Preprint. Copyrighted material. This paper will appear at ICPR 2006, Hong Kong, 20-23 August 2006

Iterative Error Bound Minimisation for AAM Alignment

Jason Saragih[1] and Roland Goecke[1,2]
[1]Dept. of Information Engineering, RSISE, Australian National University, Canberra, Australia
[2]National ICT Australia*, Canberra, Australia
Email: jason.saragih@rsise.anu.edu.au, roland.goecke@nicta.com.au

Abstract

The Active Appearance Model (AAM) is a powerful generative method used for modelling and segmenting deformable visual objects. Linear iterative methods have proven to be an efficient alignment method for the AAM when initialisation is close to the optimum. However, current methods are plagued with the requirement to adapt these linear update models to the problem at hand when the class of visual object being modelled exhibits large variations in shape and texture. In this paper, we present a new precomputed parameter update scheme which is designed to reduce the error bound over the model parameters at every iteration. Compared to traditional update methods, our method boasts significant improvements in both convergence frequency and accuracy for complex visual objects whilst maintaining efficiency.

1. Introduction

Active Appearance Models (AAM) are a popular generative method which model non-rigid shape and texture of a visual objects using a low dimensional representation obtained from applying principle component analysis (PCA) to a set of labelled data. Since its advent by Edwards *et al.* in [6] and their preliminary extension [5], the method has found applications in many image modelling, alignment and tracking problems, for example [7, 8, 11].

The power of the AAM stems from two fronts. Firstly, its compact representation as a linear combination of a small number of modes of shape and texture variation enables optimisation over a small number of parameters. Secondly, the use of a fixed linear parameter update model allows efficient calculation of parameter updates. This second point is justified in initial publications by arguing that since the error image is evaluated in the pose normalised frame, the error function around the true minimum is *similar* for all images of the class of object modelled. This allows an iterative scheme of linear updates with a simple step size selection strategy to converge.

However, for complex visual objects exhibiting large variations in shape and texture, the assumption of a fixed linear update model can be too restrictive. In light of this problem, Baker *et al.* [2] proposed the project-out inverse compositional method (POIC). By reversing the roles of the image and the model in the objective function they show that an analytic form of the Gauss-Newton update can be precomputed. Their method is more accurate than the original formulation since the assumption of a fixed linear update model is well justified. Furthermore, since the method optimises only over the shape parameters, it is also the fastest known method to date. However, it does not work well when the object in the image is not similar to the mean texture and exhibits a large number of modes of shape variation [3], restricting its application to simple objects, a person specific AAM for example. Methods which overcome this problem such as [1] and [4] require the update model to be recalculated at every iteration, an expensive operation.

In this paper, we extend the idea of pre-learning the update model as described in [6] to pre-learning the whole alignment process. In Section 2, we give a brief overview of the AAM and measures of fit. Our method for alignment and its subsequent training process is described in Section 3. To motivate our approach, we present experimental results in Section 4, comparing our method with the original AAM an POIC methods. We conclude in Section 5 with an indication of extensions and future work.

2. Background

The AAM has a compact representation of both shape and texture as a linear combination of a small number of modes of variation:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{E}_s\mathbf{p}_s \quad \text{and} \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{E}_t\mathbf{p}_t,$$

*National ICT Australia is funded by the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council

where $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ are the mean texture and shape vectors, $\mathbf{E}_s$ and $\mathbf{E}_t$ are matrices of horizontally concatenated modes of shape and texture variation, where $\mathbf{p}_s$ and $\mathbf{p}_t$ are their respective parameters. The modes of shape and texture variation are generally built by applying PCA to a set of labelled training images.

Given a current estimate of the AAM parameters $\mathbf{p}$, the updates are usually calculated using a linear update model as follows:

$$\Delta\mathbf{p} = \mathbf{R}\,\mathbf{x}, \tag{1}$$

where $\mathbf{R}$ is the linear update model and $\mathbf{x}$ is the texture error between the current texture estimate and the warped image region. The parameters are then updated either additively [6] or composed in an inverse fashion [2].

