Estimation of Human Age and Facial Expression Using Biologically Inspired Features

A. Spizhevoy, A. Bovyrin
Lobachevsky State University of Nizhni Novgord and Itseez Inc., Russia
{alexey.spizhevoy, alexander.bovyrin}@itseez.com

Abstract

In this paper we deal with the problems of human age and facial expression estimation. We propose to use Biologically Inspired Features (BIF) as a facial descriptor in the both tasks to improve performance. We developed a complete pipeline for solving those problems including geometric normalization step and CLAHE method for photometric normalization. To the best of our knowledge using BIF for facial expression classification, as well as combining BIF together with ordinal hyperplanes ranking for age prediction haven't been studied before. The proposed pipeline was tested on the standard datasets such as the FG-NET aging and the Extended Cohn-Kanade (CK+) expression databases and demonstrated high accuracy compared to the state-of-the-art methods.

Keywords: *human age estimation, face normalization, bio-inspired features, ordinal hyperplanes ranking.*

1. INTRODUCTION

Interest in solving such problems as human age estimation and facial expression classification has been growing for many years. Those tasks arise in many areas such as: human-computer interaction, surveillance monitoring, video content analysis, targeted advertising, biometrics, and entertainment.

A typical solution that recognizes facial expression or human age by image/video automatically usually includes a few basic blocks: face detection and cropping, image geometric and photometric normalization, computing face descriptors, reducing optionally feature vector dimensionality, and applying a machine learning method for estimation of age and expression.

Both problems are quite sophisticated because of high variability in face appearance due to such factors as head rotation, emotions, illumination conditions, face makeup, and many others. All of these issues should be resolved, to some extent, to build an automatic solution which would be reasonable in practice. Some of the factors are taken into account during face normalization, while the others are resolved via enriching training dataset to model age/expression function more accurately.

Contribution: the current work is application guided and describes key elements of the developed system. Our contribution includes the following items:

- For the sake of achieving time performance gain we proposed to use the same features for solving both tasks.
- We built a fully automatic system that is able to estimate human age and facial expressions in real time by frames from video.
- The developed system showed accuracy higher than the state-of-the-art methods on the standard CK+ expression and FG-NET aging databases.

2. RELATED WORK

The goal of human age estimation is to predict age or age range by person's face image. All existing approaches can be roughly split into two classes by convention, depending on what type of machine learning algorithm is used. That is either regression [8, 6] or classification [2]. The former one considers age as a continuous variable, and the latter one as a discrete variable, i.e. age group index. There are also mixed approaches, for instance [6, 7]. Such factors as illumination conditions when taking photo, face makeup, facial expressions and other might lead to age estimation errors, while a human can cope with those issues quite easily.

The problem of automatic facial expression recognition is challenging as well. The same factors that might worsen age estimation results are applicable here, i.e. illumination conditions, face makeup, genetics, etc. Moreover different human expressions can be presented at the same time which makes the problem even more difficult. As noted in [10] facial expressions in humans are dynamic in nature, consisting of an onset, peak and an offset phase. Some approaches incorporate temporal information, for instance see [10, 13, 15]. However typical methods rely on appearance features mostly, we call such methods static.

Active Appearance Models are widely used for both age estimation and expression classification as facial descriptors, see [2, 6, 11, 4] for they combine both appearance and shape information. There are some works devoted to using Local Binary Patterns for age estimation and expression classification, see [14, 5] as LBP features are quite successful in face recognition and can be computed very efficiently. One of the best results demonstrated for age estimation was achieved using descriptors based on Gabor filters, see [8, 7].

In this work we investigate behavior of Biologically Inspired Features [8, 9] on the problem of facial expression classification as such approach hasn't been studied before. We show that it provides very high accuracy compared to the best known methods. Finally we propose a pipeline that uses the same features for solving both problems: age estimation and expression classification -- that reduces the average computational time in 23%, compared to the version where facial descriptors are not shared between the problems and computed independently.

