Facial recognition and 3D non-rigid registration

A. Makovetskii[1], V. Kober[1], A. Voronin[1], D. Zhernov[1]

[1]Chelyabinsk State University, Bratiev Kashirinykh 129, Chelyabinsk, Russia, 454001

Abstract. The most efficient tool for human face recognition is neural networks. However, the result of recognition can be spoiled by facial expressions and other deviation from canonical face representation. In this paper, we propose a resampling method of human faces represented by 3D point clouds. The method is based on non-rigid Iterative Closest Point (ICP) algorithm. To improve the facial recognition performance we use a combination of the method and convolutional neural network (CNN). Computer simulation results are provided to illustrate the performance of the proposed approach.

1. Introduction

Human facial expressions describe a set of signals, which can be associated with mental states such as emotions depending on physiological conditions. There are many potential applications of expression recognition systems. They may take into account about two hundred emotional states [1].

In this paper, we use for three-dimensional facial reconstruction and face alignment a resampling method based on a non-rigid ICP algorithm [8-17] instead of network-based methods. First, we convert a 3D scan to the "canonical form". Since real 3D scans contain holes and boundary noise, we eliminate them. Second, we utilize two reference face models (the same person) with neutral and "happy" expressions. By resampling the test image with neutral expression over the undeformed reference domain, we get the test person with "happy" expression and compare it with the related real expression of the person from the database. The same algorithm is used for the "disgust" expression. The resampling method is based on the non-rigid ICP approach. Note that the proposed method gets resampling models as 3D point clouds, in contrast to [18] where a local-affine transformation is used for 2D curvature images. We use the BOSPHORUS database [19] for our experiments. Convolutional neural networks (CNN) are most often used for the processing of images. The essence of the CNN is a sequential use of alternating convolutional layers obtained from input data by a convolution operation, and subsampling layers obtained by using a pulling operation. This architecture uses the features of the perception of the visual cortex of the brain. CNN can effectively recognize and classify images while reducing the number of network parameters and allowing the parallel computations.

The paper aims to develop a convolutional neural network architecture [20] for the human faces classification and recognition tasks, taking into account facial expressions, face rotation, and brightness variation. When building architecture, we consider the effect of the hyperparameters values, such as the number of hidden layers, the number of neurons in each layer, the size of the convolution kernel, the learning rate (for various loss functions), on the learning process for convolutional neural networks. As a result, we obtain a convolutional neural network architecture, which for the considered dataset provides the most accurate classification. Computer simulation results are provided to illustrate the performance of the proposed approach.

2. Resampling method for removing facial expression

Let $P = \{p_1, \dots, p_s\}$ be a template point cloud, and $Q = \{q_1, \dots, q_s\}$ be a target point cloud in $\mathbb{R}^3$.

2.1. Non-rigid ICP

Suppose that the relationship between points in P and Q is given in such a manner that for each point p_i exists a corresponding point q_i. Denote by $S(Q)$ a surface constructed from the cloud Q, denote by $T_{q^i}(\mathrm{Q})$ a tangent plane of $S(Q)$ at point q^i. Let $J(A_1, \dots, A_s)$ be the following functional:

$$J(A_1, \dots, A_s) = J_1(A_1, \dots, A_s) + \lambda J_2(A_1, \dots, A_s), \quad (1)$$

where

$$J_1(A_1, \dots, A_s) = \sum_{i=0}^{s} ||A_i p_i - q_i||^2, \quad (2)$$

$$J_2(A_1, \dots, A_s) = \sum_{\{i,j\} \in \mathcal{E}} ||A_i - A_j||^2, \quad (3)$$

where $<\cdot,\cdot>$ denotes the inner product, A_i is a matrix of the affine transformation in the homogenous coordinates

$$A_i = \begin{pmatrix} a_i^{11} & a_i^{12} & a_i^{14} & t_i^1 \\ a_i^{21} & a_i^{22} & a_i^{23} & t_i^2 \\ a_i^{31} & a_i^{32} & a_i^{33} & t_i^3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

p_i and q_i are points from the cloud P and Q respectively, n_i is the unitary normal for $T_{q^i}(Q)$,

$$p_i = (p_i^1\ p_i^2\ p_i^3\ 1)^t, \quad q_i = (q_i^1\ q_i^2\ q_i^3\ 1)^t, \quad n_i = (n_i^1\ n_i^2\ n_i^3\ 0)^t, \quad (5)$$

$\mathcal{E}$ is the set of edges of the triangulated surface P, λ is the regularization parameter.

