*Skin Research and Technology 2005; **11:** 242–247*
Printed in Denmark · All rights reserved

Skin Research and Technology

Practical color calibration for dermoscopy, applied to a digital epiluminescence microscope

C. Grana[1], G. Pellacani[2] and S. Seidenari[2]

[1]*Department of Computer Engineering and* [2]*Department of Dermatology, University of Modena and Reggio Emilia, Modena, Italy*

Background/purpose: The assessment of colors is essential for melanoma (MM) diagnosis, both for pattern analysis on dermoscopic images, and when using semiquantitative methods. Our aim was to provide a simple, precise characterization and reproducible calibration of the color response for dermoscopic instruments.

Methods: Three processes were used to correct the non-uniform illumination pattern of the instrument, to easily estimate the camera gamma settings and to describe the color space conversion matrices required to produce standard images, in any color space. A specific color space was also developed to optimize the representation of dermatoscopic colors. The calibration technique was tested both on synthetic reference surfaces and on real images by comparing the difference between the images colors obtained with two different equipments.

Results: The differences between the images acquired by means of the two instruments, calculated on the reference patterns after calibration, were up to 10 times lower then before, while comparison of histograms referring to real images provided an improvement of about seven times on average.

Conclusions: A complete workflow for dermatologic image calibration, which allows the user to continue using his own software and algorithms, but with a much higher informative content, is presented. The technique is simple and may improve cooperation between different research centers, in teleconsulting contexts or for result comparisons.

Key words: dermoscopy – epiluminescence microscopy – color calibration – image comparison – polarized light – melanocytic lesions

Accepted for publication 13 November 2004

As COLOR analysis has proved to be an important factor in the diagnostic process of dermatoscopic images, color degradation could negatively influence the diagnostic ability of the clinician (1). Not much investigation has been conducted on the interchange of dermatoscopic images and on the differences in color reproduction between different instruments or units of the same instruments. Even if this evaluation may seem insignificant, since images appear more or less similar in successive comparisons, the problem becomes apparent in automatic analysis, that is, computer-based color measure and characterization (2, 3). Thus, the algorithms used by different image analysis programs are strictly applicable only to pigmented skin lesions acquired with the same instrument and technique, and are not adaptable to images generated by different tools, sometimes also using different acquisition methods (4). For example if an instrument allows manual light intensity tuning, so that images can be adjusted by sight to appear *bright enough*, any comparison on the dark to light variation (for instance in the search for dark areas) is influenced by this setting (5). If the light intensity tuning is continuously modified for every image, some difficulties could arise, leading to the question of whether it is possible to use those images in an automatic framework or not.

Most of the work in computerized dermatoscopy deals with the use of colors, but few studies even attempt to acquire knowledge on the color space under examination, usually only referring to an unspecified RGB color space (6, 7).

Although color calibration methods for dermatology have been previously used in cosmetology, only Haeghen and coworkers developed an approach for the acquisition of pigmented skin lesion images with reference to a well-defined color space (namely sRGB).

The authors have no conflict of interest to disclose.

It is time for the literature dealing with color images to cope with the problem of calibration, in order to generate reproducible data. In this paper we provide a detailed and practical description of a calibration framework: firstly, illumination correction is explored; secondly, an easy camera gamma estimation technique is described; and thirdly, the use of a specific color space construction to avoid low contrast effects because of the 8-bit quantization of color channels is described.

Materials and Methods

The calibration techniques described here are applicable to every video camera system, and take into account problems that are very likely found in every dermatoscopic setup. For the sake of precision we will refer in detail to the specific case of a digital epiluminescence microscope for dermatology (FotoFinder, TeachScreen software GmbH, Bad Birnbach, Germany), which consists of a probe, comprising a CCD-chip color video camera with an integrated handle and optics for epiluminescence microscopy, a processing unit and a color monitor. Optics are set in a removable conic structure with a cylindrical transparent spacer and contact plate at the end, and with six bright white LEDs, positioned at the bottom of the structure, for constant illumination of the viewing area. The digitized images offer a spatial resolution of 768×576 pixels and 16 million colors. For the epiluminescence observation, a drop of contact medium, such as alcohol in water solution, is applied between the contact plane and the skin, enabling the recognition of subsurface structures.

Illumination and border defects correction

The color calibration process begins with a first step to correct the irregular illumination of the instrument. The basic assumption here is that if we are imaging a uniform reference surface, we should obtain an almost constant reading for the whole image. This is not always so because of the instrument's characteristics, but we can measure the deviation from uniformity to invert it, thus obtaining a correction map for all the acquired images. A masking process must also occur to get rid of pixels that do not convey any data, such as top or bottom lines that are always black because of the frame grabber settings, or the black ring that some instruments present at lower magnification levels (20-fold for FotoFinder).

