Linking Sequences of Active Appearance Sub-Models via Constraints: an Application in Automated Vertebral Morphometry

M.G. Roberts, T.F. Cootes, J.E. Adams
Department of Imaging Science and Biomedical Engineering
University of Manchester
Manchester, M13 9PL, UK
martin.roberts@man.ac.uk
http://www.sys.uea.ac.uk/bmvc2003

Abstract

Statistical models of shape and appearance are powerful tools for interpreting medical and other images. However there can remain problems with under-trained models being too constrained. We have combined a global model with a sequence of partially overlapping sub-models, in a manner that exploits all the statistical information, whilst mitigating the under-training problem. Instead of applying one global model, we use a global model to apply iteratively-updated soft constraints on a sequence of sub-models. These sub-models may also partially overlap, and thus previously fit sub-models can also impose soft constraints on the next iteration. The algorithm has been applied to dual x-ray absorptiometry scans of the spine in order to automate vertebral morphometry measurements, using overlapping triplets of vertebrae as the sub-models, together with a global model of the entire spine. Combining a global model in this way with a sequence of sub-models gives substantially better results than using the former alone.

1 Introduction

1.1 Statistical models in medical imaging

Many problems in medical image interpretation require an automated system to interpret images. These images may provide noisy and possibly incomplete data, and typically deal with complex and variable structure. Model based methods offer solutions to these difficulties [13]. The great advantage of statistical shape models over more general deformable templates is that by applying simple constraints on the parameters, the generated shape always remains within physically reasonable bounds. This tends to lead to greater robustness in the presence of image noise. Furthermore by using a global shape model, correlations between different parts of the shape are implicitly modelled; thus even if portions of the shape are obscured by noise, their approximate location and shape can be inferred from other parts. In noisy images the use of every available shape constraint is helpful in improving robustness and accuracy [8].

Although the use of a global model provides helpful constraints that avoid unphysical solutions, there still remains the problem that as the model is based on a finite training set, it will fail to adequately fit unseen objects which deviate too far from the training set. Our approach steers a middle course between excessive local freedom lacking robustness in noisy/incomplete images, and training set inadequacies over-constraining a global statistical model. The approach decomposes the overall structure into a set of sub-parts, each of which has its own statistical model; but these partially overlap and thus constrain each other. The sub-models are fit in some sequence, which would normally correspond to their natural physical ordering. Furthermore an overall global model is also still used, to apply iteratively-updated soft constraints on the sequence of sub-models. Because these constraints are applied in the form of weights on a combined point and residual grey level fit, they are relatively "soft", and thus provide information without absolutely limiting the shapes that can be derived. Thus the totality of statistical information in the global model is used, but the coupling of the sub-models is less restrictive than if only a global model were used; hence training set inadequacies are mitigated. We have applied this approach to automated vertebral morphometry, which is clinically important in diagnosing osteoporosis.

1.2 Clinical background to vertebral morphometry

Osteoporosis is a progressive skeletal disease characterized by a reduction in bone mass, resulting in an increased risk of fractures, most commonly of the hip, vertebrae, and wrist. Of these hip fractures are the most serious, in terms of both morbidity and mortality [10]. Vertebral fractures are more common, and occur in younger patients. Several studies have demonstrated that the presence of vertebral fractures significantly increased the risk of further vertebral and non-vertebral fractures [2, 5]. The accurate identification of prevalent vertebral fractures is therefore important.

Vertebrae are box-like structures which form the spine, supporting the body's weight whilst allowing flexibility (see for example Figure 1). A vertebral fracture occurs when the inner structure of the vertebral body (made from spongy trabecular bone) has been weakened. A normal vertebra has an approximately rectangular appearance in lateral view, whereas a fractured vertebra tends to deform either by one of its ends collapsing (eg thus forming a wedge shape), or by appearing crushed in comparison to its neighbours.

Traditionally vertebral fractures had been detected on lateral spinal radiographs by a radiologist, who decided subjectively if any of the patient's vertebrae appeared fractured. There are a variety of methods, both quantitative and semi-quantitative for classifying vertebrae as normal or with various degrees and kinds of fractures [1, 4, 3]. Commonly used methods involve marking 6 points on each vertebra (essentially the 4 corners and two points in the middle of the top and bottom edges). From these points a posterior, a middle and an anterior height are derived along with their ratios. It is now common practice to use dual-energy absorptiometry (DXA) scans, or single-energy absorptiometry (SXA) scans for assessing vertebrae. These scans use a very low X-ray dose compared to conventional radiography, and also have the advantage of using a parallel beam to eliminate the projection effect of conventional radiography [7]. Commonly used scanners also provide software to calculate heights and ratios once the points have been marked on the digitised image, via a linked PC workstation.

