International Journal of Engineering & Technology, 7 (4.15) (2018) 539-543

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Active Appearance Model for Age Prediction: A Comparison

Musab Iqtait[1]*, Fatma Susilawati Mohamad[2], Fadi Alsuhimat[3]

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, Terengganu, Malaysia
**Corresponding author E-mail: si1950@putra.unisza.edu.my*

Abstract

Individual age gives key demographic data. It is viewed as a paramount delicate biometric characteristic for individual identification, contrasted with other pattern recognition issues. Age estimation is a complex issue particularly in relation to facial pictures with different ages, since the aging procedure varies extraordinarily across different age groups. In this work, we investigate deep learning techniques for age prediction based on Active Appearance Models (AAM) and six classifiers: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Support Vector Regression (SVR), Canonical Correlation Analysis (CCA), Linear Discriminant Analysis (LDA) and Projection Twin Support Vector Machine (PTSVM) to improve the precision of age prediction based on the present methods. In this algorithm, we extracted the traits of the facial images as traits vectors using AAM model, and the classifiers are utilized to predict the age. We were able to recognize that the accuracy of CCA algorithm is the best, the intermediate is SVR and the KNN algorithm is the lowest.

Keywords: *Age Prediction, Feature Extraction, Active Appearance Models (AAM), Age Classification.*

1. Introduction

Age prediction is a sort of delicate biometrics that provides useful data for a person's identity. The age of an individual is determined by their biometric traits. It can be recognized based on a multiclass classification problem or a regression problem [1].
Facial traits perform important functions in the applications of image processing, for example, Human-computer Interaction (HCI). It is considered one of the significant criteria of assessing the performance of frameworks and their abilities in decoding facial features progressively. Effective use of the systems helps in collecting these facial features correctly from facial pictures. Automatic age prediction systems have a variety of applications. In the security area, a well-designed Age Specific Human-Computer Interaction system (ASHCI) could help in the prevention of browsing adult web pages or obtaining age confined material from vending machine. ASHCI might also smooth our Everyday life. For example, an ASHCI introduced vehicle could prohibit youngsters from using technologies without the direction from adults. An ASHCI framework can give a caution when the driver abandons children alone in the vehicle without any whatever protective measures.
Age data might additionally be used in law enforcement. It can be utilized to search for suspects on a particular age class by restricting the search in the exhibition set. This could enhance the effectiveness of matching. Age prediction frameworks are also advantageous in the business web-domain. For instance, store managers might change decoration, styles, commercials based on their customer's demographics. Although age prediction from pictures is a critical method in real-world applications, it still faces challenges. As shown in Fig. 1, the aging procedure of different people varies incredibly. This is similarly to the aging procedures proposed in some studies [2, 3]. A significant number of factors influence the aging procedures. In general, these factors might be recognized under two different categories: internal and external factors. The internal factors are mostly dictated by physiological elements, for example, genes. External factors incorporate living environment, health condition, lifestyle and so forth. Extracting strong aging characteristics invariant to the impact brought by individual variations is still an open problem.
In the recent years, a significant number of efforts have been dedicated to the human picture age prediction process. These works can be subdivided under two issues: how could the age traits be extracted and how it could be estimated based on the elicited traits. For age traits extraction, in [2] proposed deep learning strategies based on Convolutional Neural Network (CNN). They built a new algorithm for facial traits extraction based on deep learning model. In contrast with past models developed using CNN, they utilized characteristic maps obtained in distinctive layers.
In [4] proposed additional details regarding Active Appearance Model (AAM) and Active Shape Model (ASM) algorithms for facial traits extraction. They tested their performance on one dataset of faces. This method revealed that ASM is faster and gains more accurate trait point than AAM, but AAM gains a better match to the texture.

Fig. 1: A clarification of the aging procedure for two different people. Each row indicates pictures of the same person in separate ages.

In [5] showed that for age prediction along with the appearance data, facial dynamics can be leveraged. It recommends a strategy to collect and utilize dynamic features for age estimation using a person's smile. For age prediction using extracted traits, in [6] classified age into four predefined classes. They used quick and effective machine learning methods – Extreme Learning Machines – to solve the age categorization issue. Local Gabor Binary Patterns, Biologically Inspired Feature and Gabor were incorporated to represent face image.

In [7] proposed an end-to-end learning approach to process ordinal regression issues utilizing deep Convolutional Neural Network, which could simultaneously conduct feature learning and regression modeling. The ordinal regression issue is improved under a procedure of binary classification sub-problems. Also, they proposed a multiple output CNN learning algorithm to collectively solve these classification sub-problems.

