

AGE ESTIMATION WITH REGARD FOR CLASSIFIABLE ABILITY OF EACH COMPONENT IN REDUCED DIMENSION AGE MANIFOLD

Pak DuHo, Ri KumHyok and Hyon CunGyong

Institute of Information Science, Kim Il Sung University, D.P.R. of Korea

Abstract

A new age estimation method that takes classifiable ability of each component in age manifold into account is considered. First, we analysis the age classification rate of each component in reduced dimension age manifold. Second, we apply this property to kernel function in popular method such as SVM. This is implemented by weighted kernel function. Finally, we evaluate this method in “wild” face image database. Experimental results demonstrate the effectiveness and robustness of our proposed framework.

Keywords:

Age Estimation, Support Vector Machine, Support Vector Regression

1. INTRODUCTION

The human face, as a unique identification feature, conveys a lot of representative information such as gender, age and expressions. In recent years, there has been growing interest in age estimation based on face images. Age information is useful in a variety of applications, such as human-computer interaction, surveillance monitoring, and video content analysis [24].

In general, automatic age estimation based on face image has three steps; the face detection in image, the age feature extraction, and the age prediction. Here, the age feature extraction is very important in age estimation. Age feature extraction has two stages; face feature presentation and projection to low dimension space.

The face feature presentation converts face appearance into feature vector that is computable and presents human face's trait. In the prior works, they proposed many approaches such as ASM and AAM [6]-[8], [21], Gabor [4], [13], SFP [14], LBP [11], [12], [22], BIF [15], [16], WLBP [18], GLOH [19] and others [2], [3].

Projection to low dimension space is the step that extracts factors reflecting human age, and this step decreases the information necessary for age prediction by removing the redundant information and reducing the number of unknown variables. So this can construct age manifold that presents age distribution in digital image space. Typical methods of projection to low dimension space are PCA [1], NPP [5], UDP [18], LPP [9], OLPP [10], [11], [26], CEA [20].

For age estimation, many methods are used including SVM or SVR [15]-[17], [21], [26]-[28], quadratic regression [20], and et al [18], [19], [29]-[32].

We propose the new method that uses the weight of each components of kernel function in discriminative space to do regression in typical traditional SVM or SVR methods of age estimation. This method can be applied to not only SVM but also the other regression, especially it is more effective in projection space by supervised learning.

The paper is organized as follows. In section 2, we review previous age estimation methods. In section 3, we analysis influence of each component in projection to low dimensional age manifold. The new estimation method based on each kernel function factor's weight is proposed in section 4, and the proposed methods is evaluated in section 5 and section 6 is conclusion.

2. RELATED WORK

This section reviews previous age estimation methods, especially methods using SVM or SVR. For the first time, Lanitis used PCA projection and Active Appearance Models (AAMs) to progress age estimation, where AAMs combine shape and intensity variation in face images [23]. Fu and Huang [20] generated low dimensional discriminative aging manifold by CEA projection and then applied quadratic regression to age estimation. Guo et al. [26] introduced the OLPP (Orthogonal Locality Preserving Projection) and estimated the human age by combining the SVM and SVR. In this method, a robust regression is approximately applied to the data and then identification is used for local matching.

Guo et al. [15] investigated biologically inspired features (BIF) comprised of a pyramid of Gabor filters in all positions in facial images, and used either Support Vector Machine (SVM) or SVR with Radial Basis Function (RBF) Kernels for evaluation. Xu obtained low-dimensional discriminative age manifold by UDP in WLBP face expression and then estimated the age using cosine distance [18]. K.Y Chang proposed the binary ranking method of K-1 degree that is called Ordinal Hyper-Plane Ranker based on level order, where SVM identification with loss function is used for the binary ranking of K-1 degree [17], [27]. Lu proposed a novel method based GLOH (gradient location and orientation histogram) representation and MTL (multi-task learning) feature selection along with ridge regression for global age estimation [19]. Liu et al. [25] proposed a Grouping Estimation Fusion (GEF) framework for facial age estimation.

