

Robust Active Appearance Models with Iteratively Rescaled Kernels

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@manchester.ac.uk

Abstract

Active appearance models (AAMs) are widely used to fit statistical models of shape and appearance to images, and have applications in segmentation, tracking, and classification of structures. A limitation of AAMs is that they are not robust to a large set of gross outliers. Using a robust kernel can help, but there are potential problems in determining the correct kernel scaling parameters. We describe a method of learning two sets of scaling parameters during AAM training: a coarse and a fine scale set. Our algorithm initially applies the coarse scale and then uses a form of deterministic annealing to reduce to the fine outlier rejection scaling as the AAM converges. The algorithm was assessed on two large datasets consisting of a set of faces, and a medical dataset of images of the spine. A significant improvement in accuracy and robustness was observed in cases which were difficult for a standard AAM.

1 Introduction

Active Appearance Models (AAMs) [2] are statistical models of object shape and appearance which have been used to represent and locate objects in a wide range of domains. In their basic form they match a model to a new image using an *analysis-by-synthesis* approach, in which a synthetic image is generated from the model, and then compared to the target image. The model parameters are optimised to synthesise an image as close as possible to the target. One of the strengths of AAMs is that by using a sum-of-squares measure to compare model and target, they can exploit the linear nature of the problem to perform fast parameter updates and thus are able to match to a new image very quickly.

The constraints of the underlying appearance model make AAMs reasonably robust against noise. However, sum-of-squares terms can be badly affected by large outliers. Image regions where the intensities are far from the predicted model values can have a significant deleterious effect, and may lead to poor overall matches or divergence. For example, a model of faces trained on people without beards or glasses may fail badly on images of people with beards or glasses; or on an image where part of the face is occluded. In the medical context there may be a surgical implant in part of the image. A natural method of mitigating such problems is to introduce methods from robust statistics [7]. Typically these seek to downweight larger residuals. However some scaling parameter

is then required to define which residuals are 'large'. One approach to choosing such a parameter is to use estimates of the expected variance of the residuals when matching the model to the training set. However, these tend to be underestimates, and are certainly not appropriate during early stages of search when the model may start far from the true position, leading to large initial residuals.

In this paper we introduce a novel approach in which we use a large initial estimate of the scaling, and gradually reduce the scale as the search algorithm proceeds. So as the image match becomes better, the algorithm becomes stricter as to what it considers inliers. We describe the technique, and give the results of experiments on two different datasets (facial photographs and a medical dataset), demonstrating that the approach leads to improved performance when matching to 'difficult' examples.

2 Background

2.1 Active Appearance Models

Statistical appearance models [2] are linear models of object shape and texture variation, which are learned from a training set. An appearance model is constructed by first learning the statistics of shape and texture variation, and then incorporating both into a set of combined *appearance* parameters \mathbf{c} . These represent the shape, \mathbf{x} , and texture, \mathbf{g} , (in the model frame) according to

$$\begin{aligned}\mathbf{x} &= \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \\ \mathbf{g} &= \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}\end{aligned}\tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape, $\bar{\mathbf{g}}$ the mean texture in a mean shaped patch and $\mathbf{Q}_s, \mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set.

A shape in the image frame, \mathbf{X} , can be generated by applying a suitable global transformation (such as a similarity transformation) to the points in the model frame. The texture in the image frame is generated by applying a scaling and offset to the intensities generated in the model frame. These shape pose and texture scaling parameters are concatenated with the appearance model parameters \mathbf{c} to give the full set of AAM parameters, denoted by \mathbf{p} .

