Colorimetric calibration of wound photography with off-the-shelf devices

Subhankar Bala, Ekaterina Sirazitdinova, Thomas M. Deserno

Subhankar Bala, Ekaterina Sirazitdinova, Thomas M. Deserno, "Colorimetric calibration of wound photography with off-the-shelf devices," Proc. SPIE 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment, 1013617 (10 March 2017); doi: 10.1117/12.2254721

SPIE.

Event: SPIE Medical Imaging, 2017, Orlando, Florida, United States

Colorimetric Calibration of Wound Photography with Off-the-shelf Devices

Subhankar Bala[a], Ekaterina Sirazitdinova[b], and Thomas M. Deserno[b,*]

[a]National Institute of Technology, Warangal, India
[b]Department of Medical Informatics, Uniklinik RWTH Aachen, Germany

ABSTRACT

Digital cameras are often used in recent days for photographic documentation in medical sciences. However, color reproducibility of same objects suffers from different illuminations and lighting conditions. This variation in color representation is problematic when the images are used for segmentation and measurements based on color thresholds. In this paper, motivated by photographic follow-up of chronic wounds, we assess the impact of (i) gamma correction, (ii) white balancing, (iii) background unification, and (iv) reference card-based color correction. Automatic gamma correction and white balancing are applied to support the calibration procedure, where gamma correction is a nonlinear color transform. For unevenly illuminated images, non-uniform illumination correction is applied. In the last step, we apply colorimetric calibration using a reference color card of 24 patches with known colors. A lattice detection algorithm is used for locating the card. The least squares algorithm is applied for affine color calibration in the RGB model. We have tested the algorithm on images with seven different types of illumination: with and without flash using three different off-the-shelf cameras including smartphones. We analyzed the spread of resulting color value of selected color patch before and after applying the calibration. Additionally, we checked the individual contribution of different steps of the whole calibration process. Using all steps, we were able to achieve a maximum of 81% reduction in standard deviation of color patch values in resulting images comparing to the original images. That supports manual as well as automatic quantitative wound assessments with off-the-shelf devices.

Keywords: Color card, image acquisition, mean square algorithm, photographic documentation, color calibration,color correction, lattice detection, chronic wounds, smartphone

1. INTRODUCTION

A wound is considered as chronic if it is not healing in a predictive duration of time (usually 3 month) with known order of sequential stages.[1] In the United States, an estimate of 6.5 million patients are effected by chronic wounds, and more than 25 billion US dollar along with numerous hours of vigilance and treatment is spent annually for patient care related to chronic wounds.[1] This emphasises the need for a modern software system for automatic analysis of wound images and characterization of wound tissue composition, measurement of wound size and the healing rate of different type of wounds for incoming and outgoing patients.

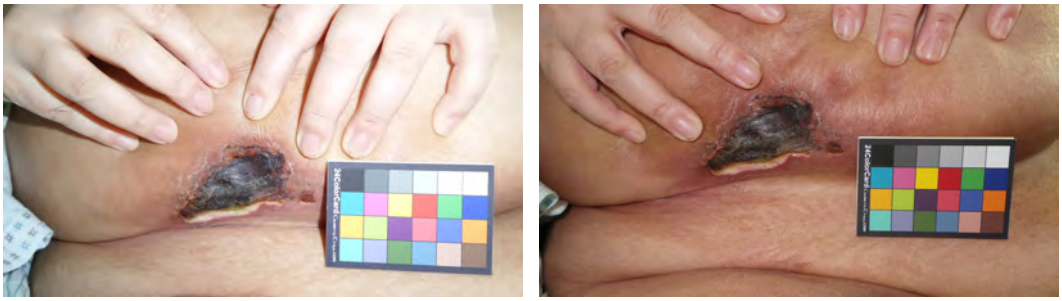
In this work, we propose documentation of wounds with off-the-shelf devices, which are quite cheap and common in the modern age. Using this technology patients do not need to visit a clinic on a daily basis for check up, as they can get the status of wounds from software, which can suggest to visit a clinic if required. This would also help in providing continuous documentation of such wounds and might be beneficial in further research related to chronic wounds.

