

3D FACE RECOGNITION UNDER PARTIAL OCCLUSIONS USING RADIAL STRINGS

Xun Yu¹, Yongsheng Gao¹, Jun Zhou²

School of Engineering¹, School of Information and Communication Technology²
Griffith University
Nathan, QLD, Australia
{xun.yu, yongsheng.gao, jun.zhou}@griffith.edu.au

ABSTRACT

3D face recognition with partial occlusions is a highly challenging problem. In this paper, we propose a novel radial string representation and matching approach to recognize 3D facial scans in the presence of partial occlusions. Here we encode 3D facial surfaces into an indexed collection of radial strings emanating from the nosetips and Dynamic Programming (DP) is then used to measure the similarity between two radial strings. In order to address the recognition problems with partial occlusions, a partial matching mechanism is established in our approach that effectively eliminates those occluded parts and finds the most discriminative parts during the matching process. Experimental results on the Bosphorus database demonstrate that the proposed approach yields superior performance on partially occluded data.

Index Terms— 3D face recognition, radial string matching, structural recognition, partial occlusions.

1. INTRODUCTION

Face recognition has attracted significant research interest in the past three decades due to its non-intrusive and friendly acquisition nature when compared to other biometrics. Despite extensive research on 2D face recognition, recognizing humans from their faces is still a challenging task under non-cooperative and uncontrolled conditions. In recent years, the research focus has shifted from 2D to 3D domain as the rapid development and dropping cost of 3D sensors. Compared to conventional 2D face recognition, 3D face recognition is expected to be less sensitive to illumination and pose variations. This is because the actual geometric information residing in the 3D data is considered to be inherently robust to changes in illumination and pose which adversely affect 2D face data [1]. However, 3D scans often suffer from the problems of expression and occlusions.

Unlike the problem of expression handling has been extensively studied in recent years, there are only a relatively few studies directly attacking the occlusion problem in the literature. Colombo et al. [2] proposed a

method to detect occlusions by analysing the difference between the input face and its eigenface approximation. By using Gappy PCA [3], the detected occlusion parts were removed and restored. Identification was performed by Fisherface algorithm. In a following up work [4], Colombo et al. introduced an occlusion detection strategy [5] into their face recognition approach. Alyuz et al. [6] used the Average Region Models (ARMs) for 3D face recognition under occlusions. The facial area was manually divided into several meaningful components and the registration of faces was performed by separate dense alignments to relative ARMs. Berretti et al. [7] represented 3D facial scans by a set of facial curves, which were generated by connecting SIFT keypoints. The most saliency curves were selected to perform partial face matching. Drira et al. [8] represented facial surfaces by radial curves emanating from the nose tips. A Riemannian framework was developed to analysis these curves and to measure facial deformations. The representation and the elastic Riemannian framework are robust to handle expressions, pose variations, missing parts and partial occlusions. Lei et al. [9] explored an efficient 3D face recognition method that used a set of local-feature descriptors called KMTS to represent a 3D face, then a TPWCRC framework was proposed to address the problem of 3D partial face recognition. In their method, only one training sample face per person was needed.

In this paper, we have proposed a novel radial string representation and matching approach to recognize 3D facial scans in the presence of partial occlusions. A 3D face is represented by an indexed collection of attributed strings in 2D space. The matching of two 3D faces is performed by the matching of two attributed string sets by a string-to-string matching scheme based on Dynamic programming (DP). The inherent partial matching mechanism can effectively eliminate those occluded parts during the matching process.

The rest of paper is structured as follows: In Section 2, we present the proposed method including radial string representation and the ensemble radial string matching. Section 3 presents the experimental results and comparative analysis on the Bosphorus database. Finally, conclusions are discussed in Section 4.