The AAM seeks to minimise a sum-of-squares problem of the form

$$F(\mathbf{p}) = |\mathbf{r}(\mathbf{p})|^2 = \mathbf{r}^T \mathbf{r}\tag{2}$$

where the vector of residuals \mathbf{r} is calculated as:

$$\mathbf{r} = \mathbf{w}(I : \mathbf{p}) - \mathbf{g}\tag{3}$$

where $\mathbf{w}(I : \mathbf{p})$ is the (normalised) texture sampled from the image I given the shape defined by model parameters \mathbf{p} . The AAM is trained (see [2]) how to solve this minimisation by learning the relationship between perturbations in parameters and the texture residuals that these induce. This relationship is then inverted to provide an update matrix that can be applied to the current residual vector. When applied iteratively this provides an efficient model fitting scheme where the necessary parameter change $\delta\mathbf{p}$ to the parameter vector \mathbf{p} , given \mathbf{r} is estimated as:

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}\tag{4}$$

where \mathbf{R} is derived from the Jacobian $\mathbf{J} = \frac{\delta \mathbf{r}}{\delta \mathbf{p}}$ thus.

$$\mathbf{R} = [\mathbf{J}^T \mathbf{J}]^{-1} \mathbf{J}^T \quad (5)$$

2.2 Robust Extensions of AAMs

Edwards et al [4] derived an exclusion threshold for each pixel based on the distribution in the training set during AAM matching. Residuals beyond their threshold are ignored in the AAM update step. Stegmann et al [17] extended this by introducing a robust measure for the comparison on AAM solutions, which changes the convergence criterion. However the AAM update steps are still performed as in [4] during the initial search. There are two potential problems with this. Firstly rejecting certain pixels early on can ignore useful information, and may even lead to convergence failure - for example if there is a large initial pose error. Secondly the AAM update matrix with a robust kernel objective function should ideally be changed. In [17] this problem is partly avoided by the use of a simulated annealing phase after conventional AAM search. Gross et al [15] propose a modification that solves this update matrix problem by using a variable Hessian that can be efficiently computed assuming a degree of spatial coherence. Theobald et al [1] experimented with using this method with a variety of kernels. However less attention was given to the problem of determining kernel scaling parameters, which is particularly important in problems where the AAM may not have a good initialisation.

A different approach was adopted by Beichel et al [5]. Their robust AAM identifies outliers by analysing the multi-modal nature of the residual histogram. This method has the advantage of avoiding the direct use of cutoff thresholds - which might be set too low for a particular target image - but at the price of some complexity in the outlier identification method.

Our approach is to recast the objective function as a weighted sum of squares in a way that allows us to apply the standard AAM update to weighted residuals. The scale parameters used for the weighting start off at a coarse scale but are gradually reduced as the AAM converges.

3 Robust Kernel Application and Scaling

We adopt an M-estimator approach adapted to the AAM framework. Given a robust kernel form ρ , the aim is to minimise the objective function:

$$S = \sum_{k=1}^n \rho(r_k, \sigma_k) \quad (6)$$

We have used the Geman-McClure [16] kernel for a residual error r_k at pixel k .

$$\rho(r_k, \sigma_k) = \frac{r_k^2}{r_k^2 + \sigma_k^2}, \quad (7)$$

where σ_k is a scale parameter that controls the convexity of the robust function.

This kernel has some history of use in computer vision, including learning and fitting eigenspace models [9, 13], and in robustly estimating optical flow [14]. The

Geman-McClure robust function is twice-differentiable, and although its influence function (derivative) tends to zero for large errors, there is no absolute cut-off point. The kernel response is always less than the L_2 -norm, and the influence is approximately linear over the inlier region $[-\frac{\sigma}{\sqrt{3}}, \frac{\sigma}{\sqrt{3}}]$, after which it declines.

The scaling of the GM kernel is typically some multiple β of the (inlier) standard deviation. The influence of inliers should typically extend to at least one standard deviation, whereas the GM function reduces a residual's influence after $\frac{\sigma}{\sqrt{3}}$. Also because training set inadequacies typically underestimate the true residuals encountered in unseen images we extend the inlier region by a further 50%, and use: $\beta = 1.5\sqrt{3}$.

