

Statistical Quality Assessment of Pre-fried Carrots Using Multispectral Imaging

Sara Sharifzadeh^{1,*}, Line H. Clemmensen¹, Hanne Løje², and Bjarne K. Ersbøll¹

¹ Department of Applied Mathematics and Computer Science,
Technical University of Denmark
{sarash, lkhc, bker}@dtu.dk

² National Food Institute, Technical University of Denmark
halo@food.dtu.dk

Abstract. Multispectral imaging is increasingly being used for quality assessment of food items due to its non-invasive benefits. In this paper, we investigate the use of multispectral images of pre-fried carrots, to detect changes over a period of 14 days. The idea is to distinguish changes in quality from spectral images of visible and NIR bands. High dimensional feature vectors were formed from all possible ratios of spectral bands in 9 different percentiles per piece of carrot. We propose to use a multiple hypothesis testing technique based on the Benjamini-Hachberg (BH) method to distinguish possible significant changes in features during the inspection days. Discrimination by the SVM classifier supported these results. Additionally, 2-sided t-tests on the predictions of the elastic-net regressions were carried out to compare our results with previous studies on fried carrots. The experimental results showed that the most significant changes occurred in day 2 and day 14.

Keywords: Multispectral imaging, Multiple hypothesis testing, Segmentation, Food quality assessment, SVM classification, Elastic-net regression.

1 Introduction

Spectral imaging has recently gained use in on-line quality monitoring of food items. This method has important privileges over the traditional assessment methods, based on skillful human experts. The imaging-based methods are automatic, fast and contact-less. Since, there is a trend toward the use of these methods for quality control of food products such as meat, dairies and vegetables; it is of importance to test the capabilities as well as the reproducibility of such.

In this paper, the multispectral images of pre-processed carrots were used to detect the effect of storage on their color and NIR characteristics. The carrots were pre-fried without oil and then frozen for about two months. Then, they were moved to the refrigerator for experiments during a period of 14 days. Generally the surface color and texture are important parameters; indicating the quality of food. Multispectral images

* Corresponding author.

provide this information in visible bands and also more information about the subsurface and chemical characteristics in NIR bands. Using the multispectral images in visible and NIR bands, we tracked the quality of carrots during the storage days. The aim was to find out in which days, significant changes occurred.

In this study, the preparation of carrots was performed in two steps. First, the vegetables were stir-fried (without oil) [1, 2]. Research findings have shown that, stir-frying produces high quality vegetables [1]. After stir-frying, the products were frozen. The concept of partial thawing during distribution which can improve shelf life was used [3, 4].

Previous studies showed that, multispectral images can be used to assess the color change over time in pre-fried vegetables [5]. The analysis of multispectral images of pre-fried carrots and celeriac (fried in oil) was carried out for change detection with in a period of 14 days in [6, 7]. In [6], the segmented vegetable pieces were used to form the high dimensional feature vectors (3249 variables). For each carrot piece, 9 different percentiles in all possible ratios of bands were calculated. Then, the elastic-net regression analysis was carried out to predict the days of analysis from these high dimensional features. Finally, a 2-sided t-test on the estimated days was applied. A significant change was detected in carrots from day 2 to 4 at a 5% level of significance. In [7], instead of the ratios of bands, the feature vectors were calculated using the percentiles of pixel intensity values within each vegetable piece. These features were used directly for unpaired t-tests to detect trends of change in reflection, as a function of days kept in the refrigerator. Again for carrot, the significant change was detected from day 2 to 4 at 5% level. The results were also compared with the sensory panel tests.

In this paper, we propose to apply a multiple hypothesis testing technique to assess the high dimensional features obtained from the ratios of spectral bands and their corresponding percentiles. The high dimensional features based on band ratios are preferred, since they are more robust toward the undesired effects such as shadows. Multiple hypothesis techniques are mostly used for genomic data [8, 9]. They involve the significance assessment of the individual features. This assessment was performed, without the use of multivariate predictive models like in [6]. Since the dimensionality of the extracted features was quite high (3078), a conventional t-test at a significance level e.g. $\alpha=0.05$ may find about 154 significant features just by chance even if, the null hypothesis of no change is true for all the features [9]. In our study, the False Discovery Rate (FDR) introduced in [11] and the expected number of significant features was used to detect the significant days of change. In addition, the Support Vector Machine (SVM) classification was employed. Although the classification results support the multiple hypotheses testing, it is difficult to use them alone, as a demonstration for significance of changes over the days. In addition, the method used in [6] was applied to our data set, and the results were compared with the findings from the multiple hypothesis tests. Finally, we found the wavelengths mostly represented the significant features over the inspection days.

