

Color Calibration of Scanners for Scanner-Independent Grain Grading¹

Muhammad A. Shahin² and Stephen J. Symons^{2,3}

ABSTRACT

Cereal Chem. 80(3):285–289

Scanner technology is emerging as a cost-effective and robust imaging alternative to camera-based systems in many applications. However, scanner technology is changing so fast that image quality can vary from model to model. It is critical that images scanned with different scanners be brought to a common basis for processing and measurement through a calibration process that eliminates scanner-to-scanner variability. The focus of this research was to investigate scanner-to-scanner variability and develop color correction or mapping functions to allow for machine-independent grain inspection. Various makes and models of scanners

were compared for optical and color characteristics. Three different color correction methods were evaluated: grayscale (GS) transformation, red-green-blue (RGB) transformation, and histogram matching. All three models of color correction worked within satisfactory tolerance for a multicolor Q60 chart. However, for grain samples of a limited color range, the histogram matching approach performed better than GS and RGB transformations for scanner calibration. The color-corrected test images matched the reference images within 3 grey values. Differences between the three models of color correction are discussed.

Color and appearance are important factors in grading grains. A machine vision system can provide an objective measure of color and appearance of grains in contrast to visual inspection, which by nature is subjective. Shahin and Symons (1999) developed a flatbed scanner-based imaging system that graded lentils by color with high accuracy. Scanner-based systems are relatively inexpensive and less sensitive to ambient conditions than camera-based systems. Images produced with different scanners exhibit wide variations in terms of brightness and color distribution. This implies that a software application developed for a particular scanner may not produce valid results for another scanner unless the images from each scanner are brought to a common basis.

The ability to be able to use different scanners is crucial from a practical standpoint. Certainly, a scanner currently being used in an application cannot function forever—a replacement will be necessary at some point in time. Moreover, scanner-to-scanner transferability will allow users to take advantage of new innovations among the many scanner manufacturers. Shahin and Symons (2000) developed an ad hoc method for matching images scanned with different scanners. Several models of scanners from different manufacturers were compared and matched using images of grain samples as the references. It was reported that images captured with any of the test scanners could be matched to those from a reference scanner. However, each scanner required a number of mapping functions for grains with different visual characteristics. For example, a mapping function generated for large-green (Laird) lentils also worked for small-green (Eston) lentils and other green lentils but brown lentils and chickpeas required separate mapping functions. This calibration scheme was not an ideal practical solution because a) several mapping functions were required for each scanner and b) the reference was a grain sample that varied from calibration to calibration. The reference grain samples can change color over time; green lentils change color from green to brown as they age. The preferred method is to use standard color charts for generating reference images because they are commercially available and do not change over time.

Shahin and Symons (2000) also tried a Q60 color chart for generating a universal mapping function for each of the test scanners. This approach worked well when tested on Q60 images, however,

it failed to match different scanners when used on grain images. This research reports another method of using Q60 reference images for scanner calibration. Such a color-correction or calibration would allow for scanner-to-scanner transferability. The focus of this research was to develop color correction or mapping functions to allow for scanner-independent grain inspection. Two specific objectives of this study were established. First, develop color calibration algorithms for matching different scanners to a reference scanner using a Q60 color chart as the test object. Second, evaluate the performance of these algorithms for images of Q60 chart and grain samples of different color characteristics.

Color images can be captured through cameras or scanners. Calibration of a color-capturing device is essential as the spectral sensitivity of a digitizing system is not always identical to the human color-matching functions. Calibration can be abstracted as a modeling problem from a device-dependent color space to a device-independent and uniform color space. Chang and Chen (2000) proposed a colorimetric model to relate acquired color data from camera-based imaging systems to the true color data (measured with a spectrophotometer). They achieved a “reasonably accurate” model for vision systems through spectral matching, reference-white luminance determination, and polynomial regression.

