3-D Facial Landmark Localization With Asymmetry Patterns and Shape Regression from Incomplete Local Features

Federico M. Sukno, John L. Waddington, and Paul F. Whelan, *Senior Member, IEEE*

***Abstract*—We present a method for the automatic localization of facial landmarks that integrates nonrigid deformation with the ability to handle missing points. The algorithm generates sets of candidate locations from feature detectors and performs combinatorial search constrained by a flexible shape model. A key assumption of our approach is that for some landmarks there might not be an accurate candidate in the input set. This is tackled by detecting partial subsets of landmarks and inferring those that are missing, so that the probability of the flexible model is maximized. The ability of the model to work with incomplete information makes it possible to limit the number of candidates that need to be retained, drastically reducing the number of combinations to be tested with respect to the alternative of trying to always detect the complete set of landmarks. We demonstrate the accuracy of the proposed method in the face recognition grand challenge database, where we obtain average errors of approximately 3.5 mm when targeting 14 prominent facial landmarks. For the majority of these our method produces the most accurate results reported to date in this database. Handling of occlusions and surfaces with missing parts is demonstrated with tests on the Bosphorus database, where we achieve an overall error of 4.81 and 4.25 mm for data with and without occlusions, respectively. To investigate potential limits in the accuracy that could be reached, we also report experiments on a database of 144 facial scans acquired in the context of clinical research, with manual annotations performed by experts, where we obtain an overall error of 2.3 mm, with averages per landmark below 3.4 mm for all 14 targeted points and within 2 mm for half of them. The coordinates of automatically located landmarks are made available on-line.**

Manuscript received November 18, 2013; revised March 20, 2014 and September 5, 2014; accepted September 9, 2014. Date of publication October 9, 2014; date of current version August 14, 2015. This work was supported in part by the Wellcome Trust under Grant WT-086901 MA, in part by the Marie Curie IEF Programme under Grant 299605 and Grant SP-MORPH, in part by Dublin City University (DCU), and in part by the Royal College of Surgeons in Ireland (RCSI) under the auspices of the 3U Partnership between DCU, Maynooth University and RCSI. This paper was recommended by Associate Editor J. Su.

F. Sukno is with the Centre for Image Processing and Analysis, DCU, Dublin 9, Ireland and also with Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland (e-mail: federico.sukno@gmail.com).

P. Whelan is with the Centre for Image Processing and Analysis, DCU, Dublin 9, Ireland.

J. Waddington is with Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland.

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes five supplementary tables that provide additional details of the main results contained within the paper. This material is 50 KB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2014.2359056

***Index Terms*—3-D facial landmarks, craniofacial anthropometry, geometric features, statistical shape models.**

I. Introduction

ACCURATE and automated detection of facial landmarks is an important problem in computer vision, with wide application to biometric identification [1]–[6] and medicine [7]–[11]. Biometric applications are typically concerned with the robustness of the algorithm (e.g., to occlusions, expressions, noncollaborative subjects) to achieve systems that can be deployed in a wide variety of scenarios. In this context, state of the art algorithms can detect the most prominent facial landmarks with average errors typically between 3 to 6 mm on large databases like the face recognition grand challenge (FRGC) [12]. These include diverse acquisition artifacts (e.g., holes, spikes) that help assess performance in challenging scenarios.

On the other hand, in medical applications such as facial surgery [11], lip movement assessment [10], or craniofacial dysmorphology [7], [8], the latter of which is the focus of our research, there is a greater focus on the highly accurate localization of landmarks, as they constitute the basis for analysis that is often aimed at detecting subtle shape differences. Depending on the author, localization and repeatability errors are considered clinically relevant when they exceed 1 mm [13] or 2 mm [14]. Acquisition conditions are therefore carefully controlled to minimize occlusions, holes and other artifacts. For example, using a hand held laser scanner it is possible to obtain a high quality ear-to-ear facial scan.[1]

The increased availability of three dimensional (3-D) scans has made it possible to overcome traditional limitations inherent to 2-D, such as viewpoint and lighting conditions. From this perspective, we can make a first distinction between methods using exclusively geometric cues (e.g., curvature) and those that analyze also texture information. While the latter have the benefit of including an additional source of information, they suffer from two shortcomings: 1) not all 3-D scanners provide texture and, even when they do, it cannot be assured that this is accurately registered to the geometry [12] and 2) they may become more sensitive to viewpoint and lighting conditions, as texture information is not invariant to these factors.

[1] See http://www.cipa.dcu.ie/videos/face3d/Scanning_DCU_RCSI.avi for an example [accessed on 20.05.2013].

Thus, there is a special interest in methods that localize facial landmarks based purely on geometric information. The most widely used feature to encode the facial geometry for landmark detection has been surface curvature. Building from early works on surface classification (using mean and Gaussian curvatures [15] or shape index [16]), several authors have explored the properties exhibited by certain facial landmarks. For example, it has been found that the nose and chin tips are peaks or caps, while the eye and mouth corners are pits or cups [17]–[21]. This classification has proved useful for rough detection of the most distinctive facial features but, in general, it does not suffice for highly accurate localization and is restricted to a very small number of specific landmarks, with little likelihood of being extended to other points.

Similar limitations are observed in the use of relief curves (or profiles), that is, a projection of the range information (depth) onto the vertical or horizontal axes. With some assumptions regarding the orientation of the head (relative to the scanner), this procedure allows the use of the resulting 1-D projections to detect some facial landmarks. This has proved helpful in detecting the nose tip, even under some variations in head pose [22], but without informing on localization accuracy. Recent extensions include the generation of multiple profiles to account for large changes in head pose [23] and combinations with curvature cues to derive heuristics for the detection of reduced sets of points on the eyes and nose [21]. Other geometric features include the response of range data when convolved with a set of primitive filters [24], Gabor wavelets [25], or combinations of features such as local volume, spin images, distance to local plane, or radial basis function (RBF) shape histograms [20], [26]–[28].

Regardless of the features that are used, it is unlikely that a unique and highly accurate detection can be achieved. Even the nose tip, so far the most successfully detected facial landmark, suffers from both false positives and negatives. Hence, the responses from feature detectors are usually combined with prior knowledge to improve performance. This leads us to a second distinction between methods that use a training set to derive these priors and those that employ heuristic rules.

Methods targeting a small subset of landmarks are often training-free. A set of carefully designed rules encodes the prior knowledge, sometimes with the help of anthropometric statistics [3]. A weakness of these methods is that they usually follow a chain of rules that depend on one another. For example, some methods [3], [21], [25], [29] start by locating the nose tip and use its location to constrain the search region of the remaining points, while others [30] first detect the inner-eye corners and use these to fit a local plane from which the nose tip is determined as the furthest point. Therefore, missing or incorrectly detecting one landmark compromises the detection of all subsequent landmarks in the chain.

Prior knowledge can also be derived from a training set. At the expense of requiring that such a set (with appropriate annotations) is available, training-based methods are more flexible than their training-free counterparts in the landmarks that can be targeted, as there is no need to derive specific rules for each point. This has been widely exploited in 2-D landmarking algorithms and is becoming more popular also in 3-D, especially since the availability of large annotated databases. Examples of this strategy include the use of graph matching [28], [31], random forests [32] or statistical shape models [19], [20], [33]–[36].

Recently, it has been shown that statistical methods can produce accurate results for diverse subsets of landmarks [19], [20], [35], [36]. The common idea behind them is to combine the responses of local feature detectors with shape constraints that ensure plausibility of the result at a global level. Since localization of landmarks is simultaneously addressed, these methods are more robust to localization errors in individual points. Nonetheless, current approaches still rely on the availability of a complete set of features, i.e., the local feature detectors are expected to always provide at least one suitable candidate position for each targeted landmark, which can prove quite difficult for most feature detectors.

One can relate this intuitively to partial occlusions, but the problem is actually more general: for example, a feature detector can fail to provide a suitable candidate for the chin tip because: 1) it is occluded (e.g., by a scarf); 2) the surface is missing (e.g., acquisition artifacts); and 3) limitations inherent to the detector itself, even though the surface of the chin was captured correctly by the scanner. In the latter case, we say the the feature detector has produced a false negative which, as discussed above, is almost impossible to avoid.

In this paper we present shape regression with incomplete local features (SRILF) for the detection of facial landmarks. It can handle any combination of missing points and allows for nonrigid deformations, while working on a global basis. Instead of trying to avoid false negatives, we provide a mechanism to handle them by using a flexible shape model that encodes prior knowledge of the facial geometry. Therefore, we withdraw the requirement of a complete set of features and try to match our set of targeted landmarks to a set of candidates that is potentially incomplete. Our matching algorithm, based on RANSAC [37], consists of analyzing reduced subsets of candidates and completing the missing information by inferring the coordinates that maximize the probability of the flexible model. Thus, despite the resulting subset possibly containing only part of the targeted landmarks, estimates for the remaining coordinates are inferred from the model priors. Subsets of candidates that fulfill the statistical constraints of the shape model are retained and additional landmarks are incorporated iteratively as long as the set remains a plausible instance of the shape model. The cost of including a new candidate is computed as the median of squared distances to the closest candidate (per landmark), which provides robustness to potential landmarks for which no nearby candidates have been found. The best solution is determined as the one with minimum inclusion cost among those with the largest number of candidates (i.e., those with the largest support).

The key contribution of SRILF is to bridge the gap between two research streams.

1) Methods based on robust point matching but restricted to rigid transformations, as done by Creusot *et al.* [27], which can handle missing landmarks but do not allow nonrigid deformation and are therefore strongly limited in their accuracy. We have shown experimentally that inability to cope with nonrigid deformations can

considerably impair accuracy even in databases without expression changes [38].

2) Methods based on statistical shape models that allow nonrigid deformation but cannot handle missing landmarks [19], [20], [35], [36].

Recent efforts to tackle these shortcomings have not provided a general and unified framework. Passalis *et al.* [19] and Perakis *et al.* [20] exploited facial symmetry to divide their shape model into left and right sub-models, but each of these is actually a separate statistical model in itself, necessitating a complete set of features and not allowing inference of the landmarks of the other sub-model. In contrast, SRILF always provides estimates for the positions of all landmarks regardless of the subset for which information is missing.

Another alternative based on statistical models is that of Zhao *et al.* [35], [36], who address a local optimization after an initial solution is provided (e.g., by a previous face detector block). Thus, even if some feature detectors produce poor responses, the search is constrained to a bounded neighborhood and is unlikely to diverge. However, we can see that the problem is actually shifted to the availability of an adequate initialization and, therefore, the solution is not global.

