# Feature Selection and Feature Manifold for Age Estimation

Shivani Kshatriya, Manisha Sawant$^{(\boxtimes)}$ , and K. M. Bhurchandi

Department of Electronics and Communication Engineering, Visvesvaraya National
Institute of Technology, Nagpur, India
bhurchandikm@ece.vnit.ac.in
http://www.vnit.ac.in

**Abstract.** In recent years, a number of manifold learning techniques
have been proposed in the literature to address the age estimation prob-
lem. In manifold methods, appearance features are projected onto a
discriminant aging subspace and the age estimation is performed on
the aging subspace. In these methods the manifold is learn from the
gray intensity images. We propose a feature based discriminant manifold
learning and feature selection scheme for robust age estimation. This
paper also presents an experimental analysis of the manifold learning
and feature selection schemes for age estimation. The exact age value is
estimated by applying regression on the resultant feature vector. Experi-
mental analysis on a large scale aging database MORPH-II, demonstrate
the effectiveness of the proposed scheme.

**Keywords:** Feature manifold · Feature selection · Age estimation

## 1 Introduction

Human face conveys significant information for human to human as well as
human-machine interaction. Estimating various facial attributes such as age,
gender and expression plays a vital role in various forensic, multimedia and law
enforcement applications. Facial aging related research is broadly classified into
three categories: age estimation, age synthesis and age invariant face recognition.
Age estimation and synthesis mainly consider aging information that change due
to aging.

Although significant research has been carried out on age invariant face
recognition, relatively few publications have been reported on age estimation
[4,11,15,31,35]. This is due to various factors such as complex biological changes,
lifestyle, ethinicity skincare. These various factors changes the shape and tex-
ture of the face. Different aging patterns are observed due to diversity in cli-
matic conditions, races and lifestyle. Due to such large variations, it is difficult
even for humans to precisely predict a person's age from the facial appearances.

In the recent years, many efforts have been devoted to identify discriminant aging subspaces for age estimation. Some representative subspace learning methods used for age estimation include principal component analysis (PCA) [9], locality preserving projections (LPP) [21], orthogonal locality preserving projections (OLPP) [1], and conformal embedding analysis (CEA) [11,16]. The basic idea of the subspace learning methods is to find a low-dimensional representation in an embedded subspace and then perform regression in the embedded subspace to predict exact age. Geng et al. [14] proposed a subspace called AGing pattErn Subspace (AGES) to learn personalized aging process from multiple faces of an individual. For age group classification, Guo et al. [16] used manifold based features and Locally Adjusted Robust Regressor (LARR).

Manifold feature descriptors are characterise by a low dimensionality. Apart from age discriminative information, it also contains other related information such as identity, expression and pose. Therefore for achieving large improvements in age estimation, it is important to figure out which feature is more appropriate and important for describing the age characteristic. Existing manifold methods extract manifold features from the gray intensity or image space. However, the image space is inefficient to model the large age variations. The texture features such as HOG [7], SIFT [27] and GOP [26] are used to capture the textural variations due to aging. But the manifold of such feature space has not been explored for age estimation. Also, selection of relevant features from these texture features is an important direction in this area. In this paper, we present analysis of manifold and feature selection methods and their quantitative impact on the age estimation. We extract age-discriminative features from the manifold subspace as well as through feature selection for age estimation. The main advantages of these features are low dimensionality, robustness under illumination variation and intensity noise resulting in improved performance.

## 2   Related Work

Early research on age estimation mainly focused on anthropometric measurements to group facial images into different age groups. Following the development of local features, instead of age classification, much attention was focused on exact age estimation. Recent research in age estimation is classified based on the feature extraction and the feature learning methods. In this section, we briefly review them based on the facial features and learning methods for age estimation.