2.1. Problem Statement

The problem of age estimation can be formulated as follows: given training dataset consisting of m samples $\{(X_i, y_i), i = 1..m\}$, where $X_i = X(I_i)$ is the i-th face descriptor, computed from the face image I_i, and $y_i = y_{age}(X_i)$ is the ground truth age value, either exact age or age group index. The goal is to estimate age $\hat{y}_{age}(X(I))$ by new unknown face image I, which is not presented in the training database.

The goal of facial expression classification is to determine facial expression by person's face image. We tract the task as a machine

learning classification problem. So it is formulated as follows: given training dataset consisting of m samples $\{(X_i, y_i) | i = 1..m\}$, where $X_i = X(I_i)$ is the i-th face descriptor, computed from the face image I_i, and $y_i = y_{ex}(X_i) \in C_{ex} = \{An, Co, Di, Fe, Ha, Sa, Su\}$ is the ground truth expression label. In our work we deal with such facial expressions as angry, contempt, disgust, fear, happy, sadness, and surprise. The facial expression labeling of that kind is provided with the Extended Cohn-Kanade (CK+) dataset [11]. The goal is to estimate facial expression label $\hat{y}_{ex}(X(I))$ by new unknown face image I, which is not presented in the training database.

3. FACE DESCRIPTORS

As a base approach for face descriptors computation we use Biologically Inspired Features (BIF) proposed in the paper [8] for both age estimation and facial expression classification problems. Using the same descriptors saves about 23% of the computational time, since we don't have to compute different features independently -- all we need is just to compute BIF once per face. See section 3.4 for details about descriptors parameters.

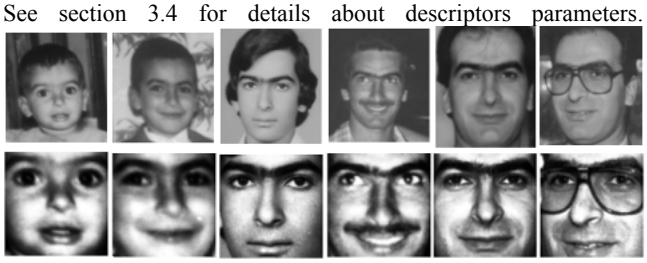

Fig. 1. Results of face normalization on the FG-NET aging database images. Top row shows original images, bottom row shows normalized images

3.1. Geometric Normalization

The goal of this step is to remove uninformative geometric transformations from images such as face scale variations and in-plane head rotations. We use eye centers and map them, using similarity transformation, into two predefined points. By doing this we eliminate in-plane head rotations and scale variations, so they don't affect facial descriptors anymore. In our experiments we used eye center positions provided with the databases.

Here is the description of the method we propose. Let $p_{le} = (x_{le}, y_{le})^T$ and $p_{re} = (x_{re}, y_{re})^T$ be the coordinates of left and right eyes respectively on source image I, where W and H are its width and height respectively. Using similarity transformation, i.e. combination of rotation, scale, and translation, we map the points p_{le}, p_{re} into fixed points $\hat{p}_{le}$, $\hat{p}_{re}$.

We use the following fixed points which were adjusted experimentally: $\hat{p}_{le} = (0.78W, 0.25H)^T$ and $\hat{p}_{re} = (0.22W, 0.25H)^T$. Image width and height (W and H) are the same for all images -- that's insured via image resizing after face detection and image cropping. We used the values $W = 66$ and $H = 66$ that gave the best results in our experiments.

Experiments showed that geometric normalization is a vital step for achieving high accuracy results. The described method decreases error on the FG-NET aging database from 5.5 to 4.56 years, see table 1.

3.2. Photometric Normalization

Input of the photometric normalization step is geometrically normalized image I. This step is important for achieving accurate results as well, and is aimed to reduce uninformative illumination variations that might worsen face recognition accuracy. We proposed two methods for experiments: Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE). For detailed descriptions we refer to [12].

