FULL-LENGTH REPRINT – differs from the original in layout but not in contents
The copyright owner of the publication is Springer.
The publication is available at https://doi.org/10.1007/978-3-319-69989-9_15. Cite as:
B. Świderski, M. Kruk, G. Wieczorek, J. Kurek, K. Śmietańska, L.J. Chmielewski, J. Górski, and A. Orłowski. Feature selection for 'orange skin' type surface defect in furniture elements. In L. Rutkowski et al., editors, *Proc. Int. Conf. on Artificial Intelligence and Soft Computing ICAISC 2018, volume 10842 of Lecture Notes in Artificial Intelligence*, pages 81-91, Zakopane, Poland, 3-7 Jun 2018. doi:10.1007/978-3-319-91262-2_8. Recompiled for technical reasons on August 3, 2018.

Feature Selection for 'Orange Skin' type Surface Defect in Furniture Elements

Bartosz Świderski[1], Michał Kruk[1], Grzegorz Wieczorek[1], Jarosław Kurek[1], Katarzyna Śmietańska[2], Leszek J Chmielewski[1], Jarosław Górski[2], and Arkadiusz Orłowski[1]

[1] Faculty of Applied Informatics and Mathematics – WZIM
[2] Faculty of Wood Technology – WTD
Warsaw University of Life Sciences – SGGW
ul. Nowoursynowska 159, 02-776 Warsaw, Poland
{bartosz_swiderski,arkadiusz_orlowski}@sggw.pl
http://www.wzim.sggw.pl, http://www.wtd.sggw.pl

Abstract. The surfaces of furniture elements having the *orange skin* surface defect were investigated in the context of selecting optimum features for surface classification. Features selected from a set of 50 features were considered. Seven feature selection methods were used. The results of these selections were aggregated and found consistently positive for some of the features. Among them were primarily the features based on local adaptive thresholding and on Hilbert curves used to evaluate the image brightness variability. These types of features should be investigated further in order to find the features with more significance in the problem of surface quality inspection. The groups of features which appeared least profitable in the analysis were the two features based on percolation, and the one based on Otsu global thresholding.

Keywords: Feature selection, surface defect, orange skin, detection, furniture, brightness variability

1 Introduction

Visual inspection seems to remain the main method of assessing quality in the furniture industry. Despite that the machine vision methods have definitely attained their maturity and became the part of routine industrial processes, in the furniture industry this is still not the case.

To our best knowledge, the literature status for orange skin or orange peel did not change since our previous review [6]. In brief, we have stated that what concerns furniture elements quality inspection, the image processing methods are very rarely if not at all mentioned in the literature. The existing references mention the defects of furniture only in the context of more general domains [11], or the raw material quality is of main concern [20]. In fact, in very few papers the *orange skin* called also *orange peel* as a surface defect is considered directly.

L. Rutkowski et al. (Eds.): ICAISC 2018, LNAI 10842, pp. 81-91, 2018.
DOI: 10.1007/978-3-319-91262-2_8

Konieczny et al. [12] generate the images of *orange peel* artificially and study the visibility of this defect for humans in various conditions. Armesto et al. [2] describe a system of moving lighting and cameras. It is designed to improve the results of quality inspection of painted surfaces in the car industry. The target of this system is to enable the *defect augmentation phenomena* as the authors call the processes the light causes on surface features of various kinds. Among the broad class of surface defects the *orange peel* is present. The system makes it possible to use local adaptive thresholding as the only detection method.

Besides the papers, there are numerous patents in which methods are shown to avoid or remove orange skin in the painting process; let us name just one by Allard et al. [1] as an example. In none of these patents the image analysis methods are recalled.

On the contrary to the furniture industry, in the timber industry the image-based analysis of structural and anatomical defects is a well established technology with broad literature (cf. the reviews [3,19]).

