Improved 3D Model Search for Facial Feature Location and Pose Estimation in 2D images

Angela Caunce
angela.caunce@manchester.ac.uk

Chris Taylor
chris.taylor@manchester.ac.uk

Tim Cootes
Tim.cootes@manchester.ac.uk

Imaging Science & Biomedical Engineering,
The University of Manchester
Manchester M13 9PT, UK

This paper tackles the problem of accurately matching a 3D deformable face model to images in challenging real-world scenarios with large amounts of head movement, occlusion, and difficult lighting conditions (Figure 2). A baseline system [1] based on the ASM [2] involves searching with a set of view-dependent local patches to locate image features and using their positions to update the face shape model parameters (Figure 1). We show here two modifications that lead to improvements in performance and can be applied in other similar systems. These are: a simple method for weighting the relative importance of each located match for model fit; and explicitly searching for occluding boundaries, which prevents the model from shrinking or rotating rather than changing shape. We demonstrate the improvements on both standard test sets and on a series of difficult in-car driver videos.

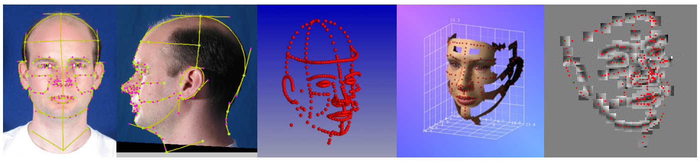

Figure 1: The model is built using manual markups. Each subject is marked in front and profile to create a 3D point set. A mesh is warped to match each subject and a subset of vertices on features of interest is extracted from each mesh. View-based patches are used to search.

Weighting: The model fitting is extended to 3D from the 2D ASM [2]. During the fitting, weights can be assigned to each point. In [1] a combination of the match value at the target and its difference from that at the current position was used as the weight. We compared this to a simpler measure based on the match value alone; also to uniform weights of 1.0; and we investigated a new method which ignores the match value and is based on d_j, the distance between the current point j and its target:

$$w_j = 1 - \frac{d_j}{\max\{d\}_{\forall j}} \qquad (1)$$

The aim of this latter scheme is to keep in place points that are already matched, rather than allowing them to be pulled away by poor matches in other parts of the model. This may be because that part of the model is less well positioned, or because the subject is occluded in that area of the image and a good match cannot be found.

Searching for Occluding Boundaries: Searching using the patches does not take advantage of important image information from the occluding boundaries. In some cases the search result was poor because the model did not expand to match the size of the face, or because the model rotated to match the features rather than translating or changing shape. To correct for this we added an explicit search for occluding boundaries. Points which are likely to be on the boundary can be identified using the surface normals of the original mesh. At each of these the search is directed, 3 pixels wide in both directions, along the normal to find the strongest edge.

DataSets: 6 data sets were used with a variety of challenges:

- XM2VTS: Least challenging type of images with a fairly uniform backdrop, neutral expressions, and near frontal poses.
- BioID: Much more natural poses (mainly near frontal) and a variety of expressions. Cluttered backgrounds in an office setting.
- Three video sequences of 3 different drivers. A great deal of lighting variation, within and between frames, and wider variation in poses.
- Artificial Images with Known Poses [1]. Artificial subjects posed at up to 90 degrees from front against a real background.

Assessment: A subset of 12 points was used for evaluation. These were: the ends of the eyebrows; the corners of the eyes; the corners of the mouth; and the top and bottom of the mouth. The results are shown in Figure 3. The error value is median average point to point distance and failure is defined by an error of over 15 percent or pixels. The errors are reported on all 12 feature points and separately for just the eye corners (bottom row in each graph). The method codes are as follows: **B**asic Method; **P**rofile Search; Equal weighting of **1**.0; Weighting using **M**atch Value; Weighting using **D**istance (1).

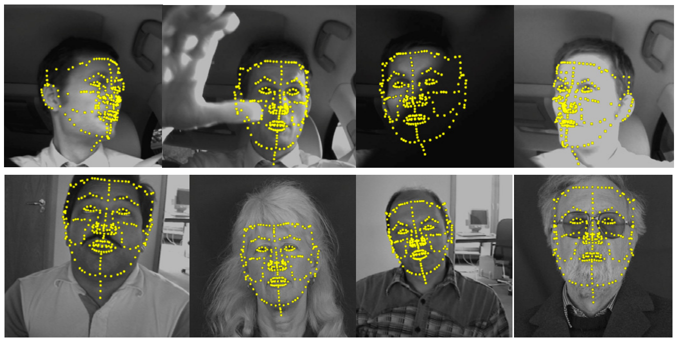

Figure 2: Search results. These example images from in-car videos and datasets show many problems including: occlusion; a non-uniform background; variable illumination; low contrast; extreme pose; glasses; and facial hair.

The improvements on the XM2VTS data set are not as dramatic as those on the BioID set which are much more difficult cases and have therefore benefitted more from the modifications. All the data sets showed improvements when both modifications were added together. It is likely that the boundary search scheme is beneficial in a smaller number of cases and this is why it has a less pronounced effect.

These simple generic modifications have an impact on search success and since they are non-specific we suggest they can be considered in other search algorithms of this type which use weighted model fitting, and where occluding boundaries may vary by view and are not explicitly modeled.

Acknowledgements: Toyota Motor Europe: funding and driver videos. Genemation Ltd.: 3D data markups and head textures.

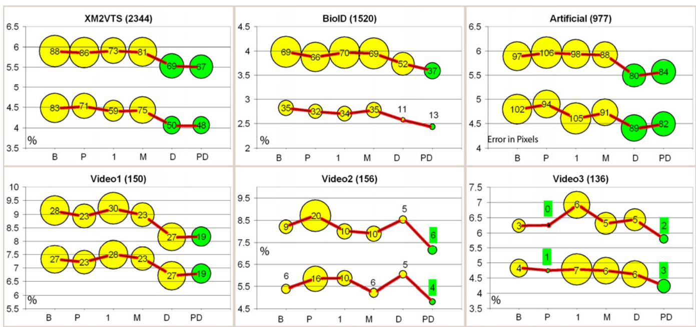

Figure 3: Assessment. See the text for the explanation of the method codes on the horizontal axis. The vertical axis is the median average error. The vertical position of the bubble indicates the error value and its size is the relative number of failures (shown inside). The top row in each graph is the result for all points, the bottom is the eyes only. The green bubbles are those results which were deemed best.

[1] A. Caunce, D. Cristinacce, C. Taylor, and T. Cootes, "Locating Facial Features and Pose Estimation Using a 3D Shape Model," *Proceedings of International Symposium on Visual Computing*, Las Vegas, 750-761, 2009.

[2] T. F. Cootes, D. H. Cooper, C. J. Taylor, and J. Graham, "Active Shape Models - Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, pp. 38-59, 1995.