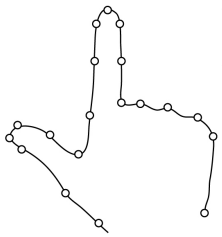
Hand Tracking - *Using Active Shape Models*

Using the hand for pointing is one of the most intuitive acts for humans and is used as a tool for communication and interaction with the surroundings. By processing images of a hand in front of a monitor it is possible to determine the position of the hand and hereby replacing the touch screen with a cheap substitute: ***The Poor Man's Touch Screen***.

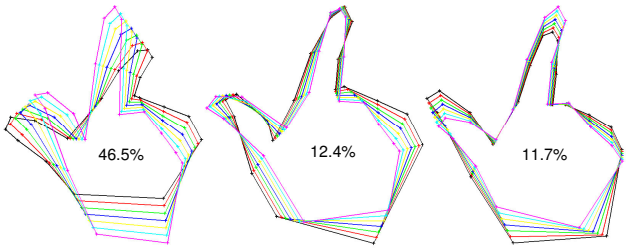
We have implemented an application in C++ that using Active Shape Models is able to track a hand through a sequence of greyscale images.

Data Pre-processing



Training images are annotated using 19 landmarks creating a shape space of 38 dimensions.

By using Principal Component Analysis the dimensionality is reduced. 95 percent of the variation can be described by the first 12 Principal Components. Each Principal Component represents a certain amount of the total variation.



Here is shown the first 3 Principal Components varying in the interval of $\pm 3\sigma$. The Statistical Shape Model is now represented by a mean shape $\mathbf{x}_{mean}$, a set of eigenvectors Φ and eigenvalues λ.

Fitting To Image Data

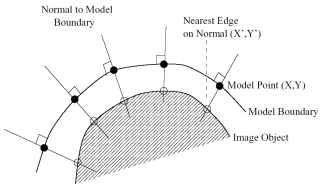
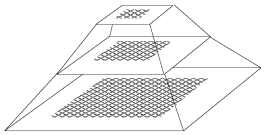


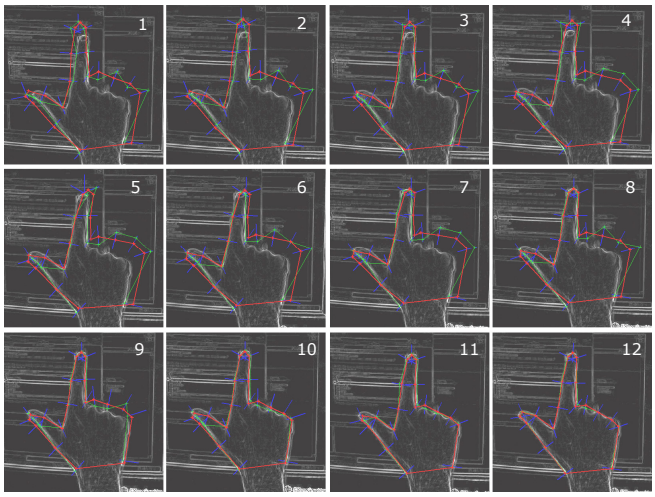
Image samples are gathered along the normals of each landmark of the model.
The maximum gradient along the normals is used as fitting measure.

To increase the effectiveness and robustness of the Active Shape Model the fitting is done in multi-resolution. Starting at low-resolution the fitting precision increases as we go through each level of the image pyramid, finalising by fitting to the original image.



The shape of the model is varied by changing the pose parameters: translation, scale and rotation, as well as the shape parameter, **b**.
$\mathbf{x} = \mathbf{x}_{mean} + \Phi \cdot \mathbf{b}$

To the right the fitting process is shown. The green outline is the result of moving the landmarks to the maximum gradient along the normals (drawn in blue). The red outline shows the model projected back into shape space.



Tracking a Hand

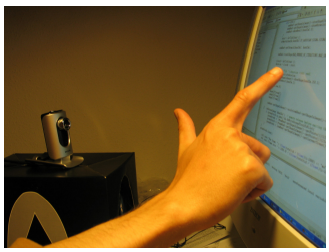
Having implemented the application in C++ it is possible to process an image in 60 milliseconds using a 2.0 GHz PC. This gives a frame rate of approximately 17 frames per seconds.

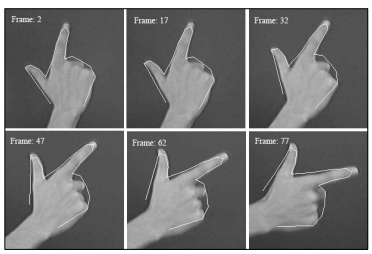
When tracking the hand through an image sequence the fitting process is repeated for each frame. As an initial pose for each frame is used the final pose from the previous frame.

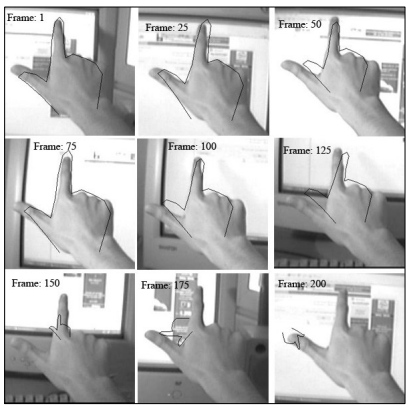
Tracking on a clutter free background the application shows good results. The deviation from ground truth is on average less than 10 percent of the length of the index finger.

However, the application shows some weakness when tracking with a monitor in the background. As is seen from the image sequence to the right the shape model eventually loses track of the hand.

This flaw might be dealt with by for instance extending the application to analyse colour images…







Andreas Rishede Hyllested (arh@itu.dk)
Martin Wallengren Nilsson (mwn@itu.dk)

02503, Advanced Image Analysis, DTU

This project is primarily based on the work of: T.F. Cootes and C.J.Taylor, Statistical Models of Appearance for Computer Vision. Imaging Science and Biomedical Engineering, University of Manchester, October 26, 2001.