Detection of 'Orange Skin' Type Surface Defects in Furniture Elements with the Use of Textural Features

Michał Kruk[1], Bartosz Świderski[1], Katarzyna Śmietańska[2], Jarosław Kurek[1],
Leszek J. Chmielewski[1(✉)], Jarosław Górski[2], and Arkadiusz Orłowski[1]

[1] Faculty of Applied Informatics and Mathematics – WZIM,
Warsaw University of Life Sciences – SGGW, ul. Nowoursynowska 159,
02-775 Warsaw, Poland
{michal_kruk,leszek_chmielewski}@sggw.pl

[2] Faculty of Wood Technology – WTD, Warsaw University of Life Sciences – SGGW,
ul. Nowoursynowska 159, 02-775 Warsaw, Poland
http://www.wzim.sggw.pl
http://www.wtd.sggw.pl

Abstract. The accuracy of detecting the *orange skin* surface defect in lacquered furniture elements was tested. Textural features and an SVM classifier were used. Features were selected from a set of 50 features with the bottom-up feature selection strategy driven by the Fisher measure. The features selected were the Kolmogorow-Smirnow-based features, some of the Hilbert curve-based features, some of the maximum subregions features and also some of the thresholding-based features. The Otsu thresholding and percolation-based features were all rejected. The images of size 300×300 pixels cut from the original, larger images were treated as objects. There were three quality classes: *very good*, *good* and *bad*. In the cross-validation process where the testing sets consisted of 90 and the training sets of 910 objects the accuracies ranged from 90% to 98% and the average accuracy was 94%. The tests revealed that more research should be done on the choice of features for this problem.

Keywords: Orange skin · Surface defect · Detection · Quality inspection · Furniture · Textural features · Classification

1 Introduction

In the furniture manufacturing industry the shape accuracy is equally important as the visual quality of the products. In our previous papers we assessed the applicability of the image processing methods to measurements in the furniture industry [1]. It occurred that the problem of the surface defect called *orange skin* requires special attention due to that the dimensional variability introduced by this defect is minute. In [2] we found that this defect seems to be easily detectable

with pure vision techniques and in [3] we stated that by using typical off-the-shelf textural features this defect is detectable in the majority of cases. In this paper we try to use some more advanced features already used by us in other, requiring applications [4,5,9].

As it was at the time of our previous publications, the literature on the application of image processing in the furniture industry seems to be extremely rare. As before, in the timber industry there is an entirely different situation: the structural and anatomical analysis is frequently performed with the image processing methods [6]. In the present paper we deal with painted surfaces, so the methods developed for raw materials [7] do not apply.

After positive experience with the detection of the *orange skin* defect treated with simple methods in [2] and some more advanced, but still well known ones in [3], now we shall assess the possibilities and limits of the application of relatively complex textural features from the groups of Kolmogorow-Smirnow features [9], maximum subregion-based features [4], features using the concept of percolation [4] and on the Hilbert curve [5].

We shall also try some features in which the main technique is a more or less advanced thresholding to finally check the usefulness of such simple features in the common setting with the advanced features, which proved their viability in demanding applications. These will be the thresholding-based features using the classical Otsu method [8] and some iteration-based features.

The remaining part of this paper is organized as follows. In the following Section we shall briefly characterize the considered surface defect. In the next Section we shall present the way we have prepared the images. Then, the features will be described together with an intermediate test aiming at feature selection with the classification accuracy as a criterion. In the following Section we shall report on the results obtained. Conclusions and outlook for further development will close the paper.

2   The Defect: *Orange Skin*

On the lacquered surfaces of furniture elements the defect called *orange skin* or *orange peel* can appear. It can emerge as small, shallow hollows, that is, an uneven structure of the hardened surface. The reasons for this defect are numerous: insufficient quantity or bad quality of dilutent, excessive temperature difference between the lacquer and the surface, bad distance or pressure of spraying, excessive air circulation during spraying or drying, and insufficient air humidity. On the lacquered surfaces the structure of wood is hidden, so *orange skin* is the only visible sign of surface unevenness.

A surface is considered good or defective for esthetic reasons. Moreover, it is not possible to point out a well-defined *defect* on the surface. The small valleys and holes are surrounded or gradually pass to the good regions. This is the presence or absence of the holes which make the whole surface *good* or *bad*. Also the *good* surface is not free from small deviations from planarity.

