


Article

# Improving Juvenile Age Estimation Based on Facial Landmark Points and Gravity Moment

Ebenezer Nii Ayi Hammond <sup>1,2</sup> , Shijie Zhou <sup>1,\*</sup>, Hongrong Cheng <sup>1</sup>  and Qihe Liu <sup>1</sup> 

<sup>1</sup> School of Information and Software Engineering, University of Electronic Science and Technology of China, Xiyuan Ave, West Hi-Tech Zone, Chengdu 611731, China; ebenhammond@std.uestc.edu.cn (E.N.A.H.); hrcheng@uestc.edu.cn (H.C.); qiheliu@uestc.edu.cn (Q.L.)

<sup>2</sup> Council for Scientific and Industrial Research, Building and Road Research Institute, P.O. Box UP40, KNUST-Kumasi AK-448-6464, Ghana

\* Correspondence: sjzhou@uestc.edu.cn

Received: 4 August 2020; Accepted: 1 September 2020; Published: 8 September 2020

**Abstract:** Facial age estimation is of interest due to its potential to be applied in many real-life situations. However, recent age estimation efforts do not consider juveniles. Consequently, we introduce a juvenile age detection scheme called LaGMO, which focuses on the juvenile aging cues of facial shape and appearance. LaGMO is a combination of facial landmark points and Term Frequency Inverse Gravity Moment (TF-IGM). Inspired by the formation of words from morphemes, we obtained facial appearance features comprising facial shape and wrinkle texture and represented them as terms that described the age of the face. By leveraging the implicit ordinal relationship between the frequencies of the terms in the face, TF-IGM was used to compute the weights of the terms. From these weights, we built a matrix that corresponds to the possibilities of the face belonging to the age. Next, we reduced the reference matrix according to the juvenile age range (0–17 years) and avoided the exhaustive search through the entire training set. LaGMO detects the age by the projection of an unlabeled face image onto the reference matrix; the value of the projection depicts the higher probability of the image belonging to the age. With Mean Absolute Error (MAE) of 89% on the Face and Gesture Recognition Research Network (FG-NET) dataset, our proposal demonstrated superior performance in juvenile age estimation.

**Keywords:** age estimation; AAM; juvenile detection; information retrieval; ordinal relationship

## 1. Introduction

Age estimation enables the automatic tagging of a person's age with a specific number or age bracket. It is relevant in real-world applications such as web access control [1], criminal investigations [2], forensics [3] and healthcare [4], where it can be particularly useful for addressing the problem of estimating the age of a separated or unaccompanied child [5].

Age estimation systems require inputs of features such as the iris [6], voice [1], teeth [7] or blood [8]. However, the majority of the systems rely on facial images [2,9,10]. This is due to the face being the most visible part of the body and natural storage for personal traits, making it the most representative part of an individual. Moreover, it is easy to acquire facial images through the use of portable, non-evasive cameras. Sharing the images is easier due to the widespread availability of the Internet. Consequently, the potential application of automatic age estimation systems continue to attract both researchers and practitioners.

However, estimating age from the facial image is challenging due to the variability of the face, which results from intrinsic (genes) and extrinsic (environmental) factors impacting the face in different ways. As a result, aging manifests differently for different individuals, or commonly for age groups. Whereas in the juvenile group, aging manifests by subtle movement and growth of facial bones

and cranial features, aging in adults is perceived through skin texture changes resulting from loss of skin elasticity and skin quality [11]. These cues broadly fall under global and local features corresponding to shape-related features on the one hand, and texture-related features on the other hand [10]. Global features include facial shape geometry and allied anthropometric characteristics that evolve from the childhood stage of growth. The evolution slows down or ceases as the child enters adulthood. Consequently, global features are more useful for juvenile age estimation or classification between juveniles and adults. Conversely, local features begin to manifest at adulthood. They include skin wrinkles and aging spots that together form the texture of the skin. They are more useful for estimating the ages of adults. Although aging introduces a unique set of features for the growing juvenile or aging adult, the frequencies of the features can influence the overall appearance of the face. The representation of the appearance of the face has been the basis for many proposals [12]. Whereas some proposals rely solely on global or local features representations, others consider both global and local features [13,14].

In general, age estimation systems operate under two different phases: image representation and age prediction. In the representation phase, facial features are extracted and used as input for the system. Some notable representation methods include anthropometric-based models [15], texture-based models [13], Active Appearance Models (AAM) [16], Active Shape Models (ASM) [17], Aging Pattern Subspace (AGES) [18], Age Manifold [19] and Multi-feature Fusion [20]. Other useful methods include Local Binary Pattern (LBP) [13], Gabor Wavelet (GW) [21], Sobel [22] and Canny [23]. The goal of feature representation is to acquire a feature vector that adequately represent the target age or age-group. Some researchers propose methods that utilize the facial landmark points, depicted in Figure 1, to represent aging features. Here, pioneer investigators include Kwon et al. [15]. They used geometrical features and wrinkle patterns derived from the locations of the landmark points to represent aging features. Additionally, Ramanathan and Chellappa [20] proposed a model to examine facial shape differences for children under 18 years of age. Recently, shape-oriented methods like the AAM are proposed for various kinds of representations [14,16,24,25]. Regardless of the adopted representation method, the obtained features should be relevant to the objectives of the proposal.

**The age prediction phase:** This phase aims to address the estimation task as a classification or regression problem. With the classification problem, facial images are organized into age-groups with defined age ranges such as juveniles or adults. In contrast, with the regression problem, ages are taken to be numeric values. In some proposals, the two prediction methods are combined hierarchically, first to classify age into the adult or juvenile group, and then estimate age as a specific number within the target age-group [26–28].

Recently, the artificial neural network approach has been gaining ground in age estimation [29]. Although they have a unique ability to combine both the feature extraction and prediction phases into a single pipeline, the approach comes with a higher dependency on very large datasets, which may not be readily available [28,30]. Consequently, traditional methods continue to attract attention [31,32].

Irrespective of the approaches considered for the age estimation proposals, Mean Absolute Error (MAE), Cumulative Score (CS), and Classification Accuracy (Acc) are the favored metrics for evaluating the performance of the system.

**Challenges:** Considering the variability of the aging face and the different aging patterns and features, age estimation continues to challenge computer vision researchers. Moreover, when the age estimation task is within the same aging subspace such as juveniles (0 to 17 years), it becomes even more difficult to measure subtle changes in features that correspond to the age-group. These challenges are compounded by the general lack of a juvenile dataset. We hope to address these problems by offering different value combinations that serve the juvenile age estimation goals.

**Contributions:** In this paper, we present the first-time use of the combination of facial landmark points and Term Frequency Inverse Gravity Moment (TF-IGM) to contribute a juvenile age estimation scheme we name LaGMO. The implementation of LaGMO took place through the following efforts:

- Utilizing the 68 facial landmark points to build high-level terms that describe the shape and appearance features. By exploiting the implicit ordinal relationships among the frequencies of the terms (features) in the various ages, we aggregated the features into a weight matrix for the entire ages.
- The weights of the features and their contributions to the age prediction task were computed by TF-IGM.
- We demonstrated the effectiveness of LaGMO on a discriminating dictionary of juvenile age descriptors, which we obtained from the FG-NET dataset.
- With a MAE of 4.42 and a SC of 89.8%, LaGMO advances juvenile age estimation.

