Objective Measurement of Erythema in Psoriasis using Digital Color Photography with Color Calibration

Abhay Raina, Ricky Hennessy, Michael Rains, James Allred, Dayna Diven, Mia K. Markey, *Senior Member, IEEE*

***Abstract*—Traditional metrics for evaluating the severity of psoriasis are highly subjective, which complicates efforts to identify effective treatments in clinical trials. We propose a method for the objective measurement of the psoriasis severity parameter of erythema (redness). This procedure is standardized for different camera systems and lighting environments through the usage of a color card with predetermined color values in order to calibrate the images. Quantitative measures based on the digital color images are shown to correlate well with subjective assessment of psoriasis severity collected using a standard numerical scale by a panel of dermatologists. Additionally, the color calibration process is shown to improve results.**

I. Introduction

Psoriasis is a chronic, inflammatory disease that affects the skin and joints. The most common form is plaque psoriasis, which presents with scaly red and white patches on the epidermal layer of the skin [1]. These plaques usually occur on the elbows and knees, but they can affect any part of the body. The cause of the condition is not fully understood, and there is no cure currently available [1].

Examination of the National Health and Nutrition Examination Survey suggests the prevalence of diagnosed psoriasis in the United States is 3.15%, which corresponds to roughly 5 million adults [2]. The survey also indicated that there is a significant amount of undiagnosed psoriasis, which ranges from 0.4% to 2.28% of the US population depending on how broadly the condition is defined. Out of these patients, 17% have moderate to severe psoriasis and 25% report that the condition poses a significant problem in their daily lives [2]. Furthermore, the condition is associated with widespread treatment dissatisfaction [3].

Abhay Raina is with the Biomedical Engineering Department, The University of Texas, Austin, TX 78712 aberaina@utexas.edu

Ricky Hennessy is with the Biomedical Engineering Department, The University of Texas, Austin, TX 78712 hennessy@utexas.edu

Michael Rains is with the Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, TX 75235 mrainsutmb@gmail.com.

James Allred is with the Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, TX 75235 JTAllred@seton.org.

Dayna Diven is with the Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, TX 75235 dgdiven@seton.org.

Mia K. Markey is with the Biomedical Engineering Department, The University of Texas, Austin, TX 78712. She is also with the Imaging Physics Department at The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 mia.markey@utexas.edu

One of the major challenges to developing more effective treatment of psoriasis is difficulty in tracking the progression of psoriasis given the subjective nature of assessing its severity. Even experienced physicians can show wide variation in evaluating the severity of psoriatic plaques [4]. The lack of an objective metric for psoriasis severity inhibits tracking of patient progress and establishment of treatment goals [5]. This is particularly problematic for studies intended to compare and evaluate different treatments, because it increases the difficulty of establishing an objective improvement [5]. The necessity of a physician's evaluation of severity parameters also significantly increases the cost and duration of these studies.

Currently, the most widespread method of evaluation of psoriasis for clinical trials is the psoriasis area and severity index (PASI) [5]. In this semi-quantitative method, the body is split up into 4 sections (head, arms, trunk and legs) and each section is given a specific weight based on the percentage of the body's total skin in that region. A physician evaluates the severity of the psoriasis in each of those regions on a 0-4 scale based on the erythema (redness), desquamation (scaling) and induration (thickness) of the plaques as well as the proportion of skin affected [5]. All of these values are entered into a formula that yields a value from 0-72 indicating the overall severity. In this study, we only consider the erythema scores.

However, PASI has significant limitations. Despite efforts to refine the PASI formula that have yielded marginal improvements, all variations of the PASI score suffer from similar drawbacks [6]. The evaluation of the severity parameters is still a relatively subjective endeavor, which repudiates the purpose of an objective score.

Consequently, there have been efforts to develop new tools for the automated evaluation of psoriasis based on clinical images. These new methods are based on the traditional PASI model, and so seek to automate the evaluation of the established parameters: erythema, desquamation, and induration. However, it is not readily feasible to determine induration of a plaque from an image, so research has focused on developing measures of erythema and desquamation.

Some prior studies classified the severity of psoriasis plaques using erythema, desquamation, or both [7], [8], [9], [10]. However, all of the existing studies used a standard camera in order to acquire the images, and little effort has been made to create a calibration method to standardize the images across cameras and lighting conditions**.** This is a key issue since a variety of cameras and lighting conditions are present in clinical practice. Moreover, prior work has

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

explored only few of the feature sets that could potentially be used to quantify properties such as redness or scaling.

