Computerized Medical Imaging and Graphics 35 (2011) 89–98

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Automated color calibration method for dermoscopy images

Hitoshi Iyatomi[a,*], M. Emre Celebi[b], Gerald Schaefer[c], Masaru Tanaka[d]

[a] Department of Applied Informatics, Hosei University, Tokyo, Japan
[b] Department of Computer Science, Louisiana State University, Shreveport, USA
[c] Department of Computer Science, Loughborough University, United Kingdom
[d] Department of Dermatology, Tokyo Women's Medical University Medical Center East, Japan

ARTICLE INFO

Article history:
Received 1 October 2009
Accepted 16 August 2010

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

ABSTRACT

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.

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.[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) [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) [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 use in

* 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).

[1] This site is temporarily redirected to http://b0112-web.k.hosei.ac.jp/DermoPerl/.

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 Grana 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 pre-defined 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 unseen 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

$$g_m(H,S,V) = \{g_m(H), g_m(S), g_m(V)\} \quad (1 \le m \le M). \tag{1}$$

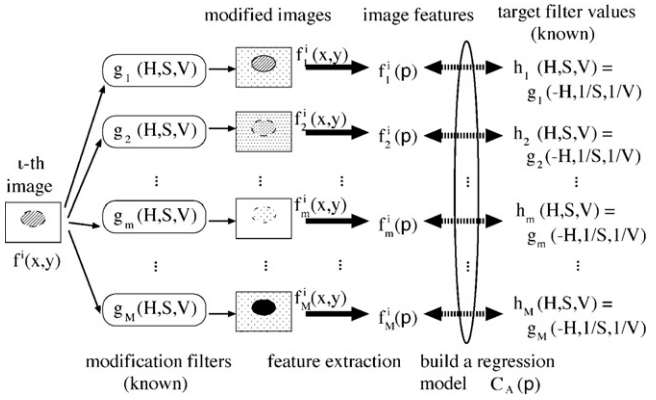

Fig. 1. Overview of building a color calibration filter.

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

In this figure, $f^i(x,y)$ denotes the ith ($1 \le i \le N$) dermoscopy image in the training set and we obtain M different images $f^i_m(x,y)$ by applying the abovementioned filters to each input image.

$$f^i_m(x,y) = f^i(x,y) \odot g_m(H,S,V). \tag{2}$$

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 $f^j_m(\boldsymbol{p})(1 \le j \le N \times M)(1 \le \boldsymbol{p} \le P)$ from each image $f^j_m(x,y)$. Now let us consider the relationship between the extracted "image features" $f^j_m(\boldsymbol{p})$ and the appropriate restoring procedure $h^j_m(H,S,V)$. For the training dataset (and their variations), the appropriate restoring factor $h^j_m(H,S,V)$ is already known as the inverse form of the color modification filter as follows:

$$h^j_m(H,S,V) = g^j_m(-H, 1/S, 1/V). \tag{3}$$

Note that hue (H) is in degrees, while saturation (S) and intensity (V) are expressed as multiplication factor and therefore the inverse form can be written as above. Thus, we can build multiple regression models $C_A(\boldsymbol{p}) = \{C_H(\boldsymbol{p}), C_S(\boldsymbol{p}), C_V(\boldsymbol{p})\}$ from the $N \times M$ combinations of "image features" $f^j_m(\boldsymbol{p})$ and the appropriate "calibration factor" $h^j_m(H,S,V)$.

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 $\boldsymbol{p}$, the regression models $C_A(\boldsymbol{p})$ can be considered as the estimated restoring factor $\hat{h}(H,S,V)$ 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 $h(H,S,V)$,

$$C_A(\boldsymbol{p}) = \begin{bmatrix} C_H(\boldsymbol{p}) \\ C_S(\boldsymbol{p}) \\ C_V(\boldsymbol{p}) \end{bmatrix} = \begin{bmatrix} \hat{h}(H) \\ \hat{h}(S) \\ \hat{h}(V) \end{bmatrix} = \hat{h}(H,S,V) \to h(H,S,V). \tag{4}$$

For this reason, we can call this regression model a "color calibration filter". The color calibration filters $C_A(\boldsymbol{p})$ are the final products of this study.

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

$$\begin{aligned} f'(x,y) &= f(x,y) \odot C_A(\boldsymbol{p}) \\ &= f(x,y) \odot \hat{h}(H,S,V) \end{aligned} \tag{5}$$

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

Table 1
172 base image features for color calibration.