### 2.1   Aging Feature

After preprocessing, facial feature extraction is first step in typical age estimation approach as shown in Fig. 1. Early age estimation approaches used Active Appearance Models (AAM) [6] for shape and texture representation. These systems utilize the shape and texture variations observed in the facial images. Lanitis et al. [25] proposed a person-specific age estimation method, wherein they
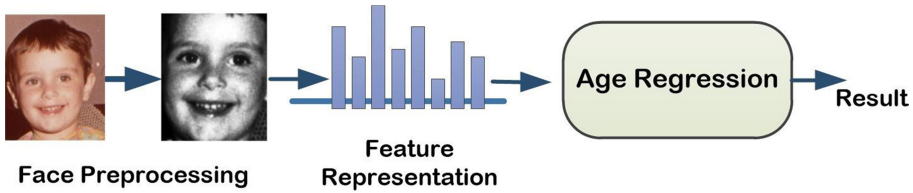
**Fig. 1.** Age estimation framework.

have used AAM to extract craniofacial growth and aging patterns at different age groups. Further, various age estimation approaches [4,13,14,24] proposed variations in AAM to capture aging patterns for age group classification. In case of AAM based methods, accurate localization of facial landmarks is a deciding factor for performance improvement.

For appearance feature extraction, apart from the earlier listed global features, histogram based local features like HOG, LBP, SIFT, BIF and Gabor are also used. Bio-Inspired Features (BIF) proposed in [19] for age estimation is based on a bank of multi-orientations and multiscale Gabour filters. Recently, for age estimation variants of BIF have been used [17,18,20]. BIF is specially designed for age estionation. Various existing local features that are also used for aging feature representation [10,22,33,36]. In [33], combination of PCA, LBP and BIF is used as aging feature. In [28] combination of global and local features is proposed. AAM is used for global feature representation whereas LPQ, LBP and Gabor for local feature extraction. Feature fusion is followed by dimensionality reduction for compact representation of the feature vector. HOG is used as aging feature in [10,22], whereas MLBP and SIFT are used as feature vectors in [36] for age estimation approach.

Besides, the local and global facial features, manifold based features are used to learn low dimensional manifolds. Various methods such as PCA, LPP, OLPP, and CEA are used in age estimation approaches. In these methods, low dimensional representation in embedded subspace is learned and age estimation is performed in the embedded subspace. Personalized aging process is learned from multiple faces of an individual using a AGing pattErn Subspace (AGES) [14]. Although, the performance of manifold based features is better than the image based features, these methods require large training data to learn the manifold.

## 2.2   Age Regression

After feature extraction, classification or regreesion methods are applied on the local features for age group classification or exact age estimation respectively. Information obtained from the facial feature has been effectively used by various learning methods for regression or classification. Age estimation from facial images falls under two categories of machine learning, classification and regression. For age group classification, an age range is treated as a class lable, whereas, for the regression it is treated as an ordered continuous value. Initial work on

age estimation in [24] compared the performance of Artificial Neural Networks (ANN), quadratic function and nearest neighbor classifier for age classification. Performance of quadratic function and ANN is found to be better than nearest neighbor. Moreover, Support Vector Regression (SVR) and Support Vector Machine (SVM) [8] are the most popular choices for age estimation. Aging patters are learned in [17] using KPLS regression. Age values represents ordered information this relative order information is encoded in Ordinal Hyperplane Ranking algorithm (OHRank) in [4]. Other than above mentioned regression methods, a Gaussian process based multitask warped gaussian process regression was developed in [35] for person specific age estimation. To reduce computational burdon during training an efficient version of WGP called Orthogonal Gaussian process (OGP) regression was proposed in [36].

Discriminant manifold subspaces are explored to encode the face for age estimation. OLPP technique is used in [12] and [11] to extract discriminant aging subspace. These methods learn the aging subspace from the raw image space, which is not able to represent the large facial variations due to aging. Various local feature descriptors such as LBP, HOG and SIFT are available in the literature which encode the facial features such as fine lines, smoothness, and wrinkles. We propose a method to extract age relevant features from the feature space instead of raw image space. Also, it is not known in advance which manifold is suitable for the age discriminative feature. We provide experimental analysis of feature manifold for age estimation. The local feature descriptors such as HOG, SIFT etc. extract important gradient and edge information and they are used for facial analysis. Hence it is important to select the age discriminative features from them. Among various machine learning approaches, feature selection is a technique which selects and rank relevant features according to their degrees of relevance and preference. In the literature of age estimation use of the feature selection method has not been identified. In this paper, we extract the age discriminative features using the feature selection methods.

