

Component-based Registration with Curvature Descriptors for Expression Insensitive 3D Face Recognition

Neşe Alyüz
Boğaziçi University
Computer Engineering Dept.
Istanbul, Turkey
nese.alyuz@boun.edu.tr

Berk Gökberk
Philips Research
Eindhoven, The Netherlands
berk.gokberk@philips.com

Hamdi Dibeklioglu, Lale Akarun
Boğaziçi University
Computer Engineering Dept.
Istanbul, Turkey
{hamdi.dibeklioglu, akarun}@boun.edu.tr

Abstract

Deformations caused by facial expression variations complicate the task of 3D face registration which is vital for successful 3D face recognition systems. In this work, we propose to use a hierarchical component-based face registration technique capable of handling the difficulties caused by non-rigid nature of faces. Local components independently registered by the Iterative Closest Point (ICP) algorithm provides a fast registration with the use of a generic face model and does not suffer from non-rigidity of human facial surface. Invariance of the proposed approach is further increased by utilizing curvature-based 3D surface descriptors. Identification experiments conducted on the multi-expression Bosphorus database reveal that the accuracy of the classical ICP-based approach can be significantly increased under extreme expression variations.

1. Introduction

Three-dimensional (3D) face recognition systems play an important role in high-security biometric applications. Traditional face recognition systems usually employ 2D texture information to infer the identity of a person. It has been shown that under realistic situations where extrinsic and intrinsic factors change, 2D face identification systems offer sub-optimal performances. Extrinsic factors such as illumination conditions and intrinsic factors such as human facial expression variations diminish the discriminative power of a face recognizer. Under such circumstances, intra-class variations usually exceed inter-class variations.

It is, however, possible to overcome majority of these problems with the use of 3D information [1]. Facial shape characteristics do not vary with different illumination conditions, and are better suited to estimate pose angle, for instance. Availability of 3D facial surface information is also useful to analyze the facial deformations caused by expressions. In biometric systems where the aim is to infer the identity, the variations caused by facial expressions should be handled effectively. It has been shown that classical 3D face recognition systems which assume rigid surfaces can only attain mediocre performances. Recently, several schemes were proposed for expression insensitive 3D face identification.

In [3], Cook *et al.* use Log-Gabor Templates (LGT) on range images to deal with expression variations. A range image is divided into multiple regions both in spatial and frequency domains. Each individual region is classified separately and these classifiers are fused at the score level. A facial image is divided into 147 regions and the LGT responses are reduced in dimension by Principal Component Analysis (PCA). With sum rule-based combination, their system achieves 94.63 per cent recognition accuracy on the Face Recognition Grand Challenge (FRGC) v.2 set with a single neutral gallery face set. In [2], Chang *et al.* propose a matching method based on overlapping multiple regions selected around the nose area. Facial surfaces are registered via ICP and similarity measures computed from individual alignments are fused using sum, min or product rules. Faltemier *et al.* [4] present an extended version of this approach by using 38 regions around the nose and combining the scores by the modified Borda count method.

