Improving Dermoscopy Image Classification Using Color Constancy

Catarina Barata, M. Emre Celebi, *Senior Member, IEEE*, and Jorge S. Marques

Abstract—Robustness is one of the most important characteristics of computer-aided diagnosis systems designed for dermoscopy images. However, it is difficult to ensure this characteristic if the systems operate with multisource images acquired under different setups. Changes in the illumination and acquisition devices alter the color of images and often reduce the performance of the systems. Thus, it is important to normalize the colors of dermoscopy images before training and testing any system. In this paper, we investigate four color constancy algorithms: Gray World, max-RGB, Shades of Gray, and General Gray World. Our results show that color constancy improves the classification of multisource images, increasing the sensitivity of a bag-of-features system from 71.0% to 79.7% and the specificity from 55.2% to 76% using only 1-D RGB histograms as features.

Index Terms—Color constancy, color features, computer-aided diagnosis system, dermoscopy images, image color normalization.

I. INTRODUCTION

IN the past decade, different groups have proposed computer-aided diagnosis (CAD) systems to identify melanomas in dermoscopy images [1]–[6]. These systems use several features, such as color, shape, and texture, to characterize the images. These features are usually inspired by medical algorithms such as the ABCD rule [7] and the seven-point checklist [8]. Most systems report highly accurate results. However, it is not easy to perform a fair comparison between their performances, since each system is trained and tested using a different dataset. Dealing with datasets acquired at different hospitals is a challenging task because they use different acquisition devices and illumination conditions. Both situations induce significant changes in the colors of the acquired images (see Fig. 1), leading to alterations in the values of the color features computed by the CAD system. Most of the proposed CAD systems do not incorporate a strategy to deal with the multisource problem. Therefore, they might not be robust in the presence of images generated by multiple sources. An exception is the internet-based system proposed by Iyatomi *et al.* [4], which incorporates a color calibration step [9].

In this paper, we investigate the problem of training and testing a color-based CAD system using multisource images. Our system is developed using a bag-of-features (BoF) model that performs well in the classification of dermoscopy images [10], [11] acquired at a single facility. To deal with the multisource problem, we assume that color variations can be corrected using color constancy algorithms [12]. The investigated calibration algorithms are Gray World [13], max-RGB [14], Shades of Gray [15], and General Gray World [16]. To the best of our knowledge, these color calibration strategies have never been applied to dermoscopy images and are significantly different from the previously proposed color calibration algorithms [9], [17]–[21]. Furthermore, the influence of color calibration in the performance of a melanoma detection system appears to have not been investigated.

The remaining paper is organized as follows. In Section II, a survey of the state-of-the-art is presented. In Section III, we describe the tested color constancy algorithms. In Section IV, we describe the system used to evaluate the performance of each color constancy algorithm, and in Section V, we show the experimental results. Section VI presents some conclusions.

II. RELATED WORK

The calibration of color in dermoscopy images is not a new problem [9], [17]–[20], [23], [24]. One of the first color calibration approaches was proposed by Haeghen *et al.* [17]. Their objective was to calibrate images by determining a set of internal camera parameters (e.g., camera offset, color gain, and aperture) and transform the image from a device output RGB color space to the standard sRGB space. The color transformation matrix is computed by acquiring images of the Gretag Macbeth color checker chart and determining the relationship between the images and the L*a*b* values of the chart, acquired with a spectrophotometer.

Grana *et al.* [18], Wighton *et al.* [19], and Quintana *et al.* [20] proposed some changes to the previously described pipeline. The major change is the use of the XYZ color space instead of L*a*b*. Furthermore, the first group tried to correct the nonuniform illumination of the dermatoscope [18], while in [19], a chromatic aberration calibration was included. Quintana *et al.* [20] also take into account the spectral reflectance of the dermatoscope lighting system when computing the transformation matrix.

The main issue with hardware-based calibration algorithms is that they require, that we know, the parameters of the cameras as well as the transformation matrices. This kind of information is not available when one is working with heterogeneous databases
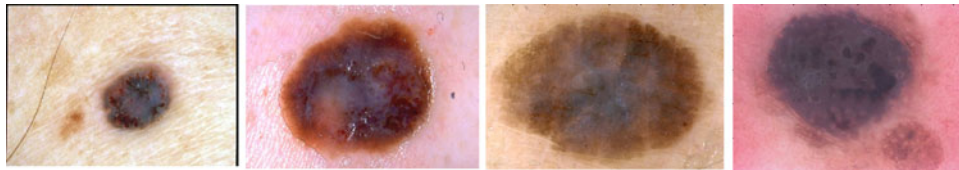
Fig. 1.   Dermoscopy images extracted from the heterogeneous EDRA database [22].

like EDRA [22] or when the images acquired at different locations are sent to an expert at a central hospital (teledermoscopy).

