CS281A Project:
Nonlinear Dimensionality Reduction on Human Facial Expressions

Ryan White
ryanw@eecs.berkeley.edu

November 26, 2003

Abstract

In this paper we explore the utility of nonlinear dimensionality reduction techniques in the realm of facial expression analysis. First, we test the ability of nonlinear techniques to describe the higher nonlinear nature of human facial expressions. We exploit the data-driven model of an embedding to create novel facial expressions. Finally, we composite the facial expressions back on the face.

1 Introduction

Within the Computer Vision community, there has been a recent push to recognize identity from digital photos of individuals. To properly identify the individual in the image, many factors need to accounted for: pose, lighting, facial expression, change in appearance et cetera.

While pose and lighting are well understood phenomenon, facial expressions are not. In a recent paper on the topic, expression and occlusion were cited as the bigest problems with facial recognition. [4] Hopefully, by removing information about identity, pose and lighting, one can characterize facial expressions independently.

Our goal is to say that the intrinsic dimensionality of a human face is quite low, and that by projecting image data down in a reasonable way, we can characterize different aspects of the face in different dimensions.

The conjecture here is that we can use these dimensionality techniques on a large and varied dataset to learn a reasonable space for expressions. First, this will give us a new attack on an existing problem of human facial expression classification. Second, it will allow us to look for faces of similar expressions in a large dataset. Finally, it is possible that this characterization will allow us to confront the larger issue of human face recognition. As shown in [4], facial expressions currently represent a substantial stumbling block in face recognition.

2 Background - Statistics

There are several methods to perform dimensionality reduction. Linear techniques include Principle Components Analysis (PCA) [13], Fisher Linear Discriminants (LDA) and Locality Preserving Projections (LPP) [5] [6]. Nonlinear techniques include Locally Linear Embedding (LLE) [9] [10], Isometric Feature Mapping (Isomap) [12] and Semidefinite Embedding (SDE).

Since this paper is not intended to be a survey of dimensionality reduction techniques, the focus of this section will be on the relevant techniques used in future sections of the paper. Specifically, the focus will be on Locally Linear Embedding, since it provides a convenient framework for creating new expressions.

2.1 Locally Linear Embedding

In LLE, the basic assumption is that the neighborhood of a given data point is locally linear. In other words, a data-point can be reconstructed as a linear combination of its neighboring points. When projecting to a low dimensional subspace, LLE preserves this locally linear structures.

To enforce the linear structure, the following reconstruction error is defined:

$$e(W) = \Sigma_i |\vec{X}_i - \Sigma_{j \in \mathcal{N}(i)} W_{ij} \vec{X}_j|^2 \tag{1}$$

In this equation, $\vec{X}_i$ are the original datapoints and W_{ij} are the weights used for reconstruction. In contrast to a linear method, the second sum is only over the neighbors of point i, denoted $\mathcal{N}(i)$. From the local reconstruction in the high dimensional space, one can define a similar reconstruction error in the low dimensional space:

$$\Phi(Y) = \Sigma_i |\vec{Y}_i - \Sigma_{j \in \mathcal{N}(i)} W_{ij} \vec{Y}_j|^2 \tag{2}$$

Putting everything together, this creates the following embedding procedure:

- First, build the neighborhood map $\mathcal{N}(i)$.
- Second, using the neighborhood map determine the reconstruction weights W_{ij}.
- Finally, using the neighborhood map and the reconstruction weights, determine the embedded values $\vec{Y}_i$.

Several key points are worth mentioning. The construction of the neighborhood map is done under the constraint $\Sigma_j W_{ij} = 1$. Also, a regularization term is typically used when calculating the weight matrix to prevent numerical errors and to allow embedding when the number of neighbors exceeds the number of dimensions. Finally, this entire method is computationally feasible and involves solving some linear equations and an eigenvalue problem.