It seems reasonable that in the preliminary stage in which the surface inspection in furniture industry is now, one of the main questions is the issue of finding proper image features which could capture the visual phenomena that make the surface look *good* or *bad*, in the context of *orange skin*. It is possible to take generic features like local spectral features, wavelet features or others, like for example the Histogram of Oriented Gradients, which we did in one of our previous papers [6]. What is more interesting, however, is to learn what features of the surface are important in the problem of our concern, and the real challenge would be to discover the meaning of these features in the problem.

In this paper we shall consider a set of features already found effective in a series of various demanding applications: classification of dermoscopic images of melanoma [13], regions in mammographic images [15,24] and images of microorganisms in soil [15]. Preliminary tests with feature selection made on this set have been previously made [14], and one of the methods of finding some order in the set which could make it easier to perform feature selection in a bottom-up manner was used. The features were ordered according to their Fisher measure of information content. They were added sequentially and the process was stopped when the classification accuracy attained its maximum. It must be admitted that drawing any conclusion from the result of this single experiment would be premature.

We shall use a set of seven feature selection methods chosen from those described by Pohjalainen et al. [22]. We shall check which features were selected the most frequently and finally we shall try to look at those features more carefully to discover how their design made them useful in the application of our interest. This will go far beyond the preliminary experiments presented previously [5] where we found that *orange skin* can be detected with simple differentiation and thresholding of the image intensity function.

The remaining part of this paper is organized as follows. In the next Section the surface defect considered will be recalled, and the way the defect is seen in the images and the method with which the objects for classification are formed

will be presented. In Sec. 3 the features and the classifier will be described. Section 4 will be devoted to the problem of feature selection. The results from seven feature selection methods will be presented in a combined way. These results will be briefly discussed in Sec. 5. Conclusions and outlook for further work will close the paper.

2 Classified objects

2.1 The defect: *orange skin*

The defect called *orange skin* or otherwise *orange peel* can appear on lacquered surfaces. In furniture elements it is one of the reasons for reduced esthetic quality of the product. It can emerge as small hollows, that is, an uneven structure of the hardened surface. The depth of hollows is smaller than the thickness of one lacquer layer so its order of magnitude is tenths parts of a millimeter. Numerous reasons can cause the defect to appear: insufficient quantity or bad quality of dilutent, large difference of temperatures between the lacquer and the surface, improper pressure or distance of spraying, excessive air circulation during spraying or drying, and insufficient air humidity. The structure of wood is hidden under the lacquer in the analysis.

Because the reason for classification of the surface as good or bad is of esthetic nature and because not only the presence of hollows but also their relations are important, it is not possible to indicate a well-defined *defect* on the surface, like it would be for example in the case of a crack or scratch. The parts of the surface which *look good* gradually pass to those which *look bad*. The *good* surface is not free from texture and has some deviations from planarity. Example images will be shown in the next Section.

2.2 Images and objects

In the present paper we have used the same set of images we analyzed in a previous paper [14]. The images were taken with the Nikon D750 24 Mpix camera equipped with the Nikon lens F/2.8, 105 mm. The distance from the focal plane to the object surface was 1 m and the optical axis of the camera was normal to the surface. The lighting was provided by a flash light with a typically small light emitting surface, located at 80 cm from the object, with the axis of the light beam deflected by 70° from the normal to the surface. In this way, the light came from a direction close to parallel to the surface, to emphasize the surface profile. The camera was fixed on a tripod and it was fired remotely to avoid vibration.