To tackle the aforementioned issues, Iyatomi *et al.* [9] proposed a calibration system that is software based. Their method performs fully automated color normalization using image content based on the HSV color space. A training step is required to build a set of normalization filters independently using the three channels of the HSV space. The filters obtained in the training step were tested on another dataset and the results showed that the color distribution of the new dataset was modified and made closer to the one of the training set. Although this method does not require hardware information, it still needs a training step. This step increases the conceptual and implementation complexities of the method. Furthermore, it has to be performed whenever the training set changes.

We are interested in exploring a different and somewhat simpler direction based only on image information, which does not require knowledge of the acquisition system properties or a training step. Color normalization is a very common problem in computer vision and image processing. Among the different strategies proposed to normalize image colors, there are approaches that try to account for the color of the light source, called color constancy algorithms [12], [25], [26]. An examination of Fig. 1 shows that part of the reason why the images look so different is the color of the light source (e.g., the fourth image was clearly acquired under a reddish light source). Thus, color constancy algorithms are suitable for the calibration of the color of dermoscopy images.

Some of the most popular color constancy algorithms require a training set acquired under a known light source and are, thus, difficult to implement (see [12] for a description and comparison of different methods). Applying these methods to the EDRA dataset would be impractical, since this would imply separating the images from different hospitals using their color content. An alternative are the much simpler statistics based algorithms. These algorithms use low-level image features, like mean value or the maximum response, to estimate the color of the light source. It has been demonstrated that with appropriate parameter values these methods achieve similar performance to the one of more complex methods [12], [15], [16]. Furthermore, they are simple to implement, fast, and only require the tuning of a few parameters. In this paper, we evaluate the performance of four algorithms: Gray World [13], max-RGB [14], Shades of Gray [15], and General Gray World [16].

III. COLOR CONSTANCY

A. Color Constancy Framework

The goal of color constancy is to transform the colors of an image I, acquired using an unknown light source, so that they appear identical to colors under a canonical light source [12], [25], [26]. Usually, it is assumed that this canonical light source is the perfect white light. Color transformation is accomplished in two separate steps. First, the color of the light source is estimated in the RGB color space $[e_R \ e_G \ e_B]^T$. Then, the image is transformed using the estimated illuminant.

Different algorithms can be used to estimate the color of the illuminant. In this paper, we compare four algorithms that use image statistics to estimate the color of the illuminant: Gray World [13], max-RGB [14], Shades of Gray [15], and General Gray World [16]. For a color image I, each component of the illuminant $e_c, c \in \{R, G, B\}$, is estimated based on the following expressions (for details refer to [12]):

1) Gray World

$$\frac{\int I_c(\mathbf{x})d\mathbf{x}}{\int d\mathbf{x}} = ke_c \tag{1}$$

2) max-RGB

$$\max_{\mathbf{x}} I_c(\mathbf{x}) = ke_c \tag{2}$$

3) Shades of Gray

$$\left(\frac{\int (I_c(\mathbf{x}))^p d\mathbf{x}}{\int d\mathbf{x}} \right)^{1/p} = ke_c \tag{3}$$

4) General Gray World

$$\left(\frac{\int (I_c^{\sigma}(\mathbf{x}))^p d\mathbf{x}}{\int d\mathbf{x}} \right)^{1/p} = ke_c \tag{4}$$

where I_c denotes the cth component of image I, $\mathbf{x} = (x, y)$ is the position of a pixel, and k is a normalization constant that ensures that $\mathbf{e} = [e_R \ e_G \ e_B]^T$ has unit length with respect to the Euclidean norm. Shades of Gray and General Gray World use the Minkowski norm to estimate the color of the illuminant. The parameter p in (3) and (4) is the degree of this norm. Finally, $I_c^{\sigma}(\mathbf{x})$ is a smoothed image, obtained by filtering $I(\mathbf{x})$ with a Gaussian low-pass filter with standard deviation σ. Both σ and p can be tuned according to the dataset.