In recent era, digital photography is used frequently for documentation of many diseases in the curative as well as research-oriented medical context. In spite of availability of qualitatively good digital cameras in the modern smartphones, reproducibility of colors in images is still poor: captured in changing illumination, same

Further author information: (Send correspondence to T.M.D.)
S.B.: E-mail: bsubhankar@student.nitw.ac.in, Telephone: +91 8142468890
E.S.: E-mail: ekaterina.sirazitdinova@rwth-aachen.de, Telephone: +49 241 80-89797
T.M.D.: E-mail: deserno@ieee.org, Telephone: +49 241 80-88793

Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment,
edited by Matthew A. Kupinski, Robert M. Nishikawa, Proc. of SPIE Vol. 10136, 1013617
 · doi: 10.1117/12.2254721

objects appear diversely in terms of RGB values (Fig. 1). Even RGB values deviate a lot for different settings of one camera and for different cameras due to their exposure time and sensitivity of the camera sensor. Each imaging system has its own device-dependent and time-varying RGB color space. If such images are used for comparison and follow-up assessment of healing and other processes, the methods fail. In case of chronic wound documentation, for understanding gradual improvement or degradation of a wound, it is necessary to compare images being in the same colorimetric standard. In this work, driven by the goal of quantitative documentation of human wounds, we combined different methods from independent authors[2,3] into a robust color correction pipeline and evaluated it with respect to its separate steps to estimate the importance and the effect of different operations in the calibration process.



(a) Wound image with flash. (b) Wound image without flash.

Figure 1: Same wound captured in different lighting conditions.

2. STATE-OF-THE-ART

There have been several studies in the field of color calibration for reconstruction of original color for different applications ranging from dermatology to chronic wound studies, pathology, and forensics. In the dermatological study of Haeghen et al.,[4] the authors have determined the calibration matrix by performing several consecutive steps. By sequential order, these are: (i) determination of the camera offset, (ii) frame grabber offset, (iii) frame grabber gain, (iv) camera aperture, (v) color gains of the camera, (vi) linearizing look-up table, and, finally, (vii) transform from the unknown device-dependent imaging system space to the standard device-dependent space. The authors have proposed a colorimetric consistent image acquisition system, where they calibrate the system and store the resulting setting in a profile. Determining a profile takes no more than 5 minutes, and that profile remains valid for weeks.

Hsu et al.[5] have presented an approach towards colorimetric correction of face images. The proposed algorithm is based on localization of top 5% illuminated pixels in the image. Those pixels are assumed to be 'white'. Based on the chromatic distance between the selected pixel color and ideal 'white' color the entire picture was corrected. The major limitation is that the method works for images with no specular reflection. However in non-controlled environment where specular reflection occurs frequently this technique fails.

Li et al.[6] have suggested a color calibration technique for colposcopic images, which is used to detect cancer and precancerous lesions of the uterine cervix. A calibration unit was mounted on the colposcope and calibration was applied prior to actual experiment. A gray balancing algorithm was applied first on the reference color chart and then a 3×4 color calibration matrix was calculated by least square approach.

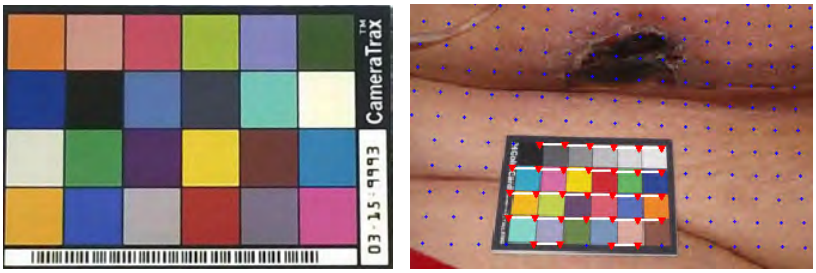
In another study, Poucke et al.[3] have transformed the color space from unknown device-dependent red-green-blue (RGB) to standard RGB (sRGB), which has a known relationship to the color-opponent space with dimensions L for lightness and a and b for the color-opponent dimensions, also known as CIELab colorimetric space. The CIELab allows computing of perceptual colour differences. The images run through inverse gamma transform, then a three times 3D polynomial transform is followed by gamma correction converting RGB to sRGB. Perceptual color differences between all possible pairs are measured for all patches before and after applying the calibration procedure and a significant improvement has been observed in the study.

