Automated color calibration method for dermoscopy images

Hitoshi Iyatomi\textsuperscript{a,}\textsuperscript{*}, M. Emre Celebi\textsuperscript{b}, Gerald Schaefer\textsuperscript{c}, Masaru Tanaka\textsuperscript{d}

\textsuperscript{a} Department of Applied Informatics, Hosei University, Tokyo, Japan
\textsuperscript{b} Department of Computer Science, Louisiana State University, Shreveport, USA
\textsuperscript{c} Department of Computer Science, Loughborough University, United Kingdom
\textsuperscript{d} Department of Dermatology, Tokyo Women’s Medical University Medical Center East, Japan

A R T I C L E   I N F O
Article history:
Received 1 October 2009
Accepted 16 August 2010

Keywords:
Dermoscopy
Computer-aided diagnosis (CAD)
melanoma
color calibration

A B S T R A C T
Accurate color information in dermoscopy images is very important for melanoma diagnosis since inappropriate white balance or brightness in the images adversely affects the diagnostic performance. In this paper, we present an automated color calibration method for dermoscopy images of skin lesions. On a set of 319 dermoscopy images, we develop color calibration filters based on the HSV color system. We determined that the color characteristics of the peripheral part of the tumors have significant influence on the color calibration filters and confirmed that the presented filters achieved satisfactory calibration performance as evaluated by cross-validation. We also confirmed that our method successfully modifies the color distribution of a given image to make it closer to the color distribution of the training image set.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction
Advanced malignant melanoma is often incurable, however early-stage melanoma can be cured in many cases, particularly before the metastasis stage. Therefore, early detection is essential for the reduction of melanoma-related deaths [1]. Dermoscopy, a non-invasive skin imaging technique, was introduced to improve accuracy in the diagnosis of melanoma. However, dermoscopic diagnosis is often subjective and is therefore associated with poor reproducibility. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is estimated to be about 75–84\% [2] or 78–88\% [3].

Several groups have developed automated analysis procedures to overcome these problems and reported high levels of diagnostic accuracy [3–9]. However, several problems have persisted with these software-based approaches. For example, results of these studies are not comparable because of the different image sets used in each one. In addition, these studies were designed to develop a screening system for new patients using standalone systems and therefore they have not been opened to the public.

To address these issues, we developed a prototype for a fully automated Internet-based melanoma screening system at our university [7]. The URL of the site has changed and it is now http://dermoscopy.k.hosei.ac.jp.\textsuperscript{1} Using an Internet connection anyone who has a dermoscopy image can use our screening system from anywhere in the world. Our latest system achieved classification performance of 0.928 in area under the ROC (receiver operating characteristics) curve (AUC), 85.9\% in sensitivity (SE) and 86.0\% in specificity (SP) on a set of 1258 non-acral dermoscopy images (1060 melanocytic nevi and 198 melanomas)\textsuperscript{10} and 0.933 in AUC, 93.3\% in SE, 91.1\% in SP on a set of acral volar 199 dermoscopy images (169 melanocytic nevi and 30 melanomas)\textsuperscript{11}. Our present system provides the final diagnosis results in the form of a malignancy score between 0 and 100 within 3–10 s.

It is well known for dermatologists and researchers in this field that color information in dermoscopy images is very important for the visual [1] as well as the computer-aided diagnosis of melanoma. Dermoscopes should therefore produce accurate color images, but unfortunately this is not always true in practice. An expert dermatologist would perform the same diagnosis on a particular case even if the image is acquired in different imaging conditions. Device calibration to compensate for various imaging conditions such as magnification factors, lighting conditions, etc. is crucial for the development of a reliable system. This deficiency of our web-based system was also pointed out by Rubegni et al. [12].

To the best of our knowledge, color calibration in dermoscopy is still an open issue and has been little investigated [13–16]. Haeghen et al. [13] reported a color calibration method for images in

\textsuperscript{1} This site is temporarily redirected to http://b0112-web.k.hosei.ac.jp/DermoPerl/.

\textsuperscript{*} Corresponding author at: 3-7-2 Kajino-cho Koganei, 184-8584, Tokyo, Japan. Tel.: +81 42 387 6217; fax: +81 42 387 6381.
E-mail address: iyatomi@hosei.ac.jp (H. Iyatomi).

