

Multispectral Image Analysis for Astaxanthin Coating Classification

Martin Georg Ljungqvist

Department of Informatics and Mathematical Modelling, and Division of Industrial Food Technology, National Food Institute, Technical University of Denmark (DTU), Lyngby, Denmark
E-mail: malj@imm.dtu.dk

Bjarne Kjær Ersbøll

Department of Informatics and Mathematical Modelling, Technical University of Denmark (DTU), Lyngby, Denmark

Michael Engelbrecht Nielsen and Stina Frosch

Division of Industrial Food Technology, National Food Institute, Technical University of Denmark (DTU), Lyngby, Denmark

Abstract. Industrial quality inspection using image analysis on astaxanthin coating in aquaculture feed pellets is of great importance for automatic production control. The pellets were divided into two groups: one with pellets coated using synthetic astaxanthin in fish oil and the other with pellets coated only with fish oil. In this study, multispectral image analysis of pellets captured reflection in 20 wavelengths (385–1050 nm). Linear discriminant analysis (LDA), principal component analysis, and support vector machine were used as statistical analysis. The features extracted from the multispectral images were pixel spectral values as well as using summary statistics such as the mean or median value of each pellet. Classification using LDA on pellet mean or median values showed overall good results. Multispectral imaging is a promising technique for noninvasive on-line quality food and feed products with optimal use of pigment and minimum amount of waste. © 2012 Society for Imaging Science and Technology.

[DOI: 10.2352/J.ImagingSci.Technol.2012.56.2.020403]

INTRODUCTION

Industrial quality inspection using image analysis is an area undergoing extensive development. Pigment inclusion in aquaculture feed pellets is an area of great interest for automatic visual analysis for statistical production control and optimization.

Astaxanthin is a naturally occurring carotenoid with high antioxidant activity essential for reproduction, growth and survival, and important for the development of color in salmonid fish.¹ The primary use of astaxanthin within aquaculture is as a feed additive to ensure that farmed salmon and trout have similar appearance to their wild counterparts;² it is the pigment that makes salmonid fish red. The color appearance of fish products is important for customers. Astaxanthin is very expensive³ and therefore

optimizing quantities used in fish feed production is important.

An automatic vision system for on-line quality control of pigment inclusion will be of great benefit to the industry, both in relation to process control and process optimization.

This article is based in part on an earlier study by Ljungqvist et al.⁴ Besides this, to the authors' knowledge no further work has previously been done on analyzing the coating of fish feed using image analysis. Multispectral image analysis has shown good results in previous biological applications, where it has been of interest to detect subtle differences in color and surface chemistry.^{5–18}

The aim of this project is to investigate the possibility of distinguishing between feed pellets coated with fish oil with and without added astaxanthin using multispectral image analysis in order to investigate what spectral features are of interest for further analysis of astaxanthin coating.

MATERIAL AND METHODS

Material

The feed types used were EcoLife20 and AquaLife R90 (BioMar A/S, Brande, Denmark), both with the radius of 4.5 mm. Each of the two types of fish feed pellets were divided into two groups: One group constitutes pellets coated with fish oil with an additional 50 ppm of a synthetic version of astaxanthin; group A (astaxanthin). The other group was the same type of pellet, coated using the same fish oil without additional astaxanthin included; group B (base). The fish oil typically contains a small amount of natural astaxanthin, but this is assumed to be less than 1 ppm and should therefore not affect the results. The distribution of the surface coating was unknown and some amount of variation was likely to have occurred.

The pellets of type EcoLife20 were all produced on the same day, while the pellets of the two groups of AquaLife

Received Feb. 3, 2011; accepted for publication Nov. 7, 2011; published online May 22, 2012.

1062-3701/2012/56(2)/020403/6/\$20.00.

R90 were produced on different days. This means that the difference between groups A and B of AquaLife R90 is not only the added synthetic astaxanthin in the coating but could also be differences in the constituent raw materials.

