International Journal of Engineering & Technology, 7 (3.28) (2018) 329-333

International Journal of Engineering & Technology

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

Research paper

Active Shape Model with Multiple Classifiers for Age Prediction

Fatma Susilawati Mohamad[1]*, Musab Iqtait[2]

[1,2]Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Besut Campus, 22200 Besut, Terengganu, Malaysia
** E-mail: si1950@putra.unisza.edu.my[1]*

Abstract

Automatic age prediction from facial images has received much attention. This is due to its various applications in security control, law enforcement, and human computer interaction. In spite of its developments, age prediction becomes more challenging. This is because the facial age procedure is specified not only by internal factors like genetic factors and external factors like lifestyle and environ-mental factors. In this paper, an enhanced age prediction algorithm using Active Shape Model (ASM) with six classifiers is suggested. 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) are adopted in order to enhance the accuracy of age prediction. In this work, traits of the facial images are extracted via ASM as the trait vector. The classifiers are utilized and compared to predict the age. The experimental results indicated that CCA has the highest accuracy while KNN has the lowest.

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

1. Introduction

Automatic age prediction receives much attention due to its extensive applications in areas such as Human Computer Interaction (HCI), electronic customer relationship management, surveillance monitoring [1]. Age prediction methods based on the usage of machine learning algorithm to predict human age rely on traits elicited from face pictures.

Age prediction undergoes several face recognition processes such as face detection, face alignment and traits elicitation. Traits elicited from facial pictures representing age connected apparent data is a key operator in the performance of age prediction methods [2]. Much research on age prediction has been conducted across age traits elicitation, such as ASM [3], Active Appearance Model (AAM)[4], age pattern subspace and age manifold [5], Age pattern Subspace (AGES) [6], and anthropometric model [7]. In these algorithms, the traits are usually elicited in low level accuracy.

Age prediction could be to develop a reliable age prediction framework dependent upon elicited traits. The age predictors can use machine learning algorithms to train a model for elicited traits and make age prediction for facial images. Age prediction is seen as a multiclass classification issue [8, 9] and a regression issue [10, 11]. In age classification, face picture is typically allocated to a class label that achieves the age value, and multi-class classification method is generally utilized. In age regression tasks, the methods yield an exact predicted age value. In this research, we concentrate on the classification task which allocates the face picture to an age label value.

In recent years, a much effort has been directed toward human picture age prediction methods. The previous works can be classified under two different issues: how the age traits could be elicited and how the age could be predicted based on the elicited traits. For age traits elicitation, Wang et al. [12] proposed deep learning strategies based on Convolutional Neural Network (CNN). They proposed a new algorithm for facial traits elicitation based on deep learning model. In contrast with past models developed using CNN, they utilized traits maps acquired in distinctive layers. Iqtait et al. [13] proposed additional details regarding AAM and ASM algorithms for facial traits elicitation. They tested their performance on one dataset of the faces. This method revealed that ASM is faster and gains more accurate trait points than AAM, but AAM gains a better match to the texture. Dibeklioğlu et al.[14] showed that for age prediction along with the appearance data, facial dynamics can be leveraged. They recommended a strategy to collect and utilize dynamic traits for age prediction using a person's smile.

