Automated Wound Identification System Based on Image Segmentation and Artificial Neural Networks

Bo Song, Ahmet Sacan
School of Biomedical Engineering, Science and Health Systems
Drexel University, 3120 Market Street
Philadelphia, PA 19104, USA
bo.song@drexel.edu, ahmet.sacan@drexel.edu

***Abstract*—A system that can automatically and accurately identify the region of a chronic wound could largely improve conventional clinical practice for the wound diagnosis and treatment. We designed a system that uses color wound photographs taken from the patients, and is capable of automatic image segmentation and wound region identification. Several commonly used segmentation methods are utilized with their parameters fine-tuned automatically to obtain a collection of candidate wound regions. Two different types of Artificial Neural Networks (ANNs), the Multi-Layer Perceptron (MLP) and the Radial Basis Function (RBF) with parameters determined by a cross-validation approach, are then applied with supervised learning in the prediction procedure for the wound identification, and their results are compared. The satisfactory results obtained by this system make it a promising tool to assist in the field of clinical wound evaluation.**

Keywords-Wound Identification; Image Segmentation; Artificial Neural Networks

I. INTRODUCTION

In chronic wounds, localized damage to the skin is an ongoing health challenge that afflicts approximately 5.7 million patients in the US [1], most of whom are elderly and bedridden patients. Identifying the wound region is the first step in assessing and treating the chronic wound condition. However in clinical practice, wound evaluation is mostly based on visual inspections and hands-on measurements which are subjective and inaccurate [2].

Digital image based and computer aided segmentation are solutions that improve the accuracy of wound assessment [3]. However, the best parameters of each image segmentation method differ from image to image, and are always decided by prior knowledge or from a large amount of trial experiments. Moreover, existing generic segmentation methods are themselves task-agnostic especially when applied to domain-specific problems [4]. In the wound identification application, semiautomatic methodology could have better performance [5], but it remains labor intensive.

This paper proposes the use of several image segmentation methods for wound segmentation from diabetic foot ulcer images and the subsequent identification of the best segmentation among them, using a machine learning approach. A parameter optimization procedure is employed to automatically fine-tune each segmentation method for specific samples with similar traits in the domain; Artificial Neural Networks (ANNs) are then applied to learn to identify and evaluate the segmentation results automatically.

II. METHODS

The methods proposed are applied in the system through two procedures. During the training procedure, wound images in the database are segmented by different methods with their corresponding optimized parameters; the resulting polygons are then filtered to remove the trivial ones for faster processing in the next step. In every image, certain properties are extracted from each of the segmented and filtered polygons to form their feature vectors; simultaneously, each polygon is compared with the manually traced wound region to obtain an overlap score. Feature vectors and overlap scores are then used to train the identification system and feed in the neural networks.

During the operation procedure, the wound region in any new wound image will be automatically recognized and evaluated by the trained system based on feature vectors of segmented and filtered polygons of that image.

A summary of our proposed methods are briefly described in Fig. 1. In the following sections, each particular step of the above algorithms will be described in detail.

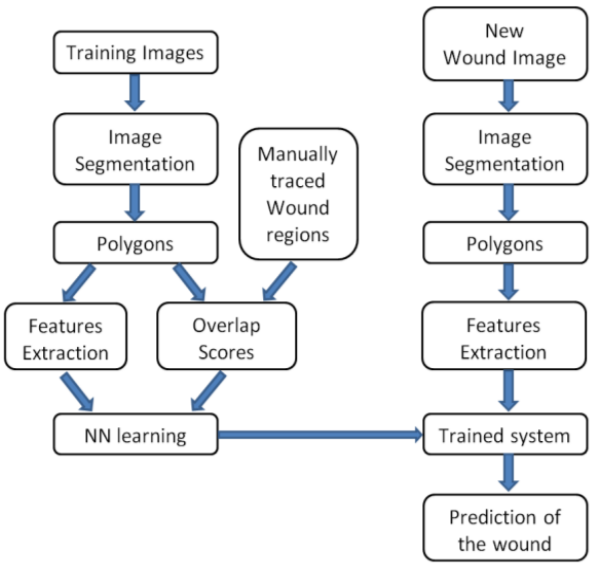

Figure 1. Flow chart of the proposed method.

A. Integrated Segmentation Approach

Accurate image segmentation is one of the fundamental precepts in numerous fields such as image analysis and pattern recognition [6], and a primary prerequisite in our system for successful automatic identification of the wound region. However, due to the variety of wounds with different features presented in the digital image, a precise automatic segmentation is not trivial. Although image segmentation has been studied for decades [7], there is still not one universal method that can be considered effective to be applied to all types of images with diverse features, as each segmentation method is designed and developed for a particular class of image whose features and conditions are specific; accordingly, not all methods will perform equally well on the same image [8].