Nevertheless the time involved in marking up scans remains significant, and further-

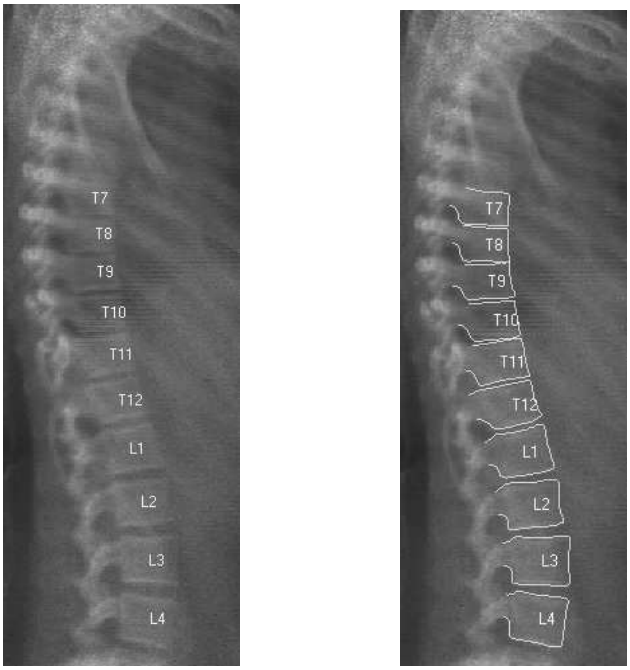

Figure 1: Example DXA Scan of normal spine

more using only 6 points may not give a complete enough description to reliably classify the early onset of osteoporosis, where fractures may be marginal and involve subtle shape changes. Smyth *et al* [9] have already demonstrated that an Active Shape Model (ASM, see below) can be used to accurately and precisely automate the measurement of normal vertebrae. The ASM covered the lumbar spine, and the thoracic spine from vertebra T12 up to T7 (see Figure 1). The vertebrae above T7 were not modelled, as the vertebrae of the upper thorax are frequently not well imaged in DXA scans [7]. Figure 1 shows a typical DXA image on the left with the superimposed model solution on the right.

2 Active Appearance and Active Shape Models

An Active Appearance Model (AAM) [13, 14] is a statistical model that describes both the shape of an object and the image texture around the shape. An AAM is created by training it with sample images on which the boundaries of the object and other important edges have been annotated. The AAM is an extension of an Active Shape Model (ASM) [13, 11]. In an ASM object shape is described by a point distribution model (PDM) which is generated by statistical analysis of the variation in landmark point coordinates in the annotated images. Firstly the training contours are aligned as closely as possible by means of scaling, rotation and translation. Then principal component analysis is performed on the residual deviations. Any new shape from the object's distribution can be linearly modelled as the sum of the mean shape plus a weighted combination of the most significant principal components.

An AAM [13, 14] typically uses a complete model of grey level texture, in which correlation between texture in different places is modelled. A PCA based model of the grey-level texture in the image around the shape is derived. The texture model PCA weights are combined with the shape model weights into a composite vector, and a further stage of PCA is conducted, to include correlations between shape and texture. The image search problem then becomes one of finding optimal parameter weights for the AAM principal components to best match the image. The AAM algorithm uses a fast linear update scheme to search, which is also learnt from the training set. Full details are given in [13, 14].

Smyth *et al* [9] located vertebrae using an ASM. We, however, use an AAM in this study. There are several reasons for this. Firstly the AAM uses a more complete model of the image grey level structure, and also models the covariance of that structure. It thus contains more information, and one would therefore expect that this would lead to more robust performance. Secondly Cootes has shown how to combine prior estimates of shape points with the AAM to perform Constrained AAM search [12], and the feeding forward of constraints is essential to our approach. Thirdly Scott *et al* have recently shown [6] that AAM reliability can be improved by using texture pre-processors, that apply the AAM approach to tuples of feature detector outputs, rather than raw pixel intensites. One of these detectors is a corner detector, which is of particular interest in vertebral morphometry, as it is the corner points that provide 4 of the 6 manually marked points. However this study only examines standard (pixel intensity) AAMs.