The rest of the paper is organized as follows: Section 2 reviews previous work and introduces the various components of our proposal. Section 3 discusses the fundamental concepts of our proposal. We present details of the proposed scheme in Section 4. In Section 5, experimentation and results are discussed. Finally, Section 6 concludes our contribution.

## 2. Related Work

Automatic age estimation continues to attract interest in research and practice. In this section, we discuss recent works that highlight our proposal.

### 2.1. Feature Extraction

Feature extraction attempts to estimate representations of the face, which are as close as possible to the ground truth. This is consistent with the knowledge that aging information is encoded in the face and can be represented by a set of facial landmarks [33]. However, extracting accurate features remains a difficult task. Consequently, various feature extraction methods are proposed for age estimation [14,15,25]. Pioneering works included that of Kwon et al. [15], who distinguished faces into baby, young-adult and senior-adult classes from geometrical features and wrinkle patterns. The authors used snakelets to detect curves and extracted wrinkle patterns from certain areas of the skin. The approach learned from a small private dataset of 47 images. Additionally, Ramanathan and Chellappa [20] proposed a shape model to examine facial shape differences for children under 18 years of age. They identified face shape by a set of facial muscle landmarks and their corresponding coordinates, which formed 48 fiducial features. The authors presented a model for measuring childhood aging deformations by warping faces per the model for rejuvenation or aging. The model enabled face age estimation based on both childhood aging features and adulthood aging features. Efraty et al. [34] proposed an automatic facial landmark detection method that analyzed image intensities with adaptive bag-of-words. Tong, et al. [35] introduced a landmark extraction method by minimizing an objective function that was trained on both labeled and unlabeled face images. Segundo et al. [36] extracted facial landmarks with a method that combined relief curves and surface curvature. Facial landmark extraction continues to attract attention. Recently, Su and Geng [37] proposed a multi-scale cascaded bivariate label distribution (BLD) learning method to improve facial landmarks. The authors used an ensemble of low and high-resolution images and an optimization mechanism to establish mappings from an input patch to the BLD for each image that most likely represented the true BLD.

From the preceding, the majority of the feature extraction methods utilize techniques that aim to locate landmarks on the face. In this regard, the Active Appearance Model (AAM) is predominately utilized [10].

#### 2.1.1. The Active Appearance Model (AAM)

Initially proposed by Cootes et al. [16], AAM is a statistical model for representing shape and appearance variations of the face. The shape representations, obtained from key landmark points on the face, result from the AMM algorithm conducting a series of Procrustean and Principal Component

Analyses (PCA). Due to its success, various extensions of the AAM have been proposed for different system objectives. Lanitis et al. [11], extended AAM to obtain person-specific features wherein the authors successfully extracted craniofacial growth-related representations for the child and adult faces.

Since AAM relies on landmark points, it is intuitive that the number of points should impact the performance of the proposed system. Although the selection of the number has yet to be reviewed comprehensively, available literature suggests the use of different numbers for different proposals; for example, 48 points were considered in [20], 17 points in [38], 79 points in [39], etc. However, the majority utilized 68 points [10,40,41]. We observe that the choice depends on the objectives of the proposal. In this proposal, we utilized the 68 facial landmark points for the following reasons:

1. They are extensively utilized for age estimation and allied systems, attesting to their ability to represent face aging features.
2. Some age estimation datasets are already annotated with the 68 facial landmark points.

Based on these advances, we utilized the 68 landmark points of the AMM to represent facial features similarly to the proposal by Chen et al. [14], wherein the features were represented as terms that described the shape and appearance of the face.

## 2.2. Age Estimation

Predicting the age from the extracted features can be achieved by classification or regression methods. However, the choice and combination of the methods depend on the system objectives. Recently, a promising proposal that alleviates the lack of dataset, known as the label distribution or soft classification, has attracted interest [42,43].

### Label Distribution

The idea of the label distribution method is to present ages as distributed labels that cover a certain number, with each number describing the extent to which each label defines an instance of the face. Consider an image  $f$ ; label distribution is represented as a vector  $d$  that describes the age  $g$ , such that the descriptions are real numbers presented as  $d_{g,f} \in [0; 1]$ . The label distribution is expected to satisfy two conditions:

1. The description degree of  $f$  should have the highest value.
2. Description degree of the neighboring ages should decrease while going away from  $f$ .

Inspired by the label distribution method, Kohli et al. [44] used an ensemble of classifiers to distinguish children and adults. Setting the child threshold at 21 years, they proceeded to estimate the age using an aging function. To mitigate dataset constraints, Geng et al. [42] proposed label distribution by learning chronological age and adjacent ages for each image. Their proposal included the image itself in the distribution and assumed the representation to be consistent with the entropy condition. Additionally, He et al. [43] combined the distribution learning of age labels with age prediction. Their approach considered context relationships by using different face samples to find the correlation of the cross-age. The method was proposed for problems that consider a preconception-less label distribution learning or for learning sample-specific, context-conscious label distribution properties capable of solving multiple tasks, including age distribution and age prediction. The label distribution method can be extended to address specific age estimation problems if the ordinal relationship is adequately exploited to represent the aging features.

From the related works, we made the following observations:

1. Aging features come in two major kinds—global and local features, corresponding to face shape-related and skin texture-related features.

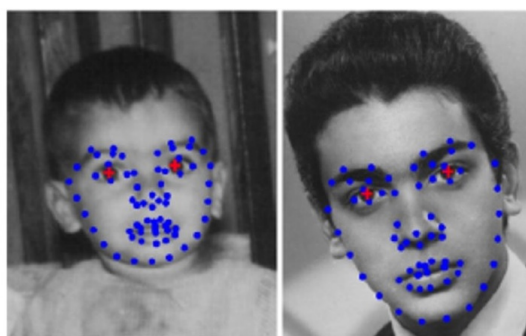
2. Shape-related features are more useful for juvenile age estimation, whereas skin texture (wrinkle) better serves the adult age estimation proposals. However, both of them can be exploited to address specific age estimation objectives.
3. Age prediction methods can be categorized as classification and regression, but considering the implicit ordinal relationship of faces within a similar aging subspace could advance the age estimation task.
4. Although some publicly available datasets are commonly used for age estimation, in general, there is a lack of datasets for juvenile age estimation.

### 3. Preliminaries

This section briefly describes the theoretical concepts and definitions that are vital for understanding our proposal.

#### 3.1. Facial Landmark-Term by AAM

AAM is a statistical shape and appearance model that presents a holistic view of the face [10,16].



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

Figure 1. 68 facial landmark points.

