# Facial Age Estimation based on Decision Level Fusion of AAM, LBP and Gabor Features

Asuman Günay and Vasif V. Nabiyev

Department of Computer Engineering

Karadeniz Technical University

Trabzon, TURKEY

*Abstract*—In this paper a new hierarchical age estimation method based on decision level fusion of global and local features is proposed. The shape and appearance information of human faces which are extracted with active appearance models (AAM) are used as global facial features. The local facial features are the wrinkle features extracted with Gabor filters and skin features extracted with local binary patterns (LBP). Then feature classification is performed using a hierarchical classifier which is the combination of an age group classification and detailed age estimation. In the age group classification phase, three distinct support vector machines (SVM) classifiers are trained using each feature vector. Then decision level fusion is performed to combine the results of these classifiers. The detailed age of the classified image is then estimated in that age group, using the aging functions modeled with global and local features, separately. Aging functions are modeled with multiple linear regressions. To make a final decision, the results of these aging functions are also fused in decision level. Experimental results on the FG-NET and PAL aging databases have shown that the age estimation accuracy of the proposed method is better than the previous methods.

*Keywords—AAM; LBP; Gabor filters; Regression; Fusion; Age estimation*

## I. INTRODUCTION

The researches on facial image processing have received considerable interest in recent decades because of the increasing need of automatic recognition systems. Face recognition, face detection, facial expression recognition and gender classification are the research topics that have been studied by many researchers in this area. Facial age estimation is a relatively new topic and the interest in this topic has significantly increased because it has many real world applications. For example, under ages can be prevented from accessing alcohol, cigarettes or obscene contents on websites using an age estimation system. In addition, age specific target advertising, face recognition and age prediction systems robust to age progression for finding the missing people and criminals are important age estimation applications.

Facial age estimation is a multi-class classification problem because an age label can be seen as an individual class. This makes age estimation much harder than other facial image processing problems such as gender classification, face detection, etc. Besides, real world age progression displayed on faces is varied and personalized as shown in Fig. 1. Aging process of a person is affected by the genetics, race, eating and



Fig. 1. Facial aging of different individuals

drinking habits, living styles, climate, etc. [1]. Extent and frequency of facial expressions, emotional stress, exposure to sunlight, extreme weight loss, smoking, usage of anti-aging products, and plastic surgery also affect the person's facial appearance [2]. Therefore, determining the type of facial features that represents the age directly is very difficult. Moreover, the accuracy of age estimation systems are insufficient, even the human skills about age estimation are limited. The lack of proper large data set including the chronological image series of individuals is another drawback in these systems.

In this paper a new hierarchical age estimation method based on decision level fusion of global and local features of facial images both in age group classification and age estimation phases is proposed as shown in Fig. 2. The global facial features which contain both the shape and appearance information of human faces are extracted using active appearance models (AAM). The local facial features are extracted using Gabor filters and local binary patterns (LBP). A set of Gabor filters capable of extracting deep and fine wrinkles in different directions are used to extract wrinkle features and LBP is used to extract the detailed skin texture features of facial images. Then dimensionality reduction is performed using principal component analysis (PCA) for each feature vector separately. After finding the lower dimensional subspaces, three distinct support vector machine (SVM) classifiers are trained using global features, wrinkle features and skin features of facial images. Then the results of these classifiers are combined to find the age group of the subject. After that, age estimation is performed in that age group in a similar way, in which three aging functions are modeled with global and local features, separately using multiple linear regression. Finally the results of these aging functions are combined to estimate the age of the subject.
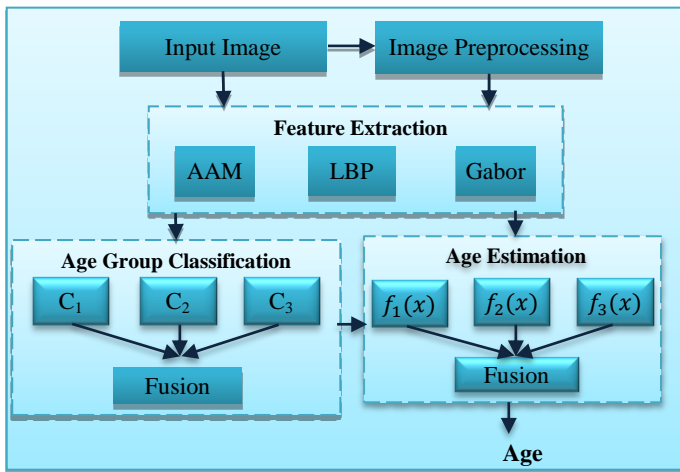
Fig. 2.   System structure

This paper is organized as follows. A survey on age estimation methods is given in Section 2. In Section 3 the proposed method including image preprocessing, global and local feature extraction, dimensionality reduction, classification, regression and decision level fusion is introduced. The experimental results are given in Section 4 and Section 5 concludes the paper.