The goal of this study was to algorithmically calculate a measure of erythema from clinical photographs and evaluate it with respect to an expert assessment of erythema by a panel of dermatologists. We also aimed to standardize the process so that any handheld camera could be used to acquire the images instead of a single specific camera system.

II. Materials and Methods

A. Data Set

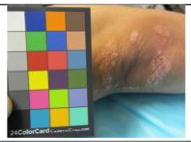
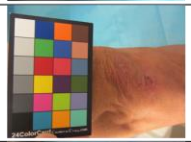
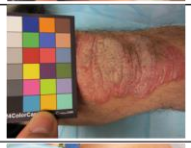
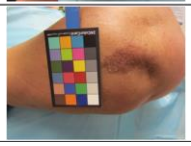
Standard clinical photographs were taken of 20 patients from local dermatology clinics exhibiting psoriatic plaques. Images were taken of both knees and both elbows for each patient regardless of whether all of those areas exhibited plaques, giving a total of 80 images. Pictures were taken with either a Canon PowerShot SX230 HS (Cannon U.S.A., Melville, New York) or a Canon PowerShot ELPH 520 HS, and the field of view was large enough to encompass the plaque as well as some surrounding skin for comparison. We also included a 4x6 color card (CameraTrax, Las Vegas, Nevada) with 24 colors in the images in order to calibrate the coloration.

Photographs were typically taken against a uniform blue background. The normal lighting of the clinic rooms was used to illuminate the target regions.

B. Manual Rating

The images were rated by a panel of five dermatologists in accordance with the traditional PASI score parameters. The raters were given a chart with examples of each score for reference during the rating, and the images were shown to each rater using the same screen. Each of the raters reported a score from 0-4 for the erythema severity of each of the images. Examples from each category are shown in Table I.

TABLE I. Erythema Scoring Examples from Data Set

Score	Description	Example Image
1	Mild	
2	Moderate	
3	Severe	
4	Very Severe	

The subjective erythema scores were analyzed for agreement and consistency. The intraclass correlation coefficient (ICC) was calculated assuming the raters were fixed and the different images represented a random sample of possible images. The ICC test for agreement indicated that r = 0.7306 and the test for consistency showed that r = 0.7492. These values indicate a high level of uniformity for both parameters of concordance. The median of the scores was taken as a composite overall assessment in order to ensure that the images were put into discrete categories.

C. Image analysis

One of our goals was to create a process that could function with any camera. This was challenging since different cameras and lighting conditions can produce differences in sharpness, coloration, and noise that can subvert image analysis.

In order to remedy the problem, we implemented a method proposed by Marguir et al. for the calibration of the photographs using a color card with pre-defined color values in order to ensure that the images appear similar [11]. They used a color card with pre-defined color values in order to standardize the coloration of different types of skin despite varying levels of illumination present and different cameras used. Their method was able to make the resulting images look very consistent despite widely varying lighting conditions. The only difference from our study was that we utilized a simpler 4x6 color card instead of the more extensive ones that they employed.

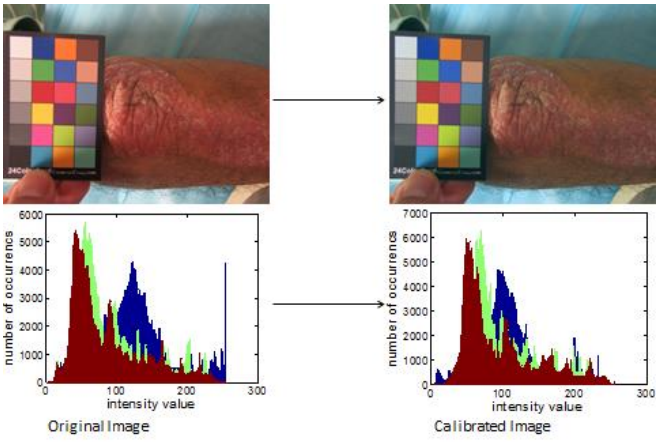


Fig. 1. The original image is calibrated with respect to the color card, and the RGB distribution of the coloration changes.

