

Age Estimation Based on AAM and 2D-DCT Features of Facial Images

Asuman Günay and Vasif V. Nabiyev
Department of Computer Engineering
Karadeniz Technical University
Trabzon, TURKEY

Abstract—This paper proposes a novel age estimation method - Global and Local feAture based Age estiMation (GLAAM) - relying on global and local features of facial images. Global features are obtained with Active Appearance Models (AAM). Local features are extracted with regional 2D-DCT (2-dimensional Discrete Cosine Transform) of normalized facial images. GLAAM consists of the following modules: face normalization, global feature extraction with AAM, local feature extraction with 2D-DCT, dimensionality reduction by means of Principal Component Analysis (PCA) and age estimation with multiple linear regression. Experiments have shown that GLAAM outperforms many methods previously applied to the FG-NET database.

Keywords—2D-DCT; AAM; Age estimation; PCA; Regression

I. INTRODUCTION

The wide-ranging topic of facial image (FI) processing has been receiving considerable interest lately because of its real world applications such as forensic art, electronic consumer relationship management, security control and surveillance, cosmetology, entertainment and biometrics. In the FI context, age recognition (or estimation) has been demanding growing attention. Age synthesis, also called age progression is defined as re-rendering FIs with natural and rejuvenating effects. Age estimation (AE) can be defined as the process of associating a FI automatically with an exact age or age group.

In order to facilitate AE, suitable facial representations are necessary. Otherwise, even the most robust classifiers will fail due to the inadequacy of the domain where the feature recognition is done [1]. Hence, the design of face recognition systems requires careful selection of the face feature recognition (FFR) domain. Some issues that should be contemplated are: (i) good discrimination of different people with tolerance to discrepancies inside a class; (ii) FFR must be effortlessly performed from raw face images to speedup processing; and (iii) the FFR must lie in a low dimensional space, in order to facilitate the implementation of the classifiers.

The FI characteristics make the FFR problem very difficult to solve. The most important hindrances are: (1) AE is not a standard classification problem; (2) a large aging database, especially a chronometrical image series of an individual is often hard to collect; and (3) real world age progression displayed on faces is uncontrollable and personalized.

Several techniques have been suggested to represent FIs for recognition purposes, but there is still no consensus on the best when it comes to age recognition/classification. Appearance-based techniques consider an FI as a 2D array of pixels and focus on deriving descriptors for face appearance without precise geometrical representations. Holistic (nonparametric) methods such as the Principal Component Analysis (PCA) [2] and the Linear Discriminant Analysis (LDA) [2, 3] along with more recent approaches like 2D-PCA [3, 4] and 2D-LDA [3] have been broadly studied. Other important approaches handle local descriptors, as for example, Scale Invariant Feature Transform (SIFT), and Affine-SIFT (ASIFT) [5, 6], and they have gained increasing awareness thanks to their robustness to problems akin to pose and illumination alterations [7, 8].

In this paper, we propose a novel Global and Local feAture based Age estiMation (GLAAM) method as shown in Fig. 1. The input images are normalized and the local features are extracted using regional 2D-DCT (2-dimensional Discrete Cosine Transform). Global features are obtained with Active Appearance Models (AAM). After feature extraction, dimensionality reduction is performed with PCA. Then, AE is cast as a regression problem. Our method uses global and local considerations and does not rely on a complex Bayesian framework [9]; besides that, it is simple and relatively fast when compared to other ones.

A survey on AE is given in the next section. Section 3 introduces the proposed method for AE including preprocessing, feature extraction, dimensionality reduction and regression modules. In Section 4, experimental results are given and Section 5 concludes the paper.

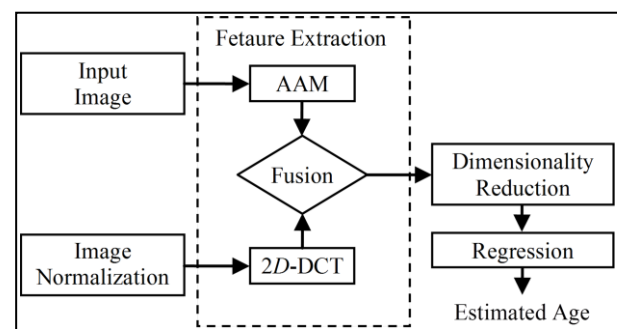


Fig. 1. System structure

II. AGE ESTIMATION METHODS

AE can be seen either as a multi-class classification problem or a regression problem. The existing AE systems typically consist of an age image representation and an AE module. Age image representation techniques rely often on shape-based and texture-based features that were extracted from FIs. They can be grouped under the topics of Anthropometric Models, AAM, AGing pattErn Subspace (AGES), Age Manifold and Appearance Models. Then, AE can be performed with age group classification or regression methods. In recent studies hybrid systems using classification and regression techniques together are presented [10]. Robust multi-instance regressor learning algorithm is also used to build a universal human age estimator, based on facial information [11].