We note that the conventional AAM is trained to minimise the residual sum of squares - the residuals are not normalised by their variance. Although this is not a maximum likelihood estimator, in practise it appears to work well because the image patch typically includes large numbers of uninteresting pixels with low variance but little structure. So we modify the objective function to make it more like a conventional AAM and optimise:

$$S' = \sum_{k=1}^n \sigma_k^2 \rho(r_k, \sigma_k) \quad (8)$$

This form of sum of robust error functions $\rho_k(r_k)$ can be recast as an Iteratively Weighted Least Squares problem:

$$\tilde{S} = \sum_{k=1}^n \sigma_k^2 w_k r_k^2 \quad w_k = \frac{1}{r_k} \frac{d\rho_k}{dr_k} \quad (9)$$

with weights determined by the kernel influence, since at the true minimum these problems have the same solution, as can readily be seen by differentiation (and regarding the weights as fixed for the current iteration). We use the AAM update matrix as a quick though approximate solver, applying it to the "kernel-ised" residuals \mathbf{r}' , with:

$$r_k' = \sqrt{w_k} \gamma_k r_k \quad (10)$$

where γ_k is a normalisation constant chosen so that $r_k' = r_k$ for a residual at a suitable small proportion ε ¹ of the residual standard deviation.

This leads to the following mapping to apply to the residual vector, which we denote by the vector function $\mathbf{K}(\mathbf{r}, \sigma)$

$$r_k' = [\mathbf{K}(\mathbf{r}, \sigma)]_k = \gamma_k \frac{r_k}{r_k^2 + \sigma_k^2} \quad \gamma_k = \frac{\varepsilon^2 + \beta^2}{\beta^2} \sigma_k^2 \quad (11)$$

Essentially our robust AAM treats an outlier residual as though it were a smaller residual, which leads to a reduced update step to the model parameters (due to this pixel). However because the implicit Jacobian being used is still $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ and not $\frac{\partial \mathbf{r}'}{\partial \mathbf{p}}$, the estimate of the required update step becomes biased for large outliers. The situation is complicated by having varying biases due to many residuals spread over many parameters. However if we consider the limiting case of the contribution of a specific residual r_k to a single parameter model, we would see by comparing the two Jacobians $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ and $\frac{\partial \mathbf{r}'}{\partial \mathbf{p}}$ that

$$\hat{\delta} p_k = \sqrt{w_k} \delta p_k^{(0)} \quad (12)$$

¹We used $\varepsilon = 0.1$.

where $\delta\hat{p}_k$ is the update step contribution we use, whereas $\delta p_k^{(0)}$ is the one really required in minimising \tilde{S} in 9. Thus we tend to underestimate the step for outliers (where the weights are small) and hence bias the step direction in favour of the inliers (which have weights near unity). The bias is mitigated by the fact that $\sqrt{w_k}$ tends to zero more slowly than w_k itself. This bias would be worse for rapidly decaying kernels, but the Geman-McClure kernel we use is relatively benign to outliers, and a degree of bias towards inliers may be acceptable. We also reduce the effect of bias by gradually reducing the scale as discussed below.

So although the AAM update matrix should ideally be adapted to the weights, the standard AAM update matrix still gives an approximate solution in the inlier region where the weights are close to unity. Nevertheless the underestimate of the update step can cause convergence failures if starting from too far away from the solution, because reducing a large residual can cause the weighting to increase in a way that makes the overall objective function increase (the weight increases faster than the residual squared decreases). In fact such local minima problems can occur in any case with other forms of Iterated Weighted Least Squares. There is always the problem of how to select correct scalings for the $\{\sigma_k\}$ so that outliers are rejected without causing convergence failures on genuine data when the AAM is initialised some way off the target shape.

