Creating Ultra Dense Point Correspondence Over the Entire Human Head

Rasmus R. Paulsen[1]([✉]), Kasper Korsholm Marstal[1,3], Søren Laugesen[2], and Stine Harder[1]

[1] DTU Compute, Technical University of Denmark, Richard Petersens Plads, Building 324, 2800 Kongens Lyngby, Denmark
{rapa,kkmars,sthar}@dtu.dk
[2] Interacoustics Research Unit, DGS Diagnostics A/S, Ørsteds Plads, Building 352, 2800 Kongens Lyngby, Denmark
slau@iru.interacoustics.com
[3] Departments of Radiology and Medical Informatics Erasmus Medical Center, Biomedical Imaging Group Rotterdam (BIGR), Rotterdam, The Netherlands
http://www.compute.dtu.dk

Abstract. While the acquisition and analysis of 3D faces has been an active area of research for decades, it is still a complex and demanding task to accurately model the entire head and ears. Having accurate models would for example enable virtual design of hearing devices. In this paper, we describe a complete framework for surface registration of complete human heads where the result is point correspondence with a very high number of points. The method is based on a volumetric and multi-scale non-rigid registration of signed distance fields. The method is evaluated on a set of 30 human heads and the results are convincing. The output can for example be used to compute statistical shape models. The accuracy of predicted anatomical landmarks is on the level of experienced human operators.

Keywords: Surface registration · Signed distance field · Human head modelling

1 Introduction

Automated analysis and identification from 2D photos is a very well-established area and for the last decade 3D face modelling, analysis, and recognition has become an established technology [6]. Acquisition has typically been done using laser scanning devices that require the subject to be in a fixed position for some time and be careful of eye exposure. Recently, a wide range of acquisition devices has become available. From multi-camera setups using multi-view stereo reconstruction algorithms [2] that can acquire a full face in a single capture to low-cost and accessible systems using the Microsoft Kinect [12]. Most of these systems provide a triangulated surface as output with either a texture map or colour values on each vertex.

Due to the complexity and optical properties of human hair, most systems can not do a realistic acquisition of hair and it is therefore often covered by a cap. 3D face acquisition is for example being used for entertainment where photorealistic faces are captured and used for computer games, in facial surgery for surgery planning [4], and for facial recognition [1]. In most applications, the focus is on the acquisition and analysis of the face.

While the acquisition of the human face is an established method, it is much more complicated to get an accurate surface scan of the entire head including the complex anatomy of the outer ear. 3D scannings of the outer ear have been used in biometrical applications [20]. Recently, accurate 3D scannings of the ear and head have been used for product design. In particular, the acoustical optimisation of hearing devices is an attractive application of head scans. The combination of 3D surface scans and advanced finite element simulations enables the computation of the so-called head related transfer function (HRTF) that enables user-specific hearing device optimisation [8]. We have previously demonstrated a method to acquire accurate full head scans that are well suited for acoustical applications [9]. The data used in this paper is similar to this type of data.

Many approaches to the analysis of human heads are based on a surface registration step where, for example, a template mesh is warped to fit a new unseen face. Surface registration has been a major research area for years [18, 19] and a variety of approaches exists. In the seminal paper on 3D morphable models [3] a modified optical flow method is used to register 3D face scans. In [13] partial scans acquired using the Microsoft Kinect are registered using a novel deformation model that potentially enables multi-level approaches [17], where a sub-sampled coarse model is initially aligned and gradually refined in further steps. However, the difference in complexity between the human face and the entire human head including ears is quite large. In our experiments we did not find an existing framework that could successfully register the entire head. To successfully do the registration, the methods need to be multi-level, so coarse features like the overall head is registered first. While the fine details of the outer ear in are registered at the final fine resolutions. In this paper, we present a method based on non-rigid volumetric registration of signed distance fields to solve the task, where the multi-level properties is given by the volumetric registration method.

