How does Computer vision compare to standard colorimeter in assessing the seed coat color of common bean (*Phaseolus vulgaris* L.)?

Common bean (*Phaseolus vulgaris* L.) exhibits a wide range of seed coat colors and this morphological trait is widely used in cultivar identification and assessment of diversity within this species. With an advancement in technology and informatics, new methods of assessing seed color are emerging in addition to traditionally used visual observation. Due to a great variety of color measuring techniques, the evaluation of the agreement between methods is needed prior to using the methods interchangeably. Seed coat color in terms of CIE L*a*b* color coordinates of 100 common bean accessions belonging to five mono-colored landraces was assessed using two methods, colorimeter and Computer vision. The percentage difference between the two methods across all samples for L* color coordinate was 5.81%, for a* color coordinate 23.32% and for b* color coordinate 44.44%. According to Bland-Altman difference plot there is a considerable lack of agreement between the two methods. However, using stepwise discriminant analysis revealed that colorimeter method correctly classified 97% of accessions into their respective landrace, while the classification success of the Computer vision was 99%.

**Keywords:** colorimetry, common bean, Computer vision, landrace
Grah (Phaseolus vulgaris L.) pokazuje veliki raspon boja sjemene ljuske te je korištenje ove morfološke karakteristike uobičajeno u identifikaciji kultivara i procjeni raznolikosti unutar vrste. Napretkom tehnologije i informatike pojavljuju se nove metode određivanja boje sjemena uz klasičnu metodu vizualnog određivanja. Zbog velikog broja tehnika određivanja boja, potrebna je ocjena usuglašenosti između metoda prije njihovog kombiniranja ili naizmjeničnog korištenja. Boja sjemena 100 primki pet jednobojnih tradicijskih kultivara graha opisana CIE L*a*b* koordinatama procijenjena je koristeći dvije metode, kolorimetar i računalni vid. Postotna razlika između dvije metode kroz sve uzorke za koordinatu L* iznosila je 5.81%, za koordinatu a* 23.32%, a za koordinatu b* 44.44%. Bland-Altmanov graf razlike pokazuje da postoji značajan nedostatak usuglašenosti između dvije metode. Koristeći postupnu diskriminantnu analizu uspješno je klasificirano 97% primki na temelju podataka dobivenih kolorimetrom dok je klasifikacijski uspjeh računalnog vida bio 99%.

**Ključne riječi:** grah, kolorimetrija, računalni vid, tradicijski kultivari

**INTRODUCTION**

Assessment of morphological traits is a traditionally used method for description and classification of plant taxa. In agriculture, cultivars are often distinguished based on shape, size and color of various plant parts. Common bean (Phaseolus vulgaris L.) is a cultivated legume of agricultural interest in many countries and the most important grain legume for direct human consumption worldwide (Gepts, 2001; Piergiovanni and Lioi, 2010). In Croatia, common bean represents the main grain legume for direct human consumption (Čupić et al., 2012). The long tradition of common bean cultivation in Croatia has led to further genetic diversification of numerous landraces differing in seed size and shape but most notably in seed coat color (Vidak et al., 2015; Carović-Stanko et al., 2017). Common bean exhibits a wide variety of seed coat colors (white, pink, red, yellow, brown, black) and intermediate hues as well as spotted and striped seed coat patterns (Lešić et al., 2004).

Visual observation to this date remains the main practice in assessing seed coat color. Descriptor lists describing size, shape, seed coat color and other relevant morphological traits have been developed for measuring and documenting diversity of a range of economically valuable plant species (Bioversity International, 2007). For common bean IPGRI descriptors (International Board for Plant Genetic Resources, 1982) and Phaseolus Database, a subset of descriptors proposed by Grain Legumes Working Group of the ECPRG are used (European Cooperative Programme for Plant Genetic Resources, 1999). Over the years many research papers have been published based on data obtained using these lists (Rodino et al., 2003; Piergiovanni et al., 2006; Stoilova et al., 2013; Scarano et al., 2014; Vidak et al., 2015). Moreover, the International Union for the Protection of New Varieties of Plants (UPOV) provides a list of morpho-agronomic traits that are used to describe newly developed cultivars. Among other traits, number of colors, main color (largest area), secondary color and its distribution on seed coat are used for description of commercial cultivars of P. vulgaris (International Union for the Protection of New varieties of Plants, 2005).