The idea of using statistical constraints to complete missing landmarks in shape models has been explored previously and has found diverse applications. These include predicting the normal shape of vertebrae from their neighbors to assess fractures [39], initializing a registration algorithm from a reduced set of manual annotations [40], [41] or reconstructing bones or facial surfaces from partial observations [42], [43]. Solutions to estimate the unknown variables were based on regularized or partial least squares [44], [45], canonical correlation analysis [46] or linear regression [39]. However, in all cases the goal of these models is to predict unknown parts of the shape based on a partial observation that is predefined statically. That is, the part of the shape that will be available is known already when the model is constructed, either at once [39] or sequentially one landmark at a time [47]. In contrast, we use a unique principal component analysis (PCA) model to handle any combination of known and unknown landmarks (as this information is not know in advance) and select the best solution based on a cost function as described above. A similar concept has been explored recently by Drira *et al.* [48] in the context of face recognition, to predict the missing information of curves extracted from facial surfaces that might be partially incomplete due to occlusions or artifacts.

We use asymmetry patterns shape contexts (APSC) [49] as feature detectors. These constitute a family of geometric descriptors based on the extraction of asymmetry patterns from the popular 3-D shape contexts (3-DSC) [50]. APSC resolve the azimuth ambiguity of 3-DSC and offer the possibility to define a variety of descriptors by selecting diverse spatial patterns, which has two important advantages: 1) choosing the appropriate spatial patterns can considerably reduce the errors obtained with 3-DSC when targeting specific types of points and 2) once an APSC descriptor is built, additional descriptors can be built incrementally at very low cost.

We experimentally demonstrate the accuracy of our approach by testing it on FRGC, the most widely used database for reporting 3-D landmark localization. We obtain an average error of approximately 3.5 mm when targeting 14 prominent facial landmarks. For the majority of these our method produces the most accurate results reported to date in this database among methods based exclusively on geometric cues. Additionally, we also show that our results compare well even with methods combining both geometry and texture, which have reported lower errors only in the case of the eye corners where texture seems to play a more prominent role. We also test our algorithm on the Bosphorus database [51], and show the suitability of SRILF to handle scans with occlusions or where large parts of the facial surface are missing.

To investigate potential limits in the accuracy that could be reached, we report experiments on a database acquired in the context of craniofacial dysmorphology research, which contains surfaces of higher quality than those from FRGC with manual annotations performed by experts. Targeting the same 14 landmarks, we obtain an average error of 2.3 mm on 144 facial scans.

We present the details of our landmark localization algorithm in Section II; experimental results are provided in Section III, followed by a discussion in Section IV and concluding remarks in Section V.

II. SRILF

The SRILF algorithm has three components: 1) selection of candidates through local feature detection; 2) partial set matching to infer missing landmarks by regression; and 3) combinatorial search, which integrates the other two components. We present each of these in separate subsections.

A. Local Feature Detection

Let $\mathcal{M}$ be a facial surface described by vertices $\mathbf{v} \in \mathcal{M}$, let $\{\mathbf{a}(\ell_k)\}_{k=1}^{L}$ be the set of manual 3-D annotations containing L landmarks and let $D(\mathbf{v})$ be a descriptor that can be computed for every vertex $\mathbf{v}$. We want to train a local descriptor model for each landmark. The objective is to compute a similarity score $s(\mathbf{v})$ based solely on the local descriptors, that correlates well with the distance to the correct position of the targeted landmark. That is, for each landmark ℓ_k we seek a function $f_k()$ such that $s_k(\mathbf{v}) = f_k\big(D(\mathbf{v})\big)$ is high for vertices close to $\mathbf{a}(\ell_k)$ and low for all other vertices of the mesh.

For example, spin images [52] or 3-DSC [50] are popular geometric descriptors; and one of the simplest options to obtain similarity scores, quite widespread both in the 2-D and 3-D landmark localization literature, is to compute the distance to a template derived as the average descriptor from a training set.

1) False Positives: For every mesh vertex, the Euclidean distance to the targeted landmark can be computed as

$$d(\mathbf{v}, \ell_k) = \|\mathbf{v} - \mathbf{a}(\ell_k)\|. \tag{1}$$

Ideally, vertices with high $s_k(\mathbf{v})$ should be close to the target and have small $d(\mathbf{v}, \ell_k)$. However, very often there are false positives, i.e., vertices with high $s_k(\mathbf{v})$ and $d(\mathbf{v}, \ell_k)$ at the same time. Whether a vertex is considered a false positive or not depends on how close to the target we require it to be, which

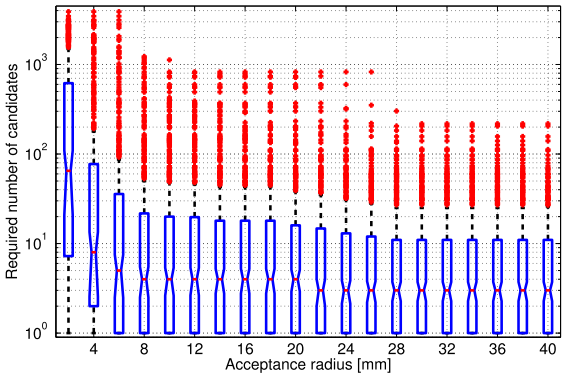

Fig. 1. Required number of candidates to be retained so that at least one of them is within the acceptance radius from the target. The boxplots indicate the results for all meshes in FRGCv1 database when targeting the right corner of the mouth using spin images.

is set by an acceptance radius r_A. To successfully locate a given landmark, we wish to retain enough candidates (the top-N_k) so that at least one of them is within our acceptance radius

$$N_k = \min_n \{n = R_D\left(s_k(\mathbf{v})\right) \mid d(\mathbf{v}, \ell_k) \leq r_A, \mathbf{v} \in \mathcal{M}\} \quad (2)$$

$$R_D\left(s_k(\mathbf{v})\right) = \#\left(\{\mathbf{w} \in \mathcal{M} \mid s_k(\mathbf{w}) \geq s_k(\mathbf{v})\}\right). \quad (3)$$

Thus, N_k is the required number of candidates, $R_D()$ is the (descending) rank function and #() is the cardinality of a set. Alternatively, if $\mathbf{v}_k^1$ is the highest scoring vertex within r_A

$$\mathbf{v}_k^1 = \operatorname*{argmax}_{\mathbf{v}} \{s(\mathbf{v}) \in \mathcal{M} \mid d(\mathbf{v}, \ell_k) \leq r_A\} \quad (4)$$

we can give a precise definition of false positives as the set of vertices $\mathcal{F}_k^+$ that are farther from the target ℓ_k than r_A but score higher than $\mathbf{v}_k^1$

$$\mathcal{F}_k^+ = \left\{\mathbf{v} \in \mathcal{M} \mid d(\mathbf{v}, \ell_k) > r_A \wedge s_k(\mathbf{v}) > s_k\left(\mathbf{v}_k^1\right)\right\}. \quad (5)$$

Thus, N_k is also the number of false positives plus one.

2) Candidate Sets: Given a mesh $\mathcal{M}$ and a landmark ℓ_k to be targeted, we define the set of candidates for that landmark, $\mathcal{C}_k$ as the ϱ_k highest scoring vertices

$$\mathcal{C}_k = \{\mathbf{v} \in \mathcal{M} \mid R_D\left(s_k(\mathbf{v})\right) \leq \varrho_k\}. \quad (6)$$

From the discussion in the previous paragraphs we can infer that the set $\mathcal{C}_k$ will contain at least one candidate within r_A if and only if $\varrho_k \geq N_k$. Clearly, we do not know N_k beforehand and trying to ensure $\varrho_k \geq N_k$ results in very high ϱ_k without a guarantee to be sufficient for all meshes.

To illustrate this, consider a set $\{\mathcal{M}_i\}_{i=1}^N$ where we compute the number of candidates required for each mesh, $N_k^{(i)}$. These values depend on our choice of r_A; the smaller r_A the larger number of candidates we need to retain. Fig. 1 shows the resulting number of candidates for r_A between 2 and 40 mm when targeting the right corner of the mouth in FRGC v1. As observed in the figure, the distributions of $N_k^{(i)}$ tend to be very skewed. Thus, setting ϱ_k based on the maximum values (which are typically outliers) is an expensive choice, as it implies retaining up to one or two orders of magnitude more candidates than needed in the majority of cases. This is a common problem to almost any geometric descriptor.

In contrast, we set ϱ_k as an outlier threshold for the distribution of $N_k^{(i)}$, as follows:

$$\varrho_k = q_3 + 1.5\,(q_3 - q_1) \quad (7)$$

which is a standard criterion to determine outliers, being q_1 and q_3 the lower and upper hinges (or quartiles) for $\{N_k^{(i)}\}_{i=1}^N$. Continuing with the example in Fig. 1, if we set $r_A = 10$ mm we get $\varrho_k \simeq 50$, while the maximum of $N_k^{(i)}$ is above 1000.

Choosing ϱ_k based on an outlier threshold for the distribution implies that, in the vast majority of cases, we will detect a candidate that is within r_A from the target, but we will miss a small proportion (the outliers). The latter will be dealt with by the partial set matching explained in the next section.

B. Partial Set Matching With Statistical Shape Models

Let $\mathbf{x} = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_L, y_L, z_L)^T$ be a shape vector, constructed by concatenating the coordinates of L landmarks.[2] By applying PCA over a representative training set [53], we get the mean shape $\bar{\mathbf{x}}$ and the eigenvector and eigenvalue matrices $\mathbf{\Phi}$ and $\mathbf{\Lambda}$, respectively, sorted in descending order ($\Lambda_{ii} \geq \Lambda_{jj}$, $\forall i < j$). Given any set of L points $\mathbf{x}$, we can obtain its PCA representation as $\mathbf{b} = \Phi^T(\mathbf{x} - \bar{\mathbf{x}})$, which will be considered to comply with the PCA model (i.e., to be a plausible instance within such a model) if it satisfies

$$\sum_{j=1}^{M} \left(\frac{b_j^2}{\Lambda_{jj}}\right) < \beta_e^2 \quad (8)$$

where M is the number of retained principal components and β_e is a constant that determines the flexibility of the model, which we set to $\beta_e = 4$ as in [38].

