

Reconstructing High-resolution Face Models from Kinect Depth Sequences Acquired in Uncooperative Contexts

Enrico Bondi, Pietro Pala, Stefano Berretti and Alberto Del Bimbo
Department of Information Engineering, University of Florence, Florence, Italy

Abstract—Performing face recognition across 3D scans of different resolution is now attracting an increasing interest thanks to the introduction of a new generation of depth cameras, capable of acquiring color/depth images over time. However, these devices have still a much lower resolution than the 3D high-resolution scanners typically used for face recognition applications. If data are acquired without user cooperation, the problem is even more challenging and the gap of resolution between probe and gallery scans can yield to a severe loss in terms of recognition accuracy. Based on these premises, we propose a method to build a higher-resolution 3D face model from 3D data acquired by a low-resolution scanner. This face model is built using data acquired when a person passes in front of the scanner, following an uncooperative protocol. To perform non-rigid registration of point sets and account for deformation of the face during the acquisition process, the Coherent Point Drift (CPD) method is used. Registered 3D data are filtered through a variant of the *lowess* method to remove outliers and build the final face model. The proposed approach is evaluated in terms of accuracy of face reconstruction and face recognition.

I. INTRODUCTION

Person identity recognition by the analysis of 3D face scans is attracting an increasing interest, with several challenging issues successfully investigated, such as 3D face recognition in the presence of non-neutral facial expressions, occlusions, and missing data [6]. Existing solutions have been evaluated following well defined protocols on consolidated benchmark datasets, which provide a reasonable coverage of the many different traits of the human face, including variations in terms of gender, age, ethnicity, occlusions due to hair or external accessories. The resolution at which 3D face scans are acquired changes across different datasets, but given a dataset it is typically the same for all the scans. Due to this, the difficulties posed by considering 3D face scans with different resolutions and their impact on the recognition accuracy have not been explicitly addressed in the past. Nevertheless, there is an increasing interest for methods capable of performing recognition across scans acquired with different resolutions. This is mainly motivated by the availability of a new generation of low-cost, low-resolution 3D dynamic scanning devices (i.e., 3D plus time, also called 4D), such as Microsoft Kinect or Asus Xtion PRO LIVE. In fact, these devices are capable of a combined color-depth (RGB-D) acquisition at about 30fps, with an optimal working distance from the sensor ranging from 40cm up to 1.5m. The spatial resolution of such devices is lower than that of

high-resolution 3D scanners, but these latter are also costly, bulky and highly demanding for computational resources. Despite the lower resolution, the advantages in terms of cost and applicability of consumer cameras motivated some preliminary works performing face detection [20], continuous authentication [19] and recognition [9], [13], [15] directly from the depth frames of the Kinect camera. However, based on the opposite characteristics evidenced by 4D low-resolution and 3D high-resolution scanners, new applicative scenarios can be devised, where high-resolution scans are likely to be part of gallery acquisitions, whereas probes are expected to be of lower resolution and potentially acquired with 4D cameras. In this context, reconstructing a higher-resolution model out of a sequence of low-resolution depth frames is a plausible way to bridge the gap between low- and high-resolution acquisitions. In fact, this could open the way to more versatile 3D face recognition methods deployable in contexts where the acquisition of high resolution 3D scans is not convenient or even possible.

Based on these premises, in this work we define an approach specifically tailored to reconstruct a higher-resolution face model from a sequence of low-resolution depth frames, and thus capable of reducing the gap between low- and high-resolution acquisitions. Some recent works explicitly addressed this problem [3], but following a cooperative acquisition protocol for the 3D dynamic sequence. Differently, in this work we focus on the problem of deriving a higher-resolution model from a dynamic sequence of 3D scans in an uncooperative scenario, where acquired persons walk passing through an access point monitored by a Kinect camera, without any specific additional requirement.

