

3D Facial Landmark Localization using Combinatorial Search and Shape Regression

Federico M. Sukno^{1,2}, John L. Waddington², and Paul F. Whelan¹

¹ Centre for Image Processing & Analysis, Dublin City University, Dublin 9, Ireland

² Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland

Abstract. This paper presents a method for the automatic detection of facial landmarks. The algorithm receives a set of 3D candidate points for each landmark (e.g. from a feature detector) and performs combinatorial search constrained by a deformable 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 deformable 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, substantially reducing the number of possible 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 a set of 144 facial scans acquired by means of a hand-held laser scanner in the context of clinical craniofacial dysmorphology research. Using spin images to describe the geometry and targeting 11 facial landmarks, we obtain an average error below 3 mm, which compares favorably with other state of the art approaches based on geometric descriptors.

1 Introduction

Accurate and automated detection of facial landmarks in 3D is an important problem in computer vision, with applications to biometric identification and medicine. Biometric applications [1, 2] are typically concerned with the robustness of the algorithm (e.g. to occlusions, expressions, non-collaborative subjects) to achieve systems that can be deployed in a wide variety of scenarios. In this context, state of the art algorithms target the most prominent facial landmarks on large databases with diverse acquisition artifacts (e.g. holes, spikes) that help in assessing performance in challenging scenarios.

In medical applications such as craniofacial dysmorphology [3], which is the focus of our research, there is greater interest in the highly accurate localization of landmarks, as they constitute the basis for analysis, often aimed at detecting quite small shape differences. Acquisition conditions are therefore carefully controlled to minimize 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.

The availability of high quality surfaces poses an important challenge to landmark localization algorithms, namely what accuracy can we obtain for facial landmarks in the presence of high quality data?

1.1 Related work

To take full advantage of three dimensional data, there is a particular interest in methods that localize facial landmarks based purely on geometric information (i.e. without including texture information). The most widely used feature to encode the facial geometry for landmark detection has been surface curvature [1, 4, 5]. Other geometric features include relief curves [6], the response of range data when convolved with a set of primitive filters [7] or Gabor wavelets [8].

Regardless of the features that are used, it is unlikely that unique and highly accurate detection can be achieved for a given landmark. Even the nose tip, so far the most successfully detected facial landmark, suffers from both false positives and negatives. Thus, responses from feature detectors are usually combined with prior knowledge to improve performance. Methods targeting a small subset of landmarks typically encode prior knowledge by a set of carefully designed rules about the human face, sometimes with the help of anthropometric statistics [1]. A weakness of these methods is that they usually follow a chain of rules that depend on one another. Therefore, missing or incorrectly detecting one landmark hampers the detection of all subsequent landmarks in the chain.

Statistical methods can derive prior knowledge from an annotated training set. At the expense of requiring that such a set is available, they 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. Examples of this strategy include the use of graph matching [9] and statistical shape models [5, 10, 11]. However, these methods still rely on the detection of all targeted points, which can prove difficult for most feature detectors.

To alleviate this problem, Creusot et al. [12] use partial graph matching and determine the final alignment by clustering transformations from triplets of points while Amberg & Vetter [13] use Branch and Bound to optimize the search of extended sets of landmarks (so that the missing ones are less important). However, in both cases a rigid shape is used, which is an important limitation for facial modeling. In contrast, Passalis et al. [14] present an algorithm that allows non-rigid deformations by using a deformable shape model. They exploit facial symmetry to account for possible occlusions, but still require the full visibility (and detection) of the landmarks of the left or right side. As a consequence, they need to retain a large number of candidates for each landmark and test nearly billions of combinations even though they target only 5 to 8 landmarks.

1.2 Contribution

In this paper we present a method for the detection of landmarks for craniofacial research that can handle missing points, allowing non-rigid deformations. It is assumed that for each targeted landmark there will be a set of *candidates* that

may or may not contain a suitable solution (i.e. one that is close enough to the *correct* position of the landmark). This is analogous to the point-matching problem found in algorithms that search for correspondences [15, 2]. However, the human face is a non-rigid object and these point-matching algorithms are typically restricted to rigid transformations³.