Another approach to determine seed coat color is the use of colorimeters commonly applied in various food and processing industries (Shanin and Symons, 2003; Whan et al., 2014). Colorimeters measure color in CIELAB color space which includes all perceivable colors. CIELAB consists of three channels: lightness (L*) ranging from darkest black at L* = 0 to brightest white at L* = 100; color channel a* representing green/red opponent colors with green at negative and red at positive a* values; and color channel b* representing blue/yellow opponent colors with blue at negative and yellow at positive b* values.
Measuring morphological traits using Computer vision has a high potential of becoming the technique of choice for characterization, taxonomic identification and quality control in the near future. The basic system is comprised of four elements: light source (different types of lamps), an image capturing device (high quality digital photographic camera or flatbed scanner), personal computer and software for photo manipulation (Myers, 1992). The measurement is usually conducted in controlled conditions requiring adequate space and some initial testing at the start. This method is increasingly used in science, as well as in agriculture and food production (Kılıç et al., 2007; Venora et al., 2007; Venora et al., 2009; Bianco et al., 2015) and its practicality is recognized in quality control applications. The procedure itself is oriented at analyzing a single, or a small number of samples at a time but at high speed. As new measurement techniques are introduced, their comparison with an established one is needed to assess the degree of agreement between the two (Bland and Altman, 1986).

The aims of this study were to (1) evaluate the agreement of measurement of colorimeter and Computer vision as well as to (2) assess the classification success of the color coordinates obtained by both methods in classification of Croatian traditional landraces.

MATERIALS AND METHODS

Data

The study was conducted on a sample of 100 common bean accessions representing five mono-colored landraces ('Tetovac', 'Biser', 'Zelenčec', 'Kukuruzar' and 'Puter') from all production areas throughout Croatia (Figure 1). The accessions are part of the Grain legume collection held at University of Zagreb, Faculty of Agriculture, Department of Seed Science and Technology. Seed coat color was determined using two different methods.

First, seed coat color was assessed by device Minolta Chroma Meter CR-410, with a color measuring area of 50 mm and standard illumination D 65, arranged for color spectrum L*, a*, b* (International Commission on Illumination, 1986). The letter L* represents lightness (pallor) of beans, with range of measurement from 0 to 100. The letter a* represents degree of redness, or spectrum of colors from green (-50) to red (50). The letter b* represents degree of yellowness, or spectrum of colors from blue (-50) to yellow (50). Sample of 30 to 40 seeds per accession was placed in a Petri dish, covering the entire bottom and three measurements were taken per sample.

Then, digital images of five seeds per accession were acquired using digital camera (Canon EOS 70D). Seed samples were placed on a white, non-reflective background within a 10 by 10 cm square frame. Photographs were taken in a dark room with no windows to eliminate external light source that could reduce the quality of the output photos. Two lamps were placed on each side of the camera to reduce the shadows of the samples. Digitized images were transferred to PC and before image analysis, all images were calibrated using ColorChecker Passport Photo. This process consisted of photographing the industry standard color reference target in same lightning conditions as the samples were captured and creating a DNG profile using ColorChecker Passport v1.0.2 that was used to match the color profiles of seed images to a standardized reference. Seed coat color of samples was measured using ImageJ 1.49v (Schneider et al., 2012). Prior to color measurement, hilum and micropyle was cut out of the images. A total of 500 bean seeds were analyzed.

Statistical analysis

Average values of CIELAB color coordinates (L*, a*, b*) of seed coat color as obtained by the two methods (Minolta and Computer vision) were calculated to assess the mean difference (bias) between the two methods. Pearson's correlation coefficients between values obtained by the two methods were computed and tested.

Bland-Altman analysis (Bland and Altman, 1986) was used to analyze the agreement between the three CIELAB color coordinates (L*, a*, b*) obtained by the two methods (Minolta and Computer vision). The mean difference (bias) and the limits of agreement (LoA) of the differences were...
The 95% confidence intervals for each limit of agreement were obtained using the method described by Carkeet (2015). The distribution of the differences obtained by the two methods for each color coordinate was tested with the Kolmogorov-Smirnov test for normality. Pearson's correlation coefficients between means and differences were computed and tested for each color coordinate.