2.2 Other Techniques

The most notable other nonlinear dimensionality reduction technique is Isomap. Isomap creates a distance matrix between points by assuming that only local distances are accurate and then enforces the triangle inequality to derive longer distances over the manifold. [12]

There has been some recent work in a technique known as Semidefinite Embedding. While a talk was given at UC Berkeley about this technique, it hasn't been published yet. [11]

3 Dataset / Preprocessing

The dataset used in this paper consists of approximately 400 unique photos of President George W. Bush extracted from the news. [2] [8]

In order to obtain somewhat reasonable representations of the data, we preprocess by looking for faces within the images. This is done using the Mikolajczyk-Schmid face finder. [7] In the News dataset, this is an absolutely essential step, because it is not known if a face is in the image.

There are a very large number of different people in the images. We selected the most frequent individual (President George W. Bush), and manually looked at all images with his name in the caption, selecting only those that actually contained his face. After that, the faces were rectified to adjust for affine differences in face orientation, specifically placing the eyes, nose and mouth in roughly the same location in every output image. [1] Only images with reasonable rectification are kept. ('Reasonably' rectified images selected by hand to include only those with features in roughly the canonical configuration)

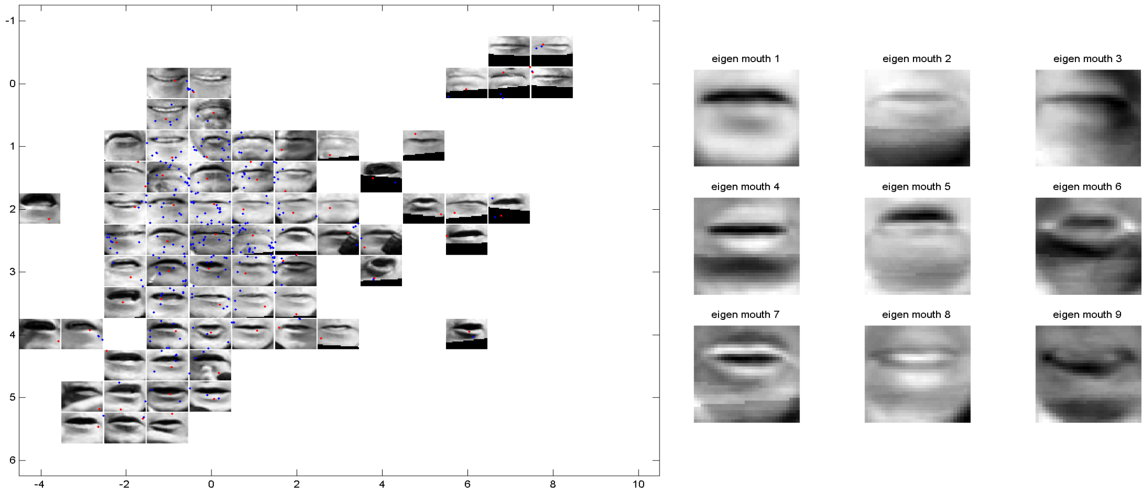


Figure 1: The results of PCA on photos of bushes mouth. The left image shows his mouth on the first two principle components. The right image represents the nine largest variance components.

4 Experiments

4.1 Linear Techniques

As an initial attempt to understand the problem, the traditional technique of PCA was used to characterize photos of Bush's mouth. Figure 1 shows the eigen-mouths and the projection of the images onto the two highest variance coordinates. [13]

Linear techniques are advantageous for a number of reasons. First, they cover the entire space. Second, they are typically easier to compute. However, it is quite reasonable to believe that the data in this case is intrinsically non-linear.