An interesting aspect of the studied color constancy algorithms is that they are related to each other. Gray World and
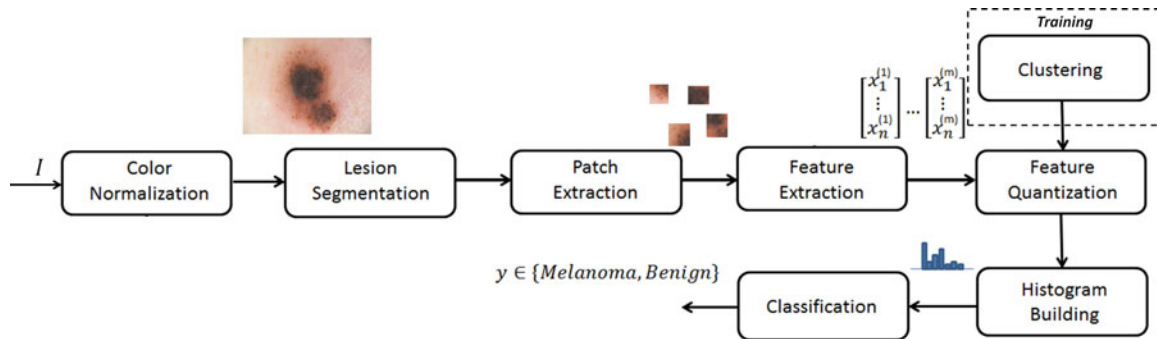
Fig. 2. Block diagram of the classification system.

max-RGB are special cases of Shades of Gray, for $p = 1$ and $p = \infty$, respectively. General Gray World is an extension of Shades of Gray where image noise is removed by low-pass filtering.

After estimating $\mathbf{e}$, we can transform the image I. A simple way to model this transformation is the von Kries diagonal model [27]

$$\begin{pmatrix} I_R^t \\ I_G^t \\ I_B^t \end{pmatrix} = \begin{pmatrix} d_R & 0 & 0 \\ 0 & d_G & 0 \\ 0 & 0 & d_B \end{pmatrix} \begin{pmatrix} I_R \\ I_G \\ I_B \end{pmatrix} \qquad (5)$$

where $[I_R, I_G, I_B]^T$ denotes the pixel value acquired under an unknown light source, and $[I_R^t, I_G^t, I_B^t]^T$ denotes the transformed pixel value, as it would appear under the canonical light source, which is assumed to be the perfect white light, i.e., $\mathbf{e}^w = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^T$. The matrix coefficients $\{d_R, d_G, d_B\}$ are related to the estimated illuminant $\mathbf{e}$ as follows:

$$d_c = \frac{1}{\sqrt{3}e_c}, \qquad c \in \{R, G, B\}. \qquad (6)$$

We have applied a gamma correction step, with γ set to the standard value of 2.2 [28], to all the images before performing color constancy.

IV. SYSTEM OVERVIEW

BoF [29] has been used with success in many computer vision applications. The main characteristic of this method is that it allows a local description of the image. This task is accomplished by sampling the image by small patches, which are separately characterized. In this section, we start by describing the general framework of a BoF model and then we discuss the application of this method in the context of dermoscopy images [6], [10].

A. BoF Framework

The BoF method aims to classify images using local features. Its operation comprises two phases: 1) model training and 2) classification [29]. In the training phase, the BoF model is learned from a training set of classified images. First, a set of local patches is detected in each image. Typically, a keypoint detector algorithm is used to extract a set of interest points associated with corners or blobs and a small patch is extracted

around each keypoint. Each patch is then characterized by a vector of features, e.g., color histogram or SIFT features [30].

It is not possible to use all the feature vectors to classify the image since the number of features would depend on the image. The BoF method solves this difficulty by defining a dictionary of *visual features*. The feature vectors of all training images are clustered into a set of groups (typically a few hundred) and a prototype (centroid) is extracted from each group. This operation is often accomplished by means of the k-means algorithm. The prototypes are called *visual words*. These *visual words* are then used to label each training patch. Each patch is associated with the closest word, i.e., the one that minimizes the Euclidean distance. It is then possible to characterize each image by counting the number of times each *visual word* is selected and compute a histogram with that information. The histogram is usually considered as a feature vector characterizing the image and it has the same size as the number of *visual words*. Finally, the histograms of the training images are used to learn a classification rule, using a supervised classification algorithm, e.g., k nearest neighbors and support vector machine (SVM).

During the test phase, each new image is sampled by small patches and these patches are characterized by local features as described earlier. Then, the extracted features are compared with the learned visual dictionary and a histogram is built for that image. Finally, the learned classification rule is applied to classify the histogram and label the image.

B. Classification of Dermoscopy Images

The classification framework used in this paper can be seen in Fig. 2. First, we start by separating each lesion from the healthy skin using manual segmentation performed by a dermatologist. This allows us to prevent classification errors associated with an incorrect segmentation. The influence of segmentation errors on the final decision has been studied in [31].

