

# COMBINED MACHINE LEARNING WITH MULTI-VIEW MODELING FOR ROBUST WOUND TISSUE ASSESSMENT

Hazem Wannous

*IMS Laboratory, University of Bordeaux, Talence, France*

Yves Lucas, Sylvie Treuillet

*PRISME Institute, University of Orleans, Orleans, France*

**Keywords:** Tissue classification, 3D modeling, Machine learning, Wound assessment.

**Abstract:** From colour images acquired with a hand held digital camera, an innovative tool for assessing chronic wounds has been developed. It combines both types of assessment, colour analysis and dimensional measurement of injured tissues in a user-friendly system. Colour and texture descriptors have been extracted and selected from a sample database of wound tissues, before the learning stage of a support vector machine classifier with perceptron kernel on four categories of tissues. Relying on a triangulated 3D model captured using uncalibrated vision techniques applied on a stereoscopic image pair, a fusion algorithm elaborates new tissue labels on each model triangle from each view. The results of 2D classification are merged and directly mapped on the mesh surface of the 3D wound model. The result is a significative improvement in the robustness of the classification. Real tissue areas can be computed by retro projection of identified regions on the 3D model.

## 1 INTRODUCTION

Wound assessment process is based on visual examination, in order to identify different tissues such as granulation, slough, and necrotic ones. The wound is usually described by its spatial measurements and the colours of its tissues, providing an important indication of its types and thus the particular stage of healing. The monitoring of the wound healing process represents a difficult task for clinicians and nurses, where it is necessary to assess the different tissue types on consecutive visits.

The clinical assessment of chronic wounds still essentially rely on manual tedious and expensive practices, which do not produce objective measurements and quantitative assessment of healing. Recently, research has focused on the analysis of the wound images in order to develop quantitative non invasive measurement with image processing techniques for monitoring. However, they addressed separately the problems of wound shape capture and tissue classification. Attempts to extract automatically the wound area using colour measurements did not completely succeeded and semi-automatic methods were preferred (Oduncu et al., 2004). Furthermore,

the results obtained on several colour spaces by direct classification on the pixels were still not acceptable, even when combining several colour and texture parameters to describe the tissues (Kolesnik and Fexa, 2004). The region-based classification approach has been discussed by (Zheng et al., 2004), but the tissue samples have been manually extracted as squared homogeneous regions of interest and finally, the assessment has been partially achieved (classification between only two types of tissues).

Concerning the spatial measurements, some prototypes based on structured light techniques (Krouskop et al., 2002) or photogrammetry (Malian et al., 2002) were tested, but these cumbersome and complex systems were not adapted to the clinical practice which requires a low cost, handy and simple tool operated by a nurse. The second version of the MAVIS system (Jones et al., 2006) which uses only a reflex digital camera equipped with special dual lens optics can record two half images from slightly different viewpoints on a single shot. But it suffers from several drawbacks: a costly digital reflex camera is required to adapt the special dual lens with extra cost, the very close viewpoints do not enable accurate 3D measurements and finally, the tissue classification

problem has not been addressed. In the Derma project (Callieri et al., 2003), wound measurements and tissue classification have been both tackled, as it enables shape and natural texture grabbing, but the classification process remains user assisted as seeds need to be manually pointed inside the wound and a similarity measure adjusted to control the merging process. Moreover, it is based on a costly Minolta 3D scanner, forbidding totally its spreading in clinical staff.

In contrast to the aforementioned methods dealing separately with wound shape and tissue classification tasks, we propose in this paper an original approach using a sharp 3D model of the wound to ensure robust classification of its tissues. Furthermore, a smart training of the classifier over tissue samples extracted automatically will be discussed. We mainly focus on the integration of geometrical structure of the wound in classification process, which improves wound assessment and gives access to real measurements. The paper is organized as follows: Section 2 presents the constitution of the sample data as a pre-processing step of the classification method described in section 3. Section 4 discusses the multi view approach. The improvement of classification is presented in section 5 before to conclude in the last section.

## 2 WOUND SAMPLE DATABASE

A database of chronic wound images has been constituted with the help of the clinical staff in several hospital centers, in order to get a exhaustive set of images for different types of tissues. Furthermore, a variety of types of chronic wounds has been collected from different care services, such as a leg ulcers, diabetic lesions, bed sores, etc. Several hundreds of colour images (3 Mpixels, 24 bits) have been taken, by different digital cameras under uncontrolled illumination conditions, with respect to a specific protocol integrating several points of views for each single wound.

