Adaptive Registration for Occlusion Robust 3D Face Recognition

Nese Alyuz[1], Berk Gokberk[2], and Lale Akarun[1]

[1] Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey,
{nese.alyuz,akarun}@boun.edu.tr
[2] Department of Electrical Engineering Mathematics and Computer Science,
University of Twente, The Netherlands,
b.gokberk@utwente.nl

Abstract. Occlusions over facial surfaces cause performance degradation for face registration and recognition systems. In this work, we propose an occlusion-resistant three-dimensional face registration method. First, the nose area is detected on a probe face using curvedness-weighted convex shape index map. Then, probable eye and mouth patches are detected and checked for validity. An adaptive model is constructed by selecting valid patches of the average face model. Finally, registration is handled with the Iterative Closest Point algorithm, where the adaptive model is used as the reference. The UMB-DB face database is used to evaluate the registration system: The nose detector has 100% and 93.90% accuracy, for the non-occluded and occluded images, respectively. A simple global depth-based recognition experiment is done to evaluate the registration performance: Our adaptive model-based registration scheme improves rank-1 recognition rate by 16%, when compared with the nose-based alignment approach.

Keywords: 3D face registration, regional face registration, face registration under occlusion.

1 Introduction

Face is a preferred biometric, due to its contactless acquisition and applicability to non-cooperative scenarios. Recent studies have shown that in the three-dimensional (3D) domain; challenges such as illumination and pose can be better handled. However, dealing with extreme occlusion variations remains a challenging task. When occlusions are present, 3D face registration algorithms fail to provide accurate facial point correspondences due to occluding surface points. The resulting alignment between facial surfaces is usually incorrect, leading to low recognition rates.

There are only a few studies dealing with the occluded 3D face recognition problem. In [1], a face recognition system composed of occlusion detection and restoration stages is proposed. However, experiments are conducted on synthetic occlusions. Initially, the non-occluded facial surfaces are registered using manually annotated landmark points, and then the artificial occlusions are added. In [2], a 3D face detection algorithm is proposed to deal with partially occluded faces where inner eye corners and the nose tip are detected based on curvature information. Then, a set of candidate triplets are

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part III, LNCS 7585, pp. 557–566, 2012.

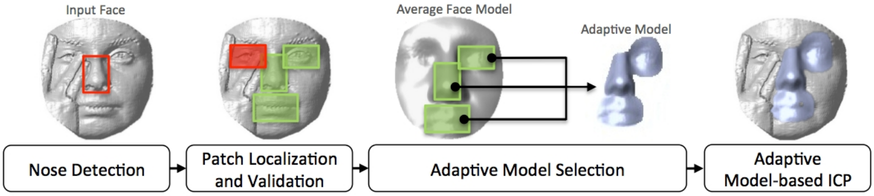

Fig. 1. Diagram of the proposed registration method

formed and registered to the average face using a variant of the ICP. In [3], an occlusion invariant 3D face recognizer is proposed, which employs a nose-based face registration. The nose region is first localized using curvature information and faces are then aligned to a generic nose model. Both in [2] and [3], registration is only dependent on the nasal area, which can be insufficient for a fine alignment. There are other studies regarding partial or regional matching of 3D surfaces: In [4], a partial ICP approach is proposed, in which a subset of nearest point pairs are utilized to calculate the alignment. In [5], a two phase registration scheme is implemented, where the faces are initially registered to a whole face model, and subsequently separate regional registrations are obtained. In [6], a similar regional approach is employed, where a large set of regional alignments are performed. In [7], a semi-rigid region composed of forehead and nose area is utilized for alignment. Although in [4–7] the aim is to improve the alignment, these methods are developed to deal with surface deformations caused by expression variations. Hence, they are not applicable to occlusion variations.

In this work, we propose an occlusion invariant 3D facial registration method. We handle registration by an adaptive model-based approach which assumes partial visibility of the nose. Prior to registration, nose detection is employed and is used to locate eye and mouth patches. Detected patches are then evaluated for their validity. The corresponding valid (occlusion-free) patches of the average face model are selected to construct an adaptive face model. ICP alignment with the adaptive model is able to discard the occluded surface points for point matching. Experiments on the UMB-DB [8] database, show that the adaptive registration attains better registration and identification accuracy under occlusion variations when compared to the nose-based scheme of [3].

2 Proposed Registration System

The proposed face registration system has three phases: (1) nose detection via curvature maps, providing an initialization for fine registration; (2) facial patch localization and validation to form an adaptive face model; (3) model based fine registration via ICP. The overall diagram of the system is given in Figure 1. Details about each phase are given in the following subsections.

