Boundary determination of foot ulcer images by applying the associative hierarchical random field framework

Lei Wang
Peder C. Pedersen
Emmanuel Agu
Diane Strong
Bengisu Tulu

Boundary determination of foot ulcer images by applying the associative hierarchical random field framework

Lei Wang,[a,*] Peder C. Pedersen,[a] Emmanuel Agu,[b] Diane Strong,[c] and Bengisu Tulu[c]

[a]Worcester Polytechnic Institute, Department of Electrical and Computer Engineering, Worcester, Massachusetts, United States
[b]Worcester Polytechnic Institute, Department of Computer Science, Worcester, Massachusetts, United States
[c]Worcester Polytechnic Institute, Foisie School of Business, Worcester, Massachusetts, United States

Abstract. As traditional visual-examination-based methods provide neither reliable nor consistent wound assessment, several computer-based approaches for quantitative wound image analysis have been proposed in recent years. However, these methods require either some level of human interaction for proper image processing or that images be captured under controlled conditions. However, to become a practical tool of diabetic patients for wound management, the wound image algorithm needs to be able to correctly locate and detect the wound boundary of images acquired under less-constrained conditions, where the illumination and camera angle can vary within reasonable bounds. We present a wound boundary determination method that is robust to lighting and camera orientation perturbations by applying the associative hierarchical random field (AHRF) framework, which is an improved conditional random field (CRF) model originally applied to natural image multiscale analysis. To validate the robustness of the AHRF framework for wound boundary recognition tasks, we have tested the method on two image datasets: (1) foot and leg ulcer images (for the patients we have tracked for 2 years) that were captured under one of the two conditions, such that 70% of the entire dataset are captured with image capture box to ensure consistent lighting and range and the remaining 30% of the images are captured by a handheld camera under varied conditions of lighting, incident angle, and range and (2) moulage wound images that were captured under similarly varied conditions. Compared to other CRF-based machine learning strategies, our new method provides a determination accuracy with the best global performance rates (specificity: >95% and sensitivity: >77%. © *2019 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMI.6.2.024002]

Keywords: conditional random field; diabetic foot ulcer; wound boundary determination; wound image analysis.

Paper 18076R received Apr. 16, 2018; accepted for publication Mar. 29, 2019; published online Apr. 21, 2019.

1 Introduction

Traditional wound assessment methods, based on visual examination and manual measurement, risk inconsistency and may not facilitate easy assessment of healing trends. In addition, frequent visits to the wound clinic are often found retrospectively to be unnecessary and represent a financial burden for the patients, as well as an avoidable workload for clinicians. Hence, in recent years, there has been a research focus on developing computer-based methods to achieve more consistent, objective, and clinically meaningful wound image analysis methods. Although many numerical indicators can be applied to describe the wound healing status (such as the geometrical measures of the wound dimension and the composition ratio for different tissues), it is still the clinical opinion that the size of the wound area is the most significant measure that provides the foundation for all other wound analysis work, including tissue classification and healing rate evaluation. Thus, this paper mainly focuses on applying computer vision techniques to determine the boundary of wound areas.

Initially, computer-based approaches for locating a wound in an image and detecting the wound boundary have been based on nonmachine learning-based methods, such as active contour models, level set-based methods, and synthetic image segmentation strategies.[1–3] However, these methods suffer from performance limitations when dealing with wound images with complicated skin textures and boundaries. Hence, most wound analysis research works in recent years have concentrated on machine learning-based computer vision approaches.

The bottom–up-based object recognition scheme has been widely used in recent years for wound boundary determination research.[4–6] This approach has a common set of processing steps: image segmentation; segment-based feature extraction; and classifier training by applying specific machine learning methods, such as support vector machine (SVM) or artificial neural network (ANN).[1,4,7,8] To improve the accuracy of wound boundary detection, more sophisticated techniques have been investigated. For example, a cascaded classifier based on ANN and Bayesian committee machine was proposed in Ref. 4 to classify different types of wound tissue. In previous work by the authors,[9] a two-stage cascaded SVM-based classifier was evaluated to determine the wound area. The SVM classifier achieved fairly accurate performance (sensitivity = 73.3% and specificity = 94.6%), provided that the wound images were captured under carefully controlled lighting and camera range, but the system did allow a significant amount of surrounding healthy skin to appear in the image which was good progress. The controlled capture conditions were achieved by using an

*Address all correspondence to Lei Wang, E-mail: wanglei19841124@gmail.com

image capture box.[9] In contrast, in other SVM-based studies,[8,10] the tested images consisted mainly of the wound areas, surrounded by only a small amount of healthy skin tissue.

Figure 1(a)–1(c) show original images captured without the aid of the image capture box, specifically images of the same moulage wound (synthetic wound) under different illumination levels and ranges. The wound areas identified by our cascaded SVM classifier are marked in red in Figs. 1(d)–1(f). The wound boundary is here defined as the outer edge of the pale yellow gel border, indicating a correct result in Fig. 1(d). We can see clearly the erroneous boundary detection on the second and third images. We can observe that the cascaded SVM classifier method totally missed the target when the wound was further away and the background information is complicated. Hence, a more robust method is needed if we wish to broaden the utility of the wound recognition algorithm by relaxing the image capture constraints, i.e., image capture without the use of the image capture box.

One possible solution to overcome the deficiencies of bottom–up-based methods is offered by a discriminative machine learning model[11] called the conditional random field (CRF), which directly models the conditional probability of different class labels (such as wound and nonwound), given a set of images. Mathematically, the CRF model directly models the posterior distribution (the distribution of the labels, conditional on the observed image data) as a Gibbs distribution.[12] This conditional probability model depends on arbitrary nonindependent characteristics of the observations, in contrast to the Markov random field model, whose generative nature obliges us to model the joint distribution of the image data and the corresponding label field, and which therefore requires strict independent assumptions to make the model inference tractable. In addition, the CRF-based approach does not model the distribution for observed image data itself, as it is not utilized when performing classification. For these reasons, there has been increasing interests in recent years in solving image-labeling problems using the CRF model.[13–17] More important, the factorization graph-based definition allows the CRF model to incorporate features in different scales and from arbitrary regions of the image.[13,17] This characteristic provides increased flexibility to meet different needs from various object recognition tasks, especially for tasks requiring scale-invariant robustness.