Colour pre-segmentation provides an automatic delineation of tissue samples and simplifies the following classification step by extracting more robust and discriminant local attributes on tissue area than with direct pixel classification. JSEG algorithm (Deng and Manjunath, 2001) has been selected as it has been proved that it is the more efficient compared to three other advanced methods for unsupervised tissue segmentation (Wannous et al., 2007). A graphical interface allows clinicians to directly label automatically pre-segmented regions. Next, a unique medical reference is elaborated by merging the region labels of the college of experts. This practical medical reference has been used exclusively for all tasks in the

following, in particular in the classifier learning step.

The pre-segmented image database has been provided to a group of clinicians in order to label it according to the classical colour code: red for granulation, yellow for slough and black for necrosis. Labelling realized by the different clinicians have been merged to get a unique and reliable medical reference by applying a majority vote criterion according to a given tissue class. We retained for each tissue the pixels with confidence level greater than or equal to 75%. Following machine learning and algorithm evaluation will be based on this medical reference.

Labeling sessions repeated one month apart by the clinicians confirm that the identification of tissues is a subjective task, as the obtained overlap scores for these tests remain moderate (58% to 85%) and also each clinician does not produce similar labels one month later (65% to 91%) (Wannous et al., 2007). Therefore, the multi expert medical reference is a solution to build a more robust and non-subjective tissue sample database from automatically segmented regions.

## 3 MACHINE LEARNING SYSTEM

To enable a sharp discrimination among the tissue classes, different types of colour and texture descriptors have been calculated on the sample database, composed of four types of tissues. Consequently, relevant descriptors have been searched for the learning step of a supervised classifier thanks to manual labeled tissue samples from the wound database.

### 3.1 Colour and Texture Features

To characterize each tissue class more accurately, a total of 850 significant tissue regions has been extracted in the wound image database from the segmentation phase of the wound images. These regions correspond to the four known types of tissue identified manually (30% Granulation, 24% Slough, 9% Necrosis and 37% Healthy). We have only tested the most common descriptors for the dermatological applications, especially those concerning wound and ulcer.

The extracted colour descriptors are: Mean Colour Descriptor (*MCD*), locally adapted Dominant Colour Descriptors (*DCD*) (Deng et al., 1999) calculated using *Mean Shift* colour clustering algorithm (Comaniciu and Meer, 2002) and 2D/3D histograms in different colour spaces. The extracted texture descriptors are: Gabor based features (*GAB*), Local Binary Pattern histogram calculated from the Gray Level wound image GL (Ojala et al., 2002), Haralick

Gray Level Co-occurrence Matrix features (*GLCM*) and the normalized texture contrast and the anisotropy (*CA*) computed from the second moment matrix, derived from the gradient of the GL image (Carson et al., 2002).

To evaluate the discriminating power of the descriptors, we measure directly the classification rates at the classifier output. This provides more consistent evaluation and more efficiency in the process. Moreover, our classifier integrates the correlation between the descriptors and avoids data reconditioning.

### 3.2 Classifier Parameters Tuning

The performance of the classifier depends strongly on the selection of appropriate kernel functions and the setting of their parameters, but also on the pertinence of the input descriptors. Consequently, we must test all feature vectors with different settings of the classifier parameters. The samples have been divided equally into two subsets for test and training, in order to evaluate the different descriptor vectors. As the training subset needs to provide a complete and representative description of the tissue classes, several iterations are applied to randomly select the training set and the final results are obtained by averaging on all iterations.

Concerning the classifier, we have selected soft-margin SVM algorithm (so-called C-SVM) tested with different classical kernels: linear, polynomial, Radial Basic Function (RBF) and perceptron kernel (Lin and Li., 2005). The adopted C-SVM is a multi-class classifier based on the *Error Correcting Output Codes* framework (Huang et al., 2006) which provides, more than the labels, the probability estimates of belonging to a class. These probability estimates are used later by the fusion algorithm to label a region of the 3D model.

