USING PARTS AND GEOMETRY MODELS TO INITIALISE ACTIVE APPEARANCE MODELS FOR AUTOMATED SEGMENTATION OF 3D MEDICAL IMAGES

Kola Babalola and Tim Cootes

Division of Imaging Sciences, University of Manchester, Manchester, UK

ABSTRACT

In recent years, statistical shape models, of which Active Appearance Models (AAMs) are a subset have been increasingly applied to the automatic segmentation of medical images. AAMs are a local search technique requiring good initialisation. In 3D automatic initialisation can be achieved by multiple initialisations, registration, template matching or by application dependent heuristics. The first three can be sub-optimal in certain situations, whilst the last is not generic.

We describe a generic, fast and automated method of initialising 3D AAMs using sparse local models of texture (the parts) together with a graph capturing their pairwise geometric relationships. Initialisation then becomes a matter of searching for the parts using the parts-and-geometry model, from which the necessary pose and shape parameters are obtained.

We demonstrate the method by applying it to the segmentation of 10 subcortical structures from 3D MRI sequences of the head.

Index Terms— Segmentation, Statistical shape models, Active appearance models, Markov Random Fields, Graphs

1. INTRODUCTION

Segmentation is a necessary step in the analysis of medical images for clinical and research purposes. As the use of imaging for diagnosis, treatment and research proliferates, the development of fully automatic 3D segmentation methods is becoming increasingly important. Automated segmentation can help to free clinician time, improve consistency in segmentations and decrease the turnaround time required to analyse images. Active appearance models (AAMs)[1] have found widespread use in computer vision and medical image analysis. In 2D they are widely used for face detection, object recognition and segmentation of medical images. In 3D medical image analysis they have been used for segmentation e.g. [2],[3] and tracking of the heart e.g. [4].

The speed of AAMs comes from their search algorithm which uses the precomputed relationship of model parameters to residuals from fitting to known examples. However, AAM search is a local optimisation and therefore needs good initialisation. Furthermore, in some applications local AAMs of sub-regions within an image are desirable. For example to model the region around the hippocampus in a MR brain scan or the kidneys in an image of the torso. Automated initialisation is therefore a necessary step in the AAM search pipeline.

In this work we propose creating sparse local models with pairwise geometric relationships for the initialisation task. We describe an automated method for parts selection and demonstrate the effect of including pairwise geometric constraints to reduce problems with false matches which can occur if the parts are used independently. Furthermore, we apply the models to initialisation in segmenting subcortical structures from MRI images of the brain.

This work was funded by EPRSC grant ...

2. BACKGROUND

A wide range of methods have been used for initialising models in 2D, including searching for the approximate global pose with a single detector (such as a face detector [5]), searching for a small set of manually chosen local features using a sparse model such as [6] or using many multiple initialisations and choosing the best [7, 8]. Donner *et al.*[9] build *Sparse MRF Appearance Models* from symmetry-based interest points and match these to query images using a Markov random field formulation. Mitchell *et al.* [10] use AAMs to segment 2D slices of the heart and perform automated initialisation by identifying the left ventricle midpoint using Hough transforms. This determines the initial position of their model and multiple instances with different orientations about this position are generated. The instance giving the lowest residual is used.

In 3D the variety of methods used for automated initialisation have been limited to application specific heuristics, generation of multiple instances and registration. Kainmueller *et al.*[11] use a generalised Hough transform method to search for line features in initialising their statistical shape model based method for segmenting the mandible. Stegmann and Pedersen [4] generate multiple instances centred at points on a regular grid within the query image. Heimann *et al.* [12] sample the parameter space. They perform a global search using an evolutionary algorithm to find the best initialisation parameters by optimising 1000 possible sets of model parameters over 40 iterations. Mitchell *et al.*[3] Extend their 2D method into 3D, identifying the same structure as in the 2D case, and performing multiple initialisations at different orientations. Generating multiple instances for automated initialisation is sub-optimal and requires a judicious choice of sampling rates for the parameter space, or a choice of positions from which to initialise the model.

Patenaude [13] in segmenting subcortical brain structures use a reference frame defined by the MNI 152 atlas. The query image is registered into this space before the application of their Bayesian Appearance Model.