The filter values are computed separately for each color channel, so if we call $R(x, y)$ the value of the red channel, the value of the filter in that point is:

$$F_{\mathrm{R}}(x,y) = \frac{R(x,y) - M_{\mathrm{R}}}{M_{\mathrm{R}}}$$

where M_{R} is the most represented value in a selected window of the image. In our application, we chose the rectangular window from (200,350) to (567,550).

This filter assessment should be carried out so that it is independent of the specific surface selected and the resulting acquisition. Thus, after computing the $M_{\mathrm{R}}^{i}$for each step of the Kodak Gray Scale, and discarding those squares whose values clip at 255, we evaluate F_{R} as the median of the measured values F_{R}^{i}. The same process is applied to the green and blue channels, obtaining F_{G} and F_{B}. Each image I is then filtered for each channel using the equation:

$$\tilde{I}_{\mathrm{R}}(x,y) = \frac{I_{\mathrm{R}}(x,y)}{1 + F_{\mathrm{R}}(x,y)}.$$

Assessment of gamma values

After obtaining light-compensated images, the next step is to estimate the non-linear relation between the luminance factor, also known as CIE tristimulus value Y, and the digital values provided by the camera.

Measuring the Kodak Gray Scale with the FotoFinder, we verified the presence of gamma correction of RGB values, which is a power relation between the digital measures and the known Y values. To give an estimate of this relation we compared the normalized values of $\hat{Y} = Y/Y_n$ and $d_{\mathrm{r}} = R/255$, $d_{\mathrm{g}} = R/255$, $d_{\mathrm{b}} = R/255$, with the following equation:

$$y = ax^{\gamma} + b$$

and estimated the three parameters (a, b, γ) separately for the three channels. The aim was to use the N measured values d_i (with reference to the currently considered color channel) and the corresponding $\hat{Y}_i$ to obtain the best triplet:

$$(a,b,\gamma) = \arg\min_{a,b,\gamma} \sum_{i=1}^{N} \hat{Y}_i - \left(ad_i^{\gamma} + b\right).$$

Following this estimation, simple inversion of the power relation allows to obtain triplets of RGB

values that can be linearly transformed by matrix multiplication into *XYZ* triplets.

Conversion from the instrument's RGB to XYZ

Many matrices can be found in text books or on web pages; however, none of these should be used. The conversion we are looking for should be assessed from the behavior of our specific instrument, which can only be observed by means of a reference object, such as the Gretag-Macbeth ColorChecker (GMCC), a target that is often used in television broadcasting in order to evaluate the color accuracy of TV cameras. Every ColorChecker square was acquired and filtered with the previously estimated filter, and then the average value of pixels from a central rectangle, whose sides were half the width and half the height of each image, was computed as the measured value for the patch. M_X, the maximum histogram value (where X is the channel), was evaluated to reveal patches that could not be correctly imaged by the instrument. We discarded all patches resulting in $M_X = 0$ or $M_X = 255$. In our experiments one patch alone (the white one) could not be used because of the extremely narrow range of the blue channel that saturated to 255 (Fig. 1).

Following a common procedure for optimization, we took the M valid triplets of RGB values, their corresponding *XYZ* ones and searched for the matrix that gave the best linear transform for all the patches. Because the relation between RGB and *XYZ* values is not always well described by a linear transform, we used a non-linear operator described by Haeghen et al. (8) that included all the covariance terms:

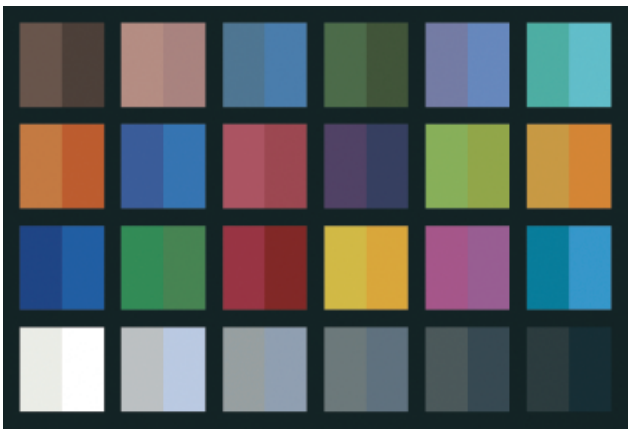

Fig. 1. Declared sRGB values of the ColorChecker paired with corresponding values as measured by the instrument.

$$\Theta_9 \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} R & G & B & RG & GB & BR & R^2 & G^2 & B^2 \end{pmatrix}^T$$

Thus, after applying this operator to the data, the transformation matrix becomes 3 rows by 9 columns, but the optimization technique is the same. For the resolution of this linear system with 27 unknowns we used the singular value decomposition, which presents the solution that gives the minimum Euclidean difference between the known *XYZ* values and those estimated from RGB ones.

