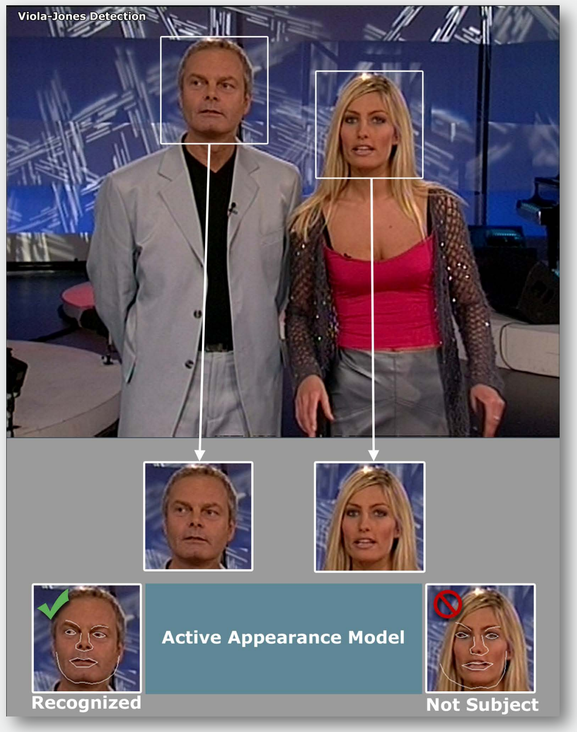
DTU Informatics
Department of Informatics and Mathematical Modelling

Face Detection and Recognition in Video-Streams

Jannik Boll Nielsen and Rasmus Larsen

Using a combination of the Viola-Jones face detection algorithm and Active Appearance Models, faces are extracted from video footage and recognized with high success-rates. This allows one to browse through large video-collections in order to recognize a given subject.



Recognition of the Danish actor Jarl Friis-Mikkelsen. Using the Viola-Jones detection algorithm, faces in the video-footage are detected. A set of snapshots are taken for each face, representing the entire time span in which the face is present. The snapshots are then matched to an Active Appearance Model describing the actor. Based on the quality of the match, i.e. the Mahalanobis distance, one is then able to determine if the actor is present in a given snapshot or not.

Introduction

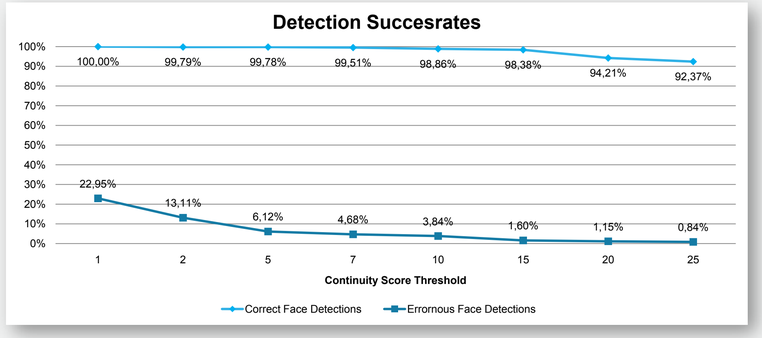
A lot of video footage exists in archives all over the world, much of it containing people. The footage is however not always properly documented and does therefore often not contain proper information about which people have been filmed. As the contents of video archives are ever increasing, it is needed to find ways of annotating relevant people included in the many footages. One way of dealing with this problem is by letting modern computers use a combination of the Viola-Jones face-detection algorithm and Active Appearance Models in order to match premade models of relevant persons to any piece of video footage.

Continuity in the Viola-Jones detections.

Even though the Viola-Jones algorithm is a very strong classifier, it does return an unwanted amount of erroneous detections from the video frames. In order remove these errors, the fact that we deal with a video sequence and not a still image is used. By only snapshotting detections that are continuous over time (frames), position and size, a very large amount of the wrong detections can be filtered out, while still maintaining a very high detection rate.

Building Active Appearance Models.

An active appearance model describing the subject being searched for, can easily be built from a set of images containing the subject's face with various expressions. One must ensure that a large amount of facial variety is apparent in the image set. The easiest way to obtain such a set, is by extracting faces from already known video footage of the subject. A model trained from no more than 30 images proved very successful .

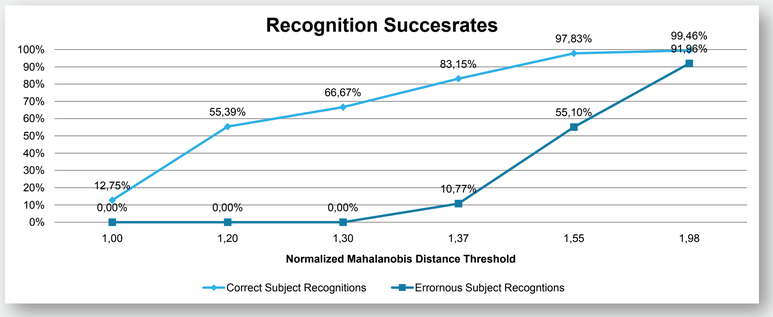


Implementation of continuity filtering: The continuity score indicates how many detections must have been present in order to classify a detection as a face. With a score around 15, most errors are removed while still maintaining a very high detection rate. The statistics are based on approximately 500 frames.

Recognition Results

As can be seen in the figure below, the recognition success-rates proved very good. Tests proved, per-frame, a successful recognition rate of up to more than 60%, with a 0% erroneous detection-rate. This resulted in per-video-recognitions with a very high recognition rate (>90%).

This method of recognizing using Active Appearance Models looks very promising. There is however one issue that needs to be solved: with a standard computer, 1 minute of interview-like footage, takes around 2-3 minutes of processing time. Being able to optimize this processing time would greatly improve this solution.



Matching using AAMs: Matching an Active Appearance Model describing the subject on face-snapshots, outputs a Mahalanobis distance, in this case normalized (divided by the number of parameters). Setting a threshold allows one to successfully distinguish persons in footage. These rates are per frame based. Integrating over entire video-sequences results much higher success rates. Statistics are based on approximately 1000 faces.

P. Viola and M. Jones. Robust real-time face detection. International Journal of Computer Vision (IJCV) 57(2) 137-154, 2004.
M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. IEEE T on Medical Imaging, 22(10):1319-1331, 2003
T.F.Cootes, G.J. Edwards and C.J.Taylor. "Active Appearance Models", IEEE PAMI, Vol.23, No.6, pp.681-685, 2001

DTU Informatics
Department of Informatics and Mathematical Modelling