Given an annotated facial image, the shape information can be considered as the dominant points defining specific landmarks of the face, including the forehead, eye corner, forehead, face cheeks, etc., or an interpolated linkage of the points around the whole face [45]. However, when considered separately, we observe that the landmark points have no inherent ability to describe the face. Their independence appears similarly to the individual units of a natural language. As with the 26 letters of the English Language, the landmark points cannot by themselves describe the age of a face. However, when represented at a higher level, they can be considered to represent age. As with the quality of work by [14], we corroborate that the 68 landmark points can be transferred into meaningful terms (features) similar to words in natural language. Then, for different term vectors (features), the vectors can represent the age of the face. For the AAM of the face image, 68 landmark points can be expanded such that the same terms for different AAMs can be treated as the same features. We denote the AAM of an image  $img$  as vector  $LM_{(img)}$  and expand it to

cover the 68 points as  $LM_{(img)} = (x_1, x_2, x_3, x_4, \dots, x_{135}, x_{136})$ , where  $x_1, x_2$  is the first landmark point,  $x_3, x_4$  the second landmark point and  $x_{135}, x_{136}$  is the last landmark point, aiming to obtain a term of the form  $F - j - LOC$ , where  $F$  represents the appearance information and  $j$  is the  $j^{th}$  element with  $LOC$  representing the location of the  $j^{th}$  point measured as  $LOC = \text{round}(\frac{x_j}{width})$ . However, as shown in Figure 1, the 68 points of the AAM do not fully represent the holistic appearance of the face, as the distribution of the points falls within the areas marked off by the points. Additionally, certain regions of the face tend to show early signs of aging and must therefore be considered for a more holistic appearance representation. Therefore, we adapt the proposal in [46] to obtain wrinkle information from the corners of the eyes and the forehead region. We annotate the wrinkles in these regions with five points representing each wrinkle's shape. The wrinkle information is obtained using a bounding box placed around the annotations and only maintaining high-frequency information by difference-of-Gaussian. The wrinkle is warped in a mean shape and then transformed in pose parameters. By using the second derivative of the Lorentzian function to fit, the average of every parameter is obtained. Thus, the process leads to the transformation of the wrinkle into a vector of meaningful parameters representing the shape and appearance of the wrinkle. The parameters of the wrinkle vector include the following:

- $cx, cy$ , the center of the wrinkle;
- $d$ , the geodesic distance between first and last points;
- $a$ , angle in degrees;
- $C$ , curvature computed as least-squares minimization using Equation (1);
- $D$ , depth of the wrinkle;
- $W$ , the width of wrinkle.

$$\min \|Y - CX^2\|_2^2 \quad (1)$$

with  $Y$  (resp.  $X$ ) the coordinates (resp. abscissa) of the wrinkle centered at the origin, and with the horizontal alignment of the first and last points.

Since there can be an unequal number of wrinkles in faces, a technique is used to obtain the wrinkle information from six equally sized non-overlapping segments of the target region. For every wrinkle, an estimate of the probability density is obtained. The strategy leads to a vector which incidentally is high dimensional. The problem is mitigated by introducing an approximation of an arbitrary joint probability of random variables by computing in pairs, every joint probability for every random variable. The computations are by means of Kernel Density Estimation (KDE) with a Gaussian and a standard deviation of 1.5. Thus, a new vector containing the following approximated information for each of the six zones of every face is obtained.

- The number of wrinkles in the current zone.
- The average wrinkle.
- Densities computed by means of KDE of the wrinkles, subtracting the average wrinkle and concatenating all vectors of the individual zones of the face to represent the wrinkles in one face image.

By fusing the wrinkle information with the shape information, we hoped to obtain a term that comprehensively describes the age of the face.

#### Fusing the Facial Shape and Wrinkle Information

Since the shapes and wrinkle vectors are different properties and sizes and cannot be directly concatenated, we employ the z-score method proposed in [22,47] to normalize into a new vector we denote as  $f_j$ . In this paper, fusion is achieved by the following expression.

$$f_j = \frac{(f'_j - \mu)}{\sigma} \quad (2)$$

where  $f_j$  is the normalized vector,  $f'_j$  represents the vector before normalization,  $\mu$  is the mean and  $\sigma$  is the standard deviation. We further applied PCA to reduce the dimension of the fused vector. Subsequently, term vector  $f_j$  is utilized in our proposal.

### 3.2. Term Frequency Inverse Gravity Moment (TF-IGM)

Perceived as a gravitational inverse problem, TF-IGM is a two-part term weighting mechanism comprising the local factor (TF) and the global factor (IGM). The main idea of TF-IGM is to measure the weight of a term (feature) in a corpus (dataset) such that the more evenly a term distributes in several classes within a dataset, the less important its inter-class discriminating power is. Thus, terms have more discriminating power if they are more concentrated in one class than other classes in the dataset. As the distribution of terms in the classes is considered uneven, the concentration of the terms can be used to:

1. Distinguish different classes when the terms with stronger class distinguishing power are assigned heavier weights than the others.
2. Measure the fine-grained inter-class distribution of a term in different classes so that the obtained weight can represent the term's contribution to the classification task.
3. Provide a co-efficient to achieve optimal performance between the local and global factors contributing to the weight.

Furthermore, since in this method weights are independent of classes, the unique IGM-based global weighting factor for a term can be obtained by one-time calculation from the dataset. To the best of our knowledge, the idea was initially proposed by Chen et al. [48] for the classification of text and information retrieval. However, the base version, known as Term Frequency Inverse Document Frequency (TFIDF), was exploited for age estimation in [14], wherein the authors represented the age features as a set of terms that described the face. For a dataset of facial images, they determined the extent to which the features corresponded the age. The proposal demonstrated state-of-the-art performance. However, TFIDF has since been improved to TF-IGM [48]. In this proposal, we leverage the class distinguishing abilities of TF-IGM and hope to utilize it for juvenile age estimation.

### 3.3. Establishing Ordinal Relationships among the Features (Terms)

To compute the inter-class distribution concentration of feature  $f_j$ , we sort in descending order all the frequencies of  $f_j$ 's occurrence in the individual age classes as follows.

$$f_{j1} \geq f_{j2} \geq \dots \geq f_{jn} \quad (3)$$

where  $f_{jr}$  ( $R = 1, 2, \dots, n$ ) is the frequency of  $f_j$  occurring in the  $r$ -th age class after sorting, and  $n$  is the number of age classes. If  $f_{jr}$  is proposed for the frequency of a feature (TF); then for the sorted list, the far-left, which is heavier, exerts more weight to the left than the far-right, which is comparatively lighter. Hence, the center position is biased to the left. As the feature keeps occurring in fewer age classes, the balance follows the left shift until it is at the starting position where  $R = 1$ . At this point, the feature has occurred in only one age class. Similarly, at the position where  $(n/2)$ , the point is considered as the "gravity center" of the overall inter-class distribution. The following expression depicts the uniform inter-class distribution of the feature.

$$f_{j1} = f_{j2} = \dots = f_{jn}$$

Clearly, it is intuitive that the position of the gravity center can reflect the inter-class distribution concentration of a feature and ultimately contribute to the inter-class distinguishing power of the feature.

