A COMPARISION OF MODEL-BASED METHODS FOR KNEE CARTILAGE SEGMENTATION

J. Cheong[1], N. Faggian[2], G. Langs[3,4], D. Suter[1] and F. Cicuttini[5]
[1]*Dept. of Electrical and Computer Systems Engineering, Monash University, Australia*
[2]*Clayton School of Information Technology, Monash University, Australia*
[3]*Institute for Computer Graphics and Vision, Graz University of Technology, Austria*
[4]*Pattern Recognition and Image Processing Group, Vienna University of Technology, Austria*
[5]*Dept. of Epidemiology and Preventive Medicine, Monash University, Australia*

Keywords: Segmentation, Model-based, Cartilage, Osteoarthritis.

Abstract: Osteoarthritis is a chronic and crippling disease affecting an increasing number of people each year. With no known cure, it is expected to reach epidemic proportions in the near future. Accurate segmentation of knee cartilage from magnetic resonance imaging (MRI) scans facilitates the measurement of cartilage volume present in a patient's knee, thus enabling medical clinicians to detect the onset of osteoarthritis and also crucially, to study its effects. This paper compares four model-based segmentation methods popular for medical data segmentation, namely Active Shape Models (ASM) (Cootes et al., 1995), Active Appearance Models (AAM) (Cootes et al., 2001), Patch-based Active Appearance Models (PAAM) (Faggian et al., 2006), and Active Feature Models (AFM) (Langs et al., 2006). A comprehensive analysis of how accurately these methods segment human tibial cartilage is presented. The results obtained were benchmarked against the current "gold standard" (cartilage segmented manually by trained clinicians) and indicate that modeling local texture features around each landmark provides the best results for segmenting human tibial cartilage.

1 INTRODUCTION

Model-based segmentation methods have become a standard and popular method for detecting structures in medical images. They meet the need to consistently identify landmarks on images with complex content by the use of a priori knowledge. Active Shape Models (ASM) (Cootes et al., 1995) and Active Appearance Models (AAM) (Cootes et al., 2001) have proven to provide reliable localisation of landmarks.

ASM is based on building a statistical shape model for a given class of objects. Since its introductory application for segmenting heart ventricles from echocardiogram, ASM has undergone many modifications and found numerous new applications for segmenting medical images. Some examples include segmenting lungs from chest radiographs (van Ginneken et al., 2001), metacarpal bone from hand radiographs (Langs et al., 2003), and human knee cartilage from MRI scans (Cheong et al., 2005). The concept of ASM was extended to include a model of the object's texture and this new method was termed Active Appearance Models, AAM. Common uses of AAM these days include segmenting brain structure from brain MRI and heart ventricles from cardiac MRI (Oost et al., 2003).

Recent modifications to speed up AAM have resulted in methods that model only relevant parts of an object's texture instead of the entire texture. For example, Patch-based AAM (PAAM) (Faggian et al., 2006) model a texture patch oriented along each landmark, while Active Feature Models (AFM) (Langs et al., 2006), represent image texture by means of local texture descriptors. These methods produce results similar to or better than AAM even with a significantly reduced amount of training information.

The motivation behind this study is to identify a model-based segmentation method that can accurately and efficiently segment human tibial cartilage, with the ultimate goal being a fully automated method. Accurate segmentation enables the cartilage volume of a patient to be estimated, and studies have shown that such measures will enable evaluation of osteoarthritis severity in the knee (Cicuttini et al., 2003; Raynauld, 2002). The current method

Cheong J., Faggian N., Langs G., Suter D. and Cicuttini F. (2007).
A COMPARISION OF MODEL-BASED METHODS FOR KNEE CARTILAGE SEGMENTATION.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IFP/IA*, pages 290-295

of cartilage volume measurement involves some form of manual segmentation performed by a trained clinician; and the process is slow, tedious, and subjective. There is thus a strong demand to develop a non-subjective and more efficient segmentation method to meet the needs of large-scale clinical trials and epidemiological studies that have been conducted to evaluate therapies that may slow down or stop cartilage degradation.

This paper is organised as follows. In Section 2, we describe the four model-based segmentation methods used. In Section 3, the experimental setup is explained and the results are presented and discussed. Concluding remarks are given in Section 4.