The anthropometric model based representations only consider the facial geometry. The earliest paper published in the area of age classification relying on facial geometry was the work by Kwon and Lobo [12]. They used craniofacial development theory [13] which uses a mathematical model to describe the growth of a person's head. They computed six ratios of distances on frontal images to separate babies from adults. This AE method can only deal with young ages since the human head shape doesn't change too much in its adulthood. So Kwon and Lobo [12] used wrinkle information to separate young adults from senior adults. They used a very small database containing 45 images in their experiments. Later on Horng et al. [14] and Dehshibi and Bastanfard [15] proposed age classification methods using distance ratios based on face anthropometry as geometric features and wrinkle information as texture features.

AAM [16, 17] based approaches consider both shape and texture rather than just the facial geometry as in the anthropometric model based methods. AAM uses a statistical model of object shape and appearance to synthesize a new image throughout a training stage which provides to the training supervisor a set of images and coordinates of landmarks existing in all of the images. AAMs represent a familiar group of algorithms for fitting shape models to images. Training a model requires labeling a database of images where a set of locations called landmarks typify the object group in question. The formulation in [17] chooses a linear and generative model, i.e. an explicit model of the input data has to be provided. This leads to an iterative Gauss-Newton type procedure, where the error between the current image features and those synthesized using the current location of the model in the image are used to derive additive updates to the shape model parameters. Nonetheless, the computational load is heavy, since an explicit image feature model must be stated and evaluated at each algorithm iteration [16]. Lanitis et al. [18] extended AAMs for aging faces by proposing an aging function, $age=f(b)$ which explains the variation in age. But they have to deal with each aging face image separately. Kohli et al. [19] extracted feature vectors from images using AAMs and used ensemble of classifiers trained on different dissimilarities to distinguish between child/teen-hood and adulthood. By using the different aging functions, accurate age of the classified image is estimated. Chao et al. [20] proposed an age estimation method using

AAM features. Their approach is based on label sensitive learning and age-oriented regression.

Geng et al. [21, 22] proposed a method called AGES that defines a sequence of personal face images of the same person sorted in temporal order. Then, a specific aging pattern is learned for each individual. AGES method can synthesize the missing age images by using an expectation maximization-like iterative learning algorithm.

Instead of learning a specific aging pattern for each individual as in AGES, age manifold [23] methods can learn a common aging trend or pattern from many individuals at different ages. This kind of aging pattern learning helps face aging representation. Age manifold utilizes a manifold embedding technique to discover the aging trend in a low dimensional domain from many face images at each age. Thus, the mapping from the image space to the manifold space can be done either by linear or by nonlinear functions [24-28], such as $Y=P(X, L)$, where X is the image space sampled by a set of face images, $X = [x_i: x_i \in R^D]_{i=1}^n$. A ground truth set $L = [l_i: l_i \in N]_{i=1}^n$ associated with images provides the age labeling. $Y = [y_i: y_i \in R^d]_{i=1}^n$ with $d \leq D$ is the low-dimensional representation of X in the embedded subspace. Compared with AGES, all ages of different individuals could be used together in age manifold. The only requirement for the age manifold representation is that the size of the training data set should be large enough in order to learn the embedded manifold with statistical sufficiency.

