

Segmentation and classification of burn images by color and texture information

Begoña Acha

Carmen Serrano

José I. Acha

Área de Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingenieros
University of Seville
Camino de los Descubrimientos s/n
41092 Sevilla, Spain

Laura M. Roa

Grupo de Ingeniería Biomédica
Escuela Técnica Superior de Ingenieros
University of Seville
Camino de los Descubrimientos s/n
41092 Sevilla, Spain

Abstract. In this paper, a burn color image segmentation and classification system is proposed. The aim of the system is to separate burn wounds from healthy skin, and to distinguish among the different types of burns (burn depths). Digital color photographs are used as inputs to the system. The system is based on color and texture information, since these are the characteristics observed by physicians in order to form a diagnosis. A perceptually uniform color space ($L^*u^*v^*$) was used, since Euclidean distances calculated in this space correspond to perceptual color differences. After the burn is segmented, a set of color and texture features is calculated that serves as the input to a Fuzzy-ARTMAP neural network. The neural network classifies burns into three types of burn depths: superficial dermal, deep dermal, and full thickness. Clinical effectiveness of the method was demonstrated on 62 clinical burn wound images, yielding an average classification success rate of 82%. © 2005 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1921227]

Keywords: color images; burn; image segmentation; burn classification.

Paper 04076 received May 12, 2004; revised manuscript received Jan. 11, 2005; accepted for publication Jan. 11, 2005; published online May 11, 2005.

1 Introduction

For a successful evolution of a burn injury it is essential to initiate the correct first treatment.¹ To choose an adequate one, it is necessary to know the depth of the burn, and a correct visual assessment of burn depth highly relies on specialized dermatological expertise. As the cost of maintaining a burn unit is very high, it would be desirable to have an automatic system to give a first assessment in all the local medical centers, where there is a lack of specialists.^{2,3} The World Health Organization demands that, at least, there must be one bed in a burn unit for each 500 000 inhabitants. So, normally, one burn unit covers a large geographic extension. If a burn patient appears in a medical center without burn unit, a telephone communication is established between the local medical center and the closest hospital with burn unit, where the nonexpert doctor describes subjectively the color, shape, and other aspects considered important for burn characterization. The result in many cases is the application of an incorrect first treatment (very important for a correct evolution of the wound), or unnecessary displacements of the patient, involving high sanitary cost and psychological trauma for the patient and family.

With the fast advances in technology, computer aided diagnosis (CAD) systems are gaining widespread acceptance. However, nowadays, the research in the field of skin color images is developing slowly due to the difficulty of translating human color perception into objective rules, analyzable by a computer. Generally speaking, one can find two main applications about skin color image processing in the literature:⁴

the assessment of the healing of skin wounds or ulcers,⁵⁻⁹ and the diagnosis of pigmented skin lesions such as melanomas.¹⁰⁻¹⁵ The analysis of lesions involves more traditional image processing techniques such as edge detection and object identification, as well as an analysis of the color, irregularity, and shape of the segmented lesion. In wound analysis, the analysis of the colors within the wound site is often more important than the detection of the wound border or the calculation of its area. Particularly, in the case of burn depth determination, focusing on the shape of the burn is irrelevant for predicting its depth. The main characteristics for this purpose are color and texture information, as they are the features observed by physicians in order to give a diagnosis.

Automatic burn wound diagnosis is still a largely unexplored field. In the related bibliography, one can find that there is a tendency to investigate objective methods for determining the depth of the burn in order to reduce the subjectivity and the high experience requirement that visual inspection demands. Some research into the relationship between depth and superficial temperature¹⁶ has been developed. There are also other works trying to evaluate burn depth by using thermographic images,¹⁷ infrared and ultraviolet images,¹⁸ radioactive isotopes¹⁹ and laser Doppler flux measurements²⁰

On the other hand, there is hardly bibliography about burn depth determination by visual image analysis and processing. Although some research groups apply segmentation algorithms to burn images,^{5,7,8,21,22} they try to give an assessment of the healing of the burn, so they focused on calculating differences among several aspects such as area, shape, and appearance in order to give a prediction of the healing evolu-

Address all correspondence to Begoña Acha. Tel: +34-954487333; E-mail: bacha@us.es



Fig. 1 Different appearances that could present a burn: (a) superficial dermal (blisters), (b) superficial dermal (red), (c) deep dermal, (d) full thickness (beige), (e) full thickness (brown).

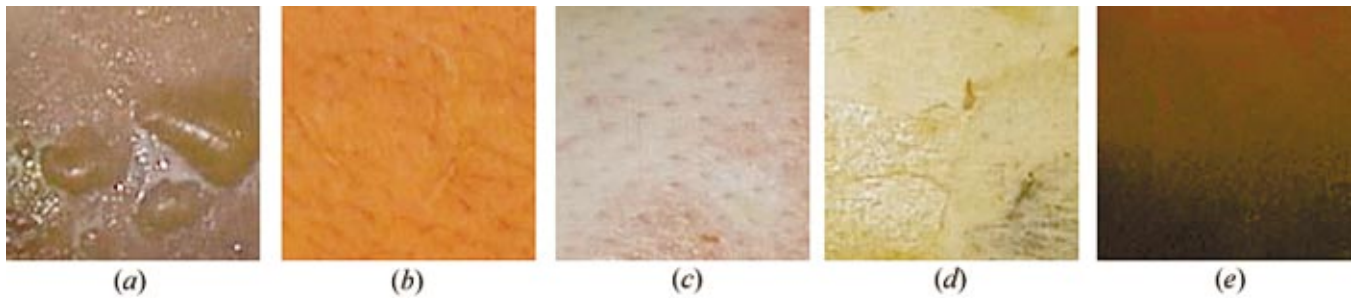


Fig. 4 Examples of the different 49×49 burn images used to train the classifier: (a) superficial dermal (blisters), (b) superficial dermal (red), (c) deep dermal, (d) full thickness (beige), (e) full thickness (brown).

tion of the wound. To our knowledge, only the group of Afromowitz et al.^{21,22} tries to give a diagnosis of the burn depth. From this assessment, they estimate the number of days that the wound will take to heal. They measure the optic reflectivity in the red, green, and infrared bands, hypothesizing that it is highly correlated with burn healing time, and they form a false color image that indicates the time of healing, or equivalently, the depth of the burn. The main disadvantage of the method is the complexity and cost of the image acquisition system (video camera, filter wheel, motor driver, etc.).

The main contribution of this work is the design of a clinically feasible system for automatic burn wound classification based on visual digital images. First, a protocol for the standardization of the burn image acquisition was designed. This first step was required due to the novelty of the application. Second, a new segmentation algorithm is proposed, which has been proven effective in segmenting burn wound images. Third, once the burnt part is segmented, representative color and texture descriptors are extracted from it. Finally, a neural network classifier processes these descriptors to give an estimation of the burn depth.