However, if the point set is incomplete, we may want to use the available points and the model statistics to infer those that are missing. Let $\mathbf{x}^f$ be the fixed (or available) landmarks, and $\mathbf{x}^g$ the unknown landmarks (the ones to guess). Without loss of generality we group the missing landmarks from 1 to $3g$

$$\mathbf{x}^g = \left(x_1, y_1, z_1, \ldots, x_g, y_g, z_g\right)^T$$
$$\mathbf{x}^f = \left(x_{g+1}, y_{g+1}, z_{g+1}, \ldots, x_L, y_L, z_L\right)^T$$
$$\mathbf{x} = \begin{pmatrix}\mathbf{x}^g \\ \mathbf{x}^f\end{pmatrix}, \qquad \mathbf{\Phi} = \begin{pmatrix}\mathbf{\Phi}^g \\ \mathbf{\Phi}^f\end{pmatrix}. \quad (9)$$

The objective is to infer the coordinates of landmarks $\mathbf{x}^g$ so that the probability of the resulting shape complying with the PCA model is maximized, ideally without modifying the coordinates in $\mathbf{x}^f$. Let $Pr(\mathbf{x})$ be the probability that shape $\mathbf{x}$ complies with the model. Assuming that $Pr(\mathbf{x})$ follows a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ in PCA-space, this probability is proportional to the negative exponential of the Mahalanobis distance, as follows:

$$Pr(\mathbf{x}) \sim e^{\left(-\mathbf{b}^T \mathbf{\Lambda}^{-1} \mathbf{b}\right)}. \quad (10)$$

We want to find its maximum with respect to $\mathbf{x}^g$, so we need to cancel the first order derivatives simultaneously for all the

[2]We assume that the shape has been aligned (e.g., by procrustes analysis) so that similarity is removed.

components of $\mathbf{x}^g$

$$\frac{\partial Pr(\mathbf{x})}{\partial \mathbf{x}^g} = \mathbf{0} \Leftrightarrow \frac{\partial}{\partial \mathbf{x}^g}\left(-\mathbf{b}^T\mathbf{\Lambda}^{-1}\mathbf{b}\right) = \mathbf{0}. \quad (11)$$

Replacing $\mathbf{b} = \Phi^T(\mathbf{x} - \overline{\mathbf{x}})$ and defining $\mathbf{y} = \mathbf{x} - \overline{\mathbf{x}}$ we obtain

$$\frac{\partial Pr(\mathbf{x})}{\partial \mathbf{x}^g} = \mathbf{0} \Leftrightarrow \frac{\partial}{\partial \mathbf{x}^g}\left(-\mathbf{y}^T\mathbf{\Phi}\mathbf{\Lambda}^{-1}\mathbf{\Phi}^\mathbf{T}\mathbf{y}\right) = \mathbf{0}. \quad (12)$$

Note that $\mathbf{y}$ and $\mathbf{x}$ differ only by a constant, so we can take derivatives directly with respect to $\mathbf{y}$. We also define a new matrix $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{\Lambda}^{-1}\mathbf{\Phi}^\mathbf{T}$ to simplify the notation

$$\frac{\partial Pr(\mathbf{x})}{\partial \mathbf{x}^g} = \mathbf{0} \Leftrightarrow \frac{\partial}{\partial \mathbf{y}^g}\left(-\mathbf{y}^T\mathbf{\Psi}\mathbf{y}\right) = \mathbf{0}. \quad (13)$$

We can explicitly separate the components related to $\mathbf{y}^f$ and $\mathbf{y}^g$, as follows:

$$\mathbf{y}^T\mathbf{\Psi}\mathbf{y} = \begin{pmatrix}\mathbf{y}^g\\ \mathbf{y}^f\end{pmatrix}^T \begin{bmatrix}\mathbf{\Psi}^{gg} & \mathbf{\Psi}^{gf}\\ \mathbf{\Psi}^{fg} & \mathbf{\Psi}^{ff}\end{bmatrix} \begin{pmatrix}\mathbf{y}^g\\ \mathbf{y}^f\end{pmatrix} \quad (14)$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{y}^g}\left(-\mathbf{y}^T\mathbf{\Psi}\mathbf{y}\right) &= -\frac{\partial}{\partial \mathbf{y}^g}\left(\left(\mathbf{y}^g\right)^T\mathbf{\Psi}^{gg}\mathbf{y}^g + \left(\mathbf{y}^f\right)^T\mathbf{\Psi}^{fg}\mathbf{y}^g\right.\\ &\quad \left. + \left(\mathbf{y}^g\right)^T\mathbf{\Psi}^{gf}\mathbf{y}^f + \left(\mathbf{y}^f\right)^T\mathbf{\Psi}^{ff}\mathbf{y}^f\right)\\ &= -\mathbf{\Psi}^{gg}\mathbf{y}^g - \left(\mathbf{\Psi}^{gg}\right)^T\mathbf{y}^g - \left(\mathbf{\Psi}^{fg}\right)^T\mathbf{y}^f - \mathbf{\Psi}^{gf}\mathbf{y}^f. \end{aligned} \quad (15)$$

The expression can be further simplified by noting that $\mathbf{\Psi}$ is symmetric (because the inverse of $\mathbf{\Lambda}$ is symmetric)

$$\frac{\partial}{\partial \mathbf{y}^g}\left(-\mathbf{y}^T\mathbf{\Psi}\mathbf{y}\right) = \mathbf{0} \Leftrightarrow \mathbf{\Psi}^{gg}\mathbf{y}^g + \mathbf{\Psi}^{gf}\mathbf{y}^f = \mathbf{0}. \quad (16)$$

Finally, as long as $\mathbf{\Psi}^{gg}$ is invertible, we can solve for $\mathbf{y}^g$

$$\begin{aligned}\mathbf{y}^g &= -\left(\mathbf{\Psi}^{gg}\right)^{-1}\mathbf{\Psi}^{gf}\mathbf{y}^f\\ &= -\left(\mathbf{\Phi}^g\mathbf{\Lambda}^{-1}\left(\mathbf{\Phi}^g\right)^T\right)^{-1}\left(\mathbf{\Phi}^g\mathbf{\Lambda}^{-1}\left(\mathbf{\Phi}^f\right)^T\right)\mathbf{y}^f. \end{aligned} \quad (17)$$

As explained in Section I, the idea of using statistical constraints to complete missing landmarks has been explored previously by other authors. The closest approach to ours is the one from de Bruijne *et al.* [39], where a closed form solution is obtained using the maximum likelihood estimate of $\mathbf{x}^g|\mathbf{x}^f$ from the covariance matrix of the training set. While results tend to be very similar, the main difference is that we maximize the probability of the shape after the projection into model space, which results in higher probability of compliance with the model at the expense of having also a higher reconstruction error for $\mathbf{x}^f$.

C. Combinatorial Feature Matching

We use RANSAC as the basis for our feature matching procedure, as described in Algorithm 1. We start from L sets of candidate points, one set for each landmark. As described in Section II-A, these candidates are the top-scoring vertices up to ϱ_k, which is determined during training. All combinations of four landmark candidates are then evaluated. In principle, we could also start from subsets of 3 points as we use similarity alignment (7 degrees of freedom), but 4 points were found to provide more robustness to estimate the initial alignment.

Algorithm 1 SRILF

1: Start from input mesh $\mathcal{M}$
2: **for** (all landmarks ℓ_k, $1 \le k \le L$) **do**
3: Compute descriptor scores $s_k(\mathbf{v})$, $\forall \mathbf{v} \in \mathcal{M}$
4: Determine landmark candidates $\mathcal{C}_k$ using (6)
5: **end for**
6: **for** (all 4-tuple combinations of candidates, $\mathbf{x}_4$) **do**
7: Initialize $\mathbf{x}^f = \mathbf{x}_4$
8: Infer $\hat{\mathbf{x}}^g$ using (17), obtaining $\hat{\mathbf{x}}$
9: **while** ($\hat{\mathbf{x}}$ fulfills the constraints in (8)) **do**
10: **for** (all other landmarks, $\ell_k \notin \mathbf{x}^f$) **do**
11: **for** (all candidates $\mathbf{c}_k$ for landmark ℓ_k) **do**
12: Add the candidate $\mathbf{c}_k$ to $\mathbf{x}^f$ to obtain $\mathbf{x}^f_{\text{test}}$
13: Infer $\hat{\mathbf{x}}^g_{\text{test}}$ from $\mathbf{x}^f_{\text{test}}$ to obtain $\hat{\mathbf{x}}_{\text{test}}$
14: Constrain $\hat{\mathbf{x}}_{\text{test}}$ to be within $\mathcal{M}$ (optional)
15: Compute the resulting cost $\gamma(\mathbf{c}_k)$ as in (18)
16: **end for**
17: Compute the landmark cost $\gamma(k) = \min_{\mathbf{c}_k}\gamma(\mathbf{c}_k)$
18: **end for**
19: Add to $\mathbf{x}^f$ the landmark with minimum $\gamma(k)$
20: Infer $\hat{\mathbf{x}}^g$ from the updated $\mathbf{x}^f$ to obtain $\hat{\mathbf{x}}$
21: **end while**
22: Compute the score for $\mathbf{x}_4$ as $\#(\mathbf{x}^f) + e^{-\gamma(k)}$
23: **end for**
24: Keep the subset that achieves the highest score

We use (17) to infer the positions of missing landmarks. As long as the generated shape fulfills the model constraints, we successively add candidates from the remaining landmarks in a sequential forward selection strategy [54]. The cost of including a new candidate $\mathbf{c}_k$ into $\mathbf{x}^f$ is computed as the median of squared distances to $\mathbf{x}^f_{\text{test}}$, taking the closest candidates to the current estimate for the missing landmarks

$$\begin{aligned}\gamma(\mathbf{c}_k) &= \text{median}\left(\Delta\hat{\mathbf{x}}_{\text{test}}\right)\\ \Delta\hat{\mathbf{x}}_{\text{test}} &= \begin{Bmatrix}\left\|\hat{\mathbf{x}}_{\text{test}}(\ell_k) - \mathbf{x}^f_{\text{test}}(\ell_k)\right\|^2, & \forall \ell_k \in \mathbf{x}^f_{\text{test}}\\ \min_{\mathbf{c}_k}\left\|\hat{\mathbf{x}}_{\text{test}}(\ell_k) - \mathbf{c}_k\right\|^2, & \forall \ell_k \notin \mathbf{x}^f_{\text{test}}\end{Bmatrix}\end{aligned} \quad (18)$$

where $\mathbf{c}_k \in \mathcal{C}_k$ are the candidates for landmark ℓ_k, $\mathbf{x}(\ell_k)$ indicates the position of the kth landmark and $\hat{\mathbf{x}}$ is the best PCA reconstruction of shape $\mathbf{x}$ in a least squares sense.

The inclusion cost in (18) is a key aspect of the algorithm and is divided in two parts from the definition of $\Delta\hat{\mathbf{x}}_{\text{test}}$. The first part is the reconstruction error for the fixed landmarks, while the second part considers the distance from the inferred landmarks to their closest candidates. Note that a possible alternative would be using $\|\mathbf{\Phi}^T(\hat{\mathbf{x}} - \overline{\mathbf{x}})\|$ as the inclusion cost, but such a choice would neglect the effect of the coordinates inferred for $\hat{\mathbf{x}}^g$. The definition of $\gamma(\mathbf{c}_k)$ based on the median implies that the landmark cost $\gamma(k)$, in line 12 of Algorithm 1, is the least median of squares [55], which provides robustness to potential outliers (e.g., landmarks for which no nearby candidates have been found).

For each set that is checked, a score is computed. The candidates successfully included in $\mathbf{x}^f$ (i.e., those which allow

completion of a shape fulfilling the PCA constraints) are considered inliers. Thus, the cardinality of $\mathbf{x}^f$ is used as the main component of the score. Upon equality of inliers, the subset with smallest $\gamma(k)$ is preferred.