2 METHODS

2.1 Active Shape Models

Active Shape Models (ASM) (Cootes et al., 1995) is a model-based segmentation technique that models the variation of object shape in images. Objects are represented as a set of n labeled points referred to as landmarks. The locations of these landmarks are extracted either manually or automatically from a set of p training images. In the segmentation process, the algorithm searches for the best candidate points for each landmark in the image based on local edge features, with the solution space constrained by a global shape model.

The shape model is constructed by stacking the landmarks $(x_1, y_1, ...x_n, y_n)$ for each training image into a shape vector, $\mathbf{s}_i$.

$$\mathbf{s}_i = (x_1, y_1, ...x_n, y_n)^T. \quad (1)$$

The shape vectors are aligned by scaling, rotation and translation using an iterative scheme known as Generalised Procrustes Analysis (GPA) (Goodall, 1991) to minimise the sum of squared distances between the landmarks. A mean shape, $\bar{s}$, is then calculated from the shape vectors $\mathbf{s}$,

$$\bar{s} = \frac{1}{p}\sum_{i=1}^{p} \mathbf{s}_i, \quad (2)$$

as well as the covariance matrix,

$$C = \frac{1}{p-1}\sum_{i=1}^{p} (\mathbf{s}_i - \bar{s})(\mathbf{s}_i - \bar{s})^T. \quad (3)$$

Principal component analysis (PCA) is then applied using eigenvalue decomposition of the covariance matrix. Eigenvectors corresponding to the r largest eigenvalues λ_i are retained in a matrix S. The number of eignevalues to retain, r, is chosen such that their sum sufficiently explains the variance in the training shapes, since the variance explained by each eigenvector is equal to the corresponding eigenvalue. r is usually set such that the explained variance ranges from 90% to 99.5%. Any shape in the training set can now be approximated by:

$$\hat{s} = \bar{s} + S\alpha, \quad (4)$$

where α is a vector of r elements containing the shape coefficients, calculated by:

$$\alpha = S^T(\hat{s} - \bar{s}). \quad (5)$$

To ensure that new shapes generated are in the allowable shape domain, the values of α are constrained to lie within the range $\pm m\sqrt{\lambda_i}$, where m has a value between 2 and 3.

The shape model is fitted onto an image by placing the mean shape, with shape coefficients initialised to 0, onto an initial location. The neighborhood of each landmark of the initial fit is then examined to find better locations for the fitting. This is implemented by examining the normals along each landmark for the strongest edge. The shape and pose of the fit is then updated, constrained by limits set to the variation of α. This procedure is iterated until convergence, i.e. there is no significant change in shape between two consecutive iterations.

There are many different improvements that have been made to ASMs, these usually deal with different implementations of better landmark position searches. The most common improvement is to build some form of texture model around each landmark (van Ginneken et al., 2001; Yan et al., 2002), very much like PAAM, discussed in Section 2.3. This extra model is claimed to produce more accurate landmark localisation compared to searching only for the strongest edge.

2.2 Active Appearance Models

Active Appearance Model (AAM) (Cootes et al., 2001) is a more powerful and more computer-intensive model based segmentation technique. Unlike ASM, it models two principle modes of object variation in images, shape and texture. The idea behind AAM is to improve segmentation results by modeling the complete texture of an object. AAM uses a fixed number of landmarks to represent objects and encodes shape variation much like ASM.

The texture variation of an object is modeled in a similar fashion. In the texture model, GPA is replaced with an object transformation step. Here, objects are transformed to the mean coordinate frame. This mean

is provided by the shape model in equation (4). PCA is then applied to the aligned objects to form a simple linear representation for a novel texture, $\hat{t}$:

$$\hat{t} = \bar{t} + T\gamma, \tag{6}$$

where $\bar{t}$ is the mean texture, T is a linear combination of the orthogonal subspace of textures, and γ consists of the texture coefficients. By combining these texture and shape models, it is possible to render a novel image:

$$I(\alpha,\gamma) = F(\bar{s} + S\alpha, \bar{t} + T\gamma). \tag{7}$$

Rendering is defined as the process of transforming a generated texture to fit a desired shape. The texture exists in a shape free representation and is bi-linearly re-sampled (F) to the desired shape using a piecewise affine transformation. The shape model is triangulated and these triangles are transformed to the mean coordinate frame through a series of affine transformation, different for each triangle. The search process now involves fitting the texture and shape model to the image. This is a non-linear optimisation task and there are many strategies to implement this. For this paper, Inverse Compositional Image Alignment (ICIA) was used for AAM fitting (Faggian et al., 2005). This approach minimises the difference, e, between the mean AAM template and the image:

$$e = \sum_x [A_0 - I(W(x;\alpha))]^2, \tag{8}$$

where A_0 is the mean shape and mean texture that AAM renders and $I(W(x;\alpha))$ is the image sampled to the mean shape coordinate frame using the shape coefficients α.

