Structural Similarity based Image Quality Map for Face Recognition across Plastic Surgery

Yunlian Sun[1,2], Massimo Tistarelli[1], Davide Maltoni[2]
[1]Department of Sciences and Information Technology, University of Sassari
Viale Mancini 5, 07100 Sassari, Italy
[2]Department of Computer Science and Engineering, University of Bologna
Via Sacchi 3, 47521 Cesena, Italy
{elaine.sun717,mtista}@gmail.com, davide.maltoni@unibo.it

Abstract

Variations in the face appearance caused by plastic surgery on skin texture and geometric structure, can impair the performance of most current face recognition systems. In this work, we proposed to use the Structural Similarity (SSIM) quality map to detect and model variations due to plastic surgeries. In the proposed framework, a SSIM index weighted multi-patch fusion scheme is developed, where different weights are provided to different patches in accordance with the degree to which each patch may be altered by surgeries. An important feature of the proposed approach, also achieving performance comparable with the current state-of-the-art, is that neither training process is needed nor any background information from other datasets is required. Extensive experiments conducted on a plastic surgery face database demonstrate the potential of SSIM map for matching face images after surgeries.

1. Introduction

Plastic surgery becomes worldwide nowadays due to the advanced surgical technologies and the affordable cost. By these medical procedures, people can correct defects of a facial feature for functionality improvement or modify the appearance for aesthetic improvement. Face recognition across plastic surgery was first introduced to the biometric community by Singh et al. [11]. In the presented work, a database of face images related to various types of plastic surgeries was publicly released. Moreover, various existing algorithms were tested on this database. The significant performance degradation concluded that the current state-of-the-art face recognition algorithms cannot provide good performance for matching faces across plastic surgery.

To handle the challenges of face recognition after plastic surgery, Bhatt et al. proposed an evolutionary granular approach to extract discriminative information from non-disjoint face granules [3]. In [7], a recognition approach which integrates information derived from local region analysis was proposed to address this problem. Aggarwal et al. developed a sparse representation based local facial characteristic matching approach [1]. In the related work, a sequestered face image set was used to fulfill the multiple image requirement of sparse representation approach. In [4], a fusion approach by combining information from both the whole face and the ocular regions [10] was proposed to deal with the challenges of matching faces across variations caused by plastic surgeries. Very recently, Liu et al. employed an ensemble of Gabor Patch classifiers via Rank-Order list fusion algorithm and achieved very promising results [6].

As presented in [11], both local and global surgeries may result in varying amount of change in relative positioning of facial features and texture. Generally, the positions of these changed features and texture are unknown to a face recognition algorithm. However, it would be of great use if the positions could be extracted automatically. In such cases, we can less consider or ignore these changed features and texture. In this work, we consider variations caused by surgeries as a variety of distortions on the pre-surgery facial images. Further, we shall attempt to exploit the quality information implicated in the pre- and post-surgery images to detect and capture these variations.

An effective image quality tool to well model variations caused by surgeries should interpret well the degradation of both texture and structural information. In [13], an objective image quality metric based on the Structural Similarity (SSIM) index was developed for localized quality measurement. Given a reference image and its distorted version, through locally computing the SSIM index, this technique can provide a spatially varing quality map of the distorted image, which delivers much information about the quality

degradation. In our work, we consider the pre-surgery image as a reference image and the post-surgery image as a distorted one. Then the SSIM quality map between the two images can be computed. Further, we employ this quality map in a patch level to control the contribution of each patch to the final matching score.

In face identification on the plastic surgery database, given a probe image y (post-surgery) and N gallery individuals (each with a pre-surgery image), the SSIM maps between y and each gallery image are first calculated. Next, we perform the matching between y and each gallery image using our proposed SSIM index weighted multi-patch fusion scheme. In this scheme, the two images being compared together with their SSIM map are first divided into the same number of patches. When matching two corresponding patches, we calculate the mean SSIM index of the corresponding SSIM map patch as the weight for controlling the contribution of the patch to the final matching score. An intuitive motivation for employing the mean SSIM index of each patch as the patch weight is that, in the SSIM map of two images, the value of each pixel (SSIM index) represents the structural similarity of the two image regions within a local window of the corresponding pixel. By directly employing the SSIM index as the weight, we can give less weights (smaller SSIM index values) to those changed features and texture, and vice versa. After matching between y and the N gallery images, we can get a total of N SSIM index weighted matching scores. The final class label is the identity owing the highest matching score (similarity score).

