Face Recognition after Plastic Surgery: a Comprehensive Study

Xin Liu[1,2], Shiguang Shan[1], Xilin Chen[1]

[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
{xin.liu, shiguang.shan, xilin.chen}@vipl.ict.ac.cn;

Abstract. It has been observed that many face recognition algorithms fail to recognize faces after plastic surgery, which thus poses a new challenge to automatic face recognition. This paper first gives a comprehensive study on Face Recognition After Plastic Surgery (FRAPS), with careful analysis of the effects of plastic surgery on face appearance and its challenges to face recognition. Then, to address FRAPS problem, an ensemble of Gabor Patch classifiers via Rank-Order list Fusion (GPROF) is proposed, inspired by the assumption of the interior consistency of face components in terms of identity. On the face database of plastic surgery, GPROF achieves much higher face identification rate than the best known results in the literature. Furthermore, with our impressive results, we suggest that plastic surgery detection should be paid more attend to. To address this problem, a partial matching based plastic surgery detection algorithm is proposed, aiming to detect four distinct types of surgery, i.e., the eyelid surgery, nose surgery, forehead surgery and face lift surgery. Our experimental results demonstrate that plastic surgery detection is a nontrivial task, and thus deserves more research efforts.

1 Introduction

In recent years, plastic surgery has become popular worldwide. People take facial plastic surgery to correct feature defects or improve attractiveness and confidence. In South Korea, plastic surgery has become an important part of medical industry and 15 to 30 percent of Korean women are said having undergone a plastic surgery [1]. According to the statistics from American Society for Aesthetic Plastic Surgery [2], from 1997 to 2011, there has been over 197% increase in the total number of cosmetic procedures. The above statistical figures lead to a practical requirement on identity authentication after plastic surgery. Especially, for face-based biometrics, plastic surgery poses a great challenge, because not only local skin texture but also face components such as eyelid and nose might be disturbed or reshaped in plastic surgery. Even the holistic appearance of face may greatly change because of the global face plastic surgery such as face lift or skin peeling.

Face Recognition after Plastic Surgery (FRAPS) was first introduced by Singh et al. [3] and was stated as a new dimension to face recognition. In their paper, a face database of plastic surgery with 900 individuals was also publicly released. In the face identification test [3], various traditional methods like PCA [4], FDA [4], LFA [5] and some recently proposed methods like CLBP [6], SURF [7] and GNN [8] were evaluated on this database. Their evaluation results showed that all the above methods suffer a serious performance decline, indicating that off-the-shelf face recognition algorithms cannot handle FRAPS effectively.

To address the challenge of FRAPS, Bhatt et al. [9] proposed an evolutional granular approach. The proposed method extracts discriminating information from non-disjoint face granules using optimized feature extractor settings. Aggarwal et al. [10] proposed a component based sparse representation approach. The sparse reconstruction errors of all face components are summed to produce the final identification. Local region analysis [11] and multimodal biometric features [12] were also applied for FRAPS, but the performance is not very impressive.

Overall speaking, the related work on FRAPS is very limited, and the identification performance is still far from satisfactory. In this paper, we do a comprehensive study on FRAPS, with the following main contributions:

1. We propose an ensemble of Gabor Patch classifiers via Rank-Order list Fusion (GPROF) algorithm to handle FRAPS. GPROF is robust to both local and global facial plastic surgery. On the face database of plastic surgery, GPROF significantly outperforms state-of-the-art methods.

2. With the face database of plastic surgery in [3], we explore the possibility of plastic surgery detection (PSD) and further propose a partial matching based method for PSD. Given two face images of the same person, our method can automatically detect the eyelid surgery, nose surgery, forehead surgery and face lift surgery. To our best knowledge, this is the first work to explore the automatic plastic surgery detection. Moreover, it is also an initial attempt to discover the interior pattern of face appearance changes caused by plastic surgery.

The rest of the paper is organized as follows. Section 2 presents an insightful analysis of the difficulty of FRAPS. Section 3 describes the proposed GPROF in detail, as well as the experimental evaluations. The partial matching based plastic surgery detection algorithm is then presented in section 4, followed by conclusions and future work in section 5.

2 The challenges of FRAPS

In this section, we review the face database of plastic surgery [3] and analyze the challenges of FRAPS. Our aim is to give a comprehensive understanding of the nature of FRAPS.