Section 1 presents introduction followed by literature survey in Sect. 2. The proposed method is presented in Sect. 3 while Sect. 4 presents experimental results. Section 5 presents the final conclusion of this paper.

## 3   Proposed Work

The proposed age estimation framework mainly incorporates four modules: face preprocessing, feature extraction, feature transformation/ selection, and regression. In the first stage, face images undergo normalizations such as pose correction and histogram equalization. Then, the histogram-of-oriented-gradient (HOG) feature is computed for each image. Being histogram based local feature, the dimension of the extracted feature vector could be very high depending on the number of scales and orientations. High dimensionality of the extracted local features is in general handled by dimensionality reduction technique. However in the dimensionality reduction, it is not analysed whether the transformed space truly represents the aging subspace. It is possible that the transformation of

the local features by dimensionality reduction technique may lead to a subspace which is not discriminative for age estimation. For analysis of facial images various local feature descriptors such as HOG, SIFT, LBP are used. These local descriptors are found suitable for both face recognition as well as age estimation task. Which implies these features carry information about both identity and age. But the dimensionality reduction techniques are not able to discriminate between aging feature and other facial features while reducing the dimension of the local features. Hence, it is highly essential to select only those features which carry the relevant aging information. Therefore, along with the analysis of manifold features, we also provide the analysis of feature selection methods for age estimation. After extracting relevant feature we apply orthogonal Gaussian process regression for estimating exact age.

### 3.1   Aging Manifold Features

Suppose the facial feature space $\mathcal{F}$ is represented as $\mathcal{F} = \left\{ f_i : f_i \in \mathbb{R}^D \right\}_{i=1}^{N}$ where $D$ is dimension of the data and $N$ is number of face images. True age labels $a_i$ are represented as $y = \{a_i : a_i \in \mathbb{N}\}_{i=1}^{N}$. We want to learn a low dimensional manifold $\mathcal{G}$ that is embedded in $\mathcal{F}$ and subsequently a manifold aging feature $\left\{ x_i : x_i \in \mathbb{R}^d \right\}_{i=1}^{N}$ with $d << D$. More specifically, our goal of learning the manifold is to find a $D \times d$ projection matrix $\mathbf{P} = [p_1, p_2 \cdots, p_d]$ such that $X = \mathbf{P}^T F$ where $F = [f_1, f_2, \cdots, f_N] \in \mathbb{R}^{D \times N}$. Various linear as well as nonlinear manifold learning techniques are available in the literature. Different from linear techniques, the nonlinear methods are designed to handle complex nonlinear data. Real world data mostly forms a highly nonlinear manifold. In such situations the solution lies with the nonlinear dimensionality reduction techniques. It should be noted that, in case of complex artificial tasks, the performance of nonlinear techniques surpasses that of linear counterparts. We adopt various following linear as well as nonlinear techniques for learning the manifold features from the feature space.

**Principle Component Analysis (PCA).** The PCA is a popular dimensionality reduction technique which constructs a low-dimensional representation of the data that is based on data variance. In PCA the projection matrix is obtained as $\underset{\|P\|=1}{\mathrm{argmax}}(P^T S P)$ where $S$ denotes the scatter matrix. In PCA, important information is extracted using data variance and further this information is expressed into a set of orthogonal basis popularly known as principal components (PCs). The principal components are obtained such that the first principal components retain maximum variance. The principal components are the eigenvectors of a data covariance matrix.

**Isometric Feature Mapping ISOMAP.** In many practical applications high-dimensional data lies on or close to a smooth low-dimensional manifold. In case of low dimensional representation, if pairwise Euclidean distances are considered

then two data points which are near in the original space may remain far in manifold space. This problem is solved by preserving pairwise geodesic distances between datapoints in Isomap [30] technique. The distance between any two points measured along the surface is called as Geodesic or curvilinear distance. For given datapoints $(x_i)_{i=1}^N$ the geodesic distances are computed by constructing a neighborhood graph. In the neighborhood graph, the connection between the each datapoint $x_i$ is established based on its $k$ nearest neighbors $(x_{ij})_{j=1}^k$ in the dataset X. Finally classical PCA is applied to the matrix of graph distances to construct an embedded manifold. Limitation of the Isomap algorithm is the high computational complexity.