The proposed approach employs the SSIM map to compute different weights for different image patches in accordance with the degree to which each patch is altered by surgeries. A significant advantage of the proposed approach is that no training process, nor any background information from other databases is required. However, when matching faces of different individuals, for regions where the two faces differ most, lower weights are also assigned. To study the effect of our approach on genuine matching (comparing faces of the same person) and impostor matching (comparing faces of different individuals), we shall illustrate the score distributions. As it can be noticed from Figure 4, the overlap region between the genuine and impostor distributions is reduced by the SSIM weighting.

To thoroughly evaluate our proposed framework, we perform face matching in both a holistic manner and a component-wise manner. The proposed approach is evaluated on the plastic surgery database introduced in [11]. With our proposed approach, a significant improvement in recognition performance is observed.

2. Related Work

To quantitatively measure the image quality, several metrics have been proposed. The Structural Similarity (SSIM), proposed by Wang et al. [13], provides a spatially varying quality map of the two images being compared. The resulting measure allows to determine the location and degree of variations of the distorted image.

SSIM provides meaningful comparisons across different types of image distortions by separating the task of similarity measurement into three comparisons: luminance, contrast and structure. Suppose x and y are two image signals, which have been aligned with each other. The SSIM index between them is computed as follows:

$$SSIM\,(x,y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \quad (1)$$

where μ_x and μ_y are the mean intensity values of signal x and y, while σ_x and σ_y are their corresponding standard deviations. σ_{xy} is the correlation coefficient between x and y. C_1 and C_2 are small positive constants used to avoid instability when the denominator is very close to zero.

For image quality assessment, it is useful to apply the SSIM index locally rather than globally. In [13], the local statistics μ_x, σ_x and σ_{xy} are calculated within a 11×11 circular-symmetric Gaussian window, which moves pixel-by-pixel over the whole image. At each step, the local statistics and SSIM index are computed within the local window. If one of the image signals being compared is regarded as perfect quality, then with such a windowing approach, a SSIM index map of the other image can be obtained. In the SSIM map, the value of each pixel is the SSIM index computed within the local window of the corresponding pixel. The SSIM index map provides a measurement of local image quality over space, in which the dynamic range of each pixel value is $[-1, 1]$. A pixel with value close to 1 means less distortions in the neighborhood of the pixel, while a pixel with a lower SSIM index value implies some variations or quality degradation within the local region of the pixel are detected in the distorted image. A Matlab implementation of the SSIM index algorithm is available online[1].

3. SSIM for face recognition across plastic surgery

In the proposed framework, the pre-surgery image is regarded as perfect quality and used for the gallery image in face identification, whereas the post-surgery image is viewed as a distorted image and used as a probe image. Denote the number of gallery individuals as N, that is we have a total of N gallery images. Given a probe image with some variations caused by surgeries, our face identification task aims to determine which gallery individual the probe image comes from. To get the final classification, we need

[1] Z. Wang, MATLAB implementation of SSIM, https://ece.uwaterloo.ca/ z70wang/research/ssim/.

to first calculate the similarity scores (matching scores) between the probe and each gallery image, and the identity owing the highest similarity score is the correct class label. Figure 1 illustrates our approach of matching the probe with one gallery individual, which can roughly consist of the following two main steps: image alignment and SSIM index weighted multi-patch fusion. After alignment of the probe and the gallery images according to the eye centers, the SSIM map between the two images is computed. Next we divide the two images as well as their SSIM map into the same number of patches of the same size. The general idea of SSIM index weighted image matching is to employ this quality map in a patch level to control the contribution of each patch to the final matching score. The outline of using our approach for face identification across plastic surgery is shown in Figure 2. Next we present details of the main steps.

3.1. Image alignment

Images from the plastic surgery database are used in this work. For each subject, there is one pre-surgery image and one post-surgery image. All the images are of the same size 273×236 pixels. Like any other common face recognition algorithms, we first need to locate some fiducial landmarks and then normalize all the images according to the positions of these landmarks. Here, we choose to use eye centers for face alignment. Considering that a number of images in the plastic surgery database present some variations in the eye region due to expressions and poses, we choose to use the four eye corners to determine the eye centers. That is the mean position of the two left (right) eye corners is computed as the position of the left (right) eye center. To locate eye corners automatically, a publicly available tool known as STASM [9] was employed [2]. After the location of eye centers, the distance between the eye centers is set as 60 pixels. Finally, all the images are normalized to the size of 160×120.

