Akinlawon Solomon
ECE 7866 Computer Vision
HW5

# Landmark Point Detection on Faces Using Active Shape Models

Given an image of an object, it is generally important to recover its underlying shape or structure. If a set of labeled training images of the object are available, then a model representing structural information observed in the set can be constructed. Therefore, to detect the shape of a new object, the strategy is to find the best fit to the constructed model of that object.

## Method:

A face can be defined to be an annotated set of 2D landmark points, such that:

$$S = \{x_1, x_2, \dots\dots\dots.., x_n, y_1, y_2, \dots\dots.., y_n\}$$

Where these points represent 2D pixel positions on the image of that face. In order to make meaningful comparisons between images of the same face, the landmark points must first be aligned in the same co-ordinate frame. That is, all the co-ordinates are adjusted until they all possess the same scaling, rotation and translation with respect to each other. These aligned vectors are then said to form a $2n - 4$ dimensional space. The goal is then to model this vector distribution, so that new faces can be examined to see if they are possible given said model, or in order for similar new faces to be created entirely. In this project, the model is created according to an algorithm known as an Active Shape Model. The steps to perform this are outlined in the following section. The algorithm used is due to Dirk Van Koon's ASM Package.

## Steps:

### Landmark Points:

- First, the 2D landmark points must be selected to represent the object. In this project, annotated markings and their corresponding images were provided.
- These landmark points were then aligned to remove rotations, translations and scaling using the Procrustes method.
- It can be observed that these landmark points are located along clear facial boundaries such as the eyes, along the border of the nose as well as the jaw, and on the contour of the mouth.

### PCA:

- Using PCA, the dimensionality of a dataset can be reduced. The shape data is analyzed using PCA to obtain principal components, or "modes", that best represent shape variation across all the landmark points. In other words, by applying PCA to the aligned landmark shape data we obtain:

$$X = \bar{X} + Pb$$

Where the matrix $X$ is an image matrix of 2D landmark points, $\bar{X}$ is the mean of all landmark points, and $P$ is the set of $k$ eigenvectors of the covariance matrix.

Finally, the vector $b$ is given by:

$$b = P^T(x - \bar{x})$$

- The eigenvectors represent the shape variation present in the set of landmark points. The parameters given in the vector $b$ are the eigenvalues that control the deformations possible in the model. This means the largest modes typically explain the global variation due to pose changes, which cause the landmark points to move relative to each other. As expected, smaller modes usually denote more local changes.
- In this project, 220 images and their corresponding landmark points were used in the training set. A plot depicting the modes is shown in the last section.

**Active Shape Models:**

- The data is first divided into a training set and a testing set. The training set is used to develop the model using PCA.
- Once the model has been created, instances of the model can be compared to a new image. This means that the parameters of the model are adjusted iteratively until the best fit to the new image is obtained.
- In order to obtain reasonable solutions, the eigenvalue parameter $b$ must lie within $\pm 3$ standard deviations of the mean shape. This ensures that only shapes similar to those in the training set can be observed.
- The model parameters are updated by minimizing a cost function between the model point and the boundary of the image we want to fit. The cost function moves along the normal of each model point in order to find the strongest edge gradient on the new image. The new model point is then said to minimize the cost function in the least squares sense.
- After less than a hundred iterations, the algorithm converges to a local minimum. This is the predicted shape of the new image. The figures on the last page show reconstructed points using images from a testing set.

**Procrustes Alignment of New Shapes:**

- The new shapes are finally aligned using Procrustes Analysis. The mean shape is shown on the last page.

**Discussion:**

Several key observations were made during the implementation of this algorithm:

- Choice of Training and Testing Set: The training set must contain enough variation in order to produce a model that scales well to new images. Otherwise, the algorithm performs badly whenever a test image differs slightly from the training set. In this project, 450 images were used to train the model. The testing set was composed of 20 images. These images were selected because they contained the most common emotional expressions i.e. neutral and smiling. Other expressions were more complicated and reduced model performance. Besides the algorithm still scales well to images with expression not seen in the training set. A possible improvement is to use RANSAC to select images that would produce a more comprehensive model.