(i) Primitive color features Value (5 kinds)	Color channel (6 kinds)	($5 \times 6 \times 4 = 120$) Region (4 kinds)
Minimum (min), average (μ), maximum (max), standard deviation (σ), skewness (κ)	Red (R), green (G), blue (B), hue (H), saturation (S), value (V)	Tumor (T), peripheral (P), tumor–peripheral (T–P), peripheral–normal (P–N)
(ii) Number of colors Quantize level (2 kinds)	Color channel (2 kinds)	($2 \times 2 \times 2 = 8$) Region (2 kinds)
8, 16	RGB, HSV	Tumor (T), peripheral (P)
(iii) Other color features Value	Color channel (6 kinds)	($1 \times 6 \times 2 = 12$) Region (2 kinds)
Average (μ)	Red (R), green (G), blue (B), hue (H), saturation (S), value (V)	Normal skin (N), peripheral–tumor (P–T)
(iv) Color features of border Value (2 kinds)	Color channel (2 kinds)	($8 \times 2 \times 2 = 32$) Size of windows S_B (8 kinds)
Ratio (inside:outside) (α), gradient (∇)	Blue (B), value (V)	$L/5$, $L/10$, $L/15$, $L/20$, $L/25$, $L/30$, $L/35$, $L/40$ [a]

[a] 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 319×75 pairs of 172-dimensional feature vectors $f_m^j(\boldsymbol{p})$ and the corresponding calibration factors $h_m^j(H, S, V)$ were used to build the color calibration filter $C_A(\boldsymbol{p})$ ($= \hat{\mathrm{h}}(H, S, V) = \hat{\mathrm{g}}(-H, 1/S, 1/V)$; 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 ($\#C_{RGB8}$, $\#C_{HSV8}$) and 16^3 ($\#C_{RGB16}$, $\#C_{HSV16}$) colors, respectively (2 regions $\times$ 2 color spaces $\times$ 2 quantization levels: subtotal 8).

(iii) Other color features: The average color of surrounding normal skin (μ^N: R, G, B, H, S, V: subtotal 6), and average color differences between the peripheral tumor area and inside of the tumor area (μ^{P-T}: R, G, B, H, S, V: subtotal 6).

(iv) Color gradient of border: The tumor area was divided into eight equi-angle regions. In each region, we defined a window of size $S_B \times S_B$ 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: α_B and α_V, gradient: ∇_B and ∇_V), respectively. These were averaged over the 8 equi-angle regions. We calculated four features for eight different window sizes S_B; 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

Table 2
Developed calibration filter functions (first 10 features.)

$\hat{g}(H)(=-C_H(\boldsymbol{p}))$		$\hat{g}(S)(=1/C_S(\boldsymbol{p}))$		$\hat{g}(V)(=1/C_V(\boldsymbol{p}))$	
Coefficient	Feature	Coefficient	Feature	Coefficient	Feature
0.140	$\min_H^P$	1.51×10^{-4}	${\kappa_S}^T$	4.07×10^{-3}	μ_R^N
0.111	σ_H^P	1.59×10^{-3}	μ_S^P	1.63×10^{-4}	${\kappa_S}^{T-N}$
−0.375	$\#C^P_{HSV8}$	5.55×10^{-5}	${\kappa_H}^P$	1.15×10^{-3}	$\max_V^P$
−0.252	μ_R^{P-N}	1.00×10^{-3}	$\#C^P_{RGB16}$	-4.58×10^{-3}	σ_S^P
−1.209	μ_G^{P-N}	-2.28×10^{-3}	σ_G^T	6.74×10^{-2}	$\nabla_V^{1/20}$
1.461	μ_V^{P-N}	2.09×10^{-3}	μ_B^{P-N}	1.73×10^{-4}	${\kappa_V}^P$
0.016	${\kappa_G}^P$	-4.92×10^{-4}	$\max_G^T$	-1.36×10^{-3}	$\max_H^P$
−0.011	${\kappa_B}^P$	5.21×10^{-4}	$\min_R^T$	-1.00×10^{-3}	μ_B^{P-N}
0.058	$\#C^P_{HSV16}$	4.39×10^{-4}	μ_R^{P-N}	-7.89×10^{-4}	μ_S^{P-N}
0.012	$\min_H^T$	-1.62×10^{-3}	$\#C^P_{RGB8}$	-4.90×10^{-4}	μ_R^{P-T}
−31.80	Const.	0.869	Const.	0.119	Const.