This paper proposes a wound detection system to determine the boundaries of foot ulcers. For the wound area determination, we apply the associative hierarchical random field (AHRF) framework, which was proposed in Ref. 16. This model can be viewed as an extension of the CRF-based approach, which augments the robust P^N model (the most widely used CRF model in the computer vision area) by incorporating segment (also referred as superpixels)-based features as higher-order potential terms into the energy formulation. There are several reasons as to why we have chosen to utilize this AHRF
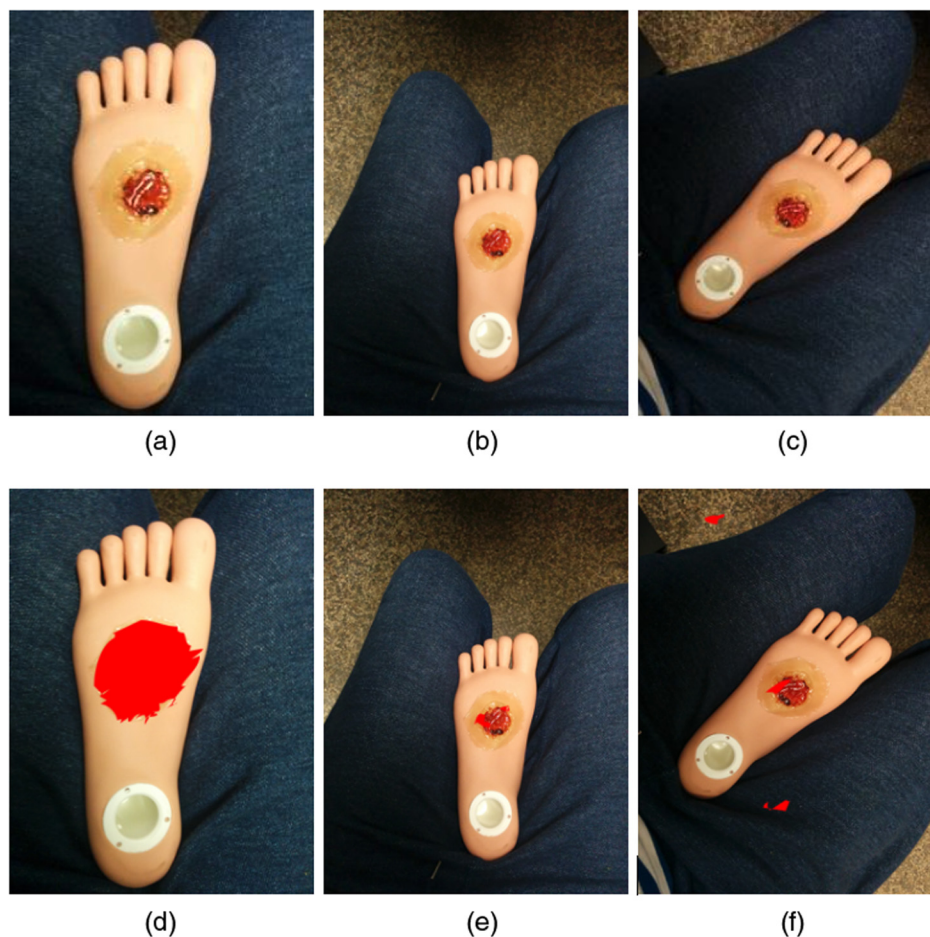
Fig. 1 An example of recognition failure by the SVM-based approach on moulage wound images of different scales (ranges) and illumination levels: (a)–(c) original wound images of the same moulage wound, (d)–(f) wound boundary determination results.

framework: (1) it provides a unification of the "top-down" and "bottom-up" approaches; (2) it allows the use of image features defined at any scale and over any arbitrary neighborhood regions, and (3) it can be solved efficiently using graph-cut-based move-making algorithms.[15,18] The wound boundary determination is completely automatic, i.e., requiring no human intervention, and can handle the recognition and wound boundary detection of wound images captured under different ranges and lighting conditions, which is a significant improvement from our previous work (our IEEE paper not issued yet). To evaluate the performance of this new wound recognition system, we use two different wound image datasets. The first image dataset is composed of images of moulage wounds placed on an artificial foot. The second dataset consists of images of real diabetic foot ulcers from recruited subjects at the Wound Clinic in UMass Medical School, Massachusetts. Images from both datasets had been captured at different ranges, illumination level, and viewpoints in order to better evaluate the robustness of the system.

The paper is organized as follows: Sec. 2 provides an overview of the foot ulcer image analysis system. Section 3 introduces basic formulation of the AHRF model. In Sec. 4, we describe in detail how this AHRF framework is applied to accomplish our foot ulcer boundary determination task; the experimental results will be presented in Sec. 5. Finally, Sec. 6 gives an overall conclusion and assessment of the proposed system.

2 Methodology Overview

We introduce the basic structure for our wound area determination approach based on the AHRF model. The entire algorithmic process is illustrated in Fig. 2. It can be seen from this figure that the complete system is divided into two subsystems: model training and wound recognition.

The wound classifier training process is shown in the left column of Fig. 2. The AHRF model is composed of several types of potential terms at different levels of image granularity: pixel-wise, pairwise, and superpixel-based terms (the pairwise terms are used to describe the relationship between two adjacent pixels). Hence, we first need to perform superpixel segmentation on the original images. Many different segmentation algorithms have been applied in the object recognition area.[5,8,19–21] In our system, we adopt the parallel version of the mean shift algorithm,[20,22] due to its good boundary adherence and efficient implementation.

The goal for our wound recognition system is to be able to accurately determine the wound boundary in images acquired under image acquisition conditions where illumination, range, and viewing angles can vary over reasonable ranges; images may also contain other background objects in the vicinity of the wound boundary. Therefore, we apply the texton map-based features, which have previously been shown to provide promising performance in object classification tasks in natural scene images.[23] These features are all required to be extracted densely (at each pixel location) and incorporated into the unary potential term in the AHRF model using the joint boost method.[24] For the pairwise potential terms, we apply the classical contrast-sensitive Potts potential form.[25] Next, the superpixel-based unary potential is also computed using a multiclass joint boost approach over the normalized histograms of multiple pixel-wise features. Finally, the pairwise potential terms are calculated at the superpixel level. More details about the feature extraction will be presented in Sec. 4.

For evaluating the ability of the CRF methods to recognize a wound and determine its boundary on a given set of images, the superpixel segmentation and feature extraction are the same as is used in the training process. We apply the learned textons to generate the texton map for each feature channels. Afterward, we evaluate the unary potential, pairwise potential, and segment-based potentials (if applicable) based on the model learned in the training process. Then we apply the CRF inference method to find the optimal labeling over the entire wound image.