A. Related Work

The idea of constructing a higher-resolution representation of an object or scene from multiple low-resolution observations, possibly altered by noise, blurring or geometric warping, has been first introduced for 2D still images. Later, this concept has been extended to 3D generic data for recovering one high-resolution model from a set of low-resolution 3D acquisitions. For example, in [26] data acquired with a time-of-flight camera are upsampled and denoised by using information from a high-resolution image of the same scene taken from a viewpoint close to the depth sensor. Time-of-flight data are processed also in [24] by using an energy minimization framework that explicitly takes into account the characteristic of the sensor, the agreement of the reconstruction with the aligned low resolution maps

This work was partially supported by Bando Ente Cassa di Risparmio di Firenze 2013

and a regularization term to cope with reconstruction of sparse data points. Some works on this topic also focus on 3D faces [14], [21], [22]. In [22], high-resolution 3D face models are used to learn the mapping between low- and high-resolution data. Given a new low-resolution face model the learned mapping is used to compute the high-resolution face model. Differently, in [21] the reconstruction process is modeled as a progressive resolution chain, whose features are computed as the solution to a MAP problem. However, in both the cases, the framework is validated just on synthetic data. In [14], an algorithm is proposed that takes a single face frame from a Kinect depth camera, and produces a high-resolution 3D mesh of the input face. In this approach, the input depth frame is divided into semantically significant regions (eyes, nose, mouth, cheeks) and a database of high-resolution scans is searched for the best matching shape per region. The input depth frame is further combined with the matched database shapes into a single mesh that results in a high-resolution shape of the input person. However, this process does not exploit any information coming from the temporal sequence of scans, rather using the similarity between parts of an individual low-resolution scan with parts of higher-resolution scans used as reference.

In the approaches above, the higher-resolution reconstruction depends on a single 3D low-resolution scan, with the additional information used for reconstruction coming from multiple high-resolution scans used as reference. This completely disregards the temporal dimension available in depth sequences acquired with a Kinect sensor. In order to exploit such temporal information, some methods approach the problem of noise reduction in depth data by fusing the observations of multiple scans [4], [10], [18]. In [18], the Kinect Fusion system is presented, which takes live depth data from a moving Kinect camera and creates a high-quality 3D model for a static scene. Later, in [11] dynamic interaction has been added to the system, where camera tracking is performed on a static background scene and a foreground object is tracked independently of camera tracking. Aligning all depth points to the complete scene from a large environment (e.g., a room) provides very accurate tracking of the camera pose and mapping [18]. However, this approach is targeted to generic objects in internal environments, rather than to faces. In [10], a 3D face model with an improved quality is obtained by a user moving in front of a low resolution depth camera. The model is initialized with the first depth image, and then each subsequent cloud of 3D points is registered to the reference one using a GPU implementation of the ICP algorithm. This approach is used in [4] to investigate whether a system that uses reconstructed 3D face models performs better than a system that uses the individual raw depth frames considered for the reconstruction. To this end, authors present different 3D face recognition strategies in terms of the used probes and gallery. The reported analysis shows that the scenarios where a reconstructed 3D face model is compared against a gallery of reconstructed 3D face models, and where one frame (1F) is compared against multiple frames in the gallery, provide better results compared to the baseline 1F-1F

approach. Although the method is not conceived to increase the resolution of the reconstructed model, it supports the idea that aggregating multiple observations enhances the signal to noise ratio, thus increasing the recognition results with respect to the solution where a single frame is used.

B. Our Method and Contribution

In this paper, we present an original solution to derive one 3D face model from low-resolution depth frames acquired with a Kinect camera. In the proposed approach, first the face is automatically detected and cropped in each depth frame; Then, the 3D face data extracted from the frames are aligned with each other, so as to build a cumulated face model; Finally, the *lowess* non-parametric regression method is used to approximate the face surface from the cumulated face model and remove outliers from the data. The proposed approach has been evaluated on a subset of the *The Florence Superface* dataset [3] which includes, for each individual, one Kinect depth sequence captured in a non-cooperative context, and one high-resolution face scan acquired with a 3dMD scanner. In summary, the main contributions of this paper are:

- A complete approach to reconstruct a 3D face model from a sequence of low-resolution depth frames of the face, acquired with a non-cooperative protocol. In so doing, the resolution of the final face model is higher than the resolution of the individual depth frames;
- An evaluation demonstrating the accuracy of the reconstructed face models with respect to the high-resolution scans, and the face recognition results obtained by using the reconstructed models as probes and the high-resolution scans in the gallery.