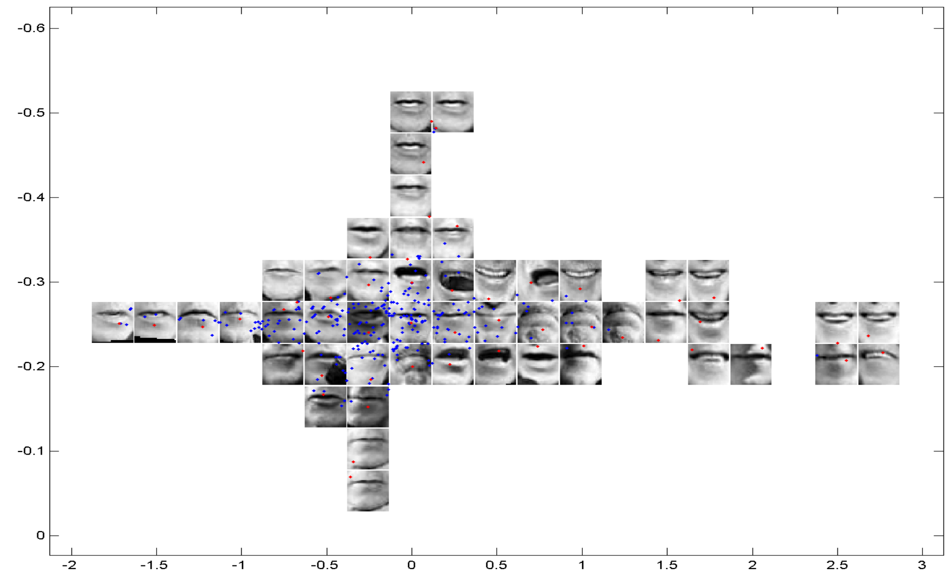
4.2 LLE on Bush photos

The first portion of this section is devoted to simply finding a low dimensional subspace for photos of Bush, similar to the low dimensional subspace found using PCA. Initial results reveal that the entire faces was too complicated to support viewing in two dimensions, so almost all images will focus on a single 'feature' such as an eye or a mouth.

In figures 2 and 5, LLE was use to project the data down onto two dimensions. To accurately characterize the effect of this projection, the original photo closest to a regularly sampled gridpoint is shown behind dots representing the actual 2D projection of the image. This allows one to easily view the data and characterize the projection. In contrast to the linear projections, the non-linear projection appears to have less of a global structure.

If the embedding structure creating by LLE is reasonable, then it should be possible to create points on the embedding surface that do not correspond to actual images. This can be done by constructing a local weight matrix similar to that used by the original LLE weight matrix. First, a neighborhood matrix is computed in the lower dimensional subspace. Second, the weight matrix is computed for these neighbors. Finally, a new image is created using the *same* weight matrix on the corresponding images (i.e. n the original dimensional space).

Since the neighborhood matrix consists of the 5 to 10 closest neighbors, it is important to regularize the weighting to reduce the overfitting of the data. (In two dimensions it takes only three points to