Upper right suffix means features in *T*: tumor area, *P*: peripheral area, *N*: normal skin area, *P–N*, *P–T*, *T–N*:difference between 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^* from all image features **x** where regression model with x^* yields best performance (lowest residual) among all. Set *BP* as x^* and $\#_{BP}=1$.
(step 2) Select one input candidate $x^\wedge$ which has the highest partial correlation coefficient from all image features without redundancy and build a linear regression model whose input elements are $(BP+x^\wedge)$. (number of input is $\#_{BP}+1$)
(step 3) Perform the statistical *F*-test to check all the selected parameters $(BP+x^\wedge)$ are significant (p<0.05) for the regression model.
(step 4) If yes in (step 3), $x^\wedge$ is added to *BP*, $\#_{BP}\leftarrow\#_{BP}+1$ and return to (step 2). Else if the developed model has a statistically negligible parameter $x^\wedge$ (p>0.10), reject $x^\wedge$ from *BP*, $\#_{BP}\leftarrow\#_{BP}-1$ 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 $C_H(\boldsymbol{p})$, $C_S(\boldsymbol{p})$, and $C_V(\boldsymbol{p})$, respectively.

The color calibration filters $C_A(\boldsymbol{p})$ built using the 10 first selected features are summarized in Table 2. In the column entitled 'feature', for instance, $\min_H^P$ denotes the minimum hue value in the peripheral part of the tumor. Note that in this study, we used $g_m^j(H,S,V)$ as a target value of the linear regression models instead of using $h_m^j(H,S,V)$ 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 and accordingly the standard error *E* in hue appears to be larger than that in the other two channels.

Table 3
Modeling performance of regression model (Dataset-A).

Criterion	# *p* [a]	R^2	*E* [b]
Hue	48	0.623	4.15 (1.2%)
	10	0.434	5.32 (1.5%)
Saturation	40	0.296	0.062 (6.2%)
	10	0.204	0.073 (7.3%)
Intensity	47	0.890	0.042 (4.2%)
	10	0.857	0.053 (5.3%)

[a] Number of constitutive image features
[b] Hue is in degrees (0–360°), while saturation and intensity are expressed as multiplying 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
Calibration performance by cross-validation test.

Criterion	MAE ± SD [a]
Hue	4.34 ± 5.21 (1.2 ± 1.5%)
Saturation	0.069 ± 0.073 (6.9 ± 7.3%)
Intensity	0.047 ± 0.053 (4.7 ± 5.3%)

[a] MAE: mean absolute error, SD: standard deviation, Hue is in degrees (0–360°), while saturation and intensity are expressed as multiplying factor.

