FULL ARTICLE

Scar tissue classification using nonlinear optical microscopy and discriminant analysis

Timothy Andrew Kelf¹, Martin Gosnell¹,², Bjornar Sandnes¹,³, Anna E. Guller⁴, Anatoly B. Shekhter⁴, and Andrei V. Zvyagin*¹

¹ MQ BioFocus Research Centre, Macquarie University, Sydney, NSW 2109, Australia
² Quantitative Pty Ltd, Sydney, New South Wales, Australia
³ Department of Physics, Norwegian University of Science and Technology, Hoegskoleringen 5, N-7491 Trondheim, Norway
⁴ Laboratory of Experimental Pathomorphology, Science-Research Center, I.M. Sechenov First Moscow Medical University, Moscow, Russia

Received 21 August 2011, revised 11 October 2011, accepted 3 November 2011
Published online 22 November 2011

Key words: SHG, NLOM, image analysis, discriminant analysis, biomedical imaging

Supporting information for this article is available free of charge under http://dx.doi.org/10.1002/jbio.201100075

This paper addresses the scar tissue maturation process that occurs stepwise, and calls for reliable classification. The structure of collagen imaged by nonlinear optical microscopy (NLOM) in post-burn hypertrophic and mature scar, as well as in normal skin, appeared to distinguish these maturation steps. However, it was a discrimination analysis, demonstrated here, that automated and quantified the scar tissue maturation process. The achieved scar classification accuracy was as high as 96%. The combination of NLOM and discrimination analysis is believed to be instrumental in gaining insight into the scar formation, for express diagnosis of scar and surgery planning.

Discriminant analysis of three tissue types.

1. Introduction

Scarring is a result of thermal or traumatic injury or surgical excisions, where the skin fails to fully recover to its initial state [1, 2]. Under normal conditions the development of any scar passes through several successive stages: from early hypertrophic [3] through to the final stage of a flat, pale scar, usually referred to as a mature scar [4–6]. This process is often assumed to take at least 1 year [4, 7–9], but depends on a number of factors, including patient’s age. During the maturation process, a remodelling of the tissue collagen framework occurs, with the ratio of the collagen types changing [10–12]; the number of strong intermolecular cross-links increasing [12–17], and a rearrangement of bundles and fibre architecture. External manifestations of scar maturation are expressed through changes in prominence, colour, density, and in scar surface modifications [18, 19]. These processes can impede the diagnosis of scar types during the maturation process, and complicate the choice of treatment and timeliness of reconstructive surgery [18, 20, 21], which is recognised to be performed not earlier than one year after the...
injury [4, 22, 23] due to possible risks connected with hypertrophic scar regrowth after the excision [24].

To our knowledge, only scarce reports are available on the histological objective criteria of scar maturity [25–27], and thus a reliable, minimally invasive, means to identify and classify post-burn scars is in demand. The use of nonlinear optical microscopy aided by discriminant analysis can be of considerable value.

NLOM relies on the use of infrared, ultra-short (~100 femtosecond) laser pulses to excite nonlinear processes, such as two-photon excited fluorescence (TPEF) and second harmonic generation (SHG), in the focal volume. This results in considerable reduction of scattering, photobleaching and absorption by the endogenous chromophores relative to that in the visible spectral range [28]. SHG occurs in polarisable materials with non-centrosymmetric structures, such as helices. Collagen fibres have a helical structure and a large hyperpolarisability [29], hence are highly visible in SHG imaging [30, 31]. Elastin, melanin and other endogenous fluorophores produce detectable TPEF signals, emitting light in the visible spectral range, but little SHG signal due to their centrosymmetric molecular structures [32–35]. Due to the exquisite sensitivity of the SHG process to collagen, SHG imaging has been used to examine collagen networks of the scar tissues towards quantitative assessment and gaining insight into the scar formation process [36–38]. To this end, discriminant analysis represents a promising approach for extending the scar collagen network quantitative assessment to the classification of scars.