Our approach bears similarities with [9] and [3]. However, unlike [3], rather than selecting a specific anatomical structure to use during initialisation, our algorithm automatically selects points which can be reliably located. Furthermore, we use more than one point, therefore do not need to determine pose separately, and our method does not involve multiple initialisations. A sparse local model framework capturing geometric relationships is used to locate the feature points in a query image similar to [9], however our implementation is in 3D.

978-1-4244-4126-6/10/$25.00

3. METHODOLOGY

Our task is to obtain pose and shape parameters required to initialise a model in a query image. The basic idea of our approach is to automatically identify a sparse set of feature points in the query image and deduce the parameters using these. Local models are used in feature detection. Each of these propose multiple locations for their feature and the optimal set of candidates is selected using the parts-and-geometry framework described in section 3.3.

3.1. Evaluating local models

Firstly, we establish correspondence across the training set to give a set of deformation fields for each training image - for example by registering the training images. This allows us to obtain an "average" reference image, together with dense deformation fields which define the mapping from the reference to each training image. Therefore, we can compute the point in each training image corresponding to a given point in the reference image.

For a particular location or node in the reference image, we can construct a feature detector based on a region of size $(2L_x + 1)$ x $(2L_y + 1)$ x $(2L_z + 1)$ centred on the node. In the following we use normalised correlation, with a mask based on the region in the reference image. Applying the feature detector to a query image gives a response image, the local peaks of which are the candidates for this particular detector.

For each training image I_i, there exists a deformation field Θ_i allowing a mapping of space between it and the reference. Applying the feature detector to I_i gives R_i, the response image. Let $\mathbf{p}_k$ specify the position of a node in the reference. Its position in I_i and R_i is $\Theta_i(\mathbf{p}_k)$. We can then define a function D that computes the distance between a local peak located at $\mathbf{p}_l$ in the space of R_i and the expected position of a node in this image:

$$D_i(\mathbf{p}_k, \mathbf{p}_l) = ||\Theta_i(\mathbf{p}_k) - \mathbf{p}_l|| \tag{1}$$

For a good detector the best response will be close to the true position $\Theta_i(\mathbf{p}_k)$ in every image in the training set. However, a detector may still be useful if one of the best peaks is close to the true position.

3.2. Feature point selection

Local feature detectors are built centred on a large number of nodes (typically over 1,000) at a range of sizes and at different resolutions of the reference image (for the purposes of this paper control points of a shape model were used - see section 3.5). Each is then evaluated on a set of images. The success rate in locating the nodes and the average value of D_i over the set are computed and used to evaluate the reliability of each node. The N most reliably located nodes can be selected to be used in building the parts-and-geometry model. An algorithmic description of the selection process is given below:

For each node, at each region size and image resolution:

1. Build a feature detector for the given node
2. For each image:
   (a) Build a pyramid and select appropriate level
   (b) Run the feature detector to give a response image
   (c) Locate local peaks and rank by them by the magnitude of their responses
   (d) For each peak compute D_i using equation 1
   (e) Record the rank, r_i, of the peak with smallest D_i.

The feature detectors are then ordered. Firstly by the average rank (average of r_i), then secondly by the average value of the smallest D_i. In the experiments below a number of detectors were found to always have their best response being the closest to the true position (average value of r_i across the training set ≈ 1). These were then ranked by the average positional error when selecting a subset of good detectors. We also desire that the detectors are spread around and not clustered in one region. We address this in a greedy manner by selecting the best detector, then iteratively selecting the next best detector not within a specified radius of the current set of chosen detectors.

3.3. Parts and geometry models

Our approach is based on a parts-and-geometry framework which is widely used in computer vision e.g. [6]. The parts are the nodes and the geometry is defined by the pairwise relationships between them (see Figure 2). Let $\mathbf{p} = (x, y, z)$ be the proposed position of a patch and $p_i(I|\mathbf{p})$ the probability that patch i matches to image I at the given location. Let $p_{ij}(\mathbf{p}_i, \mathbf{p}_j)$ be the probability that two patches i and j have the given positions. Assume that we have modelled this pairwise relationship for each pair $(i, j) \in A$, where A is a set defining the arcs in a graph representing the model.