3 Basic Knowledge of Associative Hierarchical Random Field Model

3.1 Conditional Random Field Basics

Consider an ordered set of variable $X = [X_1, X_2, \ldots X_n]$, where each variable X_i will be annotated by a label from a set L corresponding to the object classes. We write $y \in L^n$ for a labeling of X, where y_i refers to the labeling of the variable X_i. The random variables X and y are jointly distributed, but in the discriminative framework we construct a probabilistic model $P(y|X)$ to be estimated from M paired training instances $\{X^{(i)}, y^{(i)}\}_{i=1}^M$ and thus do not need to model the marginal $p(X)$. The neighborhood system N is defined by sets $N_i, \forall\ i \in V$, where N_i denotes the set of all neighbors of the variable X_i. A clique c is a set of random variables X_c, which is conditionally dependent on each other.[15] Any possible assignment of labels to the random variables will be called a labeling (denoted by y), which take its possible values from L^n. The labeling on the clique c is referred to as y_c. We use $V = \{1, 2, \ldots, n\}$ to refer to the set of valid vertices (or indices) of X. The posterior distribution $P(y|X)$ over the labeling of the CRF is a Gibbs distribution[12] and can be written as

$$P(y|X) = \frac{1}{Z} \exp \left[-\sum_{c \in C} \varphi_c(y_c) \right], \qquad (1)$$

where Z is a normalizing constant called the partition function, and C is the set of all cliques. The term $\varphi_c(y_c)$ is known as the potential function of the clique $c \subset V$, where $y_c = \{y_i : i \in c\}$. The corresponding Gibbs energy is given as

$$E(y) = -\log[P(y|X)] - \log Z = \sum_{c \in C} \varphi_c(y_c). \qquad (2)$$

Finding the most probable labeling is equivalent to solving the maximum *a posteriori* (MAP) problem. This optimal labeling y^* of the CRF is defined as

$$y^* = \arg\ \max_{y \in L} P(y|X) = \arg\ \min_{y \in L} E(y). \qquad (3)$$

According to Refs. 13, 14, and 26, labeling problems in computer vision area are typically formulated as a pairwise CRF whose energy can be written as the sum of unary and pairwise potentials[16]

$$E(y) = \sum_{i \in V} \varphi_i(y_i, \theta_u) + \sum_{i \in V, j \in N_i} \varphi_{ij}^P(y_i, y_j, \theta_p), \qquad (4)$$

where N_i is the set of neighbors of vertex i. The unary potential $\varphi_i(y_i, \theta_u)$ is computed independently for each pixel by a classifier that produces a distribution over the label assignment y_i
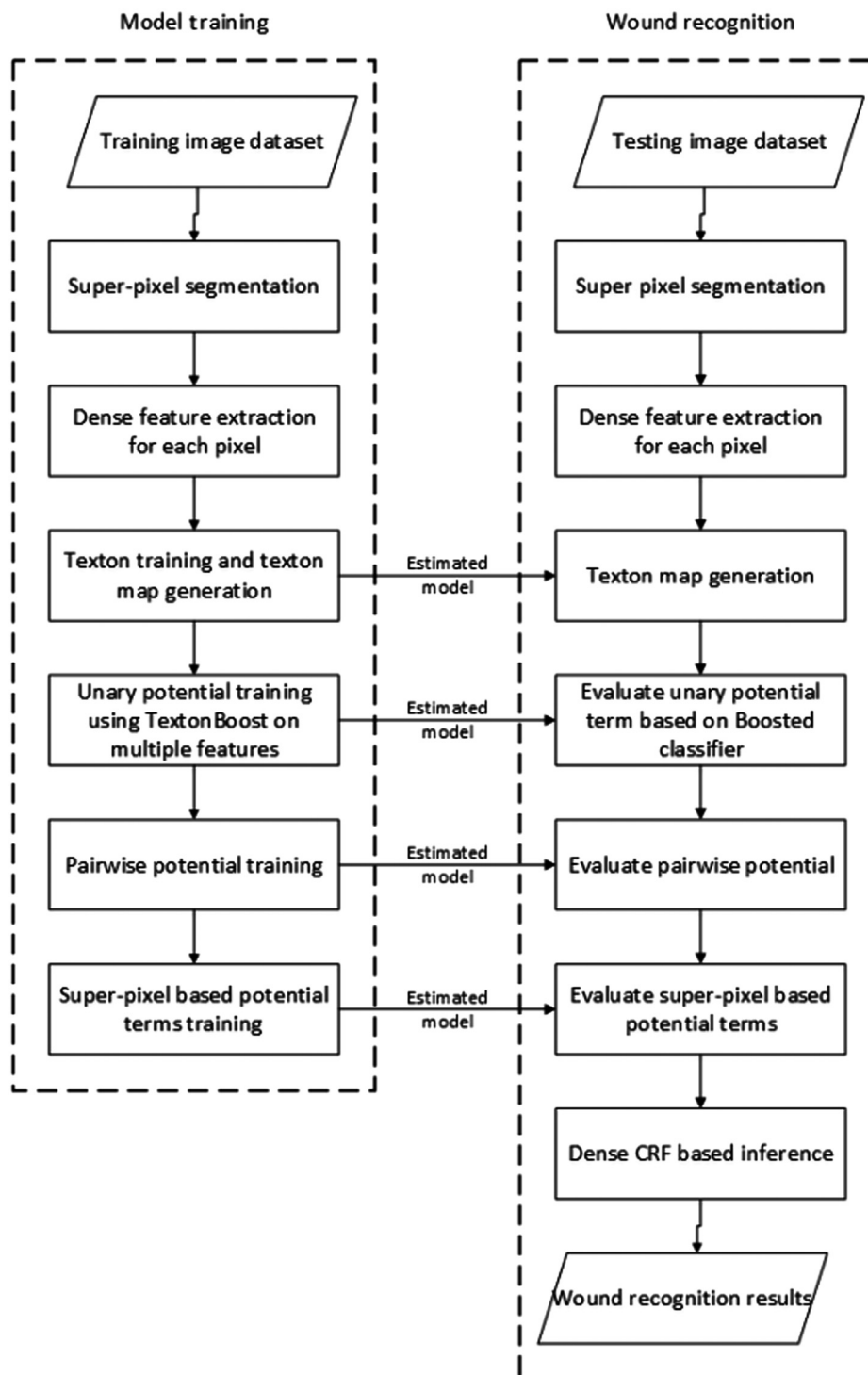
Fig. 2 Basic structure for wound recognition classifier based on AHRF model.

given image features (the features can be the pixel value or other features calculated over the neighborhood region). The pairwise potential $\varphi_{ij}^{P}(y_i, y_j, \theta_p)$ is computed between each pair of adjacent pixels in the image domain and is formulated to penalize the adjacent pixels being assigned different labels. Here, θ_u and θ_p are sets of parameters for the unary potential and pairwise potential, respectively. The details of the parameter estimation, also referred as the model learning, will be discussed in Sec. 4. The actual formulations to calculate the unary and pairwise potentials are task-independent. This flexibility allows us to incorporate different factors (color, edge, texture, or spatial position) into the CRF energy formulations.[13,17]

3.2 Associative Hierarchical Random Field Model

To include longer distance relationships within the image, higher-order potentials defined on superpixels or between pairs of superpixels are incorporated into the basic CRF models to

better describe the hierarchical connectivity. This method gives us an integration of the "top-down" and "bottom-up" approaches that are commonly used to overcome many problems in computer vision. To achieve top–down—bottom–up integration, an improved model, called the AHRF had been proposed in Ref. 16. Of practical importance is the fact that it has been shown that this model can be solved efficiently using graph-cut-based move-making algorithms mentioned earlier. It has also been proven that a new model generated by summing up two AHRFs is also an AHRF. This fact enables different potentials based on different features to be incorporated within the CRF model, while the model inference is still practical, permitting an efficient solution for the AHRF model. Here, we will introduce the AHRF model in details.