**Locally Linear Embedding (LLE).** Local Linear Embedding (LLE) also constructs a graphical representation of the datapoints. Local properties of the data are preserved by this technique. In LLE, linear combination of nearest neighbors is used to represent high dimensional datapoints.

In case of LLE, the low-dimensional representation is achieved by retaining the reconstruction weights in the linear combinations as much same as possible. To encode the local properties of the manifold around a datapoint $x_i$ linear combination of datapoint are written as a linear combination of its $k$ nearest neighbors $x_{ij}$. Due to the local linearity assumption, the reconstruction weights $w_i$ of the datapoints $x_i$ become invariant to rotation, translation and scaling. Such invariance to the transformations preserves the reconstruction weights in the lower dimensional space.

**Orthogonal Locality Preserving Projections (OLPP).** Using the LPP [21] approach, the OLPP constructs orthogonal basis for discriminative manifold. The local neighborhood distance information is used in LPP to preserve manifold structure. The weight is computed as $s_{ij} = exp\left(\frac{-\|x_i - x_j\|^2}{t}\right)$ if $x_j$ and $x_i$ are $k$ nearest neighbors else $s_{ij} = 0$. The optimal projection matrix is computed from,

$$P = \underset{P^T X D X^T}{\operatorname{argmin}} \sum_{i=1}^N \sum_{j=1}^N \left(P^T x_i - P^T x_j\right)^2 s_{ij} \tag{1}$$

### 3.2   Feature Selection

Among various machine learning approaches, feature selection is a technique which selects and rank relevant features according to their degrees of relevance and preference. Reduction of dimensionality and noise in data sets is the main objective of feature selection which further results in the improvement of the accuracy and performance of classification or regression methods. The feature selection problem is defined as: given a feature space $\mathcal{F} = \left\{f_i : f_i \in \mathbb{R}^D\right\}_{i=1}^N$ of $N$ face images and ground truth $a_i$, the feature selection problem is to find a low subspace of relevant d dimensional features from the $D$ dimensional feature

space ($d << D$) that optimally characterizes the ground truth. The optimality condition is generally given to select the best features. Feature selection methods are applied to select low dimensional discriminative feature subspace from the high dimensional indiscriminative features. Feature selection methods are categorized as wrappers, embedded methods and filter methods. Classifiers are used in wrappers to assign the scores to the feature subset. The selection process is embedded in the classifiers for feature selection in the embedded methods. The filter method ignores the classifiers and analyzes intrinsic properties of the data for feature selection. Ranking and subset selection are two important operations in the feature selection techniques. In this work, our focus is mainly on filter methods which are computationally efficient than wrapper methods [4]. Among filter methods, we adopt Laplacian Scores (LS) and maximum-Relevance Minimum-Redundancy (mRMR) as discriminative aging feature selection methods for age estimation

**Laplacian Score (LS).** For relevant feature selection Laplacian score uses an unsupervised learning method i.e. k-means clustering, hence LS is also an unsupervised method. In LS, using the locality preserving power of the feature, a score for each feature is computed. Using (2) the Laplacian Score $L_r$ of the $r^{th}$ feature $f_r$ is computed.

$$L_r = \frac{\sum_{ij} (f_{ri} - f_{rj})^2 S_{ij}}{var(f_r)} \tag{2}$$

where $var(f_r)$ represents the variance of the $r^{th}$ feature, $S_{ij}$ denotes the similarity between two nodes which is nonzero if nodes are in k-neighbor of each other. Relevant feature implies the larger value of $S_{ij}$, smaller difference $(f_{ri} - f_{rj})$ which results into smaller Laplacian Score. Therefore, the ascending order of the score represents the importance of the features.

