VOXEL CLASSIFICATION OF PERIPROSTHETIC TISSUES IN CLINICAL COMPUTED TOMOGRAPHY OF LOOSENED HIP PROSTHESES

D.F. Malan[1,2], *C.P. Botha*[2,3], *R.G.H.H. Nelissen*[1], *E.R. Valstar*[1,4]

Leiden University Medical Centre
Leiden, The Netherlands

ABSTRACT

We present an automated algorithm which classifies periprosthetic tissues in CT scans of patients with loosened hip prostheses. To our knowledge this is the first application of CT voxel classification to periprosthetic tissues of the hip. We use several image features including multi-scale image intensity, multi-scale image gradient and distance metrics. Seven classifier types were trained using five manually segmented clinical CT datasets, and their classification performance compared to manual segmentations using a leave-one-out scheme. Using this technique we are able to correctly segment the majority of each of the six tissue categories, in spite of low bone densities, metal-induced CT imaging artefacts and inter-patient and inter-scan variation. Our automated classifier forms a pragmatic first step towards eventual automatic tissue segmentation.

Index Terms— Automatic, classification, segmentation, computed tomography, periprosthetic, osteolysis.

1. INTRODUCTION

The most significant complication that threatens the long-term survival of a total hip arthroplasty (THA) is periprosthetic osteolysis [1, 2] which involves resorption of bone and replacement by soft fibrotic tissue. Once osteolysis develops it usually progresses, eventually leading to mechanical instability and prosthesis loosening.

Minimally invasive refixation of loosened prostheses is possible [3] but requires the location and extent of fibrotic lesions to be known pre-operatively. Recent studies have shown that CT is more sensitive and accurate than traditional radiographs in detecting and measuring such lesions [2, 4]. However, the steady increase in resolution offered by modern CT scanners make traditional manual segmentation extremely time-consuming, thereby limiting users' utilization of the available data.

1 Department of Orthopaedics
2 Medical Visualization Group, Department of Mediamatics, Delft University of Technology, The Netherlands
3 Department of Radiology
4 Department of Biomechanical Engineering, Faculty 3ME, Delft University of Technology, The Netherlands

CT of suffers from metal-induced artefacts [5] which drastically complicate automatic segmentation near prostheses. To make things worse patients suffering from prosthetic loosening often have very poor bone quality yielding low CT image contrast with intensity values overlapping those of other tissues. Statistical Shape Models are useful for segmenting objects from low quality image data, but fare badly when modelling pathological tissues (such as fibrotic lesions) with no generalizable geometry [6] and/or consisting of several small isolated regions.

Several papers have been published describing automatic statistical pixel- or voxel segmentation of clinical data. By combining several complementary image features, voxel classifiers deliver reasonable classification performance in spite of metal-induced CT imaging artefacts, and without resorting to explicit geometrical modelling or human intervention. Radiographs [7], MRI [8] and CT [9] have been subjected to pixel/voxel classification. Standard approaches generally make use of multi-scale image intensity as well as higher order spatial derivatives to describe local image variations and "texture". Image intensity variation between scans can complicate X-ray and MRI feature selection, but CT scanners are largely immune to this due to their well defined and calibrated output measured in Hounsfield Units (HU). The geometric position of the image pixels or voxels can be omitted [7] or incorporated [9] into the classifier's feature space, although care must be taken so that the chosen features remain invariant to inter-scan orientation and scaling offsets.

The aim of this study was to develop an automated voxel classifier that can serve as the first step in a segmentation pipeline, eventually leading to patient-specific mechanical modelling. We are interested in the 3D distribution of bone, cement and fibrotic tissue around the prosthesis, which defines the hip's mechanical stability. In this paper we present statistical voxel classifiers that classify periprosthetic tissues into six possible tissue categories, namely *cement*, *fibrotic lesion*, *trabecular bone*, *cortical bone*, *intramedullary canal* and *exterior*. To our knowledge this is the first time that such a 3D statistical voxel classifier has been applied to periprosthetic CT image data. The classifiers are trained on manually segmented CT scans of five patients with clinically loose prostheses, and evaluated in a (per patient) leave-one-out

scheme. Image features are chosen so that they can be computed fully automatically. Once trained, tissue classification can be performed automatically, delivering an approximate tissue distribution as output. This initial classification forms a good foundation for further post-processing and eventual automatic segmentation.

2. METHOD

2.1. Image parameters

We obtained data from five different patients diagnosed with loose femoral prostheses causing pain and immobility. Each patient was scanned in a Toshiba Aquilion CT scanner using its FC30 "bone kernel", yielding the highest possible resolution, at the cost of increased noise. All scans were performed with a peak tube voltage of 135kV. Since we obtained the clinical data retrospectively there was some inter-scan variation, most notably tube current (150mA to 400mA) and in-slice voxel spacing (0.44mm to 0.59mm). All scans had a slice thickness of 1mm. The scanner was set to include a single hip in its reconstruction field of view.

