Article

PhotoMetrix: An Application for Univariate Calibration and Principal Components Analysis Using Colorimetry on Mobile Devices

Gilson A. Helfer,[a] Vinícius S. Magnus,[b] Fernanda C. Böck,[a] Aline Teichmann,[a] Marco F. Ferrão[c] and Adilson B. da Costa*,[a]

[a]Programa de Pós-Graduação em Sistemas e Processos Industriais, Universidade de Santa Cruz do Sul, 96815-900 Santa Cruz do Sul-RS, Brazil

[b]Curso de Análise e Desenvolvimento de Sistemas, Universidade Luterana do Brasil, 95560-000 Torres-RS, Brazil

[c]Instituto de Química, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre-RS, Brazil

This article describes the development of a mobile colorimetric analysis tool. The application, called PhotoMetrix, employs the techniques of simple linear correlation for univariate analysis and principal components analysis (PCA) for multivariate exploratory analysis. The image data are captured by the main camera of the device and converted into red, green and blue (RGB) histograms. As regards the application, the iron determinations were performed in vitamin supplements (univariate module) and differentiation of banknotes was performed by PCA (multivariate module). For the iron determinations, three samples of vitamins at concentrations of 14, 40 and 50 mg of iron *per* tablet were tested and the results were not statistically significant ($p > 0.05$) compared to the reference method. The differentiation of banknotes was performed on Brazilian and Argentinean banknotes. The results showed clustering of the same types of banknotes, and through the loadings graph it was possible to observe the variables through the formation of clusters.

Keywords: univariate calibration, principal components analysis, colorimetry

Introduction

Recent advances in technologies for the acquisition and processing of images as well as the popularization of equipment and mobile applications for such purposes have drawn the attention of the scientific community. The use of digital images has proven effectiveness in applications that were previously restricted to specific equipment such as colorimeters, spectrophotometers, and fluorometers.[1] A digital image is composed of pixels and can be defined by a two-dimensional function of the intensity of light reflected or emitted by a scene, defined as a luminous energy distribution, in which part of that energy is absorbed, transmitted part and reflected part, and these last two parts can be captured by a device.[2]

Monochrome or color digital images are easily acquired by electronic equipment utilizing charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensors. The theory of color perception by the human eye is based on a hypothesis formulated by Thomas Young in 1801, which states that the cones (photosensitive cells in the retina along with the rods) are subdivided into three groups presenting different sensitivity peaks situated around red (R), green (G) and blue (B). Thus, all color sensations perceived by the human eye are combinations of stimulus intensities received by each of these types of cones. Although RGB is the most frequently used color model, there are other models that can be generated from it, such as hue, saturation and value (HSV), hue, saturation and lightness (HSL), and hue, saturation and intensity (HSI).[2] Hue is what most people mean by color, for instance the distinction between red and yellow. Saturation is the amount of the color that is present, for instance the distinction between red and pink. Value, lightness, or intensity is the amount of light, such as the distinction between dark red and light red or between dark gray and light gray. Value is defined as the maximum amount of R, G or B, intensity as the average of the amount of R, G, and B, and lightness

*e-mail: adilson@unisc.br

as the average of the maximum and minimum amounts of R, G or B.[3]

After the image acquisition, several processing techniques can be applied to enhance the visualization of some features of the images; there is an increasing interest in image analysis associated with multivariate analysis tools, such as principal components analysis (PCA), which are appropriate for the interpretation of the multiple types of information contained in an image.[1]

In the most recent years, different research groups have used image analysis as an analytical tool in several distinct areas of knowledge: in food science and technology for visual inspection of fruit[4] and for the development of a portable iPhone-DIC system for the quantification of tetracycline residues in milk;[5] in forensic applications, where the Simon presumptive color test is used in combination with a built-in digital camera on a mobile phone to detect methamphetamine;[6] and for the diagnosis of kidney failure through the ascertainment of creatinine in urine samples.[7]

These tools were also used in the development of a simple, fast, and safe method to identify adulteration of cow's milk based on the building of color histograms in RGB and/or HSI scales.[8] Some authors described the use of a smartphone in a Bradford assay, a method of determining protein concentrations that is used for a variety of applications in fundamental biomedical research, utilizing a smartphone spectrometer,[9] a G-Fresnel device with the dual functionality of focusing and dispersion is used to enable miniaturization.[10]

Although these studies reveal the potential of digital image analysis as an analytical tool in different areas of science, so far no image acquisition device has embedded software to enable image acquisition, processing, and presentation of results through either univariate or multivariate analysis techniques. Therefore, this study proposes the development of an application for univariate calibration and exploratory analysis of multivariate data from the decomposition of digital images acquired by the use of the main camera in Android smartphones and Windows Phone, associated with the use of simple linear regression and PCA, respectively.