Karthikeyan and Balakrishnan [28] proposed the method using hybrid filter based feature extraction. First, they used Gaussian, Gabor and Hybrid Filter to extract orientation features and local wrinkle features. The extracted features are classified separately using multi-SVM classifier and used age estimation with prefixed threshold [28].

Recently, deep learning techniques such as Convolutional Neural Networks (CNN) have been applied to human age estimation to learn aging features directly from large-scale facial data [29]-[32]. However, these methods mainly ignored the ordinal information in age labels, or over-simplified it to a linear model. Further they require relatively expensive big data, long learning time and database's equality in age range.

Peng Hou et al. [33] proposed an algorithm called Semi-supervised Adaptive Label Distribution Learning to solve the dilemma and improve the performance using unlabeled data for facial age estimation.

In preceding methods of estimating age using the SVM or SVR, regression or ranking was done without regard to influence of each factor in discriminative age manifold to age variance. Therefore we analysis and use the weight of each components of kernel function in discriminative space.

3. INFLUENCE OF EACH COMPONENT IN REDUCED DIMENSIONAL AGE MANIFOLD

Here, we consider the influence of each projection component to age dispersion in CEA (Conformal Embedding Analysis), OLPP (Orthogonal Locality Preserving Projection) that are typical dimension reduction approaches.

3.1 INFLUENCE OF EACH PROJECTION COMPONENT TO AGE DISPERSION IN CEA

Learning of discriminative age manifold uses the following models in [20].

Suppose the image space I is represented by a set of aligned face image $X = \{x_i, x_i \in R^D\}_{i=1}^n$ of m subjects in the order of subject's age. D is dimension of face image. A ground truth set $L = \{l_i, l_i \in N\}_{i=1}^n$ associated with the images provides the age label.

We define two n -node graphs g_s, g_d and their corresponding $n \times n$ affinity matrices W_s, W_d . The i^{th} node of the graph represents the data point x_i . For the graph g_s , we only consider each pair of data x_i, x_j from the same class with $l_i = l_j$. For the graph g_d , we only consider each pair of data x_i and x_j from the different class with $l_i \neq l_j$. An edge is constructed between nodes i and j if x_i is among the k_s or k_d nearest neighbours of x_i and vice versa, where parameters k_s and k_d are chosen empirically.

If node i and j are connected, the weight of the edge between x_i and x_j is set by $w_{ij} = \exp\left(\frac{-\text{Dist}(x_i, x_j)}{t}\right)$, where t is a free parameter to be tuned empirically. Otherwise, $w_{ij} = 0$ if node i and j are not connected.

$$\text{dist}(x_i, x_j) = 1 - \frac{[x_i - \text{Mean}(x_i)][x_j - \text{Mean}(x_j)]}{\|x_i - \text{Mean}(x_i)\| \cdot \|x_j - \text{Mean}(x_j)\|} \quad (1)$$

In Eq.(1), $\text{Mean}(x_i)$ means that is average of images if there are many faces for same object. Therefore CEA objective function is defined as follows:

$$\arg \max_p \sum_{i,j=1}^n y_i^T (y_i - y_j) \cdot w_{ij}^d, \sum_{i,j=1}^n y_i^T (y_i - y_j) \cdot w_{ij}^s = \delta \quad (2)$$

where, δ is a constant number. The weight matrices $W_s[i, j] = w_{ij}^s$ and $W_d[i, j] = w_{ij}^d$ are symmetric and non-negative.

Define the $D \times d$ projection matrix $P = [p_1, p_2, \dots, p_d]$, $p_i = (p_{0i}, p_{1i}, \dots, p_{Di})^T$.

The basic idea for leaning the manifold space is to find the matrix P satisfying $Y = P^T \tilde{X}$.

where $\tilde{x}_i = \frac{(x_i - \text{Mean}(x_i))}{\|x_i - \text{Mean}(x_i)\|}$ and $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n] \in R^{D \times n}$.

Following the matrix formulation Eq.(2) can be rewritten as $\arg \max_p \{P^T \tilde{X} (D_d - W_d) \tilde{X}^T P\}$,

Subject to trace

$$\{P^T \tilde{X} (D_s - W_s) \tilde{X}^T P\} = \delta \quad (3)$$

where $D_d[i, i] = \sum_j w_{ij}^{(d)}$ and $D_s[i, i] = \sum_j w_{ij}^s$.