The AHRF model is defined in Eq. (5) by incorporating higher-order potentials defined on superpixels, in addition to the labeling, as formulated in Eq. (4). This extension was proven to be valid as pixels in the same superpixel have a high probability of being assigned to the same label. The energy of the higher-order random field is of the form that can be given as

$$E(x) = \sum_{i \in V} \varphi_i(x_i) + \sum_{i \in V, j \in N_i} \varphi_{ij}^P(x_i, x_j) + \sum_{c \in S} \varphi_c^h(x_c), \quad (5)$$

where S is a set of segments (or superpixels), given by one or more superpixel segmentation algorithms,[20,27] and $\varphi_c^h(x_c)$ are the higher-order potentials defined over the cliques. The higher-order potentials can be described as a robust P^N model as

$$\varphi_c^h(x_c) = \min_{l \in L} \left[\gamma_c^{\max}, \gamma_c^l + \sum_{i \in c} w_i k_c^l \Delta(x_i \neq l) \right], \quad (6)$$

where w_i is the weight of the variable x_i, and γ satisfies $\gamma_c^l \leq \gamma_c^{\max}, \forall l \in L$. The potential has a cost of γ_c^l if all pixels in the segment are assigned with label l. The pixels that are not assigned with the same label are penalized with a cost, which is expressed as $w_i k_c^l$, and the maximum cost of the potential is truncated to $\gamma_c^{\max}$. This framework supports the integration of higher-order potentials, based on superpixels at multiple scales of the image grid.

It has been proven in Ref. 16 that the higher-order P^N of Eq. (6) is equivalent to the cost of minimal labeling of a set of pairwise potentials defined over the same clique variables x_c and a single auxiliary variable $x_c^{(1)}$ that takes the values from an extended label set $L^E = L \cup \{L_F\}$. Here, L_F is a free label, meaning that there is no dominant label in this clique. Such a segment is said to be unassigned. Finally, we can formulate the framework to incorporate the pairwise dependencies between auxiliary variables as

$$E(x) = \sum_{i \in V} \varphi_i(x_i) + \sum_{i \in V, j \in N_i} \varphi_{ij}^P(x_i, x_j)$$
$$+ \min_{x^{(1)}} \left\{ \sum_{c \in S} \varphi_c[x_c, x_c^{(1)}] + \sum_{c,d \in S} \varphi_{cd}^P[x_c^{(1)}, x_d^{(1)}] \right\}. \quad (7)$$

These pairwise terms, defined over a higher-order clique grid, impose consistency between adjacent cliques. Then, the model in Eq. (7) can be generalized to a hierarchical framework, in which the relationship between layers takes the form as

$$\varphi_c[x_c, x_c^{(1)}] = \phi_c[x_c^{(1)}] + \sum_{i \in c} \phi_c[x_i, x_c^{(1)}]. \quad (8)$$

The weights for each node in the higher layer in $\phi_c(.)$ are proportional to the sum of the weights in the "base layer" belonging to the clique c. More generally speaking, the energy of the new hierarchical model is formulated as

$$E(x) = \sum_{i \in V} \varphi_i(x_i) + \sum_{i \in V, j \in N_i} \varphi_{ij}^P(x_i, x_j) + \sum_{x^{(1)}} E^{(1)}[x, x^{(1)}], \quad (9)$$

where the third term in this energy expression is recursively defined in Eq. (10).

$$E^{(n)}[x^{(n-1)}, x^{(n)}] = \sum_{c \in S} \varphi_c^P[x^{(n-1)}, x_c^{(n)}] + \sum_{c,d \in S} \varphi_{cd}^P[x_c^{(n)}, x_d^{(n)}]$$
$$+ \min_{x^{(n+1)}} E^{(n+1)}[x^{(n)}, x^{(n+1)}].$$
$$(10)$$

In Eq. (10), $x^{(0)} = x$ represents the state of the base level, and $x^{(n)}$ where $n \geq 1$ describes the state of auxiliary variables. The interlayer relationship between two layers of auxiliary variables can be described using a weighted robust P^N potential with the unary term $\varphi_c[x_c^{(n)}]$ and pairwise term as

$$\varphi_c(x_d^{n-1}, x_c^{(n)}) = \begin{cases} 0 & \text{if } x_c^{(n)} = L_F \cup x_c^{(n)} = x_d^{(n-1)} \\ w_d k^{x_c^{(n)}} & \text{otherwise} \end{cases},$$
$$(11)$$

where the weights are summed up over the base layer as $w_d = \sum_{j \in d} w_j$. As formulated in Eq. (11), $x_c^{(n)}$ is encouraged to take the label L_F if either most of its directly connected nodes from the lower layer [we can refer to these nodes as the children of $x_c^{(n)}$] are assigned with label L_F or these children are assigned with different labels.

4 Application of Associative Hierarchical Random Field Framework to Wound Area Determination Task

Several tasks are required to utilize the AHRF-based framework for computer vision tasks: (1) extracting suitable dense features, (2) incorporating these features into different potential terms, (3) learning the optimal parameters for each term on training image set, and (4) evaluating the trained model on new test data by applying appropriate inference algorithms.

Based on the definition of the AHRF framework described above, a set of potentials used in the object-class segmentation problem has also been presented in Ref. 17. This set consists of unary potentials defined on both pixels and superpixels, pairwise potentials between pixels and between superpixels, and connective potentials between different layers in the hierarchical graph structure. According to Refs. 16 and 17, there are two different ways to incorporate these features into the CRF model.

In the first method, we can further decompose the unary potential term into a weighted summation $\varphi(x) = \sum_c \lambda_c \xi_c(x_c)$, where $\xi_c(x)$ is a feature-based potential and λ_c is its weight. We need to utilize the joint boost approach for training each feature-based potential, then estimate the weights using local search scheme on a validation set. This training method turns out to be robust, but computationally demanding as well.

The second method for training a single unary potential term is implemented by combining multiple dense extracted features together. After extracting each feature over the image grid, we perform texton generation. The textons can be defined as the fundamental microstructure elements for the visual perception of texture patterns in images. These basic structures contain more meaningful information than do individual pixels. Textons have been widely used in computer vision for a wide range of tasks, including image analysis, object recognition, and text recognition.[17,23,28] The standard texton generation process usually consists of two steps: (1) filtering and (2) clustering (more details about the texton generation can be found in Ref. 17). As a result of the texton generation process, we have NM texton channels in total, where N is the number of types of features in total and M is the cluster center number for texton generation. Before performing the texture-layout filtering, we calculate the integral image (used for efficiently calculating the sum of pixel values in a rectangular region)[29] for each channel. Then, we extract the texture layout-based features, based on these NM texton channels. Finally, we perform the joint-boost approach to determine the final unary term only for one time. Weighing the strengths and weaknesses of the first and the second method, we chose to apply the second method.

