Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Integrating Feature Extractors for the Estimation of Human Facial Age

Shahram Taheri & Önsen Toygar

To cite this article: Shahram Taheri & Önsen Toygar (2019) Integrating Feature Extractors for the Estimation of Human Facial Age, Applied Artificial Intelligence, 33:5, 379-398, DOI: 10.1080/08839514.2019.1577009

To link to this article: https://doi.org/10.1080/08839514.2019.1577009

Published online: 22 Feb 2019.

APPLIED ARTIFICIAL INTELLIGENCE
2019, VOL. 33, NO. 5, 379–398
https://doi.org/10.1080/08839514.2019.1577009

Integrating Feature Extractors for the Estimation of Human Facial Age

Shahram Taheri and Önsen Toygar

Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, Turkey

ABSTRACT

Facial feature extraction algorithms play an important role in many applications of face biometrics such as face recognition for person identification, classification of emotions by facial expression recognition and age estimation using facial images. In this paper, an integration of different type of feature extraction algorithms is applied on facial images for accurate age estimation. This integration is performed by using two-level fusion of features and scores with the help of feature-level and score-level fusion techniques. In our proposed method, the advantage of using different types of features such as biologically inspired features, texture-based features, and appearance-based features is used. Feature-level fusion of biologically inspired and texture-based methods is integrated into the proposed method and their combination is fused with an appearance-based method using score-level fusion. Experiments on the publicly available MORPH and FG-NET databases prove the effectiveness of the proposed method and the proposed method outperforms many of the state-of-the-art systems.

Introduction

One of the challenging problems in human face image understanding field is age estimation from facial images. Age estimation can be used in many real-world applications like intelligent advertisement, security surveillance monitoring, and biometrics. Some of the reasons that facial age estimation is difficult are aging process that has an uncontrollable nature and is strongly specified by the personal lifestyle. Additionally, there is a large inter-class similarity and intra-class variation of subjects' images within the age classes, and finally, it is difficult to collect a comprehensive and representative age annotated facial images dataset for training precise models. Although in the early years of emerging automatic age estimation from facial image fields, the number of existing age annotated collections was narrow and they covered a limited range of ages, but

CONTACT Önsen Toygar onsen.toygar@emu.edu.tr Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10, Turkey

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

a great progress occurred by distribution of large collections like MORPH (Ricanek and Tesafaye 2006).

The inherent difficulties in age estimation from facial image, have derived research into constructing especially complex feature descriptor approaches in which most of them are either user-defined multi-level and orientation bank of filters which were tried to mimic the behavior of animal visual cortex (primary) network, or fine-grained facial regions to perform accurate alignment by using multiple facial fiducial points. In both cases, the generated feature extractor method is difficult to reuse, and in many cases, it has high-dimension which takes considerable time to be extracted.

Recent advances in image classification and object recognition fields helped researchers to propose several very efficient and histogram-based feature extraction approaches that have interesting properties such as invariant to scale and rotation and robust to illumination and alignment variances. These schemes that determine intensity variations in small area or local neighborhood templates from spatial patches are fundamental approaches for a variety of applications which deal with extremely uncontrolled environments.

Score level fusion shows very good performance in multimodal biometric systems (Marcialis and Roli 2007; Patil and Bhalke 2016; Sim et al. 2014; Youssef, Elberrichi, and Adjoudj 2010). In a biometric recognition system, the match score is a measure of similarity between the input and template biometric feature vectors. When match scores' outputs by different biometric matchers are consolidated in order to arrive at a final recognition decision, fusion is said to be done at the match score level which is also known as score-level fusion. Apart from the raw data and feature vectors, the match scores contain the richest information about the input pattern. Additionally, it is relatively easy to access and combine the scores generated by different biometric matchers. Score-level fusion has been used in many research studies. Marcialis and Roli (2007) suggested a score-level fusion of fingerprint and face matchers for personal verification under stress conditions. Patil and Bhalke (2016) proposed a multibiometric system that used three traits such as fingerprint, palmprint, and iris that are combined by using weighted fusion technique for person identification.

In this study, we evaluate different popular local and global feature extractors in both feature-level and score-level fusion in order to investigate their efficiency and discrimination in the age estimation field. The main contributions are as follows: a brief summary of several efficient local and global feature extractors for image classification which are used for the first time in age estimation problem is given and investigated their discrimination power. Additionally, for the first time, we show that the feature-level and score-level fusion of textural descriptors, local appearance-based feature extractors, and global feature descriptors achieve superior result than the other previous works in this field.

The rest of the paper is organized as follows: previous studies in facial age estimation have been reviewed in Section 2. In Section 3, the selected feature extractors to be used in feature-level and score-level fusion are reviewed. Section 4 describes the overall structure of the proposed method. Section 5 demonstrates the evaluation results of experiments performed by combining different level fusion of local and global feature extractors. Finally, Section 6 provides the final conclusion and future works.

Related Works

The early research in facial age estimation problem was performed back 20 years ago. However, a lot of attention on this topic has been drawn from the year that the large database, namely MORPH-Album 2 (Ricanek and Tesafaye 2006), published. The number of age-annotated data in this database was increased by near 55 times with respect to the other databases at that time. Many recent works used MORPH-Album 2 and the researchers investigated different feature descriptors and classification approaches. Some of the previous methods are explained in the following subsections.

The first public domain database that has been used for age estimation is FG-NET (Lanitis, Taylor, and Cootes 2002) which contains 1002 images of 82 individuals and the age of subjects range from 0 to 69 years. Many researchers used FG-NET database for age estimation and classification (Fu and Huang 2008; Izadpanahi and Toygar 2014; Mirzaei and Toygar 2011). A comprehensive survey of research on facial aging using the FG-NET aging database had been published by Panis et al. in 2016 (Panis et al. 2016). One of the earliest steps in age estimation is to extract particular visual feature descriptors. These features should be robust within the same age and discriminative among different ages. Additionally, the features' dimensionality and computation time should be optimal. Some methods rely on appearance models and flexible shape such as Active Appearance Model (AAM) and Active Shape Model (ASM) techniques. These statistical methods try to model aging patterns (Cootes, Edwards, and Taylor 1998; Panis et al. 2016; Scandrett, Solomona and Gibsona 2006) by capturing the fundamental modes of variation in intensity and shape founded in a series of facial images and encoding the face signatures based on these characteristics. On the other hand, some approaches use various types of feature extractors, and then use a classification or regression method to estimate the age. For instance, Bio-Inspired Features (BIF) scheme which classifies the images by mimicking the process of visual cortex in recognition tasks, has frequently applied for age estimation (Geng, Yin and Zhou 2013; Riesenhuber and Poggio 1999). The BIF method is a feed-forward network structure that consists of several cascaded convolutional and pooling layers. In convolutional layer, an input image is convolved with a bank of multi-scale and multi-orientation Gabor filters. Then in the pooling layer, by using MAX operation the results will be down-sampled. (Guo et al. 2009) construct a simplified version of BIF for age estimation which has only two layers and they manually tuning the filter banks specifications. These extracted features are also used in their posterior studies (Guo and Mu 2011).

