Learning-Based Approach to Estimation of Morphable Model Parameters

Vinay Kumar

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Developing a method for estimating the parameters of a linear morphable model (LMM) directly from the images of the object class, based on a learning from examples approach.

Motivation: Amongst the many model-based approaches to modeling object classes, the Linear Morphable Model is an important one (Vetter and Poggio [6], Jones and Poggio [2]). It has been been used successfully to model faces, cars and digits. In these applications, the task of matching a novel image to the LMM is achieved through a computationally intensive analysis by synthesis approach. In Jones and Poggio [2], the matching parameters are computed by minimizing the squared error between the novel image and the model image using a stochastic gradient descent algorithm. This technique may take several minutes for matching even a single image. A technique that could compute the matching parameters with considerably less computations and using only view-based representations would make these models useful in real-time applications.

The motivation for this work comes from the use of a learning-based approach in real-time analysis of mouths (Kumar and Poggio [4]), in which it was shown that a regression function can be learnt from a Haar wavelet based input representation of mouths to hand labeled parameters denoting openness and smile. Therefore, it points to the possibility that learning may be a way for directly estimating the matching parameters of an LMM from the image.

Previous Work: Previously, morphable models of mouths have been constructed for the purpose of synthesis of visual speech (Ezzat and Poggio [3]). We will explore the morphable model as a tool in the analysis of mouth images. There has also been an attempt in the work by Cootes et al. [1] to speed up the process of analysis by synthesis for computing the matching parameters of morphable models (which they call active appearance model). The speed-up is achieved by learning several multivariate linear regressions from the error image (difference between the novel and the model images) and the appropriate perturbation of the model parameters (the known displacements of the model parameters), thus avoiding the computation of gradients. This method is akin to learning the tangent plane to the manifold in pixel space formed by the morphable model.

Approach: In this work, we propose to construct an LMM to model various mouth shapes and expressions. Following Jones and Poggio [2] the LMM is constructed from examples of mouths. However, we reduce the parameter set by performing PCA on the example textures and flows. We then use Support Vector Machine (SVM) regression (Vapnik [5]) to learn a non-linear regression function from a sparse subset of Haar wavelet coefficients to the matching parameters of this LMM directly. The training set (the outputs) for this learning problem is generated by estimating the true matching parameters using the stochastic gradient descent algorithm described in Jones and Poggio [2].

Results: Preliminary results indicate that SVM Regression along with parameter selection through cross-validation gives accurate estimates of LMM parameters. This is illustrated in figure 1 for a single flow parameter.

Future Work: So far, we have looked at LMMs that model mouths of only one person. Future work will look at the problem of extending this method to more complex LMMs, namely those that model the mouths of different people. An obvious application of estimating LMM parameters is in image syntheis (or graphics). However, recently it has been suggested that LMM parameters could also be used for higher level image analysis (or vision) such as face identification. In this paper, since we are working with mouth shapes, we explore a different application, namely, viseme recognition. Visemes are the visual analogues of phonemes (Ezzat and Poggio [3]). Recognizing visemes have potential applications in enhancing the performance of speech recognition systems.

Impact: This research is likely to contribute to the fields of facial analysis and synthesis, ultimately impacting on fields such as man-machine interfaces, teleconferencing, animation.

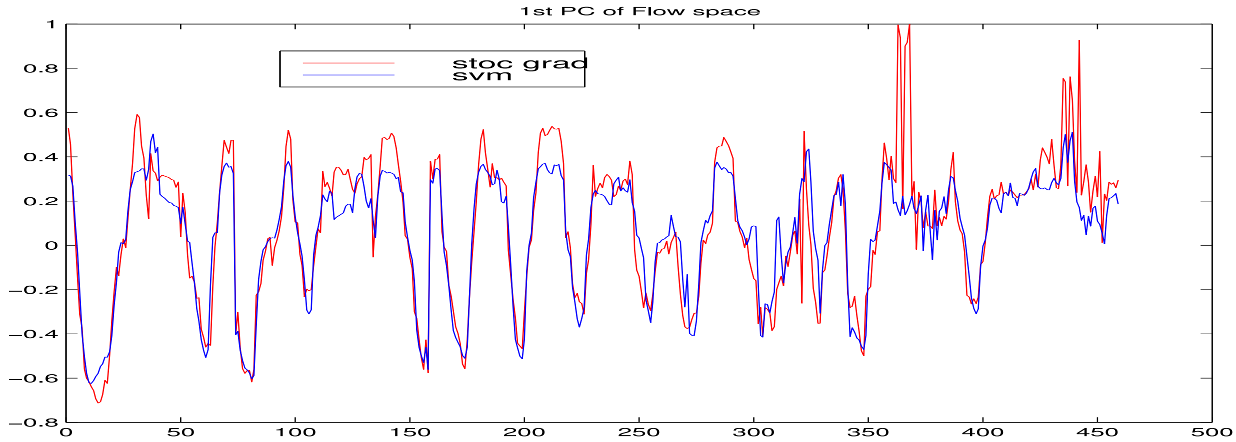

Figure 1: Estimates of LMM flow parameter using the stochastic gradient descent and support vector regression on a test sequence of 459 images.

Research Support: This research is sponsored by a grant from the National Science Foundation under contract No. IIS-9800032.

References:

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proceedings of the European Conference on Computer Vision*, Freiburg, Germany, 1998.

[2] M. Jones and T. Poggio. Multidimensional morphable models: A framework for representing and matching object classes. In *Proceedings of the International Conference on Computer Vision*, Bombay, India, 1998.

[3] T. Ezzat and T. Poggio. Visual Speech Synthesis by Morphing Visemes. *MIT AI Memo No. 1658/CBCL Memo No. 173*, 1999.

[4] V. Kumar and T. Poggio. Learning-Based Approach to Real Time Tracking and Analysis of Faces. In *Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition*, pages 96–101, Grenoble, France, 2000.

[5] V. N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[6] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:7:733–742, 1997.