### 3.4. Inverse Gravity Moment (IGM)

However, for the class-specific gravity  $f_{jr}$ , if it is ranked  $R$ , and the distance from  $R$  to the beginning 0 is observed, then the gravity moment is the product of the class-specific gravity and the rank  $R$  expressed as follows.

$$GM = f_{jr} \cdot R \tag{4}$$

In line with the TF-IGM concept, the distributed concentration of the feature is proportional to the reciprocal value of the total gravity moment. Hence, the following expressing is employed to measure the inverse gravity moment.

$$IGM(f_j) = \frac{f_{j1}}{\sum_{r=1}^n (f_{jr} \cdot R)} \tag{5}$$

where  $IGM(f_j)$  denotes the inverse gravity moment of feature  $f_j$ ;  $f_{jr} (r = 1, 2, \dots, n)$  are the frequencies of  $f_j$ 's occurrence in the various age classes sorted in descending order and  $R$  is the rank.

Typically, the inverse gravity moment of a feature's inter-class distribution is within  $2/((1+n) \cdot n)$  and 1.0. Since in the descending order the first element,  $f_{j1}$ , is the maximum in the list of  $\{f_{jr} | R = 1, 2, \dots, n\}$ , Equation (5) can further be expressed as follows.

$$IGM(f_j) = \frac{1}{\sum_{r=1}^n \frac{f_{jr}}{\max_{1 \leq j \leq n} (f_{jr})} \cdot R} \tag{6}$$

Consequently, Equation (6) depicts that the IGM of a feature is the sum of the gravity moment computed from the normalized frequencies of the feature's occurrence in the individual classes. Since the IGM value can be transformed to fall within the range of  $[0, 1.0]$ , the minimum value is close to zero. Therefore, the basic IGM model defined in Equation (5) is adopted in our proposal.

The preceding satisfies the two conditions of the label distribution approach stated in Section 2.2. Additionally, it demonstrates that the ordinal relationship is implicit in the TF-IGM mechanism.

### 3.5. Measuring Weight by TF-IGM

With TF-IGM, the weight of a feature in a sample image is determined by its frequency in the face image and its contribution to the classification, corresponding to the local factor (TF) and the global factor (IGM) respectively. Thus, a feature's contribution to classification depends on its class distinguishing power, which is reflected by its inter-class distribution concentration. The higher the level of concentration, the greater the weight assigned to the feature. The former can be measured by the IGM model expressed by Equation (5).

#### 3.5.1. Limitations of IGM

However, empirical tests revealed some cases of redundancies in the computation of the weight based on the IGM factor. For instance, in the case of five features  $f_{j1}, f_{j2}, f_{j3}, f_{j4}$  and  $f_{j5}$  having individual feature frequencies (TF) in six age classes corresponding to  $\{100, 100, 0, 0, 0, 0\}$ ,  $\{40, 40, 0, 0, 0, 0\}$ ,  $\{23, 23, 0, 0, 0, 0\}$ ,  $\{11, 11, 0, 0, 0, 0\}$  and  $\{2, 2, 0, 0, 0, 0\}$  respectively. If there are 100 samples in each age class, then standard IGM computes the weight of each feature as 0.333, regardless of the sorted distinguishing power being  $f_{j1} > f_{j2} > f_{j3} > f_{j4} > f_{j5}$ . We refer to this as case 1. Additionally, consider the case of five features  $f_{j6}, f_{j7}, f_{j8}, f_{j9}$  and  $f_{j10}$  and a dataset of five age classes each with 10 samples and corresponding feature frequencies (TF) of  $\{10, 0, 0, 0, 0\}$ ,  $\{8, 0, 0, 0, 0\}$ ,  $\{5, 0, 0, 0, 0\}$ ,  $\{3, 0, 0, 0, 0\}$  and  $\{1, 0, 0, 0, 0\}$  respectively. Intuitively, the order of the class distinguishing power should be  $f_{j6} > f_{j7} > f_{j8} > f_{j9} > f_{j10}$ . However, per Equation (5), the IGM values are all the same as 1.0. This can be considered as case 2. Obviously, computing the weight based on the standard IGM does not fully depict the distinguishing power, since in both cases, features with different frequencies are assigned the same values. In addressing these limitations, we introduced a common logarithm denoted by  $K$  to the standard IGM equation such that  $K = \log_{10}[S_{total}(f_{j(max)})/S_{f_{j(max)}}]$ , where  $S_{total}(f_{j(max)})$  is



the total face samples in the age class, in which  $f_j$  occurs the most and  $S_{f_j(max)}$  is the number of the face samples in the age class in which  $f_j$  occurs the most. For clarity, we denote the intervention as  $GMO$  expressed as follows.

$$GMO(f_j) = \frac{f_{j1}}{\sum_{r=1}^n f_{jr} \cdot R + K} \tag{7}$$

As shown in Table 1, the weights computed by  $GMO$ , using Equation (7), show unique values for features  $f_{j1}$  through  $f_{j10}$ . This shows that, whereas in case 1, the IGM values are the same for the different features, the values as computed by  $GMO$  are unique. This is indicative of the distinguishing abilities of  $GMO$  over IGM.

**Table 1.** Comparison between weights computed by IGM and  $GMO$ .

Case	F	Class/Sample	F/Frequency	Sorted Desc. Order	IGM Value	GMO Value
Case 1	$f_{j1}$	6 / 100	{100, 100, 0, 0, 0, 0}	$f_{j1} > f_{j2} > f_{j3} > f_{j4} > f_{j5}$	0.333	0.333
	$f_{j2}$	do	{40, 40, 0, 0, 0, 0}		0.333	0.332
	$f_{j3}$	do	{23, 23, 0, 0, 0, 0}		0.333	0.330
	$f_{j4}$	do	{11, 11, 0, 0, 0, 0}		0.333	0.323
	$f_{j5}$	do	{2, 2, 0, 0, 0, 0}		0.333	0.259
Case 2	$f_{j6}$	5 / 10	{10, 0, 0, 0, 0}	$f_{j6} > f_{j7} > f_{j8} > f_{j9} > f_{j10}$	1.0	1.0
	$f_{j7}$	do	{8, 0, 0, 0, 0}		1.0	0.988
	$f_{j8}$	do	{5, 0, 0, 0, 0}		1.0	0.943
	$f_{j9}$	do	{3, 0, 0, 0, 0}		1.0	0.852
	$f_{j10}$	do	{1, 0, 0, 0, 0}		1.0	0.5

Note: F = feature(term). F/frequency = frequency of the term. desc. = descending.

Consequently, Equation (8) is proposed to compute the weight as follows.

$$W(f_j) = TF(f_j, s_k) \cdot (1 + \lambda \cdot GMO(f_j)) \tag{8}$$

where  $TF$  represents the local weighting factor,  $GMO$  is the global weighting factor and  $\lambda$  is a co-efficient that maintains the balance between the two factors.  $W$  is the weight and  $s_k$  is the sample image.