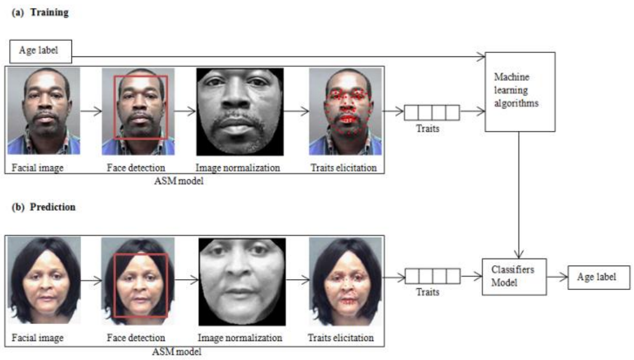

Fig. 1: Overview of the suggested age prediction algorithm

For age prediction using extracted traits, Lai et al. [15] proposed a modern technique to predict the age from facial picture under a dynamic range or a discrete age situated as opposed to solitary age or age set. They presented another measurement, i. e. Confidence Interval/Confidence Level to assess the performance of claiming recommended strategy. Furthermore, Mohamad et al. [16] proposed an enhanced age estimation framework. They utilized AAM model and three machine learning algorithms, SVM, KNN, and SVR with enhance the accuracy of age estimation dependent upon the introduced strategies. Sai et al. [17] classified age into four predefined classes. They used quick and effective machine learning methods to solve the age categorization issue. Local Gabor Binary Patterns, Biologically Inspired Feature, and Gabor were incorporated to represent face image.

This paper proposed a hierarchic algorithm that collects a solitary group of traits vectors dependent upon ASM model. For age prediction process, we utilized SVM, SVR, KNN, CCA, LDA, and PTSVM machine learning algorithms with MORPH database. The database contain 38533 pictures, we choose 2600 on a chance to be a training group to learn the proposed algorithm. Also, 400 pictures are chosen to test the proposed algorithm.

2. Age Prediction Method

This part demonstrates the strategies and algorithms utilized in this paper. Part one shows the description of the suggested algorithm. Part two depicts a database used. Part three summarizes the traits elicitation process. Part four shows the ASM model searching process while Section five demonstrates the utilized machine learning algorithms.

2.1. Suggested Method Summary

We proposed an enhanced algorithm for age estimation which combines of two stages: training stage and prediction stage. These stages are illustrated in Figure1:

The training stage consists of two main steps:

- Traits elicitation using ASM model which includes three steps:
  - Face detection: At this step, the face region is detected.
  - Image normalization: This consists of traits calibration and handling where a disparity of the picture can be improved.
  - Traits elicitation: At this step, the shape properties related to facial images are collected in traits matrices.
- Machine learning: the entered features are trained under different ages based on six machine learning algorithms.

The prediction process includes two main steps:

- Traits elicitation using ASM model.
- Classification model: the entered pictures of faces are classified under diverse ages in view of six machine learning algorithms.

2.2. Database

In this research, MORPH database was utilized, this is a facial pictures database associated with particular data in regards to the sex also age to each picture in the database. The suggested database has been transformed using Pinellas country Sherrif's office (PCSO) [18]. MORPH contains information like born date, and incorporates face portraits of the adults in distinctive ages. A bigger part in the MORPH contains pair groups. Group 1 incorporates 1690 pictures of 515 individuals while Group 2 consists of 15204 pictures of 4000 individuals. The detailed information of the portraits alongside gender, weight, ethnicity, age and tallness are all accessible. Furthermore, we choose 2600 pictures to be training set at same time we choose 400 pictures to be testing set.

2.3. Trait Elicitation and Normalization

In the suggested algorithm, ASM is used to gather facial traits from facial pictures. ASM model incorporates two processes: model construction and searching [3]. In the following section, we provide a short overview of the ASM model, which confirms its applications in facial traits elicitation.

The model structure technique for ASM aims to assign k key facial trait points to each face picture in the training group. For example, {(x_1, y1) and (x_k, yk)} are the k facial trait points marked on the picture shown in Fig. 2. A trait point is the point of the same index number in various pictures. All the k points checked on a picture develop an extraordinary shape which can be performed by a shape vector

$$s = (x_1, y_1, \dots, x_k, y_k). \quad (1)$$

After creating the shape model separate of the volume, location Furthermore direction of the face, we adjusted the shape vectors to support the rotation, scaling and interpretation. The measure of the arrangement may be on minimizing the entirety of the adjusted distances of the trait points. Subsequently, Principal Component Analysis (PCA) could be used in relation to the aligned shape vectors. The mean might be expressed as.