2 Materials and Methods

2.1 Burn Characterization

There are three main types of burn wounds.¹ (1) *Superficial dermal burn*: when the epidermis and part of the dermis are destroyed. The presence of blisters (usually brown color) and/or a bright red color characterize it. It is painful. (2) *Deep dermal burn*: it is characterized by its pink-whitish color. (3) *Full-thickness burn*: all the skin thickness is destroyed and skin grafts are needed. A beige-yellow or a dark brown color characterizes it. It is not painful.

Although a burn wound is classified in three classes, it can present five different appearances. (A) *Blisters*: they are superficial dermal burns with a bright texture and a rose-brown color. (B) *Bright red*: they are superficial dermal burns with bright red colors and wet appearance. (C) *Pink-white*: they are deep dermal burns with a dotted appearance. (D) *Yellow-beige*: first appearance of full-thickness burns. (E) *Brown*: second appearance of full-thickness burns. Examples of each appearance are shown in Fig. 1.

2.2 Image Acquisition and Calibration

The image acquisition was carried out by means of a digital photographic camera, the Canon EOS 300D (Canon Inc., Tokyo, Japan). Any nonspecialized person should be able to acquire data from the patient, because it is not possible to have an expert in each center. A digital photographic camera is easy to utilize and people are used to them.

The problems we found that had to be solved when using a digital photographic camera for this application are explained in the following subsections.

2.2.1 Illumination influence

The most important source of information for our system in order to classify burn depths is color, which is extremely influenced by the illumination. In hospitals the lighting conditions can change depending on the room where the patient is. Then, measured pixel values depend on the illuminants and with multiple illuminants the measured values cannot be ac-

curately converted to a known color space without some additional information. Therefore, a study about the influence of the different sources of illumination is needed. To perform this study, we photographed the Macbeth ColorChecker DC chart (Gretag-Macbeth GmbH, Martinsried, Germany) under three different illuminations: in a darkroom with the built-in flash (guide number=13 m at ISO 100), in a darkroom with fluorescent light, and in a room under diffused sunlight. Under these three different situations, we fixed the ISO speed to 100, the f stop (A_v) to 20 and we varied the exposure time (T_v). We define that the exposure time is optimum under a particular illuminant when it is the maximum time without saturating any channel. The ratio between the exposure times will give us the influences of the different sources of light. The optimum exposure times were 1/200, 0.6, and 1.6 s for the flash, sunlight, and fluorescent light, respectively. That means that the flash is 320 times stronger than the fluorescent and 120 times stronger than the sunlight. In other words, if we choose $T_v = 1/200$ and 8 bits per color component, the fluorescent light will not influence even the least significant bit and the sunlight will influence the two least significant bits. In fact, we took a photograph under both fluorescent and sunlight illuminations with this parameter ($T_v = 1/200$) and only these two least significant bits had values different to 0.

We can conclude that the xenon flash illumination is sufficiently strong to dominate illumination. That is an important result because in this way we only have to calibrate the images once for each camera, and not for each room where patients are treated.

2.2.2 Calibration

An additional problem we encountered is that manufacturers normally do not publish either the red (R), green (G), blue (B) primaries of the camera or the color temperature of the flash. Therefore we need to determine in some way a transformation matrix to convert from measured RGB coordinates to a device-independent color representation system.

For this purpose, we find the matrix transformation between RGB and CIE (Commission Internationale de l'Eclairage XYZ (device-independent color space). In the literature there are many transformation matrices from RGB to XYZ color space, but they are defined for specific illuminants (D65, D50, etc.) and specific RGB primaries (CCIR Rec. 709, FCC-NTSC, etc.).²³ We have developed a calibration method based on the Macbeth ColorChecker DC chart, which is specifically designed for calibration of digital cameras. The Macbeth ColorChecker DC chart has 240 color chips and it is supplied with data giving the CIE XYZ chromaticity coordinates of each chip under D50 illuminant. The 240 chips occupy an area of 12 cm×20 cm. Our method finds the transformation matrix from RGB under unknown illuminant to XYZ under D50, and corrects the nonuniformity of the illumination as well as the spatial nonuniformity of the camera sensitivity. This algorithm iteratively performs the following steps:

1. Without correcting the illumination profile and using only three color patches, we calculate the initial matrix M_1 that converts from RGB under an unknown illuminant to XYZ under D50.
2. In the i 'th step, using the 240 color patches in the chart

and the matrix M_{i-1} , we calculate the profiles, $P_{R,i}(x,y)$, $P_{G,i}(x,y)$, and $P_{B,i}(x,y)$, so that, for each patch, the R , G , B corrected with the profiles and multiplied by M_{i-1} are the X , Y , Z values specified by the manufacturer of the color chart. That is, for each patch k in the position (x_k, y_k) the following equation is performed:

$$\begin{bmatrix} P_{R,i} \\ P_{G,i} \\ P_{B,i} \end{bmatrix} = \begin{bmatrix} 1/R(x_k, y_k) \\ 1/G(x_k, y_k) \\ 1/B(x_k, y_k) \end{bmatrix} (M_{i-1})^{-1} \begin{bmatrix} X_k \\ Y_k \\ Z_k \end{bmatrix}. \quad (1)$$

3. We calculate the three fourth order surfaces, $P'_{R,i}(x,y)$, $P'_{G,i}(x,y)$, and $P'_{B,i}(x,y)$, that match best the profiles $P_{R,i}(x,y)$, $P_{G,i}(x,y)$, and $P_{B,i}(x,y)$ calculated in step 2. Previously, we have experimentally determined that a fourth order surface adequately approximates the sensitivity of the camera and the nonuniformity of the flash illumination altogether.
4. Using this profile, we calculate the matrix M_i that best maps the R , G , B values into the X , Y , Z values specified for all the patches in the color chart. To determine this optimum M_i the following mean square error is minimized:

$$\varepsilon^2 = \frac{1}{240} \sum_{k=1}^{240} (X_{t_k} - X_k)^2 + (Y_{t_k} - Y_k)^2 + (Z_{t_k} - Z_k)^2, \quad (2)$$

where X_{t_k} , Y_{t_k} , and Z_{t_k} are the X , Y , and Z values of the k 'th color patch, in the position (x_k, y_k) , specified by the manufacturer.

5. Repeat from step 2 until the mean square error ε begins to grow.

It must be emphasized that the matrix M is the product of two matrices: the transformation from RGB to XYZ under an unknown illuminant and the linear transformation to perform the chromatic adaptation from an unknown illuminant to D50. The matrix obtained with the proposed method is

$$M = \begin{bmatrix} 45 & 60 & -19 \\ 24 & 93 & -23 \\ 3 & 37 & 39 \end{bmatrix}$$

when the R , G , B values are normalized to one. It should be noted that this matrix M is specific for each camera, so calibration should be performed for every camera used.

2.2.3 Acquisition protocol

The third problem consists of fixing the acquisition protocol so that the photographs are useful for diagnosis. After fixing it we have validated its suitability.

