Pore-scale facial features matching under 3D morphable model constraint

Xianxian Zeng[1], Dong Li[1⋆], Yun Zhang[1], and Kin-Man Lam[2]

[1] Automation, Guangdong University of Technology, Guangzhou, Guangdong, China
[2] Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, HongKong, China

Abstract. Similar to iris and fingerprint, pore-scale facial features are effective features that can distinguish human identities. Recently, the local feature extraction based on deep network architecture has been proposed, which needs the large dataset to train. However, there are no large databases for the area of the pore-scale facial features. Actually, it is hard to set up a large pore-scale facial features dataset because the images of high resolution face databases are uncalibrated and non-synchronous and the human faces are nonrigid. To solve this problem, we propose a method to establish a large pore-to-pore correspondence dataset. we adopt Pore Scale-Invariant Feature Transform (PSIFT) to extract the pore-scale facial feature, and use 3D Dense Face Alignment (3DDFA) to get the fitted 3D morphable model to be a constraint of keypoints matching. From our experiment, a large pore-to-pore correspondence dataset, including 17136 classes of matched pore pairs, is established.

Keywords: Pore-scale facial features, Dataset, PSIFT, 3D morphable model, 3DDFA

1 Introduction

Pore-scale facial features include pores, fine wrinkles, and hair, which commonly appear in the whole face region. Pore-scale facial features, which are similar to the iris and fingerprint, are one of the effective features that can distinguish human identities. Recently, the local feature extraction based on deep network architecture [1], whose name is Learned Invariant Feature Transform (LIFT), has been proposed. LIFT is a deep network architecture that implement the full feature point handling pipeline, that is, detection, orientation estimation, and feature description. If LIFT is trained under a large and accuracy dataset, it can perform better than the state-of-the-art methods of the feature extraction.

[⋆] Corresponding author
Telephone: +8613226681205
This work was supported by National Natural Science Foundation of China: 61503084, U1501251 and Natural Science Foundation of Guangdong Province, China: 2016A030310348

This inspires us that maybe we can get a good pore-scale feature extraction if the LIFT is trained under the large pore-scale facial features dataset. However, there are no large and open databases for the area of the pore-scale facial features. Therefore, we should establish a large pore-to-pore correspondence dataset, firstly.

Actually, it is hard to set up a large pore-to-pore correspondence dataset, because the images of high resolution face databases are uncalibrated and nonsynchronous. Besides, human faces are nonrigid. All of this make pore-scale feature matching difficult. To the best of our knowledge, only a few studies have been reported in the literature that attempt to set up a pore-to-pore correspondences dataset using uncalibrated face images. Lin et al. [2] employed the SURF features [3] on facial images with viewpoints 45°apart, which typically obtained no more than 10 inliers (i.e. correctly matched keypoint pairs) out of a total of 30 matched candidates in 3 poses. Li et al. [4] propose a new framework, whose name is Pore Scale-Invariant Feature Transform (PSIFT), to achieve the pore-scale feature extraction, and then to generate a pore-to-pore correspondence dataset, including about 4240 classes of matched pore pairs. PSIFT is a feature that can describe the human pore patches distinctively. However, human face is symmetry, and PSIFT will get some outliers. For this problem, Li [4] use the RANSAC (Random SAmple Consensus) [14] method to get the inliers, which will reduce the number of matched keypoints. We find that the RANSAC algorithm can't perform well, if the object is nonrigid. Therefore, the method of Li [4] will get rid of many matched keypoints of facial region. In our opinion, one of the way to establish a larger pore-to-pore correspondence dataset is finding a new constraint which can perform well of the pore-scale feature matching.