$$\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i, \quad (2)$$

And the corresponding covariance can be expressed as

$$S = \frac{1}{n}\sum_{i=1}^{n}(s_i - \bar{s})^T(s_i - \bar{s}). \quad (3)$$

The eigenvalues (λ_1,λ_s) and the corresponding eigenvectors (p_1,…,p_s) of the covariance matrix S, where. $\lambda_i \geq \lambda_{i+1}$, could subsequently be acquired. The first t eigenvectors are selected based on the following equation:

$$\sum_{i=1}^{t}\lambda_i \Big/ \sum_{i=1}^{s}\lambda_i \geq \alpha. \quad (4)$$

The definition of the worth about α varies of distinctive applications. However we assign α as 0.95. The subsequent t eigenvectors could be formulated as P = (p_1,…,p_t). Then, the shape vectors can be approximated by utilizing the following equation:

$$s \approx \bar{s} + Pb, \quad (5)$$

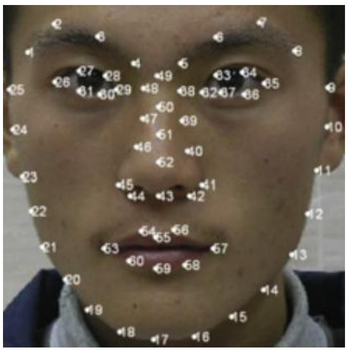

Fig. 2: Facial traits points

.

Where b is a weight vector which could be calculated as follows:

$$b = P^T(s - \bar{s}). \quad (6)$$

After updating the reshaped location of the traits landmarks in a repeated way, nearby construction ought to be structured in improvements. The local structure might be built toward calculating those pixel profiles which would be sampled around every traits landmark along the ordinary way of the shape. A point nearby (x_i,y_i) is provided based on the sampling. An M pixel on two sides of the point along the ordinary bearing of the shape prompts the profile length being 2m + 1. Also, the neighborhood structure of the points could be structured toward the initial arrangement derivatives about these profiles.

Let (g1,..,gn) is the normalized derivative profiles. Then the mean profile g Also its covariance grid Sg might a chance to be helpfully computed. We utilized the Mahalanobis distance scale to calculate the contrast the middle of the new and the mean profiles, which can be characterized as following:

$$f(g_i) = (g_i - \bar{g})S_g^{-1}(g_i - \bar{g})^T. \quad (7)$$

2.4. ASM Searching

The exploring strategy for ASM is summarized in the next two repeated stages: (1) the scan of the modern location for every point; (2) the refresh of the variable b and the variables of the relative change.

In the main step, it is necessary to sample l (l>m) pixels around jointly portions of the point and form a profile as expressed. Subsequently, a sub-profile with the length from claiming m is decided and the relating Mahalanobis distances are calculated. The focal point of the sub-profile can be the enhanced location of the point where the Mahalanobis distance achieves its level. To calculate the upgraded location repeatedly, we developed a displacement vector dS = (dS_1,…,dS_k) the affine conversion variables and b can be updated as:

$$S = M(s, \theta)[s] + S_c \quad (8)$$

2.5. Machine Learning Algorithms

In this research, we utilized six machine learning algorithms for training and prediction: KNN, SVM, SVR, CCA, LDA, and PTSVM. Let Y=(y1, y2, yi, yN) hold many training pictures; each yi has a chance to be trait matrices of dimension n. The matrices might have a chance to be modeled by linking b×c facial photograph. The number b×c=o. o is the complete assortment for pixels internal those pictures and o denotes all training pictures.