The objects were painted with white lacquer in a typical technological process. The surfaces imaged belonged to several different objects. The surfaces were classified by the furniture quality expert into three classes: *very good*, *good* and *bad* in the terms of the orange skin defect, before the experiment. The photographs were made of a part of the object which was not farther than 30 cm from the center of the image. An image of a part of an elongated object was

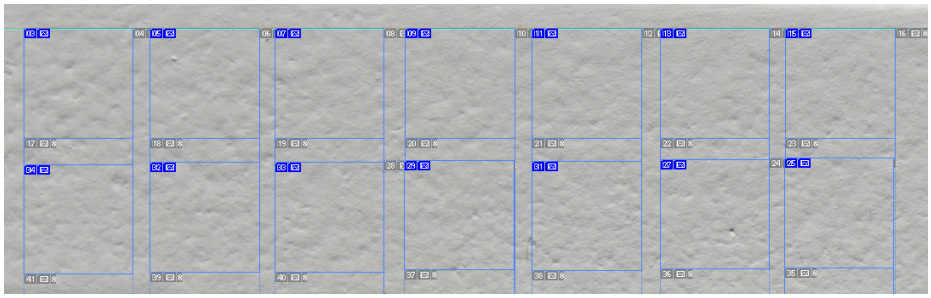


Fig. 1. Example of images of the surface of furniture elements. Small images of size 300×300 like those outlined with blue lines and marked with small dark blue icons were cut for the training and testing processes.

taken once at a time, then the object was moved and a next image was taken, to include all of the surface of the objects in the experiment. The images were made in color mode and stored losslessly.

From these images, small non-overlapping images were cut, each of them of size 300×300 pixels. There were 900 such images total. Each of these images was treated as a separate object and was classified independently of the other images.

From these objects, the training and testing sets were chosen for cross-validation. Ten cross-validation rounds were planned. In each round, there were 90 images in the testing set, selected randomly from the set of images, with equal numbers of images belonging to each class. The remaining 810 images formed the training set in the given round. The numbers of images belonging to the classes resulted from the classification made by the human expert and were close to equal, but not precisely equal.

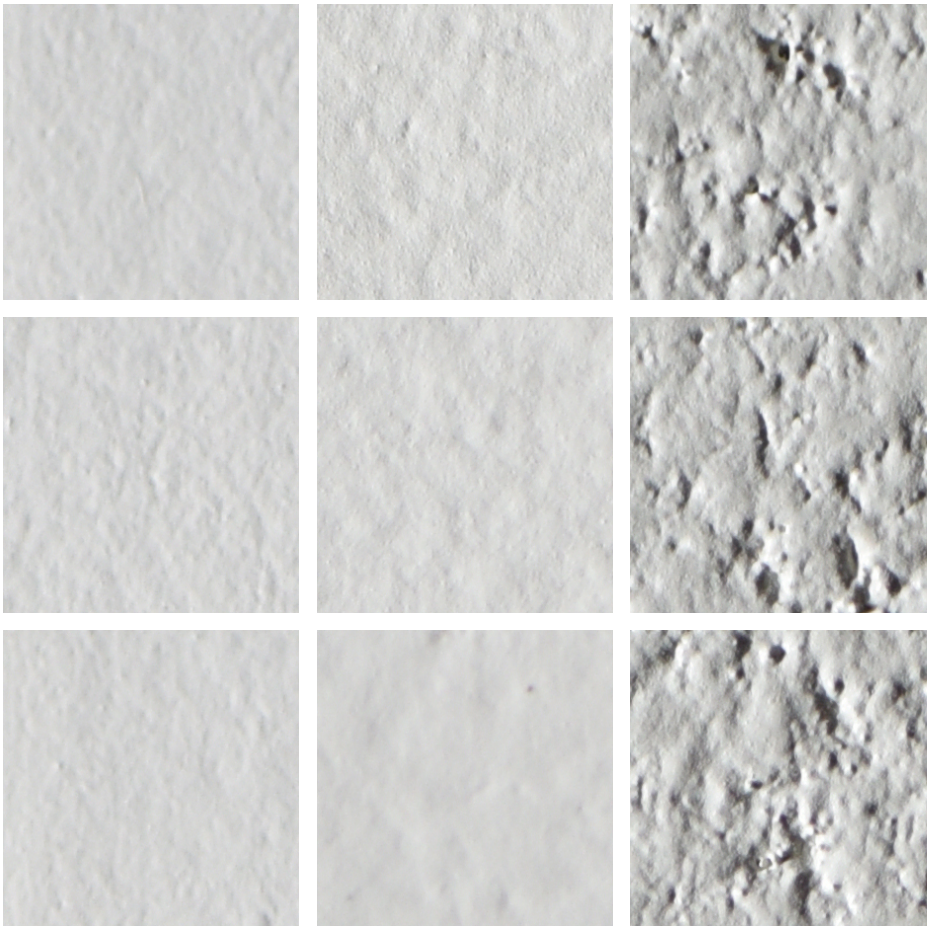
The way the small images were cut can be seen in Fig. 1. The examples of images belonging to the three classes selected in the experiment are shown in Fig. 2. A very good surface has a fine and even texture. A bad surface has an uneven and strong texture. A good surface is everything in the middle. Note that a good surface can differ from a very good one by that its texture is less even, although it can be weaker, like that in the images in the lowest row of Fig. 2.

