Impact Analysis of Nose Alterations on 2D and 3D Face Recognition

Nesli Erdogmus #1, Neslihan Kose #2, Jean-Luc Dugelay #3

Multimedia Department, EURECOM
2229 Route des Cretes,
06560 Sophia-Antipolis, France
1 nesli.erdogmus@eurecom.fr
2 neslihan.kose@eurecom.fr
3 jean-luc.dugelay@eurecom.fr

***Abstract*—Numerous major challenges in face recognition, such as pose, illumination, expression and aging, have been investigated extensively. All those variations modify the texture and/or the shape of the face in a similar manner for different individuals. However, studies on alterations applied on face via plastic surgery or prosthetic make-up which can be in countless different ways and amounts, are still very limited. In this paper, we analyze how such changes on nose region affect the face recognition performances of several key techniques. For this purpose, a simulated face database is prepared using FRGC v1.0 in which nose in each sample is replaced with another randomly chosen one. Since this is a 3D database, the impact analysis is not limited to only 2D, which is one of the novelties of this study. Performance comparisons of three 2D and four 3D algorithms are provided. In addition, differently from previous works, baseline results for the original database are also reported. Hence, the impact which is purely due to the applied nose alterations can be measured. The experimental results indicate that with the introduction of alterations both modalities lose precision, especially 3D.**

***Key Words*—** Face recognition, plastic surgery, nose alteration

I. Introduction

Plastic surgery is considered to be a relatively new challenge in face recognition when compared to pose, expression or illumination variations. With the increasing number of people resorting to plastic surgery for correction of feature defects, cosmetic reasons or even law evasion, it becomes of interest for the biometric community to investigate and prevent the impact of facial alterations on recognition performances. Yet, very few studies exist which address this problem.

An evolutionary granular approach is proposed in [3] for matching a post-surgery face image with a pre-surgery face image and 15% improvement in identification performance is reported. Furthermore, two new methods, FARO and FACE, based on fractals and a localized version of correlation index, respectively, are implemented in [4] which claims that the performance of these two algorithms compare favourably against standard face recognition methods such as PCA and LDA in case of plastic surgery changes. Singh et al. adopted the near set theory to classify facial images that have previously undergone some feature modifications in [19].

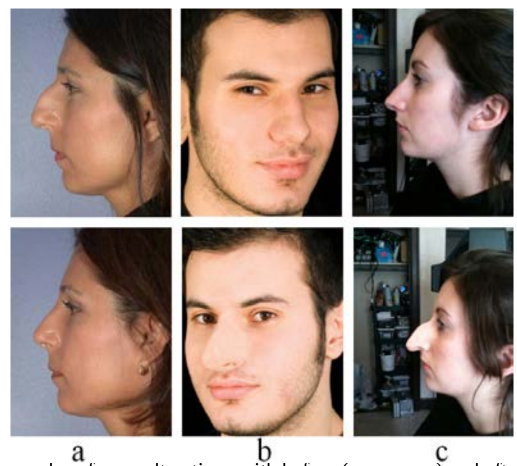

Fig. 1 Examples of nose alterations with before (upper row) and after (lower row) photos: (a) plastic surgery [20] (b) latex appliance [21] (c) makeup using wax [22]

In this paper, we focus on the nose modifications and analyse their effects on success rates of different face recognition methods. According to the statistics published by The American Society for Aesthetic Plastic Surgery in 2010 [5], nose reshaping (rhinoplasty) is the second most common surgical procedure on face after cosmetic eyelid surgery (blepharoplasty).

On the other hand, plastic surgery is just one of many ways to change the appearance of the nose. For example, latex/silicone-based prosthetic appliances can be simply purchased as off-the shelf products. Alternatively, makeup using wax or putty can also alter the nose shape very easily. Three nose alteration examples for the three aforementioned methods are given in Fig. 1.