The non-rigid ICP variational problem can be stated as follows:

$$\arg min_{A_1,\dots,A_s} J(A_1, \dots, A_s)\,. \quad (6)$$

A detailed solution to problem (6) is described in [11]. Denote by P_k and Q_k the following matrices:

$$P_k = p_k p_k^t,\ Q_k = q_k p_k^t,\ k = 1, \dots, s. \quad (7)$$

We get the following system of equations:

$$\begin{cases} A_1 P_1 + \lambda e_1 A_1 - \lambda \sum_{j=1}^{e_1} A_j = Q_1 \\ \dots \\ A_s P_s + \lambda e_s A_s - \lambda \sum_{j=1}^{e_s} A_j = Q_s \end{cases}. \quad (8)$$

The system (8) is linear and consists of $12 \times s$ equations from $12 \times s$ variables. The solution to the system is the closed form solution of the variational problem Eq. (6). The system of equations can be rewritten as

$$Ma = q, \quad (9)$$

where $a = (a_1^{00}\ a_1^{01} \dots a_1^{23}\ \ a_2^{00}\ a_2^{01} \dots a_2^{23}\ \ \dots\ \ a_s^{00}\ a_s^{01} \dots a_s^{23})^t$ is a vector of the size of $12s$, $q = (Q_1^{00}\ Q_1^{01} \dots Q_1^{23}\ \ Q_2^{00}\ Q_2^{01} \dots Q_2^{23}\ \ \dots\ \ Q_s^{00}\ Q_s^{01} \dots Q_s^{23})^t$ is a vector of the size of $12s$, matrix M of the size of $12s \times 12s$.

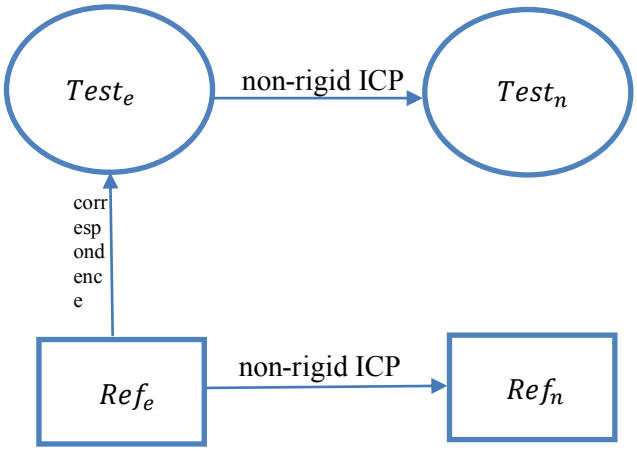


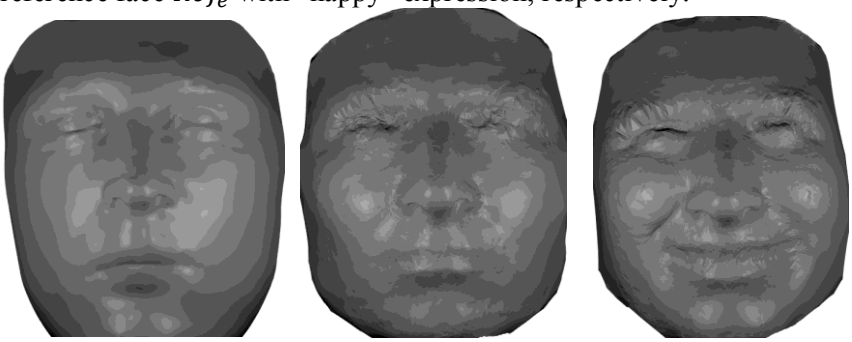
Figure 1. Flow chart of resampling.

