Face Recognition with Cheap and Expensive Data

Charles Rosenberg and Sebastian Thrun

Problem: This work explores the area of learning to learn in the context of the face recognition problem. Existing learning algorithms often need many different views of a person's face to learn a model, but in practice a single view is often all that is available. This research asks the question: Can recognition rate be improved by looking at images of other people's faces? Intuitively, this is a reasonable question to ask, since the domain of face recognition shares a large set of invariances. Some invariances are easy to describe, such as rotational or translation invariance, but others are much more difficult to model, such as invariance with respect to facial expression, or aging. The basic research conjecture is that such invariances can be learned using a "cheap" database of faces and applied to guide generalization when learning to recognize a new face. In this context, the envisioned algorithms will learn at two levels: at a conventional level that seeks to capture the people-specific features of a specific face, and a meta-level, which seeks to learn generic invariances useful for a large class of face recognition problems.

Impact: There is a large class of problems which fall under the general category of learning to learn from "cheap" and "expensive" data. In recognition scenarios, often a large database of labeled examples from many people is available, but few from the specific person that is to be recognized. For example a system that recognizes a person's handwriting should take as few samples as possible in order minimize inconvenience. If this research is successful, a learning algorithm should be obtained which, in the face domain, can perform this recognition task with high accuracy given only a small number of examples.

State of the Art: An overview of recent work in learning to learn can be found in [4]. A recent survey and evaluation of face recognition algorithms can be found in [3] and a survey of connectionist face processing algorithms can be found in [5]. The majority of these algorithms follow the same basic approach:

1. Preprocess the query image according to a model of invariants, for example: scale or shift.
2. Generate a low dimensional feature vector from the query image.
3. Compute a weighted distance measure between the query feature vector and a stored database of class exemplars and output the class of the closest examplar.

The key algorithmic choices are: invariants, features, and distance metric. In [2] the face is normalized in position and scale based on detected features. The image features are then projected onto a set of eigenfeatures and a euclidean distance metric is calculated. In [1] a reference image database is constructed in which a single reference view of each person is converted into fifteen virtual views with different head orientations. The query image is warped to match the geometry of the reference images based on eye and nose position. Pixel-wise normalized correlation with feature templates is used to evaluate the match.

Approach: The FERET face image data set described in [3] is being used for this work. In these experiments, the face data is randomly divided into a "cheap" and an "expensive" data set. The cheap face images are used to train a neural network which has as its input two face images and outputs a single value which is the posterior probability that the two faces picture the same person. No preprocessing of the face images was done aside from scaling them down to a size of 32 x 48 pixels. The neural network is trained with backpropagation and has a completely connected feedforward architecture with a single hidden layer. The goal of the training is to have the network extract a feature set and learn a distance metric from the raw pixel data appropriate to the task of distinguishing the identities of the people's faces in the images. The database

(a) (b)

Figure 1: An example of a good match set (a) and a poor match set (b). The subimage at the far left is the query image. Starting at the second subimage from the left are the matching images from the gallery set sorted in rank order. The bar at the bottom of each image is a plot of the network output, 0–1.0, a larger number corresponding to a better match. The brighter dots mark intervals 0.25 long in the 0 to 1.0 scale.

---

of images of known individuals to recognize comprises the "expensive" data set. To recognize a new face the network compares the new face to each of the faces in the gallery set. An advantage of this architecture is that information gained from multiple images of an individual can be combined outside of the network.

In experiments 69% classification accuracy was measured on the test set with a gallery size of 93 subjects. The images presented to the network included faces at various scales, between straight and quarter profile, and taken over various dates. Figure 1 contains an example of a good match set and a poor match set. In experiments with two images of each individual in the gallery set, the best performance for combining the results of the multiple similarity measures acquired for each individual was to treat each of the images as is if they had come from different individuals – in other words there was no direct exploitation of the fact that the images were of the same individuals. A possible explanation of this result is that there are distinct clusters of faces in "face space". This suggested the next logical step in this work.

Future Work: The next step in this work is to utilize an unsupervised clustering technique to create multiple classifiers each of which will specialize in operating on a specific cluster in the space of faces. The way we plan to do this is by utilizing a variant of EM. The proposed algorithm is as follows:

1. Train N neural networks, with different initial random weight values for a fixed number of epochs.
2. Evaluate the performance of the N neural networks on the entire training set.
3. Train each of the N neural networks only on the training examples they performed best on in step 2. While the performance of the networks is still increasing on a validation set, go to step 2.

The goal of this algorithm is to generate multiple networks each of which will output a similarity measure specialized to a particular cluster in face space. The hope is that these will correspond to specific invariances in this domain. This will be verified by examining which training images are assigned to each network.

Selected References:

[1] D. Beymer and T. Poggio. Face recognition from one example view. A.I. Memo 1536, Massachusetts Institute of Technology, September 1995.

[2] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[3] P. J. Phillips, H. Moon, P. Rauss, and S. Rizvi. The feret evaluation methodology for face-recognition algorithms. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, pages 137–143, June 1997.

[4] S. Thrun and L. Y. Pratt, editors. *Learning To Learn*. Kluwer Academic Publishers, Boston, 1997.

[5] D. Valentin, H. Abdi, A. O'Toole, and G. Cotterell. Connectionist models of face processing: A survey. *Pattern Recognition*, 27:1209–1230, 1994.