After the selection of a particular kernel, a regularization parameter ( $C$ ), which controls the penalty of the classification errors, must be tuned. In the case of linear or perceptron kernels, we have only to optimize this single parameter but in the case of the RBF and polynomial kernels, a second parameter has to be tuned (resp.  $\alpha$  and  $\theta$ ). For the tuning of these parameters, we used the *line search* technique for the two first kernels and the *parallel grid search* technique for the two others, combined with *k-fold* cross validation with  $k=5$  (Chapelle et al., 2001). Figure 1 illustrates the setting of hyper parameters ( $C$  and  $\alpha$ ) of C-SVM classifier with RBF kernel by parallel grid search technique and the ROC curve obtained for each kernel. The search intervals were  $[2^{-5}, 2^{15}]$  for the regularization parameter  $C$  and  $[2^{-15}, 2^3]$  for the

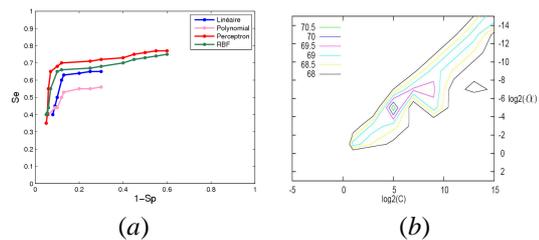


Figure 1: Classifier design: (a) ROC curve obtained by 4 different kernels (b) hyper-parameters setting of C-SVM classifier with RBF kernel. The training set is obtained by combining MCD-DCD with GLCM descriptors.

kernel parameter ( $\alpha$ ).

The selection of descriptors thanks to the evaluation based-classifier result is presented in Table 1. The classification accuracy is expressed by four predictive measures, commonly used by clinicians to evaluate the quality of a diagnostic method. These measures are sensitivity (Se), specificity (Sp), Success rate (Sr) and Overall accuracy (Oa). These results are expressed as the averaged of each predictive measure over the 4 classes. The multiple experimental tests between all color/texture descriptors show that best results can be obtained by MCD-DCD as a colour descriptor and GLCM as a texture descriptor, using C-SVM classifier with perceptron kernel (Wannous et al., 2010).

These results indicate that texture is less relevant than colour for wound tissue discrimination. However, it provides complementary information and, therefore, is significant. In this way, texture and colour information of tissue wound can be combined to achieve better classification accuracy. The quality of the tissue classification has been validated over a series of 50 wound images by computing Kappa coefficients between the medical reference provided by the experts and automatic classification (see Table 4). These coefficients are close to those obtained between clinicians and medical reference.

## 4 MULTI VIEW APPROACH

The 2D classification results still suffer from a significant drawback, as it has been established that a deviation of  $20^\circ$  of the optical axis from the normal of the wound typically leads to an underestimation of surface around 10% (Treuillet et al., 2009). This is due to lighting variations which modify the colours and perspective effects from distant viewpoints which induce significant bias in the classification results (see 2D classification in Figure 4) and then do not allow computing real surfaces. Like the clinician, who dis-

Table 1: Predictive power of colour and texture descriptors.

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

poses of many observation points to provide a reliable diagnosis, a multi view technique should allow more robust results. So, we propose to use the dimensional information captured from a multi view model because reliable wound assessment must provide reproducible results, regardless to the position and orientation of the camera.

#### 4.1 View-dependent Classification

Wound images have been taken from different points of view. 3D model has been obtained using uncalibrated vision techniques completed by original refinements to obtain semi-dense matching between widely separated views (Albouy et al., 2006).

Clinicians establish their diagnosis visually on the photographed wound, with the help of a red-yellow-black scale placed in the camera field, corresponding to the three types of tissue. However, this diagnosis is also based on their observations of the wound during the patient visit. Then, the clinician assessment of tissues can be seen as a combination of the colorimetric information (image plane) with shape information (through observation of the human eye). So, we can illustrate the dependance of the classification assessment on the point of view by a simple projection on a 3D model computed from two views of a wound. To do this, the classification result from each one of the single view have been mapped on the 3D model separately in order to label the triangular mesh. Each triangle is labeled according to its higher score and then the surface of each tissue type can be computed by summing the triangles belonging to the same tissue class. Figure 2 shows the variation of the cartography mapping on the an ulcer 3D model according to the classification results coming from single view approach.

The 3D model allows accurate comparison of single-view classifications since the differences are

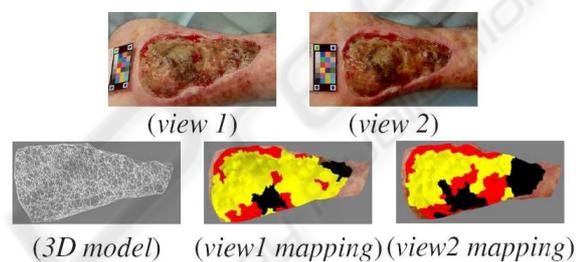


Figure 2: Dependence of the classification result on the point of view.

expressed in  $cm^2$  and not in pixel. Table 2 presents the area of the surfaces calculated in  $cm^2$  for each tissue type, when mapping the image plane on the 3D model presented in Figure 2.