The optional statement in line 14 of Algorithm 1 forces all landmarks to be on the input surface, e.g., by shifting them to the nearest vertex of $\mathcal{M}$. This is useful for discarding incorrect solutions but could be disabled to tolerate occlusions or missing parts of the surface.

1) Complexity: The loop of lines 11 to 16 of Algorithm 1 plays a central role in the overall complexity of the algorithm. Each time this loop is executed we need to compute the matrix inversion of (17). For each potential landmark to be added a new repetition of this loop is needed.[3] Thus, the complexity of the algorithm is variable and depends on how quickly we can discard implausible combinations of candidates.

An efficient way to discard implausible combinations at low cost was presented by Passalis *et al.* [19]. For each new combination, they check the distances between all pairs of landmarks and discard the combination if these are not compatible with the distances observed in the training set. The use of distances is possible due to their choice to exclude scaling from the transformation relating image and model coordinates, which is not our case. Rather, we adapt their approach to scale invariance by using ratios of distances to validate combinations of four candidates at line 7 of Algorithm 1.

2) Convergence: As we are interested in accuracy, we do an exhaustive search instead of random sampling. However, we do retain the idea of consensus as the figure of merit, hence the relation of our algorithm with RANSAC. On the other hand, an exhaustive search does not guarantee finding a plausible solution, which depends on the choice of the threshold for plausibility β_e and the number of false positives. For example, when large parts of the torso are included in the scan there might be too few candidates retained in the facial region. One can always choose to keep the best solution that was found so far, even if deemed implausible. However, in such a situation we could also benefit from the splitting of such best solution into $\mathbf{x}^f$ and $\mathbf{x}^g$ and rerun the algorithm with more candidates for the inferred landmarks, namely increasing $\varrho_k \ \forall \ell_k \notin \hat{\mathbf{x}}^f$.

The advantage of increasing the candidates for just part of the landmarks is twofold: 1) it reduces the number of combinations to test and 2) it generally results in lower proportion of combinations being plausible, which are the most expensive ones to discard. Adding candidates for landmarks in $\mathbf{x}^f$ would most likely produce additional subsets of candidate combinations that are plausible but are still geometrically similar to combinations already available before adding further candidates, thus increasing the computational cost without much benefit in accuracy.

3) Examples: Visualizing the different steps of the combinatorial search can be helpful to illustrate the process described in Algorithm 1. For this purpose, we have generated a large number of example videos showing the behavior of SRILF both for typical and extreme cases, which are available on-line.[4]

[3]Note that each execution of the loop between lines 11 to 16 of Algorithm 1 involves only one matrix inversion, as all candidates tested within the loop correspond to the same landmark and therefore produce the same split of the eigenvector matrix Φ into Φ^f and Φ^g.

TABLE I
LANDMARK DEFINITIONS AND ABBREVIATIONS

Name	Abbr	Description
Alare crest (2)	ac	Nose corner, L/R (insertion of each alar base)
Cheilion (2)	ch	Mouth corner, L/R (labial commissure)
Endocanthion (2)	en	Inner-eye corner, L/R
Exocanthion (2)	ex	Outer-eye corner, L/R
Labiale inferius	li	Middle point of the lower lip
Labiale superius	ls	Middle point of the upper lip
Nasion	n	Depressed area between the eyes, just above the nose bridge
Pogonion	pg	Chin tip (most anterior, prominent point on the chin)
Pronasale	pm	Nose tip (most anterior midpoint of the nasal tip)
Subnasale	sn	Point at which the nasal septum merges, in the midsagittal plane, with the upper lip

III. EXPERIMENTAL EVALUATION

A. FRGC Database

The FRGC database [12] is a large publicly available corpus that has been widely used to report landmark localization results, thus allowing for a direct comparison of our algorithm with state of the art methods. The 3-D part of the database provides both geometric (range) and texture information and is divided in two parts (or versions): FRGC v1 contains 943 scans from 275 subjects with only mild expression variations and without illumination changes; FRGC v2 contains 4007 scans from 466 subjects with both illumination and expression variations, some of which are very significant.

We will report experimental results using 2-fold cross-validation on each database version (v1 or v2) and results training on v1 and testing on v2, to reproduce the different experimental settings reported in the literature.

All scans were preprocessed with a median filter to remove spikes and a smoothing filter based on a bi-quadric approximation of each vertex from a 3 mm neighborhood. Finally, scans were decimated by a factor of 1:4 and converted to triangulated meshes. This resulted in an average of approximately 25 500 vertices per mesh.

Ground truth annotations for this database are also publicly available. We used annotations from [34], with the additions and corrections introduced by Creusot *et al.* [27] which are available on-line.[5] We target the 14 facial landmarks available in this set, with definitions as indicated in Table I.

B. Geometric Descriptors

We use APSC [49] as geometric descriptors [i.e., to generate the scores $s(\mathbf{v})$]. APSC descriptors are constructed by extracting asymmetry patterns from a 3-DSC. The computational cost of the latter is considerably higher than the extraction of asymmetry patterns, which allows computing several APSC descriptors at a computational cost comparable

[4]http://www.cipa.dcu.ie/face3d/SRILF_Examples.htm [15.07.2013]
[5]Available at http://clementcreusot.com/phd/ [08.07.2013]

to a single descriptor. On the other hand, the use of asymmetry resolves the azimuth ambiguity of 3-DSC, which speeds up the computation of the scores and tends to compensate the extra time needed to build the descriptors. While individual APSC descriptors can achieve comparable accuracy to other popular descriptors such as spin images or 3-DSC, using a pool of APSC to target each landmark with the most appropriate descriptor provides improved localization accuracy with a marginal increase in computation cost [49], [56].

We evaluated all APSC descriptors listed in [49] and choose the most appropriate for each landmark using default settings: $11 \times 12 \times 15$ elevation, azimuth, and radial bins covering a spherical neighborhood of $r_{\max} = 30$ mm radius and setting the smallest radial bins at $r_{\min} = 1$ mm. We select this only once, using the FRGC v1 database.

In all cases, we obtained descriptor templates for each landmark by averaging over the training set. As manual annotations for FRGC have been shown to be rather noisy, we used the least squared corrections of uncertainty algorithm [57] to build the templates. In brief, this means that we assumed an uncertainty in the manual annotations, which were allowed to move within a small neighborhood of radius r_u to enforce consistency of the extracted descriptors. Previous experiments on this database produced stable results for r_u between 5 and 20 mm, hence we adopt a conservative value and set $r_u = 5$ mm. The ground truth displacements are only used during training to derive the templates and are specific to each descriptor. We have shown that this strategy is more accurate than simply trusting the manual annotations [57].

Descriptor scores were computed as the negative Euclidean distance to the template. We also explored using the Mahalanobis distance, which generally reduced the errors, although this was significant only for the landmarks in the mouth and chin (*ch*, *ls*, *li* and *pg*). Since the dimension of the APSC descriptors is relatively high (990 bins) using the Mahalanobis distance proved computationally expensive, even though we computed it after projection into a lower dimensional space obtained by PCA. Thus, Mahalanobis distances were used only for those landmarks on the mouth and chin; Euclidean distances were used for all other landmarks.

Our evaluation of descriptors is based on the expected local accuracy $\bar{\mathbf{e}}_k$, which quantifies the expected localization error of a descriptor when it is evaluated in a local neighborhood[6] of the target [56] and the required number of candidates ϱ_k, as defined in Section II-A2. To avoid biasing the localization results, we evaluated the descriptors only within each fold of the cross-validation split. Results for the 1st fold of FRGC v1 are provided in Supplementary Table I. The descriptors finally chosen for each landmark are highlighted in blue. The criterion used was to include a new descriptor only if there were none already included that could achieve comparable performance (i.e., not significantly different from the best). This directly led to the choice of D_{AR}, A+R, and A+D_{AR} and either A_0 or A+D_{AER}. However, 1-ring APSC are faster to compute than 2-rings, therefore we choose A_0.

[6]The neighborhoods used to compute $\bar{\mathbf{e}}_k$ are determined as the nearest rings around each targeted point for which accuracy is stable. These neighborhoods can play a role when comparing descriptors but that was not the case in these experiments, hence we omit them here. Please refer to [56] for details.

TABLE II
SUMMARY OF COMPARED METHODS ON THE FRGC DATABASE

Method	# of Lmk	# scans tested	Decimation	Smoothing filter	Hole filling
Alyuz et al. [29]	5	v2: 4007	none	yes	yes
Colbry [58]	9	v1: 953	1 : 4	yes	
Creusot et al. [27]	14	v1: 943 v2: 4007	1 : 32		
Lu & Jain [59]	7	v1: 946	1 : 4		
Lu & Jain [60]	7	v1: 953	1 : 4		
Passalis et al. [19]	8	v2: 975	1 : 4	yes	yes
Perakis et al. [20]	8	v2: 975	1 : 4	yes	yes
Segundo et al. [21]	4	v1: 943 v2: 4007	none	yes	
Sukno et al. (SRILF)	14	v1: 943 v2: 4007	1 : 4	yes	
Szeptycki et al. [34]	9	v1: 462	none	yes	yes
Yu & Moon [24]	3	v1: 200	none		
Zhao et al. [35]	15	v1: 462 v2: 1400	none	yes	yes
Zhao et al. [36]	15	v1: 462 v2: 1500	none	yes	yes

Results on the 2nd fold of FRGCv1 were similar to those discussed above, hence we kept the same selection of descriptors for all experiments in this paper. Note that this relates to what descriptors were used but not to the number of candidates retained ϱ, which must be recomputed for each training set.

C. Localization Accuracy on FRGC

In this section, we compare localization errors for the 14 targeted landmarks measured as the Euclidean distance to the manual annotations. We provide results for SRILF together with results reported in the literature from other 12 methods. A summary of the experimental settings of all compared methods is provided in Table II, including the total number of landmarks targeted and the size of the test sets that were reported by their authors. Decimation is often used, with 1:4 being the preferred factor because it allows to reduce computational load without impairing accuracy. Most methods apply smoothing filters to deal with spikes and noise in the range data and a few of them apply also hole-filling. Thus, we see from Table II that our experimental settings are similar to the majority of compared methods.

Table III gathers the localization errors reported on FRGCv1. It can be seen that, among methods using only geometric information, our results are the best for all landmarks other than the nose tip, where Szeptycki *et al.* [34] and Yu and Moon [24] obtain averages about half a millimeter lower. However, in both cases the errors of these methods in the rest of landmarks make them far less accurate than SRILF.