3 Measuring Setup and Images

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 inclined by 70° from the normal to the surface. In this way, the light was falling from a direction close to parallel to the surface, to emphasize the surface unevenness. 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 photographed surfaces belonged to several different objects. The surfaces were classified by the furniture quality expert as *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. The elongated objects were moved in front of the camera between the images were taken, to encompass all of the surface of the objects in the experiment. The images were made in color mode, with lossless compression.

From these images, small non-overlapping images were cut, each of them of size 300 × 300 pix. 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 so that, in each choice, there were 90 images in the testing set, with equal numbers of images belonging to each class. The remaining 810 images formed the training set.

The way the small images were cut can be seen in Fig. 1. The examples of images belonging to the three classes are shown in Fig. 2.

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. Each of these images contained only the evenly illuminated surface of the object. (Color figure online)
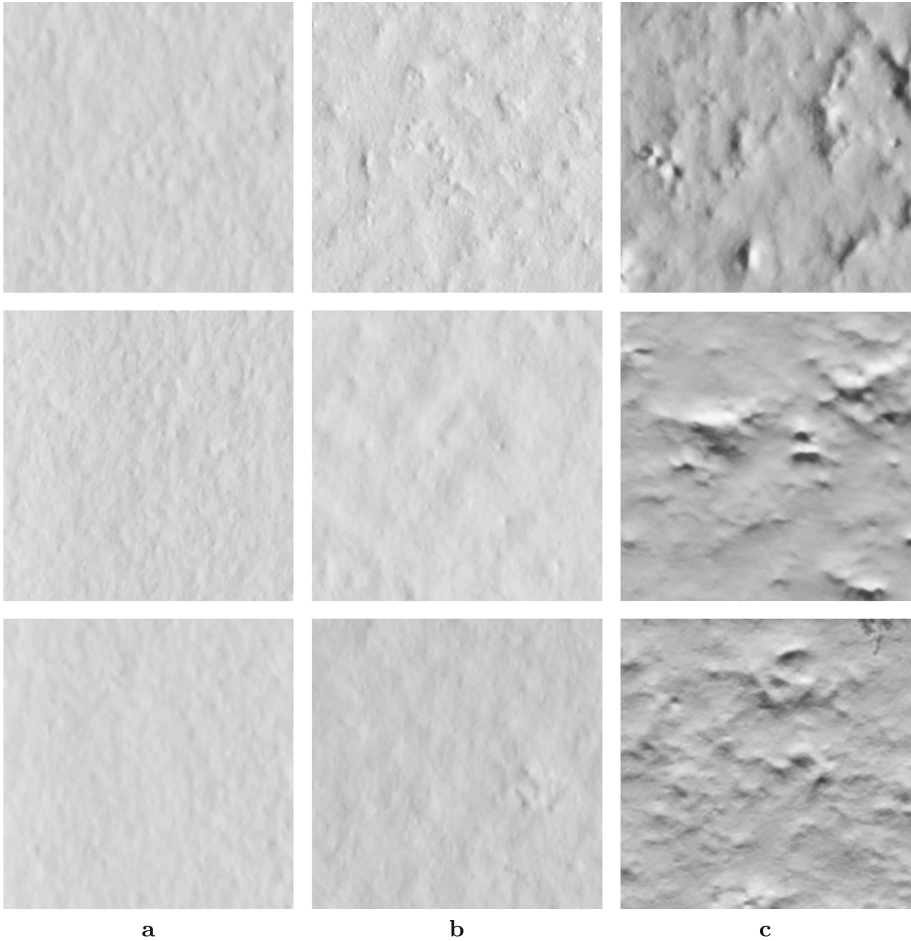
Fig. 2. Examples of images of the surfaces belonging to three classes: (**a**) very good, (**b**) good and (**c**) bad.

4   Features and Classification

4.1   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 following methods:

– number of black fields after thresholding with Otsu method – 1 feature;
– Kolmogorow-Smirnow features [9] – 7 features;
– maximum subregions features [4] – 6 features;
– features based on the percolation [4] – 2;

- features based on the Hilbert curve (cf. [5]) – 16;
- features from iterative single-valued thresholding (explained below) – 9;
- features from iterative adaptive thresholding (explained below) – 9.

The iterative 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 iterative adaptive thresholding was performed as follows. Let A be the image after applying the averaging filter with the window 20×20 pix. Then, the image I_2 is calculated nine times as $I_2 = A - Y - i, i = 1, 2, ..., 9$ and thresholded at $255/10$. The feature is the number of black regions in the image I_2, giving 9 features.