## II.   AGE ESTIMATION METHODS

Over the years a great number of approaches have been proposed in the field of age estimation from facial images. These approaches typically consist of age image representation and age estimation techniques. Age image representation techniques rely often on shape and texture-based facial features. They can be grouped under the topics of Anthropometric Models, AAM, AGing pattErn Subspace (AGES), Age Manifold and Appearance Models. Then, age group classification or regression methods are performed for age estimation. Recently, hierarchical age estimation systems combining the classification and regression techniques are presented [3]. To build a universal human age estimator, robust multi-instance regressor learning algorithm based on facial information is also used [4].

In anthropometric models only the facial geometry is considered. Kwon and Lobo published the first work using the facial geometry in the age classification area [5]. They separated the babies from adults using a few ratios of distances on frontal images. However the shape of a human's head significantly changed in the childhood, but not in the adulthood [6]. For this reason their method can be successful in young ages. So the wrinkle information is used to classify the young and senior adults. In the experiments a small database including 45 images is used. Later on other age classification methods using geometrical features based on distance ratios and texture features based on wrinkles are also proposed [7, 8]

AAM [9, 10] based methods incorporate shape and appearance information together rather than just the facial geometry as in the anthropometric model based methods. An AAM uses a statistical shape and an appearance model to represent the images [10]. 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. The mean shape is produced with taking the mean of the landmark points in the training set. Then Principal Component Analysis (PCA) is applied to the data to extract the main principal components along which the training set varies from the mean shape. To build a statistical appearance model, each image has to be normalized, so that its control points match the mean shape. Then, PCA is applied to the gray-level intensities within a pre-specified image region for learning an appearance model. Using the AAMs for age estimation was initially proposed by Lanitis et al. [11]. The relationship between the age of individuals and the parametric description of face images was defined with an aging function $age = f(b)$. Kohli et al. [12] used ensemble of classifiers trained using AAM features to distinguish between child/teenhood and adulthood. Also different aging functions are used to estimate the age of the classified image. Chao et al. [13] proposed an age estimation approach based on label sensitive learning and age-oriented regression using AAM features.

Geng et al. [14, 15] proposed a method called AGES which uses the sequence of an individual's facial images arranged in chronological order to model the aging process. The features of face images are extracted with AAM. Then, PCA is used to learn a specific aging subspace for each individual. In AGES method missing age images of individuals can be synthesized with an expectation maximization-like iterative algorithm.

Age manifold methods intend to learn a common aging trend from the images of different individuals at different ages. The aging trend is learned in a low dimensional domain using manifold embedding techniques. The mapping from the image space to the low dimensional manifold space can be done either by linear or by nonlinear functions [16-21] such as $Y=P(X, L)$. In this representation $X = [x_i : x_i \in R^D]_{i=1}^n$ is the image space, $L = [l_i : l_i \in N]_{i=1}^n$ is the vector contains the age labels associated with images and $Y = [y_i : y_i \in R^d]_{i=1}^n$ with $d \leq D$ is the low-dimensional representation of $X$ in the embedded subspace. In age manifold methods all aging images of different individuals can be used together. But 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 the extraction of global and local aging-related facial features. Fukai et al. [22] extracted aging features from facial images using Fast Fourier Transform. Then the important features are selected using genetic algorithm. As the Local Binary Patterns (LBP) are efficient texture descriptors [23], they are used in age estimation systems. In Ju and Wang's study [24] regions which vary with aging are selected using Adaboost. Then LBP histograms are extracted from these regions and used for age estimation. Wrinkle information extracted with Gabor filters have also been used as effective texture features on age estimation tasks. [25-28].

## III.   PROPOSED METHOD

This paper proposes an innovative hierarchical age estimation method based on decision level fusion of global and local facial features. This method consists of the image prepro-

cessing, global feature extraction with AAM, local feature extraction with Gabor filters and LBP, dimensionality reduction with PCA classification with SVM, aging function modeling with multiple linear regression and decision level fusion modules. These modules are explained in the following sections.

### A. Image Preprocessing

The orientation and the size of original images are different from each other. Also they have unnecessary features such as background, cloth and hair which are not related to the face and can affect the performance of the algorithm. Therefore, image preprocessing step is performed to extract only the facial regions and to adjust the size and the orientation of the faces. In the this module, the facial images are cropped, scaled and transformed to the size of 88x88, based on the eye center locations as shown in Fig. 3.

### B. Feature Extraction

The feature extraction module consists of two modules: global feature extraction with AAM, local feature extraction with Gabor filters and LBP. These modules are explained in the following subsections.

