Face Feature Localization through Statistical Models and Artificial Neural Networks

Ilaria Castelli

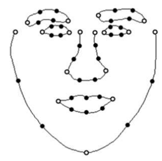
DII - University of Siena
Via Roma, 56 - 53100 Siena - Italy
castelli@dii.unisi.it

Description

Automatic face analysis is a challenging task that involves both machine learning and computer vision techniques.
Statistical shape and appearance models are particularly interesting because of their ability to provide a compact and informative description of the face. Moreover, they can be exploited for localization purposes. My major contribution is an improved version of Active Shape Models (AMSs) algorithm for face feature localization, that overcomes some limitations of the original technique.

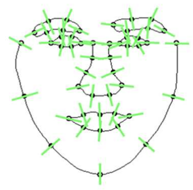
Face representation

In order to represent faces we use a set of 49 landmark points extracted by means of Catmull-Rom splines (CRSs). This technique allows us to set up, for each face feature, the number of landmark used to represent it.



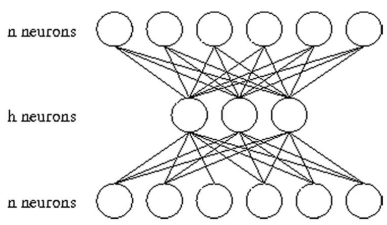
Active Shape Models

ASMs exploit both a point distribution model and a local feature model for each landmark. The shapes in the training set are aligned in a common coordinate frame and PCA is applied in order to define a generative deformable model. Thus, a shape f can be approximated as $f \approx f_m + \Phi b$ where f_m is the mean shape, Φ collects the leading eigenvectors of the covariance matrix and $b = \Phi^{-1}(f - f_m)$.
The face feature localization is performed perturbing an initial shape in order to find a better position for each landmark. Local feature models based on the gradients of the luminance along the normals to the face feature contours are exploited. During the localization phase, for each landmark a set of luminance profiles are sampled along the normal to the shape boudary, and the point corresponding to the lowest Mahalanobis distance from the model is choosen as the new landmark position.



Auto Associative Neural Networks

We propose to use Auto Associative Neural Networks (AANNs) in order to determine the optimal landmark positions, due to their proved ability to solve classification tasks that are characterized by a well-defined class (feasible positions) and by a “complement” class (unfeasible positions) whose data are particularly heterogeneous.
An AANN consists in a feed-forward multilayered perceptron, where the number n of output neurons equals the number of input neurons. The hidden layer contains $h < n$ neurons.
The Neural Network weights are adapted using the BackPropagation algorithm with the cost function



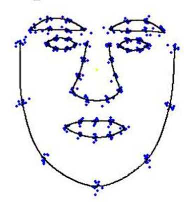
AANN architecture

$$E = \sum_{k=1}^{P} \|U_k - O(U_k)\| + \sum_{k=1}^{N} \frac{1}{\varepsilon + \|U_k - O(U_k)\|}$$

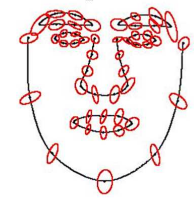
where U_k is the k-th input vector, $O(U_k)$ is the corresponding output vector, P and N are the number of positive and negative examples, respectively.
In order to classify a pattern, an input vector U is fed to the input units. Once the output vector $O(U)$ is available, the Eucledian distance between the input and the output is computed and compared with a threshold in order to classify the input pattern.

AANN-ASMs

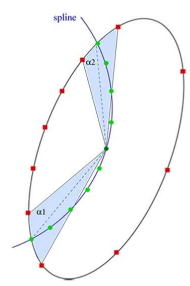
We exploit elliptical research areas, whose dimensions depend on the training data. At the end of the alignment process, the landmark positions create a set of clusters, roughly centered around the mean of each group. We can consider the landmark positions as bivariate Gaussian distributions, allowing the definition of elliptical research areas.



