Learning-based Face Synthesis for Pose-Robust Recognition from Single Image

Akshay Asthana[1]
aasthana@rsise.anu.edu.au
Conrad Sanderson[2]
conradsand@ieee.org
Tom Gedeon[1]
tom.gedeon@anu.edu.au
Roland Goecke[1,3]
roland.goecke@ieee.org

[1] RSISE and SoCS, CECS
Australian National University, Australia
[2] NICTA and University of Queensland, Australia
[3] Faculty of Information Sciences and Engineering,
University of Canberra, Australia

A major challenge for automatic face based identity inference is the shear magnitude of uncontrolled factors than can result in a considerable change of shape and appearance, such as expression, illumination and pose. The problem is complicated further when only one image per person is available for training. In this paper we present an approach that uses a data driven and computationally efficient 2D technique for the synthesis of high quality non-frontal faces from a single frontal input face, that be used for extending the training set for each person, enabling 2D face recognition algorithms to improve their performance when dealing with pose variations.

We use offline AAM fitting to obtain the locations of landmark points in a given frontal face image. Once the locations of the landmark points from the gallery images have been extracted, we use a regression-based approach to generate synthetic images at various poses. We first learn the correspondence between the landmark points of the set of frontal images exhibiting arbitrary facial expressions and their corresponding non-frontal images at arbitrary pose. Once this learner has been trained, the synthetic image of any other *unseen* face can be generated by predicting the locations of the new landmark points, followed by warping of the texture from the original frontal image.

We start by extracting 3 vectors: *Normalisation*, *Centroid*, and *Point Vector* from each of the m frontal and non-frontal training images.
Normalisation Vector is a 1D vector containing information about the normalisation distances used to normalise the feature vectors.

$$\mathbf{N} = [N_h; N_v]^T \qquad \mathbf{N}' = [N'_h; N'_v]^T$$

Centroid Vector is a 1D vector containing the location of the centroids of six individual facial features (left and right eyebrows, left and right eyes, nose and mouth) in the normalised frame.

$$\mathbf{C} = [\vec{x}_1; \vec{y}_1; \ldots; \vec{x}_6; \vec{y}_6]^T \qquad \mathbf{C}' = [\vec{x}'_1; \vec{y}'_1; \ldots; \vec{x}'_6; \vec{y}'_6]^T$$

Point Vector is a 1D vector containing the location of each of the n landmark points in the normalised frame.

$$\mathbf{P} = [x_1; y_1; \ldots; x_n; y_n]^T \qquad \mathbf{P}' = [x'_1; y'_1; \ldots; x'_n; y'_n]^T$$

Next, we construct 3 different training sets

$$\mathscr{T}_{\mathscr{N}} = \left\{ (\mathbf{N_i}, \mathbf{N'_i}) \right\} \;\; \mathscr{T}_{\mathscr{C}} = \left\{ (\mathbf{C_i}, \mathbf{C'_i}) \right\} \;\; \mathscr{T}_{\mathscr{P}} = \left\{ (\mathbf{P_i}, \mathbf{P'_i}) \right\} \; ; \; i \in (1, .., m)$$

and train a learner via regression to learn 3 different sets of regression models $\mathscr{R}_{\mathscr{N}}$, $\mathscr{R}_{\mathscr{C}}$, $\mathscr{R}_{\mathscr{P}}$ for predicting the normalisation, centroid and point vector respectively, where

$$\mathscr{R}_{\mathscr{N}} = \left\{ R_{N_h}, R_{N_v} \right\} \;\; \mathscr{R}_{\mathscr{C}} = \left\{ R_{C_i} \,|\, i \in (1, .., 12) \right\} \;\; \mathscr{R}_{\mathscr{P}} = \left\{ R_{P_i} \,|\, i \in (1, .., 2n) \right\}$$

Here, R represents the regression model learnt over a particular training set. Given an annotated frontal face image, a virtual image is reconstructed by predicting the new landmark locations and warping the texture from the frontal image via Piecewise Affine Warping (PAW). Figure 1 shows synthetic images constructed for a subset of the gallery images. The high quality of the constructed synthetic images is maintained throughout by virtue of the non-linear predictor utilised by the regression-based learning approach to predict the new locations for the landmark points.