Experimental

A PhotoMetrix application has been developed for Windows Phone and Android platforms using C# and Java programming languages, respectively, according to standard design software known as Model-View-ViewModel (MVVM). The MVVM pattern is a way of organizing the code division of a design in manageable layers, which can be independently developed, tested, and modified. This development pattern is especially effective for Windows Presentation Foundation (WPF) and Silverlight and may be adapted to the Android platform. The three main layers of the MVVM pattern are the mode of View (interface), the Model (local database), and the ViewModel (access classes to the local database, image processing algorithms, and other controls). Therefore, the agile method Scrum has been adopted, as one of its goals is to define a process for the design and development of software for environments where requirements emerge and change quickly.[11]

In general, the development best practice code reuse was employed, such as the use of static methods for conversion between color systems and operations on vectors and matrices, the separation of classes by layer, and the structure of projects for Android and Windows Phone following the same names of variables in addition to the routines and functions respecting the particularities of the programming languages C# and Java. With the aim of using this application in other countries, English has been adopted as the standard language of information within interfaces and messages with the user.

Reanalysis in other software was implemented to validate the results outputted by the applications. In order to validate the univariate analysis, Microsoft Excel 2013 (Microsoft Corporation: Redmond, WA, USA) was used, while ChemoStat software version 1.0 was used to validate the PCA.[12]

PhotoMetrix was available free[13] in Windows Phone Store and Google Play Store. In order to assist the understanding of the application usage, tutorial videos were recorded in Portuguese with subtitles in English. Both were shared on YouTube, and the addresses are available through the help buttons ("Help") on the screen of the univariate and multivariate modules and the information button ("About this app") on the main screen.

The main interface of the application includes univariate analysis options, multivariate analysis, configuration, and application information (Figure 1A). In the configuration module (Figure 1B), the user defines the number of samples, and this value can also be set before sampling the size of the region of interest (ROI) to be analyzed, which may be chosen 32×32, 48×48, 64×64 or 96×96 pixels the definition for self-nomination of samples, the optional use of flash, and the e-mail address for receiving the data analysis.

The univariate analysis module allows the user to perform the calibration, sampling (or prediction), search for a previously saved analysis, or access video tutorials for help (Figure 2A). The application performs both
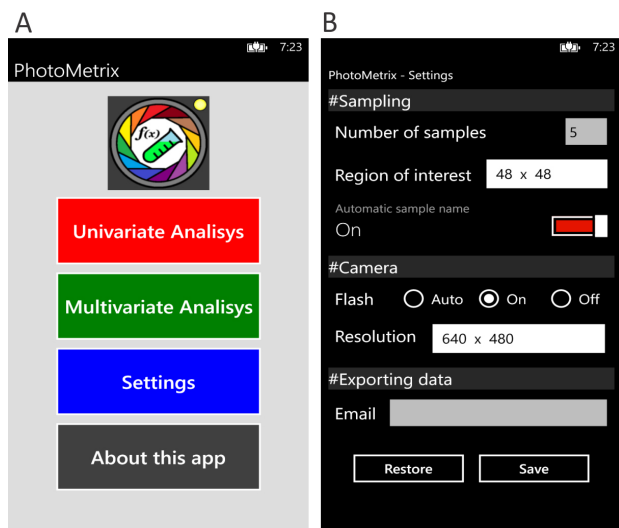
Figure 1. (A) Main graphic interface of PhotoMetrix application with analysis options; (B) configuration interface.

calibration and prediction from the camera image it has captured, where only one ROI is used (Figure 2B). Then the angular, linear, and regression coefficients are calculated. After sampling, the result is predicted in accordance with the calibration to be selected according with straight-line equation for each channel or color component ("channel") (Figure 2C).

The multivariate analysis module allows sampling, reprocessing a previous analysis (for example, using different configurations of variables or color channels), searching for a previously saved analysis, or even accessing video tutorials for help (Figure 3A). The images, which are also obtained by the camera, can be analyzed through their histograms or even the combination of color channels individually selected, having the autoscale and average

concentration as pre-processing options (Figure 3B). Then, singular value decomposition of the data matrix is performed, generating a variation of the main components' results in addition to score and loadings graphics (Figure 3C).