Our matching algorithm, based on RANSAC [16], consists of analyzing subsets of candidates and completing the missing information by inferring the coordinates that maximize the probability of a deformable shape model. Thus, despite the resulting subset possibly containing only part of the targeted landmarks, estimates for the remaining coordinates are inferred by regression from the priors encoded in the model. Subsets of candidates that fulfill the statistical constraints of the model are retained and additional landmarks are incorporated iteratively as long as the set remains a plausible instance of the shape model.

The ability of the model to work with incomplete information makes it possible to limit the number of candidates to be retained for each landmark, which substantively reduces the number of combinations to test with respect to the alternative of trying to always detect the complete set of landmarks. We experimentally demonstrate the accuracy of our approach, comparing it with two recent methods: one [4] based on heuristics on curvature and profile projections and another [14], closer to ours, which uses a statistical model.

2 Shape Regression with Incomplete Local Features

Our algorithm has three components: *i*) selection of candidates through local feature detection; *ii*) partial set matching to infer missing landmarks by regression; *iii*) combinatorial search, that integrates the other two components.

2.1 Local feature detection

Let \mathcal{M} represent a 3D surface, whose vertices we denote by $\mathbf{v} \in \mathcal{M}$. Also, let $d(\mathbf{v})$ be the (Euclidean) distance from \mathbf{v} to the ground truth (i.e. manual location of the considered landmark) and $s(\mathbf{v})$ the descriptor score (i.e. the value resulting from the evaluation of the descriptor *template* at vertex \mathbf{v}). For example, spin images [17] are a descriptor and the average per landmark over a training set can be used as a template.

Ideally, vertices with high $s(\mathbf{v})$ should be close to the target and have small $d(\mathbf{v})$. However, very often there are false positives with high $s(\mathbf{v})$ and $d(\mathbf{v})$ at the same time. We wish to retain enough candidates (the top- N_c) so that at least one of them is close enough to the target, i.e. there is some \mathbf{v} so that $d(\mathbf{v}) \leq r_A$, where r_A is the acceptance radius. Unfortunately, this derives a very large N_c . Given a training set $\{\mathcal{M}_i\}_{i=1}^N$, if we compute the candidates required for each mesh, $N_c^{(i)}$, we find a very skewed distribution, where the maxima are typically outliers.

³ While some robustness to deformations has been demonstrated experimentally in the literature, the formulation of these algorithms is constrained to rigid transformations.

Thus, we set N_c as an outlier threshold for the distribution of $N_c^{(i)}$; specifically, we used 1.5 times the inter-quartile distance from the upper quartile, which is accepted as a standard setting. We found that, for some landmarks, this choice can reduce the number of retained candidates by up to an order of magnitude.

Choosing N_c based on an outlier threshold for the distribution implies that, in the vast majority of cases, we will detect a candidate that is close enough to the target (e.g. within r_A) but we will miss a small proportion (the outliers).

2.2 Partial set matching with statistical shape models

Let $\mathbf{x} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_L, y_L, z_L)^T$ be a shape vector, constructed by concatenating the coordinates of L landmarks⁴. By applying Principal Component Analysis (PCA) over a representative training set [18], we get the mean shape $\bar{\mathbf{x}}$ and the eigenvector and eigenvalue matrices Φ and Λ , 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 valid object within such model) if

$$\sum_{m=1}^M \left(\frac{b_m^2}{\Lambda_{mm}} \right) < \beta_e^2 \quad (1)$$

where M is the number of retained principal components and β_e is a constant that determines the flexibility of the model.

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 $3g$:

$$\begin{aligned} \mathbf{x}^g &= (x_1, y_1, z_1, \dots, x_g, y_g, z_g)^T \\ \mathbf{x}^f &= (x_{g+1}, y_{g+1}, z_{g+1}, \dots, x_L, y_L, z_L)^T \\ \mathbf{x} &= \begin{pmatrix} \mathbf{x}^g \\ \mathbf{x}^f \end{pmatrix}, \quad \Phi = \begin{pmatrix} \Phi^g \\ \Phi^f \end{pmatrix} \end{aligned} \quad (2)$$

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 . Assuming that the model follows a multi-variate Gaussian distribution $\mathcal{N}(\mathbf{0}, \Lambda)$ in PCA-space, it can be shown that (see Appendix A in supplementary material):

$$\mathbf{x}^g = \bar{\mathbf{x}}^g - (\Phi^g \Lambda^{-1} (\Phi^g)^T)^{-1} (\Phi^g \Lambda^{-1} (\Phi^f)^T) (\mathbf{x}^f - \bar{\mathbf{x}}^f) \quad (3)$$