3 Features and classification

3.1 Features

All the features were generated from the luminance component Y of the YIQ color model, $Y \in \{0, 1, ..., 255\}$.

The features for each small image were formed with the methods listed below. There were 50 features in total. The ranges of indexes of features are given in boldface:



a **b** **c**

Fig. 2. Examples of images of the surfaces belonging to three classes: (**a**) very good, (**b**) good and (**c**) bad.

01-01: number of black fields after thresholding with Otsu method – 1 feature;
02-08: Kolmogorow-Smirnow features [23] – 7 features;
09-14: maximum subregions features [13] – 6 features;
15-16: features based on the percolation [13] – 2;
17-32: features based on the Hilbert curve [15,24] – 16;
33-41: features from single-valued thresholding [14] (explained below) – 9;
42-50: features from adaptive thresholding [14] (explained below) – 9.

The single-valued thresholding was performed as follows. The image was thresholded, in sequence, with thresholds: $i/10 \times 255, i = 1, 2, ..., 9$. The nine features are the numbers of black regions after each thresholding.

The adaptive thresholding was performed as follows. Let A be the image after applying the averaging filter with the window 20×20 pix. Then, the number I_2 is calculated as $I_2 = A - Y - C$, where C is a constant. The result is thresholded at 0.1×255. The feature is the number of white regions in the image I_2, giving 9 features. It remains to set the constant C. Setting the constant corresponds in fact to modifying the threshold. To scan a range of thresholds, nine values were taken, $C = i, i = 1, 2, \ldots, 9$, giving nine features. A white region corresponds to a dark blob in the image.

3.2 Classifier

As the classifier, the set of three SVM classifiers [7] for three pairs of classes was used. The classifiers voted for the final class assignment. The SVM was selected for this study due to it is one of the most widely used classifiers with very good performance in many cases. The focus of this paper is set on features, so at this stage we reduced the number of variables in the experiment and resigned from considering multiple classifiers. The version and parameters used were: radial-basis function kernel, cost $c = 300$, $\sigma = 0.1$.

4 Training and feature selection

4.1 Feature selection methods

To find the globally optimal set of features, an exhaustive search within the set of available features should be performed, which is totally impractical (in the set of 50 features there are $2^{50} - 1 > 10^{16}$ different nonempty subsets). Therefore, we used seven feature selection methods chosen from those previously described [22]. The software for these methods is publicly available [21]. The methods are:

1. method based on Chi square test (Chi2) [16],
2. Correlation-based feature selection (CFS) [10],
3. Fast Correlation-based Filter (FCBF) [17,18],
4. method based on Fisher score (FS) [9],
5. method based on Information Gain (IG) [8],
6. Sparse Multinomial Logistic Regression via Bayesian L1 Regularization (SMLR) [4],
7. Kruskal-Wallis variance test [25].

With each method, the features were selected on the basis of just one of the ten training sets. Then, these features were used in the cross-validation process to assess the accuracy of classification.