Eq.(3) can be solved in a closed form through the following reformulation.

$$[\tilde{X} (D_d - W_d) \tilde{X}^T] P = \lambda [\tilde{X} (D_s - W_s) \tilde{X}^T] P \quad (4)$$

The column vectors $\{p_i\}_{i=1}^d$ of the projection matrix P indicate the eigenvectors of $[\tilde{X} (D_s - W_s) \tilde{X}^T]^{-1} [\tilde{X} (D_d - W_d) \tilde{X}^T]$ corresponding to the d largest eigenvalues, which can be solved by SVD.

If we multiply the P^T in left side at Eq.(4) and refer to Eq.(3), we can obtain the following reformation.

$$P^T [\tilde{X} (D_d - W_d) \tilde{X}^T] P = \lambda P^T [\tilde{X} (D_s - W_s) \tilde{X}^T] P$$

$$P^T [\tilde{X} (D_d - W_d) \tilde{X}^T] P = \lambda \delta \quad (5)$$

Eq.(5) is showed that projection vector maximizing Eq.(3) is eigenvector of corresponding to the largest eigenvalues. It means that first factor projected by eigenvector of corresponding to the largest eigenvalue make an important contribution to age variance. In generally, if we arrange factors in order eigenvalues, they were arranged in order influence of age variance.

3.2 INFLUENCE OF EACH PROJECTION COMPONENT TO AGE DISPERSION IN OLPP

OLPP projection method is proposed in [11]. Optimum projection vector is obtained by solving Eq.(1) under restricted condition $p^T X D X^T p = 1$ in OLPP.

$$p_{opt} = \arg \min_p \sum_{i=0}^n \sum_{j=0}^n (p^T x_i - p^T x_j)^2 S_{ij}$$

$$= \arg \min_p p^T X L X^T p \quad (6)$$

$L = D - S$, $D_{ii} = \sum_j S_{ij}$ and D_{ii} is local density measure neighborhood of S is similarity matrix whose each element $S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$, in case that x_i is in neighborhood of x_j or x_j is in neighborhood of x_i , and otherwise $S_{ij} = 0$.

$$X L X^T p = \lambda X D X^T p \quad (7)$$

After all finding a projection vector p is equivalent to getting eigenvalue of Eq.(7). Multiplying p^T to both sides of Eq.(7) and considering Eq.(6), we can get Eq.(8).

$$p^T XLX^T p = \lambda \quad (8)$$

Hence, we can notice that projection by eigenvector corresponding to a minimum eigenvalue influences less to locality variation. Therefore arranging components with respect to least size of their eigenvalues, the influence of components to preserving locality is proportional to size of eigenvalue.

This fact also stands for PCA, UDP and LPP. Eventually, we can conclude that the influences of components in discriminative age manifold space to preserving locality must be considered in estimating age.

4. AGE REGRESSION CONSIDERING THE INFLUENCES OF EACH COMPONENT

Here, we generally consider the method covered in [26] of several papers as regard to age estimation based on SVM.

For age regression, Gaussian radial basis function kernel was adopted. A radial basis function is

$$k(y, y') = e^{-\gamma \|y - y'\|^2}$$

where γ is a constant to adjust the width of the Gaussian function.

Given the kernel mapping, the age estimation model of the nonlinear SVR is obtained as

$$\langle w, y \rangle = \sum_{i=1}^d (\alpha_i - \alpha_i^*) k(y_i, y) \quad (9)$$

and

$$f(y) = \sum_{i=1}^d (\alpha_i - \alpha_i^*) k(y_i, y) + b \quad (10)$$

where α_i, α_i^* are Lagrange multipliers and b is constant number.

Let $\lambda_i, i = \overline{1, d}$ be d eigenvalues in Eq.(4), Eq.(7).

