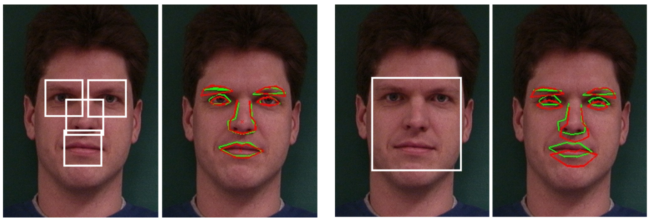
Face Alignment with Part-Based Modeling

Vahid Kazemi
vahidk@nada.kth.se
Josephine Sullivan
sullivan@nada.kth.se

CVAP
KTH Institute of Technology
Stockholm, Sweden

We propose a new method for accurate face alignment with part-based modeling. Although we focus on the class of human faces, this method could potentially be applied to any type of deformable object. We aim to learn a regression function mapping a feature representation of the appearance of the face to its shape represented by a set of connected landmarks forming contours around the major facial features. Ideally, we want to use linear regression functions to describe this mapping as they need less training data, have a lower chance of over-fitting, and are fast to compute. However, the relation between the global appearance of an object and its shape is highly nonlinear. Previously, piece-wise linear regression has been applied [8] to deal with this non-linearity. Instead, in this work, the strategy is to learn regression functions for individual parts, see figure 1. The parts are chosen to ensure the linearity of the regression mapping. The main advantage of this approach is that it requires less training data, although it necessitates good part-detection. Part-based methods, mostly



(a) with parts (b) without parts

Figure 1: This figure shows the benefit of using parts in the performance of a regression function to accurately predict the location of landmark points. In both figures green lines represent the ground truth shape, and the red lines are the result of regression function with (a) and without (b) using parts. In both cases a linear regression model is learnt to map the appearance descriptors inside the patches to the location of the landmarks.

used in recognition tasks, train different classifiers for each part of the model. Assuming each part is a rigid structure, deformation is limited to the relative transformation of each part. The optimal solution for such a multi-part based matching problem can be found efficiently by simplifying the dependency of parts to form a tree structure as is presented by Felzenszwalb et al. [5]. This method cannot be directly applied to solve dense matching problems because it requires creating individual classifiers to detect each landmark which is not only computationally expensive but impractical, since most of the landmarks don't have a distinctive local appearance.

A large group of methods have been developed which are relying on the global appearance of a deformable object to tackle the alignment problem, these include Active Shape Models(ASM) [2], and Active Appearance Model(AAM) [3]. Some attempts have been made to improve the robustness, and accuracy of these methods [1, 6, 7], but the main problem which remains unsolved in these methods is that they need a good initial estimate and are not able to adapt the model to fit a subject when the initial error is high. As an example in face tracking applications, AAM usually fails to converge when there is a sudden large deformation or motion of the subject.

One main advantage of our method compared to the global methods just described is that the regression functions directly estimate the landmark positions by-passing the need to perform iterative non-linear optimization on a complicated cost function. As a result, our method can be used on image sequences with fast motion, and it does not need any initialization. Furthermore the part-based nature of our method, and also the use of HOG descriptors [4] (in contrast to intensity vectors as in AAM), enhance the robustness and generalization ability of our algorithm. On the minus side the performance of our method is highly dependent on a proper part detection and this requires strong and distinctive features in the appearance of the object. In the case of a human face which is the

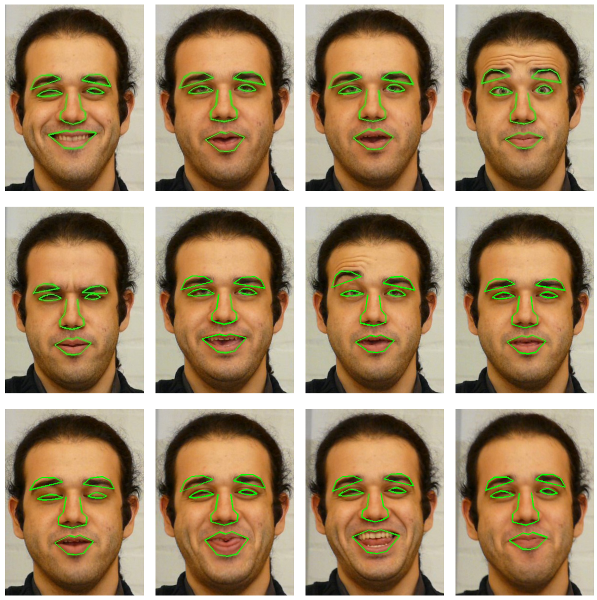


Figure 2: The result of our method, trained using IMM dataset, on an unknown subject, with unknown expressions. The results show that our model covers unseen faces with novel expressions and lighting conditions.

focus of our work, a sensible selection of parts is a division into the nose, mouth and the eyes. In a more general case we would need an automatic part selection method which can be investigated in a future work.

In our experiments we have used the IMM face database [9] to train our model, which contains 240 still images of 40 different human faces. We have tested the model on various image sequences and an example is demonstrated in figure 2. The results of our experiments prove that our method has a good generalization ability and is competitive in terms of precision with global methods such as Active Appearance Models.

As mentioned before the choice of parts is a critical factor in the performance of the algorithm. Further improvement could be achieved by designing an automated method to find the best part configuration.

[1] B. Amberg, A. Blake, and T. Vetter. On compositional image alignment, with an application to active appearance models. CVPR, 2009.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models, their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. ECCV, 1998.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. CVPR, 2005.

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1): 55–79, 2005.

[6] X. Hou, S. Z. Li, H. Zhang, and Q. Cheng. Direct appearance models. CVPR, 2001.

[7] Y. Huang, Q. Liu, and D. N. Metaxas. A component-based framework for generalized face alignment. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(1):287 –298, 2011.

[8] R. Okada and S. Soatto. Relevant feature selection for human pose estimation and localization in cluttered images. ECCV, 2008.

[9] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. Fame - a flexible appearance modelling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319–1331, 2003.