0895-6111/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2010.08.003
dermatology. The main objective of this method is inter-camera calibration by estimating internal camera parameters. This method is reported to require 5 min of manual operation. Maglogiannis et al. [14] and Graña et al. [15] reported color calibration methods for digital skin image and dermoscopy images, respectively. These methods perform color calibration using color charts that contain a predefined set of printed color squares. Since hardware-based color calibration is not feasible in a web-based system, we need to address this issue using a software-based approach.

In our experience, most of these inappropriate color conditioned images can be calibrated effectively by adjustment of the hue and intensity of the images in the HSV color system. In this paper, we develop a fully automated color calibration method for dermoscopy images using the HSV color system. The proposed method accomplishes the calibration process based on the image content alone.

2. Materials

Digital dermoscopy images of pigmented skin lesions (PSLs) were collected from three European university hospitals. We used two different data sets as described below:

- Dataset-A: 319 digital dermoscopy images (244 benign and 75 melanomas) were collected from the university hospitals of Naples and Graz as presented in the EDRA Interactive Atlas of Dermoscopy [17].
- Dataset-B: 537 digital dermoscopy images (457 benign and 80 melanomas) were collected from the University of Vienna.

All diagnoses were histopathologically or clinically confirmed. Dataset-A is used for the development of the color calibration filters and their quantitative evaluation, whereas Dataset-B is used for alternative evaluation of the developed filters.

3. Development of color calibration filters

3.1. Basic idea of the study

The requirement of this study is that all color calibration procedures are performed on the server side and accordingly we have to calibrate the image using the transmitted (obtained) image alone. This problem is referred to as “blind estimation” in control engineering. In this section, we briefly summarize the rationale behind our fully automated color calibration method.

Now we consider a dermoscopy image and modify it with certain predefined procedure (e.g. hue +10°, saturation −10%). Since the modification procedure is known, its restoring (inverse) procedure to obtain the original image from the modified image (e.g. hue −10°, saturation +10%) is also known. The main idea of this study is estimating this restoring procedure using quantitatively calculable image features extracted from the modified image. Based on a certain number of images (as training images) and their various modifications, we build linear regression models that describe the relationship between the restoring procedure (known) and features calculable from the image. Once we obtain a clear relationship between these two, these regression models can be used for the calibration of unsewn images.

3.2. Method overview

Fig. 1 shows the schematic of the proposed method. In this study, color calibration was performed based on the hue, saturation, and intensity (value) channels, independently. First, we design a total of M color modification filters
\[
\mathbf{g}_m(H, S, V) = \{g_m(H), g_m(S), g_m(V)\} \quad (1 \leq m \leq M). \tag{1}
\]

These filters modify the hue ($H$), saturation ($S$), and intensity ($V$) values of the pixels in the training image.

In this figure, \(f(x, y)\) denotes the \(i\)-th ($1 \leq i \leq N$) dermoscopy image in the training set and we obtain \(M\) different images \(f'_m(x, y)\) by applying the abovementioned filters to each input image.

\[
f'_m(x, y) = f(x, y) \odot g_m(H, S, V). \tag{2}
\]

Note that \(x\) and \(y\) denote the Cartesian coordinates of the pixels in an image and operator \(\odot\) performs the color modification.

On a data set of \(N \times M\) images (\(N\) training images and their \(M\) variations each), we extract a total of \(P\) image features \(f'_m(p)\) ($1 \leq j \leq N \times M$) ($1 \leq p \leq P$) from each image \(f'_m(x, y)\). Now let us consider the relationship between the extracted “image features” \(f'_m(p)\) and the appropriate restoring procedure \(h'_m(H, S, V)\) for the training dataset (and their variations), the appropriate restoring factor \(h'_m(H, S, V)\) is already known as the inverse form of the color modification filter as follows:

\[
h'_m(H, S, V) = g'_m(-H, 1/S, 1/V). \tag{3}
\]

The task of color calibration for an image can be considered as a problem of estimating appropriate \(h(H, S, V)\). Assuming an image \(f(x, y)\) and extracted image features from this image \(p\), the regression models \(C_m(p)\) can be considered as the estimated restoring factor \(h'_m(H, S, V)\) for \(f(x, y)\). If we build a robust regression model based on a sufficiently large training set, \(\hat{h}(H, S, V)\) will be closer to the ideal \(h(H, S, V)\),

\[
C_m(p) = \begin{bmatrix} C_H(p) \\ C_S(p) \\ C_V(p) \end{bmatrix} = \begin{bmatrix} \hat{h}(H) \\ \hat{h}(S) \\ \hat{h}(V) \end{bmatrix} = \hat{h}(H, S, V) \rightarrow h(H, S, V). \tag{4}
\]

For this reason, we can call this regression model a “color calibration filter”. The color calibration filters \(C_m(p)\) are the final products of this study.

