



Robust tissue classification for reproducible wound assessment in telemedicine environments

Hazem Wannous, Sylvie Treuillet, Yves Lucas

► To cite this version:

Hazem Wannous, Sylvie Treuillet, Yves Lucas. Robust tissue classification for reproducible wound assessment in telemedicine environments. *Journal of Electronic Imaging*, 2010, pp.023002-1-9. 10.1117/1.3378149 . hal-00648504

HAL Id: hal-00648504

<https://hal.science/hal-00648504>

Submitted on 5 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust tissue classification for reproducible wound assessment in telemedicine environments

Hazem Wannous

Sylvie Treuillet

Université d'Orléans

Institut PRISME – Polytechnique Orléans
45067 Orléans cedex 2, France

Yves Lucas

Université d'Orléans

Institut PRISME – IUT Bourges

63 avenue de Lattre

18020 Bourges cedex, France

E-mail: yves.lucas@bourges.univ-orleans.fr

Abstract. *In telemedicine environments, a standardized and reproducible assessment of wounds, using a simple free-handled digital camera, is an essential requirement. However, to ensure robust tissue classification, particular attention must be paid to the complete design of the color processing chain. We introduce the key steps including color correction, merging of expert labeling, and segmentation-driven classification based on support vector machines. The tool thus developed ensures stability under lighting condition, viewpoint, and camera changes, to achieve accurate and robust classification of skin tissues. Clinical tests demonstrate that such an advanced tool, which forms part of a complete 3-D and color wound assessment system, significantly improves the monitoring of the healing process. It achieves an overlap score of 79.3 against 69.1% for a single expert, after mapping on the medical reference developed from the image labeling by a college of experts. © 2010 SPIE and IS&T. [DOI: 10.1117/1.3378149]*

1 Visual Wound Assessment

Monitoring the wound healing process is a tedious task for clinicians and nurses as it is necessary to periodically assess the wound. All types of wounds need assessment; not only chronic wounds but also ulcers, burns, traumatic or surgical wounds, and dermatological lesions. Moreover, wound care is expensive, costing about 8 billion euros per year, as the healing process can last several months, and with the aging of the population this will necessarily increase by 25% over the next 10 yr. As health care costs must be drastically reduced, there is a growing demand for patients to be cared for at home and hence for wound monitoring to take advantage of telemedicine environments. Pioneer experiments in this area consisted simply of uploading images to a web site where a physician could view the data at his convenience.¹ In more recent studies, image processing has been added, but it provides only ulcer stage grading.² The quantitative assessment of chronic wounds

still relies^{3,4} on visual inspection and manual techniques, to describe the shape of the wound (perimeter, surface, depth, etc.) and the biological nature of the skin tissues (percentage of each class, wound severity stage, burn degree, etc.). Wound dimensions and shape are currently measured with an ordinary ruler, or sometimes through sketches on cross-ruled sheets, serum injection, or alginate moldings. While some prototypes have been developed for 3-D data capture by optical techniques, they are extremely costly and complex, making them unsuitable for routine hospital care.^{5–7} Similarly, assessing the type and proportion of tissues remains highly empirical as evaluation is performed visually and then recorded on a red-black-yellow scale corresponding, respectively, to the dominant color of the different tissues found on a wound: granulation, slough, and necrosis. Healing is a complex cascade of cellular events operating to reconstruct damaged tissue and as the different tissues may overlap and be difficult to distinguish, wound assessment is not straightforward. The lack of quantitative data affects the coordination of care staff and hinders clinical studies focused on healing. Digital cameras, though now widespread in clinical centers, are used only for ordinary patient data recording and not image processing, as wound therapeutic follow-up is carried out by nurses.

Several studies have tackled the problem of digital wound assessment, but these attempts have failed to provide a robust tool for automatic tissue classification. Results remain too dependent on image capture conditions, sample database building, region descriptor selection, tissue class learning protocol etc., preventing reproducible results from being obtained with the complete image processing chain. Several features contribute to making automatic classification difficult. Wound image acquisition requires technical skill, especially in the patient room where lighting is not controlled. At close range, the depth of field remains always limited in macro mode; ambient light is insufficient and may easily result in fuzzy images. Moreover, the patient is rarely able to maintain a convenient posture for a

Paper 09074RR received May 7, 2009; revised manuscript received Feb. 16, 2010; accepted for publication Feb. 22, 2010; published online Apr. 8, 2010.

1017-9909/2010/19(2)/023002/9/\$25.00 © 2010 SPIE and IS&T.

snapshot. After image capture, the wound area must be extracted from the image. Early attempts to accomplish this by using color measurements did not completely succeed, as reported in Ref. 8, and have been replaced by semiautomatic methods⁹ by spline fitting from a set of clicked boundary points or snakes on a manually specified closed contour.¹⁰ Color correction is often neglected as all the photos are taken with the same camera and lighting during experimentation. It will then be impossible to process correctly images from other care centers, unless the photos in each center were taken with identical digital cameras and standardized settings.¹¹ Sometimes, a white patch is included in the field of view to estimate the white balance.¹² Occasionally, a color pattern is used for enhanced color correction, giving also access to the scale factor for dimensional measurements.^{9,13} Another preprocessing step is image compression, an important requirement for data management.¹⁴ The color analysis required for tissue classification over the wound is a difficult task. As direct classification on the pixels with simple thresholds on separate color components proved to be inefficient due to the variability and nonhomogeneity of tissues, spatial continuity has been searched for through a segmentation process. The classification process is thus driven by the segmentation process.^{15,16} During the following step, a tissue database must be built, covering all wound grades, locations, and healing status. This recording task relies mainly on the care staff as the wounds are visible only after cleaning during the dressing. In many studies, the tissue samples were manually extracted by asking the clinician to pick square homogeneous regions of interest inside the images.¹⁷ Complete classification would require that during the tests these samples be automatically located in the image without *a priori* knowledge. Using photos instead of examining the patient has an impact on expert performance.¹⁸ When several experts are consulted, it appears that interobserver variability is important, exceeding intraobserver variability.¹⁹ For classification purposes, several region descriptors have been widely tested on wound tissue samples. Color histograms are typical of such statistical data gathering as tissue descriptors,^{12,20} but it has been proved that tissues could still not be classified robustly in a large collection of images. The best results were obtained when the classification was limited to two types of tissue²¹ or by multiplying the tissue classes using hybrid classes.²² To improve the results, texture parameters have finally been added.^{17,23} Several classification tools have also been addressed and supervised tools have been shown to surpass the others.^{24,25} In many studies,^{2,26,27} assessment of the tissue type is not achieved (only the grade or the surface of the wound is provided) or is only partially accomplished.