The BoF model is trained and tested as described above, using the following specifications. Each image is sampled using the Harris–Laplace keypoint detector [32]. This detector performed well in previous studies [6], [10], [11]. Then, we extract square patches around the detected keypoints and remove those that intersect the lesion in less than 50% of its area.

The description of the patches is directly related to the theme of this paper, since the color calibration step alters the color
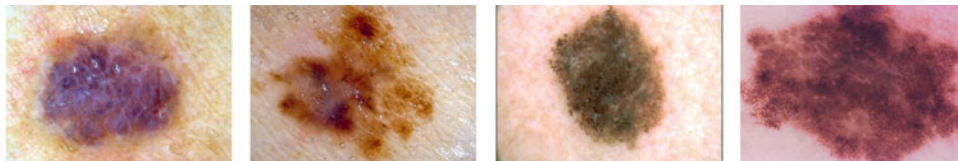
Fig. 3.   Dataset examples [22].

information of an image and, as a result, the values of the extracted features. To evaluate the influence of the color constancy algorithms, we extracted color histograms using the RGB color space. In previous studies, the BoF model with RGB histograms achieved good classification scores [6], [10] on the PH^2 dataset [33]. This suggests that, despite being perceptually nonuniform, RGB color space provides discriminative information. Furthermore, RGB is the default color space of the dermoscopy images and it is highly dependent on the color of the light source (recall Section III). Therefore, it is important to assess the performance of these histograms with and without the application of color constancy algorithms.

We use k-means to compute the dictionary of *visual words*. The histogram of visual words is classified using an SVM classifier with the χ^2 kernel

$$K_{\chi^2}(\mathbf{x}, \mathbf{y}) = e^{-\rho d_{\chi^2}(\mathbf{x}, \mathbf{y})} \tag{7}$$

where $\mathbf{x}$ and $\mathbf{y}$ are histograms of *visual words*, ρ is a width parameter, and

$$d_{\chi^2}(\mathbf{x}, \mathbf{y}) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}. \tag{8}$$

V. EXPERIMENTAL RESULTS

A. Dataset and Evaluation Metrics

In this paper, we used a heterogeneous set of 482 dermoscopy images (50% melanomas) selected from the EDRA database [22] as follows. First, we selected most of the available melanomas. Then, we randomly selected the same number of melanocytic benign lesions from the following categories: Blue nevi, Clark nevi, Spitz nevi, Combined nevi, and Dermal nevi. These images were collected from three different university hospitals: University Federico II of Naples (Italy), University of Graz (Austria), and University of Florence (Italy).

Fig. 3 shows some examples of the images used. These examples clearly show that there is a high variability among images. The camera used to acquire them is not the same and the light source under which they were acquired is significantly different.

We trained five different BoF systems. The first was trained using nonnormalized images and the remaining four were trained using one of the four color constancy methods studied in this paper. Each of the five systems was optimized in order to achieve the best possible results. This means that for each case, we varied a set of parameters. The common parameters optimized for all systems were the number of bins of the RGB histogram $\{5, 15, 25\}$, the number of centroids $\{25, 50, \dots, 300\}$,

TABLE I
MEAN AND STANDARD DEVIATION OF ILLUMINANT COLOR ESTIMATES AND
THE ASSOCIATED CPU TIMES

	e_R	e_G	e_B	CPU time
Gray World (1)	0.703±0.066	0.537±0.057	0.453±0.064	0.166 s
max RGB (2)	0.586±0.031	0.571±0.025	0.573±0.016	0.157 s
Shades of Gray (3)	0.647±0.048	0.556±0.042	0.516±0.041	0.196 s
General Gray World (4)	0.683±0.058	0.545±0.051	0.476±0.056	0.228 s

and the parameters of the SVM classifier: the cost C given to the soft margin $C \in \{2^{-5}, 2^{-4}, \dots, 2^5\}$ and the width of the kernel in (7) $\rho \in \{2^{-6}, 2^{-5}, \dots, 2^6\}$. Furthermore, we optimized specific parameters of the Shades of Gray (3) and General Gray World (4) algorithms, namely the value of $p \in \{1, 2, \dots, 10\}$ and $\sigma \in \{1, 2, \dots, 5\}$ as suggested in [16].

To evaluate the performance of each system, we computed three metrics: Sensitivity (SE), specificity (SP), and accuracy (ACC). SE corresponds to the percentage of melanomas that are correctly classified, SP is the percentage of correctly classified benign lesions, and AAC is defined as follows:

$$\text{ACC} = \frac{\text{SE} + \text{SP}}{2}. \tag{9}$$

All metrics were computed using a stratified tenfold cross validation strategy, in which the dataset is divided into ten subsets, each with approximately the same number of melanomas and benign lesions. Nine folds were used for training of the classifier and the remaining one is used for testing. The results presented for each system are the average performances of ten testing folds.