3.2. SSIM index weighted multi-patch fusion

Given a pre-surgery image and a post-surgery image, we regard the pre-surgery image as a reference image, whereas the post-surgery image is viewed as a distorted image. Next the SSIM map of the post-surgery image is computed using the approach outlined in Section 2. For the parameters C_1 and C_2, we use the same values to those used in [13]. And the standard deviation of the 11×11 circular-symmetric Gaussian weighting function is set to 1.5. More details about how to calculate the SSIM map can be found in [13].

Figure 3 shows some SSIM index maps corresponding to some pre- and post-surgery image pairs. In the SSIM map, dark regions represent smaller SSIM index and larger distortions, while bright regions mean larger structural similarity and less quality degradation. As illustrated in Figure 3, after blepharoplasty (eyelid surgery), pouches in both lower eyelid regions are nearly removed. Obviously, the correspondig regions in the SSIM index map are darker than the neighborhood regions. Similar results can also be found in images after rhinoplasty (nose surgery) and lip augmentation. For global surgeries, we take laser skin resurfacing as an example. After this surgery, the chin, cheek and mouth regions present larger variations than other regions. Fortunately, these variations are well reflected in the SSIM map. From Figure 3, we can see the effectiveness of using SSIM index map to model both local and global variations caused by plastic surgery when matching a genuine pair (i.e. where the identity of the probe and gallery faces is the same).

Next we divide the two images as well as their corresponding SSIM index map into n patches of the same size. Denote the n patches as $\{p_1, p_2, ..., p_n\}$. For each patch p_i, feature extraction is then performed on the two images being compared. After feature extraction, we can employ a classifier c_i for each patch p_i. Denote the output of the classifier c_i as sc_i^o, i.e. the matching score of the i^{th} patches of the two images. In a conventional fusion scheme, the final matching score sc^o can be computed directly using the sum rule as follows.

$$sc^o = \sum_{i \in 1,...,n} sc_i^o \tag{2}$$

Note that this provides the same weights to all the patches. In our work, we give different weights to different patches according to the degree of their quality degradation. Specifically, the mean pixel value of each patch p_i in the SSIM map is calculated as the weight of the corresponding classifier c_i. We represent the weight of c_i as w_i, then in our SSIM index weighted multi-patch fusion scheme, the final weighted matching score sc^w is calculate as follows.

$$sc^w = \sum_{i \in 1,...,n} w_i \times sc_i^o \tag{3}$$

As described in Section 2, the values of each pixel in the SSIM map locate in a dynamic range of $[-1, 1]$. In other words, the calculated weights maybe smaller than zero. This is not in accordance with our objective, in which regions with large variations will be less considered or be ignored. There are two intuitive solutions to this problem. One can either normalize all the pixel values to $[0, 1]$ or normalize all the patch weights to $[0, 1]$. However, our experiments on the plastic surgery database show that both normalization solutions do not affect the recognition performance significantly in contrast with the original might minus pixel values or weights. Thereby, in all our experiments

[2] S. Milborrow, C++ software library of STASM, http://www.milbo.users.sonic.net/stasm/.