Local neighborhood features are common descriptor choices in this field and used in many articles (Choi et al. 2011; Gunay and Nabiyev 2008; Yang and Ai 2007). In (Weng et al. 2013), four different features combined together: LBP histogram features, shape and textural features of AAM, PCA projection of the original image pixels and PCA of BIF. In (Fernández, Huerta, and Prati 2015; Huerta, Fernández and Prati 2014), HOG features have been used for age estimation.

Several classification approaches with respect to learning algorithm have been proposed for age estimation such as Support Vector Machines/Regressors (Cootes, Edwards, and Taylor 1998; Guo et al. 2009; Riesenhuber and Poggio 1999; Weng et al. 2013), projection techniques such as Partial Least Squares (PLS), Canonical Correlation Analysis (CCA) (Guo and Mu 2011, 2014), neural networks (Lanitis, Draganova and Christodoulou 2004), Conditional Probability Neural Network (Scandrett, Solomona and Gibsona 2006) and Random Forests (Montillo and Ling 2009). A comprehensive comparison of different classification approaches in the field of age estimation has been reported in (Fernández, Huerta, and Prati 2015; Huerta, Fernández and Prati 2014). As a result of this comparison, CCA accuracy and efficiency was dominated over the others. Recently, the best current result over MORPH is achieved by combining BIF features with kernel CCA (Guo and Mu 2013), however due to computational limitation, they only use 10000 samples for training and testing.

In recent years, deep learning and especially convolutional neural networks (CNN) have been used in many image processing, object recognition and computer vision applications (Druzhkov and Kustikova 2016) and outperformed state-of-the-art methods. Recently, many researchers have been using CNN for facial age estimation problems. We can categorize these deep learning approaches into two classes: shallow architecture and deep architecture. The shallow architecture consists of a limited number of convolution layers (approximately 1–5 layers) but in deep architecture, the number of convolution layers is not limited. The deep architecture suffers from over-fitting problem when there is a small number of training data. For example, the number of images in Morph-II dataset is relatively small so it is impossible to train a deep architecture from scratch by using only these samples. Recently, the researchers used additional datasets with thousands of annotated images to overcome this problem and achieved state-of-the-art results (Hu et al. 2017; Niu et al. 2016; Ranjan et al. 2017; Rothe, Timofte, and Van Gool 2018; Yi, Lei, and Li 2014). However, in this study, in order to be fair in comparison, we compare our results with the approaches that only used Morph or FG-NET dataset for training their model. The work of Yi, Lei, and Li (2014) is one of the earliest works which used CNNs for age estimation. They used several shallow multiscale CNNs on different face regions and obtained the MAE of 3.63 on MORPH-II dataset. On the other hand, Yi, Lei, and Li (2015) used CNNs and they proposed using a ranking encoding for age and gender and they reported the MAE of 3.5 on MORPH-II dataset.

The fundamental contributions of our work are the proposal of a new two-level fusion of common local and global feature extractors capturing different types of face specifications like wrinkle and facial component shapes in order to estimate subjects' age. The different characteristics of these feature descriptors allow the utilization of the advantages of each of them, bringing the discrimination power of the system to be superior compared to the performance of each separately which utilized only local or global type of features.

Feature Extraction Methods

Feature extraction is an important step in an image classification system. The selection of discriminative feature descriptors is a critical decision and it affects the whole system performance. Therefore, we have chosen different types of well-known and efficient local and global feature extractors that have been successfully used in the other image processing fields like face recognition with their discrimination ability, computational efficiency, compact feature space size, and robustness to alignment and illumination variance. These feature extraction methods, namely Histograms of Oriented Gradients, Median Robust Extended LBP, Kernel Fisher Discriminant Analysis and Biologically Inspired Features, are explained below briefly.

Histogram of Oriented Gradients (HOG)

One of the popular feature extractors that has been used in object classification (Dalal and Triggs 2005) and face recognition (Farmanbar and Toygar 2016; Eskandari and Toygar 2014) is Histogram of Oriented Gradients (HOG). This method constructs the features by counting occurrences of gradient orientation in local patches or detection window of an image. In this approach, an image is decomposed into different partitions and then the histogram of orientations for each partition is computed. Finally, all of these histograms are concatenated to construct HOG feature.

Median Robust Extended LBP (MRELBP)

Local Binary Patterns (LBP) method (Ojala, Pietikäinen and Maenpää 2002) is considered among the most computationally efficient high-performance texture descriptors. LBP is used in several recognition problems such as face, iris, palmprint and plant recognition (Dalal and Triggs 2005; Farmanbar and Toygar 2016). However, the LBP method is very sensitive to image noise and is unable to capture macrostructure information. In order to best address these disadvantages (Liu et al. 2016) introduced a novel descriptor for texture classification, namely Median Robust Extended LBP (MRELBP). Instead of using raw image intensities, MRELBP compares median of image intensities in a local region. By using a multiscale LBP type descriptor and a novel sampling scheme, this method can capture both microstructure and macrostructure texture information. MRELBP is shown to be highly robust to different types of image noises, such as salt-and-pepper noise, Gaussian noise, Gaussian blur, and random pixel corruption. (Liu et al. 2017) compared a large number of LBP variants. They designed different experiments to measure their feature descriptors' robustness against changes in rotation, viewpoint, illumination, scale, different types of image degradation, number of classes, and computational complexity. The best overall performance is obtained for the MRELBP feature. Illustration for the MRELBP descriptor and its three components is shown in Figure 1.

Kernel Fisher Discriminant Analysis (KFA)

The kernel fisher discriminant analysis (KFA) method (Liu 2006) is an extension of fisher discriminant analysis (FDA). In this approach as a first step, input space will be expanded by using a nonlinear mapping, and then in the obtained feature space, the multiclass *FDA* will be applied. By implementing nonlinear mapping the dimensionality of feature space will be increased and as a result, it improves the discriminative ability of the KFA method. The main advantage of the KFA method is that it can be applied for multiclass pattern classification problems, and its solution is unique which is its superiority to Generalized Discriminant Analysis (GDA) (Liu 2006) method which produces multiple solutions.