2.1 Nose Detection

For rigid alignment of 3D surfaces, Iterative Closest Point (ICP) algorithm [9] is a widely used method. However, like many of the other iterative approaches, performance of ICP relies greatly on the initial conditions. Therefore, an initial alignment should be provided, which will be improved in further iterations. For the surface initialization, most of the 3D face recognition systems depend on accurate localization of facial landmark points [10], [11], [12]. However, when occlusions are present over the facial surface, localization of fiducial points fails. Since facial occlusions may occur over the nose area, our nose detector assumes partial visibility of the complete nose structure with the help of local nasal surface sub-patches (See Section 2.2 for further details).

The nose detection algorithm [3] utilizes surface curvature information, which provides an advantage due to its rotation and translation invariance. Two curvature maps are computed for a given surface, namely the shape index map and the curvedness map. These measures of the local surface, was introduced in [13], computed using the maximum (κ_{max}) and the minimum (κ_{min}) curvatures. The transformation separates components that are dependent or independent of scale [14]. Scale-independent components, such as shape index, provide the distinction between spherical and cylindrical surfaces. On the other hand, the scale-dependent components, such as curvedness, give the magnitude of the curvature. The shape index value $SI(i)$ at surface point i can be computed from κ_{max} and κ_{min}:

$$SI(i) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{\kappa_{max}(i) + \kappa_{min}(i)}{\kappa_{max}(i) - \kappa_{min}(i)} \tag{1}$$

The shape index map $\boldsymbol{SI}$ takes values in $[0, 1]$ and provides a smooth transition between concave ($0 < SI(i) < 0.5$) and convex ($0.5 < SI(i) < 1$) shapes. As the scale-dependent counterpart of shape index, curvedness measures the rate of curvature at each point:

$$C(i) = \sqrt{\frac{\kappa_{min}(i)^2 + \kappa_{max}(i)^2}{2}}. \tag{2}$$

A planar surface will have a curvedness of zero, whereas a non-planar surface will have a curvedness value proportional to its rate of curvature. The nose detector first constructs shape index and curvedness maps. Since nose is a convex structure, the $\boldsymbol{SI}$ map is thresholded (by 0.5) to eliminate concave regions. The convex $\boldsymbol{SI}$ map, denoted as $\boldsymbol{SI}_{cx}$, is defined as

$$SI_{cx}(i) = \begin{cases} 0 & \text{if } SI(i) < 0.5 \\ SI(i) & \text{otherwise.} \end{cases} \tag{3}$$

After concave regions are eliminated, $\boldsymbol{SI}_{cx}$ is weighted with curvedness [15] to integrate scale-dependent and scale-independent components:

$$WSI(i) = SI_{cx}(i) * C(i) \tag{4}$$

Here, $\boldsymbol{WSI}$ denotes the curvedness-weighted convex shape index. In Figure 2, the maps constructed at each step are given for an example facial image. The maps illustrated are: $\boldsymbol{SI}$, $\boldsymbol{SI}_{cx}$, $\boldsymbol{C}$, and $\boldsymbol{WSI}$.

As illustrated in Figure 2, the nose region appears as a distinct fork-shaped structure in the $\boldsymbol{WSI}$ map. To locate the nose area, template matching is employed. For the construction of the nose template, an average face model is created by the Thin Plate Spline

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

Fig. 2. Curvature maps utilized for nose detection are illustrated on an example image: (a) depth image, (b) shape index, (c) convex shape index, (d) curvedness, and (e) weighted convex shape index.

warping, where a set of registered non-occluded neutral training images are used [16]. Then, the average nose model is obtained by manually cropping the face model. The $\boldsymbol{WSI}$ map for the nose model is constructed to serve as the nose template. Given a test image, template matching is performed by normalized cross-correlation, and the region which mostly resembles the nose structure is located.

2.2 Patch Selection and Adaptive Registration

In [3], only local nose regions were considered for occlusion invariant registration. After nose detection, the probe surface was registered using an average nose-region model. However, this approach has shortcomings. Relying solely on the nasal region for the overall face alignment might be suboptimal; especially if the borders of the nose region are affected by occlusions. Additionally, any problems on the nose surface structure, either due to acquisition errors or uncommon nose shapes, may lead to inaccurate facial surface registration. Here, we propose to utilize an adaptive face model. The idea is to adaptively detect and include other non-occluded facial regions such as eyes and mouth automatically to form an adaptive face model for registration. For instance, if the left side of a face is occluded by a hand (See Figure 1), our adaptive face model will automatically be constructed using the non-occluded regions such as right eye, mouth and nose. Then, combined regional models are used for alignment estimation instead of using only the nasal region.