The principal objective of this work is to develop a user-friendly tool for wound tissue assessment, to automatically monitor the healing process. As only an ordinary digital camera is used, the tool can be easily implemented by care staff.^{28,29} A significant improvement compared to previous work resides in the fact that all the steps of the image-processing chain have been especially designed to provide reproducible classification results, independent of the image capture conditions or expert variability. This is an essential requirement to enable telemedicine practice when images come from different care centers or from patients at

home. Other achievements are the automatic classification of all types of tissue on the wound with the accuracy of a single expert. This paper presents part of a complete wound assessment tool currently being designed. It combines wound color processing to classify the tissues with 3-D wound metric reconstruction.³⁰⁻³² In the following, we focus on the different steps of the wound image processing chain and highlight the original contributions of the method. Section 2 presents the design of the tissue sample database, incorporating colorimetric correction and scaling with a calibrated pattern and the extraction of tissue samples by merging several clinician labels. Section 3 describes the selection and tuning of adapted algorithms for automatic wound image segmentation. Section 4 details the selection of color and texture descriptors for the tissue samples, and the learning of a support vector machine (SVM) classifier. In Sec. 5 we present the classification results, before concluding with current work and perspectives.

2 Tissue Sample Extraction

2.1 Wound Image Database Building

A database of chronic wound images was constituted with the help of clinical staff in several hospitals to obtain a representative set of images for different tissue types as input for the wound image-processing chain (Fig. 1). A preliminary requirement was to acquire good-quality wound images covering all the pathologies. A variety of chronic wounds (sacrum and heel bed sores, leg ulcers, diabetic feet, etc.) was collected from different care services including gerontology, rehabilitation, diabetology, and dermatology. Several hundred color images (3 Mpixels, 24 bits) were taken with different digital cameras under uncontrolled lighting conditions, following a specific protocol that included several viewing angles for each single wound.

2.2 Color Correction

Color is the main cue for the visual assessment of wound tissues but it may be easily corrupted by the image acquisition process. Reliable tissue classification will therefore be possible only if special attention is paid to color constancy. Two factors play an important role here. First, the lighting conditions in the patient room are uncontrolled and prevent monitoring the healing of tissues over time. Even when a flash is used, the influence of lighting is not eliminated as the camera angle is different at each visit. A further problem is that each camera produces different images of the same scene. The spectral sensitivity is specific for each CCD or CMOS sensor and a camera, as with any digital imaging system, embeds a customized color-processing system. Adjustments are also made during the shooting by the white balance automatic control. Color correction therefore must address two distinct tasks: first, obtaining a constant response from the digital camera by identifying lighting conditions, and second, calibrating the digital camera color response. To achieve this, a standard Macbeth color checker pattern was introduced in the field of view, providing also the scale factor in the different images. We estimate the illumination by the retinex-white patch algorithm before adjusting the color balance. Next, the color patches enable a correct transformation of the *RGB* output values

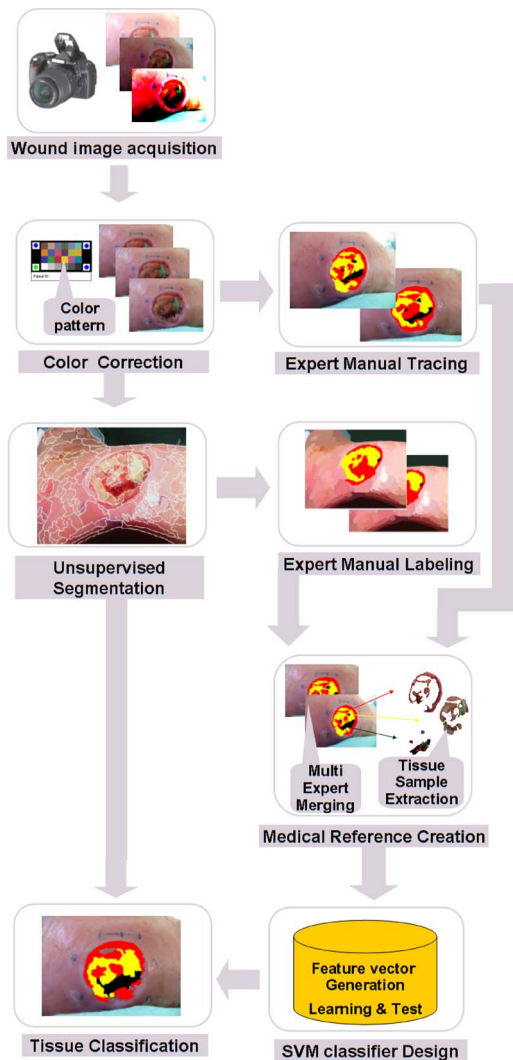


Fig. 1 Color image processing chain.

into reference ones expressed in standard *RGB* (a device-independent color space introduced by Hewlett-Packard) using a suitable polynomial model.