Conversion from XYZ to a known and standard color space

At this point we have obtained an assessment of *XYZ* color coordinates for every measured pixel. The aim is to transform these values into another known color space enabling visualization and a simpler storage. Haeghen et al. chose the sRGB color space, but a problem arises in color space conversions, because of quantization to be applied to the values, because this results in the subdivision of the 'light' range into a fixed number of steps. Indeed, if we are interested in a limited area of the sRGB color space, we lose much color detail that is set aside to describe color space regions that our images will never use. This color loss is clearly observed in dermatoscopic images, which become less clear (lower contrast and dynamic range) when converted into an sRGB color space.

For the above reasons, we decided to describe our images by a new color space, especially conceived for our instrument for daily use. The color settings usually provided by the instrument are satisfactory, so we just searched for a formal description of the color transformation from *XYZ* back to our instrument's RGB. This was performed by converting a large number of colors randomly taken from real lesions to *XYZ* using the calibration procedure previously described and then finding the best transformation enabling us to go back to the original colors. This is not the same as simply inverting the first transformation, because real images are used instead of the reference target. The color space obtained is extracted from an average characterization produced by many images; thus, it is safely applicable to dermatoscopic images, because it is designed to achieve better use of the color repre-

sentation in the spectrum area occupied by this kind of images. Its known relation with *XYZ* enables a simple conversion into any other chosen color space, and the practice of viewing the images on an sRGB-calibrated computer monitor (a common setting available in most modern monitors) allows a common evaluation of images, even if obtained from different sources. Our proposed conversion uses the following linear conversion:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.352 & -0.802 & -0.329 \\ -0.731 & 1.358 & 0.392 \\ -0.098 & -0.187 & 1.273 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{D65}$$

and the gamma conversion is given by:

$$X' = 1.46 \cdot X^{(1.0/1.85)} - 0.46$$

The choice for the gamma values was made by averaging those of the three channels. This conversion produces not too saturated images and allows simple comparisons between different units.

Results

For testing purposes, we decided to verify the effectiveness of the calibration procedure on two different FotoFinder units without carrying out a preliminary hardware calibration, i.e. attempting to subjectively adjust images. We thus decided to set the instruments at their default configuration (factory settings), choosing an Iris setting of -25 in order to avoid skin color saturation in light-skinned patients. Moreover, with this setting, all colors from the GMCC can be included, ensuring a good fitting of the data with the least squares method. To verify the effects of the resulting calibration, we used a Home Made ColorChecker (HMCC), consisting of different squares of colored paper.

The calibration procedure was performed starting from the light compensation filter, going on to the gamma functions assessment and finally to the RGB to *XYZ* matrix computation.

Looking directly at the RGB color space, we can compare the Euclidean distances before and after correction (Table 1), and this enables us to see how different values provided by different units of the same instrument can be. The comparison shows that even on a surface that is not ideal, differences obtained with the HMCC were six times lower than before calibration. It is interesting to note that we could obtain better results using the surface used for calibration; thus, for a real comparison and a serious validation of results another target should be used. In fact using the GMCC, differences obtained after calibration were 10 times lower then before.

TABLE 1. Euclidean distances in the instruments RGB color space between corresponding patches of the Home Made ColorChecker measured with the two different instruments, before and after calibration

Square number	Distances before calibration	Distances after calibration
1	76,64	29,17
2	57,17	9,68
3	56,60	3,43
4	71,00	15,70
5	62,58	10,24
6	56,10	2,24
7	76,23	12,89
8	70,71	13,08
9	82,38	14,21
10	78,25	6,40
11	61,27	6,78
12	51,08	5,68
13	65,94	0,25
14	83,82	13,01
15	72,00	10,14
16	40,47	3,33
17	73,52	5,94
18	68,11	6,23
19	74,57	19,25
20	60,20	12,51
21	70,12	15,50
22	63,87	6,37
23	47,17	11,06
24	71,75	16,61
Mean	66,31	10,40
Maximum	83,82	29,17