Two feed types were used in order to test the robustness of the astaxanthin coating prediction. A total of 2223 EcoLife20 pellets were used, and a total of 2158 AquaLife R90 pellets were used, see Table I.

Imaging Equipment

The equipment used was a camera and lighting system called VideometerLab (Videometer A/S, Hørsholm, Denmark) which supports a multispectral resolution of up to 20 wavelengths. These are distributed over the ultraviolet A, visible (VIS), and first near infrared (NIR) region. The range is from 385 to 1050 nm, as shown in Table II.

This system uses a Point Grey Scorpion SCOR-20SOM gray-scale camera, and the objects of interest are placed inside an integrating sphere (Ulbricht sphere) with uniform diffuse lighting from light sources placed around the rim of the sphere. All light sources are light-emitting diodes (LED) except for 1050 nm which is a diffused laser diode. The curvature of the sphere and its white matte coating ensure a uniform diffuse light, so that the specular effects are avoided and the amount of shadow is minimized. The device is calibrated radiometrically with a following light and exposure calibration. The system is also geometrically calibrated to ensure pixel correspondence for all spectral bands.¹⁹

The Scorpion camera has a 12 bit analogue to digital converter (ADC), and the system used 8 bit data output from the camera. The correction for calibration gives reflectance intensity output of 32 bit precision.

The image resolution is 1280×960 pixels. Each file contains 20 images, one for each spectral band. This results in a multispectral image cube with dimensions of $1280 \times 960 \times 20$. In this situation, 1 pixel represents approximately 0.072×0.072 mm.

Spectral Equipment

In order to explore further the spectral properties of astaxanthin, a spectrometer was used. Absorption spectra of synthetic astaxanthin in a solution of fish oil along with spectra of plain fish oil were recorded in the VIS and NIR range using a NIRSystems 6500 absorption spectrometer (Foss NIRSystems, Inc., USA) The absorption spectra were transformed to reflection values using the standard relation $A = -\log(R)$, where A is the absorption value and R is the reflection value. Every second nanometer (nm) was recorded in the VIS and NIR range.

Table I. Number of analyzed pellets in each group. Group A represents pellets coated with synthetic astaxanthin in fish oil, and group B is pellets coated only with fish oil.

	Size (mm)	Group A samples	Group B samples	Total samples	Production day
EcoLife20	4.5	1165	1058	2223	Same
AquaLife R90	4.5	1207	951	2158	Different

Table II. The wavelength of the light sources in the VideometerLab device and their spectral representation.

Band	Wavelength (nm)	Color
1	385	Ultraviolet A
2	430	Violet
3	450	Violet, blue
4	470	Blue
5	505	Green
6	565	Green
7	590	Yellow, orange
8	630	Red
9	645	Red
10	660	Red
11	700	Red
12	850	NIR
13	870	NIR
14	890	NIR
15	910	NIR
16	920	NIR
17	940	NIR
18	950	NIR
19	970	NIR
20	1050	NIR

Image Analysis

Initially, the pellets were segmented from the background using an intensity threshold on the multispectral images, see Figures 1 and 2. Color standard RGB image representations of the multispectral images were made only for visualization in this article, by multispectral color-mapping using penalized least square regression described in Dissing et al.²⁰

The basic pellet compound gives a spectral response which will be present in the both groups A and B. Each

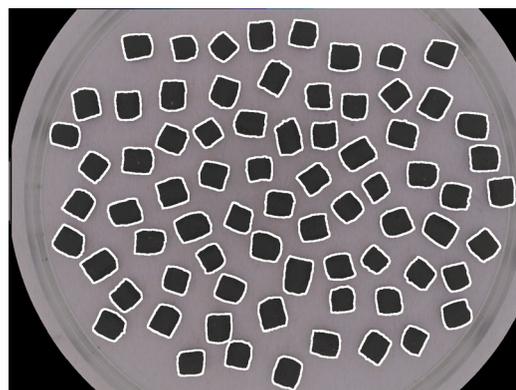


Figure 1. EcoLife 20 pellets with synthetic astaxanthin in fish oil as coating (group A) with the segmentation result overlaid (white). Standard RGB image using multispectral color-mapping using penalized least square regression.