Intuitively, the local weighting factor should be lowest since a feature occurring 68 times in an age class is generally less than 68 times as important as a feature that rarely occurs [48]. Similar to the method in [49], the TF was reduced by the introduction of square root to Equation (8), resulting in new weight expressed by Equation (9).

$$w(f_j) = [TF(f_j, s_k)]^{\frac{1}{2}} \cdot (1 + \lambda) \cdot [GMO(f_j)] \tag{9}$$

We utilize the new  $GMO$ -based weight for our proposed scheme.

#### 4. The Proposed Method

The overall illustration of our proposal is shown in Figure 2. As can be seen, facial images annotated with the 68 facial landmark points serve the system, which obtains a combination of shape and wrinkles information, and represents it as terms that can be used describe the age. In order to explore the distinguishing power of the terms, TF-IGM is used to assign weights to the term. Next, the computed weights are used to build a matrix of predictive probabilities, which is utilized for the detection of the age. Consequently, we introduced a scheme we name LaGMO, which is a combination of facial landmark points and TF-IGM for juvenile age detection.

The following is an encapsulation of the key tasks that constitute LaGMO.

**Definition 1.** Transferring Landmark points to landmark-term vectors (features):

Given the 68 facial landmark points, we denote AAM of an image  $img$  as vector  $LM_{(img)} = (x_1, x_2, x_3, x_4, \dots, x_{135}, x_{136})$ , where  $x_1, x_2$  is the first landmark point,  $x_3, x_4$  the second landmark point and  $x_{135}, x_{136}$  is the last landmark point. For the vector  $LM_{(img)}$ , we transferred the points into a string form such that the  $j^{th}$  member represents the  $j^{th}$  string of the form  $F - j - LOC$ , where  $F$  represents the appearance information,  $j$  is the  $j^{th}$  element and  $LOC = \text{round}(\frac{x_j}{width})$ . Width is hard coded. In line with the process described in Section 3.1, we obtained a compact term denoted by  $f_j$ .

**Definition 2.** Obtaining term dictionary: If for the whole space there are  $n$  different terms, then they constitute a dictionary denoted as  $TDIC = (f_{j1}, f_{j2}, \dots, f_{jn})$ . Then for any arbitrary  $f_j$ , we assign weight by Equation (9), resulting in a new vector denoted as  $LMM_{(img)} = w(f_{j1}), w(f_{j2}), \dots, w(f_{jn})$ , where the  $n^{th}$  element is between 0 and 1. The bigger the weight the higher the importance of the weight in the class. Then we define a weighted term matrix as follows.

$$wMtrx = w(f_j)(TDIC_{size}) \cdot |n| \tag{10}$$

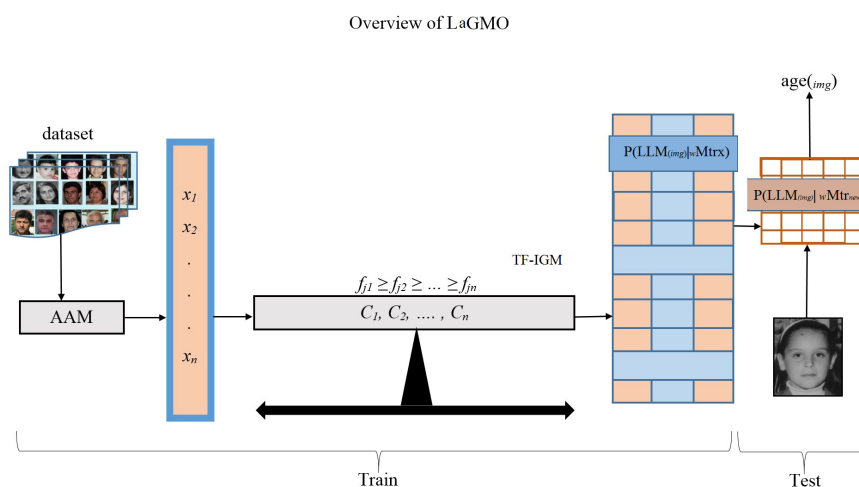
where  $w(f_j)$  is the weight of the term,  $TDIC_{size}$  is the size of the dictionary and  $|n|$  is the total number of age classes.

**Definition 3.** Establishing the relationship between the features and ages: To establish the relationship between the features and ages, we classified all  $LMM_{(img)}$  samples into different age classes according to the age denoted as  $c = \{c_1, c_2, \dots, c_n\}$  with  $n$  representing the  $n^{th}$  class. All samples in the class have the same ground-truth age  $c$ .

**Definition 4.** Restricting the matrix table to correspond to juvenile age: Since our proposal considers the relative order of vector  $LMM_{(img)}$  as age labels, we treated  $LMM_{(img)}$  as labels of the order  $LMM_{(img)} \in \{1, 2, \dots, n\}$ , where  $n$  is the number of age classes based on the entire dataset. We reduced the dataset to correspond to juveniles using Equation (11).

$$\begin{aligned} X_n^+ &= \{(LMM_{(img)}, c_n) | c_n > n\}, \\ X_n^- &= \{(LMM_{(img)}, c_n) | c_n \leq n\} \end{aligned} \tag{11}$$

Next, we resolved  $X_n^-$  into a new dataset denoted by  $tdic$  and utilized Equation (10) to build the new search table we denote  $wMtrx_{new}$ . By this strategy, we avoided the rather exhaustive search throughout the larger table, thereby reducing the overhead in both time and space.



**Figure 2.** Overview of the proposed LaGMO.  $P(LLM_{(img)}|wMtrx_{new})$  is the predictive matrix.

### Age Prediction

Generally, age estimation aims to predict age based on the mapping from an input vector to an output age space. In this proposal, for an unlabeled facial image  $img$ , we projected its landmark feature vector  $LMM_{(img)}$  into the age group space using the matrix  $wMtrx_{new}$ . Equation (12) was utilized for the projection where the projected value represents the closeness of the image belonging to the class; the bigger the value, the higher the possibility of assigning the age to the age class.

$$Project(img) = wMtrx_{new} \cdot LMM_{(img)} \quad (12)$$

The following Algorithm 1 presents the LaGMO age estimation process.

---

**Algorithm 1:** The LaGMO age estimation process.

---

```

1  $LMM_{(img)}$  complete;
2 While  $tdic$  do;
3 Train;
   Input:  $LMM_{(img)}$ 
   Output:  $wMtrx$ 
4 for ( $LMM_{(img)} \in wMtrx$ ), ...,  $n$  do;
5 compute new term matrix by Equation (10);
6 end for;
7 Save  $wMtrx_{new}$ ;

8 Test;
   Input:  $img$ 
   Output:  $age(img)$ 
9 for  $LMM_{(img)} \in wMtrx_{new}$ ), ...,  $n$  do;
10 Estimate age by Equation (12);
11 Save  $age(img)$ ;
12 end for;
   Output:  $age(img)$ 
13 end while;

```

---

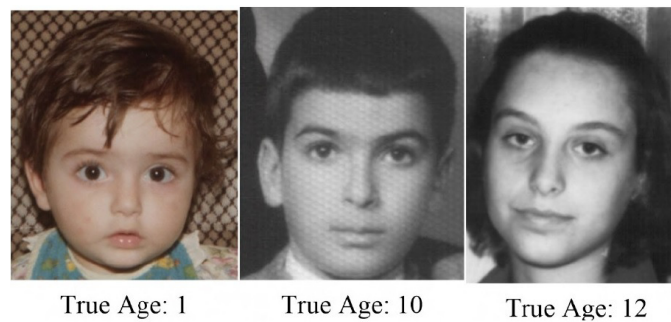
## 5. Experimentation

The image dataset, which is the data basis for experimentation, can be challenging to obtain due to privacy, quality and other issues. There are currently many facial-image datasets with age labels, but most of them are unavailable or do not contain enough juveniles. Therefore, we only used the Face and Gesture Recognition Research Network (FG-NET) database for experimentation. FG-NET which is publicly available, is already annotated, and has a large number of juvenile images.