Discriminant analysis often furnishes quantitative methodologies, such as face recognition [39], and is proven to be an excellent classification approach in biomedical applications, where large sample variations are inherent [40–43]. The general classification methodology involves a pre-processing step, which produces feature vectors from raw data, such as images. A test step employs discriminant functions to project unclassified data into a subspace, where an applied decision boundary sorts each data point or observation into a predefined class [44–46]. A brute-force search of the optimal projections and best features is computationally prohibitive even with the modern computing power, requiring trillions of tests. Instead, we used the power of a genetic algorithm to hunt for an optimal or near-optimal solution, based on the classification performance ranking.

In this paper, we report on application of the discriminant analysis to classify the SHG images of scar and normal skin tissue. A computer algorithm was trained on pre-classified tissue samples of normal skin, mature and hypertrophic post-burn scars. The trained algorithm was tested to successfully discriminate between these tissue types, and for the first time, to the best of our knowledge, demonstrated the heterogeneous composite scar tissue organisation by means of the unbiased computational methods.

2. Materials and methods

Excisional biopsies of 4 hypertrophic and 3 mature scars (from 7 patients) were performed within the boundaries of the lesions of children who were recovering from deep thermal injuries and undertaking plastic and reconstructive surgery. Surgical procedures were approved by the local Ethics Committee and informed consent was obtained from the legal representatives of all patients. The scars were evaluated by the plastic surgeon following the guideline document [4] for wide post-burn hypertrophic scars and non-hypertrophic mature scars, information listed in Table 1. Full thickness healthy skin samples were obtained from fragments of autografts from three age-matched patients undergoing post-burn plastic surgery. Biopsies were fixed in 10% neutral buffered formalin, dehydrated in a graded series of alcohol, embedded in paraffin wax and cut into

Table 1 Scar tissue origins and clinical appearance.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Etiology of burns</th>
<th>Duration after injury (years)</th>
<th>Site</th>
<th>Clinical scar appearance</th>
<th>Morphological examination results</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>7</td>
<td>Scald</td>
<td>3</td>
<td>Neck</td>
<td>Wide Hypertrophic, whitish</td>
<td>Mainly hypertrophic scar tissue</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>Scald</td>
<td>3</td>
<td>Femoral</td>
<td>Wide Hypertrophic, whitish</td>
<td>Mainly hypertrophic scar tissue</td>
</tr>
<tr>
<td>M</td>
<td>13</td>
<td>Scald</td>
<td>2</td>
<td>Elbow</td>
<td>Wide Hypertrophic, reddish</td>
<td>Mainly hypertrophic scar tissue</td>
</tr>
<tr>
<td>F</td>
<td>7</td>
<td>Flame</td>
<td>5</td>
<td>Back</td>
<td>Wide Hypertrophic, brownish, mature (wide, flat, non-pigmented)</td>
<td>Hypertrophic scar tissue</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>Scald</td>
<td>3</td>
<td>Flank</td>
<td>Mature (wide, flat, pale)</td>
<td>Mainly mature scar tissue</td>
</tr>
<tr>
<td>F</td>
<td>2.5</td>
<td>Scald</td>
<td>1</td>
<td>Thorax</td>
<td>Mature (wide, flat, pale)</td>
<td>Mature scar tissue</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>Scald</td>
<td>3</td>
<td>Thorax</td>
<td>Mature (wide, flat, pale)</td>
<td>Mature scar tissue</td>
</tr>
</tbody>
</table>
5 and 23 µm matched serial sections for conventional light microscopy histological examination and NLOM respectively. The 5 µm paraffin slices were stained with haematoxylin and eosin (H&E), van Gieson’s picro-fuchsin, toluidine blue and Unna’s fuchsin following the conventional protocols.

The stained histological preparations were examined using an upright light microscope Olympus BX51 (Olympus Optical, Tokyo, Japan) and analysed without knowledge of the clinical data and then cross-checked.

