Introduction

Seeing Unity in Diversity can resolve problems not only in life but in computer vision too. Imagine a vision system interested in detecting human faces. Human faces vary widely in color, texture and shapes. Yet they all possess certain landmark features that are common. For eg the eyebrows, nose, lips etc. By learning the boundary connecting these landmarks from a training corpora, the vision system can develop a generative model that can detect face in a given image by fitting an appropriate boundary generated from the model to the given face. Such a model is called as Active shape model (ASM).

But wait! ASM may even detect a non-human face(for eg an animal face) to be a human face! But then what if we consider not just the boundary but the texture inside the boundary while training. Then the vision system can develop an Active Appearance model (AAM) that generates a shape model as well as an appearance model which combined together can detect a human face in a given image.

This dissertation work proposes to delve deep in to the world of ASM and AAM, understand the mathematics that gives them their glory, implement a few applications that exhibits their power and if time permits investigate towards its improvisation.

The proposal here is divided in to five sections. The first section briefly explains the steps involved in ASM. Section 2 discusses about AAM. Then the aim and objectives which are basically the contents of the immediate above paragraph are listed as bullet points. Finally references are mentioned.

Active Shape Model

ASM has got a training phase and a test phase. Following are the steps in the training phase:

- Acquire the shape vectors
- Align the data
- Build a shape model
- Gather statistics across all landmark points across multiple resolutions to define a cost function for testing

In the test phase following are the steps:

- Initialize the shape by inverse aligning the mean shape depending on initial set of parameters

- Starting from coarsest resolution, iteratively best fit the shape readjusting the shape parameters using the cost function defined in training phase until convergence

All these steps are now explained briefly.

Training Phase:

Acquire the shape vectors:

Initially shape vectors corresponding to training images are acquired as a 2m x n matrix of x and y coordinates of the m landmark points across n training images. Let this matrix be called as A. For example, with respect to a human face the landmark points could be as depicted in Fig 1. Here 58 landmark points are chosen across nose, lips, eyes etc. If such landmark points are collected across 100 training images a 116 x 100 matrix can be formed. Such acquisition can be done not just with face images but even with medical images. But in case of special domain images like medical images, an expert like a radiologist has to choose the landmarks. Fig 2 shows an example of an annotated chest radiograph. It contains 166 landmarks bounding the right and left lungs, heart and the two clavicle bones. This is a sample image from JSRT database[ ] that contains 247 images divided in to two folds of 124 and 123 images respectively. Using first fold for training, a 332 x 124 matrix can be formed. It is to be noted that each column of this matrix corresponds to one training image that contains x and y coordinates of landmarks. In other words each column is a shape vector.

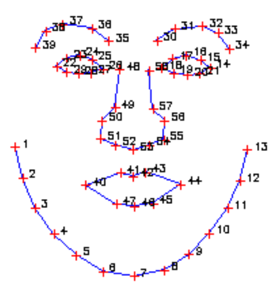

Fig 1: Landmark points on a face

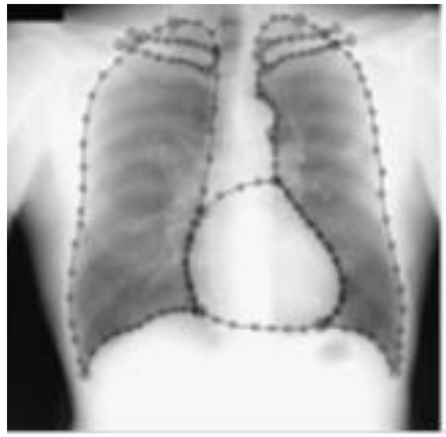

Fig 2: Landmark points on a chest radiograph

Align the data:

Once the matrix is acquired, each shape vector in the matrix is aligned i.e. the contour positions are aligned. This is achieved by

- Translating the centre of gravity of landmark points to the origin
- Subtracting the mean rotation angle from each contour position.

Translating the centre of gravity of landmark points to the origin can be done by subtracting the mean of the contour positions from each contour position.

