Effect of Plastic Surgery on Face Recognition: A Preliminary Study

Richa Singh, Mayank Vatsa
IIIT Delhi, India
{rsingh, mayank}@iiitd.ac.in
Afzel Noore
West Virginia University, Morgantown, USA
afzel.noore@mail.wvu.edu

Abstract

Variations in pose, expression, illumination, aging and disguise are considered as major challenges in face recognition and several techniques have been proposed to address these challenges. Plastic surgery, on the other hand, is considered as an arduous research issue; however, it has not yet been studied either theoretically or experimentally. This paper focuses on analyzing the effect of plastic surgery in face recognition algorithms. The preliminary study provides an experimental and analytical comparison of face recognition algorithms on a plastic surgery database of 506 individuals. The experimental results indicate that existing face recognition algorithms perform poorly when matching pre and post surgery face images. The results also suggest that it is imperative for future face recognition systems to be able to address this important issue and hence there is a need for more research in this important area.

1. Introduction

The allure for plastic surgery is experienced world-wide and is driven by factors such as the availability of advanced technology, affordable cost, and speed with which these procedures are performed. Plastic surgery is generally used for improving the facial appearance, for example, removing birth marks, moles, scars and correcting disfiguring defects. However, it can also be misused by individuals to conceal their identities with the intent to commit fraud or evade law enforcement. Face recognition after plastic surgery can lead to rejection of genuine users or acceptance of impostors. While face recognition is a well studied problem in which several approaches have been proposed to address the challenges of illumination [1, 2], pose [3, 4, 5], expression [2], aging [6, 7] and disguise [8, 9], the use of plastic surgery introduces a new challenge to designing future face recognition systems.

In general, plastic surgery can be classified into two distinct categories.

1. *Disease Correcting Local Plastic Surgery* (*Local Surgery*): This is the kind of surgery in which an individual undergoes local plastic surgery for correcting defects, anomalies, or improving skin texture. Example of disease correcting local plastic surgery would be surgery for correcting jaw and teeth structure, nose structure, chin, forehead, and eyelids. Although the global appearance may look similar, this type of surgery usually leads to varying amount of changes in the geometric distance between facial features. Such changes may cause errors in automatic face recognition and degrade the system performance.

2. *Plastic Surgery for Reconstructing Complete Facial Structure* (*Global Surgery*): Apart from local surgery, plastic surgery can be done to completely change the facial structure which is known as full face lift. This medical procedure is recommended for cases such as patients with fatal burn or trauma. In this type of surgery, the appearance, texture and facial features of an individual are reconstructed and are usually not the same as the original face. The procedure is very useful for patients, but it can also be misused by criminals or individuals who want to remain elusive from law enforcement. Thus, using this procedure, the face recognition system can be easily manipulated and made ineffective.

To the best of our knowledge, there is no study that demonstrates any scientific experiment for recognizing faces that have undergone local or global plastic surgery. The main aim of the paper is to present this important challenge to the research community and systematically evaluate the performance of existing face recognition algorithms on a face database that contains images before and after surgery. The next section presents an analytical study of

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

plastic surgery on face recognition including an experimental evaluation of six face recognition algorithms using a facial plastic surgery database of 506 individuals.

2. Plastic Surgery and Face Recognition

In general, face recognition algorithms can be classified into three categories: appearance-based, feature-based, and texture-based algorithms. Appearance-based algorithms such as Principal Component Analysis (PCA) [10], Independent Component Analysis (ICA) [11], and Fisher Discriminant Analysis (FDA) [10] usually rely on the global semblance of features. Feature-based algorithms [12] generally establish a relationship among facial features and perform matching. Texture-based algorithms [9, 13, 14] on the other hand rely on the facial texture information to recognize an individual. Most of the existing face recognition algorithms have predominantly focused on mitigating the effects of pose, illumination and expression, while the challenges of face recognition due to aging and disguise still remains.

In this section, we investigate different aspects related to plastic surgery and face recognition. When an individual undergoes plastic surgery, the facial features are reconstructed either globally or locally. Generally, this process changes the appearance of an individual. As these procedures become more and more prevalent, future face recognition systems will be challenged to recognize individuals after plastic surgery has been performed. Until now, no attempt has been made to study the effect of local and global plastic surgery on face recognition. The major reasons for the problem not being studied are:

- Due to the sensitive nature of the process and the privacy issues involved, it is extremely difficult to prepare a face database that contains images before and after surgery.
- After surgery, the geometric relationship between the facial features changes significantly and there is no technique to detect and measure such type of alterations.