2   Data and Preprocessing

The data used in this work consists of 30 3D surface scans of the entire heads including the outer ear. The outer ear is here defined as the concha, pinna, and the entrance to the ear canal. The surface scans are acquired using a Canfield Vectra M3 scanner, which is a dedicated human head scanner typically used for facial restorative surgery. Due to the very complex anatomy of the human ear it is not possible to acquire a full surface scan of the head and ear in a single acquisition. Therefore, each head was scanned from up to ten different angles by

placing the person in a rotating chair. For each scan, relevant areas are manually marked to avoid using areas influenced by motion or facial expressions. A set of sparse landmarks are also placed on each sub-scan using the template described in [9]. Using the landmarks, the marked areas are brought into rough alignment. Following the approach described in [15], a combined Markov Random Field surface reconstruction and implicit iterative closest point algorithm is used to create a triangulated surface of the entire head and ear. Finally, the colour values sampled from the scannings are transferred to the vertices of the reconstructed surface. The resulting surfaces consists of approximately 450.000 vertices and 950.000 triangles. An example of one of the 30 entire heads can be seen in Fig. 1. The green areas of the scans indicate where raw colours of the original scans have not been available. Due to the optical properties of the eye it is not possible to acquire the true outer surface of the cornea with an optical scanner. Therefore the eye region will typically be either flat or curve inwards in the used data set.
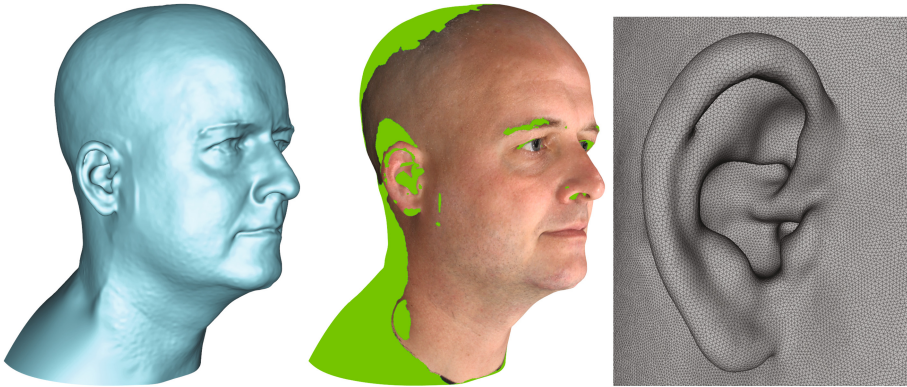

Fig. 1. A reconstructed full head and ear scan. To the left the raw mesh, in the middle with vertex colouring and to the right a close up of the right ear. (Color figure online)

3   Methods

The goal is to register two surfaces. This means creating a dense point correspondence between a source surface ($\mathcal{S}$) and a target surface ($\mathcal{T}$) so each vertex is placed on the exact same anatomical spot on both $\mathcal{S}$ and $\mathcal{T}$. In this work, $\mathcal{S}$ is deformed to fit $\mathcal{T}$. Initially, $\mathcal{S}$ is rigidly aligned to $\mathcal{T}$ using a sparse set of anatomical landmarks manually placed on both surfaces as seen in Fig. 2. The aligned source is $\mathcal{S}_a$. In this work, we use the implicit shape description embedded in signed distance fields to drive the registration. A signed distance field is computed for both $\mathcal{T}$ and $\mathcal{S}_a$ using the method described in [16]. Here the distance field is represented as a voxel volume covering the entire surface, where the value in a voxel is the Euclidean distance to the surface. The surface is implicitly defined as the zero-level iso-surface of the distance field. In order to close holes in the surface and to accommodate for missing data, a weighted

Laplacian regularisation is performed on the distance field. An example of a regularised signed distance field can be seen in Fig. 3, which shows that the shape information is well represented by the iso-curves and that the overall shape is preserved but in a smoother form in the distance field furthest from the surface.
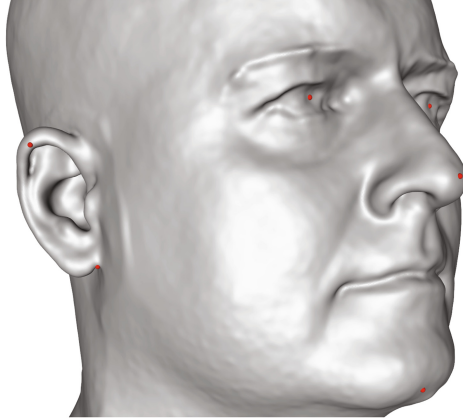

Fig. 2. The set of eight manually placed landmarks. The two landmarks placed the left ear are not shown.

