



# Discriminative aging subspace learning for age estimation

Manisha Sawant<sup>1</sup> · Kishor M. Bhurchandi<sup>1</sup>

Accepted: 22 June 2022 / Published online: 14 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Human age estimation from facial images has become an active research topic in computer vision field because of various real-world applications. Temporal property of facial aging display sequential patterns that lie on the low-dimensional aging manifold. In this paper, we propose hidden factor analysis (HFA) model-based discriminative manifold learning method for age estimation. The hidden factor analysis decomposes facial features into independent age factor and identity factor. Various age invariant face recognition systems in the literature utilize identity factor for face recognition; however, the age factor remains unutilized. The age component of the hidden factor analysis model depends on the subject's age. Thus it carries significant age-related information. In this paper, we demonstrate that such aging patterns can be effectively extracted from the HFA-based discriminant subspace learning algorithm. Next, we have applied multiple regression methods on low-dimensional aging features learned from the HFA model. Effect of reduced dimensionality on the accuracy has been evaluated by extensive experiments and compared with the state-of-the-art methods. Effectiveness and robustness in terms of MAE and CS of the proposed framework are demonstrated using experimental analysis on a large-scale aging database MORPH II. The accuracy of our method is found superior to the current state-of-the-art methods.

**Keywords** Age estimation · Hidden factor analysis (HFA) · Aging manifold · Regression

## 1 Introduction

Among various facial attributes such as identity, gender, age and ethnicity, age-related facial appearance variation has become an interesting research topic because of its emerging applications. Age estimation has potential real world applications in multimedia communication, human–computer interaction (HCI), law enforcement and security. In these applications, the estimated age value is further used to make some inferences such as age-based access control (e.g., tobacco vending machines). Human aging is a complex process that affects the shape and texture of its face. Facial shape change is a prominent factor during the developmental years. Along with the structural development, skin-related deformations are also observed due to aging. Skin-related aging

effects are dependent on both intrinsic as well as extrinsic factors such as environment, lifestyle, stress, diseases and exposure to the sun. Different aging patterns are observed even in the people of same ethnic origins due to diversity in climatic conditions and lifestyle. Hence, it is difficult to precisely predict a person's age from the facial appearances, even for humans.

Key components in automatic age estimation methods are facial feature representation and age prediction. Several techniques have been published in order to deal with these two stages. Accuracy and robustness of an age estimation system rely on quality of the extracted facial aging features. Hence, various local feature descriptors have been used for extracting appearance features from an image. In the literature, we find various local feature descriptors that are used for extracting appearance features from an image. Guo et al. Guo et al. (2009) proposed bio-inspired feature (BIF) to extract age-related facial features for age estimation. BIF and its variants have been used in Guo and Mus (2011, 2013); Han et al. (2015) for age estimation. Various existing histogram-based local features have also been widely used in age estimation (Geng et al. 2013; Pontes et al. 2016; Weng et al. 2013; Zhu

✉ Manisha Sawant  
mpparlewar@gmail.com

Kishor M. Bhurchandi  
bhurchandikm@ece.vnit.ac.in

<sup>1</sup> Department of Electronics and Communication Engineering,  
Visvesvaraya National Institute of Technology, Nagpur, India

et al. 2014). Pontes et al. combined local and global features in Pontes et al. (2016), where active appearance model (AAM) (Cootes et al. 2001) is used as a global feature and local binary pattern (LBP) (Ahonen et al. 2006), gabor and local phase quantization (LPQ) (Ahonen et al. 2008) are used as local features. Histogram of oriented gradients (HOG) is used in Huerta et al. (2015); Zhu et al. (2014) for age estimation. Scale invariant feature transform (SIFT) (Lowe 1999) and LBP are used as feature vectors in Xiao et al. (2009) for global age estimation approach. Apart from existing local feature descriptors, a few local feature descriptors specially designed for age estimation include directional age-primitive pattern (DAPP) (Iqbal et al. 2017) and local direction and moment pattern (LDMP) (Sawant et al. 2019).

Age estimation has been considered either as a classification problem or a regression problem. In classification, age label is considered as a class, whereas, in regression it is treated as a continuous value. While classification tries to learn from age variations in the training data, regression utilizes the label information. Recently proposed learning algorithms for age estimation include support vector machine (SVM)/regression (SVR) (Han et al. 2015), partial least squares (PLS) (Guo and Mus 2013) along with its kernelized version (KLPS) (Guo and Mus 2011) and canonical correlation analysis (CCA) (Guo and Mus 2013). Ordinal hyperplane ranking algorithm (OHRank) (Chang et al. 2011) utilizes the relative order information among age labels. Apart from abovementioned regression methods, multi-task warped Gaussian process (WGP) regression has been proposed (Zhang and Yeung 2010) for personalized age estimation. Computationally efficient extension of WGP has been proposed in Zhu et al. (2014) using orthogonal Gaussian process (OGP). To further improve computation complexity and age estimation performance, a two-level hierarchical Gaussian process regression was proposed in Sawant and Bhurchandi (2019).

Although there are several feature descriptors and learning techniques that are available in the literature to address age estimation problem, it is yet largely unsolved and remained a challenging research problem. In general, there are various factors that affect the overall performance of automatic age estimation, such as preprocessing and face alignment, database, number of images used for training and testing, feature extraction method and the type of age regression technique. In order to develop a robust and accurate age estimation system, it is required to focus on designing of low-dimensional discriminative aging features as well as robust regression algorithm. In this paper, we propose low-dimensional discriminative aging manifold to extract aging features.