2.1. Plastic Surgery Database

In this research, we performed an experimental study to analyze the effect of both local and global plastic surgery on face recognition. We obtained face images before and after plastic surgery from different plastic surgeons across the world and prepared a plastic surgery database. This database contains a wide variety of cases such as Rhinoplasty (nose surgery), Mentoplasty (chin surgery), Brow lift, Malar augmentation (cheek implant), Skin peeling, and Rhytidectomy (face lift). Figure 1 shows an example of such cases. Note that, we have shown only the local facial features that are reconstructed and not the complete face to protect the identity of the individuals. Further, Table 1 shows the details of images in the plastic surgery database covering different types of surgery.

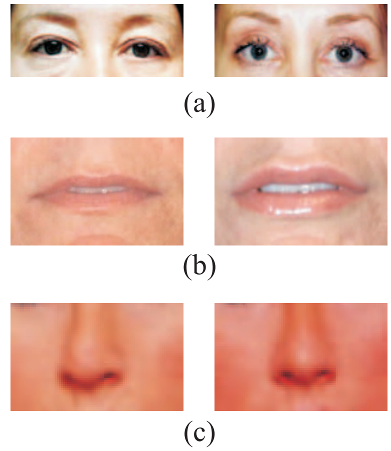

Figure 1. Examples of plastic surgeries: (a)Blepharoplasty, (b) Lip augmentation, and (c) rhinoplasty.

Type of Plastic Surgery	Number of Individuals
Rhinoplasty (Nose surgery)	71
Mentoplasty (Chin surgery)	29
Blepharoplasty (Eyelid surgery)	67
Brow lift (Forehead surgery)	54
Malar augmentation (Cheek implant)	21
Otoplasty (Ear surgery)	26
Liposhaving (Facial sculpturing)	44
Skin peeling (Skin resurfacing)	57
Rhytidectomy (Face lift)	49
Others (Craniofacial, Dermabrasion, Lip augmentation, Melasma, Mole removal, Rhytec, etc.)	88
Total	**506**

Table 1. Details of the plastic surgery database.

For each individual, there are two frontal face images with proper illumination and neutral expression: the first is taken before surgery and the second is taken after surgery. The database thus contains 1012 face images pertaining to 506 individuals. Further, the database contains 400 image pairs corresponding of local surgery and 106 cases of global surgery (i.e., skin peeling and face lift).

2.2. Face Recognition Algorithms for Evaluation

To study the effect of plastic surgery on face recognition, we selected six recognition algorithms that are based on appearance, feature, and texture.

1. Principal Component Analysis [10]
2. Fisher Discriminant Analysis [10]
3. Geometric Features (GF) [12]
4. Local Feature Analysis (LFA) [15]
5. Local Binary Pattern (LBP) [14]
6. Neural Network Architecture based 2D Log Polar Gabor Transform (GNN) [9]

PCA and FDA are appearance-based algorithms, GF and LFA are feature-based algorithms, and LBP and GNN represent texture-based algorithms. These algorithm are chosen for evaluation because they are either used as the basis for commercial systems (LBP and LFA) or have reported high accuracy in challenging scenarios (LBP, GNN, and variants of PCA, FDA).

2.3. Experimental Evaluation

In most real world applications, face verification systems are first trained on a training database and then the trained system is used to perform recognition on the gallery-probe face database. In such applications, it is highly probable that their is no overlap between the subjects used in the training database and the subjects in the gallery-probe database. To evaluate the performance of face recognition algorithms in such an application scenario, the plastic surgery database is partitioned into two groups: training database and testing database. Face images pertaining to 202 subjects (40% from the database) are used to train the face recognition algorithms and the remaining images pertaining to 304 subjects (60% from the database) are used as the test (gallery-probe) database for performance evaluation. This partition ensures that the verification is performed on unseen images. The train-test partitioning is repeated 20 times and the Receiver Operating Characteristics (ROC) curves are generated by computing the false rejection rates (FRR) over these trials at different false accept rate (FAR). The verification accuracy is computed at 0.1% FAR.