Contribution proportion w_i to changes of y_i corresponding to i^{th} eigenvalue λ_i is defined as follows.

$$\begin{aligned} \sigma &= \sum_{i=1}^d \lambda_i \\ w_i &= \lambda_i / \sigma \end{aligned} \quad (11)$$

In this paper, exponent item $\|y - y_i\|^2$ of Gaussian radial base function is defined as follows on the basis of contribution proportion that each component influences to change.

$$\|y - y_i\|^2 = \sum_i^d (y_i - y_i')^2 \cdot w_i^2 \quad (12)$$

Applying SVM based on the kernel function that reflects contribution proportion of each component in [15] - [17], [21], we can conclude that proposed method get better performance by following experimental results.

5. EXPERIMENTS

We test the proposed method using wild database from internet and criminal mug-shot database (Fig.1). We use high resolution images of 5810 mans and 6028 women that are 0 to 80 in age to perform learning. And then we are to examine age

estimation methods using the images of 5529 mans and 7764 women from 16 to 84 years that have not been learnt.



Fig.1. Sample images of the our test database

We verify the effectiveness of our method by applying it to SVM, SVR, LARR and OHR regressions on the discriminative age manifold space obtained from CEA among the PCA, LPP, CEA and OLPP projection methods.

The performance of age estimation can be measured by two different measures: the Mean Absolute Error (MAE) and the Cumulative Score (CS). The MAE is defined as the average of the absolute errors between the estimated ages and the ground truth ages, $MAE = \sum_{k=1}^N |\hat{l}_k - l_k| / N$ where l is the ground truth age for the test image k , \hat{l} is the estimated age, and N is the total number of test images. The cumulative score is defined as

$$CS(j) = \frac{N_{e \leq j}}{N},$$

where $N_{e \leq j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years.

Table.1. Performance of proposed method and others (MAE)

Method	Male	Female
SVM	5.01242	4.69897
LARR	3.65384	4.44916
OHR	4.41263	4.44952
WSVM	4.09154	3.8387
WOHR	3.17932	3.76977
WLARR	3.13947	3.92449

The Table.1 shows the results of MAE analysis on age manifold space obtained from CEA projection and The CS measures are shown in Fig.2, Fig.3 and Fig.4. In Table.1, WSVM and WOHR, WLARR is applied our proposed method to SVM and OHR, LARR.

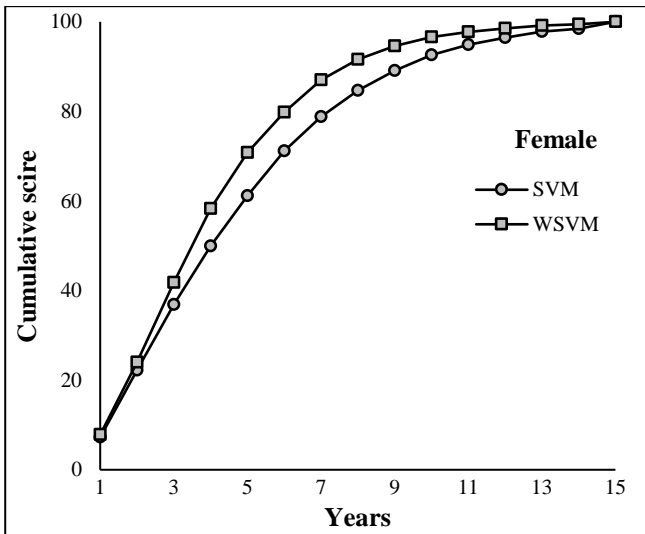
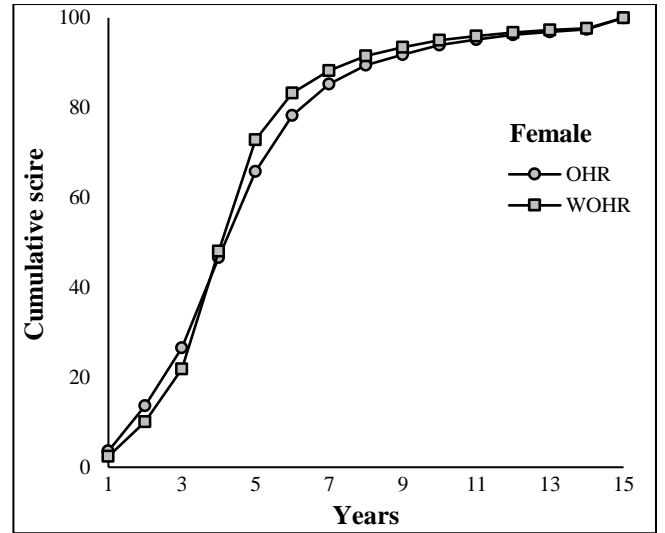
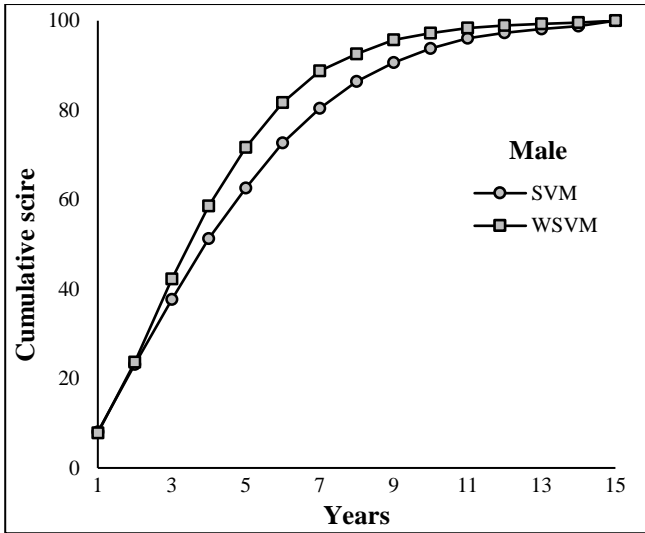