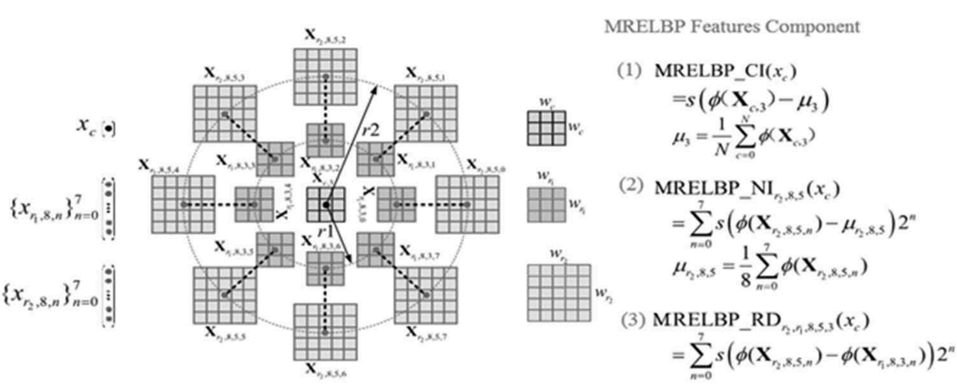

Figure 1. Illustration for the MRELBP descriptor and its three components: (1) Center pixel representation, (2) Neighbor representation and (3) Radial difference representation. In these formulas, $\varphi()$ is a median filter (Liu et al. 2016).

Biologically Inspired Features (BIF)

Biologically Inspired Features (BIF) is a method that tries to model visual processing in the cortex as a stack of increasingly sophisticated layers. Riesenhuber and Poggio (Theriault, Thome, and Cord 2011) proposed a new set of features derived from a feed-forward network of the primate visual object recognition pathway, called the "HMAX" model. The Model consists of two different types of layers: S units' neurons (simple) C units' neurons (complex) (C). A remarkable feature of this model is that it uses nonlinear maximum operation "MAX" in the Simple units' neurons instead of using the linear summation operation "SUM" in pooling inputs at the Complex layers. The benefit of using the "MAX" operation within a small neighborhood is to tolerate small changes in position and scale. Specifically, S1 layer (first layer) is constructed by convolving a set of Gabor filters over the grayscale image at four orientations and 16 scales. Then, each pair of adjacent S1 unit is combined together to generate 8 bands of units for each direction. In the next layer, which is called C1, the maximum values within local patches and across the scales within a band is computed. Therefore, C1 feature includes 8 bands and 4 orientations. The bio-inspired features (BIFs) have been investigated for object category recognition (Cord, Theriault, and Thome 2011; Thériault, Thome and Cord 2013) and face recognition (Shan 2010). An overview of BIF system is shown in Figure 2.

Proposed Method

According to the success rate of using score-level fusion in multimodal biometric recognition systems, it is believed that accuracy can be improved when the information of two different types of classifiers are consolidated. Due to large inter-class similarity and intra-class variation, we need to perform fusion in two levels with different kinds of feature descriptors: feature-level fusion of local descriptor and score-level fusion of appearance-based descriptor and the obtained results from the previous level fusion.

Different feature descriptors were investigated to select the most powerful ones for descriptor selection. Inspired by recent face recognition and texture classification works (Cord, Theriault, and Thome 2011; Shan 2010) in the computer vision community, we used Biologically Inspired Features (BIF), Median Robust Extended LBP (MRELBP) and Histogram of Oriented Gradients (HOG) for local feature description and used Kernel Fisher Analysis (KFA) for the appearance-based recognition system in which the results of recent research (Liu 2006) shows that it outperforms LDA and PCA methods.

Figure 3 illustrates the overall schematic of the proposed method. For each input image, two different kinds of features are computed. Feature-level fusion of HOG, BIF and MRELBP is performed by concatenating their feature vectors together. We also compute KFA from the original image and compare the

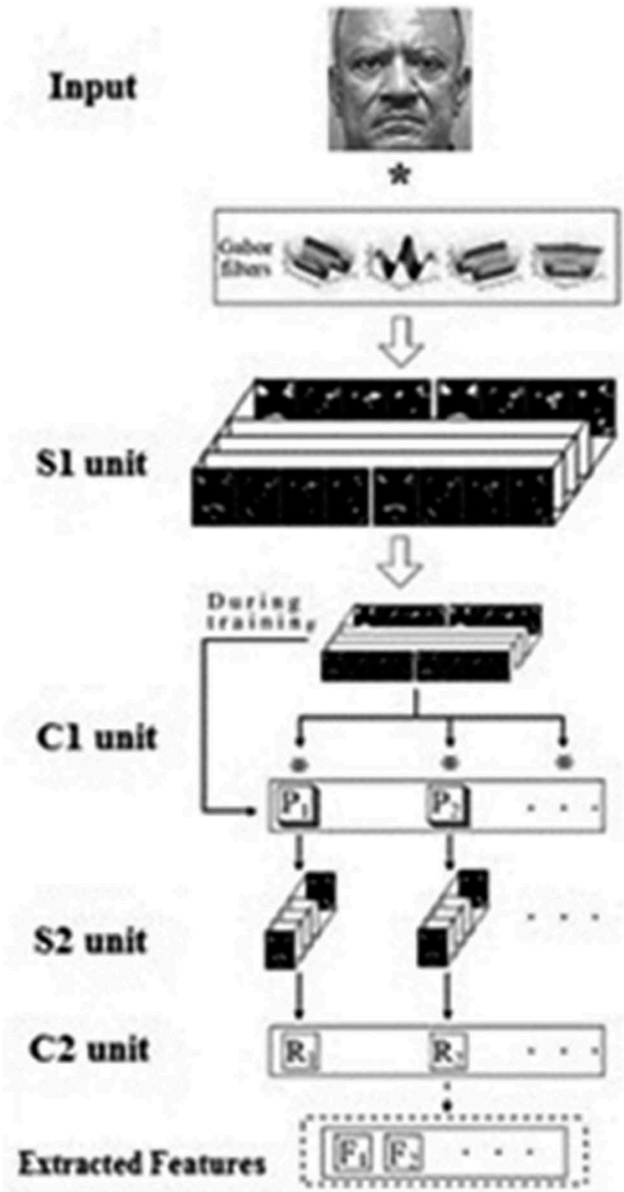

Figure 2. The system architecture of BIF (Cord, Theriault, and Thome 2011).