In the first set of experiments, we evaluated the performance of face recognition algorithms on local plastic surgery cases and global plastic surgery cases. Figures 2 and 3 show the ROC plots for face recognition algorithms using images with local and global plastic surgery respectively. Table 2 summarizes the verification accuracies at 0.1% false accept rate. Intuitively, face recognition algorithms should yield higher accuracy for images with local surgery compared to images with global surgery (i.e., skin peeling and full face lift). The experimental results in Table 2 also support this hypothesis. However, the results in Table 2 show that plastic surgery is a very challenging problem and the best verification performance for both local and global face recognition algorithms is below 50%. To analyze the results further, we performed a second set of experiments and evaluated the verification performance of face recognition algorithms on different types of plastic surgery. Table 3 shows a comprehensive break-up of the results according to the types of surgeries performed. The key observations and analysis are summarized below:

- Face recognition algorithms cannot handle global facial plastic surgery such as skin resurfacing and full face lift. With 20 times cross validation, the performance of recognition algorithms varies in the range of 0.6-16.0% which is not acceptable in real world applications. In most of the test cases, for global surgery, differences between pre and post surgery images of the same individual is very large. In other words, facial feature and texture is drastically altered after surgery and hence the algorithms do not yield good performance. For few test cases of skin resurfacing that have relatively closer resemblance in pre and post surgery images, most of the recognition algorithms are able to perform correct classification. However, with major skin resurfacing such as surgeries to look younger, none of the algorithms are able to correctly classify the faces.

- Among different types of plastic surgery, otoplasty, i.e. ear surgery has the lowest effect on the performance of face recognition. Most of the algorithms do not include the ear region for recognition. Therefore any change in ear shape, size, or texture does not affect the face recognition performance.

- Local facial regions such as nose, chin, eyelids, cheek, lips and forehead play an important role in face recognition. Any change in one of the regions, in general, affects the verification accuracy. For example, in LFA, nose and eyes play an important role and most of the local features are found close to these regions. Any change in these regions degrades the verification performance. Similarly, texture-based algorithms such as LBP and GNN yield lower accuracy for cases involving cheek and forehead regions.

- Liposhaving or facial sculpturing severely degrades the performance of all six verification algorithms. This is mainly because it removes the fat from the local facial regions and significantly changes the appearance of the face images.

- Dermabrasion and mole removal have relatively less impact on face recognition compared to melasma, thread lift and lip augmentation because the latter procedures affect the regions that are, in general, used by recognition algorithms.

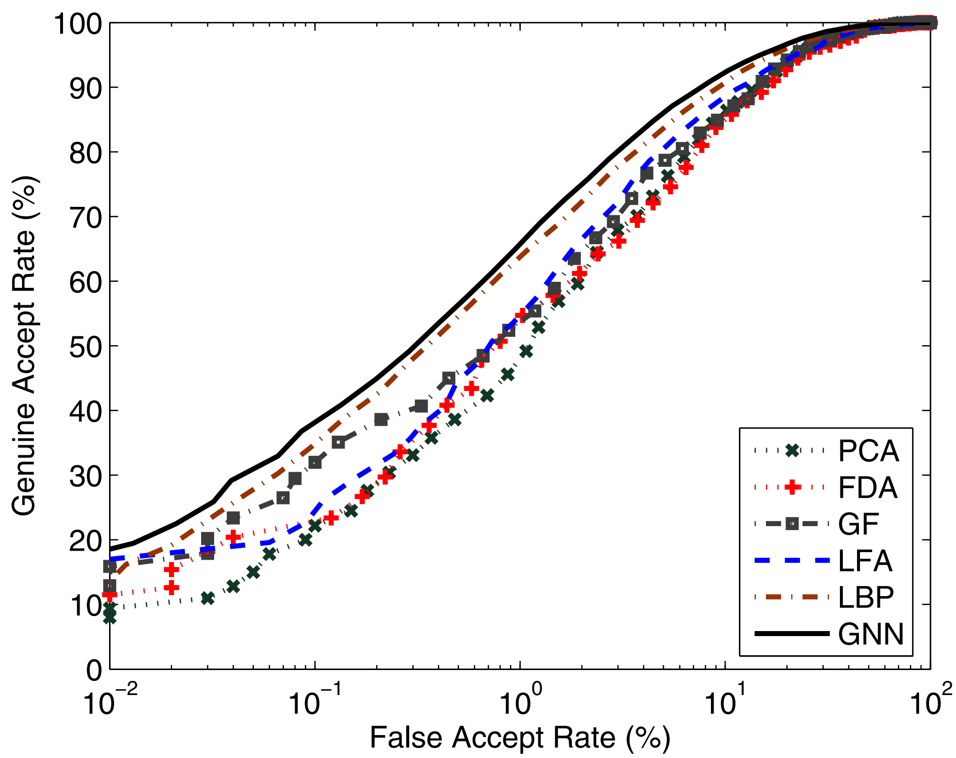

Figure 2. ROC plot demonstrating the performance of face recognition algorithms on the local plastic surgery.

