International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:11, 2012

Fusion of Colour and Depth Information to Enhance Wound Tissue Classification

Darren Thompson, Philip Morrow, Bryan Scotney, and John Winder

***Abstract*—**Patients with diabetes are susceptible to chronic foot wounds which may be difficult to manage and slow to heal. Diagnosis and treatment currently rely on the subjective judgement of experienced professionals. An objective method of tissue assessment is required. In this paper, a data fusion approach was taken to wound tissue classification. The supervised Maximum Likelihood and unsupervised Multi-Modal Expectation Maximisation algorithms were used to classify tissues within simulated wound models by weighting the contributions of both colour and 3D depth information. It was found that, at low weightings, depth information could show significant improvements in classification accuracy when compared to classification by colour alone, particularly when using the maximum likelihood method. However, larger weightings were found to have an entirely negative effect on accuracy.

***Keywords*—**Classification, data fusion, diabetic foot, stereophotogrammetry, tissue colour.

I. Introduction

ULCERATION of the lower limb is a debilitating, expensive and potentially fatal consequence of diabetes. Current diagnostic and assessment measurements are subjective and vary between clinicians [1]. Common practice is to focus on two-dimensional (2D) parameters such as measurements of perimeter or area, while research suggests that making use of the three-dimensional (3D) nature of wounds may be more useful [2]. A 3D approach which includes knowledge of the different tissues within the wounds [3] should provide a more complete wound assessment. This study investigated the fusion of depth and colour information in the classification of tissues contained within simulated wounds. The purpose of this work was to assess how this approach of including depth information would impact upon the classification process. This study made up part of a wider wound-assessment project by aiding the development of a methodology for evaluating real clinical data.

Darren Thompson is a PhD student within the School of Computing and Information Engineering at the University of Ulster, UK (phone: +44 28 7012 4698; e-mail: thompson-d6@email.ulster.ac.uk).

Dr. Philip Morrow is a Reader within the Computer Science Research Institute at the University of Ulster, UK (phone: +44 28 70124637; email: pj.morrow@ulster.ac.uk).

Prof. Bryan Scotney is the Director of the Computer Science Research Institute at the University of Ulster, UK (phone: +44 28 70124648 email: bw.scotney@ulster.ac.uk).

Dr. John Winder is a Lecturer in Clinical Physiology within the Institute of Nursing and Health Research at the University of Ulster, UK (phone: +44 28 90368440 email: rj.winder@ulster.ac.uk).

Fig. 1 An example of a diabetic foot wound featuring granulation tissue (red), slough (yellow) and necrosis (black)

II. Background

Foot ulceration (Fig. 1) tends to develop due to vascular or neuropathic deficiencies as a result of both type I and type II diabetes. Patients suffering from *diabetes mellitus* are thought to have a lifetime risk as high as 25% of developing a chronic lower limb wound, with a yearly infection rate of 36.5 per 1000 [4]. Ulceration and subsequent infection account for approximately 25% of all diabetes-related hospital admissions in the UK and USA [5], while 15% of foot ulcers result in amputation. Extreme cases of infection may even be fatal [6]. Diagnosis and treatment of diabetic wounds often lead to long hospital stays and multiple operations which have a significant economic impact on health services. Better management of wound infection would both lower morbidity and reduce costs to communities and healthcare systems [4]. The increase in prevalence of type II diabetes due to rising levels of obesity and infection from MRSA only exacerbate the problem [4].

Clinical wound measurements tend to include estimates of wound length, width or area [6]. These measurements may be made on the wound itself or using traditional photographs. Temporal changes in these parameters can indicate probable healing rates or likely clinical outcomes. They may also indicate that current treatment is unsuitable and should be changed or adjusted accordingly [3]. In the often extended process of wound healing, quantification of the healing rate is important in assessing the efficacy of treatments [7]. Measurement and assessment of wounds can vary greatly between clinicians due to the highly subjective methods available to them [1], which typically include measurements using rulers and wound tracings [8]. Additionally, Stremitzer *et al.* [1] conclude that objective assessment of wound colours is vital and that tissue evaluation is at least as important as wound size.