The rest of this paper is organized as follows; section 2 describes the data preparation for the experiments and the feature extraction step. In section 3, we explain the data analysis and section 4 presents the results. Finally, there is a conclusion for this paper in section 5.

2 Data Preparation and Feature Extraction

2.1 Data Preparation

The carrots used in this study, were prepared for the experiments a few days after harvest. First, they were cut into cube shapes of size one cm³ approximately. Then, they were wok-fried for 4 minutes at 250 °C in the continuous stir [1]. After cooling down, they were packed in 500g weight in plastic bags and frozen at -30 °C for 50-60 days. Finally, they were moved to the refrigerator (+5 °C). On days 2, 5, 8, 11 and 14 one bag was taken out from the refrigerator and the imaging experiments were conducted. The VideometeLab¹ was used for multispectral imaging like in [6, 7]. The multispectral images were obtained in 19 different wavelengths with the image resolution of 1280×960 pixels. The spectral images of a sample petri dish of carrots are shown in Fig.1 and pseudo-RGB images of samples in the five inspection days are shown in Fig 2.

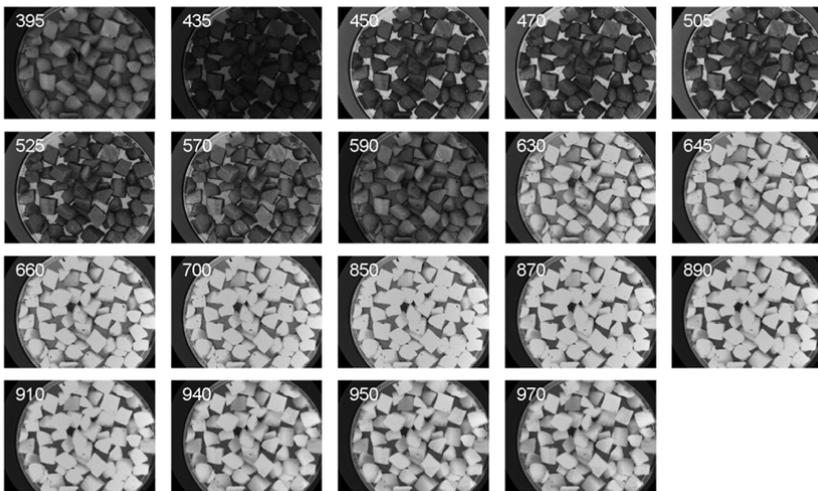


Fig. 1. Spectral images of carrot in 19 wavelengths, shown in nanometer range



Fig. 2. Pseudo-RGB images of carrot samples. From left to right: day 2, 5, 8, 11, 14.

¹ www.videometer.com

2.2 Feature Extraction

In order to form feature vectors, the first step is to segment the carrot pieces from the background and also from each other. Some of the procedures are almost similar to those in [6]. A brief description of different steps is presented in the following.

The background, which is everything except carrot, was removed in three different steps. First, a simple thresholding was performed manually, considering the histograms of the images. But, some parts of the background that had intensity variations close to the carrots, still remained. Therefore, two populations were considered (carrot and remaining background) and labeled manually. Then Canonical Discriminant Analysis (CDA) [10] was employed to improve the discrimination level between the two populations to define better thresholds [12]. Finally, a fine morphological operation (closing and erosion) was used to clean the undesired small remaining areas. For segmentation of carrot pieces, similar to [13], the Sobel edge detector was used which is based on the gradient function. Then, some morphological operations followed by a Watershed transformation were applied on the background-removed image. Using both the gradient and Watershed transform, the carrot pieces were segmented.

