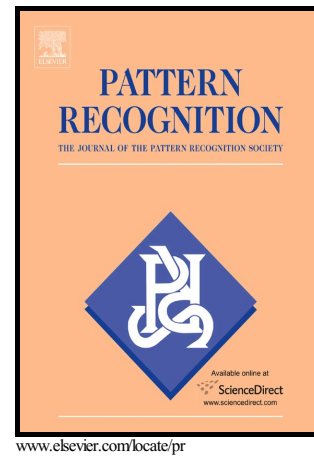


# Author's Accepted Manuscript

## A Flexible Hierarchical Approach For Facial Age Estimation Based on Multiple Features

Jhony K. Pontes, Alceu S. Britto, Clinton Fookes,  
Alessandro L. Koerich



PII: S0031-3203(15)00445-8  
DOI: <http://dx.doi.org/10.1016/j.patcog.2015.12.003>  
Reference: PR5588

To appear in: *Pattern Recognition*

Received date: 1 August 2015  
Revised date: 1 November 2015  
Accepted date: 2 December 2015

Cite this article as: Jhony K. Pontes, Alceu S. Britto, Clinton Fookes and Alessandro L. Koerich, A Flexible Hierarchical Approach For Facial Age Estimation Based on Multiple Features, *Pattern Recognition*, <http://dx.doi.org/10.1016/j.patcog.2015.12.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# A Flexible Hierarchical Approach For Facial Age Estimation Based on Multiple Features

Jhony K. Pontes, [Alceu S. Britto Jr.](#), Clinton Fookes, and Alessandro L. Koerich

## Abstract

Age estimation from facial images is increasingly receiving attention to solve age-based access control, age-adaptive targeted marketing, amongst other applications. Since even humans can be induced in error due to the complex biological processes involved, finding a robust method remains a research challenge today. In this paper, we propose a new framework for the integration of Active Appearance Models (AAM), Local Binary Patterns (LBP), Gabor wavelets (GW) and Local Phase Quantization (LPQ) in order to obtain a highly discriminative feature representation which is able to model shape, appearance, wrinkles and skin spots. In addition, this paper proposes a novel flexible hierarchical age estimation approach consisting of a multi-class Support Vector Machine (SVM) to classify a subject into an age group followed by a Support Vector Regression (SVR) to estimate a specific age. The errors that may happen in the classification step, caused by the hard boundaries between age classes, are compensated in the specific age estimation by a flexible overlapping of the age ranges. The performance of the proposed approach was evaluated on FG-NET Aging and MORPH Album 2 datasets and a mean absolute error ( $MAE$ ) of 4.50 and 5.86 years was achieved respectively. The robustness of the proposed approach was also evaluated on a merge of both datasets and a  $MAE$  of 5.20 years was achieved. Furthermore, we have also compared the age estimation made by humans with the proposed approach and it has shown that the machine outperforms humans. The proposed approach is competitive with current state-of-the-art and it provides an additional robustness to blur, lighting and expression variance brought about by the local phase features.

## I. INTRODUCTION

The human face conveys important perceptible information related to personal characteristics, including identity, gender, ethnicity and age [16]. Currently, age can play an important role in many applications such

Jhony K. Pontes and Clinton Fookes are with the Image and Video Research Laboratory, Queensland University of Technology, Brisbane, Queensland 4001, Australia (e-mail: [jhony.kaesemodelpontes@hdr.qut.edu.au](mailto:jhony.kaesemodelpontes@hdr.qut.edu.au) and [c.fookes@qut.edu.au](mailto:c.fookes@qut.edu.au))

Alessandro L. Koerich is with Department of Software Engineering and IT, École de Technologie Supérieure, Montreal, H3C 1K3, QC, Canada (e-mail: [alessandro.koerich@etsmtl.ca](mailto:alessandro.koerich@etsmtl.ca))

Alceu S. Britto Jr. is with the Pontifical Catholic University of Paraná, Postgraduate Program in Computer Science, Curitiba, 80215-901, PR, Brazil (e-mail: [alceu@ppgia.pucpr.br](mailto:alceu@ppgia.pucpr.br))

Manuscript submitted on July 2015. Revised version submitted on November 2015.

as age-based access control, age estimation in crime investigation, age-adaptive targeted marketing and age-invariant person identification. For instance, it can prevent minors from purchasing alcohol or cigarettes from vending machines, or it can deliver digital content based on age to adapt messages accordingly.

Aging is also the main risk factor for many complex diseases. As facial aging is one of the most prominent and accessible phenotypes of human aging, it is important for assessing the risks of aging diseases and for designing individualized treatments [8]. Chen *et al.* [8] found that facial morphology features are significantly correlated with health indicators in the blood. However, facial features are more reliable aging biomarkers than blood profiles and can better reflect general health status than chronological age in a non-invasive way. This is a potential application for age estimation that may help to improve our life styles and health.

Facial aging is a complex process and it is generally slow and irreversible [38]. Despite of aging shows different forms in different ages and people, there are still some general patterns that can be described. During the lifetime, there are two different stages in face growth, development, and aging forms [16]. From birth to adulthood, the greatest change is the craniofacial growth. The facial skin does not change much. From adulthood to old age, the most perceptible change becomes skin aging. Since the face ages with loss of collagen beneath skin as well as with gravity effects, the skin becomes thinner, darker, less elastic and more wrinkled [16]. Besides, intrinsic (*i.e.*, heredity, gender, ethnicity) and extrinsic factors (*i.e.*, environment, lifestyle) can influence aging. All these factors make difficult to precisely predict a person's age, even by humans. Moreover, different perturbations on facial images, such as lighting, facial expression, pose, occlusion and blur, make the age estimation more challenging.

Automatic age estimation methods are generally classified according to the age image representation and the age prediction approach. The age image representation has received a great deal of attention since it affects the estimation performance if the aging features are not very discriminative. These features are defined as global, local or hybrid features. Changes from birth to adulthood (*i.e.*, craniofacial growth) are mainly encoded in global features while the changes from adulthood to old age (*i.e.*, skin aging) in local features. The global and local features are fused to form hybrid features to have a complete facial aging representation. Several approaches have investigated different aspects of human faces such as geometric ratios between major facial features (*e.g.*, eyes, nose and mouth), shapes of facial parts or skin textures (*e.g.*, wrinkles, sagging and spots). The most common models for age image representation are Anthropometric models [27], Active Appearance Models [33], Aging Pattern Subspace (AGES) [18], and Appearance Models [9]. For age prediction, the problem is either tackled as a multi-class classification or a regression problem. The features related to aging can differ according to the age group. Wrinkles, for instance, are normally found in elderly people while geometric features change mainly during childhood and adulthood. However, facial

features that are specific to a certain age group may not be considered when treated by a single-level age estimation [9]. Although hierarchical approaches that combine the classification and regression methods have been proposed to overcome this limitation of single-level estimation and improve the age estimation performance [9], [14], there is still a drawback. For instance, a 15-year-old individual can be wrongly classified into an age group of 20 to 39 years old during the classification step and carry on such an error to the regression step affecting the age estimation due to the hard boundaries between age classes [9]. In summary, finding a robust age estimation method remains a research challenge.

This paper presents a novel hierarchical age estimation approach composed of age group classification and detailed age estimation. First, a subject is classified into a specific age group and then a specific regressor is selected to estimate a numerical age value. We propose a flexible overlapping of age ranges in the specific age estimation step to alleviate the errors due to the hard boundaries between age classes imposed by the hierarchical approach. In the proposed approach, an individual that is misclassified in a wrong age group during the classification step can still have its age well estimated by a specific regressor due to the flexible overlapping of age ranges. Furthermore, we employ global features to encode the craniofacial growth using the AAM method since it provides appearance and shape information. Holistic AAM features do not contain enough skin texture details and to overcome this, a local approach is used to deal with the second aging stage. From adulthood to old age, the skin aging is the most perceptible change, and to deal with wrinkles and skin spots the LBP, Gabor wavelets and LPQ techniques are employed to extract these local features. Blurred faces, lighting and expression variations present in the aging datasets and also in the real-world images can also affect the performance of the age estimation. Recognition of blurred faces using the LPQ operator has been proposed by Ahonen *et al.* [1]. They have shown that it reaches higher recognition rates than the LBP operator and also that it is highly tolerant to blur and sharp images. LPQ is based on quantizing the Fourier transform phase in local neighborhoods due to its blur invariant property. Since the LPQ has not been yet exploited in the age estimation problem, a new extraction method for local features based on LPQ is proposed to deal with the challenges of the imaging conditions. Finally, global and local features are fused to form a hybrid feature vector in order to have a better facial aging representation.

The performance of the proposed approach was evaluated on two publicly available datasets: the FG-NET Aging and the MORPH Album 2. Both datasets have many variations in the face images, such as expressions, light, resolution and blur. Unfortunately, no previous studies have merged both datasets for performance assessment. The combination of both datasets allows a better generalization ability and age distribution since the FG-NET covers ages from 0 to 69 years, within which 50% are between 0 and 13 years, while the MORPH covers ages from 16 to 77 years. Therefore, the proposed approach is also evaluated

under a realistic condition that accounts for a wide age span. The experimental results have shown that the performance of the proposed approach on all datasets is competitive with current state-of-the-art with increased robustness to blur, lighting and expression variance due to the local phase features. Moreover, a comparison between the age estimation made by humans and the proposed approach has shown that the machine outperforms humans.

This paper is organized as follows. First, several previous works on age estimation are briefly reviewed in Section II. Section III presents the details of the proposed approach. The experimental results are presented in Section IV. Finally, the conclusions are stated in the last section.

## II. PREVIOUS WORKS

Kwon and Lobo [27] have presented one of the first works on age estimation from facial images. They classified facial images into three groups: child, young and senior adults, based on anthropometric models and wrinkle patterns. To distinguish children from adults they computed six distance ratios (*e.g.*, distance between eyes and mouth) from a frontal facial image. Wrinkle patterns were extracted from skin areas such as forehead, cheeks, corner of eyes, etc. by using snakelets to detect curves. The experiments were conducted in a small private dataset with 47 images. Later, Dehshibi and Bastanfard [12] used the geometric ratios of facial components and texture in their approaches. The anthropometric models might be useful for young ages, but not for adults. In practice, only frontal faces can be used to measure facial geometries because the ratios of distances are computed from 2D facial images which are sensitive to pose.

Active appearance models (AAM) is a statistical model proposed initially by Cootes [10] for facial image representation. Given a set of images for training, shape and texture models are learned separately based on the principal component analysis (PCA). Lanitis *et al.* [29] extended the AAM for an age representation proposing a function,  $age = f(b)$ , to explain age variation. In such a representation,  $age$  is the actual age of a subject,  $b$  is a vector with 50 parameters learned from the AAM and  $f$  is an aging function which defines the relation between the age and the model parameters. Unlike the anthropometric models, the AAM can deal with any age. Also, the AAM does not only consider the geometry, like the anthropometric model, but also takes into consideration texture. The AAM is mainly used to estimate age from global features since it provides appearance and shape information. However, the model does not include detailed wrinkles and skin information because such information is lost in the dimensionality reduction [9]. Geng *et al.* [18] proposed the AGES, that uses a sequence of individual's facial images to model the aging process instead of dealing with each facial image separately. The problem of the AGES is that the AAM, used to represent each face image, might not encode facial wrinkles for senior people since it encodes the image intensities without using any spatial neighborhood to calculate texture patterns.

Age manifold considers learning a common aging pattern from different subjects at different ages. A nonlinear low-dimensional aging pattern that reduces the redundancy is learned from several facial images at a given age. Guo *et al.* [20] performed age estimation by a Locally Adjusted Robust Regression (LARR), where ages are first predicted by a regression function and then are locally adjusted to match the true value within a specific bound, showing that facial images can be represented on a manifold. Unlike the AGES method, the age manifold does not require images at different ages of the same subject but rather images of several subjects at many ages to learn the low-dimensional manifold.

Aging-related facial feature extraction is focused by the appearance model approaches. Both global and local features were used in current age estimation systems. Hayashi *et al.* [24] considered both texture (wrinkle) and shape (geometry) to characterize each facial image. Local binary patterns and Gabor wavelets have been exploited for appearance feature extraction in the work of Choi *et al.* [9] and Bereta *et al.* [2]. Guo *et al.* [21] investigated the Biologically Inspired Feature (BIF) derived from a feedforward model of the primate visual object recognition pathway. The advantage is that small translations, rotations and scale changes can be well handled by the new set of aging patterns.

