Robust 3D Face Shape Reconstruction from Single Images via Two-Fold Coupled Structure Learning

Pengfei Dou
bensondou@gmail.com
Yuhang Wu
yuhang@cbl.uh.edu
Shishr K. Shah
sshah@central.uh.edu
Ioannis A. Kakadiaris
ioannisk@uh.edu

Computational Biomedicine Lab
Department of Computer Science
University of Houston
Houston, TX, USA

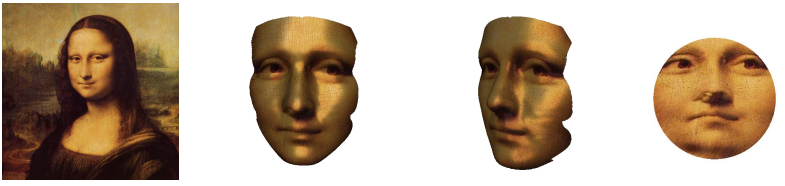
Abstract

In this paper, we propose a robust method for monocular face shape reconstruction (MFSR) using a sparse set of facial landmarks that are detected by most of the off-the-shelf landmark detectors. Different from the classical shape-from-shading framework, we formulate the MFSR problem as a Two-Fold Coupled Structure Learning (2FCSL) process, which consists of learning a regression between two subspaces spanned by 3D sparse landmarks and 2D sparse landmarks, and a coupled dictionary learned on 3D sparse and dense shape using K-SVD. To handle variations in face pose, we explicitly incorporate pose estimation in our method. Extensive experiments on both synthetic and real data from two challenging datasets using manual and automatic landmarks indicate that our method achieves promising performance and is robust to pose variations and landmark localization noise.

1 Introduction

Although research on face recognition has been progressing rapidly over the past decades, practical deployment of face recognition systems is still limited. One challenging issue preventing the practical application of most face recognition systems is the misalignment between 2D faces being matched caused by pose variation. According to Wagner *et al*. [36], without good alignment, state-of-the-art methods [40] would suffer severe degradation in performance. To solve the misalignment problem, two types of methods have been proposed. The first type of solutions propose learning a regression between subspaces spanned by different poses [5, 16, 27]. However, these methods are constrained by the consistency of pose in training and testing images. In real applications face poses are more unconstrained, and potential occlusions may introduce large bias when conducting hard regression.

As a result, the second type of methods, which propose aligning gallery and probe faces using 3D faces, is attracting increasing attention in recent years. By aligning the faces to a



(a) (b) (c) (d)

Figure 1: (a) The input 2D image; (b) frontal view of the face shape reconstruction; (c) profile view of the face shape reconstruction; and (d) lifted UV texture.

3D face model, correspondence is built, which greatly facilitates the comparison between 2D faces. For these algorithms [15, 19, 32] one critical step is acquiring a personalized 3D face shape model, which is usually captured using 3D camera system during enrollment. Until recently, the high cost of 3D cameras has limited such methods from being practically deployed. In addition, due to existing deployment of 2D cameras and their broader availability, there is a need for monocular face shape reconstruction (MFSR) [13, 18].

In this paper, we propose a Two-Fold Coupled Structure Learning (2FCSL) algorithm that is capable of reconstructing 3D face models purely based on a sparse set of 2D landmarks detected automatically by most of the recently proposed landmark detectors. The novelty of our method lies in that, instead of assuming a selective mapping, we implicitly learn the correlation between a sparse set of 3D facial landmarks and the 3D dense face model. We believe this is more effective, as 3D face landmarks are not necessarily a sparse subset of the 3D dense model. Compared with existing feature point-based methods which use a simple linear mapping to select a sparse subset of landmarks from the dense model, our method achieves better robustness to landmark localization errors that could be encountered during landmark detection. Our method is fully automatic, thus could be easily integrated in most 3D-aided face recognition systems. Our method could also serve as an independent module that generates customized 3D face models for further applications. We conducted several experiments and compared with a recent work [25]. Experimental results demonstrate that our model achieves much higher accuracy and better robustness. Fig. 1 depicts the reconstructed 3D face of Mona Lisa using the famous painting by Leonardo da Vinci and the lifted texture in a pre-registered UV space [15].

Our main contributions are:

i. We propose a novel algorithm, 2FCSL, that is capable of reconstructing a 3D human face shape model from single images using a sparse set of facial landmarks and achieves state-of-the-art performance.

ii. By explicitly incorporating 3D-2D pose estimation, our method achieves robustness to arbitrary pose of the subject's face.

iii. By formulating the MFSR problem into a two-fold coupled structure learning problem, our method achieves robustness to landmark localization noise.