### 5.1. Datasets and Evaluation

FG-NET consists of 1002 facial photographs from 82 individuals with ages ranging from 0 to 69 years. Although the number seems small, there are at least 12 age-separated images per person, and a substantial majority of the images are younger faces within the 0–40 age group. We collected images from ages 0 to 34 years to constitute our dataset. Since the FG-NET contains only 1002 faces, it could not be divided into the typical 80–20, train–test strategy. Therefore, we adopted the leave-one-person-out (LOPO) cross-validation strategy. In order to train our model, the dataset was divided into the various age classes, such that each class had approximately the same number of samples. This strategy maintained balance in the dataset. The training was first conducted on the entire set with ages ranging from 0–34 years to obtain a large reference matrix. Since the focus is on juveniles,

we reduced the set to reflect ages 0–17 years, as depicted in Figure 3. Consequently, we utilized the new dataset to construct a new reference matrix.



**Figure 3.** Sample faces from FG-NET.

To validate the performance of our scheme, the Mean Absolute Error (MAE) was used to indicate the effectiveness of our proposal. MAE is expressed as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Ag_1 - Ag_0| \quad (13)$$

where  $N$  is the total number of test data,  $Ag_0$  the actual age and  $Ag_1$  is the estimated age.

Additionally, we investigated the accuracy by another metric known as the Cumulative Score (CS) expressed as follows.

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

where  $N$  is the total number of test images and  $N_{e \leq th}$  is the number of test images whose absolute error is less than  $th$ .

## 5.2. Performance Evaluation

We conducted two tests—one to indicate the effect of reducing the local factor (TF) on the weight and the other to compare our scheme with similar approaches.

### 5.2.1. The Effect of the TF Factor on the Weight

In order to verify the effect of reducing the TF factor on the weight, we randomly selected a small sample of images and utilized Equations (8) and (9) to compute the weight. To show clarity, we denote the two computations of the weight: *standardTF*( $W$ ) and *lowTF*( $w$ ) corresponding to the weight with standard local factor and weight with reduced local factor. We noticed from the initial rough checks, a significant difference between the two, as shown in Figure 4 and Table 2. This suggests that reducing the TF factor could improve the performance of our scheme.

**Table 2.** The effect of reduced TF.

Method	CS
lowTF ( $w$ )	86.33
standardTF ( $W$ )	81.83

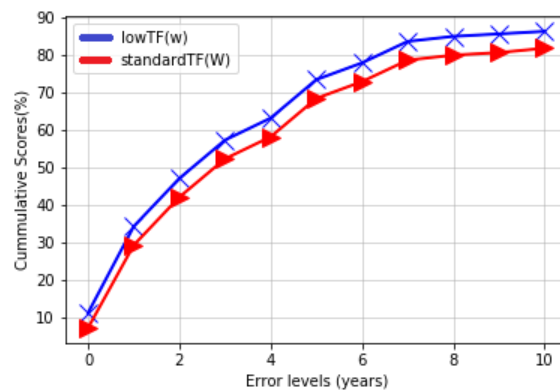


Figure 4. The effect of reduced TF.

### 5.2.2. Comparison with Similar Approaches

To validate our scheme, we investigated our proposal against the few approaches that offer similar solutions, including Chen et al. [14] (LM+TFIDF), Wang et al. [13] (LBP+SVM) and Kohli et al. [44] (EOC+CTAF). We compared LaGMO with LM+IDF based on the approach utilizing facial landmark points and the TF-IDF term weighting scheme. Regardless of LM+TFIDF focusing on age estimation in general, Table 3 and Figure 5 illustrate that LaGMO out-performed LM+IDF. We further compared our scheme with LBP+SVM. Although the LBP+SVM proposal focuses on detecting juveniles from adults, only texture-based features obtained from the coordinates of the facial landmark points were considered for feature representation. Regardless, it remains the only proposal that is fully dedicated to juvenile detection. As illustrated in Table 3, LaGMO performed better than LBP+SVM. Regarding EOC+CTAF, the authors proposed it for age estimation in general. However, the approach incorporated a channel for child age estimation and reported an impressive CS of 95.0 and a MAE of 2.69. We assumed the performance was due to the age threshold being 21 years for EOC+CTAF but 17 years for LaGMO. Therefore, we took on the challenging task of adjusting the threshold in EOC+CTAF to 17 years. Interestingly, we observed degradation in the performance of EOC+CTAF. Finally, with a CS of 89.86% and a MAE of 4.42, LaGMO demonstrated performance that was state-of-the-art for juvenile age estimation.

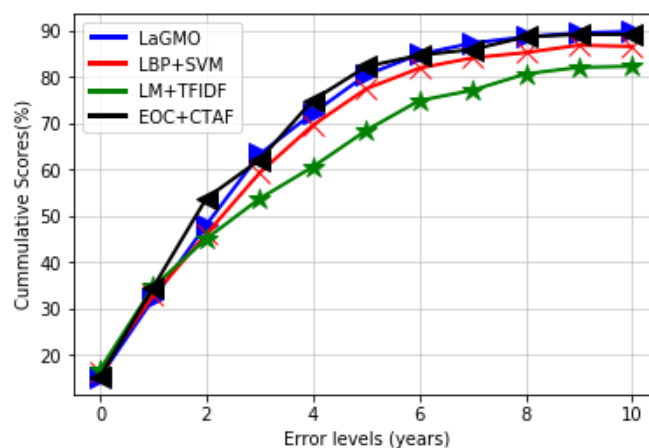


Figure 5. Comparison with similar approaches.

**Table 3.** Comparison with similar approaches (CS and MAE).

Method	CS	MAE
AAM+Indexing [15]	65	–
LM+TFIDF [14]	82.6	6.14
LBP+SVM [13]	88.6	–
EOC+CTAF [44]	89.2 (with age range 0–17 years)	–
AAM+LMFBP [25]	–	4.95
LaGMO	89.86	4.42

## 6. Conclusions and Future Work

This proposal characterized juvenile aging cues based on the 68 facial landmark points of the Active Appearance Model, where the shape and appearance features were presented as terms that described the age of the face. The scheme effectively exploited the new term weighting scheme known as Term Frequency Inverse Gravity Moment (TF-IGM), first to establish the ordinal relationship among the terms in the various age classes and ultimately to compute the weights of the terms for the classification task. The implicit ability of TF-IGM to establish ordinal relationships made it possible to demonstrate impressive performance, even with limited datasets. Therefore, this proposal demonstrates that facial landmark points can be applied to juvenile age detection. Accordingly, an age estimation scheme called LaGMO, which is the combination of facial landmark points and TF-IGM, was presented to alleviate the lack of juvenile age estimation schemes. We hope to extend the method to cover the adult aging subspace and utilize more datasets in the future.