Age estimation can be tackled as a classification or regression problem. Classification places the subject into an age group such as child, young and senior adult, or in age groups such as 0-13, 14-21, 22-39 years old and so forth. On the other hand, regression methods place a numerical value for the age of an individual, thus making it more accurate. Lanitis *et al.* [28] evaluated the performance of different classifiers to estimate an age, including k-Nearest Neighbours (kNN), neural networks and quadratic function. The facial features were extracted with the AAM technique. Guo *et al.* [20] found the SVM as a representative classifier and the SVR as a representative regressor for age estimation problems. A better way to take advantage from both methods, SVM and SVR, is combining them in a hierarchical way. Luu *et al.* [33] introduced a hierarchical approach that combines SVM and SVR to improve the accuracy of age estimation which uses AAM as holistic features and consists of two main steps. First, a binary SVM classifier was built to distinguish between youths (from 0 to 20 years) and adults (from 21 to 69 years) and two SVR aging functions to model growth and development, and the adult aging process. Then, given an output from the youth/adult classifier, an appropriate aging function is used to determine the facial age. Duong *et al.* [14] extended the automatic age estimation framework proposed in [33] by adding another step to adjust the estimated ages. They built a binary SVM classifier to distinguish youths from adults from AAM features and defined two aging functions corresponding to two stages of face aging with the given AAM feature vectors. Ages were obtained from the aging functions as in [33]. However, they calculate the differences between the real and the estimated ages. From the computed differences, they proposed to use the shape

free image extracted from AAM to extract local features with LBP. Then, the SVR was applied again to construct the age-adjusting function for the growth and development period and for the later period. They achieved a worse result than Luu *et al.* [33] even using a combination of global and local features. Indeed, this may have occurred because they applied the LBP operator in the shape free image extracted from AAM which does not contain much skin details.

Since the facial aging is perceived differently in different age groups, a hierarchical age estimation using a small and specific age group can provide an accurate result. Also, a smaller age group reduces the computational effort for age estimation [22]. However, the age estimation using a hierarchical approach that neglects age overlapping between age groups has a drawback. For instance, a 15-year-old individual could be classified into an age group from 20 to 39 years old during the classification step. This error will carry on to the specific age estimation and the final performance will be affected. By looking at these considerations, which have been overlooked in [33] and [14], Choi *et al.* [9] proposed a new hierarchical age estimation. They proposed seven age groups split into four age classes and three overlapped classes with the aim of reducing classification errors. In each age class and overlapped age class, an age estimator based on a SVR is applied and the age is estimated. Although the errors in the age group classification step are smaller than those in the hierarchical age estimation with hard boundaries, the errors can still be propagated to the specific age estimation affecting the performance. Furthermore, their method uses seven SVM classes and SVR aging functions to deal with the hard boundary issue, which increases the computational effort.

Recently, Liu *et al.* [32] proposed a hybrid constraint SVR for facial age estimation. They used fuzzy age labels annotated by human in combination with the real age labels to train an age estimation system. Fuzzy age labels are defined as the upper and lower bounds of human estimation. Geng *et al.* [17] proposed two adaptive label distribution learning algorithms, IIS-ALDL and BFGS-ALDL, to automatically learn the label distributions adapted to different ages. The age and gender estimation of faces in the wild conditions using dropout SVM approach is presented by Eidingner *et al.* [15]. They also introduced the Adience benchmark for age and gender estimation to evaluate their approach. Chang and Chen [5] proposed a cost-sensitive ordinal hyperplanes ranking algorithm for age estimation. They used the scattering transform to extract facial features, a series of binary classification results to obtain the age rank, and conducted an analysis on the cost of each individual binary classifier. In these methods, single-level age estimation is used to estimate age in the entire dataset without pre-classification. However, the aging process can differ according to the age group of a subject, and the specific age-related facial features may not be considered by such approaches.

Dibeklioglu *et al.* [13] proposed the combination of facial dynamics and appearance for age estimation. The dynamic features were derived from facial expressions and combined with the appearance features to



train the classifiers. To overcome the hard boundaries issue, they proposed a method that computes the ages which are dissimilar with their neighbors. The whole age range is divided into groups in such a way that the boundary between each two adjacent groups is discriminant. They analyzed the discrimination power of smile dynamics for age estimation and showed that it can improve the estimation accuracy.

A new hierarchical age estimation approach composed of age group classification and detailed age estimation is proposed in this work. First, a subject is classified into a determined age group and then a specific regressor is selected to estimate a numerical age value. Different from [33] and [14], smaller and specific age groups are used to provide a more accurate result, while reducing the complexities of classification and not increasing the computation effort [22]. We propose a flexible overlapping of age ranges only in the specific age estimation step to tackle the problem of hard boundaries between age classes imposed by the hierarchical approach. Different from Choi *et al.* [9], who used extra classifiers and regressors to overlap the age ranges, we do not increase the number of classifiers and regressors. In the proposed approach, an individual that is misclassified during the classification step can have its age well estimated by a specific regressor due to the flexible overlapped age range. Furthermore, we employ a local approach to overcome one of the main shortcomings of the holistic approach such as the AAM features which are not able to capture the details of small regions where wrinkles are present. To deal with the challenges of the imaging conditions such as blurred faces, lighting and expression variations a new extraction method for local features based on LPQ is proposed since it has not been yet exploited in the age estimation problem. Local and global features are fused to achieve a better facial aging representation with the aim of enhance age estimation performance.

### III. PROPOSED APPROACH

The proposed approach considers the two aging stages during the human lifetime to encode the facial aging features. From birth to adulthood the greatest change is the craniofacial growth and from adulthood to old age the most perceptible change is the skin aging. The global features are used to handle with the craniofacial growth and are extracted using the AAM since it provides appearance and shape information. However, a holistic approach, such as the AAM, is not able to capture the skin texture details. Therefore, a local approach is employed to deal with skin aging. Wrinkles and skin spots are extracted by LBP, Gabor wavelets and LPQ techniques as local features. In order to have a complete facial aging representation, the global and local features are fused to form hybrid feature vectors. In addition, a hierarchical age estimation consisting of age group classification using SVM and specific age estimation using SVR is proposed. The errors generated in the classification step caused mainly by the hard boundaries between the age groups are compensated in the regression step by using flexible overlapped classes between the age ranges.



An overview of the proposed approach is presented in Fig. 1. Given an input color or grayscale facial image, the process of age estimation starts with the conversion of the image to grayscale followed by a non-reflective similarity transformation to normalize the image. The local features are extracted from eleven skin areas cropped from the normalized image in order to perform the local analysis in the facial texture. The global features are extracted from 68 landmarks, used to build the AAM. Afterwards, the global and local features are fused and classified into one of the  $N$  age groups defined. According to the age group where an individual is classified, a specific SVR trained with flexible overlapped regions to compensate eventual errors generated during the classification step is selected. Finally, the perceived age is estimated to a numerical value. The details of all steps previously described are presented as follows.

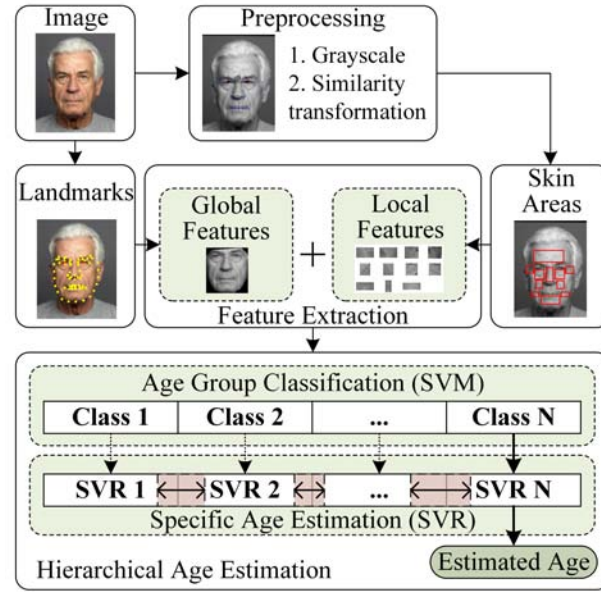


Figure 1. An overview of the proposed age estimation approach.

### A. Preprocessing

Preprocessing is only performed prior to the local feature extraction since the AAM method already takes care of the necessary preprocessing for the global feature extraction, such as similarity transformation to correct pose and scale, and lighting normalization. The images from FG-NET Aging were captured from different methods like image scanning and pictures taken from digital cameras. There are color and grayscale pictures. However, all images are first converted to grayscale in order to reduce the influence of inconsistent colors. Also, the computational cost in terms of time and memory usage is reduced.

FG-NET Aging and MORPH Album 2 provide 68 landmarks for each image, including the eye position, as shown in Fig. 2(a). In order to enhance the accuracy of skin areas localization to extract the local features, a non-reflective similarity transformation is applied to normalize each image based on two eyes and to scale

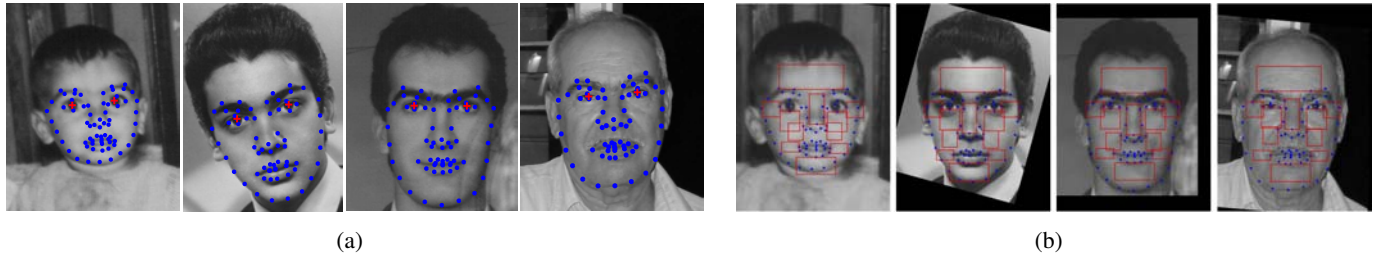


Figure 2. Sample images from FG-NET: (a) Grayscale conversion and the 68 landmarks and (b) after normalization and the eleven skin areas.

them to the same inter-pupillary distance [23]. The non-reflective similarity is a global transformation in which a single mathematical expression applies to an entire image and may include rotation, scaling, and translation. Shapes and angles are preserved, parallel lines remain parallel and straight lines remain straight. In this work, at least two control point pairs are required to perform the transformation. Therefore, the eye coordinates are used in order to have a normalized facial images with the same inter-pupillary distance as shown in Fig. 2(b).

The images from MORPH Album 2 were collected from mugshots, therefore, like in FG-NET, the images were not taken in controlled conditions. However, the previous preprocessing procedures are applied in order to have normalized images in terms of grayscale, pose and scale. Then, it is possible to apply a mask in all images to delimit and crop the skin areas of interest, which are described as follows.

### B. Skin Areas

Local features are important to describe facial components in which the wrinkles and spots appear such as eyes and mouth contour, cheeks, chin, nose and forehead. Chen *et al.* [8] identified several facial features such as the eye slope and mouth-nose distance that are significantly correlated with aging. They developed quantification methods for human facial aging features from 3D facial morphology for more than 300 individuals between 17 and 77 years old. They found that from young adulthood to old age, the mouth elongates, the nose becomes wider, the mouth-nose distance increases and eye corners droop. Also, young faces are smoother and thinner than old faces, while old faces have more face sagging, subcutaneous fat accumulation and fuller cheeks than young faces. According to the types of facial wrinkles described by Lemperle *et al.* [30], the following skin areas were chosen to capture the signatures of the aging process: corners of the eyes, mouth and nose, cheeks, chin, top of the nose and forehead, totaling eleven skin areas that are highly correlated with aging. Some landmarks were used to define the skin areas. For instance, to define the forehead area, two landmarks were used from the top of the eyebrows as reference for the width of the rectangle, while the height was manually defined in order to cover most of the forehead horizontal lines. It was preferable to define some rectangles slightly larger and with some areas overlapping others

in order to cover a wider skin area. The eleven skin areas defined are illustrated in Fig. 2(b). Once the skin areas are defined, it is necessary to crop them. This is essential for the normalization as it eliminates unnecessary areas and preserves the most relevant information for the skin analysis through local features. An example of the eleven cropped skin areas of a 69-year-old subject is shown in Fig. 3.

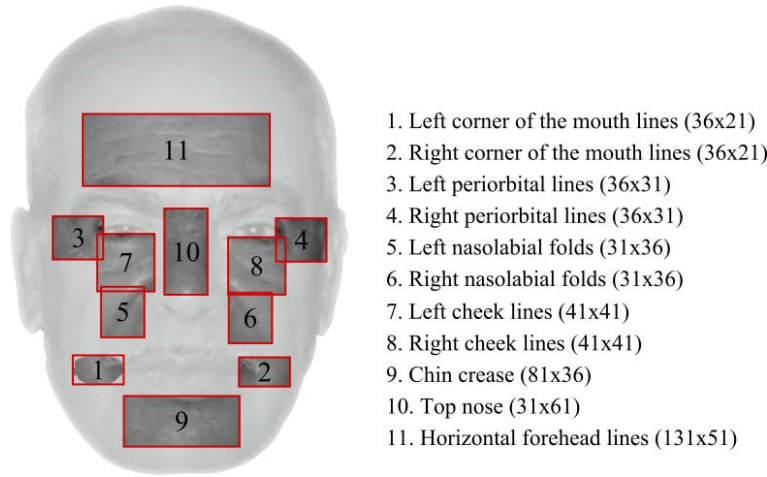


Figure 3. Example of cropped skin areas of a 69-year-old subject.