The rest of the paper is organized as follows. Section 2 reviews the literature in MFSR. Our 2FCSL method is presented in Section 3. Experiments and results are presented in Section 4. Section 5 summarizes our conclusions.

2 Related Work

The problem of estimating the 3D shape of human faces from single images is of great interest and has attracted considerable attention and research effort over the past decade. Many approaches recently proposed to solve this problem could be considered as extensions of Shape-from-Shading (SFS) methods, where a 3D shape is optimized to generate 2D renderings that match the input images. Direct application of SFS, however, has limited success in 3D face reconstruction because the human face exhibits large albedo variation and has both concave and convex regions [39]. To overcome this limitation, previously proposed algorithms tend to use either shape priors of human faces [28, 44] or certain assumptions about skin reflectance [9, 22]. Other methods in the literature propose to infer 3D face shape by fitting a set of feature points between the 2D image and the 3D model. Compared with SFS-based methods, feature-point-based methods do not make explicit assumptions about face albedo and lighting, and thus are more robust to challenging illumination conditions. Of all these methods, most propose to use a 3D morphable model (3DMM) constructed on a set of well aligned 3D facial shapes [2, 4, 7, 14, 23, 24, 29, 37, 38, 39, 42], while others require only a single reference shape [11, 17, 18, 25].

2.1 Methods based on Shape-from-Shading

Shape-from-shading is a family of methods proposed to recover the 3D shape of an object using lighting and shading cues [25]. SFS-based methods usually apply a brightness constraint that indicates the total brightness error between the rendering of reconstructed 3D shape and the input image. To render the human face from reconstructed 3D shape, SFS generally requires knowledge of the scene's lighting conditions and human faces' reflectance properties, which are complex in nature. As a result, most SFS-based methods for reconstructing 3D human face pose strict assumptions regarding lighting sources and face albedo to simplify the problem.

Dovgard and Basri [9] introduced an SFS-based method by taking into account not only statistical constraints but also a geometric constraint of facial symmetry. By assuming orthographic projection and Lambertian reflectance model, brightness constraints that incorporate unknown surface albedos and surface depths could be derived. The main drawback of the method is that it is strictly limited to frontal faces and known illumination. As the human face is not symmetric, the performance of the algorithm is intrinsically bounded. Smith and Hancock [29] proposed to embed a statistical model of surface normals instead of surface depths into the SFS framework. Their approach recovers a field of surface normals from a single intensity image by exploiting the direct relationship between surface orientations and image intensities. However, during 3D face reconstruction, an input 2D face first needs to be manually aligned. This method is also limited to frontal face view and known illumination. Rara *et al*. [23] proposed another SFS-based approach. By employing spherical harmonics (SH), the proposed method achieves the capability of dealing with arbitrary illumination. Rara *et al*. extended their method [24], where the classical SFS equation was cast as a Partial Least Squares (PLS) regression problem, thus achieving rapid computation of the solution. SFS works well for uniformly concave or convex objects; however, when it comes to the more complex human face, with both concave and convex regions, SFS would probably fail. This is because subtle changes in surface normals could cause significant changes in the corresponding integrated surface. Castelan and Hancock [6] proposed to integrate into the SFS framework a local shape-based method for enforcing the convexity of the integrated surface.

They utilized the local descriptors of shape-index and curvedness to characterize the regions and perform necessary correction to surface normal orientations to enforce surface convexity and the condition that the integrated surface has a global height maximum. An alternative to 3DMM used in SFS-based methods is using a single reference shape. One advantage of using a single reference shape is its simplicity, as constructing 3DMM requires all training 3D shapes to be point-wisely aligned. Kemelmacher-Shlizerman and Basri [18] proposed to use the input image as a guide to "mold" a single reference model to recover the corresponding 3D shape of either a different individual or a generic face. Another SFS-based approach that uses a single reference model is proposed by Kemelmacher-Shlizerman and Basri [17]. By employing a spherical harmonic approximation to model reflectance their method allows for multiple unknown light sources and attached shadows.