To match such a model to an image, we search the image for candidate positions for each patch. For each patch we select the candidate that maximises

$$C = \sum_{i=1}^{k} \log p_i(I, \mathbf{p}_i) + \sum_{(i,j)\in A} \log p_{ij}(\mathbf{p}_i, \mathbf{p}_j) \tag{2}$$

A range of discrete graph based solvers are available to find the global optimum for such a cost function. We use a variant of that used by [14], in which a network is created where each node can be thought of as having at most two parents. The optimal solution for this can be obtained with a variant of dynamic programming, in $O(NM^3)$ time.

3.4. Building a parts-and-geometry model

We construct the parts-and-geometry model using the selected feature points (see section 3.2) as nodes. A set of connecting arcs between the nodes is automatically defined based on the distances between pairs of nodes. We use a variant of Prim's algorithm, where each node has two parent nodes, rather than one. This involves creating the first arc from the two nodes which are closest together. We then repeat the following steps until all nodes are linked:

- compute the sum of the distances of each unlinked node to the closest two nodes in the current linked set
- select the node which has the minimum such distance, and link it to the two closest nodes in the linked set

This leads to a topology which allows a variant of the dynamic programming algorithm to efficiently find the global maxima of the cost function (Equation 2).

3.5. Initialising an AAM

An Active Appearance Model is a statistical model of both the shape of a structure and its appearance, together with an algorithm for

matching it to an image. The model is capable of synthesising an image of the object of interest, and the residual differences between the synthesised image and the target image are used to drive the search.

Shape is typically defined by points in corresponding locations on each training image. In our implementation this is given by the positions of control points at corresponding locations within the volume of each image in the training set. We construct the shape model by aligning the sets of control points on each image and applying Principal Component Analysis (PCA). A statistical model of texture is constructed by warping each grey-level image into the reference space and applying PCA to the resulting textures. An appearance model is a combination of the shape and texture models.

For the purposes of initialisation we are interested in the shape part of the appearance model. If we can locate a sufficiently large number of the control point nodes in a query image, we can deduce the pose and shape parameters needed to initialise an AAM close to the structure(s) of interest in the query image. The parts-and-geometry model allows us to do exactly this. We thus use the control points of our shape model as the potential feature detectors, and select a reliable subset of these to form the nodes of the parts-and-geometry model as described in the above subsections.

4. EVALUATION

We demonstrate our method by applying it to the segmentation of the subcortical structures in the brain. AAMs are well suited to this task because they incorporate information about both shape and texture. Some subcortical structures have overlapping gray level intensities and borders that do not have strong edges. Segmentation of subcortical structures is important in the investigation of a variety of neuropathological conditions.

Datasets
We had access to 275 T1-weighted MR brain images and manual labels of 10 subcortical structures (see Table 1). The MR images were provided by the Centre for Morphometric Analysis in Boston USA (see [15] for more details). The data was randomly split into a training set of 138 images and a test set of 137 images. The training set was used to build the parts-and-geometry model and the AAMs, and the test set was used for evaluation.

Processing pipeline
Figure 1 shows the segmentation pipeline used here. We have described steps 1 and 2 in detail above – AAM search is well known (see for example [1]), and an application to segmentatation of subcortical structures was described by Babalola *et al.* [2].

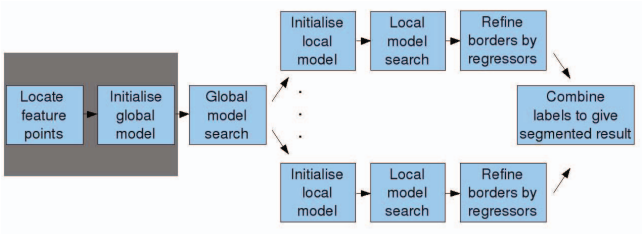

Fig. 1. The AAM pipeline for complete segmentation of an image. Our contribution is shown in the shaded region