In desktop publishing, color matching from scanner to print has been a major issue since the beginning of photographic reproduction. Trussell and Sharma (1994) used signal-processing techniques to estimate spectral sensitivity of scanners for color calibration of these devices. They concluded that the method of projections onto convex sets provided a better estimate of the sensitivity function than the principal eigenvector method. However, the observed error (δE) values were high because of scanner nonlinearity and noise in the measured reflectance spectra. Herbert (1993) used colorimetric calibration from input scanners into CIE-XYZ color space and to various output processes like color prints in a production environment. He too experienced large errors in the prints when compared with the target, which he attributed to the measurement error associated with the spectrophotometer and scanner repeatability. He remarked “keeping spectrophotometer in calibration is a whole other task.”

Many color correction algorithms have been reported in the literature for color device calibration using mathematical formulation (Vrhel and Trussell 1999a), statistical modeling (Hu and Mojsilovic 2000; Sanchez and Binefa 2000), and neural networks (Cho and Kang 1995; Vrhel and Trussell 1999b). Some researchers combined neural networks with either genetic algorithms (Watanabe et al 2001) or fuzzy logic (Mizutani and Nishio 2002) for improved performance. However, histogram comparison remains the most commonly used technique for comparing color similarity of images due to its compact representation and low complexity (Hashizume et al 1998; Yining et al 2001; Zhang et al 2001).

¹ Paper 843 of the Grain Research Laboratory, Canadian Grain Commission, 1404-303 Main Street, Winnipeg, Manitoba Canada, R3C 3G8.

² Canadian Grain Commission, Grain Research Laboratory, Winnipeg, Manitoba, Canada.

³ Corresponding author. E-mail: ssymons@grainscanada.gc.ca.

MATERIALS AND METHODS

Four scanners from three scanner manufacturers were included in this study (Table I). Scanner 1 was used as the reference scanner whereas scanners 2–4 were used as the test scanners requiring calibration. Scanners 1 and 3 were from the same manufacturer. Images of a Q60 (Kodak, Canada) color chart were acquired from all four scanners for calibration purposes. A 178- × 125-mm window enclosing the entire Q60 chart was imaged at 100 dpi. The Q60 image from the reference scanner was the reference image, whereas images from the test scanners were called test images. In each case, the reference and test images were gathered at the scanners' default settings with the contrast and color correction options disabled.

Scanner Calibration

For scanner calibration, gray values in each of the three RGB channels were measured at predefined locations on the reference image and a test image. A macro was written to make these measure-

ments using a commercial image analysis software (KS400 v. 3.0, Carl Zeiss Vision, Germany). Using these measurements, mapping functions were developed to transform RGB values of the test image such that the color distribution of the resultant corrected image closely matched the color distribution of the reference image. Three methods were used to develop mapping functions: 1) grayscale transformation, 2) color transformation, and 3) histogram transformation.

Grayscale (GS) transformation

The GS strip on the Q60 chart consisted of 24 patches changing from white to black. Average RGB values along the GS strip were measured from the reference and test images to generate the GS look-up table (LUT). The GS strip provided RGB values over the full range needed for scanner calibration. The GS-LUT contained 256 gray levels (0–255) determined from 24 measurements. Because the RGB values in the GS strip changed monotonically, a line joining 24 measurement points was assumed for interpolation and extrapolation in the LUT generation.

Color (RGB) Transformation

Following GS transformation, the mean RGB values in 228 color patches on the Q60 were measured from the GS-corrected test images. These color patches provided RGB values over a wide range of color. Corresponding RGB values were also measured from the reference image to generate the RGB-LUT. The GS-LUT would correct most differences between charts, while the RGB-LUT would correct any remaining differences not corrected during the GS transformation. Due to the wide range of RGB