In addition to normal between-patient anatomical differences we also note that different prosthesis designs and sizes were used. All prostheses were of cobalt-chrome, thereby presenting a worst case scenario, since titanium implants yield fewer artefacts.

2.2. Choosing image features

As pre-processing step we selected the upper part of each femur as a region of interest (ROI). Thanks to the artificial joint we have good separation between bony structures of the femur and pelvis. ROI extraction can therefore be performed automatically, although this falls outside the scope of this article.

We decided on using eleven image features at every voxel location. Following an approach similar to that in [7, 8, 9] we used CT grayscale values at multiple scales as our first four features. These features describe the native in-slice voxel resolution (0.5mm x 0.5mm x 1mm) along with Gaussian-smoothed versions having spherical standard deviations of 1mm, 2mm and 5mm. Features five to eight consist of the image gradient magnitude computed from at the same scales as the grayscale features.

The rationale behind using a Gaussian multi-scale approach is twofold. Firstly, by combining neighbouring pixel values, we tend to average out individual voxel noise (at the cost of resolution). Secondly, by adding information of neighbouring voxels we include neighbourhood information to every voxel (for example we can discern between an isolated bright voxel and a bright voxel in a bright neighbourhood, without doing explicit neighbourhood searches). Similarly to [10] and in constrast to [8] we decided against using second-order and higher derivatives as feature descriptors,

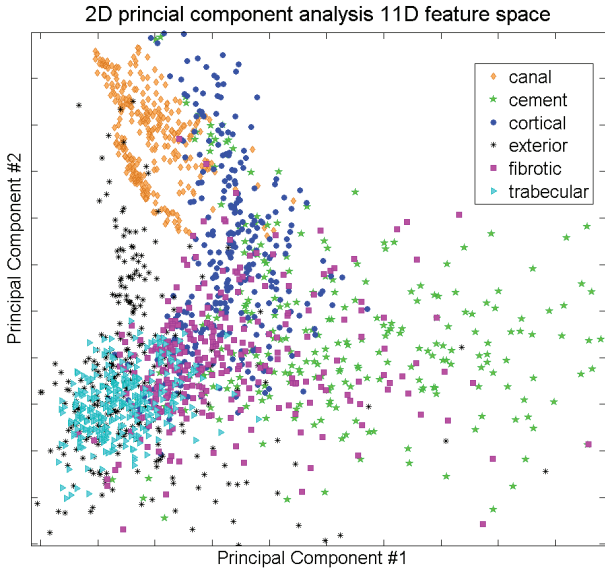

Fig. 1. Projection of the six tissue categories represented in feature space

arguing that these are excessively sensitive to image noise and contribute little additional information to our model.

The last three features (numbers nine to eleven) are distance metrics, chosen to be automatically computable and insensitive to rotational and translational offsets. In each of these three cases we compute the signed distance in millimeters, so as to be independent of the scan resolution and anisotropies.

Firstly, we compute the distance from the metal prosthesis. Metal has such a high contrast in CT that it can easily be found by performing a simple threshold at e.g. 5000 HU. This feature gives us useful information as to the "centredness" of any given voxel, which is useful since periprosthetic tissues are approximately radially distributed.

The second distance metric is computed along the scan (Z) axis from the centre of the prosthesis head. This can automatically be computed by the mean voxel location of the prosthesis head, which is easily recognizable from the previously mentioned threshold due to its increased diameter at one extremity. Due to the geometrical constraint of a patient lying on the CT gantry the prosthesis's long axis is always aligned with the scan direction, giving this distance a consistent interpretation.

Lastly we compute the the signed distance from the convex hull formed by the femur. We note that this adds information because femoral tissues have said radial distribution, with cortical bone being closest to the outer hull of the femur. We compute this feature by first thresholding the volume of interest (at 600 HU). The threshold of 600 HU was chosen so that only cortical bone, cement and the metal prosthesis fall above the threshold. We then subtract a per slice

7x7x1 voxel dilated mask of the metal prosthesis, thereby retaining only cortical bone and cement, along with possibly isolated metal-induced imaging artefacts. Next, we perform a cascaded 3x1x1 + 1x3x1 morphological opening to remove remaining metal-induced noise voxels. The result is a thresholded mask containing many islands and holes, caused by the CT shadow of the prosthesis, trabecular bone, fibrous tissue, intramedullary canal and zones of low bone density. Computing the slice-by-slice convex hull of this mask gives us a reasonable approximation to the convex hull of the femur's cortical shell.