Some studies have also highlighted the importance of a 3D aspect to wound assessment. Reduction in wound volume associated with the development of granulated tissue in the wound bed is noted as more significant to wound healing by

Flanagan [3], as compared to reduction in area by the growth of epithelial tissue. Therefore the relative amounts of different tissue types may be inherently linked to the three-dimensional nature of wounds through properties such as their surface area and volume.

Stereophotogrammetry is an imaging technique which builds 3D surfaces from 2D images. Common points are automatically tied together from one or more stereoscopic image pairs to produce a virtual 3D surface which includes colour photographic information. One such system is Di3D, developed by Dimensional Imaging, Glasgow (http://www.di3d.com). The system has been validated for linear measurements on 3D surfaces in various studies [9]-[11] and for volume measurement by Thompson *et al.* [11]. Levels of accuracy noted in each of these studies suggest that Di3D may allow rapid, relatively inexpensive, non-contact 3D assessments of wounds.

Algorithms for automated classification of images are growing in importance in medical environments. They are already employed throughout medical imaging to identify anatomical structures and regions of interest [12]. Broadly speaking, two main categories of classification algorithm exist, known as *supervised* and *unsupervised*. Supervised algorithms classify image pixels according to class parameters such as mean and variance, which are estimated from training data samples provided by the user. Unsupervised algorithms operate by iterating between classifying the pixels and updating the parameters [12], clustering data according to the distribution of pixel values within the image [13]. Thus, unsupervised algorithms effectively train themselves to extract the information which would be provided by the user in a supervised method. A traditional reliance on spectral information means that image classification tends not to make use of other available knowledge. For example, it is known that granulation tissue is a deep red colour, whilst it is also known that such tissue typically occurs in the base of a wound. 3D depth information may aid the classification of wound tissues according to their colours.

Wannous *et al.* [14][15] developed a 3D colour wound assessment tool for the imaging and classification of wound tissues as part of the ESCALE project. A single camera was used to take images at different angles, therefore using a more time-consuming form of stereophotogrammetry. Tissue classification was performed along with surface area and volume measurement. However, 3D surfaces were of relatively low resolution, reducing the accuracy of these measures.

In image classification it is practical to perform extensive evaluation of algorithms on simulation data, especially in the absence of real clinical examples. When real data are limited, it would be unwise to 'tune' algorithms to perform well on a few examples, only to find a large variation when more data are gathered. Simulated data allows the testing of algorithm performance under known and controlled conditions, which are easily manipulated and simple to reproduce [16]. Simulation also allows an exactly known segmentation, or *ground truth*, against which the results may be compared. This is rarely a feature of classification performed on real data. This study therefore required simulated data which modelled wound tissue colours as realistically as possible.

It was the aim of this work to develop a 3D wound assessment method which performed quickly and produced high resolution colour 3D surfaces without the requirement for brightness correction in a research setting. The goal was the ability to make accurate, reliable and objective measurements of tissue types, surface area and volume to provide a complete picture of wounds and their healing processes. This study explored the potential for 3D tissue classification.

III. Methodology

A. Classification algorithms

This study made use of one algorithm from each classification category: a supervised Maximum Likelihood Classifier (MLC) and a form of the unsupervised Expectation Maximisation (EM) algorithm. Of the many methods available, the two selected were chosen for their relative simplicity, popularity and established effectiveness. The chosen methods are also widely known and understood within the field of image classification. The accommodation of 3D depth information into the classification process, and its effects, could be suitably demonstrated via these methods. This study compared the performance of MLC and EM algorithms on simulated wound images while weighting the contribution made by a corresponding depth map.

MLC assumes a normal distribution of values within each class. The probability that a pixel $\mathbf{x}$ belongs to class ω_i is given by

$$p(\omega_i|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right] \quad (1)$$

where $\boldsymbol{\mu}_i$ is the spectral mean for class ω_i, $\boldsymbol{\Sigma}_i$ is the corresponding variance-covariance matrix and d is the number of dimensions to the data. The parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are estimated from training samples. From each simulated image, training sample areas of approximately 5000 pixels were selected for each tissue class, making sure to cover the full range of depth. Spectral information was classified according to this multivariate form of the MLC (d=3). Probabilities from depth were calculated from the univariate form, where $\boldsymbol{\mu}_i$ and