The dimension of the feature matrix was then 900×50: there were 900 objects (images) each described by 50 features.

4.2   Feature Selection

The 50 features were calculated for the 810 training objects. For all these objects the Fisher measure of information content was calculated as

$$S(f) = \frac{1}{3} \sum_{i,j \in \{1,2,3\}, i \neq j} S_{ij}(f), \text{ where } S_{ij}(f) = \frac{|m_i(f) - m_j(f)|}{\sigma_i(f) + \sigma_j(f)}, \quad (1)$$

and where $m_i(f)$ and $\sigma_i(f)$ are simply the mean and standard deviation of feature f in class i. The inter-class measures were averaged for all class pairs.

The Fisher measures for the classes calculated in this way are shown in Fig. 3. This measure made it possible to sort the features $f_k, k = 1, 2, ..., 50$ in the sequence of decreasing $S(f_k)$.

The support vector machine (SVM) classifier (to be described in Sect. 4.3) was used to perform the teaching process with the training objects, and the classification accuracy was checked against the testing objects. This was performed
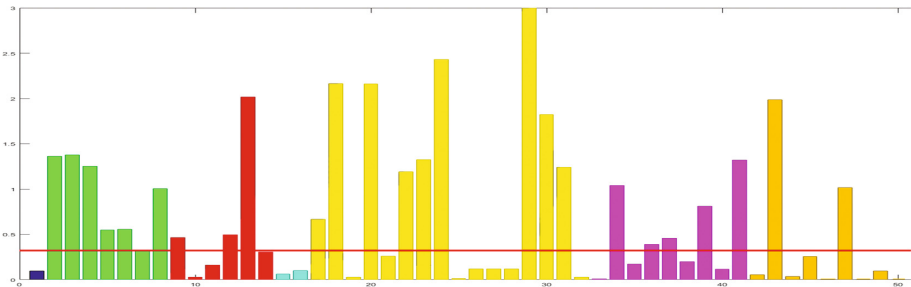

Fig. 3. Fisher measures for 50 features. Colors identify groups of features: blue: 1 Otsu feature; green: 7 Kolmogorow-Smirnow; red: 6 maximum subregions; bright blue: 2 percolation-based; yellow: 16 from Hilbert curve; magenta: 9 from single-valued thresholding; brown: 9 from adaptive thresholding. Threshold $S = 0.31$ obtained in the feature selection process marked with red line (see Sect. 4.2). (Color figure online)
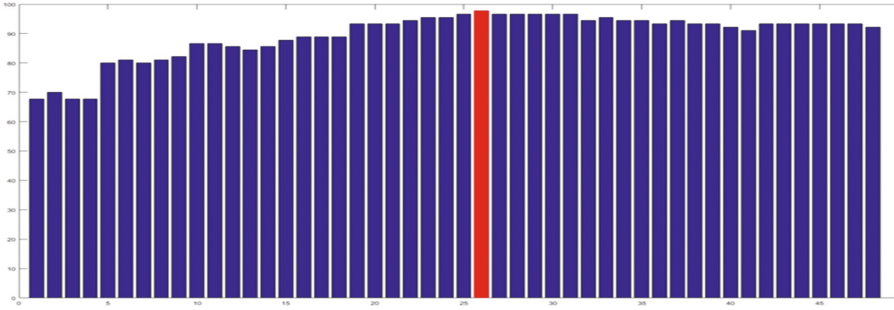
Fig. 4. Accuracy of classification for subsequent sets of features, containing from one to 50 features, chosen according to the decreasing Fisher measure S.

for different sets of features. The first feature set contained the first feature from the sequence with decreasing S (this was the 13th Hilbert feature, bearing the index 29 in the graph in Fig. 3 and having the largest S). The second set contained the first two features in the sequence (plus the 8th Hilbert feature), and so on, up to 50 features in the set. For each set, the attained classification accuracy was noted, as shown in Fig. 4. It can be seen that the accuracy increases for 26 features in the set, up to the value of 97.8%, and then decreases as the remaining 24 features are added. So, this was the bottom-up feature selection, controlled by the Fisher measure to avoid checking the accuracy attained after the addition of each of the features remaining at a given step of the feature adding process.