### C. Global Feature Extraction

Active Appearance Models is a widely used statistical model for facial modeling and feature extraction, where the variability of shape and texture are captured from a representative training set. Principal components analysis (PCA) on shape and texture data allows to produce a parametric face model that describes learned as well as new faces. A shape is described as a vector of coordinates from the landmark points. These shapes must be aligned to a common frame in order to analyze the shape variations in the dataset by its variance using the PCA. A small set of parameters is used to model the shape. A full facial representation requires modeling shape and also the texture. Similarly to the shape, the texture model also requires the alignment of texture samples to a reference texture frame. The statistical texture model is also built using the PCA. Finally, the shape and texture representations are combined in a single statistical model to represent a face in a compact way, where a single vector of appearance parameters can model shape and texture. These combined parameters are used as global features in the proposed approach in order to capture shape and appearance information related to aging. For this, 68 landmarks, provided by the FG-NET Aging and MORPH Album 2 datasets, are used to build the AAM. Lighting normalization and similarity transformation are carried out by the method itself. Therefore, it is not necessary any additional preprocessing.

#### D. Local Feature Extraction

Local features are extracted by the LBP, Gabor wavelets and LPQ operators. The LPQ is a powerful feature extractor that has been used as an alternative to the widely used LBP [1], mainly because it is very robust not only to blur but also to lighting and facial expression variations present in real-world images. Therefore, the LPQ is employed in this work to capture wrinkles through local phase features.

Also, some combinations between different techniques are performed in an attempt to capture some unique advantages of each one. The local features related to wrinkles are also extracted from a set of Gabor wavelets for being robust to noise such as hair, beard, shadows, etc. [2], [40]. The ones related to fine skin details such as fine wrinkles and spots are captured by LBP which can detect microstructures [35]. It is a simple yet very efficient texture descriptor that labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.

The  $LBP_{P,R}$  operator stands for a neighborhood of  $P$  equally spaced sampling points on a circle of radius  $R$  that form a circularly symmetric neighbor set. It produces  $2^P$  different output values corresponding to the  $2^P$  different binary patterns that can be formed by the  $P$  pixels in the neighborhood. Since certain bins contain more information than others it is possible to use only a subset of the  $2^P$  LBPs. These fundamental patterns are known as uniform patterns. An LBP is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For instance, 00000000, 00111000 and 11100001 are uniform patterns. According to Ojala *et al.* [35], nearly 90% of all patterns in the (8,1) neighborhood and about 70% in the (16,2) neighborhood are uniform in texture images. The most important characteristic of the LBP in real applications is its robustness to lighting changes and noise since it considers uniform binary patterns.

Gabor wavelets have been successfully applied to age estimation [2], [9] and for this reason they were chosen to be compared with the LPQ operator. A set of Gabor wavelets is the product of a Gaussian envelope and a plane wave, as defined in Eq. 1,

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\|k_{u,v}\|^2/\sigma^2} [e^{ik_{u,v}z} - e^{-\sigma^2/2}], \quad (1)$$

where  $z = (x, y)$  is the variable in the spatial domain and  $k_{u,v}$ , as defined in Eq. 2, is the frequency vector, that determines the scales and orientations of the wavelets.

$$k_{u,v} = \frac{k_{max}}{f^v} e^{i\phi_u}, \quad (2)$$

where  $k_{max} = \frac{\pi}{2}$ ,  $f = \sqrt{2}$  and  $\phi_u = \frac{u\pi}{8}$ . By varying the orientation  $u$  and the scale  $v$  factors, it is possible to

select different wavelets. Given an image, its Gabor transformation at a particular position can be computed by a convolution with the wavelets. All features derive from the magnitude of the resulting complex image.

In LPQ the phase is examined in local  $M$ -by- $M$  neighborhoods  $N_x$  at each pixel position  $x$  of the image  $f(x)$ . These local spectra are computed using a short-term Fourier transform that is efficiently evaluated for all image positions using simply 1D convolutions for the rows and columns successively. The local Fourier coefficients are computed at four frequency points,  $u_1 = [a, 0]^T$ ,  $u_2 = [0, a]^T$ ,  $u_3 = [a, a]^T$  and  $u_4 = [a, -a]^T$ , where  $a$  is a sufficiently small scalar. For each pixel position a vector,  $F_x = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)]$ , is generated.

The phase information in the Fourier coefficients is recorded by observing the signs of the real and imaginary parts of each component in  $F_x$  by using a scalar quantizer  $q_j(x)$  that assumes 1 in case of  $g_j(x) \geq 0$ , otherwise it assumes 0, where  $g_j(x)$  is the  $j$ th component of the vector,  $G_x = [Re\{F_x\}, Im\{F_x\}]$ . The binary coefficients  $q_j(x)$  are represented as integer values (0-255) using the coding defined by Eq. 3,

$$f_{LPQ}(x) = \sum_{j=1}^8 q_j(x) 2^{j-1}. \quad (3)$$

As a result, we get the label image  $f_{LPQ}$  whose values are the blur invariant LPQ labels. All these techniques for local skin texture description are applied in all eleven cropped and normalized skin areas. Once the local features of all of the eleven areas have been extracted independently, the feature vectors are concatenated to have a single vector with all the descriptions to represent the facial image as a whole, as described in the next section.

### E. Feature Fusion

Fusion of features at feature level has been successfully applied in biometrics [37], facial expressions [43], music [11], images [34], etc. For this reason, the feature level fusion is employed in this work. The main assumption is that different feature sets are able to extract different information from raw data which are supposed to be complimentary.

Given an input image, the global features are extracted from the combined parameters of the AAM,  $f_{AAM} = (c_1, \dots, c_k)$ . The local features are then extracted through the Gabor wavelets,  $f_{GW} = (w_1, \dots, w_l)$ , LBP,  $f_{LBP} = (t_{11}, \dots, t_{1m})$  and LPQ,  $f_{LPQ} = (t_{21}, \dots, t_{2n})$ , where  $k$ ,  $l$ ,  $m$  and  $n$  are the original vector dimensions. Because the original local feature vector has a high dimensionality, only the principal components are selected with the PCA. The z-score is employed to normalize the lower dimensionality features [37]. In statistics the z-score is used to compare the mean of different datasets homogeneously distributed.

In fact, when instances of different datasets are transformed in z-scores they become comparable. The mean and standard deviation are necessary to calculate it according to Eq. 4,

$$f'_i = \frac{f_i - \mu_i}{\sigma_i}, \quad i = AAM, GW, LBP \text{ or } LPQ, \quad (4)$$

where  $f_i$  is the feature vector (AAM, GW, LBP or LPQ),  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of the feature vector selected respectively, and  $f'_i$  is the normalized feature vector. The feature fusion is made up by the concatenation of the normalized feature vectors as follows:

$$f_{fused} = (f'_{AAM} f'_{GW} f'_{LBP} f'_{LPQ}) = (c_1, \dots, c_k, w_1, \dots, w_l, t_{1_1}, \dots, t_{1_m}, t_{2_1}, \dots, t_{2_n}), \quad (5)$$

where  $f_{fused}$  is the fused feature vector to be used in the dimensionality reduction step which will be discussed next. During the feature combination, weights are not applied in the feature vectors to avoid a dependency with the datasets. This is because the feature weights extracted from the AAM on FG-NET could not be the same as the ones extracted from the MORPH dataset. Fig. 4 shows an overview of the global and local feature fusion method.

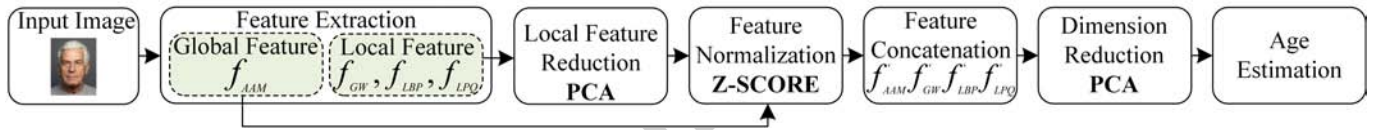


Figure 4. Overview of the global and local feature fusion method.

## F. Dimensionality Reduction

Generally, dimensionality reduction algorithms are used to reduce the number of features under consideration. In this direction, the PCA is used in this work to reduce the dimensionality of the feature vectors derived from the Gabor wavelets, LBP and LPQ techniques. PCA is a widely adopted technique to construct a subspace that captures the main variation in the dataset [26]. Also, due to the concatenation of global and local features, both the final feature space dimension and the computational cost are increased. However, the dimension of the feature vectors are reduced using PCA,  $f_{PCA} = (f_1, \dots, f_d)$ ,  $d < k + l + m + n$ , where  $d$  is the new dimension of the features fused and  $f_{PCA}$  is the feature vector to be employed in the age estimation, as discussed next.

## G. Age Estimation

The aging process can differ according to the age group of a subject. Wrinkles are normally found in elderly people, while geometric features mainly change during childhood and adulthood. Such specific



age-related facial features may not be considered when single-level classifiers are used. Hierarchical age estimation based on age-group specific classifiers has shown improved results and better ability to cope with age-related facial features [9]. However, age estimation from a hierarchical approach that neglects overlapping of ages between age groups has a drawback. For example, a 15-year-old individual could be classified into an age group from 20 to 39 years old during the classification step. This error will carry on to the specific age estimation and the performance will be reduced.

The proposed approach first classifies a facial image into a determined age group by using a SVM and then uses a SVR to define the specific age estimation (Fig. 1). In addition, flexible overlapped age ranges are considered for the specific age estimation step to reduce misclassification produced at the classification step, as illustrated in Fig. 5. The overlapped classes are defined according to the misclassification errors and can be flexible for each SVR.

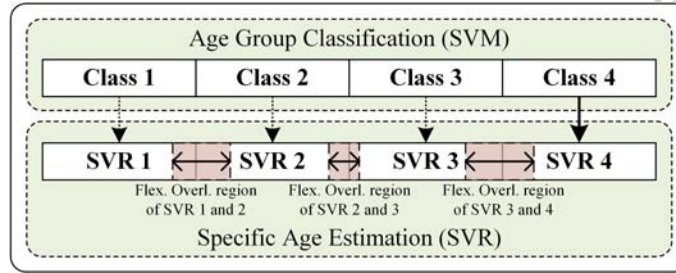


Figure 5. Hierarchical approach with flexible overlapped regions between the age classes in the specific age estimation step.

#### IV. EXPERIMENTAL RESULTS

##### A. Datasets

The performance of the proposed approach was evaluated on the publicly available FG-NET Aging [25] and MORPH Album 2 [36] datasets. The FG-NET Aging contains 1,002 facial images of 82 multiple-race subjects with ages ranging from 0 to 69 years old. The dataset is unbalanced since 50% of the subjects are between 0 and 13 years old. The images are either in color or in grayscale with an average dimension of  $384 \times 487$  pixels, and the resolution varies from 200 dpi to 1,200 dpi with large variation of lighting, pose, facial expression, blur and occlusions (*e.g.*, moustache, beard, glasses, etc.). Some facial image samples from FG-NET Aging are shown in Fig. 6(a). Each facial image of the dataset was manually annotated with 68 landmark points characterizing the subjects' shape features as shown in Fig. 7 and its description is shown in Tab. I, which shows the landmarks according to the facial elements. These points are useful for locating facial features, especially for the AAM.

The MORPH Album 2 is a dataset of mugshot images with associated metadata giving the age, ethnicity, and gender of each subject. It contains 55,608 facial images with about three aging images per person





Figure 6. Facial image samples from (a) FG-NET Aging and (b) MORPH Album 2 datasets.

ranging from 16 to 77 years old. The images are in color and have a uniform gray background. The average dimension is  $400 \times 480$  pixels and the resolution is between 81 dpi and 96 dpi with changes in lighting, pose, facial expression and occlusion. Some facial image samples from MORPH Album 2 are shown in Fig. 6(b). Chang *et al.* [4] selected 5,492 facial images for their age estimation approach and since the MORPH does not provide the point files they have annotated 68 landmarks automatically for each image. Therefore, to take advantage of the point files publicly available and to evaluate the proposed approach, the same dataset was used. However, 18 images were deleted because of the great misalignment of the landmarks, generating a dataset containing 5,474 images for this work. An analysis of the features used in this work under both datasets is discussed in the following section.

## B. Feature Analysis

1) *Active Appearance Models*: The AAMs were generated separately for the FG-NET and MORPH datasets using both pairs of 68 landmarks for each image. The annotations describe the shapes of different facial components such as eyes, eyebrows, nose, mouth and facial contour as shown in Fig. 8. They are essential to construct the statistical shape and appearance models. The Procrustes analysis is performed to align the landmark sets and build the shape model. Each training sample has its landmarks warped to the mean shape learned in order to get the shape free patch. The texture intensities of the shape free patch are normalized by a linear transformation. Thus, the PCA can be applied to build the appearance model. Finally, a correlation between the shape and the appearance models is learned to build a combined model.