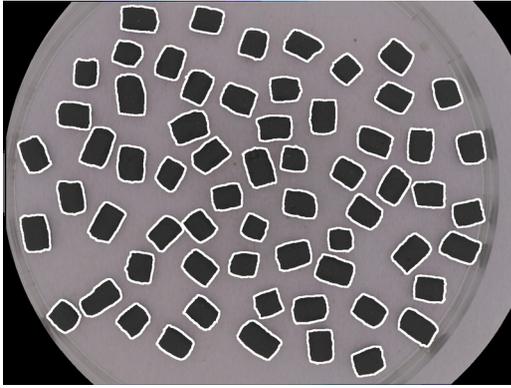


Figure 2. Ecolife20 pellets with fish oil coating (group B) with the segmentation result overlaid (white). Standard RGB image representation using multispectral color-mapping using penalized least square regression.

pixel is thus a combination of the reflectance of a set of constituents. This mix is assumed to be of equal amount for each pellet type except for the difference in the astaxanthin coating that we want to isolate in our classification.

The ground truth is that we know that certain pellets are coated with synthetic astaxanthin, but since the surface distribution is unknown, it is unclear how much synthetic astaxanthin each of these pixels contains. This gives us an uncertain one-to-many relationship situation.

One way to solve this uncertainty is to represent each pellet using the mean or median of all pixels in a pellet as sample values. In this manner, we even out the variance of all pixels in a pellet, and each pellet becomes a distinct observation.

In addition to the pellet pixel mean and median values, further summary statistical features to describe the coating distribution were extracted based on pellet pixel values: skewness, kurtosis, variance, and maximum value.

Principal Component Analysis

Our multivariate data from the images were analyzed using principal component analysis (PCA) for exploratory purposes. PCA is the most optimal method with respect to maximizing the variance²¹ and has been commonly used for dimension reduction to deal with ill-posed problems. If the relation of interest contains large variation, then PCA is a good method for analyzing the data.²²

Preprocessing

The preprocessing method standard normal variate (SNV) was used in combination with PCA. The SNV is performed by subtracting the mean and normalizing using the standard deviation of the data.²³ The SNV normalizes each sample individually and this will remove any variation in concentration level of the coating between pellets.

Discriminant Analysis

To discriminate between the two groups, we want the within group variation to be small compared to the between group's variation. Wilk's Λ consists in principle of

the ratio of the within group variation (\mathbf{W}) and the total variation (\mathbf{T}), i.e., the within group plus the between group variations, see Eq. (1). A value of Wilk's Λ which is close to zero indicates that the two groups are well separated.

$$\Lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{T})}. \quad (1)$$

For statistical discriminant analysis methods, we use linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA).²¹ These are both based on the Mahalanobis distance and assume that the observations in each group are normally distributed. The LDA and QDA are based on a distance-to-the-group mean, weighted by the variance. A training set of 70% of the samples were used here, along with a test set of 30% of the samples.

Support Vector Machine

For further discriminant analysis, we used the support vector machine (SVM), which is a supervised learning technique based on the theory of optimal separating hyperplanes.²¹ While LDA and QDA are based on the distance between a sample and the group mean, SVM is based on the distance to the nearest training data points; the margin.

The basic idea with SVM is to construct an optimal separating hyperplane for the two groups by mapping the data to a higher-dimensional space. This method uses a soft margin to handle the situation of two nonlinearly separable groups; meaning that a certain amount of the training samples are allowed to fall on the wrong side of the separating hyperplane which is defined by the slack amount.