Several laboratory tests were performed to measure precisely the influence of the practical conditions during image shooting in the hospital or at the patient's home in the case of telecare. The color shifts due to the digital camera used, the ambient lighting, and the angle between the camera optical axis and the normal of the color pattern were evaluated by computing $CIE\Delta E_{ab}$ color distance between input and corrected images, as perceptual color spaces are more appropriate prior to clinician labeling. Three digital cameras were compared: a compact Leica D-Lux3, a bridge Sony DSC-H5, and a reflex Canon EOS 350D. Six light sources were considered: flash, tungsten, direct sun, neon, halogen, and daylight. Practically, the $CIE\Delta E_{ab}$ color distance is averaged on the 24 patches of the color pattern. The results, also averaged over five view angles from 0 to 40 deg with 10-deg steps, display color shifts of about 15 units on the $CIE\Delta E_{ab}$ scale on input images, reduced to 5 units on corrected images (Figs. 2 and 3). This confirms that both the light source and the digital camera have a

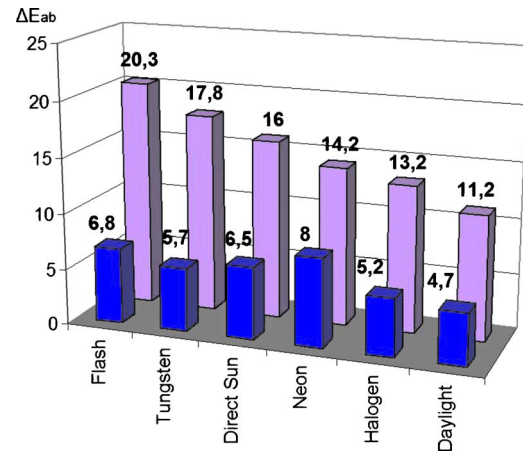


Fig. 2 $CIE\Delta E_{ab}$ image color shifts under different light sources before correction (back row) and after correction (front row).

significant impact on color shifts and it demonstrates that color correction is essential in a telecare environment, involving several care centers where the practical conditions differ widely. After correction, colorimetric stability reaches the limit of discernibility for a human observer. To complete this laboratory experiment, we analyzed in a similar fashion color shifts on photos taken during wound patient examination in hospital; $CIE\Delta E_{ab}$ was reduced to 3.5 ± 2.9 . These smaller values are explained by less varied lighting conditions. We also observed that in the patient room, different viewpoints may induce color differences exceeding 8 units between the views before correction.

2.3 Multiexpert Image Labeling

The wound image database collected was submitted to a group of clinicians to label it according to the classical color code using our graphical interface. This interface enables the user to draw closed boundaries using a pen tablet or a mouse and to fill it with color labels. In this medical imaging application, it provides the input data for building the absolute medical reference by merging expert data, as the clinician tracings are freeform. Precautions were taken concerning the lighting conditions and the color calibration of the graphical screens to ensure faithful rendering during

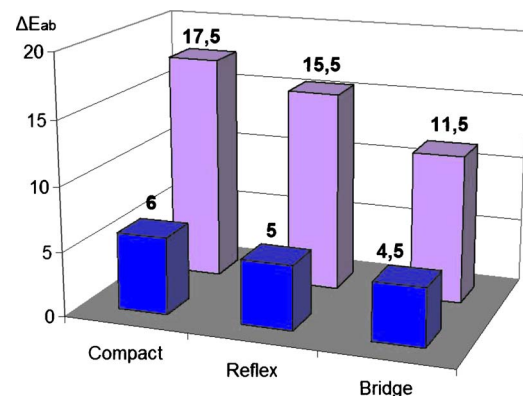


Fig. 3 $CIE\Delta E_{ab}$ image color shifts using different digital cameras before correction (back row) and after correction (front row).

Table 1 Intraexpert OSs.

Tissue Class	Expert OS (%)				Average
	1	2	3	4	
Granulation	73.9	65.0	55.9	66.7	65.4
Slough	77.4	73.8	55.3	71.4	69.5
Necrosis	81.4	91.4	59.5	81.9	78.6
Expert average	77.6	76.7	56.9	73.3	71.2

the labeling sessions. Patients corresponding to the wounds had not been previously seen by the clinicians to eliminate any external influence. Tissue labeling was carried out twice, 1 month apart, by four clinicians on 50 wound images to measure the accuracy during labeling. We used the normalized overlap score (OS), classically used to compare a segmentation S to a ground truth G by $|S \cap G| / |S \cup G|$. This was averaged over all the tissue samples weighted by their corresponding areas. As $0 \leq OS \leq 1$, it can be expressed as a percentage. The OS obtained by the same clinician labeling, one month apart, are given in Table 1. The labels of the different clinicians were merged to obtain a unique and reliable medical reference for evaluation of the algorithm. A majority vote criterion was applied to label each pixel and only those pixels for which 75% or more of wound assessments included in a single tissue class were retained. The OSs were then computed between the clinician tracings and this medical reference (Table 2).

Several observations can be made. First, the OSs between clinicians and the reference remain moderate (55 to 76%) as the labels rely on the subjective diagnosis of the clinician, who may not be familiar with wound assessment on a computer, without direct patient examination; moreover, each clinician does not produce similar labels 1 month later (57 to 78%). Scores obtained for the slough tissue class are the lowest, as this type of tissue is often mixed with necrosis and because granulation grows under slough areas. A very practical issue of these inter- and intraexpert OS is to provide a performance target to be reached by the automatic classification process.

Table 2 OSs between clinicians and medical reference.

Tissue Class	Expert OS (%)				Average
	1	2	3	4	
Granulation	73.2	72.7	39.3	71.0	64.1
Slough	58.9	76.0	42.8	70.8	62.1
Necrosis	79.3	77.1	82.6	85.6	81.2
Expert average	70.5	75.3	54.9	75.8	69.1

Table 3 Parameter tuning and OSs of four unsupervised segmentation methods.

