Unsupervised Feature Learning for Object Classification

Laxman Dhulipala, Harry Gifford, Wangzi He {ldhulipa, hgifford, wangzih}@andrew.cmu.edu

Contribution

We implement an unsupervised method for learning features from images, which are then used for image classification. Our method is robust, and we experimentally show its performance on the CIFAR-10 dataset, a standard dataset in the machine learning community. We implement and describe a feature-learning pipeline, which can be recursively applied in order to learn successively more general features.

Model

The Model was learnt from 175 subjects. We used one neutral expression scan per identity and 50 expression scans of a subset of the subjects.
The identity model is a linear model build from the neutral scans.

$$\boldsymbol{f} = \boldsymbol{\mu} + \mathbf{M}_n \boldsymbol{\alpha}_n \quad . \tag{1}$$

For each of the 50 expression scans, we calculated an expression vector as the difference between the expression scan and the corresponding neutral scan of that subject. This data is already mode-centered, if we regard the neutral expression as the natural mode of expression data. From these offset vectors an additional expression matrix $\mathbf{M}_e$ was calculated, such that the complete linear Model is

$$\boldsymbol{f} = \boldsymbol{\mu} + \mathbf{M}_n \boldsymbol{\alpha}_n + \mathbf{M}_e \boldsymbol{\alpha}_e \tag{2}$$

The assumption here is, that the face and expression space are linearly independent, such that each face is represented by a unique set of coefficients.

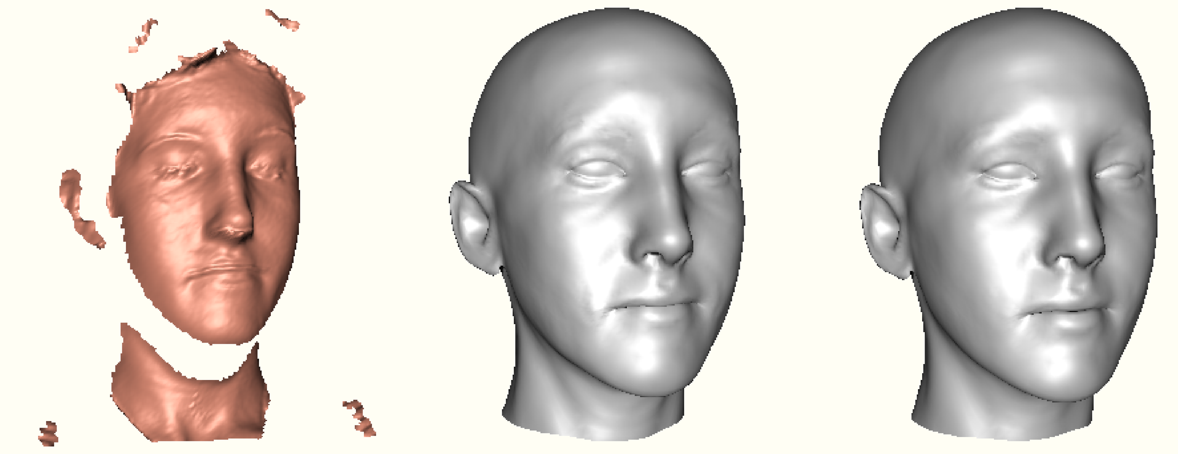
Fitting

A Robust Nonrigid ICP method was used to fit the model to the data. Robustness was achieved by iteratively reweighting the correspondences and using hard compatability test for the closest points.
Fitting was initialized by a simple nose detector and proceeded fully automatic.