The SVM includes a kernel function, responsible for the transformation of the data into a higher-dimensional space. In this way the data are mapped from its original input space into the higher-dimensional feature space. Once the data has been mapped, the aim is to train the model to define a separating hyperplane in the feature space so that the data are then mostly separable by a line. One advantage of using a kernel function is that the SVM method can perform classification without ever representing the feature space explicitly, thus reducing the computational cost. The kernel is located at the point of the dot product in the SVM algorithm.²¹

Popular choices for kernel functions can be seen in Eqs. (2) and (3). A d th degree polynomial is shown in Eq. (2), and a radial basis function (RBF) can be seen in Eq. (3). The parameters of variable choice are c (constant), s (linear scaling), d (degree of the polynomial), and γ . The samples from each group are denoted \mathbf{x}_1 and \mathbf{x}_2 , respectively.

It should be noted that the radial basis function has a Gaussian form, where $\gamma = 1/2\sigma^2$. The center of the RBF is the support vector and σ will determine the area of influence, it has over the data space.

For the kernel parameters, 120 values logarithmically distributed between 0 and 10 were tested for the parameters

γ and d , respectively, and the best result was chosen for each kernel. For the polynomial kernel, the parameters c and s were both set to 1. The value of 0.5 was used as the soft margin parameter.

$$K(\mathbf{x}_1, \mathbf{x}_2) = (s(\mathbf{x}_1 \cdot \mathbf{x}_2) + c)^d, \quad (2)$$

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma\|\mathbf{x}_1 - \mathbf{x}_2\|^2), \quad (3)$$

The SVM implementation used in this study is the SVM light package.²⁴

RESULTS AND DISCUSSION

Comparing the SNV-normalized mean spectra of the two groups of EcoLife20 shows that the largest differences between pellets coated with synthetic astaxanthin in fish oil (group A) and pellets coated only with fish oil (group B) were at 970, 950, and 565 nm (in order of magnitude), see Figure 3. Both 970 and 950 nm are in the NIR range, while 565 nm represents the green color, which is next to yellow. Also, 1050 nm shows to separate the groups. For AquaLife R90, the largest differences between the group spectra are in the visual range around 400 nm and also slightly above 600 nm.

The spectrometer results show a large deviation between synthetic astaxanthin in fish oil and plain fish oil in the range of 500–600 nm, see Fig. 3. This corresponds well with the results from the VideometerLab images and partly corresponds with previous studies of astaxanthin.^{25,26}

The mean spectra of the two groups of both EcoLife20 and AquaLife R90 are significantly different at a 0.1% level. This is promising for classification between the two coating groups. On the other hand, Wilk's Λ of the group means of EcoLife20 pellets is 0.987, and for AquaLife R90, it is 0.826.

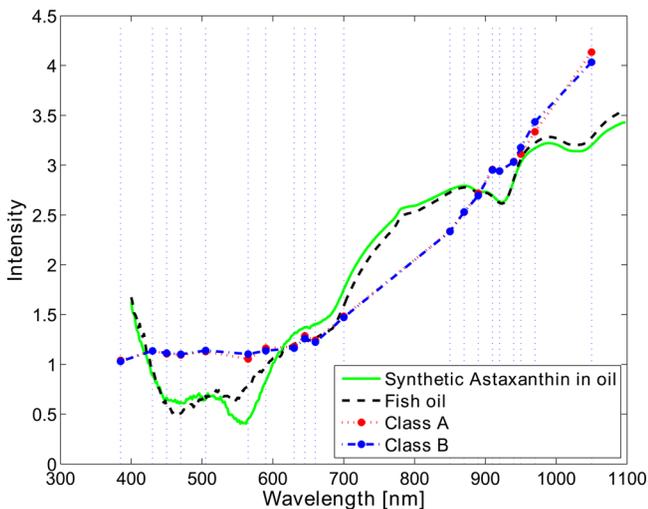


Figure 3. Spectrometer reflectance of synthetic astaxanthin in oil (green) and plain fish oil (black). Multispectral images (reflectance) mean of group A (pellets with synthetic astaxanthin in fish oil) (red) and group B (pellets with fish oil) (blue) of the EcoLife20 type. Spectrometer spectra converted from absorption to reflectance. All spectra are normalized using SNV to fit in the same plot. The wavelengths captured for the images by the VideometerLab are marked by vertical dotted lines.

