize sample matrix X in Eq. 1, where, p is the feature dimension, and n is the sample number ($n > p$).

$$X_{n \times p} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{1,1}, x_{1,2}, \dots, x_{1,p} \\ x_{2,1}, x_{2,2}, \dots, x_{2,p} \\ \vdots \\ x_{i,1}, x_{i,2}, \dots, x_{i,p} \\ \vdots \\ x_{n,1}, x_{n,2}, \dots, x_{n,p} \end{bmatrix}^T, i = 1, 2, \dots, n \quad (1)$$

Use Eq. 2 to standardize transform sample X , and then obtain standardized matrix Z .

$$z_{i,j} = \frac{x_{i,j} - \bar{x}_j}{s_j}, i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (2)$$

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{i,j}}{n}, s_j^2 = \frac{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)^2}{n-1} \quad (3)$$

Calculate covariance matrix R of standardized matrix Z .

$$R = \frac{Z^T Z}{n-1}, \text{ and } r_{i,j} = \frac{\sum_{k=1}^p \sum_{l=1}^n z_{j,k} \cdot z_{k,i}}{n-1}, k = 1, 2, \dots, n; i, j = 1, 2, \dots, p \quad (4)$$

Calculate eigenvalue λ and eigenvector α of characteristic equation $|R - \lambda E| = \vec{0}$ with covariance matrix R .

Rerank the eigenvalue by descending order and obtain $\bar{\lambda}$, calculate the contribution and cumulative contribution of each principal component by Eqs. 5 and 6.

The contribution of each principal component is Eq. 5:

$$n_i = \frac{\bar{\lambda}_i}{\sum_{j=1}^p \bar{\lambda}_j}, i = 1, 2, \dots, p \quad (5)$$

The cumulative contribution is Eq. 6:

$$m_i = \frac{\sum_{j=1}^i \bar{\lambda}_j}{\sum_{j=1}^p \bar{\lambda}_j}, i = 1, 2, \dots, p \quad (6)$$

Choose the first k principal components which cumulative contribution can reach the pattern recognition need and transform matrix E is obtained by Eq. 7:

$$E = [\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_k] \quad (7)$$

Calculate final principal components Y , and Y is the final input of the defect classifier.

$$Y = E \times X \quad (8)$$

Defect detection based on compressed sensing

In applying compressed sensing to wood classification, we used optimized feature vectoring as the sample sequence and training samples to create a data dictionary, and tested samples linearly by training the samples. When the sparse representation vector of test samples in the data dictionary was calculated by solving optimization problem under the l_1 norm, the classification result was obtained.

With compressed sensing, when a signal is sparse in certain transformation domains, an observation matrix, which is irrelevant to the transformation basis, will project the multi-dimensional information obtained by the transformation into low-dimensional space. Then, by solving a

convex optimization problem, an original signal is reconstructed from these few projections of high probability (Donoho 2006; Shi et al. 2009).

First, assume x as one-dimension discrete time signal of a real value with limited length, and can be used as a sequence vector of $n \times 1$ dimensions. If matrix ψ and vector α exists and Eq. 9 is meaningful, then x is sparse in domain ψ .

$$x = \psi \alpha \quad (9)$$

where, ψ is the orthogonal transformation basis called a sparse matrix, and $\psi \in R^{n \times n}$. The transformation coefficient of x in domain ψ , α , is $\alpha \in R^{n \times 1}$, the number of nonzero values is far less than the number of signal dimensions.

If the signal is projected onto a matrix of ϕ , which has no relationship with the transformation basis, the observed signal y is obtained through Eq. 10.

$$y = \phi x = \phi \psi \alpha \quad (10)$$

where, ϕ is the observation matrix, and $\phi \in R^{m \times n}$: y is the observation vector, and $y \in R^{m \times 1}$.