Appearance models are mainly focused on aging-related facial feature extraction. Both global and local features were used in existing AE systems. Fukai et al. [29] applied Fast Fourier Transform to extract feature spectrum from facial appearance and used genetic algorithm for feature selection. Local Binary Patterns (LBP) have been used as effective texture descriptors for appearance feature extraction [30]. Ju and Wang [31] selected regions with Adaboost that vary with the aging process and used LBP histograms of these regions for AE. Gabor features have also been tried on AE tasks and proved to be more effective than LBP [32]. Guo et al. introduced the biologically inspired features (BIF) to model the aging process on faces [33]. The first layer is created using Gabor filters on facial images. In the second layer, they have proposed to use a novel operator. Their experimental results have shown significant improvement in age estimation accuracy over previous methods. El-Dib and Onsi [34] used BIF to analyze the different facial parts: eye wrinkles, internal face and whole face. According to their analysis, the eye wrinkles contain the most important aging features compared with others. Li et al. [35] and Han et al. [36] also used BIF in their age estimation frameworks.

III. THE GLOBAL AND LOCAL FEATURE BASED AGE ESTIMATION

This paper introduces an innovative AE method – known as GLAAM – relying on local and global facial features of images. Local features are extracted using regional 2D-DCT of normalized FIs and the global features are produced by AAMs. This method consists of the following modules: (1) face normalization, (2) global feature extraction with AAM

and local feature extraction with 2D-DCT, (3) dimensionality reduction with PCA, and (4) AE with multiple linear regression.

A. Face Normalization

Since shape and local variations of images during aging suffer an evident influence from rotation, scaling and translation, all the images have to be compatible with a common shape model produced by means of a training set of samples. In order to train the shape model, each image is represented by the coordinates of 68 landmark points (Fig. 2-a). Then, the statistical shape model is trained and all images are warped to the mean shape, so that shape variations within the training set are eliminated. The warping process employs affine transformation and Delaunay triangulation (Fig. 2-b,c). Since the FIs vary in head pose, the warped images are inclined to the left as shown in Fig. 2-c. So we rotate these images and cropped the main face part and scaled to the size of 88×88 (Fig. 2-d). Thus we almost eliminate the unreasonable regions for feature extraction. In the local feature extraction phase the images are divided into blocks each having 8×8 block size. So we scale images to the size of 88×88 which is large enough for age related feature extraction. This image size is also efficient in terms of computational costs.

B. Feature Extraction

The feature extraction module consists of two phases: global feature extraction with AAM and local feature extraction with 2D-DCT computation. These steps will be explained in detail in the following sections.

1) *Global Feature Extraction with AAM*: AAM is a statistical shape and appearance model of FIs [17]. These models are generated by combining a model of shape variations with a model of the appearance variations in a shape-normalized frame. A statistical shape model can be generated with a training set of face images labeled with landmark points as shown in Fig. 2-a. Let us represent all the landmark points of training images by $X = [x_i; x_i \in R^D]_{i=1}^n$. The mean shape is produced with taking the mean of the landmark points in the training set as $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Then, PCA is applied to the data to extract the main principal components along which the training set varies from the mean shape. If the total scatter matrix S is defined as $S = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$, the projection is chosen to maximize the determinant of the total scatter matrix of the projected samples as $\arg \max_{P_s} |P_s^T S P_s|$. P_s is the set of eigenvectors of S corresponding to the d largest eigenvalues. Then a linear transformation maps the D -dimensional data space into a d -dimensional parameter space where $d \leq D$. The shape parameters are defined by linear formulation as $b_s = P_s^T X$. As a result, any training set of images can be approximated by,

$$x = \bar{x} + P_s b_s \quad (1)$$

where \bar{x} is the mean shape, P_s is a set of orthogonal principal modes of variation and b_s is a set of shape parameters.

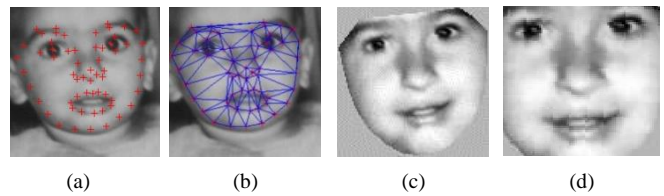


Fig. 2. a) Example of face image labeled with 68 landmark points b) Result of the Delaunay triangulation used in warping process c) Result of face normalization d) Facial image used in the local feature extraction phase

To build a statistical appearance model, each image has to be normalized, so that its control points match the mean shape using a Delaunay triangulation (Fig. 2.b, c). Then, the gray-level intensities within a pre-specified image region are stacked to form vector g are used for training an intensity model. By applying PCA to the gray level intensities, a linear model is obtained as follows:

$$g = \bar{g} + P_g b_g \quad (2)$$

where \bar{g} is the mean gray-level vector, P_g is a set of orthogonal modes of variation and b_g is a set of gray-level parameters. The shape and appearance of any image can be summarized by the vectors b_s and b_g . Since there may be correlations between the shape and gray-level variations, a further PCA is applied to them and, finally, the combined shape and appearance parameters are obtained. For the intensity model, approximately 7000 gray-level intensities in the facial region of the corresponding shape-free image are used to represent the training samples. The resulting combined shape and intensity model requires 277 model parameters to explain 95 percent of the variance in the training set. These model parameters are used as a global descriptor of FI's.