*1) Global Feature Extraction with AAM:* AAM is a statistical shape and appearance model of facial images [10]. In AAM , a model of shape variations is combined with a model of the appearance variations in a shape-normalized frame

Training samples which are labeled with landmark points are used to generate a statistical shape model. The landmark points of various facial images are given in Fig. 4. Let $X = [x_i : x_i \in R^{D_s}]_{i=1}^n$ represents all the landmark points of training images and $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ represents the mean shape of training images; the main principal components along which the training set varies from the mean shape is extracted with PCA. The projection is chosen to maximize the determinant of the total scatter matrix $S_s = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$ of the projected samples as $\arg\max_{P_s}|P_s^T S_s P_s|$. $P_s$ is the set of eigenvectors with $d_s$ largest eigenvalues which provides a linear transformation from $D_s$ dimensional shape space into a $d_s \leq D_s$ dimensional parameter space. The shape parameters are defined by linear formulation as $b_s = P_s^T X$.

To build a statistical appearance model, each image has to be warped to mean shape as shown in Fig. 4. Then, the gray-level intensities within a pre-specified image region are used to train an intensity model. Let $G = [g_i : g_i \in R^{D_g}]_{i=1}^n$ represents all the gray level intensities of training images, the main principal components along which the training set varies from
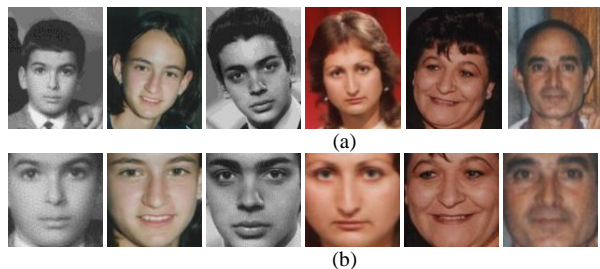


(a)



(b)

Fig. 3. Image preprocessing (a) Original images (b) Facial regions
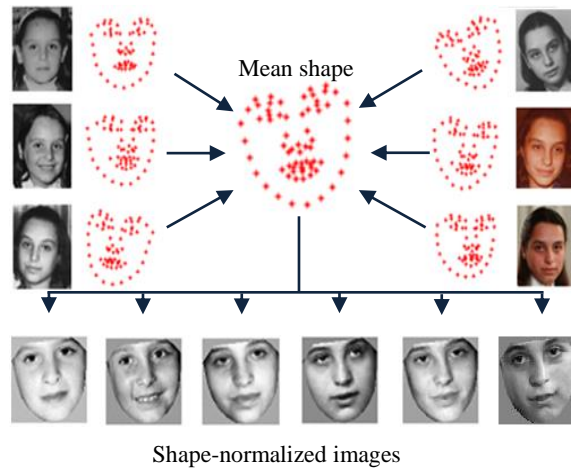


Shape-normalized images

Fig. 4. Landmark points, mean shape and normalized images used in AAM

the mean appearance is also extracted with PCA. The projection is chosen to maximize the determinant of the total scatter matrix $S_g = \sum_{i=1}^n (g_i - \bar{g})(g_i - \bar{g})^T$ of the projected samples as $\arg\max_{P_g}|P_g^T S_g P_g|$. $P_g$ is the set of eigenvectors with $d_g$ largest eigenvalues which provides a linear transformation from $D_g$ dimensional appearance space into a $d_g \leq D_g$ dimensional parameter space. The appearance parameters are defined by linear formulation as $b_g = P_g^T G$. $b_s$ and $b_g$ vectors can summarize the shape and appearance of any image. The combined shape-appearance parameters are obtained by concatenating $b_s$ and $b_g$ in a single vector and applying a further PCA in order to eliminate the correlations between them.

*2) Local Feature Extracting with Gabor Filters:* A Gabor filter is the modulation of a sinusoidal wave with a Gaussian function as shown Fig. 5. Therefore this filter will respond to the frequency which is in a localized part of the signal. 2 dimensional Gabor filters can be viewed as:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (1)$$

Where $x' = x\cos(\theta) + y\sin(\theta), y' = -x\sin(\theta) + y\cos(\theta)$. In (1), $\lambda$ is the wavelength, $\theta$ is the orientation, $\psi$ is the phase offset, $\sigma$ is the standard deviation of the Gaussian kernel and $\gamma$ is the spatial ratio of the Gabor function.

2D convolution operation is used to obtain the response of a Gabor filter to an image as follows:

$$G(x, y) = \iint I(p, q)g(x - p, y - q; \lambda, \theta, \psi, \sigma, \gamma)dpdq \quad (2)$$

Where $I(p, q)$ is the image and $G(x, y)$ is the response of a Gabor filter to the image. In the study the fine and deep wrinkles at different orientations are extracted using a Gabor filter set with 4 scales and 6 orientations as shown in Fig. 6. The responses of these filters for an image are also given in the figure.