Finally, by deriving the optimized l_0 norm of Eq. 11, we obtain α_l , which is the exact or approximate solution of α .

$$\min \|\alpha_l\|_0 \text{ s.t. } y = \phi x = \phi \psi \alpha \quad (11)$$

Equation 11 is an undetermined equation. According to compressed-sensing theory, if the signal is sparse enough, the question of minimum can be transformed into deriving the l_1 norm of Eq. 12, a process of convex optimization. By solving the problem of linear programming, the original signal is reconstructed from N observation values.

$$\min \|\alpha_l\|_1 \text{ s.t. } y = \phi x = \phi \psi \alpha \quad (12)$$

If f_i^j is the feature vector of image j of wood type i , and f_i^j is one row of training samples, then $A_i = [f_i^1, f_i^2, \dots, f_i^m]$, $A_i \in R^{n \times N}$ represents samples of wood type i . The f_i^j in this matrix is the feature vector, and $f_i^j \in R^{n \times 1}$, m is the number of training samples of wood type i .

The data dictionary matrix composed of three types of training samples is shown in Eq. 13:

$$A = [A_1, A_2, A_3] \quad (13)$$

If the number of training samples of wood type i is adequate, then the feature vector of test image y can be represented by linear combination of training samples A_i which belong to wood type i , that is:

$$y = \alpha_i^1 f_i^1 + \alpha_i^2 f_i^2 + \dots + \alpha_i^m f_i^m = A_i \alpha_i \quad (14)$$

where, y is the feature vector of the test image, and $y \in R^{n \times 1}$; the α_i here is the vector of linear representation coefficient, and $\alpha_i \in R^{m \times 1}$.

If we apply the above equations to the whole data dictionary matrix A , then:

$$y = A\alpha \quad (15)$$

where, α is sparse vector, and N is the total amount of samples.

$$\alpha = (0 \dots 0, \dots, \alpha_i^1, \alpha_i^2, \dots, \alpha_i^m, \dots, 0 \dots 0)^T, \alpha \in R^{n \times 1} \quad (16)$$

If the test samples are of type i , except for the only m data that represent the feature of wood type i , then all other data in vector α are 0. In other words, as the number of values which are not 0 in α is much less than the number of signal dimensions, α is a sparse vector, and the above process is considered the sparse decomposition of test samples.

The classification process of unknown test samples is as follows: put test sample feature y into Eq. 16 in which $y \in R^{n \times 1}$ and $A \in R^{n \times N}$, and acquire sparse vector α by solving Eq. 16. Here, Eq. 16 is an undetermined system of equations, and α vector is a sparse vector. According to compressed-sensing theory, the exact solution or approximate solution of α can be obtained by solving the optimization problem of Eq. 17 of l_1 norm. The ε in this Eq. 17 is the error threshold. In actual application, the sample type is determined by the non-zero item in α_l .

$$\alpha_1 = \arg \min \|\alpha\|_1 \quad s.t. \quad \|A\alpha - y\| \leq \varepsilon \quad (17)$$

Results and discussion

Classification steps

The defect classification experiment is shown in Fig. 2, including image collection, morphology segmentation, feature extraction, feature fusion, classifier design, and result assessment.

The defect images are read by Matlab12; for example, Figs. 3, 4, 5 represent live knot, dead knot, and crack, respectively. To reduce the computation work and increase the speed of computation, all color images are transferred into gray-scale images and changed to standard pictures of 128×128 pixels.