Using the detected location of the nose area, we find other patch locations. In Figure 3, the patch division scheme is shown on the first image. However, not all of the facial patches are beneficial for registration. Therefore, we use a subset of these patches. The patches we use are: nose, left/right eye, and mouth. We also have sub-patches such as left/right nose halves, upper/lower nose halves. Hierarchical division of patches into sub-patches enables us to discard regions where occlusion artifacts are present. To construct average patch models, initially an average face model is generated using the method of [16]. Afterwards, for each patch, an average patch model is constructed by cropping the average face model. From each model, the $\boldsymbol{WSI}$ map is computed to define the patch template. Using these templates, corresponding patch regions on a given face are detected via template matching based on normalized cross correlation. To limit the search space for the localization of each patch, we compute the probable patch center of a probe face using the relative displacements vectors between patch centers of the average face model. Additionally, a predefined bounding box around each patch center is utilized. Due to occlusions over the face, some patches will not be visible and cannot

Fig. 3. Facial patches and the adaptive models utilized in registration are given. The first image shows the division scheme utilized for patch construction. To construct the adaptive models, combination of nose, eye, and mouth patches are considered.

be located correctly. Therefore, in order to determine the validity of each patch, thresholding is applied on template matching scores. The thresholds used for patch validity are calculated from patches of a separate non-occluded neutral database, namely the neutral subset of the FRGC v.2 [17]. The probe patches that have dissimilarity scores below the threshold define the valid parts. Here, the patch localization and validation steps are not used to detect patches of the probe face to be used in registration. The validity information of patches are only used for the model selection: The respective valid patches are selected from the average face model to constitute the adaptive patch-based model for the respective probe face. In Figure 3, the 17 adaptive models utilized in the registration process are shown (the first image was included to show patch division scheme). After adaptive model construction, the whole probe surface is aligned to the adaptive model via ICP, where ICP estimates the alignment parameters using only the non-occluded regions. Hence, the overall registration approach becomes insensitive to occlusions.

3 Experimental Results

3.1 Databases

In our experiments, we have used two face databases: (1) The FRGCv2 [17], including non-occluded acquisitions; (2) The UMB-DB 3D database [8], including expression and large range of occlusion variations. The FRGCv2 is used for the construction of the average face and patch models, and for the determination of threshold values used for validity check over template matching scores. This database consists of 4007 scans and we use only the neutral faces (2365 scans of 466 subjects). The UMB-DB database is acquired from 142 subjects with a total of 1473 scans. The non-occluded acquisitions include four facial expressions. The number of non-occluded scans is 883, and 441 of

Fig. 4. Sample faces from the UMB-DB

these are neutral. The remaining 590 scans constitute the challenging occlusion subset. The occlusions can be caused by hair, eyeglasses, hands, hats, scarves, and other objects (see Figure 4 for images taken from [8] and Table 1 second column for the number of scan variations). For the occluded faces, ground truth occlusion masks are also available.

3.2 Nose Detection Accuracy

We have located nose regions automatically in the whole UMB-DB database and inspected the results visually. In Table 1, the number of correct detections are given in the third column. The nose regions in the non-occluded scans are successfully located. In the occluded subset, the nose detector obtains high performance with a detection performance of 93.90%. Our nose detector is quite robust to occlusions: In 245 of the 590 occluded images, the nose area was occluded and for 75% of the incorrectly detected 36 images, the nose area was more than 50% occluded. Even for the scarf occlusions, where the nose area is not visible, the detection rate is quite high (92.72%). In Figure 5, some detection examples of challenging occlusions are given. Our detection rates are similar to the face detection results provided in [8] (553 detections out of 590). We also applied our detection algorithm on the FRGCv2, where the nose areas are detected with 100% accuracy (by visual inspection) both on the neutral subset (2365 scans) and the non-neutral subset (1642 scans). As stated earlier, it is sufficient for us to detect the nose area approximately since the subsequent ICP registration handles fine registration using patch-based adaptive models.

Fig. 5. Correct (first row) and incorrect (second row) nose detections.