Rotation angle of each contour position can be computed as $\tan^{-1}$(y-coordinate/x-coordinate).

Let the aligned matrix be called as B.

Build shape model:

Shape model is built by applying Principal component analysis (PCA) to the aligned shape vectors. A principal component is a direction (eigen vector of the covariance of B) of maximum variance of data. If the original data is highly correlated and lives in a high dimensional space, dimensionality can be reduced by retaining only the directions where variance of the data is maximum. Of course while trying to recover the original data some loss which may not be significant will be incurred. Projecting the original data on to the retained principal components defines a set of coordinates called a shape parameter vector b. The ith component of vector b is a shape parameter corresponding to ith eigen vector. Each eigen vector captures some aspect of the original shape vector. By varying the components of b, new shapes similar to the ones seen in the training set but not exactly present in the training set can be generated. Therefore, ASM is a generative model. Clearly, to maintain similarity to the training shapes that have been observed shape parameters can vary only within 3 standard deviations of the shape vectors. Examples of generated shapes for chest radiograph example containing two lungs, two clavicle bones and heart is shown in Fig 3.

Gather Statistics:

To interpret an image using a model, we must find the set of parameters which best match the model to the image. This set of parameters defines the shape and position of the target object in an image, and can be used for further processing, such as to make measurements or to classify the object. There are several approaches which could be taken to matching a model instance to an image, but all can be thought of as optimising a cost function. For a set of model parameters an instance of the model projected into the image can be generated. This hypothesis can be compared with the target image to get a cost function. The best set of parameters to interpret the object in the image is then the set which optimises this measure.

The cost function can be defined based on the statistics gathered from the training images. Particularly, the profile normal to each landmark across training images is scanned to a length of k pixels and the mean and covariance of the derivatives along the profile assuming a multivariate Gaussian distribution is stored. The Mahalanobis distance is used as the cost function during the test phase given this mean and covariance.

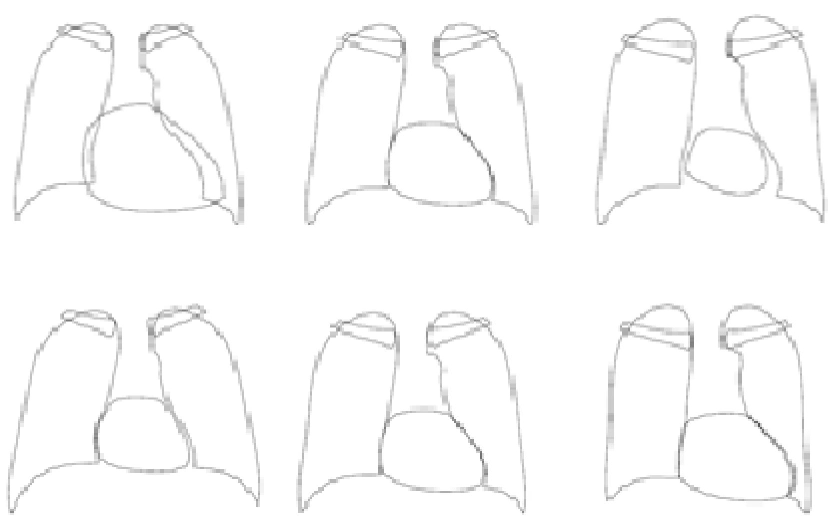

Fig 3: Generated shapes of anatomical structures in chest radiograph by varying shape parameters

Test Phase:

Here, initially the mean shape from the model space is projected to the image plane. If initial translation and rotation parameters are supplied, then the mean shape is accordingly realigned and projected on the image plane. Subsequently a decision on whether the model point needs to be moved to a new position or not is taken by evaluating the cost function described above by comparing the derivative profile along the normal to each of the current point to the already estimated ones using Mahalanobis distance. Subsequently the shape parameters are readjusted. This process continues until convergence. Fig. 4 shows an example of detection of anatomical structures in a chest radiograph.