2.1 Review of face database of plastic surgery

The face database of plastic surgery introduced in [3] is the first public database to research the influence of plastic surgery on automatic face recognition. Table. 1

Table 1. Overview of the face database of plastic surgery.

Type	Plastic Surgery Procedure	Number of Individuals
	Dermabrasion	32
	Brown lift(Forehead surgery)	60
Local	Otoplasty(Ear surgery)	74
	Blepharoplasty(Eyelid surgery)	105
	Rhinoplasty(Nose surgery)	192
	Others	56
Global	Skin peeling	73
	Face lift	308

gives the details of this database. Briefly speaking, there are 900 individuals in this distinct face database of plastic surgery. Each individual has one pre-surgery and one post-surgery face image. All these face images are in frontal view, under proper illumination and with neutral expression. Fig. 1 shows example pre-surgery and post-surgery images of four subjects selected from the database.

Fig. 1. Selected representative subjects from the face database of plastic surgery. Upper row: pre-surgery face images; Lower row: corresponding post-surgery face images. From left to right, the faces had undergone four kinds of plastic surgery: eyelid surgery, nose surgery, skin peeling and face lift.

2.2 What is the challenges of FRAPS?

Facial plastic surgery changes face appearance, which intuitively affects the robustness of appearance based face recognition. In this section, we analyze the effects of different plastic surgery procedures on face appearance.

Changes in skin texture. Some plastic surgery makes people look younger or more attractive by removing face scars, acnes or taking skin resurfacing. As a result, the skin texture will change.

Changes of face component. The main face components: forehead, eyelid, nose, lip, chin and ear can be reshaped or restructured by plastic surgery. The local skin texture around the face component may also be disturbed.

Changes of global face appearance. Global facial plastic surgery will change the global face appearance, in other words, not only part of the face component and the skin texture will change, but also the whole face geometric structure and appearance will be disturbed.

Table. 2 illustrates how different plastic surgery procedures change face appearance. To our best knowledge, no techniques are available yet to automatically detect the face appearance changes caused by plastic surgery.

In summary, the challenges of FRAPS mainly lie in the fact that faces after plastic surgery have undergone various appearance changes, but no method is available to detect or model such changes. Therefore, the key to handle FRAPS is designing a plastic surgery robust face representation.

Table 2. Effects of different plastic surgery procedures on face appearance.

Type	Plastic Surgery Procedure	Effects on Face Apperance
	Dermabrasion	Local skin texture
	Brown lift(Forehead surgery)	Face component : forehead
Local	Otoplasty(Ear surgery)	Face component : ear
	Blepharoplasty(Eyelid surgery)	Face component : eyelid
	Rhinoplasty(Nose surgery)	Face component : nose
	Others	Local skin details or face component
Global	Skin peeling	Global skin texture
	Face lift	Global face structures and skin texture

3 Ensemble of Gabor Patch Classifiers via Rank-Order List Fusion

Driven by the discussion in 2.2, we focus on finding a plastic surgery robust face representation. Human face has interior consistency, i.e., the identity information embedded in all the face components are consistent. For example, given two face images A and B of the same person, A's eyes and nose are internally look alike those of B simultaneously. As local plastic surgery only changes local face appearance, we can safely assume that it does not corrupt the interior consistency of a face. We also assume that global plastic surgery does not completely change all facial components' identity information, thus the interior consistency of face after global plastic surgery can also be reserved. Based on the above assumptions, we propose a new face recognition method, named Ensemble of Gabor Patch Classifiers via Rank-Order List Fusion. The general idea is dividing face into patches, designing one component classifier for each patch, and finally fusing the rank-order list of each component classifier.

In the proposed method, Gabor feature together with Fisher Linear Discriminant Analysis is exploited to design the component classifier. The motivation that we choose Gabor feature is as follows : Gabor feature is a bio-inspired local feature extracting multi-scale and multi-orientation local texture features. In the past 10 years, Gabor feature has been proved as one of the most efficient local descriptor for face recognition [13, 14], so we think Gabor feature is desirable evidence for the interior consistency of human face robust to plastic surgery. Furthermore, to improve the discriminative capacity of Gabor feature, we apply PCA [15] and Fisher LDA [4].