2.3 Patch-based Active Appearance Models

Patch-based AAM (PAAM) (Faggian et al., 2006) is a modification to AAM. It differs from AAM in the way it samples texture in images. Instead of using triangulation, PAAM makes use of oriented patches centered on each landmark. An oriented patch is defined at each landmark location by the shape model's connectivity using its two adjacent landmarks. For example, the second landmark in a model, v_2 and its adjacent landmarks, v_1 and v_3 are used to define a patch about v_2. A patch is now constructed by computing two principle directions. The first principle direction is the normal of v_1 to v_3, called $u_\perp$. The second principle direction is orthogonal to the first and is called u. These principle directions can be used to transform pixels from one patch into another using Barycentric coordinates, an example is shown in Figure 1.

The size of each patch is defined by a constant, k. This constant must be suitably selected by the user during PAAM construction. In practice, a larger k increases the robustness of the method with respect to alignment errors while a smaller k increases convergence when alignment is known to be good. The fitting process for PAAM is similar to AAM; ICIA is used to minimise the difference, e, between the mean AAM template and the patch-sampled image.

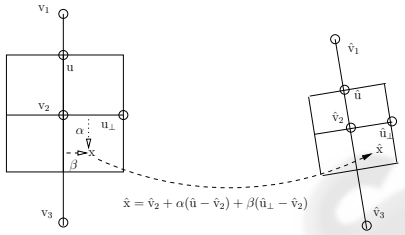


Figure 1: Transforming a pixel, x, in patch defined by v_1, v_2, v_3 to its corresponding position, $\hat{x}$, in a different patch defined by $\hat{v}_1$, $\hat{v}_2$, $\hat{v}_3$, (Faggian et al., 2006).

2.4 Active Feature Models

Active Feature Models (AFM) (Langs et al., 2006) build a model based on a set of training images for which corresponding positions of a set of landmarks are known. A statistical shape model is built based on the training shapes, much like ASM. Instead of modeling local texture around the landmarks directly, as in PAAM, AFM compacts the local texture by feature extraction using descriptors. In addition, AFM does not use landmark connectivity like PAAM, thus descriptors are independent of the shape's contour direction. Any descriptor can be used, allowing for straightforward adaptation of the algorithm to different data, especially if descriptors with favorable specificity and robustness with respect to the application are known.

During training, model parameters are perturbed randomly generating a large number of displaced model instances. A functional relation is then learned from the resulting feature vectors $\mathbf{f}$ describing local texture and the corresponding parameter displacement $\delta\mathbf{p}$ by Canonical Correlation Analysis (CCA). This is analogous to a CCA based AAM search approach proposed in (Donner et al., 2006).

The AFM search process involves extracting local texture features at the current landmark position estimates and updating the model parameters according to the trained relation

3 EXPERIMENTS

3.1 Setup

Experimental results are reported for two datasets, medial tibial cartilage and lateral tibial cartilage. Cartilage boundaries segmented by Operator 1 in July 2005 were used as the baseline "gold standard" measures for all experiments as this was the only dataset available to us until recently. Manual segmentations carried out as followup tests by Operators 1 and 2 in September 2006 provide target benchmarks. From our database of "gold standard" cartilage outlines, we randomly selected 5 patients, totalling 78 medial and 87 lateral image slices. Correspondences of 32 landmarks for images of each dataset were obtained using a Minimum Description Length (MDL) shape algorithm (Thodberg, 2003). 5-fold cross-validation was performed on the datasets for ASM, AAM, PAAM and AFM to compare segmentation results.

