# Accurate Regression Procedures for Active Appearance Models

Patrick Sauer
patrick.sauer@postgrad.man.ac.uk
Tim Cootes
tim.cootes@manchester.ac.uk
Chris Taylor
chris.taylor@manchester.ac.uk

Imaging Science and Biomedical Engineering
University of Manchester
Manchester, UK
M13 9PT

Active Appearance Models represent a class of algorithms for fitting the parameters of models which capture the morphology, i.e. shape and texture of deformable objects (e.g. faces, knees etc.). Given a particular instance of an object, the fitted model parameters allow for an optimal reconstruction of the object's appearance given the model and are low-dimensional descriptors of the object's state. The original Active Appearance Model algorithm [2] is used in conjunction with linear models of shape and texture and attempts to iteratively minimise an error measure between the image texture and the reconstruction given the current model parameters. Thus the reconstruction error is used to drive the update of the shape model parameters. In such *generative* models, the image texture is explicitly modelled and must be updated in every iteration of the algorithm. It is this 'tightening' of the error provided by updating the texture model that allows for accurate fitting, however, at substantial computational cost. In order to allow for real time model fitting, the inverse compositional framework proposed by Matthews et al. [3] allows to remove the update of the texture model from the iteration. Although this leads to highly efficient fitting procedures, the fitting accuracy is badly affected when the observed texture differs sufficiently from the mean texture of the training dataset.

**Discriminative regression models** An efficient alternative approach to model fitting is to train *discriminative* regression models which provide an explicit mapping $F(\mathbf{s})$ between the texture sampled from the current model location and updates to the (linear) shape model parameters. This is illustrated in Figure 1. Examples are [4, 5], which both use non-linear Boosting regression models of the form

$$F_i(\mathbf{s}) = \sum_{m=1}^{M} \lambda f_m(h_m(\mathbf{s})), \qquad (1)$$

where $i = 1, \ldots, d$ and $d$ is the number of shape model parameters, $h_m$ is a Haar feature, $f_m$ is a piecewise constant function and $\lambda$ is the "shrinkage" parameter. In this paper, we investigate the use of Random Forest regression models [1]. A Random Forest is a set of binary trees where at each node a feature is selected at random together with an optimal threshold. Given a texture sample, each tree predicts an update to the shape model. The output of the Random Forest is then taken to be the mean prediction of all trees.

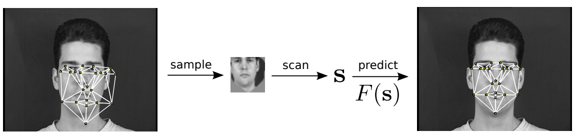



Figure 1: Shape prediction using discriminative regression models.

**Sequences** When using discriminative regression models, accuracy is affected by the lack of reinforcement provided by the texture model in the generative setting. In this paper, we attempt to emulate the role of the texture model in the generative framework by considering sequences of discriminative regression models where each stage is specialised to deal with a particular range of model displacements. Furthermore, the complexity of the models is chosen to increase throughout the sequence. This is illustrated in Figure 2.

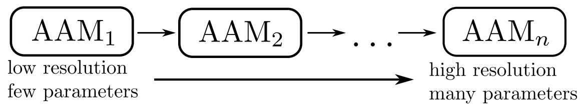
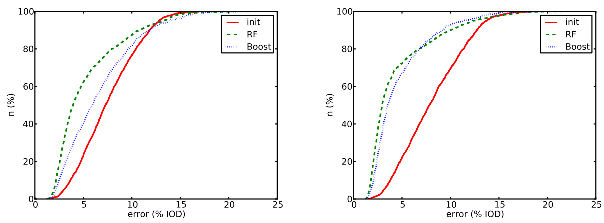



Figure 2: AAM sequences.

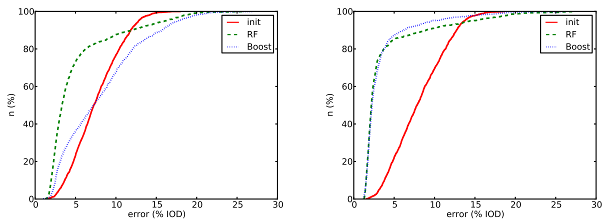
**Experiments** We evaluated the fitting performance across datasets with different properties as opposed to the more common method of splitting datasets into train/test parts. For this reason, we chose to use images taken from the BioID and XM2VTS datasets. The BioID dataset consists of low-resolution webcam images and features large changes in both pose and lighting. On the other hand, the XM2VTS dataset contains high-resolution images captured under controlled conditions. A pose-prediction experiment using a single model was used to gauge the performance of the regression methods, the result of which are shown in Figure 3. For testing the sequences, we used 5 models per sequence with each model making a one-shot prediction. Sequence results are shown in 4.

**Conclusions** In both experiments, the Random Forest clearly outperforms Boosting when trained on the XM2VTS dataset. The Random Forest may therefore be seen to be more resistant to overtraining when presented with a homogeneous database such as XM2VTS. However, the differences are much less pronounced when training on the more varied BioID database, where the Boosting method performs almost equally well.




(a) train: XM2VTS, test: BioID (b) train: BioID, test: XM2VTS

Figure 3: Pose prediction




(a) train: XM2VTS, test: BioID (b) train: BioID, test: XM2VTS

Figure 4: Sequence prediction

[1] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32–32, 2001.

[2] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[3] Iain Matthews and Simon Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[4] P. A. Tresadern, P. Sauer, and T. F. Cootes. Additive update predictors in active appearance models. In *British Machine Vision Conference*, 2010.

[5] Shaohua Kevin Zhou and Dorin Comaniciu. Shape regression machine. In *Proceedings of the 20th international conference on Information processing in medical imaging*, IPMI'07, pages 13–25, Berlin, Heidelberg, 2007. Springer-Verlag.