The proposed approach shares the idea of reconstructing a high-resolution face model from a sequence of low-resolution depth frames with our previous work in [3]. However, the method we define here differs from our previous proposal in two main aspects: *i)* in [3], subjects were asked to sit in front of the camera at a predefined distance, also requiring them to move the head to the left- and right-side in order to expose a large extent of the face to the sensor. In this paper instead, an uncooperative protocol is assumed, where subjects walk across an access gate monitored by the Kinect sensor, thus posing additional difficulties in terms of varying distance from the sensor, and different velocity; *ii)* in [3], the increased resolution of the reconstructed model was based on the up-sampling and 2D-Box splines approximation of the cumulated 3D points cloud obtained by rigid (ICP) registration of multiple 3D frames of a sequence. In this work instead, we propose non-rigid registration of point sets using the Coherent Point Drift (CPD) method, and outliers removal with the *lowess* algorithm.

The rest of the paper is organized as follows: The scenario and the problem statement are defined in Sect. II; The reconstruction of the high-resolution face model based on non-rigid registration and outliers removal is described in Sect. III. Experimental results are reported and discussed in Sect. IV. Finally, conclusions are given in Sect. V.

II. PROBLEM STATEMENT

In this work, we aim at reconstructing a *3D model* of the face, by processing a sequence of low-resolution *depth frames* (*frames* in the following). The scanner is mounted on the doorjamb at eye level, well positioned for viewing the faces of people walking through the door. In Fig. 1, some RGB and depth frames of a sample sequence are shown. The face region is cropped in each frame by feeding a face detector with the RGB data captured by the scanner.

Let $k \in \{1, \dots, K\}$ be the indexes of the frames where a face is detected and $\mathbf{x}_i^{(k)}$ the 3D coordinates (x , y and the depth value z) of the i -th facial point in the k -th frame $X^{(k)}$. Registration of 3D facial data is operated in reversed acquisition order starting from the last frame. The first order cumulated point cloud $\mathcal{C}^{(1)}$ is obtained by registering $\{\mathbf{x}_i^{(K-1)}\}_i$ to $\{\mathbf{x}_i^{(K)}\}_i$:

$$\mathcal{C}^{(1)} = \mathcal{R}^{(1)} \left(\{\mathbf{x}_i^{(K-1)}\}_i, \{\mathbf{x}_i^{(K)}\}_i \right) \cup \{\mathbf{x}_i^{(K)}\}_i. \quad (1)$$

being $\mathcal{R}(S_1, S_2)$ the registration operator that moves points in the first set S_1 , so as to match the points in S_2 . In the proposed solution, the registration operator is computed using the Coherent Point Drift (CPD) algorithm [17], a probabilistic method for non rigid registration of point sets. Data in the next frame $\{\mathbf{x}_i^{(K-2)}\}_i$ are aligned to the first order cumulated point cloud to yield the second order cumulated point cloud:

$$\mathcal{C}^{(2)} = \mathcal{R}^{(2)} \left(\{\mathbf{x}_i^{(K-2)}\}_i, \mathcal{C}^{(1)} \right) \cup \mathcal{C}^{(1)}. \quad (2)$$

This process is iterated until data from all available frames are registered to yield the K -th order cumulated point cloud $\mathcal{C}^{(K)}$. Compared to standard rigid registration methods, like ICP, CPD has some major advantages, also related to the specific characteristics of the considered acquisition process. In fact, CPD is demonstrated to be more robust to noise and outliers than ICP [17]. Furthermore, ICP performs rigid registration of point sets yielding poor results in case of non-rigid deformations, such as those that can be observed if the user changes his/her facial expression and/or is speaking during the acquisition process.

III. MANIFOLD RECONSTRUCTION

The result of the registration process described in the previous Section is a point cloud that collects a set of points in the 3D space. The generic i -th point $\mathbf{p}_i^{(3)} = (x_i, y_i, z_i)$ can be regarded as the observation, affected by some noise, of the underlying face surface that can be modeled as a 2D manifold embedded in the 3D space. In the proposed approach, reconstruction of the true face surface is formalized as a problem of manifold reconstruction from noisy data. For this purpose, we adopt and extend the approach described in [8], which is based on a combination of dimensionality reduction and local weighted regression.