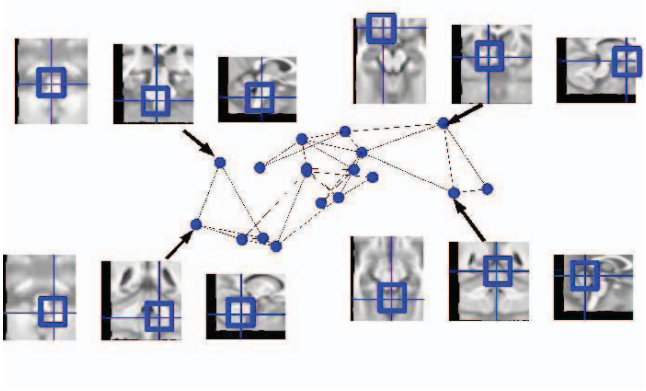

Fig. 2. Schematic illustration of the parts-and-geometry model showing pairwise connections of the parts, and axial, coronal and sagittal views of the regions used the local models in the reference image

Evaluating selected features
We established correspondence over the training set using the group-wise method of Cootes *et al.* [16] to perform non-rigid registration. The correspondences obtained here were used to construct the reference by warping each training image into the same space and taking the mean. The image pyramid of the reference was constructed from level 0, the original resolution, to level 5, $\frac{1}{16}$th original resolution. From the 1,063 nodes of the control points we obtained 12 consistently detected nodes of size 9x9x9 at level 2 of the image pyramid. These were used in the initialisation of the AAM. Although these nodes were consistently detected over the training set, taking the position that gave the highest NCC value did not always give the best candidate over the test set.

We quantify the accuracy of detection of feature points by ranking distances using D (Equation 1) and by the mean value of D. To quantify the performance of the detectors over the test set we registered the test set with the reference image. The correspondences obtained were used solely to calculate D. Using the geometric relationships from the model improved the mean ranking by D of the selected response from 1.07 ± 0.41 to 1.01 ± 0.10 (n=137) and the corresponding mean error between the expected position and the located position was reduced from $5.9 \pm 7.3mm$ to $5.6 \pm 4.8mm$.

Evaluating performance during segmentations
We run three segmentation experiments in which the global AAM (see Figure 1) is as follows.

- The global AAM is built from whole brain images, and initialisation is from the model mean
- The global AAM is built from the subregion containing the subcortical structures and initialisation is by using the best response (highest normalised correlation - NCC) of the local detectors
- The global AAM is as above, but initialisation is by using the candidates selected by the parts-and-geometry model

We quantitatively measure the performance of the segmentation using the Dice similarity metric [17] to compare the overlap of the AAM search result with that of the manually labelled version for each structure. Table 1 shows the Dice overlap values per structure for AAM search initiated in the three different ways, and Table 2

Structure	Whole brain AAM	Subcortical AAM best NCC	Subcortical AAM parts-geometry
Accumbens	70.1 (15.1)	41.1 (36.2)	72.5 (6.6)
Amygdala	71.1 (13.3)	43.7 (36.3)	74.3 (6.6)
Brain stem	89.2 (15.4)	64.8 (37.5)	91.8 (1.7)
Fourth ventricle	71.8 (16.23)	49.2 (34.7)	74.1 (10.5)
Caudate	86.4 (12.0)	52.3 (42.5)	88.1 (2.8)
Hippocampus	80.3 (13.5)	52.1 (38.7)	82.6 (3.2)
Ventricle	88.6 (11.4)	57.1 (41.5)	90.0 (3.4)
Pallidum	81.0 (12.7)	48.9 (40.2)	83.1 (3.1)
Putamen	88.7 (11.5)	54.2 (43.5)	90.3 (1.8)
Thalamus	89.1 (11.2)	58.0 (41.7)	90.6 (1.7)
All structures	81.8 (14.9)	51.6 (39.6)	83.8 (3.9)

Table 1. Mean (and sd) Dice values over structures for 137 images after AAM search using the following initialisation and global model combinations 1) AAM of the whole brain initialised from the mean, 2) subcortical AAM and feature points based on best NCC, 3) subcortical AAM and a parts-and-geometry model. Note the large standard deviations of the first two methods due to segmentation failures because of poor initialisation.

Method	$< 50\%$	$< 60\%$	$< 70\%$
Whole brain AAM	2.8%	4.5%	10.0%
Subcortical AAM & NCC	40.8%	42.1%	45.7%
Subcortical & parts geom	0.2%	1.8%	7.4%

Table 2. Percentage of results with Dice values under 50%, 60% and 70% for initialisation of AAM search by the different methods stated in column 1. The figures are percentages from all structures over the 137 test images

shows the percentage number of segmentation results that had Dice overlap values below 50%, 60%, and 70%.