Varghese et al.[7] have proposed a colorimetric calibration algorithm of high-dynamic-range (HDR) images, which are captured using a multi-exposure technique. They have calculated the color transform matrix for each of three cameras (Canon 550D, Canon 1000D, Panasonic Lumix DMC-LX7) and used those matrices to calibrate the image in two test scenes (conference room, library). They found that both of their models (model A and model B) result in a very similar calibration error. However model B gives worse prediction for scenes 'conference room' and 'library'. Their result shows that when raw images are merged into HDR images, colors can be calibrated with higher accuracy.

These examples depict the diversity of approaches. Nonetheless, the image processing chain is always similar, i.e. composed of similar steps. In this paper, we aim at standardizing the chain of processing for color normalization, and evaluating each module individually with respect to its contribution and importance for the overall results.

3. METHODS

In several studies, color calibration is performed using a reference color card with known RGB values in ideal condition. In our studies, the arrangement of color patches has been customized for a robuster automatic card detection (Fig. 2a).[2,8] Therefore, we consider reference card-based illumination corrections.



(a) Custom color arrangement.

(b) Detected corners of color patches in the original card.

Figure 2: Macbeth color card.

Recently, Jose et al.[8] have applied a lattice detection algorithm to localize a color reference card in the image. The authors use mean shift belief propagation to gradually extract the grid structure of the card to find its placement in the image. Knowing the corner coordinates of each color patch, an average color value of each patch can be computed (Fig. 2b). Note that the detected coordinates also support to correct for perspective position (geometry). Then, the colors of the card tiles are measured computing a mean within the tile.

Although none of the above-mentioned approaches actually involves non-uniform illumination correction, we have noticed already that such conditions can result in different RGB values within a single patch and may cause serious degradation in the overall performance.

3.1 Image processing chain

Combining the steps from the observed approaches with non-uniform illumination correction, we propose the following image processing chain for colorimetric calibration (Fig. 3). Composed of 5 modules, it reconstructs the original color of the scene in images with different types of colorimetric distortion.

Inverse gamma correction Gamma correction defines the relationship between the pixels numerical value and its actual luminance. Without gamma correction, shades captured by digital cameras appear different to our eyes. Cameras follow a linear relationship between actual luminance and detected light but our eyes follow a nonlinear curve. Gamma correction[9] translates between our eyes light sensitivity and that of the camera. When a digital image is saved, it is gamma encoded. Therefore, before doing the calibration operation, reverse gamma

Figure 3: General pipeline of our color correction method.

correction is needed to retrieve the original RGB values that were stored during image capturing. The output of the inverse gamma correction can be given as

$$V_{\text{out}} = AV_{\text{in}}^{1/\gamma} \tag{1}$$

where V_{out} is the output value of pixel, V_{in} is input value of pixel, A is a constant that can be defined as gain, and γ is the gamma coefficient of the system which is different for different display units. For example, a CRT screen has a gamma value within a range of 2.35 to 2.55 whereas standard LCD monitors has a gamma value of 2.2. Recently 2.2 is treated as standard display gamma for Gamma correction. We used gamma value of 2.2 in our study.

Non-uniform illumination correction A non-uniform illumination correction algorithm[10] is applied to the images with vignetting and shadow artifacts. We extract the illumination information by passing the image via a Gaussian kernel with a high value of σ ($\sigma = 300$). Then, the inverse of the illumination information is added with a proper weight to the original image to balance the intensity and making the whole image evenly illuminated.

White balancing For achieving white balance, we equalize the mean of all three of RGB channel to the same with the mean of the gray scale image of the original RGB image. So, in all three color planes, we multiply each pixel with the mean value of grayscale image divided by the mean of respective plane to equalize the mean. For example, in the red channel, each pixel is multiplied by a value which is mean of grayscale image divided by the mean of red plane image.