When considering methods that combine both geometric and texture information, we find that the two methods

TABLE III
LANDMARK LOCALIZATION ERRORS REPORTED ON FRGC V1, IN TERMS OF MEAN ± STANDARD DEVIATION. VALUES IN [MM]

Method	Eyes		Nose				Mouth and chin			
	en	ex	n	prn	ac	sn	ch	ls	li	pg
A. Approaches based on geometric cues only										
Creusot et al. [27]	4.67 ±2.26	6.25 ±3.35	4.50 ±2.48	4.07 ±2.16	4.14 ±2.37	3.39 ±1.71	4.84 ±2.94	3.62 ±2.19	4.68 ±2.40	5.46 ±2.98
Lu & Jain [60]	8.25 ±17.2	9.9 ±17.6	- -	8.3 ±19.4	- -	- -	6.1 ±17.9	- -	- -	- -
Segundo et al. [21]	4.21 ±3.33	- -	- -	2.69 ±2.14	6.69 ±2.93	- -	- -	- -	- -	- -
Sukno et al. (SRILF)	3.57 ±1.76	4.71 ±2.79	**2.76 ±1.76**	2.77 ±1.68	**3.17 ±1.83**	**2.36 ±1.24**	**3.23 ±2.19**	2.83 ±1.62	3.82 ±1.95	**4.24 ±2.46**
Szeptycki et al. [34]	3.85 ±2.03	7.96 ±3.87	- -	2.27 ±1.35	6.18 ±4.23	- -	8.56 ±7.47			- -
Yu & Moon [24]	5.17 ±13.30	- -	- -	**2.18 ±6.83**	- -	- -	- -	- -	- -	- -
B. Approaches combining both geometry and texture										
Colbry [58]	5.8 ±4.75	- -	4.8 ±6.4	4.0 ±5.4	- -	4.1 ±5.9	5.4 ±6.75	- -	- -	11.7 ±7.3
Lu & Jain [59]	5.85 ±3.15	7.5 ±5.51	- -	5.0 ±2.4	- -	- -	3.6 ±3.11	- -	- -	- -
Zhao et al. [35]	3.21 ±1.97	4.27 ±2.82	- -	2.68 ±1.85	4.47 ±3.69	- -	3.93 ±2.53	**2.72 ±1.51**	**3.76 ±2.07**	- -
Zhao et al. [36]	**3.11 ±1.49**	**3.92 ±2.02**	- -	4.11 ±2.20	4.18 ±1.75	- -	3.60 ±1.96	2.74 ±1.42	3.81 ±1.97	- -

TABLE IV
LANDMARK LOCALIZATION ERRORS ON FRGC V2 USING MODELS TRAINED ON FRGC V1, IN TERMS OF MEAN ± STANDARD DEVIATION. VALUES IN [MM]

Method	Eyes		Nose				Mouth and chin			
	en	ex	n	prn	ac	sn	ch	ls	li	pg
A. Approaches based on geometric cues only										
Creusot et al. [27]	4.30 ±2.05	5.93 ±3.08	4.22 ±2.47	3.36 ±1.95	3.72 ±1.72	3.65 ±1.61	5.57 ±3.41	4.26 ±2.63	5.47 ±3.90	6.72 ±4.15
Segundo et al. [21]	3.52 ±2.30	- -	- -	2.73 ±1.39	5.34 ±1.89	- -	- -	- -	- -	- -
Sukno et al. (SRILF)	3.35 ±1.63	4.49 ±2.64	**2.55 ±1.60**	**2.22 ±1.31**	**3.09 ±1.18**	**2.81 ±1.11**	**4.05 ±3.12**	3.40 ±1.97	4.82 ±4.04	**5.39 ±4.01**
B. Approaches combining both geometry and texture										
Zhao et al. [35]	4.07 ±2.07	5.10 ±2.99	- -	4.88 ±2.52	6.80 ±4.37	- -	5.03 ±3.07	3.53 ±1.86	6.48 ±3.16	- -
Zhao et al. [36]	**3.23 ±1.44**	**4.10 ±2.05**	- -	4.43 ±2.56	4.64 ±2.06	- -	4.22 ±2.41	**3.37 ±1.89**	**4.65 ±3.41**	- -

TABLE V
LANDMARK LOCALIZATION ERRORS WHERE BOTH TRAINING AND TEST SETS DERIVE FROM FRGC V2, IN TERMS OF MEAN ± STANDARD DEVIATION. VALUES IN [MM]

Method	Eyes		Nose				Mouth and chin			
	en	ex	n	prn	ac	sn	ch	ls	li	pg
A. Approaches based on geometric cues only										
Alyuz et al. [29]	4.98 n/a	- -	- -	3.26 n/a	4.60 n/a	- -	- -	- -	- -	- -
Passalis et al. [19]	5.25 ±2.53	5.71 ±3.46	- -	4.91 ±2.49	- -	- -	6.06 ±4.30	- -	- -	6.31 ±4.43
Perakis et al. [20]	4.28 ±2.42	5.71 ±3.38	- -	4.09 ±2.41	- -	- -	5.49 ±3.89	- -	- -	4.92 ±3.74
Sukno et al. (SRILF)	**3.54 ±1.74**	**4.63 ±2.67**	2.53 ±1.63	**2.34 ±1.70**	**2.62 ±1.35**	2.70 ±1.12	**3.87 ±2.77**	3.31 ±1.83	4.55 ±3.39	**4.91 ±3.54**

by Zhao *et al.* [35], [36] perform better than SRILF for the eye corners (both inner and outer). However, for the other seven landmarks that can be compared, SRILF produces either equivalent or better performance than all methods using texture, even though we use only geometric information. A similar trend can be observed in the results for FRGCv2 (Tables IV and V) for which we split the comparison in two, depending on whether the algorithms were trained on FRGCv1 or FRGCv2.

Training on FRGCv1 and testing on FRGCv2 is the most challenging scenario, as the training data do not contain strong facial expressions but these are present in the test set. We can analyze how this affects accuracy by comparing the results from Tables III and IV, i.e., training with FRGCv1

TABLE VI
LANDMARK LOCALIZATION ERRORS ON THE CLINICAL DATASET, IN TERMS OF MEAN ± STANDARD DEVIATION. VALUES IN [MM]

Method	Eyes		Nose				Mouth and chin			
	en	ex	n	prn	ac	sn	ch	ls	li	pg
Sukno et al. (SRILF)	1.73 ±1.07	3.21 ±1.99	1.72 ±1.20	1.90 ±1.29	2.01 ±1.27	1.83 ±1.12	2.55 ±1.69	2.19 ±1.27	2.35 ±1.38	3.35 ±2.12

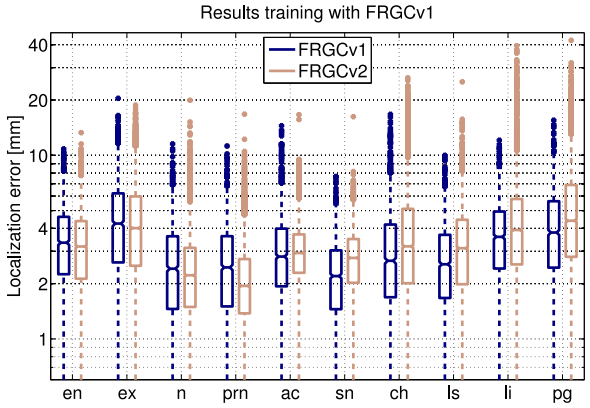

Fig. 2. Landmark localization errors of SRILF using training sets derived from FRGCv1 for test sets from FRGCv1 (cross-validation) and FRGCv2.

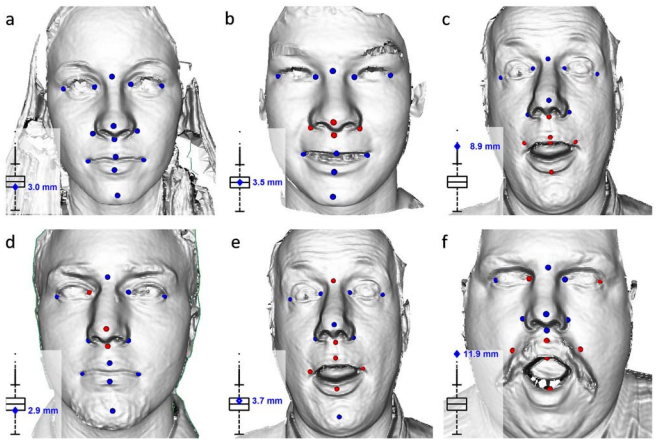

Fig. 3. Examples of landmark localization in FRGCv2 using SRILF trained on FRGCv1 (top row) and FRGCv2 with cross validation (bottom row). Landmarks identified based on candidates (included in $\mathbf{x}^f$) are displayed in blue, while inferred landmarks are displayed in red. We show also the overall error for each case and a boxplot of the overall errors for all scans in the test set where we can see the position of each example. Thus, (a), (b), (d), and (e) are representative examples while (c) and (f) are extreme cases, showing nearly worst-case performance. More examples available on-line.[4]

and testing on FRGCv1 or FRGVCv2, respectively. We can see that SRILF and the methods by Creusot *et al.* [27] and Segundo *et al.* [21] maintain their accuracy for all landmarks in the eyes and most of the nose but not for the mouth and chin landmarks (Segundo *et al.* [21] do not target them and the other two methods show increased errors). The algorithms from Zhao *et al.* [35], [36] show increased errors for most landmarks when testing on FRGCv2. In the case of [35], errors grow significantly for all landmarks while the method in [36], which incorporates an occlusion model, is able to maintain accuracy for the eyes and, to some extent, also the nose. Landmarks in the mouth and the chin clearly show higher errors.

Further details for SRILF are provided in Fig. 2. The strong facial expressions on FRGCv2 result in increased errors in the lower part of the face; the boxplots in Fig. 2 show that this is better explained by a rise in the number and the strength of outliers than by an actual change in overall accuracy (indicated by the medians). This is a rather straight-forward consequence of the mismatch between training and test sets, as illustrated in Fig. 3, top row. FRGCv1 is not a representative training set for some of the scans with strong facial expressions in FRGCv2. In those cases, landmarks in the lower part of the face cannot be identified correctly as both the local geometry around landmarks and the global shape defined from them deviate considerably from the statistics of the training set.

The above can be dealt with by deriving training and test sets from FRGCv2, which in our case was done by means of 2-fold cross validation (Fig. 3, bottom row). Now the algorithm can also tackle cases with strong facial expressions, as these are present in the training set: for example, in Fig. 3 the images in (c) and (e) correspond to the same scan, but localization results are considerably better in (e).

Some limitations of SRILF can be observed in those cases where all included candidates correspond to the upper part of the face. The latter is illustrated in Fig. 3(f): all candidates included in $\mathbf{x}^f$ correspond to landmarks above the upper lip, while all landmarks from the lower part have been inferred (i.e., they are in $\mathbf{x}^g$). As the majority of examples in FRGCv2 have a closed mouth, so does the most probable estimate unless there is image evidence that contradicts it. A possible solution would be to force that at least one of the candidates included in $\mathbf{x}^f$ corresponds to a landmark in the lower part of the face, providing the necessary constraints for a more accurate estimate [e.g., as in Fig. 3(e)]. However, this problem was limited to very few cases: as it can be appreciated in the boxplots attached to each example, Fig. 3(c) and (f) correspond to extreme cases, which are clearly outliers in terms of localization accuracy.

