

REVIEW

Open Access



# Age estimation via face images: a survey

Raphael Angulu<sup>1\*†</sup> , Jules R. Tapamo<sup>2</sup> and Aderemi O. Adewumi<sup>1</sup>

## Abstract

Facial aging adversely impacts performance of face recognition and face verification and authentication using facial features. This stochastic personalized inevitable process poses dynamic theoretical and practical challenge to the computer vision and pattern recognition community. Age estimation is labeling a face image with exact real age or age group. How do humans recognize faces across ages? Do they learn the pattern or use age-invariant features? What are these age-invariant features that uniquely identify one across ages? These questions and others have attracted significant interest in the computer vision and pattern recognition research community. In this paper, we present a thorough analysis of recent research in aging and age estimation. We discuss popular algorithms used in age estimation, existing models, and how they compare with each other; we compare performance of various systems and how they are evaluated, age estimation challenges, and insights for future research.

**Keywords:** Age estimation, Face, Anthropometry, Models, Feature, Classification

## 1 Introduction

You can never see the same face twice. This statement is true because facial appearance varies more dynamically as it is affected by several factors including pose, facial expression, head profile, illumination, aging, occlusion, mustache, beards, makeup (cosmetics), and hair style. Major factors that influence facial aging include gravity, exposure to ultraviolet (UV) rays from the sun, maturity of soft tissues, bone re-structuring, and facial muscular activities [1]. These factors cause variations in face appearance. For instance, a face seen in blue light illumination is totally different from one seen under red light illumination. Another factor that constantly and permanently causes variations in facial appearance is age. Aging is an inevitable stochastic process that affects facial appearance. Aging involves both variations in soft tissues and bony structure on the human face. A face seen at one age is totally different from the face of same individual at a different age. Therefore, these age-introduced variations could be learned and used to estimate facial age.

The human face provides prior perceptible information about one's age, gender, identity, ethnicity, and mood. Alley [2] asserts that attributes derived from human facial

appearance like mood and perceived age significantly impact interpersonal behavior as is considered as essential contextual cue in social networks [3, 4]. Information rendered by the human face has attracted significant attention in the face image processing research community. Image-based age and age-group estimation particularly has attracted enormous research interest due to its vast application areas like age-invariant face recognition and face verification across age, among other commercial and law enforcement areas [5–9]. Age estimation has been extensively studied with the aim of finding out aging patterns and variations and how to best characterize an aging face for accurate age estimation.

Age estimation research has gained significant attention in recent years with many journal and conference papers being published annually as well as Masters and PhD theses defended [10]. Age estimation is a technique of automatically labeling the human face with an exact age or age group. This age can be either actual age, appearance age, perceived age, or estimated age [11]. Actual age is the number of years one has accumulated since birth to date, denoted as a real number. Appearance and perceived age are estimated based on visual age information portrayed on the face while estimated age is a subject's age estimated by a machine from the facial visual appearance. Appearance age is assumed to be consistent with actual age although there are variations due to the stochastic nature of aging among individuals. Estimated

\*Correspondence: [rangulu@mmust.ac.ke](mailto:rangulu@mmust.ac.ke)

†Equal contributors

<sup>1</sup>University of KwaZulu-Natal, School of Mathematics, Statistics and Computer Science, Westville, Durban, 4000 South Africa

Full list of author information is available at the end of the article

age and perceived age are defined on visual artifacts of appearance age. There has been relatively few publications on age and age-group estimation [11]. This could be attributed to age estimation not being a classical classification problem. Age estimation can be approached as a multi-class classification problem or a regression problem or as an ensemble of both classification and regression in a hierarchical manner. Another reason that could be affecting research in age estimation is the difficulty in collecting a large database with chronological images for a subject. Prolific and diverse information conveyed by faces also make special attributes of aging variations not accurately captured [11]. Uncontrollable and personalized age progression information displayed on faces further complicates age estimation problem [12–14].

## 2 Facial aging

Aging is a stochastic, uncontrollable, inevitable, and irreversible process that causes variations in facial shape and texture. Although aging is stochastic with different people having different aging patterns, there are some general variations and similarities that can be modeled [15, 16]. There are two stages in human life that are distinct with regard to facial growth: formative or childhood stage and adulthood or aging stage [17].

Aging introduces significant change in facial shape in formative years and relatively large texture variations with still minor change in shape in older age groups [11, 18]. Shape variations in younger age groups are caused by craniofacial growth. Craniofacial studies have shown that human faces change from circular to oval as one ages [19]. These changes lead to variations in the position of fiducial landmarks [20]. During craniofacial development, the forehead slopes back releasing space on the cranium. The eyes, ears, mouth, and nose expand to cover interstitial space created. The chin becomes protrusive as cheeks extend. Facial skin remains moderately unchanged than shape. More literature on craniofacial development is found in [16].

As one ages, facial blemishes like wrinkles, freckles, and age spots appear. Underneath the skin, melanin-producing cells are damaged due to exposure to the sun's ultraviolet (UV) rays. Freckles and age spots appear due to overproduction of melanin. Consequently, light-reflecting collagen not only decreases but also becomes non-uniformly distributed making facial skin tone non-uniform [1]. Parts adversely affected by sunlight are the upper cheek, nose, nose bridge, and forehead.

The most visible variations in adulthood to old age are skin variations exhibited in texture change. There is still minimal facial shape variation in these age groups. Biologically, as the skin grows old, collagen underneath the skin is lost [11]. Loss of collagen and effect of gravity make the skin become darker, thinner, leathery,

and less elastic. Facial spots and wrinkles appear gradually. The framework of bones beneath the skin may also start deteriorating leading to accelerated development of wrinkles and variations in skin texture. More details about face aging in adulthood is found in [16]. These variations in shape and texture across ages could be modeled and used to automatically estimate someone's age. We refer readers to [16] for more details on facial aging. Facial aging has three unique attributes [13]:

1. *Aging is inevitable and uncontrollable.* No one can avoid aging, advance, or delay it. The aging process is slow but irreversible.
2. *Aging patterns are personalized.* People age differently. Individuals' aging pattern is dependent on her/his genetic makeup as well as various extrinsic factors such as health, environmental conditions, and lifestyle.
3. *Achieved aging patterns are temporal.* Facial variations caused by aging are not permanent. Furthermore, facial variation at a particular point in time affects future appearance and does not affect previous appearance of these faces.

These facial aging attributes, among other factors, make automatic age estimation a difficult and challenging task. Since individuals cannot voluntarily control aging, automatic age estimation data collection becomes a hard task to do. This problem was slightly alleviated by dissemination of FG-NET Aging Dataset [21] in 2002. Although this dataset has images of subjects at different ages, there are several missing images hence making the aging patterns incomplete. Fortunately, we do not need a complete aging face dataset since people, who computers try to mimic, also learn how to process face image patterns from incomplete patterns. Age estimation technique should be capable of considering various aging patterns since each individual has his/her own aging pattern.

Information rendered by the human face has attracted significant attention in the face image processing research community. Image-based age and age-group estimation has vast application areas like age-invariant face recognition, face verification across ages, commercial and law enforcement areas [5–9], security control and surveillance [11, 22], age-based image retrieval [23], biometrics [11, 24, 25] human computer interaction [26, 27], and electronic customer relationship management (ECRM) [11]. The main aim of studying age estimation is to find out aging patterns and variations in facial appearance and how to best characterize an aging face for accurate age estimation. Although this problem has attracted significant research, still automatic age estimation accuracies are far below human accuracy.

### 3 Age estimation application areas

Characterizing variations in facial appearance across age has many significant real-world applications. Computer-based age estimation is useful in situations where one's age is to be determined. There are several application areas for age estimation including the following:

#### 3.0.1 Age simulation

Characterization of facial appearance at different ages could be effectively used in simulating or modeling one's age at a particular point in time. Estimated ages at different times could help in learning the aging pattern of an individual, which could assist in simulating facial appearance of the individual at some unseen age. More details on facial aging simulation could be found in [28, 29]. By observing aging patterns at different ages, unseen appearance could be simulated and used to find missing persons. By observing aging patterns at different ages, unseen appearance could be simulated.

#### 3.0.2 Electronic customer relationship management (ECRM)

ECRM [11] is the use of Internet-based technologies such as websites, emails, forums, and chat rooms, for effective managing of distinguished interactions with clients and individually communicating to them. Customers in different ages may have diverse preferences and expectations of a product [30]. Therefore, companies may use automatic age estimation to monitor market trends and customize their products and services to meet needs and preferences of customers in different age groups. The problem here is how to acquire and analyze substantive personal data from all client groups without infringing on their privacy rights. With automatic age estimation, a camera can snap pictures of clients and automatically estimate their age groups in addition to collection of demographic data.

#### 3.0.3 Security and surveillance

Age estimation can be used in surveillance and monitoring of alcohol and cigarette vending machines and bars for preventing underage from accessing alcoholic drinks and cigarettes and restricting children access to adult websites and movies [23, 31]. Age estimation can also be significant in controlling ATM money transfer fraud by monitoring a particular age group that is apt to the vice [11]. Age estimation can also be used to improve accuracy and robustness of face recognition hence improving homeland security. Age estimation can also be used in health-care systems like robotic nurse and doctors expert system for customized medical services. For instance, a customized avatar can be automatically selected from a database for interacting with patients from various age groups depending on preferences.

#### 3.0.4 Biometrics

Age estimation via faces is a soft biometric [32] that can be used to compliment biometric techniques like face recognition, fingerprints, or iris in order to improve recognition, verification, or authentication accuracies. Age estimation can be applied in age-invariant face recognition [10], iris recognition, hand geometry recognition, and fingerprint recognition in order to improve accuracy of hard (primary) biometric system [11].

#### 3.0.5 Employment

Some government employments like the military and police consider one's age as a requirement. Age estimation systems could be used to determine age of the recruits during recruitment process. It is also a policy of several governments that employees should retire after reaching a particular age. Age estimation systems could also play a significant role in finding if one has reached retirement age.

#### 3.0.6 Content access

With the proliferation of diverse content in televisions (TV) and the Internet, age estimation can be used to control access to unwanted content to children. A camera could be mounted on a TV to monitor people looking at it such that it switches off the TV if at a particular time unwanted content is streamed and people watching are children.

#### 3.0.7 Missing persons

Age estimation role in age simulation go a step further in aiding identification of missing persons. Age simulation can be used to identify old people from their previous images for purposes of identification.

## 4 Factors affecting facial aging

Facial aging is affected by several factors ranging from lifestyle, natural, occupation, psychological, and environmental. Factors affecting facial aging can be categorized as both intrinsic and extrinsic. Extrinsic factors are those that are external to the human body like environmental and occupation factors while intrinsic are internal factors like bone structure and genetic influence which occur naturally over time [1, 33]. In childhood, facial changes are mainly caused by craniofacial development which lead to changes in facial shape [16] due to growth, modeling, and deposition of bony tissues in the face. This leads to changes in height and shape of the face [34]. The forehead slopes back releasing space on the cranium. Drifting and expansion of facial landmarks to occupy this space causes variations in facial shape in childhood. In adulthood, facial aging is mainly manifested in texture variations which are caused by a wide variety of factors.

Taister et al. [34] found that general exposure to wind and arid air influences facial aging. Arid environment and wind dehydrate the skin leading to wrinkle formation. Air pollution has also been found to affect aging by accelerating wrinkle development [35–37]. Research on air pollution and aging has shown that city dwellers who are exposed to air pollution from industries develop deep wrinkles than individuals who are not exposed to pollution. Smoking influence on aging has also been cited in [34, 38–40] although [41] asserts that smoking has negligible effect to facial wrinkling compared to effect of UV rays. However, smoking interrupts skin microvasculature which affects elastin and collagen production and functioning leading to wrinkles around the mouth, but photoaging effects lead to more facial wrinkling compared to smoking [34, 41]. It is therefore evident that facial skin aging does not provide objective analysis of cumulative exposure to UV rays. Taister et al. [34] also assert that exposure to drug and psychological stress affects skin texture and pigmentation making skin complexion spotted and blemished.

Exposure to ultraviolet (UV) rays influences production of collagen making the skin darker. UV rays dry and destroy cells and underlying skin structure, giving the skin a furrowed and thickened appearance hastening development of wrinkles especially around the eyes due to squinting effects [42]. Long exposure to UV rays leads to variations in photoaging like skin wrinkling, elastosis, actinic keratosis, and irregular pigmentation [43]. With long exposure to UV rays, skin texture and color change becoming blotchy, yellowish, leathery, loose, inelastic, and hyper-pigmented. Blood veins close to the skin surface become protrusive forming “spider vein” network in addition to overall speckled skin appearance [44]. Naturally, with lower production of collagen and elastin, the skin becomes leathery and less elastic. Fat cells begin to disappear leading to skin sagging. Fat deposits in some areas like the eye lobe region also affect skin texture. Force of gravity makes the skin leathery and less elastic hence accelerating skin wrinkling.

Internally, changes in bone structure and subsequent variations in musculature cause skin wrinkling [16]. Loss of skin elasticity makes the skin leathery leading to formation of wrinkles [45]. Aging was also found to be different between males and females with female faces tending to age faster compared to male faces [16].

Aging in males and females share many common characteristics, but there are some differences. Although it is generally acknowledged that females age faster compared to men, it is not yet clear whether these gender differences are caused by rate of aging or sexual dimorphism [16]. Investigation into differences in aging between males and females is necessary [46]. Differences in male facial aging include manifestation of facial hair like beards,

increased thickness, facial vascularity, sebaceous content, and potential differences in fat and bone absorption rates [47]. Development of deeper wrinkles around the perioral region is high in women compared to men [47] since women’s skin has few appendages compared to men [48]. Some women look younger than their actual age and have large lips and are genetically protected from wrinkle and gray hair development [49].

Other factors affecting perceived facial aging include diet, genetic makeup, ethnicity (race), skin infections, and cosmetics. Cosmetics are generally used to hide perceived age of an individual by hiding wrinkles and age spots and brightening wrinkle shadows around the eyes, mouth, and nose regions [50]. Chen et al. [50] found that facial makeup significantly impacts age estimation. Guo and Wang [51] and Nguyen et al. [52] investigated the effect of facial expression in age estimation. By quantitative evaluations on Lifespan [53] and FACES [54] datasets, Guo and Wang [51] found that facial expression influences age estimation. Same findings were reported by Nguyen et al. [52]. Voelkle and Ebner [55] investigated the effect of age, gender, and facial expression on perceived age. They found that facial expression influences age estimation with faces with happy facial expressions most underestimated. Some facial expressions like smiling, frowning, surprise, and laughing may introduce wrinkle-like lines on some regions of the face like the forehead, cheek bone area, mouth region, and nose-bridge regions. These wrinkle-like lines may be registered as wrinkles during age estimation hence having an impact on age estimation performance.

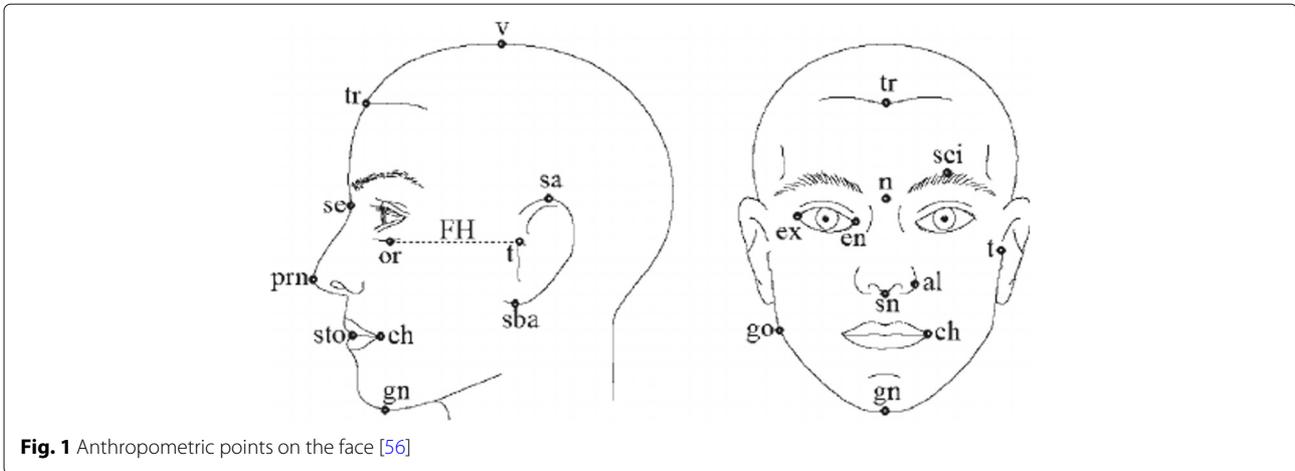
## 5 Image representation for age modeling

In this section, we present different approaches used for image representation for age estimation. Age estimation can be modelled using anthropometric data, active appearance model (AAM) parameters, aging pattern subspace (AGES), manifold learning, appearance features, or a hybrid of two or more modeling technique. We present an overview of these modeling techniques in the subsequent sections.

### 5.1 Anthropometric models

Anthropometric modeling of facial aging focuses on distance measurements between facial points. Face anthropology is the study of measuring sizes and proportions on human faces [56]. Farkas [56] defined face anthropology based on measurements taken from 57 landmark points on human faces. Figure 1 shows some of the points used to describe a face. Landmark points are identified by abbreviation of their respective anatomical names. For instance, the eye inner corner is *en* for *endocanthion* while front of the ear is *t* for *tragion*.

Farkas defined five measurements between landmarks: *shortest distance*, *axial distance*, *tangential distance*, *angle*



of inclination, and angle between locations. Figure 2 shows sample measurements of these distances.

A total of 132 facial measurements were defined by Farkas [56], whereby some corresponding measurements on the left and right of the face were paired. The measurements can be taken by hand by experienced anthropometrists or 3D scanners [56–58].

Facial measurements could be taken at different ages for instance from childhood to old age. Ratios of distances between facial landmarks like the eyes, nose, mouth, ear, chin, and forehead are measured across age. Facial measurements are used to determine the aging pattern of an individual at a particular age and hence used to discriminate between ages and age groups. This approach embraces studies in craniofacial development theory [2].

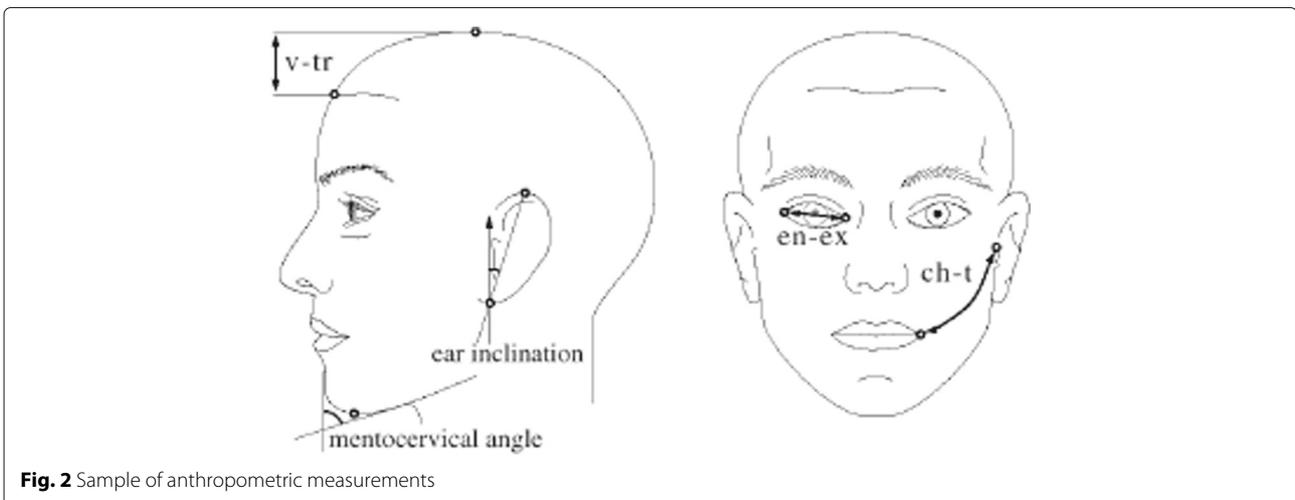
Craniofacial development theory uses cardioid strain transformation mathematical model to describe a person’s facial growth from infancy to adult age. This model defines a circle to track facial growth by tracking

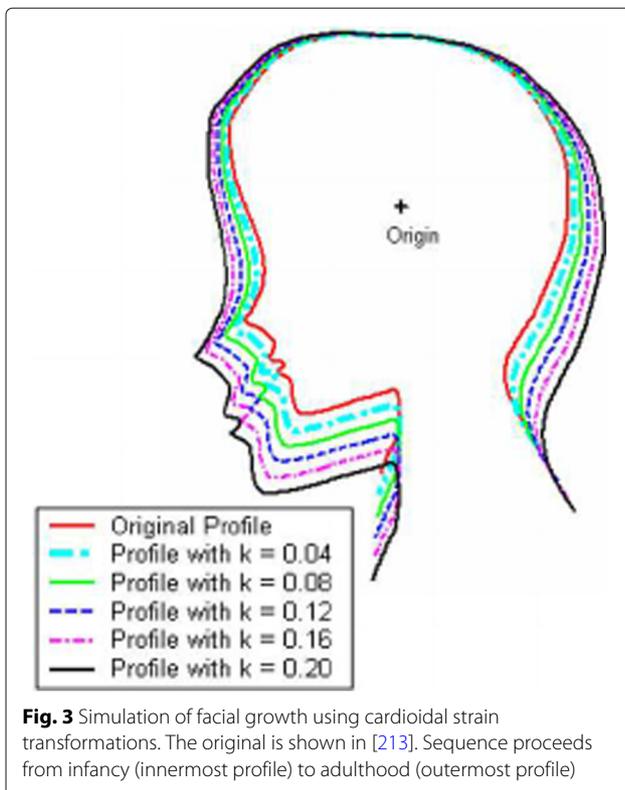
variations in radius of the circle as

$$R' = R(1 + k(1 - \cos\theta)) \tag{1}$$

where  $R$  is the initial radius of the circle,  $\theta$  is the initial angle formed with the vertical axis,  $k$  is a parameter that increases with time, and  $R'$  is the successive growth of the circle over time. Figure 3 shows simulated face profiles using cardioidal strain transformations.