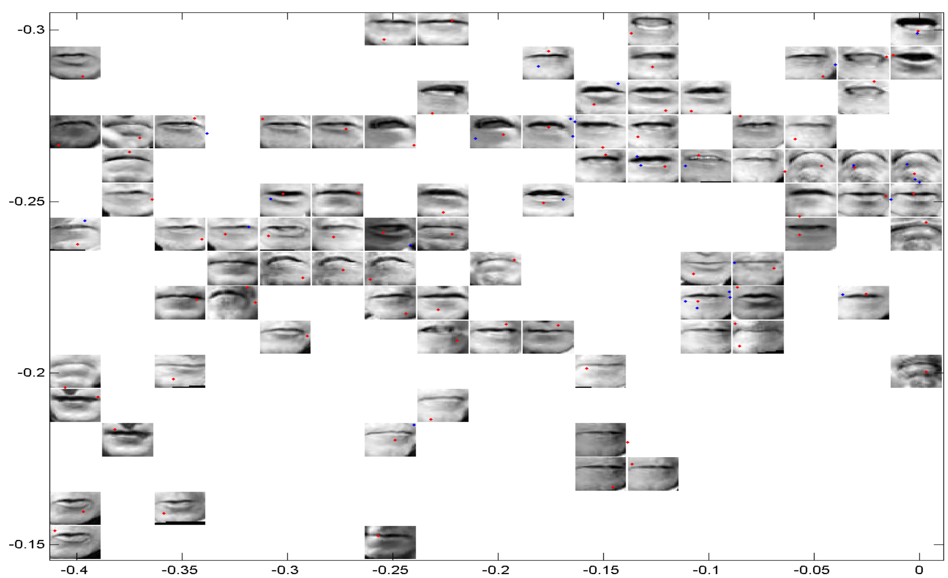


Figure 2: The results of LLE, projecting photos of President Bush's mouth onto two dimensions. Each image represents one image associated with that region of the embedding space. The points represent individual images, where the red points are the ones selected for the image display. The bottom view is a zoomed in version of the top view. The parameters used here are K=15, tol = 10^{-3}.

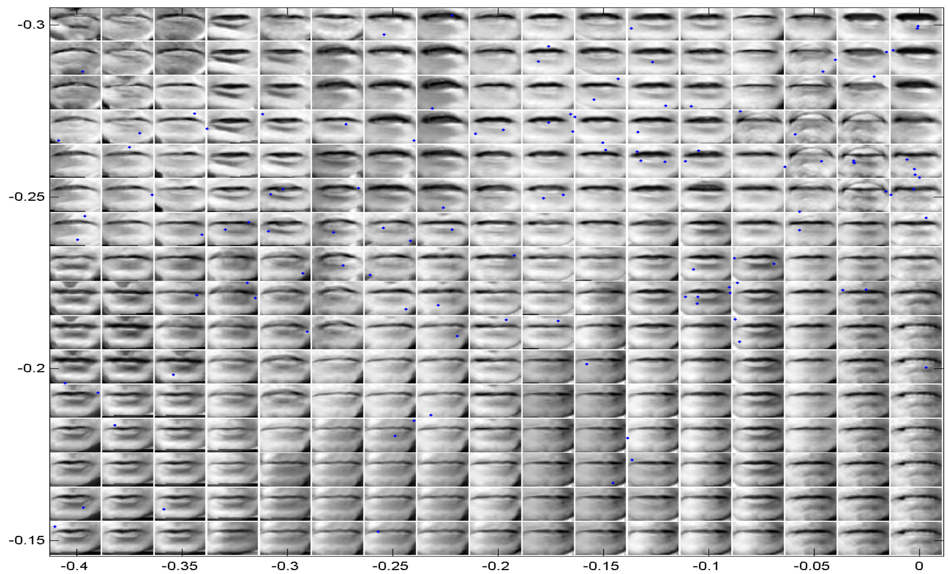


Figure 3: Here, the coordinates represent the first two dimensions that result from LLE. The images are generated by creating a local weight matrix of existing photos to establish a reconstruction in the higher space. Again, points represent the original images.

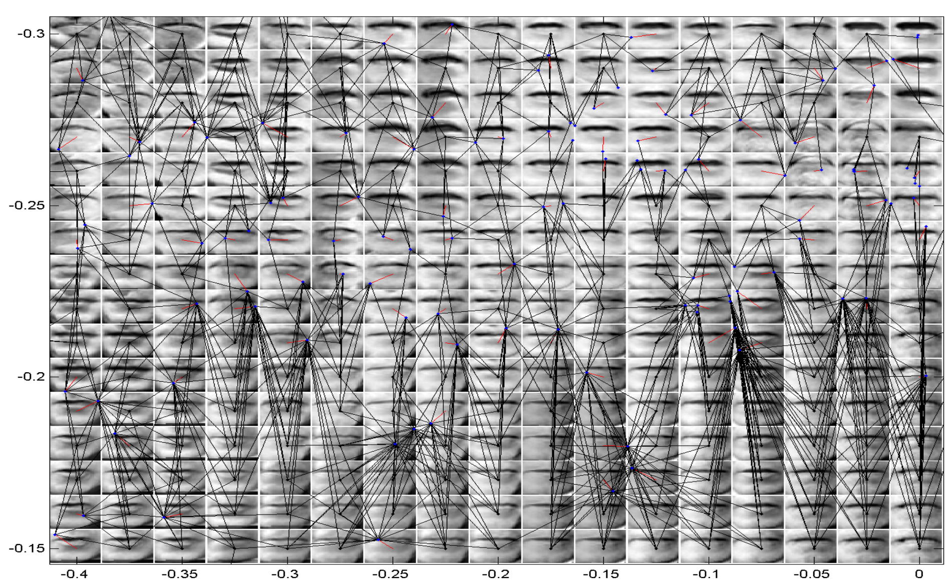


Figure 4: Again, the coordinates represent the first two dimensions that result from LLE. The black lines show the local weight matrix used to compute the new images.