While HE method improves overall contrast, it often makes sense to improve local contrast, e.g. when one part of face is much lighter/darker than the other. Our experiments showed that CLAHE provides better results than HE. You can see the accuracy achieved using HE and CLAHE for age estimation problem on the FG-NET aging database in table 1. While HE together with the geometric normalization, described in the previous section, gave 4.32 years error, using CLAHE we achieved 4.1 years error. Examples of face normalization results are shown in figure 2.

3.3. Biologically Inspired Features

In working with BIF we follow to the work [8]. The method takes as input normalized image I of some fixed predefined size. Then a number of Gabor filters are applied to the image and then the response maps are processed further. The whole feature computation algorithm can be described as follows:

1. For each angle $\theta \in [0, \pi)$ with some step:
   a. Apply a set of Gabor filters with orientation θ to image I.
   b. Split the response maps into pairs and merge them using per-pixel maximum operation.
   c. Computer statistical features (standard deviations) passing over the images with sliding window.
2. Combine all features into one vector.

The results of steps 1.1 and 1.1 form so called simple (S) layer, when results of step 1.3 form complex (C) layer from S layer elements.

3.4. Dimensionality Reduction

As the total number of features after C layer elements are computed and merged into one vector can be very high, there is a need of performing dimensionality reduction. To reduce the number of features we apply Principal Component Analysis (PCA) method. Applying this technique we reduce feature vector size to 881 -- the best value from our experiments with age estimation on the FG-NET aging database. For the problem of expression classification the reduced feature vector size is 654 -- that is the number of samples in the CK+ database after augmenting it with the vertically flipped images.

4. AGE ESTIMATION

The authors of the paper [2] presented a ranking method for age estimation. They called it Ordinal Hyperplanes Ranker -- OHRank and used it together with Active Appearance Models. In this section we describe our modification of this method that we use together with BIF descriptors.

Let X_i be the i-th feature vector from the training set, which corresponding ground truth age is $y_i \in \{0,1,...,K\}$, and K is the

maximum possible age (e.g. the maximum age observed in the database). For each possible age year k we split the training dataset into two parts according to the following rule:

$$X_k^+ = \{(X_i, 1) \mid y_i > k\},$$
$$X_k^- = \{(X_i, 0) \mid y_i \le k\}.$$

Then these sets are used to train binary classifiers $\{f_k\}$. These classifiers estimate answers (as discrete variable from the $\{0,1\}$ set) on the following queries: «is age of given person greater than k years or not»? To get age estimation all of the K binary classifiers are applied to the input feature vector X. Final value is aggregated via computing sum of all the binary classifier answers: $y(X) = \sum_{k=0}^{K-1} f_k(X)$.

5. FACIAL EXPRESSION CLASSIFICATION

For facial expression classification we propose merging texture and geometric features (i.e. face shape landmarks) via aggregation of probability distributions from different SVM models. Image samples with the face shape landmarks from the CK+ database are shown in figure 2.

Fig. 2. Sample images from the CK+ database with the face shape landmarks

Face shape features are obtained via combining 68 face landmark points resulting into 136 dimensional feature vector. Before merging points into one feature vector they are transformed (shifted and scaled) into the face rectangle system of coordinates.

In detail the process looks as follows: given $P_1(y \mid X_1)$ and $P_2(y \mid X_2)$ -- the probability distributions obtained from two different SVM models, where X_1 and X_2 are two different feature vectors corresponding to the same image, the score function is computed as their product: $S(y \mid X_1, X_2) = P_1(y \mid X_1) P_2(y \mid X_2)$. Final result of classification is determined as $y_{ex}(X_1, X_2) = \arg\max_{y \in C_{ex}} S(y \mid X_1, X_2)$.

On the CK+ database merging probability distributions from two different SVM models with RBF kernels improves the average recognition rate from 94.1% (BIF only) to 96.8% (BIF & shape).