Using *k*-fold cross validation to optimize the hyperparameters of a system works well in practice, but may lead to optimistic statistics [34]. To compensate, we apply the methodology proposed by Tibshirani and Tibshirani [35], and estimate the $\widehat{\text{bias}}$ associated with the computed ACC. The adjusted $\widehat{\text{ACC}}$ can then be defined as

$$\widehat{\text{ACC}} = ACC + \widehat{\text{bias}}. \tag{10}$$

In our study, the $\widehat{\text{bias}}$ associated with each system is in the interval $[-5.9\%, -4.2\%]$.

B. Results

We estimated the color of the illuminant for the 482 EDRA images using the four color constancy algorithms (see Section III). The average values and standard deviation of the estimates for all the images in the dataset can be seen in Table I. It is interesting to note that the Gray World (1),
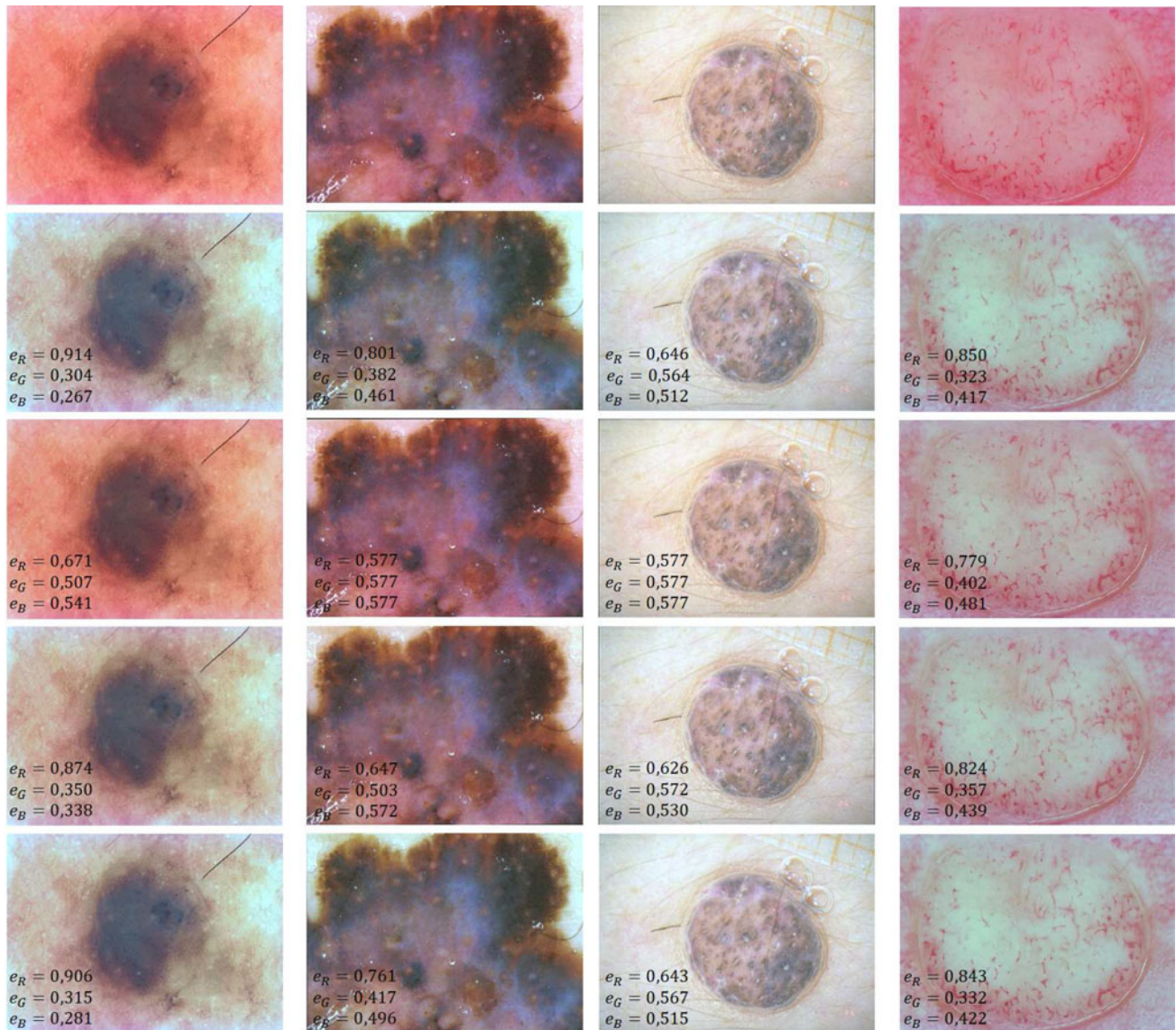
Fig. 4. Color constancy examples. From top to bottom: Original Image, Gray World, max-RGB, Shades of Gray ($p = 6$), and General Gray World ($p = 2$, $\sigma = 3$).

