

Applied Artificial Intelligence

An International Journal

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/uaai20>

A Novel Hybrid Feature Framework for Multi-View Age Estimation

A. Annie Micheal & P. Geetha

To cite this article: A. Annie Micheal & P. Geetha (2021) A Novel Hybrid Feature Framework for Multi-View Age Estimation, Applied Artificial Intelligence, 35:15, 1361-1387, DOI: 10.1080/08839514.2021.1979181

To link to this article: <https://doi.org/10.1080/08839514.2021.1979181>



Published online: 17 Sep 2021.



Submit your article to this journal [↗](#)



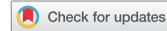
Article views: 515



View related articles [↗](#)



View Crossmark data [↗](#)



A Novel Hybrid Feature Framework for Multi-View Age Estimation

A. Annie Micheal and P. Geetha

Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

ABSTRACT

Facial age estimation has grasped the attention of numerous researchers in recent times. It is a challenging task as a consequence of illumination, pose variations, occlusion, complex background, facial expression, and facial makeup. Estimating the age of an individual with an arbitrary pose is quite a challenging job because most of the age estimation system focuses on the frontal view. In this paper, a novel framework for multi-view age estimation by amalgamating the local and global features is proposed. A novel texture feature, Median Gradient Ternary Pattern is proposed in this paper. The Pseudo Zernike Moment extracts the shape features and the View-based Active Appearance Model constructs an appearance model from the facial images. Further, all three features are combined into a feature vector by executing feature-level fusion. The dimension of the combined feature is reduced using Principal Component Analysis. Multi-class Support Vector Machine is utilized to divide the images into four poses. For each pose, a Support Vector Regression with RBF kernel is applied to train a model for estimating the actual age of an individual. The proposed methodology is performed on two databases, namely, FG-NET and CACD which showcase eminent performance.

ARTICLE HISTORY

Received 31 March 2021
Revised Vxx
Accepted 1 September 2021

Introduction

The face of a human embraces lots of data identified with individual qualities, for instance, age, identification, race, emotion, and sex. As of now, age can assume an imperative part in several applications, for instance, age-based access control, age prediction systems to find out the lost kids, internet access control, crime investigation, age-based retrieval of face images from the search engine, and vending machine. For example, it can keep away minors from getting liquor or cigarettes from vending machines, or it can convey digital content depending on age to adapt messages accordingly. Facial aging is a confounded procedure and it is irreversible and slow. The face growth and aging forms in human beings' lifetime constitute two phases namely newborn

to adulthood and from adulthood to old age. In the initial phase, craniofacial change is tremendous and the least change in skin texture. The later stage exhibits a rapid change in skin texture. Intrinsic (i.e., ethnicity, heredity, sexual orientation) and external factors (i.e., environment, lifestyle) can impact aging. All these components cause difficulty to accurately estimate a person's age, even by people. Additionally, different perturbations on facial pictures, for example, lighting, facial expression, posture, occlusion, blur, facial cosmetics, plastic surgery, make the age estimation even more difficult. Among these factors, estimating the age of a person with an arbitrary pose is the most troublesome issue. In this paper, a novel framework for multi-view age estimation by utilizing hybrid features is proposed. The hybrid features, which is a mixture of local and global features, have gotten adequate consideration since the flaws found in individual global and local feature can be made up using this strategy. The human pose is partitioned into three rotations: $[-20^\circ, +20^\circ]$ pitch, roll (360°) and yaw $[-90^\circ, +90^\circ]$. To simplify the multi-pose issue and without loss of generality, only yaw rotation is taken into account. Yaw rotation represents the face turned to the left and the right side. VAAM is a global feature and they offer greater amounts of information concerning the texture and shape of a face. VAAM (Cootes, Walker, and Taylor 2000) is utilized to construct an appearance model which embodies both the shape and texture variability in the images. Four different models are trained on labeled images for four different poses. Every single example pose can then be approximated using the suitable appearance model with a vector of parameters, c . As there is a change in orientation, the parameters approximately trace out an elliptical path. PZM (Pang, Teoh, and Ngo 2015) relies on the global information of the image and it is used as a shape feature descriptor in numerous pattern recognition applications. PZM is robust to image quantization error and provides better feature representation capability. The proposed Median Gradient Ternary Pattern (MGTP) is a local feature that amalgamates the advantage of Local Gradient Pattern (LGP) (Jun and Kim 2012) and Median Ternary Pattern (MTP) (Khan et al. 2013) in the same compact encoding scheme. In MGTP, the pixel values in the neighborhoods of the specified central pixel are replaced with the gradient values and the center pixel value is substituted by the median of the gradient values, which give rise to a better representation of the large-scale structure and more robust to random noises. Based on the local median and a threshold value defined by the user, a three-value (1, 0, -1) ternary code is obtained for every single pixel. This three-level coding pattern produces consistent texture patterns with the local image property and a certain degree of magnitude difference information is incorporated in the feature representation which makes it more robust in the presence of non-monotonic illumination variation. All the three features are fused using a feature level fusion and given as an input to PCA for dimensionality reduction. Estimating the age from a multi-view facial image is more

complex and designing a feature invariant to orientation is very difficult. For that reason, the input image is partitioned into four poses according to face orientations. Therefore, the multi-view problem is disintegrated into easier classification tasks. Multi-class SVM is used to classify the reduced feature into four poses according to its orientation. Images in the range $[0^\circ \sim +45^\circ]$ are grouped as right-half profile (PO1), images in the range $[+45^\circ \sim +90^\circ]$ are grouped as right profile (PO2), images in the range $[0^\circ \sim -45^\circ]$ are grouped as left-half profile (PO3), and images in the range $[-45^\circ \sim -90^\circ]$ are grouped as left profile (PO4). For each pose, SVR with RBF kernel is used to determine the age of an individual.

Our research is novel in the accompanying three courses, contrasted with the past works. Firstly, a new framework is proposed for multi-view age estimation. Secondly, the age estimation accuracy is incredibly enhanced by merging local and global features. Thirdly, a novel texture feature descriptor is proposed to boost the performance of the hybrid feature.

The rest of this paper is structured in this fashion: Section 2 elaborates the related work on age estimation. The proposed framework is explained in Section 3. Section 4 gives details about the experimental results of the proposed approach and Section 5 summarizes our work with future directions.

Related Works

There has been an enormous number of researches for facial age estimation. Kwon and Lobo (1999) calculated geometric distance amongst various facial features. The age was grouped into 3 categories, namely, babies, adults, and seniors depending on the size of the chosen facial characteristics and the number of facial wrinkles. However, various facial orientation, accurate estimation of top of the skull and robustness to facial hair was not explored. Thukral et al. (Thukral, Mitra, and Chellappa 2012) used five classifiers such as μ -SVC, Fisher Linear Discriminant, Partial Least Squares (PLS), Naive Bayes, and Nearest Neighbor to classify the age. The age of an individual was estimated by the Relevance Vector Machine. The framework lacks the incorporation of texture features. In the Aging Pattern Subspace (Geng, Zhou, and Smith-Miles 2007), the aging pattern is described as a grouping of face pictures from a similar individual which is arranged in a time-based order. FG-NET database was used to analyze the performance of the approach and accomplished a Mean Absolute Error (MAE) of 6.77 years. The preprocessing involved in the methodology relies on landmark points. The face size varies across ages, the methodology lacks face size information during preprocessing. However, it is difficult to collect numerous face pictures of a similar person at different ages. Li et al. (Li et al. 2012) predicted the age of an individual by concentrating on ordinal discriminative feature learning. Various feature selection methods, such as Laplacian Score, Rank Boost, Least Angle