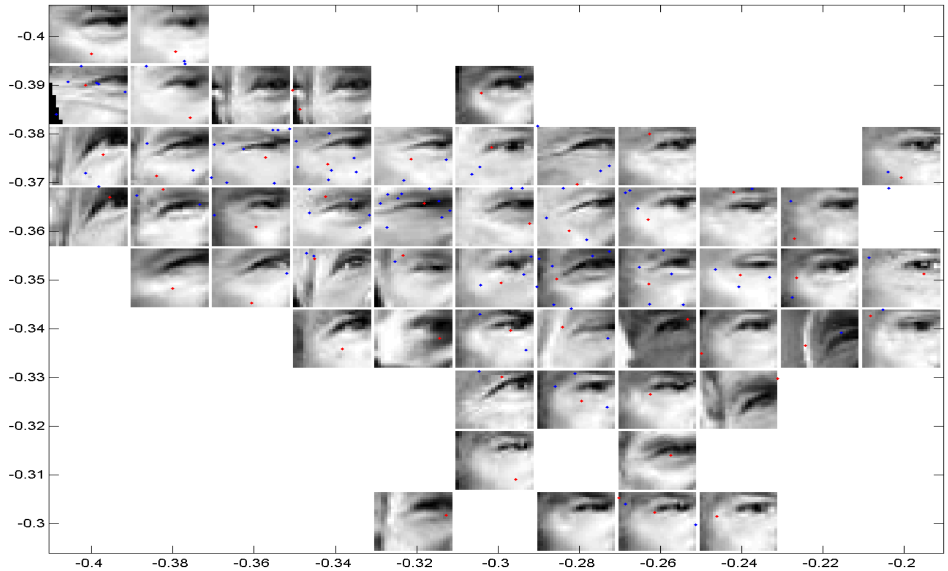


Figure 5: The results of LLE, projecting photos of President Bush's eyes onto two dimensions. The parameters used here are K=15, tol = 10^{-3}.

reconstruct the entire plane using local weights that sum to 1) Following the notaion in [10], the weights can be constructed as follows:

$$\begin{aligned} G_{jk} &= (\vec{x}-\vec{\eta}_j)\dot{(}\vec{x}-\vec{\eta}_j) \\ w_j &= \frac{\sum_k G^{-1}_{jk}}{\sum_{lm} G^{-1}_{lm}} \end{aligned}$$

To allow for regularization, one can update the Gram Matrix (G) by the following modification:

$$G_{jk} \leftarrow G_{jk} + \delta_{jk}(\frac{\Delta}{K})\mathrm{Tr}(G) \tag{3}$$

By increasing the tolarance term Δ, one can effectively regularize the weighting in the two dimensional case. Finally, one needs to enforce the convex hull constraint:

$$w_{ij} \leftarrow \frac{w_{ij}}{\sum_k w_{ik}} \tag{4}$$

A visualization of what happens on the result can be found in figures 3 and 6. For the most part, these reconstructed images actually look reasonable. Note that the same activity on the entire face fails significantly. One can view the neighborhood graph in figure 4.

Because of the heavy regularization used in computing the weight matrix, the images that appear in the same spot in the reconstructed images actually look somewhat different from the original images.

4.2.1 Compositing Faces

One way to test the structure of the LLE embedding and the image based creation of new faces is to composite a new mouth onto an existing face. For this experiment, new mouths were chosen at random from the array of mouths created in the previous section.