*3) Local Feature Extracting with LBPs:* LBPs are powerful descriptors of image texture [29]. LBP operator thresholds the center pixel with its neighbors and assigns a binary code

for it. Then the occurrence histogram of these LBP codes is used as a texture feature. Every pixel of the image is labeled with the following equation,
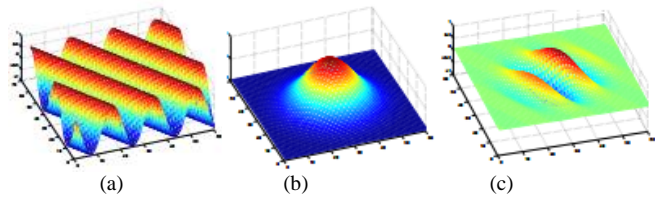


Fig. 5.   Gabor filter composition (a) 2 dimensional sinusoid oriented at 30° (b) Gaussian kernel (c) the corresponding Gabor filter
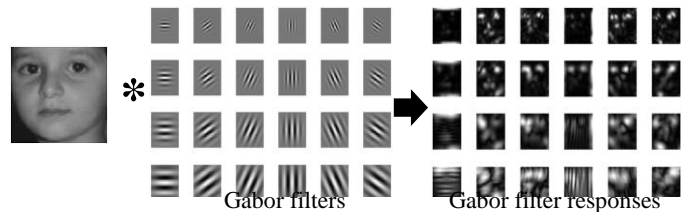


Fig. 6.   Gabor filter process

$$LBP_{P,R}(x_c) = \sum_{p=0}^{P-1} u(x_p - x_c)2^p \quad u(y) = \begin{cases} 1, & if \ y \geq 0 \\ 0, & if \ y < 0 \end{cases} \quad (3)$$

Where $x_c$ is center pixel, $x_p$ represents one of his $P$ neighbors and $R$ is the radius. In this equation $2^P$ different LBP codes can be generated for the center pixel but all of them are not used. Generally the uniform patterns are used in texture description. Uniform patterns are the ones that contain at most two bitwise transitions from 0 to 1 or vice versa when the binary pattern is considered circular. These patterns account for a bit less than 90% of all patterns when using (8,1) neighborhood [29]. But holistic descriptions of facial images are not reasonable as the texture descriptors tend to average over the image area [30]. However it is important to retain the information of spatial relations for facial images. Furthermore local representations are more robust to illumination or pose variations than holistic representations. As a result, spatial LBP histograms are extracted for an efficient representation of facial images. For this purpose image is divided into $m$ regions from which the spatial histograms are produced as follows,

$$H_{i,j} = \sum_{x_c \in R_j} f\{LBP_{8,1}(x_c) = U(i)\}, \quad f(y) = \begin{cases} 1, & y \ is \ true \\ 0, & y \ is \ false \end{cases}$$
$$i = 0,1,\dots,n-1 \quad j = 0,1,\dots,m-1 \quad (4)$$

Where $H_{i,j}$ is the $i^{th}$ value of the LBP histogram of $j^{th}$ region and $U(i)$ is the vector keeps uniform patterns. To build a global description of the image, regional histograms are concatenated in a single vector. In this work the detailed skin textures of facial images are extracted using spatial LBP histograms as shown in Fig. 7. Spatial representation of a facial image is obtained by dividing the image into 8x8 regions, producing the LBP histograms of these regions and concatenating them into a single vector.

### C. Dimensionality Reduction

After the feature extraction module, PCA is performed in order to find a lower dimensional subspace which carries sig-

nificant information for age estimation. The PCA method finds the embedding that maximizes the projected variance given below.
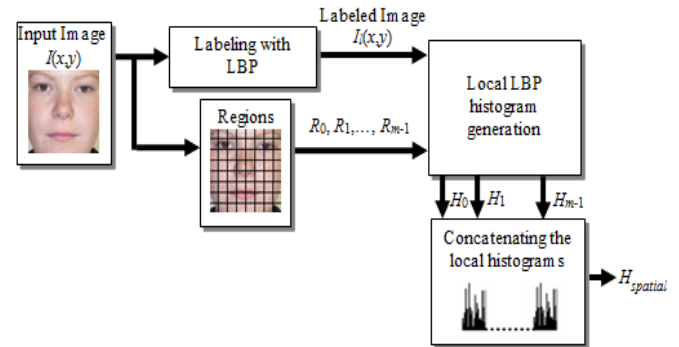


Fig. 7.   Spatial LBP histogram generation

$$W_{opt} = \arg \max_{\|W\|=1} W^T S W \quad (5)$$

In (5) $S = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$ is the scatter matrix, and $\bar{x}$ is the mean of feature vectors, $\{x_i : x_i \in R^D\}_{i=1}^{n}$. The solution of this problem is given by the set of $d \leq D$ eigenvectors associated with the largest eigenvalues of the scatter matrix. After determining the projection subspace, all the samples are projected on it using $y_i = W^T f_i$ allowing thus dimensionality reduction.