Regression (LAR), Piecewise Linear Orthonormal (PLO), and Fisher Score were utilized to reduce the dimension of the information from both the ordinal information and locality information by decreasing the rank correlation and non-linear correlation. Amongst all the feature selection techniques, the PLO was considered the best as it achieved an accuracy of 88.0%. Guo and Mu (2011) utilized the Kernel Partial Least Squares regression (KPLS) to estimate the age of a person. The dimension of the feature was reduced by KPLS and the aging function was also learned. The dimensionality of the original space was reduced by KPLS by finding a petite number of latent variables which can increase the performance of the method. A craniofacial growth model was built by Ramanathan and Chellappa (2006). This model was used to characterize the facial shape growth variations by facial landmarks. Their framework only classified a facial image into 4 categories, namely, babies, young adults, and seniors, but the exact age of a person was not estimated. The framework does not take into account texture model, facial hair, and change in baby fat. Chang et al. (Chang, Chen, and Hung 2011) separated the age estimation issue as a series of sub-problems of binary classifications by using ordinal hyperplane ranking (OH Ranker). Gao and Ai (2009) utilized the Gabor feature and fuzzy Linear Discriminant Analysis (LDA) and categorized the age into four groups like baby, child, adult, and old. 5408 pictures were used for training and 91% accuracy was achieved. Age estimation was viewed as a regression problem by Yan et al. (Yan et al. 2007) with non-negative label intervals and the issue was handled using definite programming. They solved this regression problem by introducing an EM algorithm and the optimization process was speeded up. Rather than learning a specific aging pattern for every person, a typical aging trend or pattern can be gained from numerous people of various ages. Yan et al. (Yan et al. 2008) introduced a patch-based regression method and the regression error was reduced. Each image was programmed as a group of order less coordinate patches of GMM (Gaussian Mixture Model) distribution. Then, a patch-kernel was designed which illustrate the Kullback-Leibler divergence among the derived models. To improve its discriminating power, inter-modality similarity synchronization was utilized. Kernel regression was utilized to estimate the age of an individual. The coordinate patch models geometric regions indicate head pose moments and is robust to image occlusion thereby contributing better performance. Khanmohammadi et al. (Khanmohammadi, Ghaemi, and Samadi 2013) extracted the texture and wrinkle features utilizing the Local Binary Pattern (LBP) and Histograms of Oriented Gradients (HOG). The age was categorized into 4 Groups: Underage (0–18), Young-age (19–35), Middle-age (36–59), and Old-age (60–94). This algorithm provides better accuracy with the restriction of a wide-ranging age group. Yasumoto et al. (Yasumoto, Niwa, and Koshimizu 2002), the age prediction was done by extracting the wrinkles from the face. 300 images captured under controlled conditions

ranging from 14–65 years were used for training. To enrich the wrinkles features, histogram equalization was done on the facial skin region. DTHT (Digital Template Hough Transform) was utilized to extract the wrinkles which are longer and shorter. The age was estimated by utilizing a look-up table. 72% of accuracy was achieved for age estimation. It was difficult to extract the wrinkles on females because of the presence of facial make-up. Zhang and Yeung (2010) formulated the inference of each person's age as a Warped Gaussian Process (WGP) estimation problem and developed a multi-task extension of WGP to tackle the issue. Since different individuals have different aging processes, personalization is valuable for age estimation. The major advantage of the framework is, the model parameters θ and σ are learned automatically without using model selection. Malek et al. (Malek, Azimifar, and Boostani 2017) estimated the age of a person using Zernike Moment (Mahesh, Noel, and Raj 2018). Both local and global characteristics of a face are captured by the Zernike Moment and work well even in the presence of makeup and accessories. The features extracted using Zernike Moment are fed to three classifiers, namely, K-Nearest Neighbor (KNN), Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP) neural network. The system was evaluated on the FG-NET dataset and obtained a MAE of 4.21, 6.21, and 9.35 for MLP, SVR, and KNN classifiers, respectively. The main advantage of Zernike Moment is scale, shift and rotation -invariant of orthogonal moments. Sahoo and Banka (2018) developed an age estimation system that preserved personalized aging traits and was robust to change in appearance, texture, pose, shape, expression, wrinkle, and illumination. Shape, wrinkle, and texture features are extracted using Active Shape Model (ASM), Gabor filter, and Multi-block Local Binary Pattern (MBLBP) respectively. Canonical Correlation Analysis (CCA) was used to combine two or more independent features. The age of a person was estimated by a three-level SVM and SVR-based hierarchical classifier by choosing a suitable combination of hybrid features at different levels. The experimental results show that a MAE of 4.05 years and 3.26 years were obtained for FG-NET-AD and MORPH Album-2 database. The lack of labeled facial images for age estimation was addressed by Dong et al. (Dong, Lang, and Feng 2019). The ordinal relationships among various age labels are learned with limited training samples. A structured sparse multiclass classification model was employed. The structured sparse regularization encodes the ordinal relationships among various age labels. The samples with similar age labels are close to each other in the feature space. The methodology was evaluated using two datasets namely, FG-NET and MORPH II datasets, and achieved a MAE of 4.25. Local Direction and Moment Pattern (LDMP) was proposed to utilize the textural and directional variation due to aging (Sawant, Addepalli, and Bhurchandi 2019). The orientation is encoded in eight directions. The magnitudes of higher order moments are encoded into texture information. The orientation and texture

are amalgamated into a feature descriptor. A warped Gaussian process regression is applied to estimate the age. LDMP preserves the wrinkles and shapes with the directional filters. The experiment was performed with FG-NET and MORPH II databases achieving a MAE of 4.6. Guehairia et al. (Guehairia et al. 2020) used Deep Random Forest (DRF) to address the problem of image-based age estimation. DRF is an ensemble of decision trees ensembles tied in a cascade form. FG-NET, PAL, and MORPH II databases were used to assess the performance of the system and obtained a MAE of 3.82.

The Proposed Framework

The architecture of the proposed multi-view age estimation framework is shown in Figure 1.

Face Detection

Face detection is the initial process of age estimation (V. Kumar, Namboodiri, and Jawahar 2015). Dense-SIFT is extracted from the images and a k-means based vocabulary is generated followed by feature quantization. Term frequencies (TF) and inverse document frequencies (IDF) are obtained and inverted files are constructed alike Bag-of-words (BOW) retrieval scheme. During testing, all the exemplars jointly take part in the Hough-based voting process that utilizes the spatial locations of features to locate the faces in a given image (Gall et al. 2011). The input image and face detection output are shown in Figure 2a,b. The detected face is given as an input to the feature extraction process.

Feature Extraction

View-based Active Appearance Model

The combination of shape and texture variation are encoded in an appearance model. Based on the PCA, the shape and the texture model are learned independently from the training images. PCA on texture and shape data yield a parametric face model that portrays new faces along with learned faces. 68 landmark points are generated on the eyes, chin, mouth, and nose. The facial images with 68 landmark points are shown in Figure 3a,b.

A model which depicts the shape variation is build using PCA on the whole set of landmark points. A shape is portrayed as a vector of coordinates from the landmark points. The facial shape is represented as s and texture is represented as t . The shape and the texture are controlled by the appearance model parameters, c , according to