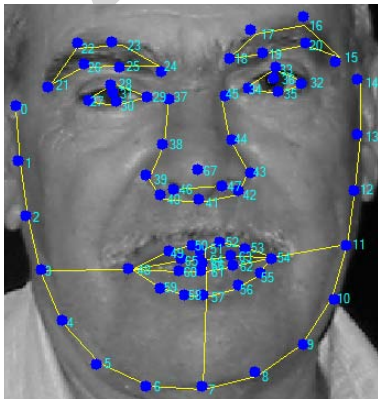


Figure 7. Facial image sample and its 68 landmarks.

Table I  
FACIAL LANDMARK DESCRIPTION ACCORDING TO FIG. 7.

Facial elements	Landmarks	Number of points
Facial contour	0-14	15
Right eyebrow	15-20	6
Left eyebrow	21-26	6
Left eye	27-30	4
Left pupil	31	1
Right eye	32-35	4
Right pupil	36	1
Nose	37-45	9
Nose center	67	1
Nostrils	46-47	2
Mouth contour	48-59	12
Lower lip	48, 60, 61, 62, 54	5
Upper lip	54, 63, 64, 65, 48	5

To handle the resolution changes present in the datasets, a multi-resolution model is built based on a Gaussian pyramid [10] that permits a multi-resolution analysis with increased robustness and performance. For each level of the pyramid, an independent appearance model is built where a proportional number of pixels to the target image area is sampled. In this work, about 20,000 intensity values were sampled from the facial image through the mean shape model learned. The first level ( $n_0$ ) has 20,000 pixels, the second level ( $n_1$ ) has 10,000 pixels and the third level ( $n_2$ ) of the pyramid has 5,000 pixels sampled. In the levels  $n_1$  and  $n_2$  is applied a Gaussian filter for smoothing before the subsampling.

For facial texture feature extraction the AAM was set to use the convex hull algorithm [39] in the facial components by using the triangulation as shown in Fig. 8. The truncation level for the details that represent the shape, texture and the combined model was set to 95%. That means that the resulting model describes 95% of the variations present in the training datasets.

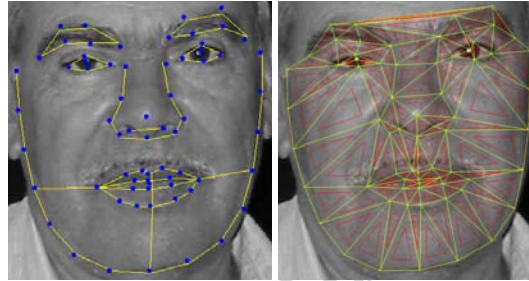


Figure 8. The 68 landmarks describing the facial components and the triangulation to describe the texture respectively.

The features related to identity variations are better represented than features related to aging because AAM was developed for facial modeling. Thus, some relevant age features such as wrinkles and other spots are removed after the reduction of the feature space by the PCA. This can be noticed in Fig. 9 which shows a comparison among the original, the reconstructed and the residual images (*i.e.*, the absolute difference between the original and the reconstructed image). The AAM does not sufficiently reconstruct the age feature, such as the wrinkles in the eyes corner and the nasolabial lines of the individual. Since the representation power of the AAM is limited, a local feature descriptor becomes necessary in order to have a better facial aging representation.

Since the combined features are sorted in descending order, the first parameter corresponds to the largest eigenvalue of the covariance matrix, which gives its variance across the training set. The first five parameters, which have the greatest variances, are enough to see the main changes that each feature parameter can cause in the facial model. The results from the synthesis of the facial model using the first five combined features in both directions of the standard deviation learned are shown in Fig. 10. The middle column shows the mean face learned by the AAM while the first four features cover mainly the head position and lighting

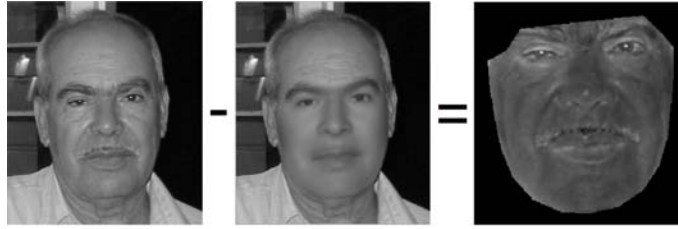


Figure 9. Original, reconstructed, and the residual image respectively. Some relevant age features such as wrinkles and other spots, as seen in the original image, are removed after the reduction of the feature space in the AAM, as seen in the reconstructed image. The residual image is the absolute difference between the original and the reconstructed image.

variations due to the great variety of non-controlled images in the dataset. The fifth feature contains more information related to age, mainly shape features, that justifies the use of AAM as global descriptor. Some examples of reconstructed faces from the FG-NET dataset are shown in Fig. 11, which shows the efficiency of the AAM for facial modeling.

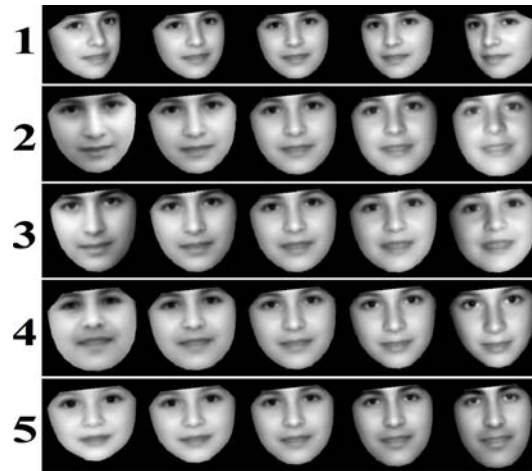


Figure 10. Effect of varying first five facial appearance model parameters in both directions of the standard deviation. The first four features cover mainly the head position and lighting variations. The fifth feature contains more information related to age, mainly shape features.

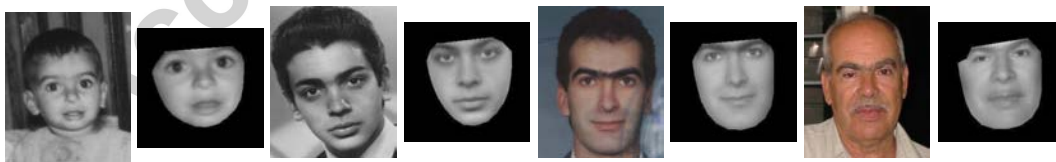


Figure 11. Examples of reconstructed faces from FG-NET by using AAM (2, 17, 28 and 69 years old respectively).

2) *Gabor Wavelets*: Facial wrinkles provide important information for age estimation. Created by repetitive movement of facial muscles and expression, they are formed perpendicularly to the facial muscles direction. For instance, the forehead wrinkles have a horizontal direction while the nasolabial folds are diagonal. Consequently, facial wrinkles have unique characteristics that depend on the direction of muscles, the frequency of use and the effects of gravity.

Gabor wavelets are evaluated in this work to detect lines, and since a face can contain wrinkles with different thicknesses it is necessary to generate wavelets with different scale factors ( $v$ ). Also, since an image can have wrinkles in many directions, the wavelets are created with different orientations ( $u$ ). To extract such features and capture wrinkles, a set of Gabor wavelets with different scales and orientations was applied in the eleven skin areas. By using a set of 40 wavelets (*i.e.*, eight orientations and five scales), it is possible to extract a great deal of information related to the intensities and orientations of wrinkles from the mean and variance of the magnitude response of each skin area. In addition, it was evaluated a Gabor wavelets with six orientations and four scales.

Fig. 12(a) and Fig. 12(b) show the forehead skin area of a 2 and a 69-year-old subject used to evaluate the Gabor wavelets with eight orientations and five scales. The magnitudes used as features and the corresponding mean values are illustrated in Fig. 12(c) for a 2-year-old, and in Fig. 12(d) for a 69-year-old subject. Since the child forehead skin is smooth, the magnitude response of the wavelets has less intensity because of few forehead lines on it. Otherwise, the forehead of a senior adult has more wrinkles and the magnitude response has more frequency content. In general, the mean magnitude values are bigger for the old subject than the younger one. This shows that Gabor wavelets are a very discriminative local descriptor for age estimation, especially when the wrinkles are very prominent on the image. However, several images from the datasets have a sort of blur, which can reduce the effectiveness of the Gabor wavelets.

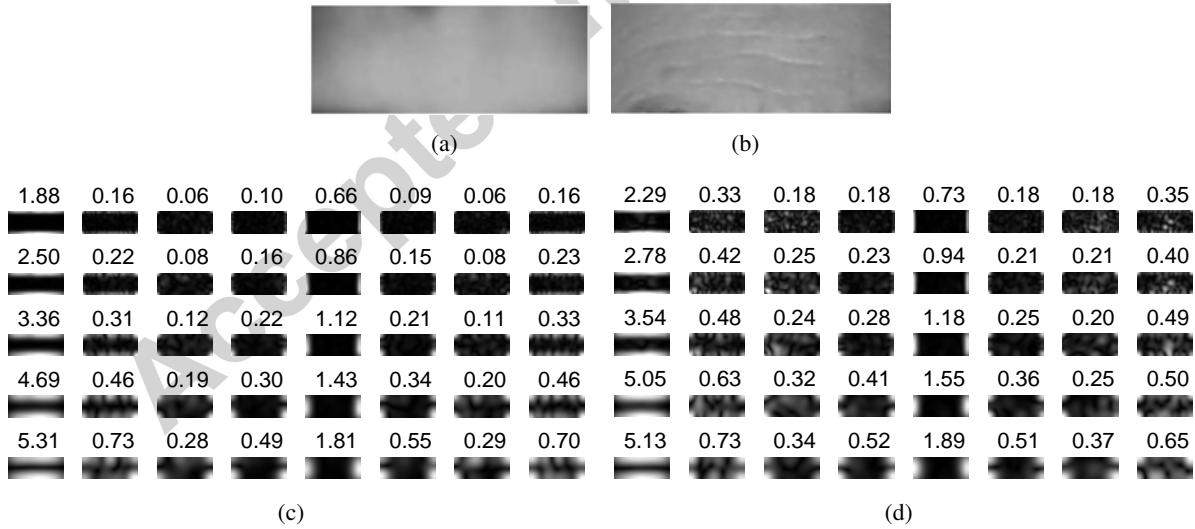


Figure 12. Forehead skin area used to apply the Gabor wavelets: (a) 2 and (b) 69 years old, and its magnitudes and mean values for eight orientations (columns) and five scales (rows) of the forehead skin area: (c) 2 and (d) 69 years old. Since the child forehead skin is normally smooth, the mean magnitude response of the wavelets has less frequency content than the forehead of a old subject which has more wrinkles.

3) *Local Binary Patterns*: The LBP was also evaluated as local feature descriptor in this work. The LBP operator extracts the local features of the eleven skin areas with variations on its parameters. The  $LBP_{P,R}^{u2}$ , where  $P$  and  $R$  are the circular neighborhood parameters, and  $P$  can assume eight or sixteen pixels while

$R$  can vary from one to three pixels. The naming  $u2$  means that only uniform patterns are used to reduce noise in the local texture structures. There are 256 patterns that can be detected by the LBP considering uniform and non-uniform patterns. Therefore, only 59 codes are used as uniform patterns, where 58 are uniform and one is used to aggregate all the non-uniform codes.

Fig. 13 shows an  $LBP_{8,2}^{u2}$  code histogram of the forehead area. An example of each age group was used: child, young, adult and senior adult. A forehead skin of a 2-year-old child is normally smooth as indicated in the histogram by the code 58, which represents a dark spot with brighter surrounding pixels. Code one in the histogram represents few spots, with all surrounding pixels having an intensity value less than the central pixel value. Some codes related to edges, corners and line ends are due to the skin area be cropped with some noise from eyebrows and hair region as indicated in the histogram. The 17-year-old subject has more codes related to spots and also a predominance of the code 59, which represents the non-uniform patterns that define many curvatures of facial lines. The 28-year-old's histogram is similar to the 17-year-old's, only changing some codes that are derived from noise caused by eyebrows in both histograms. For the last histogram, 69 years old, it is possible to find out some codes related to spots caused by aging. Various codes related to edges (*e.g.*, 11 and 15) and corners (*e.g.*, 7, 10, 15 and 25) are also observed, mainly due to the forehead horizontal lines. The LBP descriptor is very discriminative and captures mostly edges and spots information on the skin. However, the images should have good quality and blur-free, otherwise the performance can be harmed.