Procedure	PCA	FDA	GF	LFA	LBP	GNN
Local Surgery	22.2%	22.9%	32.0%	24.4%	34.5%	38.8%
Global Surgery	2.8%	8.2%	7.0%	7.8%	9.5%	10.6%

Table 2. Performance of face recognition algorithms on local and global plastic surgery. Verification accuracy is computed at 0.1% false accept rate.

- Overall, the verification accuracies are in the range of 19.2% (PCA) - 34.1% (GNN). Note that this evaluation is based on the performance with neutral expression and frontal face images under proper illumination conditions. The results thus show plastic surgery alone is a major challenge in face recognition. It is highly desirable and required to consider it as a research issue and develop algorithms to confound these effects. One possible approach would be to use thermal-infrared imagery and compute the thermal differences between pre and post surgery images. However, such an approach first requires creating a large scale face database that contains pre and post operative thermal infrared images.

- The correlation analysis of match scores from the six recognition algorithms is performed using the Pearson correlation coefficient. It is observed that the appearance-based algorithms are not correlated with the feature-based and the texture-based algorithms. Similarly, there is very little correlation between texture-based algorithms and the feature-based algorithms. This suggests that these techniques provide complementary information and if we fuse the match scores using fusion technique, then the performance should improve. In our experiments with sum rule fusion and min-max normalization [16], match scores obtained from FDA and GNN yield the best verification accuracy of 49.2% for local surgeries and 15% for global surgery.

3. Conclusion

Current face recognition algorithms mainly focus on handling pose, expression, illumination, aging and disguise. This paper formally introduces plastic surgery as another

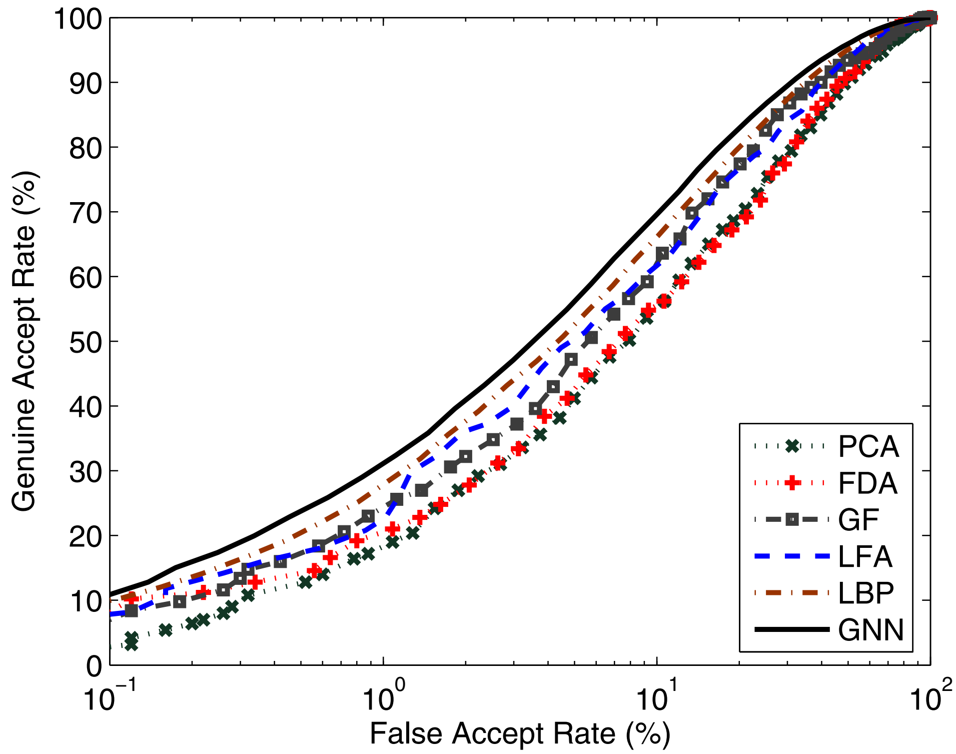

Figure 3. ROC plot demonstrating the performance of face recognition algorithms on the global plastic surgery.