In this paper, we provide a hierarchical system that extracts a single set of function vectors based on AAM model. For age group classification and aging function, we used SVM, SVR, KNN, CCA, LDA and PTSVM algorithms with database of MORPH [8]. The database has 38533 images and 2600 of them are chosen to be a training set to train our feature extraction algorithm.

2. Age Prediction Method

This section clarifies the methods and strategies utilized in this study. Section 1 provides a description of the suggested method. Section 2 summarizes the traits extraction process, while section 3 presents the classifiers utilized.

2.1. Overview of the Proposed Method

The suggested age prediction algorithm comprises facial traits extraction, traits normalization and classification. The steps are indicated in Fig. 2:

Step 1: the shape and presence variables of the face pictures are collected as traits.

Step 2: traits normalization and processing where the contrast of the picture is improved.

Step 3: the entered pictures of faces are classified under different ages based on six classifiers.

Step 4: Finally, the age is predicted as a specified age value.

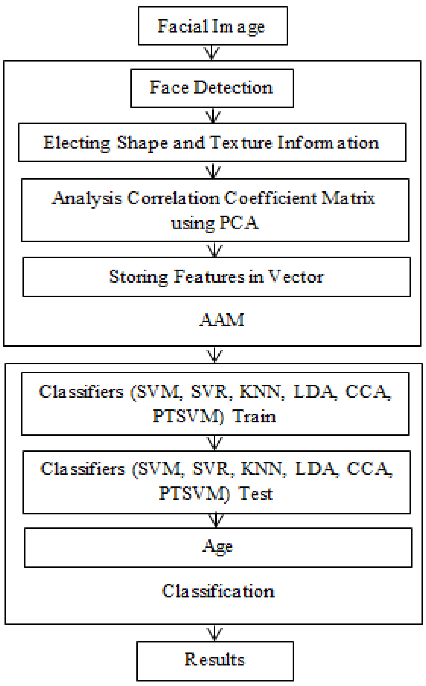

Fig. 2: Flow chart of the suggested age prediction algorithm.

2.2. Feature Extraction and Normalization

In this algorithm, AAM is used to gather facial traits from facial pictures. The AAM comprises Principal Component Analysis (PCA), which could model those facial appearances and facial shapes. It could also solve different cases using some variables [9]. Therefore, AAM model is used broadly in face recognition and face characteristics extraction.

To generate facial shape model, s points are determined in the face picture to make the facial structure $z = (x1, y1, x2, y2, xs, ys)^T$. In this approach, 68 landmarks were recognized in face photo as indicated in Fig. 3. In the experiment, we used 68 landmarks in the training and testing stages. These landmarks can be identified automatically through fitting mechanisms.

The form of face s is crossed by the linear combination of the mean shape s_0 and the n shape vectors s_i for i = 1,….,n

$$s = s_0 + \sum_{i=1}^{n} p_i s_i \tag{1}$$

where Pi, for i = 1… n, are the variables of face shape. To get the appearance model, the model normalization for the face shape of the image is performed by warping the form to the mean shape s_0. A(x) is the appearance of the face. It is exhibited through the linear combination of m appearance $A_i(x)$ and the face presence A_0.

Fig. 3: Facial traits points.

$$A(x) = A_0 + \sum_{i=1}^{m} \lambda_i A_i(x) \quad \forall x \in s_0 \tag{2}$$

where P_i is the style variable, the appearance variable is λ_i, the P_i and λ_i is utilized as face traits for age prediction. We determined the dimension of every variable to preserve the variability in the training set.

$$K(x_i, x_j) = e^{-\frac{1}{2\sigma^2}|x_i - x_j|^2}$$

When we have an unlabeled picture, AAM fitting technique needs to characteristically match the information on the new picture model. AAM fitting methodologies point will detect the shape and presence variables, which minimize errors in the organized picture. The AAM model fitting function is shown in the following equation.

$$\sum_{x \in s_0} \left[A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) - I(w(x;p)) \right]^2 \tag{3}$$

where w is the warping function required to modify the picture position start with position x in the entered picture to mean shape s_0 also i (W(x; p)) is the intensity level of the warped entered picture. To minimize (3) at the same work through for deference with the people shape variable p and the appearance variable λ, the gradient descent algorithm is utilized.