2.2 Feature-Point-Based Method

Of all these SFS-based methods, one major drawback is that lighting and face albedo should be known as a prior, or well approximated. In addition, as surface depths are computed by integrating surface normals, subtle changes in surface normals could cause significant changes in the corresponding integrated surface. As a result, the input image is required to be accurately aligned with the 3D face model. To overcome the limitation of SFS-based methods, another category of approaches based on fitting a set of sparse feature points was proposed. Blanz *et al*. [4] proposed a learning-based approach based on a 3DMM. By converting a set of training 3D shapes into a vector representation and restricting possible solutions in the span of training samples, the proposed method infers a dense 3D shape from sparse feature points based on their correlations embedded in the 3D shape model. Hu *et al*. [14] proposed an analysis-by-synthesis framework for face recognition under challenging conditions and proposed to reconstruct a personalized 3D face shape from a single frontal face image with neutral expression and normal illumination. Wang *et al*. [37] proposed a similar approach. Given an input image, facial feature points are first detected. Then, an EM framework is employed to infer the 3D shape and pose parameters iteratively. Zhang and Samaras [42] proposed to build the statistical models directly in 3D space by combining the spherical harmonics (SH) illumination representation and a 3DMM of human faces to recover 3D face shape from single images with arbitrary pose and illumination. Castelan *et al*. [7] proposed a coupled statistical model to recover 3D face shape from single images of faces. Aldrian and Smith [2] proposed a method with two steps that can be iterated and interleaved. The estimation of the 3D shape parameters is performed in a probabilistic framework which aims to find the most likely shape coefficients based on an observation of a sparse set of 2D feature points. Wang and Yang [39] proposed to learn mappings between 2D images and corresponding 3D shapes via manifold alignment.

One major limitation of feature-point-based methods is the requirement for explicit estimation of the camera view-point parameters, since coordinate-descent approaches are prone to be trapped in local minima and provide no guarantee on the optimality of the estimations. To overcome this limitation, Wang *et al*. [38] proposed a one-shot optimization approach to simultaneously determine the optimal 3D landmark model, the corresponding 2D projections, and the visibility states without explicit estimation of the camera viewpoint. Song *et al*. [30] proposed a coupled radial basis function network (C-RBF) method. Given paired 2D and 3D training data, C-RBF learns their intrinsic representations and the corresponding mapping functions. Given a new 2D face, C-RBF first computes its 2D intrinsic representation, which is then approximated by a linear combination of its K-nearest neighbors. Using

the same coefficients, the 3D intrinsic representation and the corresponding 3D face are recovered. Castelan *et al*. [8] evaluated four different Subspace Multiple Linear Regression methods for recovering 3D face shape from single images. Hassner [11] proposed a data-driven method for recovering the 3D shapes of faces viewed in single and unconstrained photos. Rara *et al*. [25] proposed a model-based approach for 3D facial shape recovery using a small set of feature points from an input image of unknown pose and illumination. Our work makes advances in this direction.

3 Method

In this section, we present our fully automatic framework for recovering a dense 3D face shape from a single image. Our approach is based on the assumption that a sparse set of fiducial landmarks have been located by a landmark detector. After successfully localizing the landmarks, two coupled relationships are learned by our algorithm: (i) 2D sparse landmarks (2DSL) to 3D sparse landmarks (3DSL), and (ii) 3DSL to 3D dense landmarks (3DDL). Our algorithm first recovers the 3DSL coefficient by learning a coupled regression model from 2DSL to 3DSL, then the dense 3D face shape can be reconstructed by recovering the coefficients of a coupled-dictionary.

Here are some notations we use in the following section: We represent the dense 3D face as a shape vector $Y^i_{3D} = [x^i_1, y^i_1, z^i_1, \cdots, x^i_L, y^i_L, z^i_L]$ that consists of the XYZ coordinates of its L vertices. With N 3D training faces, the shape model could be constructed by stacking N shape vectors together, $\gamma^d_{3D} = (Y^1_{3D}, \cdots, Y^N_{3D})$. The 3DSLs are represented by χ^s_{3D}=$(X^1_{3D}, \cdots, X^N_{3D})$, where X^i_{3D} is the vector representation of M 3D landmarks, and the 2DSLs are represented by χ^s_{2D}=$(X^1_{2D}, \cdots, X^N_{2D})$, where X^i_{2D} is the vector representation of M 2D landmarks.

3.1 Sparse 3D Shape Recovery by Coupled-Basis PLS Regression

Given a 2D image X^I_{2D} with 2DSL, we first estimate the 3D-2D projection matrix P using least squares minimization, such that $X^I_{2D} = P\bar{X}_{3D}$, where $\bar{X}_{3D}$ is the mean of 3DSLs in the training database. By projecting each 3DSL via P, we obtain the corresponding 2DSL χ^s_{2D}, which is generated on-line.