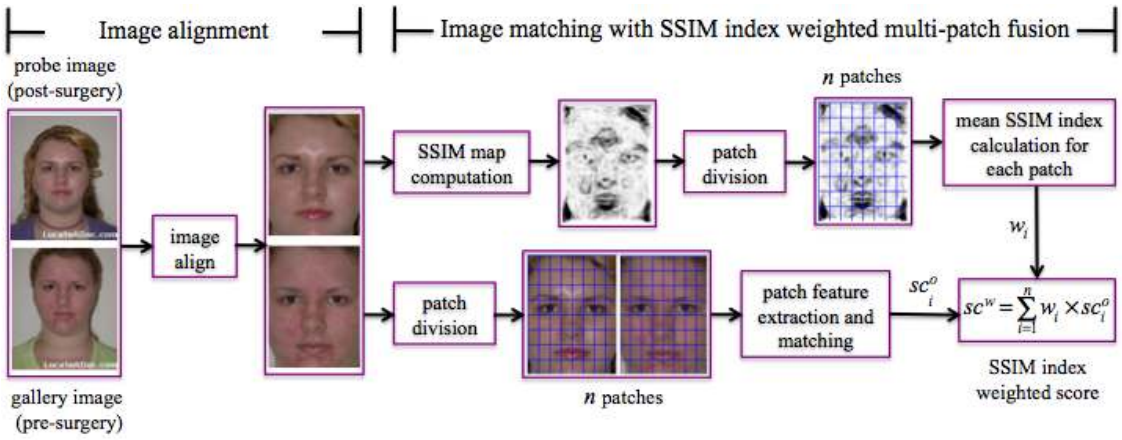

Figure 1. Outline of the process to compute the reliability weight, for face matching, from SSIM maps.

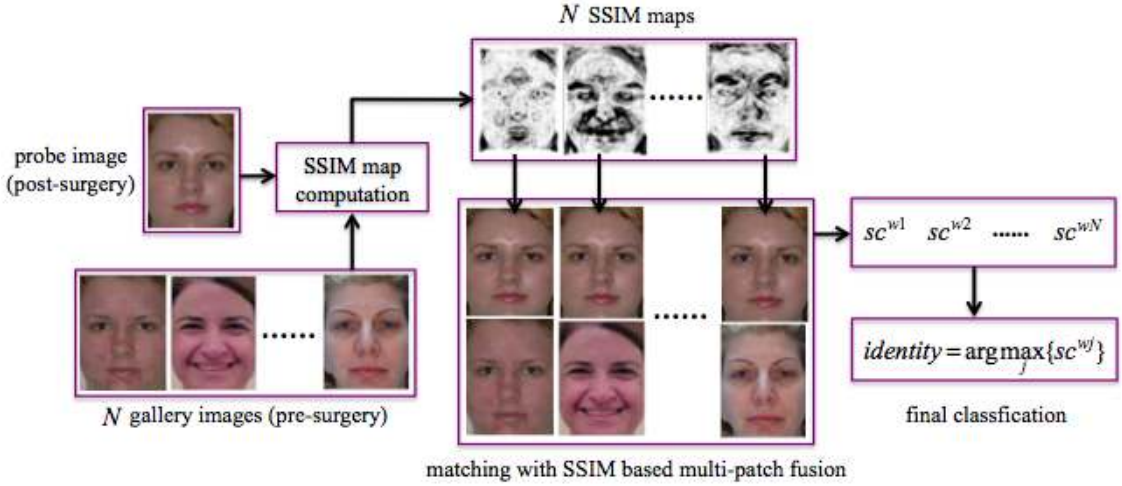

Figure 2. Outline of the process for face identification using SSIM-based image matching.

later, we directly use the original SSIM map for weight calculation.

3.3. Face identification using SSIM based patch fusion

As shown in Figure 2, given a probe image y and N gallery individuals, the SSIM maps of the probe y and each of the N gallery images are first calculated. After this, we can get a total of N SSIM maps. Next, SSIM index weighted multi-patch fusion scheme can be employed to match y with each of the N gallery images. Finally, we can get N SSIM index weighted matching scores $\{sc^{w1}, sc^{w2}, ..., sc^{wN}\}$. The final classification is performed as follows:

$$identity(y) = \arg\max_{j}\{sc^{wj}\} \quad (4)$$

4. Analysis of SSIM weighted patch fusion

The proposed approach employs the SSIM map in a patch level to provide different weights to different patches in accordance with the degree to which each patch is altered by surgeries. The advantages of the proposed approach are as follows:

1) Effectiveness for modeling variations caused by surgeries. Figure 3 well illustrated this.

2) No training cost. our approach does not employ any training procedures.