The actual surface is described as the zero-level iso-surface in the distance field, but other iso-surfaces also contain implicit information about the shape. By sub-sampling the field, a coarser description of the surface shape is obtained. Finally, the gradient field of a signed distance field is also well described since the gradient will mostly point towards the zero-level iso-surface. These properties make it attractive to use a well-established and state-of-the-art volumetric registration algorithm to do a non-rigid registration of the signed distance fields. The volumetric image registration is formulated as an optimisation problem,

$$\hat{\mathbf{T}}_\mu = \arg\min_{\mathbf{T}_\mu} = \mathcal{C}(\mathbf{T}_\mu; I_F, I_M), \tag{1}$$

where I_F is the fixed volume and I_M is the moving volume. Here I_F is the signed distance field created from $\mathcal{T}$ and I_M is the signed distance field created from $\mathcal{S}_a$. $\mathbf{T}_\mu$ is a non-rigid volumetric transformation that transforms I_M and it is parameterised by the parameter-vector μ. The goal is to find the values of μ that minimise the cost function $\mathcal{C}$. The `elastix` library [11] is used to perform the volumetric registration. The transformation used is a multi-level cubic B-spline using four resolution levels. The multi-level approach ensures that coarse anatomical structures are aligned first and then finer structures are gradually being registered. In our case, it means that the overall shape of the head is aligned first and the finer details of the ears are registered in the final resolution. Since the two volumes is of the same nature and the scale of the voxel values are very similar, the mean squared voxel value difference (MSD) is chosen as the
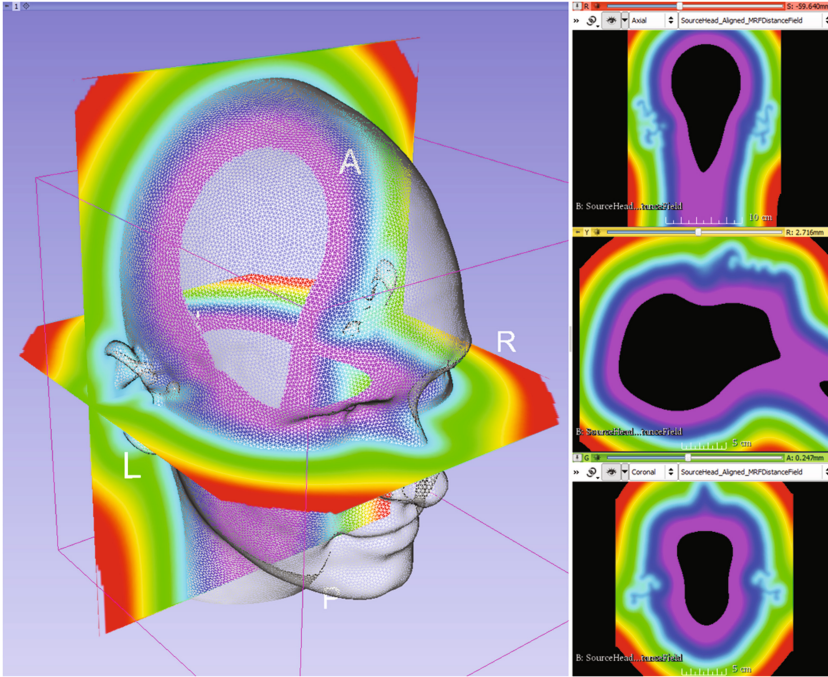
Fig. 3. A signed distance field computed for a full head. To the right the field is projected on three cuts trough the field. The distance field is thresholded to only show distances close to the surface of the head.