According to the statistical shape model, each landmark vector could be presented by the sum of the mean and its weighted orthogonal basis [12]. We decomposed the 2DSL and 3DSL into the following form:

$$X^i_{3D} = \bar{X}_{3D} + \sum_{m=1}^{N-1} a^i_m U^s_{3D} \qquad \textit{and} \qquad X^i_{2D} = \bar{X}_{2D} + \sum_{n=1}^{N-1} a^i_n U^s_{2D} \quad , \tag{1}$$

where U represents the orthogonal basis, a is the weights and $\bar{X}$ represents the mean of the training set in the corresponding 2D/3D model. The reason we do not attempt to learn a shared coupled structure as in Sec. 3.2 is because the 2D and 3D shapes may have different inner constraints due to the difference in dimensions. It is more appropriate to learn a regression model to regress the coefficients of 2DSL to 3DSL rather than force them to share the same representation.

Let $A_m = [a^1_m, a^2_m \cdots, a^{N-1}_m]$ and $A_n = [a^1_n, a^2_n \cdots, a^{N-1}_n]$ be compact representations of the corresponding shapes. A PLS regression [10] was learned from A_n to A_m. The regression

searches for the latent vectors that perform a simultaneous decomposition of A_m and A_n with the constraint to maximize the covariance between them. Since the 2DSL and 3DSL are implicitly correlated by a linear projection matrix and the landmarks in both shape models are constrained by the geometry layout, PLS regression can capture better the semantic relationships between the 2D shape and the 3D shape. In PLS regression, $A_n = TP^T$, $T^TT = I$, $\hat{A}_m = TBC^T = A_nP_{PLS}$, where $P_{PLS} = (P^{T+})BC^T$, B is a diagonal matrix with regression weights, and C is the weight matrix of dependent variables. The details of PLS regression are described in [10].

Given the 2D landmarks X^I_{2D} of the input face, we first obtain its 2D compact representation by solving for $a^I_n = {U^s_{2D}}^{-1}(X^I_{2D} - \bar{X}_{2D})$. Then the a^I_m is recovered by $a^I_m = a^I_nP_{PLS}$. The shape coefficient a_m can be subsituted to reconstruct the 3DSL through $X^R_{3D} = \bar{X}_{3D} + a^I_mU^s_{3D}$. As a result, we obtain the following equation: $X^R_{3D} = \bar{X}_{3D} + {U^s_{2D}}^{-1}(X^I_{2D} - \bar{X}_{2D})P_{PLS}U^s_{3D}$.

3.2 Dense 3D shape recovery by coupled-dictionary learning

Algorithm 1 Two-Fold Coupled Structure Learning framework for 3D dense landmark recovery

Input:

i. Input image feature points x^I_{2D}.

ii. Model database γ^d_{3D} and χ^s_{3D}.

Output: Recovered 3D dense shape Y^R_{3D}

1: Solve for the camera projection matrix, and determine P such that $x^R_{2D} = P \cdot \bar{X}^s_{3D}$.
2: Project all 3DSLs to 2DSL using the computed projection matrix: $X^i_{2D} = PX^i_{3D}$.
3: Build the 3DSL statistical shape model using PCA: $\chi_{3D} = \bar{X}_{3D} + A_mU^s_{3D}$.
4: Build the 2DSL statistical shape model using PCA: $\chi_{2D} = \bar{X}_{2D} + A_nU^s_{2D}$.
5: Regress the coupled parameters from A_n to A_m using PLS regression and obtain the projection matrix P_{PLS}.
6: Recover the 3DSL: $X^R_{3D} = \bar{X}_{3D} + U^{s'}_{2D}(X^I_{2D} - \bar{X}_{2D})P_{PLS}U^s_{3D}$.
7: Solve Eq. 2 with K-SVD to obtain $\Lambda^d_{3D}, \Lambda^s_{3D}$.
8: Solve Eq. 3 with Lasso to obtain α^*.
9: 3D dense landmark is recovered by $Y^R_{3D} = \frac{\Lambda^d_{3D}\alpha^*}{\beta_0}$.

After we obtain the 3DSL, we aim to reconstruct the 3DDL. Since the 3DSL and the 3DDL are sampled from the same 3D face, they have the same dimensions and are intrinsically correlated. By forcing them to share the same representation coefficients in a coupled subspace, their underlying relationship could be implicitly encoded. In this way, a coupled dictionary model is built (Eq. 2). The coefficient α can be treated as the identity of the 3D face. The dictionary is built with L_1 constraints, so the input will be sparsely represented by the 3DDL model. Sparse representation has been demonstrated to be effective in [40, 43], which ensures a shape represented by only a few distinct elements in the dictionary. It works especially well to cut down the noise that does not follow the constraint of the dictionary's training set.

$$\arg\min_{\alpha,\Lambda^d_{3D},\Lambda^s_{3D}} \left\| \begin{bmatrix} \beta_0\gamma^d_{3D} \\ \chi^s_{3D} \end{bmatrix} - \begin{bmatrix} \beta_0\Lambda^d_{3D} \\ \Lambda^s_{3D} \end{bmatrix} \alpha \right\|_2^2 \quad s.t. \|\alpha\|_1 \leq \beta_1 \quad . \tag{2}$$

Eq. 2 can be solved by the K-SVD [1] algorithm directly, and we obtain Λ^d_{3D}, the sparse dictionary of 3DDL, and Λ^s_{3D}, the sparse dictionary of 3DSL. β_1 is the parameter controlling the sparsity of the coefficient, and β_0 is the parameter balancing the 3D dense shape and 3D sparse shape. All parameters including the dictionary size t are tuned empirically.