Mathematical morphology is an image-processing method based on geometry with the advantages of continuous image skeleton, fewer breaking points, and rapid,



Fig. 3 Live knot



Fig. 4 Dead knot

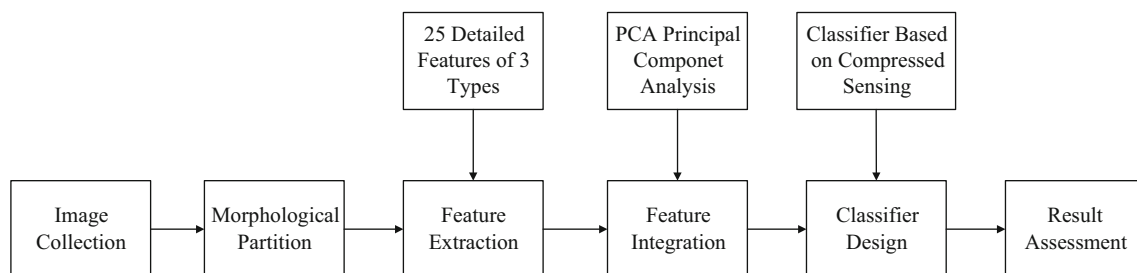


Fig. 2 The flow diagram of defect classification



Fig. 5 Crack

exact image segmentation (Zhang et al. 2014a, b). Using this method, the exact defect targets are separated from the background. The results of segmentation are shown in Figs. 6, 7, 8.

Calculate 25 features when segmentation is over and then normalize them. According to Eqs. 1–4, PCA maps high dimension features to low dimension spaces, and find eigenvalue λ and eigenvector α by covariance matrix R . $\bar{\lambda}$ is obtained by re-ranking eigenvalue λ in descending order.

The variance of data is reflected by corresponding eigenvalues. Over different spaces with the same dimensionality, the space spanned by the eigenvectors corresponding to the larger eigenvalues carries the most variance. From Eqs. 5 and 6, contribution of each principal component n_i and cumulative contribution m_i can be obtained (Table 1).

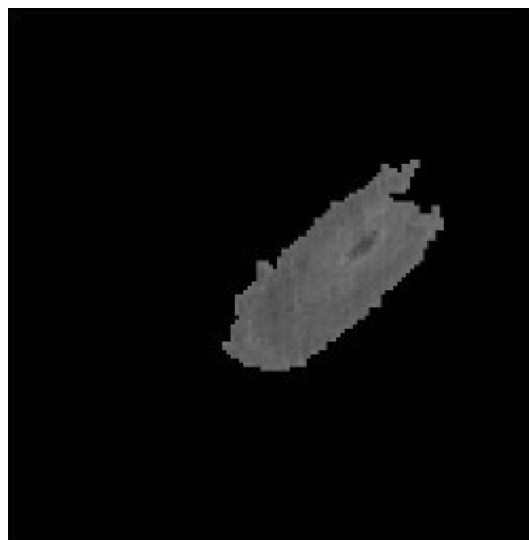


Fig. 6 Segmentation result of live knot

training samples. The data dictionary A of three types of defects is shown as follows:

$$A = \begin{bmatrix} 0.0196 & -0.0209 & 0.0026 \\ -0.0554 & 0.0953 & -0.0798 \\ 0.1942 & 0.0494 & -0.4873 \\ -0.0310 & -0.1350 & 0.3320 \\ -0.0254 & 0.0730 & -0.0951 \\ 0.0469 & -0.2613 & 0.4287 \\ 2.1204 & -1.3704 & -1.5000 \\ 0.1216 & 2.4125 & -5.0683 \end{bmatrix} \quad (18)$$

Figures 3, 4 and 5 are employed as testing samples. Extract 25 features from the segmented images of Figs. 6, 7, 8, and then calculate the principal components from defect features by PCA transformation. The principal components of the three test samples are represented by h^T , s^T , and l^T , respectively, as follows:

$$b_i^T = \begin{bmatrix} h^T \\ s^T \\ l^T \end{bmatrix} = \begin{bmatrix} -0.6562 & 1.1239 & 1.3018 & 0.4471 & -1.5050 & 0.4266 & 2.2168 & -1.3436 \\ 0.7192 & -0.2315 & 1.4465 & 0.3311 & 3.1868 & 0.8269 & -1.9511 & 2.0005 \\ 0.1405 & 0.5189 & 0.0203 & 1.0183 & 1.6230 & 2.1236 & -3.9129 & -4.8866 \end{bmatrix} \quad (19)$$

The top eight principal components can reach a cumulative contribution of more than 95 %; therefore, those components are selected as the input of the classifier.