The preparations destined for NLOM were left unstained. All samples were embedded in the mounting media and covered by cover slips. A Tsunami laser (Spectra Physics, USA) was used for the NLOM imaging, wavelength 800 nm, repetition rate 80 MHz, pulse duration 150 fs, 250 nJ/pulse at the sample, in conjunction with a Leica TCM SP2 scanning confocal microscope. Light was focused using a 40× oil-immersion objective (N.A. 1.25) and images recorded the resultant 400 nm wavelength light in transmission mode, image sizes are 375 × 375 µm. These images were split to 9, 125 × 125 µm images to increase the amount of available data; this was found to be the smallest size at which structural information was not degraded. For all tissue slices the microscope parameters where kept constant except the detection volume, which was adjusted so that each image was correctly exposed.

Image analysis was performed using Igor Pro (Wavemetrics, USA) and discriminant analysis used Matlab (Mathworks, USA).

3. Results and discussion

Representative SHG images of hypertrophic (immature) and mature scar tissues, as well as normal skin are shown in Figure 1, top row. These images were formed by collating several images together. While the SHG process is polarisation-sensitive, in turbid biological tissue and in presence of the intense off-focus SHG signals, this effect is diminished [47]. However, to ensure immunity to polarisation effects, only relative image intensities were considered.

Figure 1 insets exemplify differences between the tissue types under study. The collagen structure of hypertrophic scar tissue appeared much denser and well aligned to the skin surface than that of the normal skin tissue, which comprised more loosely packed thick bundles of collagen. In contrast, the mature scar tissue comprised relatively thin, randomly orientated, collagen fibres and bundles and exhibited a greater density than that of the normal skin tissue. Despite the distinct visual appearance of these tissues, from fragment to fragment and from patient to patient, each scar type showed a wide range of appearances, making visual classification cumbersome, and calling for an objective, machine-based classification procedure. In order to realise this procedure, multiple parameters were extracted from each image, and many images were used to ensure statistically meaningful results. Specifically, 364, 313 and 456 images sized 125 × 125 µm (totalling 1133) of the mature scar, hypertrophic scar and normal skin tissues, respectively, were used.

A professional histologist verified all images to assign their classification type on the basis of cellular and matrix features. Composition of the fibroblast population was imaged (H&E staining, representative images shown in Figure 1) and analysed in relation to the ultrastructure of the cells by means of transmission electron microscopy (data not shown). The tinctorial features of the extracellular matrix (van Gieson’s, Weigert’s Resorcin Fuchsin and toluidine blue staining) were used to detect the presence, condition and distribution of collagen fibres and bundles, elastic fibres and acid glycosaminoglycans. Also, the appearance and number of the blood vessels and inflammatory cells were noted. The above information was matched with the clinical observations to estimate the agreement between the tissue structure and the external appearance of scars, see Table 1 in supplementary information. While accurate, this procedure required much information and several tissue preparation steps, emphasising the fact that a robust algorithm using images from unstained tissue would be of considerable use.

Three image processing algorithms were employed to extract a total of 60 features, as shown in Figure 2 and tabulated in Table 2 in Supplementary Materials. Comparison of the H&E and NLOM...
The first algorithm performed a two-dimensional (2D) Fourier transform of the raw images, followed by radial averaging and fitting to a sinusoidal function to determine the direction of the maximum order relative to the tissue surfaces. The perpendicular direction was defined as the direction of minimum order. The transformed data was subsequently averaged over an angle of 10° in both directions of the maximum and minimum order; their respective ratio was calculated providing a measure of tissue order. The resultant data was averaged over the structures sized greater and smaller than a threshold value of 6 μm to yield two scalar classifiers. This threshold size was determined empirically to provide the best classification. However, this value seemed relevant to the tissue morphology delineating two collagenous structure size ranges: <6 μm pertinent to the collagen fibres, >6 μm pertinent to the fibre bundles and their separation.

In the second algorithm, a single-level discrete 2D multi-wavelet decomposition using four fitting parameters was applied to the image, yielding four output matrices that were analysed to find their mean, standard deviation, skew and entropy values. These parameters provided additional information about the overall image, and the ordering level of at several length scales.