Hidden factor analysis is a probabilistic model which separates identity and age components from the facial features. The HFA and its modified versions are widely used in many

state-of-the-art age invariant face recognition systems (Gong et al. 2013; Li et al. 2017; Wen et al. 2016). All these methods use identity factor and ignore the information present in the second hidden factor *i.e.*, age factor. The age component contains important information about the subject's age. Therefore, the age component of the HFA model is hypothesized to accelerate the age estimation performance. Also the age component on the feature space is low-dimensional. Therefore, to estimate age from the face images, we propose a novel manifold learning method derived from the HFA model. We learn the low-dimensional aging features from the learned aging subspace. The aging features learned from the aging subspace carries age discriminative information. Thus the regression algorithm can fit the data better and is expected to improve the age estimation performance.

In this paper, we propose a novel scheme for aging feature extraction from discriminative aging manifold for automatic age estimation. The objective of this method is to prove that the manifold-based method is more efficient for feature detection and subsequent age estimation from facial images compared to the state of the art. The achieved consistent minimum MAE and CF scores from the robust and explorative experiments prove the same. The basic idea is to learn a low-dimensional embedding of the aging manifold using the HFA-based subspace learning method. Further we use multiple regression techniques for subsequent age estimation. The efficiency of the proposed method has been demonstrated with extensive experiments on multiple regression algorithms using a large-scale age database. The main contributions of this paper include: 1) our method is the first to present the HFA for age estimation rather than AIRF. 2) We empirically show through the dimensionality analysis that our method provides a robust discriminative low-dimensional aging subspace. 3) The proposed method provides consistent results on various regression methods in terms of MAE and CS compared to the state-of-the-art methods.

This paper is organized as follows. We provide detailed-related work in Section 2. Proposed age estimation framework and manifold learning is presented in Section 3. The experimental setup and results are discussed in Section 4. We conclude the paper in Section 5.

## 2 Related work

Manifold-based feature extraction has gained a considerable interest over a past few years. The basic idea of manifold features extraction is that high-dimensional data lie on or close to a smooth low-dimensional manifold. In age manifold, a common aging pattern is learned from images of many individuals with different ages. Several face images are adopted

to represent an age. Each subject may be represented by one image or several images at different ages. These images make a set referred to as a manifold which makes up points in a high-dimensional vector space. Several facial images are utilized to learn a nonlinear low-dimensional aging pattern. It is assumed that the faces with close ages are located closely on the manifold, as the aging face images are distributed on to an intrinsic low-dimensional manifold. Individuals may have as low as one image at each age in a database. Once the low-dimensional aging manifold is learnt, a regression is applied on the embedded subspace to predict the exact age. Some representative works on manifold learning include locally linear embedding (LLE) (Roweis and Saul 2000), ISOMAP (Tenenbaum et al. 2000) and Laplacian eigenmaps (Belkin and Niyogi 2003). However, it has been observed that these methods can only work well on the training samples and not on test data.

In recent years, many efforts have been devoted to identify discriminant aging subspaces for age estimation. A subspace called AGing pattErn Subspace (AGES) is proposed by Geng et al. Geng et al. (2013, 2007) to learn personalized aging process from multiple faces of an individual. Discriminative subspace learning based on aging manifold and locally adjusted robust regressor (LARR) was proposed by Guo et al. Guo et al. (2008). In LARR, the ages are first predicted by a regression function. Then the predicted age value is fine tuned locally to match with the true value within a specific bound. This method shows that the facial images can be represented on a manifold. The age manifold method does not require images at different ages of the same individual, like the AGES method. But in order to learn the low-dimensional manifold, it requires images of several subjects at many ages. Later, kernel partial least square regression (KPLS) (Guo and Mus 2011) has been used that simultaneously reduces the dimension and learns the aging function.

Given an aging feature representation, the next step is to determine the individual's age from the aging features. the age estimation problem can be formulated as a regression problem, where the objective is to learn a function that best fits the mapping from the feature space to the age-value space with appropriate regularization. Typical nonlinear regression approaches such as quadratic regression, Gaussian process and support vector regression have been proposed in the literature to solve the age estimation problem.

In the early work on age estimation, Lanitis et al Lanitis et al. (2002) investigated three regression functions, linear, quadratic and cubic to estimate age from 50 raw AAM parameters. A genetic algorithm is used to learn optimal model parameters from training face images of different ages. Among the three regression methods, quadratic aging function achieved better results. Y. Fu et al. Fu and Huang (2008) also used linear, quadratic and cubic regression functions as learning algorithms, but they performed regression on

manifold features. Similar to Lanitis, Fu et al. also reported superior performance on the quadratic regression function, pointing out that cubic functions lead to over-fitting while linear functions lead to under-fitting.

Guo et al. Guo et al. (2009) proposed biologically inspired features (BIF) for age estimation. They used BIF as aging features and support vector machine for regression (SVR) (Vapnik 1999). Chao et al. Chao et al. (2013) modeled ordinal relationship between age labels and facial features by merging distance metric learning and dimensionality reduction. They presented age-oriented local regression to capture the complex aging process for age estimation. The age-oriented local regression is designed using nearest neighbor and SVR for age estimation.

Zhang et al. Zhang and Yeung (2010) formulated personalized age estimation as a multi-task warped Gaussian process (WGP) estimation problem. They developed a multi-task extension of WGP to solve the problem. Since different individuals have different aging processes, personalization is beneficial for age estimation. However, person-specific approaches rely heavily on the availability of training images across different ages for a single subject, which is usually difficult to have in practical age estimation scenario. To address this limitation, computationally efficient version of warped Gaussian process regression is proposed in Zhu et al. (2014). Gong et al. proposed orthogonal Gaussian process regression to handle global age estimation. For the WGP method, the major computational cost in each iteration step involves an inversion of matrix  $\mathbf{K}$  that is of size  $N \times N$ , where  $N$  is the number of training samples. Hence, when  $N$  is large, the computational cost would be extremely high. To alleviate this problem, Gong et al. developed an OGP for age estimation, which can optimize the model parameters in a transformed space with more efficient computations.

