

# Fully Automatic Segmentation of the Knee Joint using Active Appearance Models

Graham Vincent, Chris Wolstenholme, Ian Scott, and Mike Bowes

Imorphics Ltd., Kilburn House, Manchester Science Park, Manchester, M15 6SE, UK.

**Abstract.** We present a fully automatic model based system for segmenting bone and cartilage in magnetic resonance (MR) images of the knee. The segmentation method is based on Active Appearance Models (AAM) built from manually segmented examples from the Osteoarthritis Initiative database. High quality correspondences for the model are generated using a Minimum Description Length (MDL) Groupwise Image Registration method. A multi start and hierarchical modelling scheme is used to robustly match the model to new images.

The model has been applied to the MICCAI 2010 Grand Challenge test data with no tuning from the supplied training data, and successfully segmented all the test data automatically to a good degree of accuracy.

## 1 Introduction

Fully automated segmentation of bone and cartilage could be hugely beneficial in clinical trials and in the orthopaedic industry.

Quantitative image analysis of cartilage from knee MRI is well established method but its use in medium to large scale clinical studies has been hindered by the need for manual segmentation of the cartilage. Manual segmentation has been shown to be sufficiently sensitive to detect change [5], but is labour intensive, taking up to several hours per image, which impacts on the size of clinical trials and cost. Small trials are less well powered and less likely to detect a change. The Osteoarthritis Initiative (OAI) provides a large dataset of images, but no scalable means of performing quantitative analysis of the MR images.

In orthopaedics detailed information from MR and CT can be used to plan for the best positioning of surgical cuts, allow intra-operative plan registration and inform prosthesis design.

We have developed a statistical model based segmentation method for the analysis of bone, cartilage and other soft tissues to meet these needs.

The model is based on the Active Appearance Models (AAMs) of Cootes *et al.*, in which the statistics of shape and image information are calculated from a training set of images and used to match to new images. Variants of AAMs have been extensively developed (see [2] for a review).

The model is built from 3D DESS Sagittal Water Excitation images from part of the OAI database (available for public access at <http://www.oai.ucsf.edu/>). We used 80 subjects from this dataset to build the model.

## 2 Methods

### 2.1 Generating surface correspondences

An expert segmenter segmented the femur bone, tibia bone, femur cartilage and tibia cartilage using a combination of manual line drawing and LiveWire [3], and then converted them into surfaces.

Statistical appearance models rely on a large set of anatomically equivalent landmarks (also known as *correspondences*) across the region of interest.

To obtain the anatomical correspondences on the bone surfaces we used a variant of the Minimum Description Length approach to Groupwise Image Registration (MDL-GIR) of Cootes *et al* [1]. The MDL-GIR method finds the set of deformations which register all the images together as efficiently as possible. This idea is made concrete by the use of Information Theory to define the amount of information required to encode a model using a particular set of deformations. The method is an optimisation to find the set of deformations requiring the least amount of information to encode. The output is a reference mean image and a set of deformations which map the mean image to each example image.

We apply the MDL-GIR method to the signed distance images derived from the segmented surfaces. The output reference mean image is, like the input images, a signed distance image and can be straight forwardly segmented using the zero valued iso-surface. The mean surface is then propagated by the appropriate deformation field into the frame of each example. For each example the propagated surface lies close to the segmented surface and is projected onto it to generate correspondence points which are guaranteed to lie on the segmented surface.

To obtain correspondences on the inside and outside surfaces of the cartilages we consider the normal from each bone correspondence point normally covered by cartilage. Correspondence points are placed on the intersections of the bone normal with the inside and outside of the cartilage.

Finally the bone and cartilage correspondences are concatenated to form a superset of points to be used in building the appearance models.

The number of correspondence points output from this process are as follows: femur 60457, tibia 39239, femoral cartilage 37249, tibial cartilage 20459.

### 2.2 Active appearance models

An appearance model is a statistical model of the shape of a structure and associated imaging information. It is useful to process the imaging information further to obtain feature response images such as gradients, corners and other points of interest [4]. We refer to all such imaging information and their derivatives as texture.

An appearance model has a set of parameters which control both the shape and the texture, and are *generative* i.e. a specific parameterisation can generate a realistic looking example of the shape and texture.

An AAM can match its appearance model an image from a rough initial estimate, by optimising the model parameters to generate an example which matches the image as closely as possible (using the least squares sum of residuals). This can be made very efficient by pre-computing the Jacobian describing the average change in residuals with respect to changes in model parameters on a training set.