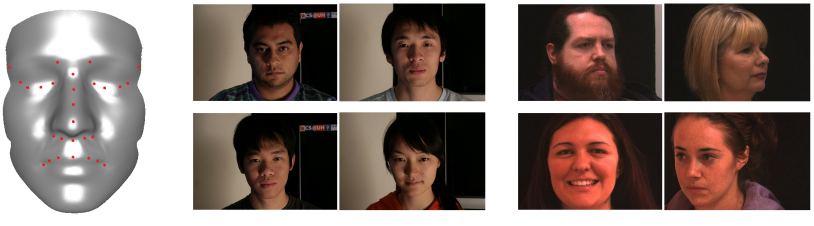
After we successfully reconstruct X^R_{3D} (Sec. 3.1), we first recover the coefficient of 3DSL by solving Eq. 3:

$$\arg\min_{\alpha^*} \left\| X^R_{3D} - \Lambda^s_{3D}\alpha^* \right\|_2^2 + \beta_2 \|\alpha^*\|_1 + \beta_3 \|\alpha^*\|_2 \quad . \tag{3}$$

The LASSO algorithm [31] is used to obtain the solution, and we force α^* to have similar sparsity as in Eq. 2 by automatically tuning β_2. We also apply L_2 regularization on the solution α^*. When we have α^* the final Y^R_{3D} is reconstructed using the coupled model's property, $Y^R_{3D} = \frac{\Lambda^d_{3D}\alpha^*}{\beta_0}$. Alg. 3.2 summarizes the major steps of our proposed method.

4 Experiments

In this section, we evaluate the proposed method using both synthetic and real data. To construct the 3DMM, we selected 165 3D facial scans from the FRGC v2.0 [21] and the UHDB31 [34] datasets. By fitting them with the Annotated Face Model (AFM) [15], which consists of 7,597 vertices, dense vertex-wise correspondence was established across the training 3D faces. On each fitted 3D face, we manually annotated 28 facial landmarks, which are illustrated in Fig. 2(a). We selected these key facial points because they are stable against facial expressions and they are detected on 2D faces by many recently proposed landmark detectors [26].

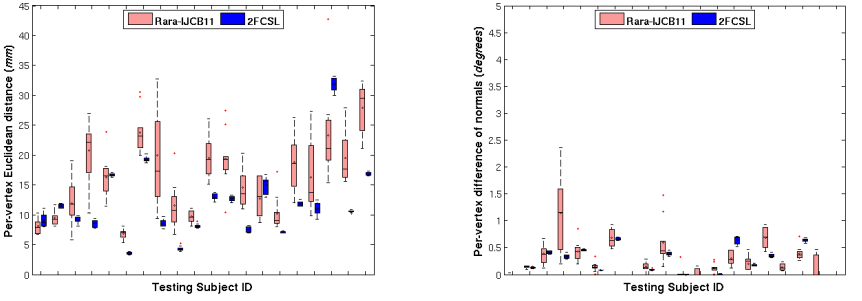


(a) (b) (c)

Figure 2: Illustration of our 3DMM and 2D faces in both databases: (a) Mean 3D face shape and mean 3D landmarks of the constructed 3DMM; (b) UHDB11 gallery set; and (c) BDCP probe set.

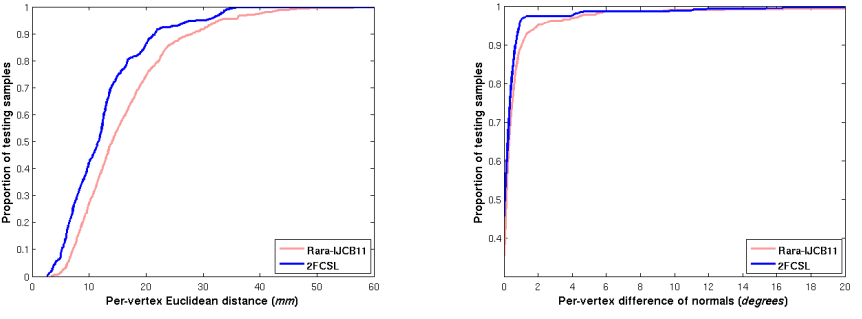
4.1 Experiments on Synthetic Data

To evaluate the reconstruction accuracy of our proposed algorithm, we first conducted experiments on synthetic data. Of the 165 collected 3D faces, we selected one as testing data



(a) (b)

Figure 3: Boxplot of reconstruction error on synthetic testing data: (a) per-vertex Euclidean distance; and (b) per-vertex difference of normals.