$$s = \bar{s} + Q_s c, \quad (1)$$

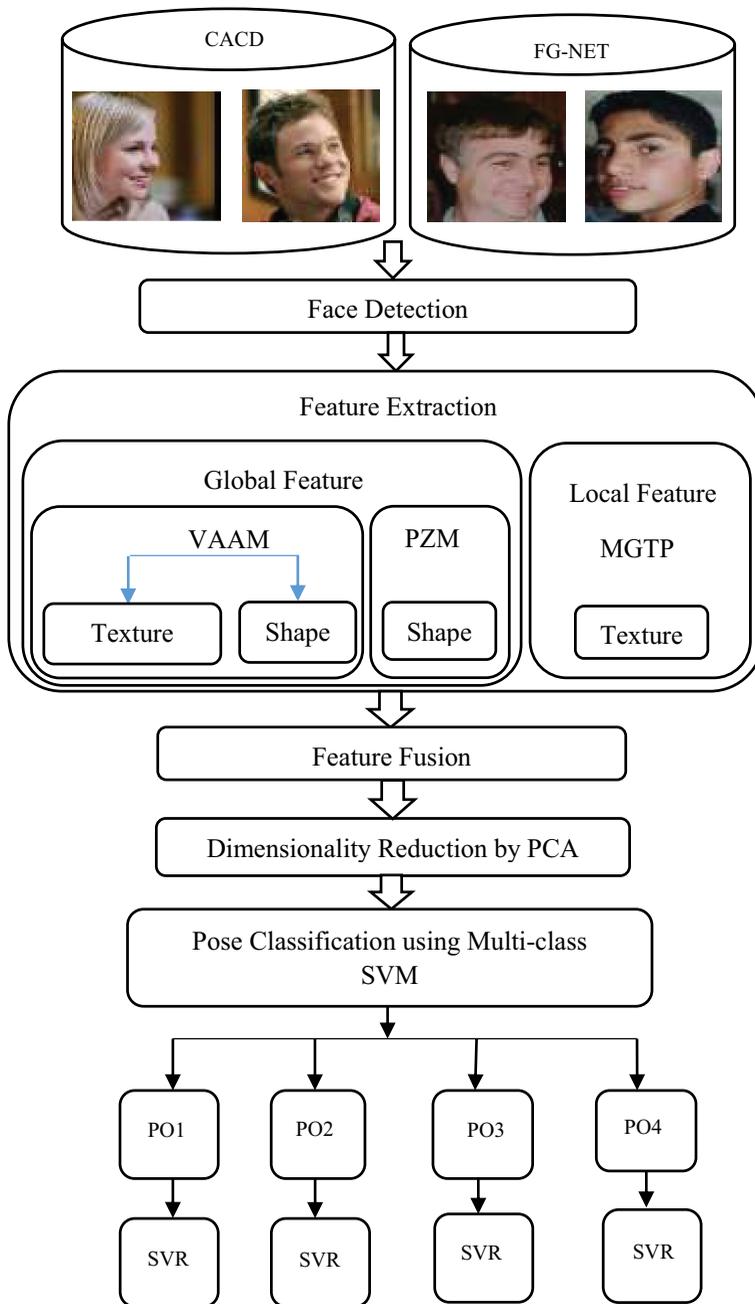


Figure 1. Architecture for the proposed multi-view age estimation.



Figure 2. (a) Input image (b) Detected face.

$$t = \bar{t} + Q_t c, \quad (2)$$

where \bar{s} indicates the mean shape, \bar{t} indicates the mean texture and Q_s and Q_t matrices portraying the types of variation from the training set. Q_s and Q_t are formed from the Eigenvectors of the covariance matrices of the data. A sample image can be synthesized for a given c by developing a texture image from the vector t and warping it by utilizing the control points portrayed by s .

Pseudo Zernike Moment

PZM is a statistical-based feature extraction technique, which relies on the global information of the image to extract the feature vector elements. The pseudo-Zernike basis is a set of an orthogonal polynomial, which is depicted as follows

$$Z_{ab}(x, y) = S_{ab}(x, y)e^{ib\theta}, \quad (3)$$



Figure 3. Facial images with 68 landmark points.

where the polynomials are well-defined over the polar coordinates inside a unit circle ($x^2 + y^2 \leq 1$), $\theta = \tan^{-1}(y/x)$, is the angle between this vector and x -axis, $a \geq 0$ is the order of PZM and b is the repetition of PZM ($|b| \leq a$), S_{ab} is called radial polynomial and it is defined as

$$S_{ab}(x, y) = \sum_{c=0}^{a-|b|} D_{a,|b|,c} (x^2 + y^2)^{\frac{a-c}{2}}, \quad (4)$$

$$D_{a,|b|,c} = (-1)^c \frac{(2a + 1 - c)!}{c!(a - |b| - c)!(a - |b| - c + 1)!}, \quad (5)$$

The PZM for an image $f(x, y)$ is calculated as

$$PZM_{ab} = \frac{a + 1}{\pi} \sum_x \sum_y f(x, y) Z_{ab}^*(x, y), \quad (6)$$

where $Z_{ab}^*(x, y) = Z_{ab}(a, b)$ and PZM is calculated for the positive value of b . PZM is robust to slight changes in shape and noise. [Figure 4\(a\)](#) shows the pseudo-Zernike polynomials with orders $a = 0$ to 30. For a more clear understanding, [Figure 4\(b\)](#) shows the first 4 pseudo-Zernike polynomials.

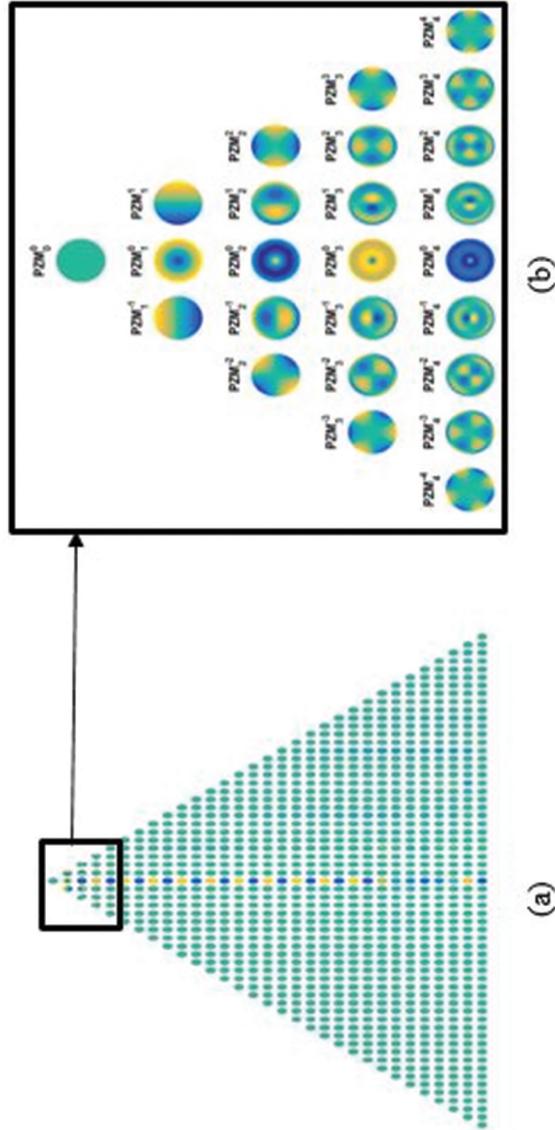


Figure 4. (a). Pseudo-Zernike polynomial for order 0 to 30 (b). Pseudo-Zernike polynomial for orders 0 to 4.

Median Gradient Ternary Pattern

A novel texture descriptor called Median Gradient Ternary Pattern is proposed in this paper. The MGTP operator makes utilization of the neighborhood gradient values of the specified central pixel. The neighborhood pixel values of the given central pixel are replaced with the gradient values and the center pixel value is substituted by the median of the gradient values. With a specified pixel as the center, a circle with radius R is considered with N sampling points on the circle. The gradient value between the center pixel i_c and its neighboring pixel i_n is characterized as $g_n = |i_n - i_c|$, and the median of N gradient values is calculated (M_G). The MGTP at a location (x, y) can be defined as

$$MGTP = \sum_{n=0}^{N-1} s(g_n)3^n,$$

$$s(g) = \begin{cases} 1, & g > M_G + t \\ 0, & M_G - t \leq g \leq M_G + t \\ -1, & g < M_G - t \end{cases} \quad (7)$$

where g is the gradient gray level, t is a threshold defined by the user and n runs over the 8 neighbors of the center pixel so that the length of the resulting histogram (including the bin-0 location) is 3^8 . Hence, $3^8 = 6561$ bins are required to represent the values in a histogram. This increases the computational complexity, so to reduce the computational complexity, each MGTP code is divided into its corresponding positive (P_{MGTP}) and negative binary codes (N_{MGTP}). Thereby, the bin number is reduced from 3^8 to $2 \times 2^8 = 512$. The “positive” code, P_{MGTP} is calculated as:

$$P_{MGTP} = \sum_{n=0}^{N-1} S_P(s(g_n))2^n,$$

$$S_P(g) = \begin{cases} 1, & g = 1 \\ 0, & otherwise \end{cases} \quad (8)$$

The “negative” code, N_{MGTP} is calculated as:

$$N_{MGTP} = \sum_{n=0}^{N-1} S_N(s(g_n))2^n,$$

$$S_N(g) = \begin{cases} 1, & g = -1 \\ 0, & otherwise \end{cases} \quad (9)$$

The histograms generated from P_{MGTP} and N_{MGTP} are then concatenated to create the final MGTP histogram. This combined histogram is then used as a final feature vector. Figure 5 shows the procedure for calculating the MGTP.