To demonstrate the potential utilization of PhotoMetrix, two different simple experiments were realized. The univariate module determined the iron in vitamin supplements and the multivariate module analyzed the differentiation of banknotes by PCA.

Application 1: determination of iron in vitamin tablets

Experiments

For the analysis of the univariate module, a Sony Xperia SP smartphone operating with Android 4.3, an 8 MP camera, and an ROI of 32×32 pixels were used in the quantification of iron in three samples of vitamin supplement tablets (ferrous sulfate) with concentrations of 14, 40 and 50 mg of iron *per* tablet.

This test method was adapted "standard test method" for iron in trace quantities using the 1,10-phenanthroline method.[14] All experiments were performed using chemical reagents of analytical grade, while distilled and deionized water (> 19 MΩ cm) was used in the preparation of all reagents and solutions.

To prepare the vitamin tablet samples, each tablet was boiled for 15 minutes in 25 mL of 6 mol L⁻¹ hydrochloric acid (Synth, Diadema, Brazil), filtered, and swelled to 100 mL of water. Subsequently, aliquots of 10, 5 and 5 mL from the solutions of 14, 40 and 50 mg tablets, respectively, were diluted to 100 mL in a volumetric flask. A 10 mL aliquot from each sample solution was transferred
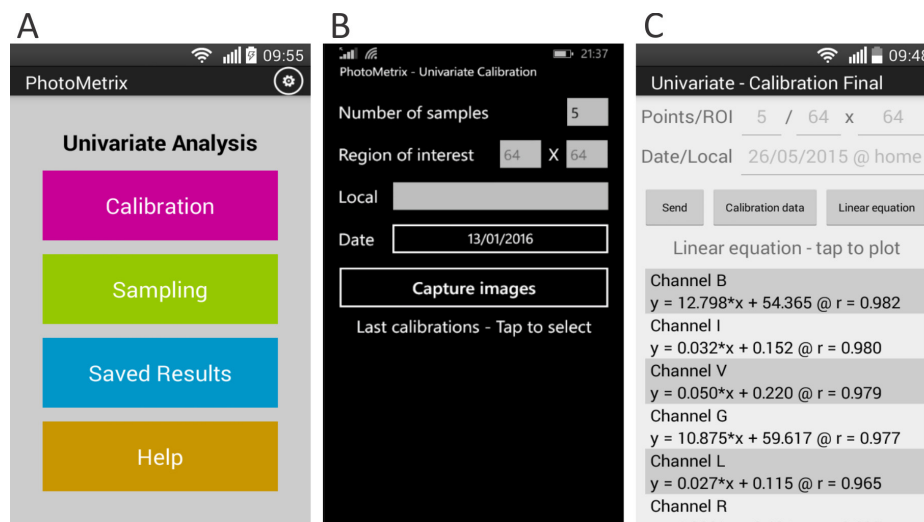
Figure 2. (A) Main interface of univariate analysis module of PhotoMetrix application; (B) configuration interface for number of samples and selection of ROI; (C) interface result of univariate analysis, with the following options: send by e-mail, save the results, and visualization by loading score and graph in the bottom bar.
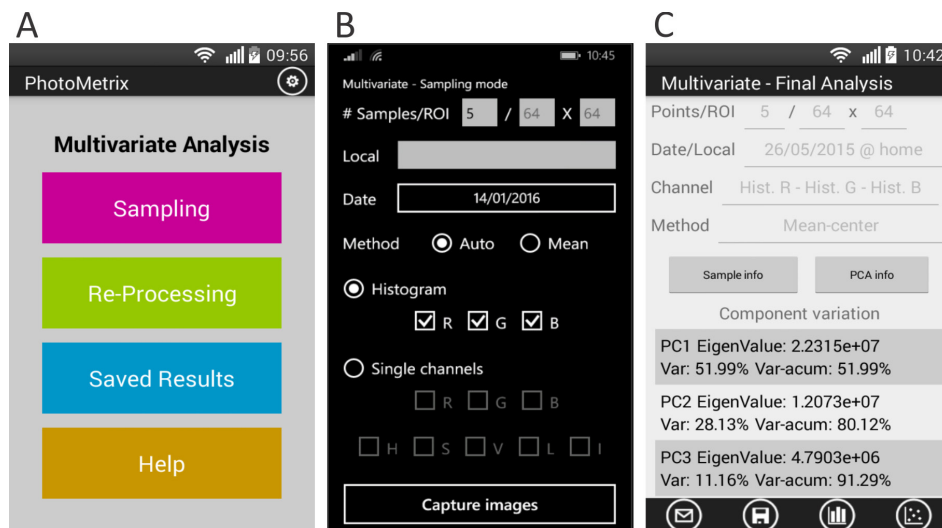
Figure 3. (A) Main PhotoMetrix interface in the multivariate analysis module; (B) multivariate analysis configuration interface; (C) interface of multivariate analysis results, with the following options: send by e-mail, save the results, and visualization by loading score and graph in the bottom bar.