The update model for most AAM alignment methods are designed to minimise a variant of a least squares loss function over the texture. If we consider pre-learning the update model, there are a number of other measures of error that can be utilised. One of these is the class of ϵ-insensitive loss functions, which we investigate in this paper:

$$|y_i - f(\mathbf{x}_i)|_\epsilon := \begin{cases} 0 & \text{if } |y_i - f(\mathbf{x}_i)| \leq \epsilon \\ |y_i - f(\mathbf{x}_i)| - \epsilon & \text{otherwise} \end{cases},$$

where y_i is the target parameter update value for the i^{th} training instance and $f(.)$ is the update model. In particular, support vector regression (SVR) [12] solves a regularised form of the ϵ-insensitive loss as a quadratic program by utilising the following form of the update model:

$$f(\mathbf{x}) = \langle \mathbf{r}, \mathbf{x} \rangle + b \text{ with } \mathbf{r} \in \chi, b \in \Re, \tag{2}$$

where χ denotes the space of the input patterns and $\langle ., . \rangle$ the inner product. If χ is the input space then the update model is that of Equation 1 with an additional bias term.

A useful variant of the basic SVR is ν-SVR [10], which integrates the hyper-parameter ν into the convex cost function such that minimisation is now over the error bound ϵ as well as the regression model. In the context of this paper, it is useful to think of the parameter ν as the upper bound on the fraction of samples outside the error bound ϵ.

3. Iterative Error Bound Minimisation

The main weakness of the original AAM formulation is that the relationship between the texture error and the parameter updates is clearly nonlinear. Thus, an adaptive linear update model is required for the general case. It is possible to train a nonlinear update model, trained on a set of perturbed images in a similar fashion to that of the original linear regression method [6]. However, this simply treats alignment as a machine learning problem, ignoring the structure of an AAM alignment. As a consequence these methods are generally computationally demanding and, hence, are rarely used in practice.

The alignment of an AAM has the peculiarity that the warped image texture for a given parameter setting extracts only a subset of the information required to completely describe the optimal parameter setting. Therefore, it is difficult to build an update model which can accurately predict updates to parameters which depend on the missing information. The main idea of this paper is to utilise information from a number of parameter settings to build the update model. Formally, we wish to find an update model of the following form:

$$\Delta\mathbf{p}_N = f_N(\mathbf{x}) \circ W\left(\mathbf{I}; \mathbf{p} + \sum_{i=1}^{N-1} \Delta\mathbf{p}_i\right), \tag{3}$$

where $\circ$ is the composition operator, $\mathbf{I}$ is the image, $W(\mathbf{I}; \mathbf{p})$ is the AAM warping function, $f_N(.)$ is the update model for the N^{th} parameter setting and $\Delta\mathbf{p}_i$ is the resulting update. The current parameters are then updated as follows:

$$\mathbf{p} \leftarrow \mathbf{p} + \sum_{i=1}^{N} \Delta\mathbf{p}_i \tag{4}$$

Since the relationship between the image pixels and the AAM parameters are generally nonlinear, any objective function which simultaneously minimises for all update functions f_i will be non-convex. In this paper, we propose a greedy approach where at each iteration the update model f_i is chosen to maximally reduce the error bounds over the parameters in the training set. We call this method the iterative error bound minimisation (IEBM). This approach is akin to boosting methods, however it differs in that the data used in calculating the weak learners changes after the addition of every weak learner to the ensemble (i.e. after every iteration). For the set of weak learners, we use the linear update model in Equation 2. For any given training set, we can always find a linear update model that does not increase the error bounds over the training set (the null update, for example). The hope is that there exists one that will significantly reduce the error bounds over the training set.

For every $f_k, k \in [1, N]$, solving for the optimal linear update model is a quadratic program:

$$\min \ \frac{1}{2}\mathbf{r}_j^T\mathbf{r}_j + \nu\epsilon_j$$
$$\text{subject to } \begin{cases} \Delta p_{ij} - \mathbf{r}_j^T\mathbf{x}_i - b_j \leq \epsilon_j \\ \mathbf{r}_j^T\mathbf{x}_i + b_j - \Delta p_{ij} \leq \epsilon_j \end{cases}$$

for every AAM parameter, where $(\mathbf{r}_j, b_j)$ is the update model for the j^{th} parameter, ϵ_j is the error bound for that parameter, $(\mathbf{x}_i, \Delta p_{ij})$ are the observation/response pair for the i^{th} training instance and ν regulates the trade-off between regularisation and empirical risk. This formulation is

known as the hard-margin SVR. As the problem is convex, the globally optimal update model can be obtained for every parameter. The linear update model for each iteration is obtained by concatenating the updates for every parameter as follows:

$$\mathbf{R} = \left[\mathbf{r}_1, \ldots, \mathbf{r}_N\right]^T \quad \mathbf{b} = \left[b_1, \ldots, b_N\right]^T \tag{5}$$