In GPROF, we make use of rank-order list to represent each face patch. The application of rank-order information has been studied in the face tagging [16] and face matching scenarios [17]. It has been proved that rank-order list is a robust description of face identity. Because of the interior consistency of face, different patches of the same input face should have similar rank-order lists against a gallery set, which is demonstrated by the same identity marked with red rectangles in the eight rank-order lists of the face with plastic surgery in Fig. 2. As a result, GPROF fuse each patch's rank-order list to compensate for the appearance changes caused by plastic surgery.

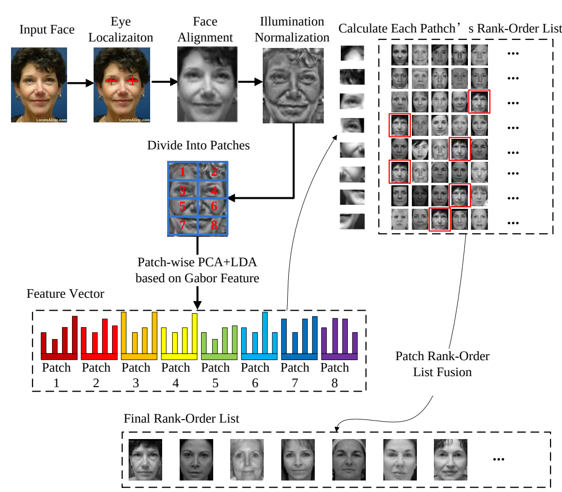

Fig. 2. Framework of the proposed GPROF.

3.1 Proposed method

The framework of GPROF is shown in Fig. 2, in which the main steps are described as follows:

1. Localize the eye centers, align the face and normalize it to 64×80 pixels.
2. Illumination preprocess with local normalization [18].
3. Divide the face into 2×4 patches of equal size. In this study, the size of each patch is 16×20 pixels without overlapping.
4. For each patch, extract Gabor magnitude features in five scale and eight orientations at each face image pixel, and use PCA and LDA to project the original Gabor features into a low-dimension discriminative subspace. Each patch is finally represented by a normalized feature vector of equal dimension.
5. Calculate each patch's rank-order list against the gallery using cosine similarity.
6. Fuse the rank-order lists of all patches via Algorithm 1, to generate the final rank-order list for the whole face.

A rank-order list is in fact a permutations of the identity labels in the gallery set, which can be seen as a signature representation for face patch. The patch rank-order list fusion algorithm is based on a voting strategy, and we put more weight on identity in the front of the rank-order list. More details can be seen in Algorithm 1.

Algorithm 1 Patch Rank-Order List Fusion.

Input:

Patch number, N;
Rank-Order list of each patch, $R_n(n = 1, 2, ..., N)$;
Patch weight, $P_n(n = 1, 2, ..., N)$;
Patch rank list length, L;
Rank list weight function $W(i)(i = 1, 2, ..., L)$.

Output:

The final Rank-Order list, FR;

```
For each rank-order list R_n : R_{n,1}, R_{n,2}, ..., R_{n,L};
Let ID(R_{n,j}) denotes the ID of rank j in rank list R_n;
Initialize an all zero rank-order list FR with the length equal to the number of
  identities in Gallery set;
for patch = 1; patch ≤ N; patch++ do
    for rank = 1; rank ≤ L; rank++ do
        FR(ID(R_{patch,rank})) += W(rank) * P_{patch}
    end for
end for
Sort FR descend;
return FR.
```

3.2 Experimental evaluation and results analysis

In this section, we evaluate the proposed GPROF on the face database of plastic surgery [3]. Comparisons with existing results on the same database are reported. We also evaluate the effects of different plastic surgery procedures on identification accuracy.

Parameter setting: empirical patch weights ={1.0, 1.0, 3.0, 3,0, 3.0, 3.0, 1.5, 1.5}, which emphasize the patches around eyes and nose. The rank list weight function is a Quadratic function:

$$W(i) = (L - i + 1)^2/L^2; \tag{1}$$

In our experiments, we empirically set L to 50. The PCA and LDA models are trained on FRGC ver2.0 [19] Exp4's training set. Dimension of PCA model is set to 600. Dimension of LDA model is set to 200. It should be noticed that the PCA and LDA models are not trained on the plastic surgery database itself. The reasons behind are: 1)the plastic surgery database itself is of small scale in terms of both subjects and images; 2)it is also desirable to evaluated how PCA/LDA generalizes across different face database.