**Author Contributions:** E.N.A.H., conceptualization, methodology, software, writing—original draft, writing—review and editing; S.Z., funding acquisition, resources, supervision; H.C. supervision, resources; Q.L., software, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Sichuan Science and Technology Program under grants 2018GZ0085 and 2019YFG0399.

**Acknowledgments:** We are immensely grateful to the associate editor and reviewers for their invaluable suggestions, and to the lead supervisor for his guidance.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Ilyas, M.; Fournier, R.; Othmani, A.; Nait-Ali, A. BiometricAccessFilter: A Web Control Access System Based on Human Auditory Perception for Children Protection. *Electronics* **2020**, *9*, 361, doi:10.3390/electronics9020361. [CrossRef]
- Rosenbloom, A.L. Age estimation based on pictures and videos presumably showing child or youth pornography. *Int. J. Leg. Med.* **2015**, *129*, 621–622, doi:10.1007/s00414-014-1048-3. [CrossRef]
- Schmeling, A.; Reisinger, W.; Geserick, G.; Olze, A. Age estimation of unaccompanied minors: Part I General considerations. *Forensic Sci. Int.* **2006**, *159*, S61–S64, doi:10.1016/j.forsciint.2006.02.017. [CrossRef] [PubMed]
- Torres, M.T.; Valstar, M.F.; Henry, C.; Ward, C.; Sharkey, D. Postnatal gestational age estimation of newborns using Small Sample Deep Learning. *Image Vis. Comput.* **2019**, *83–84*, 87–99, doi:10.1016/j.imavis.2018.09.003. [CrossRef]
- IOM. Child-and-Young-Migrants. 2019. Available online: <https://migrationdataportal.org/themes/child-and-young-migrants> (accessed on 4 July 2020)
- Machado, C.E.P.; Flores, M.R.P.; Lima, L.N.C.; Tinoco, R.L.R.; Franco, A.; Bezerra, A.C.B.; Evison, M.P.; Guimarães, M.A. A new approach for the analysis of facial growth and age estimation: Iris ratio. *PLoS ONE* **2017**, *12*, 1–19, doi:10.1371/journal.pone.0180330. [CrossRef] [PubMed]

7. Dalessandri, D.; Tonni, I.; Laffranchi, L.; Migliorati, M.; Isola, G.; Visconti, L.; Bonetti, S.; Paganelli, C. 2D vs. 3D Radiological Methods for Dental Age Determination around 18 Years: A Systematic Review. *Appl. Sci.* **2020**, *10*, 3094, doi:10.3390/app10093094. [[CrossRef](#)]
8. Daunay, A.; Baudrin, L.G.; Deleuze, J.F.; How-Kit, A. Evaluation of six blood-based age prediction models using DNA methylation analysis by pyrosequencing. *Sci. Rep.* **2019**, *9*, 8862, doi:10.1038/s41598-019-45197-w. [[CrossRef](#)]
9. Ferguson, E.; Wilkinson, C. Juvenile age estimation from facial images. *Sci. Justice* **2017**, *57*, 58–62, doi:10.1016/j.scijus.2016.08.005. [[CrossRef](#)]
10. Al-Shannaq, A.S.; Elrefaei, L.A. Comprehensive Analysis of the Literature for Age Estimation From Facial Images. *IEEE Access* **2019**, *7*, 93229–93249, doi:10.1109/ACCESS.2019.2927825. [[CrossRef](#)]
11. Lanitis, A.; Taylor, C.J.; Cootes, T.F. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 442–455, doi:10.1109/34.993553. [[CrossRef](#)]
12. Mansouri, N. Automatic Age Estimation: A Survey. *Comput. Syst.* **2020**, *24*, 877–889, doi:10.13053/CyS-24-2-3317. [[CrossRef](#)]
13. Wang, X.; Liang, Y.; Zheng, S.; Wang, Z. Juvenile detection by LBP and SVM. In Proceedings of the 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, IEEE CCIS 2012, Hangzhou, China, 30 October–1 November 2012; Volume 3, pp. 1324–1327, doi:10.1109/CCIS.2012.6664600. [[CrossRef](#)]
14. Chen, Y.W.; Lai, D.H.; Qi, H.; Wang, J.L.; Du, J.X. A new method to estimate ages of facial image for large database. *Multimed. Tools Appl.* **2016**, *75*, doi:10.1007/s11042-015-2485-9. [[CrossRef](#)]
15. Kwon, Y.H.; da Vitoria Lobo, N. Age classification from facial images. In Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR 1994, Seattle, WA, USA, 21–23 June 1994; IEEE: San Francisco, CA, USA, 1994; pp. 762–767, doi:10.1109/CVPR.1994.323894. [[CrossRef](#)]
16. Cootes, T.F.; Edwards, G.J.; Taylor, C.J. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 681–685, doi:10.1109/34.927467. [[CrossRef](#)]
17. Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J. Active Shape Models—Their Training and Application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59, doi:10.1006/cviu.1995.1004. [[CrossRef](#)]
18. Geng, X.; Zhou, Z.; Zhang, Y.; Li, G.; Dai, H. Learning from facial aging patterns for automatic age estimation. In Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, CA, USA, 23–27 October 2006; Nahrstedt, K., Turk, M., Rui, Y., Klas, W., Mayer-Patel, K., Eds.; ACM: New York, NY, USA, 2006; pp. 307–316, doi:10.1145/1180639.1180711. [[CrossRef](#)]
19. Geng, X.; Zhou, Z.; Smith-Miles, K. Automatic Age Estimation Based on Facial Aging Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2234–2240, doi:10.1109/TPAMI.2007.70733. [[CrossRef](#)] [[PubMed](#)]
20. Ramanathan, N.; Chellappa, R. Modeling shape and textural variations in aging faces. In Proceedings of the 2008 8th IEEE International Conference on Automatic Face Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–8, doi:10.1109/TPAMI.2007.70733. [[CrossRef](#)]
21. Gao, F.; Ai, H. Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method. In *Advances in Biometrics*; Tistarelli, M., Nixon, M.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 132–141, doi:10.1007/978-3-642-01793-3\_14. [[CrossRef](#)]
22. Ross, A.; Govindarajan, R. Feature level fusion of hand and face biometrics. *Proc. SPIE* **2005**, *5779*, doi:10.1117/12.606093. [[CrossRef](#)]
23. Ng, C.; Yap, M.H.; Costen, N.; Li, B. An investigation on local wrinkle-based extractor of age estimation. In Proceedings of the 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 5–8 January 2014; Volume 1, pp. 675–681.
24. Baker, A.; Ugail, H.; Connah, D. Automatic age and gender classification using supervised appearance model. *J. Electron. Imaging* **2016**, *25*, 1–11, doi:10.1117/1.JEI.25.6.061605. [[CrossRef](#)]
25. Ouloul, I.M.; Moutakki, Z.; Afdel, K.; Amghar, A. Improvement of age estimation using an efficient wrinkles descriptor. *Multim. Tools Appl.* **2019**, *78*, 1913–1947, doi:10.1007/s11042-018-6275-z. [[CrossRef](#)]
26. Luu, K.; Ricanek, K.; Bui, T.D.; Suen, C.Y. Age estimation using Active Appearance Models and Support Vector Machine regression. In Proceedings of the 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Washington, DC, USA, 28–30 September 2009; pp. 1–5, doi:10.1109/BTAS.2009.5339053. [[CrossRef](#)]