**max-Relevance Min-Redundancy (mRMR).** One of the popular approaches to feature selection is a max-Relevance Min-Redundancy. It is a supervised feature selection algorithm that selects the features with the highest relevance to target class (label). Correlation or mutual information is used to characterize the degree of relevance. For given two features $f_i$ and $f_j$ the mutual information $I(f_i, f_j)$ is computed as,

$$I(f_i, f_j) = \sum_{i,j} p(f_i, f_j) log \left( \frac{p(f_i, f_j)}{p(f_i) p(f_j)} \right) \tag{3}$$

To measure the relevance level or a similarity level among the features the mutual information is used. Our objective is to select dissimilar features so that redundancies are avoided. Dissimilarity implies minimum or less redundant feature which results into compact and relevant feature set and hence a better

and low dimensional representation of the dataset. The minimum redundancy condition to acquire the subset $S$ of selected features is defined as,

$$minW_I, \quad W_I = \frac{1}{|S|^2} \sum_{i,j \in S} I(f_i, f_j) \tag{4}$$

where $|S|$ denotes number of features in subset $S$. The maximum relevance condition is given by,

$$maxV_I, \quad V_I = \frac{1}{|S|} \sum_{i \in S} I(c, f) \tag{5}$$

where $c$ represents the target class labels $c = (c_1, c_2, \cdots, c_N)$. In mRMR, relevant features are selected by optimizing both (4) and (5).

### 3.3    Regression

After finding the low-dimensional representation of the facial image space via manifold embedding and/or feature selection, we define the age estimation as a regression problem. Various linear and nonlinear regression methods are available in the literature. For a comparative analysis of the manifold and features selection methods we use Orthogonal Gaussian Process (OGP) regression [36] on the extracted aging feature.

## 4    Experiments and Results

### 4.1    Experimental Setup

For benchmarking with the state-of-the-art methods, we have used We have MORPH II [29] facial aging database. It consist of 55,314 facial images spanning the age range 16 to 77 years. The images in this database contain large variations with respect to ethnicity, age and gender. Due to different ethnic origins different errors level are observed. To avoid this biasness on the performance [2,3,5,32] selected images from only Caucasian descent for the experimentation. We have also followed similar settings and used a randomly selected set of 10,000 images of Caucasian descent for experimentation. We have used 80% of them are used for training and 20% for testing.

### 4.2    Experiments and Results

For evaluation of the proposed method we have used two popular evaluation metrics i.e. Cumulative Score (CS) and Mean Age Error (MAE). To demonstrate the effectiveness of the manifold learning and feature selection scheme, we perform regression on the aging features extracted from the manifold and through feature selection. MAE and CS values using different manifold and feature selection techniques are listed in Table 1. We observed that feature selection methods are more suitable for age estimation than the manifold learning.

**Table 1.** Performance of the proposed method with different feature transformation and feature selection methods.

| Method | Reduced feature dimension | MAE (Years)/CS @ 5 Years (%) |
|---|---|---|
| PCA | 500 | 5.2/60.98 |
| OLPP | 500 | 4.65/71.35 |
| ISOMAP | 400 | 4.75/69.75 |
| LLE | 500 | 4.18/73.51 |
| mRmR | 500 | 3.53/81.2 |
| Laplacian score | 400 | 3.51/80.7 |

Furthermore, we have also analyzed the effect of dimensionality of various manifold and feature selection methods on MAE. Upper limit for the reduced dimensionality was selected as 500. Figure 2 compares MAE versus dimensionality of various manifold and feature selection methods. We clearly observe in Fig. 2 that feature selection methods surpass the manifold learning methods. We also compare the proposed approach with the state-of-the-art age estimation algorithms such as Scattering Transform (ST) [2], RED-SVM [3], CA-SVR [5] and Relative Attribute+SVM [32]. Benchmarking results of the proposed approach and various state of-the-art approaches in terms of CS at 5 years error and MAE are presented in Table 2 and shown in Fig. 3. Clearly, the proposed method surpasses the competing approaches in terms of MAE and CS. Therefore, the feature selection is an important direction for age estimation and discriminative manifold learning.
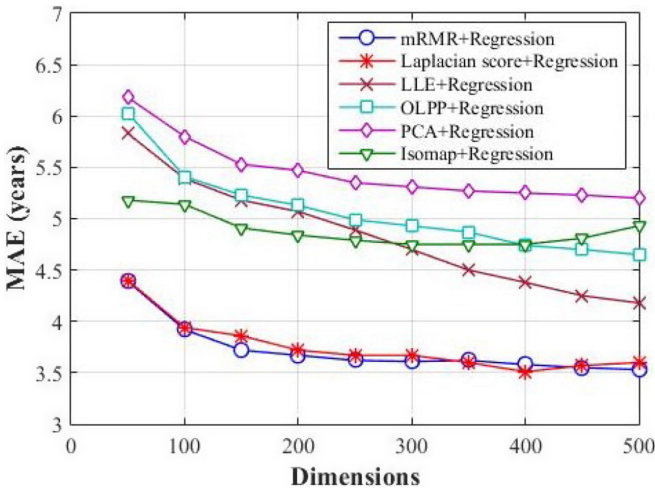