Principal Component Analysis is used to reduce the dimensionality of the manifold and compute a 2D-embedding of the point cloud. In this way, the intrinsic geometry of the manifold is preserved by mapping close points on the

manifold (that does not necessarily mean close points in the 3D space) into close points on the 2D embedding. Let $\mathbf{p}_i^{(2)} = (u_i, v_i)$ be the coordinates of the point $\mathbf{p}_i^{(3)}$ after projection onto the 2D embedding through *PCA*. Reconstruction of the manifold at point $\mathbf{p}_i^{(3)}$ is accomplished by fitting a low-dimensional polynomial to a subset of the point cloud; only those points of the cloud that are mapped close to $\mathbf{p}_i^{(2)}$ on the 2D-embedding are used to build the local subset. Operatively, the subset of data is determined by a nearest neighbors algorithm on the 2D-embedding. In the literature, the cardinality of this subset is controlled through a *smoothing parameter* $\alpha \in (0, 1)$. The points used to fit the polynomial are the $n\alpha$ closest to $\mathbf{p}_i^{(2)}$ on the 2D-embedding. This set is denoted as $N(\mathbf{p}_i^{(2)})$ and it does not include $\mathbf{p}_i^{(2)}$.

Large values of α produce smooth regression functions that wiggle the least in response to fluctuations in the data. The smaller α is, the closer the regression function will conform to the data, thus yielding poor robustness to noise.

The Weighted Least Squares method is used to fit a second order polynomial to the local subset of data:

$$\begin{aligned} x &= a_1 u^2 + a_2 v^2 + a_3 uv + a_4 u + a_5 v + a_6 \\ y &= a_7 u^2 + a_8 v^2 + a_9 uv + a_{10} u + a_{11} v + a_{12} \\ z &= a_{13} u^2 + a_{14} v^2 + a_{15} uv + a_{16} u + a_{17} v + a_{18}. \end{aligned} \quad (3)$$

Following the original approach [5], the tricube weight function is used to compute the weight associated with the point $\mathbf{p}_j^{(3)}$ of the local subset centered at $\mathbf{p}_i^{(2)}$:

$$w_i(\mathbf{p}_j^{(3)}) = \left[1 - \left(\frac{d(\mathbf{p}_i^{(2)} - \mathbf{p}_j^{(2)})}{\max_{\mathbf{p}_j^{(2)} \in N(\mathbf{p}_i^{(2)})} d(\mathbf{p}_i^{(2)} - \mathbf{p}_j^{(2)})} \right)^3 \right]^3 \quad (4)$$

the above equation holding if $\mathbf{p}_j^{(3)} \in N(\mathbf{p}_i^{(3)})$, and being the weight equal to zero otherwise.

In this way, the regression function that solves the weighted least squares problem gives higher weight to points close to $\mathbf{p}_i^{(2)}$. It should be noticed that $w_i(\mathbf{p}_i^{(3)}) = 0$, that is, the point at the center of the local subset is not used to compute the regression function.

A. Robust Locally Weighted Regression

The presence of outliers in the data badly affects the solution of the weighted least squares problem [12]. To reduce the effect of outliers and to improve robustness of the fitted model a *robust weight* is used for each data point of the local subset. Robust weights are computed based on the residuals measuring the error between the reconstructed (using LWR) and the original points in the 3D space. Let $\mathbf{g}(u, v) : \mathbb{R}^2 \mapsto \mathbb{R}^3$ be the solution of the weighted least squares problem. For each point $\mathbf{p}_j^{(3)}$, the following residual error r_j is considered:

$$r_j = \left\| \mathbf{p}_j^{(3)} - \mathbf{g}(\mathbf{p}_j^{(2)}) \right\|. \quad (5)$$

The value of the residual r_j measures the distance between the original position of a point and its reconstructed