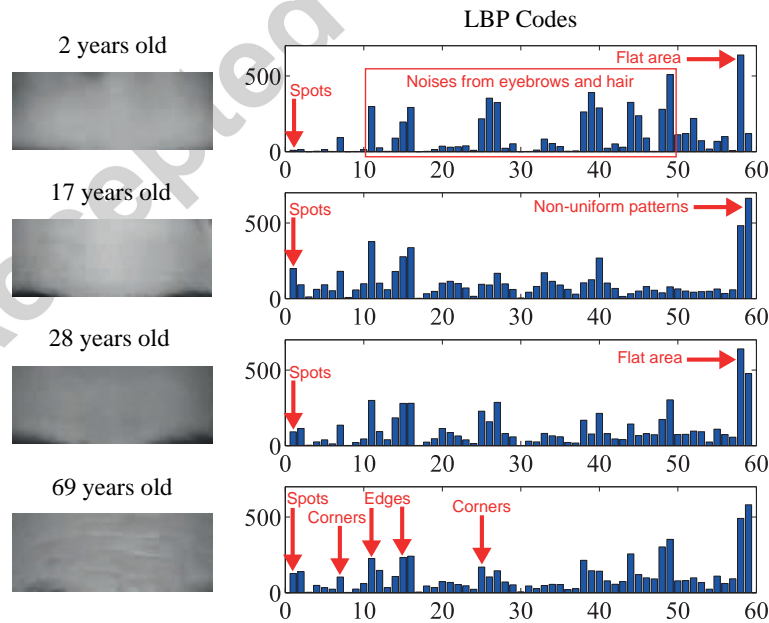


Figure 13.  $LBP_{8,2}^{u2}$  code histogram of the forehead skin area of a 2, 17, 28 and 69-year-old subjects respectively. A forehead skin of a 2-year-old child is normally flat, which represents a dark spot with brighter surrounding pixels. The 17 and 28-year-old subject has more codes related to spots and the older forehead has more codes related to spots, edges and corners, caused by aging.

4) *Local Phase Quantization*: In the experiments using the LPQ technique the features were extracted and evaluated using window sizes ( $M$ ) of 3, 5, 7, 9 and  $11 \times 11$  with  $a = 1/3$  and  $\rho = 0.9$ . The LPQ is computed locally for a window in every skin area position, providing an histogram of 256 codes as shown in Fig. 14. Such a figure gives examples of histograms for some forehead areas. The code values are generally high for young subjects (e.g., 2, 17 and 28 years old) since the phase orientations are similar due to the skin texture being more smooth. In the other hand, the 69-year-old's skin has more wrinkles and spots caused by aging. Therefore, a high diversity of phase orientations can be seen in the histogram shown in Fig. 14. As skin gets aged, the LPQ codes tend to have a more uniform distribution, with low histogram values due to the great diversity of phase orientations caused specifically by wrinkles in many directions. The LPQ is very powerful for texture description and it shows to be more discriminative than the LPB operator.

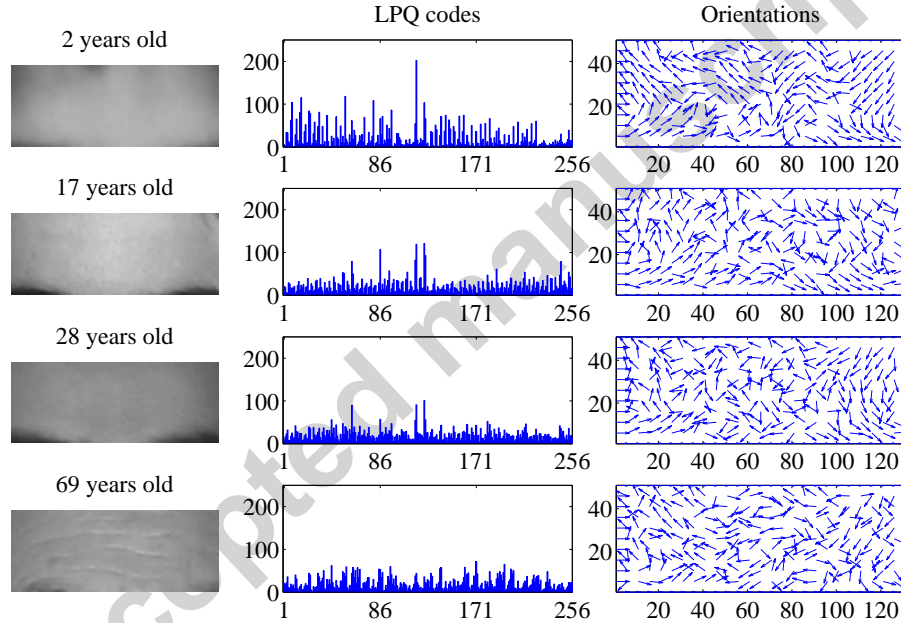


Figure 14.  $LPQ_{7 \times 7}$  code histogram for the forehead skin area and unit vectors illustrating the characteristic orientations of a 2, 17, 28 and 69-year-old subjects respectively. As skin gets aged, the LPQ codes tend to have a more uniform distribution, with low histogram values due to the diversity of phase orientations caused by wrinkles in many directions.

### C. Evaluation Metrics

The performance was evaluated by using the  $MAE$  and the cumulative score ( $CS$ ) [16]. The  $MAE$  is the average of the absolute errors between the estimated age ( $a$ ) and the ground truth age ( $\hat{a}$ ) defined as

$$MAE = \frac{\sum_{i=1}^N |\hat{a} - a|}{N}, \quad (6)$$

where  $N$  is the total number of test images. The  $CS$  allows performance comparison at different absolute



error levels and it is defined as

$$CS(l) = \frac{N_{e \leq l}}{N} \times 100\%, \quad (7)$$

where  $N_{e \leq l}$  is the number of test images on which the age estimation makes an absolute error lower or equal to  $l$  years.

#### D. Evaluation Protocols

The FG-NET and MORPH datasets were split uniformly into 75% for training and 25% for test to perform the experiments. The SVM and SVR employ a linear kernel and their parameters were found by a grid search using 5-fold cross validation. The LIBSVM library [3] was used in this work for both classification and regression tasks. The solver used for classification was the L2-regularized L2-loss support vector classification (dual) and for regression the L2-regularized L2-loss support vector regression (primal).

Since the FG-NET dataset has limited data (1,002 images), the Leave-One-Person-Out (LOPO) protocol was also used to evaluate the performance of the proposed approach. The LOPO protocol considers all samples of one subject to test the approach while the samples of all other subjects are used to train. This leads to 82 folds, since the FG-NET has 82 individuals. This scheme ensures that a person is not in the training and test set simultaneously, so the classifier does not learn individual characteristics, thus reducing the dependence on data on the experimental results.

#### E. Results on FG-NET Aging Dataset

For age group classification, four age ranges were defined according to the data distribution with the aim of having enough data for training and test. This also allowed having a group for children, young adults, adults and seniors, especially useful for the FG-NET Aging since it has a wide age range. The defined classes and data distribution are shown in Tab. II.

Table II  
CLASSES AND DATA DISTRIBUTION FOR TRAINING AND TEST SETS (FG-NET).

Class	Age group	Training set	Test set
1	0-13	385	128
2	14-21	175	58
3	22-39	140	47
4	40-69	52	17
Total	0-69	752	250



Several combinations of feature sets were evaluated, either isolated global and local features or the fusion of both to have a better facial representation. The age group classification performance on these feature sets is shown in Fig. 15. The radial basis function (RBF) kernel and the linear kernel were evaluated. The highest accuracy with the RBF kernel was 70% with the combination of global and local features (AAM+GW<sub>6,4</sub>). With the linear kernel, the highest accuracy, 68.80%, was achieved with the global and local features (AAM+LPQ<sub>9×9</sub>). In general, the classification performance by using the linear kernel was higher than the RBF kernel, as can be seen in the graph. Moreover, the linear kernel performs a faster training than the RBF kernel.

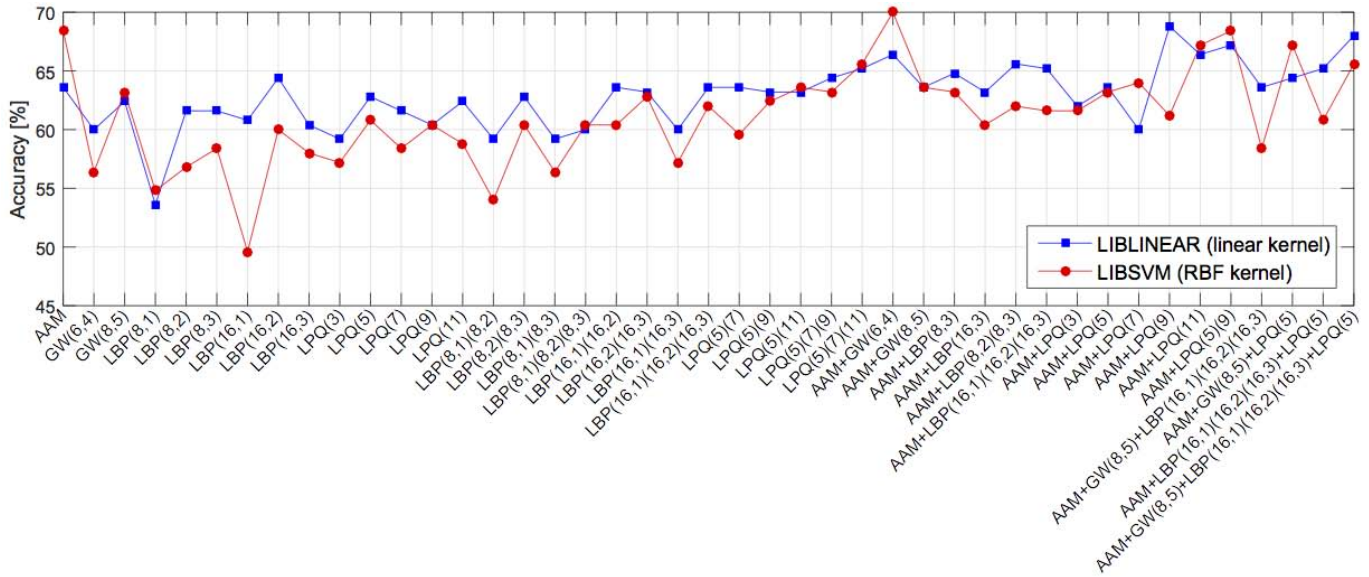


Figure 15. Comparison of the age group classification performance using RBF and linear kernel in different feature sets (FG-NET).

The confusion matrix of the age group classification using the global and local feature fusion which achieved an overall accuracy of 60% (AAM+LPQ<sub>7×7</sub>) is shown in Fig. 16. This feature combination was selected since it achieved the best *MAE* (Tab. III). The highest accuracy of 92.20% was achieved for class 1 with only 7.80% of misclassification where part of the test instances were classified as classes 2 and 3. On the other hand, classification errors above 70% were obtained for the three other classes. Most of the instances from class 2 were classified as class 1, while the instances from class 3 were misclassified as class 1 and 2. This confusion occurred because the individual of classes 1, 2 and 3 have few wrinkles, no skin spots due to good genetics, lifestyle, health, etc. The low performance on class 4 is due to the lack of data for training and test. The confusion matrix clearly shows the tendency of the FG-NET Aging dataset to be more accurate in the age group from 0 to 13 years old. In addition, the confusions also occurred due to hard boundaries between the age classes. For example, a 14-year-old subject from class 2 can easily be classified as class 1 which covers ages from 0 to 13 years old.

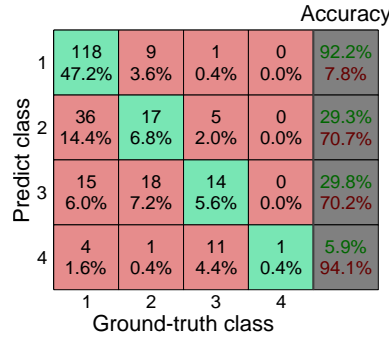


Figure 16. Confusion matrix for the age group classification using the combination of global and local features (AAM+LPQ<sub>7×7</sub>) (FG-NET).

In order to compensate the errors generated in the age group classification step, flexible overlapped age ranges in the specific age estimation step are proposed in this work. The age ranges for the SVRs were defined based on the same ranges described previously but with a flexible overlapping between age ranges which were defined experimentally on the training dataset. The classification errors were analyzed to decide on the initial values of the flexible overlapped regions, from which, through the increment or decrement of the overlapped ages, the values were chosen based on the best results in terms of *MAE* of the age estimation approach. The ages for the flexible overlapped regions between the age groups that achieved the lower *MAE* are illustrated in Fig. 17. For example, the age group used to train the SVR 1 is from 0 to 28 years old, which overlaps the SVR 3 trained with the age group from 22 to 60 years old. Such overlapped regions reduce the errors generated in the age group classification step by using regressors trained with a broader age groups.

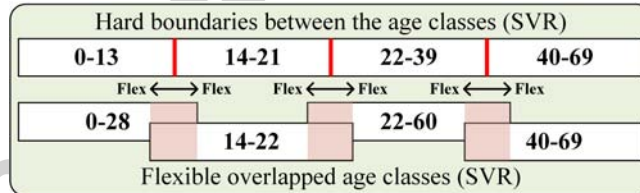


Figure 17. Ages used for the flexible overlapped regions between adjacent SVR age classes (FG-NET).