Fig.3. CS measures in OHR and WOHR (Top: male, Bottom: female)

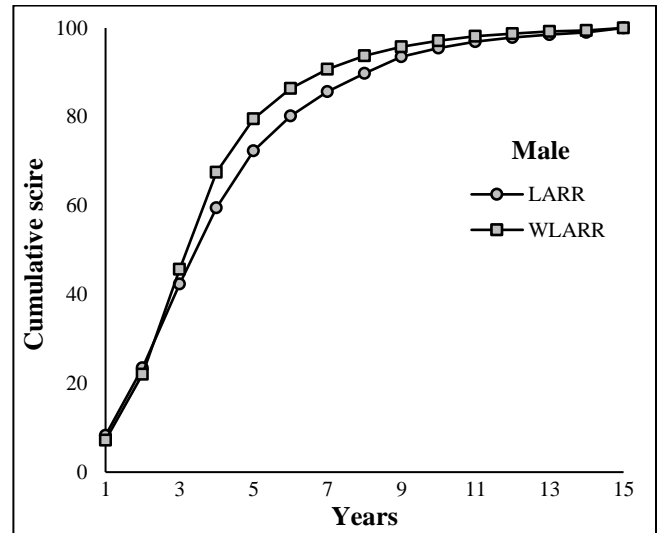


Fig.2. CS measures in SVM and WSVM (Top: male, Bottom: female)