After the determination of principal components, calculate its means and build the data dictionary A by the 50

Implement classification in accordance with Eq. 17, and obtain $\alpha_{A_i}^T$ with the least square method:

$$\alpha_{A_i}^T = \begin{bmatrix} \alpha_h^T \\ \alpha_s^T \\ \alpha_l^T \end{bmatrix} = \begin{pmatrix} 1.2905 & 0 & 0 \\ 0 & 1.1643 & 1.1361 \\ 0.2891 & 0.1522 & 1.5256 \end{pmatrix} \quad (20)$$

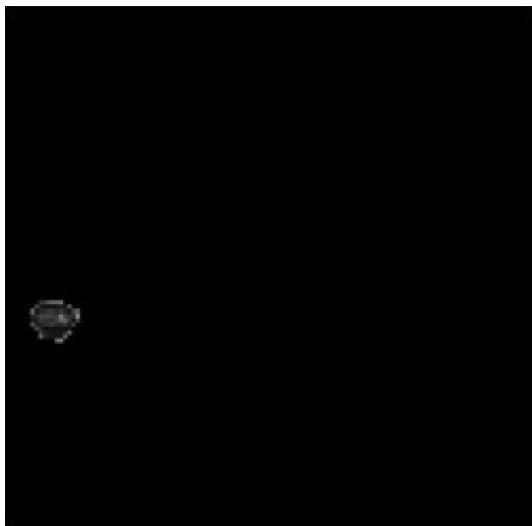


Fig. 7 Segmentation result of dead knot



Fig. 8 Segmentation result of crack

The sample type is determined by the maximum value in $\alpha_{A_i}^T$. According to $\alpha_{A_i}^T$, we see that the classification results of Figs. 3, 4, 5 are live knot, dead knot and crack, respectively.

The effective test of the PCA feature fusion

To verify the necessity of feature selection, we carried out defect-detection comparison tests between the PCA feature-fusion and variance selection methods (Peck and Devore 2005). We used 50 sample images of live knots, dead knots, and cracks for feature selection and

Table 1 Contribution of each principal component

No.	Individual contribution rate	Accumulative contribution rate
1	0.4608	0.4608
2	0.1671	0.6279
3	0.1622	0.7901
4	0.0639	0.8540
5	0.0464	0.9004
6	0.0292	0.9296
7	0.0174	0.9470
8	0.0140	0.9610
9	0.0121	0.9731
10	0.0062	0.9793
11	0.0054	0.9847
12	0.0050	0.9897
13	0.0030	0.9927
14	0.0028	0.9954
15	0.0014	0.9968
16	0.0012	0.9980
17	0.0008	0.9988
18	0.0006	0.9994
19	0.0002	0.9996
20	0.0002	0.9998
21	0.0002	1.0000
22	0.0000	1.0000
23	0.0000	1.0000
24	0.0000	1.0000
25	0.0000	1.0000

Table 2 Result of feature comparison

Feature selection method	Correct recognition	False recognition	Recognition rate (%)	Average classification time (ms)
Not selected	34	16	68	0.7125
Deviation selection	41	9	82	0.0861
PCA fusion	46	4	92	0.2015

classification. In the process of variance selection, we chose features according to their variances for the between-sample dispersion and divisibility, as determined by feature variance. The classification results of the variance selection and PCA methods are shown in Table 2.

In Table 2, the recognition rate without the feature selection step is 68 %, and the time required for recognition is 0.7125 ms. The PCA method has the best recognition rate at 92 %, and its recognition time is 0.2015 ms. Therefore, the feature selection can not only reduce identification time, but also increase the recognition rate.