Fig. 18 shows the performance in terms of *MAE* of the specific age estimation with flexible overlapped regions between the age groups on several feature sets. Looking at each SVR individually, the best SVR 1 achieved a *MAE* of 3.41 years using the AAM+GW<sub>8,5</sub> feature set. The best SVR 2 achieved a *MAE* of 4.64 years using the LBP<sub>(16,1)(16,2)(16,3)</sub> feature set. The best SVR 3 achieved a *MAE* of 6.63 years using the AAM+LPQ<sub>(5×5)(9×9)</sub> feature set. Finally, the best SVR 4 achieved a *MAE* of 0.09 years using the AAM+LPQ<sub>7×7</sub> feature set. It is possible to see through the graph of Fig. 18 that the SVR 1 has the best overall *MAE*, since most of the images from the training set are into its age group. Otherwise, the *MAE* of the SVR 4 oscillates widely across the extremes, due to the fact that there is limited data for training

the 40-69 years old class. Fig. 18 clearly shows that the  $MAE$  increases with age as expected, since the main component associated with face shape changes greatly from 0 to 18 years of age with no significant changes in subjects older than 18 years. This facial shape feature makes determining a person's age easier, rather than looking at facial skin features that only change significantly in older people.

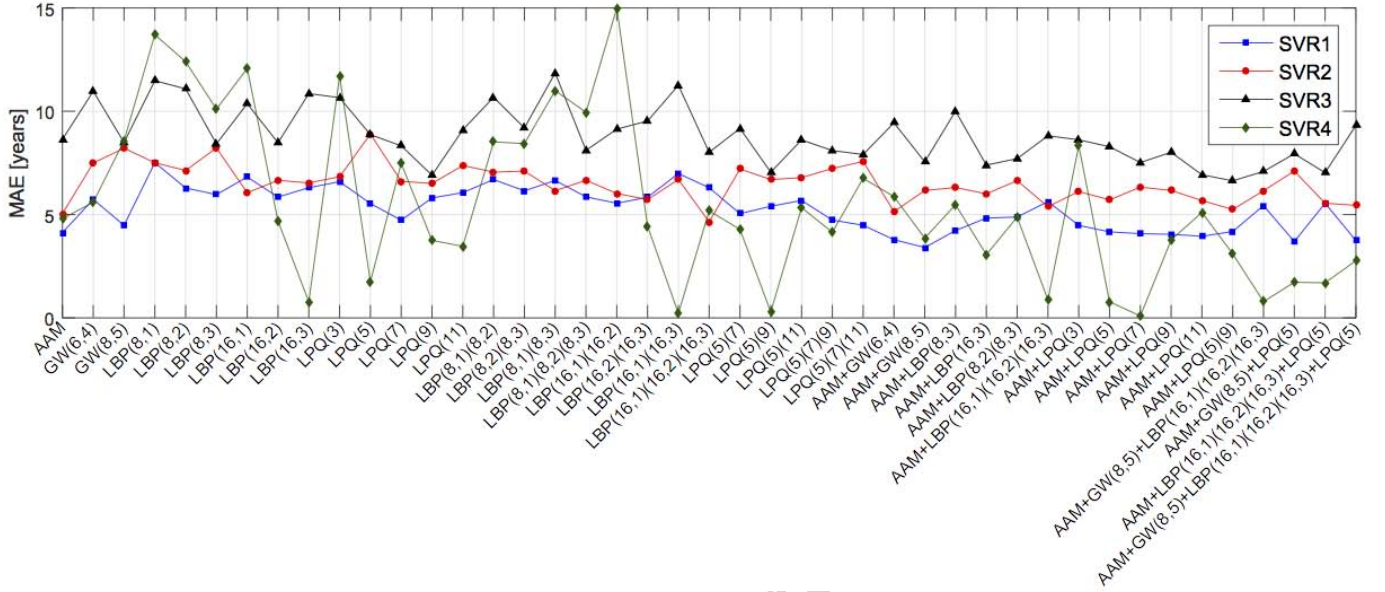


Figure 18. Performance in terms of  $MAE$  of the specific age estimation with flexible overlapped regions between the age groups in different feature sets (FG-NET).

Tab. III shows some results of the hierarchical age estimation with flexible overlapped age ranges, hard boundaries between the age classes and with fixed overlapped age ranges ( $\pm 5$  years for each class) in several feature sets. The combination of global and local features (AAM+LPQ $_{7 \times 7}$ ) achieved the best performance with a  $MAE$  of 4.50 years with flexible overlapping while a  $MAE$  of 5.30 years was achieved with fixed overlapping. Therefore, the flexible overlapping has a performance gain of 15.09% when comparing to the fixed one. Furthermore such a performance also overcomes the  $MAE$  of the hierarchical approach with hard boundaries on the FG-NET dataset. Also, it is possible to see that the combination of the global and local features has a better performance than using the features individually. The LPQ outperformed the LBP and Gabor wavelets showing that it is more discriminative for age estimation applications. A single-level approach trained in all ages using a single SVR was also evaluated in different feature sets. Using the AAM+LPQ $_{7 \times 7}$  it achieved 5.51 years of  $MAE$  as shown in Tab. III. This means that the flexible overlapping trained in a small age group had a performance gain of 18.33% over the single-level approach.

Tab. III also shows the original dimension of each feature set and the corresponding dimension after the PCA. The highest dimension comes from Gabor wavelets with eight orientations and five scales summing up 8,331 features for the eleven skin areas. With the PCA only 180 discriminant features were selected from the

GW<sub>8,5</sub>. The new dimensions were decided experimentally through the performance of the proposed approach considering each feature set. The performance was measured from incrementing the PCA dimension (*i.e.*, the number of principal components) and the dimension with the best *MAE* was chosen.

Fig. 19 shows the *CS* for the proposed approach at error levels from 0 to 15 years on the FG-NET dataset. As a result, the *CS* moved up by combining a holistic facial representation (AAM) and local features (LPQ<sub>7×7</sub>). For instance, 70% of the age estimation has an error less than or equal to 5 years old. In addition, it is possible to visualize that the combination of the global and local features have a better performance than their separate use.

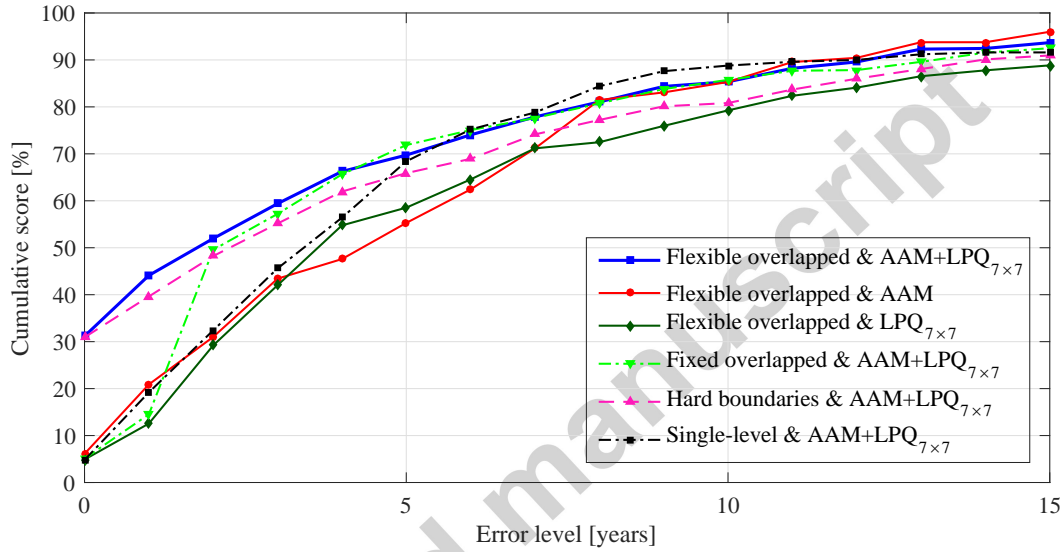


Figure 19. Cumulative scores of proposed and other age estimation methods at error levels from 0 to 15 years on FG-NET Aging.

#### F. Results on MORPH Album 2 Dataset

For age group classification, four age ranges were defined according to the data distribution with the aim of having enough data for training and test. Tab. IV shows the class and data distribution. The best overall classification accuracy of 50.70% was achieved by using the local features LPQ<sub>9×9</sub> as shown in Tab. V. It is important to notice that the best performance on the MORPH dataset was achieved using only local features. This is due to the fact that the age range in the MORPH dataset is from 16 to 77 years, and the facial shape does not change much at these ages but the texture does. Therefore, the AAM was not very discriminative in this dataset. A significant gain of 14.68% was achieved with the flexible regions relative to the hierarchical approach with hard boundaries which achieved a *MAE* of 6.72 with the same features. Tab. V also shows that the local features extracted with the LPQ operator were the most discriminant in both FG-NET and MORPH datasets with the best performance in terms of *MAE*.

Table III  
RESULTS OF DIFFERENT AGE ESTIMATION APPROACHES IN SEVERAL FEATURE SETS (FG-NET).

Feature set	Dimension Orig./PCA	Classification Accuracy [%]	Single-level [MAE]	Hierarchical Hard Boundaries [MAE]	Hierarchical Fixed ( $\pm 5$ ) Overlapped [MAE]	Hierarchical Flexible Overlapped [MAE]
AAM	-/62	63.60	5.64	5.91	5.11	5.65
GW <sub>6,4</sub>	4,999/237	60.00	7.62	7.32	7.01	7.45
GW <sub>8,5</sub>	8,331/180	62.40	6.44	7.54	7.40	7.44
LBP <sub>8,1</sub>	649/177	53.60	8.13	9.90	9.80	10.05
LBP <sub>8,2</sub>	649/200	61.60	7.69	9.51	9.68	9.22
LBP <sub>8,3</sub>	649/163	61.60	6.98	8.45	8.28	8.18
LBP <sub>16,1</sub>	2,673/171	60.80	7.79	8.95	9.14	8.83
LBP <sub>16,2</sub>	2,673/152	64.40	7.63	6.84	7.72	6.42
LBP <sub>16,3</sub>	2,673/241	60.40	7.11	6.42	6.23	6.12
LPQ <sub>3×3</sub>	2,816/132	59.20	7.90	9.16	8.53	8.94
LPQ <sub>5×5</sub>	2,816/200	62.80	6.98	6.60	6.49	6.26
LPQ <sub>7×7</sub>	2,816/197	61.60	6.62	7.31	7.35	6.79
LPQ <sub>9×9</sub>	2,816/195	60.40	6.74	6.35	5.82	5.74
LPQ <sub>11×11</sub>	2,816/192	62.40	6.78	6.25	6.82	6.50
LBP <sub>(8,1)(8,2)</sub>	377/194	59.20	7.74	8.32	9.06	8.24
LBP <sub>(8,2)(8,3)</sub>	363/169	62.80	7.07	8.10	8.37	7.72
LBP <sub>(8,1)(8,3)</sub>	340/51	59.20	7.74	8.71	8.77	8.91
LBP <sub>(8,1)(8,2)(8,3)</sub>	540/246	60.00	7.16	8.50	8.18	7.64
LBP <sub>(16,1)(16,2)</sub>	323/159	63.60	7.51	8.97	9.40	8.92
LBP <sub>(16,2)(16,3)</sub>	393/176	63.20	7.27	6.84	7.11	6.38
LBP <sub>(16,1)(16,3)</sub>	412/247	60.00	7.49	6.39	6.54	6.29
LBP <sub>(16,1)(16,2)(16,3)</sub>	564/206	63.60	7.38	6.16	6.28	6.03
LPQ <sub>(5×5)(7×7)</sub>	397/213	63.60	6.56	6.81	6.89	6.43
LPQ <sub>(5×5)(9×9)</sub>	395/212	63.20	6.42	5.44	5.49	4.86
LPQ <sub>(5×5)(11×11)</sub>	392/204	63.20	6.55	7.08	6.66	6.60
LPQ <sub>(5×5)(7×7)(9×9)</sub>	592/164	64.40	6.46	6.34	6.13	6.05
LPQ <sub>(5×5)(7×7)(11×11)</sub>	589/186	65.20	6.52	7.09	7.12	6.68
AAM+GW <sub>6,4</sub>	299/54	66.40	5.49	6.02	5.98	6.06
AAM+GW <sub>8,5</sub>	242/58	63.60	<b>5.31</b>	5.61	5.54	5.25
AAM+LBP <sub>8,3</sub>	225/201	64.80	5.99	6.32	6.02	6.50
AAM+LBP <sub>16,3</sub>	303/191	63.20	6.25	5.92	6.25	5.32
AAM+LBP <sub>(8,2)(8,3)</sub>	231/261	65.60	6.68	6.24	6.62	6.03
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub>	268/265	65.20	6.93	5.42	6.53	5.17
AAM+LPQ <sub>3×3</sub>	194/170	62.00	6.28	7.10	6.43	6.88
AAM+LPQ <sub>5×5</sub>	262/215	63.60	6.12	<b>5.13</b>	5.19	4.73
AAM+LPQ <sub>7×7</sub>	259/158	60.00	5.51	5.31	5.30	<b>4.50</b>
AAM+LPQ <sub>9×9</sub>	257/129	<b>68.80</b>	<b>5.31</b>	5.54	5.69	5.50
AAM+LPQ <sub>11×11</sub>	254/102	66.40	5.72	5.91	5.93	5.41
AAM+LPQ <sub>(5×5)(9×9)</sub>	274/222	67.20	6.08	5.51	5.31	4.79
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	448/297	63.60	6.51	5.28	6.06	4.86
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	442/137	64.40	5.78	5.36	<b>4.85</b>	5.12
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	468/329	65.20	6.38	5.33	4.96	4.95
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	648/151	68.00	5.88	5.17	5.79	5.32

Table IV  
CLASSES AND DATA DISTRIBUTION FOR TRAINING AND TEST SETS (MORPH).