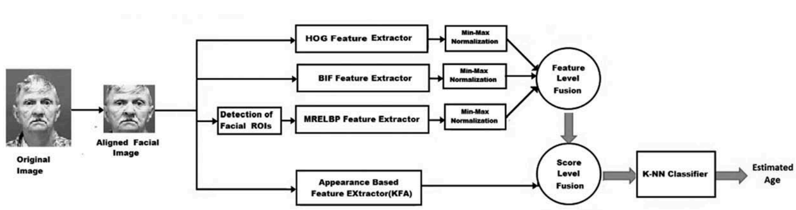

Figure 3. Schematic of the proposed method.

similarity of these two feature vectors with all of the feature vectors in the training set and then select the minimum one for each method. After normalizing these obtained scores, we add them together and make a decision by using k-Nearest Neighbour (KNN) classifier.

Experimental Results

Many experiments are conducted to evaluate the performance of the proposed method over the MORPH Album 2 and FG-NET datasets. The datasets, metrics, parameters and experimental setup details are given in the following subsections.

MORPH Album 2 Ageing Database

MORPH Album 2 is an aging database which contains more than 55,000 face images of about 13,000 subjects. These images are captured from 2003 to 2007. Age ranges in this database vary from 16 to 77 years. The main problem of this database is that the distributions of subjects gender and ethnicity are not even, for example, the number of male subjects is 5.5 times more than female subjects and the number of White subjects is about 4 times more than Black subjects' images. The age and gender information of the selected samples are shown in Table 1.

FG-NET Ageing Database

In 2004, the FG-NET aging database (the **F**ace and **G**esture recognition **NET**work) was released in order to help researchers who try to understand the effect of aging on facial appearance. After that, FG-NET was used in many studies in different domains such as in age estimation, age-invariant face recognition, gender classification, and age progression.

The FG-NET database consists of 1002 images from 82 different persons with ages varying between 0 and 69 years old. The subjects' ages are not equally distributed and most of the subjects' ages are less than 40 years old in the database. These images were collected by scanning personal photographs of subjects so they display considerable variability in resolution, image sharpness, and illumination in combination with face viewpoint and

Table 1. The age and gender information of samples randomly selected from MORPH Album 2.

	<20	20–29	30–39	40–49	>50	Total
Male	6543	13849	12322	9905	3321	45940
Female	829	2291	2886	1975	441	8422
Total	7372	16140	15208	11880	3762	54362

expression variation. Occlusions in the form of spectacles, facial hair and hats also exist in a number of images.

Metrics

The most common metric for accuracy evaluation of the age estimators is the Mean Average Error (MAE) which is also used in this study. MAE computes the average age deviation error in absolute terms as follows:

$$MAE = \sum_{i=1}^{N} \frac{|\hat{\alpha}_{\mathrm{i}} - \alpha_{\mathrm{i}}|}{N} \tag{1}$$

where $\hat{\alpha}_{\mathrm{i}}$is the computed age of the i$^{\mathrm{th}}$ subject, α_{i} is its actual annotated age and N is the total number of test subjects.

Preprocessing

Preprocessing is an essential step in image processing systems in order to enhance the quality of the input images. In this study, face images undergo a series of preprocessing steps in order to extract the region of interest of the facial image that is used for age estimation. We used face detection method described by Guo and Mu (2013) for the detection of facial images. Since MORPH contains some tattoo images, we ignore them in our experiments. After this reduction, the number of face images used as a training set and test set is 55244. After face detection step, facial image will be aligned by using geometric transformation such that the eyes have been symmetrically placed at 25% and 75% of the aligned image. The aligned images are resized to 60×60 pixels. Then the fiducial points of face images that are determined by Active Shape Model (ASM) technique (Horng, Lee and Chen 2001) and local neighborhood area around these landmarks will be cropped.

Facial Patches

The facial wrinkles provide discriminative information for age estimation, and these facial wrinkles have been considered by many researchers (Geng, Zhou, and Smith-Miles 2007; Takimoto et al. 2008). Facial fiducial points play an important role for robust face recognition systems, especially for uncontrolled environment. In order to build pose robust face descriptors, we should detect precise facial landmarks. By extracting region enclosed to these fiducial points, we can achieve some patches which have the same semantics for different subject images. For each facial image, firstly, we determine some facial fiducial points by using ASM (Horng, Lee, and Chen 2001). Figure 5 shows the locations of the founded points in a sample image. Afterwards, we crop the face patch by using these landmarks as shown in Figure 6 and extract MRELBP features separately from each of these nine regions and concatenate them together. With the help of this process, we obtain the facial patches that include the wrinkles on the facial images. These patches are then used to estimate the age of the person from his or her facial image.

Experimental Settings

In order to use the MORPH database in a systematic way, we follow the way described by Gehrig, Steiner, and Ekenel (2011) to split the database into five non-overlapped subsets randomly with very important criteria: all of the sample images from a specific subject should be in the one and only one unique fold each time. In all the experiments, the best setting for each visual descriptor is selected. For this purpose, we select the values for each parameter by using fivefold cross-validation technique such that minimize the MAE value. Distribution of the MORPH-2 database images over age in the individual folds is shown in Figure 4.

Most of the studies which evaluate systems using FG-NET, employ leave-one-person-out (LOPO) method. This approach is optimum for the databases including small number of images. In the experiments, for each of 82 persons in the dataset, a separate age estimator model is trained by using 81 remaining images and finally, the average results of these models are reported.

In order to measure the effectiveness of the selected visual descriptors and find the optimal experimental setting, we have investigated the effect of different values for each feature detection algorithm parameters.

For HOG algorithm, namely $HOG_{C,B}$, C is the patch size and B is the number of histogram bins. The optimal parameters have been obtained by

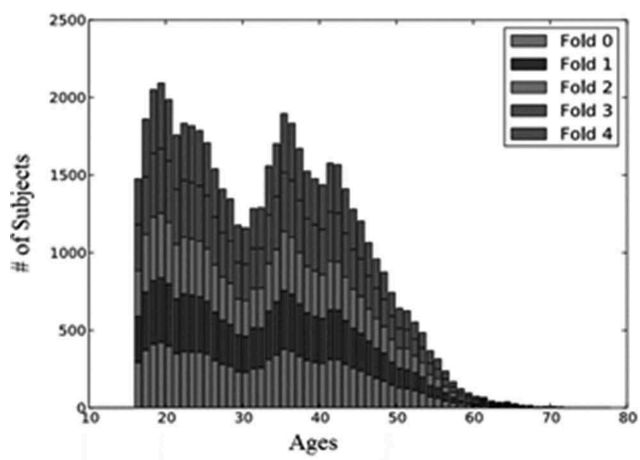

Figure 4. Distribution of the MORPH-2 database images over age in the individual folds (Gehrig, Steiner, and Ekenel 2011).