giving:

$$\Delta\mathbf{p} = \mathbf{R}\,\mathbf{x} + \mathbf{b} \tag{6}$$

In some instances, it may be beneficial to minimise the error bound only over a subset of the training data. This may be the case when there are a few training instances which are uncharacteristically *difficult*, where the reduction in error bounds may be too small to be useful if the model needs to accommodate for these cases. These samples may be outliers in the data. To allow for this, we use slack variables to capture the outliers as follows:

$$\min \ \frac{1}{2}\mathbf{r}_j^T\mathbf{r}_j + C\left(\nu\epsilon_j + \frac{1}{l}\sum_{i=1}^{l}(\xi_{ij} + \xi_{ij}^*)\right)$$

$$\text{subject to} \ \begin{cases} \Delta p_{ij} - \mathbf{r}_j^T\mathbf{x}_i - b_j \leq \epsilon_j + \xi_{ij} \\ \mathbf{r}_j^T\mathbf{x}_i + b_j - \Delta p_{ij} \leq \epsilon_j + \xi_{ij}^* \\ \xi_i, \xi_i^* \geq 0, \end{cases},$$

where ξ and ξ^* are the slack variables. This formulation is known as the ν-SVR [10]. Here the parameter ν represents the upper bound of outliers, which if sufficiently small will ensure a high frequency of convergence. The choice of the regularisation parameter C is more difficult, since it is dependent on the visual phenomenon being modelled as well as the relationship between the training and testing sets. Better generalisation is generally obtained by using a small C at the cost of a smaller reduction of the error margin at every iteration since less emphasis is placed on reducing ϵ in the quadratic program.

With this formulation, the outliers at every iteration should be removed from the training set so not to unduly influence future iterations. However, to maintain a fixed size of the training set, every discarded sample should be replaced with a new training sample. One of the advantages of training for the whole alignment process is that it can be specialised to the initialisation process used. Statistics about the distribution of samples about the optimum for the initialisation process can be collected and from this every discarded outlier can be resampled, preserving the coupling between alignment and its initialisation process. After resampling, these samples need to be propagated to the current iteration level using the appropriate linear update models. To allow for generic initialisation schemes, one can uniformly sample within predetermined initial error bounds[1].

[1] An initialisation process utilising an AAM trained in this manner must be able to initialise within the predetermined initial error bounds.

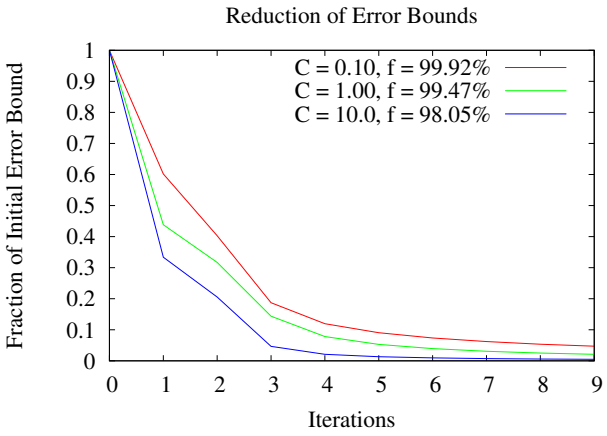

Figure 1. Reduction of error bounds in IEBM for different settings of the regularisation parameter C.

4. Experiments

To evaluate our method, we used the IMM Face Database [9] which contains 240 images of 40 individuals, each exhibiting a range of expression, pose and lighting conditions. Out of these, 30 individuals were randomly chosen for training and the others for testing. The images were downsampled to 320×240 to reduce training time.

To facilitate comparisons with the POIC alignment method, we use an independent appearance model in our experiments, however the IEBM method is not dependent on the parameter representation. The appearance model was trained, keeping 95% of the variation in shape and texture. With this model, the IEBM alignment process was trained for 10 iterations using 20 samples for each of the 180 training images. We use a generic initialisation scheme here with the initial error bound assigned to ± 10 pixels for translation, 10° rotation, 0.1 scaling and 1.0 standard deviations for all other parameters. The fraction of outliers, ν, at every iteration was set to 0.001 to ensure a high frequency of convergence. Figure 1 illustrates the effective point-to-point error bound at every iteration for two settings of the regularisation parameter C. It is clear that regularisation restricts not only the reduction of error bounds at every iteration, but also the capacity of the method.