Furthermore, some scholars solve the face alignment problem with a 3D solution. Blanz et al. [11] propose a standard 3D morphable model (3DMM), and Zhu et al. [10] present a neural network structure to fit the 3D morphable model to the image. we are inspired by the 3DDFA algorithm that we can use the fitted 3D morphable model to constrain the keypoints matching. To the best of our knowledge, 3D model constraint is one of the most effective constraint of the keypoints matching, especially for the matching of a large baseline. The proposed method framework is shown in Fig. 1. In summary, our contributions are:

1. We propose the 3D morphable model constraint, which can improve the matching accuracy.
2. Our proposed methods can establish a large number of correspondences between uncalibrated face images of the same person using the pore-scale features, which leads to many potential applications. Our work shows a way to merge face-based approaches and general computer-vision approaches.
3. Based on our framework, a pore-to-pore correspondences dataset containing 17136 classes of matched pore pairs is established by the same pore keypoints from 4 face images of the same subject with different poses.

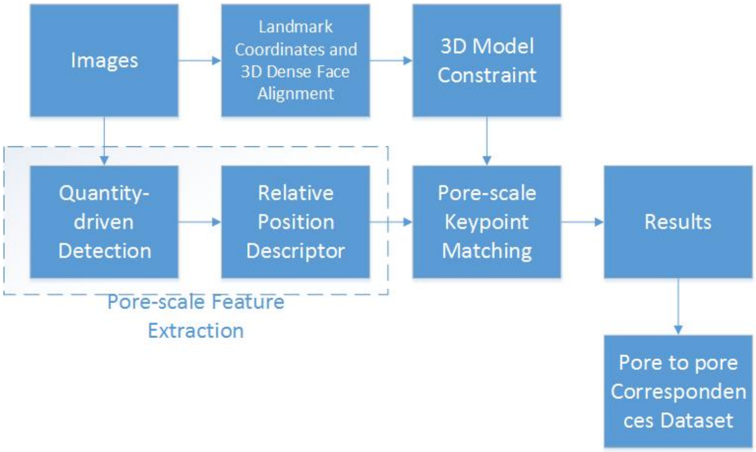

Images
Landmark Coordinates and 3D Dense Face Alignment
3D Model Constraint
Quantity-driven Detection
Relative Position Descriptor
Pore-scale Keypoint Matching
Results
Pore-scale Feature Extraction
Pore to pore Corresponden ces Dataset

Fig. 1. The structure of the proposed framework.

1 Pore-Scale Invariant Feature Transform

Pore Scale-Invariant Feature Transform (PSIFT) [4] is a modified algorithm of SIFT [9], which can generate the pore-scale feature. And the details of the PSIFT will be introduced as follow.

1.1 Pore-scale feature detection

Pore-scale facial features, such as pores and fine wrinkles, are darker than their surroundings in a skin region. Therefore, PSIFT [4] apply the DoG detector for keypoint detection on multi-scales, which is shown as follow.

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y), \quad (1)$$

where the scale space of an image $L(x,y,\sigma)$ is the convolution of the image I(x,y) and the Gaussian kernel

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} exp(\frac{-(x^2+y^2)}{2\sigma^2}), \quad (2)$$

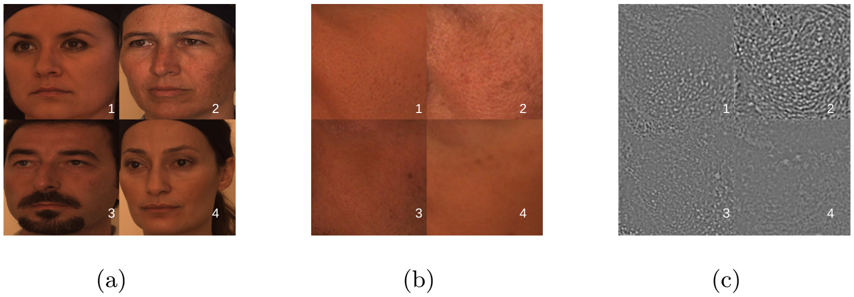
PSIFT construct the DoG in octaves, which have the σ doubled in the scale space. Li [4] find that the PSIFT detector only need the maxima of the DoG to locate the darker keypoints in face regions, and the example is shown in Fig. 2(c). Besides, a blob-shaped pore-scale keypoint is a small, darker point due to its small concavity, where incident light is likely blocked, so PSIFT model the blob-shaped skin pores using a Gaussian function, as follow:

$$pore(x,y,\sigma) = 1 - 2\pi\sigma^2 G(x,y,\sigma). \quad (3)$$

where σ is the scale of the pore model. Then, the DoG response of a pore, denoted as D_{pore}, can be computed as follows:

$$D_{pore}(x, y, \sigma_1, \sigma_2) = [G(x, y, k\sigma_1) - G(x, y, \sigma_1)] * pore(x, y, \sigma_2), \qquad (4)$$

and the pore-scale keypoints is the maxima of the D_{pore}.



(a) (b) (c)

Fig. 2. (a) Four face images with different skin conditions from the Bosphorus face database, (b) local skin-texture images, (c) the DoG of the local skin-texture image.

1.2 Pore-scale feature descriptor

The local PSIFT descriptor, which is adapted from SIFT to extract the relative-position information about neighboring pores. The keypoints from two facial-skin regions can be matched by using the PSIFT descriptor. Fig. 2 shows some sample results of DoG layer. The lighter points on the DoG, as shown in Fig. 2(c), represent the responses of the feature points. These points are very similar to each other: most of them are blob-shaped, and the surrounding region of the keypoints have almost the same color. However, the relative positions of the pores are unique. Therefore, the descriptor should extract not only the information around the keypoints, but also the information of a neighborhood wide enough to include the neighboring pore-scale features. And both the number of subregions and the support size of these subregions used in the SIFT descriptor are enlarged. Besides, Li [4] find that the keypoints are not assigned a main orientation, because most of the keypoints do not have a coherent orientation. Some parameters of the PSIFT and SIFT descriptors are shown in Table 1.

2 Matching with 3D morphable model constraint

In order to achieve a more efficient and accurate matching, we present a method of completing the local PSIFT feature matching by using the 3D model constraint. And the details will be introduced as follow.

Table 1. Different parameters of the PSIFT and SIFT descriptors

Parameters	PSIFT	SIFT
No. of subregions	8×8	4×4
Support size of each subregion	6× scale of keypoints	3× scale of keypoints
Support size of total subregion	48× scale of keypoints	12× scale of keypoints
Dimension of the feature	512	128

2.1 3D morphable model

Blanz et al [11] propose the 3D morphable model (3DMM) which describes the 3D face space with PCA:

$$S = \bar{S} + A_{id}\alpha_i d + A_{exp}\alpha_{exp}, \tag{5}$$

where S is a 3D face, $\bar{S}$ is the mean shape, A_{id} is the principle axes trained on the 3D face scans with neutral expression and α_{id} is the shape parameter, A_{exp} is the principle axes trained on the offsets between expression scans and α_{exp} is the expression parameter. For this, the A_{id} and A_{exp} come from BFM [12] and Face-Warehouse [13] respectively. The 3D face is then projected onto the image plane with Weak Perspective Projection:

$$V(p) = f * Pr * R * (\bar{S} + A_{id}\alpha_i d + A_{exp}\alpha_{exp}) + t_2 d, \tag{6}$$

where $V(p)$ is the model construction and projection function, leading to the 2D positions of model vertexes, f is the scale factor, Pr is the orthographic projection matrix $Pr = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right)$, R is the rotation matrix constructed form rotation angles $pitch$, yaw, $roll$ and t_{2d} is the translation vector. The collection of all the model parameters is $p = [f, pitch, yaw, roll, t_{2d}, \alpha_{id}, \alpha_{exp}]^T$.