Finally, color calibration process of an image \(f(x, y)\) can be described as

\[
f(x, y) = f(x, y) \odot C_m(p) = f(x, y) \odot \hat{h}(H, S, V) \tag{5}
\]

where \(f(x, y)\) is the calibrated image.

![Fig. 1. Overview of building a color calibration filter.](image-url)
Table 1
172 base image features for color calibration.

<table>
<thead>
<tr>
<th>Value (5 kinds)</th>
<th>Color channel (6 kinds)</th>
<th>Region (4 kinds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (min), average (μ), maximum (max), standard deviation (σ), skewness (κ)</td>
<td>Red (R), green (G), blue (B), hue (H), saturation (S), value (V)</td>
<td>Tumor (T), peripheral (P), tumor–peripheral (T–P), peripheral–normal (P–N)</td>
</tr>
<tr>
<td>(ii) Number of colors</td>
<td>Color channel (2 kinds)</td>
<td>Region (2 kinds)</td>
</tr>
<tr>
<td>Quantize level (2 kinds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8, 16</td>
<td>RGB, HSV</td>
<td>Tumor (T), peripheral (P)</td>
</tr>
<tr>
<td>(iii) Other color features</td>
<td>Color channel (6 kinds)</td>
<td>Region (2 kinds)</td>
</tr>
<tr>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average (μ)</td>
<td>Red (R), green (G), blue (B), hue (H), saturation (S), value (V)</td>
<td>Normal skin (N), peripheral–tumor (P–T)</td>
</tr>
<tr>
<td>(iv) Color features of border</td>
<td>Color channel (2 kinds)</td>
<td>Size of windows 5g (8 kinds)</td>
</tr>
<tr>
<td>Value (2 kinds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8 × 2 × 2 = 32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio (inside:outside) (a), gradient (V)</td>
<td>Blue (B), value (V)</td>
<td>L/5, L/10, L/15, L/20, L/25, L/30, L/35, L/40</td>
</tr>
</tbody>
</table>

* L: length of the major axis of the tumor object.

3.3. Experimental setting

We used Dataset-A as the training dataset (N = 319) and prepared a total of 75 modified images per image (M = 75) where the modification involves five different hue values H = {−10, −5, 0, 5, 10}, three different saturation values S = {0.9, 1.0, 1.1}, and five different intensity values V = {0.8, 0.9, 1.0, 1.1, 1.2}. Note that if the modified value exceeds the limit of each channel, we use the maximum possible value instead. We extracted a total of 172 image features from each image (P = 172). Accordingly, a total of 319 × 75 pairs of 172-dimensional feature vectors \( f^f_p \) and the corresponding calibration factors \( h_i^m(H, S, V) \) were used to build the color calibration filter \( C_i(p) = \hat{h}(H, S, V) = \hat{g}(-H, 1/S, 1/V) \); regression model) in each color channel.

3.4. Building the color calibration filters

In order to build robust and accurate color calibration filters, extraction of effective image features is important. The proposed method first determines the tumor area and then extracts image features from the inside, outside, and periphery of the tumor and from the entire image. Based on these extracted features, we build multiple regression models with statistical feature selection to avoid problems of multi-collinearity or over-fitting.

3.4.1. Tumor area extraction

Accurate determination of the tumor area is one of the most important steps in the computer-aided diagnosis of melanoma [18]. In this study, we used our “dermatologist-like” tumor area extraction algorithm [19]. This algorithm was shown to be highly accurate [20, 21] and has been used successfully in our previous studies [10, 11].