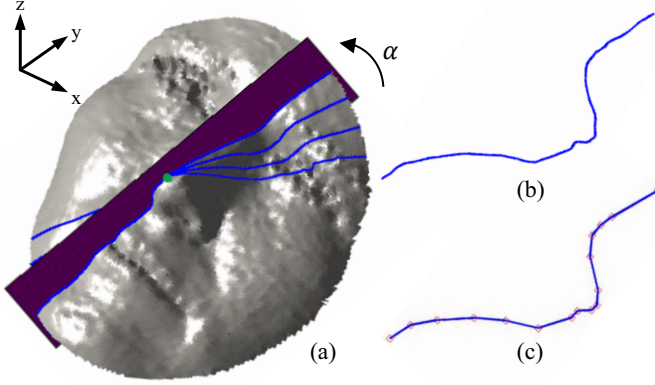


Figure 1. An illustration of the radial string extraction process. (a) Slice facial scan \mathcal{F} by 2D plane \mathcal{P} , (b) a radial curve \mathcal{C} , (c) a radial string $\tilde{\mathcal{S}}$ where red squares are the corner points.

2. PROPOSED METHOD

In this section we provide details of the radial string based face representation as well as the algorithm details of the ensemble radial string matching.

2.1. Radial String Based Facial Representation

In our study, we encode a 3D facial scan into an indexed ensemble of radial strings emanating from the nosetip. The process of extracting radial strings contains two main steps: 1) radial curve extracting, 2) radial curve polygonising. For radial curve extracting, we implement the method similar to [8]. Give a normalized facial range image \mathcal{F} , a radial curve \mathcal{C} is obtained by slicing the facial scan \mathcal{F} by a plane \mathcal{P} that rotates with the respect to the nosetip (chosen as the origin of the facial scan image) and makes an angle α with yz-plane (see Fig. 1(a)). Fig. 1(b) illustrates a radial curve in the 2D plane \mathcal{P} , where we map the radial curve \mathcal{C} in 3D space \mathbb{R}^3 to $\tilde{\mathcal{C}}$ in 2D space \mathbb{R}^2 . For each point p on $\tilde{\mathcal{C}}$, the mapping process is defined as follows:

$$p(x, y, z) \in \mathcal{C} \rightarrow p(\tilde{x}, \tilde{y}) \in \tilde{\mathcal{C}} \quad (1)$$

where $\tilde{x} = \sqrt{x^2 + y^2}$ and $\tilde{y} = z$, which means, in the 2D Cartesian system defining plane \mathcal{P} , the horizontal axis value \tilde{x} is the projection of the radial curves, and the vertical axis value \tilde{y} is the depth information.

After extracting radial curves and mapping them to 2D planes, a corner detection process based on Douglas-Peucker algorithm [10] is then applied to generate radial strings (see Fig. 1(c)) which are formed by connected line segments. The line segment $s(l, \mathbf{n}, \tilde{x}, \tilde{y})$ is the primitive in a radial string $\tilde{\mathcal{S}}$, where l, \mathbf{n}, \tilde{x} and \tilde{y} are the length, directional vector and midpoint location of the line segment, respectively. The directional vector \mathbf{n} is defined as the vector directs to the ending corner point from the starting corner point of the line segment.

2.2. Ensemble Radial String Matching

2.2.1. Edit cost functions

Let $\tilde{\mathcal{S}}^A$ and $\tilde{\mathcal{S}}^B$ be two radial strings of line segment primitives, $\tilde{\mathcal{S}}^A\langle i \rangle$ and $\tilde{\mathcal{S}}^B\langle j \rangle$ be the i th and j th primitives in $\tilde{\mathcal{S}}^A$ and $\tilde{\mathcal{S}}^B$ with attributes $(l_i, \mathbf{n}_i, \tilde{x}_i, \tilde{y}_i)$ and $(l_j, \mathbf{n}_j, \tilde{x}_j, \tilde{y}_j)$ respectively. Define $\tilde{\mathcal{S}}\langle i:j \rangle$ to be the substring of $\tilde{\mathcal{S}}$ from the i th to j th primitives of $\tilde{\mathcal{S}}$.