Table 3. Rank-1 identification rates of our method and those in literature on the face database of plastic surgery.

Method	Rank-1 Identification Rate
Our Method	**86.11%**
Result of [9]	78.61%
Result of [10]	77.90%
Result of [11]	70.00%
Result of [12]	74.40%
Best result of [3] (GNN)	53.70%

Table 4. Effects of different plastic surgery procedures on rank-1 identification rates for our method and previous methods.

Type	Surgery	Our Method	PCA	FDA	LFA	CLBP	SURF	GNN
	Dermabrasion	**81.25%**	20.2%	23.4%	25.5%	42.1%	42.6%	43.8%
	Forehead surgery	**86.67%**	28.5%	31.8%	39.6%	49.1%	51.1%	57.2%
Local	Ear surgery	**90.50%**	56.4%	58.1%	60.7%	68.8%	66.4%	70.5%
	Eyelid surgery	**89.52%**	28.3%	35.0%	40.2%	52.1%	53.9%	61.4%
	Nose surgery	**81.77%**	23.1%	24.1%	35.4%	44.8%	51.5%	54.3%
	Others	**73.21%**	26.4%	33.1%	41.4%	52.4%	62.6%	58.9%
Global	Skin peeling	**97.26%**	25.2%	31.5%	40.3%	53.7%	51.1%	53.9%
	Face lift	**86.68%**	18.6%	20.0%	21.6%	40.9%	40.3%	42.1%
	Overall	**86.11%**	27.2%	31.4%	37.8%	47.8%	50.9%	53.7%

Experimental evaluation: Table. 3 shows the comparison of our method with the related work in [3, 9–12]. These results are directly taken from the corresponding original papers to facilitate the comparison. However, we must pointed out that the results of [9, 10] are from testing on 60% of the whole database, while those of [3, 11, 12] and ours are from testing on the whole database.

Furthermore, the Cumulative Matching Curves(CMC) plots of GPROF and methods in [3] are shown in Fig. 3. Evidently, GPROF impressively outperforms both the state-of-the-art methods. Its rank-5 identification rate is as high as 95.67%, which should be a satisfactory performance for practical application.

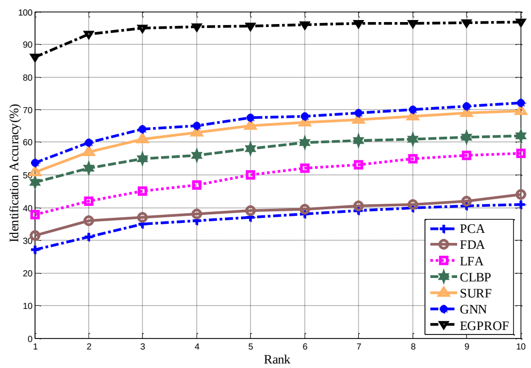

Fig. 3. CMC plots of proposed GPROF and baseline method in [3].

In Table. 4, we also report the effects of different kinds of plastic surgery procedures on rank-1 identification rates. The comparison methods are implemented by [3]. It is important to note that our method does not have a distinguished performance difference between local and global plastic surgery, which means that our assumption of the interior consistency of faces with both local and global plastic surgery is credible. The best identification performance is 97.26% on skin peeling, which is a global plastic surgery.

Overall speaking, the performance of GPROF is very impressive on the face database of plastic surgery, which is a great promotion for the research on FRAPS.

4 Plastic Surgery Detection: problem and a baseline

With the impressive identification performance of GPROF, we further explore how to detect plastic surgery. In this section, we propose a partial matching based algorithm for Plastic Surgery Detection (PSD), as an initial attempt and baseline to discover the interior pattern of appearance changes caused by plastic surgery.

4.1 Algorithm for PSD

As we have discussed, plastic surgery changes face appearance in various ways (See Table. 2). An intuitive way to detect plastic surgery is to directly match the corresponding face patches around the possible plastic surgery regions. By a component relative threshold setting, the plastic surgery region can be detected. Based on the discussion of the effects of plastic surgery on face appearance, we can reversely infer the candidate plastic surgery procedure.