2) *Local Feature Extracting with 2D-DCT*: DCT is an invertible linear transform that can express a finite sequence of data points in terms of a sum of cosine functions. The original signal is converted to the frequency domain by applying the direct DCT transform and it is possible to convert back the transformed signal to the original domain by applying the inverse DCT transform (IDCT). After the original signal has been transformed, its DCT coefficients reflect the importance of the frequencies that are present in it.

The 2D-DCT is commonly used as a pre-processing step in face recognition, because it attenuates the problems created by changes due to illumination angles, face occlusions, colors and pose [37]. Using the face images directly for recognition purposes resulted in inefficiencies because of the high information redundancy and correlation in such images. Therefore DCT is widely used as a feature extraction and compression method in various applications due to its properties such as de-correlation, energy compaction, separability and orthogonality [38]. All these properties lead us to use 2D-DCT in AE field.

De-correlation: The principle advantage of image transformation is the removal of redundancy between neighboring pixels. This leads to un-correlated transform coefficients which can be encoded independently without compromising coding efficiency.

Energy-compactness: Efficacy of a transformation scheme can be directly gauged by its ability to pack input data into as few coefficients as possible. This allows the quantizer to discard coefficients with relatively small amplitudes without introducing visual distortion in the reconstructed image. DCT exhibits large variance distribution in a small number of coefficients for highly correlated images such as face images. In other words DCT packs energy in the low frequency regions. Therefore some of the high frequency content can be discarded without significant quality degradation.

Separability: The 1D-DCT (1 dimensional DCT) transform can be represented as:

$$F(k) = \alpha(k) \sum_{n=0}^{N-1} f(n) \cos\left[\frac{(2n+1)k\pi}{2N}\right] \quad (3)$$

The 2D-DCT transform can be expressed as:

$$F(k_1, k_2) = 1D - DCT_{y-dir}(1D - DCT_{x-dir}) \quad (4)$$

This property has the principle advantage that $F(k_1, k_2)$ can be computed in two steps by successive 1D operations on rows and columns of an image.

Orthogonality: In pattern recognition techniques to make the model computationally efficient, transform orthogonality is as important as the class separation in applications like face recognition. Unlike Gabor elementary functions, which are a set of overlapping functions and not mutually orthogonal, the DCT basis functions are orthogonal. In addition to its de-correlation characteristics, this property renders some reduction in the pre-computation complexity.

The 2D-DCT definition is given in (5).

$$X_{k_1, k_2} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1, n_2} \cos\left[\frac{\pi}{N_1}\left(n_1 + \frac{1}{2}\right)k_1\right] \cos\left[\frac{\pi}{N_2}\left(n_2 + \frac{1}{2}\right)k_2\right] \quad (5)$$

In the proposed method, the normalized 88×88 images are divided into 11×11 blocks each having dimension of 8×8 and 2D-DCT is applied to them. This block size was adopted by the JPEG compression standard [39]. In the developmental phases the processing of larger blocks was seen as being prohibitively slow for the computer to execute. Also the experts observed that the use of larger blocks did not result in appreciably greater compression and quantization artifacts become more visible as the block size increases.

In practice for a wide range of images and viewing conditions, 8×8 has been found to be the optimum DCT block size and is specified in most current coding standards. After applying 2D-DCT we have 64 coefficients for each block. To eliminate the high frequency coefficients, quantization is performed. Coefficients are arranged in a vector following a zigzag fashion and the first 21 coefficients to represent that image block. Hence, the dimension of a local feature vector is 11×11×21=2541.