Table 2: Measured real surfaces for each type of tissue when mapping separately the classification results obtained from two views of the same wound on the 3D model.

Tissue class	3D surface ( $cm^2$ )	
	Accord. to view1	Accord. to view2
Granulation	12.9	18.2
Slough	36.9	27.4
Necrosis	4.2	8.4

The obtained differences reflect the effect of perspective projection in the image and the relief of skin ulcers. This experiment confirms the limitations of the single-view approach and the need to take into account the 3D aspect. So it is possible to fusion the results of tissue labeling coming from each image.

#### 4.2 Merging of the Classification Results

Based on the 3D reconstruction of the scene, the main idea is to combine the colour information of the re-

gions, the calculation of points of view and the relief in order to get a classification more robust and also access to real surfaces. To do this, pictures of the wound have been taken from different point of views and a 3D model of 3000 to 4500 matches have been obtained in  $1024 \times 768$  image pairs, allowing the matching between widely separated views. It is so possible to match homologous regions in each view and to fusion classification results.

The 3D mesh is projected on the stereo pair to provide a 2D Delaunay mesh of triangles in each image (Figure 3). Due to the point correspondences between the two images, each triangle in the left image has a homologous one in the right image. So it is possible to fusion the results of tissue labeling coming from each image.

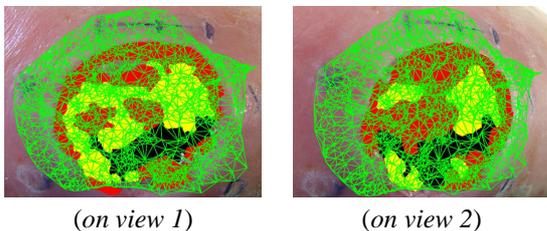


Figure 3: Triangular model projected on the classification results.

The strategy we have experimented is based on sequential multi-criterion tests with recursive splitting of the triangles<sup>1</sup>. It takes into consideration the triangle area, its dominant class, the class probabilities at the classifier output and the solid angle of the cone generated from the optical center and for which the triangle is a cross section (cf. Algorithm 1).

## 5 EXPERIMENTAL RESULTS

To analyze the management of the triangle labeling process through the fusion algorithm, we perform it on fifteen pairs of wound images. The Table 3 gives the numbers of triangles labeled at each step of the fusion algorithm applied on the 3D model of the Figure 3. It appears clearly that for more than half of model surface, the classification results dependent strongly on the viewpoint where only 40% of the total surface were labeled with the same class in both views. Only a few triangles are concerned with the splitting step; this is because of the semi-dense 3D model in our matching process. However, about 20% of the wound model area is labeled according to solid angle

<sup>1</sup>the triangle is recursively split along the median line of its longest side

```

Data: classification results on two views +
          3D model
Result: multi view classification
0.1 project the 3D triangular model on each
    view;
0.2 forall triangles of the 3D model do
0.3   find tissue percentage on each view;
0.4   if the two views have the same tissue
    class then
0.5     label triangle with common tissue
    class;
0.6   else
0.7     compute 3D triangle surface;
0.8     if surface > threshold then
0.9       split the triangle & go back to
        0.2;
0.10    else
0.11     identify dominant tissue in each
        view;
0.12     if same majority class then
0.13       label triangle with this
        majority class;
0.14     else
0.15       compute the solid angle in
        each view (S1, S2);
0.16       if  $S1 \gg S2$  then
0.17         label triangle
        corresponding to 1st
        view;
0.18       else if  $S2 \gg S1$  then
0.19         label triangle
        corresponding to 2nd
        view;
0.20       else
0.21         compute the class
        probabilities in each
        view;
0.22         label triangle by max
        probability class;
0.23       end
0.24     end
0.25   end
0.26 end
0.27 end

```

Algorithm 1: Fusion algorithm.

criteria. Finally, for about 35% of the model surface, the two criteria of dominant class and probability estimates need to be computed. Such analysis attests of the fact that the classification is view dependent, which implies to combine several viewpoints.

As the pertinence of the fusion algorithm is demonstrated. Since we are more interested in the performance of our method in practical applications, by testing its accuracy on the classification results and

Table 3: Number of triangles labeled at each step of the fusion algorithm.