Algorithm	Space	Free Parameter	Final Setting	OS
CSC	<i>RGB</i>	S_d	25	64.6
		σ	0.5	
EGBIS	<i>RGB</i>	k	200	61.3
		min	100	
		r	8	
Mean shift	<i>L*a*b*</i>	h	6.5	72.6
		min	100	
J-SEG	<i>L*u*v*</i>	m	0.2	73.1
		q	300	

3 Wound Segmentation

We selected four advanced unsupervised segmentation algorithms that are efficient on textured images to provide an automatic delineation of tissue samples and to simplify the following classification step by extracting more robust and discriminant local attributes on tissue areas than with direct pixel classification: the color structure code³³ (CSC), the efficient graph-based image segmentation³⁴ (EGBIS), the mean shift,³⁵ and the J-value segmentation (J-SEG).³⁶ Each of these algorithms is controlled by a few *ad hoc* free parameters that must be tuned precisely. In particular, the number of output regions is critical; subsegmentation leads to misclassifications of regions including several classes of tissues, while oversegmentation increases the subsequent processing time and results in classification errors on tiny regions where reliable and robust descriptors cannot be obtained without statistical evidence. As wound size may vary significantly, the color pattern placed in the field of view also captures the image scale factor, a prerequisite for extracting statistical data from the image. Therefore, the free parameters were adjusted for each algorithm by comparison with the medical reference. A unique and fixed adjustment was set for each algorithm over the whole image database (Table 3).

To assess the efficiency of the segmentation over our wound database, we matched it with the medical reference built on clinician tracings. First, we assigned to each unlabeled region the class of tissue mainly represented in the corresponding area of the medical reference. Second, the agreement between the assigned labeling and the medical reference was measured by the OS introduced in Sec. 2 (Fig. 4). The OSs averaged on 50 corrected images of various pathologies of wounds are presented in the last column of Table 3. The high OSs confirm that segmentation is a valuable preprocessing step before classification for an automatic delineation of tissue areas.

We can see that the segmentation algorithm best suited for wound image processing is J-SEG (average OS of 73.1%) followed by the mean shift. A definitive validation

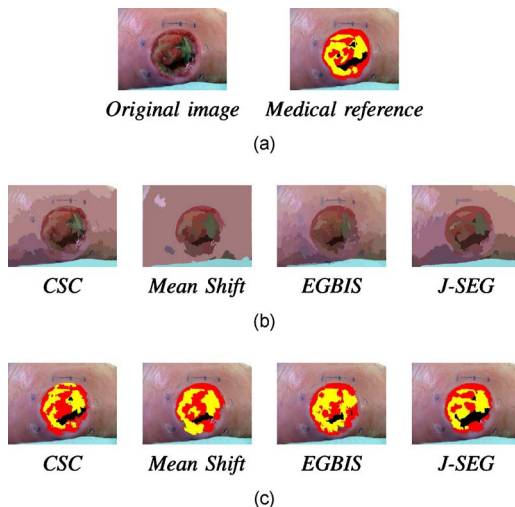


Fig. 4 Segmentation algorithm evaluation: (a) original image and medical reference (b), segmentation results, and (c) corresponding labels.

of this selection will take place after the automatic classification in Sec. 4, when comparing the classification rates obtained by each segmentation strategy. The power of the J-SEG algorithm resides in the separation of the segmentation process into two independent processing stages: color quantization adjusted by $q(0 \leq q \leq 600)$ and spatial segmentation controlled by $m(0 \leq m \leq 1)$.

As the OSs between clinician manual tracing and pre-segmented regions are quite relevant, a second interface was designed to allow clinicians to directly label pre-segmented regions automatically (Fig. 5). This interface is far less demanding for the experts as only simple mouse clicks on the regions are required. On the same basis as in Sec. 2, a practical medical reference was developed by merging the labeling results of several experts. A total of 850 significant tissue regions was selected among 1200 samples inside the database. The samples selected exhibit the usual types of tissues (48% granulation, 38% slough, and 14% necrosis on average, relative to the number of samples), with at least 75% agreement between the experts.

4 Tissue Classification

4.1 Feature Vector Generation

We focused on the color and texture descriptors already applied in dermatological imaging systems, especially for wound and ulcer tissues. Color is probably the most dominant image cue as demonstrated by the red/yellow/black healing visual assessment during clinical visits. The color

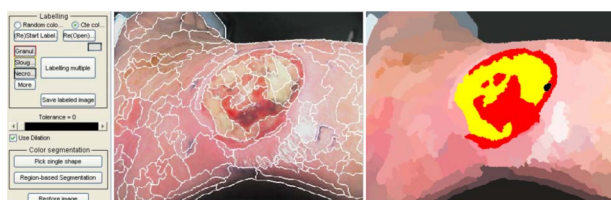


Fig. 5 Graphical interface for expert manual labeling.

Table 4 Color and texture descriptors.

Descriptor	Symbol	Space	Size	Time (s)
Histogram	h-RGB	3-D <i>RGB</i>	64	0.6
Histogram	h-LAB	3-D <i>Lab</i>	218	1.56
Histogram	h-HSV	2-D <i>HS</i>	256	1.99
Histogram	h-rg	2-D <i>rg</i>	256	1.60
Mean-dominant	MCD-DCD	<i>Lab/RGB</i>	8	0.41
LBP	h-LBP	GL	256	5.3
LBP	h-LBP	GL	59	1.5
Cooccurrence matrix	GLCM	RGB	15	0.48
Gabor filter	GAB	RGB	15	1.6
Contrast anisotropy	CA	GL	2	0.23

descriptors we have extracted are the mean color descriptor (MCD), the locally adapted dominant color descriptors (DCD) based on the Mean Shift iterative color clustering algorithm,³⁵ and 2-D and 3-D color histograms^{17,20,37} tested in different color spaces and sizes. Texture is also a relevant cue for tissue description if it is computed in a neighborhood properly scaled to the local structure. While clinical investigations indicate that it is less relevant than color, it nevertheless provides complementary data. Several well-known texture descriptors from color *RGB* components or gray-level image (GL) were extracted from the wound tissue regions. These descriptors are Gabor-based features (GAB) calculated on five scales, local binary pattern (LBP) histograms,³⁷ Haralick gray-level cooccurrence matrix features³⁸ (GLCM), and the normalized texture contrast and anisotropy (CA) (Table 4). The computation time stands for a 512×384 image processed by a PC Pentium 4 3.4-GHz CPU. It may vary slightly, depending on the number of regions and the actual size of the wound inside the image. To provide more consistent evaluation and efficiency of the image processing chain, the performance of these descriptors was also evaluated directly at the output of the classifier (see Sec. 5.3).