After obtaining the calibrated images, features were extracted from the plaques. We compiled a list of features from other similar studies that were intended to match the degree of redness in an image [7], [8], [9]. This is the complete list of features considered:

- Mean, standard deviation, skewness, and kurtosis of each of the R, G, and B components
- Mean, standard deviation, skewness, and kurtosis of the difference between the red band and the green

band (RG) and likewise for the difference between the blue and green bands (BG)

- Proportion of the image that is red (R/(R+G+B))
- Mean, standard deviation, skewness, and kurtosis of each of the L*, a*, b* components
- Differences between the means of the red, green, and blue channels of the plaque and surrounding skin
- Difference between the average hemoglobin and melanin components of the plaque and surrounding skin [10]

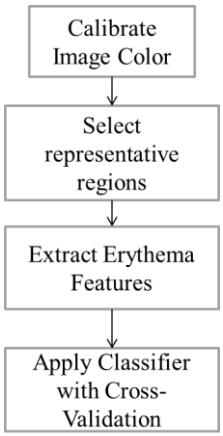


Fig. 2. Flowchart showing the overall image analysis process.

Before the features were evaluated, a human operator familiar with psoriasis images selected two representative areas on each of the images manually. One area was taken to be representative of the erythema of the image and the other was representative of the normal skin of the patient. A bounding box was drawn enclosing each selected area for isolation. This manual area selection removed the need for an automated segmentation process to isolate the plaques. This is key since psoriasis plaques are highly variable and hence not amenable to automated segmentation. Moreover, since there can be substantial variation even within a given psoriasis location, selecting an area that is typical helps focus the subsequent analysis.

The following features were extracted from the representative area associated with erythema: (1) mean, (2) standard deviation, and (3) kurtosis of the red channel; (4) mean of the proportion of redness in the RGB image; (5) mean, (6) standard deviation, and (7) kurtosis of the difference between the blue and green channels. These features were selected through a sequential trial-and-error process. Each feature from the initial list was added one at a time to the classifier, and if it improved the resulting correlation then it was retained.

The features and corresponding ratings of erythema were input into a linear discriminant analysis classifier using a leave-one-out cross-validation procedure. Specifically, the features and the associated expert ratings of all 4 images of all but one of the subjects were used to train a classifier to predict the severity of erythema for the 4 images of the one patient who was excluded; the process was repeated such that each patient was held out for testing. The results of the classification were compared with the expert ratings using a linearly weighted Cohen's kappa coefficient because this measure accounts for different degrees of concordance. If the predicted and expert rating are close together, the overall score will be penalized less than if the ratings are far apart. The accuracy was also calculated in order to compare the results with prior studies that employed this metric.

We also tested whether the calibration had a significant influence on the results of the classification. Consequently, the same process was repeated without the color calibration step in order to compare the results of using the classifier on the original and calibrated images.

III. RESULTS

The quantitative erythema scores had roughly the same distribution as the expert ratings of erythema, although the classifier tended to overestimate the severity (Table II). The agreement between the quantitative scores computed by our algorithm (with image calibration) and the subjective ratings by experienced dermatologists was $\kappa = 0.4203$ and likewise the accuracy was 48.75% which constitutes a good but not exceptional degree of agreement.

TABLE II. FREQUENCY OF DIFFERENTT PREDICTED RATING AND EXPERT RATING CATEGORIES

Predicted Rating (0-4)	Expert Rating (0-4)				
	0	**1**	**2**	**3**	**4**
0	12	5	0	1	0
1	2	12	8	1	0
2	0	8	8	2	1
3	1	3	6	7	0
4	1	1	1	0	0

In contrast, when the same algorithm was applied to the images without calibration, the agreement between the quantitative scores and the subjective ratings was $\kappa = 0.2364$ and the accuracy was 42.5%. This means that the calibration process produced a moderate improvement in the accuracy of the classifier evaluated with respect to subjective rating of erythema by experienced dermatologists.

Our method for calibrating of the coloration of the images is promising for improving automatic classification of erythema severity. Previous work has identified the influence of difference sources of noise in the coloration as an important source for error in evaluating erythema [9]. This suggests that the usage of a color card is an effective way to reduce noise introduced by environmental factors as well as the camera itself.

The accuracy of classification in this study was not as high as those reported by previous studies. For example, a recent study from Lu et al. reported a much higher overall

accuracy of 78.85% for the correct categorization of the erythema severity [8]. However, the results reported here cannot be simply compared to those in the literature since both the methodology and the data set are different. To enable direct comparison of methods, one must apply them to the same data set.

Hence, in order to directly compare our results to prior work, we implemented the analysis method described by Lu et al. and applied it to our own data set. Lu et al. used five total features, including the difference between the plaque and the skin in the mean of the red, green, and blue channels as well as the differences in what they define as the average hemoglobin and melanin components of the image [8]. They applied a k-nearest neighbors classification algorithm (k=5) with 10-fold cross-validation. When we applied their method to our data set, the accuracy was 30.0%, which is lower than the accuracy obtained using the algorithm we propose here (48.75%). This result suggests that the overall lower classification accuracy obtained in this study is likely to be due to characteristics of the data set.

