Zac Downs, MSc Natural Computation, msc22zpd@cs.bham.ac.uk

Face Detection & Recognition

Introduction

The first of three steps in identifying faces in images is to localise them with a detector. This allows certain features to be extracted for classification, to determine which people are in the image.

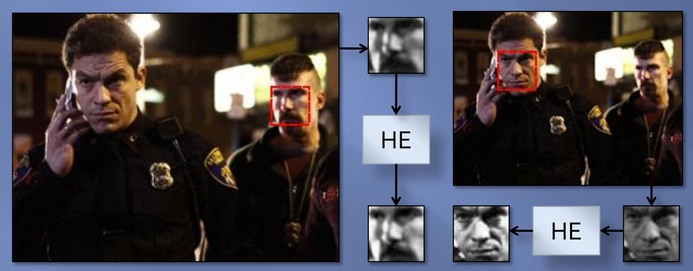
Three stage automated system for identifying faces in images

Face Detector

The face detector uses a feedforward neural network trained with normal error backpropagation. The neural network is trained to classify a 20 x 20 pixel greyscale image as either a face or a non-face.

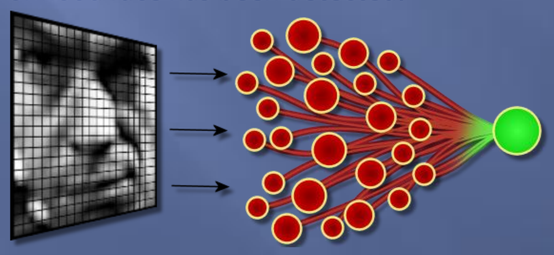
Example faces and non-faces

A 20 x 20 pixel window is scanned across the input image at various different scales, and each window is pre-processed with histogram equalisation (HE) to enhance contrast.



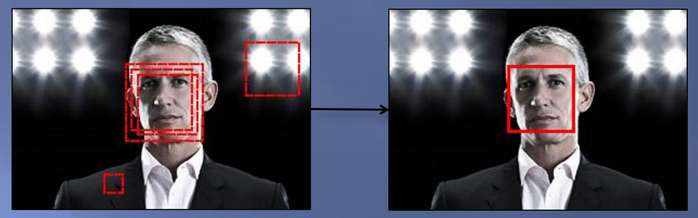
Scanning at multiple resolutions, with histogram equalisation

Each window is fed into a neural network with a single layer of hidden neurons, and a single output neuron to indicate whether or not a face has been detected.



Feedforward neural network with 25 hidden layer neurons

By grouping detections together based on size and location, a threshold can be applied in combination with removal of overlapping detections to reduce false positives and increase performance.



Similar detections are merged and thresholded

Feature Extractor

The 2-D Discrete Cosine Transform (DCT) is used to extract frequency features from detected face locations. By ignoring the majority of the highest frequency coefficients, insignificant details can be ignored and dimensionality is reduced, in much the same way as in JPEG compression. Additionally, ignoring a few of the lowest frequency coefficients provides invariance to uniform and linear changes in brightness, which eliminates confusing shadows and improves classification performance.

Example 8x8 DCT basis functions, and DCT visualisation for a face image

It is possible to identify a face with only a very small number of these low frequency coefficients, with the literature reporting top performance with as few as 0.2% coefficients selected. The lowest frequency coefficients are located in the upper left corner of the 2-D matrix, as illustrated by the highlighted red area.

Neural Network Classifier

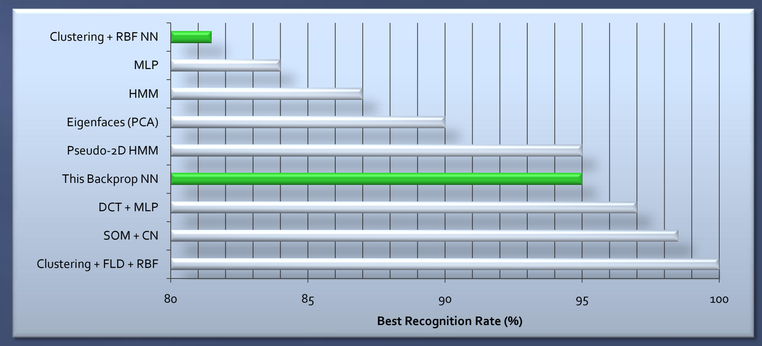
Another neural network is trained to classify individuals, with an output neuron to represent each person, and a single hidden layer of neurons. The input DCT coefficients are normalised between -1 and +1, and the neurons in the network use a sigmoid activation function.

RBF Classifier

In addition to the backpropagation neural network classifier, a radial basis function (RBF) neural network can also be used. Training data is first clustered using a supervised technique that splits the data classes based on their Euclidian distances. This creates subclasses that reduce variation between the data in each class. In other words, the training images of each person are put together in subgroups, which makes parameter estimation for the RBF neural network easier.

Recognition Performance

Two different databases are used to test the performance of the feature extractor and classifiers in their ability to identify faces. The ORL Database Of Faces is used by many researchers, and contains slight variations in pose as well as illumination variations. The two classifiers are shown below compared to a number of other implementations in the literature, all tested on the ORL database. The backprop neural network in particular performs well, achieving a 95% correct recognition rate for 40 individuals in 200 images. This is achieved by retaining just 40 in 10,000 DCT coefficients for each face image.



The second database used was created by using the face detector to automatically compile a set of test and training images for 30 famous people. The database contains difficult images, representative of real-world challenges. Faces are correctly identified in more than 50% of cases.

Example test images from custom database (Schwarzenegger & Blair)