Card-based calibration In an over determined system where there are more equations than unknown parameters, a clear solution does not exist. Therefore, we applied the least squares calibration[11] in order to minimize the error term which is defined by squared differences between output matrix and multiplication of transform with input matrices.

Gamma correction To save and display the image after all procedures, we need to perform gamma correction depending on the saving format and the type of screen for display. We apply gamma correction before sending the image to the screen for display. Let V'_{in} be the current pixel value, the result of gamma correction V'_{out} is then defined as

$$V'_{\text{out}} = AV'^{\gamma}_{\text{in}} \tag{2}$$

3.2 Evaluation

As mentioned before, different authors propose different calibration chains. Therefore, we want to assess the impact of the modules individually.

Goal In particular, we aim at answering the following questions for the particular application of mobile monitoring of chronic wounds:

1. What is the optimal pipeline for color and which modules shall be used to compose it?
2. What is the impact of camera hardware, in particular with respect to low-cost off-the-shelf devices?
3. Shall such images be taken with or without flash lights?

General design Ideally, the experimental setting allows control of all parameters such that the de-facto conditions are a-priory known. In human wounds, however, we even do not know the actual color of tissue. Therefore, a quantitative evaluation based on absolute values is not possible and relative values are considered instead of. The general assumption is that a particular color, captured in different illumination setting and with different hardware stays the same.

Ground truth and metrics Therefore, if several shots of the same probe are obtained, effective color calibration will minimize the variance or standard deviation of the RGB values. Based on that idea, we can generate colored test pattern without controlling their particular color.

Figure 4 visualizes the color test pattern to mimic different tissue types: skin, granulation, necrosis, and slough. Each tissue class is represented by four slightly different color tones. The reference card is placed in the middle of the scene and exposed it to different lighting conditions: indoor neon light, outdoor sunny and cloudy conditions, as well as blue, red, green and yellow projected light with and without using a flashlight.

(a) Human skin color. (b) Granulation. (c) Necrosis. (d) Slough.

Figure 4: Color chart containing templates of 4 color variations.

To evaluate the pipeline and its components, we have exposed the sample color patch to different extreme and moderate type environmental conditions to get a maximum deviation of the same patch RGB value due to different lighting. Then we applied our color calibration processing chain to all the images to see whether our calibration procedure can take all differently deviated colors to a single original value or not. Ideally, all pixels in a patch should result in a single RGB value, which is the true color of the patch.

3.3 Imaging hardware

For image recording, three different cameras were used:

- The *Microsoft Lumina 535* can be considered as regular off-the-shelf smartphone with no particular camera features. It provides a regular 5 MP camera and led-based flash. The images have resolution of 2592×1456 pixels.
- The *Samsung Galaxy K Zoom* provides a CMOS-based 20.7 MP camera and optical 10-fold zoom lenses. While still quite inexpensive, it can be cosidered as a smartphone with particular design for high-quality images.
- The *Sony Alpha 580* is a DSLR camera with APS-C type Exmor CMOS sensor. In our experiments, the camera was used with a Tamron zoom lens 18-270 mm. Using this device, all images were acquired with 18 mm focal length.

Image Capturing To test the efficiency of our calibration chain we manually obtained different illumination conditions using colored projection in the room. The following illumination conditions were selected: 'red', 'green', 'blue', 'yellow', 'pink', 'indoor neon', 'sunny daylight'. The images were taken using a device's built-in flash and without a flash. 42 images were acquired in total. In each image, there are four different types of tissue featuring four different tones each. Thus, in total, 672 color patches were available.

Evaluation pipeline We applied our calibration algorithm on 42 images exposed to different lighting conditions and measured the average RGB value of each of 16 color templates in the images taken before and after the calibration process. For every test case, we measured standard deviation in each color channel of each color template. The standard deviation in three dimensional plane is defined as square root of the sum of squares of standard deviation values in each of R, G and B planes. We calculate mean standard deviation of each RGB plane for each patch. We observed the resulting standard deviation values of one representative patch of each type of color template (human skin color and three types of wound tissue: necrosis, granulation and slough) considering the following settings:

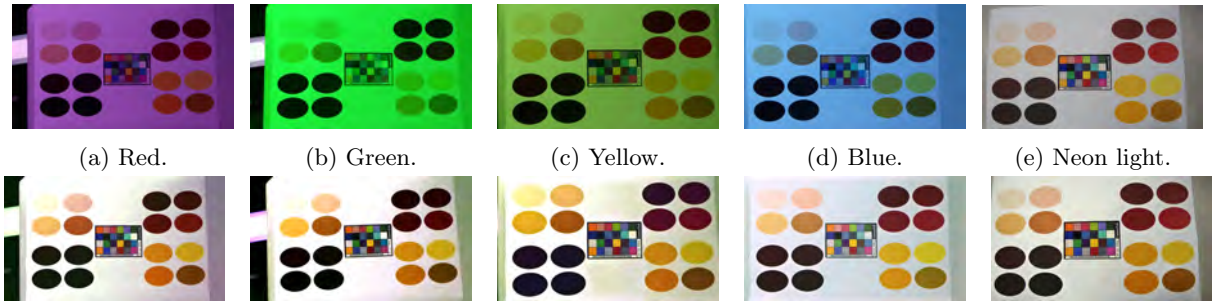
- *Original:* non-calibrated Image
- *ABCD:* image adjustment using all procedures
- *_BCD:* gamma correction is excluded from the image adjustment pipeline
- *A_CD:* white balance is excluded from the image adjustment pipeline
- *AB_D:* non-uniform illumination correction is excluded from the image adjustment pipeline
- *ABC_:* color calibration is excluded from the image adjustment pipeline

When naming different settings, by A we refer to the combination of inverse gamma and gamma corrections, because these two procedures are always evaluated together.

4. RESULT

The resulting standard deviation values are presented in the Table 1. In our initial recordings, color of each particular patch resulted in a large deviation of numerical values depending on different lighting conditions. In the original image, the standard deviation has varied from 49.6 to 80.95 for different tissues in three-dimensional RGB plane. Choosing the *ABCD* setting, which includes all of the steps in calibration pipeline, provides the best concentrated RGB values for all types of illumination resulting in the range of standard deviations from 14.7 to 19.8, which means that the chosen calibration process helps to restrict color of a patch to a shorter range for up to almost 78% in the worst case even when it is illuminated differently. The outcomes of our calibration pipeline (the setting *ABCD*) for various initial illumination conditions are shown in the Fig. 5. We can see that all resulting images are visually alike, which confirms that the chosen pipeline is able to reduce the initial spread of color values from different extremes to small deviations of original color. The importance of each step is observed through the changes in standard deviation of a patch by removing that particular step from total calibration process. Removal of the mean square calibration step causes the highest increment in the standard deviation and removal of illumination correction causes the lowest of standard deviation from the total calibration process.

Additionally, we have analyzed the effect of applying a flash in wound images and compared the result acquired from the images without flash. Due to the flash, unintentional reflections occur at the wound and at other glossy surfaces. As medical images are mostly taken from close distance, such unintentional reflections may result in saturation of colour in some particular regions, and in such cases, the original color cannot be recovered. We observed that for this application images shall be taken without flash.



(a) Red. (b) Green. (c) Yellow. (d) Blue. (e) Neon light.

(f) Red result. (g) Green result. (h) Yellow result. (i) Blue result. (j) Neon result.

Figure 5: Different illumination conditions (top row) and corresponding results (bottom row).