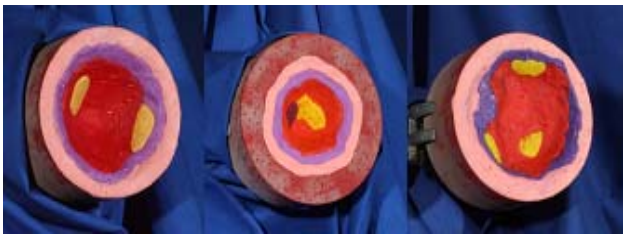

Fig. 2 Models A, B and C

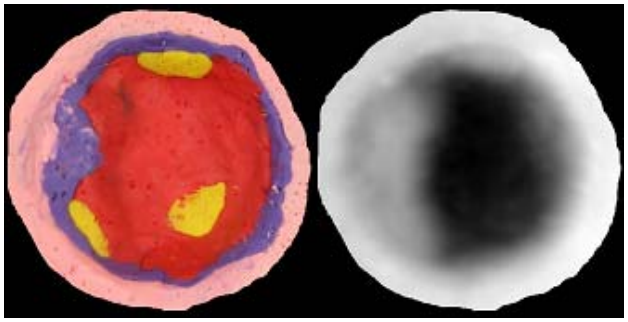

Fig. 3 Colour map and corresponding depth map from model C

$\boldsymbol{\Sigma}_i$ reduce to single scalar values (d=1). Hence two separate probabilities of a pixel belonging to each class were calculated: one each from colour and depth. Probability arrays for each class and each pixel were produced, and values were normalised so that class probabilities for each pixel would sum to unity.

The EM classification was performed similarly to the Multi-Modal EM (MMEM) algorithm outlined in Hong *et al.* [17], considering the sources of depth and colour information to be two separate modalities. However, in that study, class mean and variance values were set to initialise the algorithm before proceeding with iteration. In this study, randomly generated probabilities provided initialisation, as in the approach by Fraley & Raftery in [18]. The algorithm then iterated between the E-step of classification according to the probabilities, and the M-step of calculating new parameters, including $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$. A normal distribution was also assumed, calculating new probabilities similarly to the MLC algorithm, according to (1). Convergence was deemed to have taken place when no pixel classification changed between successive iterations.

For both algorithms, the separate probabilities from depth and colour were combined using a weighted average to obtain a final probability *Z*. In this way, the contribution of the depth information to the overall classification could be controlled. Given a spectral probability of Z_S, a depth probability Z_D, and the weight of depth contribution *W*, then we defined a final probability Z.

$$Z = WZ_D + (1-W)Z_S \tag{2}$$

B. Simulation data

Three wounds of clinically relevant volumes, A, B and C were modelled and coated with coloured plasticine in order to represent realistic wound structure and physical tissue distribution, based on previous experimental experience and discussion with experienced podiatrists (Fig. 2). Red granulation tissue tends to occur in the central base of the wound and is surrounded by purple/violet epithelial growth, which is in turn surrounded by healthy skin tissue. Yellowish slough and black necrotic tissue, which consist of dead cells, may occur anywhere throughout the wound. Model 'A' represented a wound with simple structure and 4 tissues. In order to challenge the algorithms, 'B' contained 5 tissues, and 'C' was modelled to be a more irregular shape than A or B.

The three wound models were imaged and reconstructed using the Di3D system and its accompanying software (DiCapture v6.1). The software allowed the export of the data as OBJ files, constructed from multiple triangular faces in a 3D mesh. An OBJ file contains data representing a 3D surface with colour photographic texture. These data were used to generate a two-dimensional greyscale depth map for each model, along with a corresponding 2D colour map (Fig. 3). The result was two images for each model; one containing pixels of spectral red, green and blue information, and one containing the corresponding depth co-ordinate for each colour pixel.

In order to realistically simulate typical tissue colour distributions, 2D wound images were obtained from Belfast City Hospital (e.g. Fig. 1). These were sampled to calculate estimations of spectral mean and variance for each relevant tissue type. *Analyze* software (v10.0, Lenexa, Kansas. http://www.analyzedirect.com) was used to segment the plasticine colour representations and create a binary mask for each tissue. Pixels were randomly generated according to a normal probability distribution, using the estimated parameters, for each tissue mask. The resulting regions were combined to form simulated wound images for each model.