27. Hu, Z.; Wen, Y.; Wang, J.; Wang, M.; Hong, R.; Yan, S. Facial Age Estimation with Age Difference. *IEEE Trans. Image Process.* **2017**, *26*, 3087–3097, doi:10.1109/TIP.2016.2633868. [[CrossRef](#)]
28. Lanitis, A.; Draganova, C.; Christodoulou, C. Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2004**, *34*, 621–628, doi:10.1109/TSMCB.2003.817091. [[CrossRef](#)]
29. Liu, X.; Zou, Y.; Kuang, H.; Ma, X. Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network. *Symmetry* **2020**, *12*, 146, doi:10.3390/sym12010146. [[CrossRef](#)]
30. Carletti, V.; Greco, A.; Percannella, G.; Vento, M. Age from faces in the deep learning revolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. [[CrossRef](#)]
31. Kishor, M.; Addepalli, S.; Bhurchandi, K. Age estimation using local direction and moment pattern (LDMP) features. *Multimed. Tools Appl.* **2019**, *78*, 30419–30441, doi:10.1007/s11042-019-7589-1. [[CrossRef](#)]
32. Taheri, S.; Toygar, Ö. Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images. *IET Biom.* **2019**, *8*, 124–133, doi:10.1049/iet-bmt.2018.5141. [[CrossRef](#)]
33. Wu, Y.; Ji, Q. Facial Landmark Detection: A Literature Survey. *Int. J. Comput. Vis.* **2019**, *127*, 115–142, doi:10.1007/s11263-018-1097-z. [[CrossRef](#)]
34. Efraty, B.A.; Papadakis, M.; Profitt, A.; Shah, S.; Kakadiaris, I.A. Facial component-landmark detection. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–25 March 2011; pp. 278–285, doi:10.1109/FG.2011.5771411. [[CrossRef](#)]
35. Yan, T.; Liu, X.; Wheeler, F.W.; Tu, P. Automatic facial landmark labeling with minimal supervision. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2097–2104, doi:10.1109/CVPR.2009.5206670. [[CrossRef](#)]
36. Pamplona Segundo, M.P.; Silva, L.; Bellon, O.R.P.; Queirolo, C.C. Automatic Face Segmentation and Facial Landmark Detection in Range Images. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2010**, *40*, 1319–1330, doi:10.1109/CVPR.2009.5206670. [[CrossRef](#)]
37. Su, K.; Geng, X. Soft Facial Landmark Detection by Label Distribution Learning. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, HI, USA, 27 January–1 February 2019; AAAI Press: Palo Alto, CA, USA, 2019; pp. 5008–5015, doi:10.1609/aaai.v33i01.33015008. [[CrossRef](#)]
38. Chen, H.; Gao, M.; Fang, B. An improved active shape model method for facial landmarking based on relative position feature. *Int. J. Wavelets Multiresolut. Inf. Process.* **2017**, *15*, 1–14, doi:10.1142/S0219691317500084. [[CrossRef](#)]
39. Seshadri, K.; Savvides, M. Robust modified Active Shape Model for automatic facial landmark annotation of frontal faces. In Proceedings of the 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Washington, DC, USA, 28–30 September 2009; pp. 1–8, doi:10.1109/BTAS.2009.5339057. [[CrossRef](#)]
40. Panis, G.; Lanitis, A.; Tsapatsoulis, N.; Cootes, T.F. Overview of research on facial ageing using the FG-NET ageing database. *IET Biom.* **2016**, *5*, 37–46, doi:10.1049/iet-bmt.2014.0053. [[CrossRef](#)]
41. Pontes, J.; Britto, A.; Fookes, C.; Koerich, A. A flexible hierarchical approach for facial age estimation based on multiple features. *Pattern Recognit.* **2016**, *54*, 34–51, doi:10.1016/j.patcog.2015.12.003. [[CrossRef](#)]
42. Geng, X.; Yin, C.; Zhou, Z. Facial Age Estimation by Learning from Label Distributions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2401–2412, doi:10.1109/TPAMI.2013.51. [[CrossRef](#)]
43. He, Z.; Li, X.; Zhang, Z.; Wu, F.; Geng, X.; Zhang, Y.; Yang, M.; Zhuang, Y. Data-Dependent Label Distribution Learning for Age Estimation. *IEEE Trans. Image Process.* **2017**, *26*, 3846–3858, doi:10.1109/TIP.2017.2655445. [[CrossRef](#)] [[PubMed](#)]
44. Kohli, S.; Prakash, S.; Gupta, P. Hierarchical age estimation with dissimilarity-based classification. *Neurocomputing* **2013**, *120*, 164–176, doi:10.1016/j.neucom.2012.08.069. [[CrossRef](#)]
45. Dhimar, T.; Mistree, K. Feature extraction for facial age estimation: A survey. In Proceedings of the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 23–25 March 2016; pp. 2243–2248, doi:10.1109/WiSPNET.2016.7566541. [[CrossRef](#)]
46. Martin, V.; Séguier, R.; Porcheron, A.; Morizot, F. Face aging simulation with a new wrinkle oriented active appearance model. *Multim. Tools Appl.* **2019**, *78*, 6309–6327, doi:10.1007/s11042-018-6311-z. [[CrossRef](#)]
47. Jain, A.K.; Nandakumar, K.; Ross, A. Score normalization in multimodal biometric systems. *Pattern Recognit.* **2005**, *38*, 2270–2285, doi:10.1016/j.patcog.2005.01.012. [[CrossRef](#)]



48. Chen, K.; Zhang, Z.; Long, J.; Zhang, H. Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert Syst. Appl.* **2016**, *66*, 245–260, doi:10.1016/j.eswa.2016.09.009. [[CrossRef](#)]
49. Tenenbaum, J.B.; Freeman, W.T. Separating Style and Content. In Proceedings of the Advances in Neural Information Processing Systems 9, NIPS, Denver, CO, USA, 2–5 December 1996; Mozer, M., Jordan, M.I., Petsche, T., Eds.; MIT Press: Massachusetts Avenue, Cambridge, MA, USA, 1996; pp. 662–668.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).