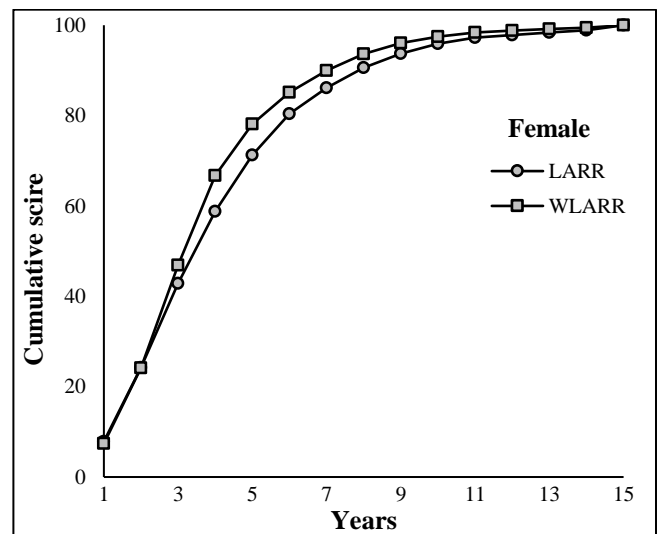
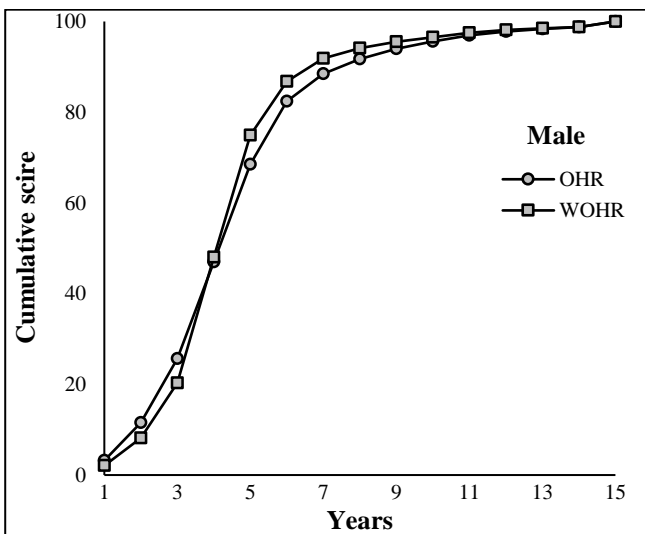


Fig.4. CS measures in LARR and WLARR (Top: male, Bottom: female)

The results show that proposed weighted method has better performances than previous ones. In Table.1, we can see that the MAE decreases by applying the weighted kernel function. From these results, WSVM and WOHR, WLARR improve MAE by about 15% compared to SVM and OHR, LARR. Although WOHR is low in age error range 1~4, others improve CS in all age error range,

6. CONCLUSION

In this paper we analyzed contribution proportion of each component variation in the reduced dimensional age manifold obtained by low dimensional projection. And we used weighted kernel function to improve the accuracy of age estimation. The weighted kernel function can be applied to different regression methods such as SVM, OHR, and LARR. Experimental results show that weighted methods consider affection of each component of reduced dimensional age manifold very well.