4.2 Accuracy of classification

Accuracies of classification attained with the methods are shown in Fig. 3. They are sorted according to the average accuracy attained. It happened that the

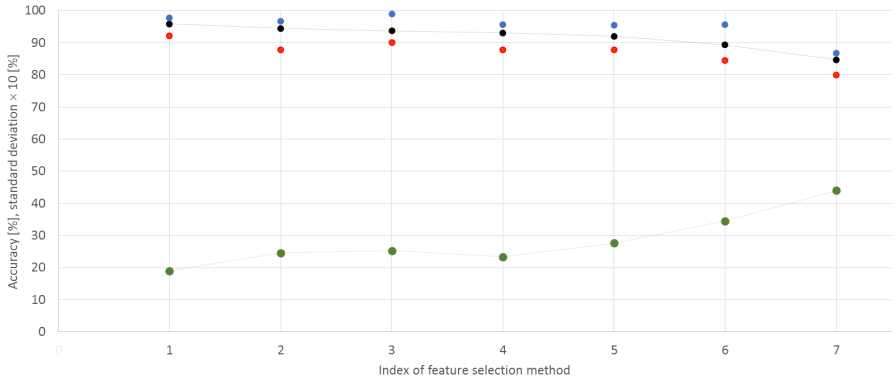


Fig. 3. Accuracy of classification for feature selection methods attained in the cross-validation process, sorted according to descending average accuracy. Indexes of feature selection methods comply with those in Tab. 1. Upper graphs: average accuracy (black ■ dots with line); maximum (blue ■) and minimum (red ■) errors attained in the cross-validation series; lower graph: standard deviation (green ■ dots with line) of these results, 10×enlarged for better visibility. Lines connecting points related to methods have no meaning except for indicating the correspondence between these points.

standard deviation of errors increased nearly monotonically together with the decrease of accuracy. Consequently, the method Chi square appeared the best both in relation to its high accuracy as well as low accuracy deviation.

The values of average, maximum and minimum accuracies obtained in the ten cross-validation rounds are shown in Table 1. The best average accuracy slightly exceeds 95 % and the minimum accuracy, which is the pesimistic estimate of actual accuracy, is slightly larger than 92 %. This indicates that indeed some more work should be done to look for better features.

Table 1. Results of feature selection and accuracy measures attained. Rows sorted according to descending average accuracy (in **bold font**).

#	Method	Acronym	Accuracy [%] avg	min	max	sdv	No. fea.	5 best features (if applicable)
1	Chi Square	Chi2	**95.88**	92.22	97.78	1.90	41	29 28 43 18 41
2	Information Gain	IG	**94.43**	87.78	96.67	2.46	26	43 29 28 41 18
3	Fisher Score	FS	**93.65**	90.00	98.89	2.53	26	29 24 18 20 13
4	Correlation-based Feature Selection	CFS	**93.09**	87.78	95.56	2.33	14	
5	Kruskal-Wallis Variance Test	KWVT	**91.98**	87.78	95.51	2.76	17	43 41 39 6 5
6	Sparse Multinomial Logistic Regression	SMLR	**89.31**	84.44	95.56	3.45	10	43 13 10 40 38
7	Fast Correlation-Based Filter	FCBF	**84.75**	80.00	86.67	4.40	5	43 28 23 14 36