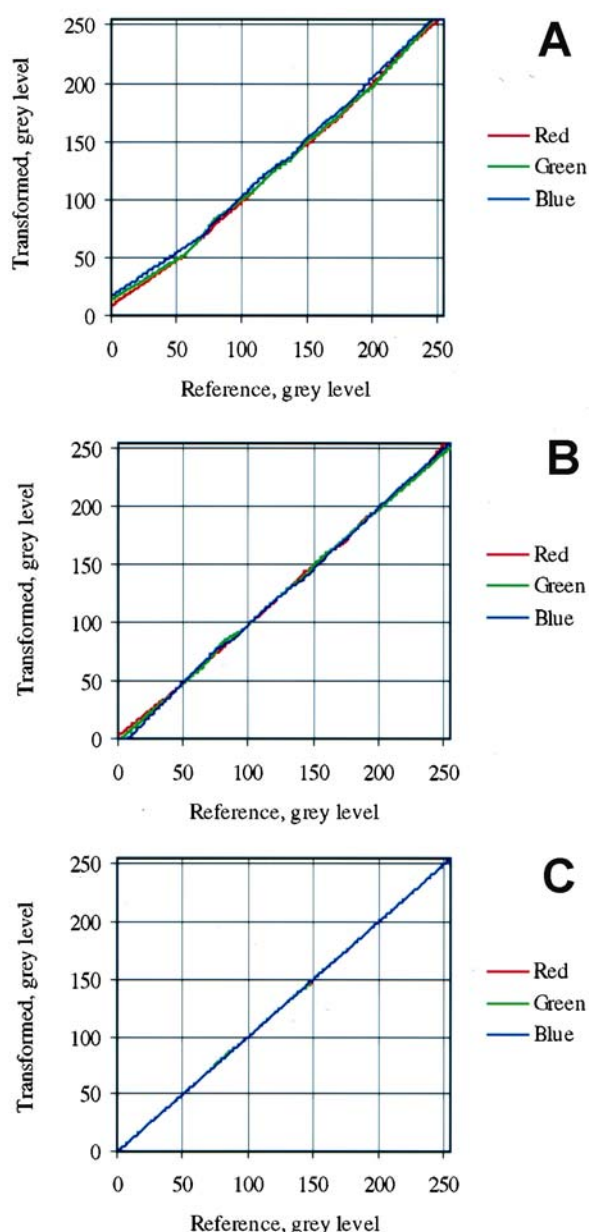


Fig. 1. Comparative performance of three calibration look-up tables (LUT) for Q60 image. Grayscale transformation (A); RGB transformation (B); histogram matching (C).

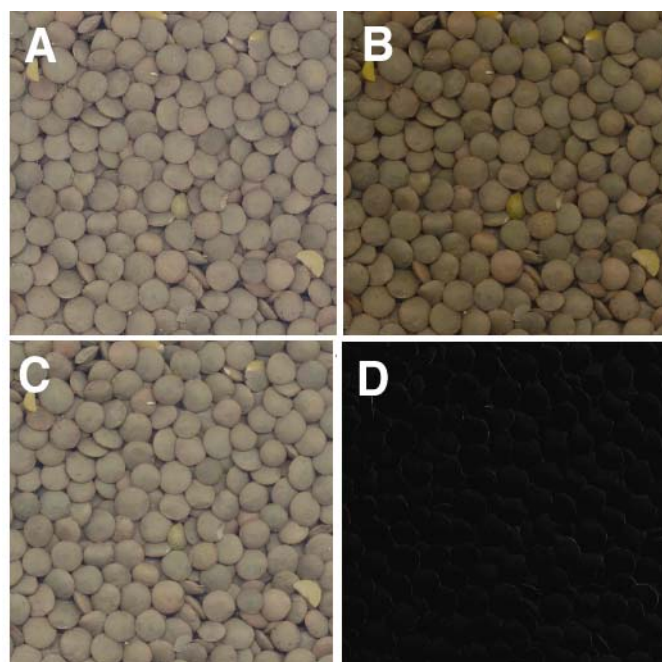


Fig. 2. Color correction results using histogram-matching transformation: A, image of a lentil sample scanned with the reference scanner; B, the same sample scanned with a test scanner; C, image from (B) after color correction; D, difference in the reference (A) and corrected (C) images.

TABLE I
Various Makes and Models of Scanners Used in This Study Along with Their Relative Cost

Scanner	Make	Model	Relative Cost ^a
1 (scanner)	Microtek	ScanMaker 4	++++
2	Umax	Astra 4000U	+++
3	Microtek	ScanMaker X6	+
4	Canon	CanoScan FB1200S	++

^a +, least expensive; +++++, most expensive

combinations in the different color patches, the mapping was not as smooth as for the GS transformation. To achieve a smooth mapping, a second-order polynomial fit was assumed for interpolation and extrapolation in generating the RGB-LUT.