K-nearest Neighbor (KNN)

KNN is a machine learning algorithm for grouping parameters as stated by the closest points of the zone inside traits [19]. KNN is a standout amongst the well-known and clear machine learning algorithms. The training approach aims to reserving trait vectors and denoting the training pictures, as well as domestic grouping processes. The unmarked location might be allocated for its k nearest neighbors. This might be regularly sorted as stated by its k nearest neighbors using controlling portion ballot respect to k=1. The variables are grouped considering the forces of the variable nearby them. These could be prerequisites for the best two segments. Subsequently, k might settle on an odd number. Also, k makes it possible to create an odd number depending upon multiclass arrangements. At this point, we varied every photo of the vector beginning to guarantee fixed-length to right values. We used the celebrated interval equation as a reachable side of the point detachment ability for KNN which is Euclidean distance.

$$d(x, y) = (\sum_{i=1}^{m}((x_i - y_i)^2))^{1/2} \quad (9)$$

The x and y are histograms with respect to X =Rm. The greater part of training data are incorporated for these facial matrices and the allotment over face matrices to each age in the age combination is portrayed.

Support Vector Machine (SVM)

In the proposed approach we used SVM to predict the ages in a particular time [20]. Eventually by applying machine learning algorithm with a specific age labels through the traits, the SVM limitation was enhanced to purposely predict the age. Throughout the training phase, we trained the SVM using the whole training information. The age classifier holds labels for age ordering from 0-69 years. The algorithm apply SVM algorithm to do a classification for the facial traits under age labels resolved using the training methods. The parameter xi is the traits matrices. For design the SVM variables, we utilized Gaussian kernel K:

$$f(x) = \sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b \quad (10)$$

Support Vector Regression (SVR)

Given a training set, L data proves X ∈ Rm and their outputs Y ∈ R1. The level fixed of formation SVR is under the awarded precise. c [21]. Also, Slack parameters ϵ are determined as:

$$\min_{w,b,\xi,\xi*} \frac{1}{2} w^t w + c\left(\sum_{i=1}^{L} \xi_i + \sum_{i=1}^{L} \xi_i^*\right)$$
$$s.t. w^T \emptyset(x_i) + \mathrm{b} - y_i \leq \epsilon + \xi_i$$
$$y_i - w^T \emptyset(x_i) - \mathrm{b} \leq \epsilon + \xi_i^* \quad (11)$$

With $\xi_i, \xi_i^* \geq 0$, i = 1 . . . L, where the kernel function $\emptyset(x_i)$ maps the traits vector x_i into a higher-dimensional space. SVR is destined to train the facial traits matrices for these age labels. Thence, it outfitted a fuzzy assembly of variables transformed throughout the age prediction recognized for a single person in the corresponding facial image. SVR might have been utilized to raise the mathematical statement f(x) starting from the training pictures and age labels (starting from 0 until 69) toward the same time traits matrices xi. SVR method for right ε-insensitive charge was utilized to train the age formula f(x) determined as Age = f(x).

Linear Discriminant Analysis (LDA)

LDA is a broad method for dimensionality reduction and arrangement. We prepared a dataset for n samples $\mathcal{T} = \{(x_i, l_i)\}_{i=1}^{N}$, where $x_i \in IR^d$ and $l_i \in \{1, 2,. , k\}$ prove independently, the vector and the co-partnered number name of the i-th sample, d is the data dimensionality, and k is the amount from claiming classes. Lesvos (•) ´ shows this transpose driver. For discriminant analysis [22], three scramble matrices are defined as follows:

$$S_w = \frac{1}{N} \sum_{j=1}^{k} \sum_{\{x \in \mathcal{T}, x:l=j\}} (x - c_j)(x - c_j)', \quad (12)$$

$$S_b = \frac{1}{N} \sum_{J=1}^{K} N_j (c_j - c)(c_j - c)', \quad (13)$$

$$S_t = \frac{1}{N} \sum_{i=1}^{K} (x_i - c)(x_i - c)', \quad (14)$$

Where N_j and c_j show the number of points and the centroid of the j-th class, and c is the selected centroid of the whole information. It takes the designation (S_w) and (S_b) that calculate the within-class union also among class divisions independently. The total scatter grid is acquired similarly 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 the vector $x_i^l \in IR^l$, Ux_i, (l<d). In the minimal dimensional space coming because of the linear transformation U, the scatter matrices turn into

$$S_w^l = U'^{S_w}U, \qquad S_b^l = U'S_bU \,, \qquad S_t^l = U'S_tU$$

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

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

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

Canonical Correlation Analysis (CCA)

CCA was first proposed in [23]. It is an algorithm for multivariate measurable analysis. It is utilized within two sets of multivariate data under a subspace, such that correlation between the anticipated data is maximized.