Firstly, we decompose the face into possible plastic surgery regions. The decomposition of the face is shown in Fig. 4.

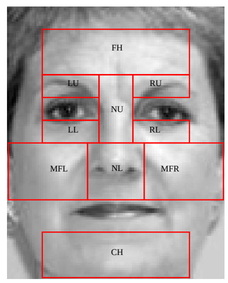

Fig. 4. Face decomposition of possible plastic surgery components. Face is decomposed into five possible plastic surgery components: Forehead : composed only one subregion FH; Eyelid, composed by LU, LU, RU, RL four subregions; Nose, composed by NU and NL two subregions; Middle Face: composed by MFL and MFR two subregions; Chin : composed by only one subregion CH.

Feature extraction and distance metric. The uniform LBP feature [20] is extracted on the five face components. For eyelid, nose and middle face, uniform LBP feature is extracted in each subregions respectively. After extracting uniform LBP feature, each subregion is represented by a 59-dimension normalized histogram.

For forehead and chin component that only have one subregion, distance is measured by Chi-square distance of the normalized LBP histogram directly. As for nose, eyelid and middle face component, the maximum Chi-square distance between corresponding subregions is reserved. We use the eye component as an example. Assume E_i and E_j are two eye components:

$$D(E_i, E_j) = Max\{d(LU_i, LU_j), d(LL_i, LL_j), d(RU_i, RU_j), d(RL_i, RL_j)\}; \quad (2)$$

For the global surgery face lift, more than one face component may be changed. For simplicity, we sum the distance of five corresponding face components. Assume F_i and F_j are two face images of the same person:

$$D(F_i, F_j) = D_{nose} + D_{forehead} + D_{eyelid} + D_{middlface} + D_{chin}; \quad (3)$$

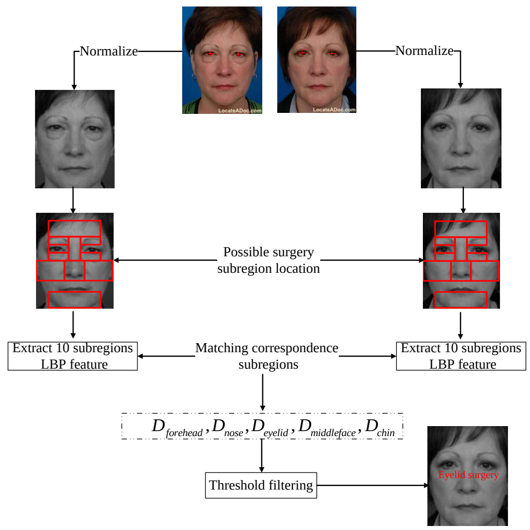

Fig. 5. PSD: a partial matching based approach

PSD approach. The PSD approach based on partial matching is shown in Fig. 5. After eye localization and normalization, face is decomposed to 10 possible plastic surgery interest subregions. The Chi-square distance of 10 corresponding subregions' normalized LBP histograms are firstly calculated, then we can get the distance of five corresponding face components. After a component relative threshold filtering, the possible plastic surgery face components are detected.

4.2 Experimental evaluation

Our method is tested on the face database of plastic surgery. We use the FRGC ver2.0 [19] Exp4's target set as the non-surgery set, because faces in this set are collected in similar imaging environment with the face database of plastic surgery. We randomly choose 10,000 within class pairs.

The performance of PSD is measured by the detection rate at given false acceptance rate. The ROC curves are shown by Fig. 6. The best performance is achieved on eyelid surgery detection. One possible interpretation is that eyelid surgery causes more distinguished appearance difference.

The experimental results show that the PSD is a nontrivial task. In real world environment, there are many factors lead to face appearance changes, such as illumination, occlusion, pose, expression variant, aging, etc. At the same time, our method is restricted to four basic plastic surgery procedures and cannot interpret various complex or minor appearance changes. Objectively speaking, our work is a previous study on PSD and provides a baseline performance. We need to study the interior pattern of different kinds of plastic surgery procedures and model the appearance changes caused by plastic surgery, and it will be the theoretical basis for both FRAPS and PSD.