In the third algorithm, images were sub-divided into two subsets forming a collagen-containing image, and its inverse comprising the tissue “voids”. These two image subsets were binarised and run through an image erosion procedure to discriminate the collagen from the background, and identify the “void” boundaries. A particle analysis procedure was subsequently run to analyse distinct pieces of tissue/“voids” and find their centroids and boundaries. This data was further processed to determine the average, maximum and skewness of the analysed area distributions, circularities, perimeter lengths and distribution of their axes with respect to the skin surface, as well as the total number of particles and nearest neighbour distances.

As a result of implementing these algorithms, 60 features were generated in total. From these, 40 features were selected that exhibited non-negligible variation between one or more tissue types, based on statistical distance and tests. A feature selection algorithm was used to find a minimum size set of the features sufficed for accurate classification. A block diagram of the selection system is shown in Figure 3. The developed selection algorithm chose features based on the classifier performance, employing a data partition cross-validation method aiming to select a final set of the features, which had a high likelihood to classify test scar/skin tissue data. Multiple discriminant analysis runs of the test data produced a set of discriminant functions by finding eigenvectors of the product of matrices describing the inter-class variance, and inverse of the intra-class variance.

The data was standardised by subtracting the mean and dividing by the standard deviation. Data was initially randomly partitioned into a feature se-
lection set, training set and testing set, a validation set was not required, since we were using a deterministic linear classifier, where control over the extent of training was not a concern. The initial feature selection portion of data was used in the feature selection algorithm, where it is further randomly partitioned into training and test sets multiple times. For each of these times, the entire population of feature sets is tested, by training the classifier on the training data and testing on the test portion. The fitness of each feature set is based on the mean of these classification results.

Once a feature set was chosen based on mean classification performance using the remaining unused partition of data. For consistency, we ensured the classifier performance was a result of the mean performance for multiple training and tests of random data partitions. Then the entire process was repeated 10 times producing the final result. For each of these final results the best performance sets of feature showed only some slight variation. It was of interest to note that selecting from only the FFT and wavelets features consistently produced a drop in performance of typically 3%.

These discriminant functions were used to project the data into a subspace, where each test data observation was classified based on the mahalanobis distance, a parameter based on the standard deviation of the total distribution. Overall classification accuracy across all categories was used as a main performance fitness criterion. This test was performed for a population, or set, of possible randomly chosen feature combinations, and each solution was fitness-ranked by the algorithm. The processes of iteratively mating and selecting from this population of solutions allowed members of the population to develop into better solutions capable of the tissue type discrimination. Random mutation of a small percentage of the population avoided the algorithm trapping in a local minimum.

Once a feature set was chosen, the overall performance was visualised by projecting the data into a new subspace by plotting each point of the new canonical variables into a two dimensional space (Figure 4a). As can be seen, the algorithm discriminated the tissue types into distinctly separate data clusters. After about 100 iterations, the algorithm performance asymptotically approached optimal performance showing no further improvements per additional iterations. Figure 4c shows the best solution performance for a range of selection tests, where limits were placed on the number of features selectable with this algorithm. The onset of the performance deterioration occurred at <15 features, while only 23 features, instead of 40 features available, were found sufficient for finding the best solution. Several observations of the final classification algorithm can be made. Of the 40 features extracted from the tissue images, those related to the overall image features obtained via Fourier and wavelet transformations were found to be most important for the classification procedure.

It is important to note that the isolated collagenous tissue and “void” image subsets (algorithm 3) exhibited little correlation. (Note that binarisation and erosion produced images that are not simply inverse). The biological rationale for separating the images into these subsets was that they contained information relating generally, to two distinct architectures in the scar tissue, i.e. the collagenous structures and the vascular networks associated with the voids.

Figure 4b shows the fitness level of the best solution for successive generations of the genetic algorithm, and shows the procedure to be effective at finding the optimal solutions. Initial features were selected as those containing the most discriminatory information quickly ramping up the algorithm efficiency.

The features related to the tissue fragments/“voids” orientation to the tissue surface (the average angle and distribution skewness) appeared to be of the greatest significance. Also, the nearest neighbour distances of both tissue and “voids” proved to be of importance. Such features as the total number of tissue fragments, their average circularity, the maximum length of “voids”, and the length distribution
skewness were worthwhile contributors to the final discrimination. The degree of the tissue alignment was an important factor in discrimination between different scar types, in agreement with the recent reports [36, 37].