Another figure of merit used to assess the performance of localization algorithms is the landmark detection rate, i.e., the percentage of landmarks that were localized within a given radius from the ground truth. While this is a weaker measure than the average errors provided in Tables III–V, it relates to robustness and is sometimes reported. Thus, we provide detection rates in Supplementary Table II. These conform SRILF as the top-performing approach for most landmarks. Additionally, the landmark coordinates obtained by SRILF for all experiments on FRGC are provided on-line,[4] together with several example videos that illustrate the behavior of SRILF in best, worst, and typical cases.

D. Landmarks and Complexity

As mentioned earlier, our primary interest is on highly accurate localization of facial landmarks. In contrast, some of the methods compared in the previous sections focus on computational complexity and can extract landmark coordinates in about 1 s per facial scan [21], [27].

While we do not target low complexity, it is important to compare SRILF with the flexible landmark models (FLMs) presented by Passalis *et al.* [19] and Perakis *et al.* [20]. These recent methods share with SRILF the use of a statistical model to validate combinations of landmark candidates but cannot handle incomplete sets. The strategy used in FLMs is to tolerate large numbers of false positives, in an attempt to avoid any false negatives in at least one side of the face. Hence, they need to retain a large number of candidates for each landmark. In contrast, we retain a smaller number of candidates and handle false negatives as missing information that is completed by inference from the statistical model.

The computational cost of both SRILF and FLMs depends on the number of landmarks that are targeted. Hence, we repeated the experiments on FRGCv1 targeting different subsets of landmarks. We started with a subset of 8 points that matches the landmarks targeted by FLMs and successively added points until reaching the full set of 14 landmarks. The results are summarized in Supplementary Table III and includes the following.

1) Localization errors per landmark, to verify whether using smaller subsets (with fewer constraints) has an impact on the accuracy of the algorithm.
2) Computational cost, measured as the average run-time on a PC equipped with an Intel Core i3-2120 CPU @ 3.30 GHz with 4 GB RAM. Reported results correspond to a C++ implementation using the Armadillo library [61] for the matrix inversions and OpenMP [62] for parallelization.

The first conclusion that can be extracted is that localization errors did not vary much for the different subsets. The largest variations were observed in the eye corners, which showed slightly higher errors when fewer landmarks were targeted. However, these differences were always within 5% of the errors obtained when targeting the full set of 14 landmarks.

SRILF required 4.7 s to locate 8 landmarks and approximately 31.5 s to target the full set. We can compare the results when targeting eight landmarks with those reported using FLMs to target the same subset, which averaged 6.68 s on a PC comparable to the one used here [20]. In both cases we can clearly isolate the time taken by the combinatorial search, thus highlighting the difference between our strategy of using incomplete sets of landmarks (0.54 s) and the one used in FLMs of trying to always find the complete set, which was reported to average 6.07 s.

In terms of scalability, an approximate analysis can be done by assuming a constant number of candidates, N_c, retained for all landmarks. Targeting an additional landmark with FLMs multiplies the number of combinations to test by N_c (or $\sqrt{N_c}$ if the extra landmark is symmetric). In SRILF we test $\binom{L}{4}N_c^4$ combinations, so targeting $L+1$ landmarks increases the combinations to test only by a factor of $(L+1)/(L-3)$.

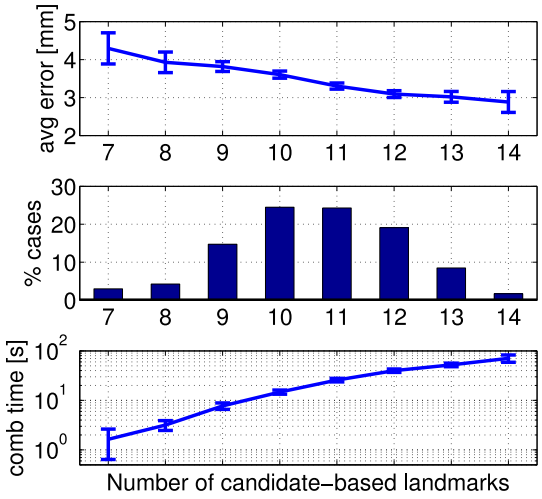

Fig. 4. Results on FRGCv1 grouped by number of candidate-based landmarks (cardinality of $\mathbf{x}^f$). Top: average error over all landmarks. Middle: percentage of scans with the number of candidate-based landmarks indicated by the horizontal axis. Bottom: average run-time of the combinatorial search. In the top and bottom plots, bars indicate a 95% confidence interval of the mean.

Therefore, SRILF not only outperforms FLMs in the concrete case of localizing eight landmarks as in [19] and [20], but it also scales better when additional landmarks are targeted, since $(L+1)/(L-3)$ quickly tends to the unit as we increase L. Also, as already mentioned, recall that SRILF needs to retain less candidates than FLMs, which results in smaller values of N_c, i.e., $N_c^{\text{SRILF}} < N_c^{\text{FLM}}$ and typically $N_c^{\text{FLM}} \gg 1$.

It is worth emphasizing that the complexity of the combinatorial search depends not only on the number of targeted landmarks but also on the number of candidates included into $\mathbf{x}^f$ case by case, as shown in Fig. 4. Having to test larger subsets of candidates increases the complexity but also reduces the number of landmarks that must be inferred and, on average, localization errors. It can be seen that in the majority of cases (82.6%) there were between 9 and 12 landmarks identified based on candidates (i.e., included in $\mathbf{x}^f$) while the remaining 2 to 5 landmarks were inferred from the model statistics.

Finally, it should be noted that both the computation of descriptors and the combinatorial search involve a large number of operations that are inherently independent. Therefore, the algorithm could in principle be accelerated substantially through parallelization (e.g., by using GPUs).

E. Occlusions and Out-of-Plane Rotations

An interesting by-product of the strategy followed by SRILF is that it can naturally handle cases with occlusions or missing data. Let us emphasize that, up to this point, we have referred to missing landmarks as those for which feature detectors did not produce suitable candidates, although the vast majority of the test surfaces did not present occlusions or missing parts.

In this section, we present tests on the Bosphorus database [51], which offers the possibility to test scans with actual occlusions (due to hair, glasses or hands covering the face) and scans where part of the surface was not captured due to self-occlusions generated by large out-of-plane rotations.

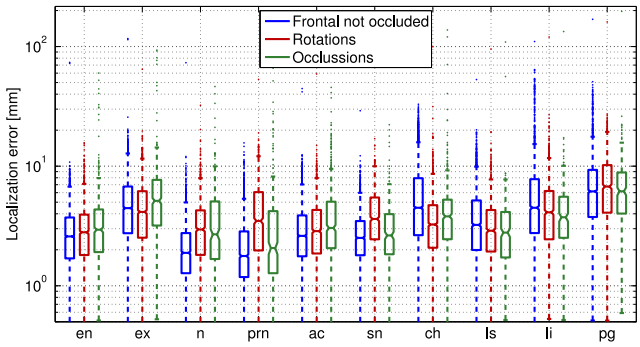

Fig. 5. Landmark localization errors of SRILF on the Bosphorus database. We show separately the errors for frontal scans without occlusions (blue), scans with rotations (red), and scans with occlusions (green).

The Bosphorus database contains scans from 105 subjects showing expressions, facial action units, and as mentioned above, rotations and occlusions. To facilitate comparison to other works, we selected the same 4339 facial scans used by Creusot *et al.* [27], namely all available scans but those with 90 degree rotations or flagged as invalid. We proceeded analogously as done with the FRGC database, including decimation by 1:4 which resulted in an average of approximately 9240 vertices per facial scan.

Fig. 5 shows the localization results for each landmark, discriminated in three sets: 2803 frontal scans without occlusions (most of which show expressions or action units), 1155 scans with out-of-plane rotations and 381 scans with occlusions. Models were constructed using exclusively scans that are frontal without occlusions, so that they could not learn from occluded or rotated scans. Experiments were carried our under 2-fold cross validation ensuring that no subject was part of both training and test sets at the same time.

Comparing the errors for the three sets in Fig. 5, we can see that the overall performance is maintained for a majority of landmarks. On the other hand, it is also clear that the presence of occlusions and rotations increases the percentage of outliers. Fig. 6 shows snapshots of rotated and occluded cases, as well as some especially challenging scans where the algorithm produces errors considerably larger than the average.

Numeric results, in terms of localization errors and detection rates, are provided in Supplementary Tables IV and V. Similarly to FRGC, comparison to other state of the art algorithms is favorable for most landmarks. The landmark coordinates obtained by SRILF for all tested scans are also provided on-line[7] together with a large collection of snapshots of the localized landmarks.

F. Localization Accuracy for Clinical Data

The experiments presented in the previous sections aimed at testing the robustness of SRILF and its performance in relation to state of the art approaches using large public databases in which there are acquisition artifacts, occlusions, strong facial expressions, and a nonnegligible degree of noise in the manual annotations, as discussed in [57].

[7]http://www.cipa.dcu.ie/face3d/SRILF_Examples_Bosphorus.htm [19.03.2014]

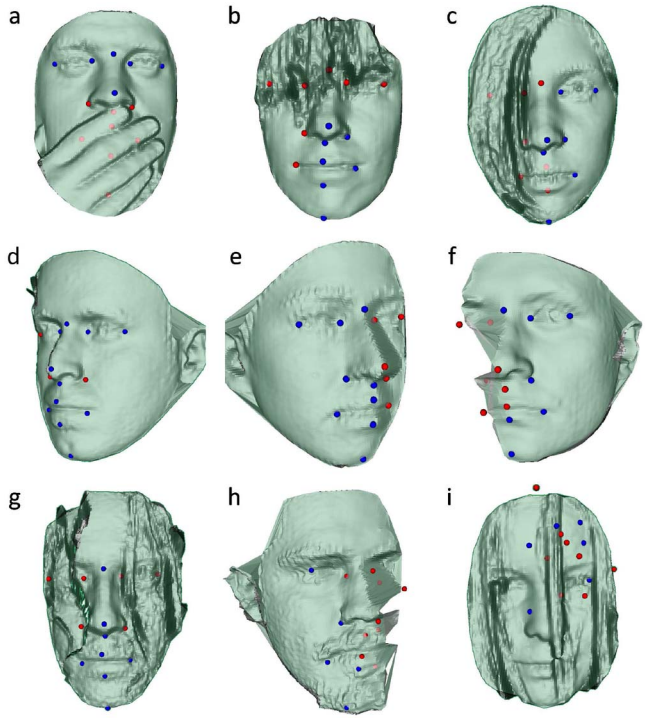

Fig. 6. Examples of landmark localization in the Bosphorus database using SRILF. (a)–(c) Scans with occlusions. (d)–(f) Scans with out-of-plane rotations that produce large missing parts of the surface. (g)–(i) Especially challenging cases. Examples (a)–(g) correspond to average performance while (h) and (i) have larger errors; in particular, (i) shows the worst result obtained in this database. Landmarks identified based on candidates (included in $\mathbf{x}^f$) are displayed in blue, while inferred landmarks are displayed in red.