The performance of these methods depends on how well the manifold is learnt from the facial images to precisely represent the age. Thus identifying a space where aging manifold truly lies is still an open research problem. The age manifold also requires large database in order to satisfactorily learn the aging information. Existing manifold-based methods extract manifold features from the gray intensity or image space. However, the image space is inefficient to model the large age variations. The texture features such as LBP, HOG, SIFT and LPQ are used to capture the textural variations due to aging. But the manifold of such feature space has not been explored for age estimation so far. In this paper, we propose a subspace learning scheme for aging feature extraction for automatic age estimation. The basic idea is to learn an aging manifold embedding using the HFA-based aging subspace learning method.

### 3 Proposed work

The traditional age estimation approaches basically use generalized local feature descriptors used in facial analysis (Geng et al. 2013; Pontes et al. 2016; Weng et al. 2013; Zhu et al. 2014), or design the age-related characteristics by extracting edge and directional information from the training samples (Guo et al. 2009; Iqbal et al. 2017; Sawant et al. 2019), without considering age group specific aging patterns. A few methods design person specific age estimation schemes (Geng et al. 2007; Zhang and Yeung 2010), which assume that every person ages in a different way. Such assumption makes accuracy of the system, identity dependent and poses restriction on the age estimation system. These facts and an objective of discriminative aging feature motivate us to design an approach to accurately learn low-dimensional identity independent aging features. The proposed method, learns low-dimensional aging features by separating person-specific component. It necessarily separates identity information and the age-related component that conveys the age from the local feature descriptor. In the proposed method, only the latter component plays an important role.

Facial aging-related research is broadly classified into three categories: age estimation, age progression and cross-age face recognition. Age estimation and progression mainly focuses on aging information that changes due to aging, whereas, in the cross-age face recognition major emphasis is on identity information that does not change with aging. Recent works on face recognition (Gong et al. 2013, 2015; Li et al. 2017; Prince and Elder 2007; Prince et al. 2008; Wang et al. 2006; Wen et al. 2016; Xu et al. 2017) adopt the hidden factor analysis (Gong et al. 2013) and the probabilistic linear discriminant analysis (PLDA) (Prince and Elder 2007) to separate within subject and between subject information from facial images. These methods build a probabilistic linear model and iteratively learn optimal model parameters using expectation–maximization (EM) (Moon 1996) algorithm. Such decomposition of facial features has also been used to separate age and identity components to achieve cross-age face recognition (Gong et al. 2013, 2015; Li et al. 2017; Wen et al. 2016; Xu et al. 2017) and age progression (Ling et al. 2007; Wang et al. 2006; Xu et al. 2017). Aging face recognition utilizes the identity factor, whereas both age and identity factors are used for rendering age progressed images. It is clear from the performance of these methods that identity factor provides identity information that is practically stable across age progression. However, the information conveyed by the aging component remains unutilized; this work utilizes the same.

In this section, the framework of the proposed manifold learning is introduced. Given a set of local features of face images from diverse age groups, for manifold learning, a

hidden factor analysis method, is introduced in this work. Training set images are used to learn the bases of the age and identity subspace. Once the parameters are well trained, by projecting the training and test faces onto the aging subspace, the age-discriminative features can be achieved for every face image. It is well known that histogram-based local feature representation often leads to high dimensionality which may result into the failure of regressors. Therefore, an appropriate manifold learning algorithm is required to reduce redundancies in the feature dimension. In this paper, we propose a discriminative aging subspace learning procedure using hidden factor analysis.

#### 3.1 Aging subspace learning based on HFA

Hidden factor analysis (Gong et al. 2013) is a linear probabilistic model that separates identity specific variations from the age for age invariant face recognition. The probabilistic linear factor analysis model presented in Gong et al. (2013) is as follows:

$$f = \mu + \mathbf{Q}x + \mathbf{P}y + \varepsilon \quad (1)$$

The major notations used are defined as follows:

- Observable feature vector  $f$  is denoted by a  $D \times 1$  vector, where  $D$  is dimension of feature vector.
- Mean of the face features  $\mu$  is denoted by a  $D \times 1$  vector.
- Hidden identity factor  $x$  is a  $r \times 1$  random vector. A prior distribution of identity factor is defined as,  $p(x) \sim \mathcal{N}(0, I_r)$ , where  $\mathcal{N}(\mu, K)$  denotes multivariate Gaussian distribution with mean  $\mu$  and covariance matrix  $K$ .
- Hidden age factor  $y$  is a  $s \times 1$  random vector. A prior distribution of age factor is defined as,  $p(y) \sim \mathcal{N}(0, I_s)$ .
- Cross-identity variations are captured by a  $D \times r$  identity factor loading matrix  $\mathbf{Q}$ .
- Cross-age variations are captured by a  $D \times s$  age factor loading matrix  $\mathbf{P}$ .
- Observation noise denoted by  $\varepsilon$  follows an isotropic Gaussian distribution *i.e.*,  $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ .

The HFA model in (1) decomposes the observed feature into identity, age and noise components. EM algorithm is used to estimate the model parameters  $\{\mu, \mathbf{Q}, \mathbf{P}, \sigma^2\}$ . Such a decomposition model has been effectively used for age invariant face recognition (Gong et al. 2013, 2015; Li et al. 2017; Wen et al. 2016; Xu et al. 2017).