Each carrot piece was considered as a single observation. Since there were lumps of pieces in the petri dishes instead of individual well separated cubes, the segmentation was not perfect. In some cases two carrots were segmented as one piece or one piece was segmented into two different pieces. However, all segmented areas include carrots bodies. Fig.3 illustrates different steps of the carrot segmentation. A feature vector was formed for each detected carrot piece. First, for each of the 19 bands, all its 18 possible ratios to the other bands were calculated (totally 342 ratio matrices). Then, in each ratio matrix, 9 percentiles (1, 5, 10, 25, 50, 75, 90, 95, 99) were calculated for each piece. Therefore, by concatenating of the 9 percentile vectors of the 342 ratios, a 3078 length feature vector per piece of carrot was obtained.

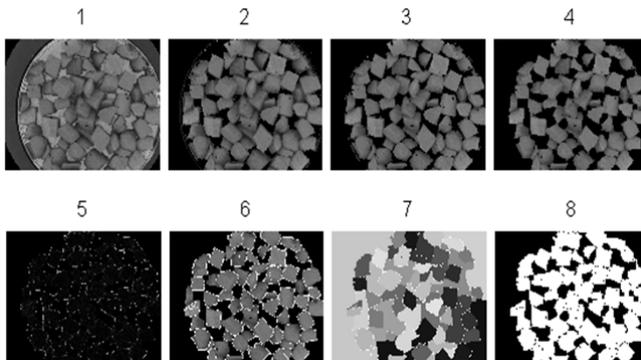


Fig. 3. Segmentation steps: 1-original image 2-two manual thresholding steps 3-thresholding using CDA 4-morphological operation 5-detected edges by Sobel 6-detected edges by Watershed 7-Watershed label matrix 8-segmented pieces

3 Data Analysis

As mentioned before, the aim of the analysis is to detect and track changes in the color and NIR characteristics of the carrots over the days of inspection. Totally, 3277 observations (carrot pieces) were segmented from the images of the 5 days of experiments. They were divided into two sets (*set1* and *set2*) randomly and all analyses were performed twice, using one of the sets as training and the other as test set and vice versa.

3.1 Elastic-Net Regression and 2-Sided t-test

Looking to the problem from a statistical point of view, a statistical test can be used for finding the significant changes. For this aim, like in [6], an elastic-net regression was performed on the ratio features to predict the number of days. 10-fold cross validation was used to generalize the prediction model. In the next step, a 2-sided t-test was applied on the predicted labels for all pairs of days to test the null hypothesis H_0 that two groups come from the same population at the 5% level of significance. This means that instead of considering the high dimensional features directly, they were first used for predicting the days and the one-dimensional prediction vector was used for the statistical test.

3.2 SVM Classification

The main problem at hand can be considered as discrimination of the data between different labels (days). If we look at the problem from this point of view, a classification approach can be used as a means of discrimination. For this aim, the powerful, support vector machine classifier was employed using LIBSVM [14]. A linear kernel with 5-fold cross validation was used for training the classifier. We expect that in days where considerable changes occurred, most of the samples be truly classified. On the other hand, for days with more similarity, misclassified samples should increase. Therefore, we will look at the confusion matrix for classification between days.

3.3 Multiple Hypothesis Testing

Another way of performing a statistical test is to perform multiple hypothesis tests for individual features of pairs of days to find the significant changes. However, not all the detected significant features are truly significant. For example, if the conventional statistical t-test with a priori significance level of $\alpha=5\%$ be used, just by chance, about 154 significantly changed features will be found out of the total 3078 features, while the null hypothesis may be true for some of them. Table 1 shows the theoretical outcomes from the M hypothesis tests [10]. Where, M is the number of features.

To address this problem, the number of falsely significant features (V in Table 1), for which the null hypothesis of no change (H_0) is true, should be found. One simple

Table 1. Possible outcomes from M hypothesis tests

	<i>Called not Significant</i>	<i>Called Significant</i>	<i>Total</i>
H0 True	U	V	M ₀
H0 False	T	S	M ₁
Total	M-R	R	M

solution is the *Bonferroni* method. In order to reduce the number of false positive features (V), it rejects H₀ if the p-value of a feature satisfies $p < \alpha/M$. It is a useful method in cases that M is small, as it is based on the assumption that the covariates are independent. However, in our case M is quit high (M=3078) and high correlation exists between the covariates. Therefore, the *Benjamin-Hachberg* (BH) method is used instead. They introduced the False Discovery Rate (FDR) as follows:

$$FDR = E\left(\frac{V}{R}\right) \tag{1}$$

It is the expected proportion of the false positive features V among the R features that are called significant. In this method, the FDR rate is bounded by a user defined level α . It is calculated based on the p-values obtained from an asymptotic approximation of the test statistic like a Gaussian or a permutation distribution.