3) No background information employed. Background information [14], also named cohort information in some literature[8], is extracted on an additional background dataset. Faces in the background dataset are disjoint by identity from the test faces. As is known,

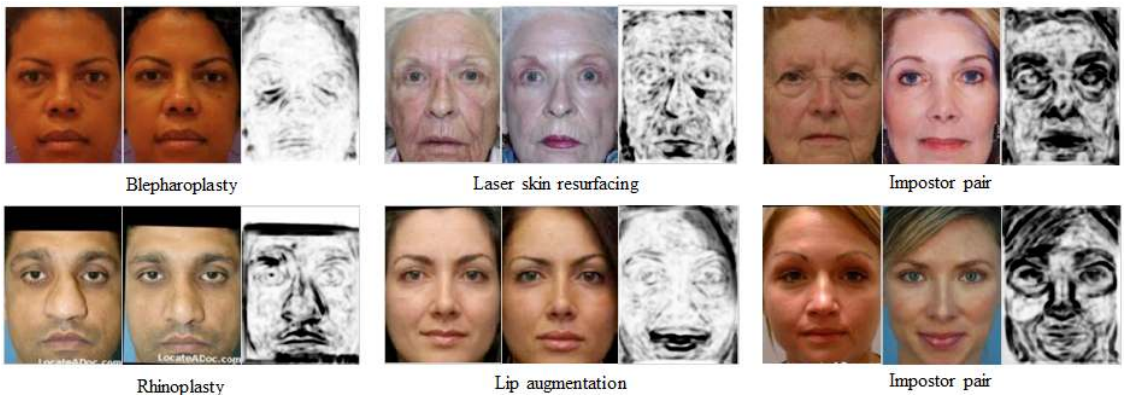

Figure 3. The SSIM maps of some pre- and post-surgery image pairs.

in most existing literature related to recognition across plastic surgery, a background dataset is collected to handle the problems caused by insufficient gallery images of each individual [1, 6]. Note that in our approach, we do not use any background information.

However, using our approach for matching faces of different individuals, also the regions where the two faces differ most are given lower weights. Two SSIM maps corresponding to two pre- and post-surgery impostor pairs (i.e. where the identity of the probe and gallery faces is not the same) are shown in Figure 3. For both impostor pairs, the two faces differ from each other significantly around the eye, nose and mouth regions. Even though these regions encode most of the discriminative information in faces, the SSIM mapping assigns lower weights in comparison with other regions, such as the forehead and cheek regions.

To well study the effect of the SSIM approach on genuine and impostor matching, the two score distributions on the plastic surgery database are shown in Figure 4. The matching scores were computed on the entire faces by means of LBP features [2]. The total number of genuine and impostor scores are respectively N and $N \times (N-1)$, in this case $N = 784$. More details about the experimental setting can be found in Section 5. The dashed plots correspond to the distributions computed without using the SSIM map, while the solid plots correspond to the distributions resulting from the SSIM weighting. After the SSIM weighting, both the genuine and impostor scores decrease. Most important, the overlap region between the genuine and impostor distributions is reduced after the SSIM weighting.

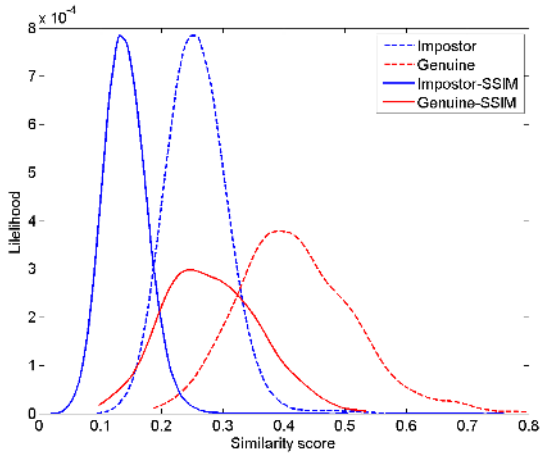

Figure 4. The score distributions of genuine and impostor before and after SSIM.

5. Experiments

The proposed approach was tested on a plastic surgery database containing $1,800$ images of 900 subjects. These images were collected from plastic surgery information websites, hence, many present a number of non-ideal factors such as duplicate entries, incomplete faces and non-frontal faces. Similarly to the pre-selection procedure used in [4], we excluded images which show the above negative factors. Finally, a dataset consisting of $1,576$ images from 784 subjects was selected, i.e., $N = 784$. In comparison with the evaluation schemes used in [11] which use 40% of the dataset for training and only the remaining 60% for testing, all our experiments were performed on the whole dataset.