Table III. The misclassification of pellet coating type for different kinds of features. Displayed values are total test error for classification of the two groups A (astaxanthin) and B (base).

EcoLife20	LDA	QDA	SVM (RBF)	SVM (Poly)
Mean	0.0646	0.0901	0.1291	0.0811
Median	0.0736	0.0931	0.1456	0.0841
Mean, SNV, PCI-5	0.1396	0.2162	0.1727	0.1637
AquaLife R90				
Mean	0.0046	0.0031	0.0093	0.0031
Median	0.0015	0.0000	0.0185	0.0031
Mean, SNV, PCI-5	0.0185	0.0201	0.0155	0.0108

The high values here reflect the situation of high variation within the groups and a low variation between the groups. So, even though the group means are well separated, there is a vast overlap of the two groups.

Classification tests of EcoLife20 show that LDA on the pellet means or pellet medians gave the best result with a classification correctness of about 93%, see Table III. for test results.

Classification tests of AquaLife R90 show that QDA on the pellet medians gave the best result with a classification correctness of 100%.

The results from LDA and QDA show for all tests that group A is misclassified into group B more often than the opposite. This could be because of the variation in the astaxanthin distribution on the pellets, in that if some pellets have less astaxanthin on the surface than others they might get misclassified.

Using LDA and QDA on the other summary statistics features (skewness, kurtosis, variance, and maximum value) gave results of lower correctness for both pellet types (results not shown).

Considering the fact that the AquaLife R90 pellets of the different groups were produced on different production days, there is a chance that this fact affects the classification rate. If the two groups also contain differences in raw material batches, this could potentially increase the differences between the two groups of this pellet type and thereby improve the classification result. This gives us a confounding situation where we are not certain whether it is the synthetic astaxanthin coating that is classified, the constituent raw material differences, or a combination of the two. In future experiments, this should be avoided by using the same production day for both groups. For further analysis in this area, the results are likely to show variation dependent on the mixture of the pellet types in relation to the variation in feed compounds that will affect the spectral response.

Having LDA train on the mean pixel values of some of the EcoLife20 pellets, 1126 pellets from group A and 1026 pellets from group B, and then using this model on pellet pixel values for the remaining 39 and 32 pellets, respectively, rendered a pixel classification correctness of only 54%. The

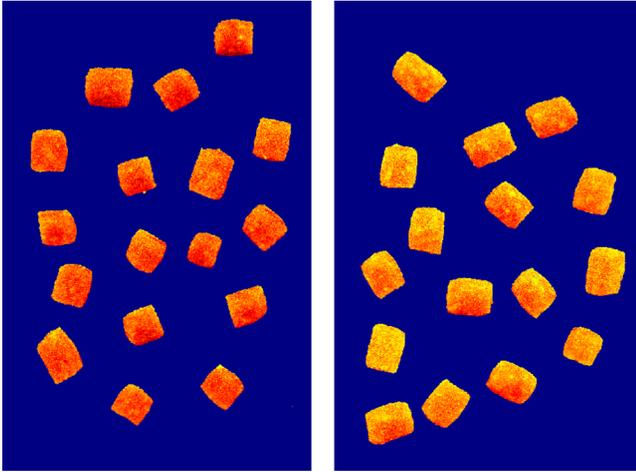


Figure 4. The second principal component of the multispectral image (reflectance) of EcoLife20 pellet pixels. Pellets coated with synthetic astaxanthin in fish oil, group A (left). Pellets coated with fish oil, group B (right).

same test for AquaLife R90 shows a pixel classification correctness of 55%. This lower value in comparison to classification on the pellet mean values is interpreted as an indication of high variance of the spatial coating distribution on the pellets and/or high variance in the feed composition at the pixel level.