The tissue masks were also used to create a ground truth segmentation against which classification results were compared.

Testing with simulation data involves challenging the algorithms with difficult circumstances. To this end, multiple simulation images were created for each model wound. In real tissues, colour distributions always overlap to varying degrees, which makes it difficult for clinicians to consistently assess wound tissues. Therefore it was appropriate to model such overlap and vary it in a controlled manner. Greater overlap was achieved by moving the colour distribution of each tissue closer to the others. The spectral means of each class were shifted towards a *ground spectral mean vector* (GSMV) by increasing their *spectral mean proximity* (SMP), using (3), according to the approach by Al Momani [19]. This process is illustrated in Fig. 4. The GSMV is calculated as the mean vector for all tissue classes. If $\boldsymbol{\mu}_i$ is the sampled class mean, M is the SMP, and $\mathbf{S}$ is the GSMV then the updated class mean vector $\boldsymbol{\mu}_i'$ is given by the equation

$$\boldsymbol{\mu}_i' = \boldsymbol{\mu}_i + M(\mathbf{S} - \boldsymbol{\mu}_i)/100 \tag{3}$$

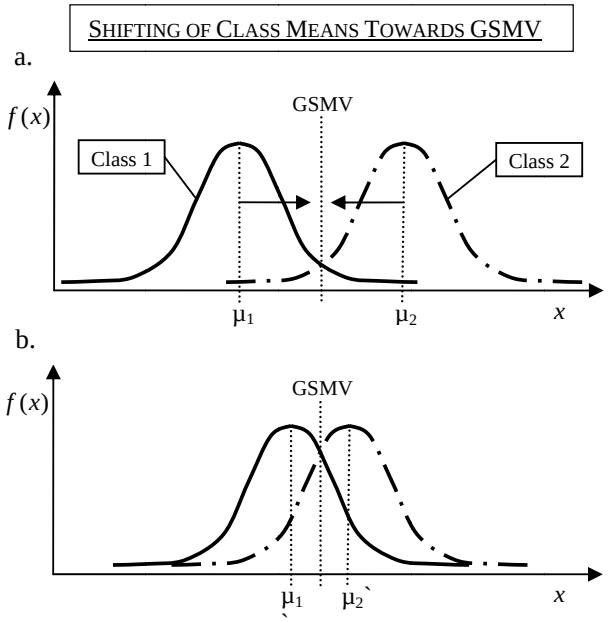

Fig. 4 The shifting of class means towards the *ground spectral mean vector* (GSMV) to create greater spectral overlap between classes

This process was used to generate images for classes with SMP of 0, 10%, 20% … 100% (Fig. 5). When at 100% the classes had identical mean vectors and were therefore distinguishable only by different variance values. As a result, 11 images were generated for each model, to a total of 33. Classification was performed by both the MLC and EM algorithms on each of the 33 simulated images, varying the contribution of depth information from 0 to 1, in increments of 0.1. The results of comparing classifications to the respective ground truths were presented in the form of confusion matrices. A confusion matrix is a tabular representation of agreement between two classifications, as in the example in Fig. 6. Results were evaluated from the confusion matrices to calculate percentage accuracy measurements for each classification.

IV. Results and Discussion

Improvements in classification accuracy were observed at lower weightings of depth information, while higher weightings led to a rapid deterioration. Results were initially represented by 3D plots to observe how algorithms performed over the full range of variables by plotting accuracy against both SMP and depth-weighting (Fig. 7). However, in all circumstances, weightings of 0.4 or higher did not show any improvements in classification accuracy. Therefore the lower weightings for each algorithm and models A, B and C were plotted two-dimensionally to show how accuracy was affected across the range of spectral mean proximities (Fig. 8). It is worth noting that there were no EM results for spectral mean proximities beyond 70%. This is because classification must be labelled by the user. At such high SMP, results became too noisy to be reliably labelled.