According to Ref. 16, the potentials are defined over a three-tier hierarchy which is organized based on the types of nodes. A three-tier hierarchy provides adequate accuracy, and increasing the number of hierarchical levels beyond three tiers has been shown not to improve the accuracy noticeably. The nodes of each tier are pixels, segments, and supersegments, respectively. In our previous work (our IEEE paper, not issued yet), we had evaluated the different superpixel segmentation algorithms in the scenario of wound recognition quantitatively. As suggested in Ref. 9, the mean shift algorithm is applied to perform the superpixel segmentation. The reason for choosing mean shift algorithm is that the spatial and range resolution parameters[20] allow us to adjust the segmentation scales easily. In this case, we can utilize finer scale segmentation for the second-tier segments and supersegments for the third tier with coarser scale segmentation.

4.1 Features

In related works,[16,17] several different types of descriptors are used and evaluated for wounds, including textons-based shape filters, local binary patterns,[30] multiscale dense scale invariant feature transformation (SIFT),[19] opponent SIFT,[31] color distribution features, and histogram of oriented gradients.[32] The textons used here are defined as clustered 17-dimensional responses to 17 different filters (Gaussian, Gaussian derivative, and Laplacian filters at different scales). The local binary pattern is an eight-dimensional binary feature, in which each element represents a comparison of the pixel value of the center pixel with its eight neighbors. The SIFT feature contains the histogram of gradients of 4×4 cells quantized into eight bins. In this case, there will be eight-element features for each cell, with one element per bin. Hence, we need 128 (eight elements for each of the 16 cells) elements in total. The resulting 128-dimensional vector is normalized to the range from 0 to 1. Opponent SIFT is another version of the traditional SIFT and is based on the histograms of gradients for three channels in the chosen color space. Similar to the work in Ref. 16, we have generated a dictionary that contains 400 words for each type of descriptor, using the K-mean clustering method, followed by quantizing the local

distribution of the descriptors for each type based on its own dictionary. Hence, for each local patch (16×16 in our case), the final features will be represented by a 400-element vector that describes the distribution, a 400-bin histogram where each bin represents one word in the dictionary.

4.2 Unary Potentials from Pixel-Wise Features

Unary potentials from pixel-wise features are derived from TextonBoost,[17] which has allowed us to perform texture-based segmentation, at the pixel level, within the same framework. The features used for constructing these potentials are computed on every pixel of the image, and are, therefore, also called dense features. TextonBoost estimates the probability of a pixel taking on a certain label by boosting weak classifiers based on a set of texture layout responses. We have observed that textons are unable to discriminate between object classes of similar textures (e.g., between the wound tissue inside the wound boundary and the skin tissue adjacent to the wound). This has motivated us to extend the TextonBoost framework by boosting the classifiers defined on multiple dense features (such as color, textons, histograms of oriented gradients, and pixel location). The results in Ref. 16 show that the boosting of multiple features together results in a significant improvement in the accuracy of scene classification (note that the improvement from 72% in Ref. 17 to 82% in Ref. 16 has been achieved based on the same image dataset). The potentials are incorporated into the framework in the standard way as a negative log-likelihood as

$$\varphi(x_i = l) = -\log \frac{e^{H_l(i)}}{\sum_{l' \in L} e^{H_l(i)}} = -H_l(i) + K_i, \qquad (12)$$

where $H_l(i)$ is the AdaBoost classifier response for a label l and a pixel i and $K_i = \log \sum_{l' \in L} e^{H_l(i)}$ is a normalizing constant.

4.3 Histogram-Based Unary Potentials for Superpixels

Unary potentials are also defined over segments and supersegments. For many object recognition problems, the distributions of pixel-wise feature responses have been found to be more discriminative than any feature alone.[16] In other words, applying a statistical form of pixel-wise features, such as the histogram, over a neighborhood region will be more powerful for visual discrimination than using these individual features directly. The unary potentials of the auxiliary segment variables are estimated using multiclass JointBoost[33] over the normalized histograms of multiple clustered pixel-wise features. The learning process is the same as for the pixel-wise unary potential introduced earlier. The unary potential defined on superpixels is incorporated into the energy as

$$\varphi_c(x^{(1)} = l) = \lambda_s |c| \min[-H_l(c) + K_c, \alpha^h]$$
$$\varphi_c(x^{(1)} = L_F) = \lambda_s |c| \alpha^h, \qquad (13)$$

where $H_l(c)$ is the response given by the AdaBoost classifier to clique c taking on label l, α^h is a truncation threshold, and $K_i = \log \sum_{l' \in L} e^{H_{l'}(c)}$ is a normalizing constant.[16] In our case, the cost of pixel labels is different from that of the associated segment labels and is set to $k_c^l = [\varphi_c(x^{(1)} = L_F) - \varphi_c(x^{(1)} = l)]/ 0.1|c|$. It can be seen that at most 10% of the pixels are allowed

to take a label that is different from the segment label without changing the state of the segment to L_F.

4.4 Histogram-Based Pairwise Potentials between Superpixels

The pairwise terms on the pixel level $\varphi_{ij}^P(.)$ take the form of the classical contrast-sensitive Potts potentials as

$$\xi^P(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j \\ g(i, j) & \text{otherwise} \end{cases}. \tag{14}$$

In Eq. (14), the function $g(i, j)$ describes the edge information based on the pixel value difference between neighboring pixels[13] as

$$g(i, j) = \theta_p + \theta_v \exp(-\theta_\beta \| I_i - I_j \|), \tag{15}$$

where I_i and I_j are the color vectors of pixel i and j, respectively. This pairwise constraint encourages neighboring pixels in the image (having a similar color) to have the same label. More details can be found in Ref. 13.