2.2 3D Dense Face Alignment

Zhu et al [10] present a network structure, whose name is 3D Dense Face Alignment (3DDFA), to compute the collection p. The purpose of 3D face alignment is estimating p from a single face image $\mathbf{I}$. 3DDFA [10] employ a unified network structure across the cascade and construct a specially designed feature PNCC (Projected Normalized Coordinate Code). In summary, at iteration k (k=0,1,...,K), given an initial parameter p^k, 3DDFA construct the PNCC with p^k and train a convolutional neutral network Net^k to predict the parameter update Δp^k:

$$\Delta p^k = Net^k(\mathbf{I}, PNCC(p^k)), \tag{7}$$

After that, a better parameter $p^{k+1} = p^k + \Delta p^k$ becomes the input of the next network Net^{k+1} which has the same structure as Net^k. The input is the $100 \times 100 \times 3$ color image of PNCC. The network contains four convolution layers, three pooling layers and two fully connected layers. The network structure is shown as Fig. 3. The output is a 234-dimensional parameter update including 6-dimensional pose parameters $[f, pitch, yaw, roll, t_{2dx}, t_{2dy}]$, 199-dimensional shape parameters α_{id} and 29-dimensional expression parameters α_{exp}. And the result of the 3DDFA of iteration 3 is shown in Fig. 4.

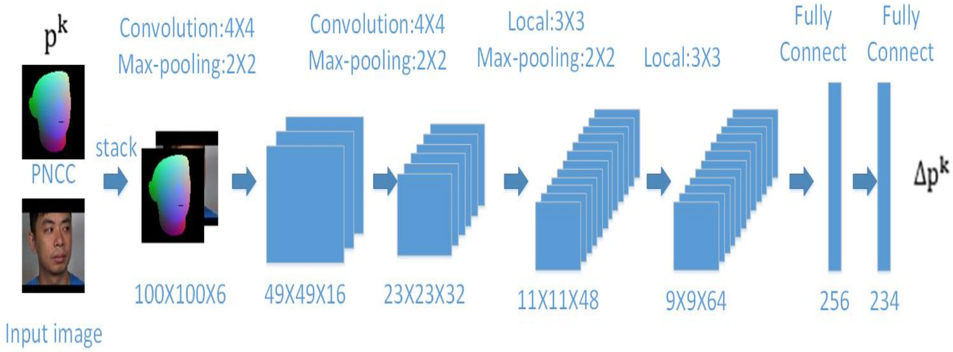
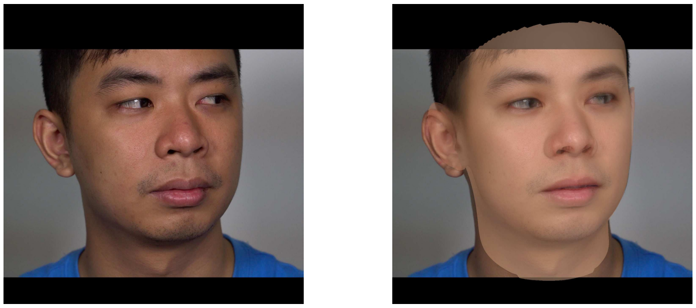


Fig. 3. An overview of 3DDFA.



(a)

(b)

Fig. 4. (a) The original image (b) the image with 3D model projection

2.3 3D morphable model constraint

The pore keypoints are one of the point of the face of the image. So we can write the equations of the probe image and the gallery image from Eq. 6.

$$V_p(pore) = f_p * Pr * R_p * (\bar{S}_p(pore) + A_{id}\alpha_{id_p} + A_{exp}\alpha_{exp_p}) + t_{2d_p}, \quad (8)$$

$$V_g(pore) = f_g * Pr * R_g * (\bar{S}_g(pore) + A_{id}\alpha_{id_g} + A_{exp}\alpha_{exp_g}) + t_{2d_g}, \tag{9}$$

where $\bar{S}_p(pore)$ and $\bar{S}_g(pore)$ are the 3D location of the pore of the mean shape. From Eq. 8 and Eq. 9, we assume that if the pore keypoint of the probe image and the pore keypoint of the gallery image are the same pore patch of the face, then $Err_{3d} = ||\bar{S}_g(pore) - \bar{S}_p(pore)||_2$ approximate to 0. And we can compute that:

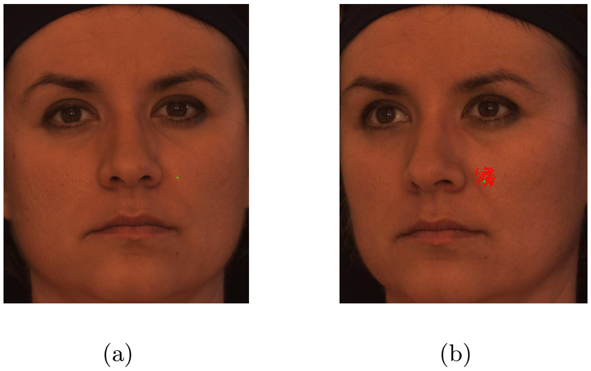
$$V_{pg}(pore) = f_g * Pr * R_g * (\bar{S}_p(pore) + A_{id}\alpha_{id_g} + A_{exp}\alpha_{exp_g}) + t_{2d_g} \tag{10}$$

$$Err_2d = ||V_{pg}(pore) - V_g(pore)||_2 < range, \tag{11}$$

where f_g, $R_(g)$, $\bar{S}_p(pore)$, α_{id_g}, α_{exp_g} and t_{2d_g} can be computed from 3DDFA. It means that if $range$ is set correctly and the same pore patch can be detected in probe image and gallery image, the Eq. 11 will be true. Then, we can only compute the nearest neighbor rate of the neighbor feature of the $V_{pg}(pore)$. If the rate less than the threshold, the matched keypoint between the probe and gallery image will be find. The estimation of the positions of the keypoints to be matched based on facial features is summarized in Algorithm 1.

Algorithm 1 Example of PSIFT improvement by using 3D model constraint

```
Given two images $I_1$, and $I_2$, we assume that there are $N_{k1}$ and $N_{k2}$ keypoints detected in $I_1$ and $I_2$, respectively. The coordinates of the $i$-th keypoint in $I_1$ are denoted as $(x_1^i, y_1^i)$. Similarly, the coordinates of the $j$-th keypoint in $I_2$ are denoted as $(x_2^j, y_2^j)$;
Using the 3DDFA method to find the best parameters $p_1$ and $p_2$ of $I_1$ and $I_2$, respectively. Afterwards we adopt Z-Buffer to project the $\bar{S}_1$ and $\bar{S}_2$ to $I_1$ and $I_2$, and denote them as $Z(\bar{S}_1)$ and $Z(\bar{S}_2)$, respectively;
Without loss of generality, assume that $N_{k1} < N_{k2}$. Matching the keypoints is established from $I_2$ to $I_1$;
for the $j$-th keypoint in $I_2$ do
    compute the $V_{pg}(j) = f_g * Pr * R_g * (Z(\bar{S}_2(j)) + A_{id}\alpha_{id_g} + A_{exp}\alpha_{exp_g}) + t_{2d_g}$;
    initial a list L includes all the keypoints in $I_1$;
    for the $i$-th keypoint in $I_1$ do
        if $||V_{pg}(j) - (x_1^i, y_1^i)|| < range$ then
            L is not updated;
        else
            remove the $i$-th keypoint from the list L of $I_1$;
        end if;
    end for;
    if the distance ratio based on the reduced list L is smaller than the threshold $\delta$, which is a constant between 0.8 and 0.9 then
        a match is established;
    end if;
end for;
```



(a) (b)

Fig. 5. (a) The neutral pose (b) the yaw rotation of 10°. The red points of (b) is the neighbor keypoints of the $V_{pg}(pore)$, and the green point of (b) is the matching point of (a)

In this paper, we don't need RANSAC [14] to identify those inliers, because the 3D morphable model constraint can get the inliers accurately and save more matched keypoints, and some example will be shown in Fig. 5. From Fig. 5, the green point of (a) is one of the pore keypoints. And the red points of (b) are the neighbor of the green point of (a) by using the 3D model constraint. Besides, the green point of (b) is the matched pore keypoint of the green point of (a).