to a 50 mL volumetric flask, and were added 4 mL of 5% solution (m/v) of hydroxylamine hydrochloride (HONH$_2$. HCl, Vetec, Duque de Caxias, Brazil), 4 mL of 2 mol L^{-1} sodium acetate solution (CH$_3$COONa, Vetec, Duque de Caxias, Brazil), and 8 mL of 0.25% (m/v) 1,10-fenantrolina solution (C$_{12}$H$_8$N$_2$.H$_2$O, Sigma-Aldrich, St. Louis, US), and the volume was completed to 50 mL with water. After 15 minutes of reaction, 5 mL aliquots of the samples were respectively transferred to a polypropylene 24-well plate (Nest Biotech Co., Wuxi, China) for the acquisition of images by PhotoMetrix software.

The calibration curve (between 2.4 and 8.8 mg L^{-1}) was prepared from a stock solution of 20 mg L^{-1} ferrous ammonium sulfate (Fe(NH$_4$)$_2$(SO$_4$)$_2$.6H$_2$O, Synth, Diadema, Brazil) and submitted to the same procedure as the samples. The calibration samples and the medication samples were analyzed in triplicate. Images were acquired at a focal length of 6 cm and ambient light of 98 ± 4 Lux (Lux meter LD-500, ICEL, Manaus, Brazil) (Figure 4A). The results obtained by the PhotoMetrix application were compared to those obtained by a FEMTO 600 Plus spectrophotometer (FEMTO, São Paulo, Brazil) at 510 nm using an optical path of 10 mm.

The results obtained through the methodology developed in this research were compared to the results obtained by a reference method employing the one-way analysis of variance (ANOVA) test using the software GraphPad InStat®, version 3.0. The limits of detection (LOD) and quantification (LOQ) were calculated using the mean blank + 3s and mean blank + 10s, where s is the standard deviation of 10 consecutive measurements of the blank. The root mean square difference (RMSD) was used to indicate the average error in the analysis (equation 1),

and the mean relative error (MRE) was determined using equation 2.[15]

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\bar{x}_i - x_i)^2} \qquad (1)$$

$$MRE = \frac{1}{n} \left(\sum_{i=1}^{n} \frac{|\bar{x}_i - x_i|}{x_i} \right) \times 100 \qquad (2)$$

where x_i is the mean of the result obtained by the proposed method (n = 3), $\bar{x}$ is the mean of the true concentration (obtained by the reference method), and n is the total number of replicates.

Application 2: principal component analysis applied to images of banknotes

Experiments

To illustrate the multivariate analysis module using the PCA of digital images, pictures of banknotes presenting different colors, tones, and textures were chosen.

For the analyses that applied the multivariate module it was used a smartphone Motorola Moto X second generation smartphone operating with Android 6.0, a 13 MP camera, an ROI of 48 × 48 pixels, and a focal length of 12 cm was used for the image capture.

A total of 16 banknotes were employed, all from the year 2015 and in good condition without damage; 12 were from Brazil (four of each value: 20, 50, and 100 reais) and four were from Argentina (four of 100 pesos). Only the top right part of the Brazilian bills, which presents the numerical value, was used to acquire the images. For the Argentinean bills, the equivalent part containing the value
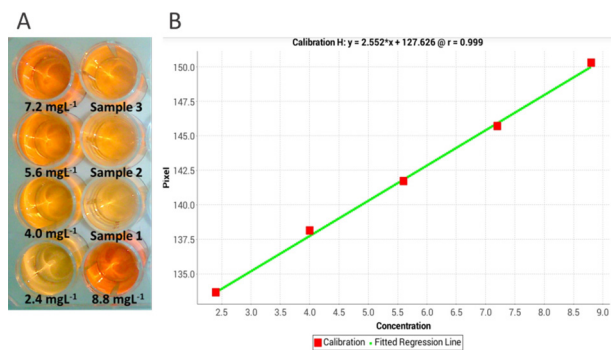
Figure 4. (A) The calibration samples and the medication samples in a polypropylene 24-well plate to acquire the images; (B) result obtained from the application of the calibration curve [amount of hue × iron concentration (mg L^{-1})].