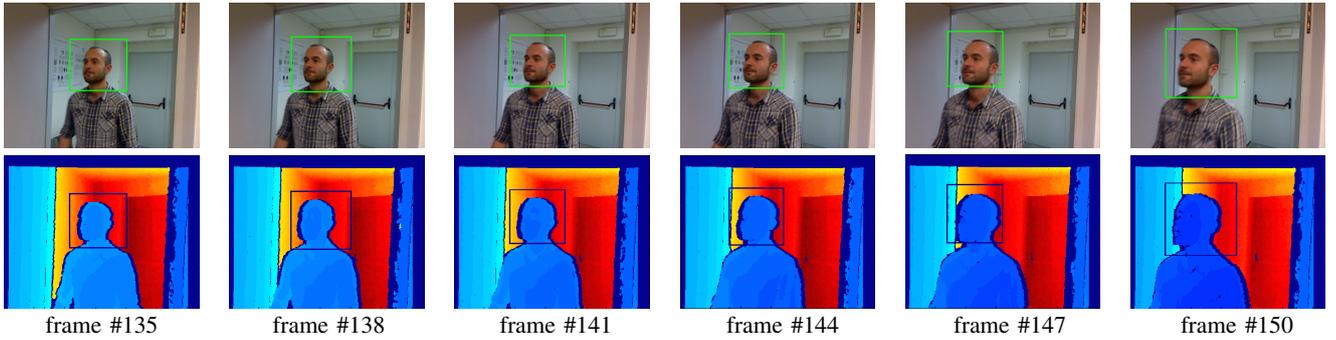


Fig. 1. Sample RGB and colored depth frames acquired by the Kinect. The face region identified by the face detector is evidenced in each frame

counterpart. Thus, it can be used to introduce a correction to the weight w_j . This correction results in a severe decrease of the weight if the value of the residual is high, meaning that the point is likely to be an outlier. For this purpose, the bisquare weight function is used:

$$B(r_j) = \begin{cases} \left[1 - \left(\frac{r_j}{6\hat{r}}\right)^2\right]^2 & \text{if } |r_j/\hat{r}| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

being \hat{r} the median value of the residuals of points of the local subset.

IV. EXPERIMENTAL RESULTS

The proposed approach has been evaluated considering two aspects: The effectiveness of the reconstruction process, by computing the error between the reconstructed models and the corresponding high-resolution scans (Sect. IV-B); The possibility of using the reconstructed models in a face identification experiment, where the reconstructed models are compared against a gallery of high-resolution scans (Sect. IV-C). These experiments have been performed on a subset of the *The Florence Superface* dataset, which is extended to include new acquisitions captured according to the protocol described in this work (see Sect. IV-A).

A. Dataset

Some public datasets exist for face analysis from consumer cameras like Kinect (see for example the *EURECOM Kinect Face* dataset [16], or the *The 3D Mask Attack* database specifically targeted to detect face spoofing attacks [7]). In the experiments reported hereafter, we use and extend a subset of the *The Florence Superface* dataset (UF-S) [3]. This dataset has been originally designed to include 3D high resolution face scans, and 2D videos of the face acquired in different conditions [1]. Successive extensions of the dataset have been done to include depth video sequences of the face, acquired with the Kinect camera according to a cooperative protocol. In this work, we further extend this dataset by capturing depth video sequences for a subset of the subjects according to an uncooperative protocol. In particular, the subset of UF-S used in the experiments includes 16 subjects, each with the following data:

- A 3D high-resolution face scan, with about 40,000 vertices acquired with a 3dMD scanner (see Fig. 2(c)

for some examples). The geometry of the mesh is highly accurate with an average RMS error of about 0.2mm;

- A *Kinect* video sequence (RGB-D), where a person goes through an access gate monitored by the camera. The imaged 3D frames capture the subject while s/he is walking towards the camera at a distance varying from about 1.2m to 40cm (some frames acquired for a sample subject are shown in Fig. 1). This results in video sequences lasting approximately 2 to 4 seconds on average, at 30fps.

Since just a part of these 16 subjects were originally included in the UF-S, we extended the 3D part obtaining a total of 65 high-resolution scans of different subjects.

B. Reconstruction Accuracy

This experiment aims to evaluate the error of the reconstructed 3D model with respect to the 3D high-resolution scan of the same subject, compared to the same measure of error computed between the depth frame used as reference for a sequence (*reference frame*) and the 3D high-resolution scan. In the following, we use the last frame where the face is detected in the RGB data as *reference*.

For each of the 16 subjects used in the experiments we considered: The high-resolution scan; The reconstructed model using the proposed approach; and the low-resolution scan obtained from the reference frame. In all these cases, the 3D facial data are represented as a mesh and cropped using a sphere of radius $95mm$ centered at the nose tip (the approach in [25] is used to detect the nose tip). To measure the error between the high-resolution scan and the reconstructed model of the same subject, they are first aligned through ICP registration [23]. Then, for each point of the reconstructed model its distance to the closest point in the high-resolution scan is computed to build an error-map. As an example, Fig. 2 shows for some subjects (one column per subject), the cropped 3D mesh of the reference frame, the reconstructed model, the high-resolution scan and the error-map between the reconstructed model and the high-resolution scan.