In the case of MLC results (Fig.8.1-8.3), weightings of 0.1 to 0.3 provided improvement in classification accuracy for all spectral mean proximities in models A and C, when compared to classification by colour alone (W=0). At low SMP values, the improvement is small but becomes much more significant as SMP increases. This is as might be expected, as a greater difficulty in distinguishing classes by colour would lead to a greater reliance on depth information. However, depth classification of model B provided very little improvement in any circumstance and started to actually harm the accuracy by weighting as low as W=0.3. This may have been due to the presence of a fifth tissue class of black necrotic tissue which,

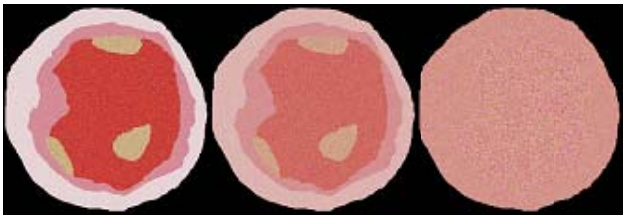

Fig. 5 Simulated colour for model C, showing Spectral Mean Proximities of 0, 50, 100%

		Ground Truth				Row total
		Gran	Epith.	Slough	Skin	
Algorithm classification	Gran	83289	43	158	0	83490
	Epith.	109	46228	1901	5	48243
	Slough	753	3563	11656	41	16013
	Skin	0	1	23	60545	60572
Column total		84151	49835	13741	60591	208318
Correct = 201718 / 208318 = 96.83 %						

Fig. 6 Confusion matrix for MLC performed on model A with a depth weighting of zero