The cost function for change operation from $\tilde{\mathcal{S}}^A\langle i \rangle$ to $\tilde{\mathcal{S}}^B\langle j \rangle$ is defined as

$$\text{Cost}(\tilde{\mathcal{S}}^A\langle i \rangle \rightarrow \tilde{\mathcal{S}}^B\langle j \rangle) = |l_i - l_j| + f(\Delta(\mathbf{n}_i, \mathbf{n}_j)) + \sqrt{(\tilde{x}_i - \tilde{x}_j)^2 + (\tilde{y}_i - \tilde{y}_j)^2} \quad (2)$$

where $\Delta(\mathbf{n}_i, \mathbf{n}_j) \in [0, 90^\circ]$ is the angle difference between two primitives, and $f(x) = x^2/w$ is a non-linear function to penalize large angle deviation resulted from inter-class difference, but ignore small variation derived from data noise or intra-class difference. w is a weight to be determined experimentally (See Section 3.1).

A merge operation [11] with modified merge cost is used to tackle the problem of adding, missing and shifting of feature points. The cost of merging a sequence of k primitives $\tilde{\mathcal{S}}^A\langle i-k+1:i \rangle = \tilde{\mathcal{S}}^A\langle i-k+1 \rangle \dots \tilde{\mathcal{S}}^A\langle i \rangle$ into one merged primitive $\tilde{\mathcal{S}}_k^A\langle i \rangle$ is defined as

$$\begin{aligned} \text{Cost}(\tilde{\mathcal{S}}^A\langle i-k+1:i \rangle \rightarrow \tilde{\mathcal{S}}_k^A\langle i \rangle) \\ = f\left(\frac{k-1}{l^k} \sum_{q=i-k+1}^i \Delta(\mathbf{n}^k, \mathbf{n}_q) \times l_q\right) \end{aligned} \quad (3)$$

where l^k and \mathbf{n}^k are the length and the directional vector of merged primitive $\tilde{\mathcal{S}}_k^A$, l_q and \mathbf{n}_q are the length and the directional vector of primitive $\tilde{\mathcal{S}}^A\langle q \rangle$ in $\tilde{\mathcal{S}}^A\langle i-k+1:i \rangle$ before merging. So the cost function for a change operation after merge can be rewritten as

$$\begin{aligned} \text{Cost}(\tilde{\mathcal{S}}_k^A\langle i \rangle \rightarrow \tilde{\mathcal{S}}_l^B\langle j \rangle) = |l_i^k - l_j^l| + f(\Delta(\mathbf{n}_i^k, \mathbf{n}_j^l)) \\ + \sqrt{(\tilde{x}_i^k - \tilde{x}_j^l)^2 + (\tilde{y}_i^k - \tilde{y}_j^l)^2} \end{aligned} \quad (4)$$

If $k = 1$ and $l = 1$, no merge is performed and the above change operation reduces to the conventional one-to-one change operation (see Eq. 2).

2.2.2. Radial string matching

The similarity matching between the two radial strings can be characterized by the edit cost functions using Dynamic Programming (DP). Let $\tilde{\mathcal{S}}^A = \tilde{\mathcal{S}}^A\langle 1 \rangle \tilde{\mathcal{S}}^A\langle 2 \rangle \dots \tilde{\mathcal{S}}^A\langle m \rangle$ and $\tilde{\mathcal{S}}^B = \tilde{\mathcal{S}}^B\langle 1 \rangle \tilde{\mathcal{S}}^B\langle 2 \rangle \dots \tilde{\mathcal{S}}^B\langle n \rangle$ be string representation of corresponding radial strings from probe and gallery facial scans. Here, the corresponding radial strings means these two radial strings are generated by the slicing plane of same angle α with yz-plane as described in Section 2.1. A

similarity matrix \mathbf{S} with a size of $(m+1) \times (n+1)$ is generated through the DP, which is defined as follows:

$$\mathbf{S} = \begin{pmatrix} s(0,0) & s(0,1) & \cdots & s(0,n) \\ s(1,0) & s(1,1) & \cdots & s(1,n) \\ \vdots & \vdots & \ddots & \vdots \\ s(m,0) & s(m,1) & \cdots & s(m,n) \end{pmatrix} \quad (5)$$

where each element $s(i, j)$ stores the maximum accumulated similarity value between the strings (substrings) $\tilde{\mathcal{S}}^A\langle 1:i \rangle$ and $\tilde{\mathcal{S}}^B\langle 1:j \rangle$, as defined in Eq. 6.