3 Experiment

In this section, we will evaluate the performances of our proposed method in terms of accuracy for pore matching. The face images used in the experiments are the original size in Bosphorus database [15].

3.1 Skin matching based on the Bosphorus dataset

In this section, we estimate the performance of each stage of our algorithm in terms of skin matching. We use 105 skin-region pairs cropped from 420 face images, which were captured at 10°, 20°, 30°and 45°to the right of the frontal view in the Bosphorus database, as shown in the Fig. 2 and Fig. 6. Considering the fact that the dataset is uncalibrated and unsynchronized, Li [4] set the distance threshold used in RANSAC at 0.0005, so he can get some limited number of accurate matching. Conversely, our method uses 3D model constraint, so we can get more matched keypoints than the method of Li [4]. Table 2 illustrates the number of inliers of the methods. From the Table 2, it shows that our method can get more matched keypoints, so we can use this method to generate a larger pore-to-pore dataset.

Table 2. Skin matching results

Method	Avg. no. of inliers	total inliers
PSIFT + RANSAC	40.4	4240
PSIFT + 3D model constraint	163.2	17136

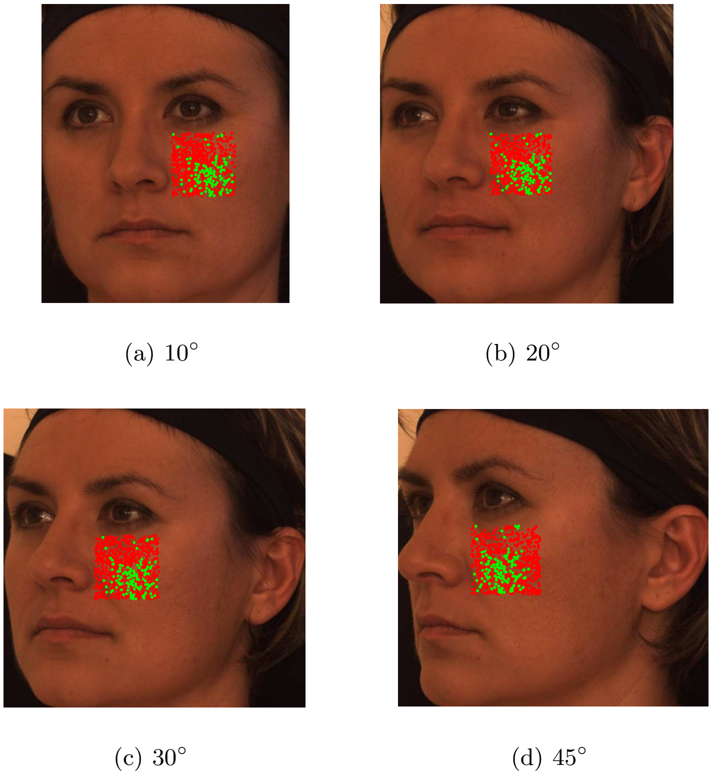
3.2 Pore-to-pore correspondences dataset

With the improvement of PSIFT, we also have built a larger pore-to-pore correspondences dataset so that the learning for pore-pair matching can be conducted. For each subject, its pore keypoints at one pose are matched to the corresponding pore keypoints at an adjacent pose. And we establish three sets of matched keypoint pairs of 10°and 20°, 20°and 30°, 30°and 45°. After finding a set of matches between each image pair, we use the matched keypoints to tracks. A track is a set of matched keypoints across the face images of the same subject at different poses. If a track contains more than one keypoint in the same image, it is considered to be inconsistent and is removed. We choose only those consistent tracks containing 4 keypoints conrresponding to the 10°, 20°, 30°and 45°pose, as shown in Fig. 6. Finally, 17136 tracks are established, which is larger than the pore-to-pore correspondences dataset of Li [4]. Besides, we also generate another large pore-to-pore correspondences dataset with whole face of the subjects of Boshorus dataset, including 80236 tracks.