	<i>Common class</i>	<i>After splitting</i>	<i>Dominant class</i>	<i>Solid angle</i>	<i>Probability estimates</i>
<i>Nbr of triangles</i>	1280	77	540	700	561
<i>Percentage</i>	40.5 %	2.4 %	17.0 %	22.2 %	17.9 %

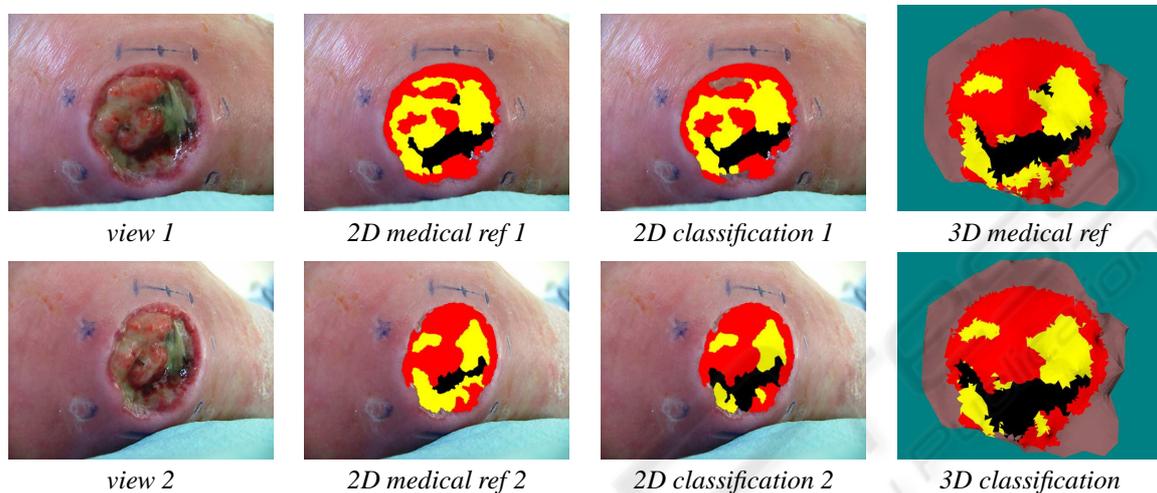


Figure 4: Multi view classification.

its advances for tissue wound assessment application. To evaluate the improvement due to the fusion of 2D classification, we compared the scores between 2D medical reference and 2D automatic classification on one part and the overlap scores between 3D medical reference and 3D classification results on the other part (Table 4). The 3D medical reference is simply the result of the fusion of the medical references coming from the left and right images, mapped on the 3D model. In this case, the class probability is replaced by the level of confidence obtained from manual expert labeling.

Table 4: 2D/3D Overlap scores and Kappa coefficients averaged over the wound database.

Tissue class	Overlap score (%)		Kappa coef.	
	2D	3D	2D	3D
Granulation	79.8	81.4	0.82	0.84
Slough	69.3	72.0	0.75	0.77
Necrosis	60.7	67.9	0.73	0.77
Average	69.9	<b>73.8</b>	0.77	<b>0.79</b>

The Kappa coefficient is also commonly used to compute the degree of agreement between two medical judgements (Landis and Koch, 1977). This statistical indicator, varying between 0 and 1, can be calculated by  $K = \frac{P_o - P_e}{1 - P_e}$  where  $P_o$  is the relative observed agreement and  $P_e$  the hypothetical probabil-

ity of chance agreement, using the observed data to calculate the probabilities of each observer randomly voting for each category. We therefore calculated the Kappa coefficient in 2D approach (between classifier and 2D medical reference) and 3D approach (between 3D classification and 3D medical reference) Table 4.

We observe that the agreement between medical reference and automatic classification is globally improved after the fusion step. The convergence of results from 2D to 3D approach is verified not only globally but separately on each type of tissue. As everything depends on the relevance of our 3D medical reference, combining assessments from various points of view leads logically greater robustness. The improved performance of multi view classification is visible in Figure 4, where some areas of the wound were misclassified in one of two views before the fusion. We are observing a convergence of the classifications from the single view to the multi view case, which is expressed by a better agreement between manual labeling and automatic classification.

These tests show that the fusion of 2D classifications enables more accurate tissue classification. Moreover, as the results can be mapped on the mesh surface of the wound 3D model, real tissue surfaces can be computed on it.