Table 3 The comparison with SOM neural network method

Classification method	Recognition time (ms)				Recognition rate (%)
	Image segmentation	Feature extraction	Classification	Total	
Neuron network	27.476	19.745	0.7125	51.839	87
Compressed sensing	28.507	17.002	0.202	54.120	92

The classification test of compressed sensing classifier

To test the performance of the classification method proposed in our study, we used the neural network classifier (Candes 2006). As the SOM neural network requires fewer training samples with higher classification accuracy, SOM was compared with compressed sensing in our experiment. Live knots, dead knots, and cracks for 50 test images were classified, and the accuracy and time of classification are shown in Table 3.

As shown in Table 3, step-by-step iterative computation is necessary in the process of SOM classification: each step may influence the computation results, so the SOM classifier has limits on recognition and time consuming. However, the wood defect recognition based on compressed sensing doesn't require complex computation. The time required for recognition is significantly reduced, while the exactness of recognition has improved by 5 % over the SOM classifier.

Conclusion

Focusing on the complexity of wood-board surface defect information, we proposed a new defect feature fusion method by performing PCA on the high-dimension features. Then, we built a compressed sensing classifier to construct the data dictionary of typical samples, and obtained an optimized solution using the least square method. The results of simulation experiments reveal that the PCA fusion method can give a more complete representation of the defect information. Compared with the SOM neural network algorithm, the compressed sensing classifier has several advantages: fewer parameters, better flexibility, less computation time, and higher classification exactness. Therefore, the defect classification algorithm based on PCA fusion and compressed sensing can effectively increase the speed and exactness of wood-defect detection.

Acknowledgments This work was financially supported by the Fund of Forestry 948 Project (2011-4-04), the Fundamental Research Funds for the Central Universities (DL13CB02, DL13BB21), and the Natural Science Foundation of Heilongjiang Province (C201415)

References

- Candes E (2006) Compressive sampling. *Proc Int Congr Math* 3:1433–14521
- Castellani M, Rowlands H (2009) Evolutionary artificial neural network design and training for wood veneer classification. *Eng Appl Artif Intell* 22:732–741
- Donoho D (2006) Compressed sensing. *IEEE Trans Inf Theory* 52(4):1289–1300
- Estévez PA, Perez CA, Goles E (2003) Genetic input selection to a neural classifier for defect classification of radiata pine boards. *For Prod J* 53(7/8):87–94
- Gu IYH, Andersson H, Vicen R (2008) Automatic classification of wood defects using support vector machines. In: *Computer vision and graphics: lecture notes in computer science*, vol 5337. Springer, Berlin, pp 356–367
- Peck R, Devore JL (2005) *The exploration and analysis of data*. Duxbury Press, Belmont, pp 611–662
- Pham DT, Alcock RJ (1998) Automated grading and defect detection: a review. *For Prod J* 48(3):34–42
- Pham DT, Alcock RJ (1999) Automated visual inspection of wood boards selection of features for defect classification by a neural network. *Proc Inst Mech Eng Part E* 213(4):231–245
- Ruz GA, Estevez PA, Ramirez PA (2009) Automated visual inspection system for wood defect classification using computational intelligence techniques. *Int J Syst Sci* 40(2):163–172
- Shi GM, Liu DH, Gao DH, Liu Z, Lin J, Wang LJ (2009) Advances in theory and application of compressed sensing. *Acta Electron Sin* 37(5):1070–1081
- Zhang YZ, Cao J, Xu L, Yu HL (2013) Wood floor defects segmentation and recognition based on morphological and SOM. *Electr Mach Control* 17(4):116–120
- Zhang YZ, Liu SJ, Cao J, Li C, Yu HL (2014a) A rapid, automated flaw segmentation method using morphological reconstruction to grade wood flooring. *J For Res* 25(4):959–964
- Zhang YZ, Xu L, Ding L, Cao J (2014b) Defects segmentation for wood floor based on image fusion method. *Electr Mach Control* 18(7):113–118