The mathematical formulation in Eq. 1 is not commonly used for age estimation because it does not encode head profile, especially in adults [59], and head profiles are hard to estimate from 2D facial images [11]. Furthermore, anthropometric models cannot be used for age modeling in adult and old age face images since there are no significant changes in facial shape at these stages. This approach is also only appropriate for frontal face images since distance between landmarks are sensitive to head poses. This modeling technique has not been experimented on a large publicly available database, with few studies reported in the literature working on small private datasets. Another





**Fig. 3** Simulation of facial growth using cardioidal strain transformations. The original is shown in [213]. Sequence proceeds from infancy (innermost profile) to adulthood (outermost profile)

limitation of this approach is that it only considers distance between facial landmarks with no consideration for facial appearance. Measurements and landmark points defined by Farkas in [56], which often guide anthropometric modeling, are from people in one ethnic group (European) and may not be representative of all other races.

## 5.2 Active shape models

Active shape model (ASM) [60] is a statistical model that characterizes shape of an object. ASM builds a model by learning patterns of variability from a training set of correctly annotated images. ASMs are able to capture natural variability of images of the same class unlike active contour models (ACMs) [61]. ASMs are specific to images of the class of objects they represent. Face image shape is denoted by a collection of landmark points. Good choices for landmark points are points at clear corners of the face and facial landmark boundaries. These points can be determined by use of appropriate 2D landmarking algorithm like the one proposed in [62]. The sets of points are automatically aligned to reduce the variance in distance between equivalent points. The number of landmark points must be adequate enough to show overall shape of the face images. Each face is then represented by a predefined number of landmark points depending on complexity of the facial shape and the desired level

of descriptive information. A point distribution model (PDM) is derived by examining spatial statistics of labeled points. PDM gives mean locations of points and a set of parameters that control main variability modes found in the training set.

Given such a model and test image, image interpretation involves choosing values for each of the parameters such that the best fit of the model to the image is found. ASM allows initial rough guess of best shape, orientation, scale, and position which is refined by comparing hypothesized model instance to image data and using difference between model and image to deform to shape. ASM is more similar to AAM but differs in the sense that instances in ASM can only deform according to variations found in the training set. ASM is not commonly used in age estimation; hence, more investigations adopting this modeling strategy are necessary.

Active shape model has the following limitations [63]:

1. Results into poor matching of boundaries in an image due to parametric description of shape. It is not robust when new images are introduced. These lead to problems during subsequent image analysis
2. Active shape model needs many landmark points and training samples to represent shape and its variations. Makes ASM costly and time consuming during training
3. Active shape model segmentation results are sensitive to local search region around landmarks

## 5.3 Active appearance model

Active appearance models (AAMs) [64] are statistical facial image coding models. Using principal component analysis (PCA), AAM learns shape model and intensity model from a set of training images. AAMs have been used extensively in modeling facial shape for face recognition, face verification, age estimation, and gender estimation among other tasks. AAM considers both facial shape and texture unlike anthropometric models that consider shape parameters only. This makes AAMs appropriate for age estimation modeling at all stages from infancy to old age. Labeling each test image with a definite age label from continuous age range makes AAM approaches give precise age estimations [11].

Annotated sets of training images marked with points defining facial main features are needed to build AAM. Figure 4 shows a sample of annotated face and points used for annotation.

These points can be determined by use of appropriate 2D landmarking algorithm like the one proposed in [62]. These sets of points are represented as a vector and aligned before a statistical shape model built. Each training image is then warped so that the annotated points match points of mean shape and obtain a shape-free

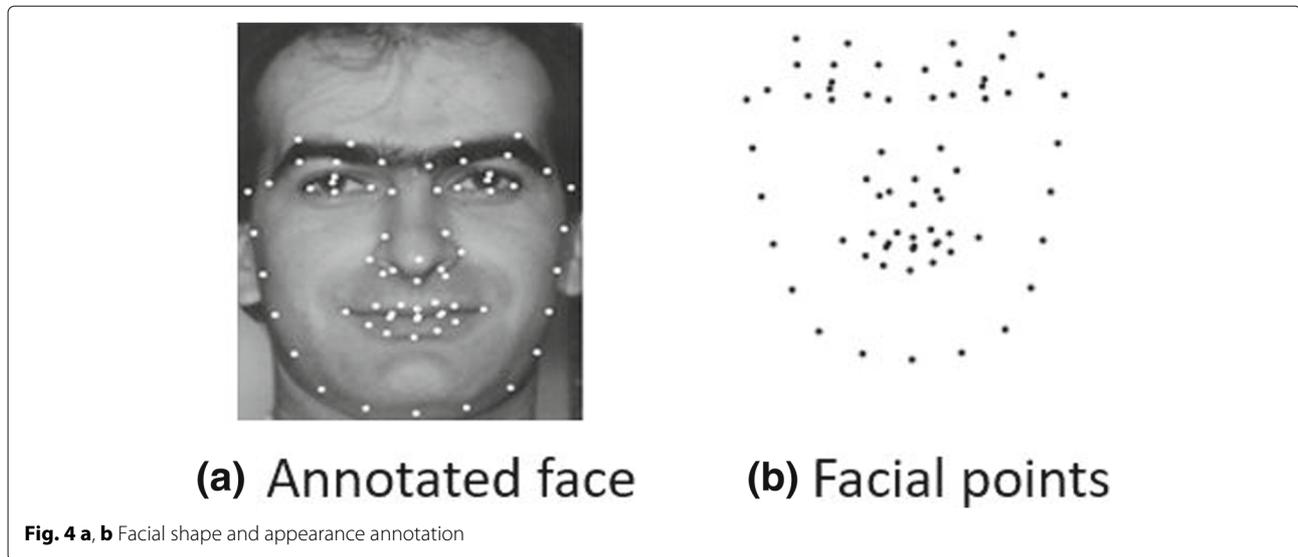


image patch. The shape-free raster is pushed into a texture vector,  $g$ , which is normalized by applying a linear transformation,  $g \leftarrow \frac{(g - \mu_g \mathbf{1})}{\sigma_g}$ , where  $\mathbf{1}$  is a vector of ones and  $\mu_g$  and  $\sigma_g^2$  are the mean and variance of elements of  $g$ , respectively. After normalization,  $g^T \mathbf{1} = 0$  and  $|g| = 1$ . Principal component analysis (PCA) is then used to build a texture model. Finally, connections between shape and texture are learned to produce a combined appearance model as detailed in [65].

The generated appearance model has parameters,  $c$ , controlling the shape and texture according to:

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \tag{2}$$

where  $\bar{x}$  is the mean shape,  $\bar{g}$  is the mean texture in a mean-shaped patch, and  $Q_s$  and  $Q_g$  are matrices describing modes of variation derived from training set. AAM are slower compared to active shape models (ASMs) [60]. Details of AAM implementation could be found in [64].

Lanitis et al. [66] extended AAM by proposing and aging function  $age = f(b)$ . In this function,  $age$  is the real subject's age,  $b$  is AAM-learned vector of 50 raw model parameters, and  $f$  is aging function. The function  $f$  describes the association between an individual's age and vector of parameters.

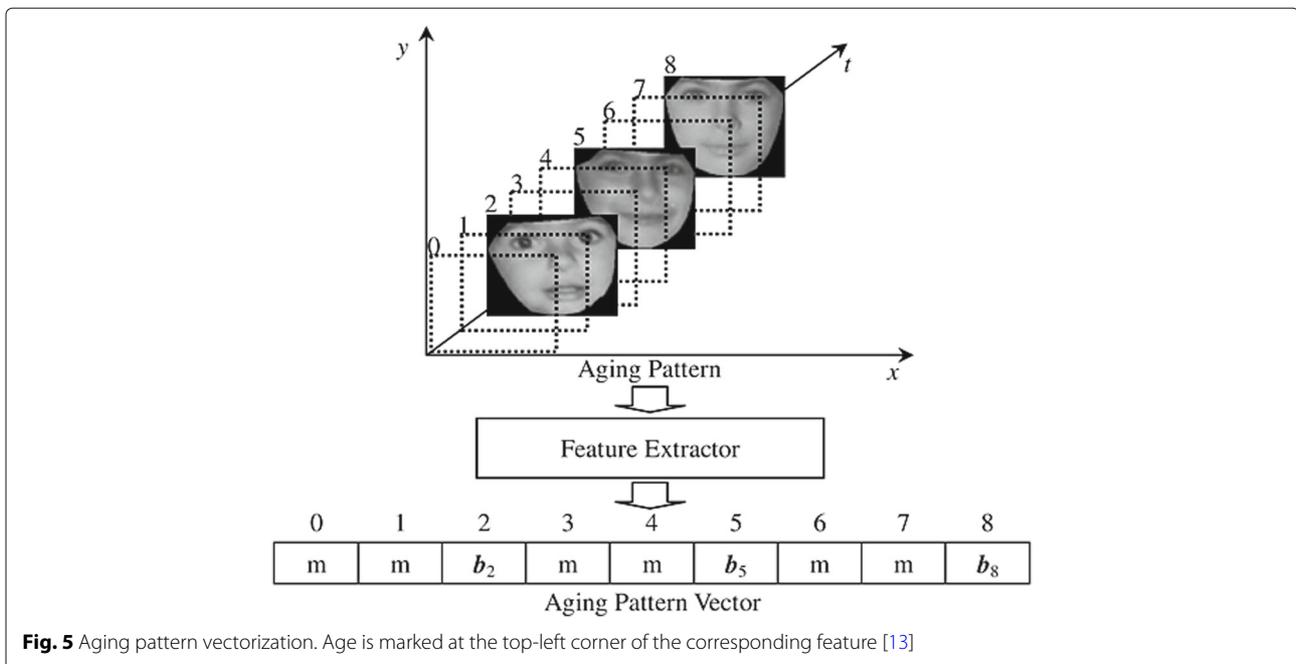
AAM face encoding considers both shape and texture unlike anthropometric techniques that only represent shape. This makes AAM approaches appropriate for age estimation since both texture and shape features necessitate precise age estimation. However, evidence is needed to show that aging patterns can be modelled as a quadratic function and highlight effect of outliers in age estimation. Active appearance model is computational intensive. Training phase requires a substantive number of images

for the model to learn robust shape and appearance features. Active appearance model uses gray-level intensities of the image to train an intensity model. Gray-level intensities may be affected by noise hence leading to a weak intensity model. Performance of AAM depends on the quality of images used. Images with significantly different background and scale inhibit model fitting, resulting in poor performance of AAM-based systems.

#### 5.4 Aging pattern subspace

Geng et al. [13, 26] proposed aging pattern subspace (AGES) for automatic age estimation using appearance of face images. A series of individual images arranged in temporal order make up aging pattern. Aging pattern is defined in [13] as "... a sequence of personal face images sorted in time order." All images in a pattern must come from the same individual and must be ordered by time. This aging pattern is called a complete pattern if images at all ages for an individual are available or else it is referred to as an incomplete pattern. AGES compensate missing ages by learning a subspace representation of one's images when modeling a series of a subject's aging face. To estimate age, test image is positioned at each possible location in the aging pattern to find a point that can best reconstruct it. Aging subspace that minimizes reconstruction error determines age of the test image. Figure 5 shows vectorization of aging pattern with missing images in the aging pattern vector marked with  $m$ . Available face images in the pattern (ages 2, 5, and 8) are placed at their respective positions and ages at which images are not available if their positions are left blank.

After vectorization of the aging pattern, face images at ages 2, 5, and 8 are represented by feature vectors  $b_2, b_5,$  and  $b_8$ , respectively. Representing aging pattern using AGES ensures that label  $age(I)$  and  $id(I)$  are integrated



**Fig. 5** Aging pattern vectorization. Age is marked at the top-left corner of the corresponding feature [13]

into the data whereby each pattern implies an ID and each age is fixed at a particular time-ordered position in the aging pattern.

The first step of AGES is learning, where aging pattern is learned then followed by age estimation. Subspace representation is obtained in the learning stage using PCA. Due to the possibility of missing age images, reconstruction error between available age and reconstructed face image is minimized by expectation maximization (EM) iterative learning technique. Average of the available face images is used to initialize values for missing faces. Thereafter, mean, covariance matrix, and eigenvectors of all face images are computed. Faces are then reconstructed using mean face and eigenvectors. This process is repeated until the reconstruction error is significantly small. During age estimation, the test image finds aging pattern subspace and position in that pattern that can minimize its reconstruction error. The position that gives minimal reconstruction error is returned as the estimated age of the probe image. Ghost-like twisted faces are reconstructed when test image is positioned at a wrong location in the aging pattern subspace [13, 26].

AGES was evaluated on FG-NET [21] and a MAE of 6.77 years was reported [13, 26]. This performance was superior to previously used approaches reported in literature. In AGES, face images are first encoded with AAM. AGES undertakes existence of multiple images of the same person at various ages or aging pattern of the face is similar in a given training dataset. This assumption may not be satisfied in aging datasets like Yamaha gender and age (YGA) [12]. Collecting face dataset with individuals'

face images at several ages with some image quality may not be possible. AAM cannot encode wrinkles on the face since AAM only encodes image gray values without spatial neighborhood information for texture pattern calculation. Intensities of individual pixels cannot describe local texture. This affects applicability of AGES for age and age-group estimation since single pixel values cannot represent local texture. Techniques like Gabor filter [67] may be appropriate to encode wrinkle features on elderly faces.

### 5.5 Age manifold

In age manifold, a common aging pattern is learned from images of many individuals and 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 make up points in a high-dimensional vector space. Age manifold learning face representation offers flexible means of face representation as compared to AGES [13]. Age manifold [68] can be used to learn aging pattern by learning low-dimensional aging pattern from several faces at every age. Individuals may have as low as one image at each age in the dataset which makes it simpler to collect enormous facial aging dataset. Scherbaum et al. [69] proposed statistical age estimation using manifold learning on 3D morphable model. Isosurfaces of non-linear support vector regression (SVR) function formed the manifold, and aging pattern was found by identifying a trajectory orthogonal to the isosurfaces. Discriminative subspace learning based on manifold criterion for low-

dimensional representation of aging manifold was proposed by Guo et al. in [31]. Coded face representation and age is learned by applying regression on aging manifold patterns. This approach consisted of two support vector regression (SVR) with one used for rough age-group estimation followed by refined age estimation within the initially obtained age group.

Given age-ordered image space  $X = \{x_i : x_i \in \mathbb{R}^D\}_{i=1}^n$  with image dimension  $D$  and a vector  $L = \{l_i : l_i \in \mathbb{N}^D\}_{i=1}^n$  of labels associated with the images in the image space, the objective is to learn a low-dimensional manifold in the embedded subspace, data distribution, and its representation  $Y = \{y_i : y_i \in \mathbb{R}^d\}_{i=1}^n$  with  $d \leq D$ , which is a direct mapping to  $X$ . Therefore, image space to manifold space projection can be modelled as  $Y = P(X, L)$ , where  $P(\cdot)$  denotes the projection function which can be linear or nonlinear. Figure 6 shows a simple nonlinear projection function that models an image space into a 2D age manifold. Respective ages are shown on the top-left corner of each image.

The objective of manifold embedding is to find  $n \times d$  matrix  $P$  that satisfies  $Y = P^T X$  or directly find  $Y$  where  $Y = \{y_1, y_2, \dots, y_n\}$ ,  $X = \{x_1, x_2, \dots, x_n\}$ ,  $P = \{p_1, p_2, \dots, p_n\}$ , and  $d \leq n$ . PCA, locally linear embedding (LLE), and orthogonal locality preserving projections (OLPP) are examples of techniques used for dimensionality reduction and embedding manifold. PCA finds the embedding that maximizes the projected variance  $P = \arg \max_{|p|=1} P^T p$  where  $S = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$  is the scatter matrix and  $\bar{x}$  is the mean of vector  $\{x_i\}_{i=1}^n$ . LLE technique seeks a nonlinear embedding in a neighborhood-preserving way by using local linear image class reconstruction symmetries while seeking local reconstruction optimal weights. Based

on linear preserving projections (LPP), OLPP technique produces orthogonal basis functions [70, 71] to find additional discerning information for embedding. LPP looks for the embedding that will preserve essential manifold structure by measuring distance information in local neighborhood. Affinity weights are defined as  $s_{ij} = \exp\left(\frac{|x_i - x_j|^2}{t}\right)$  where  $x_i$  and  $x_j$  are  $k$  nearest neighbors of each other; otherwise,  $s_{ij} = 0$  and  $s_{ij}$  is a symmetric matrix. LPP similarly defines diagonal matrix  $D(i, j)$  and a Laplacian matrix  $L = D - S$ . LPP represents age manifold well and performs better in age estimation compared to traditional PCA.

There is a connection between age manifold and subspace analysis for aging patterns. This technique finds embedded low-dimensional when each age is represented by many faces in the database. By using LPP for manifold embedding, age labels can be incorporated to the embedding process in a supervised manner which improves results compared to PCA embedding. Age manifold, unlike AGES [13], does not learn subject-specific aging pattern; rather, it uses all available ages from different individuals. However, age manifold requires a large dataset in order to satisfactorily learn the embedded manifold.

Huang et al. [72] proposed a multi-manifold metric learning (MMML) for face recognition based on image sets. In MMML, several person-specific distance metrics in different manifolds are learned by modeling each image set as a manifold minimizing intra-class variations and maximizing inter-class manifold variations. Figure 7 shows the multi-manifold metric learning.

MMML could be applied to age estimation by grouping images at the same age into one set and learn distance

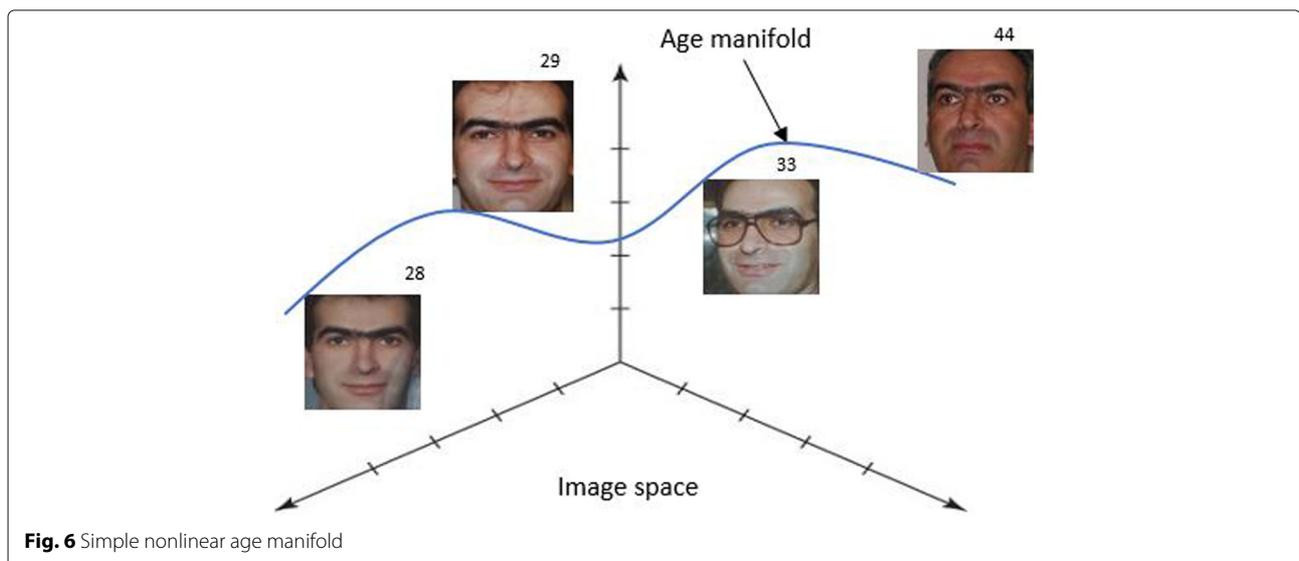


Fig. 6 Simple nonlinear age manifold

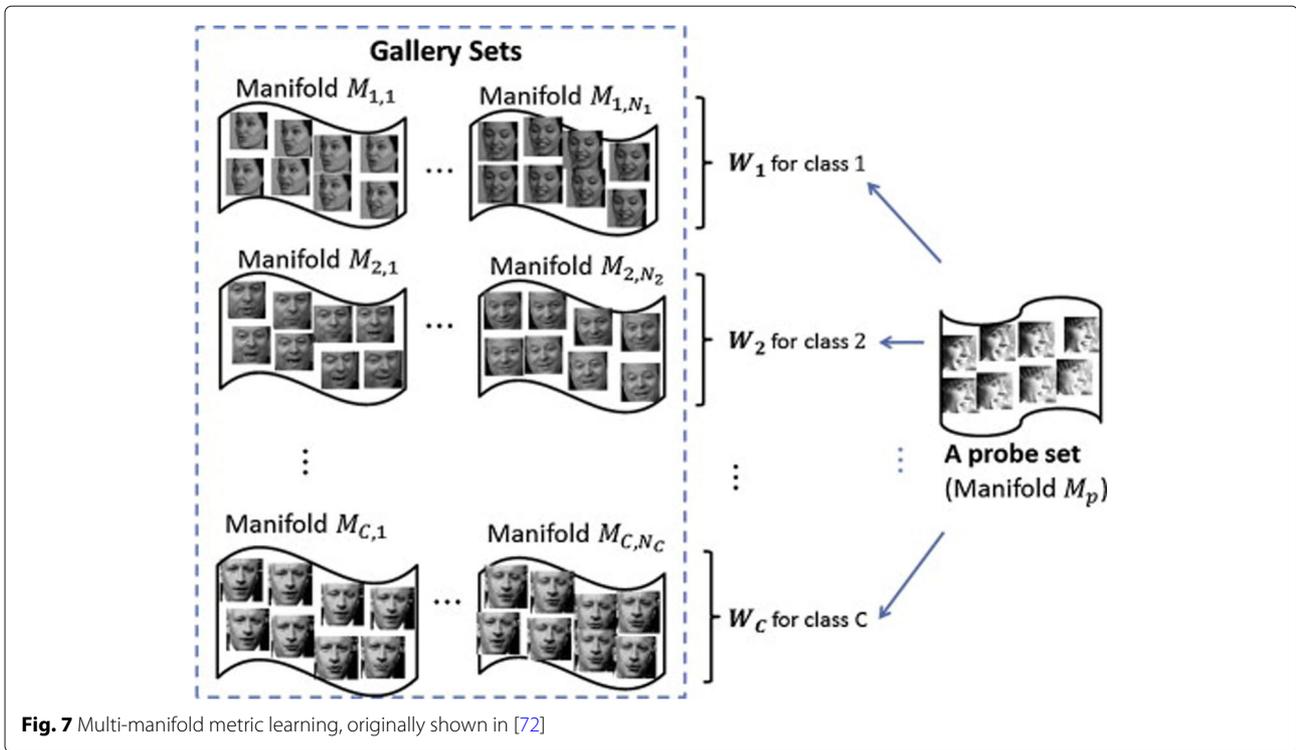


Fig. 7 Multi-manifold metric learning, originally shown in [72]

metrics between these sets. Each class (as shown in Fig. 7) could consist of images at a particular age. The limitation of age manifold models is that they are computationally intensive.