We use the pairwise potential in the segment level defined in Eq. (16). This potential forces the superpixels with the same texture and color features to be assigned the same label. The term $g(c, d)$ is defined as $g(c, d) = \min(|c|, |d|) | h[x_c^{(1)}] - h[x_d^{(1)}]|_2^2$, where $h(.)$ is the normalized histogram for color values of a given segment,[16] and where the operator $|.|_2^2$ represents the Euclidean distance of histograms between two given neighboring segments, indexed as c and d

$$\xi_{cd}^P[x_c^{(1)}, x_d^{(1)}] = g \begin{cases} 0 & \text{if } x_c^{(1)} = x_d^{(1)} \\ g(c, d)/2 & \text{if } [x_c^{(1)} = L_F \,\&\, x_d^{(1)} \neq L_F] \cup [x_c^{(1)} \neq L_F \,\&\, x_d^{(1)} = L_F] \\ g(c, d) & \text{otherwise} \end{cases}. \tag{16}$$

4.5 Model Inference

As stated in Ref. 18, a CRF framework will not be useful without an efficient method for optimization. In the object recognition scenario, this optimization task is defined as finding the optimal label that can minimize the energy function defined in Eq. (7). In the artificial intelligence area, we also refer to the optimization of the CRF framework as inference. In Ref. 16, the suitability of various inference algorithms for AHRF has been analyzed. It is claimed that graph-cut-based move-making algorithms (such as alpha-expansion and alpha-beta swap methods) are the most suitable algorithms for solving the inference problem of minimizing the pairwise energy function defined over densely connected networks, which are commonly used in the computer vision field.

The move-making algorithms usually first assign an arbitrary initial solution (in our case, one solution is a label vector for all pixels in the image) where the goal is to find the optimal solution that minimizes the energy function defined in Eq. (7). Hence, a sequence of changes will be made to the initial solution toward the direction for energy minimization. According to Ref. 13, only the alpha-beta swap method can be directly applied to the AHRF framework. Other move-making algorithms require that the interlayer costs either form a metric[18] or are truncated convex. This property requires that the pairwise potential terms consists of two parts: (1) the "convex" part, which encourages the smoothness and (2) the "truncated" part, which ensures that the edge information will be preserved with respect to some ordering of the labels.[18] Thus, we decided to use the alpha-beta swap algorithm for the inference of the AHRF framework for wound boundary determination. The general idea of the alpha-beta swap algorithm is presented below. More details can be found in Ref. 18.

After the initial solution was randomly determined, we chose two labels (label alpha and label beta) from the label set. In our case, the wound recognition task was a binary labeling problem so there were only two labels: wound or nonwound. We can call either one the alpha label or the other one the beta label. Then we applied a max-flow-based algorithm[34] to find the optimal swap of alpha-beta label pairs for all pixels by treating this subtask as

a $s - t$ min-cut problem,[18] which was the basic binary form of a min-cut problem in graph theory, to find the optimal binary classification over a connected two-dimensional grid. This procedure was run until the approximate global optimal swap was identified.

5 Experimental Results

5.1 Experimental Setup

To evaluate the performance of the AHRF-based wound recognition system, we utilized two different wound image datasets. The first image dataset was composed of images of moulage wounds placed on an artificial foot. The second dataset consisted of images of actual diabetic foot ulcers from recruited subjects at the Wound Clinic in UMass Medical School. To better evaluate our system, the wounds in images of the first dataset were captured at different ranges, illumination levels, and viewing angles. Specifically, we collected 162 images of six moulage wounds for the first dataset. Twenty-seven images for each wound were captured, at three different ranges, three different viewing angles, and three different illumination conditions. In the second training dataset, 100 images were captured from 15 subjects where most of them were acquired using an image capture box, as described in our previous work.[9]

To evaluate the performance of the wound recognition over the entire dataset, we divided both datasets equally into 10 folders. Then a 10-folder validation method was carried out as follows. We performed the "train and test" operation for 10 rounds. In each round, we trained the model on nine folders and tested the model on the remaining folder. The average specificity and sensitivity were evaluated by combining the test results from 10 rounds. For the moulage image dataset, we first manually segmented the image into four different labels: (1) the wound, (2) gel which is the transparent material that surrounds the moulage wound, (3) the healthy skin, and (4) the background, as shown in Figs. 3(a) and 3(c). For all the images in the real-wound image dataset, we will segment each image into three different labels, which are identical to three of the moulage wound labels; eliminated is the surrounding gel label, as shown in Figs. 3(b) and 3(d).

(a)   (b)   (c)   (d)
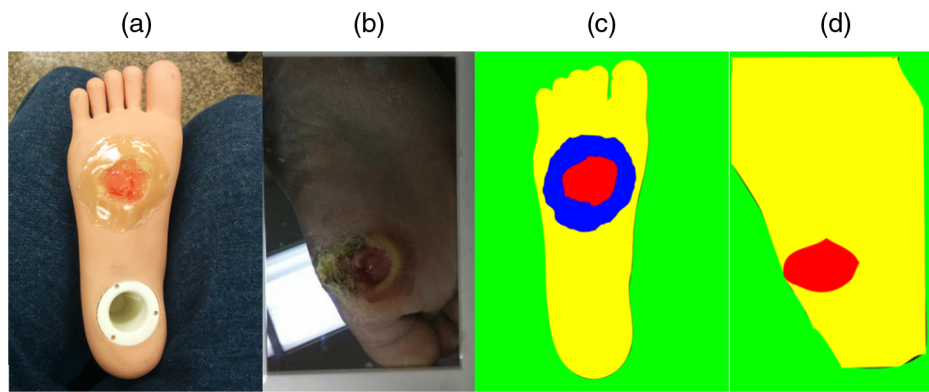
Fig. 3 Single sample of original image and ground truth from two dataset: (a) original image of the dataset 1, (b) original image of the dataset 2, (c) ground truth labeling of dataset 1 (green for background, yellow for healthy skin, blue for artificial gel, and red for wound), and (d) ground truth labeling of dataset 2 (the same labeling color fashion as for dataset 1 except for the absence of artificial gel category).

To evaluate the performance of the AHRF-based wound area determination approach more completely, we have also compared it to the performance obtained with two other CRF-based strategies introduced in Ref. 17. The first strategy (referred to as CRF model 1), which is described in Ref. 17, is based on an ordinary pairwise CRF model. The pixel-wise unary term and pairwise term are designed in the same way as the AHRF model. However, there are no potential terms based on superpixels. For the second reference strategy (referred as CRF model 2), we applied the fully connected CRF model introduced in Ref. 14. Compared to the other two models, the most distinguishing characteristic of the AHRF model, i.e., our model, is that each pair of pixels in the image is connected by an edge, which is further associated with the pairwise potential. We applied these models one by one independently to the same two datasets in the above-mentioned 10-folder validation approach.

The two most important parameters are the cluster center N for the texton generation and the boosting iteration number M for joint boost training scheme. To achieve better parameter estimation, we performed a grid search method to select the best parameter pair (N, M). We performed the AHRF model on the moulage image dataset using the above-described 10-folder validation method. The Matthews correlation coefficient (MCC)[35] results (the values of MCC score ranges from -1 to 1, where the higher the value, the better is the classification results) are shown in Table 1, whereas the wound recognition computation time evaluation results are shown in Table 2. The algorithm is implemented on the following PC: Intel quad-core CPU, 4GB RAM. We did not evaluate the training efficiency evaluation as the model training is assumed to be performed offline. In our current

Table 1 MCC results using different (N, M) parameter settings.