Table 1. Nose detection performances on the UMB-DB database

Acquisition Type	Sample Count	Detected Noses (Detection Rates)
Neutral (Gallery)	142	142 (100.00%)
Neutral (Probe)	299	299 (100.00%)
Non-neutral	442	442 (100.00%)
Occlusion	590	554 (93.90%)
Occlusion Type		
Scarf	151	140 (92.72%)
Glasses	75	74 (98.67%)
Hair	33	27 (81.82%)
Hand	165	152 (92.12%)
Hat	183	181 (98.91%)
Other	38	33 (86.84%)

3.3 Patch Validation and Selection Accuracy

After nose detection, the patches of a probe face are estimated and checked for validity. Using validated patches, the corresponding model of the probe face is constructed adaptively. The thresholds used for patch validation are determined from the template matching scores of the FRGCv2 neutral subset. These thresholds are used to set patch validity flags of the UMB-DB scans. When the model selection results are analyzed, it is seen that in 77 out of 590 occlusion scans, the model selection is erroneous. In 36 of the 77 errors, the nose detection prior to patch validation fails. In the other 41 scans, some of the patches are selected incorrectly: These errors appear mostly in the mouth area.

3.4 Registration Accuracy

To evaluate the registration performance, we have constructed a simple recognition experiment. As facial features, we have used depth images: The depth images are obtained by resampling the surfaces from a regular grid, which enables sufficient computation of distances using only the z coordinates. To evaluate the performance of registration, ground truth masks are employed, where non-occluded parts are annotated for the UMB-DB occlusion variations. Using these masks, the occluding parts are discarded from the registered depth images. Then, a depth-based classifier is employed: the averaged l_1-norm between occlusion-removed probe and gallery depth images are computed and 1-nearest neighbor classification is performed. The identification experiment is conducted with three different registration approaches: (1) global face model-based ICP, as a baseline approach; (2) nose model-based ICP, which was previously used in [3]; and (3) the proposed adaptive model-based registration. In Table 2, recognition rates for the UMB-DB database are given in the columns labeled by “manual occlusion removal”. The gallery set contains first neutral scan of each of the 142 subjects.

When the identification results (with manual occlusion removal) in Table 2 are analyzed, it is clear that using a larger model is beneficial for the non-occluded scans: For the neutral and non-neutral subsets, best performances are obtained when the whole face model is utilized. The adaptive model based registration has comparable results

Table 2. Identification performances on the UMB-DB database. The reported baseline identification accuracy for the occlusion subset of this database is 56.50% [8].

	Recognition Rates(%)			
Acquisition	Manual Occlusion Removal			Automatic Identification
Type	*Face Model*	*Nose Model*	*Adaptive Model*	*Adaptive Model*
Neutral (Probe)	98.66	88.63	97.32	96.99
Non-neutral	71.49	67.87	70.14	82.35
Occlusion	44.58	48.98	65.08	67.63
Occlusion Type				
Scarf	21.85	28.48	41.72	41.72
Glasses	84.00	64.00	80.00	88.00
Hair	54.55	48.48	66.67	66.67
Hand	16.36	32.12	58.79	63.64
Hat	63.40	73.77	83.61	81.42
Other	28.95	52.63	60.53	68.64

with the facial model, even though the adaptive model has at least 47.7% fewer surface points. This shows that the patch regions provide sufficient information for registration. It is also clear that the adaptive model based registration is superior to the face or nose model based ICP, when faces have occlusions. The face model based registration is not successful on occluded faces (44.58%). By comparing classification results, we clearly see the advantage of the adaptive model (65.08%) over the nose model (48.98%). Furthermore, analysis of performances for different occlusion types are included in Table 2. In most of the scarf occlusions, the lower half of the face including the nose area is occluded. Therefore using a face or a nose model cannot provide acceptable registration. However, for the adaptive approach, the valid eye patches are used and the identification rate is improved. In the hair, hand, and hat occlusions, the adaptive model is always better than face and nose model registrations. In comparison, the nose model covers a much smaller area, and is less prone to occlusions. However, even a small portion of an occlusion appearing in the nasal area will affect the final registration significantly. When valid eye and mouth regions are included in the model, alignment disruptions will be corrected. For the eyeglasses case, the registration scheme depending on a face model is slightly better than the adaptive method since glasses can sometimes invalidate the eye regions.