4.2 SVM Classifier Design

Initially designed for binary classification after supervised learning, support vector machines (SVMs) are also used for multiclass problems through one against all and one against one strategies.³⁹ Nonlinear class separation in low-dimensional space may result in smart separation in higher dimensional space, using a suitable kernel function. The key point of SVM classifier design remains the choice of the kernel function, mainly because no universal kernel will fit all applications, depending on the image database and input descriptors.⁴⁰ The SVM classifier selected here is a soft-margin algorithm (so-called C-SVM) available online at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. It has been tested by computing receiver operating characteristic (ROC)

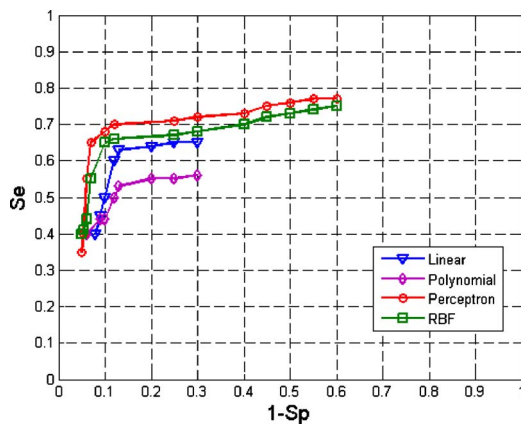


Fig. 6 ROC curves obtained by four different kernels.

curves for several classical kernels: linear, polynomial, radial basic function (RBF), and perceptron (Fig. 6). After the selection of a particular kernel, its parameters must be tuned. In the case of the perceptron kernel finally selected, there is only one free parameter; it controls the penalty of a classification error and it has been adjusted by a line search technique.

During the classifier learning stage, each set of samples was divided equally into training and test subset, then several iterations were applied through a k -fold cross-validation with $k=5$ to randomly select the training set before averaging the results. Each feature vector was tested with different settings of the classifier parameters during a backward and forward process. To evaluate the quality of a diagnostic method, it is common for the clinicians to compute predictive measures such as sensitivity (Se), specificity (Sp), success rate (Sr) and overall accuracy (Oa), which are reliable performance estimators.⁴¹ This enabled us to compare the discriminating power of the input descriptors (Table 5).

The multiple tests show that the best results are finally obtained by combining 23 attributes including MCD-DCD as the color descriptor and GLCM as the texture descriptor as input of the C-SVM classifier implemented with a perceptron kernel (parameter $C=100$). To evaluate the performance of the SVM approach more completely, we also compared it to other strategies, with the same descriptors as input. We selected other classical supervised methods, such as K -NN (K nearest neighbor) and fuzzy K -NN, and k -Means as an unsupervised one. As before, we divided the same data set randomly into training and test sets to program the classifier (Table 6).

We can see that unsupervised learning is inefficient in classifying the tissue samples into four different classes, whereas the supervised approach appears to be quite relevant for this kind of problem and thanks to its generalization ability, the discriminating power of SVM is unequaled.

5 Experimental Results

To validate this classification tool applied to chronic wound assessment, global statistical scores are not sufficient. To come closer to clinical measurements, two series of tests are described in the following subsections. In the first one, we examine separately the classifier performance in rela-

Table 5 Predictive power of several color and texture descriptors.

Feature	Descriptor	Se (%)	Sp (%)	Sr (%)	Oa (%)
Color	h-RGB	58	87	72	80
	h-LAB	66	87	76	82
	h-HSV	62	87	75	81
	h-rgb	57	86	72	80
	MCD-DCD	67	89	78	84
Texture	h-LBP	30	78	54	66
	m-LBP	29	77	53	66
	GLCM	54	82	68	72
	GAB	47	81	64	71
	CA	32	79	55	68
Col. + Tex.	MCD-DCD+GLCM	77	92	84	88

tion to each class of tissue by counting well-classified and misclassified regions. In the second one, we take into account the surface of each classified tissue region to compute the total percentage of well-classified areas over complete wound images.

5.1 Specific Tissue Classification

The experiment reported here was performed on the entire database. Ten iterations were again applied to select randomly the training set and the final results were averaged. To evaluate the degree of agreement between two judgments, the kappa statistical coefficient is widely used in the medical field, avoiding nonsignificant high scores when randomly voting.⁴² The results of Table 7 demonstrate the efficiency of the classifier against individual experts compared to the medical reference. Furthermore, it produces high scores on necrotic regions, which are critical in medical decisions, as immediate and invasive clinical treatment may be required if such regions are detected. An average kappa coefficient of 0.81 was obtained, which is considered as quite good agreement.

Table 6 Comparison of performance for several classifiers.

Classifier	Se (%)	Sp (%)	Sr (%)	Oa (%)
K-NN	63	86	75	80
Fuzzy K-NN	66	87	77	81
k-Means	39	77	58	68
SVM	77	92	84	88

Table 7 Kappa coefficient for tissue sample classification.

Tissue Class	Experts				SVM
	1	2	3	4	
Granulation	0.87	0.74	0.68	0.82	0.80
Slough	0.75	0.78	0.54	0.80	0.79
Necrosis	0.65	0.87	0.44	0.79	0.85
Average	0.75	0.79	0.55	0.80	0.81

5.2 Global Assessment on Wound Images

In the previous tests, the classification performance was examined globally for each type of tissue. Neither the wound that the sample came from nor the surface area of each tissue in the wound were taken into account. However, in clinical applications, the misclassification of a small area does not have the same impact as that of a large one. It is therefore important to measure the quality of assessment directly on several wounds, with area-weighted scores (Table 8).