Local feature descriptor of any subject contains information about identity, age, pose, ethnicity and other related facial attributes. We decompose facial features  $f_i^k$  into linear combination of identity, age and noise component using HFA model. All other facial attributes are considered in the noise



component. Suppose that  $f_i^k$  represents the feature vector of the individual  $i$  at age-group  $k$ , then the discriminative aging subspace is learned from (2).

$$f_i^k = \mu + \mathbf{Q}x_i + \mathbf{P}y_k + \varepsilon \tag{2}$$

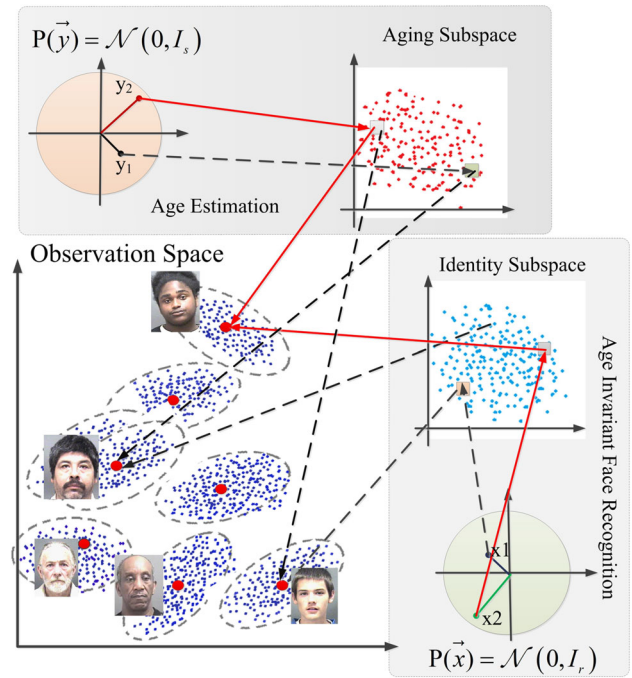
Since the observed feature space does not provide efficient representation for age estimation, we offer a robust representation with fine properties. The second component of HFA model in (2) depends only on the 's identity, whereas the third component depends only on the age of the subject. Matrix  $\mathbf{P}$  in (2) is age factor loading matrix that captures the cross-age variations. The columns of the matrix  $\mathbf{P}$  are the bases for cross-age variations; therefore we term matrix  $\mathbf{P}$  as aging subspace. The age factor  $y_k$  can be viewed as the position of the  $f_i^k$  i.e., observation variable in the aging subspace. The relationship between the observation space  $f_i^k$  and the identity and aging subspace is indicated in Fig. 1. The observation space represents feature space of the local feature descriptor, whereas the aging and the identity subspace are learned from the HFA model. In (1)  $p(x) \sim \mathcal{N}(0, I_r)$  and  $p(y) \sim \mathcal{N}(0, I_s)$  hence their representation in Fig. 1 follows Gaussian distribution. In Fig. 1, every subject's feature vectors are clustered. In Fig. 1, we have shown two subjects of similar age but different ethnicity are located at some distance in the observations space. These two subjects are discriminated in the identity space. However, these subjects with different identity but similar age are located close in the aging subspace. Hence, the proposed approach is not a person specific. And when we project the high-dimensional local features on the aging subspace, we get a low-dimensional age discriminative aging feature vector. These low-dimensional feature vectors are further used to learn regression.

The learning objective is to estimate the model parameter  $\theta = \{\mu, \mathbf{Q}, \mathbf{P}, \sigma^2\}$  using training data. We learn the set of parameters by maximizing joint likelihood  $p_\theta(f_i^k, x_i, y_k)$  of the observed feature vector and associated identity and age factors. Both the identity and age factors are not observed directly. It is possible to only infer them through their posterior distributions for fixed set of model parameters. EM is used to solve the optimization problem. EM algorithm is a maximum likelihood method that iteratively updates set of parameters. The objective function for optimization is,

$$L_c = \sum_{i,k} \ln(p_\theta(f_i^k, x_i, y_k)). \tag{3}$$

Given the initial estimate of parameters  $\theta_0$ , in expectation step (i.e., E-step), first two moments of the age factor are computed as

$$\mu = \frac{1}{N} \sum_{i,k} f_i^k \tag{4}$$



**Fig. 1** Illustration of the proposed aging subspace learning scheme. Latent identity and age variables in (1) have a prior distribution  $p(x) \sim \mathcal{N}(0, I_r)$  and  $p(y) \sim \mathcal{N}(0, I_s)$ . Circles in the observation space represent observed data for a person at different age. Identity factor is used for age invariant face recognition (Gong et al. 2013, 2015; Li et al. 2017). We propose aging subspace for age estimation

$$E(y_k) = \frac{\mathbf{P}^T \sum_{i=1}^{N_k} (f_i^k - \mu)}{N_k} \tag{5}$$

$$E(y_k y_k^T) = \frac{I - \mathbf{P}^T \sum_{i=1}^{N_k} \mathbf{P}}{N_k} + E(y_k) E(y_k)^T \tag{6}$$

where  $\sum = \mathbf{Q}\mathbf{Q}^T + \mathbf{P}\mathbf{P}^T + \sigma^2 I$ ,  $N$  is total training images,  $N_k$  is number of images at  $k^{th}$  age group and  $E[\cdot]$  is expectation. The estimate of the first two moments of the identity factor is calculated in a similar way. In maximization step (i.e., M-step), the model parameters are computed using the estimate of age and identity factors. The complete log likelihood in (3) is rewritten as,

$$L_c = \sum_{i,k} \ln(p(f_i^k | x_i, y_k)) + \ln(p(x_i, y_k))$$

where  $p(f_i^k | x_i, y_k) = \mathcal{N}(\mu + \mathbf{Q}x_i + \mathbf{P}y_k, \sigma^2)$  and  $p(x_i, y_k) = \mathcal{N}(0, I)$ .