similarity metric. The surface of interest is by definition close to the zero-level of the distance field and therefore a binary sampling mask is applied to the moving volume. The mask is generated by only including the voxels that have a value in the range $[-20, 20]$ (measured in mm) in the distance field. Due to the large shape variation around the ears we found it necessary to aid the registration in a few case by adding a set of eight manually placed landmarks as seen in Fig. 2. This is included in the registration by adding

$$\mathcal{S}_{\mathrm{CP}} = \frac{1}{P} \sum_i \|\mathbf{y}_i - \mathbf{T}_\mu(\mathbf{x}_i)\| \tag{2}$$

as a metric that penalises distances between corresponding landmarks. Here $\mathbf{y}_i$ and $\mathbf{x}_i$ are P corresponding points on $\mathcal{S}$ and $\mathcal{T}$. The final similarity metric then becomes

$$\mathcal{S} = \omega_1 \mathrm{MSD}(\mu; I_F, I_M) + \omega_2 \mathcal{S}_{\mathrm{CP}}, \tag{3}$$

where $\omega_1 = 1.0$ and $\omega_2 = 0.001$ are weights that are experimentally set. The degree of smoothness is implicitly regularised by the knot spacing in the B-spline. This means that the final cost function used in the registration is equal

to the similarity metric: $\mathcal{C} = \mathcal{S}$ [11]. The optimisation was done using adaptive stochastic gradient descent [10] with 2048 random samples per iteration for a maximum of 500 iterations.

The result of using the non-linear registration on the signed distance fields is that a transformation, $\hat{\mathbf{T}}_\mu$, that brings the distance field representing $\mathcal{S}_a$ in alignment with the distance field representing $\mathcal{T}$. By applying $\hat{\mathbf{T}}_\mu$ to the vertices of $\mathcal{S}_a$, the vertices are propagated to $\mathcal{T}$ and thus creates point correspondence. The transformed mesh is $\mathcal{S}_{\mathrm{NR}}$.

Since the registration is based on a large part of the distance field it is not guaranteed that the zero-level iso-surfaces match exactly. Therefore some propagated vertices do not fall exactly on $\mathcal{T}$. We apply the point propagation method originally described in [14] to fix the vertices to $\mathcal{T}$. Here the vertices are first projected to the closest position on $\mathcal{T}$. There is now a correspondence vector for each vertex in $\mathcal{S}_{\mathrm{NR}}$. The correspondence vector goes from the position computed in the registration step, stored in $\mathcal{S}_{\mathrm{NR}}$, to the projected position on $\mathcal{T}$. Together these correspondence vectors represents a correspondence vector field (CVF).

The CVF is now cast into a Markov Random Field regularisation (MRF) framework, where each vector is penalised by deviations from its neighbours. In each iteration the CVF is MRF regularised and each vector is reprojected onto $\mathcal{T}$. The final result of applying this method is that all vertices from $\mathcal{S}$ have been propagated directly onto $\mathcal{T}$. In this work, we use a mesh with 600.000 vertices and 1.200.000 triangles as $\mathcal{S}$ and thereby creating an ultra dense point correspondence over the entire head and outer ear. This mesh has been selected from the set of full head scannings and remeshed into this mesh resolution.

In order to validate the registration, the template mesh was registered to all the other full heads in the data set, thereby creating a full correspondence over the data set. Following the steps of building a point distribution model [5], a Procrustes alignment of the registered meshes is performed. To avoid the bias introduced by using a specific shape as template, the Procrustes average shape was used as the template in a second registration step.