To represent the average error of the reconstructed models and reference frames with respect to high-resolution scans, the *Root Mean Square Error* (RMSE) between their surfaces S and S' is computed considering the vertex correspondences defined by the ICP registration, which associates each vertex

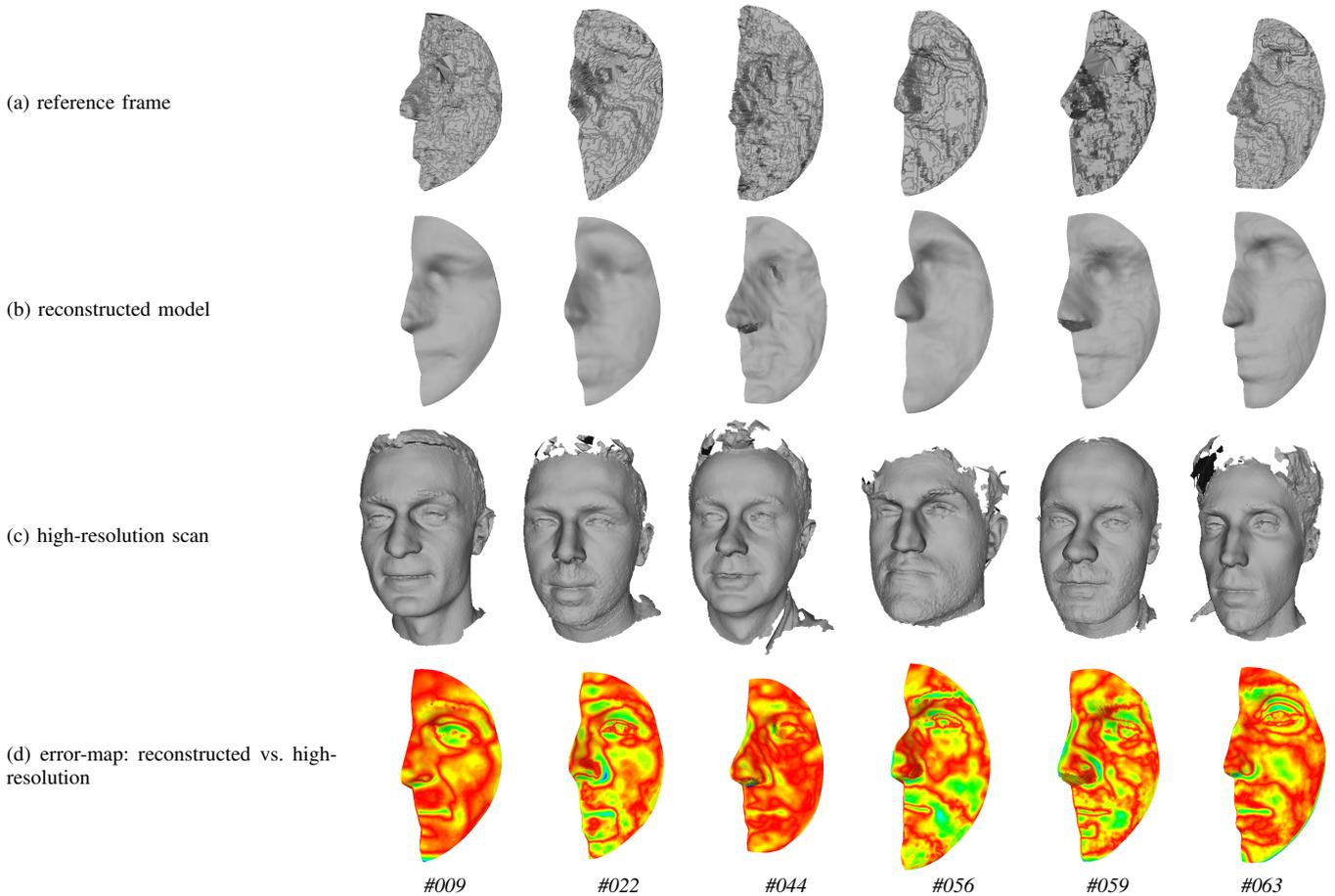


Fig. 2. Each column corresponds to a different subject and reports: (a) The low resolution 3D scan of the reference frame; (b) The reconstructed 3D model; (c) The high-resolution 3D scan. The error-map in (d) shows, for each point of the reconstructed model, the value of the distance to its closest point on the high-resolution scan after alignment (distance increases from red/yellow to green/blue)

$p \in S$ to the closest vertex $p' \in S'$:

$$RMSE(S, S') = \left(\frac{1}{N} \sum_{i=1}^N (p_i - p'_i)^2 \right)^{1/2}, \quad (7)$$

being N the number of correspondent points in S and S' .