was employed; however, it was the back part of the bill due to the existence of a mirror part, which can interfere with the image acquisition. Two experiments were conducted: the first with only the 12 Brazilian bills and the second including all 16 banknotes.

Results and Discussion

Application 1: determination of iron in vitamin tablets

The best calibration curve was obtained by using the channel H (hue) with a linear correlation coefficient (0.999) and the calibration curve equation y = 2.552x + 127.626 (Figure 4).

The parameters of the calibration curves, LOD, LOQ, and results obtained for the iron determination in vitamin tablets using the reference method and PhotoMetrix method are presented in Table 1. The comparative analysis between the iron concentrations in the samples of vitamins obtained by the reference method and the ones presented by the application for the three samples of medicines demonstrated that even though the application showed a

standard deviation higher than the one presented by the reference method, no significant differences between the two methods of quantification were identified ($p > 0.05$). Moreover, the results of RMSD and MRE were 0.37 mg and 1.04%, respectively; these are similar to the results presented by Zarei *et al.*,[15] who obtained RMSD = 0.17 to 0.37 mg and MRE = 3.6 to 5.3% using chromogenic mixed reagent for principal component artificial neural network (PC-ANN) and multivariate calibration by principal component regression (PCR) and partial least squares (PLS) for determination of iron in pharmaceutical formulations.

It is important to highlight that three factors have a direct influence on the results: lighting, distance between the object and the lens, and the nonlinearities of CMOS sensors. The first case can be mitigated by using flash, when available, and the second by the use of a support (tripod, selfie-stick). The third indicates that a better quality camera (CCD sensor) will allow more accurate estimations of values.

The proposed method uses the same analytical procedure as the reference method, and thus the use of PhotoMetrix does not represent a significant reduction in the consumption of reagents and materials. However, there is a significant gain (over 50%) in the analysis time and mobility, and thus the proposed method is suitable for use as a screening method.

### Application 2 : principal component analysis applied to images of banknotes

Experiment I

Satisfactory results for the PCA of images obtained for Brazilian banknotes were generated when the average values from the R, G, B, H, S, V and I channels were used with the autoscaled data. The first three principal components accounted for 98.65% of the variance of data,

Table 1. Main characteristics of the reference method and the PhotoMetrix method, and the results obtained for the determination of iron in vitamin tablets

	Reference method	PhotoMetrix
Calibration curve equation	y = 0.1844x + 0.0063	Y = 2.552x + 127.628
Linear correlation coefficient (R)	0.999	0.999
Limit of detection (LOD) / (mg L^{-1})	0.02	0.19
Limit of quantification LOQ / (mg L^{-1})	0.04	0.23
Sample 1 (14 mg)	14.3 ± 0.1, n = 3	14.1 ± 0.3, n = 5
Sample 2 (40 mg)	39.5 ± 0.1, n = 3	39.7 ± 0.2, n = 5
Sample 3 (50 mg)	48.9 ± 0.1, n = 3	48.4 ± 1.4, n = 5
Root mean square difference (RMSD) / mg		0.37
Mean relative error (MRE) / %		1.04

and the first two were sufficient to identify the three groups of bills analyzed. In scores from PC1 × PC2, it can be observed that PC1, with 76.14% of the variance, separates the 100 reais samples (samples 1-4) with negative scores, while PC2, with 17.14% of the variance, separates the 20 reais samples (samples 9-12), which present positive scores for PC2, from the 50 reais samples (samples 5-8), which have negative scores for PC2 (Figure 5).
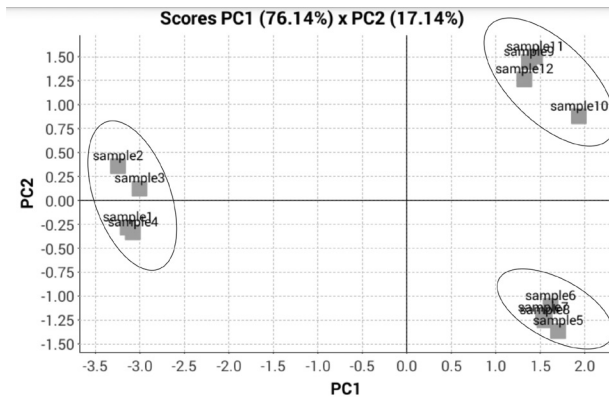

Figure 5. Score graph for image data PCA of Brazilian banknotes.