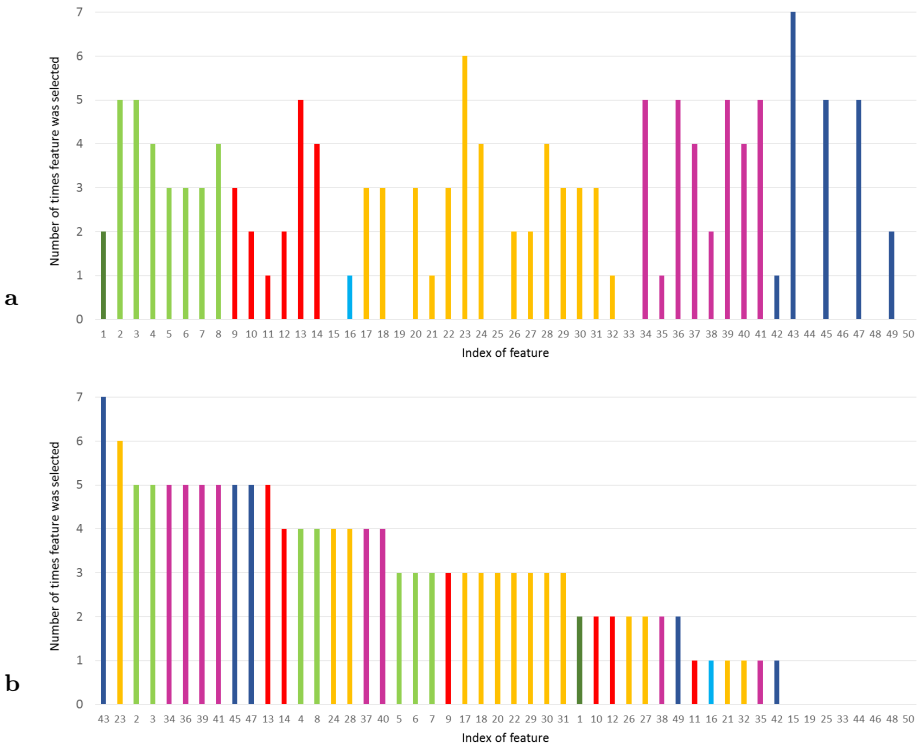


Fig. 4. Cumulated results of feature selection. (**a**) Features sorted according to the groups of features; (**b**) features sorted according to the number of times of being selected. Groups of features marked with colors (cf. enumeration in Sect. 3.1): **01-01:** dark green, Otsu, 1 feature; **02-08:** green, Kolmogorow-Smirnow, 7 features; **09-14:** red, maximum subregions, 6; **15-16:** blue, percolation, 2; **17-32:** yellow, Hilbert, 16; **33-41:** violet, single-valued thresholding, 9; **42-50:** dark blue, adaptive thresholding, 9 features.

4.3 Feature selection results

In Table 1, alongside with the accuracies, the five best features are shown for those methods in which the single features are assigned a measure of quality in a natural way[3]. More insight in the results of feature selection can be gained by examining the histograms in Fig. 4 which show the number of times a given feature was selected in the seven feature selection methods used. It can be seen that some features were selected in more feature selection algorithm, and some in less of them. Some features were not selected at all.

[3] This does not concern CFS, where features are not sequenced; in this method, the following features were selected: $\{2, 3, 13, 14, 23, 24, 28, 34, 39, 40, 41, 43, 45, 47\}$.

5 Discussion

The graphics in Fig. 4 indicate that the feature 43 was always chosen (7 times) and the next most frequently chosen feature was 23. Other features belonging to nearly all groups of features were chosen 5 times. On the opposite end, features like $15, 19, 25$, etc. were not chosen at all. Both the most frequently chosen features and the least frequently chosen ones belong to various groups. No general conclusion concerning the groups of features can be drawn.

The most frequently chosen feature with the index 43 is the second feature from the set of adaptive thresholding-based features. This indicates that adaptive thresholding is a viable method in looking for good features in our task. Other features from this group were chosen as well, but the thresholds used in them differed by more than one (except index 1 and 2). This means that serial thresholding with different threshold has its merit.

The second best feature has the index 23 and it is the sevenths feature based on Hilbert curves. The features from this group are using the following scheme. Two strings of pixels located on subsequent fragments of the Hilbert curve are compared. In the comparison, the Kolmogorow-Smirnow statistics S_{KS} is used together with its minimum significance level p. The larger S_{KS} and p, the more the pixels belonging to the two strings are different. The strings are moved along the curve. In each location a new pair (S_{KS}, p) is found. Several measures of the amount of variation in the image can be built for these pairs. The sevenths Hilbert feature is $\mathrm{std}(p)/\mathrm{mean}(p)$. The fact that this feature was selected so frequently seems to indicate that the statistical measures of brightness variability in which the mapping between the image surface and some function with which this surface can be mapped into a 1-dimensional curve perform well in texture analysis. This can be interpreted as an encouragement to investigate more the features based on such a concept.