The acquisition protocol was developed by an interdisciplinary group formed by burn specialized physicians and technicians.²⁴ The main points of the acquisition protocol were the following: distance between camera and patient should be about 40–50 cm (to fix this parameter, physicians carried out a careful analysis of photographs taken of different burn wounds from different distances; in the end, they chose 40–50 cm because they could distinguish texture from this distance and, at the same time, they usually had a global vi-

sion of the burn), healthy skin should appear in the image when possible, the background should be a green/blue sheet (the ones used in hospitals, because as the blue/green color is so different from the skin colors, the background can be easily rejected by the segmentation algorithm), the flash must be on and the camera should be placed parallel to the burn. The parameters of the camera were set to: ISO speed 100, exposure time 1/200 s and aperture (f stop) 20.

In order to validate the acquisition protocol, a survey was done.^{24,25} For this survey, 38 photographs of all etiologies, locations, and characteristics of the most frequent lesions were taken following the specified protocol. They were presented to a panel of 12 experts in burn diagnosis. The experts had to answer about the certainty in diagnosis (1–5): 1=minimal, 3=moderate, 5=maximum, certainty. A mean of 4.26 in sureness in diagnosis and 84.6% of diagnostic accuracy was answered, whereas diagnostic accuracy of a trained plastic surgeon when looking live at the same 38 burn wounds was 84.3%.

2.3 Burn Wound Segmentation

The segmentation approach used here is a supervised pixel-based algorithm based on measures in the CIE $L^*u^*v^*$ color coordinate space. $L^*u^*v^*$ and $L^*a^*b^*$ color representation systems are called uniform systems because Euclidean distances between colors measured in these spaces are very much correlated with color differences according to human perception. They are particularly useful in color image segmentation of natural scenes using histogram-based techniques, in which our method is included. They are slightly different because of the different approaches to their formulation. Nevertheless, both spaces are equally good in perceptual uniformity and provide very good estimates of color difference (distance) between two color vectors.²³ Therefore, we could have chosen any of these two spaces, but we preferred the $L^*u^*v^*$ one, because the color components a^* and b^* do not depend on the luminance, and it is known that color perception is strongly influenced by the luminance.²⁶

The following steps show the scheme proposed:

2.3.1 Selection of a small region in the burn wound by the user and preprocessing of the image

For a nonexpert physician (in fact, for most of the people) it is easy to differentiate burnt skin from normal one. Therefore, the burn wound will be segmented using the color information of a 5×5 pixel area around the point that the user selects with the mouse.

Before segmenting the image, it is convenient to preprocess it in order to get more homogeneous regions eliminating noise and small structures. To perform this task, an anisotropic diffusion is applied to the color image.^{27,28} The aim of the diffusion is to make the regions more homogeneous but preserving the edge information. In order to perform the anisotropic diffusion, the approach of separating the diffusion of the chromatic and achromatic information was followed²⁸ as is shown in Fig. 2. First, the image is converted into $L^*u^*v^*$ color coordinate system according to²³

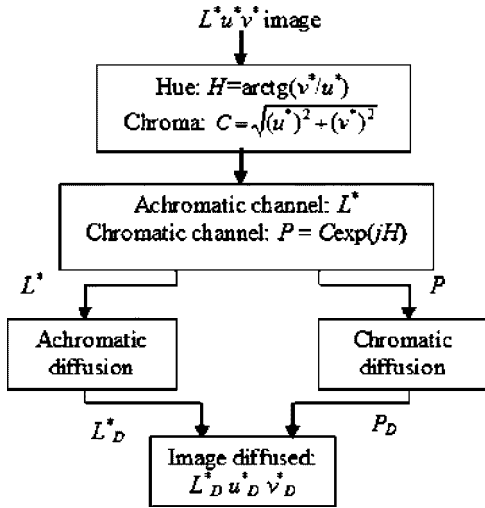


Fig. 2 Diffusion filtering separating chromatic and achromatic information.

$$L^* = \begin{cases} 116 \left(\frac{Y}{Y_0}\right)^{1/3} - 16, & \text{if } \frac{Y}{Y_0} > 0.008856 \\ 903.3 \left(\frac{Y}{Y_0}\right), & \text{otherwise} \end{cases} \quad (3)$$

Computation of u^* and v^* involves intermediate u' , v' , u'_0 , and v'_0 quantities defined as

$$u' = \frac{4X}{X + 15Y + 3Z}, \quad (4)$$

$$v' = \frac{9Y}{X + 15Y + 3Z}.$$

Finally,

$$u^* = 13L^*(u' - u'_0), \quad (5)$$

$$v^* = 13L^*(v' - v'_0).$$

Y_0 , u_0 , and v_0 correspond to the white reference point, which depends on the illuminant (D50 after the calibration).

From these coordinates, the hue and chroma components are calculated as $H = \arctan(v^*/u^*)$ and $C = \sqrt{(u^*)^2 + (v^*)^2}$, respectively. A complex quantity is calculated that relates the hue and the chroma as $P = C \exp(jH)$.

The achromatic anisotropic diffusion, applied to L^* , is carried out by means of the discrete formulation²⁷ of the partial differential equation

$$\frac{\partial}{\partial t} L^*(x,y,t) = \text{div}[\alpha(x,y,t) \nabla L^*(x,y,t)], \quad (6)$$

where div and ∇ denote the divergence and the gradient operators, respectively, and $\alpha(x,y,t)$ is a monotonically decreasing function of the image gradient magnitude called the *conductance coefficient* and is given by

$$\alpha(x,y,t) = \frac{1}{1 + \left(\frac{|\nabla L^*(x,y,t)|}{\gamma_p}\right)^2}. \quad (7)$$

The *diffusion constant* γ_p was selected as the 5% of the maximum value of $|\nabla L^*(x,y,t)|$ at each t , an artificial time parameter that denotes the number of diffusion iterations, which was fixed to 20.

The chromatic anisotropic diffusion is performed by applying Eq. (6) to the complex quantity P

$$\frac{\partial}{\partial t} P(x,y,t) = \text{div}[\alpha(x,y,t) \nabla P(x,y,t)], \quad (8)$$

where $\nabla P(x,y,t)$ is²⁸

$$\nabla P(x,y,t) = [\nabla C(x,y,t) + jC \nabla H(x,y,t)] \exp[jH(x,y,t)] \quad (9)$$

and separating real and imaginary parts of Eq. (8) it follows that

$$\frac{\partial}{\partial t} C = \text{div}(\alpha \nabla C) - \alpha C |\nabla H|^2, \quad (10)$$

$$\frac{\partial}{\partial t} H = \text{div}(\alpha \nabla H) + 2 \frac{\alpha}{C} \nabla C \cdot \nabla H,$$

where the spatial and temporal dependencies have been omitted for convenience.