AAMs require an initial estimate of the model parameters including position, rotation and scale. We provide multiple initial estimates at a grid of starting points across the image. The grid of starting points are typically 30mm apart in all directions. This is done at a low image and model resolution with a small number of measured residuals to make it reasonably fast. The results of these searches are ranked according to the sum of squares of the residual, and a proportion (typically 75%) removed from consideration. The remaining search results are used to initialise models at a higher resolution, and so on. Finally, the single best result at the highest resolution gives the segmentation result.

In addition we use a hierarchical scheme whereby a single model of femur and tibia generates an interim segmentation. This is then used to initialise individual models of smaller regions e.g. the femur bone, tibia bone, femur cartilage, tibia cartilage.

### 2.3 Extrapolation of shaft

The OAI model, developed to segment the articulating areas of the knee joint, is restricted to a region roughly 70mm from the knee centre and therefore did not segment all the shaft in the Grand Challenge images.

We apply an additional step which used a long shaft shape model (derived from a small number of CT images) to *extrapolate* the shaft from a given result from the short shaft OAI model.

The correspondence between the short shaft OAI shape model and the long shaft shape model are calculated by aligning the mean shapes from the two models. For a given result from the short shaft OAI model the positions of the equivalent long shaft points are then estimated using these correspondences. A fit of the long shaft shape model is applied, the fit having zero weights in the region not covered by the short shaft model.

It is obviously not ideal that the shaft is not fitted to image information, nor that the long shaft shape model was learnt over different and restricted training set compared to the OAI AAM. However this time-saving compromise was necessary in order to compete in the Grand Challenge. Ideally we would have extended the shaft in our OAI model.

### 2.4 Segmentation pipeline

In summary, the segmentation proceeds according to the following pipeline:

- For each image:

- Run N (typically 10s-100s) AAMs of the knee joint from a grid of starting positions across the image at low resolution.
  - \* Run the 25% best results at increased resolution
  - \* Repeat until at highest resolution
- Choose best result
- Initialise and run separate AAMs of femur, tibia and cartilage
- Extrapolate final AAM surfaces with long shaft shape model

### 3 Results and Discussion

In this section we present the results of the model on test datasets provided for the MICCAI 2010 Grand Challenge workshop (see Table 1 and Figure 1)

It was extremely encouraging that the OAI model worked "out of the box" - that is, not tuned to the Grand Challenge training data. The model segmented every example reasonably well (as indicated by a maximum mean distance 1.4 mm).

Interestingly, it looks like there is a systematic bias in the OAI model towards thinner cartilage than for the Grand Challenge protocol, particularly on the femoral side (see Table 1). This could be because of training data selection effects or because of systematic segmentation protocol effects.

We should point out that the OAI data has mild OA whereas the Grand Challenge data contained a number of endpoint OA cases with many osteophytes which are not well captured by the OAI model. We have seen from work on other datasets that osteophytic areas can actually be quite well modelled as they are more systematic than might be expected.

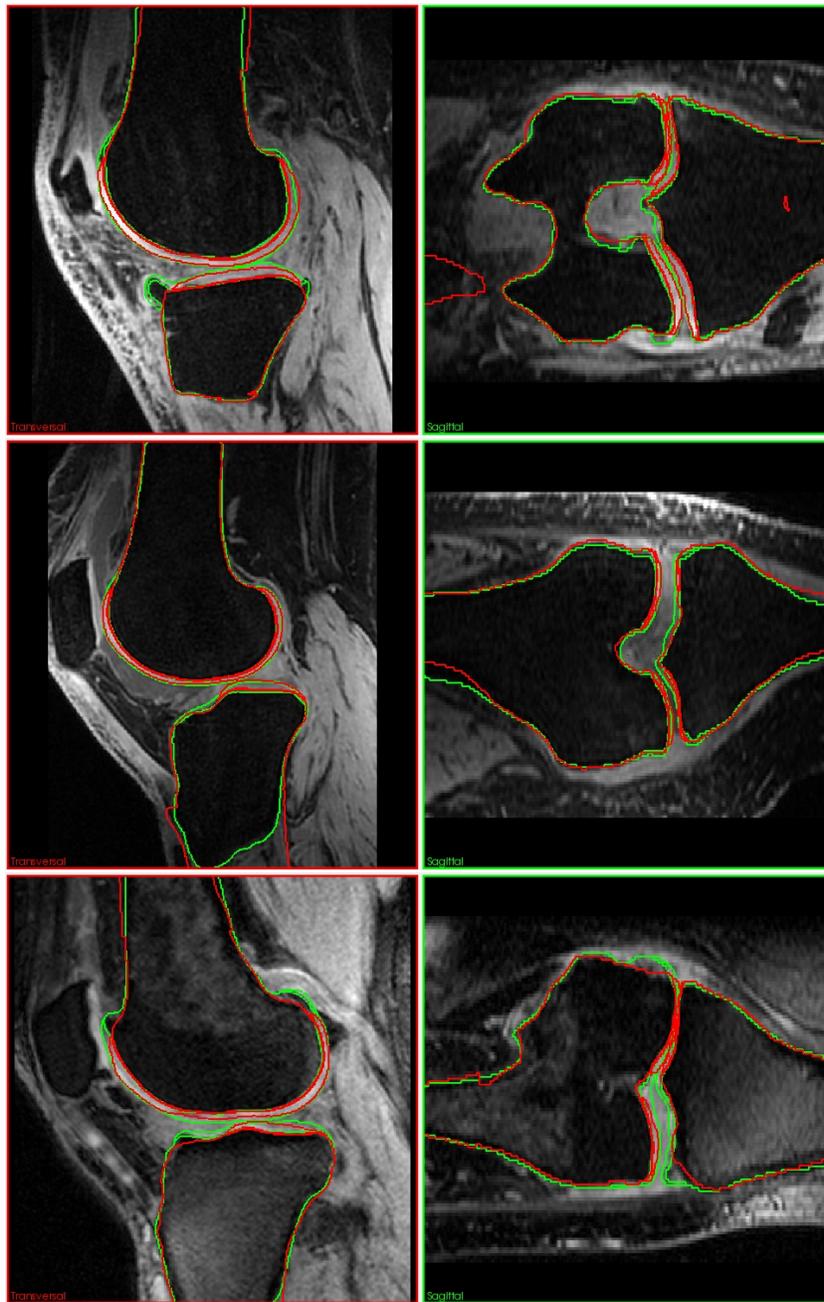
These last two points argue for including the Grand Challenge training data in the model in order to capture the specific imaging and segmentation protocols, the long shaft and more disease states.