In this section, we explore how much can we reduce localization errors by testing SRILF on a clinical dataset where special care has been taken to minimize the presence of artifacts and manual annotations have been performed by experts. The dataset consists of 144 facial scans acquired by means of a hand-held laser scanner[8] with an average of approximately 44 200 vertices per mesh.

The dataset contains exclusively healthy volunteers who acted as controls in the context of craniofacial dysmorphology research. All scans are from different individuals (i.e., one scan per person) and volunteers were asked to pose with neutral facial expressions. Each scan was annotated with a number of anatomical landmarks [63], among which we target the same 14 points as in the previous experiments. Due to the moderate size of the dataset, we used 6-fold cross-validation so that training sets would always contain 120 scans. All parameters were kept as in Section III-B.

Results are shown in Fig. 7 and Table VI. Average errors are below 3.5 mm for all landmarks and within 2 mm for half of them. However, we can also see that averages are still importantly affected by the presence of outliers and the median errors are at or below 2 mm for the majority of landmarks. Recent work in the clinical domain suggests that human observers could annotate facial landmarks with errors between

[8]Polhemus FastSCAN^{TM}, Colchester, VT, USA. Example available at http://www.cipa.dcu.ie/videos/face3d/Scanning_DCU_RCSI.avi [20.05.2013].

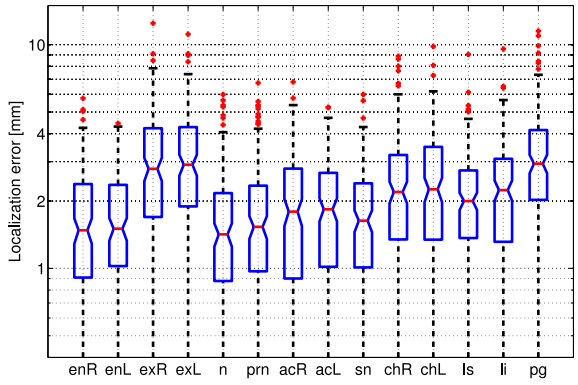

Fig. 7. Landmark localization errors of SRILF on the clinical dataset.

1 and 2 mm [13], [14], which would be an acceptable accuracy for craniofacial dysmorphology applications [7].

Comparing these results to those in FRGCv1 (the part of FRGC with less variations due to facial expressions), the overall localization error is more than 1 mm lower: 3.44 mm on FRGCv1, 2.31 mm on the clinical dataset. Looking at each landmark individually, all of them have lower average errors in the clinical dataset. In relative terms, the reduction in average errors ranged from slightly above 20% (*sn*, *ch*, *ls*, *pg*) to more than 50% (*en*). In both datasets the chin tip and outer-eye corners were the most difficult points to locate.

IV. Discussion

The experiments presented in Section III have shown that our algorithm can locate facial landmarks with an accuracy that is comparable or better than state of the art methods. The methods compared can be divided into two categories.

1) Approaches based on geometric cues only: these are the most direct competitors, as our algorithm belongs to this category. A combined analysis of Tables III–V shows that SRILF always obtained lower localization errors than all other geometric methods for 12 out of the 14 tested landmarks (*en*, *ex*, *n*, *ac*, *sn*, *ch*, *ls*, and *li*). For the remaining two landmarks, SRILF was the most accurate method for the nose tip in FRGCv2 but not in FRGCv1; while it was the most accurate for the chin tip in all three experiments, results were similar to the method by Perakis *et al.* [20] in FRGCv2.
2) Approaches combining both geometry and texture: in principle, these methods have an advantage over SRILF, not only because they incorporate an additional source of information but also because manual annotations for FRGC have been derived from 2-D images and could therefore have some bias toward texture. However, the results reveled that our algorithm was still as accurate or better than texture-based methods for the majority of compared landmarks and it only produced consistently higher errors for the eye corners. Nonetheless, this increase was in all cases below 20%.

In terms of average errors over all targeted landmarks, SRILF obtains the best results (3.4 to 3.7 mm), followed by the method from Zhao *et al.* [36] (3.7 to 4.1 mm). Interestingly, these two methods share the concept of using partial sets of landmarks if there is no information available for the complete set. In the case of Zhao *et al.* [36] this is achieved by using an occlusion detection block, which indicates whether the image information for a given landmark should be used or discarded (presumably due to an occlusion). Comparison to prior work of the same authors without occlusion detection [35] yields similar errors in FRGCv1 but considerably higher errors in FRGCv2. However, the number of scans with occlusions (or missing parts of the surface) in FRGC is limited and affects a rather small percentage of the data, suggesting that the information that is discarded is not restricted to occluded data but also includes regions where the image features are not reliable (e.g., do not match the statistics from the training set).

Among remaining methods that target landmarks in all facial regions, those from Creusot *et al.* [27], Perakis *et al.* [20], and Passalis *et al.* [19] are the most accurate, although their overall errors are above 4.5 mm. These three methods also include strategies to handle partial information: Creusot *et al.* [27] use combinatorial search based on RANSAC but constrained to a rigid model that can be scaled but not deformed, while Passalis *et al.* [19] and Perakis *et al.* [20] exploit bilateral symmetry to account for cases where information is complete only for one side of the face, but without providing estimates for the other side. The method by Fanelli *et al.* [32] is yet another recent work using partial information to target facial landmarks; this, unfortunately, has not yet been reported on FRGC.

V. Conclusion

In this paper, we present SRILF for the automatic detection of facial landmarks. The algorithm generates sets of candidate points from geometric cues extracted by using APSC descriptors and performs combinatorial search constrained by a flexible shape model. A key assumption of our approach is that some landmarks might not be accurately detected by the descriptors, which we tackle by using partial subsets of landmarks and inferring those that are missing from the flexible model constraints.

We evaluated the proposed method in the FRGC database, where we obtained average errors of approximately 3.5 mm when targeting 14 prominent facial landmarks. For the majority of these our method produces the most accurate results reported to date in this database. This was verified even for methods combining both geometry and texture, which outperformed SRILF only when targeting eye corners, suggesting that texture information might be of special importance in the localization of the eyes. It was also shown that smaller subsets of landmarks could be targeted while keeping accuracy essentially constant and reducing computational cost.

From the 12 methods that we included for comparison, those that achieved the best results shared with SRILF some ability to use partial information. There seems to be a trend indicating that most successful methods for landmark localization are those that can dynamically determine (on a case by case basis) what information to rely on and what information to discard or ignore. In this sense, SRILF provides a general framework that integrates nonrigid deformation with the ability to handle any combination of missing points.

We also investigated the performance of our algorithm on data with occlusions and out-of-plane rotations, as well as potential limits in the accuracy that could be reached. Testing our algorithm against expert annotations in a clinical dataset, we found that SRILF could localize facial landmarks with an overall accuracy of 2.3 mm, with typical errors below 2 mm for more than half of the targeted landmarks. Nonetheless, this relates only to overall performance and cannot be guaranteed for all individual cases. Thus, further efforts should concentrate on reducing the number and the strength of outliers.

ACKNOWLEDGMENT

The authors would like to thank their colleagues in the Face3D Consortium (www.face3d.ac.uk).

REFERENCES

[1] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1–15, 2006.

[2] O. Celiktutan, S. Ulukaya, and B. Sankur, "A comparative study of face landmarking techniques," in *Proc. EURASIP J. Image Video Process.*, 2013, pp. 1–13.

[3] S. Gupta, M. Markey, and A. Bovik, "Anthropometric 3D face recognition," *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 331–349, 2010.

[4] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1927–1943, Nov. 2007.

[5] E. Vezzetti and F. Marcolin, "3D human face description: Landmarks measures and geometrical features," *Image Vis. Comput.*, vol. 30, no. 10, pp. 696–712, 2012.

[6] X. Zhao, G. Evangelopoulos, D. Chu, S. Shah, and I. A. Kakadiaris, "Minimizing illumination differences for 3D to 2D face recognition using lighting maps," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 725–736, May 2014.

[7] K. Chinthapalli *et al.*, "Atypical face shape and genomic structural variants in epilepsy," *Brain*, vol. 135, no. 10, pp. 3101–3114, 2012.

[8] R. Hennessy, P. A. Baldwin, D. J. Browne, A. Kinsella, and J. L. Waddington, "Frontonasal dysmorphology in bipolar disorder by 3D laser surface imaging and geometric morphometrics: Comparison with schizophrenia," *Schiz. Res.*, vol. 122, nos. 1–3, pp. 63–71, 2010.

[9] T. Mutsvangwa *et al.*, "Design, construction and testing of a stereophotogrammetric tool for the diagnosis of fetal alcohol syndrome in infants," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1448–1458, Sep. 2009.

[10] H. Popat, S. Richmond, A. I. Zhurov, P. L. Rosin, and D. Marshall, "A geometric morphometric approach to the analysis of lip shape during speech: Development of a clinical outcome measure," *PLoS ONE*, vol. 8, no. 2, p. e57368, 2013.

[11] A. Sharifi *et al.*, "How accurate is model planning for orthognathic surgery?" *Int J. Oral Max. Surg.*, vol. 37, no. 12, pp. 1089–1093, 2008.

[12] P. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. San Diego, CA, USA, 2005, pp. 947–954.

[13] J. Plooij *et al.*, "Evaluation of reproducibility and reliability of 3D soft tissue analysis using 3D stereophotogrammetry," *Int J. Oral Max. Surg.*, vol. 38, no. 3, pp. 267–273, 2009.

[14] N. Aynechi, B. E. Larson, V. Leon-Salazar, and S. Beiraghi, "Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition," *Angle Orthod.*, vol. 81, no. 2, pp. 245–252, 2011.

[15] P. Besl and R. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 2, pp. 167–192, Mar. 1988.

[16] C. Dorai and A. Jain, "COSMOS–A representation scheme for 3D free-form objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 10, pp. 1115–1130, Oct. 1997.

[17] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recogn.*, vol. 39, no. 3, pp. 444–455, 2006.

[18] H. Dibeklioglu, A. Salah, and L. Akarun, "3D facial landmarking under expression, pose, and occlusion variations," in *Proc. 2nd IEEE Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, 2008, pp. 1–6.

[19] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3D face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1938–1951, Oct. 2011.

[20] N. Perakis, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection under large yaw and expression variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1552–1564, Jul. 2013.

[21] M. Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1319–1330, Oct. 2010.

[22] T. Faltemier, K. Bowyer, and P. Flynn, "Rotated profile signatures for robust 3D feature detection," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Amsterdam, The Netherlands, 2008, pp. 1–7.