$$s(i, j) = \max_{k,l} (T(k, l)) \quad (6)$$

where $T(k, l)$ is the accumulated similarity resulting from edit operations as defined below.

$$T(k, l) = \begin{cases} s(i-k, j-l) + S_\Delta & \text{if } S_\Delta \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

To seamlessly include the partial matching mechanism in our approach, the similarity parameter S_Δ in Eq. 7 is defined as

$$S_\Delta = Y - [\text{Cost}(\tilde{\mathcal{S}}_k^A\langle i \rangle \rightarrow \tilde{\mathcal{S}}_l^B\langle j \rangle) + \text{Cost}(\tilde{\mathcal{S}}^A\langle i-k+1:i \rangle \rightarrow \tilde{\mathcal{S}}_k^A\langle i \rangle) + \text{Cost}(\tilde{\mathcal{S}}^B\langle j-l+1:j \rangle \rightarrow \tilde{\mathcal{S}}_l^B\langle j \rangle)] \quad (8)$$

where Y is a threshold to decide whether $\tilde{\mathcal{S}}_k^A\langle i \rangle$ and $\tilde{\mathcal{S}}_l^B\langle j \rangle$ are matched substrings. If the combined cost of merging and changing operations is greater than Y (which gives $S_\Delta < 0$), these primitives are considered as outliers. This is achieved by setting $T(k, l) = 0$ when $S_\Delta < 0$.

The detailed algorithm of creating the \mathbf{S} matrix is given in Algorithm 1, where limit_i^A and limit_j^B are the controlling upper limits on the number of primitives to be merged into a new one in $\tilde{\mathcal{S}}^A$ and $\tilde{\mathcal{S}}^B$, respectively. In our experiments the controlling upper limits are set as 5 to balance the recognition performance and computing cost.

In order to calculate the similarity scores between the corresponding radial strings from the similarity matrix \mathbf{S} . A traceback process is implemented as follows: The pair of substrings with maximum similarity is found by first locating the maximal element in \mathbf{S} . The other matrix elements leading to this maximal value are then sequentially determined with a traceback procedure ending with an element of \mathbf{S} equaling to zero. The pair of substrings with the next best similarity is found by applying the traceback procedure to the second largest element in \mathbf{S} not associated with the first traceback. The similarity of associating a group of substrings from $\tilde{\mathcal{S}}^A$ with a group of substrings from $\tilde{\mathcal{S}}^B$ is computed as

$$s(\tilde{\mathcal{S}}^A, \tilde{\mathcal{S}}^B) = \sum_{i=1}^f S_i \quad (9)$$

where S_i is the similarity (the i th maximal element in matrix \mathbf{S}) of the i th best similar substrings and f is the number of best similar substrings.

The similarity between the probe scan $P = \cup^\alpha \tilde{\mathcal{S}}_\alpha^A$ and gallery scan $G = \cup^\alpha \tilde{\mathcal{S}}_\alpha^B$ is calculated by the fusion of all the similarity values between corresponding radial strings, which is defined as:

$$s(P, G) = \sum_{i=1}^N s(\tilde{\mathcal{S}}_\alpha^A, \tilde{\mathcal{S}}_\alpha^B) \quad (10)$$