Class	Age group	Training set	Test set
1	16-25	1,065	520
2	26-35	1,115	369
3	36-45	1,295	266
4	46-77	632	212
Total	16-77	4,107	1,367

Table V  
RESULTS OF DIFFERENT AGE ESTIMATION APPROACHES IN DIFFERENT FEATURE SET (MORPH).

Feature set	Classification Accuracy [%]	Hierarchical Hard Boundaries [MAE]	Hierarchical Flexible Overlapped [MAE]
AAM	31.31	11.61	10.84
GW <sub>8,5</sub>	42.36	8.02	7.26
LBP <sub>8,1</sub>	47.92	7.14	6.36
LBP <sub>8,2</sub>	48.57	7.07	6.38
LBP <sub>8,3</sub>	49.23	7.16	6.39
LBP <sub>16,1</sub>	47.48	7.10	6.31
LBP <sub>16,2</sub>	47.55	7.18	6.29
LBP <sub>16,3</sub>	47.99	7.20	6.31
LPQ <sub>5×5</sub>	50.77	<b>6.67</b>	5.90
LPQ <sub>7×7</sub>	50.69	6.72	<b>5.86</b>
LPQ <sub>9×9</sub>	<b>51.06</b>	6.74	5.91
LBP <sub>(16,1)(16,2)(16,3)</sub>	48.87	7.04	6.17
AAM+LPQ <sub>5×5</sub>	47.84	7.15	6.27
AAM+LPQ <sub>7×7</sub>	48.35	7.27	6.31
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	47.62	7.19	6.49
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	48.50	6.87	6.29
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	50.48	6.79	5.93

The confusion matrix of the age group classification using the LPQ<sub>7×7</sub> which achieved an overall accuracy of 50.70% is shown in Fig. 20. The highest accuracy of 80.30% was achieved for class 1 with 19.70% of misclassification where part of the test instances were classified as classes 2 and 3. On the other hand, classification errors above 60% were obtained for the classes 2 and 3. Most of the instances from class 2 were classified as class 1, while the instances from class 3 were misclassified as class 1, 2 and 4. This confusion occurred because the individual of classes 1, 2 and 3 have few wrinkles, no skin spots, etc. Class 4 properly classified 61.40% of its instances and misclassified some subjects as classes 1, 2 and 3 for the same reasons described previously. The confusion matrix clearly shows the tendency to achieve a

better accuracy in the age groups from 16 to 25 and 46 to 77 years old where the aging patterns are more prominent. In addition, the confusions also occurred due to the hard boundaries between the age classes which are compensated in the specific age estimation by the flexible overlapping of age ranges.

		Accuracy				
Predict class	1	285 20.8%	52 3.8%	18 1.3%	0 0.0%	80.3% 19.7%
	2	166 12.1%	122 8.9%	74 5.4%	9 0.7%	32.9% 67.1%
	3	68 5.0%	114 8.3%	157 11.5%	92 6.7%	36.4% 63.6%
	4	12 0.9%	22 1.6%	47 3.4%	129 9.4%	61.4% 38.6%
		1	2	3	4	
		Ground-truth class				

Figure 20. Confusion matrix for the age group classification using local phase features  $LPQ_{7 \times 7}$  (MORPH).

Fig. 21 shows the  $CS$  for the proposed approach at error levels from 0 to 15 years on the MORPH dataset. The fusion of global and local features (AAM+ $LPQ_{7 \times 7}$ ) leads to a lower performance than using only local features ( $LPQ_{7 \times 7}$ ), particularly in error levels between 5 and 10 years. This underperformance is due to the age range of MORPH be from 17 to 77 years and the AAM is more efficient to describe the facial shape which is more prominent in the age group from 0 to 18 years. Therefore, global features are not very relevant on MORPH dataset. The hierarchical age estimation with flexible overlapped regions using the local phase features  $LPQ_{7 \times 7}$  has about 57% of the age estimation with a  $MAE$  less than or equal to 5 years and 86% less than or equal to 10 years old.

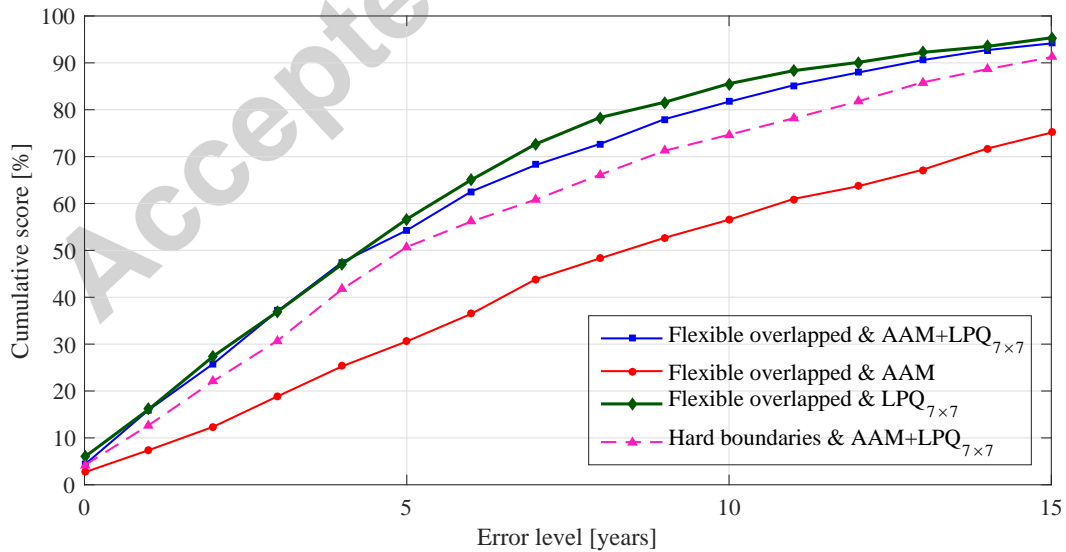


Figure 21. Cumulative scores of proposed and hierarchical age estimation with hard boundaries at error levels from 0 to 15 years on MORPH.



### G. Results on the Combination of FG-NET and MORPH Datasets

The FG-NET Aging and MORPH Album 2 were merged in order to have a better data distribution into specific age groups since the age group from 0 to 16 years is only present in FG-NET and from 17 to 77 years in the MORPH dataset. The new classes and data distribution are shown in Tab. VI. The best overall classification accuracy of 64.46% was achieved by using the combination of global and local features (AAM+LPQ<sub>7×7</sub>) as shown in Tab. VII. The proposed approach achieved a *MAE* of 5.20 years with the same features which shows the efficiency of the proposed approach when using the fusion of global and local features. When using the hierarchical approach with hard boundaries framework, a *MAE* of 5.55 years was achieved. Therefore, the flexible overlapped approach had a performance gain of 6.73%.

Table VI  
CLASSES AND DATA DISTRIBUTION FOR TRAINING AND TEST SETS (FG-NET+MORPH).

Class	Age group	Training set	Test set
1	0-13	385	113
2	14-21	788	95
3	22-39	2,288	1,139
4	40-77	1,397	271
Total	0-77	4,858	1,618

The *CS* for the proposed approach at error levels from 0 to 15 years on the merged FG-NET and MORPH datasets is shown in Fig. 22. The proposed approach with the combination of holistic (AAM) and local features (LPQ<sub>7×7</sub>) achieved the best result where more than 85% of the age estimation was made with an error level less than or equal to 10 years old.

Table VII  
RESULTS OF DIFFERENT AGE ESTIMATION APPROACHES IN DIFFERENT FEATURE SET (FG-NET+MORPH).

Feature set	Classification Accuracy [%]	Hierarchical Hard Boundaries [ <i>MAE</i> ]	Hierarchical Flexible Overlapped [ <i>MAE</i> ]
AAM	47.16	11.73	10.95
AAM+LPQ <sub>5×5</sub>	63.66	5.88	5.61
AAM+LPQ <sub>7×7</sub>	<b>64.46</b>	<b>5.55</b>	<b>5.20</b>
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	63.16	5.79	5.73
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	61.93	6.14	5.66
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	63.16	5.69	5.43

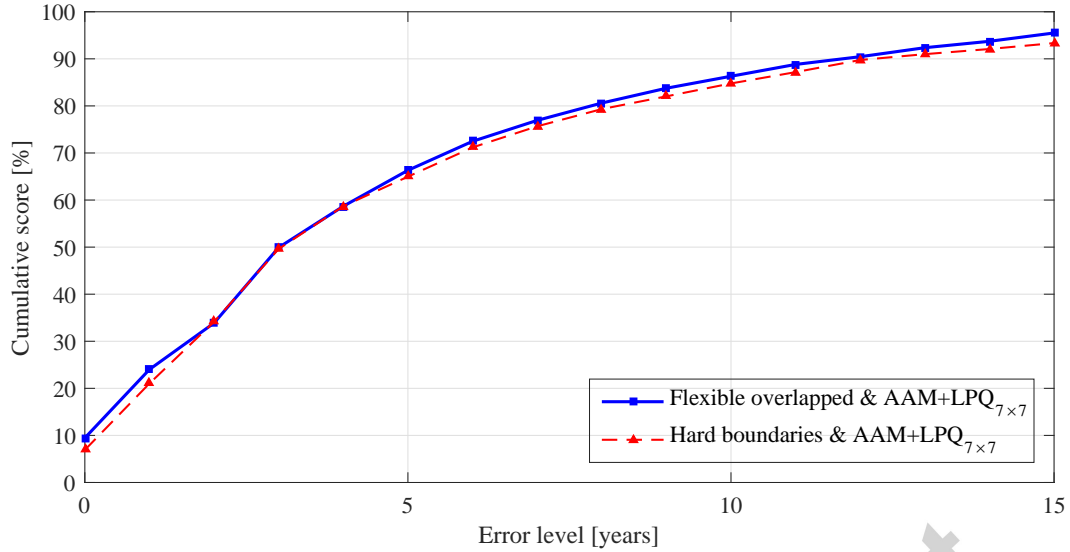


Figure 22. Cumulative scores of proposed and hierarchical age estimation with hard boundaries at error levels from 0 to 15 years on the combination of the FG-NET and MORPH datasets.

#### H. Human Age Estimation

The performance of the proposed automatic age estimation approach was compared with age estimation by made humans using 250 facial images. Twenty five people of different age and gender were selected from the FG-NET Aging dataset to estimate the age of ten facial images. The 250 images are the same used to test the proposed approach on FG-NET. Tab. VIII shows the performance comparison between the human age estimation and the proposed approach on FG-NET Aging. The perceived ages of the facial images were divided into age groups according to the hierarchical age estimation with flexible overlapped age classes in order to have a comparison in the same conditions. It is notable that the human age estimation had a lower performance than the computational approach. A slightly better performance is observed in the classes 0-28 and 14-22 years old. Only one image was used in the class 40-69 years where the human made an error of 9 years and the approach properly estimated the age.

Table VIII  
PERFORMANCE COMPARISON BETWEEN THE HUMAN AGE ESTIMATION AND THE PROPOSED APPROACH ON FG-NET AGING.

Age group	Human	Hierarchical
Flexible overlapped	Age Estimation [ <i>MAE</i> ]	Flexible Overlapped [ <i>MAE</i> ]
0-28	3.10	4.08
14-22	6.22	6.33
22-60	7.77	7.52
40-69	9.00	0.09
<b>Overall mean</b>	<b>6.52</b>	<b>4.50</b>

Some results of the human age estimation and the machine age estimation are shown in Fig. 23. The

actual, perceived and estimated ages are compared. In general, the estimated ages by the proposed approach are closer to the actual age. Some images had higher estimation errors. For instance, a facial image of a 29-year-old female was estimated to be seventeen years old by the machine. However, it is noticed that there are some occlusions caused by make-up which hide important features for a more accurate estimation. Another example is the woman of 47 years which the system underestimated by thirteen years, perhaps due to head position, expression, make-up, etc. A counter-example includes the human overestimation of 32 years of a facial image of an 18-year-old male, while the machine properly estimated the age. Human over or underestimation can be explained by the influence of appearance. In this example, the moustache made the man look older. In the other hand, the computer performs a shape and microtextural skin analysis to have a more accurate age estimation and is not influenced by appearance features that influences humans.