After the global and local features are extracted, they are combined in a single vector in order to perform dimensionality reduction with PCA. To combine the global and local features,

a feature level fusion approach is used. For this purpose, the feature vectors are normalized by the z-score normalization as:

$$\hat{f}_{i,j} = \frac{f_{i,j} - \mu_j}{\sigma_j} \quad \text{with } j=1,2 \text{ and } i=1,\dots,n \quad (6)$$

where n is the number of images, $f_{i,j}$ is the j -th feature vector of i -th image and μ_j, σ_j are the mean and standard deviation of feature vector $f_{i,j}$, respectively. Then, the fused feature vector is created by concatenating the normalized global and local feature vectors as follows:

$$f_{i,fused} = [\hat{f}_{i,1} \hat{f}_{i,2}] \quad i=1,\dots,n \quad (7)$$

C. Dimensionality Reduction

After the feature extraction module, PCA is performed in order to find a lower dimensional subspace which carries significant information for AE. Then, high-dimensional feature vectors are projected onto a low-dimensional subspace in order to improve the efficiency. Using this technique the p -dimensional feature vector f is transformed into a d -dimensional vector y with $d \leq p$.

The PCA method finds the embedding that maximizes the projected variance given below.

$$W_{opt} = \arg \max_{\|W\|=1} W^T S W \quad (8)$$

In (8) $S = \sum_{i=1}^n (f_i - \bar{f})(f_i - \bar{f})^T$ is the scatter matrix, and \bar{f} is the mean vector of $\{f_i\}_{i=1}^n$. The solution of this problem is given by the set of d eigenvectors associated to the d largest eigenvalues of the scatter matrix. Once the projection subspace is determined, training and testing images were projected on it. The low dimensional representation of feature vectors is calculated with $y_i = W^T f_i$ allowing thus dimensionality reduction.

D. Regression

After finding a lower dimensional representation of facial images, we recast the AE problem as a multiple linear regression as follows:

$$age = F(M): \leftrightarrow \hat{L} = \hat{F}(Y) \quad (9)$$

where \hat{L} denotes the estimated age label, $F(\cdot)$ the unknown regression function, and $\hat{F}(\cdot)$ is the estimated regression function. The corresponding matrix formulation is

$$L = \tilde{Y}B + e, \quad Var(e) = \sigma^2 I \quad (10)$$

where L is the age label vector. \tilde{Y} is a known matrix including a column of 1s for the intercept and observed values. The vector B is the unknown parameter vector which we need to estimate during the learning stage. The error vector e consists of unobservable random variables, and assumed to have zero mean and uncorrelated with common variance σ^2 . In fitting model, B is estimated by ordinary least squares $\hat{B} = (\tilde{Y}'\tilde{Y})^{-1}\tilde{Y}'L = HL$ or robust regression, and the fitted value of L is given by,

$$\hat{L} = \tilde{Y}\hat{\beta}. \quad (11)$$

The vector of residuals is $\hat{e} = L - \hat{L}$, with $E(\hat{e})=0$ and $Var(\hat{e}) = \sigma^2(I - H)$. The age regression function used in this study is a linear function given by

$$\hat{l} = \hat{\beta}_0 + \hat{\beta}_1^T y \quad (12)$$

where \hat{l} is the estimate of age, $\hat{\beta}_0$ is the offset, $\hat{\beta}_1$ is the weight vector and y is the extracted feature vector.

IV. EXPERIMENTS AND RESULTS

In this paper, the FG-NET Aging Database [40] is used to train and test the proposed method. This database contains 1,002 face images from 82 subjects with approximately 10 images per subject. The ages in the database are distributed in a wide range from 0 to 69. The age distribution of the FG-NET database is given in Table 1. One can see from the table that the images are not distributed uniformly.

A typical aging sequence from the FG-NET database is shown in Fig. 3-a. Besides the aging variation, most aging sequences display variations in pose, illumination, facial expression, occlusion, etc. Although these variations may increase computational complexity, all the images have been used in the experiments to avoid restrictions.

The normalization phase determines the mean shape from the 68 landmarks of training samples. Next, all images are warped to the mean shape (Fig. 3.b) using affine transformation and Delaunay triangulation and scaled to the size of 88x88. Furthermore, each image is represented with 277 AAM model parameters that are used as global face features.