2.3. Age Classification

In this paper, we used six algorithms for classification: KNN, SVM, SVR, CCA, LDA and PTSVM. Let X = (x1, x2, xi, xN) hold numerous training photos; each xi has a chance to be characteristic vectors with dimension n. These vectors might have a chance to be modeled by connecting l×m facial photograph. The constant l × m = n. n is the full assortment for pixels inside those facial pictures. Also, n denotes all training pictures.

2.3.1. K-nearest Neighbor (KNN)

This is a technique of grouping parameters according to the closest samples of the zone inside features [10]. KNN is one of the well-known and clear arrangement algorithms. Learning approach only incorporates saving characteristic vectors and marks of the learning pictures, as well as inward grouping operations. The unmarked position might be truly allocated for its k closest neighbors. This item is normally categorized according to the marks of its k closest neighbors by using dominant part polling on k = 1. The parameters are categorized based on the power of the parameter closest to them. Assuming that there will be requirements for best two segments, then k would settle on an odd number. K might have a chance to be an odd number depending upon multiclass arrangement. Then, we transformed each photo of the vector starting with guaranteeing fixed-length for correct numbers. We used the celebrated distance equation as a relatable point separation capacity for KNN, which is Euclidean distance:

$$d(x,y) = \left(\sum_{i=1}^{m}((x_i - y_i)^2)\right)^{1/2} \quad (4)$$

Those x and y are histograms on X = Rm. The greater part of training pictures is included in these facial vectors and the appropriation over face vectors for each age in the age combination is portrayed.

2.3.2. Support Vector Machine (SVM) Algorithm

This is prepared to assess ages in a particular period [11]. By applying classification algorithm with a specific age values through the traits, we improved the limit of the classifier to carefully evaluate the distinctive age. Throughout the training procedure, we trained an age classifier using the entire training data. It holds objects for age ordering from 0 to 69 years. The framework of age prediction used the SVM algorithm shown in Fig. 2 to classify the face picture under age values determined using training methods. The inputs x_i are the traits vectors. To configure the SVM parameters, we utilized Gaussian Kernel K:

$$f(x) = \sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b \quad (5)$$

2.3.3. Support Vector Regression (SVR)

Provided a training group with L information illustrates $X \in R^m$ and their outputs $Y \in R^1$. The level specified from figuration SVR is under the provided systematic charge c. Also, Slack parameters ϵ are determined as:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w^t w + c\left(\sum_{i=1}^{L}\xi_i + \sum_{i=1}^{L}\xi_i^*\right)$$
$$s.t.\, w^T \emptyset(x_i) + b - y_i \le \epsilon + \xi_i$$
$$y_i - w^T \emptyset(x_i) - b \le \epsilon + \xi_i^* \quad (6)$$

With ξ_i , $\xi_i^* \ge 0$, i = 1 . . . L where the kernel function $\emptyset(x_i)$ maps the traits vector x_i into a higher-dimensional space.

SVR is prepared to learn the group facial characteristic vectors with the related ages. Subsequently, it furnished an obscure assembly of variables produced throughout the age prediction identified with the single person in the identical facial picture. SVR might have been used to build the equation f(x) beginning with the training images and recognized ages (starting with 0 until 69) together with traits vectors x_i. SVR algorithm with straight ε-insensitive expense was used to learn the age equation f(x) defined as follows.

$$\text{Age} = f(x) \quad (7)$$

2.3.4. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis is a widespread technique of dimensionality reduction and arrangement. We prepared a dataset for n samples, where $x_i \in IR^d$ and $l_i \in \{1, 2,. , k\}$ demonstrate $\mathcal{T} = \{(x_i, l_i)\}_{i=1}^{N}$ independently, the traits vector and the copartnered number name of the i-th sample, d is the information dimensionality and k is the number of classes. Lesvos (·)´ demonstrate the transpose driver. For discriminant analysis [12], three scramble matrices are defined as follows:

$$S_w = \frac{1}{N}\sum_{j=1}^{K}\sum_{\{x\in\mathcal{T},x:l=j\}}(x - c_j)(x - c_j)' , \quad (8)$$