where $\alpha = (i/N) * 2\pi$, N is the number of radial strings in each scan. In our experiments, N is equal to 40.

```

Input:  $\tilde{\mathcal{S}}^A = \tilde{\mathcal{S}}^A\langle 1 \rangle \cdots \tilde{\mathcal{S}}^A\langle m \rangle$ ,  $\tilde{\mathcal{S}}^B = \tilde{\mathcal{S}}^B\langle 1 \rangle \cdots \tilde{\mathcal{S}}^B\langle n \rangle$ 
Outputs:  $\mathbf{S}(\tilde{\mathcal{S}}^A, \tilde{\mathcal{S}}^B)$ 
Initialization:  $s(0,0) = 0$ ,  $s(i,0) = 0$ ,  $s(0,j) = 0$ ;
for  $i = 1$  to  $m$ ,  $j = 1$  to  $n$ 
    for  $k = 1$  to  $\min(\text{limit}_i^A, i)$ ,  $l = 1$  to  $\min(\text{limit}_j^B, j)$ 
         $S_\Delta = Y - [\text{Cost}(\tilde{\mathcal{S}}_k^A\langle i \rangle \rightarrow \tilde{\mathcal{S}}_l^B\langle j \rangle) + \text{Cost}(\tilde{\mathcal{S}}^A\langle i-k+1:i \rangle \rightarrow \tilde{\mathcal{S}}_k^A\langle i \rangle) + \text{Cost}(\tilde{\mathcal{S}}^B\langle j-l+1:j \rangle \rightarrow \tilde{\mathcal{S}}_l^B\langle j \rangle)]$ 
         $T(k, l) = \begin{cases} s(i-k, j-l) + S_\Delta & \text{if } S_\Delta \geq 0 \\ 0 & \text{otherwise} \end{cases}$ 
    end
     $s(i, j) = \max_{k,l} (T(k, l))$ ;
end

```

Algorithm 1. Generate Similarity Matrix \mathbf{S} .

3. EXPERIMENTAL RESULTS

In order to evaluate the proposed approach in coping with real partial occlusions, the Bosphorus database [12] is used in our experiment, which includes 4652 scans collected from 105 subjects. The images are in various expressions, poses and partial occlusions. In Fig. 2, four types of occlusions presented in the Bosphorus database are shown: (1) occlusion of mouth with hand, (2) occlusion of eye with hand, (3) occlusion of eye with eyeglasses, and (4) occlusion of the face with hair. All the facial scans are preprocessed in the same manner as in [13]. A template with three manually selected feature points (the nose tip and two inner eye corners) is used to initially align the cropped and smoothed 3D facial scans before implementing a recursive-ICP algorithm [8] to finely align those facial scans.



Figure 2. Scans of one subject in the Bosphorus database with four occlusion types.

3.1. Effect of Parameter w

In this section, we assess the effect of parameter w on the recognition accuracy to choose a value of w which provides a better recognition performance. For the training experiments, 105 subjects with 2 neutral 3D faces per person from the Bosphorus database created a training database. The first neutral scan of each subject was used to construct the gallery set, while the other neutral scans in the second session formed the probe set. The recognition rate is plotted against the values of w in Fig. 3. It can be observed that the recognition rate increases quickly when w ranges from 20 to 60. Then, it reaches the optimal value of 99% when w is 80 and remains stable till 120. Performance degradations are noticed if w continues to increase beyond this range. In the rest of experiments in this study, w is set as 100.

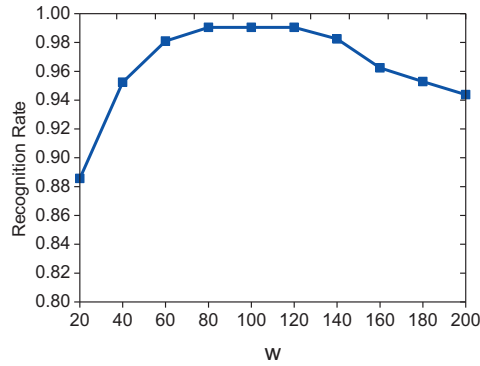


Figure 3. The effect of w on recognition rate.

3.2. Effect of Parameter γ

In this section, we analyse the recognition performance when parameter γ changes. In theory, the parameter γ is a threshold to decide the strength of partial matching. When γ close to 0, except the identical string pairs, all other string pair would be considered as outliers and discarded in the matching mechanism. And as γ increases, the percent of outlier strings will decrease.

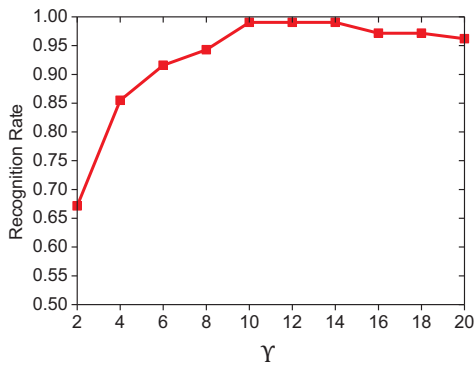


Figure 4. The effect of γ on recognition rate.