In the local feature extraction step, the normalized 88x88 images are divided into 11x11 blocks each having 8x8 size and 2D-DCT is applied to them. After 2D-DCT computing, we have 64 coefficients for each block. To eliminate the high frequency coefficients, quantization is performed. Coefficients are arranged in a vector according to a zigzag fashion. In this phase the determination of the number of DCT coefficients is done experimentally. For this purpose, the AE performance of different number of coefficients is calculated and the results are listed in Table 2. We can see from Table 2 that using the 21 DCT coefficients gives better results than other ones. So the first 21 coefficients are selected to represent each image block. Hence, the dimension of the local feature vector is 11x11x21 (121 blocks with 21 entries/block). Then global and local feature vectors are normalized according to their mean

TABLE I. AGE RANGE DISTRIBUTION OF THE IMAGES IN FG-NET DATABASE

Age Range	FG-NET (%)
0-9	37.03
10-19	33.83
20-29	14.37
30-39	7.88
40-49	4.59
50-59	1.50
60-69	0.80

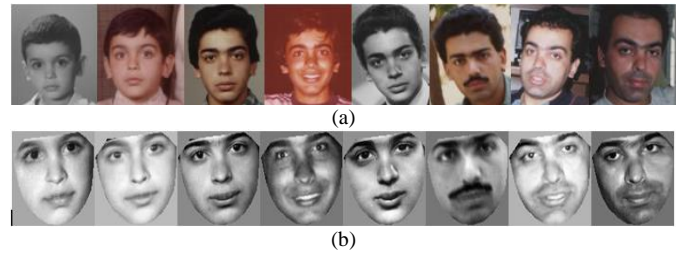


Fig. 3. (a) Typical aging face sequence in FG-NET Aging Database; and (b) Normalized face sequence

TABLE II. THE ESTIMATION RESULTS OF DIFFERENT NUMBER OF DCT COEFFICIENTS

Number of DCT Coefficients	MAE
1	7,25
3	6,85
6	6,65
10	6,43
15	6,26
21	6,18
28	6,20
36	6,19

and standard deviation and concatenated into a single vector. After that, a low dimensional age manifold is learned with PCA. AE is performed with multiple linear regression in the low dimensional space.

Performance evaluation is done by means of a cross-validation variant known as the leave-one-person-out (LOPO), i.e., in each fold the images of one person are used as test set and those of the others are used as the training set. As FG-NET contains face images from 82 subjects, after 82 folds, each subject has been used as test set once, and the final results are calculated based on all estimations. This scenery is very close to real life applications and, hence, it is very adequate for testing. The Mean Absolute Error (MAE) has been chosen as a metric for performance comparison. MAE is defined as the average of the absolute error between the recognized labels and the ground truth labels as follows:

$$MAE = \sum_{i=1}^N |\hat{l}_i - l_i| / N_t \quad (13)$$

where \hat{l}_i is the recognized age for the i th testing sample, l_i is the corresponding ground truth, and N_t is the total number of the testing samples.

The estimation results of earlier methods and GLAAM are listed in Table 3. As one can infer from Table 3, GLAAM achieves better results than earlier methods like WAS, AAS, KNN and AGES on the FG-NET database. DCT encodes facial texture and edge information in the frequency domain. Moreover local appearance information is captured using the block based DCT, but the global ones are ignored. So we use AAM, because it encodes the geometrical and global facial texture information in spatial domain. These feature sets capture differential complementary information. The combination of these feature vectors outperforms the AE accuracy of each one of the feature vector alone.

We also investigate the AE performance of our method in various age ranges. The estimation results are given in Table 4. From Table 4 we can observe that, GLAAM outperforms the AE accuracy of global features and local features alone, almost in all age ranges. As the age variation in the age range increased, the effectiveness of our method became more outstanding as shown in Fig. 4.

V. CONCLUSION

In this paper, an AE method relying on an AAM model named GLAAM has been introduced. Its main contribution is a set of parameters accounting for both global texture features as well as local features of FIs. Locality is preserved by regional DCT coefficients and this is the main advantage/contribution of GLAAM over its competitors because DCT captures more accurately local features in FIs. Moreover, the proposed method is simple and relatively fast when compared to other ones used as benchmark, because 2D-DCT is recast and computed by means of 1D-DCT operations.