In the method of age estimation, provided pictures from the database and appearance traits with dimension are first elicited from the pictures. These trait vectors are sorted out under two data matrices: S= $\{(x_{11}, x_{12}),\ldots,(x_{n1}, x_{n2})\}$, where $x_{ij} \in R^{pj}$, j=1,2 represents the ith sample starting with the jth view of p_j dimension. Two matrices $X_1 = [x_{11}, \ldots, x_{n1}]$ and $X_2 = [x_{12}, \ldots, x_{n2}]$ are recognized to exhibit the majority of the data from the two perspectives. Two straight transforms w1, w2 are provided to project the samples, beginning from two perspectives under the general subspace and at last by developing the connection $w^T{}_1X_1$ *and* $w^T{}_2X_2$ as:

$$\max_{w1,w2} w_1^T X_1 X_2^T w_2 \tag{15}$$

$$s.t. w_1^T X_1 X_1^T w_1 = 1, w_2^T X_2 X_2^T w_2 = 1 \tag{16}$$

For the Lagrange multiplier, mathematical equation (16) can be utilized to solve the eigenvalue decomposition. For w1 and w2, the tests from two perspectives may be compared, projecting the relatable point space. In unconfirmed approaches, CCA could be identified as two-view advancement to PCA. CCA is prepared for two-view case, and the pairwise method might be applied at the point of the multi-view circumstance. In turn, the prerequisite 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.

Projection Twin Support Vector Machine (PT-SVM)

The primary function of linear projection twin support vector [24] machine is to search for a projection axis for every class, so that within-class contrast of the predicted information points is minimized after the anticipated data points of the distinctive classes are identified. Thus, the primal issues in regards to straight PTSVM are from QPPs.

$$\min_{w1} \frac{1}{2} w_1^T S_1 W_1 + c_1 e_2^T \xi_2 \tag{17}$$

$$s.t.\ Bw_1 - \frac{1}{m_1} e_2 e_1^T A W_1 + \xi_2 \geq e_2, \qquad \xi_2 \geq 0$$

$$\min_{w2} \frac{1}{2} w_2^T S_2 W_2 + c_2 e_1^T \xi_1 \tag{18}$$

$$s.t. - \left(Aw_2 - \frac{1}{m_2} e_1 e_2^T B W_2\right) + \xi_1 \geq e_1, \qquad \xi_1 \geq 0$$

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, and 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

$$S_1 = \sum_{i=1}^{m_1} \left(x_i^{(1)} - \frac{1}{m_1} \sum_{j=1}^{m1} x_j^{(1)}\right) \left(x_i^{(1)} - \frac{1}{m_1} \sum_{j=1}^{m1} x_j^{(1)}\right)^T \tag{19}$$

$$S_2 = \sum_{i=1}^{m_2} \left(x_i^{(2)} - \frac{1}{m_2} \sum_{j=1}^{m2} x_j^{(2)}\right) \left(x_i^{(2)} - \frac{1}{m_2} \sum_{j=1}^{m2} x_j^{(2)}\right)^T \tag{20}$$

From equations (19) and (20), it is clear that the target function to TWSVM and PTSVM is not considered as neighborhood geometrical structure between the samples.

