# Live 3D facial scanning and landmark detection using Shape Regression with Incomplete Local Features

Federico M. Sukno* †, John L. Waddington† and Paul F. Whelan*
*Centre for Image Processing & Analysis, Dublin City University, Dublin 9, Ireland
†Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland

## I. Introduction

This demo is focused on the automatic detection of facial landmarks in surfaces obtained from a hand held laser scanner. The objective is to demonstrate the effectiveness of the algorithm by detecting the landmarks on the facial surface of any person that volunteers to be scanned.

A hand held laser scanner allows acquisition of a 3D surface by gathering measurements made by sweeping the scanning wand over an object (in a manner similar to spray painting). The final surface is obtained by merging the different sweeps which can be from various viewpoints, allowing a complete reconstruction of the facial surface, irrespective of the head pose and possible self-occlusions. For this demo we use a Cobra Wand 298, Polhemus FastSCAN$^{TM}$, Colchester, VT, USA)[1]. The reconstruction from multiple viewpoints, together with portability and price, are important advantages with respect to single-view scanners.

Landmark localization is accomplished by using SRILF (Shape Regression with Incomplete Features) [2]. This algorithm works by calculating a set of candidate points for each landmark and performing combinatorial search, with the key assumption that some landmarks might be missed (i.e. no candidates detected) which is tackled by using partial subsets of landmarks and inferring those that are missing by maximizing their plausibility based on a statistical shape model. Such assumption is crucial for the generalizability of the model for live scanning scenarios, where pre-processing is not possible but to a minimum extent and the quality of the resulting surfaces can vary considerably.

## II. Landmark localization algorithm

The SRILF algorithm [2] combines the response from local feature detectors for each of the targeted landmarks with statistical constraints that ensure the plausibility of landmark positions on a global basis. The algorithm has three components: $i)$ selection of candidates through local feature detection; $ii)$ partial set matching to infer possibly missing landmarks; $iii)$ combinatorial search, which integrates the other two components.

### A. Selection of candidates

The selection of candidates is performed independently for each targeted landmark; a similarity score is computed for every vertex and the top-scoring ones are retained as candidates for the considered landmark. As in many other algorithms, it is expected that one of these candidates will be close enough to the correct position of the landmark. Nonetheless, the number of false positives (i.e. vertices that produce high similarity scores even though they are far from the correct landmark location) can change considerably for different landmarks, as well as from one facial scan to another, making it difficult to choose the number of candidates that should be retained.

While many approaches try to retain large numbers of candidates to make sure that at least one will be *reasonably close* to the desired landmark position, SRILF determines the number of candidates as an upper outlier threshold from the distribution of false positives over a training set. This implies that, in the vast majority of cases, a candidate that is close enough to the target landmark will be detected, but a small proportion will be missed. Hence, for each targeted landmark there will be an initial set of candidates that may or may not contain a suitable solution and we need to match our set of target landmarks to a set of candidates that is potentially incomplete. This is analogous to the point-matching problem found in algorithms that search for correspondences. However, the human face is a non-rigid object and these point-matching algorithms are typically restricted to rigid transformations.

### B. Partial set matching

The second component of the algorithm aims at dealing with the above problem. Based on the priors encoded in a statistical shape model, it uses a subset of the landmarks (i.e. those with suitable candidates) to infer the most likely position of the ones that are missing.

Let $\mathbf{x} = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_L, y_L, z_L)^T$ denote a shape of $L$ landmarks in 3D, and let $\overline{\mathbf{x}}$, $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ be the mean shape, eigenvector and eigenvalue matrices of a representative training set of such shapes. Given a shape for which we only know part of its landmarks, we could split it in the known (or fixed) part $\mathbf{x}^f$ and the unknown part $\mathbf{x}^g$. The objective is to infer the coordinates of landmarks $\mathbf{x}^g$ so that the probability that the resulting shape complies with the PCA model is maximized, ideally without modifying the coordinates in $\mathbf{x}^f$. Assuming a multi-variate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ in PCA-space, it can be shown that:

$$\mathbf{x}^g = \overline{\mathbf{x}}^g - \left(\mathbf{\Phi}^g \mathbf{\Lambda}^{-1} (\mathbf{\Phi}^g)^T\right)^{-1} \left(\mathbf{\Phi}^g \mathbf{\Lambda}^{-1} (\mathbf{\Phi}^f)^T\right) (\mathbf{x}^f - \overline{\mathbf{x}}^f) \quad (1)$$