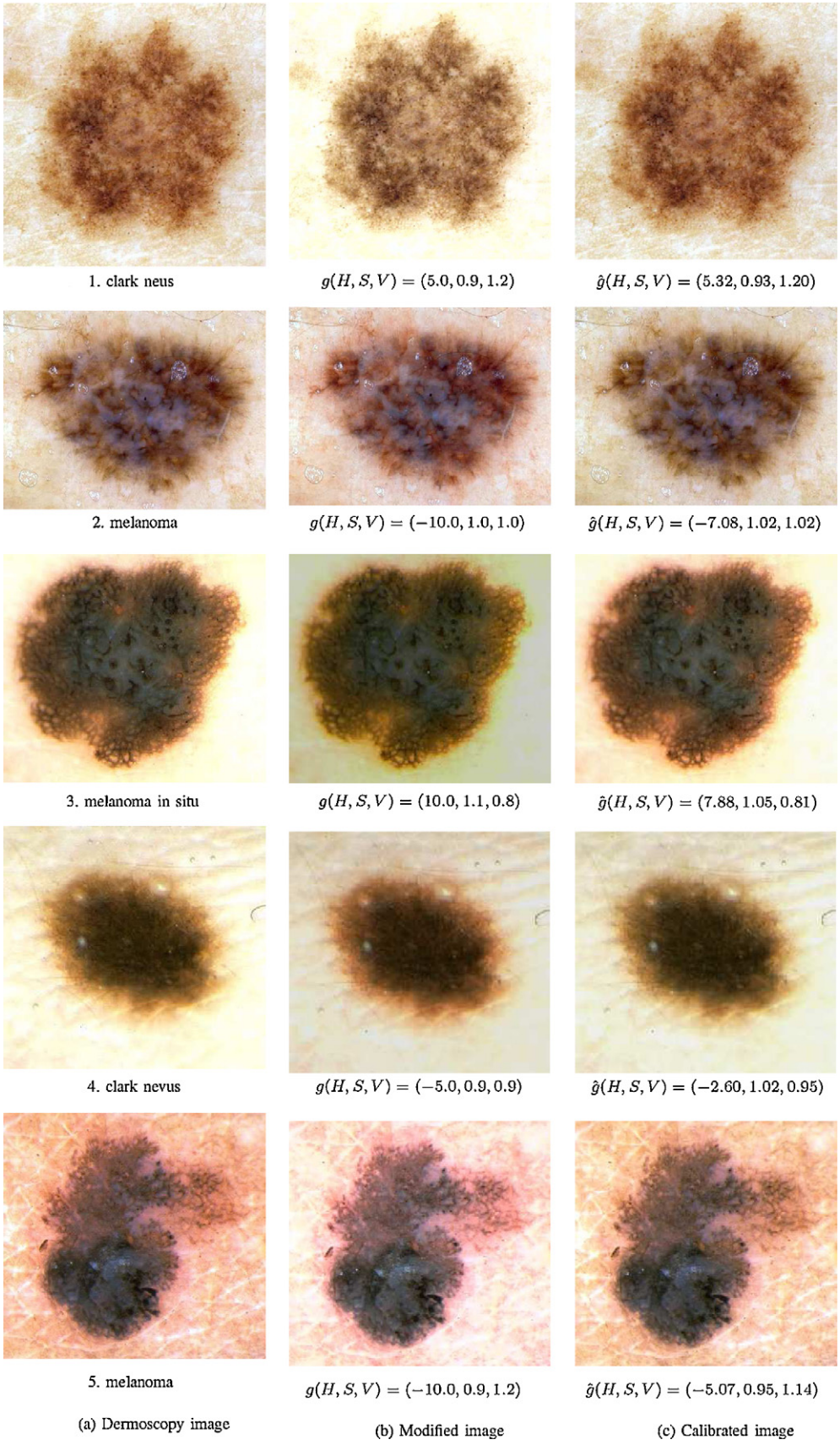

Fig. 2. Sample results of color calibration (Dataset-A: known data).

