Introduction

Problem:
How can we build a statistical deformable model (e.g. Active Appearance Model (AAM)) if we only have frontal face images of a person?
- Statistical approaches of non-rigid deformable models (ASM, AAM, 3DMM) popular
- Require a labelled dataset of training images → tedious manual annotation
- Propose an approach for automatic annotation of face images at any arbitrary pose and expression that only requires annotated frontal images
- Firstly, learn the correspondence between manually annotated landmark points of frontal and varying viewpoint images of a face
- Secondly, synthesize images of any arbitrary pose and expression from frontal view images
- Thirdly, create View-Based AAMs using the Simultaneous Inverse Compositional fitting method and apply to automatically annotate unseen images

Learning a Regression Model of Correspondences

- Require a labelled dataset of training images
- Determine the normalisation, centroid and point vector, resp., with each pair representing a gallery and corresponding probe image
- Next, construct 3 different training sets:
  \[ T_N = \{ (N_1, N_1') \} \quad T_C = \{ (C_1, C_1') \} \quad T_P = \{ (P_1, P_1') \} \quad \text{where } i \in \{ 1, 2, \ldots, m \} \]
- Train a learner via regression (Support Vector Regression (SVR), Boosted SVR, Gaussian Process Regression (GPR)) to learn 3 different sets of regression models \( R_N, R_C, R_P \) for predicting the normalisation, centroid and point vector, resp.:
  \[ R_N = \{ R_{N_1}, R_{N_1'} \} \quad R_C = \{ R_{C_1}, R_{C_1'} \} \quad R_P = \{ R_{P_1}, R_{P_1'} \} \quad \text{for } i \in \{ 1, 2, \ldots, m \} \]

Generation of Virtual Images

- Given an annotated frontal image, virtual images with the same pose as the probe images are generated by predicting the new landmark locations via the learnt regression models and warping the texture from the frontal image via Piecewise Affine Warping

Experimental Results

1. CMU PIE database subset
- Regression models for generation of virtual images were trained for 6 different poses (4 per viewpoint: Up and Down, 22.5° Left and Right, 45° Left and Right)
- Four View-Based AAMs were created \( (\text{L1}, \text{L4}, \text{R12.5} \text{ and R45}) \) and used to automatically annotate unseen images
- Compared automatic annotations with manual annotations (‘ground truth’) → GPR performed best, average pixel error \( \pm 2\text{mm} \)

2. Evaluate generalisability of the framework on Face Pointing and FERET databases, using the regression models learned for CMU PIE database
- Similar results for Face Pointing and FERET databases, highlighting the good generalisability of the proposed framework

Automatic Face Annotation for Arbitrary Poses and Expressions from Frontal Images Only

Akshay Asthana\(^1\), Roland Goecke\(^1,2\), Novi Quadrianto\(^3\), Tom Gedeon\(^4\)
\(^1\)Australian National University, Canberra, Australia \(^2\)University of Canberra, Australia \(^3\)NICTA, Canberra, Australia
Email: aasthana@ruse.anu.edu.au, roland.goecke@canberra.org, novi.quadrianto@gmail.com, tomgedeon@anu.edu.au

Figure 1: Overall architecture of the Auto Annotator

Figure 2: Extracting Normalised, Centroid and Point Vectors
- **Point Vector**: Location of each of \( n \) landmark points in normalised frame:
  \[ P = [x_1, y_1; \ldots; x_n, y_n]'^T \]
- **Normalisation Vector**: Distances used to normalise the feature vectors
  \[ N = [N_1; N_1']^T \]
- **Centroid Vector**: Location of centroids of six individual facial features (left / right eyebrows, left / right eye, nose, mouth) in normalised frame
  \[ C = [c_1; c_2; \ldots; c_6; c_0]'^T \]

Figure 3: Step-by-step reconstruction of virtual image
- Using Simultaneous Inverse Compositional fitting in a view-based AAM setting - to overcome poor initialisation problems - build AAMs for automatic annotation of unseen face images

Figure 4: Experimental Results
- Regression errors for different viewpoints for annotation of CMU PIE dataset
- Regression errors for CMU PIE dataset with different viewpoints

Figure 5: Generated virtual images. Column: Different poses. Rows: Samples from different databases with arbitrary facial expressions (1-3 CMU PIE, 4-5 Face Pointing, 6-7 FERET)