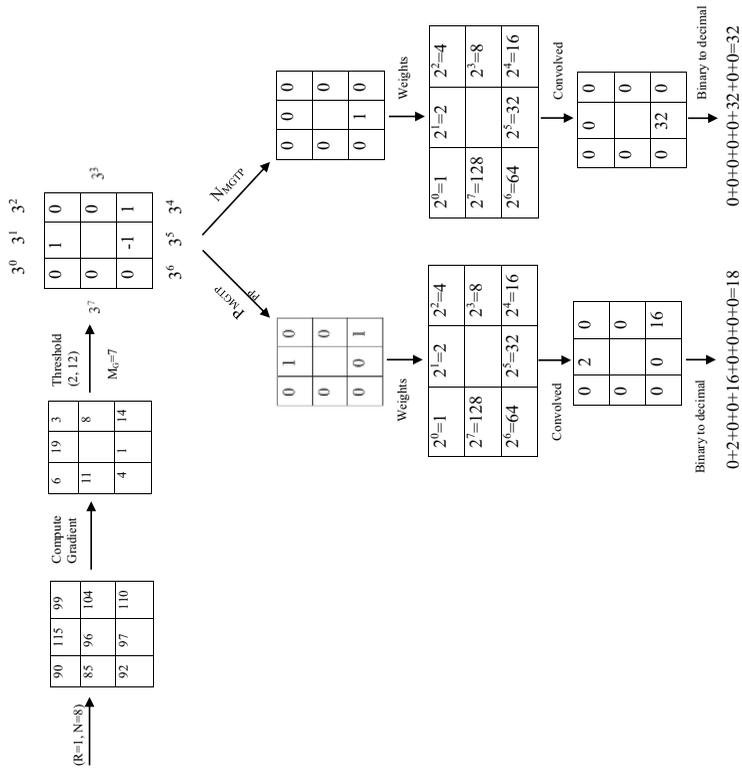


Figure 5. Illustration of MGTP for a 3×3 neighborhood.

Feature Fusion

Feature level fusion, decision level fusion, and score level fusion are used to amalgamate multiple features in biometrics (Ross and Jain 2003). Combining features using feature-level fusion comprises richer information about the raw data (Ross and Govindarajan 2004). Hence, feature-level fusion is used to amalgamate the local and global features. After extracting the three features, namely, VAAM (f_{VAAM}), PZM (f_{PZM}), and MGTP (f_{MGTP}), feature normalization is carried out using z-score normalization as,

$$f'_i = \frac{f_i - \mu_i}{\sigma_i}, \quad (10)$$

where f_i indicates the i^{th} feature vector, μ_i indicates the mean and σ_i indicates the standard deviation of the i^{th} feature vector and f'_i indicates the normalized feature vector. The normalized features are concatenated to create the fused feature as

$$f = f_{VAAM}f_{PZM}f_{MGTP}, \quad (11)$$

The feature dimension increases due to concatenation, therefore PCA is used to reduce the feature dimension.

Principle Component Analysis

The PCA finds a lower-dimensional subspace that gives promising details of age estimation. The PCA method find out the embedding that maximizes the projected variance given underneath

$$Z_{opt} = \arg \max_{Z=1} Z^T S Z, \quad (12)$$

where $S = \sum_{i=1}^n (f_i - \bar{f})(f_i - \bar{f})^T$ is the scatter matrix, \bar{f} is the mean vector of training data and f_i is the i^{th} feature vector with D dimension. A set of d D eigenvectors (Z) related to the d largest eigenvalues of S is attained. The dimensionality reduction is performed by calculating $U_i = Z^T f_i$ where $U = [u_1, u_2, \dots, u_n]$ is the projected new feature with dimension d .

Multi-class SVM

Support Vector Machine (SVM) (Cortes and Vapnik 1995) was initially designed to perform binary classification. Later, the classification was expanded to multi-class classification by using methods such as one-against-all, one-against-one, etc (Hsu and Lin 2002). Multi-class SVM is designed as a pose classifier based on the “one-against-all” technique to classify the images into four poses. Four binary SVM classifiers are formed with a non-linear radial basis kernel, where each classifier distinguishes one pose class from the

remaining three pose classes. The m training points of the form $(u_1, v_1), \dots, (u_m, v_m)$ where $u_i \in R^n$ represents the feature vector of the i^{th} sample, and $v_i \in \{1, \dots, k\} (k = 4)$ represents the class label of u_i , the k^{th} SVM solves the following problem:

$$\min_{w_k, b_k, \xi^k} \frac{1}{2} \|w_k\|^2 + C \sum_{i=1}^m \xi_i^k$$

,

$$\text{s.t } \tilde{y}_i (w_k^T u_i + b_i) \geq 1 - \xi_i^k,$$

$$\xi_i^k \geq 0, i = 1, \dots, m \tag{13}$$

where $\tilde{y}_i = 1$ if $y_i = k$ and $\tilde{y}_i = -1$ otherwise. The decision function is given by $f_k(u) = w_k^T \cdot u + b_k$. A new sample u will be classified in the class which attains the highest value of $f_k(u)$, i.e. u is in k^* - th class when $f_{k^*}(x) = \max\{f_k(x) : k = 1, 2, 3, 4\}$. If more than one class attains this maximum, then the new sample u will be classified in the class related to the lowest index k^* by convention.

Support Vector Regression (SVR) With RBF Kernel

Given the m training points $\{(u_1, v_1), \dots, (u_m, v_m)\}$, $u_i \in R^n, v_i \in R, i = 1, \dots, m$ where u_i denotes sample i with label v_i . The SVR predicts a function $v = f(u)$, which is defined as:

$$f(u) = w\phi(u) + b, \tag{14}$$

where $\phi((u))$ is the feature vector of an input image, which is mapped from space u to a higher-dimensional space, b is the bias and w is the vector for the regression coefficient. SVR concurrently tries to maximize the margin width and minimize the regression error. The optimal regression function is termed as

$$\min_{w, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i^+ + \xi_i^-)$$

,

$$\text{Subject to } \begin{cases} v_i - w_i\phi(u_i) + b \leq \varepsilon + \xi_i^+ \\ w_i\phi(u_i) + b - v_i \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \leq 0 \end{cases} \tag{15}$$

where $C > 0$ finds out the trade-off between the flatness of f and data deviations, and ξ_i^+, ξ_i^- are slack variables. A radial basis function is defined in Equation. (16)

$$k(u, u') = e^{-\gamma u - u'^2}, \quad (16)$$

where γ is a constant to adjust the width of the Gaussian function. Given the kernel mapping, the solution of the non-linear SVR is attained as

$$\langle w, u \rangle = \sum_{i=1}^m (\alpha_i - \alpha_i^*), k(u_i, u), \quad (17)$$

and

$$f(u) = \sum_{i=1}^m (\alpha_i - \alpha_i^*), k(u_i, u) + b, \quad (18)$$

where α_i, α_i^* are the Lagrange multipliers.

Algorithm 1 explains the steps involved in the proposed multi-view age estimation framework.

Algorithm 1: Multi-view age estimation algorithm

Input: Face image from the dataset

Output: Estimated age of the input image

- 1: **for** $i = 1 \rightarrow n$ **do** \blacktriangleright n is the total number of images in the dataset
- 2: The face region is detected using an exemplar-based method
- 3: Crop the face and convert it into a grayscale image
- 4: Compute the VAAM feature (f_{VAAM})
- 5: Landmark points are marked on each face
- 6: The training images are divided into four pose sets.
- 7: **For each pose**
- 8: Align landmark points with Procrustes Analysis
- 9: Apply PCA to the aligned landmark points and obtain the mean shape and shape model
- 10: Wrap each training image landmarks to the learned mean shape, acquiring a “shape-free patch.”
- 11: Apply PCA to the warped images and obtain the mean texture and texture model.
- 12: Apply PCA to the concatenated shape and texture parameters to acquire a combined appearance model.

$$s = \bar{s} + Q_s c$$

$$t = \bar{t} + Q_t c$$

13: **end for**

14: The PZM (f_{PZM}) of order a and repetition b for an image $f(x, y)$ is formulated as

$$PZM_{ab} = \frac{a+1}{\pi} \sum_x \sum_y f(x,y) Z_{ab}^*(x,y)$$

15: Extract the MGTP feature (f_{MGTP})

16: Compute the gradient value between the center pixel i_c and its neighboring pixel i_n as $g_n |i_n - i_c|$

17: The MGTP descriptor is computed as $MGTP = \sum_{n=0}^{N-1} s(g_n) 3^n$

18: The MGTP code is divided into “positive ($P_{MGTP}^{n=0}$)” and “negative (N_{MGTP})” binary codes.