For feature extraction, we employ two widely used facial features LBP [2] and Gabor [5]. In all our experiments, the size of each patch is set to be 8×8. When extracting LBP features, for each of the $(160 \div 8) \times (120 \div 8) = 300$ patches, we extract a 59-bin uniform LBP histogram. To compute the Gabor feature, we adopt five scales and eight orientations of the Gabor filters. After this, we can get a 40-dimension Gabor jet for each pixel. Directly using Gabor jets from all pixels in the 8×8 patch as the feature

representation will result in a very large dimension feature vector ($64 \times 40 = 2,560$), thus having a high computational cost. To address this problem, we use responses at a smaller set of pixels selected uniformly with a 8×8 down-sampling rate. The SSIM map is down-sampled using the same down-sampling rate. The down-sampled responses have only $(160 \div 8) \times (120 \div 8) = 300$ pixels, each of which corresponds to a 40-dimension Gabor jet. Here we use the down-sampled pixels to simulate the patches described above, and use the 40-dimension Gabor jet at each down-sampled pixel as the patch feature representation, thus having a much lower computational cost. For the matching score, we compute the cosine similarity between the two descriptor vectors.

5.1. Experimental results

To thoroughly validate the proposed approach, we perform face matching in both a holistic manner and a component-wise manner. We shall first treat the case of the holistic manner and return our discussion to the case of component-wise manner later.

Holistic matching. The whole face image is divided into a number of the above mentioned 8×8 patches. The accuracy is reported in terms of Cumulative Match Characteristic (CMC) curves. Figure 5 shows the CMC curves for: (a) holistic LBP without SSIM quality; (b) holistic LBP with SSIM quality; (c) holistic Gabor without SSIM quality; (d) holistic Gabor with SSIM quality. As expected, when using LBP and Gabor features in a holistic manner, SSIM weighted multi-patch fusion significantly outperforms approaches without using SSIM quality information. The Rank-1 accuracies of these scenarios were observed to be: (a) 65.05%; (b) 73.85%; (c) 59.95%; (d) 69.52%; SSIM quality improves the Rank-1 accuracy of LBP feature about 8.80%, while with Gabor feature, we can get an increased accuracy of almost 9.57%.

Component-wise matching. Seven facial regions are extracted, including: forehead, leftocolar, rightocular, nose, leftcheek, right cheek and mouth. Figure 6 shows the seven components and their size. The remaining process, including feature extraction and matching, is performed individually for each facial component. For the component-wise approach, we perform experiments using only LBP features. Figure 7 and Figure 8 show the individual CMC curves corresponding to the seven components. As we can see that, the proposed approach improves the recognition accuracy significantly. The increased Rank-1 accuracy for all the seven components is here listed: forehead (8.80%), leftocular (16.46%), rightocular (17.35%), nose (8.29%), leftcheek (15.95%), rightcheek (15.94%) and mouth (3.57%).

Next we fuse the outputs of the seven component classifiers for the final decision. Considering that matching faces using a component-wise manner might lose some useful information of the whole face geometric structure, we auto-

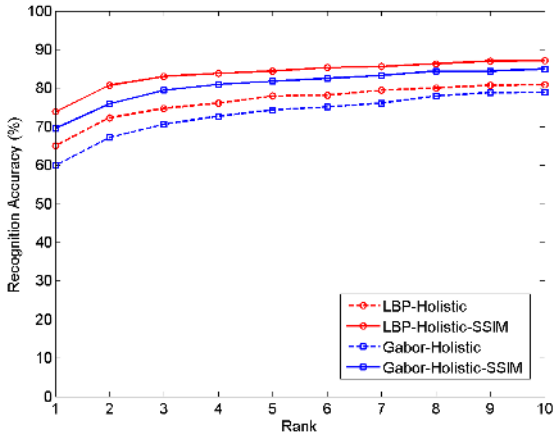

Figure 5. The CMC plots illustrating the performance of holistic matching algorithms.

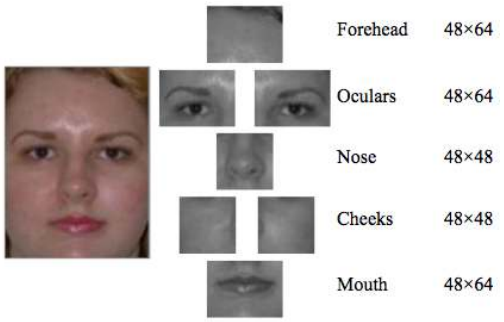

Figure 6. The 7 components used in the component-wise matching mechanism and their size.