Using PCA before doing LDA or QDA on the pellet mean values did not improve the results, see Table III. This may be an indication that maximizing the variance is not a well-suited method for this particular problem, which was also indicated by the high variation within groups in comparison to the variation between groups. PCA maximizes the variance without specifically considering the variance between the two groups in the data. PC2 shows the largest difference between the two groups, see Figure 4. The first five principal components explain 98% of the total variance of the pellet mean values, and still the result of the discriminant analysis on these five components rendered worse classification in comparison to using all 20 variables in the plain data.

In most of the tests LDA performed better than QDA, this is interpreted such that the problem is linear. This could also be explained by the fact that QDA uses two covariance matrices, instead of LDA's one matrix, and in combinations with 20 dimensions this could make QDA a bit numerically unstable. LDA also showed better results than SVM for the EcoLife20 pellets. For AquaLife R90, LDA or QDA showed better results than SVM for classification using the pellet mean or pellet median values, while the classification using PC1–5 on the pellet means showed better results using SVM than the LDA and QDA. The LDA and QDA are quite quick to compute, while the SVM takes significantly longer time.

To sum up, the results show that it is possible to distinguish between feed pellets with and without inclusion of synthetic astaxanthin in the coating using multispectral image analysis. However, more work is needed in order to

make the method robust for different pellet types and also for different amounts of astaxanthin. Since astaxanthin is expensive, it is important to have a good accuracy in the method. This will later be of importance for developing rapid and noninvasive on-line quality food and feed products with optimal use of pigment and minimum amount of waste.

ACKNOWLEDGMENTS

This work has received funding from BioMar A/S and the EU under the Seventh Framework Programme FP7/2007–2013 under Grant Agreement No. 214505.10.