Landmark clusters



Estimated ellipses



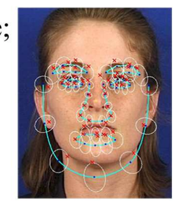
Target association

The landmark positions are represented by means of the Haar-like features defined by Viola and Jones.

Each training example has an associated target extracted exploting CRSs. In particular:
- positive examples are sampled on the part of spline that lies inside the research area;
- negative examples are sampled on the research area contour.

AANN-ASM algorithm

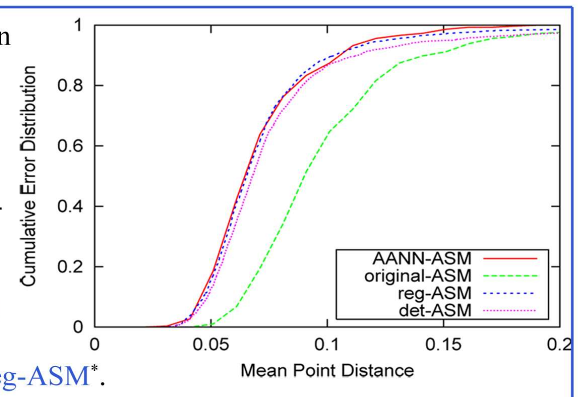
1) roughly locate the face region and place f_m on the image;
2) for each landmark, a certain number of points are sampled inside the elliptic research area and, for each point, the Haar-like features are extracted;
3) each sample is processed by the corresponding AANN and the point that obtain the lowest distance is chosen as the new position;
4) the shape is fitted to the candidate landmark positions;
5) the algorithm is repeated from step 2) until the convergence is reached.



AANN predictions (1st iteration)

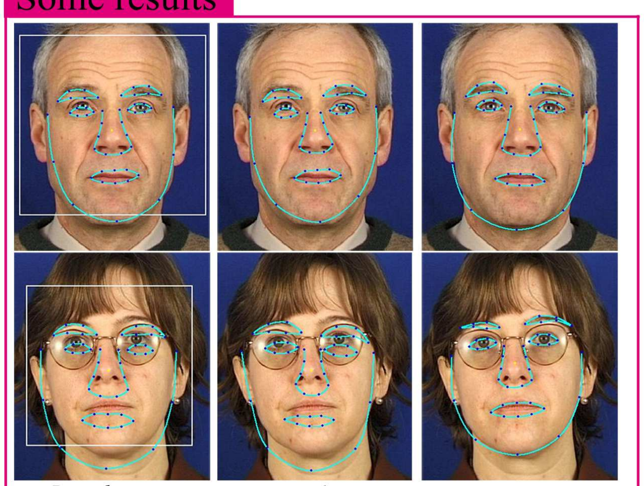
Performances

The experimental evaluation has been carried out on 295 images extracted from the XM2VTS dataset. The *mean point distance* metric has been used to evaluate the performances of our method. The proposed technique outperforms both original ASMs, det-ASM*, and it is comparable with reg-ASM*.



*D.Cristinacce and T.Cootes. Boosted Regression Active Shape Models. In18th British Machine Vision Conference, Warwick, UK, pages 880-889, 2007.

Some results



Initialization *Original ASM* *AANN-ASM*

Future work

Although the research community gave increasing attention to the field of face analysis in the last years, robust approaches to this task are still missing. Moreover, this is a very large field that involves many subproblems, such as localization, recognition, head pose estimation, 3D model reconstruction, etc.
Face recognition and identification are particularly interesting because of their application in security systems.
Anyway, scientific literature currently shows that there is a lack in the ability to recognize faces when the dimension of the dataset of known subjects becomes large.

This observation leads to some suggestions for future work:
- the need to explore new face representation spaces, able to catch peculiarities that guide the human cognitive and perceptive recognition mechanisms;
- the need to study classification systems able to deal with hundreds of different classes.