A. Related Work

To the best of our knowledge, the impact of facial alterations, specifically due to plastic surgeries, on face

MMSP’12, September 17-19, 2012, Banff, Canada.

Fig. 2 Examples of facial hair, expression and makeup variations on the facial images between before (upper row) and after (lower row) plastic surgery procedure

recognition was first analysed by Singh et al. in [1] where the effect of plastic surgery is evaluated on six recognition algorithms. The database used consisted of 506 subjects with 2 images: before and after the plastic surgeries. Later, this work was extended in [2] by augmenting the database up to 900 subjects and additionally including a different non-surgery database for performance comparison. The results showed that the evaluated appearance, feature and texture-based algorithms were unable to effectively mitigate the decline caused by plastic surgery procedures.

Three shortcomings of these studies, which will be addressed throughout this paper, can be identified as follows:

- Due to the fact that a single image is provided before the plastic surgery procedure, a non-surgery vs. non-surgery recognition experiment had to be conducted on a separate database with different subjects. Unfortunately, for face recognition algorithms, the accuracy can vary widely depending on the difficulty of the database. Hence, an authentic comparison is not possible.
- In the plastic surgery database, the before and after images differ not only as a result of the procedure, but also due to expressions, makeup and facial hair variations (Fig. 2). This leads to an additional decrease in the performances which clouds the true measurement of the plastic surgery effect.
- Since this is an image database, the analyses are restricted to 2D. However, 3D face recognition gains a rising popularity as it offers superiority over to its 2D counterpart by being intrinsically robust against illumination and pose variations. For this reason, the impact of the facial alterations on 3D algorithms should also be investigated.

In this paper, these limitations are eliminated by creating a synthetic database using FRGC v1.0 [6] for which nose regions are randomly exchanged between subjects. In this way, a 2D+3D database is obtained for nose alterations and since the conditions and the subjects are identical for the original and the simulated databases, measuring the exact impact of the applied changes is possible.

The rest of the paper is organized as follows: Section 2 describes the method for simulated nose alterations. Section 3 gives the experimental results to show the effect of the applied changes in both 2D and 3D face recognition and finally, the conclusions are provided in Section 5.

II. Simulating Nose Alterations

The nose region can be altered in many ways using plastic surgery, prosthetic appliances or makeup and it can be made bigger, smaller, wider or thinner. In order to simulate these changes and preserve the authenticity of the facial shape, noses in the database are replaced by randomly chosen ones from different subjects. For this purpose, a Thin Plate Spline (TPS) based method is implemented.

A metamorphosis technique for 3D plastic surgery simulation was proposed in [7], where three morphing operations: augmentation, cutting and lacerating were simulated. Later in [8], an automatic virtual plastic surgery system was presented which similarly to our approach, replaced an individual's facial features with corresponding features of another individual and fused the replaced features with the original face, but only in 2D. In a more recent work [9], effective patient-specific improvements for facial attractiveness are automatically suggested using 3D scans of the patients and the results are simulated by merging the target feature of the most similar face in the 3D database of attractive faces with the patient’s face.

In this paper, beautification of the faces is not a concern. What we aim is to change nasal regions in the database as realistically as possible and create nose variations for all subjects. Therefore, a target list is randomly generated to transfer noses (from another person) for each sample in the database.

For this purpose, firstly, nose regions of all facial scans are automatically segmented in a similar manner to [10] where an annotated generic face model is deformed to fit the target models and the annotations are transferred.

Next, nose deformations are applied using TPS method [11] in 3D. Prior to warping, the target model is aligned with the source model using 4 of 5 landmark points around the nose (Fig. 3), excluding the nose tip. A linear transformation that includes rotation, translation and isotropic scaling is computed in a least square sense, based on the two sets of landmarks and applied onto the source model. Subsequently, using all 5 point pairs a coarse TPS approximation is computed.

In the final step, for one-fifth of the vertices on the target nose, the closest vertices on the source nose are found and coupled to be utilized in a second and denser TPS warping, which results in the source nose completely transforming into the target nose. The proposed method is illustrated in Fig. 3.