### 5.6 Appearance models

Appearance models mainly model facial appearance using texture, shape, and wrinkle features for age estimation, face recognition, face verification, and gender estimation among other tasks. Image is represented by vectoring both shape and texture [73]. Appearance models are more like AAM [64] that builds a statistical model using the shape and texture of the face. Both global and local texture, shape and wrinkle features are extracted and modelled for age estimation. Texture and shape have been used for age and gender estimation [74, 75]. Age estimation using appearance features can be improved by performing gender estimation prior since males and females exhibit varied aging patterns.

Given a set of facial images  $X = \{x_i : x_i \in \mathbb{R}\}_{i=1}^n$  and a vector of age labels  $X = \{l_i : l_i \in \mathbb{N}\}_{i=1}^n$ , facial features are extracted from vector  $\{x_i\}_{i=1}^n$  of images at a particular age. Every feature  $F_i$  has a one-to-one mapping with one of the age label  $l_i$ . After features are extracted and associated with age label, they are used for age estimation either using a regression model or classification. Effectiveness of LBP [76] in texture characterization has made it popular

in extraction of appearance features for age estimation. LBP has been used in [77] and achieved 80% accuracy in age estimation with nearest neighbor classifier and 80–90% accuracy with AdaBoost classifier [78]. Gao and Ai [79] used Gabor filter [67] appearance feature extraction technique for age estimation and reported better results compared to LBP technique. BIF [80, 81] is also used in appearance-based models as used in [82]. Using age manifold, BIF and SVM classifier, MAE of 2.61 and 2.58 years for females and males, respectively, can be achieved on YGA database [11]. This shows BIFs' superior performance in age estimation. Spatially flexible patch (SFP) proposed in [83, 84] is another feature descriptor that can be used for characterizing appearance for age estimation. Other techniques that can be used to build appearance models for age estimation are linear discriminant analysis (LDA) and principal component analysis (PCA). Detailed description of these techniques is presented in Section 6.

### 5.7 Hybrid models

What is the best modeling approach for age estimation? It is hard to certainly answer this question since each of the modeling approaches discussed have their inherent strengths and limitations. To get the answer to the question, one may try different modeling approaches on the representative images and compare their performance. By comparing different modeling approaches, strengths and limitations of each of the models can be

found. Modeling approaches that are complementary of each other can be combined to form a hybrid modeling approach. Hybrid age estimation modeling combines several modeling techniques to take advantage of the strengths of each technique used. By combining different modeling techniques, age estimation accuracies are expected to not only improve but also be robust. These models could be combined in a hierarchical manner or parallel and results from different models combined for final age estimation.

## 6 Aging feature extraction techniques

### 6.1 Gabor filters

Originally introduced by Denis Gabor in 1946 [67], Gabor filters have been extensively used for wrinkle, edge, and texture feature extraction due to its capability of determining orientation and magnitude of wrinkles [70]. Gabor filter has been regarded as the best texture descriptor in object recognition, segmentation, tracking of motion, and image registration [71]. Gabor features have been used in age estimation [27] and demonstrated to be an effective texture descriptor compared to LBP. Since wrinkles appear as edge-like components with high frequency, Gabor edge analysis technique has been commonly used for wrinkle feature extraction. Sobel filter [85, 86], Hough transform [74], and active contours [87] are among the most commonly used texture edge descriptors. Though edges in a face image also consist of noise such as beards, mustache, hairs, and shadows, to reduce the effect of this noise, [70] proposes use of predominant orientation of wrinkles to be considered in wrinkle feature extraction. 2D spatial domain Gabor is defined as:

$$g(x, y) = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi jWx \right] \quad (3)$$

where  $\sigma_x$  and  $\sigma_y$  are the standard deviations of the distribution along  $x$  and  $y$  axes, respectively, and  $W$  is the sinusoidal radial frequency.

The general equation for creating Gabor filter bank could be expressed as:

$$g_b(x, y) = a^{-m} g(\bar{x}, \bar{y}) \quad (4)$$

where  $\bar{x} = x \cos \theta + y \sin \theta$  and  $\bar{y} = -x \sin \theta + y \cos \theta$  where  $\theta_k = \pi \frac{(k-1)}{n}$ ,  $k = 1, 2, 3 \dots n$  where  $n$  is the number of orientations used and  $a^{-m}$  is filter scale for  $m = 0, 1, 2 \dots S$  for  $S$  scales. Redundancy in the frequency domain is prevented by designing Gabor wavelets as:

$$\sigma_u = \frac{\left( \left( \frac{U_h}{U_l} \right)^{\frac{1}{(s-1)}} - 1 \right) U_h}{\left( \left( \frac{U_h}{U_l} \right)^{\frac{1}{(s-1)}} + 1 \right) \sqrt{2 \ln 2}} \quad (5)$$

$$\sigma_v = \tan \left( \frac{\pi}{2k} \right) \left[ U_h - 2 \ln \left( \frac{\sigma_u^2}{U_h} \right) \right] \left[ 2 \ln 2 - \frac{(2 \ln 2)^2 \sigma_u^2}{U_h^2} \right]^{0.5}$$

where  $U_l$  and  $U_h$  denote lower and higher average frequencies, respectively, and  $W = U_h$ . We refer readers to [71] and [88] for more details on Gabor wavelets.

### 6.2 Linear discriminant analysis

Linear discriminant analysis (LDA) [89, 90] is a feature extraction technique that searches for features that best discriminate between classes. Given a set of independent features, LDA creates a linear combination of these features such that the largest mean differences between classes are achieved. LDA defines two measures: within class scatter matrix, given by

$$S_w = \sum_{j=1}^c \sum_{i=1}^{N_j} (x_i^j - \mu) (x_i^j - \mu)^T \quad (6)$$

where  $x_i^j$  is  $i$ th sample of class  $j$ ,  $\mu_j$  is the mean of class  $j$ ,  $c$  is number of classes, and  $N_j$  is the number of samples in class  $j$ , and between-class scatter matrix, given by

$$S_b = \sum_{j=1}^c (\mu_j - \mu) (\mu_j - \mu)^T \quad (7)$$

where  $\mu$  is the mean of all classes. The LDA main objective is to maximize between-class scatter matrix while minimizing within-class scatter matrix.

One way of doing this is maximizing the ratio  $\frac{\det|S_b|}{\det|S_w|}$ . Given that  $S_w$  is nonsingular, it has been proven [89] that this ratio is maximized when column vectors of projection matrix are the eigenvectors of  $S_w^{-1} S_b$ .  $S_w$  maximum rank is  $N - c$  with  $N$  samples and  $c$  classes. This therefore requires  $N = t + c$  samples to guarantee that  $S_w$  does not become singular, where  $t$  is the dimensionality of input data. The number of samples  $N$  is almost always smaller than  $t$ , making the scatter matrix  $S_w$  singular. To solve this problem, Belhumeur [91] and Swets and Weng [92] propose projecting input data to PCA subspace, to reduce dimensionality to  $N - c$ , or less, before applying LDA. PCA and LDA are widely used appearance feature extraction methods in pattern recognition [93]. Consequently, we adopt LDA for extraction of global face appearance features for age-group estimation.

### 6.3 Local binary patterns

Texture features have been extensively used in age estimation techniques [10]. Local binary pattern (LBP) is a

texture description technique that can detect microstructure patterns like spots, edges, lines, and flat areas on the skin [76]. LBP is used to describe texture for face recognition, gender classification, age estimation, face detection, and face and facial component tracking. Gunay and Nabiyevev [94] used LBP to characterize texture features for age estimation. They reported accuracy of 80% on FERET [77] dataset using nearest neighbor classifier and 80–90% accuracy on FERET and PIE datasets using AdaBoost classifier [78]. Figure 8 shows a sample of  $3 \times 3$  LBP operation.

Concatenating all 8 bits gives a binary number. The resulting binary number is converted to a decimal and assigned to center pixel as its LBP code.

Ojala et al. [95] found that when using eight neighbors and radius 1, 90% of all patterns are made up of uniform patterns. The original LBP operator had limitation in capturing dominant features with large-scale structures. The operator was latter extended to capture texture features with neighborhood of different radii [95]. A set of sampling pixels distributed evenly along the circle circumference centered at the pixel to be labeled defines the neighborhood. Bilinear interpolation of points that do not fall within the pixels is done to allow any radii and any number of sampling pixels.

Uniform patterns may represent microstructures as line, spot, edge, or flat area. Figure 9 shows microstructure pattern representation.

Ojala et al. [76] further categorized LBP codes as uniform and non-uniform patterns. LBP pattern with utmost two bitwise transition from 0 to 1 or 1 to 0 is categorized as a uniform pattern. For instance, 00000000, 00010000, and 11011111 patterns are uniform while 01010000, 11100101, and 10101001 are non-uniform patterns. For  $n$ -bit pattern representation, there is  $n(n - 1) + 2$  uniform patterns. Figure 9 shows LBP codes for sample uniform patterns in  $LBP(8, 1)$  neighborhood. In order to extract rotational invariant features using LBP, the generated LBP code is circularly rotated until its minimum value is obtained [96].

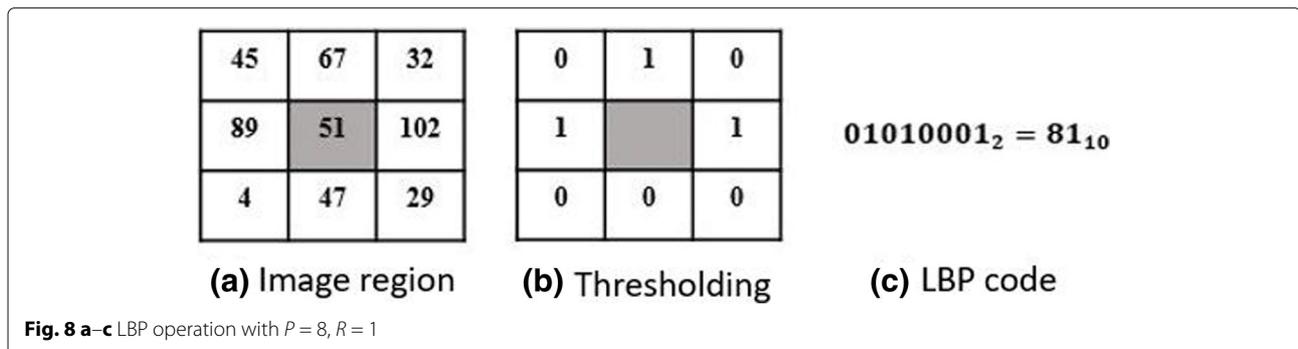
Extended LBP operator could capture more texture features on an image but still it could not preserve spatial information about these features. Ahonen et al. [97] proposed a technique of dividing a face image into  $n$  cells. Histograms are generated for each cell then concatenated to a single spatial histogram. Spatial histogram preserves both spatial and texture descriptions of an image. Image texture features are finally represented by histogram of LBP codes. LBP histogram contains detailed texture descriptor for all structures on the face image like spots, lines, edges, and flat areas. More details on the use of LBP on facial image analysis could be found in [76, 96–98].

### 6.4 Local directional pattern

Local binary patterns (LBP) [99] were found to be unstable to image noise and variations in illumination. Jabid et al. [100] proposed local directional pattern (LDP) which is robust to image noise and non-monotonic variations in illumination. Figure 10 shows robustness of LDP operator to noise compared to LBP.

LDP computes 8-bit binary code for each pixel in the image by comparing the edge response of each pixel in different orientations instead of comparing raw pixel intensities as LBP. Kirsch edge detector [101], Prewitt edge detector [102], and Sobel edge detector [103] are some of the edge detectors that can be used [104]. Among them, the Kirsch edge detector has been known to detect different directional edge responses more accurately than others because the Kirsch edge detector considers all eight neighbors [105]. Figure 11 shows Kirsch edge detector response masks (kernels) for eight orientations.

Given a center pixel in an image  $P(i, j)$ , 8-directional responses are computed by convolving the neighboring pixels,  $3 \times 3$  image region, with each of the Kirsch masks. For each center pixel, there will be eight directional response values. The presence of an edge or a corner will show high (absolute) response values in that particular direction. The interest of LDP is to determine  $k$  significant directional responses and set their corresponding bit value to 1 and set the rest of  $8 - k$  bits to 0. These binary



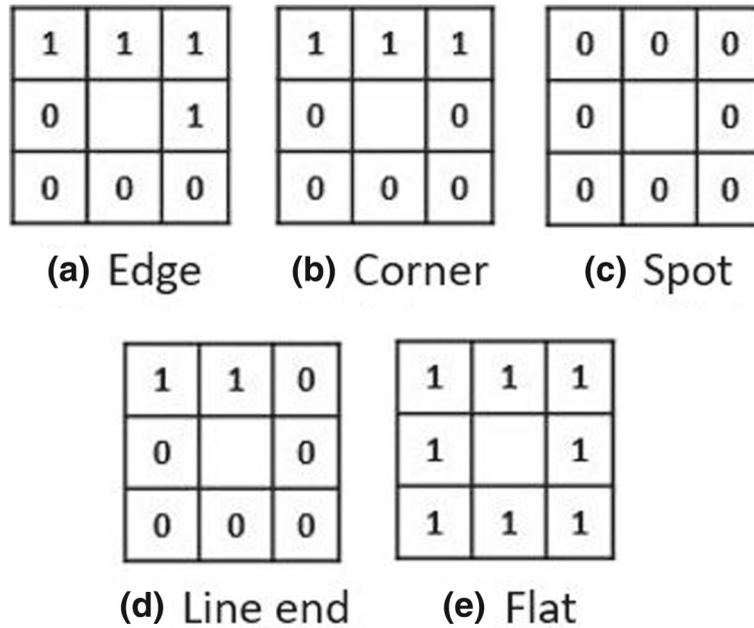


Fig. 9 a–e Microstructure pattern LBP code with  $P = 8, R = 1$

bits are converted to decimal and assigned to the center pixel. This process is repeated for all pixels in an image to obtain LDP representation of the image. Figure 12 shows the process of encoding an image using LDP operator.

Given an image region as shown in Fig. 12a, LDP response in the east direction is obtained by convolving the  $3 \times 3$  image region shown in Fig. 10 with the East  $M_0$  mask shown in Fig. 11 top-left corner as:

$$M_0 = (85 \times -3) + (32 \times -3) + (26 \times 5) + (10 \times 5) + (53 \times 5) + (38 \times -3) + (60 \times -3) + (45 \times -3) \quad (8)$$

$$= -399$$

The absolute values of the directional responses are arranged in descending order. For  $k = 3$  significant responses, the binary response bit for each of the eight neighboring pixels shown in Fig. 12b is calculated as:

$$LDP_k = \sum_{i=0}^{i=7} b_i ((m_i - m_k) \times 2^i) \quad (9)$$

$$b_i(a) = \begin{cases} 1, & \text{if } a \geq 0 \\ 0, & \text{if } a < 0 \end{cases}$$

where  $m_k$  is the  $k$ th significant directional response, example in Fig. 12  $m_k = |-399|$ , and  $m_i$  is response of Kirsch mask  $M_i$ .

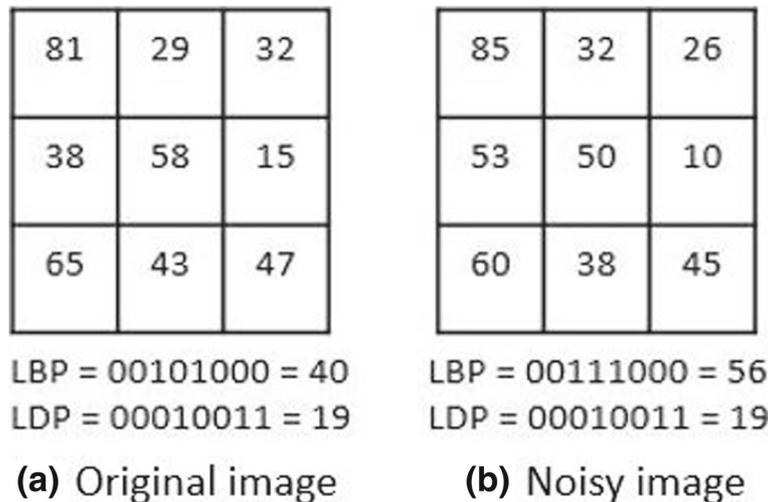
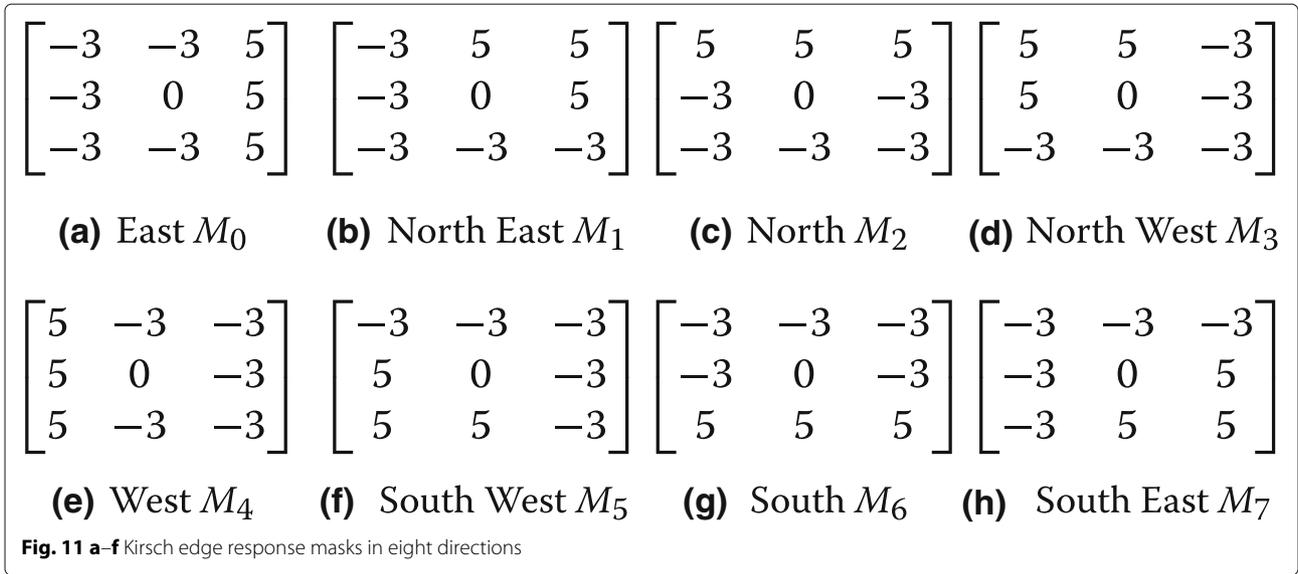


Fig. 10 a, b Robustness of LDP compared to LBP



For  $k = 3$ , LDP operator generates  $C_3^8 = \frac{8!}{3! \times (8-3)!} = 56$  distinct values in the LDP encoded image. The resultant histogram will have values between 0 and 56. A histogram  $H(i)$  with  $C_k^8$  bins can be used to represent the input image of size  $M \times N$  as:

$$H(i) = \sum_{m=0}^M \sum_{n=0}^N f(LDP_k(m, n), i) \tag{10}$$

$$f(p, i) = \begin{cases} 1 & \text{if } p = i \\ 0 & \text{if } p \neq i \end{cases}$$

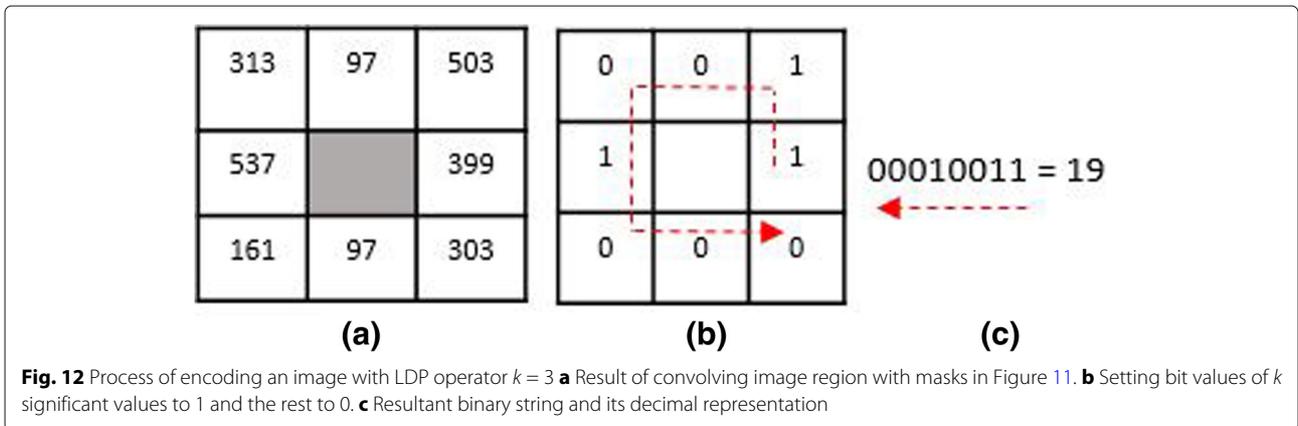
where  $f(p, i)$  is a logical function that compares if the LDP code at location  $p(m, n)$  of the LDP-encode image is equal to the current LDP pattern  $i$  for all  $i$  in the range  $0 \leq i \leq C_k^8$ . The resultant histogram has dimensions  $1 \times C_k^8$  and is used to represent the image. The resultant feature has spots, corners, edges, and texture information about the image [106].