The result of this feature selection process is that the features having the Fisher measure $S > 0.31$ were chosen. This threshold is marked with the red line in Fig. 3. The chosen features were:

– All 7 Kolmogorow-Smirnow features (green in Fig. 3);
– 3 maximum subregions features (red);
– 9 features based on Hilbert curve (yellow);
– 5 features from single-valued thresholding (thresholds: $0.2, 0.4, 0.5, 0.7, 0.9$);
– 2 features from adaptive thresholding (thresholds: $0.2, 0.6$).

4.3   SVM Classifier

The used version and parameters of the SVM classifier were: radial-basis function kernel, cost $c = 300$, $\sigma = 0.1$. In this paper we paid attention rather to the problem of feature selection than to the choice of the classier. Therefore, we have used one of the classifiers typically applied to the problems of this kind. Extending the set of classifiers will be one of the next steps in our studies.

4.4   Accuracy of Classification

As described above, in the feature selection process, based upon one choice of the training and the testing set, the accuracy attained the highest value equal to 97.8%. The rates of a priori errors received in this case are shown in Table 1.

Table 1. A priori confusion matrix from the SVM classifier with 26 features chosen.

		Classified		
		very good	good	bad
Actual	very good	1.00	0.00	0.00
	good	0.0(6)	0.9(3)	0.00
	bad	0.00	0.00	1.00

The only error made was that two images actually belonging to the class *good* were misclassified and assigned to the class *very good*. In relation to 30 objects in each class (in the testing set) this gives the error rate 0.0(6). In relation to the whole 90 testing object this gives the already mentioned accuracy value of 97.8%.

The two images for which the errors were made are shown in Fig. 5a and b. In the case of these images the classification result was a false positive with respect to the class of the best quality, which should be avoided in quality inspection. However, it can be argued if each of these images really represents inferior quality, but this is beyond the scope of this technical study. It is positive that there are no errors between the two good classes and the bad one.

This result was further verified with the cross-validation with ten different divisions of the set of objects were used. As previously, in each division, 90 objects were chosen for testing, and the remaining 810 ones were used for training. The features were not changed with respect to those selected as described before. The first division was the one analyzed above. The results are shown in Fig. 6 and Table 2.

It can be clearly seen that for the first division the results were the most optimistic. In the remaining ones the results were worse. The average accuracy attained was 94.2%, but in the worst case it was 90%.
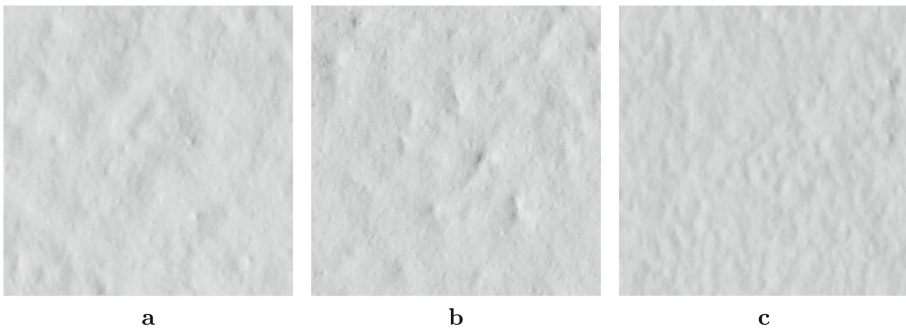

a   b   c

Fig. 5. Images for which the classifier made an error – subfigures (**a**) and (**b**). These are *good* objects but they were classified as *very good*. A truly *very good* object is shown for comparison in (**c**).
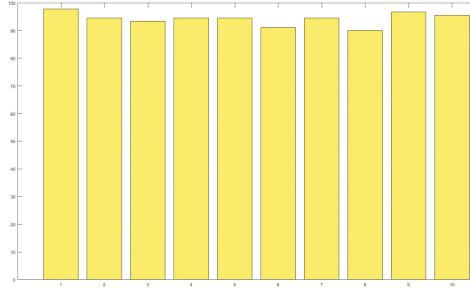
Fig. 6. Accuracy of classification for ten cross-validation cases.

Table 2. Accuracies and numbers of errors for ten cross-validation cases.