All methods were initialised with the true center of gravity (CoG) of the corresponding cartilage, calculated by finding the mean of the 32 landmark locations. An optimal normal length of 3 was used to search for landmark positions with ASM. For the PAAM experiments, the patch size, k, was set to 14 and for the AFM experiments, steerable filters (Freeman and Adelson, 1991) were used due to their reliability and low dimensionality. A gabor jet with filter frequency $\phi = 1$ and directions $\theta \in \{\pi/2, 3\pi/4\}$ proved to give the best descriptions for cartilage segmentation.

The following measures were employed to compare the different segmentation results, Goodness of Fit (GOF), sensitivity, and the difference in area between the segmentation results and the "gold standard". The GOF measure is defined as:

$$\text{GOF} = \frac{TP}{TP+FP+FN}, \tag{9}$$

where TP represents true positive (segmented area correctly classified as cartilage), FP represents false positive (segmented area incorrectly classified as cartilage), and FN represents false negative (cartilage area incorrectly classified as background). A GOF of 1 represents a perfect overlap with the "gold standard" and a GOF of 0 represents no overlap with the "gold standard". Sensitivity is defined:

$$\text{Sensitivity} = \frac{TP}{\text{No. of pixels in "gold standard" seg.}}. \tag{10}$$

Sensitivity measures the proportion of the "gold standard" that is segmented and provides a useful indication of oversegmentation or undersegmentation when combined with the GOF measure. GOF and sensitivity provide a measure of the segmentation accuracy while the area difference provides an estimation of measurement accuracy.

Table 1: Mean and standard deviation results of the Goodness of Fit (GOF), sensitivity, and difference in area when compared to the "gold standard" for different segmentation methods.

Medial Tibial Cartilage	**GOF** $\mu \pm \sigma$	**Sensitivity** $\mu \pm \sigma$	**Area Diff. (mm^2)** $\mu \pm \sigma$
ASM	0.60 ± 0.14	0.74 ± 0.14	10.27 ± 7.85
AAM	0.41 ± 0.15	0.56 ± 0.20	21.31 ± 14.22
PAAM	0.64 ± 0.15	0.74 ± 0.13	9.95 ± 7.38
AFM	0.54 ± 0.11	0.69 ± 0.10	9.08 ± 7.27
Manual Operator 1	0.82 ± 0.07	0.89 ± 0.06	3.47 ± 2.79
Manual Operator 2	0.79 ± 0.09	0.88 ± 0.08	5.17 ± 4.42
Lateral Tibial Cartilage	**GOF** $\mu \pm \sigma$	**Sensitivity** $\mu \pm \sigma$	**Area Diff. (mm^2)** $\mu \pm \sigma$
ASM	0.60 ± 0.16	0.74 ± 0.17	18.68 ± 16.85
AAM	0.48 ± 0.13	0.61 ± 0.19	30.35 ± 20.27
PAAM	0.72 ± 0.09	0.79 ± 0.10	15.76 ± 14.50
AFM	0.54 ± 0.20	0.69 ± 0.20	17.65 ± 17.77
Manual Operator 1	0.85 ± 0.06	0.92 ± 0.05	3.09 ± 2.67
Manual Operator 2	0.80 ± 0.08	0.88 ± 0.06	4.92 ± 3.64

3.2 Results

The results of all experiments are given in Table 1, and Figures 2 and 3 show typical result for each experiment. PAAM produces the best segmentation results for both medial and lateral tibial cartilage. It is the only method from the four that models background texture locally around the landmarks, taking into account the connectivity of the landmarks. The texture patch model includes both foreground and background, thus enabling PAAM to correctly locate the cartilage boundaries with higher success.

AAM produces the worst results because it models only the texture within the cartilage. This leads to poor performance because there is inherently a large variation in texture due to the biochemically heterogeneous property of cartilage. Without modeling any background information, it is difficult for AAM to locate the cartilage boundaries.

AFM produces reasonable results even though the data in Table 1 ranks AFM as third best behind ASM and PAAM. The segmented shapes of PAAM and AFM are smoother than ASM and examples can be observed in Figures 2 and 3. Background texture is not used directly by AFM but represented by means of features extracted by descriptors using steerable filters. AFM tends to oversegment images because the

filter blurs the edges of the cartilage, thus making it harder to locate the correct boundary.