3.4.2. Image feature extraction

After extracting the tumor area, we rotated the tumor object to align its major axis with the Cartesian x-axis. We then extracted a total of 172 color related objective features from the image. These were (i) 120 primitive color features (30 × 4 areas), (ii) 8 polychrome features, (iii) 12 other color related features, and (iv) 32 color gradient features in the peripheral areas. Table 1 summarizes the image features used in this study. Their detailed explanations are follows:

(i) Primitive color features: minimum (min), average (μ), maximum (max), standard deviation (σ) and skewness (κ) values in the RGB and HSV color spaces, respectively (subtotal 30) for the whole tumor area (T: tumor), periphery of the tumor area (P: peripheral), difference between the tumor area and the surrounding normal skin (T–N: tumor–normal skin) and difference between peripheral and normal skin (P–N: peripheral–normal skin). Note that the peripheral part of the tumor is defined as the region inside the border that has an area equal to 30% of the tumor area and determined by a recursive dilation process applied to the outer border, working inward from the border of the extracted tumor. The ratio of 30% was determined in our preliminary experiments based on the visual assessment provided by several dermatologists.

(ii) Polychromatic features: The number of colors in the tumor area and peripheral tumor area in the RGB and HSV color spaces quantized to \( 8^3 \) (#CRGB, #CHSV) and \( 16^3 \) (#CRGB16, #CHSV16) colors, respectively (2 regions × 2 color spaces × 2 quantization levels: subtotal 8).


(iv) Color gradient of border: The tumor area was divided into eight equi-angle regions. In each region, we defined a window of size \( S_p \times S_g \) pixels that is centered on the border of the tumor. In each window, the ratio of the color intensity inside and outside of the tumor and the gradient of color intensity were calculated in the blue and luminance channels (ratio: \( \alpha_B \) and \( \alpha_V \); gradient: \( V_B \) and \( V_V \)), respectively. These were averaged over the 8 equi-angle regions. We calculated four features for eight different window sizes \( S_p: 1/5, 1/10, 1/15, 1/20, 1/25, 1/30, 1/35 \) and \( 1/40 \) of the length of the major axis of the tumor object (L).

3.4.3. Feature selection

We used many color related image features and some of which may be correlated. It is well known that building a regression model with highly correlated parameters is adversely affected by the so-called multicollinearity, and, in such a case, the model loses accuracy and generality.

In this study, incremental stepwise feature selection was performed to select only the most significant image features for each regression model (e.g. H, S, and V). This feature selection
Developed calibration filter functions (first 10 features.)

Table 2

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Feature</th>
<th>Coefficient</th>
<th>Feature</th>
<th>Coefficient</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.140</td>
<td>min(^N)</td>
<td>1.51 \times 10^{-4}</td>
<td>(k_i)</td>
<td>4.07 \times 10^{-3}</td>
<td>(k_i)</td>
</tr>
<tr>
<td>0.111</td>
<td>(\sigma^p)</td>
<td>1.59 \times 10^{-4}</td>
<td>(\mu^p)</td>
<td>1.63 \times 10^{-4}</td>
<td>(\mu^p)</td>
</tr>
<tr>
<td>-0.375</td>
<td>(\mu^N)</td>
<td>5.55 \times 10^{-5}</td>
<td>(k_i)</td>
<td>1.15 \times 10^{-3}</td>
<td>(k_i)</td>
</tr>
<tr>
<td>-0.252</td>
<td>(\mu^N)</td>
<td>1.00 \times 10^{-3}</td>
<td>#4/33a</td>
<td>-4.58 \times 10^{-3}</td>
<td>#4/33a</td>
</tr>
<tr>
<td>-1.209</td>
<td>(\mu^N)</td>
<td>-2.28 \times 10^{-3}</td>
<td>(\sigma^p)</td>
<td>6.74 \times 10^{-2}</td>
<td>(\sigma^p)</td>
</tr>
<tr>
<td>1.461</td>
<td>(\mu^N)</td>
<td>2.09 \times 10^{-3}</td>
<td>(\mu^N)</td>
<td>1.73 \times 10^{-4}</td>
<td>(\mu^N)</td>
</tr>
<tr>
<td>0.016</td>
<td>(k_i)</td>
<td>-4.92 \times 10^{-4}</td>
<td>max(^S)</td>
<td>-1.36 \times 10^{-4}</td>
<td>max(^S)</td>
</tr>
<tr>
<td>0.011</td>
<td>(k_i)</td>
<td>5.21 \times 10^{-4}</td>
<td>min(^N)</td>
<td>-1.00 \times 10^{-4}</td>
<td>min(^N)</td>
</tr>
<tr>
<td>0.058</td>
<td>#4/33a</td>
<td>4.39 \times 10^{-4}</td>
<td>(\mu^N)</td>
<td>-7.89 \times 10^{-4}</td>
<td>(\mu^N)</td>
</tr>
<tr>
<td>0.012</td>
<td>min(^N)</td>
<td>-1.62 \times 10^{-3}</td>
<td>#4/33a</td>
<td>-4.90 \times 10^{-4}</td>
<td>#4/33a</td>
</tr>
<tr>
<td>-3.180</td>
<td>Const.</td>
<td>0.869</td>
<td>Const.</td>
<td>0.119</td>
<td>Const.</td>
</tr>
</tbody>
</table>