4   Results

The registration framework is applied to 30 entire head scans. We use a template mesh as the source mesh and apply it to all the other heads (targets). An example can be seen in Fig. 4. The template mesh seen in Fig. 4 is the average shape from the Procrustes alignment [5] of the initial registration of the data set, where an arbitrary head was used as template. As can be seen, the average shape is smooth but contains all important facial features including detailed outer ears. The result of the registration is that the template mesh is placed exactly on top of the target meshes and it is not possible to visualise the potential discrepancies by overlaying the registration results on the target. Instead, the template mesh was manually annotated with 93 landmarks as defined in [9] (green points in Fig. 4). The same set of landmarks was also manually annotated on all the heads
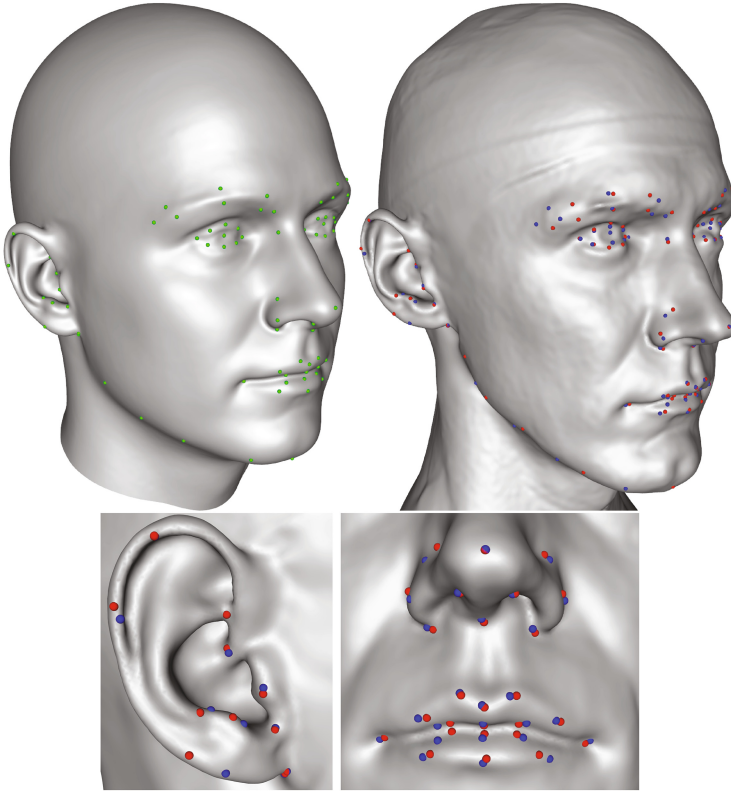
Fig. 4. Upper left:.eps The green landmarks are manually placed on the template. Blue landmarks are the results of the registration and the red landmarks are manually placed on the target. (Color figure online)

in the data set (red points in Fig. 4). In order to validate the accuracy of the registration, the landmarks from the template mesh are propagated to the target mesh (blue points in Fig. 4). An estimate of the accuracy can be computed as the distance between the annotated (red) and predicted (blue) landmarks. However, it is well known that manual annotation is error-prone and that each annotated landmark has a spatial uncertainty [7]. We have therefore chosen to only validate the accuracy with landmarks that can be accurately placed manually - well knowing that this is not a truly neutral estimate of the registration accuracy. The results are shown in Table 1.

It can be seen that the average error is in the range of 1.2–2.6 mm which is comparable to the errors from manual annotators [7]. The landmarks that are difficult to place by a human operator had errors in the range of 4 to 8 mm.

After the template mesh has been applied to the target it is possible to sample the colours from the original scans. In Fig. 5, a registered mesh with a colour assigned to each vertex can be seen. The average vertex colour over the entire set

Table 1. Error (in mm) between annotated and predicted landmarks (LM).

LM	Place	Mean	Median	Standard deviation
17	Right lateral canthus	1.9	1.4	1.4
21	Right medial canthus	1.9	1.8	1.0
25	Left lateral canthus	2.6	2.4	1.4
29	Right medial canthus	2.1	2.0	1.1
40	Columellar connection to upper lip	1.2	1.0	0.8
47	Right oral commissure	1.7	1.8	0.9
53	Left oral commissure	2.0	1.7	1.1
75	Left ear bottom of concha	1.6	1.3	0.9
76	Left ear tragus	1.6	1.6	0.8
77	Left ear crus of helix	2.0	2.0	0.9
85	Right ear bottom of concha	1.3	1.2	0.9
86	Right ear tragus	1.6	1.6	0.9
87	Right ear crus of helix	1.8	1.5	1.0

can then be computed due to the vertex correspondence. Applying the average vertex colours to the Procrustes average can be seen in Fig. 5. It can clearly be seen that the eye colours have been smoothed out and that the outline of the

Fig. 5. Left: A registered mesh with vertex colours sampled from the original scans. **Right:** The Procrustes average with average vertex colours.

eye is blurry. This can be caused by the problems in acquiring the shape and texture of the eye correctly and the fact that the eye also changes shape and texture when exposed to light.

The running time of one full head registration is in the order of 20 minutes on a standard Windows-7 laptop computer with 8 GB RAM. Some parts of the algorithm are implemented as parallel processes using the 8 processing cores of the laptop otherwise the algorithm was not optimised for speed.