## 6 CONCLUSIONS

Machine learning based on SVM classifier with region descriptors as input has been improved by multi view management. We have combined the region classification results coming from several 2D images of a skew surface, using the matched vertices of the reconstructed 3D model. This approach has been applied to the design of a complete 3D and colour wound assessment tool. Experimental results show that the fusion of 2D classification enables more accurate tissue classification. Moreover, as the results can be mapped on the mesh surface of the wound 3D model, real tissue surfaces and volumes can be computed on it. Future works include several tests on a larger image database. We also intend to improve these results by matching regions from more than two views and by testing colour descriptors invariant to viewpoint and lighting conditions.

## REFERENCES

- Albouy, B., Koenig, E., Treuillet, S., and Y.Lucas: (2006). Accurate 3d structure measurements from two uncalibrated views. In *ACIVS*, pages 1111–1121.
- Callieri, M., Cignoni, P., Coluccia, M., Gaggio, G., Pingi, P., Romanelli, M., and Scopigno, R. (2003). Derma : monitoring the evolution of skin lesions with a 3d system. In *8th Int. Work. on Vis. Mod. and Visualization*, pages 167–174.
- Carson, C., Belongie, S., Greenspan, H., and Malik, J. (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. PAMI*, 24(8):1026–1038.
- Chapelle, O., Vapnik, V., Bousquet, O., and Mukherjee, S. (2001). Choosing kernel parameters for support vector machines. *Machine Learning*, pages 131–160.
- Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Trans. PAMI*, 24(5):603–619.
- Deng, Y., Kenney, S., Moore, M., and Manjunath, B. S. (1999). Peer group filtering and perceptual color image quantization. In *IEEE Inter. Symp. on Circ. and Sys. VLSI (ISCAS'99)*, volume 4, pages 21–24, Orlando, FL.
- Deng, Y. and Manjunath, B. S. (2001). Unsupervised segmentation of colour-texture regions in images and video. *IEEE Trans. PAMI '01*, 23:800–810.
- Huang, T., Weng, R. C., and Lin, C.-J. (2006). Generalized bradley-terry models and multi-class probability estimates. *Journal of Machine Learning Research*, 7:85–115.
- Jones, C. D., Plassmann, P., Stevens, R. F., Pointer, M. R., and McCarthy, M. B. (2006). Good practice guide to the use of mavis ii. Technical report, Medical Imaging Research Unit TR-07-06, Univ. of Glamorgan.
- Kolesnik, M. and Fexa, A. (2004). Segmentation of wounds in the combined color-texture feature space. *SPIE Medical Imaging*, 5370:549–556.
- Krouskop, T. A., Baker, R., and Wilson, M. S. (2002). A noncontact wound measurement system. *J Rehabil Res Dev*, 39(3):337–345.
- Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.
- Lin, H.-T. and Li, L. (2005). Infinite ensemble learning with support vector machines. *J. Gama et al., eds., Machine Learning: ECML '05, Lecture Notes in Artificial Intelligence*, 3720:242–254, Springer-Verlag.
- Malian, A., Heuvel van den, F., and Azizi, A. (2002). A robust photogrammetric system for wound measurement. In *International Archives of Photogrammetry and Remote Sensing*, volume 34, pages 264–269, Corfu, Greece.
- Oduncu, H., Hoppe, H., Clark, M., Williams, R., and Harding, K. (2004). Analysis of skin wound images using digital color image processing: a preliminary communication. *Int J Low Extrem Wounds*, 3(3):151–156.
- Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. PAMI*, 24(7):971–987.
- Treuillet, S., Albouy, B., and Lucas, Y. (2009). Three-dimensional assessment of skin wounds using a standard digital camera. *IEEE Trans. on Medical Imaging*, 28:752–762.
- Wannous, H., Treuillet, S., and Lucas, Y. (2007). Supervised tissue classification from color images for a complete wound assessment tool. In *29th Inter. Conf. of the IEEE Eng. in Med. and Bio. Soc. EMBS'07*, pages 6031–6034.
- Wannous, H., Treuillet, S., and Lucas, Y. (2010). Robust tissue classification for reproducible wound assessment in telemedicine environments. *to appear in the Journal of Electronic Imaging*.
- Zheng, H., Bradley, L., Patterson, D., Galushka, M., and Winder, J. (2004). New protocol for leg ulcer tissue classification from colour images. In *26th Inter. Conf. of the IEEE Eng. in Med. and Bio. Soc. EMBS'04*, volume 1, pages 1389–1392.