Additive Update Predictors in Active Appearance Models

Philip A Tresadern
philip.tresadern@manchester.ac.uk

Patrick Sauer
patrick.sauer@postgrad.manchester.ac.uk

Tim F Cootes
tim.cootes@manchester.ac.uk

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

Images of deformable objects from a specific class that vary in appearance (*e.g.* members of a biological group) can be described by a set of parameters that define their shape and texture. If we find the parameter values that describe a particular instance, we can use these values to analyse and interpret that example (*e.g.* in face analysis, texture strongly indicates identity and shape can indicate pose or expression). Active Appearance Models (AAMs) [1] find optimal parameter values by repeatedly improving the current estimate based on the corresponding image error (*i.e.* the difference between the model texture and the image data sampled with respect to the model shape). The usual way to do this is via linear regression that predicts parameter updates, $\delta\mathbf{b}$, from the image residual, $\delta\mathbf{g}$:

$$\delta\mathbf{b} = \mathbf{R} \cdot \delta\mathbf{g} \qquad (1)$$

where $\mathbf{R}$ can be estimated from training data by inverting the linear relationship [2]. Since this is typically under-constrained, more robust alternatives involve regressing over PCA coefficients of the residuals [4] or computing a Gauss-Newton approximation to the Jacobian matrix [1]; we refer to these latter two approaches as *lin_pcr* and *lin_jac*, respectively.

The aim of this study, however, was to investigate the implications of replacing linear regression with *additive* (or *boosted*) predictors – strong learners expressed as a weighted sum of weak learners,

$$\delta\mathbf{b} = \sum_{m=1}^{M} \lambda \mathbf{f}_m(h_m(\mathbf{g})), \qquad (2)$$

where: $h_m()$ computes a scalar feature (*e.g.* a Haar-like feature) of the sampled image data, $\mathbf{g}$; the *weak learner*, $\mathbf{f}_m()$, is a function (*e.g.* piecewise constant [5]) of h_m that predicts each of the parameter updates; and $0 \le \lambda \le 1$ is a 'shrinkage' parameter [3] that scales each weak learner's contribution to the final output. Learning additive models, however, is notoriously slow [6] and parameters (particularly λ) are tricky to determine. Using the publicly available XM2VTS and BioID datasets, we present: a method for speeding up additive model training; an investigation into how linear additive models compare with linear regression; and a hybrid model that marries the strengths of linear and non-linear models.

Pooling Weak Learners   Training (boosting) an additive model proceeds at each of the M rounds by adding the 'best' weak learner (weighted by λ) to the ensemble and updating the predicted output (or residual). The shrinkage parameter, λ, prevents any one learner from dominating [3] and thus promotes a diverse ensemble that is less prone to overfitting. Since $\lambda \ll 1$, however, the predicted output changes by only a small amount at each round and successive weak learners are often highly correlated. To retain these properties while improving efficiency, we instead propose adding the best $K > 1$ learners, weighted by λ/K, at each of the M rounds:

$$\delta\mathbf{b} = \sum_{m=1}^{M} \frac{\lambda}{K} \sum_{k=1}^{K} \mathbf{f}_{m,k}(h_{m,k}(\mathbf{g})). \qquad (3)$$

For an ensemble of a given size, the number of rounds of boosting can be reduced by a factor of K at little extra computational cost. Results that compare combinations of $M \times K$ and λ suggest that faster training incurs a small loss in accuracy (Figure 1) that can be compensated for by increasing K (though this means slower testing since the total number of learners has increased). The apparent sensitivity to λ is also reduced.

Linear Additive Models   When features, $h_m()$ and weak learners, $f_m()$ are themselves linear the additive model becomes equivalent to a linear regression. In a second experiment, we compared the three linear models with respect to accuracy and the basis images generated. The outcomes
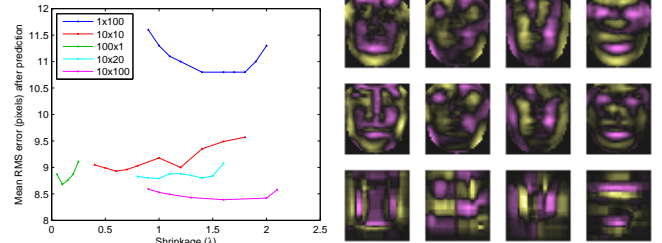

Figure 1: (left) Accuracy with respect to $M \times K$ and λ; (right) Basis images for three linear methods, corresponding to (l-r) scale, orientation, x-translation and y-translation.
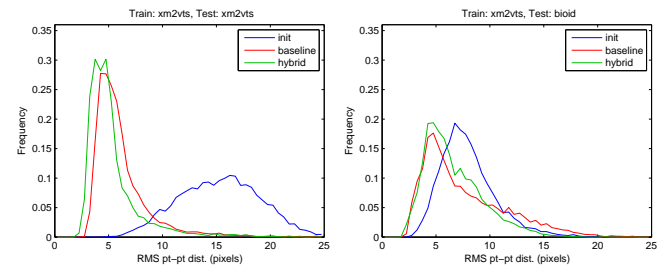

Figure 2: Distribution of RMS error when trained on XM2VTS and tested on (left) XM2VTS and (right) BioID.

showed that the linear additive model could perform as well as the other linear models and also highlighted the relatively poor performance of the Gauss-Newton approximation. Qualitatively, the basis images have similar characteristics such as a 'checkerboard' pattern for rotation (Figure 1). One observation from this experiment is that the additive linear model may be more efficient to apply than linear regression.

Hybrid Sequential Model   The added flexibility of non-linear additive models [5, 6] makes them less prone to local minima when far from the true solution. They are, however, no more (and potentially less) accurate that linear (regression or additive) models when close to the true solution; an experiment that examines point-to-point errors for both linear and non-linear predictors, initialized with displacements of varying size, confirms this intuition. We therefore propose a hybrid system that uses non-linear prediction to get close to the true solution then continues using linear additive models until convergence for accuracy; this outperforms a 'baseline' AAM when training and testing with the same set but improvement is less pronounced for cross-dataset evaluations where we train on XM2VTS and test on BioID (Figure 2).

[1] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, 1998.

[3] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.

[4] X. Hou, S. Z. Li, H. Zhang, and Q. Cheng. Direct appearance models. In *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.*, 2001.

[5] J. Saragih and R. Goecke. A nonlinear discriminative approach to AAM fitting. In *Proc. IEEE Int'l Conf. on Comp. Vis.*, 2007.

[6] S. K. Zhou and D. Comaniciu. Shape regression machine. In *Proc. Int'l Conf. on Information Processing in Medical Imaging*, 2007.