To obtain the coefficient α for the complex quantity P we need to calculate $|\nabla P(x,y,t)|$, which is

$$|\nabla P(x,y,t)| = \sqrt{|\nabla C(x,y,t)|^2 + C^2(x,y,t) |\nabla H(x,y,t)|^2}. \quad (11)$$

2.3.2 Conversion to single channel image

In this step a gray scale image is obtained from the diffused color image. In this gray scale image, differences between the burnt skin selected by the user and other parts of the image are emphasized. Based on the observation that doctors segment burn wounds by measuring differences among colors, the selection box selected by the user is slid as a mask of size 5×5 pixels along the image and, for each pixel in the image under the center of the sliding mask, the following operation is performed:²⁹

$$f(n,m) = \frac{1}{MAX} \sum_{i=n-\Delta}^{n+\Delta} \sum_{j=m-\Delta}^{m+\Delta} d_E[p(i,j), w(i,j)], \quad (12)$$

where MAX is $\max_{n,m} (\sum_{i=n-\Delta}^{n+\Delta} \sum_{j=m-\Delta}^{m+\Delta} d_E(p(i,j), w(i,j)))$, $\Delta = (L-1)/2$ with $L=5$, $p(i,j)$ represents a pixel in the diffused image to be segmented in $L^*u^*v^*$ color space, $w(i,j)$ is a pixel of the mask selected by the user, and $d_E(\cdot)$, the Euclidean distance between pixels $p(i,j)$ and $w(i,j)$, is defined as

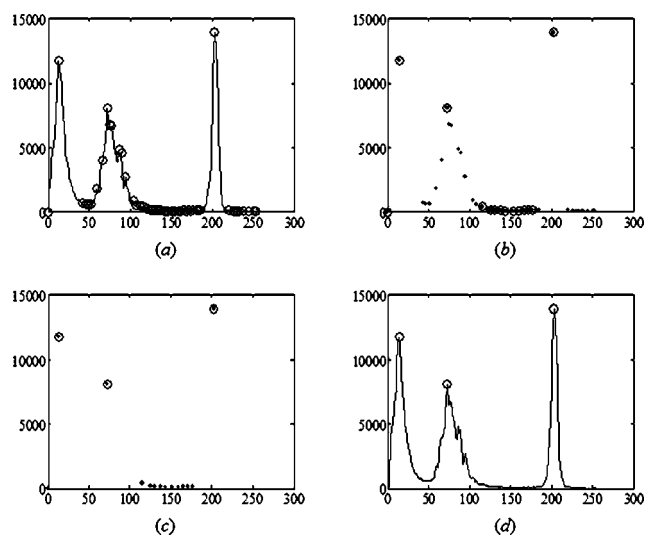


Fig. 3 Process of detecting the main peaks in the histogram. (a) Detection of the peaks in the histograms: peaks are marked with circles. (b) Finding the peaks in the histogram of the peaks: peaks from the original histogram are marked with dots and new peaks with circles. (c) Rejection of nonsignificant peaks: peaks from Fig. (b) are marked with dots and peaks selected in this step are marked with circles. (d) Final peaks in the original histogram after the rejection of peaks without a significant valley between them. In this case the three peaks in the former step are accepted.

$$d_E(p(i,j),w(i,j)) = \{ [L_p^*(i,j) - L_w^*(i,j)]^2 + [u_p^*(i,j) - u_w^*(i,j)]^2 + [v_p^*(i,j) - v_w^*(i,j)]^2 \}^{1/2}. \quad (13)$$

2.3.3 Thresholding operation and postprocessing

The result of the above step is a gray-scale image where pixels with lowest values are those in the region to be segmented. This image has been carefully designed to emphasize the burnt regions, and a thresholding operation should suffice to get a good segmentation. The histogram of this distance image is multimodal so a method to find a threshold to select the mode in the left of the histogram should be found. This task is carried out in two steps: (1) the peaks (maximum values) of the different modes present in the histogram are found, and (2) the threshold which separates the two modes closest to the left of the histogram is calculated applying Otsu's thresholding method.³⁰

To perform the first step, the following algorithm is applied to the histogram of the gray-scale image: (1) find all peaks in the histogram, that is, all the values in the histogram which are higher than their two neighbors; (2) form a new curve with the peaks found in the previous step and then select again the peaks in the new curve; (3) remove nonsignificant peaks, i.e., those peaks whose values are less than 1% of the maximum peak value are rejected; (4) remove nonsignificant valleys, that is, if two peaks have not a significant valley between them we maintain only the highest of the two peaks. To check if a valley is significant or not, the minimum value between two peaks is found. If this minimum value is greater than 75% of the lowest peak out of the two peaks, then

the valley is considered nonsignificant. These four steps are illustrated in Fig. 3.

Once we have localized the main modes in the histogram, we have to find the threshold which separates the two modes closest to the left part of the histogram. This task is carried out by applying Otsu's method,³⁰ which is an adaptive thresholding technique to split a histogram into two classes, c_1 with gray levels $[1, \dots, k]$, and c_2 with gray levels $[k+1, \dots, K]$. Let $m_i(k)$ and m_T be the mean intensities for the class c_i and for the whole image, respectively. The between-class variance was defined by Otsu as

$$\sigma_b^2(k) = \omega_1(k)(m_1(k) - m_T)^2 + \omega_2(k)(m_2(k) - m_T)^2, \quad (14)$$

where $\omega_1(k)$ and $\omega_2(k)$ are cumulative sums of the probabilities in each class, that is, $\omega_1(k) = \sum_{j=1}^k p_j$, $\omega_2(k) = \sum_{j=k+1}^K p_j$, and $p_j = x_j / N_{\text{pixels}}$, where x_j is the number of pixels with gray level j in an image and N_{pixels} is the number of pixels with gray levels from 1 to K in the whole image, that is, the total number of pixels in the image. The optimal threshold \hat{k} is chosen so that the between-class variance σ_b^2 is maximized.

The election of Otsu's method, among many existing thresholding methods, is due to its simplicity in computation.³¹ In fact, many modern segmentation algorithms are based in Otsu's method or use it for comparison.³²⁻³⁴

Finally, by the application of a 3×3 median filter, the segmentation result is improved by removing spurious points (1-4 pixel sized), that is, points that have been segmented and do not actually belong to the burn.

2.4 Classification

Once the burn is segmented, its depth must be estimated for classification purposes. It has been proven that physicians determine the depth of a burn based on color perception, as well as on some texture aspects. As it has been previously said, $L^*u^*v^*$ space is a perceptually uniform color representation system. Also, the hue and the chroma coordinates are intimately related to the way human beings perceive chromaticity. That is why, in this study, a set of descriptors formed by statistical moments of the histograms obtained for each coordinate of the $L^*u^*v^*$ color space, as well as for the hue and chroma image planes derived from them, have been used. More specifically, the descriptors chosen are: mean of lightness (L^*), mean of hue (H), mean of chroma (C), standard deviation of lightness (σ_L), standard deviation of hue (σ_H), standard deviation of chroma (σ_C), mean of u^* , mean of v^* , standard deviation of u^* (σ_u), standard deviation of v^* (σ_v), skewness of lightness (s_L), kurtosis of lightness (k_L), skewness of u^* (s_u), kurtosis of u^* (k_u), skewness of v^* (s_v) and kurtosis of v^* (k_v).

Afterwards it has been necessary to apply a descriptor selection method to obtain the optimum set for the subsequent classification.