### D. Classification

In the age group classification module, the subject is classified into one of the age groups using a SVM classifier. SVM is a supervised learning method which uses support vectors to build a classification or regression model [30]. SVM finds a linear and optimal hyperplane which can separate two classes, providing the lowest separation error and maximum margin between the classes. Consider a two class classification problem with $M$ training points $x_i$ and assigned labels $y_i$ is defined as,

$$\{(x_i, y_i) | x_i \in R^n, y_i \in \{-1, +1\}\}_{i=1}^{M} \quad (6)$$

Linear SVM assumes that there exists a hyper plane separating two classes. The function of this hyper plane can be formulated as,

$$f(x) = w^T x + b \quad x \in R^n, w \in R^n, b \in R \quad (7)$$

Where $w$ is the normal vector and $b$ is the distance from the origin. This function can be used as a decision rule for a data point $x$ with label $y$ is as follows:

$$y = \begin{cases} +1, & f(x) \geq 0 \\ -1, & f(x) < 0 \end{cases} \quad (8)$$

In the training phase the $w$ and $b$ are found such that this decision rule is valid for all training and test data points. In real world applications, data can rarely be separated by a linear hyper plane. Thus the basic version of SVM only allowing linear classification is changed by applying a so called kernel

trick. The non-linear separable data is transformed into higher dimensional space using a kernel function $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$, which behaves like a scalar product and keeps the computational costs low. The conventional SVM assumes that there exists a linear hyperplane separating two classes. To extend this to nonlinear data, radial basis function kernel as shown in (9) is used in this paper.

$$K(x_i, x_j) = e^{\|x_i - x_j\|^2 / 2\sigma^2} \qquad (9)$$

SVM is originally designed for binary classification. To solve multi-class classification problems with SVM, different implementations like one-against-all and one-against-one are used. In the proposed approach one against-one method is used for multi-class classification. For this purpose, $k(k$-1$)$ binary SVMs representing all possible pairs of $k$ classes are constructed. Each of these classifiers is trained to discriminate only two of the $k$ classes. Then majority voting strategy is used to predict the final output. The data point is assigned to the class that has maximum votes. The optimal parameters for SVM were selected experimentally from the training set.

*E. Regression*

After finding the age groups of facial images, the age estimation problem is recast as a multiple linear regression as follows:

$$L = YB + e, \ \ Var(e) = \sigma^2 I \qquad (10)$$

Where $Y$ is the data matrix including a columns of 1s, $B$ is the unknown parameter vector, $L$ is the age label vector and $e$ is the error vector with zero mean and common variance $\sigma^2$. During the learning stage the unknown parameters are estimated using least squares, or robust regression. The regression function used in this study is a quadratic function given by,

$$\hat{l} = \hat{\beta}_0 + \hat{\beta}_1^T y + \hat{\beta}_2^T y^2 \qquad (11)$$

Where $\hat{l}$ is the estimate of age, $\hat{\beta}_0$ is the offset, $\hat{\beta}_1$, $\hat{\beta}_2$ are the weight vectors and $y$ is the low dimensional representation of the extracted feature vector.

*F. Decision Level Fusion*

In the proposed method decision level fusion is performed in both classification and regression modules. In the classification module, the age class labels are produced with three distinct classifiers which are trained with global and local features. Then the results of these classifiers are combined to determine the age group of the subject.

In the age estimation module three aging functions are modeled separately in that age group, using global and local features. The results of these aging functions are also combined to make a final decision for the age of the test sample as follows:

$$age = \sum_{i=1}^{N} age_i \Big/ N \qquad (12)$$

Where $age$ is the final age of the test sample, $age_i$ is the age estimated by the $i^{th}$ aging function and $N$ is the total number of aging functions.

## IV. EXPERIMENTS AND RESULTS

In this paper, the performance of the proposed method is evaluated using FG-NET [31] and PAL [32] aging databases. FG-NET database comprises of 1,002 images in the age range of 0-69 years.

