Muhammad Atif #127958

Yasir Hayat #124072

Ziyab Ikram #121923

Mujtaba Shahid Faizi #131818

**Problem:** Face alignment is a vision technology for identifying the geometric structure of human faces in digital images. Given the location and size of a face, it automatically determines the shape of the face components such as eyes and nose.

**Tool**: The tools we’ll be using are OpenCV+Python/C++.

**Intro**: The main aim of the project is an implementation of an excellent paper from this year’s Computer Vision and Pattern Recognition Conference: One Millisecond Face Alignment with an Ensemble of Regression Trees by Vahid Kazemi and Josephine Sullivan.

**Solution/Algorithm:** Using regression trees, we will construct a new algorithm performing face alignment in very less of a time (minimum goal being milliseconds) that achieves accuracy superior or comparable to state-of-the-art methods on standard datasets. The speed gains over previous methods is a consequence of identifying the essential components of prior face alignment algorithms and then incorporating them in a streamlined formulation into a cascade of high capacity regression functions learnt via gradient boosting.

Face alignment can be solved with a cascade of regression functions. Each regression function in the cascade efficiently estimates the shape from an initial estimate and the intensities of a sparse set of pixels indexed relative to this initial estimate. learnt regression functions carry two key elements that are present in several of the successful algorithms cited and we detail these elements in following.

The first revolves around the indexing of pixel intensities relative to the current estimate of the shape. The extracted features in the vector representation of a face image can greatly vary due to both shape deformation and nuisance factors such as changes in illumination conditions. This makes accurate shape estimation using these features difficult. The dilemma is that we need reliable features to accurately predict the shape, and on the other hand we need an accurate estimate of the shape to extract reliable features. We will use an iterative approach to deal with this problem. Instead of regressing the shape parameters based on features extracted in the global coordinate system of the image, the image is transformed to a normalized coordinate system based on a current estimate of the shape, and then the features are extracted to predict an update vector for the shape parameters. This process is usually repeated several times until convergence.

The second considers how to combat the difficulty of the 1 inference/prediction problem. At test time, an alignment algorithm must estimate the shape, a high dimensional vector, that best agrees with the image data and our model of shape. The problem is non-convex with many local optima. Successful algorithms handle this problem by assuming the estimated shape must lie in a linear subspace, which can be discovered, for example, by finding the principal components of the training shapes. This assumption greatly reduces the number of potential shapes considered during inference and can help to avoid local optima. Recent work uses the fact that a certain class of regressors are guaranteed to produce predictions that lie in a linear subspace defined by the training shapes and there is no need for additional constraints.

**Applications:**

* Training face landmark detector
* Face landmark detection in an image
* Face landmark detection in a video
* Face swapping

**Conclusion:** Crucially, our regression functions have the two elements described in the solution part of the proposal. Allied to these two factors is our efficient regression function learning. We optimize an appropriate loss function and perform feature selection in a data-driven manner. We learn each regressor via gradient boosting with a squared error loss function, the same loss function we want to minimize at test time. The sparse pixel set, used as the regressor’s input, is selected via a combination of the gradient boosting algorithm and a prior probability on the distance between pairs of input pixels. The prior distribution allows the boosting algorithm to efficiently explore many relevant features. The result is a cascade of regressors that can localize the facial landmarks when initialized with the mean face pose.