Through the analysis of the loading graphs it can be seen that the information on the G, B and S channels was responsible for the separation of 100 reais samples (Figure 6A), while channel I was determinant for the separation of the 20 and 50 reais samples (Figure 6B).

Experiment II

Satisfactory results for the PCA of the images acquired when considering the sets of Brazilian and Argentinean banknotes were obtained when the average values of the R, G, B, H, S, V, L and I channels were used with autoscaled data. The first three principal components accounted for 98.75% of the data variance, since any of these combinations allowed the four groups containing the analyzed banknotes to be identified.

Based on the scores of PC1 × PC2 (Figure 7A), PC1 × PC3 (Figure 7B), and PC2 × PC3 (Figure 7C) in PCA based on peso bills (Figure 8A), it was observed that PC1 is responsible for the separation of the Argentinean banknotes from the Brazilian ones, with channels R, G, V, L, and I being most representative for the Argentinean banknotes (samples 1-4). The combination of channels R, S and V presents a higher loading contribution to PC2 (Figure 8B), while the combination of channels B, S, and V does so for PC3 (Figure 8C). Therefore, the determinant channel for the separation of the 20 reais bills (samples 13-16) is the contribution of channel R, while for the 100 reais bills (samples 5-8) it is channel B. The 50 reais bills (samples 9-12) were separated from the
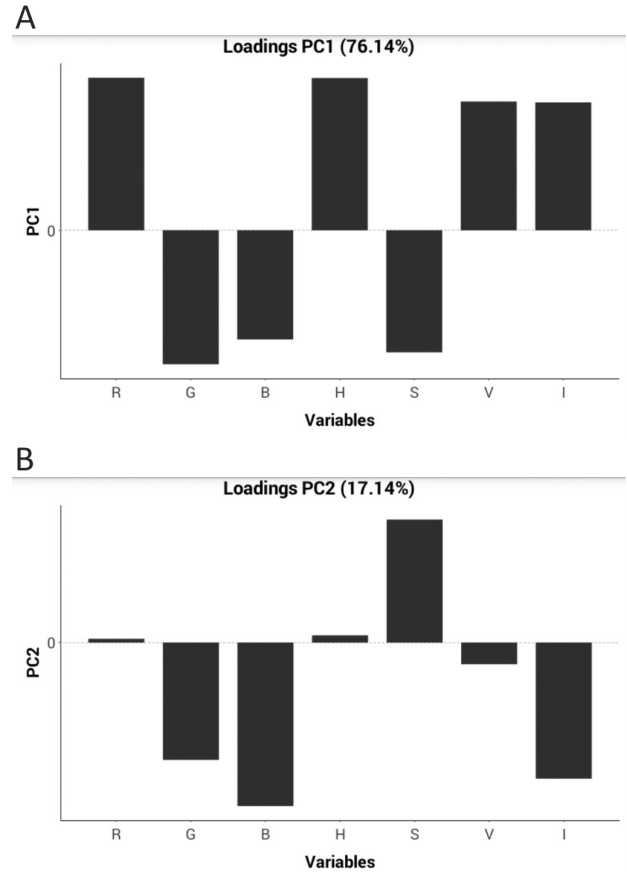
**Figure 6.** PC1 and PC2 loading graphs for the PCA of the image data of Brazilian banknotes at relative scale.

others because their only significant contributions were from channels H and S.

These results do not characterize a forensic application, but clearly demonstrate the potential of PhotoMetrix to differentiate the four sets of banknotes of different colors, tones, and textures through the PCA algorithm.