$$S_t = \frac{1}{N}\sum_{i=1}^{K}(x_i - c)(x_i - c)' , \quad (9)$$

$$S_b = \frac{1}{N}\sum_{j=1}^{K}N_j(c_j - c)(c_j - c)' , \quad (10)$$

where N_j and c_j demonstrate the number of points and the centroid to the j-th class, and c is the enrolled centroid of the whole data. It takes the designation (S_w) and (S_b) that measure those within-class union and between class divisions independently. The total scatter matrix is obtained as $S_t = s_b + s_w$. LDA computes a straight change for $U \in IR^{l\times d}$, mapping the vector $x_i \in IR^d$ to a vector $x_i^l \in IR^l$, Ux_i, (l < d). In the low dimensional space resulting from the linear transformation U, the scatter matrices become

$$S_w^l = U' S_w U, \quad S_b^l = U' S_b U, \quad S_t^l = U' S_t U \quad (11)$$

The better transformation U^{LDA} is computed by the following equation:

$$U^{LDA} = max_u trace\left(S_b^l (S_t^l)^{-1}\right) \quad (12)$$

The matrix U^{LDA} is identified by the vectors $S_t^{-1} S_B$ corresponding to the target k-1 value.

2.3.5. Canonical Correlation Analysis (CCA)

First introduced in [13], CCA is a tool for multivariate statistical analysis. It is used in projecting two sets of multivariate information under a subspace, such that correlation between the anticipated information is maximized.
In the process of age estimation, given images from the database, appearance features with dimension are first extracted from the images. These feature vectors are organized into two data matrices: S = {(x_{11}, x_{12}),…,(x_{n1} ,x_{n2})}, where $x_{ij} \in R^{pj}$, j = 1,2 represent the ith sample starting with the jth view of p_j dimension. Two matrices $X_1 = [x_{11},\dots, x_{n1}]$ and $X_2 = [x_{12},\dots, x_{n2}]$ are recognized to demonstrate the information from those two perspectives. Two straight transforms w1, w2 are provided to project the samples, starting with two perspectives under the regular subspace and finally by developing the connection between $w^T{}_1X_1$ *and* $w^T{}_2X_2$ as below:

$$\max_{w1,w2} w_1^T X_1 X_2^T w_2 \tag{13}$$

$$s.t. \, w_1^T X_1 X_1^T w_1 = 1, w_2^T X_2 X_2^T w_2 = 1 \tag{14}$$

With the Lagrange multiplier, in (14) can be used in resorting of the eigenvalue decomposition. For w1,w2, the tests from two perspectives might be compared, projecting the relatable point space. Similarly, an unsupervised approach, CCA might be identified as two-view advancement for PCA. CCA is prepared for two-view case, and the pairwise strategy might have a chance to be applied at the point of the multi-view circumstance. In turn, requirement for asserting CCA is that the prepared data for CCA must be given previously. View-pair mode, i. e. the number of specimens for both views has a chance to be on $X_1X^T{}_2$ process.

2.3.6. Projection Twin Support Vector Machine (PT-SVM)

The main function of linear projection twin support vector [14] machine is to search for a projection axis for each class, so that within-class contrast of the predicted data points is minimized after the anticipated data points of the different classes are identified. Thus, the primal issues regarding straight PTSVM are from requesting QPPs.

$$\min_{w1} \frac{1}{2} w_1^T S_1 W_1 + c_1 e_2^T \xi_2 \qquad s.t. \;\; Bw_1 - \frac{1}{m_1} e_2 e_1^T A W_1 + \xi_2 \geq e_2, \qquad \xi_2 \geq 0 \tag{15}$$

$$\min_{w2} \frac{1}{2} w_2^T S_2 W_2 + c_2 e_1^T \xi_1 \qquad s.t. - (Aw_2 - \frac{1}{m_2} e_1 e_2^T B W_2) + \xi_1 \geq e_1, \qquad \xi_1 \geq 0 \tag{16}$$

where $c_1 > 0$ and $c_2 > 0$ are trade-off constants, $e_1 \in R^{m1}$ and $e_2 \in R^{m2}$ are both vectors of ones, ξ_1 and ξ_2 are both non negative slack variable vectors. S_1 and S_2 are within-class variance matrix which are expressed as

$$S_1 = \sum_{i=1}^{m_1} \left(x_i^{(1)} - \frac{1}{m_1} \sum_{j=1}^{m1} x_j^{(1)} \right) \left(x_i^{(1)} - \frac{1}{m_1} \sum_{j=1}^{m1} x_j^{(1)} \right)^T \tag{17}$$

$$S_2 = \sum_{i=1}^{m_2} \left(x_i^{(2)} - \frac{1}{m_2} \sum_{j=1}^{m2} x_j^{(2)} \right) \left(x_i^{(2)} - \frac{1}{m_2} \sum_{j=1}^{m2} x_j^{(2)} \right)^T \tag{18}$$

From (17) and (18), it is clear that the target function to TWSVM and PTSVM is not considered as neighborhood geometrical structure between the samples.