[23] X. Peng, M. Bennamoun, and A. Mian, "A training-free nose tip detection method from face range images," *Pattern Recogn.*, vol. 44, no. 3, pp. 544–558, 2011.

[24] T. Yu and Y. Moon, "A novel genetic algorithm for 3D facial landmark localization," in *Proc. 2nd IEEE Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, Arlington, VA, USA, 2008, pp. 1–6.

[25] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise localization of landmarks on 3D faces using Gabor wavelets," in *Proc. 1st IEEE Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, Crystal City, VA, USA, 2007, pp. 1–6.

[26] C. Conde, L. Rodriguez-Aragon, and E. Cabello, "Automatic 3D face feature points extraction with spin images," in *Proc. 3rd Int. Conf. Image Anal. Recognit. (ICIAR)*, 2006, pp. 317–328.

[27] C. Creusot, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3D meshes," *Int. J. Comput. Vis.*, vol. 102, pp. 146–179, Mar. 2013.

[28] M. Romero-Huertas and N. Pears, "Landmark localization in 3D face data," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Surveillance (AVSS)*, Genova, Italy, 2009, pp. 73–78.

[29] N. Alyuz, B. Gokberk, and L. Akarun, "Regional registration for expression resistant 3-D face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 425–440, Sep. 2010.

[30] Y. Sun and L. Yin, "Automatic pose estimation of 3D facial models," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Tampa, FL, USA, 2008, pp. 1–4.

[31] S. Jahanbin, A. Bovik, and H. Choi, "Automated facial feature detection from portrait and range images," in *Proc. IEEE Southwest. Symp. Image Anal. Interpret. (SSIAI)*, 2008, pp. 25–28.

[32] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Gool, "Random forests for real time 3D face analysis," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 437–458, 2012.

[33] P. Nair and A. Cavallaro, "3-D face detection, landmark localization and registration using a point distribution model," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 611–623, Jun. 2009.

[34] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *Proc. 3rd IEEE Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, 2009, pp. 1–6.

[35] X. Zhao, P. Szeptycki, E. Dellandrea, and L. Chen, "Precise 2.5D facial landmarking via an analysis by synthesis approach," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, Snowbird, UT, USA, 2009, pp. 1–7.

[36] X. Zhao *et al.*, "Accurate landmarking of three-dimensional facial data in the presence of facial expression and occlusions using a three-dimensional statistical facial feature model," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 5, pp. 1417–1428, Oct. 2011.

[37] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 5, pp. 381–395, 1981.

[38] F. Sukno, J. Waddington, and P. Whelan, "3D facial landmark localization using combinatorial search and shape regression," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 32–41.

[39] M. de Bruijne, M. T. Lund, L. B. Tankó, P. P. Pettersen, and M. Nielsen, "Quantitative vertebral morphometry using neighbor-conditional shape models," *Med. Image Anal.*, vol. 11, no. 5, pp. 503–512, 2007.

[40] J. Baileul, S. Ru, and J.-M. Constans, "Statistical shape model-based segmentation of brain MRI images," in *Proc. IEEE Conf. Eng. Med. Biol. Soc. (EMBS)*, 2007, pp. 5255–5258.

[41] J. Hug, C. Brechbuhler, and G. Szekely, "Model-based initialization for segmentation," in *Proc. 6th Eur. Conf. Comput. Vis. (ECCV)*, Dublin, Ireland, 2000, pp. 290–306.

[42] Y. Yang, D. Rueckert, and A. Bull, "Predicting the shapes of bones at a joint: Application to the shoulder," *Comput. Methods Biomech. Biomed. Eng.*, vol. 11, no. 1, pp. 19–30, 2008.
[43] M. Luthi, T. Albrecht, and T. Vetter, "Probabilistic modeling and visualization of the flexibility in morphable models," in *Proc. Math. Surfaces.*, York, U.K., 2009, pp. 251–264.
[44] R. Blanc, E. Syrkina, and G. Székely, "Estimating the confidence of statistical model based shape prediction," in *Proc. Inf. Process. Med. Imag. (IPMI)*, 2009, pp. 602–613.
[45] A. Rao, P. Aljabar, and D. Rueckert, "Hierarchical statistical shape analysis and prediction of sub-cortical brain structures," *Med. Image Anal.*, vol. 12, no. 1, pp. 55–68, 2008.
[46] T. Liu, D. Shen, and C. Davatzikos, "Predictive modeling of anatomic structures using canonical correlation analysis," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, 2004, pp. 1279–1282.
[47] K. Lekadir and G.-Z. Yang, "Optimal feature point selection and automatic initialization in active shape model search," in *Proc. 11th Int. Conf. Med. Image Comput. Comput.-Assist. Intervention (MICCAI)*, New York, NY, USA, 2008, pp. 434–441.
[48] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions and pose variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, Sep. 2013.
[49] F. Sukno, J. Waddington, and P. Whelan, "Rotationally invariant 3D shape contexts using asymmetry patterns," in *Proc. 8th Int. Conf. Comput. Graphics Theory and App.*, 2013, pp. 7–17.
[50] A. Frome *et al.*, "Recognizing objects in range data using regional point descriptors," in *Proc. 8th Eur. Conf. Comput. Vis. (ECCV)*, 2004, pp. 224–237.
[51] A. Savran *et al.*, "Bosphorus database for 3D face analysis," in *Proc. 1st Eur. Workshop Biometrics Identity Manage. (BIOID)*, 2008, pp. 47–56.
[52] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.
[53] T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," Wolfson Image Analysis Unit, Univ. Manchester, U.K., Tech. Rep., 2001.
[54] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recogn.*, vol. 33, no. 1, pp. 25–41, 2000.
[55] P. Rousseeuw, "Least median of squares regression," *J. Amer. Stat. Assoc.*, vol. 79, no. 388, pp. 871–880, 1984.
[56] F. Sukno, J. Waddington, and P. Whelan, "Comparing 3D descriptors for local search of craniofacial landmarks," in *Proc. 8th Int. Symp. Adv. Vis. Comput. (ISVC)*, Crete, Greece, 2012, pp. 92–103.
[57] F. Sukno, J. Waddington, and P. Whelan, "Compensating inaccurate annotations to train 3D facial landmark localization models," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, Shanghai, China, 2013, pp. 1–8.
[58] D. Colbry, "Human face verification by robust 3D surface alignment," Ph.D. Dissertation, Dept Comput. Sci., Michigan State Univ., East Lansing, MI, USA, 2006.
[59] X. Lu and A. Jain, "Multimodal facial feature extraction for automatic 3D face recognition," Dept. Comput. Sci., Michigan State Univ., East Lansing, MI, USA, Tech. Rep. MSU-CSE-05-22, 2005.
[60] X. Lu and A. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, Southampton, U.K., 2006, pp. 585–590.
[61] C. Sanderson, "Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments," NICTA, Tech. Rep., 2010.
[62] L. Dagum and R. Menon, "OpenMP: An industry standard API for shared-memory programming," *IEEE Comput. Sci. Eng.*, vol. 5, no. 1, pp. 46–55, Jan./Mar. 1998.
[63] R. Hennessy, A. Kinsella, and J. Waddington, "3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis: Initial application to sexual dimorphism," *Biol. Psychiat.*, vol. 51, no. 6, pp. 507–514, 2002.

Federico M. Sukno received the degree in electrical engineering from La Plata National University, La Plata, Argentina, and the Ph.D. degree in biomedical engineering from Zaragoza University, Zaragoza, Spain, in 2000 and 2008, respectively.

From 2010 to 2013, he was a Post-Doctoral Researcher with a joint appointment at Dublin City University, Dublin, Ireland, and the Royal College of Surgeons in Ireland, Dublin. He joined Pompeu Fabra University, Barcelona, Spain, as a Research Fellow, promoted to him in a competitive and public international call (UPFellow program), in 2014. His current research interests include the analysis of 3-D facial dysmorphology with applications into neuropsychiatric disorders from developmental origin and reconstructive surgery. He has published 16 journal papers plus a number of conference papers and has participated in several national and international Research and Development projects in facial biometrics and cardiac imaging.

Dr. Sukno was the recipient of the Marie Curie Intra-European Fellowship Award from FP7, in 2011.

John L. Waddington received the M.A. degree in natural sciences from the University of Cambridge, Cambridge, U.K., and the Ph.D. and D.Sc. degrees in neuroscience from the University of London, London, U.K.

He is currently a Professor of Neuroscience with the Royal College of Surgeons in Ireland, Dublin, Ireland. He was elected to the Royal Irish Academy, Dublin, Ireland, in 2003. His work has been supported by Science Foundation Ireland, the Health Research Board of Ireland, the Irish Research Council for Science, Engineering and Technology, Dublin, the Wellcome Trust, U.K., and the Stanley Medical Research Institute, USA. His current research interests include application of new technologies to identify indices of early developmental disturbances in brain structure and function. He has published over 350 articles (including papers in Nature, Science, and Proceedings of the National Academy of Sciences) and book chapters relating to the neuroscience of mammalian behavior in health and disease, and has edited six books. His group has pioneered the application of 3-D laser surface imaging and geometric morphometrics to craniofacial dysmorphology as an index of early brain dysmorphogenesis in neuropsychiatric disorders of developmental origin, including schizophrenia and bipolar disorder.

Paul F. Whelan (M'85–SM'01) received the B.Eng. degree in electronic engineering from Dublin City University (DCU), Dublin, Ireland (then National Institute for Higher Education Dublin), the M.Eng. degree in machine vision from the University of Limerick, Limerick, Ireland, and the Ph.D. in machine vision from Cardiff University, Cardiff, U.K.

From 1985 to 1990, he was with Industrial and Scientific Imaging Ltd., Limerick, and later, Westinghouse, Monroeville, PA, USA, where he was involved in the Research and Development of industrial vision systems. He joined the School of Electronic Engineering, DCU, in 1990, where he established the Vision Systems Group to focus on machine vision research. He is a fellow with the Institution of Engineering and Technology, a Chartered Engineer, and was an Elected Member of the DCU Governing Authority, from 2006 to 2011. He is a Principal Investigator in the RINCE Engineering Research Institute and the National Biophotonics and Imaging Platform. In 1999, he began to broaden his research into the field of biomedical image analysis and was reflected by the founding of the Centre for Image Processing and Analysis, Dublin, in 2006. He became an Associate Professor, in 2001, and a Professor in Computer Vision (Personal Chair), in 2005. His current research interests include image segmentation, and its associated quantitative analysis (specifically mathematical morphology, color-texture analysis) with applications in computer/machine vision, and medical imaging (specifically computer aided detection and diagnosis focusing on translational research).

Prof. Whelan was the recipient of the DCU Presidents Research Award, in 2011, and DCU Alumni Award, in 2014. He served as the First National Representative and a member of the governing board of the *International Association for Pattern Recognition* from 1998 to 2007, and the Inaugural President of the Irish Pattern Recognition and Classification Society, from 1998 to 2007.