In M-step, first the derivative of the objective function is computed and the optimal parameters  $\phi = \{\mathbf{Q}, \mathbf{P}, \sigma^2\}$  are obtained at  $E\left(\frac{\partial L_c}{\partial \phi}\right) = 0$ . To calculate the optimal parameters, we solve (7), (8) and (9).

$$\mathbf{Q} \sum_{i,k} E(x_i x_i^T) = \sum_{i,k} (f_i^k - \mu) E(x_i) - \mathbf{P} E(y_k x_i^T) \quad (7)$$

$$\mathbf{P} \sum_{i,k} E(y_k y_k^T) = \sum_{i,k} (f_i^k - \mu) E(y_k) - \mathbf{Q} E(x_i y_k^T) \quad (8)$$

The optimal solution for the noise variance is represented as

$$\sigma^2 = \frac{1}{DN} \sum_{i,k} (f_i^k - \mu)^T (f_i^k - \mu - \mathbf{Q} E(x_i) - \mathbf{P} E(y_k)) \quad (9)$$

where  $D$  is the dimension of the mean feature and  $N$  denotes total number of training images. After learning parameters of the HFA model and corresponding low-dimensional representation of feature space, we can use a regression technique to estimate the age of a query image.

### 3.2 Regression for age estimation

Let the local facial feature space  $\mathcal{F}$  is represented as  $\mathcal{F} = \{f_i : f_i \in \mathbb{R}^D\}_{i=1}^N$  where  $N$  is number of face images and  $D$  is dimension of the data. The ground truth age labels  $a_i$  are represented as  $a = \{a_i : a_i \in \mathbb{N}\}_{i=1}^N$ . Our objective is to learn a low-dimensional discriminative manifold  $\mathcal{G}$  embedded in  $\mathcal{F}$  and subsequently a low-dimensional aging feature  $\{t_i : t_i \in \mathbb{R}^d\}_{i=1}^N$  with  $d \ll D$ . Once the aging subspace  $\mathbf{P}$  is learned from the HFA training, our next task is to find low-dimensional age discriminative features. The low-dimensional aging features are obtained by projecting local feature onto a  $D \times d$  projection matrix *i.e.*,  $T = \mathbf{P}^T F$  where  $F = [f_1, f_2, \dots, f_N] \in \mathbb{R}^{D \times N}$  and  $\mathbf{P} = [p_1, p_2, \dots, p_d]$ . Having obtained the low-dimensional feature representations from the HFA model, we define the age estimation as a multiple linear regression problem in the low-dimensional manifold space. Various linear and non-linear regression methods are available in the literature. To explore the discriminative power of the aging feature, we use multiple regression techniques such as support vector regression, Gaussian process regression (Zhu et al. 2014) and hierarchical age estimation methods on the proposed aging feature.

### 3.3 Age estimation framework

Our age estimation framework mainly incorporates four modules: face preprocessing, feature extraction, discriminative aging subspace learning and regression. In the training stage, face images undergo normalizations such as geometric alignments and illumination normalization (basically

histogram equalization). Then, the histogram-of-oriented-gradient (HOG) (Dalal and Triggs 2005) feature is extracted on each image. The extracted feature is of very high dimension. Next, the age manifold is learned from the hidden factor analysis model (presented in section 3.1), to map the features into a low-dimensional subspace. We further learn a regression function to fit the manifold data. In the test stage, an input face image goes through the same preprocessing and local feature extraction stage. Then the high-dimensional local feature is transformed to the low-dimensional aging feature by projecting it on the learned aging manifold. Finally, the age of the input face image is estimated by fitting the regression on the learned low-dimensional aging feature.

## 4 Experiments

In this section, we first present evaluation metric and evaluation protocol used for performance analysis of the proposed work. Next, in the result section we first demonstrate the robustness of the proposed discriminative aging feature by using it along with the multiple regression schemes. Next, we perform age estimation experiments to verify the effectiveness of the proposed approach and to compare the proposed work with the state-of-the-art methods.

Our approach of aging manifold-based age estimation has been evaluated on MORPH-II (Ricanek and Tesafaye 2006) database. It contains 55,314 face images of 12,936 subjects in the age range 16 to 77 years. For training HFA model using MORPH-II dataset, we followed similar settings as in the recent works (Gong et al. 2013, 2015; Li et al. 2017) where 12,000 images of 6,000 subjects are used for training HFA model. Remaining images from the MORPH-II dataset are used for age estimation experiments. Note that we first train HFA model to learn discriminative aging subspace and then extract aging feature for regression from the learned subspace. In the preprocessing, an input face image is first aligned and then cropped to size  $200 \times 160$ . The cropped face image is converted to a gray-scale image and then histogram equalized. We divide preprocessed face image into overlapping patches and extract HOG feature from each patch. The extracted feature results into a very high dimension therefore we apply widely used principle component analysis (PCA) + linear discriminant analysis (LDA) prior to learn the HFA model. As per our settings the feature space dimension is 6000.

### 4.1 Evaluation metrics and protocol

Performance of age estimation techniques is assessed using two evaluation metrics, mean absolute error (MAE) and cumulative score (CS). The smaller the MAE, the better the age estimation performance. MAE shows the average perfor-

mance of the age estimation technique and is an appropriate measure when the training data has many missing images. MAE indicates the mean absolute error between the predicted result and the ground truth for testing set and is given by,

$$MAE = \sum_{k=1}^N \frac{|\hat{y}_k - y_k|}{N} \tag{10}$$

In addition to MAE, CS is another performance metric which computes the overall accuracy of the estimator and is defined as:

$$CS(k) = \frac{N_{e < k}}{N} \times 100\% \tag{11}$$

where  $N_{e < k}$  is the number of test images for which the absolute error by the age estimation algorithm is not higher than  $k$  years. Various state-of-the-art methods provides CS at 5 years; hence, we have also selected  $k = 5$ . The higher the CS value, the better the age estimation performance. CS is a useful measure of performance in age estimation when the training dataset has samples at almost every age. However, in age estimation, due to imbalanced and skewed databases both MAE and CS are used for evaluation.