Upper right suffix means features in T: tumor area, P: peripheral area, N: normal skin area, P-N: peripheral-normal, peripheral-tumor area, tumor area-normal skin area, respectively. Lower right suffix indicates color channel.

method rejected statistically negligible features during incremental selection and therefore, these highly correlated features were automatically excluded from the model. The details of the feature selection procedure are given below:

(Step 0) Set base parameter BP to NULL and number of base parameter # BP = 0.

(Step 1) Search one input parameter \(x^c\) from all image features \(x\) where regression model with \(x^c\) yields best performance (lowest residual) among all. Set BP as \(x^c\) and # BP = 1.

(Step 2) Select one input candidate \(x^c\) which has the highest partial correlation coefficient from all image features without redundancy and build a linear regression model whose input elements are \((BP + x^c)\). (Number of input is # BP + 1)

(Step 3) Perform the statistical F-test to check all the selected parameters \((BP + x^c)\) are significant (p<0.05) for the regression model.

(Step 4) If yes in (step 3), \(x^c\) is added to BP, # BP ← # BP + 1 and return to (step 2). Else if the developed model has a statistically negligible parameter \(x^c\) (p>0.10), reject \(x^c\) from BP, # BP ← # BP − 1 and return to (step 2). Otherwise Terminate input selection process.

Based on these selected features, we build a linear multiple regression model for each of the H, S, and V channels.

4. Results

4.1. Modeling performance

The incremental stepwise feature selection method selected regression models with 48, 40 and 47 parameters for \(C_H(p), C_S(p), \) and \(C_V(p), \) respectively.

The color calibration filters \(C_i(p)\) built using the 10 first selected features are summarized in Table 2. In the column entitled ‘feature’, for instance, \(\text{min}_{\text{N}}\) denotes the minimum hue value in the peripheral part of the tumor. Note that in this study, we used \(g^i_{\text{im}}(H, S, V)\) as a target value of the linear regression models instead of using \(h^i_{\text{im}}(H, S, V)\) as explained in Section 3.2 and accordingly, the developed calibration models are in negative or division form. Note that all selected parameters for regression models are shown in Table 7 in Appendix.

Table 3 summarizes the modeling performance of the linear regression models. From left to right, the columns in the table correspond to the number of constituent image features of the regression model (# p), determination coefficients adjusted by degree of freedom \(R^2\), standard error of the regression model (E). Note again that hue (H) is in degrees (0–360 ), while saturation (S) and intensity (V) are expressed as multiplication factor.

4.2. Calibration performance

We evaluated the color calibration performance with 75-fold cross-validation test using Dataset-A. Note that this evaluation involves separate training and test data sets. The results are summarized in Table 4.

Fig. 2 shows examples of color calibration results. Note that here the modified image (b) is obtained from the original image (a) using the pre-defined modification filters \(g(H, S, V)\). In (c), we show estimated modified value \(\hat{g}(H, S, V)\) as a reference. In these examples, the test image is excluded from the training set during the development of the corresponding calibration filter. The result of cross-validation is almost equivalent to the training error and this error range is acceptable from a visual perspective.

Now we would like to test the performance of the proposed color calibration method for the case where the test images are different (Dataset-B) from those in the training data set (Dataset-A). Figs. 3 and 4 show examples of color calibration results. Based on visual assessment, all images seem to be appropriately calibrated as well as those in Dataset-A.

5. Discussion

5.1. HSV color system

We used the HSV color system for the color calibration task. The HSV color system is capable of handling brightness (luminance, lightness and intensity) and chromaticity information, separately. The V (value) of the HSV is also called intensity and it provides brightness information independent from the chromatic one. The brightness of a captured image is highly dependent on both shutter speed and aperture of the camera. Using the V channel, it is much easier to calibrate the brightness of the image rather than using the RGB color system. In a similar way, the H (hue) and S (saturation) of the HSV color system are also useful for calibrating the white-balance of the camera. For the reasons mentioned above, we used the HSV color system in our study.