The idea of using statistical constraints to complete missing landmarks has been explored previously by other authors [19, 20]. While Blanc et al. [19] use

⁴ We assume that the shape has been aligned (e.g. by Procrustes analysis) so that Similarity is removed.

an iterative approach to solve the resulting system of equations, de Bruijne et al. [20] obtain a closed form solution by applying linear regression. They use the maximum likelihood estimate of $\mathbf{x}^g | \mathbf{x}^f$ from the covariance matrix of the training set, which produces results very similar to ours. The 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.

2.3 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. All combinations of 4 landmarks are then evaluated⁵. 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.

We use equation (3) to complete the shape by inferring the 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 [21]. 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}_{test}^f , taking the closest candidates to the current estimate for the missing landmarks:

$$\begin{aligned} \gamma(\mathbf{c}_k) &= \text{median}(\Delta \hat{\mathbf{x}}_{test}) & (4) \\ \Delta \hat{\mathbf{x}}_{test} &= \left\{ \begin{array}{ll} \|\hat{\mathbf{x}}_{test}(\ell_j) - \mathbf{x}_{test}^f(\ell_j)\|^2, & \forall \ell_j \in \mathbf{x}_{test}^f \\ \min_{\mathbf{c}_j} \|\hat{\mathbf{x}}_{test}(\ell_j) - \mathbf{c}_j\|^2, & \forall \ell_j \notin \mathbf{x}_{test}^f \end{array} \right\} \end{aligned}$$

where ℓ_j is the j -th landmark, and \mathbf{c}_j are each of the candidates for landmark j . We use $\mathbf{x}(\ell_j)$ to indicate the j -th landmark of the shape, and $\hat{\mathbf{x}}$ is the best PCA reconstruction of shape \mathbf{x} in a least squares sense.

The inclusion cost in (4) is a key aspect of the algorithm and is divided in two parts, from the definition of $\Delta \hat{\mathbf{x}}_{test}$. 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 $\|\Phi^T(\hat{\mathbf{x}} - \bar{\mathbf{x}})\|$ as the inclusion cost. However, this option neglects the effect of the coordinates inferred from $\hat{\mathbf{x}}^g$. We have found that it is important to constrain the solution to be within the mesh surface and using the landmark candidates is a convenient alternative to achieve this. The definition of $\gamma(\mathbf{c}_k)$ based on the median implies that the landmark cost $\gamma(k)$, in line 12, is the least median of squares [22], which provides robustness to potential outliers (e.g. landmarks for which no nearby candidates have been found).

Finally, 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

⁵ Note that, as we are interested in accuracy, we do an exhaustive search instead of random sampling but we do retain the idea of consensus as the figure of merit.

Algorithm 1 SRILF: Shape Regression with Incomplete Local Features

```

1: Start from a set of candidates for each landmark
2: for (all 4-tuple combinations of landmarks and candidates  $\mathbf{x}_4$ ) do
3:   Initialize  $\mathbf{x}^f = \mathbf{x}_4$ 
4:   Infer  $\hat{\mathbf{x}}^g$  using (3), obtaining  $\hat{\mathbf{x}}$ 
5:   while ( $\hat{\mathbf{x}}$  fulfills the constraints in (1)) do
6:     for (all other landmarks,  $\ell_k \notin \mathbf{x}^f$ ) do
7:       for (all candidates  $\mathbf{c}_k$  for landmark  $\ell_k$ ) do
8:         Add the candidate  $\mathbf{c}_k$  to  $\mathbf{x}^f$  to obtain  $\mathbf{x}_{test}^f$ 
9:         Infer  $\hat{\mathbf{x}}_{test}^g$  from  $\mathbf{x}_{test}^f$  to obtain  $\hat{\mathbf{x}}_{test}$ 
10:        Compute the resulting cost  $\gamma(\mathbf{c}_k)$  as in (4)
11:      end for
12:      Compute the landmark cost  $\gamma(k) = \min_k \gamma(\mathbf{c}_k)$ 
13:    end for
14:    Update  $\mathbf{x}^f$  adding the landmark with minimum  $\gamma(k)$ 
15:    Infer  $\hat{\mathbf{x}}^g$  from the updated  $\mathbf{x}^f$  to obtain  $\hat{\mathbf{x}}$ 
16:  end while
17:  Compute the score for  $\mathbf{x}_4$  as  $\#(\mathbf{x}^f) + e^{-\gamma(k)}$ 
18: end for
19: Keep the subset that achieved the highest score

```

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.