### 6.5 Local ternary patterns

LBP is sensitive to noise and illumination especially in nearly uniform image blocks. Local ternary patterns (LTP) [107] seek to improve robustness of image features in a fairly uniform region. LTP extends LBP to a three-value code by comparing pixel values of the neighboring pixels with a preset threshold value  $\tau$ . Values that lie within  $\pm \tau$  are set to 0, values above  $\tau$  are set to +1 while values below  $\tau$  are set to -1. The thresholding function is defined as

$$f(x_i, x_c, \tau) = \begin{cases} 1 & \text{if } x_i \geq x_c + \tau \\ 0 & \text{if } |x_c - x_i| < \tau \\ -1 & \text{if } x_i \leq x_c - \tau \end{cases} \tag{11}$$

where  $\tau$  is a preset threshold,  $x_c$  is the value of the central pixel, and  $x_i$  for  $i = 0, 1, 2 \dots 7$  are the neighboring pixels of  $x_c$ . Although this extension makes LTP robust to



noise and encode more patterns, it is not easy to practically select an optimum  $\tau$  for all images in a dataset or for all datasets, and the resultant code is not invariant to pixel value transformations. LTP can encode  $3^8$  patterns. The LTP codes are split into its positive and negative parts and two histograms are generated, one for the negative part and the other for the positive part. These histograms are concatenated and used as feature descriptor for pattern recognition. Figure 13 shows LTP codes for a  $3 \times 3$  sample image region.

### 6.6 Gray-level co-occurrence matrix

Statistical moments of histogram intensities of an image are commonly used to describe texture of an image [108]. Use of histograms to describe texture results to texture descriptors that convey information about gray-level intensity distribution with no spatial relative information of pixel with each other. Haralick et al. [109] introduced gray-level co-occurrence matrix (GLCM) back in 1973.

GLCM describes image texture by comparing each pixel with its neighboring pixel at a specified distance and orientation. This technique extracts second-order statistical texture features from grayscale images. GLCM is a square matrix whose rows and columns are equal to the number of quantized gray levels,  $N_g$ . The entry  $p(i, j)$  is the second-order statistical probability for changes between gray level values  $i$  and  $j$  at a particular distance  $d$  and orientation  $\theta$ .

Supposed we have an  $N \times N$  image  $I(i, j)$ , with  $N_x$  columns and  $N_y$  rows.  $N_g$  is quantization of gray level appearing at each pixel in the image. Let the rows of the image be  $N_y = (1, 2, \dots, N_y)$ , the columns be  $N_x = (1, 2, \dots, N_x)$ , and set of  $N_g$  quantized gray levels be  $G_x = (1, 2, 3, \dots, N_{g-1})$ . The image can be represented as a function that assigns some gray level in  $G$  to each pixel or pair of coordinates in  $L_y \times L_x; G \leftarrow L_y \times L_x$ . Texture information is specified by GLCM matrix of relative frequencies  $C(i, j)$ . The value at  $GLCM(i, j)$  represents the number of occurrences of gray-level value  $i$  at reference pixel and gray-level value  $j$  at a neighbor pixel, a certain distance  $d$ , and orientation  $\theta$ . The probability measure can be defined as:

$$P_{d,\theta} = p(i, j) \tag{12}$$

where  $p(i, j)$  is defined as:

$$p(i, j) = \frac{GLCM(i, j)}{\sum_{i=0}^N \sum_{j=0}^N GLCM(i, j)} \tag{13}$$

The sum in the denominator represents total number of gray-level pairs  $(i, j)$  within the image and is bounded by  $N_g \times N_g$ . Dividing every pixel in the GLCM matrix with the denominator results into a normalized GLCM matrix. Figure 14 shows an example of calculating GLCM from an image region at distance 1 and angle  $\theta = 0^\circ$ , and Fig. 15 shows an example of calculating GLCM from an image region at distance 1 and angle  $\theta = 45^\circ$ .

The orientation of the neighbor pixel from reference pixel can be  $\theta = (0^\circ, 45^\circ, 90^\circ, 135^\circ)$ , and distance can vary from  $d = (1, 2, 3 \dots n)$  where  $n$  is any reasonable distance bounded by  $M_x$  and  $M_y$ .

Haralick et al. [109] defined 14 statistical features that can be used to describe texture. Table 1 shows some of the Haralick features used for texture description [110] where:

$$\mu_x = \sum_i \sum_j ip(i, j)$$

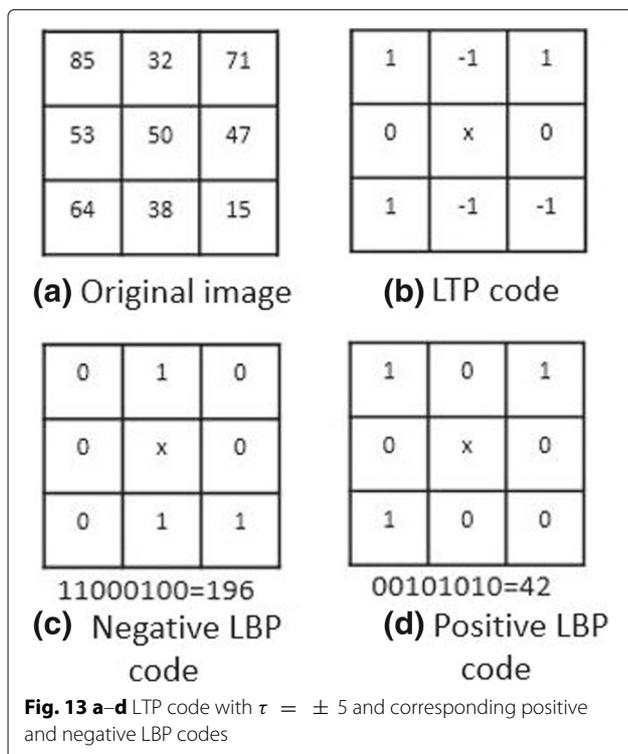
$$\mu_y = \sum_i \sum_j jp(i, j)$$

$$\sigma_x = \sqrt{\sum_i \sum_j (i - \mu_x)^2 p(i, j)}$$

and

$$\sigma_y = \sqrt{\sum_i \sum_j (j - \mu_y)^2 p(i, j)} \tag{14}$$

Harlick features have been successfully used in brain tumor classification [111], texture description [112], and remote sensing [113] among other fields. GLCM has not been investigated in aging feature extraction. Haralick features like homogeneity, variance, and correlation could be extracted from age-separated faces and used for age estimation.



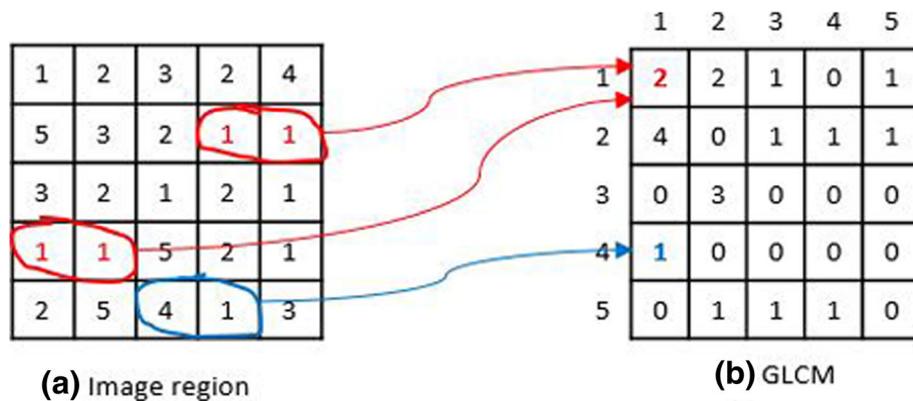


Fig. 14 a, b GLCM calculation with  $d = 1, \theta = 0^\circ$ . The figure shows how GLCM at angle 0 is calculated. The figure is supplied as glcm0.jpg

### 6.7 Spatially flexible patch

The spatially flexible patch (SFP) proposed in [83] and [84] is another feature descriptor that can be used for feature extraction for age estimation. SFP is effective for capturing local variations in facial appearance as one ages. SFP encodes local appearance and its spatial information. SFP solves the problem of local variations in appearance during aging since SFPs similar in appearance and slightly different in position can provide similar confidence for age estimation. By considering local patches and their spatial information, SFP can effectively characterize facial images with slight disorientation, occlusion, and head pose disparities. Another advantage of SFP is that it alleviates the problem of insufficient samples by enriching the discriminating characteristics of the feature vector.

### 6.8 Grassmann manifold

Grassmann manifold is the space  $G(k, n)$  of all  $k$ -planes through the origin in  $\mathbb{R}^n, k \leq n$  that generalizes real projective spaces [114]. It consists of a set of all  $k$ -dimensional subspaces of  $\mathbb{R}^n$ . To each  $k$ -plane  $v$  in  $\mathbb{R}^n$ , a matrix  $n \times k$  can be associated with orthogonal matrix  $Y$ , such that

columns of matrix  $Y$  form an orthonormal basis vector that spans the same subspace. Therefore, each  $k$ -plane  $v$  in  $G(k, n)$  is connected with a correspondence class of  $n \times k$  matrices  $YR$  in  $\mathbb{R}^{n \times k}$ , for  $R \in SO(k)$ , where  $Y$  is an orthonormal basis for the  $k$ -plane.  $G(k, n)$  is not a vector space, but points on  $G(k, n)$  can be projected onto the tangent space at mean-point, and standard vector-space methods can be used on tangent space. Geodesic distance between points on the manifold are used for classification or regression problems. Wu [115] used Grassmann manifold tangent-space regression approach for age estimation.

Grassmann manifold can be used in age estimation by representing each face by a deformation that warps an average face to a given face. This requires defining what an average face is and how to quantify the deformation between the average face and the given face. Average face can be represented by computing a mean point from all the (landmark) points on  $G(k, n)$ . This can be done by calculating Karcher mean [116]. Age estimation can be performed using the Grassmann nearest neighbor (GNN) classification approach. In GNN, Karcher mean is computed for every age. During testing, compare the Karcher

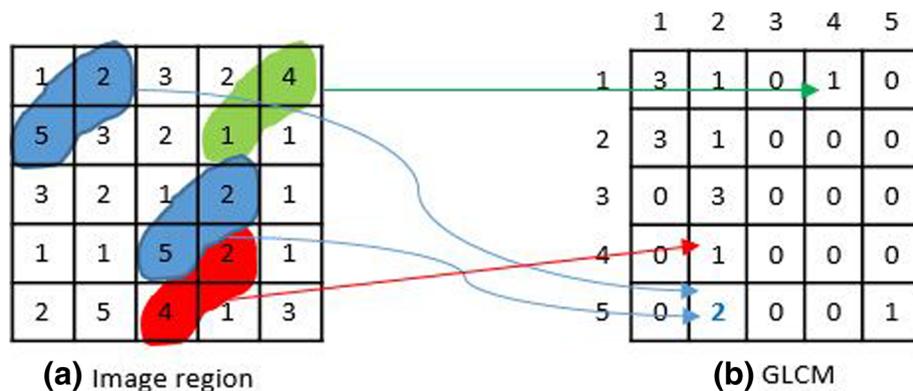


Fig. 15 a, b GLCM calculation with  $d = 1, \theta = 45^\circ$ . This figure shows how GLCM is calculated. The figure is supplied as glcm45.jpg file

**Table 1** Summary of Haralick features

Feature name	Feature description	Feature formula
Angular second moment (ASM)	Shows how uniform a texture is by measuring local homogeneity	$ASM = \sum_i \sum_j p(i,j)^2$
Energy (E)	Measures homogeneity	$E = \sqrt{\sum_i \sum_j p(i,j)^2}$
Contrast (C)	Shows variation in texture	$C = \sum_i \sum_j  i - j ^2 p(i,j)$
Dissimilarity (D)	Variation in texture	$D = \sum_i \sum_j  i - j  p(i,j)$
Homogeneity (H)	Uniformity of non-zero entries	$H = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$
Entropy (En)	Spatial disorder of texture	$En = \sum_i \sum_j p(i,j) \log(p(i,j))$
Correlation (Cr)	Linear relationship of texture	$Cr = \sum_i \sum_j p(i,j) \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y}$
Autocorrelation (ACr)	Measure repeating patterns	$ACr = \sum_i \sum_j (i \cdot j) p(i,j)$
Variance (V)	Measure of texture heterogeneity	$V = \sum_i \sum_j (i - \mu_x)^2 \cdot p(i,j) + \sum_i \sum_j (i - \mu_y)^2 \cdot p(i,j)$
Cluster shade (Cs)	Measure perceptual uniformity	$Cs = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i,j)$
Cluster prominence (Cp)	Measure image symmetry	$Cp = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i,j)$
Maximum probability (Mp)	Maximum co-occurrence	$Mp = \max p(i,j)$

mean of the probe image with the mean of every age using one defined distance on Grassmann manifold. The closest mean to the probe gives the target age.

**6.9 Biologically inspired features**

Biologically inspired features (BIFs) were first proposed in 1999 by Riesenhuber and Poggio (R and P model) [80]. These BIF features are derivative of primates feed-forward model of visual object recognition pipeline, referred to as HMAX model [117]. Primates are known to be able to recognize visual patterns with high accuracy. Recent studies in computer vision and brain cognition show that biologically inspired models (BIM) improve face identification performance [118], object recognition [119], and scene classification [120]. Visual cortex application in age estimation tasks saw some improvement in age estimation accuracies.

The visual model of primates contains alternating layers of simple (S) and complex (C) cell units. Complexity of these cells increase as layers advance from primary visual cortex (V1) to inferior temporal cortex (IT). In primary visual cortex, S units use a bell-shaped tuning function to combine input intensities to increase scale and orientation selectivity. Using MAX, STD, AVG, or any other pooling operation, C units pool inputs from S units, thereby introducing gradual invariance to scale, rotation, and translation.

Gabor functions [121, 122] are used to model simple cells (S) in the visual cortex of mammalian brains. Frequencies and orientation illustration in Gabor filters are alike to frequencies and orientations in human visual system. It is therefore thought that Gabor filter image analysis is similar to perception in visual system of humans. BIFs have demonstrated success in age estimation tasks [82, 123, 124]. BIF feature extraction encompass two layers of

computational units with simple cell units (S1) in layer one followed by complex cell units (C1) in the subsequent layer.

**S1 units—simple cells:** They represent the receptive field in primary visual cortex (V1) [121] which has basic attributes of multi-orientation, multi-frequency, and multi-scale selection [125]. S1 units are commonly described by a bank of Gabor filters [81]. Gabor filters are appropriate for modeling of cortical simple-cell receptive fields. 2D spatial domain Gabor is defined as:

$$G(x,y) = \exp\left(-\frac{X^2 + \gamma^2 Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right) \quad (15)$$

where  $X = x \cos \theta + y \sin \theta$  and  $Y = -x \sin \theta + y \cos \theta$  are angle of rotations of Gabor filters,  $\theta$  varies from 0 to  $\pi$ ,  $\gamma$  and  $\sigma$  are aspect ratio and standard deviation of the Gaussian envelop, respectively, and  $\lambda$  is the wavelength and determines spatial frequency  $1/\lambda$ .

Useful discriminating features are extracted using Gabor filters with different orientation and frequencies [126]. Consequently, previous studies [126, 127] suggest that spatial frequency processing is done in primary visual cortex. Spatial frequency analysis extracts discriminative features that are more robust to distortions [128]. Daugman [129] found that visual system in primates extracts information both in 2D spatial and frequency domains, and Shapley [38] proved that spatial frequency analysis help the brain understand an image.

**C1 units—cortical complex cells:** These units receive responses from S1 units and perform linear feature integration. C1 units represent complex cells that are shift invariant. Lampl et al. [130] proposed that spatial integration of complex cell in visual cortex can be described by a series of pooling operations. Riesenhuber and Poggio [80] demonstrated merits of using MAX pooling operator

compared to SUM while Guo et al. [82] showed that standard deviation (STD) pooling operator outperforms MAX operator. Cai et al. [125] improved on STD by using a cell grid of  $4 \times 4$  in normalization. The MAX operator returns maximum values at each index  $i$  of the two consecutive scale features. Given a feature at scale  $S_x$  and scale  $S_{x+1}$ , the maximum value  $F_i$  at index  $i$  is given by:

$$F_i = \begin{cases} S_x^i, & \text{if } S_x^i \geq S_{x+1}^i \\ S_{x+1}^i, & \text{if } S_x^i < S_{x+1}^i \end{cases} \quad (16)$$

where  $S_x^i$  and  $S_{x+1}^i$  are the filtered values at the position  $i$  of features from scale  $x$  and  $x + 1$  respectively.

Guo et al. [82] defined the STD operator to incorporate mean of values in a particular neighborhood. The STD operator was defined as:

$$STD = \sqrt{\frac{1}{n_s \times n_s} \sum_{i=1}^{n_s \times n_s} (F_i - \bar{F})^2} \quad (17)$$

where maximum value at index  $i$  between two consecutive  $S_1$  scales is represented by  $F_i$  and  $\bar{F}$  is the mean of filtered values within  $n_s \times n_s$  neighborhood. Given two  $N \times N$  features at scales  $S_x$  and  $S_{x+1}$ , STD operator with  $n_s \times n_s$  grid returns  $\lfloor N/n_s \rfloor \times \lfloor N/n_s \rfloor$  features. STD operator captures local texture and wrinkle variations which are significant for subtle age estimation.

Serre et al. [81, 131] extended the HMAX model [80] to include two layers,  $S_2$  and  $C_2$  for object recognition. In  $S_2$ , template matching is done to match patches from  $C_1$  layer with some pre-learned patches extracted from images. The  $S_2$  layer gets more selective intermediate features capable of discriminating between object classes. The  $S_2$  units are convolved over an entire image, and maximum response values of  $S_2$  are assigned to  $C_2$  units. Mutch and Lowe [132] extended the model in [81] by reducing the number of output units in  $S_1$  and  $C_1$  and picking features that are highly weighted by support vector machines (SVMs) [133].

## 7 Age estimation algorithms

Once aging features are extracted and represented, the subsequent phase is age estimation. Age estimation is a special pattern recognition task where age labels can be viewed as a class or a set of sequential value. When age labels are viewed as classes, age estimation is approached as a classification problem, whereas when age labels are viewed as sequential chronological series, regression approach is used for age estimation. Hybrid approach can also be employed for age estimation where both classification and regression techniques integrated, mostly hierarchically, to find the relationship between extracted feature vectors and age labels. We present an analysis of

existing approaches and suggest an effective approach in our opinion.

### 7.1 Classification

Lanitis et al. [23] explored the performance of nearest neighbor, artificial neural network (ANN), and quadratic function in age estimation tasks. Although the quadratic function used to relate face representations to face labels is a regression function, the authors referred to it as a quadratic function classifier [23]. The quadratic function reported MAE of 5.04, which was superior to MAEs reported by nearest neighbor. ANN and self-organizing maps (SOMs) reported better performance compared to quadratic function. The authors proposed clustering and hierarchical age estimation for improving performance. The error rates in the extended techniques reduced although evaluations were done on small datasets. Comparison between humans and computers in age estimation was also done and found that computers can estimate age almost as reliable as humans.

Ueki et al. [134] built 11 Gaussian models in low-dimensional 2DLDA and LDA feature space using expectation maximization (EM). Age-group estimation was determined by fitting probe image to each cluster and comparing the probabilities. They reported a higher accuracy, 82% male and 74% female, with wide age groups of 15 years as compared to 50% male and 43% female in age groups of a 5-year range. This demonstrates that this approach can only post better accuracies where age groups have wide ranges and hence not applicable in a narrow-range age-group estimation.

Fusing texture and local appearance, Huerta et al. [135] used a deep learning classification for age estimation. Using LBP [95], speeded-up robust features (SURF) [136], and histogram of oriented gradients (HOG) [137], he evaluated the performance of deep learning on two large datasets and achieved MAE of 3.31. Hu et al. [138] used Kullback-Leibler/raw intensities for face representation before using convolutional neural network (CNN) for age estimation. Their approach achieved MAE of 2.8 on FG-NET and 2.78 on MORPH II. This demonstrates that deep learning (deep neural networks or CNN) achieves better MAE compared to traditional classification methods.

### 7.2 Regression

Using 50 raw model parameters, Lanitis et al. [66] investigated linear, quadratic, and cubic formulation of aging function. Genetic algorithm is used to learn optimal model parameters from training face images of different ages. Quadratic and cubic aging function achieved better MAE 0.86 and 0.75, respectively, compared to 1.39 of linear function. This suggests that quadratic function offers the best alternative since its MAE was not significantly different from that of cubic function and it is not

computationally intensive as cubic function. Guo et al. [31, 139] used linear support vector regression (SVR) on age manifold for age estimation. They reported MAE of 7.47 and 7.00 years for males and females, respectively, on YGA dataset and MAE of 5.16 on FG-NET dataset. Yan et al. [140] formulated a regression problem for age estimation using semidefinite programming (SDP). The regressor was learned from uncertain nonnegative labels. They reported MAE of 10.36 and 9.79 years for males and females, respectively, on YGA. They further demonstrated that age estimation by SDP formulation achieves better results compared to ANN. The limitation of SDP is that it is computationally expensive especially when the training set is large.