We find the matching pore scale keypoints of the same subject from different perspectives, which relies of PSIFT features. We extract training patches according to the scale σ of the point, for both feature point image regions. Patches P are extract from a $24\sigma \times 24\sigma$ support region at these locations, and standardized into $S \times S$ pixels where $S = 128$. Some data of the pore-to-pore dataset are shown in Fig. 7.

4 Conclusion

In this paper, we have proposed the 3D model constraint to improve the performance of pore-scale feature matching, which is a method to improve the matching performance when the face images have a larger baseline. Afterward, a larger pore-to-pore correspondences dataset, including 17136 classes of matched pore pairs, is established. In our future work, we will use the larger pore-to-pore correspondences dataset to train a deep neural network, which may be a better pore-scale feature extraction. Besides, we will use our method to test for different expressions and different light condition, and then to generate a pore dataset with different condition.



(a) 10°

(b) 20°

(c) 30°

(d) 45°

Fig. 6. Different poses images of the same subject. The red points are the keypoints of the skin region, and the green points are the correspondences keypoints of different pose

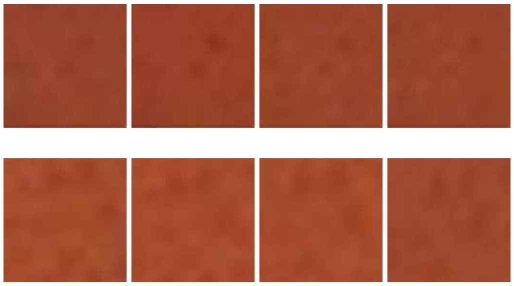


Fig. 7. Some patches of the subject

References

1. Yi K. M., Trulls E., Lepetit V., et al.: LIFT: Learned Invariant Feature Transform. European Conference on Computer Vision. Springer. 467–483 (2016)
2. Lin Y., Medioni G., Choi J.: Accurate 3d face reconstruction from weakly calibrated wide baseline images with profile contours. Computer Vision and Pattern Recognition. IEEE. 1490–1497 (2010)
3. Bay H., Ess A., Tuytelaars T., Van Gool L.: Speeded-up robust features. Computer Vision and Image Understanding. 110, 3, 404–417 (2008)
4. Li D., Lam K. M.: Design and learn distinctive features from pore-scale facial keypoints. Pattern Recognition. 48, 3, 732–745 (2015)
5. Matthews I., Baker S.: Active appearance models revisited. International Journal of Computer Vision. 60, 2, 135–164 (2004)
6. Tzimiropoulos G, Zafeiriou S, Pantic M.: Robust and Efficient Parametric-FaceAlignment. IEEE International Conference on Computer Vision. IEEE. 1847–1854 (2012)
7. Spaun N A.: Facial Comparisons by Subject Matter Experts: Their Role in Biometrics and Their Training. Advances in Biometrics. Springer Berlin Heidelberg. 161–168 (2009)
8. Lin D, Tang X.: Recognize High Resolution Faces: From Macrocosm to Microcosm. Computer Vision and Pattern Recognition, IEEE. 1355–1362 (2006)
9. Lowe, David G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60, 2, 91–110 (2004)
10. Zhu X, Lei Z, Liu X, et al.: Face alignment across large poses: A 3d solution. Computer Vision and Pattern Recognition, IEEE. 146–155 (2016)
11. Blanz V, Vetter T.: Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence. 25, 9, 1063–1074 (2003)
12. Paysan P., Knothe R., Amberg B., et al.: A 3D Face Model for Pose and Illumination Invariant Face Recognition International Conference on Advanced Video and Signal Based Surveillance. IEEE. 296–301 (2009)
13. Cao C., Weng Y., Zhou S., et al.: Facewarehouse: a 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 20, 3, 413–425 (2014)
14. Fischler M. A., Bolles R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. ACM. 24 726–740 (1981)
15. Savran A, Alyz N, et al.: Bosphorus Database for 3D Face Analysis Biometrics and Identity Management. Springer-Verlag. 47–56 (2008)