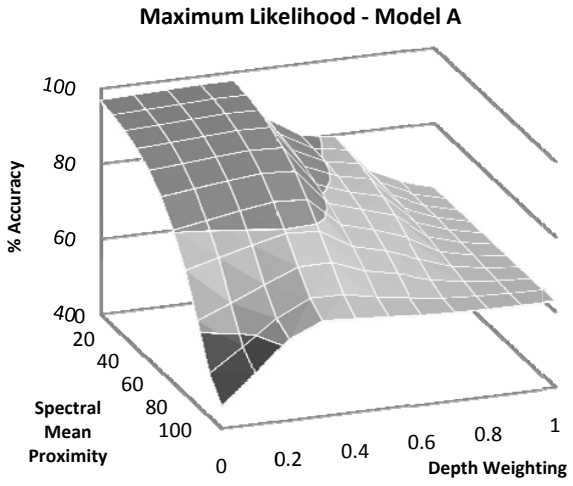

Fig. 7 3D plot of all data for MLC performed on model A

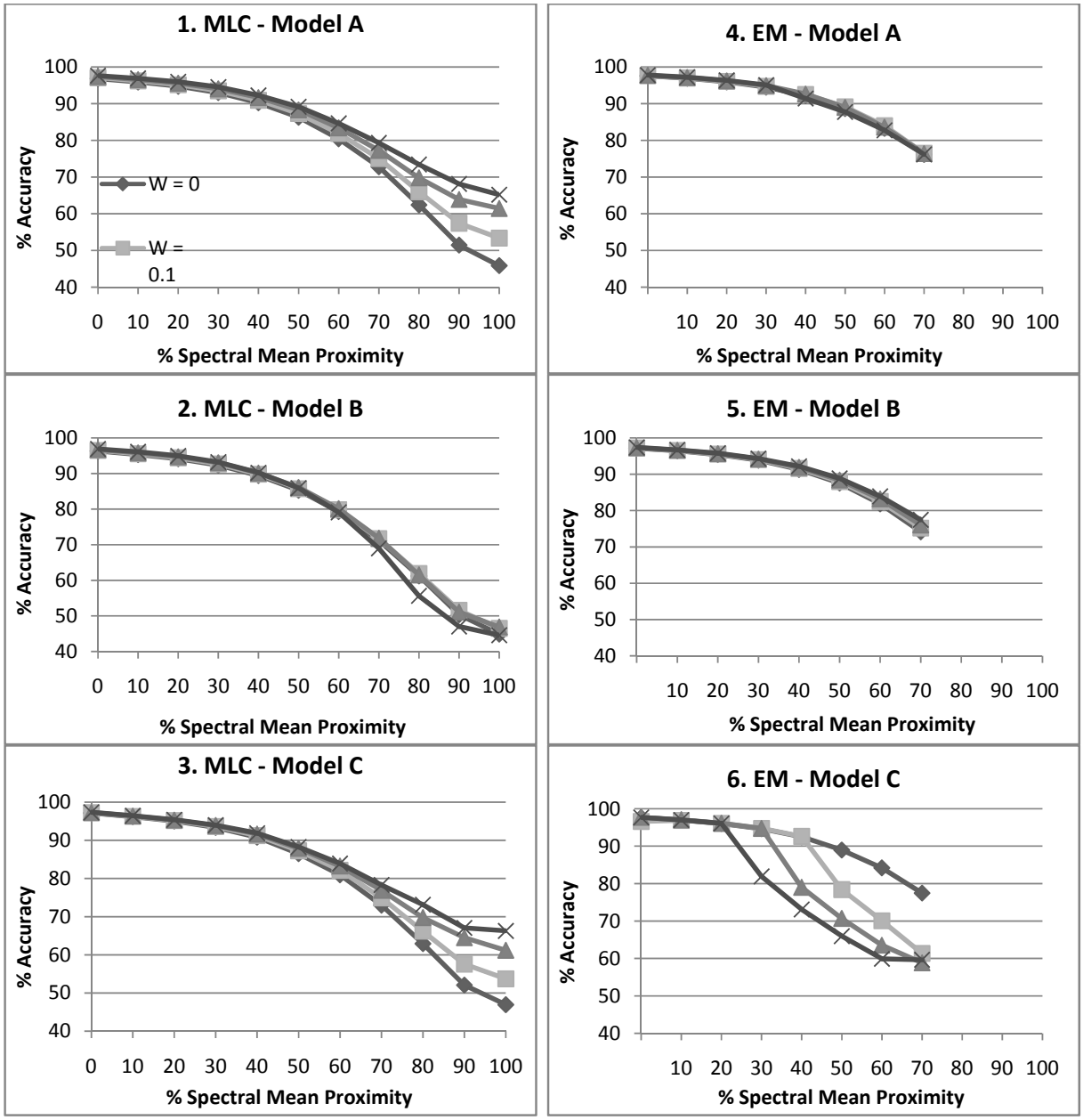

Fig. 8. Results for both MLC (1-3) and EM (4-6) on all models, for W = 0.0, 0.1, 0.2, 0.3.

like slough, has little dependence on depth.

The MMEM algorithm provided much more variable results (Fig. 8.4-8.6). In contrast to the MLC, model B showed the only substantial improvements for the lower weightings, again when SMP values were high. Model A was largely unaffected, although a weight of *W*=0.3 provided some small reduction in accuracy. Model C showed the only case where including depth information was actually detrimental to classification accuracy. Fig. 8 shows that at all weightings for this model, accuracy quickly drops off compared to colour classification alone.

For all six sets of results, the consistent pattern was a rapid drop in accuracy when weighting was increased beyond a certain point. In the MLC cases this occurred beyond *W*=0.3, indicating a consistency not present in the EM results, which as Fig. 8 shows, were variable and therefore unpredictable. This may have been due to the user-controlled, supervised nature of MLC. The choice of training samples for MLC made use of knowledge of the models which the user possessed, while EM trained itself without such knowledge.

V. Conclusion

The work presented in this paper shows that the incorporation of depth information into the colour classification of simulated wounds could be of significant benefit, albeit in limited circumstances. The inclusion of depth

was particularly effective in cases in which classes were of similar colour and therefore difficult to distinguish by spectral measurements alone. However, this study has also illustrated the importance of optimising the weight given to the depth probabilities. The choice of algorithm is also a major concern. While results were largely positive for the MLC algorithm, EM results were significantly poorer. It was noted however that for colour classification alone (W=0), the EM accuracy was actually higher, in the order of 1-5%. Therefore another choice of method for the depth classification may improve both the accuracy and reliability of results.

Further investigation could overcome many of the problems encountered in this study. Future work may involve the exclusion of some tissues from the depth classification process as it may be counter-productive to include tissues which have little or no dependency on depth. The use of other algorithms may improve accuracy, whether used alone or combined with those tested in this work. It is also possible that non-parametric models could assist classification, as the assumption of a normal distribution may be inappropriate. The results of this study do not show obvious overall benefit to the inclusion of a weighted depth probability for classification. However, important issues and areas for improvement have been revealed. Ethical approval has been granted for the collection of real wound data, which will allow for more thorough evaluation of algorithm performance on a wider range of more realistic data. The simulation results from this study will inform and influence study design when assessing real data.