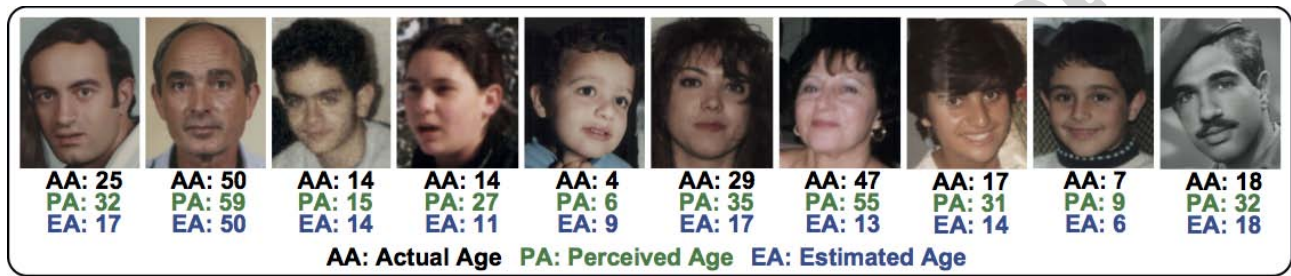


Figure 23. Some results of the human age estimation (perceived) and the machine age estimation made by the proposed approach (estimated).

### I. Comparisons to Other Works

The  $MAE$  for the proposed approach and other methods reported on the FG-NET dataset are given in Tab. IX. The proposed approach outperforms several state-of-the-art approaches [5]–[7], [14], [18], [19], [29], [31], [32], [42], [44]. However, Luu *et al.* [33] has achieved slightly better  $MAE$  of 4.37. They used only a holistic facial representation using AAM, which might not encode facial wrinkles well for older people. On the other hand the proposed approach is robust to facial wrinkles since the LPQ features capture the texture patterns at local regions. Our approach compares favorably to the literature regardless of differences in the experimental protocols, such as the dataset partition and images used for training and test. For instance, Luu *et al.* [33] have partitioned the dataset into 80% for training and the remaining to test. Through the LOPO protocol a  $MAE$  of 4.78 years was achieved for the proposed approach. It is slightly worse than using 75% of the data to train and 25% to test but it assures that a person is not in the training and test set at the same time, thus eliminating data dependency.

Tab. X shows a comparison of the proposed approach to the previous works reported on the MORPH Album 2 dataset. Han *et al.* [23] and Geng *et al.* [17] have achieved better  $MAEs$  of 5.10 and 4.37 respectively. However, the  $MAEs$  cannot be compared directly due to the different amount of data used.

Table IX

COMPARISON OF THE PROPOSED APPROACH TO THE PREVIOUS WORKS REPORTED ON THE FG-NET AGING DATASET.

Methods of age estimation	<i>MAE</i>	Methods of age estimation	<i>MAE</i>
WAS [29]	8.06	Duong <i>et al.</i> [14]	4.74
AGES [18]	6.77	Choi <i>et al.</i> [9]	4.65
RUN [42]	5.78	PLO [31]	4.82
Ranking [41]	5.33	CA-SVR [7]	4.67
LARR [19]	5.07	HC-SVR [32]	5.28
SVR [19]	5.66	ST+CSOHR [5]	4.70
Luu <i>et al.</i> [33]	4.37	<b>Human age estimation</b>	<b>6.52</b>
MTWGP [44]	4.83	<b>Flexible overlapped &amp; AAM+LPQ<sub>7×7</sub></b>	<b>4.50</b>
OHRank [6]	4.85	<b>Flexible overlapped &amp; AAM+LPQ<sub>7×7</sub> (LOPO)</b>	<b>4.78</b>

Both of them have used the complete commercial version of the MORPH dataset with more than 55,000 images to develop and evaluate their methods. Moreover, both of them used holistic biologically inspired features (BIF) as local features implemented with the real components of Gabor wavelets. Guo *et al.* [21] investigated the BIF for age estimation from faces using a regression-based approach. Although the results are very promising when compared to the literature, the main drawback is that it requires a large amount of time when extracting the aging features. The proposed approach has achieved competitive results (5.86 years of *MAE*) using only 7% of the amount of data and a faster feature extractor based on the LPQ when compared to these works. In a recent work, Chang and Chen [5] achieved a *MAE* of 3.74 by proposing an age ranking approach algorithm and using scattering transform as facial descriptor. Despite their impressive result, they have not reported any result in a dataset with larger age span (*e.g.*, 0-77 years) as we have did by merging the FG-NET and the MORPH datasets to reduce the overfitting to a single dataset.

Table X

COMPARISON OF THE PROPOSED APPROACH TO THE PREVIOUS WORKS REPORTED ON THE MORPH ALBUM 2 DATASET.

Methods of age estimation	<i>MAE</i>	Methods of age estimation	<i>MAE</i>
WAS [29]	9.32	CA-SVR [7]	5.88
AGES [18]	8.83	BFGS-ALDL [17]	4.34
MTWGP [44]	6.28	ST+CSOHR [5]	3.74
Rank [4]	6.49	<b>Flexible overlapped &amp; LPQ<sub>7×7</sub></b>	<b>5.86</b>
Han <i>et al.</i> [23]	5.10		

Unfortunately it is not possible to compare the performance of the proposed approach with other works on the merged FG-NET and MORPH datasets because they are reported for the first time in this paper.

## V. CONCLUSION

In this paper, a novel approach that combines global features extracted by AAM and local features extracted by LPQ to age estimation has been proposed. The proposed hierarchical approach uses a multiclass SVM classifier with hard boundaries to classify the individuals into four age classes at a first stage and SVRs with flexible overlapping age classes at a second stage to specific age estimation. As a result, the classification errors produced at the first stage due to the hard boundaries are somewhat compensated by the flexible overlapped boundaries of the second stage.

Extensive experiments employing global features extracted by AAM and local features extracted by Gabor wavelets, LBP and LPQ as well as several features combinations with different settings have been conducted in this work. These features or their combinations have not been deeply exploited in the FG-NET Aging and MORPH Album 2 datasets by previous works. Also, in order to have a better generalization and age distribution and coverage, a combination of the FG-NET and MORPH datasets was proposed in this work. This also contributes to reduce the overfitting of the proposed approach to a single dataset.

The LPQ features have showed that they are very robust not only to blur but also to other challenges such as lighting and expression variations present in the FG-NET and MORPH datasets. Despite its simplicity and robustness, LPQ was shown to be highly discriminative, producing very good results for age estimation either using it as local features only or fused with global features. This descriptor has also outperformed the other widely used feature extraction methods such as LBP and Gabor wavelets.

In addition, the performance of the proposed age estimation approach was compared to human age estimation. Through this research, it was possible to conclude that the machine outperforms human in the age estimation task. Some factors affected the performance of the proposed approach such as facial wrinkles which are generally removed by the photographer or through make-up. Similarly, some images from the datasets were taken using very bright lighting sources, occlusions like glasses, beards, etc. All these factors can have a negative impact on age estimation.

One of the key issues in age estimation is to cope with the lack of data. Indeed, some ages are either absent or poorly distributed within the FG-NET and MORPH datasets, making it difficult to establish a universal age estimator. Future work will address new techniques such as deep learning and fine-grained classification to enhance the age estimation performance.

## REFERENCES

- [1] T. Ahonen, E. Rahtu, V. V. Ojansivu, and J. Heikkilä, "Recognition of blurred faces using local phase quantization," in *19th International Conference on Pattern Recognition*, Tampa, USA, 2008, pp. 1–4.

- [2] M. Bereta, P. Karczmarek, W. Pedrycz, and M. Reformat, "Local descriptors in application to the aging problem in face recognition," *Pattern Recognition*, vol. 46, no. 10, pp. 2634–2646, 2013.
- [3] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 27, pp. 1–27, 2011.
- [4] K. Chang, C. Chen, and Y. Hung, "A ranking approach for human ages estimation based on face images," in *20th International Conference on Pattern Recognition*, 2010, pp. 3396–3399.
- [5] K.-Y. Chang, , and C.-S. Chen, "A learning framework for age rank estimation based on face images with scattering transform," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 785–798, 2015.
- [6] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, 2011, pp. 585–592.
- [7] K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, USA, 2013, pp. 2467–2474.
- [8] W. Chen, W. Qian, G. Wu, W. Chen, B. Xian, X. Chen, Y. Cao, C. D. Green, F. Zhao, K. Tang, and J.-D. J. Han, "Three-dimensional human facial morphologies as robust aging markers," *Cell Research*, vol. 25, pp. 574–587, May 2015.
- [9] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, no. 6, pp. 1262–1281, 2011.
- [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, 2001.
- [11] Y. M. G. Costa, L. E. S. Oliveira, A. L. Koerich, F. Gouyon, and J. G. Martins, "Music genre classification using LBP textural features," *Signal Processing*, vol. 92, no. 11, pp. 2723–2737, April 2012.
- [12] M. M. Dehshibi and A. Bastanfard, "A new algorithm for age recognition from facial images," *IEEE Transactions on Signal Processing*, vol. 90, no. 8, pp. 2431–2444, 2010.
- [13] H. Dibeklioglu, F. Alnajar, A. A. Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1928–1943, 2015.
- [14] C. N. Duong, K. G. Quach, K. Luu, H. B. Le, and K. Ricanek, "Fine tuning age estimation with global and local facial features," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, 2011, pp. 2032–2035.
- [15] E. Eiding, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.
- [16] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.
- [17] X. Geng, Q. Wang, and Y. Xia, "Facial age estimation by adaptive label distribution learning," in *IEEE International Conference on Pattern Recognition*, 2014, pp. 4465–4470.
- [18] X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.
- [19] 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, 2008.
- [20] G. Guo, Y. Fu, T. Huang, and C. Dyer, "Locally adjusted robust regression for human age estimation," in *IEEE Workshop on Applications of Computer Vision*, Cooper Mountain, USA, 2008, pp. 721–724.
- [21] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, 2009, pp. 112–119.
- [22] G. D. Guo, G. Mu, Y. Fu, C. R. Dyer, and T. S. Huang, "A study on automatic age estimation using a large database," in *IEEE International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 1986–1991.



- [23] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *IEEE International Conference on Biometrics*, Madrid, Spain, 2013, pp. 1–8.
- [24] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, "Age and gender estimation from facial image processing," in *41st SICE Annual Conference*, vol. 1, 2002, pp. 13–18.
- [25] INRIA, "The FG-NET aging database," [http://www.aaronstonemd.com/Facial\\_Aging\\_Rejuvenation.shtml](http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtml), 2010.
- [26] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer Verlag, 2002.
- [27] Y. H. Kwon and N. D. V. Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.
- [28] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.
- [29] 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, 2002.
- [30] G. Lemperle, R. E. Holmes, S. R. Cohen, and S. M. Lemperle, "A classification of facial wrinkles," *Plastic and Reconstructive Surgery*, vol. 108, no. 6, pp. 1735–1750, 2001.
- [31] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, 2012, pp. 2570–2577.
- [32] J. Liu, Y. Ma, L. Duan, F. Wang, and Y. Liu, "Hybrid constraint SVR for facial age estimation," *Signal Processing*, vol. 94, pp. 576–582, 2014.
- [33] K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in *IEEE 3rd International Conference on Biometrics*, Washington DC, USA, 2009, pp. 1–5.
- [34] A. Marini, A. J. Turatti, A. S. Britto Jr., and A. L. Koerich, "Visual and acoustic identification of bird species," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2309–2313, April 2015.
- [35] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.
- [36] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *IEEE International Conference on Automatic Face and Gesture*, 2006, pp. 341–345.
- [37] A. Ross and R. Govindarajan, "Feature level fusion in biometric systems," in *Proc. of the Biometric Cons. Conf.*, 2004.
- [38] A. Stone, "The Aging Process of the Face & Techniques of Rejuvenation," [http://www.aaronstonemd.com/Facial\\_Aging\\_Rejuvenation.shtml](http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtml), 2010.
- [39] A. Vedrana and A. Renee, "Active shape and appearance models," IT University of Copenhagen, Tech. Rep., 2005.
- [40] K. B. Vinay and B. S. Shreyas, "Face recognition using Gabor wavelets," in *14th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, 2006, pp. 593–597.
- [41] S. Yan, H. Wang, T. Huang, Q. Yang, and X. T. Xiaoou, "Ranking with uncertain labels," in *IEEE International Conference on Multimedia and Expo*, 2007, pp. 96–99.
- [42] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning autostructured regressor from uncertain nonnegative labels," *Proceedings of the IEEE Conference on Computer Vision*, pp. 1–8, 2007.
- [43] T. H. H. Zavaschi, A. S. Britto Jr., L. E. S. Oliveira, and A. L. Koerich, "Fusion of feature sets and classifiers for facial expression recognition," *Expert Systems with Applications*, vol. 40, no. 2, pp. 646–655, 2013.
- [44] Y. Zhang and D.-Y. Yeung, "Multi-task warped gaussian process for personalized age estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, 2010, pp. 2622–2629.