3 Methods

3.1 Vertebral Triplet Modelling

Smyth *et al* [9] showed that good performance could be obtained with an ASM using "miss-one-out" tests on normal vertebrae. Nevertheless if the aim were to automatically classify vertebrae as fractured or normal, the considerably greater range of variability in fractured vertebrae would pose a problem for the model. Not only can individual fractured vertebrae vary in a much wider manner than normal ones, but there may well be combinations of different fractured vertebrae not present in any one image in the training set. Thus the global model is likely to be over-constrained. We obtained the dataset of Smyth *et al* [9], and used it to build both a global model of the whole spine, and overlapping triplets of vertebrae. We applied our algorithm to these triplets. The vertebral triplet is a good compromise between the two extremes of individual vertebrae and the entire spine. Each vertebra (except the bottom and top, L4 and T7) is fitted using the sub-model in which it is central, and the two neighbouring vertebrae provide constraints and additional shape and texture information which assist the fit. L4 and T7 vertebrae are just fit as part of the triplet centred on their neighbours. It is intended to apply the approach to include also fractured vertebrae once a suitable dataset has been obtained, as this approach allows for fracture combinations that were not present in any one image in the training set. Although potential training set inadequacies might remain at the individual vertebra level, problems due to unseen fracture combinations should be mitigated. The dataset consisted of 78 lateral spine DXA images in women (mean age 61 years, age range 44-80), obtained from a Hologic QDR2000plus scanner. The dimensions of each pixel of the scan were 0.9 x 1.0 mm.

3.2 Algorithm

The sub-model sequence fit algorithm has been coded in a general framework described next, and in principle can be applied to any linked sequence of sub-models where there is some natural ordering of the shape sub-parts. It is assumed that there is typically some overlap between each sub-part, and thus the previous stage partly determines the starting solution for each sub-part's image fit. An initialisation of the global model (i.e. the entire shape) from a few user-defined landmarks provides further information for the sub-model starting solutions. At each sub-model fit iteration the current solution is *projected* (see next paragraph) from the global model into the sub-model, and then the sub-model fit is projected back into the global model. Each sub-model fit uses a constrained fit of an AAM (see [12]). Constraint weights are also adjusted as the search proceeds along the chain of sub-models, with higher constraint weights applied to points that have already been determined by a previous sub-model when fitting the next stage sub-model.

The projection between the global solution and each sub-model and back consists of the following. The current points and constraint weights are maintained in global vectors. However each sub-model in effect has its own local copy of only the subset of points and weights which it models. The specific application supplies to the algorithm the mapping between the global vector indices, and the sub-model indices. This mapping is actually maintained in 3 sub-maps corresponding to: the central portion of the sub-model's shape which it is primarily responsible for determining; the overlapping sub-portion determined by the previous iteration; and the overlapping sub-portion which will be finally determined by the next sub-model. In the vertebral application these 3 sub-portions simply correspond to individual vertebrae. For example when fitting the L3/L2/L1 triplet, this iteration determines the points of L2; using significantly weighted constraints applied to the points projected in from L3 (*previous* portion); and it additionally feeds forwards updated points and weights for L1 (*next* portion). All values are projected in, but the weights will be high for L3 (because of the previous iteration); moderate for L2; and low for L1. After the sub-model fit is complete only the points for L2 and L1 are projected back out into the global vectors; the weights for L2 are increased to high; and the weights for L1 increased to moderate. The latter L1 weights will in turn become high at the next iteration, when L1 is the central portion of the next sub-model.

In general a model may well be initialised by the user manually setting a small number of landmarks, e.g. by clicking with a mouse. For example in the vertebral morphometry application, the user initialises the search by clicking on 3 points corresponding to the mid-points of the bottom of L4, top of T12 and top of T7. This follows earlier work by Smyth *et al* [9]. The algorithm as such then proceeds as follows.

1. Fit the global shape model to the user input (or otherwise defaulted) landmarks as a starting solution. Note that this does not search the image. Initialise low weights on the predicted points (any higher weights on those landmark points input by the user are retained).