## 4.2 Result

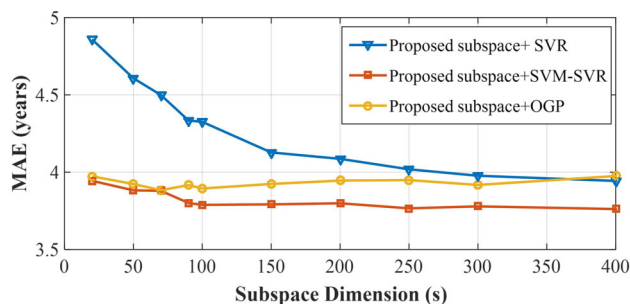
### 4.2.1 Analysis of the proposed subspace

To demonstrate the effectiveness of the proposed manifold learning scheme, we perform regression on the aging features extracted from the discriminative aging subspace. We selected standard regression techniques such as support vector regression (SVR), orthogonal Gaussian process (OGP) regression (Zhu et al. 2014) and hierarchical SVM-SVR (Han et al. 2013, 2015) for age estimation. In SVR and SVM learning, LIBSVM (Chang and Lin 2011) is used to evaluate the approaches. The SVM and SVR employ radial basis function kernel and their parameters were found by a grid search using fivefold cross-validation. Table 1 lists the MAE and CS of three regression methods on the proposed aging manifold. We observed that hierarchical regression yields best performance on the proposed subspace. Note that the performance of proposed subspace along with SVR and OGP are also competent with the state-of-the-art age estimation methods.

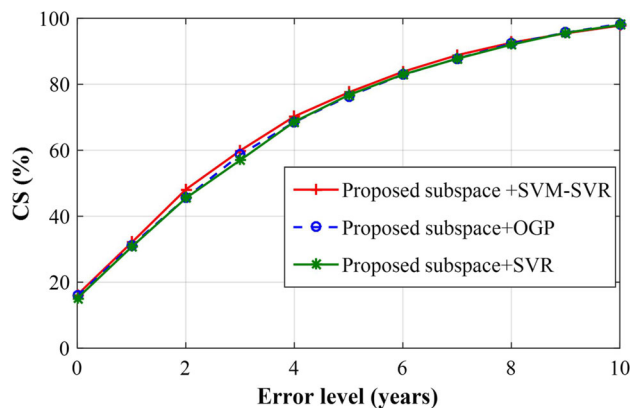
Furthermore, we have conducted experiments to demonstrate the effect of subspace dimensionality on MAE of three regression methods. Large parameter  $s$  implies more age information of a face image. We choose 400 as the upper limit for the dimension of learned manifold subspace. Note that the original feature space has dimension 6000. Figure 2 compares MAE versus dimensionality of the proposed aging

**Table 1** Performance of The Proposed Method with Different Regression Techniques

Regression Technique	MAE	CS
Proposed subspace+SVR	3.93	77
Proposed subspace+OGP	3.88	77
Proposed subspace+(SVM-SVR)	3.75	79



**Fig. 2** MAE versus dimensionality of proposed aging subspace



**Fig. 3** Age estimation comparison of three regression methods on proposed aging manifold in terms of CS versus error levels

subspace on multiple regression methods. We observe in Fig. 2 that when subspace dimensionality reaches 300, the age discriminative information extracted by the aging manifold saturates. Comparative analysis in terms of CS is shown in Fig. 3.

To analyze the effect of identity subspace dimension on the age estimation performance, we conducted one more experiment. In that, we fixed the aging subspace and varied identity subspace dimension and recorded computation time (HFA training + age estimation) and MAE for that setting. The details of identity subspace dimension, aging subspace dimension and MAE are shown in Table 2. As far as age estimation is concerned, as already discussed the aging subspace dimension plays major role. However, the identity subspace dimension has hardly any role in the age estimation. It is clear from Table 2 that for any given aging subspace dimension, the MAE corresponding to identity subspace dimension of 10,

**Table 2** Effect of aging subspace and identity subspace dimension on accuracy

Aging subspace dimension (10)		Aging subspace dimension (50)		Aging subspace dimension (100)	
Identity Subspace Dimension	MAE	Identity subspace dimension	MAE	Identity subspace dimension	MAE
10	4.11	10	3.86	10	3.83
20	4.05	20	3.91	20	3.84
50	4.05	50	3.92	50	3.84
100	4.01	100	3.88	100	3.83
150	4.13	150	3.88	150	3.82
200	4	200	3.92	200	3.88
250	3.98	250	3.92	250	3.88
300	4.05	300	3.94	300	3.88
Aging subspace dimension (200)		Aging subspace dimension (250)		Aging subspace dimension (300)	
Identity subspace dimension	MAE	Identity subspace dimension	MAE	Identity Subspace Dimension	MAE
10	3.78	10	3.82	10	3.74
20	3.81	20	3.75	20	3.79
50	3.79	50	3.75	50	3.79
100	3.78	100	3.74	100	3.82
150	3.80	150	3.76	150	3.84
200	3.86	200	3.79	200	3.79
250	3.86	250	3.83	250	3.79
300	3.85	300	3.85	300	3.80