19: Concatenate P_{MGTP} and N_{MGTP} to form a feature vector

20: Combine f_{VAAM} , f_{PZM} and f_{MGTP} by feature level fusion

21: Reduce feature dimension using Principle Component Analysis.

22: Multi-class SVM classifies the reduced feature into four poses ($PO1$ ($[0^\circ \sim +45^\circ]$), $PO2$ ($[+45^\circ \sim +90^\circ]$), $PO3$ ($[0 \sim -45^\circ]$), $PO4$ ($[-45^\circ \sim -90^\circ]$),).

23: **for** each pose, $PO = 1 \rightarrow 4$

24: Estimate the age using SVR Based RBF kernel

25: **end for**

26: **end procedure**

Experimental Details and Result Analysis

FG-NET (Lanitis and Tsapatsoulis 2015) and Cross-Age Celebrity Dataset (CACD) (Chen, Chen, and Hsu 2015) are the benchmark databases for age estimation and it comprises pictures caught in the wild condition. FG-NET database comprising of 1,002 face pictures belonging to 82 subjects with an age range of 0–69 years. All the images are annotated with pose information, landmark points, and age information and are subjective to pose variation, illumination, and facial expression. Cross-Age Celebrity Dataset (CACD) is the largest public cross-age database of 2000 celebrities were collected from the internet. This database contains 163,446 pictures with an age range of 16–62 years. The images are subjected to pose variation, facial makeup, and illumination variation. The celebrities’ age can be ascertained by subtracting the birth year from the time at which the picture was taken. The summary of the database used in our experiment is shown in Table 1. Sample facial images from FG-NET, CACD with their age for yaw rotation are shown in Figure 6a, b. MATLAB is used for implementing the proposed methodology. Table 1. The database used in our experiment

The face detection process is done using an exemplar-based method, which utilizes an accumulation of discriminatively trained exemplars for detection. Every single exemplar casts a vote utilizing retrieval framework and generalized Hough voting, to discover the faces in the test picture. The detected face is cropped and changed into a grayscale image. The feature extraction strategy

Table 1. The database used in our experiment.

Database	Images	Subjects	Age Range
FG-NET	1,002	82	0–69
CACD	163,446	2000	16–62

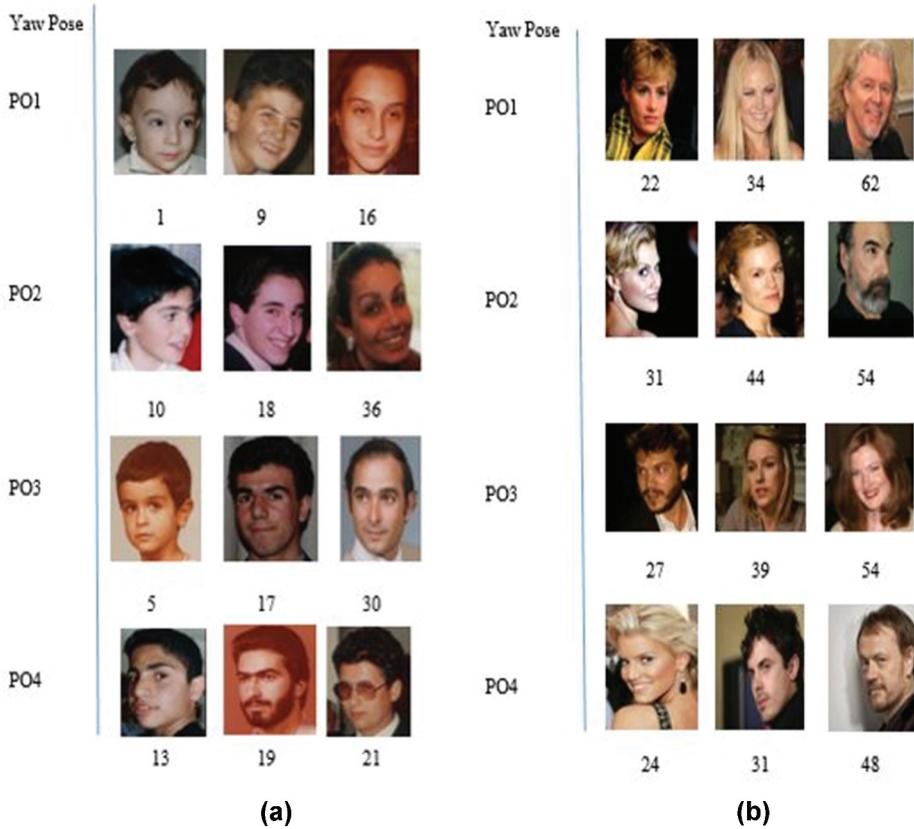


Figure 6. Images with Yaw rotation (a) FG-NET (b) CACD.

Table 2. Explanation of training and test data from the FG-NET and CACD database.

Dataset	Pose category	Total	Training	Test
FG-NET	PO1	345	249	96
	PO2	116	77	39
	PO3	413	300	113
	PO4	128	86	42
CACD	PO1	2048	1536	512
	PO2	1753	1315	438
	PO3	2159	1620	539
	PO4	1721	1291	430

Table 3. MAE (in years) for LGP using SVR.

Feature Extraction	Dataset	Parameter	Pose category				
		Sub-region	PO1	PO2	PO3	PO4	AP
LGP	FG-NET	3 × 3	9.67	10.23	9.32	9.96	11.72
		5 × 5	9.43	10.02	9.16	9.74	11.51
		7 × 7	9.21	9.86	8.92	9.48	11.24
	CACD	3 × 3	11.25	11.74	11.04	11.53	14.61
		5 × 5	11.16	11.62	10.94	11.38	14.37
		7 × 7	10.93	11.32	10.76	11.18	13.93

Table 4. MAE for MTP using SVR.

Feature Extraction	Dataset	Parameter	Pose category				
		Sub-region	PO1	PO2	PO3	PO4	AP
MTP	FG-NET	3 × 3	7.36	7.25	6.75	6.99	9.03
		5 × 5	7.16	7.01	6.58	6.78	8.84
		7 × 7	7.10	6.97	6.34	6.53	8.36
	CACD	3 × 3	8.72	9.57	8.97	9.24	10.98
		5 × 5	8.43	9.32	8.68	8.97	10.64
		7 × 7	8.29	9.04	8.47	8.78	10.21

Table 5. MAE for MGTP using SVR.

Feature Extraction	Dataset	Parameter	Pose category				
		Sub-Region	PO1	PO2	PO3	PO4	AP
MGTP	FG-NET	3 × 3	3.54	3.21	3.58	3.96	5.05
		5 × 5	3.27	3.08	3.32	3.70	4.78
		7 × 7	3.03	2.90	3.16	3.28	4.41
	CACD	3 × 3	4.39	4.17	5.48	5.69	7.01
		5 × 5	4.15	3.92	4.36	4.52	6.88
		7 × 7	3.97	3.76	4.09	4.37	6.74

takes this grayscale picture as an input. For VAAM feature representation, the face landmark points have to be given in advance. FG-NET database has been provided with the 68 landmark points for each face image but for the CACD database, the 68 landmark points were annotated manually. These 68 landmark feature points were utilized for VAAM training. Four distinct models were trained for four different poses. The Procrustes analysis is performed to align the landmark sets and to construct a shape model. To acquire a shape-free patch, each training sample landmarks are warped to the learned mean shape. The shape-free patch's texture intensities are normalized using a linear transformation. Now, PCA is utilized to construct a texture model. Finally, a correlation between the shape and the texture models are learned to build a combined model. Next, the pseudo-Zernike moment for the different order of moment (a) is calculated. In MGTP, the face images were divided into 3 × 3, 5 × 5, 7 × 7 sub-region, and the Mean Absolute Error (MAE) was computed

Table 6. MAE for VAAM using SVR.