Surgery	PCA	FDA	GF	LFA	LBP	GNN
Rhinoplasty (Nose surgery)	21.4%	22.1%	31.4%	23.3%	32.0%	37.3%
Mentoplasty (Chin surgery)	21.8%	22.5%	31.2%	24.5%	34.2%	38.5%
Blepharoplasty (Eyelid surgery)	24.3%	25.0%	34.7%	27.6%	36.4%	40.7%
Brow lift (Forehead surgery)	20.5%	20.8%	31.6%	22.8%	31.5%	37.0%
Malar augmentation (Cheek implant)	21.0%	22.5%	32.0%	24.5%	33.0%	36.5%
Otoplasty (Ear surgery)	100%	100 %	100%	100 %	100 %	100%
Liposhaving (Facial sculpturing)	12.0%	12.9%	12.3%	12.7%	18.2%	19.1%
Skin peeling (Skin resurfacing)	5.2 %	11.5%	10.8%	11.2%	14.8%	16.0%
Rhytidectomy (Facelift)	0.6 %	1.0 %	1.4 %	1.4%	1.8 %	2.0 %
Others	22.4%	23.1%	31.4%	25.6%	34.8%	39.0%
Overall	19.2%	20.4%	27.8%	21.6%	30.3%	34.1%

Table 3. Analyzing the effect of different types of plastic surgeries on face recognition algorithms.

major challenge for face recognition algorithms. Plastic surgery is becoming prevalent due to advances in technology, affordability, and the speed with which these procedures can be performed. The procedures can significantly change the facial regions both locally and globally, altering the appearance, facial features and texture. Existing face recognition algorithms generally rely on these information and any variations can affect the recognition performance. In this paper, we present an experimental study to quantitatively evaluate the performance of face recognition algorithms on a plastic surgery database that contains face images with both local and global surgeries. The study shows that PCA, FDA, GF, LFA, LBP and GNN algorithms are unable to effectively mitigate the variations caused by the plastic surgery procedures. Based on the results, we believe that more research is required in order to design an optimal face recognition algorithm that can also account for the challenges due to plastic surgery. It is our assertion that the results of this work would stimulate further research in this important area.

References

[1] S. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007. 1

[2] R. Singh, M. Vatsa, and A. Noore, "Improving verification accuracy by synthesis of locally enhanced biometric images and deformable model", *Signal Processing*, vol. 87, no. 11, pp. 2746–2764, 2007. 1

[3] V. Blanz, S. Romdhami, and T. Vetter, "Face identification across different poses and illuminations with a 3d morphable model", in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 202–207. 1

[4] X. Liu and T. Chen, "Pose-robust face recognition using geometry assisted probabilistic modeling", in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 502–509. 1

[5] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition", *IEEE Transactions on Systems, Man and Cybernetics - Part B*, vol. 37, no. 5, pp. 1212–1225, 2007. 1

[6] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Toward automatic simulation of aging effects on face images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–450, 2002. 1

[7] N. Ramanathan and R. Chellappa, "Face verification across age progression", *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3349–3362, 2006. 1

[8] N. Ramanathan, A.R. Chowdhury, and R. Chellappa, "Facial similarity across age, disguise, illumination and pose", in *Proceedings of International Conference on Image Processing*, 2004, vol. 3, pp. 1999–2002. 1

[9] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images", *Image and Vision Computing*, vol. 27, no. 3, pp. 245–257, 2009. 1, 2, 3

[10] P.N. Bellhumer, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 711–720, 1997. 2, 3

[11] M.S. Bartlett, J.R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis", *IEEE Transaction on Neural Networks*, vol. 13, pp. 1450–1464, 2002. 2

[12] I.J. Cox, J. Ghosn, and P.N. Yianilos, "Feature-based face recognition using mixture-distance", in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 1996, pp. 209–216. 2, 3

[13] L. Shen and L. Bai, "A review on gabor wavelets for face recognition", *Pattern Analysis and Applications*, vol. 9, no. 2, pp. 273–292, 2006. 2

[14] T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: application to face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006. 2, 3

[15] P. Penev and J. Atick, "Local feature analysis: a general statistical theory for object representation", *Network: Computation in Neural Systems*, vol. 7, pp. 477–500, 1996. 3

[16] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of multibiometrics*, Springer, New York, 1st edition, 2006. 4