2.4.1 Feature selection

The discrimination power of these 16 features is analyzed using the sequential forward selection (SFS) method and the

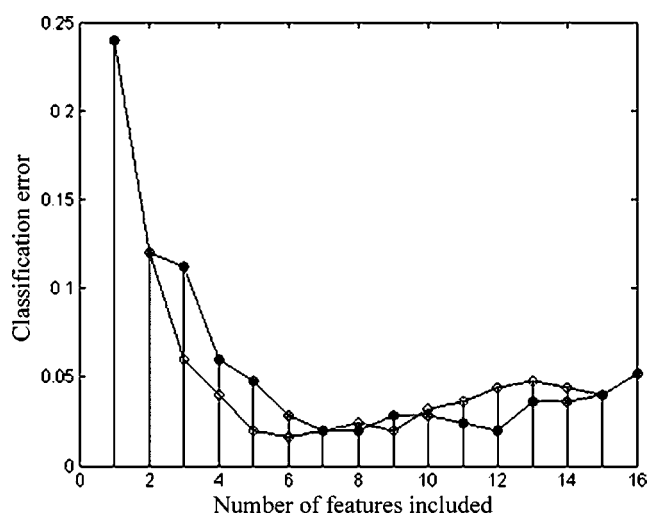


Fig. 5 Evolution of the classification error for SFS method (●) and SBS method (○).

sequential backward selection (SBS) method^{35,36} via the Fuzzy-ARTMAP neural network which is detailed in the following subsection.

SFS is a bottom-up search procedure where one feature at a time is added to the current feature set. At each stage, the feature to be included in the feature set is selected among the remaining available features which have not been added to the feature set. So the new enlarged feature set yields a minimum classification error comparing to adding any single feature. The algorithm stops when adding a new feature yields an increase of the classification error. The SBS is the top-down counterpart of the SFS method. It starts from the complete set of features and, at each stage, the feature which shows the least discriminatory power is discarded. The algorithm stops when removing another feature implies an increase of the classification error.

To apply these two methods, 50 49×49 pixel images for each burn appearance have been used (see Fig. 4). As there are five appearances, in all we have 250 49×49 pixel images.* One photograph has been taken per burn wound. In general, we selected only one 49×49 pixel image per photograph, unless there were different appearances in the same wound. In this case, one 49×49 image per appearance was selected.

The selection performance is evaluated by fivefold cross validation (XVAL).¹⁵ In this sense, the disadvantage of sensitivity to the order of presentation of the training set, that the SBS and SFS methods present,³⁵ is diminished. To perform the XVAL method the 50 images per burn appearance are split into five disjoint subsets. Four of these subsets (that is, 40 images per appearance) serve as a training set for the neural network, while the other one (ten images) is used as validation set. Then, the procedure is repeated interchanging the validation subset with one of the training subsets, and so on

*The 250 49×49 pixel images are small images showing each one only one burn appearance (no healthy skin or background). Each 49×49 pixel image has been validated by two physicians as belonging to a particular depth. Therefore, these 250 images form a database used only for the feature selection step.

till the five subsets have been used as validation sets. The final classification error is calculated as the mean of the errors for each XVAL run.

In Fig. 5 the evolution of the classification error is presented for both selection methods. It can be observed that both curves coincide at the beginning and at the end, but then they separate obtaining a minimum classification error with seven or eight descriptors (2% error) for the SFS method, and six descriptors (1.6% error) for the SBS method. In fact, this minimum error is again reached with 12 descriptors, although it is reasonable to choose the set of six, because it will imply less complexity in the neural network and shorter processing time. The six descriptors provided by SBS method were chosen as the best feature set: lightness, hue, standard deviation of the hue component, u^* chrominance component, standard deviation of the v^* component, and skewness of lightness.

2.4.2 Fuzzy-ARTMAP neural network

The classifier used is a Fuzzy-ARTMAP neural network. This type of network is based on the Adaptive Resonance Theory developed by Grossberg and Carpenter. Fuzzy-ARTMAP is a supervised learning classification architecture for analog-value input pairs of patterns.³⁷ The reasons for this choice are that Fuzzy-ARTMAP offers the advantages of well-understood theoretical properties, an efficient implementation, clustering properties that are consistent with human perception, and a very fast convergence. It has also a track record of successful use in industrial and medical applications.³⁸ Other strong-points of this type of neural network are the small number of design parameters (the vigilance parameter, $\rho_a \in [0,1]$, and the selection parameter, $\alpha > 0$), and that the architecture and initial values are always the same, independent of the application.

When the input parameters are the features selected by the SBS method above, the network classifies the burn depth of the segmented region into five types: the first and the second belonging to superficial dermal depth, the third to deep dermal, and the fourth and fifth to full thickness. So, the network has six neurons in the input layer and five neurons in the output layer. In the Fuzzy-ARTMAP neural network the architecture is dynamic, so the number of neurons in the hidden layer is fixed during the training and according with the vigilance parameter.

3 Experimental Results

The images used to test the burn CAD tool were 62 digital photographs taken by physicians following the acquisition protocol. All the images were diagnosed by a group of plastic surgeons, affiliated with the burn unit of the Virgen del Rocío Hospital, from Seville (Spain). The assessments were validated one week later, as is the common practice when handling burnt patients. The images were 1536×1024 pixels and they were stored as JPEG (high quality) files.

The computer used was a Pentium IV, 1.7 GHz and 256 MB of random access memory. The average run time was 4 min for an image and the programming tool was MATLAB 6.1 (The Mathworks Inc., Natick, Massachusetts).

Table 1 Quantification of segmentation results (PPV: positive predictive value; S: sensitivity).

Image	PPV	S	Image	PPV	S
Image 1	0,9309	0,8093	Image 19	0,8303	0,9280
Image 2	0,9314	0,6969	Image 20	0,9627	0,9005
Image 3	0,9391	0,8684	Image 21	0,9196	0,7418
Image 4	0,9302	0,8324	Image 22	0,8752	0,7789
Image 5	0,9614	0,9015	Image 23	0,9559	0,9725
Image 6	0,9741	0,8853	Image 24	0,8622	0,9107
Image 7	0,8807	0,7297	Image 25	0,9082	0,9069
Image 8	0,8984	0,8108	Image 26	0,9646	0,7989
Image 9	0,9618	0,7772	Image 27	0,9320	0,9364
Image 10	0,9737	0,8206	Image 28	0,8711	0,8457
Image 11	0,7928	0,8190	Image 29	0,9569	0,9482
Image 12	0,9624	0,7452	Image 30	0,9571	0,8814
Image 13	0,9806	0,7248	Image 31	0,9134	0,8318
Image 14	0,9424	0,7820	Image 32	0,9327	0,8698
Image 15	0,9384	0,8457	Image 33	0,6990	0,7588
Image 16	0,8327	0,8066	Image 34	0,9192	0,5174
Image 17	0,6420	0,8539	Image 35	0,7701	0,8530
Image 18	0,8788	0,9646	Average	0,9023	0,8301

3.1 Segmentation Results

The segmentation algorithm proposed in this paper was tested with 35 out of the 62 images of the database. These 35 images were manually segmented by five physicians.