3 Experimental evaluation

Our test dataset consisted of 144 facial scans acquired by means of a hand-held laser scanner (FastSCANTM, Colchester, VT, USA). Special care was taken to avoid occlusions due to facial hair. There is some heterogeneity regarding the extent to which neck and shoulders were included. The extracted surfaces were subsampled by a factor of 4 : 1, to facilitate comparison to [14]. As a result, there were, on average, approximately 21.3 thousand vertices per mesh.

The dataset contains exclusively healthy volunteers who acted as controls in the context of craniofacial dysmorphology research. Each scan was annotated with a number of anatomical landmarks [3], among which we target the following 11: the nose root or *nasion* (n); the nose tip or *pronasale* (prn); the chin tip or *pogonion* (pg); the inner-eye corners or *endocanthion* (en, left & right); the outer-eye corners or *exocanthion* (ex, left & right) the nose corners or *alare crest* (ac, left & right) and the mouth corners or *cheilion* (ch, left & right).

All experiments were performed using 6-fold cross-validation. For each fold, the 120 surfaces composing the training set were used to build the PCA model and to determine the feature candidates to be retained. For the PCA model, we kept 99% of the total variance, with the flexibility parameter set to $\beta_e = 4$. We chose spin images [17] as our local geometry descriptor, adopting the parameters

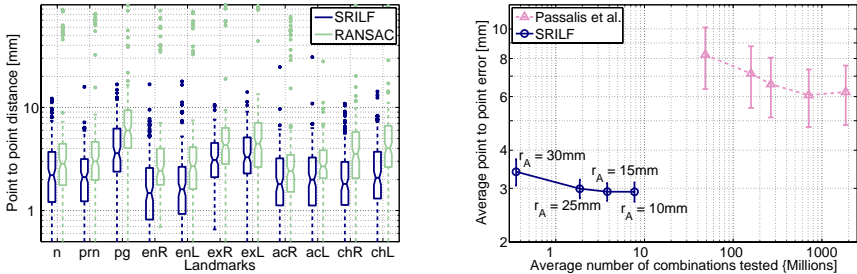


Fig. 1. Left: accuracy of SRILF and RANSAC with respect to ground truth annotations. For symmetric landmarks, Left or Right are additionally indicated. Right: Accuracy and number of combinations tested for SRILF and the method by Passalis et al. [14] when varying the parameters that control the number of candidates retained. Error bars indicate a 95% confidence interval of the mean.

indicated in [14] (16 bins of 2 mm each). For each targeted landmark we compute a descriptor template as the median over the training set and determine the number of candidates N_c as described in Section 2.1. Unless otherwise stated, reported results correspond to $r_A = 20$ mm.

3.1 Localization accuracy

We measured the accuracy of our method by comparing the automatic results with manual annotations from [3] that were used as independent ground truth.

To assess the necessity of a deformable model, the same experiment was repeated with a rigid template (the mean-shape of the PCA model). The resulting algorithm is much like RANSAC. The *fixed* landmarks are now simply those in the 4-tuple being tested, which are used to align the mean-shape to the test shape. Since deformation is not allowed, the shape is automatically completed and we have a rigid guess⁶ for the remaining landmarks. Then we can use the same cost as in SRILF, defined in (4). Such a choice is equivalent to defining a minimum support of 50% in the RANSAC algorithm.

The localization results are shown in Fig. 1 (left). As expected, the deformable model is superior to the rigid one and this is verified for all tested landmarks. Since variations in our dataset are mainly due to identity with only residual expression changes, this is a remarkable result in favor of non-rigid modeling *vis-à-vis* the rigid registration used, for example, by Creusot et al. [12] and Ambert & Vetter [13]. The median errors of SRILF were below 3 mm for all landmarks except the chin tip and outer eye corners. Less than 2% of landmarks had errors above 10 mm and only in two cases did errors exceed 20 mm.