Shades of Gray (3), and General Gray World (4) algorithms consider that most of the images need correction, since there is a significant deviation between the estimated light and the values for the white light, where $e_R = e_G = e_B = 1/\sqrt{3} \simeq 0.577$. On the other hand, according to the max-RGB algorithm, there are fewer images that require normalization because the values are closer to those of the white light. This might be explained by that fact that while the other algorithms use all the pixels in the image to compute the value of the illuminant in each color channel, max-RGB uses only the pixel with the highest value in each color channel. The last column of Table I gives the average CPU time for each color constancy algorithm on the EDRA dataset with average image size of 510×765 pixels (programming environment: MATLAB 2013a, CPU: Intel Core i7-2600).

Some examples of normalized images as well as the estimated values of the corresponding light sources are shown in Fig. 4. These results clearly show that the color constancy algorithms alter the appearance of the images, making them look more similar. This is more noticeable in images illuminated by a reddish light source (see Fig. 4, first and fourth columns), where the colors of the lesions and surrounding skin become much more distinguishable. Furthermore, color constancy also seems to enhance the contrast inside the lesion (see Fig. 4, second column) and between the lesion and the surrounding skin, as can be seen in Fig. 4 first column. This last image is very interesting because without the color constancy it would be almost impossible to notice the light brown area that surrounds the darker center of the lesion, which might lead to an incorrect segmentation and consequent loss of color information. Separately analyzing each of the color constancy algorithms, it is possible to notice that Gray World (second row) and General Gray World (fifth row, $p = 2$, $\sigma = 3$) are the ones that most alter the images, giving them a grayish color.

TABLE II
CLASSIFICATION RESULTS WITH AND WITHOUT COLOR CONSTANCY

Algorithm	SE	SP	ACC	$\widehat{ACC}$
None	71.0%	55.2%	63.1%	58.8%
Gray World	78.8%	75.2%	77.0%	72.7%
max-RGB	79.2%	75.5%	77.3%	73.1%
Shades of Gray	**79.7%**	**76.0%**	**77.8%**	**73.4%**
General Gray World	79.6%	75.6%	77.6%	71.7%

TABLE III
CLASSIFICATION RESULTS FOR THE PH^2 (SINGLE SOURCE) DATASET WITH AND WITHOUT COLOR CONSTANCY

COLOR CONSTANCY	SE	SP	ACC	$\widehat{ACC}$
None	92.5%	75.6%	84.1%	79.1%
Shades of Gray	92.5%	76.3%	84.3%	79.3%

TABLE IV
CLASSIFICATION RESULTS FOR SIFT FEATURES WITH AND WITHOUT COLOR CONSTANCY

Algorithm	SE	SP	ACC	$\widehat{ACC}$
None	78.3%	63.9%	71.1%	66.9%
Gray World	80.4%	62.7%	71.6%	67.2%
max-RGB	78.8%	64.8%	71.8%	67.3%
Shades of Gray	**80.2%**	**65.6%**	**72.9%**	**68.7%**
General Gray World	75.9%	67.3%	71.6%	66.5%

Shades of Gray (fourth row, $p = 6$) mitigates this effect giving the images a more normal coloration. Max-RGB seems to be the algorithm that least alters the aspect of the lesions. This was already observed in Table I. Recall that this algorithm is the limit of Shades of Gray for $p = \infty$, which explains its performance.

Table II shows the performance of the classification systems trained using the original images and the color corrected ones. By inspecting this table, it can be noticed that all four color correction algorithms significantly improve the performance of the classification system, Shades of Gray being slightly better than the others. This, combined with the information provided by Fig. 4, suggests that this algorithm performs slightly better than the other three in this experiment.

We also applied color constancy to the PH^2 database with 200 images (40 melanomas and 160 nonmelanomas). With this new dataset, we are able to evaluate the influence of color constancy on a single-source dataset. The results are shown in Table III. These results were obtained using Shades of Gray, with $p = 6$. It can be seen that color constancy does not alter the performance of the classifier when the images come from a single source (there is even a marginal improvement in SP).