Our proposed solution is to apply a form of deterministic annealing similar to that used in [9, 13] (which used quasi-Newton non-linear optimisation). We calculate two sets of scaling parameters: $\{\tilde{\sigma}_{Ik}\}$ and $\{\sigma_{Fk}\}$ for the initial and final scale parameters respectively. In the initial AAM updates the scaling is such that the vast majority of residuals are inliers, and so the AAM update equation can validly be used to find an approximate solution to the weighted least squares problem. Then the scale is gradually reduced. The initial scale parameters are learned during the standard (kernel-free) AAM training. We learn the Jacobian of the induced residual vector \mathbf{r} as parameters \mathbf{p} are displaced from their correct values. This Jacobian is used to derive the initial set $\{\tilde{\sigma}_{Ik}\}$ from the modelled parameter variances. However if the AAM has really converged successfully then the final residuals should be substantially lower, and the final residual variances should be approximated by those of the underlying appearance model. These can be learnt by refitting the appearance model back to its own training set². We denote the underlying appearance model ideal standard deviations as $\sigma^{(0)}$. Our approach reduces the kernel scaling parameters from one set to the other. AAMs are typically used with coarse-to-fine search using Gaussian image pyramids, and this reduction is applied at each stage of the image pyramid before moving onto the next.

In [6], the median absolute deviation (MAD) over the entire image dataset was used as a lower bound on each pixel's scale parameter. On the assumption that the training set represents outlier-free data with zero-mean uncorrelated Gaussian distributed residuals,

the MAD over all pixels has an expected value of: $MAD(\sigma) = \frac{\sqrt{\sum_{k=1}^n \sigma_k^2}}{1.4826\sqrt{n}}$ where n is the number of pixels.

We obtain estimates of the initial residual standard deviations by summing over the m_p AAM parameters and using the Jacobian \mathbf{J} and the AAM parameter variances $\{\sigma_{j_p}^2\}$. For the appearance model parameters \mathbf{c} these variances are known, and determined via the eigenvalues of the appearance model modes. For the pose parameters (e.g. similarity transform from model to image reference frame) we assume the associated parameter

²or preferably a jackknifed miss-1-out estimate can be derived

standard deviation is the positive displacement range used in AAM training (i.e. when deriving the estimate of \mathbf{J}). Then

$$\sigma_{Ik} = \sqrt{\sum_{j_p=1}^{m_p} J_{kj_p}^2 \sigma_{j_p}^2} \quad (13)$$

Applying the MAD lower bound and rescaling by β gives:

$$\tilde{\sigma}_{Ik} = \beta \max(\sigma_{Ik}, \hat{MAD}(\sigma_{\mathbf{I}})) \quad (14)$$

Whereas for the converged kernel parameters:

$$\sigma_{Fk} = \beta \max(\sigma_k^{(0)}, \hat{MAD}(\sigma^{(0)})) \quad (15)$$

We now summarise the application of the kernel and its rescaling more formally.

For each image pyramid level loop:

- 1 Set $\sigma = \tilde{\sigma}_l$
- 2 While the AAM update step improves the solution loop:
 - 2.1 Calculate the raw residual error vector $\mathbf{r}(\mathbf{p})$
 - 2.2 Calculate $\mathbf{r}'(\mathbf{p}) = \mathbf{K}(\mathbf{r}, \sigma)$
 - 2.3 Compute current kernel-ised error norm $E = \|\mathbf{r}'\|^2$
 - 2.4 Set $\Delta = 1$
 - 2.5 Update the parameter vector $\tilde{\mathbf{p}} = \mathbf{p} - \Delta \mathbf{R}\mathbf{r}'$
 - 2.6 Calculate a new error vector $\tilde{\mathbf{r}}$ using the updated parameters $\tilde{\mathbf{p}}$
 - 2.7 Compute the updated kernel-ised error vector $\tilde{\mathbf{r}}' = \mathbf{K}(\tilde{\mathbf{r}}, \sigma)$
 - 2.8 If $\|\tilde{\mathbf{r}}'\|^2 < E$, then accept the new parameters: $\mathbf{p} = \tilde{\mathbf{p}}$
 - 2.9 Else retry for $\Delta = 1.5, \Delta = 0.5, \Delta = 0.25$ etc and go to step 2.5; or exit this loop when no more improvement results
 - 2.10 End loop
- 3 Calculate the median reduction ratio μ of $\{\frac{\sigma_{Fk}}{\tilde{\sigma}_{Ik}}\}$
- 4 Given annealing decay constant α (e.g. $\alpha = 0.9$), set the number of kernel rescaling steps $N = \left\lceil \frac{\log(\mu)}{\log(\alpha)} \right\rceil$
- 5 for $N + 1$ rescalings do
 - 5.1 if final rescaling set $\sigma = \sigma_{\mathbf{F}}$
 - 5.2 else: for each k: $\sigma_k = \max(\sigma_{Fk}, \alpha \sigma_k)$
 - 5.3 Set $M = 4$ (or other suitable update count)
 - 5.4 Set $\Delta = \frac{1}{M}$
 - 5.5 for M forced updates do
 - 5.5.1 Recalculate the raw residual error vector $\mathbf{r}(\mathbf{p})$
 - 5.5.2 Calculate $\mathbf{r}'(\mathbf{p}) = \mathbf{K}(\mathbf{r}, \sigma)$
 - 5.5.3 Update the parameter vector $\tilde{\mathbf{p}} = \mathbf{p} - \Delta \mathbf{R}\mathbf{r}'$
 - 5.5.4 $\mathbf{p} = \tilde{\mathbf{p}}$
 - 5.6 end loop