The original database FRGC v1.0 consists of 943 multimodal samples from 275 subjects and the simulated 3D database is of the same size.

For simulated samples in 2D, the synthesized 3D models with the corresponding original texture mapped on. However, due to some mismatches in 2D and 3D samples in the original database, 39 samples had to be removed, leaving us with a database of 904 samples from 268 subjects.

In order to evaluate the visual plausibility of the created database, an online survey was conducted, for which the

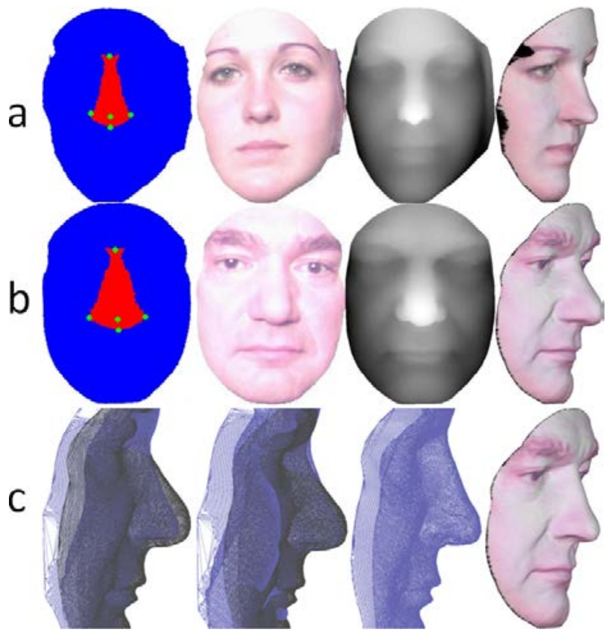

Fig. 3 From left to right: (a) Nose region with landmark points, color map, depth map and profile view for the target model (b) Same images for the source model (c) Two models superimposed before and after alignment, resulting mesh after warping and profile view for the synthesized model

participants were asked to classify the randomly displayed facial images (with or without texture) as original or simulated. According to a total number of 81 participations, success rate is found to be 60.68% for the images displayed with texture. For the ones without texture, the performances deteriorate as expected (58.77%) since the texture gives a better hint about originality. Being very close to the average performance of a random classifier (50%), this result indicates very low distinguishability. This means that a highly realistic look for the simulated noses is achieved, which is our aim while preparing this database.

For the sake of clarity, the original databases in 2D and 3D will be referred as DB-o2 and DB-o3, while the simulated nose alteration databases will be referred as DB-s2 and DB-s3, in the rest of this paper.

III. Experimental Evaluation

The effect of the applied nose alterations on face recognition performances are evaluated with three different scenarios in both 2D and 3D which are determined according to the study of Singh et al. [2] for comparison purposes.

Initially, all four databases, DB-o2, DB-o3, DB-s2 and DB-s3 are partitioned in non-overlapping training and testing datasets. This is done by randomly selecting 40% of the subjects and assigning their samples to the training set, while the rest is used for testing. The partitioning is repeated 10 times and verification and identification performances are computed over these 10 trials. The scenario for non-overlapping partitioning of training and testing datasets are also determined according to the study in [2] for comparison purposes. Many face recognition algorithms require training data. For example, the face recognition algorithm PCA need training data to compute the eigenspace. In our study, training data is used for the experiments with PCA and LDA. As it is explained in [23], in face recognition, training and testing datasets are selected as no overlaps between them.

In face recognition, performance can be recorded on both verification and identification [23]. In face verification, a claimed identity is validated based on the image of a face, and it either accepts or rejects the identity claim. It is a one-to-one matching. If the similarity score is above some threshold, the user's identity is verified. On the other hand, in face identification, the references for all faces in the database are examined and the one with the best match-score denotes the class of the input, which is a one-to-many matching.