The time taken for an image search depends on the number of AAM starting points. Given reasonable assumptions about the placement of the centre of the knee (e.g. that at least 50mm of femur and tibia are within the image) the average run time on a Dell Vostro 420 is approximately 15 minutes.

### 4 Conclusion

In this paper we have presented a fully automatic model based method to segment the bone and cartilage at the knee joint built from DESS images from the Osteoarthritis Initiative database.

The results of the method without tuning to the Grand Challenge protocol are very encouraging and demonstrate the robustness of the system. However the model dataset is quite different from the Grand Challenge data and it would be interesting to include the training data in the model for a reanalysis.



**Fig. 1.** Different views on selected test cases 8 (top), 16 (center), and 31 (bottom). The outline of the reference segmentation is displayed in green, the outline of the automatic method described in this paper in red.

## 5 Acknowledgements

The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners. We would like to thank Chris Taylor, Tim Cootes and Tomos Williams for a long collaboration.

## References

1. T. F. Cootes and C. Taylor C V. Petroviand R. Schestowitz. Groupwise construction of appearance models using piece-wise affine deformations. *16th British Machine Vision Conference. Volume, 2*:879888, 2005.
2. Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543 – 563, 2009.
3. Eric N. Mortensen and William A. Barrett. Interactive segmentation with intelligent scissors. In *Graphical Models and Image Processing*, pages 349–384, 1998.
4. I. M. Scott, T. F. Cootes, and C. J. Taylor. Improving appearance model matching using local image structure. In *In Information Processing in Medical Imaging, 18th International Conference*, pages 258–269. Springer, 2003.
5. Tomos G. Williams, Andrew Holmes, John C. Waterton, Rose A. Maciewicz, Anthony F. P. Nash, and Christopher J. Taylor. Regional quantitative analysis of knee cartilage in a population study using mri and model based correspondences. *ISBI*, pages 11–314, 2006.