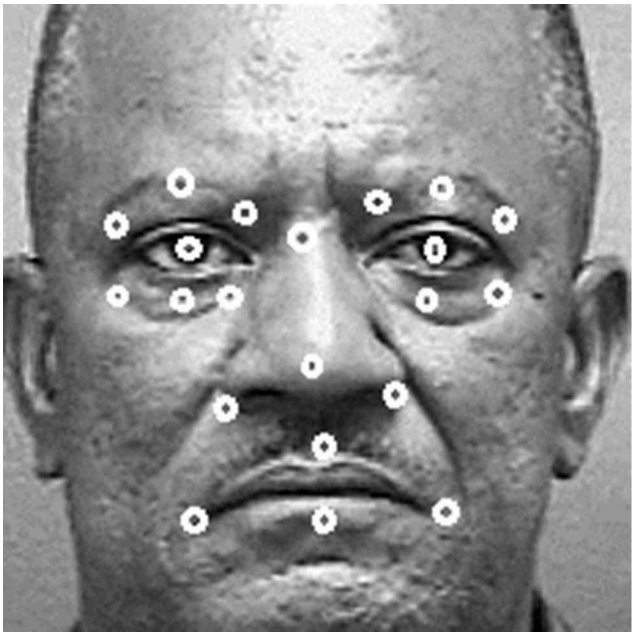

Figure 5. The positions of the detected landmarks on an image.

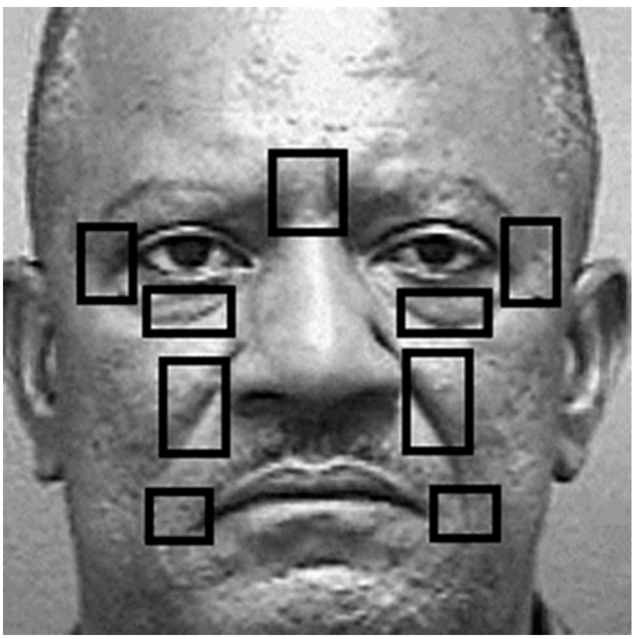

Figure 6. Wrinkle regions used for textural feature extraction.

testing different values for patch size and number of histogram bins with fivefold cross-validation for MORPH dataset and LOPO for FG-NET dataset. Figure 7 shows the MAE of HOG visual descriptors for MORPH dataset, and

	Cx=Cy	number of histogram bins (B)											
		6	7	8	9	10	11	12	13	14	15	16	17
Grid Size (C)	7	5.41	5.15	5.05	4.92	4.90	4.84	4.85	4.83	4.77	4.80	4.82	4.79
	8	5.10	4.90	4.90	4.80	4.74	4.76	4.70	4.68	4.62	4.64	4.60	4.65
	9	4.90	4.72	4.67	4.53	4.55	4.52	4.50	4.45	4.45	4.42	4.43	4.47
	10	4.84	4.65	4.60	4.49	4.50	4.46	4.48	4.43	4.40	4.40	4.40	4.44
	11	4.63	4.53	4.49	4.41	4.40	4.34	4.37	4.31	4.31	4.31	4.37	4.32
	12	4.62	4.51	4.46	4.42	4.39	4.39	4.35	4.30	4.33	4.32	4.34	4.34
	13	4.55	4.44	4.37	4.35	4.37	4.32	4.28	4.26	4.29	4.26	4.31	4.29
	14	4.49	4.39	4.31	4.32	4.34	4.30	4.27	4.25	4.24	4.27	4.29	4.28
	15	4.41	4.29	4.29	4.28	4.27	4.25	4.24	4.19	4.22	4.24	4.28	4.27
	16	4.42	4.31	4.39	4.26	4.33	4.27	4.25	4.23	4.26	4.31	4.31	4.29
	17	4.34	4.29	4.27	4.27	4.26	4.20	4.21	4.21	4.21	4.29	4.28	4.26
	18	4.27	4.22	4.25	4.25	4.22	4.18	4.18	4.20	4.20	4.22	4.21	4.24
	19	4.28	4.20	4.19	4.21	4.19	4.17	$\boxed{4.16}$	4.18	4.18	4.19	4.20	4.23
	20	4.44	4.31	4.27	4.35	4.35	4.34	4.33	4.35	4.34	4.39	4.38	4.42

Figure 7. Results for $HOG_{C,B}$ with varying patch size $C_x$andC_y and number of bins B (columns). The bolded value indicates the optimal result.

the best result is achieved when $C_x = C_y = 19$ and B = 12 in which MAE = 4.16 years. The best result for FG-NET dataset is achieved when $C_x = C_y = 13$ and B = 12 in which MAE = 5.29 years.

The optimal parameters for MRELBP descriptor are found by performing the same procedure. In the case of $MRELBP^{u2}_{R,P}$, the investigation has been performed by testing different values for the number of sampled neighbors (P) and radius (R), limiting the number of neighbors to either 8 or 16. The best result for Morph dataset was achieved by P = 8 and R = 2 and it was MAE = 4.85 years. On the other hand, the best result for FG-Net dataset is achieved by P = 8 and R = 1 and it is MAE = 5.68 years.

Feature-Level and Score-Level Fusion

In order to enhance the system performance and utilize the distinct characteristics of different feature extractors, comprehensive experiments of different fusion level of various pairs of selected visual descriptors have been performed. We tested a different type of feature descriptors for both local feature based and appearance-based classifiers. We used HOG, CLBP, LBP-HF, Haralick features and MRELBP as local feature descriptors and tested LDA and KFA as appearance-based feature extractors. In each method, we used cross-validation approach to find the optimal parameter settings. After feature extraction, in the case of one-descriptor and feature-level fusion, we used a linear SVR for age estimation and in the case of score-level fusion, we used k-Nearest Neighbours (k-NN) classifier.