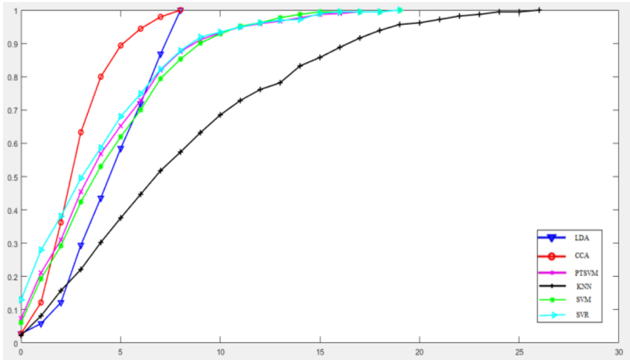

Fig. 3: Cumulative Score curves for Machine Learning algorithms

3. Experimental Results

In this section, results of the selected algorithms are displayed. Section 1 describes the precision and the execution of the machine learning algorithms while section 2 demonstrates the efficiency of the machine learning algorithms.

3.1. Experimental Setup

We utilized ASM model as traits elicitation method because of its structure elicitation from facial pictures. ASM has been utilized in a lot of methodologies as main or subaltern trait elicitation methods. However, a significant age prediction algorithm proposed LDA and CCA over their frameworks. In this experiment, the algorithms used are KNN, SVM, SVR, LDA, CCA, and PT-SVM. Two standardization were utilized to determine the execution of the proposed algorithm: com cumulative score and Mean Absolute Error (MAE). MAE $=\sum_{i=1}^{N}\left|\overline{l_i} - l_i\right| /N$, The N is the total of examine images, and l_i is the fact value while $\overline{l_i}$ is related to the predicted age. The cumulative score equation is specified as prosecute: Cumulative-Score (L) = $(N_{e<L}/N) \times 100\%$, $N_{e<L}$ is the number of examine photos that have absolute error minimal than the error level L.

Table 1 demonstrates the results of the proposed algorithm using MAE. We find that CCA get results better than other machine learning algorithms. This can be an association between marks due to regression. The suggested algorithm attains the lowest MAE of 3.50 in CCA algorithm and the medium MAE of 4.64 in PLSVM method, as well as the highest MAE of 8.43 in KNN algorithm.

Table 1: Mean Absolute Error for machine learning methods.

Method	SVM	SVR	KNN	LDA	CCA	PT-SVM
MAE	4.85	4.61	8.43	5.13	3.50	4.64

Figure 3 demonstrates the cumulative score outcomes in the multiple error levels which could stretch with error level. At the time that cumulative score is settled, more diminutive error level will be superior. It has been observed that the CCA machine learning algorithm has the highest precision contrasted with other techniques in exclusive error levels.

3.2. Efficiency

The effectiveness has been made In light of time-measurements to apply all algorithms to 400 pictures consecutively. Table 2 indicates the time calculations for machine learning algorithms:

Table 2: Efficiency for the used machine learning algorithms.

Meth-od	SVM	SVR	KNN	LDA	CCA	PT-SVM
MAE	0.047014	0.049475	0.019070	0.086501	0.166884	0.070886

Table 2 summarizes the running time estimation for the machine learning algorithms. Each technique run independently and has been used for 400 pictures consecutively. the test have been directed on MATLAB R2016a, computer with processor Intel(R) Core(TM) i5, 4.00 GB RAM memory and Operating System Windows 7 64-bit. The consequence indicates that KNN techniques get the most reduced time in execution.

4. Conclusion

This paper proposed an age prediction algorithm for face images based on traits elicited from facial pictures. The prescribed method utilized ASM model for trait elicitation from facial images. The proposed algorithm relied on each of Shape transforms that happened during aging moreover to the texture transformation which are apparent and the greater part powerful in the age of person. We also utilized six machine learning algorithms suitable for age prediction from facial pictures: CCA, LDA, PTSVM, SVM, SVR, and KNN.

For the experiment, CCA has the best accuracy results, despite the reality that KNN was shown to have the best efficiency results.

Acknowledgements

This research is financially supported by the Ministry of Education Malaysia with Fundamental Research Grant Scheme (FRGS), Project Code RR227 and supported academically by Faculty of Informatics and Computing, Universiti Sultan Zainal Abdin (UniSZA), Terengganu, Malaysia.

