Incremental learning of dynamical models of faces

Cyril Charron[1]
CharronC@cardiff.ac.uk
Yulia Hicks[1]
HicksYA@cardiff.ac.uk
Peter Hall[2]
pmh@cs.bath.ac.uk
Darren Cosker[2]
dpc@cs.bath.ac.uk

[1] School of Engineering
Cardiff University
Cardiff, UK

[2] Department of Computer Science
University of Bath
Bath, UK

Active Appearance Models (AAM) [1] are a useful and popular tool for modelling facial variations. They have been used in face tracking, recognition and synthesis applications. For modelling facial dynamics of speech, they have been used in conjunction with Hidden Markov Models (HMM) [2]. However, the high dimensionality of the training data and of the resulting AAMs leads to long learning time of HMMs and thus imposes serious limitations on their joint use.

Here, we propose a new method for learning HMMs of facial dynamics incrementally. Our algorithm is fully unsupervised and can be used for on-line learning as new data becomes available. Another important feature of our algorithm is the automatic choice of the number of states in the model. We show in experiments an improvement in learning speed of three orders of magnitude. Finally, we demonstrate the quality of the learned HMMs by generating video footage of a talking face.

There are two major stages in our algorithm. In the first stage, we update the observation distribution (represented by a Gaussian Mixture Model (GMM) of the existing HMM by merging it with that of the incoming HMM. This is done using a new principled Expectation-Maximisation (EM) procedure which does not require any recourse to the training data, but simply relies on the state descriptions. Thus, this stage is both efficient in terms of memory and fast. Due to the built-in selection mechanism, the updated states describe the data optimally in the Minimum Message Length (MML) sense, thus being compact. In the second stage of our algorithm, we use an approach proposed by Hicks and Hall [5] to update the transition matrix and the initial state priors of the new HMM, which is done in time linearly dependent on the number of states.

We present quantitative assessment of our algorithm on a widely used data set: the shrinking spiral. An interesting property of this dataset is that it forms a one dimensional manifold embedded in an higher dimensional space $\mathbb{R}^3$, which is similar to the trajectories of the faces we analyse in this article. We want to mimic online incremental learning of a GMM describing the above dataset. We generate the main test set consisting of $1,250$ points that we split (over 20 trials) into a randomly chosen number, ranging from two to ten, of smaller subsets. A standard batch method developed by Figueiredo and Jain [3] with a built-in selection mechanism is applied to the main set resulting in a model which constitutes our baseline. The same batch method is also applied to every subset producing local models (as opposed to the global model learned on the whole set).

To start our incremental learning procedure, we concatenate the descriptions of the first two local models and use the result as the input of our procedure. We then concatenate the result of our method to the next local model and apply our merging procedure again. This is repeated until all local models have been merged. As a quick check, we also look at the effect of just concatenating the local models.

Table 1 presents different measurements obtained with the three different models (batch, our incremental and the simple concatenation). The measurements are averaged over 20 trials. One of the important characteristic of our method is its automatic selection mechanism, thus we present the number of components of the models. As can be seen, our incremental method and the batch method produce similar models in terms of complexity. The models resulting from the concatenation overfit in the MML sense the data as they have considerably more components. We also look at the Bayesian criterion proposed by Roberts (RBC) [6], which takes into account the likelihood and penalises the complexity of the models. As expected, the concatenated model has the worst RBC, whilst our method is comparable with the batch method.

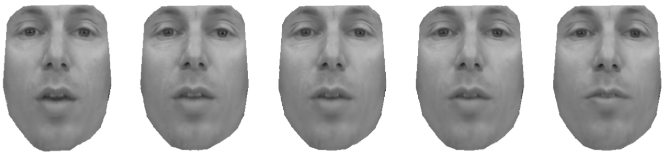

Figure 1: Five synthetised frames 340 to 390 (10 frames steps).

	batch mixtures	concat mixtures	incr mixtures
Nb of components	17 ± 0	31 ± 0	$17 \pm 0.$
log likelihood	-7.77 ± 0.03	-7.78 ± 0.02	-7.75 ± 0.03
RBC ($\times 10^4$)	4.56 ± 0.03	4.64 ± 0.02	4.55 ± 0.01
Time (s)	53 ± 2	4.2 ± 0.7	15 ± 3

Table 1: Comparison of the batch and incremental methods for learning observation distribution. The concatenation is given for comparison.

Next, we apply our method to learning HMMs modelling the dynamics of a video realistic human face. In [2], Active Appearance Models (AAM) were trained on videos of talking people in order to produce new realistic synthetic videos for computer graphics applications. A continuous HMM was learned from the same training data to model their dynamics. The dimensionality of the training data and the number of samples necessary to learn such models is usually very large. Thus, in the original paper [2], incremental methods [4] were used for constructing the AAM to circumvent the memory issues. Nonetheless, in [2], the HMMs are still learned on the whole dataset projected onto the AAM eigenspace. Here, we apply the incremental procedure presented in this article to learn an HMM of facial dynamics on the same dataset as in [2].

An example of a typical video produced by our algorithm can be found in the supplementary material and some frames are shown in Fig 1. The animation is comparable to what Cosker *et al.* obtained with the same data set (with exclusion of the results produced by hierarchical models [2]).

The full quantitative and qualitative assessment of our method as well as its mathematical derivation is presented in the paper.

Acknowledgements

This work has been funded by EPSRC grant EP/E017576/1.

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

[2] D. Cosker, D. Marshall, P. Rosin, and Y.A. Hicks. Speech driven facial animation using a hidden markov co-articulation model. *Proc. of IEEE International Conference on Pattern Recognition (ICPR)*, 1: 128–131, 2004.

[3] M. A. F. Figueiredo and A. K. Jain. Unsupervised learning of finite mixture models. *IEEE Transactions on pattern analysis and machine intelligence*, 24(3):381–396, 2002.

[4] P. Hall, D. Marshall, and R. Martin. Merging and splitting eigenspace models. *IEEE Transactions on pattern analysis and machine intelligence*, 22(9):1042–1049, 2000.

[5] Y.A. Hicks, P.M. Hall, and A.D. Marshall. A method to add hidden markov models with application to learning articulated motion. *British Machine Vision Conference*, 2003.

[6] S. J. Roberts, D. Husmeier, I. Rezek, and W. Penny. Bayesian approaches to gaussian mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intellingence*, 20(11):1133–1142, 1998.