- Landmark Intensity Profile: This feature determines the number of pixels the algorithm searches for along the normal. That is, given a model point, it determines how far the search algorithm goes to look for the nearest boundary in the new image. The author of the package recommends using a profile value between 3 and 7 but values upward of 10 also seem to perform well on this dataset.
- Initial Contour Position: The algorithm's performance is severely limited by the choice of a good starting position. This is because it searches for the nearest local minimum, which means it may diverge depending on how many iterations are used and the quality of the initial starting position. If a good starting position is chosen, then the algorithm converges quickly and diverges as the number of iterations is increased. For a poor staring position, the algorithm performs better as the number of iterations are increased. See the plots below for further description.
- Overall, the algorithm performs reasonably on this dataset provided the training set contains enough instances for each expression. The results of the algorithm are included on the last page.

Akinlawon Solomon
ECE 7866 Computer Vision
HW5

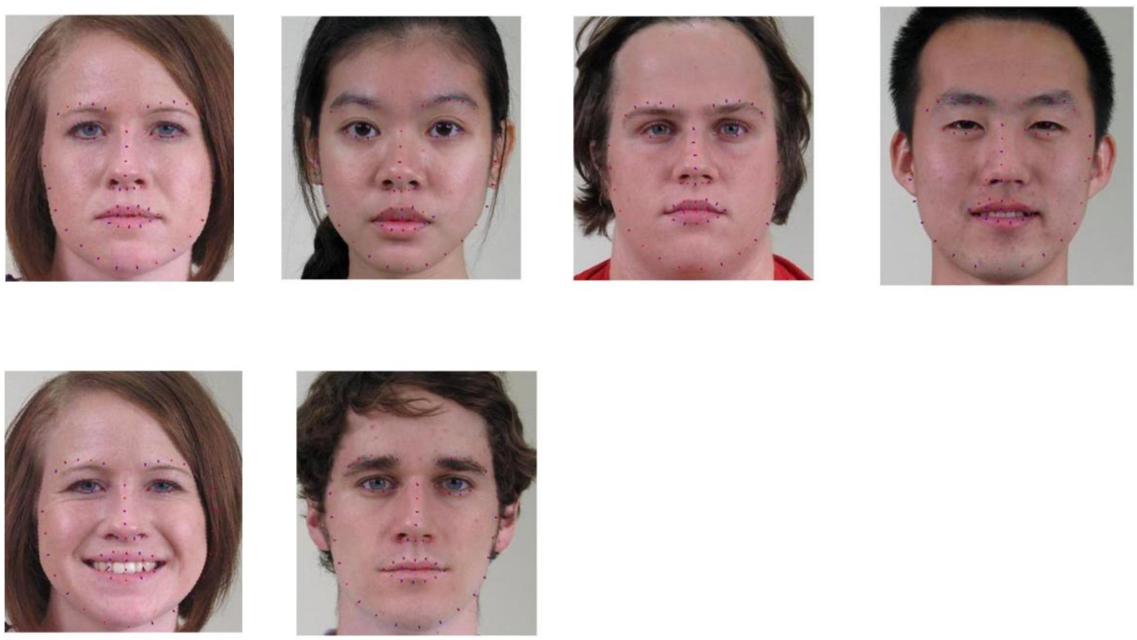


*Figure 1. Reconstructed Images from the Training set*

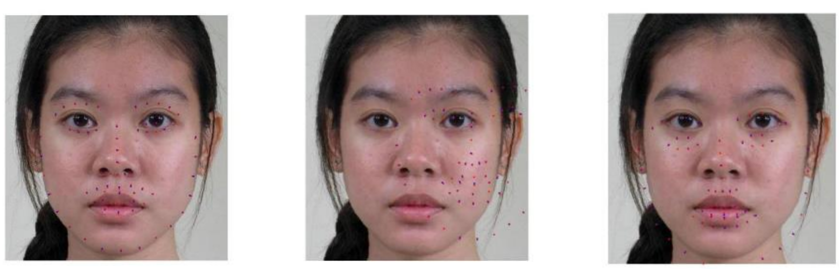


*Figure 2. Result after a poor initialization*

Akinlawon Solomon
ECE 7866 Computer Vision
HW5

*Figure 3. Good result despite using only 10 iterations; effect of a good starting point*