REFERENCES

- A. P. Simopoulos, "Omega-3 fatty acids in health and disease and in growth and development," *Am. J. Clin. Nutr.* **54**, 438–463 (1991).
- O. J. Torrissen, R. W. Hardy, and K. D. Shearer, "Pigmentation of salmonids—Carotenoid deposition and metabolism," *Rev. Aquat. Sci.* **1**, 209–225 (1989).
- R. T. M. Baker, A.-M. Pfeiffer, F.-J. Schöner, and L. Smith-Lemmon, "Pigmenting efficacy of astaxanthin and canthaxanthin in fresh-water reared Atlantic salmon, *Salmo salar*," *Anim. Feed Sci. Technol.* **99**, 97–106 (2002).
- M. G. Ljungqvist, S. Frosch, M. E. Nielsen, and B. K. Ersbøll, "Analysis of astaxanthin in fish feed pellets," *Proc. West European Fish Technologists Association* (Ege University, Izmir, 2010), Vol. 40, pp. 59–60.
- D. D. Gomez, L. H. Clemmensen, B. K. Ersbøll, and J. M. Carstensen, "Precise acquisition and unsupervised segmentation of multi-spectral images," *Comput. Vis. Image Underst.* **106**, 183–193 (2007).
- L. H. Clemmensen and B. K. Ersbøll, "Multispectral recordings and analysis of psoriasis lesions," *Proc. 9th MICCAI 06—Workshop on Biophotonics Imaging for Diagnostics and Treatment*, October 6 (Technical University of Denmark, Kgs Lyngby, 2006).
- L. H. Clemmensen, M. E. Hansen, J. C. Frisvad, and B. K. Ersbøll, "A method for comparison of growth media in objective identification of penicillium based on multi-spectral imaging," *J. Microbiol. Methods* **69**, 249–255 (2007).
- B. S. Dissing, L. K. H. Clemmensen, B. K. Ersbøll, H. Løje, and J. Adler-Nissen, "Temporal reflectance changes in vegetables," *2009 IEEE 12th Int. Conference on Computer Vision Workshops (ICCV Workshops)* (IEEE, Washington DC, 2009), pp. 1917–1922.
- L. H. Clemmensen, M. Hansen, and B. K. Ersbøll, "A comparison of dimension reduction methods with application to multi-spectral images of sand used in concrete," *Mach. Vis. Appl.* **21**, 959–968 (2010).
- B. S. Dissing, M. E. Nielsen, B. K. Ersbøll, and S. Frosch, "Multispectral imaging for determination of astaxanthin concentration in salmonids," *PLoS ONE* **c6**, e19032 (2011).
- H. Løje, B. S. Dissing, L. K. H. Clemmensen, B. K. Ersbøll, and J. Adler-Nissen, "Multispectral imaging of wok fried vegetables," *Proc. Scandinavian Workshop on Imaging Food Quality*, May 27 (Technical University of Denmark, Kgs Lyngby, 2011), pp. 59–62.
- O. Unay, B. Gosselin, O. Kleynen, V. Leemans, M.-F. Destain, and D. Debeir, "Automatic grading of bi-colored apples by multispectral machine vision," *Comput. Electron. Agric.* **75**, 204–212 (2011).
- K. Kobayashi, Y. Matsui, Y. Maebuchi, T. Toyota, and S. Nakachi, "Near infrared spectroscopy and hyperspectral imaging for prediction and visualisation of fat and fatty acid content in intact raw beef cuts," *J. Near Infrared Spectrosc.* **18**, 301–315 (2010).
- J. P. Wold, F. Westad, and K. Heia, "Detection of parasites in cod fillets by using SIMCA classification in multispectral images in the visible and NIR region," *Appl. Spectrosc.* **55**, 1025–1034 (2001).
- T. Brosnan and D.-W. Sun, "Improving quality inspection of food products by computer vision—A review," *J. Food Eng.* **61**, 3–16 (2004).
- K. C. Lawrence, W. R. Windham, B. Park, D. P. Smith, and G. H. Poole, "Comparison between visible/NIR spectroscopy and hyperspectral imaging for detecting surface contaminants on poultry carcasses," *Proc. SPIE* **5271**, 35–42 (2004).
- S. K. Stormo, A. H. Sivertsen, K. Heia, H. Nilsen, and E. Elvevoll, "Effects of single wavelength selection for anisakid roundworm larvae

- detection through multispectral imaging”, *J. Food Prot.* **70**, 1890–1895 (2007).
- ¹⁸ B. Park, Y. R. Chen, and M. Nguyen, “Multi-spectral image analysis using neural network algorithm for inspection of poultry carcasses”, *J. Agric. Eng. Res.* **69**, 351–363 (1998).
- ¹⁹ J. Folm-Hansen, “On chromatic and geometrical calibration”, Ph.D. thesis (Technical University of Denmark, 1999).
- ²⁰ B. S. Dissing, J. M. Carstensen, and R. Larsen, “Multispectral colormapping using penalized least square regression,” *J. Imaging Sci. Technol.* **54**, 0304011 (2010).
- ²¹ T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, 2009).
- ²² P. Geladi, H. Isaksson, L. Lindqvist, S. Wold, and K. Esbensen, “Principal component analysis of multivariate images”, *Chemom. Intell. Lab. Syst.* **5**, 209–220 (1989).
- ²³ A. Rinnan, F. van den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra”, *TrAC Trends Anal. Chem.* **28**, 1201–1222 (2009).
- ²⁴ T. Joachims, *Making Large-Scale SVM Learning Practical. Advances in Kernel Methods—Support Vector Learning* (MIT, Cambridge, MA, 1999).
- ²⁵ M. Buchwald and W. P. Jencks, “Optical properties of astaxanthin solutions and aggregates”, *Biochemistry* **7**, 834–843 (1968).
- ²⁶ J.-P. Yuan and F. Chen, “Identification of astaxanthin isomers in *Haematococcus lacustris* by HPLC-photodiode array detection”, *Biotechnol. Tech.* **11**, 455–459 (1997).