For verification tests, all vs. all verification is applied. This means that each image in the test set are compared with all other images in the test set, and a similarity score is obtained for each image pair. The Receiver Operating Characteristic (ROC) curves which plot Verification Rates (VR) as a function of False Acceptance Rates (FAR) are reported together with the verification rates at 0.001 FAR.

For identification tests, the first sample of every individual in the test set is used in the gallery set and the rest is used in the probe set. In [23], the gallery and the probe set terminologies are determined as follows: The gallery is the set of reference images of the people to be recognized. These images are given to the algorithm as examples of each person who might need to be recognized. The probe is the set of the test images. The images in the probe set are presented to the system to be classified with the identity of the person with the image. The rank-1 recognition rates are reported for identification tests.

- **Experiment 1 – Performance on the original database**: It is important to compute the performances on the original datasets in terms of having a baseline performance. In this way, the impact of the applied changes can be measured accurately. For this purpose, 2D and 3D algorithms are evaluated on DB-o where the similarities are calculated between each original image pair.

- **Experiment 2 – Performance on the simulated database:** In this scenario, the similarity scores between every DB-o and DB-s sample pairs are calculated and used to evaluate recognition performances. For each subject selected for the training set, half of the corresponding images are taken from DB-o and the rest from DB-s. The selection of training set partitioning is in accordance with the study in [2].

Experiment 2 is identical to Experiment 1, except the probe images are now replaced by their modified versions.

- **Experiment 3 - Performance on the simulated database with training on an external database:** Face recognition algorithms are usually trained using different databases. In fact, face recognition algorithms are unlikely to be trained using before and after facial alteration images. Therefore in this scenario, Experiment 2 is repeated, but the training partition is composed of samples from an external database, namely Texas 3D Face Recognition Database [12] [13] [14]. Briefly, the Texas 3D Face Recognition (Texas 3DFRD) database is a collection of 1149 pairs of facial color

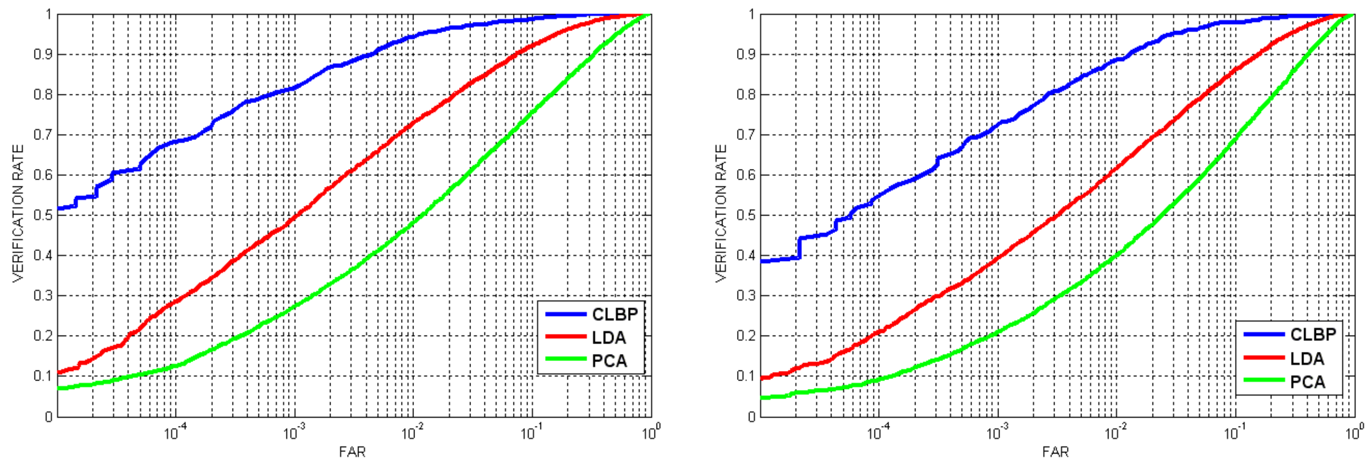

Fig. 4 Verification rates for all 2D FR algorithms by Experiment 1 (left) and Experiment 2 (right)

and range images of 105 subjects. In order to obtain a training set of similar subject and sample numbers as in experiments 1 and 2, a subset of Texas 3DFRD is compiled with 350 color and range images of 103 subjects, without expressions.