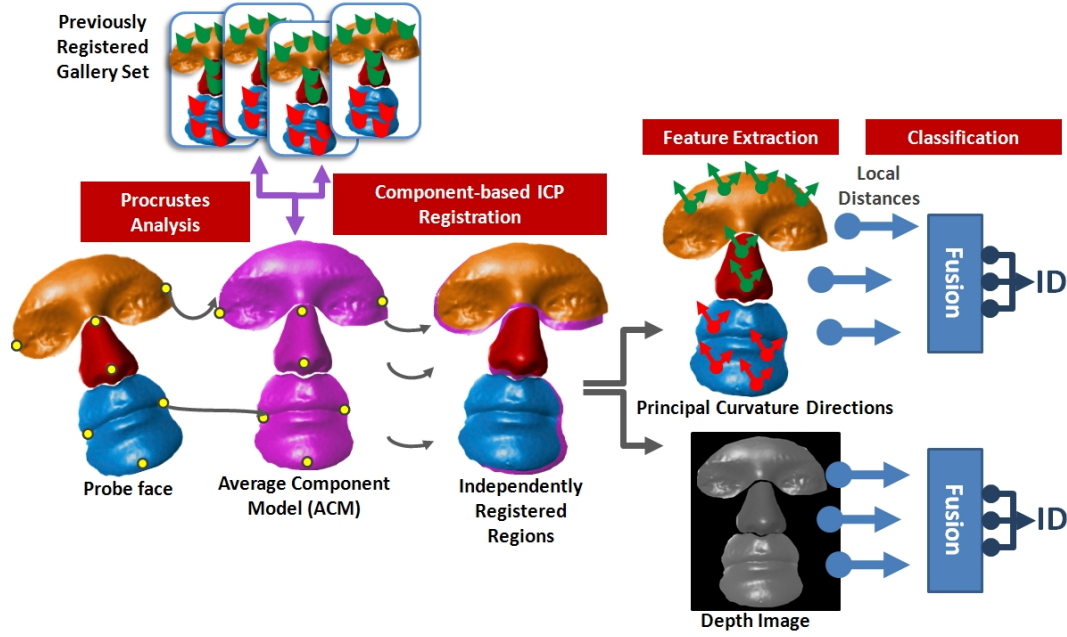


Figure 1. Overall approach.

In [6], Kakadiaris *et al.* fit a deformable facial model to describe a facial surface. 2D geometry images and normal maps are constructed for multiple regions of the facial surface. Both representations of each region are analyzed with a wavelet transform and the classifiers were combined using a weighted sum. Mian *et al.* [7] develop a multi-modal algorithm which combines 2D and 3D. In 3D space the inflection points around the nose tip are automatically extracted and these points are used for the segmentation of the face into eye-forehead and nose regions. These facial regions are considered because they are less affected under expression variations. Regions are separately matched with ICP and the similarity measures are fused at the metric level.

In this paper, we propose a hierarchical piecewise representation of faces where local building blocks, *patches*, are grouped into higher level structures, called *regions*. Registration of faces are carried out at the region level to avoid global misalignments due to deformed parts of a face. This way it is possible to approximate complex deformation-based registration schemes by multiple locally rigid registrations. The ICP algorithm is used as the region-level registration module. Novel contribution of the proposed scheme is the use of invariant 3D shape descriptors together with the component-based registration scheme. Minimal and maximal principal curvature directions were employed as 3D features: these do not change if the surface translates or rotates. Combination of local similarity values by information fusion techniques reveals that each facial component contributes to the overall performance. We show that

the hierarchical component-based facial registration framework, aided by invariant curvature descriptors, significantly increases the identification power of a 3D face recognition system. Comparative analysis of the proposed method is provided on the Bosphorus 3D face database [8] which contains multiple expressions.

2. Face Recognition Methodology

A 3D recognition system is usually composed of the following steps: 1) alignment/registration, 2) 3D feature extraction, and 3) dissimilarity calculation and pattern recognition. The alignment phase transforms faces in such a way that it is later possible to calculate dissimilarities between faces efficiently. Alignment step itself has two phases: coarse and fine alignment. At the coarse alignment stage, several fiducial landmarks are used to transform faces into a common coordinate system. Coarse alignment stage is crucial for the success of the fine alignment/registration step. Iterative techniques used for fine registration usually converge better if the initial conditions are proper. Indeed, the quality of the registration found by the ICP algorithm heavily relies on the initial positions determined by the coarse alignment step. In our work, we use 22 manually determined facial landmark points at the coarse alignment phase. These landmark points are shown in Figure 3(a). Given two landmark sets (l_A and l_B) for face A and B , the best transformation that maps l_A to l_B is found by the Procrustes algorithm, which determines scale, translation and rotation parameters of this mapping.

After coarse alignment, fine registration is handled via the use of the ICP algorithm. The ICP algorithm does not assume registered facial points, as opposed to Procrustes algorithm, and tries to determine best translation and rotation parameters. The output of the ICP algorithm is the set of parameters of this mapping and the point correspondence between two facial surfaces. Classical ICP-based 3D face recognition systems register probe face with all of the gallery faces and pick the gallery image where the point cloud difference of the established correspondence is minimal (called *one-to-all ICP method*). This scheme is slow since it requires N registrations, per probe face, if the gallery set has N samples. We solve this problem by the use of a generic face model. During the training phase, we construct an average face model (AFM) and establish point correspondences between the AFM and all of the gallery set. At the identification phase, a given probe face is just registered to the AFM. Let M_g be the mapping between a gallery face and the AFM, and M_p be the mapping between the probe and the AFM, the final transformation that maps the probe face to the gallery face can be inferred by the combined mapping $M = M_g(M_p)$. This method significantly reduces the time complexity of the identification phase while retaining comparable registration accuracy. We refer to this method as the *AFM-based registration method*.