6. EXPERIMENTS

6.1. Standard Datasets

6.1.1. The FG-NET Aging Database

For human age estimation validation we used the standard FG-NET aging database [1]. The database contains 1002 face photos of 82 persons taken under uncontrolled conditions. The images have some light intensity and head pose variations. For each person there are more than 10 images in the range from 0 to 69 years. For each face image there are 68 landmarks available, from which we use only eyed coordinates for geometric normalization.

6.1.2. The Extended Cohn-Kanade Database

For facial expression validation we used the Extended Cohn-Kanade (CK+) database [11]. The database contains 593 face image sequences, but only 327 are labeled with expression classes. Those 327 images cover 118 out of 123 persons. Each sequence begins with the neutral face and ends with the peak intensity expression. We used only images with the peak expression intensity. The database provides the 68 face landmarks labeling.

6.2. Accuracy Metrics

There are two widely used age estimation accuracy metrics. The first one is the mean absolute error, which is computed as follows: $MAE = \frac{1}{m} \sum_{i=1}^{m} | y_i - \hat{y}_i |$, where m is the total number of the test samples, y_i is the ground truth age value for the i-th test sample, and $\hat{y}_i$ is the predicted age value. The second accuracy metric is the cumulative score, which reflects how prediction errors are distributed over years. For every error level L it is computed as $CS(L) = \frac{m_{e \le L}}{m} 100\%$, where $m_{e \le L}$ is the number of the test sample prediction errors less or equal to L.

Method	MAE	CS(5)
BIF+SVR [8]	4.77	≈69
AAM+OHRank1 [2]	4.48	74.4
AAM+OHRank2 [2]	4.56	74.7
C-lsRCS+lsLPP [3]	4.38	≈74
Ours (no norm)	5.5	68.7
Ours (geom)	4.56	74.3
Ours (geom+HE)	4.32	74.7
Ours (geom+CLAHE)	4.1	76.4

Table 1. Comparison of the age estimation results on the FG-NET aging database. The proposed approach outperforms the state-of-the-art methods on the same database. Also the table shows how geometric and photometric normalization steps are important.

6.3. Results

In table 1 we compare the results of our human age estimation approach with the best reported ones so far on the FG-NET aging database. From the table you can see that our approach shows high accuracy in comparison to the best methods. The table also shows superior accuracy of the CLAHE based photometric normalization over the HE based one. Leave-one-person-out (LOPO) cross-validation technique was used to get accuracy estimate.

Method	Avg. Rec. Rate, %	Static	Valid. Protocol
Baseline [11]	83.3	Yes	LOPO
CLM [4]	74.4	Yes	LOPO
Shape+SVM [10]	84.06	Yes	4-fold
Shape+LDCRF [10]	95.79	No	4-fold
Cov3D [13]	92.3	No	5-fold
CPL [16]	88.42	Yes	10-fold
ITBN [15]	86.3	No	15-fold
Ours (no norm)	91.9	Yes	LOPO

Ours (geom)	92.7	Yes	LOPO
Ours (geom+HE)	94.4	Yes	LOPO
Ours (geom+CLAHE)	96.8	Yes	LOPO

Table 2. Comparison of recognition rates for 7 facial expressions classification on the CK+ database. The 3rd column denotes whether only one image is used in the method to estimate expression (we call such methods static) or a few frames are used. Here again the proposed approach outperforms the state-of-the-art methods on the same database.

For facial expression classification problem we compute the confusion matrix in exactly the same way as the authors of paper [11] did, i.e. LOPO cross-validation and 7 facial expressions (anger, contempt, disgust, fear, happy, sadness, surprise).

7. CONCLUSION

We proposed to use a combination of BIF together with ordinal hyperplanes ranking for human age estimation that hasn't been studied before. The developed approach was tested on the standard FG-NET aging database and MAE of 4.1 years was achieved. We investigated face normalization methods and the results showed that a huge improvement in accuracy can be made using a combination of geometric normalization and CLAHE as photometric normalization, see table 1. Haven't been applied to the problem of facial expression classification before, the proposed pipeline was tested on the the CK+ database and the result of 96.8% average recognition rate was achieved, see table 2.