Canonical discriminant analysis was used to test the utility of the three CIELAB color coordinates to discriminate among five common bean landraces based on seed coat color. The color coordinates were evaluated for its performance as the discriminant criterion to classify common bean accessions correctly into their respective landraces by estimating error rates (probabilities of misclassification) using cross-validation. All the statistical analyses were performed using SAS software version 9.3 (SAS Institute, 2011).
RESULTS AND DISCUSSION

The percentage difference between the two methods across all samples for $L^*$ color coordinate was 5.81%, for $a^*$ color coordinate 23.32% and for $b^*$ color coordinate 44.44%. When observing individual landraces, the percentage differences ranged from 3.45% for ‘Kukuruzar’ to 14.09% for ‘Puter’ landrace ($L^*$ color coordinate), from 17.65% for ‘Zelenčec’ to 81.32% for ‘Tetovac’ landrace ($a^*$ color coordinate), and from 41.52% for ‘Tetovac’ to 51.4% for ‘Biser’ landrace (Table 1). The measurements for $L^*$ and $a^*$ color coordinates obtained with Minolta were consistently higher than the ones obtained by Computer vision, while measurements for $b^*$ color coordinate were consistently lower than the ones obtained by Computer vision. The two methods of measurement appeared to be highly correlated in $a^*$ and $b^*$ color coordinates ($r(a^*)=0.95$, $r(b^*)=0.97$) and moderately correlated in $L^*$ color coordinate ($r(L^*)=0.64$) with all correlations highly significant (Figure 2).

However, high correlation between the two measurements does not imply that there is a good agreement between the two measuring methods (Giavarina, 2015).

Bland-Altman difference plot revealed considerable lack of agreement between the two methods. The $L^*$ color coordinate measurement obtained using Minolta would be between 8.26 less and 17.12 more than the measurement made with Computer vision. The $a^*$ color coordinate measurement made with Minolta would be between 0.61 less and 3.1 more than the measurement made with Computer vision. The $b^*$ color coordinate measurement made with Minolta would be between 19.84 and 4.25 less than the measurement made with Computer vision (Figure 3). Comparing different methods of measurement is a standard procedure when trying to replace an old or more expensive method with newer or cheaper method in different fields, most notably in medicine (Downie, 2015). It also raises a possibility of combining data from different sources (institutions using different measuring methods or different measuring methods applied in laboratory and in field research) if the methods are in an agreement.

The stepwise discriminant analysis revealed that all three color coordinates, $b^*$, $a^*$ and $L^*$ (in this order in accordance to partial $R^2$ statistic), were useful in discriminating common bean landraces using both methods (Minolta and Computer vision) (Table 2).

Table 1. Average values of CIE $L^*$/$a^*$/$b^*$ color coordinates of seed coat color in 100 common bean accessions belonging to five landraces as obtained by the two methods (Minolta and Computer vision). The mean ($\Delta$) and percentage (%Δ) difference between the two methods are given.

<table>
<thead>
<tr>
<th>No.</th>
<th>Landrace</th>
<th>n</th>
<th>$L^*$ Minolta</th>
<th>$L^*$ Computer vision</th>
<th>Δ (%Δ)</th>
<th>$a^*$ Minolta</th>
<th>$a^*$ Computer vision</th>
<th>Δ (%Δ)</th>
<th>$b^*$ Minolta</th>
<th>$b^*$ Computer vision</th>
<th>Δ (%Δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>‘Tetovac’</td>
<td>22</td>
<td>88.75</td>
<td>83.5</td>
<td>5.25</td>
<td>2.26</td>
<td>0.95</td>
<td>1.31</td>
<td>11.49</td>
<td>17.51</td>
<td>-6.02</td>
</tr>
<tr>
<td>L2</td>
<td>‘Biser’</td>
<td>8</td>
<td>83.42</td>
<td>79.17</td>
<td>4.25</td>
<td>3.43</td>
<td>2.35</td>
<td>1.08</td>
<td>13.31</td>
<td>22.52</td>
<td>-9.21</td>
</tr>
<tr>
<td>L3</td>
<td>‘Zelenčec’</td>
<td>26</td>
<td>75.12</td>
<td>70.24</td>
<td>4.87</td>
<td>4.68</td>
<td>3.92</td>
<td>0.76</td>
<td>21.73</td>
<td>35.91</td>
<td>-14.19</td>
</tr>
<tr>
<td>L4</td>
<td>‘Kukuruzar’</td>
<td>37</td>
<td>72.96</td>
<td>70.49</td>
<td>2.48</td>
<td>8.58</td>
<td>7.18</td>
<td>1.41</td>
<td>28.02</td>
<td>42.76</td>
<td>-14.73</td>
</tr>
<tr>
<td>L5</td>
<td>‘Puter’</td>
<td>7</td>
<td>81.43</td>
<td>70.71</td>
<td>10.72</td>
<td>11.58</td>
<td>9.36</td>
<td>2.22</td>
<td>21.06</td>
<td>33.13</td>
<td>-12.08</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>78.43</td>
<td>74</td>
<td>4.43</td>
<td>5.97</td>
<td>4.73</td>
<td>1.25</td>
<td>21.08</td>
<td>33.13</td>
<td>-12.05</td>
</tr>
</tbody>
</table>

n - number of accessions
Table 2. Stepwise discriminant analysis summary for three color coordinates allowing discrimination among common bean landraces based on seed coat color