Figure 2a displays a common problem that ASM exhibits. Landmarks are attracted to regions with strong edges, therefore in slices where the tibial and femoral cartilage touch, there is no edge information for the upper boundary of the tibial cartilage and ASM locates the upper boundary of the femoral cartilage instead. The resulting shape estimate from ASM also tends to be uneven, compared to PAAM and AFM. This is because when new landmark locations are found, the pose and shape parameters are updated to best match the shape model to these landmark points. This step can potentially shift the shape model away from the true landmark positions.

All methods perform very poorly on a few particular slices, namely slices located at the very ends of the cartilage in the middle of the knee. On these slices, the cartilage areas are very small and also display very poor contrast and high shape variability. Even the manual operators exhibit some discrepancy in their results for these slices.

4 CONCLUSION

In this paper, we have compared four different model-based segmentation methods (ASM, AAM, PAAM and AFM) for the purpose of tibial cartilage segmentation. All four methods have been fine tuned so that only the optimal settings for cartilage segmentation were compared. In conclusion, methods that work on local texture perform better for cartilage segmentation due to a restriction to more relevant information being used for regression and fitting. In addition, the use of landmark connectivity for orientation consistency results in an even more specific description of the texture.

ACKNOWLEDGEMENTS

The authors would like to thank Fahad Hanna and Yuanyuan Wang for their time and contribution with performing the manual cartilage segmentation used in this paper, and also René Donner for providing parts of the AFM implementation. Georg Langs has been supported by the Austrian Science Fund(FWF) under the Grant P17083-N04 (AAMIR).

REFERENCES

Cheong, J., Suter, D., and Cicuttini, F. (2005). Development of semi-automatic segmentation methods for measuring tibial cartilage volume. In *Proc. DICTA 2005*, pages 307–314.

Cicuttini, F., Wluka, A. E., Forbes, A., and Wolfe, R. (2003). Comparison of tibial cartilage volume and radiologic grade of the tibiofemoral joint. *Arthritis and Rheumatism*, 48:682–688.

Cootes, T., Edwards, G., and Taylor, C. (2001). Active appearance models. *IEEE TPAMI*, 23(6):681–685.

Cootes, T., Taylor, C., Cooper, D., and Graham, J. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1).

Donner, R., Reiter, M., Langs, G., Peloschek, P., and Bischof, H. (2006). Fast active appearance model search using canonical correlation analysis. *IEEE TPAMI*, 28(10):1690 – 1694.

Faggian, N., Paplinski, A., and Sherrah, J. (2006). Local texture patches for active appearance models. In *Proc. IVCNZ 2006*.

Faggian, N., Romdhani, S., Paplinski, A., and Sherrah, J. (2005). Color active appearance model analysis using a 3d morphable model. In *Proc. DICTA 2005*, pages 407–414.

Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE TPAMI*, 13(9):891–906.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shapes. *Journal of the Royal Statistical Society*, 53:285–339.

Langs, G., Peloschek, P., and Bischof, H. (2003). Asm driven snakes in rheumatoid arthritis assessment. In *Proc. SCIA 2003*, pages 454–461.

Langs, G., Peloschek, P., Donner, R., Reiter, M., and Bischof, H. (2006). Active feature models. In *Proc. ICPR 06,*, volume 1, pages 417–420.

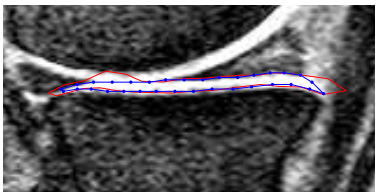
Oost, C. R., Lelieveldt, B. P. F., Üzümcü, M., Lamb, H. J., Reiber, J. H. C., and Sonka, M. (2003). Multi-view active appearance models: Application to x-ray lv angiography and cardiac mri. In *IPMI 2003*, pages 234–245.

Raynauld, J. (2002). Magnetic resonance imaging of articular cartilage: toward a redefinition of ”primary” knee osteoarthritis and its progression. *The Journal of Rheumatology*, 29:1809–1810.

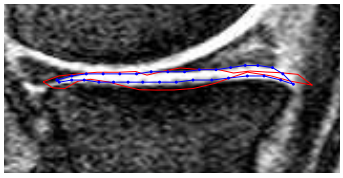
Thodberg, H. (2003). Minimum description length shape and appearance models. In *IPMI 2003*, pages 51–62.

van Ginneken, B., Frangi, A. F., Staal, J., ter Haar Romeny, B. M., and Viergever, M. A. (2001). A non-linear gray-level appearance model improves active shape model segmentation. In *Proc. MMBIA 2001*, pages 117–127. IEEE Computer Society Press.