	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$	$N = 600$
$M = 1000$	0.393	0.438	0.471	0.523	0.532	0.538
$M = 2000$	0.469	0.498	0.547	0.596	0.602	0.606
$M = 3000$	0.550	0.582	0.617	0.648	0.651	0.655
$M = 4000$	0.598	0.632	0.668	0.699	0.694	0.703
$M = 5000$	0.707	0.738	0.769	0.813	0.816	0.821

Table 2 Wound recognition time using different (N, M) parameter settings (unit: seconds).

	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$	$N = 600$
$M = 1000$	10.2	10.3	10.8	11.0	11.2	11.3
$M = 2000$	18.1	18.5	19.9	20.2	22.0	22.4
$M = 3000$	27.7	28.2	30.0	30.9	31.3	33.0
$M = 4000$	38.8	39.9	41.3	41.5	42.9	42.9
$M = 5000$	46.3	47.2	49.3	50.1	50.5	51.2

PC-based environment, the training might consume >2 h if we set the iteration number as 5000.

Based on the results shown in Tables 1 and 2, the MCC gives the best performance when $N = 600$ and $M = 5000$; moreover, we can see that the MCC value increases as we increase the boosting iteration number (M), but obviously there is a corresponding increase in the computational time. On the other hand, when the cluster center number N becomes >400, there is no obvious improvement in the MCC result. In contrast, increasing the cluster center will substantially increase the computation burden for the model training. In conclusion, we set $N = 400$ and $M = 3000$ empirically for the best trade-off between accuracy and speed.

5.2 Wound Area Determination Results

Sample wound recognition results are shown in Figs. 4 and 5 for datasets 1 and 2, respectively. The specificity and sensitivity evaluation results for the three CRF models on the two datasets are given in Tables 3 and 4. Finally, the computational times for wound recognition are presented in Table 5. We can see that model 1 did not recognize the wound very well in a multiscale scenario, as it is a pairwise model in which the pairwise potential terms have only been evaluated on pairs of pixels in the same clique. Model 2 outperformed model 1 on wound recognition accuracy as it generated the pairwise potentials on each pair of pixels in the image. In this case, the long-range connections are incorporated into the CRF formulation. The AHRF model provides even better wound recognition performance than

Fig. 4 Samples of wound recognition results on dataset 1. (a) The original images; images 1 to 3 and 4 to 6 represent the two different wound simulations in different scales and viewpoints, respectively. (b)–(d) Wound recognition results provided by CRF models 1 and 2 and AHF model; the wound areas are labeled with red color.

model 2, especially when dealing with images of the same wound captured from different ranges (scales), viewing angles, and illumination conditions, due to its hierarchical structure involving superpixel-based higher-order potential terms. As mentioned earlier,[16] the potentials defined over a three-tier hierarchy provide the best trade-off between the time performance and recognition performance, although the hierarchy can be extended indefinitely. It is also found that the wound boundary recognition accuracy reaches a plateau when the number of hierarchy level is increased beyond three. However, the AHRF model requires longer computing times than those of the other two models due to the superpixel segmentation required and the increased number of potential terms to be evaluated, as can be observed from the computation times given in Table 5.

6 Discussion and Conclusion

An automatic wound boundary determination system for foot ulcer images has been presented in this paper. We tracked 15 patients in the Wound Clinic at UMASS over a 2-year period resulting in 100 high-resolution foot ulcer images. To better evaluate the robustness of our system, we also designed another dataset with images of moulage wounds, captured at different ranges, illumination levels, and viewing angles. We utilized the AHRF framework as the wound recognition model in our system. Higher-order potentials defined on superpixels or between pair of superpixels were incorporated into the basic CRF models to better describe the connectivity using a hierarchical structure. Therefore, the proposed wound boundary determination method is expected to be more robust when the wound image capture range, illumination, and angles are variable. To apply AHRF framework, we first performed superpixel segmentation using the mean shift algorithm. Second, we generated texton maps densely (for each pixel position) for several well-known features, and third, we incorporated these feature maps into both the pixel-wise and segment-wise unary potential terms using the joint boost method (the parameters for each term were learned at the same time). For the pairwise potential term, we applied the classical contrast-sensitive Potts form. Finally, the optimal label inference was performed by applying the alpha-beta swap method.

Fig. 5 Samples of wound recognition results on dataset 2. (a) The original images; images 1 to 6 represent three different wounds imaged in different scales, viewpoints, and illumination, respectively. (b)–(d) Wound recognition results provided by CRF models 1 to 3; the wound areas are labeled with red color.

Table 3 Wound recognition specificity using different CRF models on our two datasets.

	CRF model 1	CRF model 2	AHRF model
Dataset 1	0.927	0.984	0.992
Dataset 2	0.898	0.911	0.955

Table 4 Wound recognition sensitivity using different CRF models on our two datasets.

	CRF model 1	CRF model 2	AHRF model
Dataset 1	0.674	0.767	0.844
Dataset 2	0.618	0.703	0.769

Table 5 Wound recognition computation time using different CRF models on our two datasets (unit: seconds).

	CRF model 1	CRF model 2	AHRF model
Dataset 1	36.7	30.9	57.4
Dataset 2	37.4	35.9	60.3

To evaluate the AHRF-based binary wound classification system, we compared its performance to the performance of two other CRF-based classification strategies, which also have been widely used in object recognition research. Based on the experimental results, we found that the AHRF framework provided the best wound recognition accuracy, especially in dealing with images of the same wound captured at different ranges (scales), viewing angles, and illumination conditions, due to its hierarchical structure involving superpixel-based higher-order potential terms. However, the performance enhancement required more parameters to be estimated for more potential terms. As a result, the wound recognition time for this model was longer than for the other two CRF strategies.

The results indicate that chronic wounds can be correctly located in an image, and the wound boundary determined, without requiring tightly controlled range and lighting conditions. This implies that wound images can be captured and correctly processed in the lighting conditions likely to be encountered in a clinic, which broadens its clinical utility significantly.

There are a number of directions for future research work. First, to further improve the robustness of the wound boundary determination, we need to expand the diversity of the real wound images in the database, in terms of wound type, shape, color composition, surrounding tissues, skin color, and texture. To convert pixel-based area measures to actual units (e.g., square millimeter), image calibration should be done first to determine the range ratio (e.g., square millimeter per pixel). Another potential direction might be to improve the efficiency of the CRF-based approach. The results presented in our paper show that the computational time of the proposed approach is nearly 60 s even when implemented on powerful PCs. Owing to the iterative nature, the potential evaluation step is the computationally most expensive part. As described in Sec. 5, the iteration number has to be >3000 to acquire near-optimal results. Hence, the best option to reduce the computational burden may be to remove the extraction of less effective features (feature selection), which may require more detailed evaluation of the effectiveness of features. Last, although this paper compares AHRF with two other CRF models, we have not carried out a similar comparison to deep learning methods, such as convolutional neural

networks. However, based on what is required to obtain a good performance with machine learning, we can make some general observations: (i) when working with a small number of wound images, AHRF is likely to outperform deep learning, but as the number of wound images increases, the AHRF performance is likely to plateau; (ii) as deep learning in principle has many more trainable parameters as compared to AHRF, deep learning is likely to outperform AHRF when a large number of wound images is available.