References

[1] Dong, Y., Y. Liu, and S. Lian, *Automatic age estimation based on deep learning algorithm.* Neurocomputing, 2016. **187**: p. 4-10.

[2] Liu, K.-H., S. Yan, and C.-C.J. Kuo, *Age estimation via grouping and decision fusion.* IEEE Transactions on Information Forensics and Security, 2015. **10**(11): p. 2408-2423.

[3] Cootes, T.F., et al., *Active shape models-their training and application.* Computer vision and image understanding, 1995. **61**(1): p. 38-59.

[4] 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.

[5] Fu, Y. and T.S. Huang, *Human age estimation with regression on discriminative aging manifold.* IEEE Transactions on Multimedia, 2008. **10**(4): p. 578-584.

[6] Geng, X., et al. *Learning from facial aging patterns for automatic age estimation*. in *Proceedings of the 14th ACM international conference on Multimedia*. 2006. ACM.

[7] Kwon, Y.H. and N. da Vitoria Lobo, *Age classification from facial images.* Computer vision and image understanding, 1999. **74**(1): p. 1-21.

[8] Niu, Z., et al. *Ordinal regression with multiple output cnn for age estimation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[9] Pontes, J.K., et al., *A flexible hierarchical approach for facial age estimation based on multiple features.* Pattern Recognition, 2016. **54**: p. 34-51.

[10] [Guo, G. and G. Mu. *Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression*. in *Computer vision and pattern recognition (cvpr), 2011 ieee conference on*. 2011. IEEE.

[11] Guo, G. and G. Mu. *Joint estimation of age, gender and ethnicity: CCA vs. PLS.* in *Automatic face and gesture recognition (fg), 2013 10th ieee international conference and workshops on*. 2013. IEEE.

[12] Geng, X., Q. Wang, and Y. Xia. *Facial age estimation by adaptive label distribution learning*. in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. 2014. IEEE.

[13] Iqtait, M., F. Mohamad, and M. Mamat. *Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM).* in *IOP Conference Series: Materials Science and Engineering*. 2018. IOP Publishing.

[14] Dibeklioğlu, H., et al., *Combining facial dynamics with appearance for age estimation.* IEEE Transactions on Image Processing, 2015. **24**(6): p. 1928-1943.

[15] Lai, D., et al., *Age estimation with dynamic age range.* Multimedia Tools and Applications, 2017. **76**(5): p. 6551-6573.

[16] Mohamad, F.S., M. Iqtait, and F. Alsuhimat. *Age prediction on face features via multiple classifiers*. in *2018 4th International Conference on Computer and Technology Applications (ICCTA)*. 2018. IEEE.

[17] Sai, P.-K., J.-G. Wang, and E.-K. Teoh, *Facial age range estimation with extreme learning machines.* Neurocomputing, 2015. **149**: p. 364-372.

[18] Ricanek, K. and T. Tesafaye. *Morph: A longitudinal image database of normal adult age-progression*. in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. 2006. IEEE.

[19] Cover, T. and P. Hart, *Nearest neighbor pattern classification.* IEEE transactions on information theory, 1967. **13**(1): p. 21-27.

[20] Burges, C.J., *A tutorial on support vector machines for pattern recognition.* Data mining and knowledge discovery, 1998. **2**(2): p. 121-167.

[21] Smola, A.J. and B. Schölkopf, *A tutorial on support vector regression.* Statistics and computing, 2004. **14**(3): p. 199-222.

[22] Yan, Y., et al., *Multitask linear discriminant analysis for view invariant action recognition.* IEEE Transactions on Image Processing, 2014. **23**(12): p. 5599-5611.

[23] Hotelling, H., *Relations between two sets of variates.* Biometrika, 1936. **28**(3/4): p. 321-377.

[24] Chen, X., et al., *Recursive projection twin support vector machine via within-class variance minimization.* Pattern Recognition, 2011. **44**(10-11): p. 2643-2655.