Feature Extraction	Dataset	Pose category				
		PO1	PO2	PO3	PO4	AP
VAAM	FG-NET	3.30	4.29	3.94	3.98	5.12
	CACD	4.72	5.25	4.93	4.99	7.01

Table 7. MAE for PZM using SVR.

Feature Extraction	Dataset	Parameter Order of moment(<i>a</i>)	Pose category				
			PO1	PO2	PO3	PO4	AP
PZM	FG-NET	26	4.14	4.87	4.56	4.38	5.83
		27	3.98	4.74	4.41	4.23	5.77
		28	3.83	4.61	4.30	4.11	5.61
		29	3.70	4.57	4.21	4.06	5.46
		30	3.59	4.46	4.17	4.03	5.35
	CACD	26	5.19	5.79	5.48	5.31	7.97
		27	5.13	5.66	5.37	5.24	7.72
		28	5.03	5.52	5.22	5.16	7.59
		29	4.98	5.47	5.17	5.10	7.36
		30	4.85	5.38	5.11	5.01	7.11

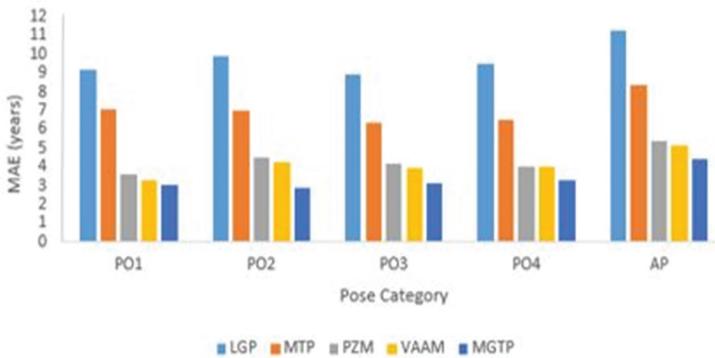


Figure 7. The comparison of MAE for LGP, MTP, VAAM, PZM, and MGTP on the FG-NET database.

for all the sub-region. The threshold value t was set to 5 and the sub-region is chosen as 7×7 as it gives less MAE. All three features are fused and fed into PCA for reducing the dimension of the feature. The parameters $C = 36$ and

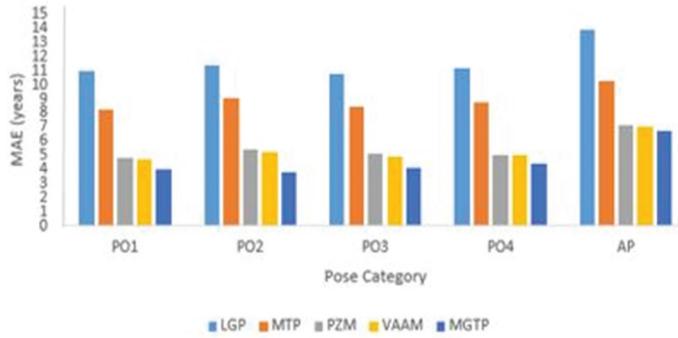


Figure 8. The comparison of MAE for LGP, MTP, VAAM, PZM, and MGTP on the CACD database.

Table 8. MAE for two feature analysis using SVR.

Feature Extraction	Dataset	Parameter	Pose category				
			PO1	PO2	PO3	PO4	AP
VAAM+ PZM	FG-NET	a = 30	3.10	4.13	3.51	3.96	4.94
	CACD		4.58	5.10	4.77	4.96	6.54
VAAM+ MGTP	FG-NET	7 × 7	2.78	2.83	2.94	3.02	4.19
	CACD		3.36	3.65	3.81	4.07	5.78
PZM+ MGTP	FG-NET	a = 30	2.85	2.88	2.96	3.12	4.25
	CACD	7 × 7	3.59	3.67	3.89	4.19	5.92

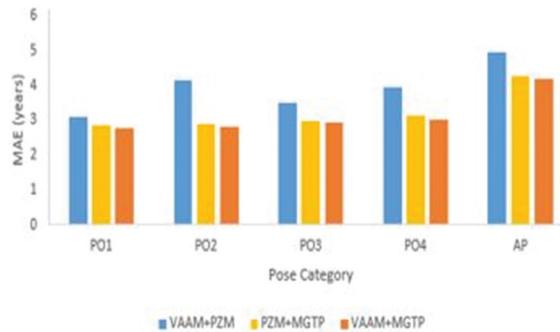


Figure 9. The comparison of MAE for two feature combinations on the FG-NET database.

$\gamma = 12$ are considered a good choice for SVR. Mean Absolute error (MAE) is used as the criteria for evaluating the performance of the proposed approach. MAE is defined in Equation. (19)

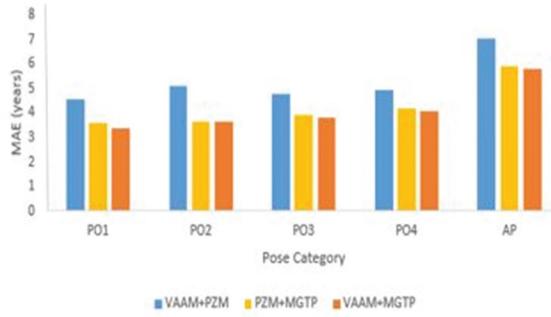


Figure 10. The comparison of MAE for two feature combinations on the CACD database.

Table 9. MAE for three feature analysis with and without PCA.

PCA	Feature Extraction	Dataset	Pose category				
			PO1	PO2	PO3	PO4	AP
Without PCA	VAAM+ PZM+ MGTP	FG-NET	2.65	2.54	2.73	2.82	3.13
		CACD	3.18	3.90	3.45	3.65	4.67
With PCA	VAAM+ PZM+ MGTP	FG-NET	2.02	2.26	2.14	2.39	2.98
		CACD	2.48	3.49	2.93	3.33	4.04

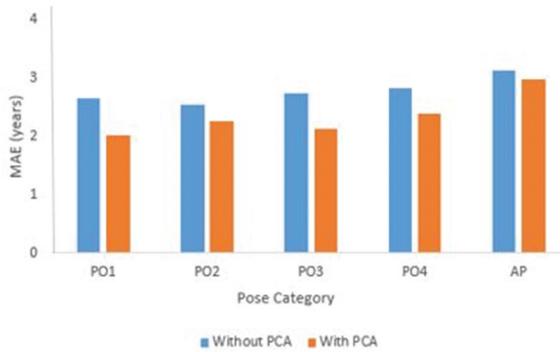


Figure 11. Comparison of MAE for the combination of three features with and without PCA on the FG-NET database.

$$MAE = \frac{\sum_{i=1}^N |\hat{x}_i - x_i|}{N}, \tag{19}$$

where \hat{x}_i and x_i are the estimated age and the true age of the i^{th} image and N is the total number of the test images. The distribution of training and test data chosen from the FG- NET and CACD database for various pose categories is shown in Table 2.

Table 2. Explanation of training and test data from the FG-NET and CACD database.

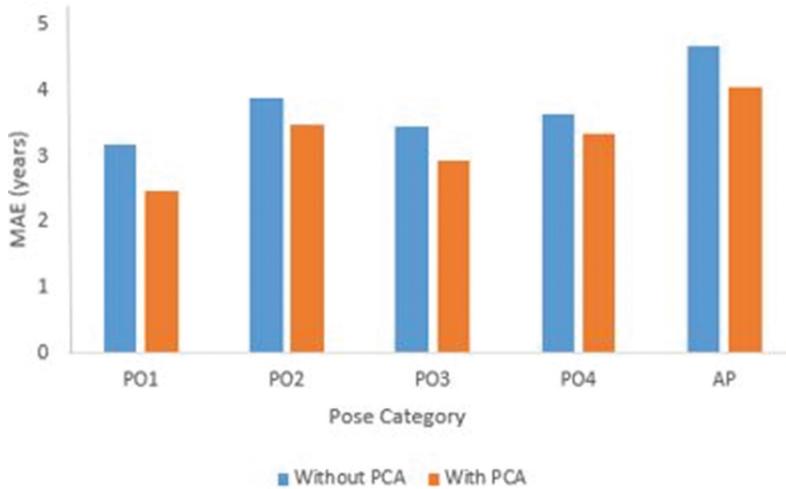


Figure 12. Comparison of MAE for the combination of three features with and without PCA on the CACD database.