Histogram Matching (HM) Transformation

Instead of measuring RGB values at prespecified locations on the Q60 image, the entire Q60 was measured as a single object. Histogram percentiles from the reference and test images were used to generate the HM-LUT. RGB values corresponding to 5, 10, . . . , and 95 percentiles extracted from the test image were mapped to respective RGB levels in the reference image. The histogram matching would cause the color distribution of the corrected image closely resemble that of the reference image. Histogram matching used here slightly differed from the method of histogram specification described by Gonzalez and Woods (1992). Instead of using histogram-equalized images for generating mapping functions, raw images were used because of totally different color ranges of the Q60 and grain images.

Performance Evaluation

Performance of all three calibration LUT for each of the test scanners was evaluated using the Q60 images. Histogram percentiles of a corrected test image were plotted against those of the reference image. A 1:1 correlation between the reference and the

corrected test image would indicate perfect color matching (Fig. 1C). Parameters of a straight line fit to the data (slope, y-intercept, R^2), as well as the mean squared error (MSE), maximum absolute error (MaxAE), and number of gray values mapped incorrectly (Count) were computed as numerical measures of performance (Table II). An error of >3 gray values was considered incorrect mapping based on system repeatability tests. Repeated scans of different commodities (lentils, chickpeas, and peas) showed that the mean values for the RGB and hue light saturation (HLS) channels in the images were within 3 GS (unpublished data).

These LUT were also tested using images of different grain samples (lentils, chickpeas, and peas). In the grain images, corrected test images were visually observed and compared with the reference images. The reference and a corrected image were subtracted to form a difference image and used as a qualitative measure of color matching (Fig. 2). A great deal of care was taken to maintain the position and orientation of seeds in the images gathered with different scanners. However, slight variations in the seed arrangement were unavoidable. For quantitative comparison, the mean values of the RGB and HLS channels were measured in both the reference and corrected test images using the best LUT for each of the test scanners. Comparison of the RGB and HLS measurements to determine the goodness of match between corrected and reference images was made using Tukey's multiple comparisons (Table III). To minimize the effect of within sample variability, image areas corresponding to shadows and off-colored seeds were excluded from the measurements. Details of image processing technique for separating damaged and peeled seeds from lentil images can be seen in Shahin and Symons (2001).

Considering the complexity of the color space, it would be unrealistic to expect perfect color matching for all grain samples. Color corrected images would always represent the best approximation of the reference. The fact that matters the most in imaging-based grain grading is how does the remapping of colors affect the decision process for grade determination. To ensure this, samples of lentils for two color grades were scanned with different scanners and the color (hue) of the corrected test images was compared with that of the reference image. Analysis of variance (ANOVA) and Tukey's multiple comparisons were applied to test for statistical differences among scanners (Table IV).

TABLE II
Error in Color Matching for Look-Up Tables
Generated with Different Methods of Color Correction

Error Measured ^a	Color Channel		
	Red	Green	Blue
Grayscale transformation			
MSE	9.2	16.8	39.5
MaxAE	9	13	17
Count	70	65	126
Slope	0.9998	0.9859	0.9851
y-intercept	0.1844	3.1793	6.7687
R ²	0.9983	0.9974	0.9973
RGB transformation			
MSE	2.7	4.2	6.9
MaxAE	4	4	7
Count	4	9	44
Slope	0.9985	0.995	1.0127
y-intercept	0.6432	-0.9646	-3.3481
R ²	0.9996	0.9997	0.9995
Histogram transformation			
MSE	0.01	0.09	0.02
MaxAE	1	1	1
Count	0	0	0
Slope	1.0	0.9995	1.0
y-intercept	-0.0042	0.1315	0.0149
R ²	1.0	1.0	1.0

^a MSE, mean squared error; MaxAE, maximum absolute error; Count, number of gray levels incorrectly mapped with an error of more than 3 gray values.