2. For each sub-model in the required processing order loop:[1]

[1]There are no particular rules on the ordering, which is implicitly supplied to the algorithm via the ordering of the map container holding the sub-models. The ordering used is a matter of specific-application strategy, but should be consistent with the previous/next projection sub-maps. The software also allows a reverse mode which swaps the meaning of previous and next in a sub-model's projection sub-maps.

(a) Project points and weights into the sub-model from the global vectors, including normally an overlap portion from a previous sub-model.

(b) Initialise a sub-model solution by performing a weighted least squares fit of the shape model to these projected point positions.

(c) Perform a weighted constrained AAM fit of this sub-model to the image.

(d) Project back the updated points into the global vectors. Note that the updated points comprise the central portion of the sub-model, and the next sub-model overlap.[2] However the subset of points primarily determined at the prior stage are not projected back again.[3]

(e) Increase the global weights for the central portion to high, and those of the next overlap portion to moderate.

(f) Re-perform a weighted least squares fit of the global shape model to the global points vector, thus making the globally modelled shape as consistent as possible with the sub-model solutions so far. Copy out from the global model any as yet unfitted (by a sub-model) points into the global points vector. This step thus adjusts some of the initial points of the next sub-model for the next iteration. Thus maximal feed-forward of information is maintained.

3. End loop

Although our specific application used triplets of vertebrae, the algorithm really deals with a generalised "triplet", whose three sub-parts are defined purely by the domains of the 3 projection sub-maps for: points previously determined; points to be determined now; and further overlap points to be fed forward as a soft constraint.

3.3 Experiments

The 78 images were split into 3 groups of 26. "Leave-one-group-out" experiments were then performed (i.e. 52 training images and 26 test images, repeated thrice). On each experiment the 3-point user initialisation was emulated by using the known equivalent marked points and adding random offsets to them. These were zero-mean Gaussian errors with SD of 1 pixel in the y-direction (along the spine) and 3 pixels in the x-direction. The figure of 1 pixel in y corresponds to other published manual precision figures of around 1mm [9, 7]. As the mid-points are used for initialisation, and there is a greater degree of arbitrariness in the lateral position of this than for corners, an SD of around 3 times as much in lateral precision seemed reasonable. Previous work by Smyth *et al* [9, 8] has in any case demonstrated high precision for ASMs, so it is not expected that the exact form of initial randomisation will have much effect.

[2] Additional checks are made to ensure that overlap soft constraint points are fed forward at most once, and in particular these can never overwrite the assignments from a prior iteration in which these points are in the central "to be determined now" subset.

[3] But in the special case of the first iteration any points defined as prior are in fact copied out.

Single Model				Sub-Model Sequence		
Vertebra ID	Mean Acc	SD Acc	%ge Failure	Mean Acc	SD Acc	%ge Failure
T7	1.53	1.08	22.10	0.97	0.65	1.23
T8	1.50	0.90	12.31	0.90	0.52	0.00
T9	1.37	0.70	4.94	0.86	0.57	1.19
T10	1.13	0.71	6.14	0.73	0.31	4.85
T11	1.08	0.73	8.61	0.78	0.27	4.85
T12	1.33	0.83	9.79	0.93	0.31	6.09
L1	1.33	0.68	8.64	0.92	0.46	7.32
L2	1.18	0.52	8.65	0.82	0.24	8.56
L3	1.22	0.62	9.88	0.89	0.51	2.47
L4	1.42	0.70	13.59	1.07	0.45	3.66
Overall	1.28	0.74	9.17	0.88	0.43	4.33

Table 1: Global Model Search Error and Failure Rate (pixels, approximately equivalent to mm)

4 Results

The accuracy and robustness of the search was characterised by calculating the mean absolute point-to-line distance error for each vertebra. The error is the distance from each point on the located vertebra contour to the nearest point on the "true" contour (i.e. as annotated during training). Then the mean and SD of the errors were calculated over all the images in which the search was successful. However if on a particular image, the mean error for a vertebra was less than 5mm, then this was classed as a search failure (for that vertebra). The figure of 5 mm corresponds to around 25% of a typical vertebral height and would be likely to cause mis-classification between normal and fractured vertebrae (a grade 1 fracture can be classed as a 20% reduction in height [4]). Table 1 compares the results of using the single global model in a standard AAM search to the sub-model sequence approach, whilst table 2 shows the result of using a standard AAM search on a somewhat larger training set using a "miss-one-out" train/test cycle.