5   Conclusion

We have presented a method for registration of full heads including ears that successfully computes an ultra dense correspondence from a template mesh to an arbitrary mesh. It is also demonstrated that the method accurately maps anatomical meaningful landmarks from a template mesh to an arbitrary target mesh. The results in terms of accuracy of single landmark placement is comparable to what trained human operators can do. The method is based on matching the pure shape information implicitly described by a signed distance field. It is possible that including the surface colours in the registration could further increase the accuracy, in particular in regions with little shape information. The method can for example be used to build statistical shape models that can further be used in modelling, analysis, and design-related applications.

References

1. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D face recognition: a survey. Pattern Recogn. Lett. **28**(14), 1885–1906 (2007)
2. Beeler, T., Bickel, B., Beardsley, P., Sumner, B., Gross, M.: High-quality single-shot capture of facial geometry. ACM Trans. Graph. (Proc. SIGGRAPH) **29**(3), 40:1–40:9 (2010)
3. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of 26th Annual Conference on Computer Graphics and Interactive Techniques, pp. 187–194 (1999)
4. Chang, J.B., Small, K.H., Choi, M., Karp, N.S.: Three-dimensional surface imaging in plastic surgery: foundation, practical applications, and beyond. Plast. Reconstr. Surg. **135**(5), 1295–1304 (2015)
5. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
6. Daoudi, M., Srivastava, A., Veltkamp, R.: 3D Face Modeling, Analysis and Recognition. Wiley, UK (2013)
7. Fagertun, J., Harder, S., Rosengren, A., Moeller, C., Werge, T., Paulsen, R.R., Hansen, T.F.: 3D facial landmarks: inter-operator variability of manual annotation. BMC Med. Imaging **14**(1), 1 (2014)

8. Harder, S., Paulsen, R.R., Larsen, M., Laugesen, S., Mihocic, M., Majdak, P.: A framework for geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions with a focus on hearing-assistive devices. Comput. Aided Des. **75**, 39–46 (2016)
9. Harder, S., Paulsen, R.R., Larsen, M., et al.: A three dimensional children head database for acoustical research and development. In: Proceedings of Meetings on Acoustics, vol. 19, p. 050013. Acoustical Society of America (2013)
10. Klein, S., Pluim, J.P., Staring, M., Viergever, M.A.: Adaptive stochastic gradient descent optimisation for image registration. Int. J. Comput. Vis. **81**(3), 227–239 (2009)
11. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2010)
12. Li, B.Y., Mian, A.S., Liu, W., Krishna, A.: Using kinect for face recognition under varying poses, expressions, illumination and disguise. In: 2013 IEEE Workshop on Applications of Computer Vision (WACV), pp. 186–192. IEEE (2013)
13. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. Eurographics Symp. Geom. Process. **27**(5), 1421–1430 (2008)
14. Paulsen, R.R., Hilger, K.: Shape modelling using markov random field restoration of point correspondences. In: Proceedings of Information Processing in Medical Imaging, pp. 1–12 (2003)
15. Paulsen, R.R., Larsen, R.: Anatomically plausible surface alignment and reconstruction. In: Theory and Practice of Computer Graphics, pp. 249–254 (2010)
16. Paulsen, R., Bærentzen, J., Larsen, R.: Markov random field surface reconstruction. IEEE Trans. Vis. Comput. Graph. **16**(4), 636–646 (2010)
17. Sumner, R.W., Schmid, J., Pauly, M.: Embedded deformation for shape manipulation. In: ACM Transactions on Graphics (TOG), vol. 26, p. 80. ACM (2007)
18. Tam, G.K., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X.F., Rosin, P.L.: Registration of 3D point clouds and meshes: a survey from rigid to nonrigid. IEEE Trans. Vis. Comput. Graph. **19**(7), 1199–1217 (2013)
19. Van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D.: A survey on shape correspondence. In: Computer Graphics Forum, vol. 30, pp. 1681–1707. Wiley Online Library (2011)
20. Yan, P., Bowyer, K.W.: Biometric recognition using 3D ear shape. IEEE Trans. Pattern Anal. Mach. Intell. **29**(8), 1297–1308 (2007)