Shape variations are eliminated via normalization with respect to mean shape obtained from a training set consisting of FIs. Furthermore, these local features are combined with

TABLE III. THE COMPARISON OF ESTIMATION RESULTS ON FG-NET DATABASE

Methods	MAEs
WAS [17]	8.06
AAS [17]	14.83
KNN [18]	8.24
BP [18]	11.85
SVM [18]	7.25
AGES [17]	6.77
AGES _{da} [18]	6.22
AAM	6.07
2D-DCT	6.18
AAM+2D-DCT	5.39

TABLE IV. AGE ESTIMATION RESULTS AT DIFFERENT AGE RANGES IN FG-NET DATABASE

Age Group	#img.	AAM	2D-DCT	AAM+2D-DCT
0-9	371	2,63	1,35	1,62
0-19	710	2,81	2,46	2,34
0-29	854	3,77	3,45	3,24
0-39	933	4,77	4,59	4,20
0-49	979	5,51	5,48	4,91
0-59	994	5,89	5,85	5,23
0-69	1002	6,07	6,18	5,39

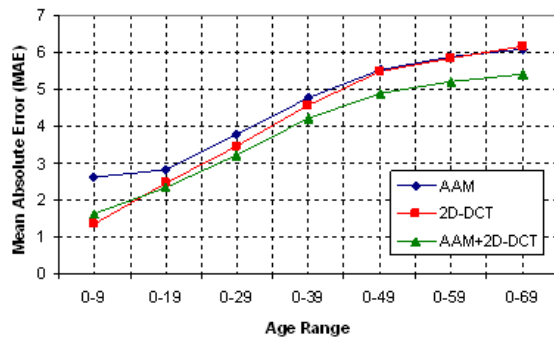


Fig. 4. MAEs at different age ranges in FG-NET

global features of images. Experimental results using the FG-NET aging database show that GLAAM is better than earlier methods. However, there is plenty of room for research, since there are methods that do not require normalization such as SIFT and ASIFT. An extra improvement in GLAAM would be the use of Principal Component Regression, since it combines PCA and Regression in the same stage.

REFERENCES

- [1] A. Hadid, M. Pietikäinen and T. Ahonen, "A discriminative feature space for detecting and recognizing faces", Proc. of Computer Vision and Pattern Recognition, pp.797-804, 27 June-2 July 2004.
- [2] W. Chen, M. J. Er and S. Wu, "PCA and LDA in DCT domain", Pattern Recognition Letters, vol. 26, no. 15, pp. 2474-2482, 2005.
- [3] M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix", Pattern Recognition Letters, vol. 26, no. 5, pp. 527-532, 2005.
- [4] H. Yu and M. Bennamoun, "1D-PCA, 2D-PCA to nD-PCA", 18th International Conference on Pattern Recognition (ICPR'06), pp. 181-184, 2006.
- [5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [6] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison", SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 438-469, 2009.
- [7] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kalviainen and J. Matas, "Feature-based affine-invariant localization of faces", IEEE Transactions on Pattern Recognition and Machine Intelligence, vol. 27, no. 9, pp. 1490-1495, 2005.
- [8] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli, "On the use of SIFT features for face authentication", Proc. of the Computer Vision and Pattern Recognition Workshop (CVPRW'06), pp. 17-22, June 2006.
- [9] P. Paalanen, J.-K. Kamarainen, J. Ilonen and H. Kälviäinen, "Feature representation and discrimination based on Gaussian mixture model probability densities-practices and algorithms", Pattern Recognition, vol. 39, no. 7, pp. 1346-1358, 2006.
- [10] S. E. Choi, Y. J. Lee, S. J. Lee and K. R. Park, "Age estimation using a hierarchical classifier based on global and local facial features", Pattern Recognition, vol. 44, no. 6, pp. 1262-1281, June 2011.
- [11] B. Ni, Z. Song and S. Yan, "Web image and video mining towards universal and robust age estimator", IEEE Transactions on Multimedia, vol. 13, no. 6, pp. 1217-1229, December 2011.
- [12] Y. H. Kwon and N. V. Lobo, "Age classification from facial images", Computer Vision and Image Understanding, vol. 74, no. 1, pp. 1-21, April 1999.
- [13] T. R. Alley, Social and Applied Aspects of Perceiving Faces, Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.
- [14] W.-B. Horng, C.-P. Lee and C.-W. Chen, "Classification of Age Groups Based on Facial Features", Tamkang Journal of Science and Engineering vol. 4, no.3, pp. 183-192, 2001.
- [15] M. M. Dehshibi and A. Bastanfard, "A new algorithm for age recognition from facial images", Signal Processing, vol. 90, no.8, pp. 2431-2444, 2010.
- [16] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework", International Journal of Computer Vision, vol. 56, no. 3, pp. 221-255, 2004.
- [17] T. Cootes, G. Edwards and C. Taylor, "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, Jun 2001.
- [18] A. Lanitis, C. Taylor and T. Cootes, "Toward automatic simulation of aging effects on face images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442-455, April 2002.
- [19] S. Kohli, S. Prakash and P. Gupta, "Hierarchical age estimation with dissimilarity-based classification", Neurocomputing, vol. 120 pp. 164-176, 2013.