(a) (b)

Figure 4: Cumulative distribution of reconstruction error on synthetic testing data: (a) per-vertex Euclidean distance; and (b) per-vertex difference of normals.

and used all others as training data. By projecting the 3D landmarks of the selected 3D face into 2D space using randomly generated projection matrix corresponding to pan rotations within $[-20°, +20°]$, we synthesized multiple (ten) 2D faces with arbitrary poses. To simulate landmark localization error, we added Gaussian noise with variance $\sigma = 5$ *pixels* to the landmark location. Based on the synthesized 2D data, 3D faces were reconstructed and reconstruction error was computed in terms of per-vertex Euclidean distance (*mm*) and per-vertex difference of normals (*degrees*). We repeated this process 80 times and empirically selected the optimal values for our parameters to be, $t = 164$, $\beta_0 = 0.15$, $\beta_1 = 20$, $\beta_2 = 0.3$, and $\beta_3 = 0.35$. We also implemented a recently proposed algorithm of [25], which we will refer to as Rara-IJCB11, and compared it with our method.

Fig. 3 depicts the boxplot (randomly selected 20 out of 80 tested subjects) of the reconstruction error. Fig. 4 depicts the cumulative distribution of all testing samples w.r.t. the reconstruction error. Results indicate that, when compared with [25], our 3D reconstruction is more accurate and our algorithm is more robust to pose variation.

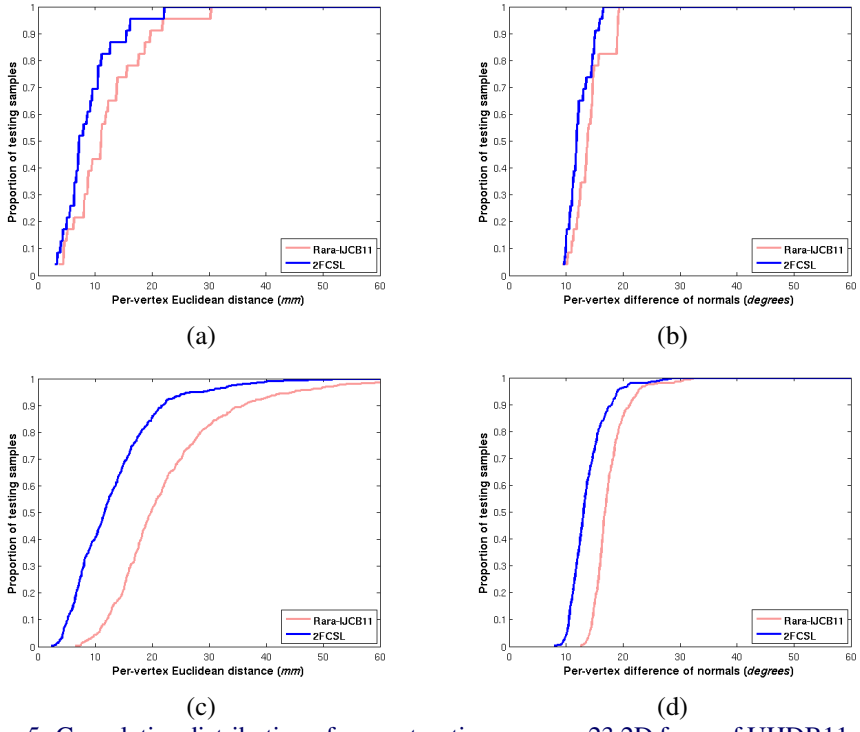


Figure 5: Cumulative distribution of reconstruction error on 23 2D faces of UHDB11 gallery set (a-b), and 381 2D faces of BDCP probe set (c-d): (a,c) per-vertex Euclidean distance; and (b,d) per-vertex difference of normals.

4.2 3D Face Reconstruction on Real Datasets

In the second experiment, 3D face reconstruction accuracy on two real datasets is systematically evaluated. The UHDB11 dataset [33] and the BEST Development Challenge Problem (BDCP) face dataset [35] are used in this experiment. The size of face region in UHDB11 is around 1,000*1,000 pixels and as for BDCP is about 250*250 pixels.