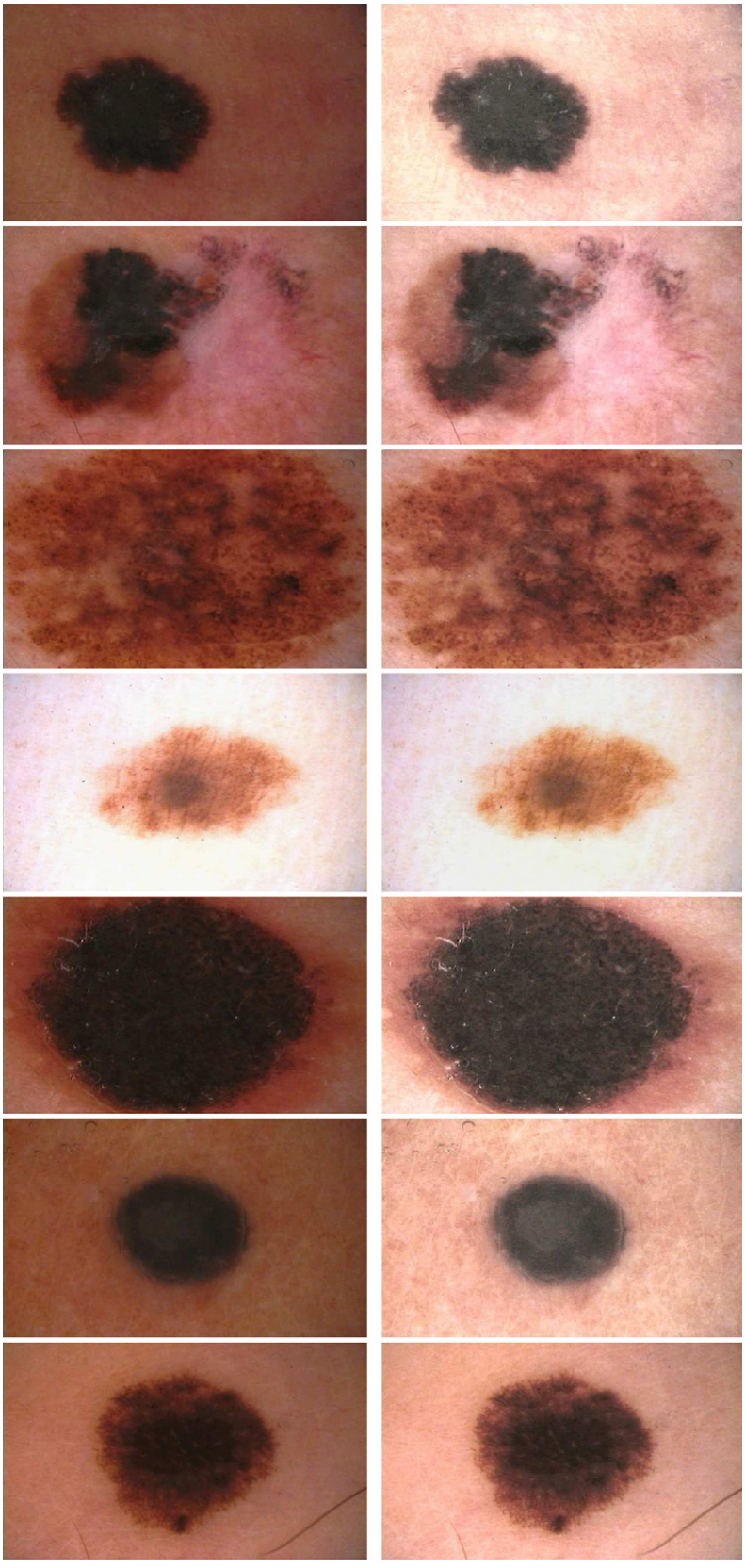

Fig. 3. Sample results of color calibration (Dataset-B: unknown data (benign)).