Table 10. The processing time comparison for training and testing.

	FG-NET		CACD	
	Training	Testing	Training	Testing
Feature Extraction				
MGTP	1043.93 s	132.93 s	3311.81 s	1179.30 s
VAAM	4423.94 s	865.87 s	9639.26 s	5845.56 s
PZM	1650.65 s	156.70 s	4954.76 s	2034.85 s
MGTP+VAAM+PZM	8465.76 s	1558.54 s	19045.43 s	10143.76 s

Table 11. Comparison of MAE of the state-of-the-art methods and the proposed method on the FG-NET database.

Paper	Method	FG-NET
Tokola et al. (2014)	Gabor+LBP+GLCM	5.38
Han and Jain (2014)	BIF	4.5
Zou et al. (2016)	Cost-sensitive SVM and Random Forests	4.34
T. Kumar and Haider (2018)	Golden ratio, feature fusion using CCA, multi-feature-based pattern filling	4.05
Dong, Lang, and Feng (2019)	Structured sparse learning	4.25
Sawant and Bhurchandi (2019)	Warped Gaussian Process	4.41
Sawant and Bhurchandi (2019)	LDMP with probability	4.6
Guehairia et al. (2020)	Deep Random forest	3.82
Akbari et al. (2020)	DNN	3.26
Proposed Method	MGTP+PZM+VAAM	2.98

Table 12. Comparison of MAE of the state-of-the-art methods and the proposed method on the CACD database.

Paper	Method	CACD
Liu et al. (2015)	DFDNet	5.57
Tan et al. (2017)	Soft Softmax	5.19
Rothe, Timofte, and Van Gool (2018)	DEX	4.79
Tan et al. (2019)	DHAA	4.34
Shen et al. (2021)	DRF	4.61
J. Han et al. (2021)	Wasserstein-GAN	4.41
Proposed Method	MGTP+PZM+VAAM	4.04

A total of 7681 images are chosen randomly from the CACD face database. Tables 3, 4 and 5 shows the performance of the LGP, MTP, and MGTP for the sub-region (3×3 , 5×5 , 7×7) with threshold value $t = 5$ on FG-NET and CACD databases for various pose category and putting all poses in one model (AP). The bold value in the table shows the less MAE in each dataset.

The sub-region 7×7 gives better results for age estimation. The performance of VAAM on FG-NET and CACD database for various pose categories and all poses in one model is shown in Table 6. Table 7 shows the performance of PZM on the FG-NET and CACD database for various pose categories and all poses in one model. The outcomes indicate that the PZM descriptor with an order of moment ($a = 30$) provides better results. The comparison of MAE for LGP, MTP, VAAM, PZM, and MGTP on FG-NET and CACD databases are shown in Figure 7 and 8.

Table 8 shows the two feature analysis using SVR with $a = 30$ for PZM and sub-region (7×7) for MGTP. We observe that the MAE for the combination of VAAM and MGTP is less when compared to the other combination of features. The comparison of MAE for two features combination on FG-NET and CACD are shown in Figure 9 and 10.

Three features are combined and the performance of the proposed framework is analyzed with PCA and without PCA and the analysis is shown in Table 9. The experimental result demonstrates that the system performs better with PCA for FG-NET and CACD databases. The MAE is compared for the combination of three features with and without PCA on CACD and FG-NET database are shown in Figure 11 and 12.

The training and testing time comparison for FG-NET and CACD databases is shown in Table 10. Table 11 and Table 12 shows the performance of our methodology compared with the state-of-the-art age estimation algorithms on FG-NET and CACD databases respectively. Our approach outperforms the state-of-the-art methods on both FG-NET and CACD databases. Figure 13a,b shows the age estimation results for the CACD and FG-NET databases.

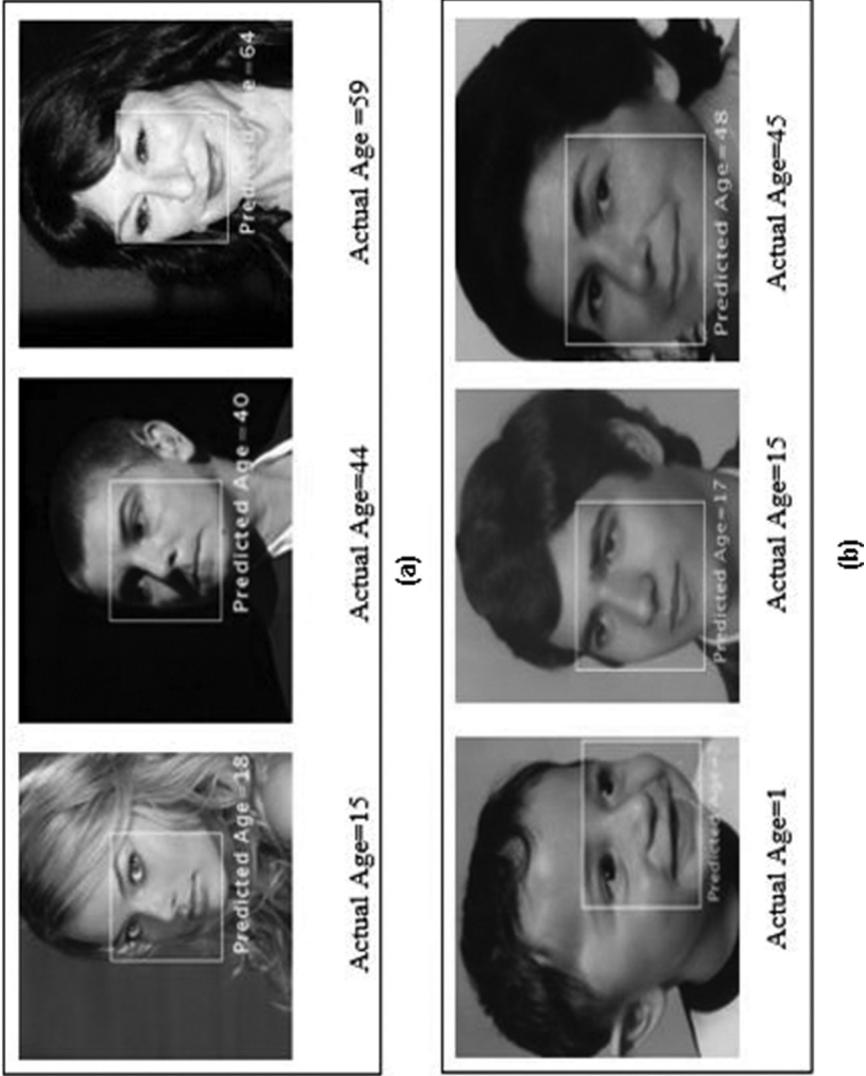


Figure 13. Age estimation output (a) CACD (b) FG-NET.

Conclusion

In this paper, a novel framework for multi-view age estimation by combining the local and global features is proposed. Also, a novel descriptor termed MGTP is proposed to extract the texture features which give rise to a better representation of the large-scale structure and more robust to random noises and non-monotonic illumination variation. The shape feature of a facial image is extracted using PZM. An appearance model which embodied both shape and texture variation is constructed using VAAM. Multi-class SVM is used to partition the images into four poses and for each pose, SVR is used to find the exact age of an individual. The performance of the proposed framework is analyzed for various pose categories by trying different parameters for MGTP and PZM. The experimental outcomes indicate that partitioning the images into different poses effectively estimates the ages for the facial images from the FG-NET and CACD database using a combination of local and global features than putting all the poses in one model. In future work, real-time database will be used and the effect of age will be studied on gender recognition and age can be estimated for different factors like race, facial expression, etc.