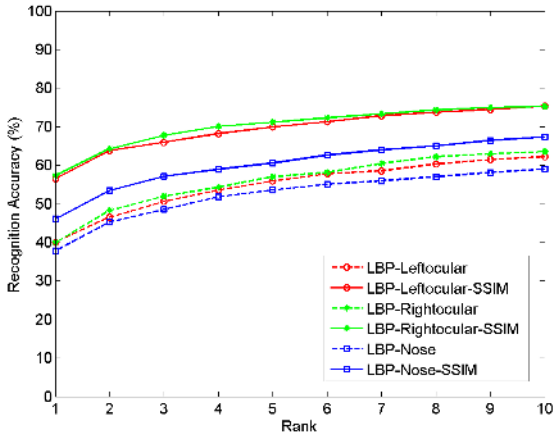

Figure 7. The CMC plots illustrating the performance of leftocular, rightocular and nose using LBP as features.

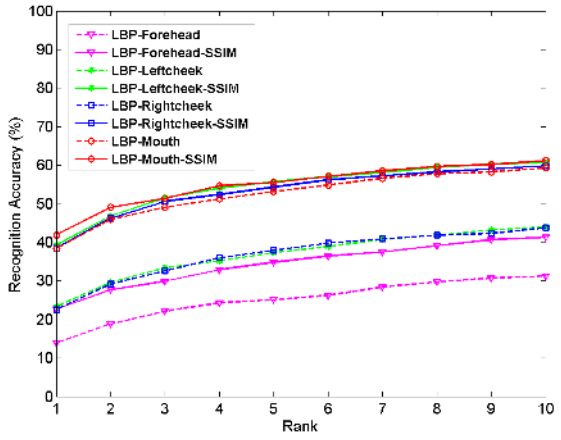

Figure 8. The CMC plots illustrating the performance of forehead, leftcheek, rightcheek and mouth using LBP as features.

matically locate 11 landmarks using STASM and construct 83 triangles using these 11 landmarks. For each triangle, we compute the radian values of the three angles. Further, we concatenate the three radian values of all the triangles to form a triangle descriptor representing the whole face geometric structure. Next the cosine similarity of two triangle descriptors from two faces are calculated for the geometric matching score. It is worth nothing that the triangle descriptor for each face is computed on the original image instead of the aligned one due to the scale-invariant property of congruent triangles. The 11 landmarks and the 83 triangles are illustrated in Figure 9. Finally, the rank-order list based fusion scheme proposed in [6] is employed to generate the final rank-order list by fusing the seven component classifiers and the triangle classifier. The weights for the seven component classifiers and the triangle classifiers are respectively: forehead (2); leftocular (4); rightocular (4); nose (2.5); leftcheek (3); rightcheek (3); mouth (3.5); triangle (4); The fusion results of component LBP features and triangle classifier as well as the result of using only the triangle classifier are illustrated in Figure 10. It can be observed that triangle descriptor shows some useful information. With SSIM index, a significant improvement is achieved in the component-wise manner. The Rank-1 accuracy increased from 69.13% to 77.55%.

Comparison with the state-of-the-art Note that our experiments are performed without exploiting any outside dataset for providing auxiliary information [1, 6]. Hence, it is more reasonable to compare our approach with those existing approaches which do not use any background information. Table 1 shows the comparison results. For each algorithm, we show not only the Rank-1 accuracy but also whether it employs training and feature fusion or not. Furthermore, the number of gallery subjects and probe images used in each algorithm is illustrated for a better comparison. In [4], 661 pre-surgery images from the plastic surgery database and 568 images from the FRGC database[3] are used to form the gallery set, while the query images are the corresponding 661 post-surgery images. Verilook 3.2 is a commercial software from Neurotechnology[4]. From these results, we can see the effectiveness of our approach for face recognition across plastic surgery.

Figure 9. The 11 landmarks and the 83 triangles for the calculation of the triangle descriptor.