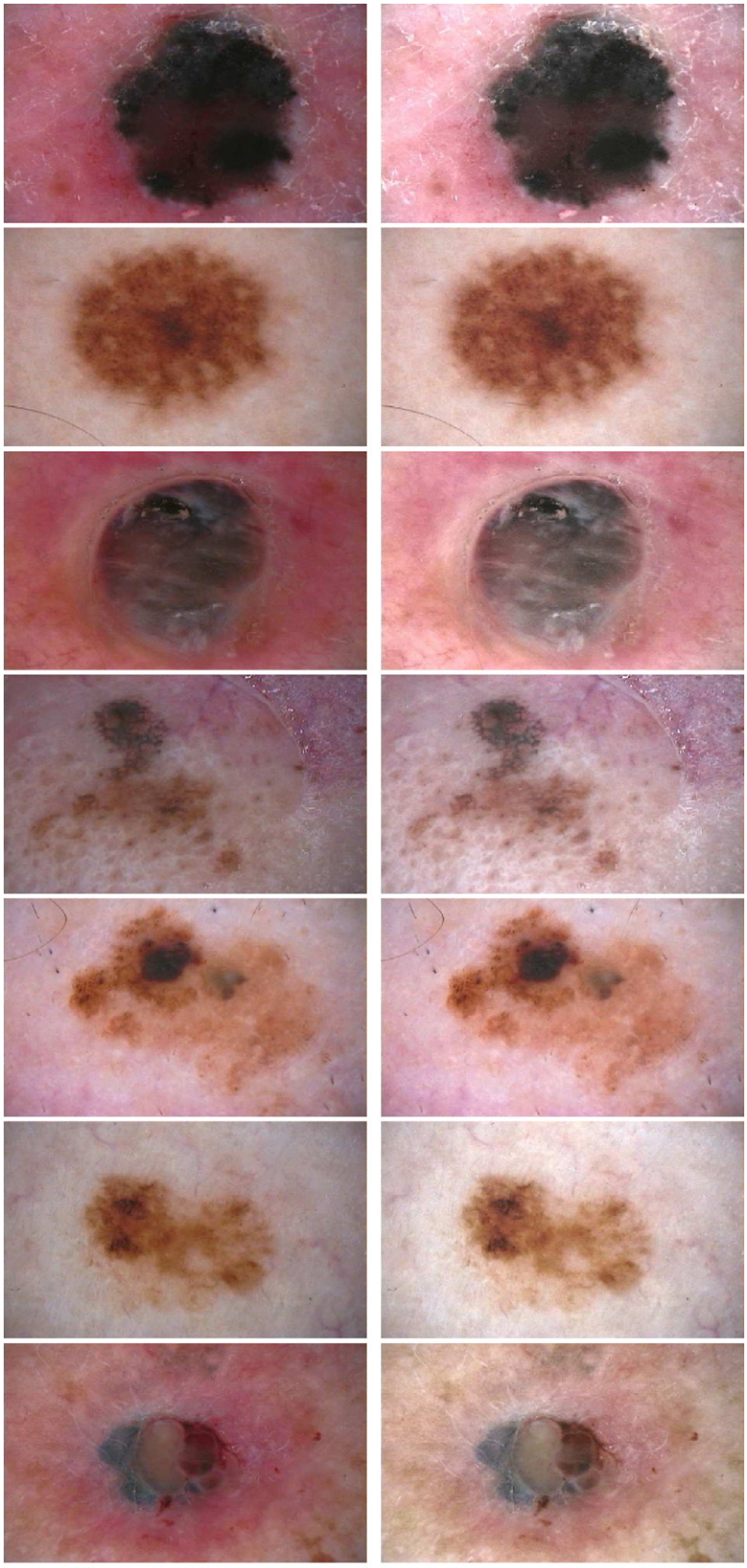

Fig. 4. Sample results of color calibration (Dataset-B: unknown data (malignant)).

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% [16]).

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 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.

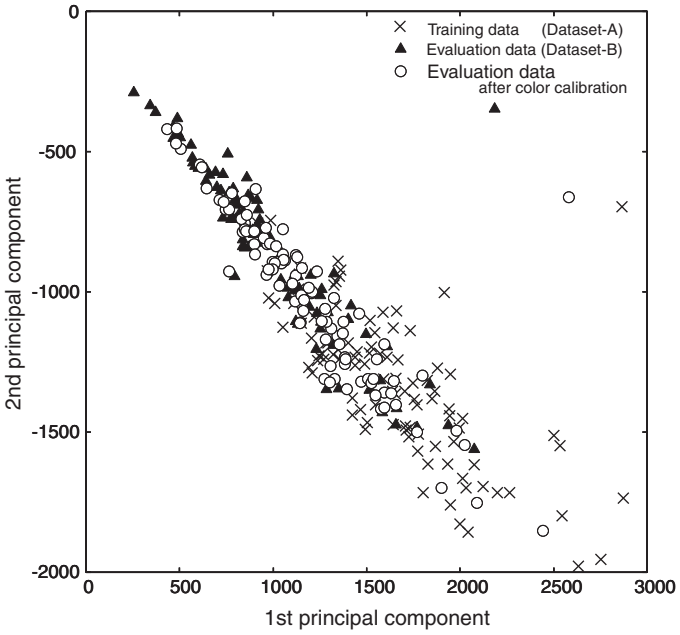

Fig. 5. Transition of feature distribution after color calibration.

Table 5
Comparison of distribution of color related image features.

Dataset	1st PC [a]	2nd PC
Dataset-A (training)	1891 ± 687	-1205 ± 430
Dataset-B	999 ± 390	-774 ± 307
Dataset-B with calibration	1206 ± 456	-945 ± 298

[a] Principal component.

5.4. Calibration of intensity

As for calibration of intensity, a linear regression model with only one factor μ_R^N (average red value of surrounding skin) shows especially high modeling performance (R^2=0.781, E=0.066) compared to the other two and it is described as

$$C_V(\boldsymbol{p}) = \frac{1}{0.0486\mu_R^N - 0.0795}. \quad (6)$$

This means that the intensity of a dermoscopy image can be calibrated using the red value of the surrounding normal skin which meets our intuition.

Now let us discuss calibration performance of intensity using this equation. Average and standard deviation of μ_R^N and μ_V^N (mean intensity of surrounding normal skin) in Dataset-A are 228.01 ± 12.98 and 211.95 ± 16.81, respectively. Correlation between μ_R^N and μ_V^N is 0.810. Since these two are strongly correlated, we assume here μ_V^N can be estimated by

$$\mu_V^N = \frac{211.95}{228.01}\mu_R^N \quad (7)$$

for simplicity. Table 6 shows the relationship among μ_R^N, estimated μ_V^N using Eq. (7), calibration factor $C_V(\boldsymbol{p})$ by Eq. (6), and

Table 6
Calibration of intensity using only one factor μ_R^N.