Studies have shown that color constancy can also be applied before converting RGB images to other color spaces such as HSV [36]. We tested this hypothesis and trained two classification systems using 1-D HSV color histograms. In the first case, we computed the 1-D HSV histograms without prior color constancy and the results were the following: SE = 73.8%, SP = 76.8%, and ACC = 75.3%. The second system was trained using HSV histograms of normalized images using Shades of Gray with $p = 6$. The classification results were better: SE = 73.9%, SP = 80.1%, and ACC = 77.0%, which means that color constancy can also be used to increase the performance of classifiers based on the HSV.

Texture features can also be used to characterize the patches in a BoF model. In previous studies, various texture features have been investigated in the context of dermoscopy image classification [6], [10], [37]. To determine if color constancy can be used with features besides color related ones, we applied color constancy to systems trained with SIFT features [30]. These features have already been used in the classification of dermoscopy images [37]. The results can be seen in Table IV. As expected, color constancy does not significantly influence the performance of texture features. Nonetheless, there is a marginal improvement, more pronounced in the case of Shades of Gray.

VI. CONCLUSION

In this paper, we investigated four color constancy algorithms and their influence on a BoF system. The results were obtained using a heterogeneous dataset, acquired at three different hospitals [22]. The performance of the system was significantly improved from SE = 71% and SP = 55.2% to SE = 79.7% and SP =76.8%, using the Shades of Gray algorithm. This algorithm was also the one that led to the more realistic colors. Preliminary tests also showed that color constancy algorithms can be used to improve color features derived from other color spaces such as HSV. Furthermore, we have showed that color constancy can be used with texture features and be applied to single-source datasets.

REFERENCES

[1] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, Mar. 2001.
[2] P. Rubegni, G. Cevenini, M. Burroni, R. Perotti, G. Dell'Eva, P. Sbano, C. Miracco, P. Luzi, P. Tosi, P. Barbini, and L. Andreassi, "Automated diagnosis of pigmented skin lesions," *Int. J. Cancer*, vol. 101, no. 6, pp. 576–580, 2002.
[3] M. E. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, and R. Moss, "A methodological approach to the classification of dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 31, pp. 362–373, 2007.
[4] H. Iyatomi, H. Oka, M. E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Comput. Med. Imag. Graph.*, vol. 32, no. 7, pp. 566–579, 2008.
[5] Q. Abbas, M. Celebi, C. Serrano, I. Fondón García, and G. Ma, "Pattern classification of dermoscopy images: A perceptually uniform model," *Pattern Recog.*, vol. 46, no. 1, pp. 86–97, 2013.

[6] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," to be published in IEEE Systems Journal, 2014.

[7] W. Stolz, A. Riemann, and A. B. Cognetta, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma," *Eur. J. Dermatol.*, vol. 4, pp. 521–527, 1994.

[8] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and E. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Arch. Dermatol.*, vol. 134, pp. 1563–1570, 1998.

[9] H. Iyatomi, M. E. Celebi, G. Schaefer, and M. Tanaka, "Automated color calibration method for dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 35, pp. 89–98, 2011.

[10] C. Barata, M. Ruela, T. Mendonça, and J. S. Marques, "A bag-of-features approach for the classification of melanomas in dermoscopy images: The role of color and texture descriptors," in *Computer Vision Techniques for the Diagnosis of Skin Cancer*, J. Scharcanski and M. E. Celebi, Eds. New York, NY, USA: Springer-Verlag, 2014, pp. 49–69.

[11] C. Barata, J. S. Marques, and J. Rozeira, "The role of keypoint sampling on the classification of melanomas in dermoscopy images using bag-of-features," in *Proc. Iberian Conf. Pattern Recog. Image Anal.*, 2013, pp. 715–723.

[12] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.

[13] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 210, pp. 1–26, 1980.

[14] E. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, pp. 108–128, 1977.

[15] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. 12th Color Imag. Conf.: Color Sci. Eng. Syst., Technol., Appl.*, 2004, pp. 37–41.

[16] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.

[17] Y. V. Haeghen, J. M. A. D. Naeyaert, and I. Lemahieu, "An imaging system with calibrated color image acquisition for use in dermatology," *IEEE Trans. Med. Imag.*, vol. 19, no. 7, pp. 722–730, Jul. 2000.

[18] C. Grana, G. Pellacani, and S. Seidanari, "Pratical color calibration for dermoscopy applied to a digital epiluminescence microscope," *Skin Res. Technol.*, vol. 11, pp. 242–247, 2005.

[19] P. Wighton, T. K. Lee, H. Lui, D. McLean, and M. S. Atkins, "Chromatic aberration correction: An enhancement to the calibration of low-cost digital dermoscopes," *Skin Res. Technol.*, vol. 17, pp. 339–347, 2011.