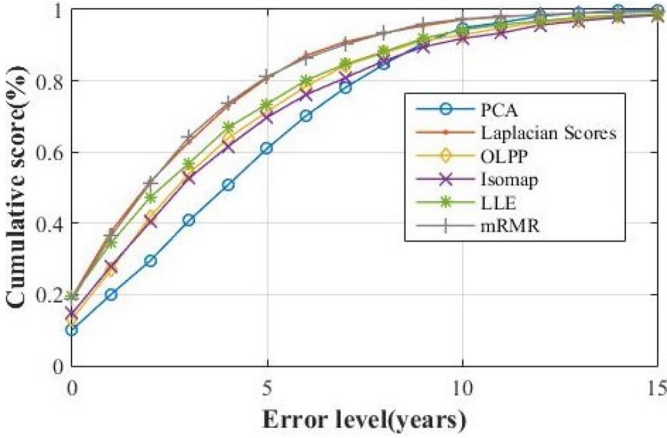**Fig. 2.** MAE versus dimensionality of aging subspace.

**Fig. 3.** MAE versus CS.

**Table 2.** Comparison of the proposed age estimation method with previous works in terms of MAE (Years) and CS (% ).

| Method | Face representation | MAE (Years)/CS @ 5 Years (%) |
|---|---|---|
| ST [2] | Scattering transform | 3.7/56.03 |
| RED-SVM [3] | AAM | 6.4/- |
| CA-SVR [5] | AAM | 5.8/- |
| Relative attribute+SVM [32] | SIFT | 4/- |
| Han et al. [20] | Demographic informative features | 5.10/− |
| CNN [23] | Deep features | 3.88/− |
| Multi-scale CNN [34] | Deep features | 3.63/− |
| CPNN [13] | Deep features | 4.87/− |
| Proposed method | Manifold lerning | 4.18/73.51 |
| Proposed method | Feature selection | **3.51/80.7** |

## 5   Conclusion

In this paper, we proposed a different perspective of dimensionality reduction, i.e. manifold feature and feature selection for age estimation. The technique learns discriminative aging features from these two techniques. A feature based discriminant manifold learning and feature selection both these approaches have not so far been previously considered for age estimation. In our opinion instead of looking for a new feature descriptor for age estimation, extracting or selecting relevant features from the existing feature representations is an important future direction. Our experimental analysis validates our claim and shows that the performance of the proposed method using feature selection surpasses the-state-of-the-art methods.