Skin color in red channel μ_R^N	250	240	230	220	210	200	190	180	170	160	150	140	130
Estimated skin intensity μ_V^N [a]	232.4	223.1	213.8	204.5	195.2	185.9	176.6	167.3	158.0	148.7	139.4	130.1	120.8
Calibration ratio $C_V(\boldsymbol{p})$	0.880	0.920	0.963	1.010	1.062	1.120	1.184	1.257	1.339	1.432	1.539	1.663	1.810
Calibrated intensity [b]	204.6	205.2	205.8	206.5	207.3	208.2	209.2	210.3	211.5	213.0	214.6	216.5	218.7

[a] $\mu_R^N \times (211.95/228.01)$ (from Eq. (7)).
[b] $\mu_V^N \times C_V(\boldsymbol{p})$.

Table 7
All selected features for color calibration filter $C_A(\boldsymbol{p})$.

	$\hat{g}(H)(=-C_H(\boldsymbol{p}))$	$\hat{g}(S)(=1/C_S(\boldsymbol{p}))$	$\hat{g}(V)(=1/C_V(\boldsymbol{p}))$
1	σ_H^P	μ_S^P	κ_S^{T-N}
2	$\#C_{HSV8}^P$	κ_H^P	$\max_V^P$
3	μ_R^{P-N}	$\#C_{RGB16}^P$	σ_S^P
4	μ_G^{P-N}	$\min_R^T$	$\alpha_V^{1/20}$
5	μ_V^{P-N}	μ_R^{P-N}	$\max_H^T$
6	κ_G^P	$\#C_{RGB8}^P$	μ_S^{P-N}
7	κ_B^P	μ_H^N	κ_V^{T-N}
8	$\#C_{HSV16}^P$	κ_R^T	σ_V^{T-N}
9	$\min_H^T$	μ_R^{T-N}	μ_H^{P-T}
10	$\nabla_B^{1/20}$	μ_S^T	$\max_V^{P-N}$
11	$\max_G^T$	$\min_H^T$	κ_S^{P-N}
12	$\max_V^T$	μ_S^{T-N}	σ_S^{P-N}
13	$\max_G^P$	μ_S^{P-N}	μ_G^P
14	σ_G^P	σ_S^P	$\alpha_V^{1/15}$
15	μ_R^N	σ_S^{T-N}	$\min_S^P$
16	κ_H^P	κ_B^T	$\#C_{RGB8}^T$
17	μ_S^N	$\nabla_V^{1/10}$	$\max_V^T$
18	σ_S^T	$\max_B^T$	$\max_G^T$
19	σ_S^{P-N}	$\max_B^P$	$\max_B^P$
20	κ_S^P	μ_V^N	$\#C_{HSV16}^P$
21	μ_B^N	κ_S^{P-N}	$\#C_{RGB16}^T$
22	$\nabla_V^{1/5}$	μ_B^T	σ_V^T
23	$\min_S^T$	$\max_V^T$	$\max_B^T$
24	$\min_V^{T-N}$	$\max_G^{P-N}$	$\alpha_V^{1/40}$
25	κ_B^{P-N}	$\min_S^{P-N}$	$\alpha_V^{1/30}$
26	σ_S^{T-N}	$\nabla_B^{1/20}$	σ_R^T
27	$\min_G^T$	$\nabla_B^{1/30}$	σ_R^{T-N}
28	$\min_R^{T-N}$	κ_S^T	σ_R^P
29	$\nabla_B^{1/15}$	κ_V^T	κ_V^{P-N}
30	$\alpha_B^{1/40}$	σ_S^{P-N}	κ_R^P
31	$\alpha_B^{1/5}$	κ_S^{T-N}	μ_S^N
32	μ_R^{P-T}	$\nabla_B^{1/10}$	μ_B^N
33	κ_R^{P-N}	$\nabla_B^{1/5}$	μ_H^N
34	κ_V^{P-N}	μ_R^T	μ_H^{P-N}
35	κ_R^T	σ_V^P	μ_H^{T-N}
36	μ_S^{T-N}	κ_G^P	$\max_R^T$
37	μ_B^{T-N}	σ_H^T	μ_S^T
38	μ_G^{P-T}	κ_B^P	$\nabla_V^{1/40}$
39	κ_G^{T-N}	$\alpha_V^{1/5}$	κ_V^P
40	κ_V^{T-N}	$\alpha_V^{1/10}$	σ_V^P
41	κ_V^T		$\alpha_V^{1/25}$
42	μ_S^P		μ_H^P
43	σ_R^{P-N}		$\min_H^P$
44	$\alpha_V^{1/40}$		μ_B^T
45	μ_S^{P-N}		κ_B^T
46	μ_B^{P-N}		κ_B^{T-N}
47	μ_S^{P-T}		κ_G^P
48	κ_R^{T-N}		