Table 4

<table>
<thead>
<tr>
<th>Criterion</th>
<th>MAE ± SD</th>
<th># p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hue</td>
<td>4.34 ± 2.51 (1.2 ± 1.5%)</td>
<td>48</td>
</tr>
<tr>
<td>Satiation</td>
<td>0.069 ± 0.073 (6.9 ± 7.3%)</td>
<td>40</td>
</tr>
<tr>
<td>Intensity</td>
<td>0.047 ± 0.053 (4.7 ± 5.3%)</td>
<td>47</td>
</tr>
</tbody>
</table>

Note: MAE: mean absolute error, SD: standard deviation, Hue is in degrees (0–360 ), while saturation and intensity are expressed as multiplying factor.
Fig. 2. Sample results of color calibration (Dataset-A: known data).
Fig. 3. Sample results of color calibration (Dataset-B: unknown data (benign)).
Fig. 4. Sample results of color calibration (Dataset-B: unknown data (malignant)).
5.2. Evaluation of color calibration – Dataset-A

Our regression model achieved good modeling and calibration performance for hue (modeling error = 1.2%, error under cross-validation = 1.2%) and intensity (modeling error = 4.2%, error under cross-validation = 4.7%). Calibration performance for saturation was inferior to others, however, and error of about 6.2% (6.9% in cross-validation) in saturation does not have a large impact compared with that in hue under visual assessment (see Fig. 2 case 4: 12%, case 5: 5%). The calibration performance is comparable with previously published hardware-based calibration results (mean error in RGB channel = 3.1%).

In this study, we highlighted 10 features each for calibration model in hue, saturation and intensity obtained by input selection (Table 3). Note that parameters chosen early in the stepwise feature selection are considered to be more important for the classification because the most statistically significant parameters are selected in each step. From Table 3, it can be seen that the calibration model developed using only 10 features shows reasonable calibration performance, especially for intensity. We can see that similar calibration performance is obtained using only these features.

Next, we discuss important image features for calibrating the image using Table 2. Interestingly, the peripheral part of the tumor is crucial for the estimation of the color calibration filter. Especially in the hue channel, 9 out of 10 features are related with ones from the peripheral part of the tumor. This selectivity is much higher than that observed in classification models. On the other hand, the gradient of these areas was hardly ever selected.

5.3. Evaluation of color calibration – Dataset-B

As for Dataset-B, since we have no ground truth information defining the appropriate calibration, we cannot perform a quantitative evaluation directly. Alternatively, we can compare the distribution of image features among the training dataset (Dataset-A), the test dataset (Dataset-B), and the test dataset after color calibration. This evaluation is based on the assumption that if color calibration works properly, the color distribution of the test images would be closer to that of the training set.

We conducted principal component analysis (PCA) and orthogonalized a total of 172 color related image features described above. Fig. 5 compares the parameter distribution with two principal components for the sake of simplicity (accumulated contribution ratio of the first two components was 72.2%). Note that we randomly reduced the number of points in the plot to 100 in each dataset for visualization purposes. We can see that the color calibration process makes the color distribution of the test set closer to that of the training set. The centroid and standard deviation of the data distribution are summarized in Table 5.

From Figs. 3 and 4, we can see that the darker images are appropriately calibrated and images that require no calibration are left unchanged. Based on these observations, we can conclude that our method also has the ability to perform color calibration for unseen data.

Table 6

<table>
<thead>
<tr>
<th>Skin color in red channel $\mu^R$</th>
<th>250</th>
<th>240</th>
<th>230</th>
<th>220</th>
<th>210</th>
<th>200</th>
<th>190</th>
<th>180</th>
<th>170</th>
<th>160</th>
<th>150</th>
<th>140</th>
<th>130</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated skin intensity $\mu^R$</td>
<td>232.4</td>
<td>223.1</td>
<td>213.8</td>
<td>204.5</td>
<td>195.2</td>
<td>185.9</td>
<td>176.6</td>
<td>167.3</td>
<td>158.0</td>
<td>148.7</td>
<td>139.4</td>
<td>130.1</td>
<td>120.8</td>
</tr>
<tr>
<td>Calibration ratio $\hat{C}_V(p)$</td>
<td>0.880</td>
<td>0.920</td>
<td>0.963</td>
<td>1.010</td>
<td>1.062</td>
<td>1.120</td>
<td>1.184</td>
<td>1.257</td>
<td>1.339</td>
<td>1.432</td>
<td>1.539</td>
<td>1.663</td>
<td>1.810</td>
</tr>
<tr>
<td>Calibrated intensity $\hat{I}$</td>
<td>204.6</td>
<td>205.2</td>
<td>205.8</td>
<td>206.5</td>
<td>207.3</td>
<td>208.2</td>
<td>209.2</td>
<td>210.3</td>
<td>211.5</td>
<td>213.0</td>
<td>214.6</td>
<td>216.5</td>
<td>218.7</td>
</tr>
</tbody>
</table>