Finally, we had to test the results on real-world images, i.e. skin lesions. The problem here was that we could not just compare the average values of the images, because this would not be sufficiently accurate (images are not uniform surfaces). So we decided to provide an initial measure distinguishing the skin from the lesion, and then we tried to compare the distribution of colors in the RGB color space. To this aim, the 16 million colors RGB cube was reduced, allowing each channel to have values from 0 to 7, i.e. a 3 bit per channel representation, that led to 512 possible colors. A three-dimensional color histogram was produced, and then a common histogram comparison technique, called *histogram intersection* (9), was used to provide a metric to assess color similarity. The similarity measure is given by:

$$HI(h_1, h_2) = \sum_{r,g,b=0}^{7} \min\{h_1(r,g,b), h_2(r,g,b)\}$$

TABLE 2. Euclidean distances between average lesion and skin values and histogram intersection percentages (referring to the whole image), in the instruments RGB color space, between corresponding images assessed by the two different instruments, before and after calibration: mean and maximum values

	Lesion		Skin		Histogram	
	Uncalibrated	Calibrated	Uncalibrated	Calibrated	Uncalibrated	Calibrated
Mean	79,70	14,91	116,41	11,13	11,20%	76,18%
Maximum	100,86	37,61	133,30	25,50	23,14%	90,56%

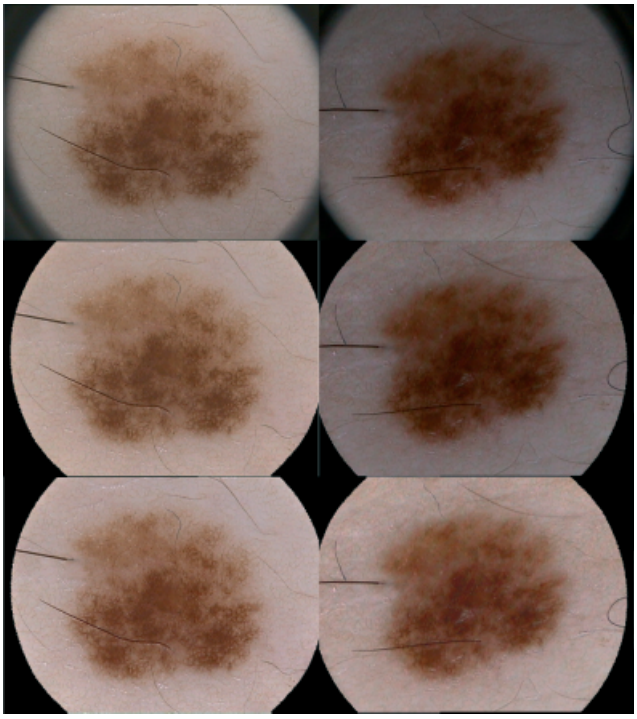

Fig. 2. Example calibration result: the first line shows the original images, the second line shows the results of lighting correction and the third line shows the calibrated images.

The value is computed over the normalized versions of the two histograms (each value of the histogram is divided by the image size), so that *HI* is limited between 0 (no correspondence) and 1 (perfect histogram match). In Table 2 and Fig. 2, the resulting comparisons are shown.

Discussion

Future developments of dermoscopy will deal with the progress of image analysis systems, tied to automatic diagnosis, and with tele-dermatology, to obtain a remote diagnostic consultation. In both cases, color calibration is of fundamental importance for two reasons: to make the developed algorithms applicable to various instruments, i.e. diagnostic center independent, and on the other hand to allow expert evaluators to have reproducible diagnoses, not biased by color degradation.

Unfortunately, in the digital dermoscopy area, color standardization is very seldom taken into account, even if working in conjunction with image analysis programs. Most references are made to a generic RGB space that is implicitly assumed to be the RGB color space provided by their instrument. For instance, Vannoorenberghe et al. (6) describe a system that relies on learned probabilities to detect lesion contours directly in the RGB color space, Faziloglou et al. (7) use color histograms to detect differences between melanoma and nevus colors: to get rid of color influences, they subtract the average skin color from each lesion pixel, without, however, considering gamma or contrast differences. In Gerger's study (10) a completely different approach was followed, but unspecified RGB statistics were still present, disallowing any possible systematic reproduction of the published results.