Nguyen et al. [141] used a regression model for age estimation. The face image was represented by a multi-level local binary pattern (MLBP). Their approach achieved a MAE of 6.6. Guo and Mu [124] achieved a MAE of 4.0 by using BIF to model a regression model for age estimation. Using manifold of raw pixel intensities to represent face image, Lu and Tan [142] evaluated their regression model on MORPH II dataset and obtained a MAE of 5.2 for White ethnic group and 4.2 for Black ethnic group. Onifade et al. [143] applied a boosted regressor on age-rank local binary patterns (arLBP). They reported a MAE of 2.34 on FG-NET using LOPO validation protocol. Their approach demonstrated that age ranking with correlation of aging patterns across age groups improves performance of age estimation. Using raw pixel features, Akinyemi and Onifade [144] investigated ethnic-specific age group ranking for age estimation. This approach learns ethnic parameters in addition to the parameters learned in [143]. They evaluated this technique on FG-NET and FAGE datasets and reported a MAE of 3.19 years. Their findings show that incorporating ethnic parameters improves performance of age estimation approaches. This could be attributed to the fact that people in different ethnic groups age differently.

### 7.3 Hybrid approach

As discussed in the preceding sections, age estimation task can be approached as either a classification or a regression problem. To choose between the two, one may perform an experiment by selecting representative classifiers and regressors to compare their performance on the same dataset using the same features. Guo et al. [31, 139] compared SVM classifier to SVR regressor. This experiment showed that SVM performs better compared to SVR on YGA dataset with SVM achieving a MAE of 5.55 for females and 7.00 for males while SVR achieving 5.52 for females and 7.47 for males. It was also reported that SVM performed poorly on FG-NET compared to SVR (MAE 7.16 against 5.16 years). This experiment shows that classification approach to age estimation may perform better

or worse than regression approach depending on other aspects like quality of images in the dataset used, feature selection and feature extraction techniques used, and distribution of images across ages among other factors.

Combining classification and regression may result into robust and more accurate age estimation systems. Guo et al. [31, 139] therefore proposed age estimation using locally adjusted robust regression (LARR). LARR first performs regression using all existing aging images. Regression results are then used to limit a classifier with small search range. They demonstrated that better age estimation performance can be achieved by combining classification and regression schemes. By combining regression and classification, the MAE improved to 5.30 and 5.25 years for females and males, respectively, on YGA dataset and 5.07 on FG-NET dataset. The limitation of LARR method [139] is that it cannot automatically determine local search range for a classifier. The range is determined by heuristically trying different ranges and requires the user to experimentally choose the best solution. To automatically determine limited search range, Guo et al. [145] proposed a likelihood-based approach for combining classification and regression outcomes. Using a uniform distribution, regression results are transformed into likelihoods, then likelihoods from classification outcome are cut off by the uniform distribution. This further improved accuracies by achieving MAE 5.12 and 5.11 for males and females, respectively, on YGA and 4.97 on FG-NET.

Gunay et al. [146] represented aging face by fusing AAM, LBP, and Gabor features. They used an ensemble of three SVMs arranged in a hierarchical manner to build an age estimation model. The first step of their model was to perform age-group estimation by SVM classification. A linear regression was then performed to estimate age within the age group. Their approach achieved a MAE of 4.13 on FG-NET. These results show that feature and decision fusion used in a hybrid hierarchical age estimation can improve estimation errors compared to classification approaches.

Han et al. [147] performed hierarchical demographic estimation and compared machine and human performance. They extracted BIF features and demographic informative features using a boosting algorithm. They then perform a hierarchical age estimation using between-group classification followed by within group regression. Evaluating this technique on MORPH II and FG-NET, they achieved MAE of 3.6 and 3.8 on MORPH II and FG-NET datasets, respectively. Choi et al. 2011 [70] used AAM, Gabor, and LBP to represent face image. Their hybrid age estimation model achieved a MAE of 4.7 on FG-NET, 4.3 on PAL, and 4.7 on BERC datasets.

Hybrid approach to age estimation demonstrates better performance compared to regression and classification

when used alone. To combine classification and regression, one may test extracted features on both techniques separately before combining them. Arrange regression and classification in an arbitrary hierarchical order and compare performance when regression is done before classification and when done after classification.

**8 Facial aging databases**

Precise age and age-group estimation requires a database with good quality facial images at different ages. It is hard to collect a large aging database with a series of chronometric images from an individual. Age and age-group estimation often uses databases early collected and published. Brief descriptions of these databases are found in [11]. Table 2 gives the summary of some of the aging databases available.

FG-NET, MORPH, and web-collected Gallagher’s databases are publicly available. Other databases can be found by contacting the owners. MORP, Ni’s, YGA, LHI, and Gallagher’s web-collected databases are large databases and well suited for regression-based age estimation using statistical algorithms like AAM and age manifold. FG-NET is a suitable database for evaluations with several age estimation methods like AGES. AI & R, LHI, and Iranian datasets comprise comparatively high resolution 2D face images. Other datasets stated here were not extensively used but may be appropriate for some application areas.

**8.1 FG-NET aging database**

FG-NET [21] contains 1002 both color and grayscale images of 82 individuals from age 0 to 69 years. Each individual has averagely 12 images. Images are collected from

multi-race subjects and have great inconsistencies in head pose, facial expression, and illumination. Some images have adverse condition because they were scanned. There are 68 landmark points provided which can be used to model facial shape. Age features can be modelled as AAM or as appearance model using texture and wrinkle features.

**8.2 MORPH database**

MORPH [148] is a publicly available aging database created by the Face Aging Group at the University of North Carolina. This dataset is split into two sets. Album 1 has 1724 images collected between 1962 and 1998 from 515 individuals. Images in this dataset range from 27 to 68 years. There are 1430 images for males and 294 images for females with age gap ranging from 46 days to 29 years. Set 2 contains 55,134 images of 13,000 individuals collected over 4 years. Both albums contain metadata for race, gender, date of birth, and date of acquisition. The eye coordinates of the dataset can be requested. A commercial version of album 2 contains a larger set of images collected over a longer time span and includes information like the height and weight of individual.

**8.3 Yamaha gender and age (YGA) database**

YGA [12, 68] database has 8000 high-resolution colored images of 1600 individuals consisting of 800 males and 800 females of Asian race, aged between 0 and 93 years. Each subject has approximately five nearly frontal face images at the same age and a label of his or her approximated age. The images have high variations in expression, illumination, and facial expression. Haar cascade face

**Table 2** Summary of facial aging databases

Database	No. of subjects	Database size	Age range (years)
FG-NET [21]	82	1002	0–69
MORPH [148]	13,618	55,134	27–68
Yamaha gender and age (YGA) [12, 68]	1600	8000	0–93
Waseda human-computer interaction technology [134]	26,222	5500	3–85
AI & R Asian [150]	17	34	22–61
Burt’s Caucasian Face database [151]	–	147	20–62
Lotus Hill Research Institute (LHI) database [152]	–	50,000	9–89
Human and object interaction processing (HOIP) [11]	300	306,600	15–64
Iranian face database [153]	616	3600	2–85
Gallagher’s Web-Collected database [4]	–	28,231	0–66
Ni’s Web-Collected database [154, 155]	–	219,892	1–80
Kyaw’s Web-Collected Database [156]	–	963	3–73
BERC database [214]	95	5910	3–83
3D morphable database [69, 157]	438	–	–

detector [149] is used to crop and resize images to  $60 \times 60$  grayscale patches.

#### 8.4 WIT-DB database

Waseda human-computer interaction technology [134] dataset consists of 12,008 face images of 2500 females and about 14,214 images of 3000 males from the Japanese race, with age ranging between 3 and 85 years. The ages are arranged in 11 non-overlapping age groups. The dataset has wide variations in illumination on unoccluded frontal view faces with neutral facial expression. Face images are cropped and resized to  $32 \times 32$  grayscale patches.

#### 8.5 AI & R Asian face database

AI & R Asian [150] dataset contains images of different expressions, ages, poses, and illuminations. There are 34 frontal-view images collected from 17 individuals with ages ranging from 22 to 61 years. There are averagely two images per individual making this database not suitable for age or age-group estimation.

#### 8.6 Burt's Caucasian face database

This was collected and used in [151] by Burt and Perrett to investigate visual cues to age by blending color and shape of facial components. The database contains 147 images of European males aged between 20 and 62 years. Faces had neutral expression with beards shaved with no glasses and makeups. There are 208 landmark points placed manually in standardized positions. These points can be used to encode facial shape.

#### 8.7 LHI face database

Lotus Hill Research Institute (LHI) database contains 50,000 images of Asian adults at different ages. The images have slight dissimilarities in pose and lighting. Part of this database was used in [152] by Suo et al. to model a hierarchical face model for age estimation. The part used consists of 8000 color images of individuals aged between 9 and 89 years with one image per person. This database could not be appropriate for subject-based age estimation since it does not provide multiple face images of the same individual at different ages.

#### 8.8 HOIP face database

Human and object interaction processing (HOIP) database consists of 306,600 images of 300 individuals aged between 15 and 64 years. The database is divided in 10 age groups. Each age group has got 30 subjects, 15 females and 15 males [11].

#### 8.9 Iranian face database

Iranian face database [153] has 3600 color images from 616 individuals aged between 2 and 85 years of which 487 are males and 129 females. The images have variations in

pose and facial expression. At least one image with glasses was also taken. Majority of the images are of subjects in the age group of 1–40 years. This database can therefore be appropriate in modelling aging and age estimation in formative and middle-age years.

#### 8.10 Gallagher's web-collected database

This database was collected by Gallagher and Chen [4] from Flickr.com image search engine. The database has 28,231 faces in 5080 images. It divided into seven age groups as 0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+. This dataset is suitable for age-group estimation although the age groups are wider in older ages.

#### 8.11 Ni's web-collected database

This database was collected from the web by Ni et al. [154, 155] using Google.com and Flickr.com image search engines. The database has 219,892 faces in 77,021 images with age range between 1 and 80 years. This is the largest aging database ever reported. The wide age range in this database makes it suitable for age estimation in child, adult, and old age groups.

#### 8.12 Kyaw's web-collected database

This database was collected from the web by Kyaw et al. [156] using API services provided by Microsoft Search Engine Bing. The images in the collected database are aligned with eye corner points captured manually and cropped to  $65$  by  $75$  patches. The database contains 963 images divided in four age groups of 3–13, 23–33, 43–53, and 63–73. The database is not appropriate for age-group estimation since there are missing images between age groups.

#### 8.13 BERC database

BERC database [70] was collected by the Biometric Engineering Research Center (BERC). The database contains images of 390 subjects with age ranging from 3–83 years. Images are of high resolution  $3648 \times 2736$  pixels. There are no variations in light and facial expression on all the images, and subjects are uniformly distributed with respect to age and gender. These make the database suitable for age estimation, although it is comparatively small.

#### 8.14 3D morphable database

The database contains 3D scans of 100 male adults and 100 female adults' faces and 238 teenage faces aged between 8 and 16 years consisting of 113 females and 125 males [69, 157]. All faces were without makeup, accessories, and facial hair. In 3D morphable face models, individual faces are represented as face vector in 3D. By caricaturing texture and shape feature vectors, the model can transform one's face. As one ages, each face will transform along a curved trajectory in a high dimensional

space. Faces are represented by shape and texture vectors such that each linear combination of different faces is a new realistic face.

### 8.15 Summary

FG-NET, MORPH, and web-collected Gallagher's databases are publicly available. Other databases can be found by contacting the owners. MORP, Ni's, YGA, LHI, and Gallagher's web-collected databases are large databases and well suited for regression-based age estimation using statistical algorithms like AAM and age manifold. FG-NET is a suitable database for evaluations with several age estimation methods like AGES. AI & R, LHI, and Iranian datasets comprise comparatively high-resolution 2D face images. Other datasets stated here were not extensively used but may be appropriate for some application areas.

## 9 Age estimation evaluation protocols

Evaluation protocol determines system test, criteria for test data selection, and system performance measure. A good validation strategy should be independent of training data and representative of the population from which it has been drawn [158]. Age estimation technique needs to be validated using previously unseen data to avoid overfitting age estimation technique and improve its generalization capability. Cross-validation is a popular strategy for age estimation evaluation. In cross-validation, data is split into two subsets; one segment is used to train or learn age estimation model and the other segment is used to validate or evaluate the model. In classic cross-validation, training and validation datasets must cross-over in consecutive rounds such that every data point has equal chance of being validated or evaluated against the other. The basic form of validation is holdout.

Holdout strategy is the simplest and computational efficient strategy [159] used for validating age estimation techniques. The dataset is randomly split into two sets: training subset and validation subset. Commonly, training subset consists of two thirds of the original data, and the remaining one-third samples constitute validation subset. Age estimation model is then fitted using the training subset and validated on the test subset. In this strategy, the model is trained and validated only once. Although this method is preferred and takes a shorter time to compute, its evaluation depends on the data in respective subsets and results into high variance hence making this strategy give different evaluation results depending on how the dataset is divided [160]. Another validation strategy commonly used is repeated random sub-sampling (RSS) [161, 162]. In RSS validation technique, the holdout strategy is iterated a number of times and results averaged. The dataset is randomly split into two subsets (train and validation) with a fixed number of samples for each phase

of validation. For each data split, age estimation model is retrained on train subset and validated using test subset. The advantage of this strategy over  $k$ -fold validation is that the size of training and validation is independent to the number of validation iterations. However, this strategy has a limitation such that some samples may never be selected for validation while other samples may be selected repetitively leading to overlapping of validation subsets [163]. But with a significantly large number of iterations done, RSS is likely to achieve better results as  $k$ -fold validation [164].

Cross-validation [163] is a standard statistical technique used for model generalization ability with wide application in classification and regression problems [165]. It involves dividing dataset into two subsets, one subset is used to train an estimator while the other subset is used to test an estimator [166]. Cross-validation is used to assess how a model generalizes to initially unseen data [163, 167]. Cross-validation strategies can be categorized into two: (i) exhaustive (compute all possible ways of data splitting) and (ii) non-exhaustive (does not compute all possible ways on data splitting). Exhaustive cross-validation algorithms include leave-one-out (LOO) and leave-p-out (LPO) while non-exhaustive include  $k$ -fold and repeated random subsampling (RSS) [160, 168]. Cross-validation [169] consists of averaging multiple holdout validation results from different subsets of data.

$k$ -fold cross-validation is the basic form of cross-validation. Other forms of cross-validation are just but special cases of  $k$ -fold cross-validation or involve repeated rounds of  $k$ -fold validation. In  $k$ -fold cross-validation [169], original data is randomly split into  $k$  equal subsets. Then,  $k$  iterations of training and validation are performed such that in every iteration, a different fold of data is reserved for validation while the remaining  $k - 1$  are used to learn a model. The estimated error is the mean of all validation errors. Standard deviation of these errors can be used to approximate the confidence range of the estimate. The main advantage of  $k$ -fold cross-validation is that eventually all samples will be used for both learning and validating models. The common value of  $k$  used in various techniques is 10 as a compromise between efficiency and accuracy. A stratified cross-validation is commonly used in order to improve accuracy of the estimation [163].

Leave-one-out (LOO) [166, 169, 170] is a special type of cross-validation that given a dataset with  $C$  classes,  $C - 1$  validation experiments are performed. For each experiment, data from  $C - 1$  classes is used for training and data from one class that was left out is used for validation. Therefore, given a dataset of  $S$  subjects from age  $0 \rightarrow A_n$ , LOO cross-validation will perform  $S - 1$  validation experiments. In each experiment  $i$ , facial images of subject  $S_i$  are used for validation while images of the rest  $S - 1$  subjects

are used for learning a model. In this approach, images of each subject will be used for both training and validation. This way, the technique is validated in the same way as its application scenario where the subject whose age is to be estimated is previously unseen in the system. Although LOO is almost unbiased, it may give unreliable estimates due to its high variance [171]. Leave- $p$ -out (LPO) [172] with  $p \in \{1, 2, 3, \dots, n-1\}$  successively leaves out every possible subset of  $p$  data samples to be used for validation. In age estimation, given a set of images of  $N$  subjects, LPO can be used by leaving out images of  $p$  where  $p \leq (N-1)$  subjects to be used for validation and use images of  $N-p$  subjects for training. Elisseeff and Pontil [173] showed that LPO cross-validation is less biased compared to LOO. LPO will have  $\binom{n}{k}$  iterations where  $n$  is the number of images. These iterations are almost always much higher compared to  $n-1$  iterations in LOO, leading to high computation time. LPO with  $p=1$  is same as LOO. LOO and LPO are exhaustive cross-validation strategies compared to other methods. Further information on LPO can be found in [174]. Detailed information on cross-validation can be found in [172] and [175].

Bootstrap is a strategy introduced by Efron and Tibshirani [176, 177]. Bootstrap is commonly used when working on a small dataset [159]. In this strategy, a bootstrap set is created by uniformly sampling, with replacement,  $n$  instances from the original data to make a training set. The remaining samples not selected are used as testing set. The value  $n$  of selected samples is likely to change from fold to fold. Since data is sampled with replacement, the probability of any data sample not being selected is given by  $(1 - \frac{1}{n})^n \approx e^{-1} \approx 0.368$ . Chances of a data sample being selected into a train set is  $(1 - 0.368) = 0.632$ . Therefore, the expected number of distinct samples appearing in the train set is  $0.632 \times n$ . Since error estimate obtained by using test data will be too pessimistic (since only 62.3% of instances are used for training), error is calculated as  $error = 0.632 \times e_0 + 0.368 \times e_{bs}$  where  $e_0$  is rate of error obtained from bootstrap sets not having the instance being predicted (test set error) and  $e_{bs}$  is the error obtained on bootstrap sets themselves, both averaged over all data samples and bootstrap samples. Estimate accuracy is directly proportional to number of times the process is repeated. More details on bootstrap validation technique can be found in [177]. Bootstrapping increases the variance that can occur in each fold which makes this strategy more realistic of the real application situation [177]. This validation strategy is rarely used in age estimation.

In most cases, a dataset is split into three subsets: validation subset, training subset, and testing subset [167]. In this approach, the validation subset is used to tune the system to determine the termination point of the training phase when overfitting starts occurring on the

training subset. The testing subset is used to validate the trained model using data samples not initially in validation and training subsets. Kiline and Uysal [164] proposed a technique of splitting the dataset with samples from specific subjects rotationally left out of training and validation sets. Budka and Gabrys [158] proposed a density-preserving sampling (DPS) technique that eliminates the need for repeating error estimation procedures by dividing the dataset into subsets that are guaranteed to be representative of the population the dataset is drawn from. These new proposed approaches of model validation could be experimented in age estimation problem and results compared with other common methods. Cross-validation and bootstrap strategies are commonly used when one has limited data such that holdout strategy cannot be sufficient for data representativeness in both training and test sets. With abundant data with stable distribution over time, single stratified random split is able to provide required representativeness [158].

For purposes of comparing the performance metric of two or more learning algorithms, Salzberg [178] proposed the use of  $k$ -fold cross-validation followed by appropriate hypothesis testing instead of comparing their average accuracies. This strategy can be used to compare two age estimation techniques.

In each iteration of validation, absolute error (AE) for each estimated age is defined as:

$$AE = |a_i - \bar{a}_i| \tag{18}$$

where  $a_i$  is the ground truth age and  $\bar{a}_i$  is the estimated age. After all validation iterations, mean absolute error (MAE) is defined as the average of all absolute errors between estimated and ground truth age as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |a_i - \bar{a}_i| \tag{19}$$

where  $N$  is the total number of test images,  $a_i$  is the ground truth age of image  $i$ , and  $\bar{a}_i$  the estimated age of image  $i$ . Although this performance evaluation is commonly used, it does not give age estimation performance for specific age but rather gives general performance of the technique for all ages. This approach could be slightly modified such that it gives MAE for every age and general MAE of the technique.

Given a set of testing images  $a_1^{n_1}, a_2^{n_2} \dots a_k^{n_k}$  belonging to  $k$  ages to be estimated with  $n_i$  representing number of test images known to belong to age  $a_i$ , MAE for every age can be defined as:

$$MAE_k = \frac{1}{n} \sum_{i=1}^n |a_k - \bar{a}_i| \tag{20}$$

where  $\bar{a}_i$  is the estimated age for image  $i$  of age  $a_k$  and  $n$  is the number of test images belonging to age  $a_k$ . This

will give age-specific performance of age estimation technique. Overall, MAE can be found by summing all the MAE for all ages tested and dividing by the sum of the number of test images in each age as:

$$MAE_{TOTAL} = \sum_{i=1}^k \frac{(MAE_i \times n_i)}{N} \quad (21)$$

where  $N = n_1 + n_2 + \dots + n_k$ .

Age estimation technique performance is evaluated based on MAE. The smaller the MAE, the better the age estimation performance. MAE only shows average performance of the age estimation technique. MAE is the appropriate measure of age estimation when the training data has missing images [10]. The overall accuracy of the estimator is given by cumulative score (CS) [12, 31] which is defined as:

$$CS(x) = \frac{N_{e \leq x}}{N} \times 100\% \quad (22)$$

where  $N_{e \leq x}$  is the number of images on which the age estimation technique makes an absolute error no higher than  $x$  years error tolerance and  $N$  is the total number of test images.

In age-group estimation, the age-group label represents a range of ages; hence, the cumulative scores are compared at error level 0, i.e., the percentage of exactly correct age-group estimation. Therefore, the CS equation becomes:

$$CS(x) = \frac{n_x}{N_x} \times 100\% \quad (23)$$

where  $n_x$  is the number of test images correctly recognized as belonging to age group  $x$  and  $N_x$  is the total number of test images in age group  $x$ . Therefore, CS is used as an indicator of accuracy of age-group estimator [13]. CS is a useful measure of performance in age estimation when the training dataset has samples at almost every age [11]. MAE is a good evaluation technique when the training set has a lot of missing ages. However, in age estimation, both MAE and CS are used since different techniques, datasets, and systems may be extremely imbalanced or skewed for evaluation.