[1] http://polhemus.com/?page=Scanning_Fastscan

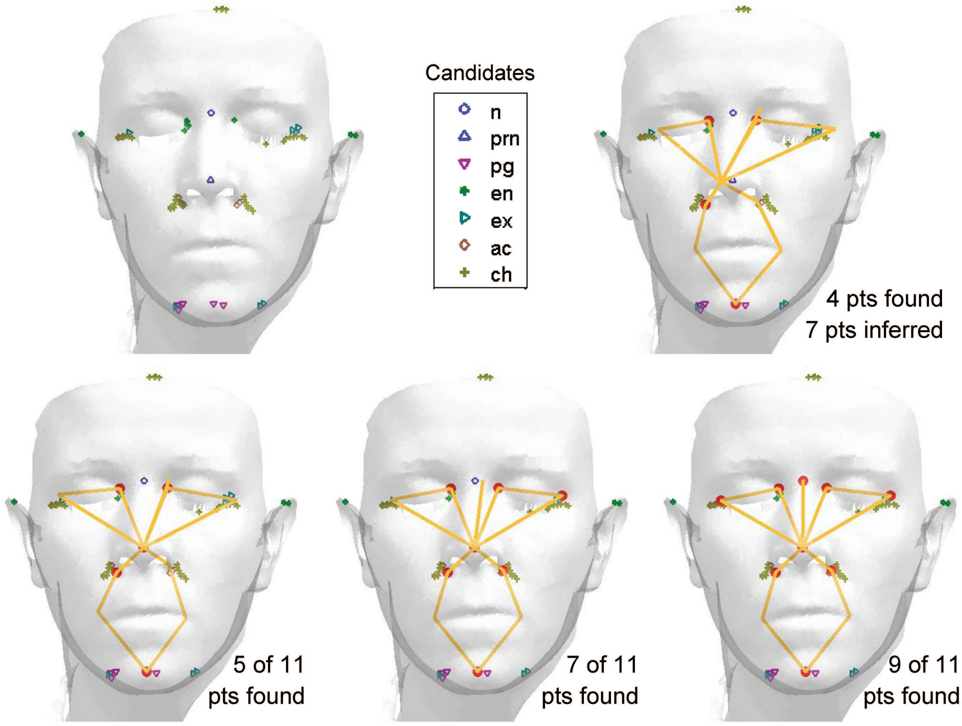


Fig. 1. Example of the SRILF algorithm targeting 11 landmarks: nose root (n) and tip (prn), chin tip (pg), inner (en) and outer (ex) eye corners, nose corners (ac) and mouth corners (ch). The top row shows the facial surface with the retained candidates for each landmark (left), as described by the legend box and a subset of 4 candidates identified as a plausible instance of the shape model (right). The identified candidates are highlighted by red circles and the resulting shape (completed by inferring all other landmarks) is indicated with solid lines. The bottom row shows further additions to the initial set until reaching 9 points.

### C. Combinatorial search

Finally, the third component of the algorithm integrates the two previous steps into a combinatorial search. It consists of analyzing subsets of candidates and completing the missing information by inferring the coordinates that maximize the probability of a deformable shape model.

We start from $L$ sets of candidate points, one per landmark. All possible combinations of $4$ candidates are then evaluated: the selected candidates are hypothesized to be correct and constitute fixed landmarks $\mathbf{x}^f$ and the shape is completed by using equations (1). As long as the generated shape constitutes a plausible shape (i.e. it fulfills the model constraints based on eigenvalue limits), we successively add candidates to $\mathbf{x}^f$ from the remaining landmarks in a sequential forward selection strategy, and repeat the shape completion process.

The maximum number of landmarks that can be included in $\mathbf{x}^f$ while keeping a plausible shape is used as figure of merit of the subset being tested (analogously to the support set in RANSAC [1]). Upon equality of supports, an inclusion cost is used which penalizes the reconstruction error of the fixed landmarks $\mathbf{x}^f$ and the distance from the inferred landmarks $\mathbf{x}^g$ to the closest candidates available.

An indicative example of the different steps is provided in Fig. 1. The first step showed corresponds to a subset of $4$ candidates that fulfills the model constraints. Note that, although the resulting shape is plausible, the inferred locations of the remaining 7 points are not very accurate. The next step is to try including candidates from the remaining landmarks. The nose tip is the one that achieves the lowest cost of inclusion, and is therefore added. This considerably improves the accuracy of the inferred shape. Inclusions continue, one at a time, until $9$ landmarks are placed in $\mathbf{x}^f$. All remaining candidates are checked, but in this case none of them produces a plausible instance with $10$ points in $\mathbf{x}^f$.

### Acknowledgment

The authors would like to thank their colleagues in the Face3D Consortium (www.face3d.ac.uk), and the financial support provided from the Wellcome Trust (grant 086901/Z/08/Z) and the Marie Curie IEF programme (grant 299605, SP-MORPH).

### References

[1] Fischler, M., Bolles, R.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM **24** (1981) 381–395.

[2] F.M. Sukno, J.L Waddington, and P.F. Whelan. 3D facial landmark localization using combinatorial search and shape regression. In *Proc. ECCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, LNCS vol. 7583*, pages 32–41, 2012.