The UHDB11 dataset consists of 23 subjects. The gallery set contains a 2D frontal face and a 3D facial scan of each subject. Fig. 2(b) depicts several examples of the 2D faces in the gallery of the UHDB11 dataset. We used the 2D frontal faces with manual landmarks in the gallery to reconstruct the 3D faces for all the 23 subjects. The 3D facial scans in the gallery were then used as groundtruth to compute the error of the reconstructions. Figs. 5(a-b) depict the cumulative distribution of reconstruction error for both methods. Results show that our method outperforms [25] and achieves more accurate reconstruction of 3D face shapes.

The probe set of the BDCP database consists of 381 2D and 3D faces of 95 subjects [20]. Different from the gallery set of the UHDB11 database, the 2D faces in BDCP's probe set, as illustrated in Fig. 2(c), have arbitrary pose. Instead of using manual annotations, we applied a recently proposed 2D landmark detector [41] to detect the 28 facial landmarks on the 381 2D faces. We have also manually checked our 3DMM to make sure it did not include the 3D face of any subject in the BDCP subset. Figs. 5(c-d) depict the cumulative distribution

of reconstruction error for both methods. Our method outperforms Rara *et al.*'s method [25] with much smaller reconstruction error. This result also indicates that our method is more robust against 2D face pose and landmark localization noise.

5 Conclusions

We have presented a robust method for monocular face shape reconstruction using a sparse set of facial landmarks. We proposed to recover the 3D face shape from a single 2D image by learning a two-fold coupled structure that consists of the regression between two subspaces spanned by 3D sparse landmarks and 2D sparse landmarks, respectively, and a coupled dictionary of 3D dense and sparse shape. Our experiments on synthetic data indicate that our method is robust to face pose variations. Experiments on real datasets with both manual annotations and automatic landmarks detected by an off-the-shelf landmark detector suggest that our algorithm is robust to landmark localization noise and achieves promising performance.

6 Acknowledgment

This research was funded in part by the US Army Research Lab (W911NF-13-1-0127) and the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the sponsors.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54:4311–4321, 2006.

[2] O. Aldrian and W.A.P Smith. A linear approach of 3D face shape and texture recovery using a 3D morphable model. In *Proc. British Machine Vision Conference*, pages 75.1–10, Aberystwyth, Wales, UK, Aug. 30-Sep. 2 2010.

[3] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, Feb. 2003.

[4] V. Blanz, A. Mehl, T. Vetter, and H.-P. Seidel. A statistical method for robust 3D surface reconstruction from sparse data. In *Proc. International Symposium on 3D Data Processing Visualization and Transmission*, pages 293–300, Thessaloniki, Greece, September 6-9 2004.

[5] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou. Regularized latent least square regression for cross pose face recognition. In *Proc. International Joint Conference on Artificial Intelligence*, pages 1247–1253, Beijing, August 2013.

[6] M. Castelán and E.R. Hancock. Acquiring height data from a single image of a face using local shape indicators. *Computer Vision and Image Understanding*, 103(1):64 – 79, 2006.

[7] M. Castelán, W.A.P. Smith, and E.R. Hancock. A coupled statistical model for face shape recovery from brightness images. *IEEE Transactions on Image Processing*, 16 (4):1139 –1151, April 2007.

[8] M. Castelán, G.A. Puerto-Souza, and J. Van Horebeek. Using subspace multiple linear regression for 3d face shape prediction from a single image. In *Advances in Visual Computing*, pages 662–673. Springer, 2009.

[9] R. Dovgard and R. Basri. *Statistical symmetric shape from shading for 3D structure recovery of faces*. Springer, 2004.

[10] P. Geladi and B.R. Kowalski. Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185:1–17, 1986.

[11] T. Hassner. Viewing real-world faces in 3D. In *Proc. IEEE International Conference on Computer Vision*, Sydney, Australia, December 1-8 2013.

[12] T. Heimann and H. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical image analysis*, 13(4):543–563, August 2009.

[13] G. Hsu and H. Peng. Face recognition across poses using a single 3D reference model. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 869–874, Portland, OR, June 23-28 2013.

[14] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang. Automatic 3D reconstruction for face recognition. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 843–848, Seoul, South Korea, May 17-19 2004.

[15] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.

[16] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen. Multi-view discriminant analysis. In *Proc. European Conference on Computer Vision*, pages 808–821, Firenze, Italy, October 7 - 13 2012.

[17] I. Kemelmacher-Shlizerman and R. Basri. Molding face shapes by example. In *Proc. European Conference on Computer Vision*, pages 277–288, Graz, Austria, May 7 - 13 2006.