2.3. Training the voxel classifiers

To provide training and validation labels to the image voxels, an experienced user manually segmented each femur using the interactive MITK software tool [11]. The segmented masks were then used to select the relevant voxels for the eleven image feaures. The statistical classifiers were constructed and trained on these features using PRTools [12], a pattern recognition toolbox for MATLAB.

Our classification task involves separating six distinct tissue types as collections of points in an eleven-dimensional space. Figure 1 shows a 2D projection along the computed axes of greatest separation of the 11 dimensional training voxel feature space. We see poor separation between classes indicating a very challenging classification problem. Selecting an appropriate classifier for this task is not a straightforward choice. Authors of recently published medical voxel classifiers have opted for a colourful mix of k-nearest-neighbour (kNN) [10, 8], linear- and quadratic discriminant (LDC & QDC) [9], decision trees [13] and neural networks, to name a few. In this paper we chose to compare several available classifiers namely LDC, QDC, Parzen, kNN, back-propagation neural network as well as a combined "voting" classifier composed of simple LDC and kNN classifiers.

The five different patient CTs were use to test the classifiers in a rotating leave-one-out scheme, where each time the classifiers were trained on four of the CT datasets and tested on the remaining set. At each step in the leave-one-out scheme we transform all features such that the training features have a zero mean and unit variance. Depending on the classifier we use a suitable sized random subset of training voxels. We use a subset of all available training voxels to keep the training time in check – a kNN classifier, for example, needs to store all training data internally. For each classifier we use an equal number of training samples per tissue class, along with equal priors.

3. RESULTS & DISCUSSION

The different classifiers' performance is shown in fig. 2. We see that the very fast LDC and QDC classifiers are generally less capable than the more complex alternatives. We were surprised at the relatively poor results obtained with the 3-layer back propagation neural network. The combined classifier delivered good classification results, although we found the Parzen classifier to have the most stable response across all tissue classes and test cases.

Our small dataset of five patients is a limitation in assessing the true potential of these methods. The limitation lies not in the number of data points available during training, but rather in their ability to represent the variation in human femora and scan parameters. However, it can be expected that segmentation performance will *increase* as larger and therefore more general training sets become available.

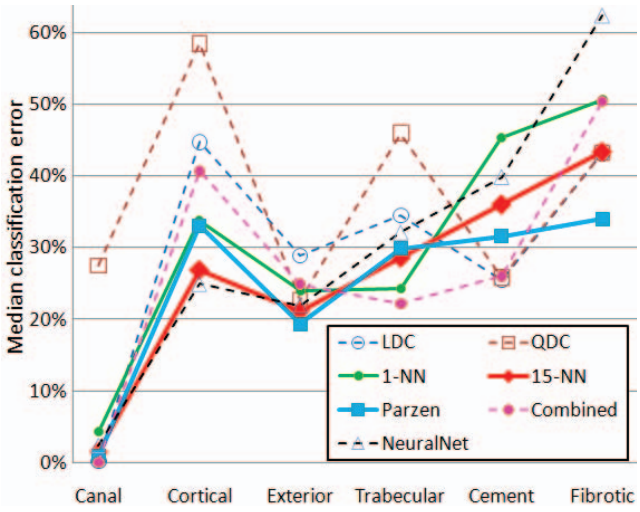

Fig. 2. Median classification error for different tissues and classifiers.

Figure 3 illustrates that automatic voxel segmentation correctly identifies the general distribution of the separate tissues, even before any additional post-processing. Looking at table 1, we see that the most problems occur when classifying cement and fibrotic tissue. Both of these exist close to the metal prosthesis where they are strongly affected by metal artefacts. Voting between each voxel and its neighbours' soft (continuous) classification can improve filtering of misclassifications by incorporating more geometrical coherence.

Manual seg.	Automatic classification					
	Canal	Cem	Cort	Ext	Fibr	Trab
Canal	**98.5**	0.4	0.3	0.0	0.8	0.0
Cem	1.8	**67.3**	6.3	0.4	21.5	2.7
Cort	1.5	6.0	**78.2**	1.1	6.1	7.0
Ext	0.2	2.0	1.7	**78.4**	4.9	12.8
Fibr	1.6	15.9	6.4	2.0	**61.4**	12.7
Trab	0.0	1.6	3.8	12.8	9.7	**72.0**

Table 1. Confusion matrix for the Parzen classifier over all test femora. (see also fig. 2).