# References

1. Cai, D., He, X., Han, J., Zhang, H.J.: Orthogonal laplacianfaces for face recognition. IEEE Trans. Image Process. **15**(11), 3608–3614 (2006)
2. Chang, K.Y., Chen, C.S.: A learning framework for age rank estimation based on face images with scattering transform. IEEE Trans. Image Process. **24**(3), 785–798 (2015)
3. Chang, K.Y., Chen, C.S., Hung, Y.P.: A ranking approach for human ages estimation based on face images. In: 2010 20th International Conference on Pattern Recognition, pp. 3396–3399. IEEE (2010)
4. Chang, K.Y., Chen, C.S., Hung, Y.P.: Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 585–592. IEEE (2011)
5. Chen, K., Gong, S., Xiang, T., Change Loy, C.: Cumulative attribute space for age and crowd density estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2467–2474 (2013)
6. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
8. Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J., Vapnik, V.: Support vector regression machines. In: Advances in Neural Information Processing Systems, pp. 155–161 (1997)
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2012)
10. Fernández, C., Huerta, I., Prati, A.: A comparative evaluation of regression learning algorithms for facial age estimation. In: Ji, Q., B. Moeslund, T., Hua, G., Nasrollahi, K. (eds.) FFER 2014. LNCS, vol. 8912, pp. 133–144. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-13737-7_12
11. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. IEEE Trans. Multimedia **10**(4), 578–584 (2008)
12. Fu, Y., Xu, Y., Huang, T.S.: Estimating human age by manifold analysis of face pictures and regression on aging features. In: 2007 IEEE International Conference on Multimedia and Expo, pp. 1383–1386. IEEE (2007)
13. Geng, X., Yin, C., Zhou, Z.H.: Facial age estimation by learning from label distributions. IEEE Trans. Pattern Anal. Mach. Intell. **35**(10), 2401–2412 (2013)
14. Geng, X., Zhou, Z.H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2234–2240 (2007)
15. Geng, X., Zhou, Z.H., Zhang, Y., Li, G., Dai, H.: Learning from facial aging patterns for automatic age estimation. In: Proceedings of the 14th ACM international conference on Multimedia, pp. 307–316. ACM (2006)
16. Guo, G., Fu, Y., Dyer, C.R., Huang, T.S.: Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Trans. Image Process. **17**(7), 1178–1188 (2008)
17. Guo, G., Mu, G.: Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In: 2011 IEEE Conference on Computer Vision and Pattern recognition (CVPR), pp. 657–664. IEEE (2011)
18. Guo, G., Mu, G.: Joint estimation of age, gender and ethnicity: CCA vs. PLS. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–6. IEEE (2013)

19. Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 112–119. IEEE (2009)
20. Han, H., Otto, C., Jain, A.K.: Age estimation from face images: Human vs. machine performance. In: 2013 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2013)
21. He, X., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing Systems, pp. 153–160 (2004)
22. Huerta, I., Fernández, C., Prati, A.: Facial age estimation through the fusion of texture and local appearance descriptors. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8926, pp. 667–681. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16181-5_51
23. Huerta, I., Fernández, C., Segura, C., Hernando, J., Prati, A.: A deep analysis on age estimation. Pattern Recogn. Lett. **68**, 239–249 (2015)
24. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) **34**(1), 621–628 (2004)
25. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Mach. Intell. **24**(4), 442–455 (2002)
26. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: A study of face recognition as people age. In: 2007 IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8. IEEE (2007)
27. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157. IEEE (1999)
28. Pontes, J.K., Britto Jr., A.S., Fookes, C., Koerich, A.L.: A flexible hierarchical approach for facial age estimation based on multiple features. Pattern Recognit. **54**, 34–51 (2016)
29. Ricanek, K., Tesafaye, T.: Morph: A longitudinal image database of normal adult age-progression. In: 2006 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 341–345. IEEE (2006)
30. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
31. Wang, S., Tao, D., Yang, J.: Relative attribute SVM+ learning for age estimation. IEEE Trans. Cybern. **46**(3), 827–839 (2015)
32. Wang, S., Tao, D., Yang, J.: Relative attribute SVM+ learning for age estimation. IEEE Trans. Cybern. **46**(3), 827–839 (2016)
33. Weng, R., Lu, J., Yang, G., Tan, Y.P.: Multi-feature ordinal ranking for facial age estimation. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–6. IEEE (2013)
34. Yi, D., Lei, Z., Li, S.Z.: Age estimation by multi-scale convolutional network. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9005, pp. 144–158. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16811-1_10
35. Zhang, Y., Yeung, D.Y.: Multi-task warped gaussian process for personalized age estimation. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2622–2629. IEEE (2010)
36. Zhu, K., Gong, D., Li, Z., Tang, X.: Orthogonal gaussian process for automatic age estimation. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 857–860. ACM (2014)