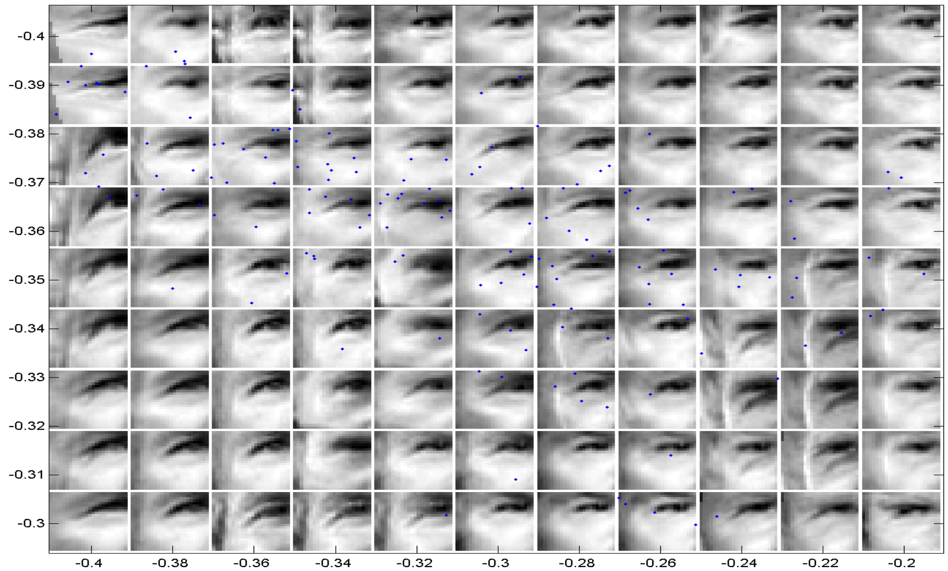


Figure 6: Here, the coordinates represent the first two dimensions that result from LLE. The images are generated by creating a local weight matrix of existing photos to establish a reconstruction in the higher space. Again, points represent the original images.

The nearest face was selected for compositing. To make sure that the new mouth matched the intensity values of the previous mouth in a reasonable way, the intensity values of the new image were globally linearly scaled to match the existing image as closely as possible:

$$\min_k \Sigma_{xy} |\mathcal{I}_{xy}^{\text{original}} - k\mathcal{I}_{xy}^{\text{new}}| \tag{5}$$

Here $\mathcal{I}^{\text{original}}$ is the original mouth and $\mathcal{I}^{\text{new}}$ is the new mouth. Results of this process are shown in figures 7 and 8. These images are some of the more promising composites. A better stitching technique could easily improve results. [3]

5 Discussion

Using LLE to project portions of Bush's face provided a somewhat reasonable way to characterize the variations inherent in his expression. Though images of an embedding of his entire face do not appear in this paper, they proved unsatisfying. We suspect that there are two causes for this problem: not enough data (400 images of Bush) and too low of an embedding dimension (can't easily view anything more than two dimensions).

In [10], they present an image of mouths projected onto the first two coordinates in a similar technique. Unlike the results here, their embedding maintains a strong global structure. There are two key differences. First, there dataset is significantly larger: instead of approximately 400 mouth images, they had 15960. Second, their dataset seemed to include many fewer artifacts of imaging in real world circumstances. (examples include lighting, occlusion, et cetera)

While the embedding presented here still leaves something to be desired, it is consistant enough to build new images of the patches. Furthermore, there is some promise of an image based approach to generation of novel facial expressions. In contrast to techniques like [14], this technique does not require in depth understanding of facial geometry.