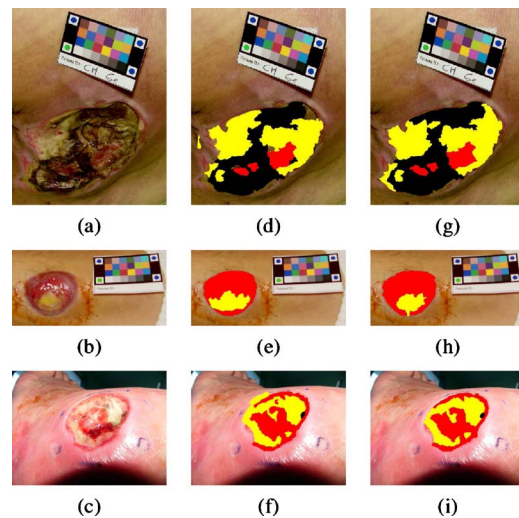
After averaging on the college of experts, it appears that the classifier performs better than a single expert. This is possible because the learning step relies on a reference established from several experts, and is thus based on stronger evidence. Figure 7 shows the classification results obtained on different wound images. Several observations can be made. First, the power of separation of tissue classes is constant over various locations and pathologies (sacrum pressure sore, leg ulcer, ankle wound). Second, it is not affected by the skin tones of the different patients or the lighting conditions encountered. Moreover, the classifier is not confused by the reflections on moist granulation regions.

5.3 Influence of the Preprocessing on Classification

It is interesting to verify that the adopted settings of the algorithms and the tuning of parameters do indeed optimize the output of the classifier. In the next subsection, we assess the influence of two major preprocessing options on the final classification results: color correction and segmentation algorithm. We compared the classification results for each tissue type, both with and without a color correction

Table 8 Averaged OSs of experts and classifier in percent compared to the medical reference.

Tissue Class	Experts	Algorithm
Granulation	64.1	85
Slough	62.1	70
Necrosis	81.1	83
Average	69.1	79.3

**Fig. 7** Three examples of automatic wound assessment compared to the medical reference: (a) to (c) original image, (d) to (f) automatic classification, and (g) to (i) medical reference.

step (Table 9). It is clear that the contribution of color correction is significant. Its effect is particularly noticeable for slough and necrosis tissues, which are very sensitive to color shifts. As we can see in Fig. 8, the automatic white balance control of the camera does not correctly compensate the color shift introduced by the ambient lighting and colored background. The segmentation results are clearly enhanced when this color shift is removed.

To select the most efficient segmentation algorithm, we calculated in Sec. 3 a virtual OS after assigning to each segmented region the class of tissue mainly represented in the corresponding area of the medical reference, due to the lack of automatic classification. A high OS was a valuable indication in helping to select one particular segmentation method, but it remained difficult to choose between methods with similar scores. The real influence of the segmentation step may be directly gauged at the output of the classifier. It demonstrates that the J-SEG algorithm is indeed the best choice on these wound images, followed by the mean shift (Table 10).

At a lower level, the tuning of the parameters of the segmentation algorithm may also have a significant impact on the classifier performance, in particular the average size of segmented regions (Fig. 9). This is controlled in the

Table 9 OSs in percent without and with color correction.

Tissue Class	No. Correction	Correction
Granulation	85	85
Slough	52	70
Necrosis	50	83
Healthy skin	84	92
Average	67	83

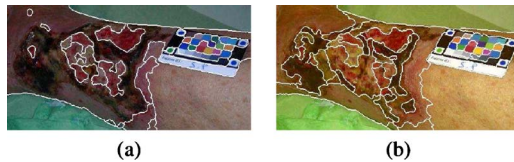


Fig. 8 Wound image segmented by the J-SEG algorithm (a) before and (b) after color correction. (Color online only.)

J-SEG algorithm by adjusting the two free parameters m and q (see Sec. 3). We compared the classifier performance obtained with different levels of segmentation and observed that the sensitivity (Se) rises from 66.2 to 80.6 ($m=0.4$ and $q=400$) when the number of regions increases from 350 to 1000 and that it drops slightly to 78.2 when the number of regions reaches 1500. The statistical measures needed to compute tissue descriptors are not relevant on tiny regions. A slight oversegmentation is preferable because in our strategy of classification-driven segmentation, subsegmentation inevitably leads to the misclassification of large areas. We found experimentally that a threshold of about 50 pixels ($\approx 1 \text{ cm}^2$ of tissue) is a good compromise.

6 Conclusion

For the development of remote wound assessment in telemedicine environments, a system based on image processing was designed to replace visual examination by clinicians with automatic tissue classification. The system design enables reproducible and standardized results required for such practice. First, particular attention was paid to ensuring independence from practical image acquisition conditions such as digital camera type, lighting, and viewpoint. This was achieved by color correction on input images and unsupervised image segmentation, combining color and texture analysis. Second, the automatic assessment reaches a human expert level of confidence. This was obtained by combining several expert results during the learning step and by the choice of discriminating tissue descriptors as input for the SVM classifier. By direct measurements at the output of the tissue classifier, we demonstrate that the preceding choices are relevant and enable a global enhancement of the system performance. An OS of 79.3% is obtained between classified tissues and the medical reference, which compares favorably with the average score of 69.1% obtained by a single clinician. Several perspectives of this work can be mentioned here. First, a specific

Table 10 OSs in percent for the four segmentation methods.