The reason of using 35 photographs instead of 62 is that, although the protocol says that it should appear as healthy and burnt skin, very often the extension of the burn wound is so large that there is only burnt skin in the image. Therefore, in these cases it is not meaningful to compare the segmentation results performed by the physicians and by the algorithm.

The segmentation gold standard was obtained by applying the voting method to the regions segmented by the five specialists. In other words, one pixel was considered to belong to the segmented region in the gold standard if most of the physicians had considered it in this way.

Once a gold standard was obtained, two parameters were calculated to measure the performances of the segmentation algorithm. The first parameter was the *positive predictive value* (PPV), which measures the ratio between the number of pixels segmented by the algorithm which fit the segmentation gold standard and the total amount of pixels segmented. The second parameter is called *sensitivity* (S), and it is the ratio between the number of pixels segmented by the algorithm which fit the segmentation gold standard and the total amount of pixels in the segmentation gold standard. Intuitively it can

be seen that the first parameter measures the over segmentation, which would be null if PPV were 1. Likewise, S measures the under segmentation. In Table 1 the results for the 35 images are presented. As is shown in this table, almost all the photographs are properly segmented. It must be emphasized that, although the sensitivity tends to be only around 0.8, this is because doctors tend to over segment the burnt region. Therefore, this should not be interpreted as a poor performance of the algorithm.

Figures 6–8 show the segmentation results for some images of the three types of depth. Figures (a) represent original images and Figs. (b) represent the segmented ones. In the segmented images we have marked with yellow color the segmented region. In all the cases, the burn wound was segmented correctly from the normal skin.

3.2 Classification Results

To test the classification part we employed the 62 images of the database used for validation (different from the one used for training). The neural network was trained with the 250 49×49 pixel images previously cited. The training was performed with $\rho_a = 1$ and $\alpha = 0.001$. At the end of the training the weights were fixed for the subsequent classification test. For this test the six features were extracted from the segmented part of the 62 images. Classification results are summarized in Table 2. We have used 22 images with superficial dermal burns, 18 with deep dermal burns, and 22 with full-thickness burns. The average success percentage was 82.26%. All superficial dermal burns misclassified were classified by the network as deep dermal ones. All deep dermal burns were misclassified as superficial dermal ones. And, in the case of misclassified full-thickness burns, 80% of them were classified as superficial dermal and 20% as deep dermal.

4 Discussion and Conclusions

The classification of burn depths based on visual inspection is a difficult task, which needs a lot of training. That is why in burn related literature there is a constant search for objective methods to determine the depth of a burn. A prototype of one invasive technique is the acquisition of biopsies and their histological study for the burn depth diagnosis.³⁹ This technique, although it can be considered as “gold standard,” is not exempt from problems related to loss of dermis in the burn, to the existence of considerable variability depending on where the biopsy was acquired, and to the fact that this technique is a snapshot view of the lesion, apart from the residual scars provoked by the biopsy acquisition. These inconveniences have directed efforts towards the design of noninvasive procedures. Some noninvasive techniques analyze the perfusion of the burn wound based on the fact that tissue damage is inversely proportional to the vascularization after the lesion.^{40–42} Nevertheless, in these procedures it is necessary to supply a vital colorant to the patient by intravenous method and it is essential to have an emergency system. Other experimental techniques analyze the changes in optical properties of the skin related to the changes of its vascularization,⁴³ although their application environment is, for the moment, exclusively experimental. In another type of approximation to the problem being studied, the remission-optical measurement exploits the different spectral backscattering effects of burned

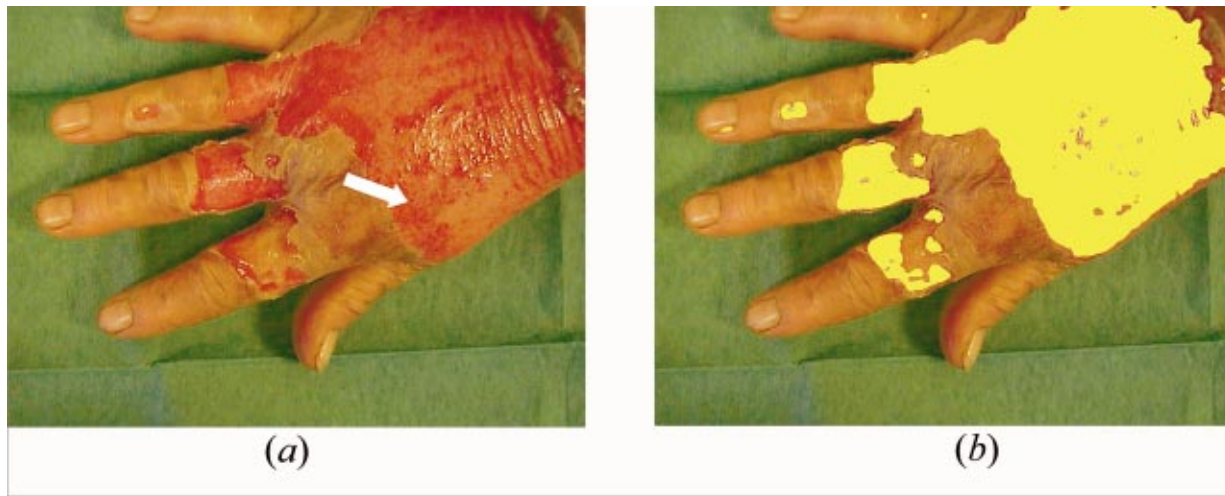


Fig. 6 Segmentation result for a superficial dermal burn. (a) Original image where the selection made by the user is shown with an arrow. (b) Segmented image.

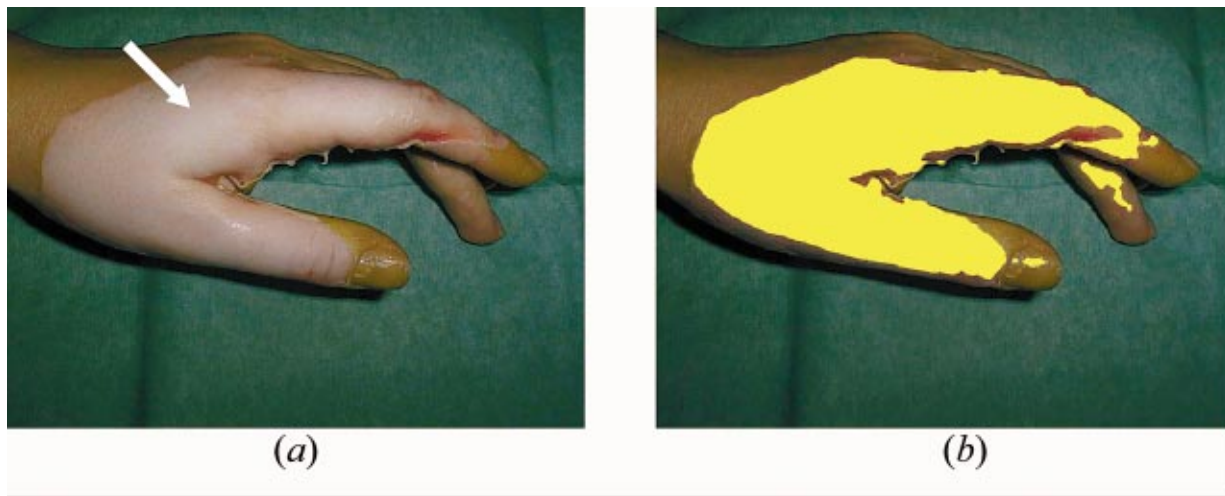


Fig. 7 Segmentation result for a deep dermal burn. (a) Original image where the selection made by the user is shown with an arrow. (b) Segmented image.