<table>
<thead>
<tr>
<th>Color Coord</th>
<th>Minolta</th>
<th>Computer vision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partial R²</td>
<td>F-value</td>
</tr>
<tr>
<td>b*</td>
<td>0.926</td>
<td>296.85</td>
</tr>
<tr>
<td>a*</td>
<td>0.762</td>
<td>75.04</td>
</tr>
<tr>
<td>L*</td>
<td>0.497</td>
<td>22.92</td>
</tr>
</tbody>
</table>

P-value significant level: *** P<0.001

Figure 2. Correlation between values of CIE L*a*b* color coordinates of seed coat color in 100 common bean accessions as obtained by the two methods (Minolta and Computer vision)
Figure 3. The Bland-Altman plots showing the paired differences (Δ) against the mean values of CIE L*a*b* color coordinates obtained by the two methods (Minolta and Computer vision). Mean differences (bias) and limits of agreement are shown by the solid lines, while the 95% confidence intervals are shown by the dashed lines. The samples indicated by arrows lay outside the 95% confidence interval.
The discriminant function, based on three color coordinates as obtained by Minolta, correctly classified 97 out of 100 accessions (97% classification success) into their respective landrace using cross-validation. The classification success of the Computer vision method was 99%, correctly classifying all 100 accessions except for one. The classification success would be further increased by including other morphological descriptors for the species since some landraces have similar seed coat color but differ in other morphological attributes.

For example, landraces ‘Biser’ and ‘Tetovac’ (two out of three misclassified accessions using Minolta and the only misclassified accession using Computer vision) have similar seed coat color but the former is characterized by small, round seeds while the later has seeds that are large, and kidney shaped. Similarly, ‘Kukuruzar’ and ‘Zelenčec’ landraces (1 misclassified accession using Minolta) are similar in seed coat color but differ in seed shape and color of micropile (Vidak et. al., 2015). In addition to measuring color coordinates of mono-colored seed samples, Computer vision can also be used to measure size and shape of the seeds as well as to measure color coordinates of multicolored seed samples thus eliminating the need for additional equipment (e.g. calipers) and reducing the time needed to take additional measurements (Venora et al., 2009). The canonical discriminant analysis based on three color coordinates showed that the first two canonical variables (CV) were significant jointly explained 99.8% (Minolta) and 97.6% (Computer vision) of the total variation among accessions (Figure 4). Although the methods are not comparable, both managed to successfully classify a large number of accessions into their respective landrace.

Determining seed coat color using descriptor lists is largely qualitative in nature and depends greatly on the skill of the observer and the conditions in which the analysis is being conducted. While human observers are often highly trained and experienced, it is difficult to standardize the results obtained by this method for technological purposes. In addition, this can be a long and slow procedure when the sample size is large, and samples differ in small degree with higher probability of human error. With advancements in technology, colorimeters and Computer vision provide more accurate information on seed coat color, all with minimal costs (Bassett et al., 2002; Ozturk et al., 2009; Cobb et al., 2013; Konzen and Tsai, 2014).

**CONCLUSIONS**

Having the same discriminatory power, Computer vision can be considered as a superior method of obtaining not only data on seed coat color, but also other important morphological traits important for classification of bean landraces such as height and width of the seed sample. The whole process can also be automated, enabling the analysis of a great number of samples in short time.
Another advantage of this method is the possibility using it in analysis of multicolored accessions (accessions that have a base color, and a pattern of another color). Using Computer vision enables us to obtain color coordinates for both base color and pattern as well as their share on the seed. This is especially important in Croatia where the majority of bean production is based on ‘Trešnjevac’, a multicolored landrace with several subgroups that could be defined better using Computer vision rather than human observer.

ACKNOWLEDGEMENTS

This work has been supported by Croatian Science Foundation under the project UIP-11-2013-3290 Genetic basis of bioactive nutrient content in Croatian common bean landraces (BeanQual), and under the project KK.01.1.1.01.005 CoE CroP-BioDiv Centre of Excellence for Biodiversity and Molecular Plant Breeding.

REFERENCES