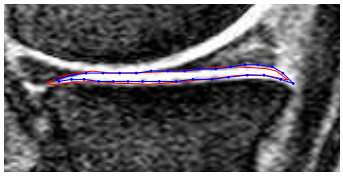
Yan, S., Liu, C., Li, S., Zhang, H., Shum, H.-Y., and Cheng, Q. (2002). Texture-constrained active shape models. In *Proc. GMBV 2002*, pages 107–113.



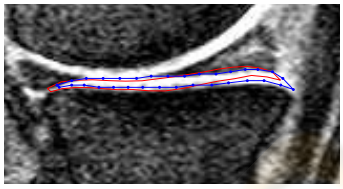
(a) ASM
GOF = 0.70, Sens = 0.92, Area Diff. = 11.43mm^2



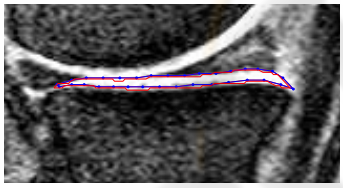
(b) AAM
GOF = 0.53, Sens = 0.76, Area Diff. = 10.94mm^2



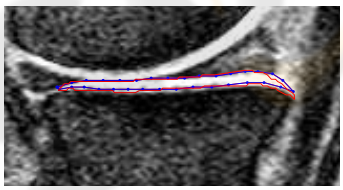
(c) PAAM
GOF = 0.77, Sens = 0.87, Area Diff. = 0.78mm^2



(d) AFM
GOF = 0.59, Sens = 0.70, Area Diff. = 0.92mm^2

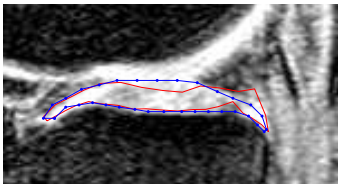


(e) Manual Operator 1
GOF = 0.88, Sens = 0.94, Area Diff. = 0.68mm^2

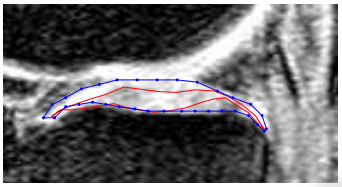


(f) Manual Operator 2
GOF = 0.79, Sens = 0.91, Area Diff. = 4.54mm^2

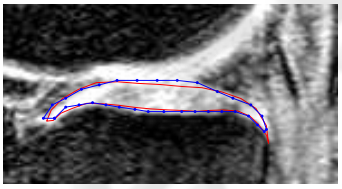
Figure 2: Example results for medial tibial cartilage. Segmentations are given by the solid line and the "gold standard" by the dotted line. NOTE: Images are best viewed in colour.



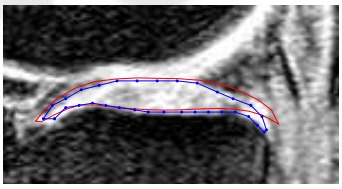
(a) ASM
GOF = 0.65, Sens = 0.72, Area Diff. = 16.70mm^2



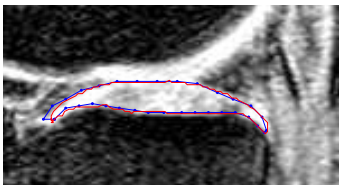
(b) AAM
GOF = 0.43, Sens = 0.44, Area Diff. = 51.56mm^2



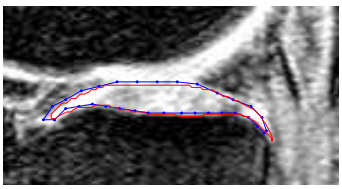
(c) PAAM
GOF = 0.79, Sens = 0.82, Area Diff. = 13.28mm^2



(d) AFM
GOF = 0.69, Sens = 0.86, Area Diff. = 15.72mm^2



(e) Manual Operator 1
GOF = 0.92, Sens = 0.95, Area Diff. = 1.42mm^2



(f) Manual Operator 2
GOF = 0.88, Sens = 0.90, Area Diff. = 7.08mm^2

Figure 3: Example results for lateral tibial cartilage. Segmentations are given by the solid line and the "gold standard" by the dotted line. NOTE: Images are best viewed in colour.