It is natural to assume that color calibration technologies and methods are still being studied or are as yet unknown in the dermatologic field, but this is not the case, because for example Korichi et al. (11) illustrated an on-skin lipstick color measurement system with full calibration, and very recently in a work by Miyamoto et al. (12) a quick but effective conversion was estimated from RGB to $L^*a^*b^*$, by means of a third-order multiple regression analysis. The authors were also careful in referring to the obtained values as *quasi-*$L^*a^*b^*$ to stress the fact that this was an approximation of the full calibration approach. Setaro and Sparavigna (13) obtained a simple calibration by using a standard reference based on three colors and then adjusting the images comparing the measured values of the marker to the image colors by simple difference. Even this first-order model is reported to produce good reproducibility. A systematic work for color measurement from video camera in dermatology was presented by Herbin et al. (14), but one of the most detailed and precise methods was provided

by Haeghen et al. (8). In this study, a full calibration approach was followed and the problem of obtaining standardized images was successfully resolved. However, the final choice of sRGB color space for image exchange and handling led to the production of low contrast images and loss of details, owing to the limited area of the abovementioned color space, when applied to pigmented skin lesions acquired by digital dermoscopy.

We presented a complete workflow for dermatologic image calibration, taking into account some practical, but very important issues such as illumination correction, easy camera gamma estimation and a specific color space generation. The system is simple, and after calibration, allows the user to continue using his own software and algorithms, but with a much higher informative content. Conveyance of the final selected color space to whoever receives the corrected images should be ensured, to enable their use in color-calibrated contexts.

Encouraging the widespread use of color calibration in this field will not only improve the quality of dermatoscopic digital libraries but will also open the way to teleconsulting, remote analysis and result comparisons between different computer algorithms, in order to stimulate joint development or knowledge sharing.

References

1. Seidenari S, Pellacani G, Righi E, Di Nardo A. Is JPEG-compression of videomicroscopic images compatible with telediagnosis? Comparison between diagnostic performance and pattern recognition on uncompressed TIFF images and JPEG compressed ones. Telemed J E-Health 2004; 10(3): 294–303.
2. Seidenari S, Pellacani G, Giannetti A. Digital videomicroscopy and image analysis with automatic classification for detection of thin melanoma. Melanoma Res 1999; 9: 163–171.
3. Pellacani G, Grana C, Seidenari S. Automated description of colours in polarized-light surface microscopy images of melanocytic lesions. Melanoma Res 2004; 14: 125–130.
4. Pellacani G, Seidenari S. Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized light videomicroscopy. Clin Dermatol 2002; 20: 222–227.
5. Pellacani G, Grana C, Cucchiara R, Seidenari S. Automated extraction and description of dark areas in surface microscopy melanocytic lesion images. Dermatology 2004; 208: 21–26.
6. Vannoorenberghe P, Colot O, De Brucq D. Dempster-Shafer's theory as an aid to color information processing application to melanoma detection in dermatology. Proc Int Conf Image Anal Process 1999; 774–779.
7. Faziloglou Y, Stanley RJ, Moss RH, Van Stoecker W, McLean RP. Colour histogram analysis for melanoma discrimination in clinical images. Skin Res Technol 2003; 9: 147–155.
8. Haeghen YV, Naeyaert JMAD, Lemahieu I, Philips W. An imaging system with calibrated color image acquisition for use in dermatology. IEEE Trans Med Imaging 2000; 19: 722–730.
9. Swain MJ, Ballard DH. Color indexing. Int J Comput Vision 1991; 7: 11–32.
10. Gerger A, Stolz W, Pompl R, Smolle J. Automated epiluminecence microscopy-tissue counter analysis using CART and 1-NN in the diagnosis of melanoma. Skin Res Technol 2003; 9: 101–110.
11. Korichi R, Provost R, Heusèle C, Schnebert S. Quantitative assessment of properties of make-up products by video imaging: application to lipsticks. Skin Res Technol 2000; 6: 222–229.
12. Miyamoto K, Takiwaki H, Hillebrand GG, Arase S. Development of a digital imaging system for objective measurement of hyperpigmented spots on the face. Skin Res Technol 2002; 8: 227–235.
13. Setaro M, Sparavigna A. Quantification of erythema using digital camera and computer-based colour image analysis: a multicentre study. Skin Res Technol 2002; 8: 84–88.
14. Herbin M, Venot A, Devaux JY, Piette C. Color quantitation through image processing in dermatology. IEEE Trans Med Imaging 1990; 9: 262–269.

Address:
Stefania Seidenari
Department of Dermatology
University of Modena and Reggio Emilia
41100 Modena, Italy
Tel: +39-59 4222464
Fax: +39-59 4224271
e-mail: seidenari.stefania@unimore.it