In this paper, a *plug-in* version of this method is followed [10]. The FDR rate is bounded at $\alpha=0.15$ and the permutation distribution is used. Then, the number of truly significant features $\widehat{E}(S)$ is estimated and used for making decision about the days that significant changes occurred in the vegetable data. The procedures are as follows:

1. The t statistics are calculated for all the features of both days: $t_j, j = 1, \dots, M$
2. For K=1000 times, the sample’s labels (days) are permuted and the t statistic for the features at each permutation round is calculated. $t_j^k, j = 1, \dots, M, k = 1, \dots, K$
3. For a cut-point C, the $R = \sum_{j=1}^M I(|t_j| > C)$, as well as the $\widehat{E}(V) = \frac{1}{K} \sum_{j=1}^M \sum_{k=1}^K I(|t_j^k| > C)$ are calculated. (the I function shows the number of times the inequality is satisfied)
4. The $\widehat{FDR} = \widehat{E}(V)/R$ and the number of truly significant features $\widehat{E}(S) = R - \widehat{E}(V)$ are computed.

In this study, the cut-point C is chosen equal to the t statistic of the critical point of the BH method. The critical value controls the FDR to be around $\alpha=0.15$. It is calculated in the following steps:

1. The corresponding pooled p-value of the t statistics for each feature is computed: $p_j = \frac{1}{MK} \sum_{j'=1}^M \sum_{k=1}^K I(|t_{j'}^k| > |t_j|)$
2. The p-values for the features are ranked in ascending order $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(M)}$
3. The BH critical point is $L = \max\{j: p_{(j)} < \alpha \times j/M\}$ and so the cut point is $C=|t_L|$.

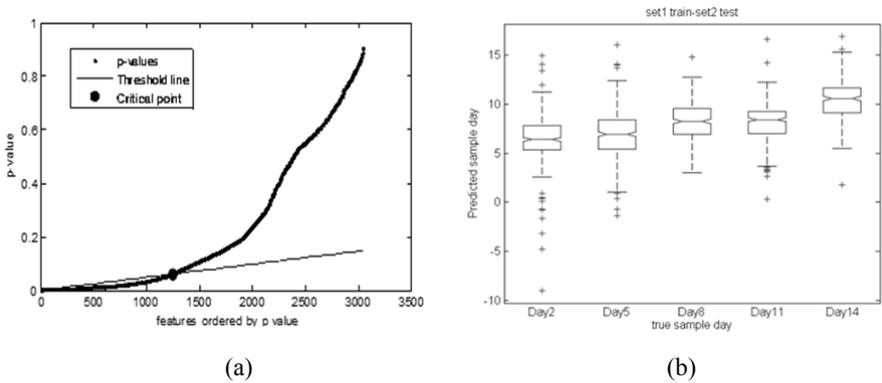


Fig. 4. (a) A plot of the ordered p-values $p_{(j)}$, the threshold line $(\alpha \times j/M)$ as well as the critical point of the BH method. (b) The box-plots of the elastic-net prediction, set1 training set2 test.

Fig.4 (a) shows a sample plot of the ordered p-values, the threshold line $(\alpha \times j/M)$ as well as the critical point. Using this method, the number of significantly changed features is obtained for all pairs of days.

4 Results and Discussion

In the first experiment, like in [6], an elastic-net prediction was performed using the spectral data of set1 to train the prediction model. Then, set2 was used for test. Fig. 4(b) shows the boxplot of the result. The same experiment was repeated, when the training and test sets were swapped. The prediction error was high (similar to [6]) and the test MSEs were 13.55 and 13.66 for the two experiments respectively. Although the prediction accuracy was poor, a linear trend could be seen for the predictions over the 5 days. A pairwise 2-sided t-test (at a 5% level of significance) was performed for all pairs of days using the prediction output. The results of both sets showed significant change almost between all pairs of days.