A. Evaluation on 2D Face Recognition Algorithms

Three key methods are chosen to be evaluated for 2D face recognition: Principal Component Analysis (PCA) [15], Linear Discriminant Analysis (LDA) [15] and Circular Local Binary Pattern (CLBP) [16]. PCA and LDA are appearance-based approaches which are widely used for dimensionality reduction and feature extraction. On the other hand, CLBP is a texture-based algorithm for describing local structures.

In this part, features of images are extracted using these three techniques. Then, similarity between each image pair is computed by using 'cosine' distance metric for PCA and LDA and by using 'chi-square' distance metric for CLBP. Verification and identification tests are done using these similarity scores between image pairs.

The plastic surgery database in [2] has images after several types of surgery operations such as forehead surgery, ear surgery, eyelid surgery, nose surgery, skin resurfacing, face lift etc. However, since the results are reported for each surgical operation separately, we can compare our results with the reported rank-1 identification accuracies for nose surgery (rhinoplasty) as shown in Table I.

According to this comparison, it is observed that even the evaluated databases are completely different (one is the simulated nose alteration database, the other one is the real plastic surgery database [2]), very similar identification results are achieved with both PCA and LDA for Exp.3 for which training is done using an external database.

TABLE I
RANK-1 IDENTIFICATION ACCURACIES FOR 2D FR ALGORITHMS FOR EXPERIMENT 1, 2 AND 3

Algorithm	Exp. 1	Exp. 2	Exp. 3	Exp.3[2]
PCA	40.24%	30.02%	24.74%	23.1%
LDA	64.74%	51.56%	27.94%	24.1%
CLBP	91.90%	86.52%	86.52%	44.8%

Observing such consistent results with a real plastic surgery database indicates high accuracy for our synthetic database.

However, this is not the case for CLBP. Very different identification rates are obtained, mainly due to two main reasons: The significant variance between the pre-surgery and post-surgery images in [2] (as shown in Fig. 2) and the fact that in our case, the only variation is due to the nose alteration and hence the change in the image texture is minimal.

TABLE II
VERIFICATION RATES AT 0.001 FAR FOR 2D FR ALGORITHMS FOR EXPERIMENT 1, 2 AND 3

Algorithm	Exp. 1	Exp.2	Exp. 3
PCA	27.50%	21.18%	11.66%
LDA	50.69%	40.11%	15.30%
CLBP	81.51%	71.72%	71.72%

The verification rates at 0.001 FAR and the ROC curves for all three algorithms are given in Table II and Fig. 4, respectively. For CLBP, the rates are identical for Experiments 2 and 3, since no training is required.

According to the results in Table I and II, best performance is obtained using CLBP method for both identification and verification with a marked difference. This shows that being a texture based method, CLBP is much more appropriate than appearance based methods, PCA and LDA in case of nose alterations. With CLBP, the percentage change between the results of Experiment 1 and Experiment 2 ($(result_{Exp1} - result_{Exp2}) / result_{Exp1}$) for identification and verification are found as 5.85% and 12.01%, respectively.

Robustness of LDA is observed to be higher than PCA, with ~20% decrease for both verification and identification scenarios. Whereas, PCA suffers 25.40% and 22.98% loss in identification and verification accuracies, respectively.

Utilization of an external database worsens the results even further for both identification and verification experiments.

B. Evaluation on 3D Face Recognition Algorithms

For the evaluation of 3D face recognition systems, 4 algorithms are selected where the facial surfaces are represented as depth maps or point clouds.