Conclusions

The development of this study allowed the creation of an application for quantitative univariate and qualitative multivariate analyses, both from the decomposition of digital images acquired by using the main camera of a mobile device. Simple linear regression techniques were used for univariate analysis and PCA was used for multivariate analysis. Highlighting of the main features: (*i*) marking the region of interest in real time on the camera image; (*ii*) adjusting the focus by touching the screen, when available; (*iii*) presenting the results clearly and objectively; (*iv*) the possibility of sending the image data (histograms), analysis, and graphs by e-mail.
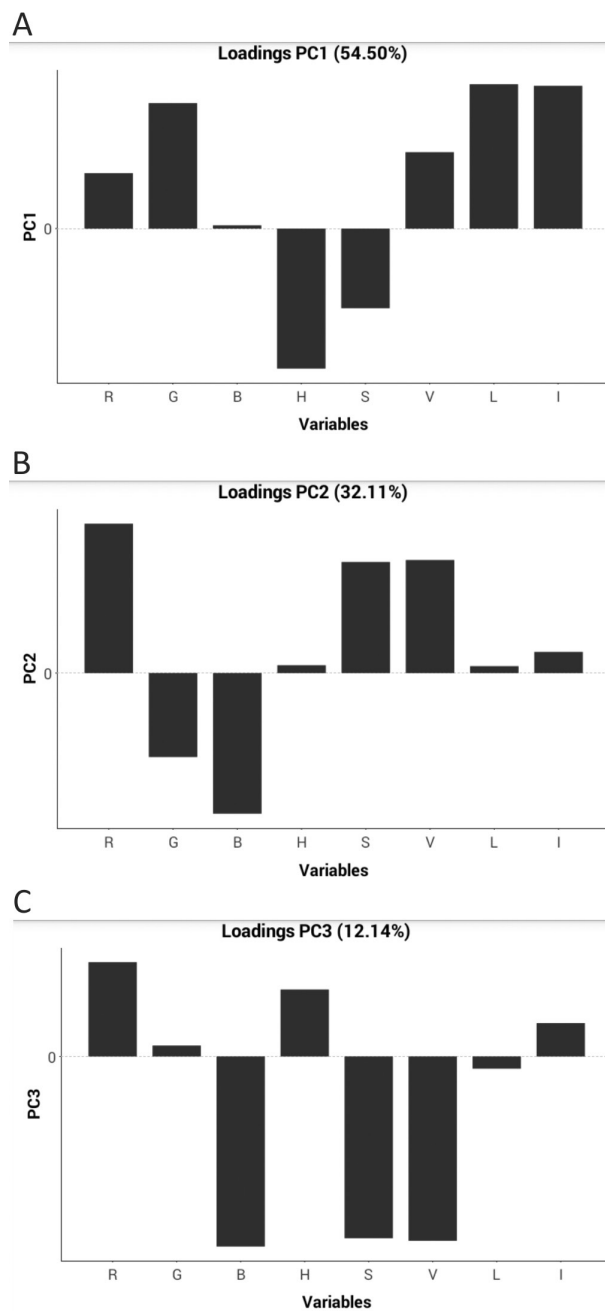
In order to continue the work and increase the functionalities of the developed application, the

A

B

C
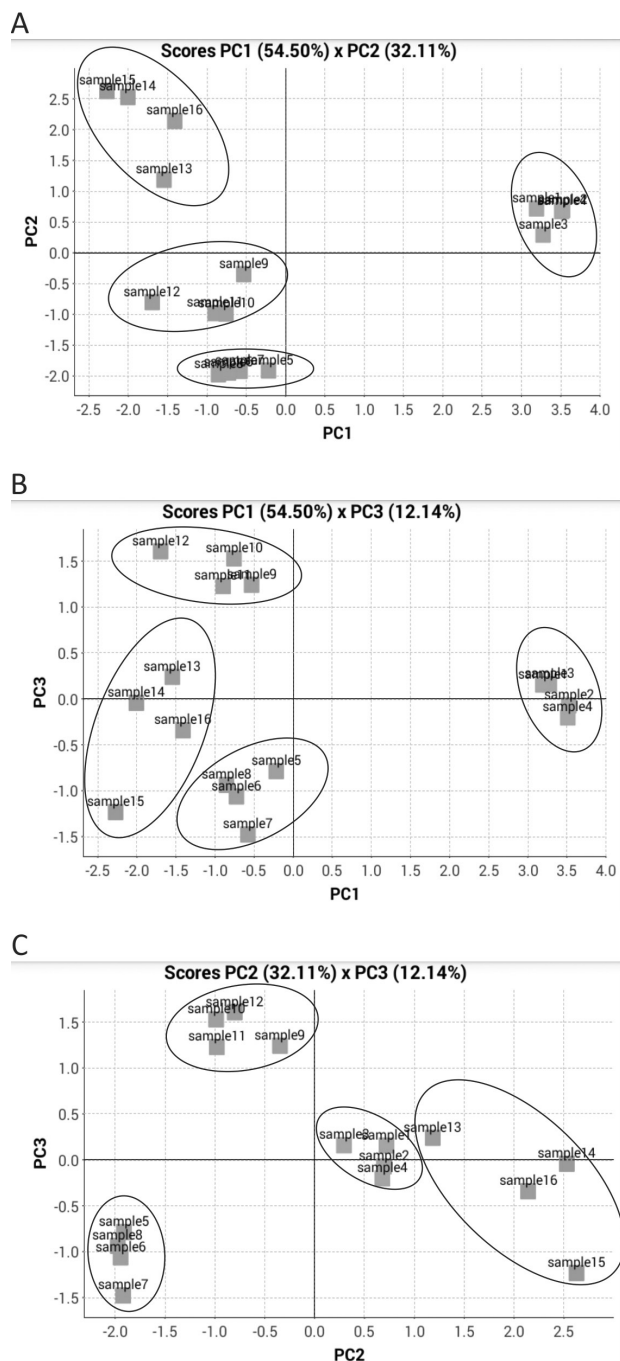
Figure 7. Score graph for the image data PCA of Brazilian and Argentinean banknotes. (A) PC1 × PC2; (B) PC1 × PC3; (C) PC2 × PC3.