2.2. The resampling method

Let us denote by Ref_n, Ref_e, $Test_n$, $Test_e$ the following point clouds: the point cloud Ref_n corresponds to the "reference face" with neutral face expression; Ref_e corresponds to the reference face with some face expression; $Test_n$ corresponds to the reference face with neutral face expression; $Test_e$ corresponds to the reference face with the same type face expression that Ref_e. Figure 1 shows flow chart of the proposed resampling method. We obtain the output $Test_n$ of the resampling method utilize the non-rigid geometrical transformations between Ref_e and Ref_n point clouds.

2.3. Preliminary 3D scans processing

scans can contain holes, boundary noise, other artifacts, and also scans of human faces can have different scale. To remove these phenomena, we map (by the non-rigid ICP algorithm) of a scan to the canonical point cloud, see Figure 2 (a). Figures 2(b) and 2(c) show reference face Ref_n with neutral expression and reference face Ref_e with "happy" expression, respectively.



a) b) c)

Figures 2. a) Canonical point cloud, b) and c) show reference face.

3. Neural network architectures

We describe here the network scheme, utilized data base and training method of the network.

3.1. The database and training method

When training a network, we use the depth maps corresponded to 3D clouds. Since the BOSPHORUS database contains a correspondent depth map for every face point cloud, we use the approach that utilizes both data types. In the training set there are 2007 depth maps. The test set with neutral faces has 283 images. In the test set with "happy" facial expression, there are 105 images (for a person).

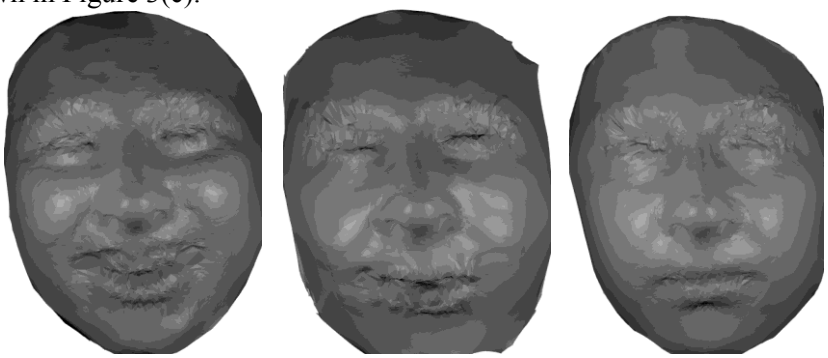
3.2. The network architectures

The output data of CNN for both data types is an array that contains 105 values of probabilities for all classes of the database. Let we assign P^i_{depth} as the probability for the *i*-th class for depth data. When using depth maps the input data for the CNN is a depth matrix size of 100×100. The network has four convolutional layers. The convolution kernel has a size of 3×3 for every layer. The number of feature maps is equal to 32, 64, 128, 200 correspondently. The activation function is ReLU (for each layer). After using the activation function, the Max Pooling function is used (for 2×2 squares). The output of the convolutional layers is the vector that contains 3200 elements. After the convolutional layers, the network contains two fully connected layers. The first of them consists of 1000 neurons with activation function ReLU. The second layer is the output consists of 105 neurons with SoftMax activation function. It is used the algorithm Adam of Keras for training of CNN, the error function is categorical crossentropy. Total number of the trainable parameters of the network is 3629953.

4. Computer simulation

The network yields the precision of face recognition of 88.34% on the test set with neutral facial expressions only. The network yields the precision of face recognition of 81.13% on the test set with the "happy" faces. In particular, the "happy" faces of persons No. 15, No. 88, No. 101 from the database were incorrectly recognized by the network.

Figure 3(a) shows the “happy” face No. 15 represented by the triangulated surface. We use for the correspondent point cloud our resampling method and obtain the point cloud (represented as triangulated surface) that is shown in Figure 3(b). The true point cloud No. 15 with neutral facial expression is shown in Figure 3(c).



a) b) c)

Figure 3. a) The person No. 15 with “happy” facial expression; b) The result of the resampling method; c) The person No. 15 with neutral facial expression.