2.1. Component-based Face Registration

The proposed component-based face registration method operates on local facial parts independently. A facial surface is first divided into patches which are the basic building blocks. A collection of patches forms regions, which are higher level components. As patches, we select facial parts which have salient information such as eyes, nose, central/left/right forehead, upper/lower mouth, upper/lower/middle cheeks, and chin parts. Facial patch division is illustrated in Figure 2. As an example of a region, the upper face region is shown on the right in Figure 2 where it covers left/right eye patches, nose and central forehead patch.

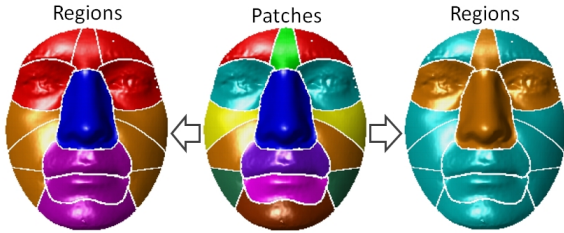


Figure 2. Division of a facial surface into two-level components: central image displays first level components, *patches*, and the left/right images depicts two different alternatives for higher-level components, *regions*, formation.

Patches are manually determined on the AFM. We call the divided AFM as the *Average Component Face Model (ACM)*. Component-based registration starts with the coarse alignment stage, as in the global ICP-based registration. Here, each patch or region to be registered is transformed independently of each other. Given the landmark points of a test face and the ACM, we first apply Procrustes analysis to coarsely align a selected patch over the ACM to the test face. For instance, the nose patch of the ACM is first transformed to the test face using its nose tip and upper nose bridge coordinates. Then, the ACM patch is finely registered to the test facial surface by the ICP algorithm. This process is repeated for each patch defined over the ACM. Each patch produces different transformation parameters. For instance, if the test face has an open mouth expression, the ACM patch responsible for the chin component produces completely different rotation parameters compared to the ACM nose patch. Through this scheme, it is possible to obtain a better registration if faces exhibit significant deformations.

2.2. 3D Shape Descriptors

After the alignment phase, 3D facial surfaces can be compared since they lie on the same coordinate system. A simple method is to use the coordinate differences between two surfaces. This method is an estimate of the volumetric difference between two given surfaces. Given an erroneous alignment, point cloud representation-based similarity calculation may not be optimal. It is therefore necessary to consult better shape descriptors. In our system, we propose to use *normal curvature* descriptors since they measure intrinsic characteristics of a surface, and are invariant to translations and rotations.

Normal curvatures measure the bending degree of a surface. For a two dimensional surface, specifically, a *Monge patch*, s , characterized by a height function $f(u, v)$ defined over a support plane parameterized by (u, v) , the intersection of s with planes defined by two orthogonal vectors in the tangent space produces plane curves. The direction at which the curvature of the plane curve is maximal or minimal determines the principal directions p_1, p_2 . We use an analytical method outlined in [5] to estimate the principal directions. It is based on fitting a quadratic order surface of the form $z = f(x, y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2$ in a neighborhood of the point of interest. The eigenvectors e_1, e_2 of the Weingarten matrix $W = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$ can then be transformed by $p_1 = e_1 \cdot [X_u \ X_v]$ and $p_2 = e_2 \cdot [X_u \ X_v]$ to obtain principal directions p_1, p_2 in $\mathbb{R}^3$. Coefficients A, B and C are estimated by the least-squares technique. Using this method, we represent each point by a (p_1, p_2) pair. Distance between corresponding points can then be calculated by the sum of the angle differences of maximal and minimal

principal direction pairs.

With piecewise registration and feature extraction, each local component produces independent dissimilarity values when compared with a gallery face. We consider these dissimilarity values as scores calculated from different classifiers and use information fusion to compute the combined score. In our experiments, we found out that arithmetic combination via the product rule performs the best. Therefore, we use the product rule when combining each component’s scores. As a classifier, 1-nearest neighbor rule is employed.