## 10 A review of age estimation studies

### 10.1 Age-group estimation

Global, local, and hybrid features have been previously used in age and age-group estimation. Ramanathan et al. [179] present a recent survey in automated age estimation techniques.

Age group is a range of ages. Persons whose real age are within the defined ranges are said to be in the same age group. Significant amount of research has been done to automatically extract visual artifacts from faces and group persons in respective age groups. Kwon and Lobo [87] estimated age group based on anthropometry and

density of wrinkles. They separated adults from babies using distance ratios between frontal face landmarks on a small dataset of 47 images. They also extracted wrinkle features from specific regions using snakes. Young adults were differentiated from senior adults using these wrinkle indices. Baby group classification accuracy was lower than 68%, but overall performance of their experiments was not reported. Furthermore, ratios used were mainly from baby faces. Horng et al. [85] used geometric features and Sobel filter for texture analysis to classify face images into four groups. They used Sobel edge magnitude to extract and analyze wrinkles and skin variance. They achieved an accuracy of 81.6% on subjectively labeled age-groups.

Ramanathan and Chellappa [59] computed eight distance ratios for modelling age progression in young faces like 0 to 18 years. Their objective was to predict one's appearance and face recognition across age progression. Using 233 images of which 109 were from FG-NET aging dataset, and the rest from their private dataset, they reported improvement in face recognition from 8 to 15%. Dehshibi and Bastanfard [20] used distance ratios between landmarks to classify human faces in various age groups. Using a back propagation neural network with distance ratios as inputs, they classified face images into four age groups of 15, 16–30, 31–50, and above 50. Using a private dataset, they reported 86% accuracy. Thukral et al. [180] used geometric features and decision fusion for age-group estimation. They achieved 70% overall performance for 0–15, 15–30, and above 30 age groups. Farkas et al. [181] used 10 anthropometric measurements of the face to classify individuals in various ethnic groups. They analyzed these measurements and identified ones that contribute significantly to diversity in facial shape in different ethnic groups. They also found that horizontal measurements differed between ethnic groups than vertical measurements.

Tiwari et al. [182] developed a morphological-based face recognition technique using Euclidean distance measurements between fiducial facial landmarks. Using morphological features with back propagation neural network, they reported superior recognition rate than performance of principal component analysis (PCA) [90] with back propagation neural network. This technique recognized faces but it was independent of aging factor due to variations in these distances as one ages. This signifies that distances between facial landmarks differ at different age, especially in young age-groups, and therefore, it could be used in age estimation. Gunay and Nabiye [94] used spatial LBP [76] histograms to classify faces into six age groups. Using nearest neighbor classifiers, they achieved accuracy of 80% on age groups  $10 \pm 5, 20 \pm 5, 30 \pm 5, 40 \pm 5, 50 \pm 5$ , and  $60 \pm 5$ . In [146], Gunay and Nabiye trained three support vector machine (SVM) models for age-group estimation using

AAM [64], LBP, and Gabor filter [67] features. They fuse decisions from these classifiers to obtain final decision. Although they reported 90% accuracy of subsequent age estimation, overall performance of age-group estimation was not reported.

Hajizadeh and Ebrahimnezhad [183] represented facial features using histogram of oriented gradients (HOG) [137]. Using probabilistic neural network (PNN) to classify HOG features extracted from several regions, they achieved 87% accuracy in classifying face images into four groups. Liu et al. [184] build a region of certainty (ROC) to link uncertainty-driven shape features with particular surface features. Two shape features are first designed to determine face certainty and classify it. Thereafter, SVM is trained on gradient orient pyramid (GOP) [185] features for age-group classification. Testing this method on three age groups, 95% accuracy was reported. They further used GOP in [186] with analysis of variance (ANOVA) for feature selection to classify faces into age groups using linear SVM [187] to model features from the eyes, nose, and mouth regions. They achieved 91% on four age groups on FG-NET dataset and 82% on MORPH dataset. It was also found that the overall performance of age estimation decreases as the number of age groups increase. This is because the number of images in each age group reduces drastically as the number of groups increase.

Lanitis et al. [66] adopted AAM to represent face image as a vector of combined shape and texture parameters. They defined aging as a linear, cubic, or quadratic function. For automatic age estimation, they further evaluated quadratic function, nearest neighbor, and artificial neural network (ANN) in [23]. They found that hierarchical age estimation achieves better results with quadratic function and ANN classifiers. Although AAM has been extensively used, it does not extract texture information. This problem is avoided by using hybrid feature extraction techniques to combine both shape and texture features for age and age-group estimation.

Sai et al. [188] used LBP, Gabor, and biologically inspired features for face representation. They used extreme learning machines (ELM) [189] for age-group estimation. Their approach achieved accuracy of about 70%. Using LBP and a bank of Gabor filters, Wang et al. [190] classified images into four age groups. They used SVM, error-correcting output codes (ECOC) and AdaBoost for age-group estimation. Table 3 shows the summary of age and age-group estimation studies.

## 10.2 Age estimation

Age is a real number that signifies the number of years elapsed since one's birth to a point in life. Age estimation is the process of estimation one's actual age using visual artifacts on the face. These visual artifacts are extracted and used to estimate one's age.

Lanitis et al. [66] adapted active appearance model (AAM) for aging face by proposing aging function. They defined age as a function  $age = f(b)$  to cater for age-introduced variations. In this function, age is the real estimated age of a subject,  $b$  consists of 50 AAM-learned-parameters feature vector, and  $f$  is the aging function. They performed experiments on 500 images of 60 individuals of which 45 subjects had images at different ages. Focusing on small age variations, they demonstrated that simulation of age improves performance of face recognition from 63 up to 71% and from 51 to 66% when training and testing datasets are used interchangeably.

Adopting aging pattern subspace (AGES), Geng et al. [13, 26] proposed automatic age estimation using appearance of face images. Evaluating AGES on FG-NET aging database, they used 200 AAM parameters to characterize each image for age estimation. They reported 6.77 years mean absolute error (MAE). Fu and Huang [12] used age-separated face images to model a low-dimensional manifold. Age was estimated by linear and quadratic regression analysis of feature vectors derived from respective low-dimensional manifold. The same approach of manifold learning was used by Guo et al. in [31]. They extracted face aging features using age learning manifold scheme and performed learning and age prediction using locally adjusted regressor. Their approach reported better performance than support vector regression (SVR) and SVM.

Guo et al. [31] used locally adjusted robust regression (LARR) to estimate age. Evaluating their approach on a large dataset, they reported MAE of 5.30 and 5.07 years on FG-NET. Guo et al. [82] further proposed age estimation using biologically inspired features (BIF) [80, 81]. BIF features with support vector machine (SVM) achieved MAE of 4.77 years on FG-NET aging dataset and 3.91 and 3.47 years on females and males, respectively, on YGA dataset. Combining gender and age estimation, Guo et al. [191] used BIF and age manifold feature extraction with SVM classifier. They reported superior MAE of 2.61 for females and 2.58 for males on YGA database. Yan et al. [192] performed person-independent age image encoding using synchronized submanifold embedding (SME). SME considers both individuals' identities and age labels to improve generalization ability on age estimation. Evaluating this technique on FG-NET, they reported a MAE of 5.21 years. Yan et al. [83, 84] used spatially flexible patch (SFP) for feature description. SFP does not only consider local patches only but also their spatial information. With SFP, slight misalignment, pose variations, and occlusion can be effectively handled. Furthermore, this technique can improve discriminating characteristics of the feature vector when limited samples are available. Adopting Gaussian mixture model (GMM), they achieved a MAE of 4.95 years on FG-NET aging dataset and 4.94 and 4.38 years on females and males, respectively, on

**Table 3** Summary of age and age-group estimation studies

Publication	Feature representation	Face database/database size	Algorithm	Evaluation protocol	Performance/accuracy (MAE, CS)
Kwon and Lobo 1999 [87]	Anthropometric model	Private/47	Classification	15 images for testing	CS 100%
Kanno et al. 2001 [215]	Appearance model	Private/440	Classification	N/A	CS 80%
Lanitis et al. 2002 [66]	2D shape, raw pixel values	Private/500	Regression	500 train, 65 test	MAE 4.3
Iga et al. 2003 [216]	GWT, 2D shape	Private/101	Classification	N/A	CS 58.4%
Lanitis et al. 2004 [23]	Active appearance model (AAM)	Private/400	Classification and regression	50% train, 50% test	MAE 3.82–5.58
Zhou et al. 2005 [217]	Anthropometric model	FG-NET/1002	Regression	5-fold cross-validation	MAE 5.81
Ueki et al. 2006 [134]	Anthropometric model	WIT-DB	Classification	2-fold cross-validation	CS 50%(M), 43%(F)
Takimoto et al. 2006 [218]	Anthropometric model	HOIP	Classification	Leave-one-out	CS 57.3%(M), 54.7%(F)
Geng et al. 2006 [26]	AGES	FG-NET/1002	Regression	Leave-one-person-out	MAE 6.77; CS 81%
Takimoto et al. 2007 [219]	Anthropometric model	HOIP	Regression	Leave-one-out	MAE 3.0(M), 4.4(F)
Gen et al. 2007 [13]	AGES	FG-NET/1002 MORPH	Regression	Train on FG-NET, Test on MORPH	MAE 8.83; CS 70%
Yan et al. 2007 [220]	Active appearance model	FG-NET/1002	Regression	Leave-one-person-out	MAE 5.33
Yan et al. 2007 [220]	Appearance model (raw image)	YGA/8000	Regression	4-fold cross-validation	MAE 6.95(M), 6.95(F); CS 79%(M), 78%(F)
Yan et al. 2007 [140]	Active appearance model	FG-NET/1002	Regression	Leave-one-person-out	MAE 5.78
Yan et al. 2007 [140]	Appearance model (raw image)	YGA/8000	Regression	1000 train, 3000 test	MAE 10.36 (M), 9.79 (F); CS 61%(M), 63%(F)
Fu et al. 2007 [12]	Age manifold	YGA	Regression	50% test, 50% train	MAE 8.0 (M), 7.8 (F); CS 70% (M), 70% (F)
Yan et al. 2008 [83]	Appearance model (patches)	FG-NET/1002	Regression	Leave-one-person-out	MAE 4.95
Yan et al. 2008 [83]	Appearance model (patches)	YGA/8000	Regression	1000 train, 3000 test	MAE 4.38 (M), 4.94 (F); CS 88% (M), 85% (F)
Guo et al. 2008 [31]	Active appearance model	FG-NET/1002	Regression	Leave-one-person-out	MAE 5.08
Guo et al. 2008 [145]	Active appearance model	FG-NET/1002	Hybrid	Leave-one-person-out	MAE 4.97; CS 88%
Guo et al. 2008 [145]	Age manifold	YGA/8000	Hybrid	4-fold cross-validation	MAE 5.12 (M), 5.11 (F); CS 83% (M), 82% (F)
Yan et al. 2008 [84]	Appearance model (patches)	YGA/8000	Regression	1000 train, 3000 test	MAE 7.82 (M), 8.53 (F); CS 75% (M), 70% (F)
Fu and Huang 2008 [12]	Age manifold	YGA/8000	Regression	50% test, 50% train	MAE 6.0 (M), 5.5 (F); CS 82% (M), 83% (F)
Suo et al. 2008 [152]	Active appearance model, photometric	FG-NET/1002, private/8000	Regression	N/A	MAE 6.45 (private), 6.97 (FG-NET)
Zhuang et al. 2008 [221]	Appearance model (patches)	YGA/8000	Regression	50% test, 50% train	MAE 5.40 (M), 6.33 (F); CS 82% (M), 76% (F)
Ni et al. 2009 [154]	Appearance model (patches)	MORPH, web images	Regression	Train on web images, test on MORPH	MAE 8.60

**Table 3** Summary of age and age-group estimation studies (Continued)

Yan et al. 2009 [192]	Active appearance model	FG-NET/1002	Classification	Leave-one-person-out	MAE 5.21
Xiao et al. 2009 [222]	Active appearance model	FG-NET/1002	Classification	Train 300, test 702	MAE 5.04
Guo et al. 2009 [82]	Age manifold, BIF	FG-NET/1002	Regression	Leave-one-person-out	MAE 4.77
Guo et al. 2009 [82]	Appearance model, BIF	YGA/8000	Classification	4-fold cross-validation	MAE 3.47 (M), 3.91 (F); CS 88% (M), 85% (F)
Guo et al. 2009 [191]	Appearance model (BIF) + age manifold	YGA/8000	Classification	4-fold cross-validation	MAE 2.58 (M), 2.61 (F)
Guo and Mu 2011 [123]	BIF	MORPH II/55;132	Regression	50% train, 50% test	MAE 4.2
Choi et al. 2011 [70]	Active appearance model, Gabor, LBP	FG-NET/1002, PAL/430, BERC/390	Hybrid	Leave-one-person-out	MAE 4.7 (FG-NET), 4.3 (PAL), 4.7 (BERC); CS 73%, 70%, 65%, respectively
Luu et al. 2011 [197]	Contourlet appearance model	FG-NET/1002 PAL/443	Hybrid	Leave-one-person-out	MAE 4.1 (FG-NET), 6.0 (PAL); CS 74% (FG-NET)
Chang et al. 2011 [195]	Active appearance model	FG-NET/1002 MORPH II/5492	Classification	20% test, 80% train	MAE 4.5 (FG-NET), 6.1 (MORPH); CS 74.4%, 56.3%, respectively
Wu et al. 2012 [115]	Grassmann manifold of 2D shape	FG-NET/1002, passport/223	Hybrid	Leave-one-person-out	MAE 8.84 (passport), 5.9 (FG-NET); CS 40% (passport), 62% (FG-NET)
Thukral et al. 2012 [180]	Grassmann manifold of 2D shape	FG-NET/1002	Hybrid	Leave-one-person-out	MAE 6.2
Chao et al. 2013 [194]	Active appearance model with distance metric adjustment	FG-NET/1002	Regression	Leave-one-person-out	MAE 4.4
Lu and Tan 2013 [142]	Manifold of raw pixel intensities	MORPH II/20000	Regression	50% test, 50% train	MAE 5.2 (white), 4.2 (black); CS 67% (white), 59% (black)
Hadid	Pietikainen 2013 [200]	Raw intensities, volume LBP	Internet videos/2000	Classification	10% test, 90% train; CS 83.1%

**Table 3** Summary of age and age-group estimation studies (Continued)

Guo and Mu 2013 [124]	BIF	MORPH II/55132	Regression	N/A	MAE 4.0
Geng et al. 2013 [201]	Active appearance model, label distribution	FG-NET/1002 II/55132	MORPH Classification	Leave-one-person-out (FG-NET), 10-fold cross-validation (MORPH II)	MAE 4.8(FG-NET), 4.8(MORPH II)
Han and Jain 2014 [223]	BIF	LFW/143, FG-NET/1002	Classification	98.8% train, 1.2% test	MAE 4.5(FG-NET); CS 66.7% (LFW), 68.1% (FG-NET)
Han et al. 2015 [147]	Demographic informative features	FG-NET/1002 II/78, 207, PCSO/100,012, LFW/4,211	MORPH Hybrid	N/A	MAE 3.8(FG-NET), 3.6(MORPH II), 4.1(PCSO), 7.8(LFW); CS 78% (FG-NET), 77.4% (MORPH II), 72.6% (PCSO), 42.5% (LFW)
Nguyen et al. 2015 [141]	MLBP, Gabor	PAL/1160	Regression	50% test, 50% train	MAE 6.6
Li et al. 2015 [224]	2D age manifold, BIF	FACES/1,026 FG-NET/1002	Classification	N/A	MAE 1.3(FG-NET), 8.2(FACES)
Onifade et al. 2015 [143]	Age-rank LBP	FG-NET/1002	Regression	Leave-one-person-out	MAE 2.34
Yang et al. 2015 [225]	Raw intensities	Private/3615	Classification	2479 train, 1136 test	MAE 0.31
Huerta et al. 2015 [135]	HOG, LBP, SURF	FG-NET/1002	Classification	Leave-one-person-out	MAE 3.31
Hu et al. 2016 [138]	Kullback-Leibler/raw intensities	LFW/150,000 II/55,132 FG-NET/1002	MORPH Classification (CNN)	N/A	MAE 2.8(FG-NET), 2.78(MORPH)
Akinyemi and Onifade 2016 [144]	Raw pixel intensities	FG-NET/1002 FAGE/238	Regression	Leave-one-person-out	MAE 3.19

YGA dataset. Combining BIF and age manifold features and SVM for age estimation achieves MAE of 2.61 and 2.58 years for males and females, respectively, on YGA dataset [11].

Suo et al. [152] designed graphical facial feature topology based on hierarchical face model [193]. They used particular filters to diverse features at various stages of their hierarchical feature extraction design. Using multi-layer perceptron (MLP), they reported MAE of 5.97 years on FG-NET and 4.68 years on their private dataset.

Craniofacial aging model that combines psychophysical and anthropometric evidences was proposed by Ramanathan et al. [59]. The model was used to simulate perceived age of a subject across age for improving accuracy of face recognition. Choi et al. [70] proposed age estimation approach using hierarchical classifiers with local and global facial features. Using Gabor filters for wrinkle extraction and LBP for skin feature extraction, they classified face images into age groups with SVM. This approach is error prone because it only depends on a single classifier. Wrong age group classification leads to wrong age estimation. For accurate age estimation, age group classification must be robust, and this can be achieved by use of an ensemble of classifiers. Chao et al. [194] determined the relationship between age labels and facial features by merging distance metric, learning, and dimensionality reduction. They used label-sensitive and nearest neighbor (KNN) and SVR for age estimation. Chang et al. [195] proposed ordinal hyperplane ranker for age estimation. Using AAM and SVM, their approach achieved 4.48 MAE on FG-NET and MORPH II datasets. Guo et al. [123] build a regression model using BIF and partial least squares (PLS) for age estimation. Their approach achieved 4.43 MAE on MORPH II dataset and showed that learning label distribution improves age estimation. Lu and Tan [142] investigated age estimation using ordinary preserving manifold analysis approach. They found that gait can be used as an effective cue for age estimation at a distance for purposes of enhancing understanding capabilities of existing visual surveillance systems. They further found that discriminating age information can be better exploited in the low-dimensional manifold for achieving better age estimation performance.

Using uniform ternary patterns (UTP) and AAM, Tan et al. [107] and Luu et al. [196] proposed a spectral regressor for age estimation. Evaluating their technique, they achieved a MAE of 6.17. Further work by Luu et al. [197] using contourlet transform achieved a MAE of 6.0 on FG-NET and PAL datasets which was better compared to using UTP. Using Gabor wavelets and orthogonal locality preserving projections (OLPP), Lin et al. [198] developed an automatic age estimation system. They evaluated their technique on FG-NET dataset and SVM as a classifier and achieved a MAE of 5.71 years. Wu et al.

[115] used 2D points to model facial shape for age estimation. Choobert et al. [199] proposed use of an ensemble of classifiers for improving automatic age estimation. The limitation of this work is that only neural network was used to make the ensemble. An ensemble can be made robust if different classifiers are used so as each acts as a complimentary to the other. Guo and Mu [124] compared canonical correlation analysis (CCA) and partial least squares (PLS) performance in age, gender, and ethnicity estimation. Using BIF as a feature extractor, they found that CCA performs better compared to PLS. Hadid and Pietikainen [200] experimented manifold learning on age and gender estimation. They reported 83.1% accuracy age estimation on images extracted from video. Geng et al. [201] learned label distribution and used them for age estimation. Their technique was evaluated on both FG-NET and MORPH datasets.

Guo et al. [82] first introduced BIF in image-based age estimation domain. They reported that using Gabor bank starting from smaller sizes like  $5 \times 5$  can characterize aging. Later, Guo and Mu [123] used  $k$ -partial least squares (KPLS) for simultaneous dimensionality reduction of BIF features for age estimation using a regressor. They also showed that partial least squares (PLS) performs better in dimensionality reduction compared to traditional dimensionality reduction techniques like principal component analysis (PCA). They later [124] used canonical correlation analysis (CCA) for modelling age estimation as multiple-label regression problem. They reported that CCA-based methods work better compared to KPLS-based methods. Spizhevoi and Bovyryn [202] used RBF SVM to learn BIF features for age estimation. Han et al. [203] proposed a hierarchical age estimation and analyzed how aging affects distinct facial components. They used SVM for both classification and regression to classify each face component. Their component localization was not accurate, thereby affecting subsequent features extracted from these components. They later [147] compared human and machine performance on demographic (age, gender, and ethnicity) estimation. They modelled age estimation in particular as a hierarchical problem that consists of between-class classification and within class regression of boosted BIF and demographic informative features extracted from a face image.

Deep learning schemes, especially convolutional neural network (CNN), have been successfully used in face analysis tasks including face detection, face alignment [204], face verification [205], and demographic estimation [206]. Wang et al. [207] extracted feature maps obtained in different layers as age features based on deep learning model. Huerta et al. [135] provide a thorough evaluation on deep learning for age estimation using fused features and compare it with hand-crafted fusion features. CNN have been used in different recent studies on age estimation and have

demonstrated superior performance compared to other methods. Niu et al. [208] used ordinal regression and multiple output CNN for age estimation and reported a MAE of 3.27 on MORPH II and a private Asian Face Age Dataset (AFAD). Chen et al. [209] presented a cascaded CNN that had 0.297 Gaussian error on age estimation. As further demonstrated in [210–212], CNN have posted better results in age estimation tasks. Although CNN performs better than other traditional methods, their applicability is limited by high processing demand required for their implementation. Table 3 shows a summary of studies in age-group and age estimation.