\[ a \mu^R = (211.95/228.01) \text{ (from Eq. (7)).} \]

\[ b \mu^V = \hat{C}_V(p). \]
calibrated intensity value ($\mu_{V}^{N} \times C_{V}$). We can confirm from this table that the pixel with intensity value out of normal range is appropriately calibrated to the certain value. (i.e. Average of $\mu_{V}^{N} \approx 212$)

Intensity of images in Dataset-B is generally lower than those in Dataset-A. Note that the EDRA data set (source of Dataset-A) is widely used in this field. In Figs. 3 and 4, we can see several examples in which the color distribution of darker images are corrected appropriately.

### 5.5. Calibration for vascular coloration

The vascular coloration (peripheral erythema, vascular blush) which characterizes malignancies of all types appears to be diminished in some cases by this method, and we are conducting ongoing studies to preserve this feature in color-calibrated images.

### 6. Conclusions

In this paper we developed a new color calibration method for dermoscopy images. The key feature of our method is that color calibration can be performed based on image content alone. Color calibration for dermoscopy images is especially important for computer-aided diagnosis systems. We are currently investigating the influence of color calibration on the final diagnosis.

### Acknowledgements

This research was partially supported by the Ministry of Education, Culture, Science and Technology of Japan (Grant-in-Aid for Young Scientists program (B), 21791101, 2009–2010) and the Louisiana Board of Regents (LEQSF2008-11–RD-A-12).

### Appendix A.

Table 7 shows all the image features of developed color calibration filter $C_{A}(p)$ selected by incremental stepwise method. Note that some of the features are different from those shown in Table 2. This is because some features are eliminated during the increment steps.

### References


Hitoshi Iyatomi is a Lecturer of Hosei University, Japan. He was born in Tokyo, Japan in 1976. He received his B.E. and M.E. degrees in Electrical Engineering in 1998 and 2000 and Ph.D. degree in Open and Environmental Systems (Engineering) in 2004 from Keio University, respectively. From 2000–2004 he was employed by Hewlett Packard Japan. His research interests include image understanding, machine learning, and medical image analysis.

M. Emre Celebi received the B.Sc. degree in computer engineering from Middle East Technical University, Ankara, Turkey, in 2002 and the M.Sc. and Ph.D. degrees in computer science and engineering from the University of Texas at Arlington in 2003 and 2006, respectively. He is currently an Assistant Professor in the Department of Computer Science, Louisiana State University, Shreveport. His research interests include medical image analysis, color image processing, content-based image retrieval, and open-source software development.

Gerald Schaefer gained his BSc. in Computing from the University of Derby and his PhD in Computer Vision from the University of East Anglia. He worked at the Colour & Imaging Institute, University of Derby (1997–1999), in the School of Information Systems, University of East Anglia (2000–2001), in the School of Computing and Informatics at Nottingham Trent University (2001–2006), and in the School of Engineering and Applied Science at Aston University (2006–2009) before joining the Department of Computer Science at Loughborough University. His research interests are mainly in the areas of colour image analysis, image retrieval, physics-based vision, medical imaging, and computational intelligence.

Masaru Tanaka is Professor of Dermatology, Tokyo Women’s Medical University Medical Center East from 2007. He was born in Tokyo, Japan in 1958. He received his M.D. and Ph.D. degrees at Keio University in 1984 and 1992, respectively. He studied abroad in Cardiff, UK from 1993–1995. He was Asistant Professor from 1996 and Associate Professor from 1999 at the Department of Dermatology, Keio University School of Medicine. His research interests include dermatopathology, dermatoscopy and image analysis.