3. Experimental Results

3.1. Bosphorus 3D Face Database

In our identification simulations, we used the Bosphorus database which is a multi-expression and multi-pose 3D face database [8]. The richness of expressions makes the Bosphorus database attractive for both expression understanding and identification studies. The Bosphorus database contains two different types of facial expressions: 1) expressions that are based on facial *action units* (AU) of the Facial Action Coding System (FACS) and 2) basic emotional expressions. In the first type, the selected action units are grouped into three sets: i) 20 lower face AUs, ii) five upper face AUs and iii) three AU combinations. In the second type, we consider the following six universal emotions: happiness, surprise, fear, sadness, anger and disgust. Figure 3(b) shows all different types of expressions. To the best of our knowledge, this is the first database where ground-truthed action units are available. In order to achieve more natural looking expressions, professional actors and actresses were scanned.

Images in the Bosphorus database are acquired by the Inspeck Mega Capturor II 3D sensor which has about $x = 0.3mm$, $y = 0.3mm$ and $z = 0.4mm$ sensitivity. A typical pre-processed scan consists of approximately 35K points. The locations of several fiducial points are determined manually (see Figure 3(a)). The database contains 3396 facial scans of 81 subjects, 51 men and 30 women. Majority of the subjects are Caucasian and aged between 25 and 35. The Bosphorus database has two parts: v.1 and v.2. In our simulations we used the v.2 set which has more expression variations. In Bosphorus v.2, there are 47 subjects having approximately 34 scans for different expressions. 30 of these 47 subjects are professional actors/actresses. For each subject, a single neutral image is used as gallery and the other scans are put into the probe set. In total, there are 47 gallery images and 1507 probe images.

3.2. Recognition Results

In our experiments, we choose two baseline systems for comparative analysis of the proposed approach. The first

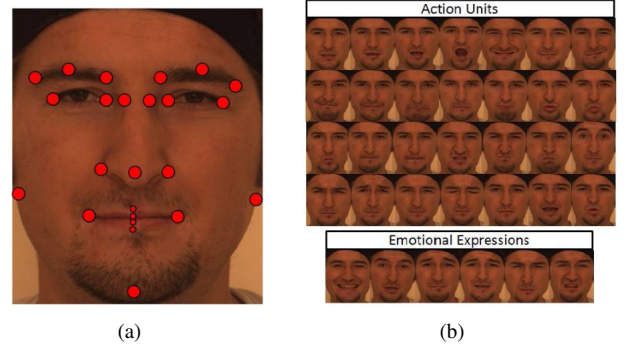


Figure 3. a) Manually located landmark points and b) expressions for the Bosphorus database.

baseline system is One-to-All ICP which registers the probe image with all gallery images and selects the identity of the one which produces the smallest ICP alignment error. One-to-All ICP method uses the whole facial surface and employs 3D point sets as features. The second baseline system is based on the AFM-based registration of faces which is a fast variant of the One-to-All ICP method. The identification accuracies of the baseline methods on the Bosphorus face database are shown in Table 1. One-to-All ICP baseline obtains 72.19% correct classification accuracy by misclassifying 419 images out of 1507 probes. The AFM-based ICP method performs slightly better by attaining 75.45% accuracy. These results show that both baseline systems have a mediocre performance on the expression-variant Bosphorus database. In addition, storing a single neutral face as the enrollment template makes the identification experiments challenging. However, forcing a difficult experimental protocol is instrumental in measuring the relative performances of the compared methods.

The effect of incorporating 3D curvature descriptors is next analyzed. To design a fair benchmarking for the comparison of feature sets, we modify the AFM-based ICP approach by using principal directions as point features after global alignment. This method is referred to as G_{curv} , emphasizing *global* registration with *curvature* features. Note that except for the One-to-All ICP method, we always use the AFM-based registration. Principal curvature directions increase the identification rate significantly compared to the AFM-based ICP algorithm by getting 90.78% accuracy (see Table 1, fourth row). This gain validates the motivation of using surface intrinsic features for 3D face recognition. Since non-rigid deformations cause sub-optimal ICP alignments, point set differences around the misaligned facial parts usually lead to greater intra-class variations. However, principal directions are more robust to such situations. To analyze the sensitivity of point set approach, we calculated the difference maps of the point set after the AFM-based alignments. Figure 4 displays difference maps obtained by