## 11 Conclusions

Comprehensive survey of various techniques and approaches used for age estimation has been presented. There has been enormous effort from both academia and industry dedicated towards modelling age estimation, designing of algorithms, aging face dataset collection, and protocols for evaluating system performance. Table 3 summarizes the findings of recent studies in age estimation, evaluation protocol used, dataset used, age estimation approach used (regression, classification, or hybrid), and feature extraction or age face representation used.

The main issues to consider in age estimation via faces are image representation and estimation techniques. AAM provides a parametric modelling for face representation. A face is represented as a set of shape and texture parameters learned from a face image. AAM can represent both young and old faces since model parameters encode both facial shape and texture. AAM is often used in line with regression-based age estimation approaches. Anthropometric face representation encodes change in facial shape. Anthropometric approaches to facial representation can be very significant in capturing change in facial shape in young faces. AGES can be used to extract subjects' aging patterns when a dataset has sequential aging face images while age manifold is convenient when a dataset has missing aging face images in a large age dataset with wide age ranges. Age manifold learning entwines aging feature extraction and dimensionality reduction. Age manifold can be used both in classification- and regression-based approaches. Appearance models often extract facial features that can be used in regression- or classification-based age estimation approach. These features represent facial appearance. These features could be texture, shape, or wrinkle. Feature extraction techniques like LBP, Gabor, BIF, LDA, PCA, and LDP have been often used for appearance face modeling.

Age estimation can be either approached as age-group estimation or exact age estimation. Age-group estimation approaches approximate age range in which a face image can fall. Exact age estimation approaches estimate a

single label (value) that represents the age of a face image. Both exact age and age-group estimations can be either classification-based, regression-based, or hybrid of both classification and regression. Choice between regression and classification may be guided by face image representation and size and age distribution of the dataset. For big datasets with sequential age labels, both classification and regression can be used, while for datasets with only age-group labels or significantly missing images at some ages, classification-based approach may be more appropriate. Both classification and regression can be combined in a hierarchical manner. In this hybrid approach, often classification is used for age-group estimation followed by exact age estimation within the age-group using regression techniques.

Age estimation techniques can be evaluated using mean absolute error (MAE) or cumulative score (CS). MAE is appropriate when the training set has a lot of missing ages while CS is used when the training dataset has samples at almost every age. Overall performance of the system is represented by CS. In practice, both MAE and CS are used because different techniques and datasets may be biased for evaluation. The most often used evaluation protocols are LOPO and Cross-Validation.

There are a number of promising future directions for age estimation. The following are some of the future research directions that may see improvement in age estimation performance:

- *Fusion*—Feature and decision fusion for age estimation has not been extensively investigated. Fusing shape, wrinkle, and texture features may result into a rich feature set that can distinguish faces in different ages or age groups. Decisions from multiple classifiers or regressors could also be fused to see how they impact age estimation performance.
- *Multi-instance*—Facial landmarks can be extracted and considered as an instance for age estimation. Which parts of the face age faster and how? A face can be broken down into its components (eyes, forehead, nose, nose bridge, mouth, and cheeks) and aging investigation done on each component. Both geometric and anthropometric appearance face modeling can be used on each component.
- *Ethnic*—Faces of subjects from different ethnic groups age differently. Incorporating ethnic parameters as in [144] improves age estimation performance. This approach has not been fully investigated due to lack of large datasets with images from different ethnic groups like African, Asian, and Caucasian.
- *Lifestyle*—One's lifestyle affects how the face ages. Faces of individuals of the same age but with different lifestyles will appear different. Research has shown that

smoking has an influence in facial aging [34, 38–41]. It may be interesting to investigate aging and age estimation among a smoking population and how it compares to non-smoking population. Taister et al. [34] asserts that exposure to drug and psychological stress affects skin texture and color making skin complexion spotted and blemished. Drug use and stress could also be investigated to determine their effect on age estimation.

- **Environment**—Taister et al. [34] found that general exposure to wind and arid air influence facial aging. Arid environment and wind dehydrates the skin leading to wrinkle formation. An investigation of age estimation in populations in different environments is an interesting direction for further research.
- **Databases**—A large multi-racial database is needed for effective investigation of aging in different ethnic groups and gender. Collecting a large database with well-distributed age labels is essential. Web image collection is an efficient way of achieving this [154, 155].
- **Profile face aging**—How do non-frontal parts of the face age? How to estimate age from non-frontal face images? Investigations to answer these two questions could be necessary though are based on availability of such databases (non-frontal face images). 3D face modelling could be vital in investigating profile face aging and age estimation.
- **Multi-sensor**—Image collection from multiple imaging sensors could be appropriate for mitigating degrading factors from uncontrollable and personalized attributes. Fusion could be done on the image features for age estimation.

#### Abbreviations

AAM: Active appearance model; ASM: Active shape model; AGES: Aging pattern subspace; ACM: Active contour model; ANN: Artificial neural network; ANOVA: Analysis of variance; BIF: Biologically inspired feature; CS: Cumulative score; CCA: Canonical correlation analysis; CNN: Convolutional neural network; 2D: Two dimensional; ECRM: Electronic customer relationship management; EM: Expectation maximization; ELM: Extreme learning machines; FG-NET: Face and gesture network; GLCM: Gray-level co-occurrence matrix; GNN: Grassmann nearest neighbor; GMM: Gaussian mixture models; HOG: Histogram of oriented gradients; KNN: K-nearest neighbour; KPLS:  $k$ -partial least squares; LBP: Local binary patterns; LDP: Local directional patterns; LDA: Linear discriminant analysis; LLE: Locally linear embedding; LPP: Linear preserving projections; LTP: Local ternary patterns; LOPO: Leave-one-person-out; LARR: Locally adjusted robust regression; LOO: Leave-one-out; LPO: Leave-p-out; MML: Multi-manifold metric learning; MAE: Mean absolute error; MLBP: Multi-level local binary pattern; MLP: Multilayer perceptron; OLPP: Orthogonal locality preserving projections; PDM: Point distribution model; PCA: Principal component analysis; PLS: Partial least squares; RSS: Random sub-sampling; ROC: Region of certainty; SVM: Support vector machines; SVR: Support vector regression; SFP: Spatially flexible patch; STD: Standard deviation; SOM: Self-organizing maps; SURF: Speeded-up robust features; SDP: Semidefinite programming SME: Submanifold embedding; TV: Television; UV: Ultraviolet; UTP: Uniform ternary patterns; YGA: Yamaha gender and age

#### Authors' contributions

All authors had equal contribution towards this work. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>University of KwaZulu-Natal, School of Mathematics, Statistics and Computer Science, Westville, Durban, 4000 South Africa. <sup>2</sup>University of KwaZulu-Natal, School of Engineering, Howard, Durban, 4041 South Africa.

Received: 13 June 2017 Accepted: 15 May 2018

Published online: 06 June 2018

#### References

1. MS Zimpler, MS Kokosa, JR Thomas, Anatomy and pathophysiology of facial aging. *Facial Plast. Surg. Clin. N. Am.* **9**, 179–187 (2001)
2. R Alley, *Social and Applied Aspects of Perceiving Faces*. (Lawrence Erlbaum Associates, Inc, Hillsdale, 1998)
3. A Gallagher, T Chen, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Estimating age, gender and identity using first name priors (IEEE, Anchorage, 2008)
4. A Gallagher, T Chen, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Understanding images of groups of people, (2009)
5. N Ramanathan, R Chellappa, S Biswas, Computational methods for modeling facial aging: a survey. *J. Vis. Lang. Comput.* **20**, 131–144 (2009)
6. MJ Raval, P Shankar, Age invariant face recognition using artificial neural network. *Int. J. Advance Eng. Res. Dev.* **2**, 121–128 (2015)
7. A Sonu, K Sushil, K Sanjay, A novel idea for age invariant face recognition. *Int. J. Innov. Res. Sci. Eng. Technol.* **3**, 15618–15624 (2014)
8. SN Jyothi, M Indiramma, Stable local feature based age invariant face recognition. *Int. J. Appl. Innov. Eng. Manag.* **2**, 366–371 (2013)
9. S Jinli, C Xilin, S Shiguang, G Wen, D Qionghai, A concatenational graph evolution aging model. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2083–2096 (2012)
10. G Panis, A Lanitis, N Tsapatsoulis, TF Cootes, Overview of research on facial ageing using the FG-NET ageing database. *IET Biometrics.* **5**, 37–46 (2016)
11. Y Fu, G Guo, T Huang, Age synthesis and estimation via faces: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 1955–1976 (2010)
12. Y Fu, TS Huang, Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimedia.* **10**, 578–584 (2008)
13. X Geng, Z Zhou, K Smith-miles, Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 2234–2240 (2007)
14. N Ramanathan, R Chellappa, Face verification across age progression. *IEEE Trans. Image Process.* **15**, 3349–3361 (2006)
15. LA Zebrowitz, *Reading Faces: Window to the Soul*. (Westview Press, Washington DC, 1997)
16. AM Alberta, K Ricanek, E Patterson, A review of the literature on the aging adult skull and face: implications for forensic science research and applications. *Forensic Sci. Int.* **172**, 1–9 (2007)
17. LS Mark, JB Pittenger, H Hines, C Carello, RE Shaw, JT Todd, Wrinkling and head shape as coordinated sources for age-level information. *Percept. Psychophys.* **27**, 117–124 (1980)
18. U Park, Y Tong, AK Jain, Age invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 947–954 (2010)
19. N Ramanathan, R Chellappa, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Face verification across age progression (IEEE, San Diego, 2005), pp. 462–469
20. MM Dehshibi, A Bastanfard, A new algorithm for age recognition from facial images. *Signal Process.* **90**, 2431–2444 (2010)
21. FG-NET: Face and Gesture Recognition Working Group (2002). <http://www-prima.inrialpes.fr/FGnet/>. Accessed 10 Apr 2017
22. Z Song, B Ni, D Guo, T Sim, S Yan, in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*. Learning universal multi-view age estimator using video context (IEEE, Barcelona, 2011), pp. 241–248
23. A Lanitis, C Draganova, C Christodoulou, Comparing different classifiers for automatic age estimation. *IEEE Trans. Man Syst. Cybern.* **34**, 621–628 (2004)

24. E Patterson, A Sethuram, M Albert, K Ricanek, M King, in *Proceedings of 1ST IEEE Conference on Biometrics, Theory and Application Systems*. Aspects of age variation in facial morphology affecting biometrics (IEEE, Crystal City, 2007), pp. 1–6
25. K Ricanek, E Boone. The effect of normal adult aging on standard PCA face recognition accuracy rates (IEEE, Montreal, 2005), pp. 2018–2023
26. X Geng, Z-H Zhou, Y Zhang, G Li, H Dai, in *Proceedings of ACM Conference on Multimedia*. Learning from facial aging patterns for automatic age estimation (ACM, Santa Barbara, 2006), pp. 307–316
27. F Gao, H Ai, in *Proceedings of 3rd International Conference on Advances in Biometrics*. Face age classification on consumer images with gabor feature and fuzzy lda method: lecture notes in computer science (Springer, Alghero, 2009), pp. 132–141
28. H Yang, D Huang, Y Wang, H Wang, Y Tang, Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE Trans. Image Process.* **25**(6), 2493–2507 (2016). <https://doi.org/10.1109/TIP.2016.2547587>
29. H Wang, D Huang, Y Wang, H Yang, Facial aging simulation via tensor completion and metric learning. *IET Comput. Vis.* **11**(1), 78–86 (2017). <https://doi.org/10.1049/iet-cvi.2016.0074>
30. B Bruyer, J-C Scaillquin, Person recognition and ageing: the cognitive status of addresses—an empirical question. *Int. J. Psychol.* **29**, 351–366 (1994)
31. G Guo, Y Fu, C Dyer, T Huang, Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Process.* **17**, 1178–1188 (2008)
32. AK Jain, SC Dass, K Nandakumar, in *Proceedings of International Conference on Biometric Authentication*. Soft biometrics traits for personal recognition systems (Springer, Berlin, 2004), pp. 731–738
33. I Macleod, B Hill, *Heads and Tales: Reconstructing Faces*. (National Museums of Scotland, Edinburgh, 2001)
34. MA Taister, SD Holliday, HIM Borman, Comments in facial aging in law enforcement investigation. *Forensic Sci. Commun.* **2**, 1–11 (2000)
35. E Drakaki, C Dessinioti, CV Antoniou, Air pollution and the skin. *Environ. Sci.* **2**(11), 1–6 (2014)
36. D Zoe, MD Draelos, Aging in a polluted world. *J. Cosmet. Dermatol.* **13**, 85 (2014)
37. A Vierkotter, T Schikowski, U Ranft, D Sugiri, M Matsui, U Kramer, J Krutmann, Airborne particle exposure and extrinsic skin aging. *J. Invest. Dermatol.* **130**(12), 2719–2726 (2010)
38. FG Fedok, The aging face. *Facial Plast. Surg.* **12**, 107–115 (1996)
39. WC Leung, I Harvey, Is skin ageing in the elderly caused by sun exposure or smoking? *Br. J. Dermatol.* **147**, 1187–1191 (2002)
40. HB López, J Tercedor, JM Ródenas, LFRM Simón, OS Serrano, Skin aging and smoking. *Rev. Clin. Esp.*, 147–149 (1995)
41. PM O'Hare, AB Fleischer, RB D'Agostino, SR Feldman, MA Hinds, AA Rassette, A McMichael, PM Williford, Tobacco smoking contributes little to facial wrinkling. *J. Eur. Acad. Dermatol. Venereol.* **12**, 133–139 (1999)
42. RB Shaw, EB Katzel, PF Koltz, DM Kahn, JA Girotto, HN Langstein, Aging of the mandible and its aesthetic implications. *Plast. Reconstr. Surg.* **125**, 332–342 (2010)
43. M Situm, M Buljan, V Cavka, V Bulat, I Krolo, LL Mihic, Skin changes in the elderly people—how strong is the influence of the UV radiation on skin aging? *Coll. Anthropol.* **34**, 9–13 (2010)
44. R Neave, in *Proceedings of Forensic Medicine (J.G. Clement and D.L. Ranson, eds)*. Age Changes to the Face in Adulthood (Oxford University Press, New York, 1998), pp. 225–231
45. S Coleman, R Grover, The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Aesthet. Surg. J. Am. Soc. Aesthet. Plast. Surg.* **26**, 4–9 (2006)
46. PL Leong, Aging changes in the male face. *Facial Plast. Surg. Clin. N. Am.* **16**, 277–279 (2008)
47. K Sveikata, I Balciuniene, J Tutkuvieni, Factors influencing face aging. Literature review. *Stomatologija Baltic Dental Maxillofac. J.* **13**, 113–116 (2011)
48. EC Paes, HJ Teeppen, WA Koop, M Kon, Perioral wrinkles: histologic differences between men and women. *Aesthet. Surg. J.* **29**, 467–472 (2009)
49. DA Gunn, H Rexbye, CE Griffiths, PG Murray, A Fereday, SD Catt, CC Tomlin, BH Strongtharm, DI Perrett, M Catt, AE Mayes, AG Messenger, MR Green, F van der Ouderaa, JW Vaupel, K Christensen, Why some women look young for their age. *PLoS ONE.* **4**(12), e8021 (2009)
50. C Chen, A Dantcheva, A Ross, in *Proceedings of 9th International Conference on Computer Vision Theory and Applications*. Impact of facial cosmetics on automatic gender and age estimation algorithms (IEEE, Lisbon, 2014), pp. 182–190
51. G Guo, X Wang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. A study on human age estimation under facial expression changes (IEEE, Providence, 2012), pp. 2547–2553
52. DT Nguyen, SR Cho, KY Shin, JW Bang, KR Park, Comparative study of human age estimation with or without preclassification of gender and facial expression. *Sci. World J.* 1–15 (2014)
53. M Minear, D Park, A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* **36**, 630–633 (2004)
54. N Ebner, M Riediger, U Lindenberger, FACES—a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods.* **42**, 351–362 (2010)
55. MC Voelkle, NC Ebner, Let me guess how old you are: effects of age, gender, and facial expression on perceptions of age. *Psychol. Aging.* **27**, 265–277 (2012)
56. L Farkas, *Anthropometry of the Head and Face*. (Raven Press, New York, 1994)
57. K Bush, O Antonyshyn, 3-dimensional facial anthropometry using a laser-surface scanner-validation of the technique. *Plast. Reconstr. Surg.* **98**(2), 226–235 (1996)
58. J Kolar, E Salter, *Craniofacial Anthropometry: Practical Measurement of the Head and Face for Clinical, Surgical and Research Use*. (Charles C. Thomas Publisher LTD, 1996)
59. N Ramanathan, R Chellappa, in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*. Modeling age progression in young faces (IEEE, New York, 2006), pp. 384–394
60. TF Cootes, CJ Taylor, DH Cooper, J Graham, Active shape models—their training and application. *Comp. Vision Image Underst.* **61**, 38–59 (1995)
61. M Kass, A Witkin, D terzopoulos, Snakes: active contour models. *Int. J. Comput. Vis.* **1**(321), 321–331 (1988)
62. N Dutta, AK Jain, MP Dubuisson-Jolly, Automatic construction of 2d shape model. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 433–446 (2001)
63. J Liu, JK Udupa, Oriented active shape models. *IEEE Trans. Med. Imaging.* **28**(4), 571–584 (2009)
64. TF Cootes, GJ Edwards, CJ Taylor, Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 681–685 (2001)
65. G Edwards, A Lanitis, C Taylor, T Cootes, Statistical models of face images—improving specificity. *Image Vis. Comput.* **16**, 203–211 (1998)
66. A Lanitis, J Taylor, TF Cootes, Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 442–455 (2002)
67. D Gabor, Theory of communication. *J. Inst. Electr. Eng.* **93**, 429–457 (1946)
68. Y Fu, Y Xu, S HT, in *Proceedings of IEEE Conference Multimedia and Expo*. Estimating human ages by manifold analysis of face pictures and regression on aging features (IEEE, Beijing, 2007), pp. 1383–1386
69. K Scherbaum, M Sunkel, H-P Seidel, V Blanz, Prediction of individual non-linear aging trajectories of faces. *Computer Graphics Forum.* **26**(3), 285–294 (2007)
70. SE Choi, YJ Lee, JL S, RP K, J Kim, Age estimation using hierarchical classifier based on global and local features. *Pattern Recogn.* **44**, 1262–1281 (2011)
71. BS Manjunathi, WY Ma, Texture features for browsing, retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 837–842 (1996)
72. L Huang, J Lu, Y-P Tan, Multi-manifold metric learning for face recognition based on image sets. *J. Vis. Commun. Image Represent.* **25**(7), 1774–1783 (2014)
73. D Beymer, T Poggio, Image representations for visual learning. *Science.* **272**, 1905–1909 (1996)
74. J Hayashi, M Yasumoto, H Ito, H Koshimizu, in *Proceedings of the 7th International Conference on Virtual Systems and Multimedia*. A method for estimating and modeling age and gender using facial image processing (IEEE, Berkeley, 2001)
75. J Hayashi, M Yasumoto, H Ito, Y Niwa, H Koshimizu, in *Proceedings of SICE Annual Conference*. Age and gender estimation from facial image processing (IEEE, Osaka, 2002), pp. 13–18
76. T Ojala, M Pietikainen, D Harwood, A comparative study of texture measures with classification based on featured distribution. *Pattern Recogn.* **29**, 51–59 (1996)