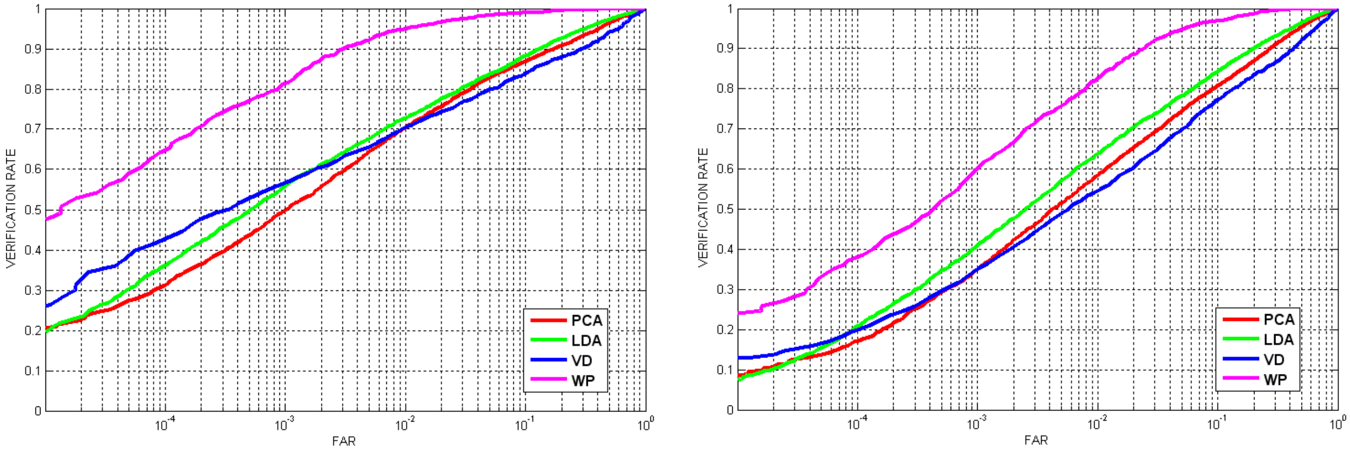

Fig. 5 Verification rates for all 3D FR algorithms by Experiment 1 (left) and Experiment 2 (right)

Depth maps can be involved in most of the existing 2D techniques, including subspace methods. In this part of the study, similar to the 2D evaluations, PCA and LDA are selected to be evaluated.

Additionally, two different approaches are implemented for 3D face recognition using point clouds. In this representation, faces are required to be registered prior to similarity measurements.

For this reason in the first technique, the faces are aligned with a generic face model using 3 landmark points (2 outer eye corners and the nose tip) and then the depth values are regularly sampled. The similarity (in this case distance) between two faces is obtained by averaging the z-distances of all vertices. In this way, the volume difference (VD) between two facial surfaces is approximated.

For the second approach, a method that we previously introduced in [17] was adopted. Again, initially, the facial surfaces are aligned with a generic face model using the same landmark points. Additionally in this approach, the alignment is further improved by Iterative Closest Point (ICP) method [18]. Afterward, 140 previously selected points on the generic model are coupled with the closest vertices on the face under analysis and TPS warping is applied resulting in warping parameters (WP) of size 140x3. Finally, the distance between two face models is computed by taking the median of the cosine distances between the corresponding feature vectors (WP).

The achieved rank-1 identification rates and the verification rates at 0.001 FAR by all 3D algorithms on databases DB-o3 and DB-s3 are given in Table III and Table IV. As is the case with CLBP, since the two point cloud methods do not require any training, rates for Experiments 2 and 3 are the same.

For identification, the best performing and most robust algorithm is observed to be WP, followed by LDA on range images. Both PCA and VD suffer a drastic decline (~25%) when nose alterations introduced. In this study, we compute the percentage change between the experiment results to obtain an idea on robustness of the methods to nose alterations.