The images were retrieved from real-life albums of 82 subjects, so the dataset includes uncontrolled variations of occlusion, facial expressions, head pose, illumination, etc. The data distribution of the FG-NET database according to age is shown in Fig. 8-(a). It can be seen from the figure that the image distribution is not uniform which can adversely affect the system performance.
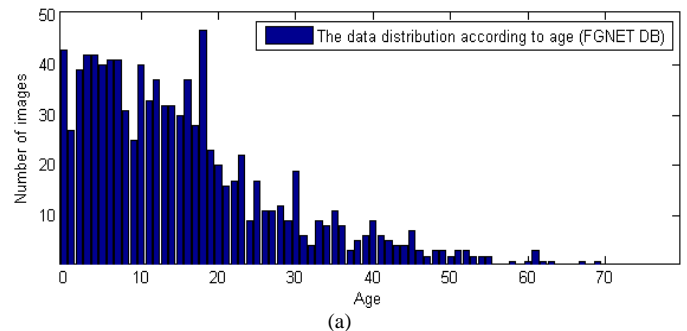
The PAL aging database contains 580 images of different individual in the age range of 18-93 years. The images were captured under natural lighting conditions using a digital camera. This database includes various expressions such as neutral faces, anger, sadness or smiling. The distribution of images in this database according to age is shown in Fig. 8-(b).

Performance evaluation is done using leave-one-person-out (LOPO) for FG-NET database. In this method the images of a person are used as test set and all the other images are used as training set in each fold. This procedure is iterated for 82 folds, which is the number of subjects in the database. After 82 folds final estimation is calculated by taking the mean of all estimations. In the experiments 3-fold cross validation mode is used for PAL database in which the 1/3 of the images are selected randomly as test set and the rest are used as training set. After 3 folds the mean of all estimations is determined as estimation performance of the system.

The Mean Absolute Error (MAE) and Cumulative Score (CS) metrics are used for performance comparison in the study. MAE is defined as the average of the absolute error between the recognized labels and the ground truth labels as follows:

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

Where $\hat{l}_i$ is the estimated age for $i^{th}$ test sample, $l_i$ is the corresponding ground truth, and $N_t$ is the total number of the test
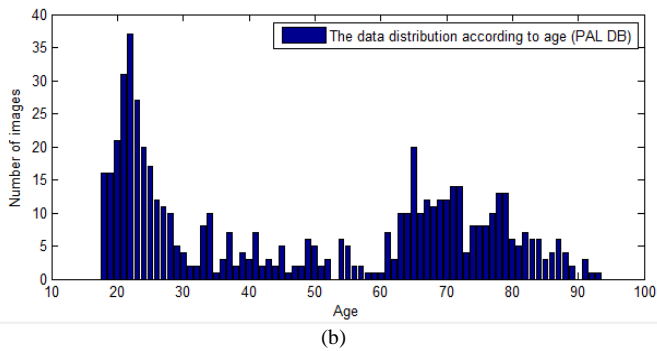


(a)

Fig. 8. The data distributions of (a) FG-NET and (b) PAL databases according to age

samples. CS enables performance comparison at different absolute error levels. It is the ratio of the number of images, whose absolute errors are lower than a threshold value to the total number of images. It is expressed by,

$$CS(th) = \frac{N_{e \leq th}}{N} \times 100 (\%) \qquad (14)$$

Where $N_{e \leq th}$ is the number of images with the absolute estimation error is less than $th$, and $N$ is the number of test images.

In the global feature extraction step, the coordinates of 68 landmark points on the training samples are used to train the shape model. Also the mean shape is determined from these points. Next, affine transformation is used in the warping process of all images to the mean shape. Then approximately 7000 gray-level intensities in the facial region of the corresponding shape-normalized images are used to train the appearance model. Finally, 277 AAM model parameters are used as global features to represent the images.

In the local feature extraction, wrinkle information of the facial images is extracted using Gabor filters. The fine and deep wrinkles at different orientations are extracted with Gabor filters applied in 4 scales and 6 orientations. The responses of these filters are concatenated into a single vector and dimensionality reduction is performed using. Furthermore the detailed skin textures of facial images are extracted using spatial LBP histograms. For this purpose LBP histograms are produced from 8x8 sub-regions of facial images and concatenated into a single vector, resulted a spatial representation of the facial image. Also PCA is applied to learn a low dimensional representation of this feature vector.

In the spatial LBP histogram generation phase, the number of sub-regions is determined experimentally. For this purpose, the age estimation performances of the spatial LBP histograms produced with different number of sub-regions are calculated and the results are shown in Fig. 9. It can be seen from the figure that using the 8x8 sub-regions gives better results for age estimation.