77. JP P, H Moon, SA Rizvi, PJ Rauss, The feret evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1090–1104 (2000)
78. Z Yang, H Ai, in *Proceedings of International Conference on Biometrics*. Demographic classification with local binary patterns (Springer, Seoul, 2007), pp. 464–473
79. F Gao, H Ai, in *Proceedings of International Conference on Advances in Biometrics*. Face age classification on consumer images with gabor feature and fuzzy lda method (Springer, Alghero, 2009), pp. 132–141
80. M Riesenhuber, T Poggio, Hierarchical models of object recognition in cortex. *Nature Neuroscience*. **2**, 1019–1025 (1999)
81. T Serre, L Wolf, S Bilechi, M Riesenhuber, T Poggio, Robust object recognition with cortex-like mechanism. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 411–426 (2007)
82. G Guo, G Mu, Y Fu, TS Huang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Human age estimation using bio inspired features (IEEE, Miami, 2009), pp. 112–119
83. S Yan, X Zhou, M Liu, M Hasegawa-Johnson, TS Huang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Regression from patch-kernel (IEEE, Anchorage, 2008)
84. S Yan, M Liu, TS Huang, in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*. Extracting age information from local spatially flexible patches (IEEE, Las Vegas, 2008), pp. 737–740
85. WB Horng, CP Lee, CW Chen, Classification of age groups based on facial features. *Tamkang J. Sci. Eng.* **4**, 183–192 (2001)
86. H Takimoto, Y Mitsukura, M Fukumi, N Akamatsu, Robust gender and age estimation under varying facial poses. *Electron. Commun. Jpn.* **91**, 32–40 (2008)
87. Y Kwon, N Lobo, Age classification from facial images. *Comp. Vision Image Underst.* **74**, 1–21 (1999)
88. HG Jung, J Kim, Constructing a pedestrian recognition system with a public open database, without the necessity of re-training: an experimental study. *Pattern. Anal. Applic.* **13**, 223–233 (2010)
89. RA Fisher, The statistical utilization of multiple measurements. *Ann. Eugenics.* **8**, 376–386 (1938)
90. K Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd edn. (Academic Press, San Diego, 1990)
91. PN Belhumeur, JP Hespanda, DJ Kriegman, Eigenfaces vs fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 711–720 (1997)
92. DL Swets, JJ Weng, Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 71–86 (1996)
93. AM Martinez, AC Kak, PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 228–233 (2001)
94. A Gunay, VV Nabiyev, in *Proceedings of 23rd International Symposium of Computer and Information Sciences*. Automatic age classification with LBP (IEEE, Istanbul, 2008), pp. 1–4
95. T Ojala, M Pietikainen, T Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971–987 (2002)
96. T Maenpaa, M Pietikainen, *Texture Analysis with Local Binary Patterns: Handbook of Pattern Recognition and Computer Vision*. (World Scientific, 2005)
97. T Ahonen, A Hadid, M Pietikainen, in *Proceedings of European Conference on Computer Vision*. Face recognition with local binary patterns (Springer, Prague, 2004), pp. 469–481
98. D Huang, C Shan, M Ardabilian, Y Wang, L Chen, Local binary patterns and its application to facial image analysis: a survey. *IEEE Trans. Syst. Man Cybernet. Part C (Appl. Rev.)* **41**(6), 765–781 (2011). <https://doi.org/10.1109/TSMCC.2011.2118750>
99. T Ojala, M Pietikainen, Maenpää, in *Proceedings of 2nd ICAPR*. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification (Springer, Rio de Janeiro, 2001), pp. 397–406
100. T Jabid, MH Kabir, O Chae, in *Proceedings of 2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*. Local directional pattern (LDP) for face recognition (IEEE, Las Vegas, 2010). <https://doi.org/10.1109/ICCE.2010.5418801>
101. RA Kirsch, Computer determination of the constituent structure of biological images. *Computers Biomed. Res.* **4**(3), 315–328 (1971). [https://doi.org/10.1016/0010-4809\(71\)90034-6](https://doi.org/10.1016/0010-4809(71)90034-6)
102. JMS Prewitt, *Object Enhancement and Extraction*. (Academic Press, New York, 1970)
103. I Sobel, F G, in *Presented at the Stanford Artificial Intelligence Project (SAIL)*. A 3 X 3 isotropic gradient operator for image processing, (1968)
104. WK Pratt, *Digital Image Processing*. (Wiley, New York, 1978)
105. S-W Lee, Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(6), 648–652 (1996). <https://doi.org/10.1109/34.506416>
106. T Jabid, MH Kabir, O Chae, in *Proceedings of 2010 20th International Conference on Pattern Recognition*. Gender classification using local directional pattern (LDP) (IEEE, Istanbul, 2010). <https://doi.org/10.1109/ICPR.2010.373>
107. X Tan, B Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **19**, 1635–1650 (2010)
108. RC Gonzalez, RE Woods, *Digital Image Processing*, 3rd edn. (Pearson, 2008)
109. MH Pratt, K Shanmugam, I Dinstein, Texture features for image classification. *IEEE Trans. Syst. Man Cybernet.* **3**, 610–621 (1973)
110. M Unser, Sum and difference histograms for texture classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 118–125 (1986)
111. N Zulpe, V Pawar, GLCM texture features for brain tumor classification. *Int. J. Comput. Sci. Issues.* **3**(3), 354–359 (2012)
112. FR Siquira, WR Schwatz, H Pedrini, Multi-scale gray level co-occurrence matrices for texture description. *Neurocomputing.* **120**, 336–345 (2013)
113. LK Soh, C Tsatsoulis, Texture analysis of SAR sea ice imagery using grey level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sensing.* **37**, 780–795 (1999)
114. A Edelman, TA Arias, ST Smith, The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **20**, 303–353 (1999)
115. T Wu, P Turaga, R Chellappa, Age estimation and face verification across aging using landmarks. *IEEE Trans. Inf. Forensics Secur.* **7**, 1780–1788 (2012)
116. H Karcher, Riemannian center of mass and mollifier smoothing. *Commun. Pur. Appl. Math.* **30**, 509–541 (1977)
117. T Serre, M Riesenhuber, Realistic modeling of simple and complex cell tuning in the hmax model and implications for invariant object recognition in cortex. Technical report, Massachusetts Institute of Tech Cambridge Computer Science Artificial Intelligence Lab DTIC Washington DC USA Tech. Rep. AI-MEMO-2004-017 (2004)
118. B Ma, Y Su, F Jurie, Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image Vis. Comput.* **32**, 379–390 (2014)
119. Y Huang, K Huang, D Tao, T Tan, X Li, Enhanced biologically inspired model for object recognition. *IEEE Trans. Syst. Man Cybern.* **41**, 1668–1680 (2011)
120. D Song, D Tao, Biologically inspired feature manifold for scene classification. *IEEE Trans. Image Process.* **19**, 174–184 (2010)
121. GD J, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Optic. Soc. Am.* **2**, 1160–1169 (1985)
122. S Marcelja, Mathematical description of the responses of simple cortical cells. *J. Optic. Soc. Am.* **70**, 1297–1300 (1980)
123. G Guo, G Mu, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression (IEEE, Providence, 2011), pp. 657–664
124. G Guo, G Mu, in *Proceedings of IEEE Conference on Face and Gesture Recognition*. Joint estimation of age, gender and ethnicity:cca vs pls (IEEE, Shanghai, 2013), pp. 1–6
125. B Cai, X Xu, X Xing, BIT: Biologically inspired tracker. *IEEE Trans. Image Process.* **25**, 1327–1339 (2016)
126. M Haghighat, S Zonouz, M Abdel-Mottaleb, in *Proceedings of Computer Analysis of Images and Patterns Conference*. Identification using encrypted biometrics (Springer, York, 2013), pp. 440–448
127. MR Turner, Texture discrimination by gabor functions. *Biol. Cybernet.* **55**, 71–82 (1986)
128. L Shen, L Bai, A review on gabor wavelets for face recognition. *Pattern Anal. Appl.* **9**, 273–292 (2006)
129. JG Daugman, Spatial visual channels in the Fourier plane. *Vis. Res.* **24**, 891–910 (1984)

130. I Lampl, D Ferster, T Poggio, M Riesenhuber, Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex. *J. Neurophysiol.* **92**, 2704–2713 (2004)
131. T Serre, L Wolf, T Poggio (eds.), *Object Recognition with Features Inspired by Visual Cortex* (IEEE, San Diego, 2005)
132. J Mutch, D Lowe, Object class recognition and localization using sparse features with limited receptive fields. *Int. J. Comput. Vis.* **80**, 45–57 (2008)
133. VN Vapnik, An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**(5), 988–999 (1999). <https://doi.org/10.1109/72.788640>
134. K Ueki, T Hayashida, T Kobayashi, in *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*. Subspace-based age group classification using facial images under various lighting conditions (IEEE, Southampton, 2006), pp. 43–48
135. I Huerta, C Fernandez, C Segura, J Hernando, A Prati, A deep analysis on age estimation. *Pattern Recognit. Lett.* **68**, 239–249 (2015)
136. H Bay, T Tuytelaars, LV Gool, Surf: Speeded up robust features. *Comput. Vis.-ECCV.* **3951**, 404–417 (2006)
137. B Triggs, N Dalal, in *Proceedings of IEEE on Computer Vision and Pattern Recognition*. Histograms of oriented gradients for human detection (IEEE, San Diego, 2005), pp. 886–893
138. Z Hu, Y Wen, J Wang, M Wang, R Hong, S Yan, Facial age estimation with age difference. *IEEE Trans. Image Process.* 1–13 (2016)
139. G Guo, Y Fu, TS Huang, C Dyer, in *Proceedings of IEEE Workshop on Applications of Computer Vision*. Locally adjusted robust regression for human age estimation (IEEE, Copper Mountain, 2008)
140. S Yan, H Wang, X Tang, TS Huang, in *Proceedings of IEEE Conference on Computer Vision*. Learning auto-structured regressor from uncertain non-negative labels, (2007)
141. DT Nguyen, SR Cho, KR Park, Age estimation-based soft biometrics considering optical blurring based on symmetrical sub-blocks for mlbp. *Symmetry*, 1882–1913 (2015)
142. J Lu, Y Tan, Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Trans. Hum. Mach. Syst.* **43**, 249–258 (2013)
143. OFW Onifade, DJ Akinyemi, A groupwise age ranking framework for human age estimation. *Int. J. Image Graphics Signal Process.* **5**, 1–12 (2015)
144. JD Akinyemi, OFW Onifade, in *Proceedings of IEEE Symposium on Technologies for Homeland Security*. An ethnic-specific age group ranking approach to facial age estimation using raw pixel features (IEEE, Waltham, 2016), pp. 1–6
145. G Guo, Y Fu, TS Huang, C Dyer, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition-Semantic Learning and Applications in Multimedia Workshop*. A probabilistic fusion approach to human age prediction, (2008), pp. 1–6
146. A Gunay, WV Nabiyev, Facial age estimation based on decision level fusion of amm, lbp and gabor features. *Int. J. Adv. Comput. Sci. Appl.* **6**, 19–26 (2015)
147. H Han, XO CLiu, AK Jain, Demographic estimation from face images: human vs machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1148–1161 (2015)
148. K Ricanek, T Tesafaye, in *Proceedings of IEEE 7th International Conference on Automatic Face and Gesture Recognition*. Morph: A longitudinal image database of normal adult age-progression (IEEE, Southampton, 2006), pp. 341–345
149. P Viola, M Jones, Robust real-time object detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
150. Y Fu, N Zheng, M-face: An appearance-based photorealistic model for multiple facial attributes rendering. *IEEE Trans. Circ. Syst. Video Technol.* **16**, 830–842 (2006)
151. DM Burt, DI Perrett, in *Proceedings of Royal Society of London Series B Biological Sciences*. Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information, (1995), pp. 137–143
152. J Suo, T Wu, S Zhu, S Shan, X Chen, W Gao, in *Proceedings of IEEE Conference in Automatic Face and Gesture Recognition*. Design sparse features for age estimation using hierarchical face model (IEEE, Amsterdam, 2008)
153. A Bastanfard, MA Nik, MM Dehshibi, in *Proceedings of International Conference on Machine Vision*. Iranian face database with age, pose and expression (IEEE, Islamabad, 2007), pp. 50–55
154. B Ni, Z Song, S Yan, in *Proceedings of ACM International Conference on Multimedia*. Web image mining towards universal age estimator (ACM Press, Beijing, 2009), pp. 85–94
155. B Ni, Z Song, S Yan, Web image and video mining towards universal and robust age estimator. *IEEE Trans. Multimedia.* **13**, 1217–1229 (2011)
156. SP Kyaw, J-G Wang, EK Teoh, in *Proceedings of IEEE International Conference on Information, Communication and Signal Processing*. Web image mining for facial age estimation (IEEE, Tainan, 2013)
157. V Blanz, T Vetter, in *Proceedings of ACM Conference SIGGRAPH*. A morphable model for the synthesis of 3D faces (ACM Press, New York, 1999), pp. 187–194
158. M Budka, B Gabrys, Density-preserving sampling: robust and efficient alternative to cross-validation for error estimation. *IEEE Trans. Neural Netw. Learn. Syst.* **24**, 22–34 (2013)
159. S Weiss, C Kulikowski, *Computer Systems that Learn*. (Morgan Kaufmann, 1991)
160. Y Bengio, Y Grandvalet, No unbiased estimator of the variance of k-fold cross validation. *J. Mach. Learn. Res.* **5**, 1089–1105 (2004)
161. A Jain, R Duin, J Mao, Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 4–37 (2000)
162. RR Picard, RD Cook, Cross validation of regression models. *J. Am. Stat. Assoc.* **79**, 575–583 (1984)
163. R Kohavi, in *Proceedings of International Joint Conference on Artificial Intelligence*. A study of cross-validation and bootstrap for accuracy estimation and model selection (Morgan Kaufmann Publishers Inc., San Francisco, 1995), pp. 1137–1143
164. O Kiline, I Uysal, in *Proceedings of 14th International Conference on Machine Learning and Applications*. Source-aware partitioning for robust cross-validation (IEEE, Miami, 2015)
165. R Duda, P Hart, D Stork, *Pattern Classification*, 2nd edn. (Wiley, Menlo Park, 2001)
166. M Stone, Cross-validated choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B (Methodological)*. **36**, 111–147 (1974)
167. T Hastie, R Tibshirani, J Friedman, J Franklin, The elements of statistical learning: data mining inferences and prediction. *Math. Intell.* **27**, 83–85 (2005)
168. P Rafaeilzadeh, L Tang, H Liu, Cross-validation. *Encycl. Database Syst.*, 532–538 (2009)
169. S Geisser, The predictive sample reuse method with applications. *J. Am. Stat. Assoc.* **70**, 320–328 (1975)
170. DM Allen, The relationship between variable selection and data augmentation and a method for prediction. *Technometrics.* **16**, 125–127 (1974)
171. B Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Stat. Assoc.* **78**, 316–331 (1983)
172. J Shao, Linear model selection by cross-validation. *J. Am. Stat. Assoc.* **88**, 486–494 (1993)
173. A Elisseeff, M Pontil, *Leave-One-Out Error and Stability of Learning Algorithms with Applications*. (IOS Press, 2003)
174. A Celisse, S Robin, Nonparametric density estimation by exact leave-p-out cross-validation. *Comput. Stat. Data Anal.* **52**, 2350–2368 (2008)
175. S Arlot, A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010)
176. B Efron, R Tibshirani, Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–77 (1986)
177. B Efron, R Tibshirani, *An Introduction to the Bootstrap*. (Chapman & Hall, New York, 1993)
178. S Salzberg, On comparing classifiers: pitfalls to avoid and a recommended approach. *Data Mining Knowl. Discov.* **1**, 317–328 (1997)
179. N Ramanathan, R Chellapa, W Biswas, Age progression in human faces: a survey. *J. Vis. Lang. Comput.* **15**, 3349–3361 (2009)
180. P Thukral, K Mitra, R Chellappa, in *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*. A hierarchical approach for human age estimation (IEEE, Kyoto, 2012), pp. 1529–1532
181. LG Farkas, MJ Katic, CR Forrest, International anthropometric study of facial morphology in various ethnic groups/races. *J. Craniofacial Surg.*, 615–646 (2005)
182. R Tiwari, A Shukla, C Prakash, D Sharma, R Kumar, S Sharma, in *Proceedings of IEEE International Advance Computing Conference*. Face recognition using morphological methods (IEEE, Patiala, 2009), pp. 529–534

183. MA Hajizadeh, H Ebrahimnezhad, in *Proceedings of 7th Iranian Machine Vision and Image Processing Conference*. Classification of age groups from facial image using histogram of oriented gradient (IEEE, Tehran, 2011), pp. 1–5
184. K-H Liu, S Yan, JC-C Kuo, in *Proceeding of IEEE Winter Conference on Applications of Computer Vision (WACV)*. Age group classification via structured fusion of uncertainty-driven shape features and selected surface features (IEEE, Steamboat Springs, 2014), pp. 445–452
185. H Ling, S Soatto, N Ramanathan, D Jacobs, Face verification across age progression using discriminative methods. *IEEE Trans. Inf. Forensics Secur.* **5**, 82–91 (2010)
186. K-H Liu, S Yan, JC-C Kuo, Age estimation via grouping and decision fusion. *IEEE Trans. Inf. Forensics Secur.* **10**, 2408–2423 (2015)
187. C-C Chang, C-J Lin, Libsvm: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27 (2011)
188. P Sai, J Wang, E Teoh, Facial age range estimation with extreme learning machines. *Neurocomputing*. **149**, 364–372 (2015)
189. G-B Huang, Q-Y Zhu, C-K Siew, Extreme learning machines: theory and applications. *Neurocomputing*. **70**, 489–501 (2006)
190. J Wang, W Yau, HL Wang, in *Proceedings of IEEE Applications of Computer Vision (WACV)*. Age categorization via ecoc with fused gabor and lbp features, (2009), pp. 313–318
191. G Guo, G Mu, Y Fu, C Dyer, TS Huang, in *Proceedings of IEEE Conference on Computer Vision*. A study on automatic age estimation using a large database (IEEE, Kyoto, 2009)
192. S Yan, H Wang, Y Fu, J Yan, X Tang, TS Huang, Synchronized submanifold embedding for person independent pose estimation and beyond. *IEEE Trans. Image Process.* **18**, 202–210 (2009)
193. J Suo, F Min, S Zhu, S Shan, X Chen, in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*. A multi-resolution dynamic model for face aging simulation (IEEE, Minneapolis, 2007)
194. W Chao, J Liu, J Ding, Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recogn.* **46**, 628–641 (2013)
195. K-Y Chang, C-S Chen, Y-P Hung, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Ordinal hyperplane ranker with cost sensitivities for age estimation (IEEE, Providence, 2011), pp. 585–592
196. K Luu, TD Bui, CY Suen, in *Proceedings of IEEE Conference on Automatic Face & Gesture Recognition and Workshops*. Kernel spectral regression of perceived age from hybrid facial features (IEEE, Santa Barbara, 2011)
197. K Luu, K Seshadri, M Savvides, T Bui, C Suen, in *Proceedings of IJCB*. Contourlet appearance model for facial age estimation (IEEE, Washington, DC, 2011), pp. 1–8
198. C-T Lin, D-L Li, J-H Lai, M-F Han, J-Y Chang, Automatic age estimation system for face images. *Int. J. Adv. Robot. Syst.* **9**, 1–9 (2012)
199. AK Choobier, Improving automatic age estimation algorithms using an efficient ensemble technique. *Int. J. Mach. Learn. Comput.* **2**, 118–122 (2012)
200. A Hadid, M Pietikainen, Demographic classification from face videos using manifold learning. *Neurocomputing*. **100**, 197–205 (2013)
201. X Geng, C Yin, Z-H Zhou, Facial age estimation by learning from label distribution. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 2401–2412 (2013)
202. AS Spizhevoi, AV Bovyriin, Estimating human age using bio-inspired features and the ranking method. *Pattern Recognit. Image Anal.* **25**, 547–552 (2015)
203. H Han, AK Jain, in *Proceedings of IAPR International Conference on Biometrics*. Age estimation from face images: human vs. machine performance (IEEE, Madrid, 2013)
204. Y Sun, X Wang, X Tang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Deep convolutional network cascade for facial point detection (IEEE, Portland, 2013), pp. 3476–3483
205. Y Taigman, M Yang, M Ranzato, L Wolf, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Deepface: closing the gap to human level performance in face verification. (Columbus, 2014), pp. 1701–1708
206. M Yang, S Zhu, F Lv, K Yu, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Correspondence driven adaptation for human profile recognition (IEEE, Colorado Springs, 2011), pp. 505–512
207. X Wang, R Guo, C Kambhamettu, in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*. Deeply-learned feature for age estimation (IEEE, Waikoloa, 2015), pp. 534–541
208. Z Niu, M Zhou, L Wang, X Gao, G Hua, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Ordinal regression with multiple output cnn for age estimation (IEEE, Las Vegas, 2016), pp. 4920–4928
209. J-C Chen, A Kumar, R Ranjan, VM Patel, A Alavi, R Chellappa, in *Proceedings of IEEE Conference on Biometrics, Theory, Applications and Systems*. A cascaded convolutional neural network for age estimation of unconstrained faces (IEEE, Niagara Falls, 2016)
210. Z Huo, X Yang, C Xing, Y Zhou, P Hou, J Lv, X Geng, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Deep age distribution learning for apparent age estimation (IEEE, Las Vegas, 2016), pp. 722–729
211. RC Malli, M Aygun, HK Ekenel, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Apparent age estimation using ensemble of deep learning models (IEEE, Las Vegas, 2016), pp. 714–721
212. B Hebda, T Kryjak, in *Proceedings of IEEE Conference on Computer Science and Information Systems*. A compact deep convolutional neural network architecture for video based age and gender estimation (IEEE, Gdansk, 2016), pp. 787–790
213. T Todd, SM Leonard, RE Shaw, JB Pittenger, The perception of human growth. *Sci. Am.* **242**(2), 132–144 (1980)
214. S k. Kim, H Lee, S Yu, S Lee, in *Proceedings of 2006 1ST IEEE Conference on Industrial Electronics and Applications*. Robust face recognition by fusion of visual and infrared cues (IEEE, Singapore, 2006), pp. 1–5. <https://doi.org/10.1109/ICIEA.2006.257072>
215. T Kanno, M Akiba, Y Teramachi, H Nagahashi, T Agui, Classification of age-group based on facial images of young males using neural networks. *IEICE Trans. Inf. Syst.* **E84-D**, 1094–1101 (2001)
216. R Iga, K Izumi, H Hayashi, G Fukano, T Ohtani, in *Proceedings of SICE Annual Conference*. A gender and age estimation system from face images (IEEE, Fukui, 2003), pp. 756–761
217. SK Zhou, B Georgescu, XS Zhou, D Comaniciu, in *Proceedings of IEEE International Conference on Computer Vision*. Image based regression using boosting method (IEEE, Beijing, 2005), pp. 541–548
218. H Takimoto, Y Mitsukura, M Fukumi, N Akamatsu, in *Proceedings of ICASE International Joint Conference*. A design of gender and age estimation system based on facial knowledge (IEEE, Busan, 2006), pp. 3883–3886
219. H Takimoto, T Kuwano, Y Mitsukura, H Fukai, M Fukumi, in *Proceedings of SICE Annual Conference*. Appearance-age feature extraction from facial image based on age perception (IEEE, Takamatsu, 2007), pp. 2813–2818
220. S Yan, H Wang, TS Huang, X Tang, in *Proceedings of IEEE Conference on Multimedia and Expo*. Ranking with uncertain labels (IEEE, Beijing, 2007), pp. 96–99
221. X Zhuang, X Zhou, M Hasegawa-Johnson, TS Huang, in *Proceedings of International Conference on Pattern Recognition*. Face age estimation using patch-based hidden markov model supervectors (IEEE, Tampa, 2008)
222. B Xiao, X Yang, Y Xu, in *Proceedings of ACM International Conference on Multimedia*. Learning distance metric for regression by semidefinite programming with application to human age estimation (ACM Press, Beijing, 2009)
223. H Han, AK Jain, *Age, gender and race estimation from unconstrained face images*. Technical report MSU-CSE-14-5. (Michigan State University, 2014)
224. C Li, Q Liu, W Dong, X Zhu, J Liu, H Lu, Human age estimation based on locality and ordinal information. *IEEE Trans. Cybernet.* **45**, 2522–2534 (2015)
225. X Yang, B-B Gao, C Xing, et al, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Deep label distribution learning for apparent age estimation (IEEE, Santiago, 2015), pp. 344–350