The group of features which should probably not be investigated more are the two features based on percolation, and the one based on Otsu global thresholding.

The results were obtained with a single classifier.

6 Summary and prospects

Features selected from a set of 50 features were used to classify the surfaces affected by the *orange skin* surface defect. The features were selected with seven methods. The results of these selections were consistently positive for some of the features. Among them were primarily the features based on local adaptive thresholding and on Hilbert curves used to evaluate the image brightness variability. These types of features should be investigated further in order to find the features with more significance in the problem of surface quality inspection.

It is planned to extend the experiments with the *orange skin* defect in furniture by studying more images taken in slightly varying conditions to further test the stability of the obtained results. The extension to more than one classifier is also considered.

References

1. Allard, M., Jaecques, C., Kauffer, I.: Coating material which can be thermally cured and hardened by actinic radiation and use thereof (Sep 27 2005), US Patent 6,949,591
2. Armesto, L., Tornero, J., Herraez, A., Asensio, J.: Inspection system based on artificial vision for paint defects detection on cars bodies. In: 2011 IEEE International Conference on Robotics and Automation. pp. 1–4 (May 2011), doi:10.1109/ICRA.2011.5980570
3. Bucur, V.: Techniques for high resolution imaging of wood structure: a review. Measurement Science and Technology **14**(12), R91 (2003), doi:10.1088/0957-0233/14/12/R01
4. Cawley, G.C., Talbot, N.L.C.: Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. Bioinformatics **22**(19), 2348–2355 (2006), doi:10.1093/bioinformatics/btl386
5. Chmielewski, L.J., Orłowski, A., Śmietańska, K., Górski, J., et al.: Detection of surface defects of type 'orange skin' in furniture elements with conventional image processing methods. In: Huang, F., Sugimoto, A. (eds.) Image and Video Technology – PSIVT 2015 Workshops. Lecture Notes in Computer Science, vol. 9555, pp. 26–37. Springer, 2016, Auckland, New Zealand (23-27 Nov 2015), doi:10.1007/978-3-319-30285-0_3
6. Chmielewski, L.J., Orłowski, A., Wieczorek, G., Śmietańska, K., Górski, J.: Testing the limits of detection of the 'orange skin' defect in furniture elements with the HOG features. In: Nguyen, N., Tojo, S., et al. (eds.) Proc. 9th Asian Conference on Intelligent Information and Database Systems ACIIDS 2017, Part II. Lecture Notes in Artificial Intelligence, vol. 10192, pp. 276–286. Springer, Kanazawa, Japan (3-5 Apr 2017), doi:10.1007/978-3-319-54430-4_27
7. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3), 273–297 (1995), doi:10.1007/BF00994018
8. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley (1991)
9. Duda, R., Hart, P., Stork, D.: Pattern Classification. John Wiley & Sons, New York, 2 edn. (2001)
10. Hall, M.A., Smith, L.A.: Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In: Proc. 12th Int. Florida AI Research Society Conf. FLAIRS '99. AAAI (May 1-5, 1999), https://www.aaai.org/Library/FLAIRS/1999/flairs99-042.php
11. Karras, D.A.: Improved defect detection using support vector machines and wavelet feature extraction based on vector quantization and SVD techniques. In: Proc. Int. Joint Conf. on Neural Networks, 2003. vol. 3, pp. 2322–2327 (July 2003), doi:10.1109/IJCNN.2003.1223774
12. Konieczny, J., Meyer, G.: Computer rendering and visual detection of orange peel. Journal of Coatings Technology and Research **9**(3), 297–307 (2012), doi:10.1007/s11998-011-9378-2
13. Kruk, M., Świderski, B., Osowski, S., Kurek, J., et al.: Melanoma recognition using extended set of descriptors and classifiers. EURASIP Journal on Image and Video Processing **2015**(1), 43 (2015), doi:10.1186/s13640-015-0099-9
14. Kruk, M., Świderski, B., Śmietańska, K., Kurek, J., Chmielewski, L.J., Górski, J., Orłowski, A.: Detection of 'orange skin' type surface defects in furniture elements with the use of textural features. In: Saeed, K., Homenda, W., Chaki, R.