Feature-level fusion causes increase in the dimensionality of feature space and this expansion causes the curse of dimensionality problem and over-fitting. Therefore, it is not possible to combine all of the features in this level. Table 2 shows the most successful combinations. Feature-level fusion has been obtained by simply concatenating the individually extracted features of separated descriptors. The concatenation of BIF, MRELBP and HOG descriptors achieved the best results in feature level fusion. This combined feature has the advantage of mixing local appearance-based features and textural ones.

In order to solve the aforementioned problem and also to benefit from the other feature descriptors, we performed score-level fusion. According to the success rate of using score-level fusion in multimodal biometric recognition systems, it is believed that accuracy can be improved when the information of two different types of classification systems are consolidated.

The distance between each test sample and its nearest training samples is assumed to be the score of that test sample in the corresponding classification/regression system. These scores are normalized by Min-Max normalization (He et al. 2010) method as follows:

Table 2. Summary of results for different level fusion of various feature extractors that achieve the optimal value.

	Feature Extraction Method					MORPH	FG-NET
Experiment #	MRELBP	HOG	BIF	KFA	Fusion method	MAE	MAE
1	×				N/A	4.93	5.68
2		×			N/A	4.47	5.89
3			×		N/A	4.31	4.61
4				×	N/A	5.21	4.84
5	×	×			Feature-level	4.25	5.59
6	×		×		Feature-level	4.34	4.47
7		×	×		Feature-level	4.29	4.38
8	×	×	×		Feature-level	4.09	4.16
9	×		×	×	Feature-level	4.43	5.01
10			×	×	Feature-level	4.37	4.89
11	×	×		×	Feature-level	4.31	4.91
12	×	×	×	×	Feature-level	4.48	5.36
13	×	×			Score-level	4.30	5.33
14	×		×		Score-level	4.36	4.49
15		×	×		Score-level	4.24	4.52
16	×	×	×	×	Score-level	4.88	5.16
17	×			×	Score-level	4.67	5.08
18		×		×	Score-level	4.33	4.89
19			×	×	Score-level	4.21	4.32
20	*	*		×	Feature-level and score-level	4.19	4.28
21	*		*	×	Feature-level and score-level	4.22	4.19
22		*	*	×	Feature-level and score-level	4.11	4.13
23	*	*	*	×	Feature-level and score-level	**3.89**	**4.06**

In the Experiments 20–23, first, the features denoted by * are concatenated with feature-level fusion method and then the obtained results are combined with the features denoted by × with score-level fusion method.

$$x' = \frac{x - Min(x)}{Max(x) - Min(x)} \quad (2)$$

where x is the raw score, $Max(x)$ and $Min(x)$ are the maximum and minimum values of the raw scores, respectively, and x' is the normalized score. After score normalization, the multimodal score vector$\langle x1, x2 \rangle$ is constructed, with $x1$ and $x2$ corresponding to the normalized scores of two different systems. The next step is fusion at the matching score level. The score vector is combined by Sum rule-based fusion method (He et al. 2010) to generate a single scalar score which is then used to make the final decision as follows:

$$fs = w_1 x_1 + w_2 x_2 \quad (3)$$

The notation w_i stands for the weight which is assigned to one of the two systems and we decided to use equal weights in all of the experiments in order to give equal chance to each feature extractor.

In order to show that score-level fusion can improve the accuracy of age estimation system, we combined different types of feature descriptor scores. The experimental results show that, in all the cases, the score-level fusion causes meaningful improvement in MAE. When we perform both feature-level and score-level fusion together, the MAE is reduced to 3.89 years which is better than all the other methods presented in Table 3. Therefore, it can be claimed that the performance of our method outperforms the other local and appearance-based methods and all the possible combination pairs of these methods with feature-level and score-level fusion.

On the other hand, we compare the proposed method accuracy with the other state-of-the-art methods that presented the results on MORPH Album 2 and FG-NET datasets. Tables 3 and 4 exhibit the MAE of our proposed method and state-of-the-art methods on these two datasets. Recently, in some CNN-based approaches, the authors use some additional datasets (these methods are denoted by * in Table 3) which we believe that it is unfair to consider them in comparison with the methods that only use the original dataset (without any additional images). By this consideration, it is clearly seen that our proposed method outperforms many other methods and its results are comparable with CNN-based approaches which use original dataset. The results demonstrate that combining different features in different fusion levels (feature-level and score-level) enhances the performance of the age estimation system.

Table 3. Comparison with the state-of-the-art methods on Morph-II dataset (* denotes an additional dataset was used).

References	Method/Feature	MAE
Lanitis, Taylor, and Cootes (2002)	WAZ/AAM+BIF	9.21
Geng, Yin, and Zhou (2013)	AAS/AAS+BIF	10.10
Geng, Zhou, and Smith-Miles (2007)	AGES/AAM+BIF	6.61
Chang, Chen, and Hung (2011)	SVM/AAM	6.49
Chang, Chen, and Hung (2011)	OHRank/AAM	6.07
Chang, Chen, and Hung (2011)	OHRank/AAM+BIF	6.28
Guo and Mu (2011)	PLS/BIF	4.56
Guo and Mu (2011)	kPLS/BIF	4.04
Geng, Yin, and Zhou (2013)	IIS-LLD/AAM+BIF	5.67
Geng, Yin, and Zhou (2013)	CPNN/AAM+BIF	4.87
Guo and Mu (2013)	CCA/BIF	5.37
Guo and Mu (2013)	rCCA/BIF	4.42
Guo and Mu (2013)	kCCA/BIF	3.98
Geng, Yin, and Zhou (2013)	MFOR/PCA+LBP+BIF	4.20
Han, Otto, and Jain (2013)	SVM+SVR/BIF+ASM	4.20
Fernández, Huerta, and Prati (2015)	SVR.HOG	4.83
Huertaa et al. (2015)	rCCA/Fusion	4.25
Huertaa et al. (2015)	CNN/CNN	3.88
Yang et al. (2015)	Deeprank/Deep Network	3.57
Han et al. (2015)	DIF/Demographic	3.80
Huerta, Fernández, and Prati (2014)	Fusion	4.25
Niu et al. (2016)*	OR-CNN/CNN	3.27
Rothe, Timofte, and Van Gool (2018)*	DEX/CNN	3.25
Wang, Guo, and Kambhamettu (2015)	DLA/CNN	4.77
Yi, Lei, and Li (2014)*	CNN	3.63
Hu et al. (2017)*	CNN	2.78
Ng et al. (2017)*	CNN	3.88
Proposed Method	**Two-Level Fusion**	**3.89**

Table 4. Comparison with the state-of-the-art methods on FG-NET dataset.