Case	Accuracy	No. of errors
1	97.8%	2
2	94.4%	5
3	93.3%	6
4	94.4%	5
5	94.4%	5
6	91.1%	8
7	94.4%	5
8	90.0%	9
9	96.7%	3
10	95.6%	4
Worst	90.0%	9
Average	94.2%	5.2

5 Discussion

Similar problem has been investigated in our previous publications [2,3]. Results obtained in those works suggested that the choice of features was not crucial to the classification result. Therefore, we did not pay much attention to this choice. However, the problem we investigated here was different from that described in [2], where as the classified object the whole product was considered, and there were only two classes: *good* and *bad* object. Now, we consider a small fragment of the product as the object to be classified. This is a more precise approach; accordingly, the choice of the feature should be more careful. Therefore, we have retained the features in which thresholding was used, but we augmented the set of features with a number of other ones, which were known for being useful in the problem of classification of the state of a surface.

The results obtained in this work did not reach the level of accuracy we expected. The average overall classification accuracy attained was 94% but the

smallest one was 90%. Besides the change in the definition of the object of classification mentioned before, this can have the origin in that the objects cut from larger surfaces were investigated, so the defects could manifest themselves in a larger variety of ways.

The groups of features selected as useful were the Kolmogorow-Smirnow based features, some of the Hilbert curve-based features, some of the maximum subregions features and some of the thresholding-based features. The Otsu thresholding-based and the percolation-based features were rejected. Other groups of features attained less univocal results. It is interesting that among the simplest features, namely those using the plain single-valued thresholding, some were actually selected.

6 Summary and Prospects

The SVM classifier with features selected from a set of 50 features was successfully used to classify the surfaces affected by the *orange skin* surface defect. This defect appears at lacquered surfaces of furniture and causes these surfaces to belong to three classes: *very good*, *good* and *bad*, according to the furniture quality experts. The bottom-up feature selection strategy driven by the Fisher measure was applied. The features which performed the best were the Kolmogorow-Smirnow based features, some of the Hilbert curve-based features, some of the maximum subregions features and some of the thresholding-based features. The Otsu thresholding-based and the percolation-based features were all rejected. However, some of the simple thresholding-based features were selected.

This tend to confirm the usefulness of very simple methods in solving the problem of orange skin detection. However, such simple tools alone do not provide a sufficiently good result. It is clear that the problem of features should be investigated in depth. The variety of ways in which the surface defects manifest themselves in the lacquered furniture elements makes it necessary to broaden the range of features used in the classification. This will be the subject of our forthcoming papers.

References

1. Chmielewski, L.J., et al.: Defect detection in furniture elements with the hough transform applied to 3D data. In: Burduk, R., Jackowski, K., Kurzyński, M., Woźniak, M., Żołnierek, A. (eds.) Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015. AISC, vol. 403, pp. 631–640. Springer, Cham (2016). doi:10.1007/978-3-319-26227-7_59
2. Chmielewski, L.J., Orłowski, A., Śmietańska, K., Górski, J., Krajewski, K., Janowicz, M., Wilkowski, J., Kietlińska, K.: Detection of surface defects of type 'Orange Skin' in furniture elements with conventional image processing methods. In: Huang, F., Sugimoto, A. (eds.) PSIVT 2015. LNCS, vol. 9555, pp. 26–37. Springer, Cham (2016). doi:10.1007/978-3-319-30285-0_3

3. 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.T., Tojo, S., Nguyen, L.M., Trawiński, B. (eds.) ACIIDS 2017. LNCS, vol. 10192, pp. 276–286. Springer, Cham (2017). doi:10.1007/978-3-319-54430-4_27

4. Kruk, M., Świderski, B., Osowski, S., Kurek, J., et al.: Melanoma recognition using extended set of descriptors and classifiers. EURASIP J. Image Video Process. **2015**(1), 43 (2015). doi:10.1186/s13640-015-0099-9

5. 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. Proceedings of 5th Eccomas Thematic Conference on Computational Vision and Medical Image Processing VipIMAGE 2015, pp. 187–191. CRC Press, Tenerife, Spain, 19–21 October 2015. doi:10.1201/b19241-32

6. 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. Comput. Electron. Agric. **85**, 77–89 (2012). doi:10.1016/j.compag.2012.03.013

7. 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

8. Otsu, M.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979). doi:10.1109/TSMC.1979.4310076

9. Świderski, B., Osowski, S., Kruk, M., Kurek, J.: Texture characterization based on the Kolmogorov-Smirnov distance. Expert Syst. Appl. **42**(1), 503–509 (2015). doi:10.1016/j.eswa.2014.08.021