It should be stressed that depth-based identification performances with manually removed occlusions are only provided to indicate the relative standing of the registration approaches. A recognition approach based on a more advanced representation method is expected to give better recognition performance. We are continuing our studies to develop such an approach. However, we have obtained preliminary occlusion-invariant automatic classification results, where the adaptive model of the probe face is used to define the validity mask for classification. The respective valid points on the probe and gallery faces are then used to compute dissimilarity values. This system is automatic, since no manually labeled occlusion masks are considered. The identification rates of the fully automatic system are given in the last column of Table 2. It is clear that, for the neutral scans, it is beneficial to use the whole face. However, for non-neutral and occlusion scans, automatically defining valid regions and using them at the classification phase by the adaptive model is beneficial: It achieves even better identification rates

(67.32%) than by using manually removed occlusions (65.08%). Our automatic identification results are also better than the results presented in [8] where the PCA-based classifier attains 56.50% identification rate on restored faces after occlusion removal.

4 Conclusion

In this work, we have proposed a 3D face registration system which is robust to occlusions: For the experiments, we have used the challenging UMB-DB, which is reported to have a baseline identification accuracy of 56.50% [8]. Our experiments show that the adaptive model based registration is beneficial for occluded faces: Noses on the non-occluded scans can be detected with 100% accuracy, whereas for the occluded scans, the performance of the nose detector is still very high (93.90%). With an identification experiment, we have shown that, under extreme occlusions, face and nose model-based registrations fail. The proposed scheme, on the other hand, is able to cope with occlusions: The depth-based classifier on occlusion-removed faces shows an improvement of 16%: from 48.98% (nose model) to 65.08% (adaptive model). The preliminary automatic identification results show that it is beneficial to use the adaptive model regions for classification.

References

1. Colombo, A., Cusano, C., Schettini, R.: Three-dimensional occlusion detection and restoration of partially occluded faces. Journal of Mathematical Imaging and Vision 40, 105–119 (2011)
2. Colombo, A., Cusano, C., Schettini, R.: Gappy PCA Classification for Occlusion Tolerant 3D Face Detection. Journal of Mathematical Imaging and Vision 35, 193–207 (2009)
3. Alyuz, N., Gokberk, B., Spreeuwers, L., Veldhuis, R., Akarun, L.: Robust 3D Face Recognition in the Presence of Realistic Occlusions. In: International Conference on Biometrics (2012)
4. Wang, Y., Pan, G., Wu, Z., Wang, Y.: Exploring Facial Expression Effects in 3D Face Recognition Using Partial ICP. In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006, Part I. LNCS, vol. 3851, pp. 581–590. Springer, Heidelberg (2006)
5. Alyuz, N., Gokberk, B., Akarun, L.: A 3D face recognition system for expression and occlusion invariance. In: International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–7 (2008)
6. Faltemier, T., Bowyer, K., Flynn, P.: A region ensemble for 3-D face recognition. IEEE Trans. on Information Forensics and Security 3, 62–73 (2008)
7. Al-Osaimi, F., Bennamoun, M., Mian, A.: An expression deformation approach to non-rigid 3D face recognition. International Journal of Computer Vision 81, 302–316 (2009)
8. Colombo, A., Cusano, C., Schettini, R.: UMB-DB: A Database of Partially Occluded 3D Faces. In: ICCV Workshops, pp. 2113–2119 (2011)
9. Besl, P.J., McKay, H.D.: A method for registration of 3D shapes. IEEE Trans. PAMI 14, 239–256 (1992)
10. Lu, X., Jain, A., Colbry, D.: Matching 2.5D face scans to 3D models. IEEE Trans. PAMI 28, 31–43 (2006)
11. Chang, K., Bowyer, W., Flynn, P.: Multiple nose region matching for 3D face recognition under varying facial expression. IEEE Trans. PAMI 28, 1695–1700 (2006)

12. Colombo, A., Cusano, C., Schettini, R.: 3D face detection using curvature analysis. Pattern Recogn 39, 444–455 (2006)
13. Koenderink, J., van Doorn, A.: Surface shape and curvature scales. Image and Vision Computing 10, 557–564 (1992)
14. Tittle, J., Perotti, V.: The perception of shape and curvedness from binocular stereopsis and structure from motion. Attention, Perception, & Psychophysics 59, 1167–1179 (1997)
15. Lo, T., Siebert, J.: Sift keypoint descriptors for range image analysis. In: Annals of the BMVA X, pp. 1–18 (2009)
16. Salah, A.A., Alyuz, N., Akarun, L.: Registration of 3D face scans with average face models. Journal of Electronic Imaging 17 (2008)
17. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: CVPR, vol. 1, pp. 947–954 (2005)