References	Method/Feature	MAE
El Dib and El-Saban (2010)	BIF	3.17
Hong et al. (2013)	BIF,AAM	4.18
Chao, Liu, and Ding (2013)	Label-sensitive learning	4.38
Han, Otto, and Jain (2013)	Component and holistic BIF	4.6
Geng, Yin, and Zhou (2013)	Label distribution(CPNN)	4.76
Guo et al. (2009)	BIF	4.77
Liang et al. (2014)	Hierarchical framework	4.97
Ylioinas et al. (2013)	LBP kernel density estimate	5.09
Günay and Nabiyev (2013)	Local radon features	6.18
Zhang et al. (2013)	Hierarchical model	4.89
Proposed method	**Two-Level Fusion**	**4.06**

Conclusions

A novel age estimation method based on different level of information fusion is proposed in this paper. Biologically inspired features (BIF) and texture-based features such as MRELBP and HOG were involved during the first level of information fusion (Feature-level) process and then the obtained concatenated feature vector was fused with appearance-based method of KFA in the second level of information fusion (score-level). Compared with the state-of-the-art methods, our proposed approach obtained significant lower MAE on MORPH-2 and FG-NET datasets. The experimental results demonstrate that the feature-level and score-level fusion of local features and appearance-based features provide a higher accuracy than the other algorithms. Moreover, the proposed method will be applied on other aging databases under different conditions as future work.

References

Chang, K.-Y., C.-S. Chen, and Y.-P. Hung. 2011. Ordinal hyperplanes ranker with cost sensitivities for age estimation. CVPR, IEEE, 585–92, Colorado Springs, CO.

Chao, W.-L., J.-Z. Liu, and -J.-J. Ding. 2013. Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognition* 46 (3):628–41. doi:10.1016/j.patcog.2012.09.011.

Choi, S. E., Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim. 2011. Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognition* 44 (6):1262–81. doi:10.1016/j.patcog.2010.12.005.

Cootes, T., G. Edwards, and C. Taylor. 1998. Active appearance models. *Proceedings European Conference Computer Vision* 2:484–98.

Cord, M., C. Theriault, and N. Thome 2011. HMAX-S: Deep scale representation for biologically inspired image categorization. Proceedings of IEEE International Conference on Image Process, 1261–64, Brussels, Belgium.

Dalal, N., and B. Triggs 2005. Histograms of oriented gradients for human detection, computer vision and pattern recognition, CVPR 2005. Computer Society Conference on IEEE, 1:886–93, San Diego, CA.

Druzhkov, P. N., and V. D. Kustikova. 2016. A survey of deep learning methods and software tools for image classification and object detection. *Pattern Recognition and Image Analysis* 26 (1):9–15. doi:10.1134/S1054661816010065.

El Dib, M. Y., and M. El-Saban.2010. Human age estimation using enhanced bio-inspired features (ebif). In 2010 IEEE International Conferenc on Image Processing, 1589–92. Hong Kong.

Eskandari, M., and Ö. Toygar. 2014. Fusion of face and Iris biometrics using local and global feature extraction methods. *Signal, Image and Video Processing* 8 (6):995–1006. doi:10.1007/s11760-012-0411-4.

Farmanbar, M., and Ö. Toygar. 2016. Feature selection for the fusion of face and Palmprint biometrics. *Signal, Image and Video Processing* 10 (5):951–58. doi:10.1007/s11760-015-0845-6.

Fernández, C., I. Huerta, and A. Prati. 2015. A comparative evaluation of regression learning algorithms for facial age estimation. In *Face and facial expression recognition from real world Videos*, 133–44. Cham: Springer.

Fu, Y., and T. Huang. 2008. Human age estimatin with regression on discriminative aging manifold. *IEEE Transactions on Multimedia* 10 (4):578–84. doi:10.1109/TMM.2008.921847.

Gehrig, T., M. Steiner, and H. K. Ekenel. 2011. Draft: Evaluation guidelines for gender classification and age.

Geng, X., C. Yin, and Z. H. Zhou. 2013. Facial age estimation by learning from label distributions. *Tpami, IEEE* 35:2401–12. doi:10.1109/TPAMI.2013.51.

Geng, X., Z.-H. Zhou, and K. Smith-Miles. 2007. Automatic age estimation based on facial aging patterns. *Tpami* 29 (12):2234–40. doi:10.1109/TPAMI.2007.70733.

Gunay, A., and V. V. Nabiyev. 2008. Automatic age classification with lbp, 23rd International Symposium on Computer and Information Sciences, ISCIS'08, 1–4, Istanbul.

Günay, A., and V. V. Nabiyev. 2013. Age estimation based on local radon features of facial images. In *Computer and information sciences III*, 183–90. London: Springer.

Guo, G., and G. Mu. 2011. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. CVPR, IEEE, 657–64, Colorado Springs, CO.

Guo, G., and G. Mu. 2013. Joint estimation of age, gender and ethnicity: CCA vs. PLS'. 10th International Conference on Automatic Face and Gesture Recognition, IEEE, 1–6, Shanghai.

Guo, G., and G. Mu. 2014. A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing* 32 (10):761–70. doi:10.1016/j.imavis.2014.04.011.

Guo, G., G. Mu, Y. Fu, and T. S. Huang. 2009. Human age estimation using bio-inspired features, CVPR, IEEE, 112–19, Miami, FL.

Han, H., C. Otto, and A. K. Jain. 2013. Age estimation from face images: Human vs. machine performance. International Conference on Biometrics (ICB), IEEE, 1–8, Madrid.

Han, H., C. Otto, X. Liu, and A. Jain. 2015. Demographic estimation from face images: Human vs. machine performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (6):1148–61. doi:10.1109/TPAMI.2014.2362759.

He, M., S. Horng, F. Fan, and R. Run. 2010. Performance evaluation of score level fusion in multimodal, biometric systems. *Pattern Recognition* 43:1789–800. doi:10.1016/j.patcog.2009.11.018.

Hong, L., D. Wen, C. Fang, and X. Ding.2013. A new biologically inspired active appearance model for face age estimation by using local ordinal ranking. Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, ACM, 327–30, Huangshan, China.

Horng, W., C. Lee, and C. Chen. 2001. Classification of age groups based on facial features. *Tamkang Journal of Science and Engineering* 4 (3):183–92.

Hu, Z., Y. Wen, J. Wang, M. Wang, R. Hong, and S. Yan. 2017. Facial age estimation with age difference. *IEEE Transactions on Image Processing* 26 (7):3087–309. doi:10.1109/TIP.2016.2633868.

Huerta, I., C. Fernández, and A. Prati. 2014. Facial age estimation through the fusion of texture and local appearance descriptor. In *European conference on computer vision*, 667–81. Cham: Springer.