In the next experiment, SVM classification was performed. Table 2 shows the confusion matrix of the SVM test results, where set2 was used for training the model and set1 was used for test. It shows the percentage of each day’s samples that were classified into one of the 5 days of inspection. Therefore, all rows sum to 100 percent. As can be seen, at each day, more than half of the observations were classified to the same class (inspection day) correctly, and less observations were misclassified. This implies the capability of the multispectral imaging to distinguish subtle changes in carrots over days that are even difficult to be observed by eyes, as we have seen in Fig. 2.

On the other hand, the number of misclassified samples was also considerable in many cases which illustrate the similarity of the samples in different days. Day 14

gained the minimum misclassification and 78.66% of the samples were just truly classified in that day. For the second experiment, where the training and test sets were swapped, the same holds true for day 14. Day two gained the second highest classification performance. However, in the second experiment its percentage was around 60.29. Therefore, more information about the significance of the changes seemed necessary to confirm a change in quality of carrots between days.

At this step, we used the results from the multiple hypothesis tests explained in the previous section. Tables 3, shows the number of significantly changed features in all pairs of days for set1. These results were obtained using the BH method for the FDR rate bounded at 0.15. The Table is symmetric to the diagonal. The pairwise analysis between day 14 and the all other days showed the highest number of significantly changed features. This is compatible with the previous results from the SVM classification. The next highest number of significant features was detected from the pairwise analysis of day 2 with the other days. This was seen in one of the SVM test results as well. The same analysis was performed on set2 and the results were similar to set 1. Therefore, we can conclude from these analyses that, the most important change in carrot samples occurred after 2 weeks. While with less significance level, they also changed after 2 days being kept in the refrigerator.

Comparison of the results obtained from the SVM and BH methods with those from the t-test on the elastic-net predictions, showed differences in the significant days of change. However, we believe that the SVM and BH methods results are more reliable. First, since they were obtained by the direct use of the 3078 features of spectral image than the prediction results. Second, the elastic-net regression prediction error was high.

Table 2. The percentage of each day data assigned to the 5 classes by SVM

	<i>Class 2</i>	<i>Class 5</i>	<i>Class 8</i>	<i>Class 11</i>	<i>Class 14</i>
Day 2	73.19	2.41	4.22	15.36	4.82
Day 5	4	64	20	2.15	9.85
Day 8	4.19	18.39	62.26	4.84	10.32
Day 11	24.92	1.25	5.61	62.30	5.923
Day 14	4.73	6.80	10.35	2.66	75.45

Table 3. Number of significantly changed features in all pairs of inspection days for set1

<i>Set 1</i>	<i>Day 2</i>	<i>Day 5</i>	<i>Day 8</i>	<i>Day 11</i>	<i>Day 14</i>
Day 2	0	592.47	503.99	332.48	1288.67
Day 5	592.47	0	4.98	22.94	1118.72
Day 8	503.99	4.98	0	22.01	1025.24
Day 11	332.48	22.94	22.01	0	1026.94
Day 14	1288.67	1118.72	1025.24	1026.94	0

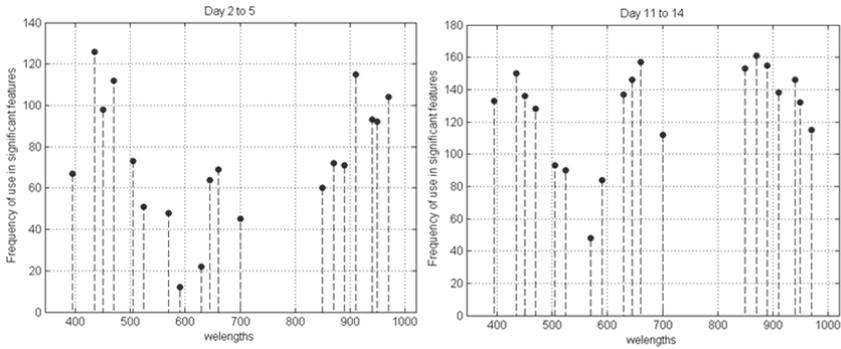


Fig. 5. The Frequency map for contribution of the wavelengths in significant features

In addition, considering the requirements of an industrial level vision system, it is interesting to know which wavelengths contributed most in the significant features. For this reason, we examined the frequency at which a wavelength was included in the features below the critical point. Fig. 5 shows the frequency of each wavelength being used in those features. The frequency maps of the pairwise analysis between days 2 to 5 and 11 to 14 are shown. For day 2 to 5, the three mostly used wavelengths were 435 nm (blue), 910 nm (NIR) and 470 nm (blue). In case of day 11 to 14, the 850-890 nm NIR bands as well as 660 nm (red) and 435 nm (blue) had the highest frequency. Similar analysis for other cases showed that, NIR bands as well as the blue and red wavelengths were among the top frequent bands.