Using the optimised set of features, the discrimination procedure was tested using 10 full cross-validation runs, with the result shown in Figure 4d. This procedure performed a validation comparison using much of the available data comparing to *priori* histology-based classification results. It was found that the classifier was accurate to better than 96%, verifying the algorithms effectiveness. Of the 4% of the algorithms inaccuracies, most were found to be skin and mature scar tissue samples wrongly labelled as hypertrophic scars (as histologically classified). This is important if the algorithm is to be of clinical use as the classification errors were skewed to the side of caution.

The initial univariate feature selection was performed, as follows. The first stage of feature selection disregarded those features that performed poorly using statistical tests chosen to help making an informed decision regarding the class discriminating power of each feature on its own merit. These tests were conducted on a pair-wise basis to cover the three classes of data, the minimum p-value for each of the three tests was used, and only values below 0.05 were accepted, indicating that this feature might be useful for separating, at least, two of the classes.

After optimisation the classified tissue clusters exhibited a measurable distribution, as it can be seen in Figure 4a. At the macroscopic scale, scars are classified clinically by their visual appearance and clinical history, while at the microscopic scale, the cellular population and the extracellular matrix features provided an approach to distinguish different scar tissues. However, classification of the scar tissue architecture at the sub-millimetre level, as inferred from our results, may be a valuable source of diagnostic information. The reported procedure enables not only accurate scar tissue classification, but also quantification of its key parameters. For example, it is possible to assign a classification value to each tissue image fragment (termed “elementary classification cell”, ECC) that is large enough to contain the tissue image fragment (termed “elementary classification cell”), ECC, which is now in order. The original images were sub-saturated process of scar tissue in humans.

Discussion of the potential sources of inaccuracies is now in order. The original images were sub-

![Figure 5](online color at: www.biophotonics-journal.org) (a) Mature scar tissue sample. (b) False-colour image based on the discrimination procedure. (c) Discrimination analysis results for the different regions superimposed on a colour palette.
partitioned into a set of smaller images and these sub-images might have been correlated. Hence, it may be reasonable to conclude that partitioning the data such that some of these sub-images were included in the training and test sets might have been giving the classifier an undesirable advantage.

In order to test this possible source of inaccuracy, we modified the data partitioning algorithm to accept a sub-grouping variable, which provided a constraint on the partitioning disallowing sub-images spreading across the feature selection, training and testing partitions. The entire selection and testing system was re-run to gauge the impact of this. The resultant accuracy was slightly decreased, estimated as 5–6%, which can be explained using the following reasoning. The employed image selection constraint provided effectively only ~100 data points, which were further broken into three sets yielding the smallest possible number of observations needed for a full cross-validated test. It was, therefore, expected that the performance of a generalising test suffered by virtue of the minimal size of the data set alone.

The presented methodology of the collagenous tissue classification based on the discrimination analysis is believed to hold promise for medical applications. Firstly, accurate diagnosis of scars is deemed useful, especially in pediatric practices, where scar assessment over extended periods of time may not be an option due to possible face/body disfigurement consequences. As we found, a tissue fragment (ECC) as small as 125 × 125 μm sufficed for the reliable scar tissue classification, which opens an attractive opportunity to perform minimally invasive needle biopsy.

We also envisage broader-scope applications. Among these, analysis of the skin cancer collagen network is believed to provide a strong base for diagnosis and surgery planning. Besides, we observed profound collagen modification in chondrosarcoma biopsies (unpublished), which should enable useful diagnostic opportunity.

In conclusion, we demonstrated that the application of nonlinear optical microscopy in combination with discriminant analysis has potential for delivering fully-automated classification of scar tissues, with the validated accuracies as high as 96%. The tissue gross morphology and the collagen structure angular alignment with respect to the tissue surface were found to be the main discrimination parameters. Using second harmonic imaging and discriminant analysis we demonstrated the scar spatial heterogeneity at the sub-millimetre level, in accordance with the algorithm quantitative classifier. The reported methodology is adaptable to minimally invasive biopsies, from that stage it will be of considerable value for express diagnosis of scars in humans.