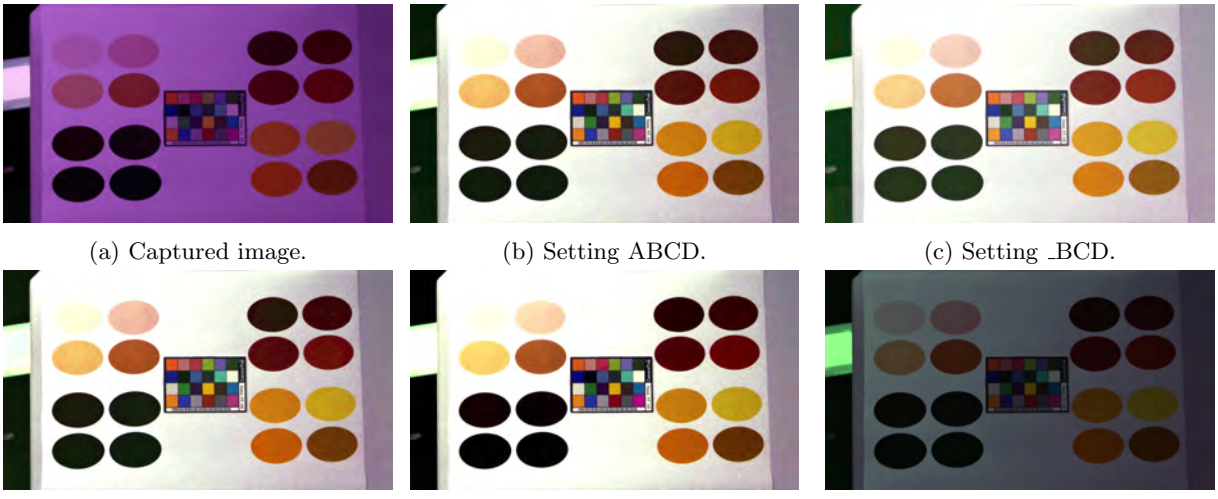
Table 1: Standard deviation of each color tile depending on the combination of procedures applied

Tissue type	Original	ABCD	_BCD	A_CD	AB_D	ABC_
Normal skin	69.28	18.12	25.8	27.96	22.11	60.88
Granulation	49.6	19.8	24.18	26.03	23.5	30.43
Necrosis	20.83	15.16	25.41	15.18	15.37	11.0
Slough	80.95	14.7	22.2	20.55	16.01	50.15

5. DISCUSSION

From our experiments, we can conclude that the reference card-based mean square calibration contributes most in the color correction process. White balance and gamma correction have a lower impact. Fig. 6 illustrates one of the extreme test cases, when the captured scene was exposed to the intense red light. The resulting images show that, in some combinations of algorithmic settings, the colors became quite close to the original. In Fig. 7, the result of calibration is shown when exposed to indoor neon light, which might be considered as a condition when the colors are closest to the typical for stationary inpatients. Again the complete pipeline yields the best result.

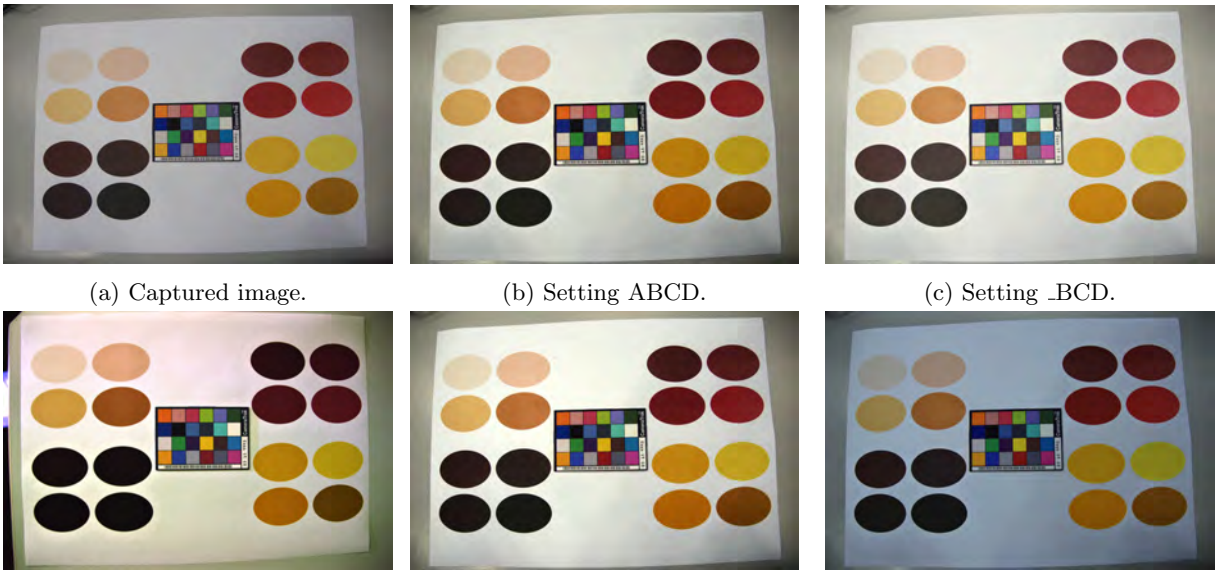
In the chosen calibration approach, a linear dependency between each of the three RGB planes of the image is assumed. It would be also beneficial to observe color correction in terms of other type of dependency between



(a) Captured image. (b) Setting ABCD. (c) Setting _BCD.