Tissue Class	J-SEG	Mean shift	CSC	EGBIS
Granulation	85	77	61	26
Slough	70	75	69	59
Necrosis	83	72	24	34
Healthy skin	92	86	70	66
Average	83	77	60	46

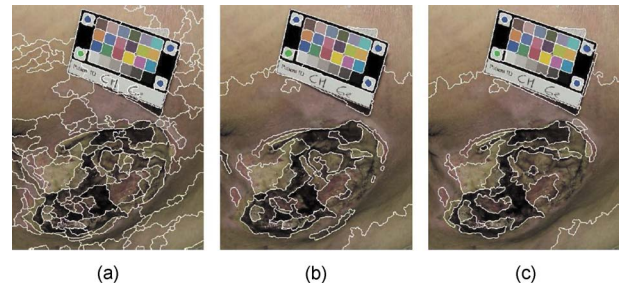


Fig. 9 Tuning of the segmentation level: (a) oversegmentation, (b) intermediate segmentation, and (c) subsegmentation.

color pattern limited to wound tones should enable sharper color correction with the same number of patches. The extension of the current wound database is also in progress for enhanced tissue characterization and clinical staff education. In practical terms, a website is now ready for wound image uploading by clinical staff. Another improvement is made possible by the wound 3-D reconstruction that we achieve by combining images from several camera viewpoints. This advanced tool for wound assessment enables us to compute the real tissue areas by mapping tissue classification results on wound 3-D surfaces.

Acknowledgments

This research program is supported by the European Social Fund and by the French Delegation for Research and Technology.

References

1. J. Lowery, J. Hamill, E. Wilkins, and E. Clements, "Technical overview of a web-based telemedicine system for wound assessment," *Adv. Skin Wound Care* **15**(4), 165–169 (2002).
2. D. Kosmopoulos and F. Tzevelekou, "Automated pressure ulcer lesion diagnosis for telemedicine systems," *IEEE Eng. Med. Biol. Mag.* **26**(5), 18–22 (2007).
3. P. Unger, C. Fife, and D. Weir, "Capturing the essence of the wound evaluation," *Today's Wound Clin.* **2**(2), 38–42 (2008).
4. M. Romanelli, G. Gaggio, M. Coluccia, F. Rizzello, and A. Piaggese, "Technological advances in wound bed measurements," *Wounds* **14**(2), 58–66 (2002).
5. T. A. Krouskop, R. Baker, and M. S. Wilson, "A noncontact wound measurement system," *J. Rehabil. Res. Dev.* **39**(3), 337–345 (2002).
6. A. Malian, A. Azizi, and F. A. Van Den Heuvel, "Medphos: a new photogrammetric system for medical measurement," *Int. Arch. Photogramm. Remote Sens.* **35**, 929–933 (2004).
7. M. Callieri, P. Cignoni, M. Coluccia, G. Gaggio, P. Pingi, M. Romanelli, and R. Scopigno, "Derma: monitoring the evolution of skin lesions with a 3-D system," in *Proc. 17th Int. Workshop on Vision, Modeling and Visualization*, pp. 167–174, Munich (2003).
8. B. Jones and P. Plassmann, "An instrument to measure the dimensions of skin wounds," *IEEE Trans. Biomed. Eng.* **42**(5), 464–470 (1995).
9. H. Oduncu, A. Hoppe, M. Clark, R. J. Williams, and K. G. Harding, "Analysis of skin wound images using digital color image processing: a preliminary communication," *Int. J. Low. Extrem. Wounds* **3**(3), 151–156 (2004).
10. P. Plassmann and T. Jones, "Improved active contour models with application to measurement of leg ulcers," *J. Electron. Imaging* **12**(2), 317–326 (2003).
11. J. Molnar, W. Lew, D. Rapp, E. Gordon, D. Voignier, S. Rushing, and W. Willner, "Use of standardized quantitative digital photography in a multicenter web-based study," *Eplasty* **9**(4), 20–45 (2009).
12. P. Plassmann and T. Jones, "Mavis: a non-invasive instrument to measure area and volume of wounds," *Med. Eng. Phys.* **20**, 332–338 (July 1998).
13. Y. V. Haeghen, J. M. Naeyaert, I. Lemahieu, and W. Philips, "An imaging system with calibrated color image acquisition for use in dermatology," *IEEE Trans. Med. Imaging* **19**(7), 722–730 (2000).