It should be noted that the images in our study require a larger field of view in order to include the entire plaque, the color card, and surrounding skin. However, a larger field of view does not permit as much detail in in the image of the plaque itself, and also introduces a greater risk of error from shadows and uneven illumination. In some of the images in our data set, the illumination was not completely uniform and shadows were visible in the image. This served to introduce significant noise into the calculations particularly in the case where the shadow was over the plaque but not on the color card because the color calibration could not account for the shadows.

The images from other studies appear to have been taken with a smaller field of view and more consistent illumination, which could help explain why higher accuracy was obtained [7], [8]. A more detailed image of the plaque could make it easier for the algorithms to distinguish the images.

IV. Conclusion

In conclusion, we present a quantitative method for measuring erythema in psoriasis plaques and show that color calibration using a color card leads to improved results. This technique could serve to reduce error from environmental factors such as illumination and could theoretically standardize image collection from different camera systems.

However, the larger field of view needed to include the color card in the images can also make classification more challenging by reducing the amount of information specifically from the plaque.

Acknowledgments

We would like to thank Dr. Ashley Brown and Dr. Donald Warren for helping with the rating of the images as well as the Seton hospital system for accommodating our data collection.

References

[1] A. R. Mason, J. Mason, M. Cork, G. Dooley, and G. Edwards, "Topical treatments for chronic plaque psoriasis," *Cochrane Database Syst. Rev.*, no. 2, p. CD005028, 2009.

[2] S. K. Kurd and J. M. Gelfand, "The prevalence of previously diagnosed and undiagnosed psoriasis in US adults: results from NHANES 2003-2004," *J. Am. Acad. Dermatol.*, vol. 60, no. 2, pp. 218–224, Feb. 2009.

[3] R. S. Stern, T. Nijsten, S. R. Feldman, D. J. Margolis, and T. Rolstad, "Psoriasis is common, carries a substantial burden even when not extensive, and is associated with widespread treatment dissatisfaction," *J. Investig. Dermatol. Symp. Proc. Soc. Investig. Dermatol. Inc Eur. Soc. Dermatol. Res.*, vol. 9, no. 2, pp. 136–139, Mar. 2004.

[4] R. Marks, S. P. Barton, D. Shuttleworth, and A. Y. Finlay, "Assessment of disease progress in psoriasis," *Arch. Dermatol.*, vol. 125, no. 2, pp. 235–240, Feb. 1989.

[5] I. Schäfer, J. Hacker, S. J. Rustenbach, M. Radtke, N. Franzke, and M. Augustin, "Concordance of the Psoriasis Area and Severity Index (PASI) and patient-reported outcomes in psoriasis treatment," *Eur. J. Dermatol. EJD*, vol. 20, no. 1, pp. 62–67, Feb. 2010.

[6] T. Henseler and K. Schmitt-Rau, "A comparison between BSA, PASI, PLASI and SAPASI as measures of disease severity and improvement by therapy in patients with psoriasis," *Int. J. Dermatol.*, vol. 47, no. 10, pp. 1019–1023, Oct. 2008.

[7] M. H. Ahmad Fadzil, D. Ihtatho, A. Mohd Affandi, and S. H. Hussein, "Objective assessment of psoriasis erythema for PASI scoring," *J. Med. Eng. Technol.*, vol. 33, no. 7, pp. 516–524, 2009.

[8] J. Lu, E. Kazmiercazk, J. H. Manton, and R. Sinclair, "Automatic Scoring of Erythema and Scaling Severity in Psoriasis Diagnosis," in *AI 2012: Advances in Artificial Intelligence*, M. Thielscher and D. Zhang, Eds. Springer Berlin Heidelberg, 2012, pp. 73–84.

[9] H. Iyatomi, H. Oka, M. Hagiwara, A. Miyake, M. Kimoto, K. Ogawa, and M. Tanaka, "Computerized quantification of psoriasis lesions with colour calibration: preliminary results," *Clin. Exp. Dermatol.*, vol. 34, no. 7, pp. 830–833, Oct. 2009.

[10] J. Lu, E. Kazmierczak, J. H. Manton, and R. Sinclair, "A quantitative technique for assessing the change in severity over time in psoriatic lesions using computer aided image analysis," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2380–2383.

[11] J. Marguier, N. Bhatti, H. Baker, M. Harville, and S. Süsstrunk, "Assessing human skin color from uncalibrated images," *Int. J. Imaging Syst. Technol.*, vol. 17, no. 3, pp. 143–151, 2007.