References

- Akbari, A., M. Awais, Z.-H. Feng, A. Farooq, and J. Kittler. 2020. A flatter loss for bias mitigation in cross-dataset facial age estimation. <http://arxiv.org/abs/2010.10368>.
- Chang, K. Y., C. S. Chen, and Y. P. Hung. 2011. Ordinal hyperplanes ranker with cost sensitivities for age estimation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 585–92, Colorado Springs, USA.
- Chen, B. C., C. S. Chen, and W. H. Hsu. 2015. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia* 17 (6):804–15. doi:10.1109/TMM.2015.2420374.
- Cootes, T. F., K. Walker, and C. J. Taylor. 2000. View-based active appearance models. Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, 227–32, Grenoble, France.
- Cortes, C., and V. Vapnik. 1995. SUPPORT-VECTOR NETWORKS. *Machine Learning* 20 (3):273–97. doi:10.1007/BF00994018.
- Dong, Y., C. Lang, and S. Feng. 2019. General structured sparse learning for human facial age estimation. *Multimedia Systems* 25 (1):49–57. doi:10.1007/s00530-017-0534-0.
- Gall, J., A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky. 2011. Hough forests for object detection, tracking, and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (11):2188–202. doi:10.1109/TPAMI.2011.70.
- Gao, F., and H. Ai. Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method. International Conference on Biometrics, ICB 2009. Lecture Notes in Computer Science, Springer, 5558, 132–41, Berlin, Heidelberg
- Geng, X., Z.-H. Zhou, and K. Smith-Miles. 2007. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (12):2234–40. doi:10.1109/TPAMI.2007.70733.

- Guehairia, O., A. Ouamane, F. Dornaika, and A. Taleb-Ahmed. 2020. Deep random forest for facial age estimation based on face images. *1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, 305–09, El Oued, Algeria.
- Guo, G., and G. Mu. 2011. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 657–64, Colorado Springs, USA.
- Han, H., and A. K. Jain. 2014. Age, Gender and Race Estimation from Unconstrained Face Images. MSU Technical Report (MSU-CSE-14-5):1–9.
- Han, J., W. Wang, S. Karaoglu, W. Zeng, and T. Gevers. 2021. Pose invariant age estimation of face images in the wild. *Computer Vision and Image Understanding* 202:103123. doi:10.1016/j.cviu.2020.103123.
- Hsu, C., and C. Lin. 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* 13 (2):415–25. doi:10.1109/72.991427.
- Jun, B., and D. Kim. 2012. Robust face detection using local gradient patterns and evidence accumulation. *Pattern Recognition* 45 (9):3304–16. doi:10.1016/j.patcog.2012.02.031.
- Khan, A., F. Bashar, F. Ahmed, and H. Kabir. 2013. Median ternary pattern (MTP) for face recognition. *International Conference on Informatics, Electronics and Vision (ICIEV)*, 6–10, Dhaka, Bangladesh.
- Khanmohammadi, S., S. Ghaemi, and F. Samadi. 2013. Human age group estimation based on ANFIS using the HOG and LBP features. *Electrical and Electronics Engineering: An International Journal (ELEIJ)* 2 (1):21–29.
- Kumar, T., and S. Haider. 2018. Multi-feature-based facial age estimation using an incomplete facial aging database. *Arabian Journal for Science and Engineering* 43 (12):8057–78. doi:10.1007/s13369-018-3293-0.
- Kumar, V., A. Namboodiri, and C. V. Jawahar. 2015. Visual phrases for exemplar face detection. *Proceedings of the IEEE International Conference on Computer Vision, 1994–2002*, Santiago, Chile.
- Kwon, Y. H., and V. Lobo. 1999. Age classification from facial images. *Computer Vision and Image Understanding* 74 (1):1–21. doi:10.1006/cviu.1997.0549.
- Lanitis, A., and N. Tsapatsoulis. 2015. An overview of research on facial aging using the FG-NET aging database. *Iet Biometrics* 5 (2):37–46.
- Li, C., Q. Liu, J. Liu, and H. Lu. 2012. Learning ordinal discriminative features for age estimation. *IEEE Conference on Computer Vision and Pattern Recognition*, 2570–77, RI, USA.
- Liu, T., Z. Lei, J. Wan, and S. Z. Li. 2015. DFDnet: Discriminant face descriptor network for facial age estimation. *Chinese conference on biometric recognition. Lecture Notes in Computer Science, Springer* 9428:649–58.
- Mahesh, V. G. V., A. Noel, and J. Raj. 2018. Zernike moments and machine learning based gender classification using facial images. *International Conference on Soft Computing and Pattern Recognition*, 398–408. Cham: Springer.
- Malek, M. E., Z. Azimifar, and R. Boostani. 2017. Facial age estimation using zernike moments and multi-layer perceptron. *International Conference on Digital Signal Processing (DSP)*, 1–5, London.
- Pang, Y., A. B. J. Teoh, and D. C. L. Ngo. 2015. A discriminant pseudo zernike moments in face recognition. *Journal of Research and Practice in Information Technology* 38 (2):197–211.
- Ramanathan, N., and R. Chellappa. 2006. Modeling age progression in young faces. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 387–394. New York, USA.
- Ross, A., and A. Jain. 2003. Information fusion in biometrics. *Pattern Recognition Letters* 24 (13):2115–25. doi:10.1016/S0167-8655(03)00079-5.

- Ross, A., and R. Govindarajan. 2004. Feature level fusion in biometric systems. Proceedings of Biometric Consortium Conference (BCC),1-2. <http://www.nws-sa.com/biometrics/ear/featureFusion.pdf>.
- Rothe, R., R. Timofte, and L. Van Gool. 2018. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision* 126 (2-4):144-57. doi:10.1007/s11263-016-0940-3.
- Sahoo, T. K., and H. Banka. 2018. Multi-feature-based facial age estimation using an incomplete facial aging database. *Arabian Journal for Science & Engineering (Springer Science & Business Media BV)* 43 (12):8057-8078
- Sawant, M., S. Addepalli, and K. Bhurchandi. 2019. Age estimation using local direction and moment pattern (LDMP) features. *Multimedia Tools and Applications* 78 (21):30419-41. doi:10.1007/s11042-019-7589-1.
- Sawant, M. M., and K. Bhurchandi. 2019. Hierarchical facial age estimation using gaussian process regression. *IEEE Access* 7:9142-52. doi:10.1109/ACCESS.2018.2889873.
- Shen, W., Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. Ile. 2021. Deep differentiable random forests for age estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2):404-19. doi:10.1109/TPAMI.2019.2937294.
- Tan, Z., S. Zhou, J. Wan, Z. Lei, and S. Z. Li. 2017. Age estimation based on a single network with soft softmax of aging modeling. *Asian Conference on Computer Vision, ACCV. Lecture Notes in Computer Science, Springer, 10113, 203-16.*
- Tan, Z., Y. Yang, J. Wan, G. Guo, and S. Z. Li. 2019. Deeply-learned hybrid representations for facial age estimation. *International Joint Conference on Artificial Intelligence (IJCAI)*, 3548-54.
- Thukral, P., K. Mitra, and R. Chellappa. 2012. A hierarchical approach for human age estimation, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1529-32, Kyoto, Japan.
- Tokola, R., D. Bolme, D. Barstow, and K. Ricanek. 2014. Discriminating projections for estimating face age in wild images. *IEEE International Joint Conference on Biometrics*, 1-8, Clearwater, FL, USA.
- Yan, S., H. Wang, T. S. Huang, Q. Yang, and X. Tang. 2007. Ranking with uncertain labels, *IEEE International Conference on media and Expo*, 96-99, Beijing, China.
- Yan, S., X. Zhou, M. Hasegawa-johnson, and T. S. Huang. 2008. Regression from Patch-Kernel. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8, AK, USA.
- Yasumoto, J. H. M., H. Y. Niwa, and H. Koshimizu. 2002. Age and gender estimation from facial image processing. *Proceedings of the 41st SICE Annual Conference*,5, Osaka, Japan.
- Zhang, Y., and D. Yeung. 2010. Multi-task warped Gaussian process for personalized age estimation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2622-29, San Francisco, CA, USA.
- Zou, M., J. Niu, J. Chen, Y. Liu, and X. Zhao. 2016. Facial ages Estimation with images in the wild, in multimedia modeling, MMM 2016. *Lecture Notes in Computer Science, Springer* 9516:454-65.