Registration	3D Feature	Region of Interest	Method Name	Accuracy
Global	Point Set	Holistic	<i>Baseline #1: One-to-All ICP</i>	72.19
Global	Point Set	Holistic	<i>Baseline #2: AFM-based ICP</i>	75.45
Global	Curvature	Holistic	G_{curv}	90.78
Global	Point Set	Best Patch Set	G_{pset}^{best}	87.19
Global	Curvature	Best Patch Set	G_{curv}^{best}	93.17
Component	Point Set	Eye, nose, cheek, chin	C_{pset}	96.02
Component	Point Set	Upperface (eye, nose, central forehead)	$C_{pset}^{upperface}$	92.10
Component	Curvature	Upperface (eye, nose, central forehead)	$C_{curv}^{upperface}$	97.28

Table 1. Rank-1 recognition accuracies of 3D face recognizers.

computing the distances between probe faces of subject #1 and the gallery face. Darker regions demonstrate the facial regions having greater dissimilarities. It is clear from these figures that after global registration, point set differences cause large variations especially around non-rigid parts of the faces such as mouth and cheeks.

In order to further verify this observation and reduce the intra-class variations under expression changes, we decided to locate the most discriminative facial parts after global ICP-based alignment. More formally, we formulate our aim as finding the best subset of all the patches around the whole facial surface that gives the best identification accuracy. As shown in Figure 2, there are 15 patches. We have carried out an exhaustive search of all possible combinations of these surface patches. The performances of the best patch subset methods for both point set and curvature descriptors are given in Table 1, referred to as G_{pset}^{best} and G_{curv}^{best} , respectively. Both methods produce the best identification accuracies when the eyes, nose and the central forehead regions are selected. Using only these regions, G_{curv}^{best} improves the G_{curv} method by 2.39 per cent, obtaining 93.17% accuracy. As expected, the improvement for the point set method is significantly higher: G_{pset}^{best} method correctly classifies 87.16% of the probe images. Compared to G_{pset} , the improvement is 11.74%. However, even in the case of using the best subset of local patches, principal directions perform better than point sets.

The experimental findings mentioned so far support the benefits of using local components. Therefore, we now can proceed to the comparative analysis of component-based registration techniques. As a starting scheme for component-based registration, we have formed four regions by combining neighboring patches around the eye/forehead, nose, cheek, and chin regions. These four regions can be seen in Figure 2. As explained in Section 2.1, each of these regions are located on the probe face and are registered to the ACM independently. Faces in the gallery set were already registered to the ACM prior to the identification phase. Therefore, after component-based registration, we can estimate the best alignment between each region for

Region	Eye	Nose	Cheek	Chin	Fusion, C_{pset}
Acc(%)	88.25	85.93	52.16	35.3	96.12

Table 2. Local region performances for the point set approach and their fusion by product rule.

each gallery-probe face pair through ACM. Given the four point set-based alignment errors for each region, we compute the final dissimilarity by multiplying individual dissimilarity scores. The classification rate of this method, C_{pset} , is 96.02% which is notably better than even the best point set-based subset selection after the global registration (see the seventh row of Table 1). This finding demonstrates the utility of the component-based registration scheme under extreme expression variations. Table 2 shows individual identification powers of eye/forehead, nose, cheek and chin parts used in the C_{pset} approach. It is obvious from Table 2 that combining the decisions of individual parts is very beneficial. It is also important to point out that exclusion of any region in the fusion step causes performance degradations. Therefore, it is vital to use each region's score in the computation of the final dissimilarity score regardless of whether they perform sub-optimally or not. Component-based registration handles these parts robustly and provides complementary information in the fusion setting.