- 6 end loop
 - 7 Repeat AAM update loop (2) at reduced scale now $\sigma = \tilde{\sigma}_F$
 - 8 Move to next image level
- End loop over image pyramid hierarchy

4 Experiments

We used a standard set of face images [10], of size 720x576 pixels. An AAM was trained using a subset of the data containing 162 subjects free of facial hair (beards) and glasses. The form of appearance model used was a feature AAM using sigmoidally renormalised corner and edge features combined with similarly renormalised cartesian gradients. See [8] for details. This has been shown to generally outperform a standard texture AAM. The renormalisation to the local image feature statistics also means that there are less likely to be problems with the kernel scaling being inappropriate to a particular image.

The test set used was the remainder of 136 subjects having either facial hair or wearing glasses. Thus the appearance model would be expected to be unable to accurately fit to significant portions of the test cases. This will typically cause a number of overall convergence failures. We tested running the AAM with and without the robust kernel, to see if the robust kernel improves convergence reliability in cases which significantly depart from the trained model. There were 3 images used for each subject, giving a total test set size of 408. The AAM was fitted to each test case using a fixed grid of initial position and scale offsets. The initial pose parameters were set by performing an initial best fit of the shape model to the (known) test case points. Then the appearance parameters were reset to zero, and the x,y and scale pose parameters varied as follows. The x and y values were stepped respectively between [-20,20] pixels with a step-size of 5 pixels; the scale was stepped between 1/1.05,1.0,1.05 respectively. This gives a total intialisation grid of $9 \times 9 \times 3 = 243$ searches for each test image.

For the medical images miss-8-out experiments were performed on a dataset of 308 DXA images of the spine. These used 10 initialisations for each image, based on random perturbations on the mean shapes implied by the vertebral centres. See [12] for details. DXA images use a very low X-ray dose, which makes them quite noisy. The aim was to locate the positions of the 10 vertebrae from lumbar L4 to thoracic T7. The segmentation algorithm uses multiple AAMs, each of which fits a triplet of vertebrae. These are overlapped, which establishes linkage between the various sub-models. This gives good accuracy on healthy vertebrae [12]. However the main aim of the DXA imaging is in the diagnosis of osteoporosis, as the the weakening of bone leads to fracture of vertebrae as one of the early symptoms of the disease. Accuracy results on severely fractured vertebrae are much poorer, partly because the AAM is prone to miss the fractured vertebrae and instead fit to neighbouring vertebrae. For example the top of a fractured T8 may be fitted to the bottom of the neighbouring T7, which provides a locally similar edge. See Fig 1c for an example. In [11] we used an alternative fractured AAM initialisation which partly resolved this problem. Also the low bone density in fractured vertebrae can give them a very poor signal to noise ratio, which means that noise outliers have a more significant overall effect. We re-ran the experiments of [11] to see if the robust kernel might improve the accuracy on these difficult fractured vertebrae.