**Timothy A. Kelf** received his M.Sc. in Physics from The University of Southampton in 2002. He completed his Ph.D., studying Plasmonics and Surface Enhanced Raman, in 2006, also at The University of Southampton, UK. From 2006 to 2008 he worked as a post doctoral researcher at Hokkaido University, Japan, studying the ultra-fast vibrational properties of nano-structures. From 2008 he has worked at a research associate at Macquarie University, Australia, studying the interactions of nano-particles with cells and tissues, along with developing novel optical detection and characterisation systems.

**Martin E. Gosnell** received his Bachelor of Engineering in Electrical engineering with first class honours from the University of Technology at Sydney in 2001. He has worked as an engineer and scientist for CSIRO and by contract to the university of Queensland and Newcastle, and as a design engineer and research scientist for consumer electronics manufacturing companies in the areas of computer brain interface, intelligent recognition systems, laser applications and audio. His interests include applied mathematics, control, information theory, hyperspectral and image analysis, algorithm, and classification system design.

**Bjornar Sandnes** received his Ph.D. from the Department of Physics, University of Oslo in 2004. A postdoc period followed in the Advanced Materials and Complex Systems Group, 2004–2007. In 2008 Sandnes was appointed Research Fellow at the Optical Biomedical Imaging and Sensing Group at Macquarie University, Australia, and later in 2008 he was awarded a fellowship by the Norwegian Research Council, working jointly at the University of Sydney and Macquarie University. From 2011 he is Associate Professor at the Norwegian University of Science and Technology, NTNU. Sandnes’ research background includes soft and complex materials, photonics and biomedical imaging.
Anna E. Guller received her degree in Biology at the Department of Human and Animal Physiology and Biophysics of the Taurida National V.I. Vernadsky University, Ukraine, 1995. In 1995–1999 she has been working as a Junior Researcher in the Russian Disaster Medicine Center in Moscow. From 1999 to 2003 she was a Ph.D. student in the Experimental Pathomorphology Department of Sechenov First Moscow State Medical University (Russia), focusing on functional morphology and development of fibrous connective tissue. Currently Anna is working at the same Department as a principal researcher. She is interested in pathomorphology of wound healing, scarring and regeneration, tissue engineering of bone, cartilage and skin, conventional and non-linear optical microscopy, numerical methods in pathology and non-invasive morphological diagnostics.

Anatoly B. Shekhter received his M.D., Ph.D. and D.Sc. degrees in pathological anatomy from Sechenov First Moscow State Medical University, Russia, in 1957, 1962 and 1983, respectively. From 1962 Dr. Shekhter has been working as a senior scientist in the Central Science-Research Laboratory of the same University, and in 1969 he became the Head of the Experimental Pathomorphology Department of the Sechenov University. In 1988 Dr. Shekhter has been awarded an academic title of Professor. His current research interests include wound healing and regeneration in skin, bone and cartilage, guided transformations of pathologically changed connective tissue under external physical factors, implantation and biocompatibility of materials.

Andrei V. Zvyagin received Diploma from the Moscow Engineering Physics Institute, Russia, and the Ph.D. degree in Engineering from the Tokyo Institute of Technology in 1997. He spanned a range of positions from an Engineer to Senior Research Scientist in the Institute of Metrology for Time and Space, Moscow, Russia. From 1998 to 2004, he was an Australian Postdoctoral Fellow in the Optical & Biomedical Engineering Laboratory, The University of Western Australia. From 2004–2007, he joined The University of Queensland as a Lecturer/Senior Lecturer. Since 2007, Dr. Zvyagin is Associate Professor at the Macquarie University, Sydney, Australia. His current research interests include various areas of Biophotonics and Nanotechnology: nonlinear optical microscopy, luminescent nanomaterials, as molecular reporters for diagnosis and targeted delivery.

References