After the feature extraction and dimensionality reduction phase, age group classification is performed using three SVM classifiers. The age ranges of age groups are selected as: 0-12 childhood, 13-19 adolescence, 20-39 young adulthood, 40-64 middle adulthood and ≥65 late adulthood. Then age estimation

is performed in the specified age group using multiple linear regression. For this purpose, three aging functions are modeled separately using global features, wrinkle features and skin features for age estimation. Then the results of these aging functions are combined and a final decision is made for the test sample. In the experiments, first the age estimation performances of the global, local and fused features are determined using a single level age estimation scheme. In this scheme all the images are used to train the aging functions and the decision level fusion is performed for a final decision for the age. The experimental results on FG-NET and PAL databases are listed in Table 1. It can be seen from the table that age estimation performance of the AAM features is better than Gabor and LBP features on FG-NET and PAL databases as the AAM features both include the shape and appearance information of facial images. But when they are fused with local features including wrinkle and skin texture information at decision level, age estimation performance noticeably increased. As a result MAE of 4.87 years on FG-NET database and MAE of 5.38 years on PAL database is achieved when using single level age estimation based on decision level fusion of facial features.
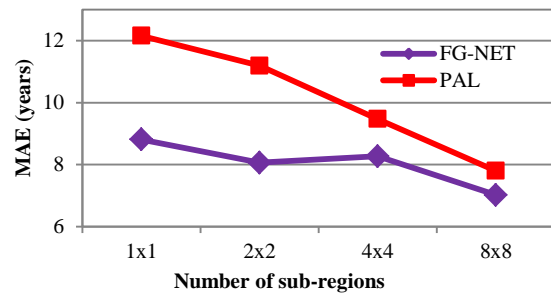


Fig. 9. MAE's of different number of sub-regions used in spatial LBP histogram generation

TABLE I. SINGLE LEVEL AGE ESTIMATION RESULTS (MAE) BASED ON DECISION LEVEL FUSION OF FEATURE VECTORS

| Feature | FG-NET | PAL |
|---|---|---|
| AAM | 6.02 | 6.61 |
| LBP | 7.02 | 7.80 |
| Gabor | 6.55 | 7.37 |
| AAM+LBP | 5.36 | 5.96 |
| AAM+Gabor | 5.23 | 5.70 |
| **AAM+LBP+Gabor** | **4.87** | **5.38** |

The Cumulative Scores of the single level age estimation scheme on FG-NET and PAL databases at error levels from 0 to 15 years are shown in Fig. 10. Age of approximately 8.08% of the subjects in the FG-NET database and 5.12% of the subjects in PAL database can be estimated with zero error level. As the error level increases the estimation accuracy also increases for all feature extraction methods. This single level age estimation approach is able to achieve cumulative scores of 89.92% and 85.86% for an absolute error of 10 years for FG-NET and PAL databases, respectively.

Age estimation performance of the proposed hierarchical age estimation approach based on decision level fusion of global and local features both in classification and detailed age estimation phases are given in Table 2. One can see from the table that proposed method achieves the MAE of 4.13 on FG-

NET database and MAE of 4.67 on PAL database. Cumulative scores for FG-NET and PAL databases for an absolute error of 10 years are increased to 90.88% and 92.17%, respectively as shown in Fig. 11.

As the FG-NET database is the most common database used in age estimation works, the performance of the proposed method and the previous works on FG-NET aging database are compared in Table 3. One can see from Table 3 that proposed method has an MAE of 4.13 years which is lower than the previous methods. This result also shows that decision level fusion of global and local features in a hierarchical system improves the age estimation performance. Gabor filters and block-based LBP histograms encode the texture information of the facial images. Global features are extracted with AAMs which encodes the shape and appearance information of facial images. These feature sets capture differential complementary information. The decision level fusion of these features estimates the age better when compared with the age estimation accuracies obtained using these features alone.

TABLE II. HIERARCHICAL AGE ESTIMATION RESULTS BASED ON DECISION LEVEL FUSION OF FEATURE VECTORS

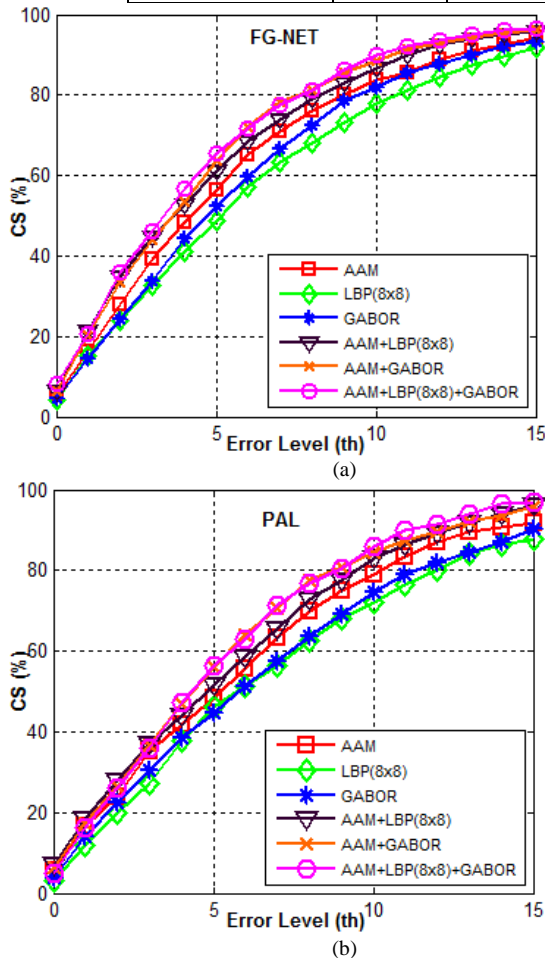| Database | FG-NET | PAL |
|----------|--------|-----|
| MAE | 4.13 | 4.67 |