In order to tackle this problem, we selected four popular image segmentation methods which are K-means Clustering, Edge Detection, Thresholding, and Region Growing, and then packed these four segmentation methods as one integrated segmentation approach. Parameters of each method will be regulated automatically for one class of wound images grouped by similar features. Under this solution, different image containing wound with different features can always be effectively segmented by at least one fit segmentation method in the integrated segmentation approach. This design greatly increases the robustness of the system.

Furthermore, while all the segmentation algorithms have been well-developed in gray-scale, we applied them in color space with more information utilized and hence will generate more accurate segmentation results [9].

B. Pre-processing

In order to guarantee and improve the efficiency, the image or data in the preceding and succeeding stages of the segmentation will need to be preprocessed and prepared.

The wound region of each image from the database will be all manually traced and saved as a standard. To facilitate this procedure, we created a graphical user interface (GUI) with multiple tracing tools.

As our system is based on MEGA-pixel digital photography, image size is a big concern for the processing time, especially when implementing pixel-based segmentation. Thus adaptive rescaling is applied on the images prior to the segmentation to reduce time consumption, and the segmented results will be rescaled back to the original size.

Since each of the four segmentation methods could result in numerous polygons on one image, each image will get a considerable number of polygons in total. In order to accelerate the processing, we created a filter to reduce the number of polygons to be selected as candidate wound regions, by eliminating trivial polygons that possess too large or too small areas in relation to the total size of the image, as well as those located too close to the border of the image.

For each remaining candidate polygon, an overlap score is calculated as its performance criterion for evaluation, parameter optimization, and ANNs training. Every polygon of an image is compared with the manually traced wound region of that image by the Matthews Correlation Coefficient (MCC) measure algorithm to score the agreement in between. MCC will output a score between -1 and +1, where -1 represents a worst agreement and +1 indicates the best agreement between two classifications.

MCC can be represented by the formula:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \quad (1)$$

Where, TP (true positive) represents the pixels of overlapping between P (manually traced wound region) and Q (segmented polygon), TN (true negative) represents the pixels that do not belong to either P or Q, FP (false positive) represents the pixels belonging to Q but not to P, and FN (false negative) represents the pixels in P but not Q.

The higher the resulting overlap score of a polygon, the better the chance of it being considered a representative of the expected wound region.

C. Parameter optimization

Different segmentation method fits different images with distinct features, and we packed four methods for different group of wound images with similar features. However the best parameters for a suitable segmentation method could still be variable and sensitive to the slight diversity of features in target images even when the features are similar. Conventionally, selecting the proper parameters for one image relies on prior knowledge or on conducting many trial experiments which is non-trivial, and it is also difficult to implement quantifiable evaluations, let along decide the optimal parameters for a group of images with similar features while also considering the effects due to the interactions among the parameters.

We proposed a solution to this dilemma which automatically optimizes the parameters of each segmentation method. Firstly, training images are grouped by features' similarity to focus on different segmentation methods. Then in each group, the corresponding segmentation method with parameters set to their initial values will produce multiple polygons and their according overlap scores. The mean value of all the highest scores collected from each image in the group is used as an optimization criterion. After that, parameters values will be altered and a new optimization criterion will be generated. This process is iterative to search for the best criterion, and the corresponding parameters will be the optimal parameters for that segmentation method. Parameters of all four segmentation methods are optimized in the same way.

As there are two data types for the segmentation parameters, we applied a combination of two optimization techniques for the best criterion searching: one is the exhaustive Grid search [10] for integer-based parameters, the other one is the computationally compact Nelder-Mead simplex algorithm [11] for continuous non-integer parameters.

D. Feature Extraction

Feature vectors representing the candidate polygons, which are generated by the optimized segmentation methods and the following filtering process, are prepared as training inputs for the ANNs.

We extracted 49 features out of a polygon to form its representative feature vector, which consists of the information of pixel value measurements, shape measurements, and geometry measurements of the polygon. These features included: area, perimeter, centroid, weighted centroid, axes lengths, min/max/mean intensities of each color channel, and the centroidal principal moments.

III. Experiments & Results

The wound images applied in our project are acquired from diabetic foot ulcer patients recruited for a chronic wound assessment study [12]. There are 2 to 10 digital color images recorded during the wound healing procedure for each patient. We ruled out images of 5 patients from 19 which are of low quality and less value, leaving 14 patients with 92 available wound images. One image is picked up randomly from each of the patients to form a testing database which consists of 14 wound images, and the remaining 78 images will serve as training database for the system. Wound region in each image from both databases is manually traced in advance by experts and realized on our GUI.