We used the CMU PIE database for training the regression model and the b subset of FERET database to evaluate the usefulness of the synthetic images in a face recognition scenario. To objectively verify the quality of the synthesised images, we augmented two frontal face recognition systems: a baseline PCA-based face recogniser and the recently proposed Multi-Region Histograms (MRH) based face recogniser, by extending the single image training set for each person with synthesised non-frontal faces. Experiments on the FERET dataset show that the robustness of a baseline PCA approach as well as the MRH method can be considerably increased when dealing with faces at $\pm40°$ and $\pm60°$ views, while maintaining high recognition rates for $\pm15°$ and $\pm25°$ views. Moreover, the approach is 2D, entirely data-driven, computationally inexpensive and can easily be combined with other face recognition algorithms than the ones shown here without requiring any major modification.

We are currently working on extending the approach into the 3D domain to enable better handling of occlusions as well as the construction of synthetic faces at extreme poses (such as profile views).

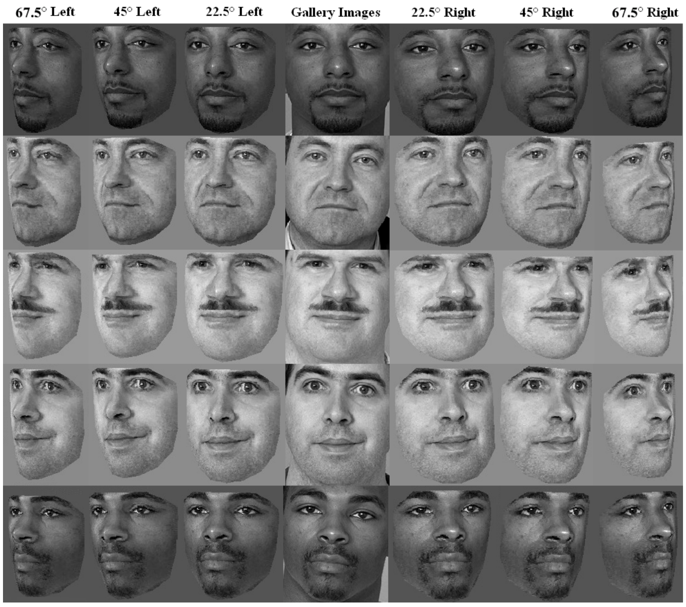
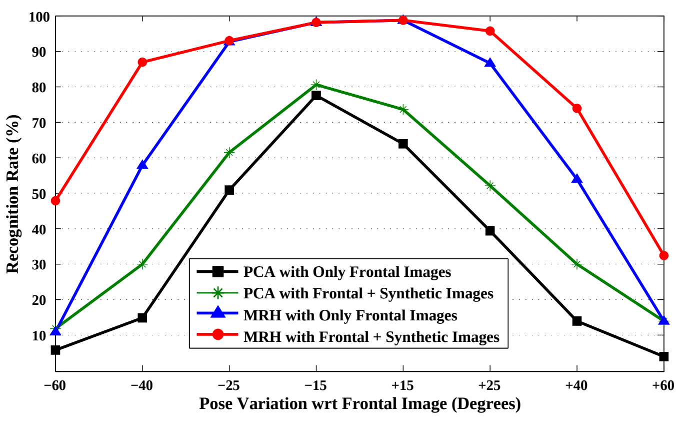


Figure 1: Synthetic images for sample gallery (frontal) images



(a) Rank-1 recognition rates

Figure 2: Comparison of recognition rates obtained by PCA and MRH face recognition systems with and without the use of synthesised faces.