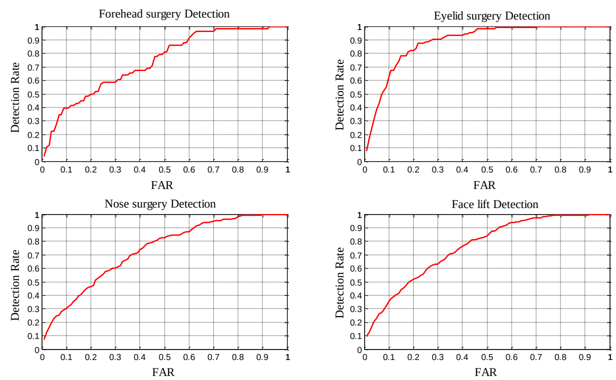

Fig. 6. Roc curves for PSD

5 Conclusion and future work

This paper gives a comprehensive study on FRAPS. The proposed GPROF to handle FRAPS makes use of each patch's rank-order information and proves to be efficient for FRAPS. Moreover, PSD is explored the first time. Experimental results show that PSD is a nontrivial task, and more research efforts should be put on discovering the interior pattern of plastic surgery procedures.

As for future work, a much bigger face database of plastic surgery need to be gathered, and the number of face images for different plastic surgery procedures should be balanced. Moreover, we believe that the key step to push forward the research on FRAPS is to model the face appearance changes caused by plastic surgery.

6 Acknowledgements

This work is partially supported by Natural Science Foundation of China (NSFC) under contract Nos. 61173065, and U0835005; and Beijing Natural Science Foundation (New Technologies and Methods in Intelligent Video Surveillance for Public Security) under contract No. 4111003.

References

1. Asian Plastic Surgery Guide: Cosmetic plastic surgery in south korea. http://www.asianplasticsurgeryguide.com/korea/a_korean.html (2012)
2. American Society for Aesthetic Plastic Surgery: 2011 ASAPS statistics: Complete charts [including national totals, percent of change, gender distribution, age distribution, national average fees, economic, regional and ethnic information]. http://www.surgery.org/media/statistics (2012)
3. Singh, R., Vatsa, M., Bhatt, H., Bharadwaj, S., Noore, A., Nooreyezdan, S.: Plastic surgery: a new dimension to face recognition. Information Forensics and Security, IEEE Transactions on **5** (2010) 441–448

4. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions on **19** (1997) 711–720
5. Penev, P., Atick, J.: Local feature analysis: A general statistical theory for object representation. Network: computation in neural systems **7** (1996) 477–500
6. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006) 2037–2041
7. Dreuw, P., Steingrube, P., Hanselmann, H., Ney, H.: Surf-face: Face recognition under viewpoint consistency constraints. In: British Machine Vision Conference. Volume 2. (2009) 7
8. Singh, R., Vatsa, M., Noore, A.: Face recognition with disguise and single gallery images. Image and Vision Computing **27** (2009) 245–257
9. Bhatt, H., Bharadwaj, S., Singh, R., Vatsa, M., Noore, A.: Evolutionary granular approach for recognizing faces altered due to plastic surgery. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE (2011) 720–725
10. Aggarwal, G., Biswas, S., Flynn, P., Bowyer, K.: A sparse representation approach to face matching across plastic surgery. IEEE Workshop on Applications of Computer Vision (2012)
11. De Marsico, M., Nappi, M., Riccio, D., Wechsler, H.: Robust face recognition after plastic surgery using local region analysis. Image Analysis and Recognition (2011) 191–200
12. Lakshmiprabha, N., Bhattacharya, J., Majumder, S.: Face recognition using multimodal biometric features. In: Image Information Processing (ICIIP), 2011 International Conference on, IEEE (2011) 1–6
13. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. Image Processing, IEEE Transactions on **11** (2002) 467–476
14. ngel Serrano, de Diego, I.M., Conde, C., Cabello, E.: Recent advances in face biometrics with gabor wavelets: A review. Pattern Recognition Letters **31** (2010) 372 – 381
15. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE (1991) 586–591
16. Zhu, C., Wen, F., Sun, J.: A rank-order distance based clustering algorithm for face tagging. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 481–488
17. Schroff, F., Treibitz, T., Kriegman, D., Belongie, S.: Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 2494–2501
18. Xie, X., Lam, K.: An efficient illumination normalization method for face recognition. Pattern Recognition Letters **27** (2006) 609–617
19. 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: Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on. Volume 1., IEEE (2005) 947–954
20. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 971–987