TABLE IV
Mean Hue Values Measured from Images of Two Lentil Samples
of Different Color Grades Acquired with Different Scanners

Scanner	Color Grade	
	Good	Fair
1 (reference)	24.86a	15.66b
2	25.31a	16.75b
3	25.87a	17.06b
4	26.10a	16.85b

^a Values followed by the same letter are not significantly different ($P < 0.05$).

TABLE III
Mean RGB and HLS Values Measured from Reference and Corrected Images Acquired with Different Scanners

Grain	Scanner	Red	Green	Blue	Hue	Light	Saturation
Lentil	1 (reference)	140.55a	128.06b	110.03c	24.86e	125.04f	31.53h
	2	138.20a (97.20)	125.12b (80.88)	105.62d (54.67)	25.31e (25.85)	121.66g (75.69)	34.20hi (71.75)
	3	141.19a (117.47)	127.72b (96.70)	106.80cd (66.56)	25.87e (24.74)	123.75fg (91.76)	36.30i (71.85)
	4	137.93a (97.70)	126.29b (86.36)	107.34cd (59.73)	26.10e (29.72)	122.38fg (78.16)	31.97h (63.56)
Chickpea	1 (reference)	174.32jk	149.31lm	127.32n	19.38o	150.55pq	58.40r
	2	172.48j (125.72)	147.28i (95.36)	123.71n (68.24)	20.03o (19.35)	147.84p (96.73)	59.17r (78.62)
	3	176.52k (146.10)	151.85m (111.50)	125.51n (80.49)	19.83o (19.32)	152.77q (113.05)	60.9r (86.32)
Pea	1 (reference)	149.65s	146.73t	132.89u	35.13w	141.06x	18.92y
	2	150.34s (101.98)	147.33t (95.23)	128.00v (72.84)	36.47w (32.69)	139.00x (87.17)	21.73y (65.69)
	3	151.55s (116.05)	148.13t (107.39)	130.55uv (85.84)	36.26w (30.62)	140.97x (100.71)	21.43y (39.79)

^a Numbers in parentheses represent mean grey-values of test images measured before color correction. Values followed by the same letter are not significantly different ($P < 0.05$).

RESULTS AND DISCUSSION

For each of the test scanners, all three calibration LUT performed well on the Q60 images. When observed visually, the corrected image in each case looked very similar to the reference image. Figure 1 presents a qualitative measure of performance of the three LUT on the Q60 image for one of the test scanners (scanner 2). Table II lists quantitative measures of performance of the three LUT. For both the GS and RGB transformations, the corrected test images apparently matched the reference image closely with very high R^2 values and slopes nearing the ideal value of 1.0. The GS transformation corrected most of the differences between the reference and the test images, although deviated at low gray values as shown by the curve (Fig. 1A) and y -intercept values (Table II). MaxAE for the red, green, and blue channels was as high as 9, 13, and 17, respectively. A large number of the 256 gray levels in all three color channels (70 for red, 65 for green, 126 for blue) were mapped to values that differed from the desired value by >3 gray levels. The RGB transformation improved the linearity of the color correction, but minor deviations from the reference remained (Fig. 1B). MaxAE for the red, green and blue channels (4, 4, and 7 gray levels, respectively) was lower than for the GS transformation (Table II). Some of the gray levels in all three color channels (4 for red, 9 for green, 44 for blue) were mapped to values that differed from the desired value by more than 3 gray levels. For both the GS and RGB transformations, the blue channel had the largest error. The histogram matching method gave a 1:1 correlation ($R^2 = 1.0$ in R, G, and B channels) between the corrected test image and the reference indicating perfect color matching (Fig. 1C). Only a few gray levels in each of the color channels were incorrectly mapped giving MaxAE of 1 gray level (Table II). Reflecting a high degree of color matching, the MSE was also much lower than the other two transformations. A similar behavior was observed for all the test scanners.