*Figure 4. Improved performance after 60, 80 and 100 iterations respectively; increasing iterations sometimes improves performance*

*Figure 5. Performance after 60, 80 and 120 iterations; increasing iterations may also worsen performance when a global minimum may not be possible.*

Akinlawon Solomon
ECE 7866 Computer Vision
HW5

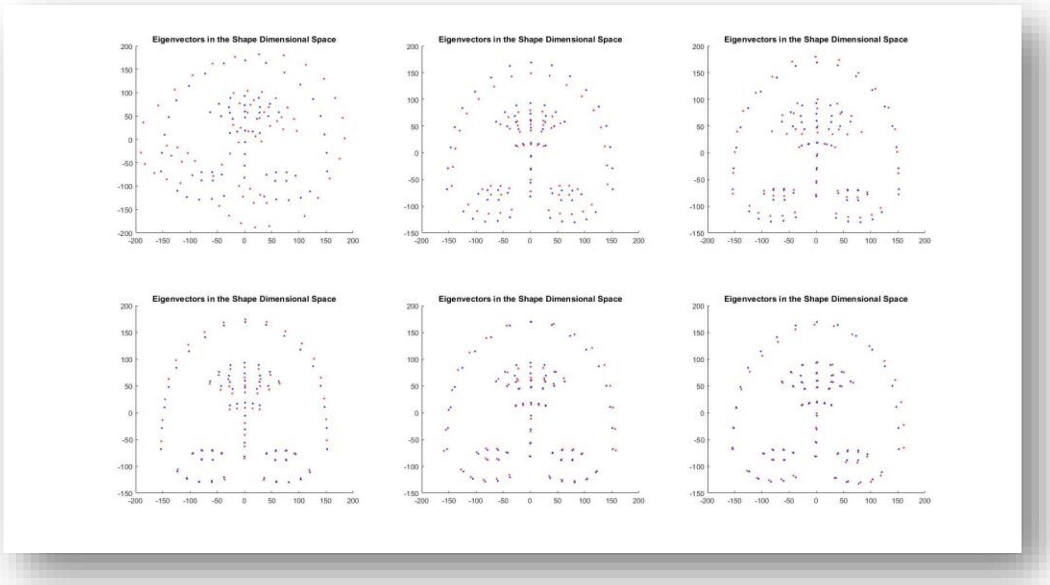
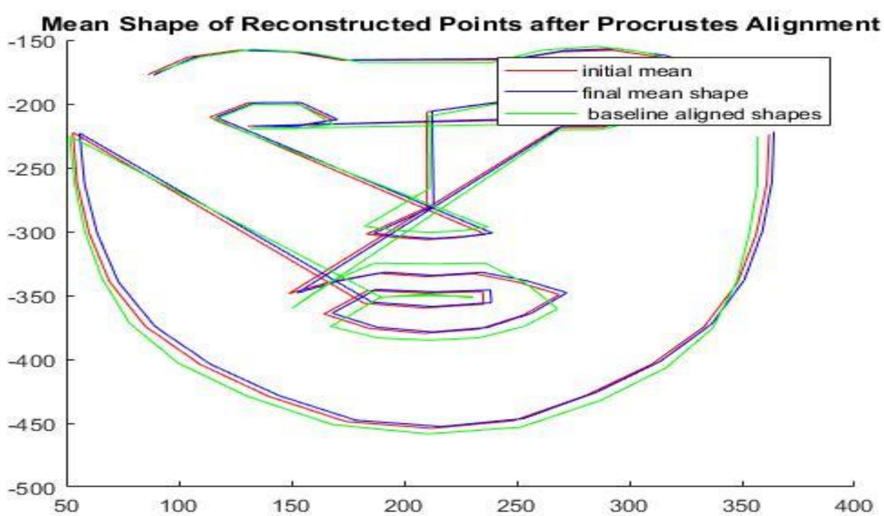



*Figure 6. Eigenvectors plotted in the Shape Space*



Akinlawon Solomon
ECE 7866 Computer Vision
HW5