**Table 3** Comparison of the proposed age estimation method to previous works in terms of MAE (years) and CS (%)

Technique	Facial feature	MAE / CS
AGES (Geng et al. 2007)	AAM	8.8 / 46%
OGP (Zhu et al. 2014)	SIFT	3.92 / -
KPLS (Guo and Mus 2011)	BIF	4.18 / -
Flexible overlap (Pontes et al. 2016)	LPQ <sub>7×7</sub>	5.86 / 68%
Guo et. al Guo et al. (2009)	BIF	4.2 / -
LU and Tan Lu and Tan (2013)	Manifold of raw intensity	White: 5.2/- Black : 4.2/-
Han et. al. Han et al. (2013)	Demographic informative features	5.10 / -
CNN (Huerta et al. 2015)	Deep features	3.88 / -
CPNN (Geng et al. 2013)	Deep features	4.87 / -
Proposed <b>Method</b>	Age discriminative features	3.7 / 79%

50, 100, 200, 250 and 300 is hardly dependent on the identity subspace dimension. For the given aging subspace dimension, the MAE does not vary drastically with respect to the corresponding identity subspace dimension. Thus the MAE is practically independent of the identity subspace dimension. However a weak phenomenon is observed in the Table 2 that for 2 cases, best performance was observed for identity subspace dimension 100. From Table 2, by varying the identity subspace dimension we do not found any particular trend in MAE. Thus finally we conclude that the age estimation performance hardly dependent on identity subspace dimension endorsing the robustness of the only aging subspace-based proposed age estimation scheme.

#### 4.2.2 Age estimation experiments

We also compare the proposed approach with the state-of-the-art age estimation algorithms such as appearance and age specific (AAS) (Lanitis et al. 2002), KPLS (Guo and Mus 2011), hierarchical age estimation using SVM-SVR (Han et al. 2015; Liang et al. 2014; Pontes et al. 2016; Thukral et al. 2012) and a few deep learning methods (Huo et al. 2016; Wang et al. 2015). Table 3 shows the comparison study among proposed approach and various state-of-the-art approaches in terms of MAE and CS at 5 years error. Clearly, the proposed method has a better MAE and CS than the competing approaches. This indicates that the pro-



posed aging subspace provides discriminative aging feature and about 79% of the age predictions from it do not differ more than 5 years from the ground truth. Therefore, the proposed subspace learning from HFA model represents a robust aging manifold. In the proposed work, we have found two components from the local feature vector, *i.e.*, identity and age factor, while all other facial attributes are considered as noise components. Analysis of effect of other facial attributes such as ethnicity, pose and gender on the age estimation is a viable future scope of this work. Various optimization techniques such as Abualigah et al. (2022, 2021a, b); Abualigah (2019) are available in the literature. Such optimization techniques can be further explored for age estimation.

## 5 Conclusion

In this paper, we have proposed a probabilistic HFA approach to address the challenging problem of age estimation. The HFA model has been widely used in many AIFR systems, it the first time we have proposed the HFA model to learn a low-dimensional aging manifold. The basic idea of the proposed model is to pursue a robust aging manifold and corresponding aging feature descriptor from the age component of the HFA model. Experimentally we have shown that the proposed age estimation scheme is hardly dependent on the identity subspace dimension. This implies, the HFA-based aging manifold is not person dependent. Extensive experiments conducted on MORPH Album II demonstrate that proposed subspace can extract discriminative aging features for age estimation. Experimental results also demonstrate the effectiveness and robustness of the proposed aging manifold on different regression algorithms. It is encouraging and also interesting to see that our method surpasses the state-of-the-art methods including deep learning approaches. Analyzing impact of ethnicity and other facial attributes using ethnicity and other components of revised HFA model is a viable extension of the proposed work.

**Funding** The authors have not disclosed any funding.

**Data availability** Enquiries about data availability should be directed to the authors.

## Declarations

**Conflict of interest** The authors have not disclosed any conflict interest.

## References

Abualigah LMQ et al (2019) Feature selection and enhanced krill herd algorithm for text document clustering. Springer

- Abualigah L, Diabat A, Mirjalili S, Abd Elaziz M, Gandomi AH (2021) The arithmetic optimization algorithm. *Comput Methods Appl Mech Eng* 376:113609
- Abualigah L, Yousri D, Abd Elaziz M, Ewees AA, Al-Qaness MA, Gandomi AH (2021) Aquila optimizer: a novel meta-heuristic optimization algorithm. *Comput Ind Eng* 157:107250
- Abualigah L, Abd Elaziz M, Sumari P, Geem ZW, Gandomi AH (2022) Reptile search algorithm (rsa): a nature-inspired meta-heuristic optimizer. *Expert Syst Appl* 191:116158
- Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Transact Pattern Anal Mach Intell* 28(12):2037–2041
- Ahonen T, Rahtu E, Ojansivu V, Heikkilä J (2008) Recognition of blurred faces using local phase quantization. In: 2008 19th international conference on pattern recognition, pp 1–4. IEEE
- Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6):1373–1396
- Chang KY, Chen CS, Hung YP (2011) Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: Computer vision and pattern recognition (cvpr), 2011 IEEE conference on, pp 585–592. IEEE
- Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. *ACM Transact Intell Syst Technol (TIST)* 2(3):27
- Chao WL, Liu JZ, Ding JJ (2013) Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognit* 46(3):628–641
- Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. *IEEE Transact Pattern Anal Mach Intell* 23(6):681–685
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol 1, pp 886–893. IEEE
- Fu Y, Huang TS (2008) Human age estimation with regression on discriminative aging manifold. *IEEE Transact Multimed* 10(4):578–584
- Geng X, Zhou ZH, Smith-Miles K (2007) Automatic age estimation based on facial aging patterns. *IEEE Transact Pattern Anal Mach Intell* 29(12):2234–2240
- Geng X, Yin C, Zhou ZH (2013) Facial age estimation by learning from label distributions. *IEEE Transact Pattern Anal Mach Intell* 35(10):2401–2412
- Gong D, Li Z, Lin D, Liu J, Tang X (2013) Hidden factor analysis for age invariant face recognition. In: Computer Vision (ICCV), 2013 IEEE International Conference on, pp 2872–2879. IEEE
- Gong D, Li Z, Tao D, Liu J, Li X (2015) A maximum entropy feature descriptor for age invariant face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5289–5297
- Guo G, Fu Y, Dyer CR, Huang TS (2008) Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transact Image Process* 17(7):1178–1188
- Guo G, Mu G (2011) Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In: Computer vision and pattern recognition (cvpr), 2011 IEEE conference on, pp 657–664. IEEE
- Guo G, Mu G (2013) Joint estimation of age, gender and ethnicity: Cca versus pls. In: 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp 1–6. IEEE
- Guo G, Mu G, Fu Y, Huang T.S (2009) Human age estimation using bio-inspired features. In: Computer vision and pattern recognition, 2009. CVPR 2009. IEEE Conference on, pp 112–119. IEEE
- Han H, Otto C, Liu X, Jain AK (2015) Demographic estimation from face images: human versus machine performance. *IEEE Transact Pattern Anal Mach Intell* 37(6):1148–1161