As suggested by the example in Fig. 2, our method does not actually locate a suitable candidate for every landmark. Therefore, only some of the targeted

⁶ The alignment is performed using a Similarity transformation and hence uniform re-scaling is possible, but the model shape is not allowed to deform.

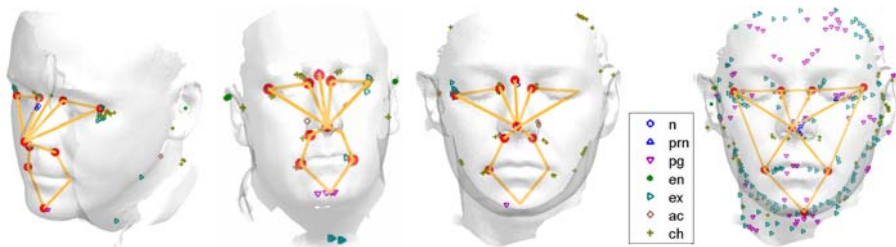


Fig. 2. Three examples of SRILF results (left) and one of the method by Passalis et al. (right). Solid lines show the obtained shape and red circles indicate successfully detected candidates. In SRILF the shape is completed by inference. All retained candidates for each landmark are also indicated according to the markers in the key.

landmarks are detected with the remainder inferred using statistics from the shape model. In our experiments, an average of 8.67 landmarks were found for each shape ($\sim 78.8\%$). The most frequently detected point was the nose tip (found in more than 93% of faces) and the least frequent was the chin tip (only found in 48% of faces). Note that all errors reported for our method are averages over all 144 shapes including both detected and inferred landmarks.

3.2 Comparison to other methods

In this section we provide a quantitative comparison with two recently published methods. The first method, from Segundo et al. [4], has reported state of the art accuracy (among methods purely based on geometry) using a training-free approach, completely different from ours. In contrast, the second method, from Passalis et al. [14], is based on statistical shape models and is closer to the approach proposed here. Indeed, some of the settings used in our experiments were chosen following [14] to facilitate comparison. The key difference is that their approach is based on *fixed* models of 5 landmarks that need to be detected in all cases (i.e. a *suitable* candidate for every landmark is always required). This leads them to define heuristics to keep large sets of candidates which, in addition to a potential increase of mismatches, considerably raises the computational load, as illustrated by the examples in Figure 2.

Table 1 summarizes the comparison results. Statistically significant differences with respect to SRILF are indicated by an asterisk⁷. It can be observed that our method performs consistently best for all compared landmarks except the nose tip, where [4] obtained similar accuracy. The average run time for our algorithm was 4.1 seconds⁸.

Fig. 1 (right) provides further comparison to [14]: we found its performance very dependent on the number of candidates retained. While Passalis et al. specify an upper limit, the actual number kept depends on a threshold on the

⁷ $p < 0.05$ both in a paired t-test and a paired Wilcoxon signed rank test.

⁸ Non-optimized C++ implementation on an Intel Xeon E5320 @1.86 GHz.

Table 1. Localization accuracy [mm] per landmark (mean \pm standard error) for the different methods.

Landmark	n	prn	pg	en	ex	ac	ch
Passalis et al. [14]	n/a	2.89(*) ± 0.15	9.19(*) ± 0.97	3.42 ± 0.66	6.98(*) ± 1.35	n/a	5.88(*) ± 0.96
Segundo et al. [4]	n/a	2.63 ± 0.13	n/a	5.64(*) ± 0.61	n/a	4.93(*) ± 0.21	n/a
SRILF	3.08 ± 0.22	2.43 ± 0.15	4.52 ± 0.25	2.26 ± 0.20	3.67 ± 0.18	2.45 ± 0.22	2.69 ± 0.19

spin-image scores below which candidates are discarded. The much larger number of candidates retained in [14] not only increases computational load but also impairs accuracy, as it is more likely to find combinations of points that, while being off-target, still comply with the statistical constraints. The need of a suitable candidate for every targeted landmark makes it necessary to retain those large sets, which might even be insufficient in some cases. For example, in order to always detect both mouth corners in our dataset, we would need to retain more than 1000 candidates, while the heuristics defined in [14] limit this value to 256. A similar situation was observed for the eye corners and the chin tip.