[18] I. Kemelmacher-Shlizerman and R. Basri. 3D face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):394–405, 2011.

[19] X. Lu and A.K. Jain. Deformation modeling for robust 3D face matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1346–1357, 2008.

[20] P. Moutafis and I.A. Kakadiaris. Can we do better in unimodal biometric systems? A novel rank-based score normalization framework for multi-sample galleries. In *Proc. 6^{th} IARP International Conference on Biometrics*, pages 1–8, Madrid, Spain, June 4-7 2013.

[21] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 947–954, San Diego, CA, June 20–25 2005.

[22] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of a convex lambertian object. *Journal of the Optical Society of America A*, 18(10):2448–2459, 2001.

[23] H. Rara, S. Elhabian, T. Starr, and A. Farag. Model-based shape recovery from single images of general and unknown lighting. In *Proc. IEEE International Conference on Image Processing*, pages 517–520, Cairo, November 2009.

[24] H. Rara, S. Elhabian, T. Starr, and A. Farag. 3D face recovery from intensities of general and unknown lighting using partial least squares. In *Proc. IEEE International Conference on Image Processing*, pages 4041–4044, Hong Kong, September 2010.

[25] H.M. Rara, A.A. Farag, and T. Davis. Model-based 3D shape recovery from single images of unknown pose and illumination using a small number of feature points. In *Proc. International Joint Conference on Biometrics*, pages 1 –7, Washington, DC, October 2011.

[26] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *IEEE International Conference on Computer Vision Workshops*, pages 397–403, Darling Harbour, Sydney, December 1-8 2013.

[27] A. Sharma, A. Kumar, H. Daume, and D.W. Jacobs. Generalized multiview analysis: A discriminative latent space. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2160–2167, Providence, RI, June 16-21 2012.

[28] I. Shimshoni, Y. Moses, and M. Lindenbaum. Shape reconstruction of 3D bilaterally symmetric surfaces. *International Journal of Computer Vision*, 39(2):97–110, 2000.

[29] W.A.P. Smith and E.R. Hancock. Recovering facial shape using a statistical model of surface normal direction. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 28(12):1914–1930, December 2006.

[30] M. Song, D. Tao, X. Huang, C. Chen, and J. Bu. Three-dimensional face reconstruction from a single image by a coupled RBF network. *IEEE Transactions on Image Processing*, 21(5):2887–2897, 2012.

[31] R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[32] G. Toderici, G. Passalis, S. Zafeiriou, G. Tzimiropoulos, M. Petrou, T. Theoharis, and I.A. Kakadiaris. Bidirectional relighting for 3D-aided 2D face recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2721–2728, San Francisco, CA, June 13-18 2010.

[33] G. Toderici, G. Evangelopoulos, T. Fang, T. Theoharis, and I.A. Kakadiaris. UHDB11 database for 3D-2D face recognition. In *Proc. 6^{th} Pacific-Rim Symposium on Image and Video Technology*, pages 73–86, Guanajuato, Mexico, Oct. 28–Nov. 1 2013.

[34] University of Houston, Computational Biomedicine Lab. UHDB31: 3D face recognition database. URL http://cbl.uh.edu/pages/research/projects.

[35] University of Notre Dame, Computer Vision and Research Laboratory. Biometrics Datasets. URL http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html.

[36] A. Wagner, J. Wright, A. Ganesh, Zihan Zhou, H. Mobahi, and Yi Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372 –386, February 2012.

[37] C. Wang, S. Yan, H. Li, H. Zhang, and M. Li. Automatic, effective, and efficient 3D face reconstruction from arbitrary view image. In *Advances in Multimedia Information Processing - PCM 2004*, pages 553–560. Springer, 2005.

[38] C. Wang, Y. Zeng, L. Simon, I.A. Kakadiaris, D. Samaras, and N. Paragios. Viewpoint invariant 3D landmark model inference from monocular 2D images using higher-order priors. In *Proc. IEEE International Conference on Computer Vision*, pages 319–326, Barcelona, Spain, Nov. 6-13 2011.

[39] X. Wang and R. Yang. Learning 3D shape from a single facial image via non-linear manifold embedding and alignment. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 414 –421, June 13-18 2010.

[40] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, February 2009.

[41] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 532–539, Portland, Oregon, June 25-27 2013.

[42] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):351–363, 2006.

[43] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D.N. Metaxas, and X.S. Zhou. Sparse shape composition: A new framework for shape prior modeling. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1025 –1032, June 21-25 2011.

[44] W. Zhao and R. Chellappa. Symmetric shape from shading using self-ratio image. *International Journal of Computer Vision*, 45(1):55–75, October 2001.