[20] J. Quintana, R. Garcia, and L. Neumann, "A novel method for color correction in epiluminescence microscopy," *Comput. Med. Imag. Graph.*, vol. 35, pp. 646–652, 2011.

[21] G. Schaefer, M. I. Rajab, M. E. Celebi, and H. Iyatomi, "Colour and contrast enhancement for improved skin lesion segmentation." *Comput. Med. Imag. Graph.*, vol. 35, pp. 99–104, 2011.

[22] G. Argenziano, H. P. Soyer, V. De Giorgi, D. Piccolo, P. Carli, M. Delfino, A. Ferrari, V. Hofmann-Wellenhog, D. Massi, G. Mazzocchetti, M. Scalvenzi, and I. H. Wolf. (2000). *Interactive Atlas of Dermoscopy*. [Online]. Available: http://www.dermoscopy.org/atlas/

[23] A. Madooei, M. S. Drew, M. Sadeghi, and M. S. Atkins, "Automated preprocessing method for dermoscopic images and its application to pigmented skin lesion segmentation," in *Proc. Color Imag. Conf.*, 2012, Vol. 2012, no. 1, pp. 158–163.

[24] R. Amelard, J. Glaister, A. Wong, and D. A. Clausi, "Melanoma decision support using lighting-corrected intuitive feature models," in *Computer Vision Techniques for the Diagnosis of Skin Cancer*, J. Scharcanski and M. E. Celebi, Eds. New York, NY, USA: Springer-Verlag, 2014, pp. 193–219.

[25] S. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.

[26] G. Klinker, S. Shafer, and T. Kanade, "A physical approach to color image understanding." *Int. J. Comput. Vis.*, vol. 4, no. 1, pp. 7–38, 1990.

[27] J. von Kries, "Influence of adaptation on the effects produced by luminous stimuli," *Sources of Color Vision*. Cambridge MA, USA: MIT Press, pp. 109–119, 1970.

[28] C. Poynton, *Digital Video and HD: Algorithms and Interfaces*. San Mateo, CA, USA: Morgan Kaufman, 2012.

[29] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp.1470–1477.

[30] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[31] C. Barata, J. S. Marques, and M. E. Celebi, "Towards an automatic bag-of-features model for the classification of dermoscopy images: The influence of segmentation," in *Proc. 8th Int. Symp. Image Signal Process. Anal.*, 2013, pp. 274–279.

[32] K. Mikolajczyk, and C. Schmid., "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.

[33] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "Ph2–A dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*, 2013, pp. 5437–5440.

[34] M. Stone, "Asymptotics for and against cross-validation," *Biometrika*, vol. 64, no. 1, pp. 29–35, 1977.

[35] R. J. Tibshirani and R. Tibshirani, "A bias correction for the minimum error rate in cross-validation," *Ann. Appl. Stat.*, 2009, pp. 822–829.

[36] J. van de Weijer and C. Schmid, "Coloring local feature extraction," in *Proc. ECCV*, 2006, pp. 246–268.

[37] C. Barata, J. S. Marques, and J. Rozeira, "Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model," *Adv. Vis. Comput.*, 2013, pp. 40–49.

Catarina Barata received the B.Sc. and M.Sc. degrees in biomedical engineering from the Technical University of Lisbon, Lisbon, Portugal, in 2009 and 2011, respectively. She is currently working toward the Ph.D. degree at the Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon.

Her research interests include image processing, pattern recognition, and dermoscopy.

M. Emre Celebi (S'05–M'07–SM'11) received the B.Sc. degree in computer engineering from the 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, Arlington, TX, USA, in 2003 and 2006, respectively.

He is currently an Associate Professor with the Department of Computer Science, Louisiana State University in Shreveport, Shreveport, LA, USA. He has pursued research in the field of image processing and analysis. He has published more than 120 articles in journals and conference proceedings. His recent research is funded by grants from the National Science Foundation.

Jorge S. Marques received the E.E., Ph.D., and Aggregation degrees from the Technical University of Lisbon, Lisbon, Portugal, in 1981, 1990, and 2002, respectively.

He is currently an Associate Professor with the Electrical and Computer Engineering Department, Instituto Superior Técnico, Lisbon, and a Researcher at the Institute for Systems and Robotics. His research interests include the areas of image processing, shape analysis, and pattern recognition.

Dr. Marques was the Cochairman of the IAPR Conference IbPRIA 2005, President of the Portuguese Association for Pattern Recognition (2001–2003), and an Associate Editor of *Statistics and Computing* (Springer).