All three transformations appeared to be similar when a 1st order polynomial was fit to the data plotted in Fig. 1. An $R^2 \geq 0.9973$ was found for each of the RGB channels, indicating a good fit for all the transformations that did not conform to the measures of error (MSE, MaxAE, and Count) listed in Table II. The value for the slope was also very similar in all cases. The y -intercept, on the other hand, generally correlated well with measures of error in color matching with one exception. The green channel after histogram matching showed larger than expected y -intercept values despite low error values. This suggests that the MSE, MaxAE, Count, and y -intercept can be used (individually or in some combination) as a quantitative indicator of the performance of the calibration process. Figure 1 provided a quick visual indication of the accuracy of the calibration achieved. Deviations from the desired 1:1 correlation between the reference and transformed gray levels could be readily observed from the colored graph as an indication of "not so good" color correction. Indistinguishable overlapping straight lines at $\approx 45^\circ$ indicated a perfect color correction.

When tested on the grain images, both the GS and RGB transformations collapsed. The corrected image in both cases looked quite different in color from the reference. The difference between the corrected and reference images did not seem so severe for chickpeas (cream color), but the difference was strikingly obvious for lentils and peas (green color). Similar results were observed for each of the test scanners. Poor performance of the GS-LUT and RGB-LUT on the grain images may be attributed to the dramatic difference in the color ranges between the images of Q60 and grain samples. Both LUT matched the broad bandwidth Q60 images from which they were generated but subsequently failed to correct subtle color variations in the narrow bandwidth grain images. The HM-LUT did not appear to suffer from this limitation.

Histogram matching worked as well for grain images as it did for the Q60 images. The images of lentils, chickpeas, and peas corrected with the HM-LUT looked similar to the respective refer-

ence images. Images of a lentil sample obtained from two scanners are shown in Fig. 2. The reference image (Fig. 2A) and the corrected test image (Fig. 2C) looked very similar, for the most part. However, slight differences along the seed boundaries were observed (Fig. 2D). Optical and physical factors may be responsible for these differences. In practice, images produced with different scanners may vary slightly in sharpness due to inherent differences in the light detectors used and differences in the depth of focus. A sharp image would have crisp seed boundaries as opposed to fuzzy boundaries, leading to edges in the difference image as seen in Fig. 2D. Slightly different seed arrangement due to movement of the samples between scanners may also contribute to differences found. A similar phenomenon was observed for chickpea and pea images as well (not shown). All the test scanners showed similar results.

The mean values of the R, G, B, H, L, and S channels measured from corrected grain images closely matched the values measured from the reference images when the HM-LUT was applied (Table III). Although the off-colored seeds and shadows were excluded from the measurements, minor deviations from the reference were observed, as expected. These deviations are much smaller than the difference between the reference and the uncorrected images (numbers in parentheses in Table III represent uncorrected gray values). Tukey's multiple comparisons at $\alpha = 0.05$ showed that, for the most part, all the test scanners matched the reference scanner after color correction. Scanner 2 slightly differed from the reference in blue (two instances) and light (one instance) channels. Scanner 3 differed in saturation channel (one instance). These deviations in color, however, were within acceptable instrumental variations. Data in Table IV shows that scanner-to-scanner variations after color matching were significantly smaller than grade-to-grade color variation for lentil grading ($P < 0.0001$). This indicates that the differences between scanners after color correction were not of practical significance. Statistically, all the test scanners after calibration were found to be the same as the reference scanner ($\alpha = 0.05$). This means that, after color matching, all four scanners would correctly assign the same grade to a sample.

These results show that images captured with any of the test scanners can be matched to the reference images. An imaging application developed using a scanner system can be replicated onto other scanner platforms. Scanner selection for a particular application will not be restricted to model or manufacturer. For online grain inspection applications, factors such as desired accuracy, repeatability, speed of operation, and compatibility with the application software will dictate the selection of the scanner. For offline applications, however, any less expensive scanner fulfilling accuracy and repeatability specifications could be used.

The scanner-matching scheme presented in here is simple and practical as a standard color chart is used for the reference images. The use of a graphical tool such as Fig. 1 makes it very easy to monitor and assess whether or not a scanner is properly calibrated. A 1:1 correlation between the reference and corrected test images of the Q60 chart will indicate perfect matching with the reference. Any deviations from the desired matching can be detected at a glance. In addition, quantitative measures of error discussed here can provide a summary of deviations from the target. Tolerance limits on these deviations still need to be defined and will be covered in a planned future study.