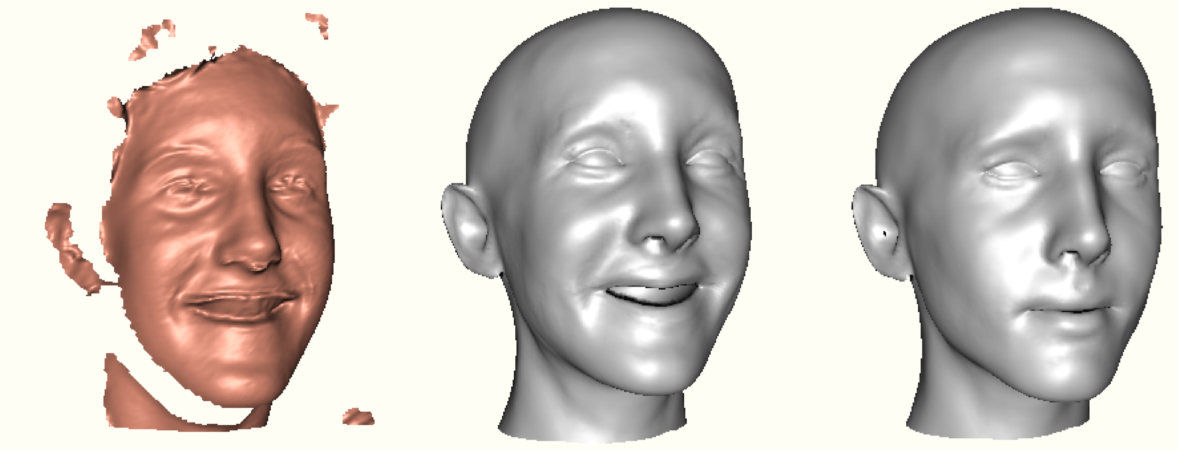
Distance Measure

The Mahalanobis angle between the identity coefficients $\boldsymbol{\alpha_n}$ was used for classification.

Expression Neutralization



a) Target b) Fit c) Normalized

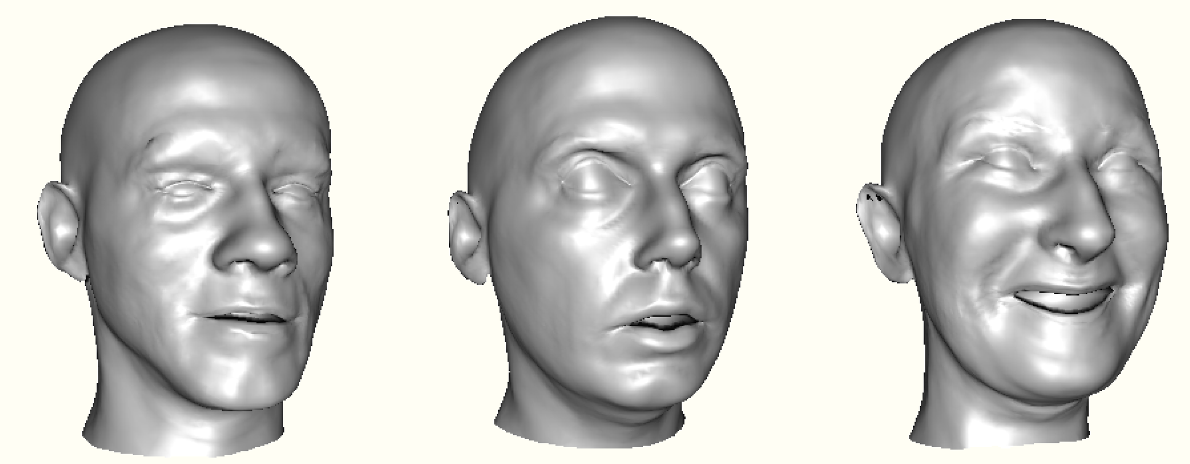


a) Target b) Fit c) Normalized

Expression normalisation for two scans of the same individual. The robust fitting gives a good estimate (b) of the true face surface given the noisy measurement (a). It fills in holes and removes artifacts using prior knowledge from the face model. The pose and expression normalized faces (c) are used for face recognition.

Robustness

a) Targets



b) Fits

The reconstruction (b) is robust against scans (a) with artifacts, noise, and holes.

Results

The method was evaluated on the GavabDB expression dataset which contains 427 Scans, with 3 neutral scans and 4 expression scans per ID. To test the impact of expression invariance on neutral data we used the UND Dataset from the Face Recognition Great Vendor Test, which contains 953 neutral scans with one to eight scans per subject.

Expression neutralization improves results on the expression dataset without decreasing the accuracy on the neutral testset. Plotted is the ratio of correct answers to the number of possible correct answers.

Plotted are precision and recall for different retrieval depths. The lower precision of the UND database is due to the fact that some queries have no correct answers.

Open Questions

While the expression and identity space are linearly independent, there is some expression left in the identity model. This is because a "neutral" face is interpreted differently by the subjects. We investigate the possibilty to build an identity/expression separated model without using the data labelling, based on a measure of independence.

References

Funding

This work was supported in part by Microsoft Research through the European PhD Scholarship Programme.