We classified training images into four groups by their feature similarities for the parameters of four segmentation methods in the integrated segmentation approach to be optimized respectively. The optimal values of two parameters in K-means Clustering, four parameters in Edge Detection, one parameter in Thresholding and two parameters in Region Growing are then automatically determined

After all the training images being segmented with the optimized parameters, the resulting polygons are filtered. Finally 1451 polygons are considered as candidates of wound regions. Then a feature vector is formed to represent each of the candidate polygons, whose overlap score is also calculated based on the MCC algorithm.

For the identification system, we applied two types of feed forward ANNs conducting supervised learning, and provided them with feature vectors as training inputs together with their corresponding overlap scores as training outputs. One is Multi-Layer Perceptron (MLP) and the other is the Radial Basis Function (RBF) [13, 14], both of which are implemented with the MATLAB R2012a Neural Network Toolbox™ 7.

The connection weight and bias values were updated until the Mean Squared Error (MSE) between the prediction outputs and the desired outputs (overlap scores) is minimized. The number of nodes and spread parameters were identified by minimizing the MSE on the training data. 5 hidden neurons were used in MLP, and 40 neurons were used in RBF.

After both ANNs are created, 1451 feature vectors and their corresponding overlap scores are fed into the networks. 14 wound images in the testing database are then put through the procedure of rescaling, image segmentation, filtering, and feature extraction, resulting in representative feature vectors to be fed into the identification system, which will generate predicted overlap scores correspondingly. The polygon with highest predicted score of each image will be selected as the predicted wound region.

In order to test and compare the predictions by the two trained ANNs in the identification system, the testing images are all manually traced to provide the overlap scores of their candidate polygons. As the wound region of each image might be segmented out by more than one of the four segmentation methods, using the overlap scores together with the visualization of the segmentation results, we are able to choose the desired polygons of each image that are supposed to be identified by the system.

Take one particular testing image as an example, its visualized candidate polygons are shown in Fig. 2, and the desired polygons are the second (from K-means Clustering), the tenth (from Thresholding) and the thirteenth (from Region Growing).

If the predicted polygon is the same as the desired polygon, the wound region of that image is considered to be successfully identified by the prediction system. The statistical results of the testing from MLP and RBF are shown in the Tab. I, where the prediction result is 10 out of 14 images correct for MLP, and 12 out of 14 images correct for RBF. While MCC values ranged from -1 to 1, the MSE values produced by both networks were approximately 0.01, which is quite satisfactory.

In order to further analyze the efficacy of the system, we examine the ranking of the predicted score of the desired polygons. If the desired polygons have no top ranked score, but are all ranking greater than most of the other candidate polygons in an image, the prediction can still be considered to be highly effective. The ranking analysis of the 14 testing images is detailed in Tab. II.

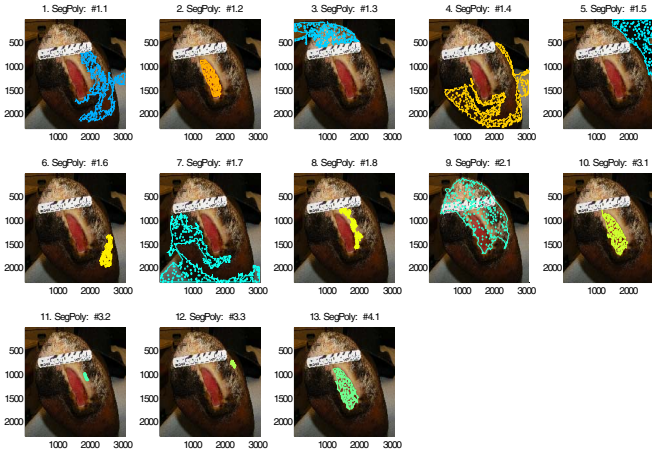

Figure 2. The visualization results of candidate polygons by four segmentation methods on one test image. The desired polygon number is the second, tenth and the thirteenth (from left to right and top to bottom).

TABLE I. COMPARATIVE RESULTS OF THE TWO ANNS

	Time used for training (second)	Correct Prediction (%)	training MSE
MLP (5)*	12.6636	71.4 (10/14)	0.010681
RBF (40)**	1.7175	85.7 (12/14)	0.019551

*Number of neurons in MLP, **Number of neurons in RBF.