The calibration technique presented here can also be applied to calibrate camera-based imaging systems where light changes relatively frequently. In many cases, this calibration technique may be satisfactory. For applications demanding more accurate mapping of each of the RGB and HLS channels, a second-step calibration may be needed to correct for hue and chroma. Desktop publishing applications would require a third step gamut mapping to CMYK color space for printing. Microscopy and side-lit macro systems may also benefit from this calibration technique with some minor alterations to suit a particular application.

CONCLUSIONS

Based on the results of this study, it can be concluded that 1) different scanners can be matched to a reference scanner by histogram-matching-based mapping functions using a standard Q60 color chart for reference images; and 2) the histogram-matching-based mapping function developed from Q60 performed equally well on the images of different grain samples.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of Saskatchewan Pulse Growers and Pulse Canada for funding this study.

LITERATURE CITED

- Chang, G. W., and Chen, Y. C. 2000. Colorimetric modeling for vision systems. *J. Elec. Imag.* 9:432-444.
- Cho, M., and Kang, B. 1995. Device calibration of a color image scanner digitizing system by using neural networks. *IEEE Con. Proc. Neural Networks* 1:59-62.
- Gonzalez, R. C., and Woods, R. E. 1992. *Digital image processing*. Addison-Wesley: New York.
- Hashizume, C., Vinod, V. V., and Murase, H. 1998. Robust object extraction with illumination-insensitive color description. *IEEE Conf. Proc. Image Processing* 3:50-54.
- Herbert, F. 1993. CIE color space and IT* at work: A quantitative analysis of color matching from scanner to print in a production environment. *SPIE* 1909:168-177.
- Hu, J., and Mojsilovic, A. 2000. Optimal color composition matching of images. *IEEE Conf. Proc. Pattern Recognition* 4:47-50.
- Mizutani, E., and Nishio, K. 2002. Multi-illuminant color reproduction for electronic cameras via CANFIS neuro-fuzzy modular network device characterization. *IEEE Trans. Neural Networks* 13:1009-1022.
- Sanchez, J. M., and Binefa, X. 2000. Color normalization for appearance based recognition. *IEEE Conf. Pattern Recognition* 1:1815.
- Shahin, M. A., and Symons, S. J. 2001. A machine vision system for grading lentils. *Can. Biosystems Eng.* 43:7.7-7.14.
- Shahin, M. A., and Symons, S. J. 2000. Comparison of scanners for grain grading by image analysis. Paper No. 00-3096. ASAE: St. Joseph, MI.
- Shahin, M. A., and Symons, S. J. 1999. A computer vision system for color classification of lentils. Paper No. 99-3202. ASAE: St. Joseph, MI.
- Trussell, H. J., and Sharma, G. 1994. Signal processing methods in color calibration. *SPIE* 2170:18-23.
- Vrhel, M. J., and Trussell, H. J. 1999a. Color device calibration: A mathematical formulation. *IEEE Trans. Image Processing* 8:1796-1806.
- Vrhel, M. J., and Trussell, H. J. 1999b. Color scanner calibration via neural network. *IEEE Conf. Proc. Acoustic, Speech, and Signal Processing*. 6:3465-3468.
- Watanabe, T., Kojima, A., Kuwahara, Y., and Kurosawa, T. 2001. High quality color correction method combining neural networks with genetic algorithms. *IEEE Conf. Proc. Image Processing* 1:553-556.
- Yining, D., Manjunath, B. S., Kenny, C., Moore, M. S., and Shin, H. 2001. An efficient color representation for image retrieval. *IEEE Trans. Image Processing* 10:140-147.
- Zhang, M., Xie, J., Li, Y., and Wu, D. 2001. Color histogram correction for panoramic images. *IEEE Conf. Proc. Virtual Systems and Multimedia* 328-331.

[Received March 13, 2002. Accepted November 18, 2002.]