The order in which sub-models are fit will have some effect on the final solution. We experimented with a variety of orderings, and found that the best sequence was to start from the lowest vertebral triplet (L4/L3/L2) and proceed up the lumbar and into the mid-thorax as far as the triplet centred on T11. However because the vertebra of the upper thorax are frequently poorly imaged, it is better to then proceed to fit T7 and T8, then alternate back to T10 concluding with the now heavily constrained T9. This is because a large constraint weight is applied to the user's initialisation point on the top of T7. Applying this constraint, and using this user-supplied information relatively early in the fit of the thorax, tends to increase overall reliability, and substantially improves performance on the difficult T7 and T8. This lessens the chance that the search drifts off into clutter from the ribs and lungs around T10/T9, which would of-course then mal-position the starting solution for subsequent vertebrae. This point also emphasizes that our approach can naturally lend itself to application-specific strategies for fitting the best parts of the image first, or emphasizing user-supplied information in the fit sequence.

Vertebra ID	Mean Acc	%ge Failure
T7	1.82	10.26
Rest of spine	1.21	4.23

Table 2: Global Model Search Error (pixels, approximately equivalent to mm) and Failure Rate on miss-one-out train-test cycle for Standard AAM

5 Discussion

The performance of the sub-model sequence algorithm appears far better at T7, reducing a poor 25% failure rate to 1.23%. Over the rest of the spine there appears to be a reduction in failures from 9.17% to 4.33%. With the current sample size this is not actually statistically significant, although the fact that the sub-model approach performs better on every single vertebra offers some further support that the apparent improvement is real. The mean point-to-line distance error is reduced on average from 1.28 to 0.88, (excluding T7 because of its high failure rate with the global model). The reduction of 0.4mm is significant assuming the non-T7 data can be pooled, and corresponds to over 4 standard errors. The variance of the accuracy is also significantly lower for the sub-model sequence approach.

Table 2 shows that the accuracy attained with a single global model using a somewhat larger (77 image) training set is not significantly better than in the former case of a 52 image training set. The search failure appears to be less on a maximally trained model, but again the difference is not statistically significant, except at T7 which is marginally so (at the 5% level assuming Binomial convergence to Gaussian has been attained). However the results at least suggest that the smaller training set might be causing some additional failures. These miss-one-out results for a standard AAM can be compared to results published by Smyth for an ASM applied to the same dataset [8]. On the whole the simpler ASM appeared to perform better on this data than a standard AAM, especially at T7. The comparable results for the ASM were a mean error of 1.39 pixels at T7 and 1.04 pixels over the rest of the spine. There are no variance figures published for the ASM results so it is not possible to assess the statistical significance of the difference. Cootes *et al* [13, 15] concluded that although an AAM may be more robust than an ASM, nevertheless an ASM tends to have a larger capture range if started from a poorly initialised solution. This is because an ASM searches around the current location, whereas the AAM only examines the image directly under its current area. Our results may be indicating this also, but our new algorithm still gives better results than the ASM.

Figure 2 shows an example of an image where the spine is more curved than is typical in the training set, and which also contains a vertebra on the margins of a grade 1 wedge deformity (L1). The overall global model is too constrained to fit to this image, and the solution can be seen to be poor. However the sub-model sequence, by exploiting the looser coupling of the sub-models, and their independent pose, still provides a good fit.

In conclusion, the results show that combining a global model in this way with a sequence of sub-models gives better results than using the former alone. We have generalised from the idea of a triplet of vertebrae to the general threesome of: previously determined points; points to be determined now; and further overlapping points to be fed

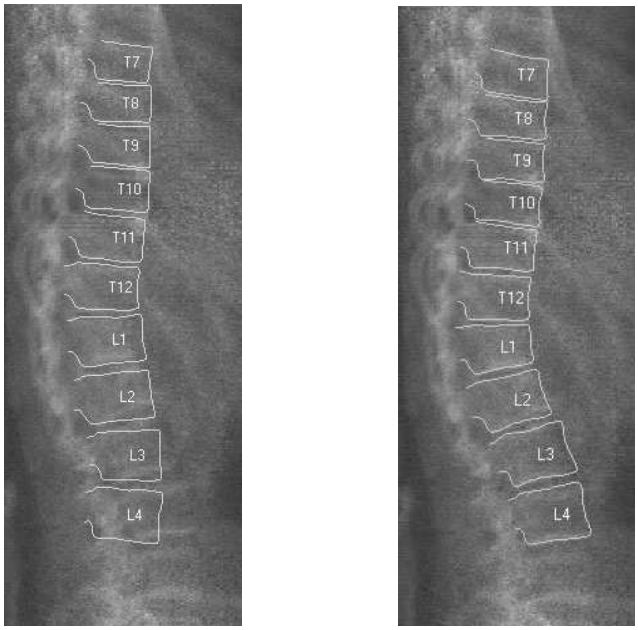

Figure 2: Example of global model fit failure (left) and sub-model fit success (right)