4 Conclusions

We present a method for the localization of facial landmarks that uses regression from a deformable shape model to tackle the potential false negatives in the detection of some landmarks. This allows an important reduction in the number of candidates to test, hence reducing the space of possible solutions.

We compared our method with two state of the art approaches targeting 11 landmarks on a dataset of 144 facial surfaces acquired with a hand-held laser scanner, in the context of craniofacial dysmorphology research. Despite the moderate size of the database, there were statistically significant differences in favor of the proposed approach for the majority of targeted points. We also showed that the capability to tackle non-rigid deformations through the deformable shape model clearly outperforms the alternative of using a rigid template.

The method is general and not constrained in any way to the landmarks selected for our experiments. The only requirement, as a learning-based method, is the availability of an annotated training set. Furthermore, the reduction in computational complexity with respect to similar approaches suggests that a larger number of landmarks might be targeted. Nonetheless, this would be linked to the accuracy of the utilized descriptor, so that the number of candidates that must be retained is kept within acceptable bounds.

Acknowledgments. The authors would like to thank their colleagues in the Face3D Consortium (www.face3d.ac.uk), and the financial support provided for it from the Wellcome Trust (grant 086901/Z/08/Z).

References

1. Gupta, S., Markey, M., Bovik, A.: Antopometric 3D face recognition. *Int. J. Comput. Vis.* **90** (2010) 331–349
2. Zhang, G., Wang, Y.: Robust 3D face recognition based on resolution invariant features. *Pattern Recogn. Lett.* **32** (2011) 1009–1019
3. Hennessy, R., Kinsella, A., Waddington, J.: 3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis: initial application to sexual dimorphism. *Biol. Psychiatry.* **51** (2002) 507–514.
4. Segundo, M., Silva, L., Bellon, O.P., et al.: Automatic face segmentation and facial landmark detection in range images. *IEEE Trans. Syst., Man, Cybern. B* **40** (2010) 1319–1330
5. Szeptycki, P., Ardabilian, M., Chen, L.: A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking. In: *BTAS (2009)* 1–6
6. Faltemier, T., Bowyer, K., Flynn, P.: Rotated profile signatures for robust 3D feature detection. In: *FG (2008)* 1–7
7. Yu, T., Moon, Y.: A novel genetic algorithm for 3D facial landmark localization. In: *BTAS (2008)* 1–6
8. D’Hose, J., Colineau, J., Bichon, et al.: Precise localization of landmarks on 3D faces using Gabor wavelets. In: *BTAS (2007)* 1–6
9. Romero-Huertas, M., Pears, N.: Landmark localisation in 3D face data. In: *Int. Conf. on Adv. Video and Signal Based Surveillance.* (2009) 73–78
10. Zhao, X., Szeptycki, P., Dellandrea, et al.: Precise 2.5D facial landmarking via an analysis by synthesis approach. In: *Workshop on Appl. of Comput. Vis.* (2009) 1–7
11. Nair, P., Cavallaro, A.: 3-D face detection, landmark localization and registration using a point distribution model. *IEEE Trans. Multimedia* **11** (2009) 611–623
12. Creusot, C., Pears, N., Austin, J.: 3D face landmark labelling. In: *ACM Workshop on 3D object retrieval.* (2010) 23–32
13. Amberg, B., Vetter, T: Optimal landmark detection using shape models and branch and bound . In: *ICCV (2011)* 455–462
14. Passalis, G., Perakis, N., Theoharis, T., et al.: Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **33** (2011) 1938–1951
15. Jiang, H., Yu, S., Martin, D.: Linear scale and rotation invariant matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **33** (2011) 1339–1355
16. 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
17. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **21** (1999) 433–449
18. Cootes, T., Taylor, C.: Statistical models of appearance for computer vision. Technical report, Wolfson Image Analysis Unit, University of Manchester, UK (2001)
19. Blanc, R., Syrkinina, E., Székely, G.: Estimating the confidence of statistical model based shape prediction. In: *IPMI (2009)* 602–613
20. de Bruijne, M., Lund, M., Tanko, et al.: Quantitative vertebral morphometry using neighbor-conditional shape models. *Med. Image Anal.* **11** (2007) 503–512
21. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. *Pattern Recogn.* **33** (2000) 25–41
22. Rousseeuw, P.: Least median of squares regression. *J. Am. Stat. Assoc.* **79** (1984) 871–880