14. C. Serrano, A. Begona, R. Rangaraj, and L. M. Roa, "Segmentation-based lossless compression of burn wound images," *J. Electron. Imaging* **10**(3), 720–726 (2001).
15. M. Celebi, Y. Aslandogan, and P. Bergstresser, "Unsupervised border detection of skin lesion images," in *Proc. Int. Conf. on Information Technology: Coding and Computing (ITCC05)*, Vol. 2, pp. 123–128, IEEE Computer Society, Washington, DC (2005).
16. M. Celebi, H. Kingravi, H. Iyatomi, Y. Aslandogan, W. Stoecker, R. Moss, J. Malters, J. Grichnik, A. Marghoob, H. Rabinovitz, and S. Menzies, "Border detection in dermoscopy images using statistical region merging," *Skin Res. Technol.* **14**(3), 347–353 (2008).
17. M. Galushka, H. Zheng, D. Patterson, and L. Bradley, "Case-based tissue classification for monitoring leg ulcer healing," in *Proc. 18th IEEE Symp. on Computer-Based Medical Systems*, pp. 353–358 (2005).
18. T. Defloor and L. Schoonhoven, "Inter-rater reliability of the epup pressure ulcer classification system using photographs," *J. Clin. Nurs.* **13**, 952–959 (2004).
19. D. Beeckman, L. Schoonhoven, J. Fletcher, K. Furtado, L. Gunningberg, H. Heyman, C. Lindholm, L. Paquay, J. Verdu, and T. Defloor, "Epuap classification system for pressure ulcers: European reliability study," *J. Adv. Nurs.* **60**(6), 682–691 (2007).
20. P. Berris, "Acquisition of skin wound images and measurement of wound healing rate and status using color image processing," PhD thesis, The University of Reading, UK (2000).
21. A. Perez, A. Gonzaga, and J. Alves, "Segmentation and analysis of leg ulcers color images," in *Proc. Int. Workshop on Medical Imaging and Augmented Reality*, pp. 262–266, IEEE Computer Society, Washington, DC (2001).
22. H. Zheng, L. Bradley, D. Patterson, M. Galushka, and J. Winder, "New protocol for leg ulcer tissue classification from colour images," in *Proc. 26th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Vol. 1, pp. 1389–1392 (2004).
23. O. G. Cula, K. J. Dana, F. P. Murphy, and B. K. Rao, "Skin texture modeling," *Int. J. Comput. Vis.* **62**(1–2), 97–119 (2005).
24. B. Belem, "Non-invasive wound assessment by image analysis," PhD dissertation, University of Glamorgan, Wales (2004).
25. H. Mesa, L. Morente, and F. Veredas, "Tissue recognition for pressure ulcer evaluation," in *ECIFMBE Proceedings*, S.-V. Berlin, Ed., Vol. 22, pp. 1524–1527 (2008).
26. H. Sanada, T. Moriguchi, Y. Miyachi, T. Ohura, T. Nakajo, K. Tokunaga, M. Fukui, J. M. Sugaa, and A. Kitagawa, "Reliability and validity of DESIGN, a tool that classifies pressure ulcer severity and monitors healing," *J. Wound Care* **13**(1), 13–18 (2004).
27. G. Hansen, E. Sparrow, J. Kokate, K. Leland, and P. Iaizzo, "Wound status evaluation using color image processing," *IEEE Trans. Med. Imaging* **16**(1), 78–86 (1997).
28. H. Wannous, S. Treuillet, and Y. Lucas, "Supervised tissue classification from color images for a complete wound assessment tool," in *Proc. 29th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 6031–6034 (2007).
29. H. Wannous, Y. Lucas, and S. Treuillet, "Efficient SVM classifier based on color and texture region features for wound tissue images," in *Medical Imaging 2008: Computer-Aided Diagnosis*, M. L. Giger and N. Karsseneijer, Eds., *Proc. SPIE* **6915**, 69152T (2008).
30. H. Wannous, Y. Lucas, S. Treuillet, and B. Albouy, "Fusion of multi view tissue classification based on wound 3-D model," *Lect. Notes Comput. Sci.*, **5259**, 924–935, (2008).
31. H. Wannous, Y. Lucas, S. Treuillet, and B. Albouy, "A complete 3-D wound assessment tool for accurate tissue classification and measurement," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pp. 2928–2931 (2008).
32. B. Albouy, Y. Lucas, and S. Treuillet, "3-D modeling from uncalibrated color images for a complete wound assessment tool," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.* Vol. **2007**, pp. 3323–3326 (2007).
33. L. Priese and V. Rehrmann, "Fast and robust segmentation of natural color scenes," in *Proc. Asian Conf. on Computer Vision, Lect. Notes Comput. Sci.* **1351**, 598–606 (1998).
34. P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.* **59**(2), 167–181 (2004).
35. 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).
36. 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).
37. M. Kolesnik and A. Fexa, "Multi-dimensional color histograms for segmentation of wounds in images," *Lect. Notes Comput. Sci.* **3656**, 1014–1022 (2005).
38. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973).
39. C.-W. Hsu and C.-J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Trans. Neural Netw.* **13**(2), 415–425 (2002).
40. O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing kernel parameters for support vector machines," *Mach. Learn.* **46**, 131–160 (2001).
41. A. J. Alberg, J. W. Park, B. W. Hager, M. V. Brock, and M. Diener-West, "The use of 'overall accuracy' to evaluate the validity of screening or diagnostic tests," *J. Gen. Intern. Med.* **19**, 460–465 (2004).
42. J. Landis and G. Koch, "The measurement of observer agreement for categorical data," *Biometrics* **33**, 159–174 (1977).



Hazem Wannous received his DEA postgraduate degree in vision for robotic applications from the University of Clermont-Ferrand, France, in 2003, his MS degree in image processing and artificial intelligence from the University of Bourgogne, France, in 2005, and his PhD degree in image processing at the University of Orléans, France, in 2008. He then spent one year as an assistant professor under contract with the Ecole Polytechnique, University of Orléans. He is currently an associate researcher at the IMS Laboratory, University of Bordeaux. His research interests include machine learning, pattern recognition in both images and video, and geometric vision.



Sylvie Treuillet received her MS degree in electronics from the University of Clermont-Ferrand in 1988. For two years, she was with a private company in Montpellier developing image processing and pattern recognition methods in automatic cytogenetics. She then joined the research laboratory LASMEA and received her PhD degree in computer vision in 1993 from the University of Clermont-Ferrand. Since 1994, she has been assistant professor in computer sciences and electronic engineering at the Ecole Polytechnique of the University of Orléans, France. Now a member of the PRISME Institute, her research interests include color image processing, and computer vision for 3-D modeling, object recognition, and tracking for biomedical or industrial applications. She has supervised about fifteen master's theses or PhD dissertations and has published over 50 papers.



Yves Lucas received his MS degree in discrete mathematics from Lyon 1 University, France, in 1988 and his DEA postgraduate degree in computer science and automatic control from the National Institute of Applied Sciences in Lyon, France, in 1989. He then focused on the field of CAD-based vision system programming and received his PhD degree from INSA Lyon, France, in 1993. He subsequently joined Orléans University, France, where he was placed in charge of the vision group at the Vision and Robotics laboratory, which is centered on 3-D object reconstruction and color image segmentation. In 2008 he became a member of the PRISME Institute at Orléans University, where he is currently part of the Image and Signals for Systems research group. He has authored over 40 international journal and conference proceedings papers. His research interests include vision system design, learning and tuning, and pattern recognition and image analysis for medical, industrial, and robotic applications.