	An	**Co**	**Di**	**Fe**	**Ha**	**Sa**	**Su**
Recogn. Rate, %	95.6	88.9	98.3	96	100	100	98.8

Table 3. Recognition rates for each of 7 facial expressions on the CK+ database obtained using our approach.

The obtained results show that our framework provides high accuracy for both age estimation and expression classification problems compared to the best known methods. The pipeline is also perspective for using in applications as the same features are used for solving the both problems, hence BIF must be computed only once per image. Sharing BIF descriptors between both age estimation and expression classification saves about 23% of the computational time, compared to the version where facial descriptors are not shared between the problems and computed independently.

8. REFERENCES

[1] The FG-NET Aging Database. http://www.fgnet.rsunit.com, http://www-prima.inrialpes.fr/FGnet/.

[2] Chang, K.-Y., Chen, C.-S., and Hung, Y.-P. Ordinal hyperplanes ranker with cost sensitivities for age estimation. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* – P. 585 - 592. IEEE (2011).

[3] Chao, W.-L., Liu, J.-Z., and Ding, J.-J. Facial age estimation based on label-sensitive learning and age-specific local regression. *In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. – P. 1941 - 1944. IEEE (2012).

[4] Chew, S.~W., Lucey, P., Lucey, S., Saragih, J., Cohn, J.~F., Sridharan, S. Person-independent facial expression detection using constrained local models. *In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. – P. 915 - 920. IEEE (2011).

[5] Gunay, A., Nabiyev, V.~V. Automatic age classification with LBP. *In Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on* – P. 1 - 4. IEEE (2008).

[6] Guo, G., Fu, Y., Dyer, C.~R., and Huang, T.~S. Image-based human age estimation by manifold learning and locally adjusted robust regression. *Image Processing, IEEE Transactions on 17*, 7, 1178 - 1188 (2008).

[7] Guo, G., Fu, Y., Huang, T.~S., and Dyer, C.~R. Locally adjusted robust regression for human age estimation. *In Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*. – P. 1 - 6. IEEE (2008).

[8] Guo, G., Mu, G., Fu, Y., and Huang, T.~S. Human age estimation using bio-inspired features. *In Computer Vision and Pattern Recognition, 2009*. CVPR 2009. IEEE Conference on p. 112 - 119. IEEE (2009).

[9] Spizhevoy, A.S., Ogolikhina, A.I., Bovyrin, A.V. Automatic Facial Age Estimation Using Adaptive Brightness Equalization and Biologically Inspired Features. *Vestnik of Lobachevsky State University of Nizhni Novgorod*, Issue 1. – P. 273 - 279 (2014).

[10] Jain, S., Hu, C., and Aggarwal, J.~K. Facial expression recognition with temporal modeling of shapes. *In Computer Vision Workshops (ICCV Workshops), 2011. IEEE International Conference on,* pp. 1642--1649. IEEE (2011).

[11] Lucey, P., Cohn, J.~F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. *In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, p. 94 - 101. IEEE (2010).

[12] Pizer, S.~M., Amburn, E.~P., Austin, J.~D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar~Romeny, B., Zimmerman, J.~B., and Zuiderveld, K. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing 39*, 3. – P. 355 - 368 (1987).

[13] Sanin, A., Sanderson, C., Harandi, M.~T., and Lovell, B.~C. Spatio-temporal covariance descriptors for action and gesture recognition. *arXiv preprint arXiv:1303.6021* (2013).

[14] Shan, C., Gong, S., and McOwan, P.~W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing 27*, 6. – P. 803 - 816. (2009).

[15] Wang, Z., Wang, S., and Ji, Q. Capturing complex spatio-temporal relations among facial muscles for facial expression recognition.

[16] Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., and Metaxas, D.~N. Learning active facial patches for expression analysis. *In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, p. 2562 - 2569. IEEE (2012).