- Han H, Otto C, Jain A.K (2013) Age estimation from face images: human versus machine performance. In: Biometrics (ICB), 2013 international conference on, pp 1–8. IEEE
- Huerta I, Fernández C, Segura C, Hernando J, Prati A (2015) A deep analysis on age estimation. *Pattern Recognit Lett* 68:239–249
- Huo Z, Yang X, Xing C, Zhou Y, Hou P, Lv J, Geng X (2016) Deep age distribution learning for apparent age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 17–24
- Iqbal MTB, Shoyaib M, Ryu B, Abdullah-Al-Wadud M, Chae O (2017) Directional age-primitive pattern (dapp) for human age group recognition and age estimation. *IEEE Transact Inf Forensics Secur* 12(11):2505–2517
- Lanitis A, Taylor CJ, Cootes TF (2002) Toward automatic simulation of aging effects on face images. *IEEE Transact Pattern Anal Mach Intell* 24(4):442–455
- Li H, Zou H, Hu H (2017) Modified hidden factor analysis for cross-age face recognition. *IEEE Signal Process Lett* 24(4):465–469
- Liang Y, Wang X, Zhang L, Wang Z (2014) A hierarchical framework for facial age estimation. *Mathematical Probl Eng* **2014**
- Ling H, Soatto S, Ramanathan N, Jacobs DW (2007) A study of face recognition as people age. In: Computer vision, 2007. ICCV 2007. IEEE 11th international conference on, pp 1–8. IEEE
- Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision, vol 2, pp 1150–1157. IEEE
- Lu J, Tan YP (2013) Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Transact Human-Machine Syst* 43(2):249–258
- Moon TK (1996) The expectation-maximization algorithm. *IEEE Signal Process Mag* 13(6):47–60
- Pontes JK, Britto AS Jr, Fookes C, Koerich AL (2016) A flexible hierarchical approach for facial age estimation based on multiple features. *Pattern Recognit* 54:34–51
- Prince SJ, Elder JH (2007) Probabilistic linear discriminant analysis for inferences about identity. In: Computer vision, 2007. ICCV 2007. IEEE 11th international conference on, pp 1–8. IEEE
- Prince SJ, Elder JH, Warrell J, Felisberti FM (2008) Tied factor analysis for face recognition across large pose differences. *IEEE Transact Pattern Anal Mach Intell* 30(6):970–984
- Ricanek K, Tesafaye T (2006) Morph: a longitudinal image database of normal adult age-progression. In: Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on, pp 341–345. IEEE
- Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326
- Sawant M, Bhurchandi K (2019) Hierarchical facial age estimation using gaussian process regression. *IEEE Access* 7:9142–9152
- Sawant M, Addepalli S, Bhurchandi K (2019) Age estimation using local direction and moment pattern (ldmp) features. *Multime Tools Appl* 78(21):30419–30441
- Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):2319–2323
- Thukral P, Mitra K, Chellappa R (2012) A hierarchical approach for human age estimation. In: Acoustics, speech and signal processing (ICASSP), 2012 IEEE international conference on, pp 1529–1532. IEEE
- Vapnik V (1999) The nature of statistical learning theory. Springer science & business media
- Wang X, Guo R, Kambhamettu C (2015) Deeply-learned feature for age estimation. In: Applications of computer vision (WACV), 2015 IEEE winter conference on, pp 534–541. IEEE
- Wang J, Shang Y, Su G, Lin X (2006) Age simulation for face recognition. In: Pattern recognition, 2006. ICPR 2006. 18th international conference on, vol. 3, pp 913–916. IEEE
- Weng R, Lu J, Yang G, Tan YP (2013) Multi-feature ordinal ranking for facial age estimation. In: 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp 1–6. IEEE
- Wen Y, Li Z, Qiao Y (2016) Latent factor guided convolutional neural networks for age-invariant face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4893–4901
- Xiao B, Yang X, Xu Y, Zha H (2009) Learning distance metric for regression by semidefinite programming with application to human age estimation. In: Proceedings of the 17th ACM international conference on Multimedia, pp 451–460
- Xu C, Liu Q, Ye M (2017) Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing* 222:62–71
- Zhang Y, Yeung DY (2010) Multi-task warped gaussian process for personalized age estimation. In: Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, pp 2622–2629. IEEE
- Zhu K, Gong D, Li Z, Tang X (2014) Orthogonal gaussian process for automatic age estimation. In: Proceedings of the 22nd ACM international conference on multimedia, pp 857–860. ACM

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.