TABLE II. COMPARATIVE RANKING EVALUATION OF THE TWO ANNS

Test image number	Number of segmented polygons	Number of desired polygons	Ranking of desired polygons					
			MLP			*RBF*		
			# 1	*# 2*	*# 3*	*# 1*	*# 2*	*# 3*
1	13	3	1	2	3	1	2	3
2	21	1	1	N/A	N/A	1	N/A	N/A
3	21	1	2	N/A	N/A	20	N/A	N/A
4	20	2	1	3	N/A	1	2	N/A
5	22	1	1	N/A	N/A	1	N/A	N/A
6	21	2	1	2	N/A	1	3	N/A
7	14	2	2	4	N/A	1	5	N/A
8	19	1	1	N/A	N/A	2	N/A	N/A
9	21	2	1	3	N/A	1	2	N/A
10	18	2	1	2	N/A	1	2	N/A
11	17	1	1	N/A	N/A	1	N/A	N/A
12	15	3	2	3	4	1	2	3
13	21	1	3	N/A	N/A	1	N/A	N/A
14	11	2	1	2	N/A	1	2	N/A
Average	18	1.7	1.4	2.5	3.5	2.4	2.5	3

IV. DISCUSSION & CONCLUSION

While prediction and comparison results of the two ANNs presented in Tab. I and II indicate that both regulated MLP and RBF have decent efficiency, the two types of networks also have different advantages and disadvantages.

MLP has a slightly better performance in respect of the training MSE, and a better generalization capability as the desired polygons are more stable and are higher ranking, although its exact accuracy rate is less competitive with the RBF network in a testing database of fourteen images, which could make a difference in a much larger database. Another drawback of MLP is its training time, especially when the training database is large. However, because the training procedure is a one-time cost process, this could be a tolerable issue. RBF, on the other hand, needs significantly less training time and has more competitive prediction accuracy. However as shown in Tab. II, the deviation of the prediction is worse, and this is mainly due to the local nature of its approximation capability.

In summary, we proposed and designed a system to automatically identify the wound region from clinical digital wound images. There is a one-time cost of training with a set of manually traced images; once trained, the system is fully automated. The satisfactory results of the identification system demonstrate that the proposed methods and implemented system are a promising way for automated clinical wound assessment, and the reduction of the human error and intensity guarantees the greater accuracy. Future works may include further optimization of the system based on these preliminary results.

REFERENCES

[1] C. K. Sen, G. M. Gordillo, S. Roy, R. Kirsner, L. Lambert, T. K. Hunt, F. Gottrup, G. C. Gurtner, and M. T. Longaker, "Human skin wounds: a major and snowballing threat to public health and the economy," Wound Repair Regen, vol. 17, pp. 763-71, Nov-Dec 2009.

[2] D. H. Keast, C. K. Bowering, A. W. Evans, G. L. Mackean, C. Burrows, and L. D'Souza, "MEASURE: A proposed assessment framework for developing best practice recommendations for wound assessment," Wound Repair and Regeneration, vol. 12, pp. S1-S17, May-Jun 2004.

[3] H. A. Thawer, P. E. Houghton, M. G. Woodbury, D. Keast, and K. Campbell, "A comparison of computer-assisted and manual wound size measurement," Ostomy Wound Manage, vol. 48, pp. 46-53, Oct 2002.

[4] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," Annual review of biomedical engineering, vol. 2, pp. 315-337, 2000.

[5] E. S. Papazoglou, L. Zubkov, X. Mao, M. Neidrauer, N. Rannou, and M. S. Weingarten, "Image analysis of chronic wounds for determining the surface area," Wound Repair Regen, vol. 18, pp. 349-58, Jul-Aug 2010.

[6] R. C. González and R. E. Woods, Digital Image Processing: Prentice Hall, 2008.

[7] K. S. Fu and J. K. Mui, "A survey on image segmentation," Pattern Recognition, vol. 13, pp. 3-16, 1981.

[8] N. R. Pal and S. K. Pal, "A Review on Image Segmentation Techniques," Pattern Recognition, vol. 26, pp. 1277-1294, Sep 1993.

[9] K. N. Plataniotis and A. N. Venetsanopoulos, Color Image Processing and Applications. Berlin: Springer, 2000.

[10] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," 2011.

[11] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," Siam journal of optimization, vol. 9, pp. 112-147, 1998.

[12] E. S. Papazoglou, M. S. Weingarten, L. Zubkov, M. Neidrauer, and K. Pourrezaei, "Assessment of diabetic foot ulcers with diffuse near infrared methodology," 2008, pp. 1-5.

[13] M. T. Hagan, H. B. Demuth, M. H. Beale, and B. University of Colorado, Neural network design: PWS Pub, 1996.

[14] J. E. Dayhoff and J. M. DeLeo, "Artificial neural networks," Cancer, vol. 91, pp. 1615-1635, 2001.