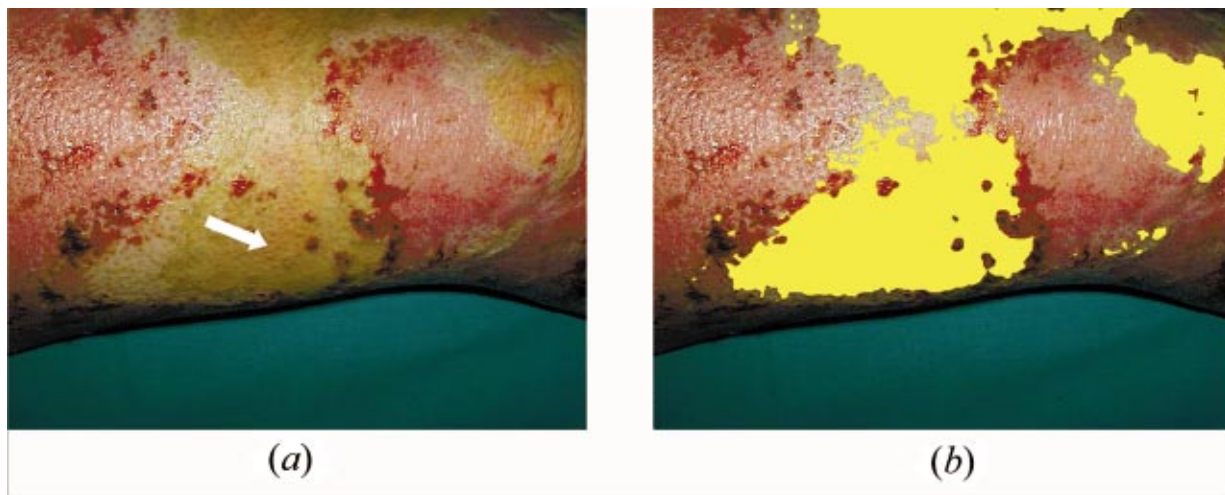


Fig. 8 Segmentation result for a full thickness burn. (a) Original image, which has both superficial dermal burn (the red part) and full-thickness burn (the creamy part). (b) Segmented image. In this case the user has made the selection in the creamy part in order that the algorithm segments all the full-thickness part of the burn. It segments correctly all the full-thickness parts of the image regarding what physicians said.

Table 2 Classification results.

Burn depth	Success percentage
Superficial dermal	86.36%
Deep dermal	83.33%
Full thickness	77.27%
Average	82.26%

skin at different burn degrees,⁴⁴ although again having an acquisition method and an image processing system are necessary and the majority of the emergency unities do not have them.

In this paper, we present an objective way of determining burn depth without the problems of invasive techniques and without the requirement of specialized acquisition equipments. The great advantage of this system is its facility of implementation in any local medical center, where there is a lack of experts and sophisticated equipments. The only resources that it needs are a digital photographic camera and a computer. Furthermore, this system does not demand users trained in this technique or in burn diagnosis, achieving a success rate of 82.26% in classifying the depths of the burns, which is comparable with experts' assessment. Since there are no experts in burn treatment, but general practitioners, in an emergency unit, this rate will diminish unnecessary displacements and mistreatments.

The system starts with a segmentation step, where the aim is to isolate the burn wound from the rest of the scene (healthy skin and background). It is important to note, that the user has to select with the mouse a small selection box of the region (color) to be segmented. It is not possible to solve the problem without the help of the user due to the overlapping existing among different healthy skin colors and different burn depth colors. Although it is difficult for a nonexpert to assess the depth of the burn, it is not difficult to know which part of the skin is burnt or not.

Once the burn is isolated, we extract from it six color and texture descriptors that will be the inputs to the classifier. The six descriptors are the inputs to a Fuzzy-ARTMAP artificial neural network which classified them as one of the possible depths a burn can present. We tested 62 photographs, yielding a classification average success percentage of 82.26%. The 55% of the total number of misclassifications were considered as superficial dermal types while they actually were deep dermal ones, or vice versa. In general, this is also common among physicians; actually some burns are diagnosed as "intermediate depth," when they are neither clearly superficial dermal nor deep dermal.

As this work is in a research stage, the programming has been done in MATLAB. In order to reduce the computational time, it will be implemented in C programming language. As with MATLAB the computational time has been short, it is expected to have a negligible computational cost when using C language.

Acknowledgments

The authors thank Dr. Gómez-Cía, Dr. Torre, and the burn unit of Virgen del Rocío Hospital, Seville (Spain) for providing us with the burn wound photographs and their medical advice, and the CICYT Spain (Project No. TIC-2002-01401) and the Junta de Andalucía for financial support.