Using the same training set, we compared our method both with the original AAM formulation [6] and the POIC method. Figure 2 shows the histogram of point-to-point RMS errors at convergence for each of the methods. The graph shows a significant improvement in convergence accuracy of the IEBM method compared to the other two (note that an error of zero is unattainable since only 95% of the shape variation was used). Furthermore, the convergence

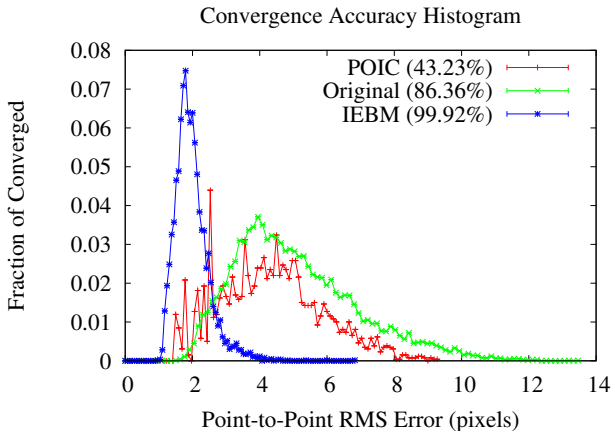

Figure 2. Histogram of point-to-point errors after search from displaced positions.

frequency for IEBM (99.92%) is far superior to the other two methods (86.36% for the original and only 43.23% for the POIC method), achieving convergence in almost every trial. Here we define convergence as having a final point-to-point error smaller than at initialisation.

Despite this significant improvement in accuracy of the IEBM method, it retains its computational efficiency. In fact, it is slightly faster than the POIC per iteration, currently the fastest known method, as it updates its parameters additively rather than in an inverse compositional fashion. It should be noted however, that the IEBM method requires all trained iterations to be performed whilst POIC terminates using some other condition.

5. Conclusion

A new linear update scheme for AAM alignment is proposed which utilises the optimality properties of support vector regression to find the best model for every iteration. The method is motivated by the computational complexity of adaptive update methods in problems with large appearance variation. Since our method precomputes the update model and no steps size adaptation is required, our method is one of the most efficient alignment algorithms to date. A summary of the contribution of this method are as follows:

- Convergence frequency is predictable and independent of initialisation (as long as within initial bounds)
- Convergence bounds can be directly controlled by computational resources available.
- Generalisation is directly integrated into the training process to reflect confidence over the training set.
- Alignment process can be specialised to initialisation process.

Through experiments we have shown that an iterative error bound minimisation process can obtain better convergence frequency and accuracy compared to the original fixed linear regression method. However, since our method aims to minimise error over a training set, its accuracy cannot be better than one which adapts the update model to the problem at hand.

As our problem is formulated as a machine learning problem, various extensions already in use in other fields can be applied here. Examples include automatic feature selection, selective warp updates and nonlinear update models.

References

[1] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical report, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, November 2003.

[2] S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms. In *CVPR 2001*, December 2001.

[3] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 1. Technical report, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, July 2002.

[4] A. Batur and M. Hayes. Adaptive active appearance models. *IEEE Transactions on Image Processing*, 14(11):1707–1721, November 2005.

[5] T. F. Cootes, G. Edwards, C. J. Taylor, H. Burkhardt, and B. Neuman. Active appearance models. In *Proc. Eur. Conf. Computer Vision*, volume 2, pages 484–489, 1998.

[6] G. Edwards, C. J. . Taylor, and T. F. Cootes. Interpreting face images using active appearance models. In *IEEE Int. Conf. Automatic Face and Gesture Recognition*, pages 300–305, 1998.

[7] T. Lehn-Schiøler, L. K. Hansen, and J. Larsen. Mapping from speech to images using continuous state space models. In *Lecture Notes in Computer Science*, volume 3361, pages 136 – 145. Springer, Jan 2005.

[8] P. Mittrapiyanuruk, G. N. DeSouza, and A. C. Kak. Accurate 3D tracking of rigid objects with occlusion using active appearance models. In *WACV/MOTION*, pages 90–95, 2005.

[9] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann. The IMM face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Lyngby, May 2004.

[10] B. Schoelkopf, A. Smola, R. Williamson, and P. L. Bartlett. New support vector algorithms. *Neural Computation*, 12:1207–1245, 2000.

[11] M. B. Stegmann and H. B. Larsson. Fast registration of cardiac perfusion MRI. In *International Society of Magnetic Resonance In Medicine*, page 702, Toronto, Canada, 2003.

[12] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.