Fig. 10. Cumulative scores of global, local and fused features using single level age estimation scheme on (a) FG-NET and (b) PAL databases
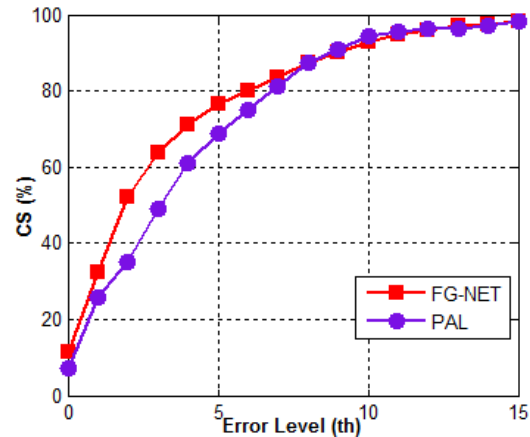


Fig. 11. Cumulative scores of the proposed method on FG-NET and PAL databases

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

| Methods | MAEs |
|---------|------|
| AAS [14] | 14.83 |
| WAS [15] | 8.06 |
| AGES [15] | 6.77 |
| AGES$_{lda}$ [15] | 6.22 |
| LARR[20] | 5.16 |
| RMIR[4] | 8.37 |
| Ju and Wang[24] | 6.85 |
| Lu and Tan [21] | 5.75 |
| Choi et al. [3] | 4.32 |
| **Proposed** | **4.13** |

## V. CONCLUSION

In this paper, a hierarchical age estimation method relying on decision level fusion of AAM, Gabor and LBP features of facial images is proposed. Its main contribution is decision level fusion of global texture features and local texture features of facial images. Locality is preserved by regional LBP histograms and Gabor filters. Furthermore, these local features are combined with global features of images extracted with AAMs. Experimental results using the FG-NET and PAL aging databases have shown that the proposed method is better than previous methods.

REFERENCES

[1] M. Albert, K. Ricanek and E. Patterson, A review of the literature on the aging adult skull and face: implications for forensic science research and applications, Forensic Science International 172 (1) (2007) 1-9.

[2] M. Gonzalez-Ulloa and E. S. Flores, Senility of the face-Basic study to understand its causes and effects. Plastics & Reconstructive Surgery 36 (2) (1965) 239-246.

[3] 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.

[4] 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.

[5] 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.

[6] T. R. Alley, Social and Applied Aspects of Perceiving Faces, Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

[7]  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.

[8]  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.

[9]  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.

[10] 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.

[11] 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.

[12] S. Kohli, S. Prakash and P. Gupta, "Hierarchical age estimation with dissimilarity-based classification", Neurocomputing, vol. 120 pp. 164-176, 2013.

[13] W. -L. Chao, J. -Z. Liu and J. -J. Ding, "Facial age estimation based on label-sebsitive learning and age oriented regression", Pattern Recognition, vol. 43, pp. 628-641, 2013.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] J. Lu and Y. -P. Tan, "Ordinary Preserving Manifold Anaysis for Human Age and Head Pose Estimation", IEEE Transactions on Human-Machine Systems, vol.43, no.2, pp. 249-258, 2013.

[22] 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.

[23] 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.

[24] 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.

[25] F. Gao and H. Ai, "Face age classification on consumer images with gabor feature and fuzzy LDA method", Proc. of 3$^{rd}$ International Conference on Advances in Biometrics (LNCS'5558), pp. 132-141, 2-5 June 2009.

[26] 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.

[27] 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.

[28] 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.

[29] T. Ojala, M. Pietikainen and T. Maenpaa, "Multi-resolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[30] C. Cortes and V. Vapnik, "Support-vector network", Machine Learning, vol. 20, no.3, pp. 273-297, 1995.

[31] FG-Net aging database. Available: http://sting.cycollege.ac.cy /~alanitis/fgnetaging.

[32] M. Minear and D. C. Park, "A lifespan database of adult stimuli", Behavior Research Methods, Instruments and Computers, vol.36, no.4, pp.630-633, 2004.