forward as soft constraints. We believe that the combination of this conceptual triplet with the use of further (iteratively updated) soft-constraints from a global model, provides a powerful framework for exploiting all the information in statistical shape models, whilst mitigating fit-problems due to under-training of the model. The framework also naturally lends itself to semi-automatic fitting, where a user can manually adjust part of the fit in noisy or atypical regions. Then large weights can be attached to the user-adjusted point(s), and the primarily affected sub-models only need be refit. It is intended to continue with this approach in the vertebral morphometry application by including fractured vertebrae, which have a much greater range of variability than normal vertebrae. There is also scope for applying this approach to conventional radiological scans, where X-ray fan beam distortion means that there is a variable dimensional scaling across the image. As the sub-models are allowed to have independent pose, the variable scale could be incorporated.

References

[1] Guermazi A, Mohr A, Grigorian M, Taouli B, and Genant HK. Identification of vertebral fractures in osteoporosis. *Seminars in Musculoskeletal Radiology*, 6(3):241–252, 2002.

[2] Black DM, Arden NK, Palermo L, Pearson J, and Cummings SR. Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures. *Study of Osteoporotic Fractures Research Group, J Bone Miner Res*, 14:821–828, 1999.

[3] McCloskey E, Spector T, Eyres K, and *et al*. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporosis Int*, 3:138–147, 1993.

[4] Genant HK, Wu CY, van Kuijk C, and Nevitt MC. Vertebral fracture assessment using a semi-quantitative technique. *J Bone Miner Res*, 8:1137–1148, 1993.

[5] Melton LJ III, Atkinson EJ, Cooper C, O'Fallon WM, and Riggs BL. Vertebral fractures predict subsequent fractures. *Osteoporosis Int*, 10:214–221, 1999.

[6] Scott IM, Cootes TF, and Taylor CJ. Improving active appearance model matching using local image structure. In *18th Conference on Information Processing in Medical Imaging*, page To Be Published, 2003.

[7] Rea JA, Li J, Blake GM, and et al. Visual assessment of vertebral deformity by x-ray absorptiometry : a highly predictive method to exclude vertebral deformity. *Osteoporosis Int*, 11:660–668, 2000.

[8] Smyth PP. *Measurement of osteoporosis using computer vision,PhD Thesis*. Department of Medical Biophysics, University of Manchester, 1997.

[9] Smyth PP, Taylor CJ, and Adams JE. Vertebral shape: automatic measurement with active shape models. *Radiology*, 211:571–578, 1999.

[10] Cummings SR, Kelsey JL, Nevitt MC, and O'Dowd KJ. The epidemiology of osteoporosis and osteporotic fractures. *Epidemiol Rev*, 7:178–208, 1985.

[11] Cootes TF, Hill A, Taylor CJ, and Haslam J. The use of active shape models for locating structures in medical images. *Image Vision Comput*, 6:276–285, 1994.

[12] Cootes TF and Taylor CJ. Constrained active appearance models. In *8th International Conference on Computer Vision*, volume 1, pages 748–754. IEEE Computer Society Press, July 2001.

[13] Cootes TF and Taylor CJ. Statistical models of appearance for medical image analysis and computer vision. *Proc SPIE Medical Imaging*, 3:138–147, 2001.

[14] Cootes TF, Edwards GJ, and Taylor CJ. Active appearance models. In Burkhardt H and Neumann B, editors, *5th European Conference on Computer Vision*, volume 2, pages 484–498. Springer (Berlin), 1998.

[15] Cootes TF, Edwards GJ, and Taylor CJ. Comparing active shape models with active appearance models. In T. Pridmore and D. Elliman, editors, *10th British Machine Vision Conference*, volume 1, pages 173–182, Nottingham UK, 1999. BMVA Press.