Img	Femur bone			Tibia bone			Fem. cartilage			Tibial cartilage			Total Score
	AvgD	RMSD	Scr	AvgD	RMSD	Scr	VOE	VD	Scr	VOE	VD	Scr	
	[mm]	[mm]		[mm]	[mm]		[%]	[%]		[%]	[%]		
1	0.81	1.39	55	0.59	0.97	61	27.7	-24.3	47	28.9	-1.0	88	62.6
2	1.09	1.62	44	0.75	1.12	52	40.0	-37.7	35	29.9	-13.3	66	49.3
3	0.69	1.12	63	0.47	0.73	69	31.7	-13.8	64	42.9	0.0	84	70.1
4	0.74	1.12	61	0.69	1.01	56	45.9	-40.5	33	44.2	-43.9	34	46.2
5	0.90	1.54	50	0.70	1.19	53	34.2	-24.2	45	38.1	-13.6	62	52.3
6	1.11	1.83	39	0.77	1.26	49	35.5	-24.7	44	29.6	-18.5	57	47.0
7	0.92	1.42	51	0.77	1.29	48	34.2	-30.9	38	23.3	-10.5	73	52.5
8	1.24	2.07	32	0.68	1.14	54	43.5	-41.8	34	29.1	-15.6	62	45.5
9	1.05	1.64	44	0.62	0.99	59	48.7	-36.9	32	38.6	2.3	82	54.4
10	1.14	1.76	40	0.76	1.22	50	43.6	-41.6	34	31.1	-24.3	46	42.4
11	1.12	1.90	38	0.84	1.39	44	30.8	-25.3	44	30.0	0.8	88	53.4
12	0.43	0.77	76	0.49	0.82	67	40.5	-30.8	35	43.9	-3.3	78	64.0
13	1.21	2.19	31	0.77	1.33	47	36.7	-5.5	77	28.2	11.7	69	56.0
14	0.65	1.03	65	0.84	1.37	44	31.8	-8.5	73	20.1	-9.4	76	64.6
15	0.66	1.18	63	0.67	1.22	53	29.2	-18.6	57	32.8	-29.0	38	52.5
16	0.46	0.79	75	0.68	1.26	51	33.1	-25.4	43	34.8	-8.2	73	60.5
17	1.20	1.99	34	0.68	1.13	54	41.1	-21.1	48	29.0	-21.3	52	47.1
18	0.86	1.52	51	0.89	1.37	42	36.5	-29.4	37	27.1	-18.1	58	47.2
19	0.69	1.31	60	0.61	1.03	59	28.7	-19.8	55	38.2	-17.0	56	57.2
20	0.84	1.34	55	0.64	1.03	58	31.4	-24.0	46	25.6	-15.2	64	55.7
21	0.49	0.77	74	0.56	0.88	63	39.1	-36.2	36	32.6	-22.7	48	55.2
22	0.81	1.24	57	1.05	1.71	30	37.8	-30.4	36	30.7	-25.0	45	42.1
23	0.75	1.25	59	0.91	1.39	41	35.5	-23.3	46	48.1	47.7	32	44.6
24	0.94	1.52	49	0.69	1.16	53	36.4	-31.0	37	41.9	-12.9	62	50.3
25	1.30	2.52	23	0.98	1.62	34	35.8	-17.8	56	45.1	7.5	70	45.8
26	0.94	1.57	48	0.67	1.05	56	33.5	-29.3	38	36.7	-24.4	44	46.5
27	0.87	1.33	54	0.73	1.20	51	47.5	-37.7	33	28.3	-24.9	46	46.0
28	1.14	2.40	29	0.62	0.99	59	34.8	-11.5	67	36.4	-34.4	37	48.1
29	0.80	1.31	56	0.74	1.12	53	43.6	-39.3	34	41.8	-8.7	69	53.1
30	0.95	1.71	46	0.94	1.70	34	39.6	-31.3	36	54.3	-36.0	30	36.4
31	1.11	1.75	41	0.73	1.05	54	39.7	-28.6	35	33.7	-12.8	65	48.9
32	0.74	1.25	59	0.60	0.93	61	31.1	-15.6	61	21.6	5.6	82	65.9
33	0.57	0.99	68	0.49	0.87	66	39.3	-29.0	36	41.5	-24.6	41	52.8
34	0.68	1.37	59	0.49	0.79	68	34.4	-0.8	86	36.3	9.2	71	70.8
35	1.07	1.81	41	0.82	1.37	45	35.8	-20.5	51	33.1	-22.1	49	46.3
36	0.52	0.96	70	0.66	1.10	56	26.6	-22.4	51	24.9	-6.7	79	63.9
37	1.21	2.00	34	1.47	2.10	8	32.5	-11.4	68	41.4	46.5	35	36.2
38	0.88	1.61	50	0.60	0.95	61	30.5	-22.6	49	29.2	-3.0	84	60.8
39	1.13	1.98	36	1.37	2.43	5	37.8	-14.7	60	32.1	17.3	58	39.9
40	0.56	0.92	69	0.71	1.14	53	35.5	-28.2	37	48.3	-7.0	70	57.5
Avg	0.88	1.49	51	0.74	1.21	51	36.3	-25.2	47	34.6	-9.5	61	52.3
	$\pm 0.24$	$\pm 0.44$	$\pm 14$	$\pm 0.21$	$\pm 0.34$	$\pm 14$	$\pm 5.3$	$\pm 10.1$	$\pm 14$	$\pm 7.9$	$\pm 18.8$	$\pm 17$	$\pm 8.6$

**Table 1.** Results of the comparison metrics and scores for all 40 test cases. AvgD and RMSD are the average and RMS surface distance, respectively, VOE is the volumetric overlap error and VD indicates the volumetric difference.