The experiment is performed using the same experimental setup as the one discussed in Section 3.1. We observe that, the recognition accuracy remains higher than 95% when w ranges from 10 to 20 (See Fig. 4). The best performance is achieved when γ ranges from 10 to 14. In the following experiments, γ is set to 12.

3.3. Comparative Evaluation on Bosphorus Database

In this section, to compare with the benchmark methods, we follow the same experiment protocol used by Colombo et al. [4]. A neutral scan for each person is selected to form a gallery set of size 105. The probe set contains 360 scans that have occlusions.

The rank-1 recognition rate is reported in Table 1 for different approaches under various types of occlusions. In the four experiments under varying occlusions, the proposed method has significantly outperformed other methods in one type of occlusion, and match or close to the best results in three other cases. The overall recognition rate is 8.7%, 7.1%, and 2.1% higher than the methods from Colombo et al. [4], Drira et al. [8] and Lei et al. [9], respectively. This has confirmed that our method can effectively handle partial occlusions.

Approaches	eye	mouth	glasses	hair	all
Colombo [4]	91.1%	74.7%	94.2%	90.4%	87.6%
Drira [8]	98.9%	78.5%	94.2%	85.2%	89.2%
Lei [9]	90.2%	94.3%	96.2%	98.0%	94.2%
Ours	97.2%	96.1%	94.2%	98.0%	96.3%

Table 1. Comparison of rank-1 recognition rates on the subset of occlusion on the Bosphorus database.

4. CONCLUSIONS

In this paper, we have presented a new 3D recognition approach using a radial string based representation and matching strategy to effectively tackle the challenges of partial occlusions. The proposed method encodes a 3D facial scan into a set of radial strings in 2D space that integrates local structural information with spatial information of a 3D surface. The matching of two 3D facial scans is achieved by matching two radial string sets through string-to-string matching, which is able to effectively find the most similar substrings for each corresponding radial string pair. The performance of the proposed approach has been evaluated and compared with several state-of-the-art approaches. Preliminary experimental results demonstrated the feasibility and effectiveness of our approach, showing a new way for face representation and recognition.

5. REFERENCES

- [1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, pp. 1885-1906, 2007.
- [2] A. Colombo, C. Cusano, and R. Schettini, "Detection and restoration of occlusions for 3D face recognition," *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 1541-1544, 2006.
- [3] R. Everson and L. Sirovich, "Karhunen–Loeve procedure for gappy data," *JOSA A*, vol. 12, pp. 1657-1664, 1995.
- [4] A. Colombo, C. Cusano, and R. Schettini, "Three-dimensional occlusion detection and restoration of partially occluded faces," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 105-119, 2011.
- [5] A. Colombo, C. Cusano, and R. Schettini, "Recognizing faces in 3d images even in presence of occlusions," *Proc. of IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1-6, 2008.
- [6] N. Alyuz, B. Gokberk, and L. Akarun, "A 3D face recognition system for expression and occlusion invariance," *Proc. of IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1-7, 2008.
- [7] S. Berretti, A. Del Bimbo, and P. Pala, "Sparse matching of salient facial curves for recognition of 3-D faces with missing parts," *IEEE Trans. on Information Forensics and Security*, vol. 8, pp. 374-389, 2013.
- [8] 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*, vol. 35, pp. 2270-2283, 2013.
- [9] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, and X. Zhou, "A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample," *Pattern Recognition*, vol. 52, pp. 218-237, 2016.
- [10] M. Visvalingam and J. D. Whyatt, "The Douglas- Peucker Algorithm for Line Simplification: Re- evaluation through Visualization," *Computer Graphics Forum*, vol. 9, pp. 213-225, 1990.
- [11] W.-H. Tsai and S.-S. Yu, "Attributed string matching with merging for shape recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 453-462, 1985.
- [12] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, et al., "Bosphorus database for 3D face analysis," *Biometrics and Identity Management*, pp. 47-56, 2008.
- [13] A. S. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1927-1943, 2007.