Huertaa, I., C. Fernándezb, C. Segura, J. Hernando, and A. Prat. 2015. A deep analysis on age estimation. *Pattern Recognition Letters* 68 (2):239–49. doi:10.1016/j.patrec.2015.06.006.

Izadpanahi, S., and Ö. Toygar. 2014. Human age classification with optimal geometric ratios and wrinkle analysis. *International Journal of Pattern Recognition and Artificial Intelligence* 28 (2):1456003. doi:10.1142/S0218001414560035.

Lanitis, A., C. Draganova, and C. Christodoulou. 2004. Comparing different classifiers for automatic age estimation. *TSMC-B* 34 (1):621–28.

Lanitis, A., C. J. Taylor, and T. F. Cootes. 2002. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Mchine Intelligence* 24 (4):442–55. doi:10.1109/34.993553.

Liang, Y., X. Wang, L. Zhang, and Z. Wang. 2014. A hierarchical framework for facial age estimation. *Mathematical Problems in Engineering* 2014:1–8.

Liu, C. 2006. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (5):725–37. doi:10.1109/TPAMI.2006.40.

Liu, L., P. Fieguth, Y. Guo, M. X. Wang, and M. Pietikäinen. 2017. Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition* 62:135–60. doi:10.1016/j.patcog.2016.08.032.

Liu, L., S. Lao, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen. 2016. Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing* 25 (3):1368–81. doi:10.1109/TIP.2016.2522378.

Marcialis, G., and F. Roli. 2007. Score-level fusion of fingerprint and face matchers for personal verification under stress conditions', 14th IEEE International Conference on Image Analysis and Processing ICIAP 2007, 259–64, DC, USA.

Mirzaei, F., and Ö. Toygar. 2011. Facial age classification using subpattern-based approaches. The 2011 International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV2011), Nevada, USA.

Montillo, A., and H. Ling. 2009. Age regression from faces using random forests, ICIP, IEEE,2465–68, Cairo.

Ng, C. C., Y. T. Cheng, G. S. Hsu, and M. H. Yap. 2017. Multi-layer age regression for face age estimation, 15th IAPR International Conference on Machine Vision Applications (MVA), 294–97, Nagoya. doi:10.1016/j.jenvman.2017.06.049

Niu, Z., M. Zhou, L. Wang, X. Gao, and G. Hua.2016. Ordinal regression with multiple output cnn for age estimation. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada.

Ojala, T., M. Pietikäinen, and T. Maenpää. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7):971–87. doi:10.1109/TPAMI.2002.1017623.

Panis, G., A. Lanitis, N. Tsapatsoulis, and T. F. Cootes. 2016. Overview of research on facial ageing using the FG-NET ageing database. *IET Biometrics* 5 (2):37–46. doi:10.1049/iet-bmt.2014.0053.

Patil, A., and D. Bhalke. 2016. Fusion of fingerprint, Palmprint and Iris for person identification. International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) International Institute of Information Technology (I^2IT), Pune.

Ranjan, R., S. Sankaranarayanan, C. D. Castillo, and R. Chellappa. 2017. An all-in-one convolutional neural network for face analysis. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 17–24, Washington, DC.

Ricanek, K., and T. Tesafaye. 2006.MORPH: A longitudinal image database of normal adult age-progression'. FG, 341–45, Southampton, UK.

Riesenhuber, M., and T. Poggio. 1999. Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2:1019–25. doi:10.1038/14819.

Rothe, R., R. Timofte, and L. Van Gool. 2018. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision* 126 (2-4):144–57.

Scandrett, C. M., C. J. Solomona, and S. J. Gibsona. 2006. A person-specific, rigorous aging model of the human face. *Pattern Recognition Letters* 27 (15):1776–87. doi:10.1016/j.patrec.2006.02.007.

Shan, C. 2010. Learning local features for age estimation on real-life faces. ACM International Workshop on Multimodal Pervasive Video Analysis, Florence, Italy.

Sim, H., A. Hishammuddin, H. Rohayanti, and O. Razib. 2014. Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images. *Expert Systems with Applications* 41:5390–404. doi:10.1016/j.eswa.2014.02.051.

Takimoto, H., Y. Mitsukura, M. Fukumi, and N. Akamatsu. 2008. Robust gender and age estimation under varying facial pose. *Electronics and Communications in Japan* 91 (7):32–40. doi:10.1002/ecj.v91:7.

Theriault, C., N. Thome, and M. Cord. 2011. HMAX-S: Deep scale representation for biologically inspired image categorization. 18th IEEE International Conference on Image Processing, 1261–64, Brussels.

Thériault, C., N. Thome, and M. Cord. 2013. Extended coding and pooling in the HMAX model. *IEEE Transactions On Image Processing* 22 (2):764–77. doi:10.1109/TIP.2012.2222900.

Wang, X., R. Guo, and C. Kambhamettu. 2015. Deeply-learned feature for age estimation. IEEE Winter Conference on Applications of Computer Vision (WACV), 534–41, Hawaii.

Weng, R., J. Lu, G. Yang, and Y.-P. Tan. 2013. Multi-feature ordinal ranking for facial age estimation. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1–6, Shanghai.

Yang, H.-F., B.-Y. Lin, K.-Y. Chang, and C.-S. Chen. 2015. Automatic age estimation from face images via deep ranking. In *Proceedings of the British Machine Vision Conference (BMVC)*, eds. X. Xie, M. W. Jones, and G. K. L. Tam, 55.1–55.11. BMVA Press. September.

Yang, Z., and H. Ai. 2007. *Demographic classification with local binary patterns, advances in biometrics*, 464–73. Heidelberg, Berlin: Springer.

Yi, D., Z. Lei, and S. Z. Li. 2014. Age estimation by multi-scale convolutional network. *Proceedings of Asian Conference on Computer Vision*, 9005, 144–58, Singapore.

Ylioinas, J., A. Hadid, X. Hong, and M. Pietikäinen. 2013. Age estimation using local binary pattern kernel density estimate. Proceedings Image Analysis and Processing, ICIAP 2013. Lecture Notes in Computer Science, Springer, 8156, 141–50, Naples, Italy.

Youssef, E., Z. Elberrichi, and R. Adjoudj. 2010. Score level fusion based multimodal biometric identification (fingerprint & voice). 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, 146–50, Sousse. doi:10.1177/1753193408098907

Zhang, L., X. Wang, Y. Liang, and L. Xie. 2013. A new method for age estimation from facial images by hierarchical model. Proceedings of 2nd International Conference on Innovative Computing and Cloud Computing, Wuhan, China