A

B

C

Figure 8. PC1, PC2, and PC3 loading graphs for the image data PCA of Brazilian and Argentinean banknotes at relative scale.

implementation of classification techniques, such as soft independent modeling of class analogy (SIMCA) and hierarchical cluster analysis (HCA), and multivariate quantitative analysis, such as classical least squares (CLS) and PLS, is suggested.

Moreover, another suggestion is to promote other functions in the univariate analysis such as editing the calibration curve and lot sampling. For the multivariate analysis it is also proposed the biplot graph as an option, in which the scores and loadings are shown in the same Cartesian plane and detection of anomalous samples through Hotelling's T^2.

Acknowledgments

References

1. Antonellia, A.; Cocchib, M.; Favaa, P.; Focab, G.; Franchinia, G. C.; Manzinib, D.; Ulricia, A.; *Anal. Chim. Acta* **2004**, *515*, 3.

2. Gonzalez, R. C.; Woods, R. E.; *Digital Image Processing*; Prentice Hall: New Jersey, 2008, p. 954.

3. Russ, J. C.; *The Image Processing Handbook*; CRC Press: Boca Raton, 2011.

4. Butz, P.; Hofmann, C.; Tauscher, B.; *J. Food Sci.* **2005**, *70*, 131.

5. Masawat, P.; Harfield, A.; Namwong, A.; *Food Chem.* **2015**, *184*, 23.

6. Choodum, A.; Parabun, K.; Klawach, N.; Daeid, N. N.; Kanatharana, P.; Wongniramaikul, W.; *Forensic Sci. Int.* **2014**, *235*, 8.

7. Debus, B.; Kirsanov, D.; Yaroshenko, I.; Sidorova, A.; Piven, A.; Legin, A.; *Anal. Chim. Acta* **2015**, *895*, 71.

8. Santos, P. M.; Pereira-Filho, E. R.; *Anal. Methods* **2013**, *5*, 3669.

9. Yang, C.; Shi, K.; Edwards, P.; Liu, Z.; *Opt. Express* **2010**, *18*, 23529.

10. Zhang, C.; Cheng, G.; Edwards, P.; Zhou, M. D.; Zheng, S.; Liu, Z.; *Lab Chip* **2016**, *16*, 246.

11. Schwaber, K.; Beedle, M.; *Agile Software Development with Scrum*; Prentice Hall: New Jersey, 2001, p. 158.

12. Helfer, G. A.; Bock, F.; Marder, L.; Furtado, J. C.; Costa, A. B.; Ferrão, M. F.; *Quim. Nova* **2015**, *38*, 575.

13. www.photometrix.com.br, accessed in June 2016.

14. ASTM E394: *Standard Test Method for Iron in Trace Quantities Using the 1,10-Phenanthroline Method*, Philapelphia, 2004.

15. Zarei, K.; Atabati, M.; Kazemi, L.; *Farmaco* **2005**, *60*, 37.