- [20] W. -L. Chao, J. -Z. Liu and J. -J. Ding, "Facial age estimation based on label-sensitive learning and age oriented regression", *Pattern Recognition*, vol. 43, pp. 628-641, 2013.
- [21] X. Geng, Z. H. Zhou, Y. Zhang, G. Li and H. Dai, "Learning from facial aging patterns for automatic age estimation", *Proc. of ACM Conference on Multimedia*, pp. 307-316, 2006.
- [22] X. Geng, Z. H. Zhou and K. S. Miles, "Automatic age estimation based on facial aging patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234-2240, December 2007.
- [23] Y. Fu, Y. Xu and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging feature's", *Proc. IEEE International Conference on Multimedia and Expo*, pp. 1383-1386, 2-5 July 2007.
- [24] D. Cai, X. He, J. Han and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition", *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3608-3614, 2006.
- [25] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold", *IEEE Transactions on Multimedia*, vol. 10, no. 4, pp. 578-584, June 2008.
- [26] G. Guo, Y. Fu, T. S. Huang and C. R. Dyer, "Locally adjusted robust regression for human age estimation", *IEEE Workshop on Applications of Computer Vision (WACV'08)*, pp. 1-6, 7-9 Jan 2008.
- [27] G. Guo, Y. Fu, C. R. Dyer and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression", *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178-1188, July 2008.
- [28] J. Lu and Y. -P. Tan, "Ordinary Preserving Manifold Analysis for Human Age and Head Pose Estimation", *IEEE Transactions on Human-Machine Systems*, vol.43, no.2, pp. 249-258, 2013.
- [29] H. Fukai, H. Takimoto, Y. Mitsukura and M. Fukumi, "Apparent age estimation system based on age perception", *Proc. SICE 2007 Annual Conference*, pp. 2808-2812, , 17-20 Sept 2007.
- [30] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no.12, pp. 2037-2041, December 2006.
- [31] C. H. Ju and Y. H. Wang, "Automatic age estimation based on local feature of face image and regression", *2009 International Conference on Machine Learning and Cybernetics*, pp. 885-888, 12-15 July 2009.
- [32] F. Gao and H. Ai, "Face age classification on consumer images with gabor feature and fuzzy LDA method", *Proc. of 3rd International Conference on Advances in Biometrics (LNCS'5558)*, pp. 132-141, 2-5 June 2009.
- [33] G. Guo, G. Mu, Y. Fu and T. S. Huang, "Human Age Estimation Using Bio-Inspired Features", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 112-119, 2009.
- [34] M. Y. El Dib and H. M. Onsi, "Human age estimation framework using different facial parts", *Egyptian Informatics Journal*, vol.12, no.1, pp. 53-59, 2011.
- [35] H. Han, C. Otto, X. Liu and A. Jain, K., "Demographic Estimation from Face Images: Human vs. Machine Performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.
- [36] C. Li, Q. Liu, W. Dong, X. Zhu, J. Liu and H. Lu, "Human Age Estimation Based on Locality and Ordinal Information", *IEEE Transactions on Cybernetics*, 2014.
- [37] W. Chen, M. J. Er and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain", *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 36, pp. 458-466, 2006.
- [38] K. Rao and P. Yip, *Discrete cosine transform, Algorithms, Advantages, Applications*, Academic Press, Boston, MA, 1990.
- [39] W. Pennebaker and J. Mitchell, *JPEG: Still image data compression standard*, Kluwer Academic Publishers, Norwell, MA, 1992.
- [40] FG-NET aging database. [Online]. Available: <http://sting.cycollege.ac.cy/alanitis/fgnetaging/>