3. Results and Discussion

In this section, we present the results of our algorithm. In section 1, we portray the database which has been utilized. Section 2 summarizes the accuracy and the performance of the classifiers while section 3 indicates the efficiency of the classifiers.

3.1. Database

In our approach, we utilized MORPH database which is a face picture database connected with detailed information regarding the gender and age of each picture inside the database. The suggested database has been transformed, starting with Pinellas Country Sherrif's Office (PCSO) [8]. MORPH incorporates information of the date of birth. The database incorporates face portraits of grown-ups in different ages. A larger part of the databases comprises two collections. Collection 1 incorporates 1690 pictures of 515 people, while Collection 2 incorporates 15204 pictures of 4000 people. Those metadata of the portraits along with sex, weight, ethnicity, age, height are all accessible. Also, 2600 pictures were selected to train the set while 400 pictures were selected to test our algorithm.

3.2. Experimental Setup

We used AAM model as facial traits extraction approach due to its structure elicitation from facial images. AAM has been used in a number of methodologies as primary or secondary traits extraction methods. However, a significant measure of age prediction routines proposed LDA and CCA in their frameworks. In this experiment, the methodologies which we analyzed comprise KNN, SVM, SVR, LDA, CCA and PT-SVM.
Two predominant calibrations were used to assess the execution of age prediction: Cumulative score and Mean Absolute Error (MAE). MAE $=\sum_{i=1}^{N} |\bar{l}_i - l_i| \, /N$, where N is the amount of testing pictures, and l_i is the ground truth. While $\bar{l}_i$ is related to the predicted age. The Cumulative score formulation is defined as follows: Cumulative-Score (L) = $(N_{e<L}/N) \times 100\%$, $N_{e<L}$ is the amount of testing pictures with absolute error less than the error level L.
Table 1 indicates the testing results about our experiment by using MAE. We found that CCA performed better than other classifier algorithms. This can be an association between marks due to regression. Our suggested age-prediction algorithm accomplishes the lowest MAE of 4.17 in CCA method and the medium MAE of 4.68 in SVR method, as well as the highest MAE of 10.20 in KNN method