TABLE I

RMSE COMPUTED BETWEEN REFERENCE SCANS AND RECONSTRUCTED MODEL WITH RESPECT TO 3D HIGH-RESOLUTION SCANS

models	RMSE			
	min	max	mean	std dev
reference vs. high-resolution	0.96	2.02	1.56	0.31
reconstructed vs. high-resolution	0.79	1.30	1.07	0.14

Results obtained using this distance measure are summarized in Table I. In particular, we reported the average values for the $RMSE$ computed between the high-resolution scan and, respectively, the reconstructed model and the reference scan. On the one hand, values in Table I measure the magnitude of the error between the reconstructed model and the high-resolution scan of the same subject; On the other, they give a quantitative evidence of the increased quality of

the reconstructed model with respect to the reference scan. This latter result is indeed an expected achievement of the proposed approach, since the reconstructed models combine information of several frames of a sequence. However, it is interesting to note the substantial decrease of the error with respect to the reference frame (more than 30% decrease of the mean $RMSE$).

C. Face Identification Results

In this experiment, we consider a subject identification task in which the gallery is composed of high-resolution scans, whereas reconstructed models are used as probes. Description and matching of gallery and probe models is obtained according to the face recognition approach proposed in [2], which is based on the extraction and comparison of local features of the face.

We included 65 high-resolution scans in the gallery, and considered the reconstructed models as probes. Recognition accuracy is evaluated through the Cumulative Matching Characteristic (CMC) curves. Fig. 3 reports the CMC curve in the case the reconstructed models are used as probes. The curve clearly shows that reconstructed models achieve a reasonable high recognition rate of 75% at rank-1, with 100% recognition achieved at rank-4. The green dashed line

in the plot also reports the cumulative recognition curve for the case in which the comparison between different instances of high-resolution scans is performed.

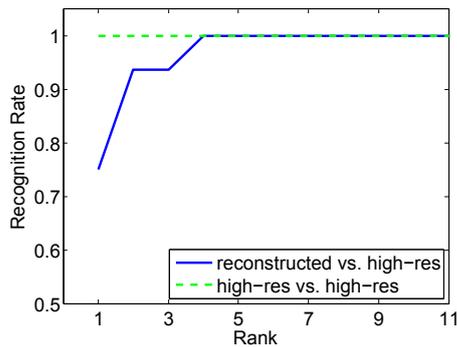


Fig. 3. CMC curves obtained by using the reconstructed models as probes, and the high-resolution scans in the gallery. The plot also reports the case in which different instances of high-resolution scans are used as probes (dashed line)

V. CONCLUSIONS

In this paper, we have defined an approach that permits the construction of a higher-resolution face model starting from a sequence of low-resolution 3D scans acquired with a consumer depth camera. In particular, values of the points of the higher-resolution model are constructed by iteratively aligning the low-resolution 3D frames to a reference frame using the Coherent Point Drift (CPD) algorithm, and filtering the registered 3D data through a variant of the *lowess* method to remove outliers and build the final face model. Qualitative and quantitative experiments have been performed by extending a subset of the *The Florence Superface* dataset with sequences of low-resolution 3D frames acquired with a Kinect camera according to an uncooperative protocol. Results of the reconstruction process of high-resolution models are evaluated by measuring the distance error between the reconstructed models and the high-resolution 3D scan used as the ground truth data of a subject's face. Results support the idea that constructing higher-resolution models from consumer depth cameras can be a viable approach to make such devices deployable in real application contexts that also include identity recognition using 3D faces.

REFERENCES

- [1] A. D. Bagdanov, A. Del Bimbo, and I. Masi. The Florence 2D/3D hybrid face dataset. In *Joint ACM Work. on Human Gesture and Behavior Understanding (J-HGBU)*, pages 79–80, Arizona, USA, Dec. 2011.
- [2] S. Berretti, A. Del Bimbo, and P. Pala. Sparse matching of salient facial curves for recognition of 3D faces with missing parts. *IEEE Trans. on Information Forensics and Security*, 8(2):374–389, Feb. 2013.
- [3] S. Berretti, P. Pala, and A. Del Bimbo. Face recognition by super-resolved 3D models from consumer depth cameras. *IEEE Trans. on Information Forensics and Security*, 9(9):1436–1449, 2014.
- [4] J. Choi, A. Sharma, and G. Medioni. Comparing strategies for 3D face recognition from a 3D sensor. In *Proc. IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1–6, Gyeongju, Korea, Aug. 2013.