Our previous investigation on finding the best discriminative parts under global registration pointed out the importance of eye, nose and central forehead regions. Therefore, we decided to consider these parts as a single region in the component-based registration framework. Using the combination of these parts and discarding other regions, we have performed alignment between face pairs. This scheme can be considered as a single component-based registration, as opposed to independently registering local components such as eye and nose regions. We call this method as upper face-based component registration. Again, we employ both point set and curvature-based shape descriptors in dissimilarity calculation. The last two rows of Table 1 shows the identification power of these two methods, $C_{pset}^{upperface}$ and $C_{curv}^{upperface}$, for point set and principal directions, respec-

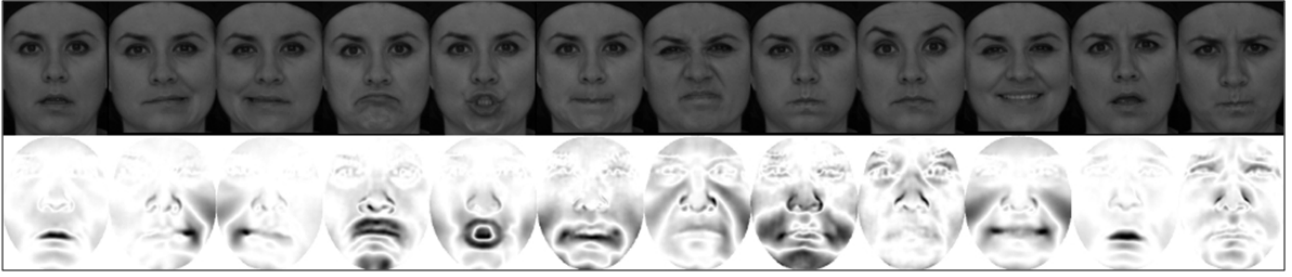


Figure 4. Dissimilarity maps computed for globally registered faces using the point set features. Each image is constructed by subtracting the image from that subject's neutral gallery image. Darker regions illustrate greater dissimilarity.

tively. With the use of these techniques, we achieve the best results so far: $C_{\text{curv}}^{\text{upperface}}$ method correctly identifies 97.28% of the probe faces by misclassifying only 41 images. Figure 5 shows several faces misclassified by the $C_{\text{curv}}^{\text{upperface}}$ algorithm. Out of the 41 misclassified faces, 17 have a nose wring that causes curvature variations around the upper nose part. $C_{\text{pset}}^{\text{upperface}}$ is also superior to its global counterpart, AFM-based ICP, and attains 92.10 per cent accuracy. Overall, we see that component-based registration of faces equipped with the maximal/minimal curvature direction features delivers the best identification performances under significant expression variations.



Figure 5. Faces misclassified by the $C_{\text{curv}}^{\text{upperface}}$ algorithm.

4. Conclusion

In this paper, we show that hierarchical subdivision of facial surfaces enables accurate registration and thus increased identification performance. Under extreme expression changes, multiple locally rigid registrations handle surface deformations efficiently. Additionally, incorporation of coordinate system independent 3D shape descriptors, such as principal curvature directions, leads to better discriminative information around individual facial regions. We show that it is possible to achieve 97.28% rank-1 correct classification rate by focusing on a region comprised of eye, nose, and central forehead parts. Considering the performance of the classical global ICP techniques (72%-75%), the improvement is substantial. In addition to the improvement due to the component-based registration and analysis,

the improvement due to using curvature-based features is also considerable. Using the same regions, point set performance (92.10%) increases to 97.28% when curvature direction features are utilized. As future work, we plan to study the robustness of our approach if the facial landmarks are found automatically.

References

- [1] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, 101:1–15, 2006.
- [2] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 157–164, 2005.
- [3] J. Cook, V. Chandran, and C. Fookes. 3D face recognition using Log-Gabor templates. In *British Machine Vision Conference*, pages 83–92, 2006.
- [4] T. Faltemier, K. W. Bowyer, and P. J. Flynn. A region ensemble for 3D face recognition. *IEEE Transactions on Information Forensics and Security*, 3(1):62–73, 2007.
- [5] J. Goldfeather and V. Interrante. A novel cubic-order algorithm for approximating principal direction vectors. *ACM Trans. Graph.*, 23(1):45–63, 2004.
- [6] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murduza, 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.
- [7] A. S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.
- [8] A. Savran, O. Çeliktutan, A. Akyol, J. Trojanova, H. Dibeklioğlu, S. Esenlik, N. Bozkurt, C. Demirkır, E. Akagündüz, K. Çalıskan, N. Alyüz, B. Sankur, İ. Ulusoy, L. Akarun, and T. M. Sezgin. 3D face recognition performance under adversarial conditions. In *Proc. eINTERFACE07 Workshop on Multimodal Interfaces*, 2007.