(eds.) Proc. 16th IFIP TC8 Int. Conf. Computer Information Systems and Industrial Management Applications CISIM 2017. Lecture Notes in Computer Science, vol. 10244, pp. 402–411. Springer, Cham, Białystok, Poland (16-18 Jun 2017), doi:10.1007/978-3-319-59105-6_34

15. Kurek, J., Świderski, B., Dhahbi, S., Kruk, M., et al.: Chaos theory-based quantification of ROIs for mammogram classification. In: Tavares, J.M.R.S., Natal, J.R.M. (eds.) Computational Vision and Medical Image Processing V. Proc. 5th Eccomas Thematic Conf. on Computational Vision and Medical Image Processing VipIMAGE 2015. pp. 187–191. CRC Press, Tenerife, Spain (19-21 Oct 2015), doi:10.1201/b19241-32
16. Liu, H., Setiono, R.: Chi2: Feature selection and discretization of numeric attributes. In: Vassilopoulos, J. (ed.) Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence. pp. 388–391. IEEE Computer Society, Herndon, Virginia (Nov 5-8, 1995), doi:10.1109/TAI.1995.479783
17. Liu, H., Yu, L.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proc. 20th Int. Conf. Machine Leaning ICML-03. pp. 856–863. ICM, Washington, D.C. (2003), http://www.aaai.org/Papers/ICML/2003/ICML03-111.pdf
18. Liu, H., Hussain, F., Tan, C.L., Dash, M.: Discretization: An enabling technique. Data Mining and Knowledge Discovery **6**(4), 393–423 (2002), doi:10.1023/A:1016304305535
19. Longuetaud, F., Mothe, F., Kerautret, B., et al.: Automatic knot detection and measurements from X-ray CT images of wood: A review and validation of an improved algorithm on softwood samples. Computers and Electronics in Agriculture **85**(0), 77–89 (2012), doi:10.1016/j.compag.2012.03.013
20. Musat, E.C., Salca, E.A., Dinulica, F., et al.: Evaluation of color variability of oak veneers for sorting. BioResources **11**(1), 573–584 (2016), doi:10.15376/biores.11.1.573-584
21. Pohjalainen, J.: Feature selection code (2015), http://users.spa.aalto.fi/jpohjala/featureselection/, http://users.spa.aalto.fi/jpohjala/featureselection/ [Online; accessed 25 Apr 2017]
22. Pohjalainen, J., Räsänen, O., Kadioglu, S.: Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. Computer Speech & Language **29**(1), 145 – 171 (2015), doi:10.1016/j.csl.2013.11.004
23. Świderski, B., Osowski, S., Kruk, M., Kurek, J.: Texture characterization based on the Kolmogorov-Smirnov distance. Expert Systems with Applications **42**(1), 503–509 (2015), doi:10.1016/j.eswa.2014.08.021
24. Świderski, B., Osowski, S., Kurek, J., Kruk, M., et al.: Novel methods of image description and ensemble of classifiers in application to mammogram analysis. Expert Systems with Applications **81**, 67–78 (2017), doi:10.1016/j.eswa.2017.03.031
25. Wei, L.J.: Asymptotic conservativeness and efficiency of Kruskal-Wallis test for K dependent samples. Journal of the American Statistical Association **76**(376), 1006–1009 (1981), doi:10.1080/01621459.1981.10477756