Error Measure	Standard AAM			Robust Kernel AAM		
	Mean	Median	90%-ile	Mean	Median	90%-ile
Point-to-Curve	4.93	3.36	10.54	4.64	3.28	8.86
CoG X	1.95	1.05	4.63	1.48	0.94	2.96
CoG Y	3.79	1.22	12.55	3.17	1.08	9.51

Table 1: Search accuracy results (pixels) for the face data. CoG X and CoG Y are the absolute errors in the centre of gravity of the face

Fracture Type	Standard AAM				Robust Kernel AAM			
	Mean (mm)	Median (mm)	90%-ile (mm)	%ge Edges Misfit	Mean (mm)	Median (mm)	90%-ile (mm)	%ge Edges Misfit
Normal	0.73	0.55	1.57	0.20%	0.78	0.59	1.62	0.20%
Grade 1	0.83	0.57	1.77	4.07%	0.83	0.60	1.79	1.61%
Grade 2	1.01	0.67	2.34	4.37%	0.98	0.67	2.26	1.96%
Grade 3	1.65	0.88	4.52	16.64%	1.28	0.78	3.05	8.42%

Table 2: Search accuracy results by vertebral fracture status for spine DXA images. NB Fracture grades 1-3 correspond to height reduction thresholds of 20%, 25% and 40% respectively. Errors are in mm but the %ge edges misfit columns give the percentage ratio of superior or inferior edges fitted erroneously onto a neighbouring vertebra

5 Results

For the face experiments we calculated the mean, median and 90%-ile for: a) mean absolute point-to-curve error over each face image; b) face centre of gravity x and y positions. For the vertebral data we similarly calculated the mean, median and 90%-ile on raw point to curve errors, and also enumerated the proportion of vertebral edges that were erroneously fitted to a neighbouring vertebra. These were analysed according to their classification as normal (non-osteoporotic), or one of 3 fracture severity grades according to a standard system used by radiologists. Results for the face and vertebral datasets are summarised in tables 1 and 2 respectively.

6 Discussion and Conclusions

In the face experiments a statistically significant improvement occurs in mean point-to-curve error, and also in the mean error in the cartesian coordinates of the face centre. Moreover the tail of the distribution appears to be reduced by the use of the robust kernel. The 90th percentile is reduced from 10.54 pixels to 8.86. The median error is also marginally lower, indicating that the improvement in mean error is not at the expense of a general deterioration in the better fitting cases. There is a larger improvement in the accuracy of the estimate of the centre of the face, both in terms of mean and the 90th percentile tail. Overall in the face data there is a modest but significant improvement in accuracy. An example is shown in Figure 1, which shows a partial search failure without the use of the kernel. The AAM appears to have been confused by the facial hair and does not correctly locate the mouth and chin, whereas in the corresponding case with the



Figure 1: Examples. The left (Fig 1 a) and middle (Fig 1b) figures show an example from the face data using a standard AAM (leftmost) and with the kernel AAM (middle). The facial hair appears to confuse the non-robust AAM (left) as to the location of the mouth and chin, whereas with the robust kernel (middle) this problem is avoided. The right-hand figure (Fig 1c) shows an edge misfit (standard non-robust AAM) on a vertebra for a grade 3 fracture. The fitted standard 6-point morphometric corner and mid-points are shown. The arrows indicate where the AAM search fitted the top of T8 to the bottom of T7, and the anterior bottom corner of T8 to the top of T9.

kernel applied the solution has converged with reasonable overall accuracy.