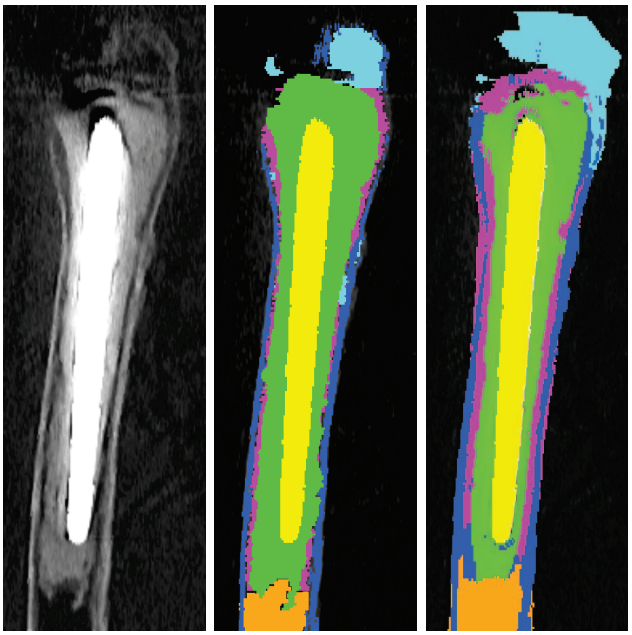

Fig. 3. a) Sagittal CT slice and tissue classification using b) manual segmentation and c) our automatic Parzen classifier

4. CONCLUSION & FUTURE WORK

The voxel classifier presented in this paper offers an automatic tool for performing an initial segmentation of 3D CT scans of loosened hip prostheses. We achieve a correct classification rate ranging between 66% and 70% for fibrotic lesions, bone and cement, the tissues we are most interested in. The result obtained represents a useful first step towards automated segmentation, and a significant improvement above simple threshold-based segmentation. Future work will include post-processing the initial classification result by incorporating neighbourhood voting and the the classification certainty associated with each voxel. We see this solution as the first step in a fully automatic tissue segmentation pipeline.

5. REFERENCES

[1] Thomas W. Bauer and Jean Schils, "The pathology of total joint arthroplasty - II. Mechanisms of implant failure," *Skeletal Radiol*, vol. 28, pp. 483–497, 1999.

[2] Tim A Walde et al., "Comparison of CT, MRI, and radiographs in assessing pelvic osteolysis: a cadaveric study.," *Clin Orthop Relat Res*, vol. 437, no. 437, pp. 138–144, Aug 2005.

[3] Jolanda J de Poorter, Rob C Hoeben, Simone Hogendoorn, Vivien Mautner, John Ellis, Wim R Obermann, Tom W J Huizinga, and Rob G H H Nelissen, "Gene therapy and cement injection for restabilization of loosened hip prostheses.," *Human Gene Therapy*, vol. 19, no. 1, pp. 83–95, Jan 2008.

[4] Eduardo Garcia-Cimbrelo, Mar Tapia, and Carmen Martin-Hervas, "Multislice Computed Tomography for evaluating acetabular defects in revision THA," *Clin Orthop Relat Res*, vol. 463, pp. 138–143, 2007.

[5] C. Lemmens, D. Faul, and J. Nuyts, "Suppression of metal artifacts in ct using a reconstruction procedure that combines map and projection completion," *IEEE Transactions on Medical Imaging*, vol. 28, no. 2, pp. 250–260, Feb. 2009.

[6] Tobias Heimann and Hans-Peter Meinzer, "Statistical shape models for 3d medical image segmentation: A review," *Medical Image Analysis*, vol. 13, pp. 543–563, 2009.

[7] Marco Loog and Bram van Ginneken, "Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification," *IEEE Transactions on Medical Imaging*, vol. 25, no. 5, pp. 602–611, May 2006.

[8] J Folkesson, E.B. Dam, O.F. Olsen, P.C. Pettersen, and C. Christiansen, "Segmenting articular cartilage automatically using a voxel classification approach.," *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 106–115, Jan 2007.

[9] E. M. van Rikxoort, B. de Hoop, S. van de Vorst, M. Prokop, and B. van Ginneken, "Automatic segmentation of pulmonary segments from volumetric chest CT scans," *IEEE Transactions on Medical Imaging*, vol. 28, no. 4, pp. 621–630, Apr. 2009.

[10] Bram van Ginneken, Mikkel B. Stegmann, and Marco Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, February 2006.

[11] D. Maleike, M. Nolden, H.-P. Meinzer, and I. Wolf, "Interactive segmentation framework of the medical imaging interaction toolkit," *Computer Methods and Programs in Biomedicine*, vol. 96, no. 1, pp. 72–83, Oct. 2009.

[12] R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and S. Verzakov D.M.J. Tax, *PRTools4 - A Matlab Toolbox for Pattern Recognition*, TU Delft, 2007.

[13] Ayelet Akselrod-Ballin, Meirav Galun, John Moshe Gomori, Massimo Filippi, Paola Valsasina, Ronen Basri, and Achi Brandt, "Automatic segmentation and classification of multiple sclerosis in multichannel MRI," *IEEE Transactions on Bio-Medical Engineering*, vol. 56, no. 10, pp. 2461–2469, Oct. 2009.