There were some differences between the previous work in [6] and this study. The carrots were fried in oil in that work while no oil was used for this study. The inspection days weren't exactly the same compared to this work and the freezing period was two months more. The elastic-net regression model was built using leave-one-out cross validation that we believe may cause over fitting regarding the limited samples and high number of features. The significant change was found between days 2 and 4 in that work.

5 Conclusion

In this paper, multispectral images of pre-fried carrots were used to detect the changes in their quality within 14 days of inspection. The pre-fried carrots were kept around two months in the freezer and then were moved into the refrigerator. The use of multispectral images helps to extract the features representing the surface color as well as NIR characteristics of carrots. The Benjamin-Hachberg (BH) multiple hypothesis testing method was used to find the most significantly changed features over the storage days. The most important change in carrot samples occurred after 2 weeks. While with less significance level, they also changed after 2 days. Classification results obtained by SVM supported this. However, the elastic-net regression results had high MSEs. As a result, the 2-sided t-tests on the regression predictions of any set of 2 days at a 5% level were significant.

Acknowledgement. The authors would like to thank Peter Stubbe (National Food Institute, Technical University of Denmark) and Helene Carlsen (student) for their help in VideometerLab experiments.

This work was (in part) financed by the Centre for Imaging Food Quality project which is funded by the Danish Council for Strategic Research (contract no 09-067039) within the Program Commission on Health, Food and Welfare.

References

1. Adler-Nissen, J.: The Continuous Wok - a New Unit Operation in Industrial Food Processes. *Food Process Engine* 25, 435–453 (2002)
2. Adler-Nissen, J.: Continuous Wok-Frying of Vegetables. Process parameters influencing scale up and product quality. *Food Engineering* 83, 54–60 (2007)
3. Bao, H.N.D., Arason, S., Porarinsdottir, K.A.: Effects of Dry Ice and Superchilling on-Quality and Shelf Life of Arctic Charr (*SalvelinusAlpinus*) Fillets. *Food Engineering* 3(3), art.7 (2007)
4. Adler-Nissen, J., Akkerman, R., Frosch, S., Grunow, M., Løje, H., Risum, J., Wang, Y., Ørnholt- Johansson, G.: Improving the supplychain and foodquality ofprofessionallyprepared meals. *Trends in Food Science & Technology* 29, 74–79 (2013)
5. Løje, H., Dissing, S.B., Clemmensen, H.L., Ersbøll, K.B., Adler-Nissen, J.: Multispectral Imaging of Wok Fried Vegetables. In: 18th Scandinavian Workshop on Imaging Food Quality, Technical Report, Ystad, pp. 59–62 (2011)
6. Dissing, S.B., Clemmensen, H.L., Løje, H., Ersbøll, K.B., Adler-Nissen, J.: Temporal Reflectance Changes in Vegetables. In: 12th IEEE International Workshop on Computer Vision, pp. 1917–1922. IEEE Press, Kyoto (2009)
7. Clemmensen, H.L., Dissing, S.B., Hyldig, G., Løje, H.: Multispectral Imaging of Wok fried vegetables. *Imaging Science and Technology* 56(2), 20404-1-6 (2012)
8. Dudoit, S., Shaffer, P.J., Boldrick, C.J.: Multiple Hypothesis Testing in Microarray Experiments. *Statist. Sci.* 18, 71–103 (2003)
9. Diz, A.P., Carvajal-Rodríguez, A., Skibinski, D.O.F.: MultipleHypothesisTesting in Proteomics: a Strategy for Experimental Work. *Molecular & Cellular Proteomics* 10, M110.004374 (2011)
10. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Springer, New York (2009)
11. Benjamini, Y., Hochberg, Y.: Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Royal Statistical Society* 57, 289–300 (1995)
12. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9, 62–66 (1979)
13. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, New Jersey (2001)
14. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 27:1–27:27 (2011), Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>