In the medical image data, there was no improvement with normal vertebra. In fact there was a mild deterioration when applying the kernel, from 0.73mm to 0.78mm, though this difference is on the margins of statistical significance, and is far smaller than the precision of the underlying manual annotation. There is little difference in mean accuracy on mildly fractured vertebrae, although the number of incorrectly converged solutions on neighbouring vertebra does appear to reduce. But when we come to the challenging case of the severely fractured vertebra there is a substantial reduction in the number of incorrect solutions on neighbouring edges, and this is accompanied by a significant improvement in mean accuracy from 1.65mm to 1.28mm, and the 90%ile error is reduced from 4.52mm to 3.05mm. Where the AAM was already performing well (i.e. the normal vertebrae), there is no advantage to using the kernel on this dataset. However in this application the marginal deterioration in well localised healthy vertebra is an acceptable price for a substantial improvement in accuracy for the more difficult severely osteoporotic cases. The marginal deterioration on normals may be because the use of the standard AAM update matrix on the transformed residuals leads to small biases. In fact variations in \mathbf{R} across the population is a general problem for AAMs. Cootes introduced the Updating AAM [3] to address this. In future we intend combining our form of robust AAM with the Updating AAM, which should lead to better optimisation of the objective function \tilde{S} (equation 9).

In conclusion the use of a Geman-McClure robust kernel prior to applying the standard AAM update matrix leads to a modest but significant improvement in accuracy, when applied to test images which would be expected to provide difficult cases. Initial coarse kernel scaling parameters can be derived using the AAM Jacobian, and we have shown a

method of reducing these to the finer scale of the appearance model itself. This generally avoids the trap of local minima that would be caused by the premature rejection of too many useful pixels. This method could be used with other kernels, but approximations in the update step mean it would work best with kernels having a gradual outlier rejection - e.g. inverse polynomial decay rather than exponential.

References

- [1] Theobald BJ, Matthews I, and Baker S. Evaluating error functions for robust active appearance models. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 149–154, 2006.
- [2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685, june 2001.
- [3] TF Cootes and CJ Taylor. An algorithm for tuning an active appearance model to new data. In *British Machine Vision Conference*, pages 919–928, 2006.
- [4] G.J. Edwards, T. F. Cootes, and C. J. Taylor. Advances in active appearance models. In *7th International Conference on Computer Vision*, pages 137–142, 1999.
- [5] Beichel R et al. Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging*, 24:1151–1169, 2005.
- [6] Black MJ et al. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7:421–432, 1998.
- [7] JP Hampel et al. *Robust Statistics: an Approach based on Influence Functions*. Wiley, 1986.
- [8] 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*, pages 258–269, 2003.
- [9] De la Torre F and Black MJ. A framework for robust subspace learning. *International Journal of Computer Vision*, 54:117–142, 2003.
- [10] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In *Proc. 2nd Conf. on Audio and Video-based Biometric Personal Verification*. Springer Verlag, 1999.
- [11] Roberts MG, Cootes TF, and Adams JE. Improving the segmentation accuracy of fractured vertebrae with dynamically sequenced active appearance models. In *MICCAI Joint Disease Workshop*, pages 1–8, 2006.
- [12] Roberts MG, Cootes TF, and Adams JE. Vertebral morphometry: semi-automatic determination of detailed shape from dual-energy x-ray absorptiometry images using linked active appearance models. *Investigative Radiology*, 41(12):849–859, 2006.
- [13] Black MJ and Jepson AD. Eigentracking: Robust matching and tracking of objects using view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998.
- [14] Black MJ and Anandan P. A framework for the robust estimation of optical flow. In *IEEE International Conference on Computer Vision*, pages 231–236, 1993.
- [15] Gross R, Matthews I, and Baker S. Constructing and fitting active appearance models with occlusion. In *IEEE Workshop on Face Processing in Video*, pages 72–79, 2004.
- [16] Geman S and McClure D. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, LII:4–5, 1997.
- [17] M. B. Stegmann, R. Fisker, and B. K. Ersbll. Extending and applying active appearance models for automated, high precision segmentation in different image modalities. In *Scandinavian Conference on Image Analysis*, pages 90–97, 2001.