Table 1: Mean absolute error of age prediction algorithm on the MORPH database

Method	SVM	SVR	KNN	LDA	CCA	PT-SVM
MAE	5.71	4.68	10.2	5.67	4.17	4.67

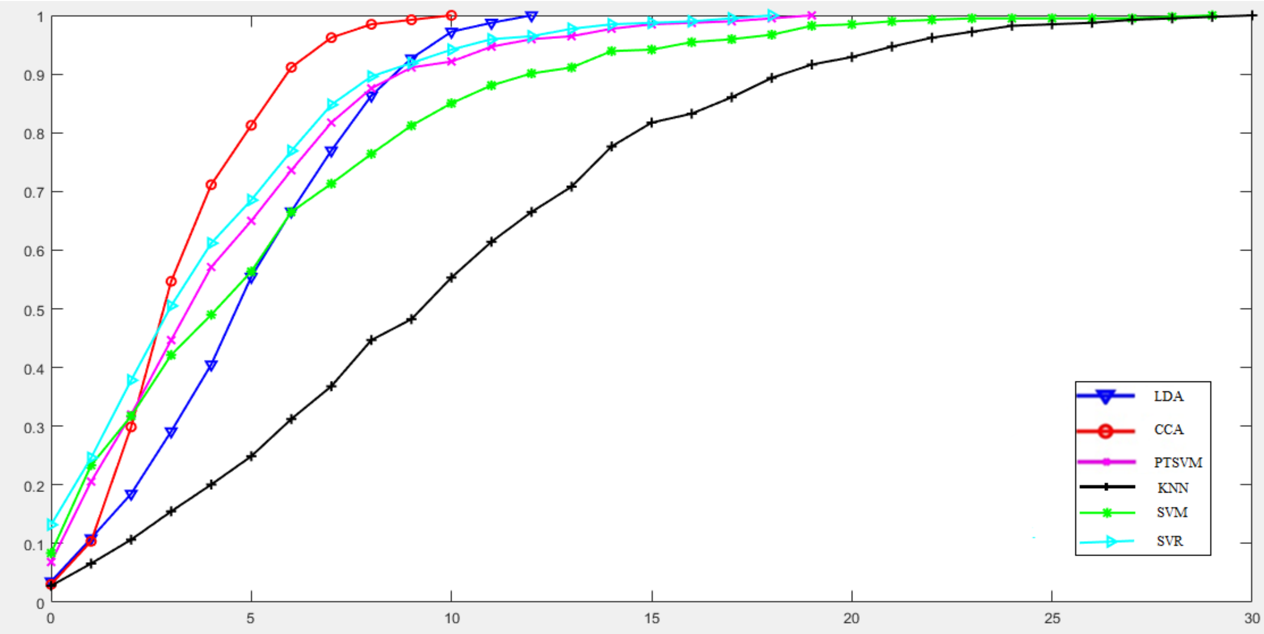

Fig. 4: Cumulative Score curves of the used techniques for MORPH Database

Fig. 4 indicates the joined score results in the multiple error levels which could expand with error level. When the cumulative score is stable, more diminutive error level will be better. It has been recognized that the CCA classifiers have the greater precision compared with different methods in exclusive error levels.

3.3. Efficiency

The efficiency was taken regarding time-measurements for running each algorithm on 400 images sequentially. Table 2 shows time measurements for classification algorithms:

Table 2: Time measurements for classification algorithms

Meth-od	SVR	SVM	KNN	CCA	LDA	PT-SVM
Time	0.0594 75	0.0553 51	0.0708 86	0.0506 61	0.2457 44	0.0708 86

Table 2 compares the time measurements for the six algorithms. Each algorithm performed separately and was applied on 400 images sequentially. The experiment was conducted on MATLAB R2016a, PC with processor Intel(R) Core(TM) i5, 4.00 GB RAM memory and Operating System Windows 7 64-bit. The result shows that CCA method takes the lowest time to perform full process.

4. Conclusion

We proposed an age prediction approach to face pictures based on the traits extracted from facial pictures. The prescribed technique used AAM model for traits extraction from the facial pictures. Measures depend on changes that occur in each shape throughout the aging process and the texture modifications, which are noticeable in the period of individual aging. We also used six machine learning algorithms suitable for age prediction from facial pictures: CCA, LDA, PTSVM, SVM, SVR and KNN.

Throughout the experiment, CCA furnished the best accurate results, despite the reality that CCA was indicated to be the most computationally powerful. Furthermore, we identified MORPH database as a suitable database for age prediction.

References

[1] Grd, P., Introduction to human age estimation using face images. Research Papers, Faculty of Materials Science and Technology Slovak University of Technology, 2013. 21(Special Issue): 24-30.
[2] Geng, X., Q. Wang, and Y. Xia. Facial age estimation by adaptive label distribution learning. Proceedings of the IEEE 22nd International Conference on Pattern Recognition, 2014.
[3] Wang, X., R. Guo, and C. Kambhamettu. Deeply-learned feature for age estimation. Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 2015.
[4] Iqtait, M., F. Mohamad, and M. Mamat. Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM). Proceedings of the IOP Conference Series: Materials Science and Engineering. 2018. IOP Publishing.
[5] Dibeklioğlu, H., et al., Combining facial dynamics with appearance for age estimation. IEEE Transactions on Image Processing, 2015. 24(6): 1928-1943.
[6] Sai, P.-K., J.-G. Wang, and E.-K. Teoh, Facial age range estimation with extreme learning machines. Neurocomputing, 2015. 149: p. 364-372.
[7] Niu, Z., et al. Ordinal regression with multiple output cnn for age estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
[8] Ricanek, K. and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. in IEEE 7th International Conference on Automatic Face and Gesture Recognition, 2006.
[9] Cootes, T.F., G.J. Edwards, and C.J. Taylor, Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001. 23(6): p. 681-685.
[10] Cover, T. and P. Hart, Nearest neighbor pattern classification. IEEE transactions on information theory, 1967. 13(1): p. 21-27.
[11] Burges, C.J., A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998. 2(2): 121-167.
[12] Yan, Y., et al., Multitask linear discriminant analysis for view invariant action recognition. IEEE Transactions on Image Processing, 2014. 23(12): 5599-5611.
[13] Hotelling, H., Relations between two sets of variates. Biometrika, 1936. 28(3/4): 321-377.
[14] Chen, X., et al., Recursive projection twin support vector machine via within-class variance minimization. Pattern Recognition, 2011. 44(10-11): 2643-2655.