TABLE III
RANK-1 IDENTIFICATION ACCURACIES FOR 3D FR ALGORITHMS FOR EXPERIMENT 1, 2 AND 3

Algorithm	Exp. 1	Exp. 2	Exp. 3
PCA	64.11%	48.40%	33.96%
LDA	68.47%	58.15%	42.03%
VD	68.26%	51.95%	-
WP	94.46%	86.64%	-

Likewise, analysis concerning the verification rates reveals that LDA and WP are least affected from nose alterations. However in verification, deteriorations are much more visible for all four methods.

TABLE IV
VERIFICATION RATES AT 0.001 FAR FOR 3D FR ALGORITHMS FOR EXPERIMENT 1, 2 AND 3

Algorithm	Exp. 1	Exp. 2	Exp. 3
PCA	49.85%	35.22%	17.42%
LDA	56.67%	42.18%	17.74%
VD	56.97%	35.23%	-
WP	81.18%	60.79%	-

Similar to the case observed in 2D experiments, utilization of an external database has a negative effect on the recognition accuracies. Algorithms have better performances when they are trained on both pre- and post-alteration images.

IV. CONCLUSION

As means of altering facial shape proliferate, its impact on recognition performances becomes crucial to measure and prevent. Today, more and more people undergo plastic surgeries (From 2009-2010, there was almost a 9% increase in the total number of cosmetic surgical procedures and since 1997, there has been over 155% increase in the total number of cosmetic procedures [5].) not only for medical reasons but also to improve their appearance or even to hide their true identity. Easy-to-use appliances and makeup products are within reach of everyone who seeks ways to evade recognition.

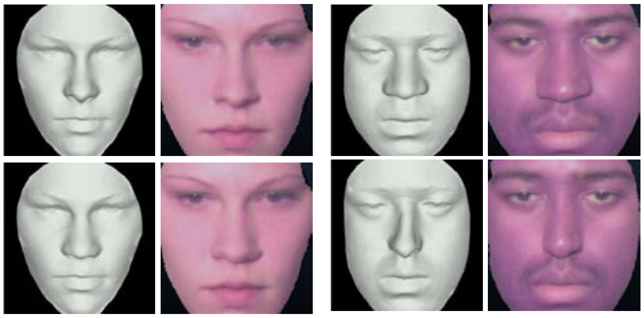

Fig. 6 Two examples of nose alterations with and without textures (upper row: originals lower row: altered)

In this study, a synthetic nose alteration database is obtained for which the nose of every subject in FRGC v1.0 is transfigured by replacing it with another randomly selected one. It is utilized to evaluate the performances of face recognition algorithms in presence of nose alterations.

The novelty of this contribution is that the analyses are not restricted to 2D images. Thanks to the nature of the simulated database, the effect of the applied modifications can be determined also in 3D. Additionally, since it is possible to measure the original performances on FRGC v1.0, an authentic comparison between pre- and post-alteration performances can be provided, which is a significant advantage of this study when compared to the previous ones.

The results reveal that the evaluated algorithms are not robust to the variations caused by nose alterations, especially for the purpose of verification. Furthermore, comparing verification performances of 2D and 3D algorithms show that 3D is much more vulnerable against the nose variations. On the other hand, robustness in identification is observed to be more method dependant then modality.

Robust face recognition algorithms are necessary to mitigate the effects of facial modifications. Our future research direction is to develop such face recognition methods. Additionally, we would like to measure the efficiency of nose alterations for face spoofing purposes.

ACKNOWLEDGMENT

This work has been partly performed by the TABULA RASA project 7th Framework Research Programme of the European Union (EU), grant agreement number: 257289. The authors would like to thank the EU for the financial support and the partners within the consortium for a fruitful collaboration. For more information about the TABULA RASA consortium please visit http://www.tabularasa-euproject.org.

REFERENCES

[1] R. Singh, M. Vatsa, *Effect of Plastic Surgery on Face Recognition: A Preliminary Study*, IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, pp.72-77, 20-25 June 2009.