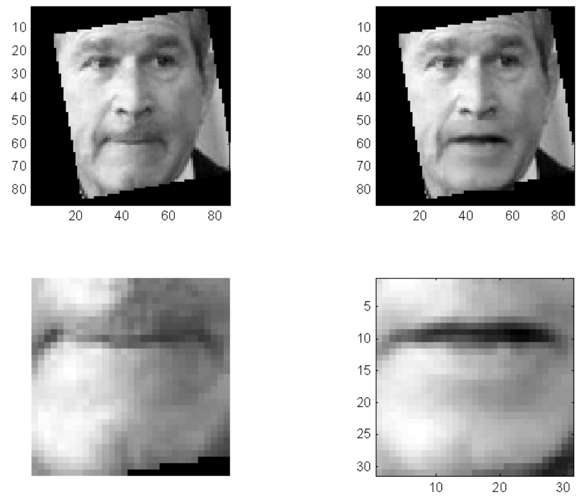


Figure 7: An example of image compositing. The top-left is the original face. The bottom-left is the extracted mouth. The bottom-right is the new mouth created using LLE, and the top-right is the mouth composited back onto the face. While the borders aren't perfect around the mouth, the image is still fairly impressive.

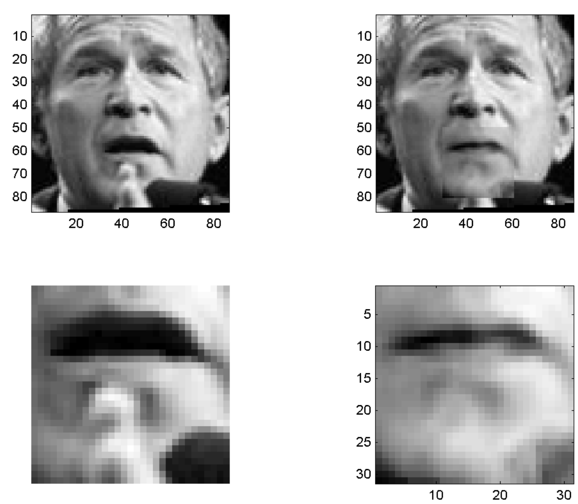


Figure 8: Another example of image compositing. This example is somewhat less impressive than the first, but given a proper stitching technique, it still shows promise.

The most promising aspect of this technique is the image compositing. One avenue of future research would be to build a better embedding by using more images. One could imagine building the embedding using images obtained from video. It would make sense to use a higher dimensional embedding as well, to capture the many aspects of variation. Evaluation of the embedding error could provide a determination of the intrinsic dimensionality of the data. Also, stitching techniques could easily improve the visual believability of the composite images.

References

[1] Alexander Berg and Jitendra Malik. Geometric blur for template matching. In *Computer Vision and Pattern Recognition*, 2001.

[2] J. Edwards, R. White, and David Forsyth. Words and pictures in the news. In *Proc. HLT-NAACL03 Workshop on Learning Word Meaning from Non-Linguistic Data*, 2003.

[3] Alexei A. Efros and William T. Freeman. Image quilting for texture synthesis and transfer. *Proceedings of SIGGRAPH 2001*, pages 341–346, August 2001.

[4] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition? In *Proc. Third Workshop on Empirical Evaluation Methods in Computer Vision*, 2001.

[5] X. He and P. Niyogi. Locality preserving projections. In *Proc. Advances in Neural Information Processing Systems*, 2003.

[6] X. He, S. Yan, Y. Hu, and H.J. Zhang. Learning a locality preserving subspace for visual recognition. In *Proc. International Conference on Computer Vision*, 2003.

[7] K. Mikolajczyk. *Detection of local features invariant to affine transformations*. PhD thesis, INPG Grenoble, 2002.

[8] Tamara Miller, Alexander Berg, Jaety Edwards, Michael Maire, Ryan White, Yee-Whye Teh, Erik Learned-Miller, and D.A. Forsyth. Faces and names in the news. In *submitted, Computer Vision and Pattern Recognition*, 2004.

[9] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, December 2000.

[10] S. Roweis and L. Saul. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, pages 119–155, 2003.

[11] Lawrence Saul. Nonlinear dimensionality reduction by semidefinite programming. talk given at UC Berkeley, October 2003.

[12] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, December 2000.

[13] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[14] T. Vetter. Synthesis of novel views from a single face image. *International Journal of Computer Vision*, 28:2:103–11, 1998.