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 $\mu_V^N \cong 212$)

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 $C_A(\boldsymbol{p})$ 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

[1] Stolz W, Braun-Falco O, Bilek P, Landthaler M, Burgfort WHC, Cognetta AB. Color atlas of dermatoscopy—2nd enlarged and completely revised edition. Berlin: Blackwell Publishing; 2002.
[2] Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sara F, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of American Academy of Dermatology 2003;48(5):679–93.
[3] Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C. Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. British Journal of Dermatology 2004;151:1029–38.
[4] Ganster H, Pinz A, Rohrer R, Wilding E, Binder M, Kitter H. Automated melanoma recognition. IEEE Transactions on Medical Imaging 2001;20(3):233–9.
[5] Rubegni P, Cevenini G, Burroni M, Perotti R, Dell'Eva G, Sbano P, et al. Automated diagnosis of pigmented skin lesions. International Journal of Cancer 2002;101:576–80.
[6] Hoffmann K, Gambichler T, Rick A, Kreutz M, Anschuetz M, Grunendick T, et al. Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. British Journal of Dermatology 2003;149:801–9.
[7] Oka H, Hashimoto M, Iyatomi H, Tanaka M. Internet-based program for automatic discrimination of dermoscopic images between melanoma and Clark Nevi. British Journal of Dermatology 2004;150(5):1041.
[8] Seidenari S, Pellacani G, Grana C. Pigment distribution in melanocytic lesion images: a digital parameter to be employed for computer-aided diagnosis. Skin Research and Technology 2005;11:236–41.
[9] Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV, et al. A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics 2007;31(6):362–71.
[10] Iyatomi H, Oka H, Celebi ME, Hashimoto M, Hagiwara M, Tanaka M, et al. An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Computerized Medical Imaging and Graphics 2008;32(7):566–79.
[11] Iyatomi H, Oka H, Celebi ME, Ogawa K, Argenziano G, Soyer HP, et al. Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. Journal of Investigative Dermatology 2008;128:2049–54.
[12] Rubegni P, Burroni M, Sbano P, Andreassi L. Digital dermoscopy analysis and Internet-based program for discrimination of pigmented skin lesion dermoscopic images. British Journal of Dermatology 2006;152(2):395–6.
[13] Haeghen YV, Naeyaert JMAD, Lemahieu I. An imaging system with calibrated color image acquisition for use in dermatology. IEEE Transactions on Medical Imaging 2000;19(7):722–30.
[14] Maglogiannis I, Kosmopoulos DI. A system for the acquisition of reproducible digital skin lesions images. Technology and Health Care 2003;11:425–41.
[15] Grana C, Pellacani G, Seidenari S. Practical color calibration for dermoscopy applied to a digital epiluminescence microscope. Skin Research and Technology 2005;11:242–7.
[16] Maglogiannis I, Pavolopoulos S. An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images. IEEE Transactions on Information Technology in Biomedicine 2005;9(1):86–98.
[17] Argenziano G, Soyer HP, De Giorgi V. Dermoscopy: a tutorial. EDRA Medical Publishing & New Media; 2002.

[18] Celebi ME, Iyatomi H, Schaefer G, Stoecker WV. Lesion border detection in dermoscopy images. Computerized Medical Imaging and Graphics 2009;33(2):148–53.
[19] Iyatomi H, Oka H, Saito M, Miyake A, Kimoto M, Yamagami J, et al. Quantitative assessment of tumour area extraction from dermoscopy images and evaluation of the computer-based methods for automatic melanoma diagnostic system. Melanoma Research 2006;16(2):183–90.
[20] Celebi ME, Aslandogan YA, Stoecker WV, Iyatomi H, Oka H, Chen X. Unsupervised border detection in dermoscopy images. Skin Research and Technology 2007;13(4):454–62.
[21] Celebi ME, Kingravi H, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH, et al. Border detection in dermoscopy images using statistical region merging. Skin Research and Technology 2008;14(3):347–53.

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 Assitant Professor from 1996 and Associate Professor from 1999 at the Department of Dermatology, Keio University School of Medicine. His research interests include dermatopathology, dermoscopy and image analysis.