REFERENCES

[1] S. Stremitzer, T. Wild and T. Hoelzenbein, "How precise is the evaluation of chronic wounds by health care professionals?" *Int Wound J*; vol 4(2): pp. 156-61, 2007.

[2] X. Liu, W. Kim, R. Schmidt, B. Drerup and J. Song, "Wound measurement by curvature maps: a feasibility study." *Physiol Meas*; vol 27(11): pp. 1107-23, 2006.

[3] M. Flanagan, "Wound measurement: Can it help us to monitor progression to healing?" *J Wound Care*; vol 12(5): pp. 189-94, 2003.

[4] P.C. Matthews, A.R. Berendt and B.A. Lipsky, "Clinical management of diabetic foot infection: diagnostics, therapeutics and the future." *Expert Rev Anti Infect Ther*; vol 5(1): pp. 117-27, 2007.

[5] D.G. Armstrong and B.A. Lipsky, "Diabetic foot infections: stepwise medical and surgical management." *Int Wound J;* vol 1(2): pp. 123-32, 2004.

[6] M.J. Levy and J. Valabhji, "The diabetic foot." *Surgery (Oxford)*; vol 22(12): pp. 338-41, 2004.

[7] E.S. Papazoglou, L. Zubkov, X Mao, M. Neidrauer, N. Rannou and M.S. Weingarten. "Image analysis of chronic wounds for determining the surface area." *Wound Repair Regen*; vol 18(4): pp. 349-58, 2010.

[8] J. Shaw, C.M. Hughes, K.M. Lagan, P.M. Bell and M.R. Stevenson, "An evaluation of three wound measurement techniques in diabetic foot wounds." *Diabetes Care*; vol 30(10): pp. 2641-2, 2007.

[9] R.J. Winder, T.A. Darvann, W. McKnight, J.D. Magee and P. Ramsay-Baggs, "Technical validation of the Di3D stereophotogrammetry surface imaging system." *Br J Oral Maxillofac Surg*; vol 46(1): pp. 33-7, 2008.

[10] B. Khambay, N. Nairn, A. Bell, J. Miller, A. Bowman and A.F. Ayoub, "Validation and reproducibility of a high-resolution three-dimensional facial imaging system." *Br J Oral Maxillofac Surg*; vol 46(1): pp. 27-32, 2008.

[11] D.P. Thompson, J.H Cundell, D.A. McDowell and R.J. Winder, "Simulated wound volume measurement using 3D stereophotogrammetry." *Proc. 14th annual Irish Machine Vision and Image Processing (IMVIP) Conference*, Limerick, 2010, pp. 30-43.

[12] D.L. Pham, C. Xu and J.L. Prince, "Current Methods in Medical Image Segmentation." *Annual Review of Biomedical Engineering*, vol 2(1), pp. 315-337, 2000.

[13] R.O. Duda, P.E. Hart and D.G. Stork. *Pattern Classification*, 2nd Edn, 2001. New York: John Wiley & Sons.

[14] H. Wannous, Y. Lucas, S. Treuillet and B. Albouy. "A complete 3D wound assessment tool for accurate tissue classification and measurement." *15th IEEE International Conference on Image Processing*, 2008; pp. 2928-31.

[15] H. Wannous, Y. Lucas and B. Albouy. "Enhanced assessment of the wound-healing process by accurate multiview tissue classification." *IEEE Transactions on Medical Imaging*; pp. 315-26, 2011.

[16] Y.J. Zhang, "A review of recent evaluation methods for image segmentation", *Signal Processing and its Applications, Sixth International, Symposium on;* pp. 148-51, 2001.

[17] X. Hong, S.I. McClean, B. Scotney and P.J. Morrow, "Model-based Segmentation of Multimodal Images", *Lecture Notes in Computer Science*, vol 4673, pp. 604-61, 2007.

[18] C. Fraley and A.E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis." *The Computer Journal*; vol 41(8): pp. 578-88, 1998.

[19] B. Al Momani, "Classification of Remotely Sensed Imagery Using a Knowledge-Based Approach." *PhD thesis, University of Ulster*, 2008.