Remark. The experiments show that the proposed combined algorithm that utilizes the resampling method with neural network improves the facial recognition performance of the network. The precision of face recognition increased from 81.13% to 83.98%.

5. Conclusion

The most efficient tool for human face recognition is neural networks. However, the result of recognition can be spoiled by facial expressions and other deviation from canonical face representation. We proposed a resampling method of human faces represented by 3D point clouds. The method based on the non-rigid ICP algorithm. We showed that the combining this method and convolutional neural network for the face recognition task improves the facial recognition performance of the system.

6. Acknowledgments

The work was supported by the RFBR (grant № 18-07-00963).

7. References

[1] Rusu, B. Fast Point Feature Histograms (FPFH) for 3D registration / B. Rusu, B. Beetz // IEEE International Conference on Robotics &Automation, 2009. – P. 3212-3217.

[2] Voronin, S. Aregularization algorithm for registration of deformable surfaces / S. Voronin, A. Makovetskii, A. Voronin, J. Diaz-Escobar // Proc. Applications of Digital Image Processing. – 2018. – Vol. 10752. – P. 107522S.

[3] Voronin, S. Non-rigid ICP and 3D facial models / S. Voronin, V. Kober, A. Makovetskii, A. Voronin // Applications of Digital Image Processing XLII. – 2019. – Vol. 11137. – P. 111372K.

[4] Tihonkih, D. The iterative closest points algorithm and affine transformations / D. Tihonkih, A. Makovetskii, V. Kuznetsov // CEUR Workshop Proceedings. – 2016. – Vol. 1320. – P. 349-356.

[5] Makovetskii, A. An efficient point-to-plane registration algorithm for affine transformations / A. Makovetskii, S. Voronin, V. Kober, D. Tihonkih // Proc. SPIE’s Applications of Digital Image Processing. – 2017. – Vol. 10396. – P. 103962J.

[6] Makovetskii, A. Affine registration of point clouds based on point-to-plane approach / A. Makovetskii, S. Voronin, V. Kober, D. Tihonkih // Procedia Engineering. – 2017. – Vol. 201. – P. 322-330

[7] Makovetskii, A. A non-iterative method for approximation of the exact solution to the point-to-plane variational problem for orthogonal transformations / A. Makovetskii, S. Voronin, V.

Kober, A. Voronin // Mathematical Methods in the Applied Sciences. – 2018. – Vol. 41(18). – P. 9218-9230.

[8] Makovetskii, A. A point-to-plane registration algorithm for orthogonal transformations / A. Makovetskii, S. Voronin, V. Kober, A. Voronin // Applications of Digital Image Processing. – 2018. – Vol. 10752. – P. 107522R.

[9] Ruchay, A. Fusion of information from multiple kinect sensors for 3D object reconstruction / A. Ruchay, K. Dorofeev, V. Kolpakov // Computer Optics. – 2018. – Vol. 42(5). – P. 898-903. DOI: 10.18287/2412-6179-2018-42-5-898-903.

[10] Ruchay, A. Impulsive noise removal from color images with morphological filtering / A. Ruchay, V. Kober // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – 2018. – Vol. 10716 LNCS. – P. 280-291.

[11] Ruchay, A. An efficient detection of local features in depth maps / A. Ruchay, K. Dorofeev, A. Kober // The International Society for Optical Engineering. – 2018. – Vol. 10752. – P. 1075223.

[12] Savran, A. Non-rigid registration based model-free 3D facial expression recognition / A. Savran, B. Sankur // Computer Vision and Image Understanding. – 2017. – Vol. 162. – P. 146-165.

[13] Bosphorus 3d face database [Electronic resource]. – Access mode: http://bosphorus.ee.boun.edu.tr/default.aspx.

[14] Leonov, S. Analysis of the convolutional neural network architectures in image classification problems / S. Leonov, A. Vasilyev, A. Makovetskii // Applications of Digital Image Processing XLII. – 2019. – Vol. 11137. – P. 111372E.