- [5] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.
- [6] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(9):2270–2283, 2013.
- [7] N. Erdogmus and S. Marcel. Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect. In *IEEE Int. Conf. on Biometrics: Theory, Applications and Systems, (BTAS)*, Washington DC, USA, Sept. 2013.
- [8] P. Frasconi, L. Silvestri, P. Soda, R. Cortini, F. Pavone, and G. Iannello. Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images. *BIOINFORMATICS*, 30:i587–i593, 2014.
- [9] G. Goswami, S. Bharadwaj, M. Vatsa, and R. Singh. On RGB-D face recognition using Kinect. In *Proc. IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, Washington DC, USA, Sept. 2013.
- [10] M. Hernandez, J. Choi, and G. Medioni. Laser scan quality 3-d face modeling using a low-cost depth camera. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 1995–1999, Bucharest, Romania, Aug. 2012.
- [11] S. Izadi, R. Newcombe, D. Kim, O. Hilliges, D. Molyneaux, S. Hodges, P. Kohli, J. Shotton, A. Davison, and A. Fitzgibbon. Kinectfusion: realtime dynamic 3D surface reconstruction and interaction. In *Proc. ACM SIGGRAPH*, page 1, Vancouver, Canada, Aug. 2011.
- [12] W. Jacoby. Loess: A nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, 19(4):577–613, 2000.
- [13] B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna. Using kinect for face recognition under varying poses, expressions, illumination and disguise. In *Proc. IEEE Work. on Applications of Computer Vision (WACV)*, pages 186–192, Clearwater, Florida, Jan. 2013.
- [14] S. Liang, I. Kemelmacher-Shlizerman, and L. G. Shapiro. 3d face hallucination from a single depth frame. In *Int. Conf. on 3D Vision*, pages 1–8, Tokyo, Japan, Dec. 2014.
- [15] R. Min, J. Choi, G. Medioni, and J.-L. Dugelay. Real-time 3D face identification from a depth camera. In *Proc. Int. Conf. on Pattern Recognition (ICPR)*, pages 1739–1742, Tsukuba, Japan, Nov. 2012.
- [16] R. Min, N. Kose, and J.-L. Dugelay. Kinectfacedb: A kinect database for face recognition. *IEEE Trans. on Systems, Man, and Cybernetics: Systems*, 44(11):1534–1548, Nov 2014.
- [17] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, Dec 2010.
- [18] R. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Proc. IEEE Int. Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–10, Basel, Switzerland, Oct. 2011.
- [19] M. Pamplona Segundo, S. Sarkar, D. Goldgof, L. Silva, and O. Bellon. Continuous 3d face authentication using rgb-d cameras. In *Proc. IEEE Work. on Biometrics*, pages 1–6, Portland, Oregon, USA, June 2013.
- [20] M. Pamplona Segundo, L. Silva, and O. Bellon. Real-time scale-invariant face detection on range images. In *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, pages 914–919, Anchorage, Alaska, USA, Oct. 2011.
- [21] G. Pan, S. Han, Z. Wu, and Y. Wang. Super-resolution of 3D face. In *Proc. European Conf. on Computer Vision (ECCV)*, pages 389–401, Graz, Austria, May 2006.
- [22] S. Peng, G. Pan, and Z. Wu. Learning-based super-resolution of 3D face model. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, volume II, pages 382–385, Genoa, Italy, Sept. 2005.
- [23] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. Int. Conf. on 3D Digital Imaging and Modeling (3DIM)*, pages 145–152, Quebec City, Canada, May 2001.
- [24] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. Lidarboost: Depth superresolution for ToF 3D shape scanning. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 343–350, Miami, Florida, USA, June 2009.
- [25] C. Xu, T. Tan, Y. Wang, and L. Quan. Combining local features for robust nose location in 3D facial data. *Pattern Recognition Letters*, 27(13):1487–1494, 2006.
- [26] Q. Yang, R. Yang, J. Davis, and D. Nister. Spatial-depth super resolution for range images. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Minneapolis, Minnesota, USA, June 2007.