[2] R. Singh, M. Vatsa, H. S. Bhatt, S. Bharadwaj, A. Noore, S. S. Nooreyezdan, *Plastic Surgery: A New Dimension to Face Recognition*, IEEE Trans. on Information Forensics and Security, vol. 5, no. 3, pp. 441-448, Sep. 2010

[3] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, and A. Noore *Evolutionary granular approach for recognizing faces altered due to plastic surgery*, IEEE International Conf. on Automatic Face & Gesture Recognition and Workshops, pp.720-725, 2011.

[4] M. De Marsico, M. Nappi, D. Riccio, H. Wechsler, *Robust Face Recognition after Plastic Surgery Using Local Region Analysis*, Lecture Notes in Computer Science - Image Analysis and Recognition, pp.191-200, Springer Berlin / Heidelberg, 2011.

[5] The American Society for Aesthetic Plastic Surgery, *Cosmetic Surgery National Databank Statistics, 2010* [Online]. Available (Last accessed on 03/12/2011.): http://www.surgery.org/sites/default/files/Stats2010.pdf

[6] 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*, IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 947-954, 2005.

[7] T.-Y. Lee, Y.-N. Sum, Y.-C. Lin, L. Lin and C. Lee, *Three-dimensional facial model reconstruction and plastic surgery simulation*, IEEE Trans. on Information Technology in Biomedicine, vol.3, no.3, pp.214-220, 1999.

[8] S.A. Rabi, and P. Aarabi, *Face Fusion: An Automatic Method for Virtual Plastic Surgery*, International Conf. on Information Fusion, pp.1-7, 2006.

[9] A. Bottino, M. De Simone, and A. Laurentini, *A computer-aided technique for planning plastic surgery based on 3D face scans: preliminary results*, Proceedings of International Conf. on 3D Body Scanning Technologies, 2010.

[10] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Yunliang Lu, N. Karampatziakis, and T. Theoharis, *Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.29, no.4, pp.640-649, April 2007.

[11] F. L. Bookstein, *Principal warps: Thin-Plate Splines and Decomposition of Deformations*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 11(6):567–585, 1989.

[12] S. Gupta, M. K. Markey, and A. C. Bovik, *Anthropometric 3D Face Recognition*, International Journal of Computer Vision, Volume 90, 3:331-349, 2010.

[13] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, *Texas 3D Face Recognition Database,* IEEE Southwest Symposium on Image Analysis and Interpretation, p. 97-100, Austin, TX, 2010.

[14] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, *Texas 3D Face Recognition Database* [Online]. Available: http://live.ece.utexas.edu/research/texas3dfr/index.htm

[15] P. N. Bellhumer, J. Hespanha, and D. Kriegman, *Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17, no. 7, pp. 711–720, 1997.

[16] T. Ahonen, A. Hadid, and M. Pietikäinen, *Face recognition with local binary patterns*, Computer Vision, ECCV 2004 Proceedings, Lecture Notes in Computer Science 3021, Springer, 469-481.

[17] N. Erdogmus, J-L. Dugelay, *On Discriminative Properties of TPS Warping Parameters for 3D Face Recognition*, International Conf. on Informatics, Electronics & Vision 2012, ICIEV 2012, May 2012.

[18] P. Besl and N. McKay, *A Method For Registration of 3-D Shapes*, IEEE Trans. on Pattern Analysis and Machine Intelligence,vol.14, no. 2, pp. 239-256, 1992.

[19] K. R. Singh, R. S. Khedgaonkar, and S. P. Gawande, *A New Approach to Local Plastic Surgery Face Recognition Using Near Sets*, International Journal of Engineering Science and Technology, NCICT Special Issue, pp.71-75, 2011.

[20] http://behind-the-knife-cosmetic-surgery.blogspot.com

[21] http://www.cinemasecrets.com

[22] http://good-times.webshots.com

[23] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, Second Edition, Springer, 2011.