5. DISCUSSION AND CONCLUSIONS

We have described a completely automatic method for selecting a set of feature detectors which can be used to initialise a 3D deformable model. This involves testing a large number of candidate detectors and selecting those which are most reliable. Our method is generic and avoids the need to manually select points in an arbitrary or otherwise fashion. We have demonstrated its use by applying it to the segmentation of the subcortical structures in brain images. It gives results on par with those of other leading methods (see e.g. [15],[2]).

Tables 1 and 2 shows that including the geometrical constraints improves the performance of the initialisation of the models in the subcortical region (without any noticable effect on speed). We believe the value of this work lies more in applications which require the location of subregions within a larger volume – such as the segmentation of the kidney in MR images of the torso. Although we focussed on the application to segmentation, the parts-and-geometry models described here can also be used to initialise registration.

In future work we will explore other methods of using combinatoric algorithms to include multiple candidates. We intend to explore the use of more sophisticated features such as 3D SIFT or those returned by 3D corner detectors. We also intend to investigate other methods of matching the local models other than normalised correlation.

Acknowledgements

We would like to thank David Kennedy and the Center for Morphometric Analysis, Boston, for providing the MR images used.

6. REFERENCES

[1] T F Cootes *et al.*, "Active appearance models," *IEEE PAMI*, vol. 23, no. 6, pp. 681–685, 2001.

[2] K O Babalola *et al.*, "3D brain segmentation using active appearance models and local regressors," in *Proc. MICCAI, LNCS 5241*, 2008, pp. 401–408.

[3] S C Mitchell *et al.*, "3-D active appearance models: Segmentation of cardiac MR and ultrasound images," *IEEE TMI*, vol. 21, no. 9, pp. 1167–1178, 2002.

[4] M B Stegmann and D Pedersen, "Bi-temporal 3D active appearance models with applications to unsupervised ejection fraction estimation," in *SPIE*, 2005, vol. 5747, pp. 336–350.

[5] P Viola and M Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001, vol. 1, p. 511.

[6] P F Felzenszwalb and D P Huttenlocher, "Pictorial structures for object recognition," *Int. Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[7] G J Edwards, T F Cootes, and C J Taylor, "Advances in active appearance models," in *Proc. ICCV*, 1999, pp. 137–142.

[8] H H Thodberg and A Rosholm, "Application of the Active Shape Model in a commercial medical device for bone densitometry," in *Proc. BMVC*, 2001, vol. 1, pp. 43–52.

[9] R Donner *et al.*, "Sparse MRF appearance models for fast anatomical structure localisation," in *Proc. BMVC*, 2007, vol. 2, pp. 1080–1089.

[10] S C Mitchell *et al.*, "Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images," *TMI*, vol. 20, no. 5, pp. 415–423, 2001.

[11] D Kainmueller *et al.*, "Automatic extraction of mandibular nerve and bone from cone-beam CT data," in *Proc. MICCAI, LNCS 5762*, 2009, pp. 76–83.

[12] T Heimann, H-P Meinzer, and I Wolf, "A statistical deformable model for the segmentation of liver CT volumes," in *3D Segmentation in the clinic: A grand challenge*, Workshop Proceedings, MICCAI 2007, Brisbane, pp. 161–166.

[13] B Patenaude *et al.*, "Bayesian shape and appearance models," *Tech. rep. TR07BP1, FMRIB - University of Oxford*, 2007.

[14] P F Felzenszwalb and D P Huttenlocher, "Representation and detection of deformable shapes," *IEEE PAMI*, vol. 27, no. 2, pp. 208–220, 2005.

[15] K O Babalola *et al.*, "An evaluation of four automatic methods of segmenting the subcortical structures in the brain," *Neuroimage*, vol. 47, no. 4, pp. 1435–1447, 2009.

[16] T F Cootes *et al.*, "Computing accurate correspondences across groups of images," *To appear in IEEE PAMI*, 2010.

[17] L R Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297–302, 1945.