Disclosures

The authors have no relevant financial interests in the article and no other potential conflicts of interest to disclose.

References

1. L. Wang et al., "Smartphone-based wound assessment system for patients with diabetes," *IEEE Trans. Biomed. Eng.* **62**(2), 477–488 (2015).
2. L. Wang et al., "Wound image analysis system for diabetics," *Proc. SPIE* **8669**, 866924 (2013).
3. T. D. Jones and P. Plassmann, "An active contour model for measuring the area of leg ulcers," *IEEE Trans. Med. Imaging* **19**(12), 1202–1210 (2000).
4. F. Veredas, H. Mesa, and L. Morente, "Binary tissue classification on wound images with neural networks and Bayesian classifiers," *IEEE Trans. Med. Imaging* **29**(2), 410–427 (2010).
5. H. Wannous, Y. Lucas, and S. Treuillet, "Enhanced assessment of the wound-healing process by accurate multiview tissue classification," *IEEE Trans. Med. Imaging* **30**(2), 315–326 (2011).
6. B. Song and A. Sacan, "Automated wound identification system based on image segmentation and artificial neural networks," in *IEEE Int. Conf. Bioinf. and Biomed.* (2012).
7. K. Wantanajittikul, "Automatic segmentation and degree identification in burn color images," in *4th Biomed. Eng. Int. Conf.*, pp. 169–173 (2011).
8. M. Kolesnik and A. Fexa, "Segmentation of wounds in the combined color-texture feature space," *Proc. SPIE* **5370**, 549–556 (2004).
9. L. Wang et al., "Area determination of diabetic foot ulcer images using a cascaded two-stage SVM based classification," *IEEE Trans. Biomed. Eng.* **64**, 2098–2109 (2017).
10. D. Nouri et al., "Colour and multispectral imaging for wound healing evaluation in the context of a comparative preclinical study," *Proc. SPIE* **8669**, 866923 (2013).
11. S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vision* **68**(2), 179–201 (2006).
12. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**(6), 721–741 (1984).
13. L. Ladický et al., "Associative hierarchical CRFs for object class image segmentation," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 739–746 (2009).
14. P. Krahenbuhl, V. Koltun, and P. Krahenbuhl, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Adv. Neural Inf. Process. Syst. 24 (Proc. NIPS)*, no. 4, pp. 1–9 (2011).
15. C. Sutton and A. McCallum, "An introduction to conditional random fields for relational learning," in *Introduction to Statistical Relational Learning*, L. Getoor and B. Taskar, Eds., pp. 93–126, MIT Press, Cambridge (2007).
16. L. Ladický et al., "Associative hierarchical random fields," *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(6), 1056–1077 (2014).
17. J. Shotton et al., "{TextonBoost} for image understanding: multi-class object recognition and segmentation by jointly modeling appearance, shape and context," *Int. J. Comput. Vision* **81**(1), 2–23 (2009).
18. Y. Boykov et al., "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001).
19. B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *IEEE 12th Int. Conf. Comput. Vision* (2009).
20. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002).
21. R. Achanta, A. Shaji, and K. Smith, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012).
22. C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," in *Object Recognit. Supported User Interact. Serv. Rob.*, Vol. **4**, pp. 150–155 (2002).
23. J. Malik et al., "Contour and texture analysis for image segmentation," *Int. J. Comput. Vision* **43**(1), 7–27 (2001).
24. R. E. Schapire and Y. Singer, "BoosTexter: a boosting-based system for text categorization," *Mach. Learn.* **39**, 135–168 (2000).
25. P. Kohli, L. Ladický, and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," *Int. J. Comput. Vision* **82**(3), 302–324 (2009).
26. X. He, R. S. Zemel, and M. A. Carreira-Perpinan, "Multiscale conditional random fields for image labeling," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, Vol. **2**, pp. 695–702 (2004).
27. Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(8), 800–810 (2001).
28. T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision* **43**(1), 29–44 (2001).
29. F. C. Crow, "Summed-area tables for texture mapping," in *Proc. 11th Annu. Conf. Comput. Graphics Interact. Tech.*, Vol. **18**, pp. 207–212 (1984).
30. B. S. Manjunath et al., "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.* **11**(6), 703–715 (2001).
31. H. Fujiyoshi, "Gradient-based feature extraction -SIFT and HOG," *Computer* (Long. Beach. Calif.) **107**(206), 211–224 (2007).
32. D. Sen and S. K. Pal, "Gradient histogram: thresholding in a region of interest for edge detection," *Image Vision Comput.* **28**(4), 677–695 (2010).
33. A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: efficient boosting procedures for multiclass object detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 762–769 (2004).
34. Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Comput. Vision* **70**(2), 109–131 (2006).
35. B. W. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta* **405**(2), 442–451 (1975).

Lei Wang is a research software engineer in the Image Search Dept., Google Inc. He received his PhD in electrical and computer engineering from Worcester Polytechnic Institute (WPI), Massachusetts, in 2016, and then worked in the Ultrasound Imaging R&D Department for Philips until January 2018. His research areas include machine learning-based image ranking algorithm design, medical imaging and image processing, and the design of a user-friendly image analysis for chronic wounds using a smartphone.

Peder C. Pedersen is a professor emeritus in ECE at WPI. He received his PhD in bioengineering from the University of Utah in 1976, and then has been a faculty in ECE at Drexel University, Pennsylvania, before coming to WPI in 1987. His research areas include elastography methods for imaging of the Young's modulus in soft tissues, development of a low-cost, portable personal ultrasound training simulator, and the design of user-friendly image analysis for chronic wounds using a smartphone.

Emmanuel Agu is a professor of computer science at WPI. He received his PhD in electrical and computer engineering from the University Massachusetts Amherst in 2001. He has been involved in research in mobile and ubiquitous computing, computer graphics, and imaging for over 20 years. Currently, he is working on mobile health projects focusing on smartphone wound image analysis, and the discovery of smartphone biomarkers for traumatic brain injury and infectious diseases.

Diane Strong: Biography is not available.

Bengisu Tulu is an associate professor in the Foisie Business School at WPI. She is one of the founding members of the Healthcare Delivery Institute at WPI. She received her PhD in management of information systems and technology from Claremont Graduate University, California. Her research interests include development and implementation of health information technologies and the impact of these implementations on healthcare organizations and consumers.