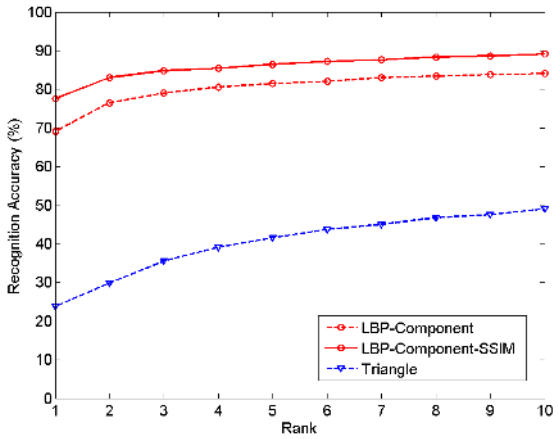

Figure 10. The CMC plots illustrating the performance using the component-wise manner and using solely the triangle descriptor.

6. Conclusions and Future Work

In this paper, we introduced SSIM index weighted multi-patch fusion to face recognition across plastic surgery. Experimental results on the plastic surgery database highlight the effectiveness of our approach. In comparison with the existing approaches, neither training process is needed nor any background information is required. The good potential of SSIM quality map for face recognition after plastic surgery can be seen through our work. However, there are

[3]NIST, Face Recognition Grand Challenge (FRGC) Database, http://www.nist.gov/itl/iad/ig/frgc.cfm.

[4]Verilook 3.2, Neurotechnology, http://www.neurotechnology.com/.

Algorithms	Rank-1	Training	Feature fusion	# of gallerys	# of probes
GNN [12, 11]	54.20%	yes	no	540	540
Verilook 3.2 [4]	73.90%	no	no	661PS + 568FRGC	661
Our approach (LBP)	**77.55%**	**no**	**no**	**784**	**784**
Evolutionary Granular [3]	78.61%	yes	yes	540	540
Verilook 3.2 + Ocular [4]	81.50%	no	yes	661PS + 568FRGC	661

Table 1. The Rank-1 accuracies of our approach and those of the existing approaches on the plastic surgery database.

still several issues deserving more study on this direction.

1) Demand for more discriminative weights. As discussed in Section 4, the SSIM weighting can induce lower weights for regions where two different faces differ most. Hence, it is important to change the SSIM weights to make those regions more discriminative. A function may be learned mapping the SSIM weights to more discriminative values.

2) Extraction of holistic geometric information. For some plastic surgeries modifying the skin texture such as skin resurfacing and rhytidectomy, the geometric structure of the whole face will be of great importance. For instance, in our approach, the use of the triangle descriptor might be improved by employing more landmarks located in the outline of the face.

Acknowledgments.

We want to thank the anonymous reviewers for their valuable comments and suggestions. This work has been partially supported by the European Commission COST Action IC1106.

References

[1] G. Aggarwal, S. Biswas, P. Flynn, and K. Bowyer. A sparse representation approach to face matching across plastic surgery. In *IEEE Workshop on Applications of Computer Vision*, pages 113–119, 12.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *European Conference on Computer Vision*, pages 469–481, 2004.

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

[4] R. Jillela and A. Ross. Mitigating effects of plastic surgery: fusing face and ocular biometrics. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 402–411, 2012.

[5] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image processing*, 11(4):467–476, 2002.

[6] X. Liu, S. Shan, and X. Chen. Face recognition after plastic surgery: a comprehensive study. In *Asian Conference on Computer Vision*, pages 565–576, 2012.

[7] M. D. Marsico, M. Nappi, D. Riccio, and H. Wechsler. Robust face recognition after plastic surgery using local region analysis. In *Image Analysis and Recognition*, pages 191–200, 2011.

[8] A. Merati, N. Poh, and J. Kittler. User-specific cohort selection and score normalization for biometric systems. *IEEE Transactions on Information Forensics and Security*, 7:1270–1277, 2012.

[9] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *European Conference on Computer Vision*, pages 504–513, 2008.

[10] U. Park, R. Jillela, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, 2011.

[11] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreyezdan. Plastic surgery: a new dimension to face recognition. *IEEE Transactions on Informtion Forensics and Security*, 5(3):441–448, 2010.

[12] R. Singh, M. Vatsa, and A. Noore. Face recognition with disguise and single gallery images. *Image and Vision Computing*, 27:245–257, 2009.

[13] Z. Wang, B. A.C, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[14] L. Wolf, T. Hassner, and Y. Taigman. Similarity scores based on background samples. In *Asian Conference on Computer Vision*, pages 88–97, 2009.