References

1. J. A. Clarke, *A Colour Atlas of Burn Injuries*, Chapman & Hall Medical, London (1992).
2. C. Serrano, L. Roa, and B. Acha, "Evaluation of a telemedicine platform in a burn unit," in *Proc. IEEE Int. Conf. on Information Technology Applications in Biomedicine*, 121–126 (1998).
3. L. Roa, T. Gómez-Cía, B. Acha, and C. Serrano, "Digital imaging in remote diagnosis of burns," *Burns* **25**(7), 617–624 (1999).
4. W. P. Berriss and S. J. Sangwine, "Automatic quantitative analysis of healing skin wounds using colour digital image processing," available: <http://www.smtl.co.uk/World-Wide-Wounds/1997/july/Berriss/Berriss.html#perednial> (1997).
5. M. Herbin, F. X. Bon, A. Venot, F. Jeanlouis, M. L. Dubertret, L. Dubertret, and G. Strauch, "Assessment of healing kinetics through true color image processing," *IEEE Trans. Med. Imaging* **12**(1), 39–43 (1993).
6. F. X. Bon, E. Briand, S. Guichard, B. Couturaud, M. Revol, J. M. Servant, and L. Dubertret, "Quantitative and kinetic evolution of wound healing through image analysis," *IEEE Trans. Med. Imaging* **19**(7), 767–772 (2000).
7. G. L. Hansen, E. M. Sparrow, J. Y. Kokate, K. J. Leland, and P. A. Iaizzo, "Wound status evaluation using color image processing," *IEEE Trans. Med. Imaging* **16**(1), 78–86 (1997).
8. M. Burns, J. Enderle, E. Rosow, and Q. Zhu, "Development of a wound assessment system for quantitative chronic wound monitoring," *Proc. IEEE 28th Annual Northeast Bioengineering Conf*, Philadelphia, PA, pp. 7–8 (2002).
9. J. R. Mekkes and W. Westerhof, "Image processing in the study of wound healing," *Clin. Dermatol.* **13**(4), 401–407 (1995).
10. R. A. Fiorini, M. Crivellini, G. Codagnone, G. F. Dacquino, G. Libertini, and A. Morresi, "DELM image processing for skin-melanoma early diagnosis," *Proc. SPIE* **3164**, 359–370 (1997).
11. J. P. Thira and B. Macq, "Morphological feature extraction for the classification of digital images of cancerous tissues," *IEEE Trans. Biomed. Eng.* **43**(10), 1011–1020 (1996).
12. G. A. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Unsupervised color image segmentation with application to skin tumor borders," *IEEE Eng. Med. Biol. Mag.* **15**(1), 104–111 (1996).
13. Z. Zhang, W. V. Stoecker, and R. H. Moss, "Border detection on digitized skin tumor images," *IEEE Trans. Med. Imaging* **19**(11), 1128–1143 (2000).
14. D. H. Chung and G. Sapiro, "Segmenting skin lesions with partial-differential-equations-based image processing algorithms," *IEEE Trans. Med. Imaging* **19**(7), 763–767 (2000).
15. H. Ganster, A. Pinz, R. Röhler, E. Wilding, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imaging* **20**(3), 233–239 (2001).
16. F. J. Wyllie and A. B. Sutherland, "Measurement of surface temperature as an aid to the diagnosis of burn depth," *Burns* **17**(2), 123–127 (1991).
17. R. P. Cole, S. G. Jones, and P. G. Shakespeare, "Thermographic assessment of hand burns," *Burns* **16**(1), 60–63 (1990).
18. R. E. Barsley, M. H. West, and J. A. Fair, "Forensic photography: ultraviolet imaging of wounds on skin," *Am. J. Forensic Med. Pathol.* **11**(4), 300–308 (1990).
19. J. E. Bennett and R. O. Kingman, "Evaluation of burn depth by the use of radioactive isotopes—an experimental study," *Plast. Reconstr. Surg.*, **20**(4), 261–272 (1957).
20. Z. B. M. Niazi, T. J. H. Essex, R. Papini, D. Scott, N. R. McLean, and J. M. Black, "New laser doppler scanner, a valuable adjunct in burn depth assessment," *Burns* **19**(6), 485–489 (1993).
21. M. A. Afromowitz, G. S. Van Liew, and D. M. Heimbach, "Clinical evaluation of burn injuries using an optical reflectance technique," *IEEE Trans. Biomed. Eng.* **34**(2), 114–127 (1987).
22. M. A. Afromowitz, J. B. Callis, D. M. Heimbach, L. A. DeSoto, and M. K. Norton, "Multispectral imaging of burn wounds: A new clinical

- cal instrument for evaluating burn depth," *IEEE Trans. Biomed. Eng.* **35**(10), 842–850 (1988).
23. K. N. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*, Springer, Berlin (2000).
 24. C. Serrano, L. Roa, B. Acha, and T. Gómez-Cía, "Aspectos metodológicos para la realización de una plataforma de telemedicina en una unidad de quemados," *Int. Telemedicine* **12**, 24–32 (1999).
 25. B. Acha, "Segmentación y clasificación de imágenes en color. Aplicación al diagnóstico de quemaduras," PhD thesis, University of Seville, Spain (2002).
 26. P. Schmid, "Segmentation and symmetry measure for image analysis: Application to digital dermatoscopy," PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland (1999).
 27. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990).
 28. L. Lucchese and S. K. Mitra, "Color segmentation based on separate anisotropic diffusion of chromatic and achromatic channels," *IEE Proc. Vision Image Signal Process.* **148**(3), 141–150 (2001).
 29. C. Serrano, B. Acha, and J. I. Acha, "Segmentation of burn images based on color and texture information," *Proc. SPIE* **5032**, 1543–1550 (2003).
 30. N. Otsu, "A threshold selection method from Gray-level histograms," *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979).
 31. K. C. Lin, "Fast image thresholding by finding the zero(s) of the first derivative of between-class variance," *Mach. Vision Appl.* **13**(5–6), 254–262 (2003).
 32. P. S. Liao, T. S. Chen, and P. C. Chung, "Fast algorithm for multi-level thresholding," *J. Inf. Sci. Eng.* **17**(5), 713–727 (2001).
 33. O. Virmajoki and P. Fränti, "Fast pairwise nearest neighbor based algorithm for multilevel thresholding," *J. Electron. Imaging* **12**(4), 648–659 (2003).
 34. M. L. G. Althouse and C. I. Chang, "Image segmentation by local entropy methods," *IEEE Int. Conf. on Image Proc.*, **3**, 61–64, Washington, DC (1995).
 35. K. Fukunaga, *Introduction to statistical pattern recognition*, 2nd ed., Academic, Press, San Diego, CA (1990).
 36. S. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Med. Imaging* **19**(2), 115–126 (2000).
 37. G. A. Carpenter, S. Grossberg, S. Markuzon, and J. H. Reynolds, "Fuzzy-ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Netw.* **3**(5), 698–713 (1992).
 38. G. W. Donohoe, S. Nemeth, and P. Soliz, "ART-based image analysis for pigmented lesions of the skin," *11th IEEE Symp. Computer-based Med. Systems*, Lubbock, TX, 293–298 (1998).
 39. A. M. Watts, M. P. Tyler, M. E. Perry, A. H. Roberts, and D. A. McGrouther, "Burn depth and its histological measurement," *Burns* **27**, 154–160 (2001).
 40. M. R. Jerath, K. T. Schomacker, R. L. Sheridan, and N. S. Nishioka, "Burn wound assessment in porcine skin using indocyanine green fluorescence," *J. Trauma* **46**, 1085–1088 (1999).
 41. L. P. Kamolz, H. Andel, W. Haslik, A. Donner, W. Winter, G. Meissl, and M. Frey, "Indocyanine green video angiographies help to identify burns requiring operation," *Burns* **29**, 785–791 (2003).
 42. F. W. Kloppenberg, G. I. Beerthuisen, and H. J. Duis, "Perfusion of burn wound assessed by laser Doppler imaging is related to burn depth and healing time," *Burns* **27**, 359–363 (2001).
 43. L. T. Vo, P. Anikijenko, W. J. McLaren, P. M. Delaney, D. H. Barkla, and R. G. King, "Autofluorescence of skin burns detected by fiber-optic confocal imaging: evidence that cool water treatment limits progressive thermal damage in anesthetized hairless mice," *J. Trauma: Inj., Infect., Crit. Care* **51**, 98–104 (2001).
 44. W. Eisenbeiss, J. Marotz, and J. P. Schrade, "Reflection-optical multispectral imaging method for objective determination of burn depth," *Burns* **25**, 697–704 (1999).