Face tracking using an extended Active Appearance model on Time-of-Flight data

DTU Vision days, 28-30 May 2008.

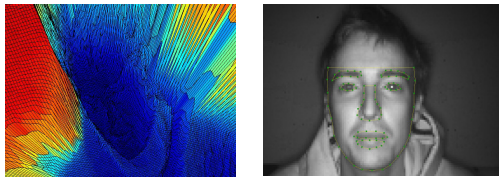
Abstract

This project describes statistical methods for modelling the shape, texture and appearance of human faces using three-dimensional *Time-of-Flight* data. The traditional 2D Active Appearance Model is implemented and extended to incorporate a model of the depth.

Data

The data was recorded using a *Swissranger SR-3000* Time-of-Flight camera and three datasets was created:

1. Casual talk
2. Wild gesticulate
3. Lots of movement – to be used for later tracking.



Building the model: To the left a depth image is shown and to the right the corresponding amplitude image. Both types are generated by the TOF-camera

Upper row contains example images from the first dataset. Lower row is the second dataset, which contains quite a lot more variation.

Building the Model

Three separate models was created:

- Shape model
- Texture model
- Depth model

Principal Component Analysis is applied to each of the models to generate a parameterized model with a reduced dimensionality. Each model is of the form:

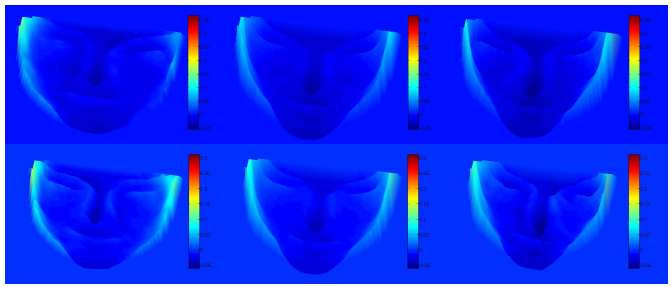
$$x \approx \bar{x} + \phi b$$

where x is either pixel coordinates, texture intensities or depth information. Before the models are created the shape-data is aligned using *Procrustes* analysis and the texture/depth are considered in a *shapefree* context.

The *combined appearance model* incorporates all three models into one common model. A PCA is applied to the following vectors of combined parameters:

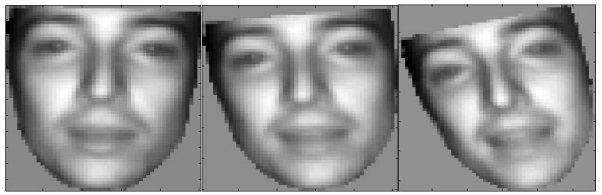
$$b = \begin{bmatrix} W_s b_s \\ W_t b_t \\ W_z b_z \end{bmatrix}$$

where the W's are scaling factors since the three models are measured in different units.



Up: The first principal component variation of the depth model.

Down: The second principal component variation..



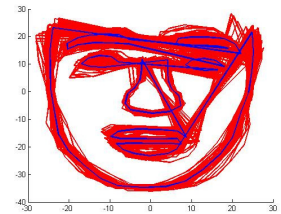
An example of the combined appearance model of shape and tezture.. This figure shows variations of the first. principal component. To the left is -3 standard deviations and to the right +3. The middle image is the mean.

The *combined model* is then:

$$b = \phi_c c$$

where c is controlling both shape, texture and depth.

Finally an *Active Appearance Model* is implemented to locate the model that best fits a new image. The goal is then to examine if the depth information gives better segmentation results.



The Procrustes aligned shapes shown red and the Procrustes meanshape in blue.

Further work

The Active Appearance Model is to be fully implemented and the impact of the depth information are going to be quantified. If time allows it a larger database of several persons of different gender and looks should be considered. Furthermore a comparison of the AAM and the new Constrained Local Models(CLM) could be interesting.

Informatics and Mathematical Modelling

DTU

Author: Christian Hjorth Møller (s022054)