(d) Setting A_CD. (e) Setting AB_D. (f) Setting ABC_.

Figure 6: Calibration result of an extreme case using red illumination for different settings.



(a) Captured image. (b) Setting ABCD. (c) Setting _BCD.

(d) Setting A_CD. (e) Setting AB_D. (f) Setting ABC_.

Figure 7: Calibration result of a case of indoor neon light illumination for different settings.

the planes (e.g., square, logarithmic or exponential) which may help to study the nature of noise degrading the quality of the image. In addition, the evaluated approach is not robust towards non-uniform illumination in the cases if it is confined to a very small region or if some large portion of the image is darker than the shaded region. These issues will be addressed in the future work. Finding a more robust approach towards non-uniform illumination correction and consideration of human skin tones diversity could be another possible future task.

6. CONCLUSION

In our work, we have captured simulated wound colors of four variations and exposed to different regular as well as extreme lighting conditions with a reference color card in the scene of the image field. Then, we applied the lattice detection algorithm[8] to detect the card with reference colors and run the color calibration pipeline in order to bring possible wound tissue colors to ideally a single RGB value in all lighting condition. In this approach, we have managed to restrict the deviation of each simulated wound color patch to a low value after applying the proposed processing pipeline.

Considering the overall good performance of the composed approach even in extreme conditions, we can recommend the analysis of this method in clinical trial. It could be particularly helpful for the situations when off-the-shelf devices (like smartphones) are used for documentation.

REFERENCES

[1] Fauzi M, Khansa I, Catignani K, Gordillo G, Sen C, Gurcan M. Computerized segmentation and measurement of chronic wound images. Comput Biol Med. 2015;60:74–85.

[2] Deserno TM, Sárándi I, Jose A, Haak D, Jonas S, Specht P, et al. Towards quantitative assessment of calciphylaxis. In: Proc SPIE Med Imag. vol. 9035. International Society for Optics and Photonics; 2014. p. 90353C.

[3] Van Poucke S, Vander Haeghen Y, Vissers K, Meert T, Jorens P. Automatic colorimetric calibration of human wounds. BMC Med Imag. 2010;10(7):1–11.

[4] Haeghen Y, Naeyaert J, Philips W. An imaging system with calibrated color image acquisition for use in dermatology. IEEE Trans Med Imag. 2000 July;19(7):722–730.

[5] Hsu RL, Abdel-Mottaleb M, Jain AJ. Face detection in color images. IEEE Tran on Pattern Analys and Mach Intell. 2002 May;24(5):696–706.

[6] Li W, Soto-Thompson M, Xiong Y, Lange H. A new image calibration technique for colposcopic images. In: Proc SPIE Med Imag. vol. 6144. International Society for Optics and Photonics; 2006. .

[7] Varghese D, Wanat R, Mantiuk R. Colorimetric calibration of high dynamic range images with a ColorChecker chart. In: Proc HDRi; 2014. p. 17–22.

[8] Jose A, Haak D, Jonas S, Brandenburg V, Deserno TM. Human wound photogrammetry with low-cost hardware based on automatic calibration of geometry and color. In: Proc SPIE Med Imag. vol. 9414. International Society for Optics and Photonics; 2015. p. 94143J.

[9] Poynton CA. Digital video and HDTV: Algorithms and interfaces. Morgan Kaufmann Publishers Inc; 2003.

[10] Leong F, Brady M, McGee J. Correction of uneven illumination (vignetting) in digital microscopy images. Journ of Clinic Pathol. 2003 February;56(8):619–621.

[11] Lawson CL, Hanson RJ. Solving least squares problems. SIAM; 1974.