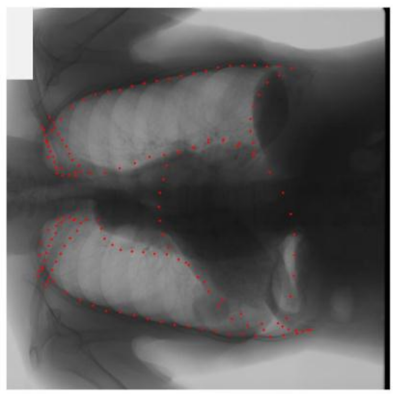

Fig 4: Result of ASM on a chest radiograph

Active Appearance Model:

In AAM, along with the shape the texture inside the shape is also modeled. This is also a generative model but not only generates shapes but synthesizes new images in photo realistic quality with shape and texture. The following are the steps in AAM:

- Build a shape model as in ASM
- Build a texture model
- Build a combined model
- Model Training
- Model Fitting

Since shape model is already described the other steps are briefly covered:

Build a texture model:

The basic idea is to build a texture model that will enable to synthesize new textures. First, all the texture vectors from the training set are aligned to the mean shape vector. That is the control points on each of the training data is warped to the control points of the mean shape. This is achieved by a deluanay triangulation, an inverse piecewise affine warping followed by bilinear interpolation. Forward warping is avoided as it will introduce holes. Bilinear interpolation is required as warping will not generally produce coordinates on integral grid of pixels. Then a histogram equalization is done on the training image. Subsequently texture vector is extracted and normalized using a photometric normalization method to make the algorithm robust to illumination variance. Then, similar to ASM, PCA is employed to build a texture model. Similar to shape parameter vector in ASM, a texture parameter vector of size n corresponding to n modes of variations (eigen vectors) is obtained. By changing these components within three standard deviations, textures can be synthesized.

Build a combined model:

To remove correlations between shape and texture model parameters a combined model can be built using PCA. The resultant model parameter is a single vector whose components when varied affects shape as well as texture. Compared to independent shape and texture models the combined model requires much lesser parameters to describe both shape and texture. But now shape and texture parameters are no longer orthogonal restricting the application of AAM for model fitting.

Model Training:

The AAM fitting process seeks to minimize texture residuals, i.e. the texture difference between the target image and the generated AAM model instance. This texture difference will drive the model on additive parameter update scheme to better estimate a new model instance until convergence. The training consists in learning the correlations between AAM model instances and texture residuals. In this stage, a set of experiences are required, perturbing a set of ground truth appearance parameters, that describe each model instance, by a known amount and recording the texture difference between the sampling image and the model. From this process results two huge matrices, one that holds the model perturbations and other matrix that keeps the texture residuals. With this information, two approaches to estimate a correction matrix could be applied: the Multivariate Linear Regression (MLR) and by differentiation, estimating the Jacobian matrix.

Model Fitting:

Fitting an AAM model to an image is a nonlinear optimization problem where the texture residuals are minimized by updating the model parameters. Several nonlinear solver methods can be used, such as Steepest Descent, Levenberg Marquardt, Powell's method, Genetic Algorithms, Simulated Annealing, Simplex, etc.

Aim and Objectives of this Dissertation Work:

Aim is to achieve a comprehensive understanding of ASM and AAM particularly focusing on the mathematical foundations behind these algorithms.

A few objectives towards realizing the aforementioned aim are listed as follows:

- Complete the mathematical derivations wherever it is incomplete in the original papers [1] and [2] covering ASM and AAM.
- Implement ASM and AAM in C with matlab wrappers.
- Apply the implementions to face recognition problem and detection of anatomical structures in medical images.

References:

[1]. T.F.Cootes, C.J.Taylor, D.H.cooper, J. Graham, "Active Shape Models-their training and application", Journal of computer vision and image understanding, vol 61, issue 1, pg 38-59, 1995.

[2]. Timothy F. Cootes , Gareth J. Edwards , Christopher J. Taylor, "Active Appearance Models", IEEE transactions on Pattern Analysis and machine intelligence, pg 484-498, 1998.