REFERENCES

- [1] M.A. Turk and A.P. Pentland, "Face Recognition using Eigenfaces", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586-591, 1991.
- [2] K.H. Liu, T.J. Liu, H.H. Liu and S.C. Pei, "Facial Makeup Detection Via selected Gradient Orientation of Entropy Information", *Proceedings of IEEE International Conference on Image Processing*, pp. 4067-4071, 2015.
- [3] T.J. Liu, K.H. Liu, H.H. Liu, and S.C. Pei, "Comparison of Subjective Viewing Test Methods for Image Quality Assessment", *Proceedings of IEEE International Conference on Image Processing*, pp. 3155-3159, 2015.
- [4] P.K. Sai, J.G. Wang and E.K. Teoh, "Facial Age Range Estimation with Extreme Learning Machines", *Neurocomputing*, Vol. 149, pp. 364-372, 2015.
- [5] Y.W. Pang, L. Zhang, Z.K. Liu, N.H. Yu and H.Q. Li, "Neighborhood Preserving Projections (NPP): A Novel Linear Dimension Reduction Method", *Proceedings of International Conference on Intelligent Computing*, pp. 117-125, 2005.
- [6] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Toward Automatic Simulation of Aging Effects on Face Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp. 422-455, 2002.
- [7] A. Lanitis, C. Draganova and C. Christodoulou, "Comparing Different Classifiers for Automatic Age Estimation", *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, Vol. 34, No. 1, pp. 621-628, 2004.
- [8] Y. Zhang and D. Yeung, "Multi-Task Warped Gaussian Process for Personalized Age Estimation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2622-2629, 2010.
- [9] X. He and P. Niyogi, "Locality Preserving Projections", Available at: <https://papers.nips.cc/paper/2359-locality-preserving-projections.pdf>.
- [10] D. Cai, X. He, J.W. Han and H.J. Zhang, "Orthogonal Laplacian Faces for Face Recognition", *IEEE Transactions on Image Processing*, Vol. 15, No. 11, pp. 3608-3614, 2006.
- [11] Hui Fang, Phil Grant and Min Chen, "Discriminant Feature Manifold for Facial Aging Estimation", *Proceedings of International Conference on Pattern Recognition*, pp. 12-16, 2010.
- [12] Z. Yang and H. Ai, "Demographic Classification with Local Binary Patterns", *Proceedings of International Conference on Biometrics*, pp. 464-473, 2007.
- [13] F. Gao and H. Ai, "Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method", *Proceedings of International Conference on Biometrics*, pp. 132-141, 2009.
- [14] S. Yan, T. S. Huang, H. Wang and X. Tang, "Ranking with Uncertain Labels", *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 96-99, 2007.
- [15] G. Guo, G. Mu, Y. Fu and T.S. Huang, "Human Age Estimation using Bio-Inspired Features", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 112-119, 2009.
- [16] Mohamed Y. El Dib and Hoda M. Onsi, "Human Age Estimation Framework using Different Facial Parts", *Egyptian Informatics Journal*, Vol. 12, No. 1, pp. 53-59, 2011.
- [17] Kang-Yu Chang, Chu-Song Chen and Yi-Ping Hung, "Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 131-135, 2011.
- [18] F. Xu, K. Luu, M. Savvides, Tien D. Bui and Ching Y. Suen, "Investigating Age Invariant Face Recognition Based on Periocular Biometrics", *Proceedings of International Joint Conference on Biometrics*, pp. 1-4, 2011.
- [19] Y. Liang, L. Liu, Y. Xu, Y. Xiang and B. Zou, "Multi-Task GLOH Feature Selection for Human Age Estimation", *Proceedings of IEEE International Conference on Image Processing*, pp. 241-245, 2011.
- [20] Y. Fu and T.S. Huang, "Human Age Estimation with Regression on Discriminative Aging Manifold", *IEEE Transactions on Multimedia*, Vol. 10, No. 4, pp. 578-584, 2008.
- [21] N.S. Lakshmi Prabha, J. Bhattacharya and S. Majumder, "Age Estimation using Gender Information", *Proceedings of IEEE International Conference on Computer Networks and Intelligent Computing*, pp. 211-216, 2011.
- [22] M. Gayathri and K. Chandra, "Face Verification with Aging using AdaBoost and Local Binary Patterns", *Proceedings of 7th Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 101-108, 2010.
- [23] A. Lanitis, C. Draganova and C. Christodoulou, "Comparing Different Classifiers for Automatic Age Estimation", *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, Vol. 34, No. 1, pp. 621-628, 2004.
- [24] J. Zeng, H. Ling, L.J. Lakecki and S. Fitzhugh, "Analysis of Facial Image across Age Progression by Humans", *ISRN Machine Vision*, Vol. 2012, pp. 505974-505977, 2012.
- [25] K.H. Liu, S. Yan and C.C. J. Kuo, "Age Estimation via Grouping and Decision Fusion", *IEEE Transactions on Information Forensics and Security*, Vol. 10, No. 11, pp. 2408-2423, 2015.
- [26] G. Guo, Y. Fu, C. Dyer and T. Huang, "Image-based Human Age Estimation by Manifold Learning and Locally Adjusted

- Robust Regression”, *IEEE Transactions on Image Processing*, Vol. 17, No. 7, pp. 1178-1188, 2008.
- [27] K.Y. Chang and C.S. Chen, “A Learning Framework for Age Rank Estimation based on Face Images with Scattering Transform”, *IEEE Transactions on Image Processing*, Vol. 24, No. 3, pp. 785-798, 2015.
- [28] D. Karthikeyan and G. Balakrishnan, “A Comprehensive Age Estimation on Face Images using Hybrid Filter based Feature Extraction”, *Biomedical Research*, Vol. 2017, pp. 610-618, 2017.
- [29] Z. Niu, M. Zhou, L. Wang, X. Gao and G. Hua, “Ordinal Regression with Multiple Output CNN for Age Estimation”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 503-507, 2016.
- [30] Shixing Chen, Caojin Zhang, Ming Dong, Jialiang Le and Mike Rao, “Using Ranking-CNN for Age Estimation”, Available at: <http://www.cs.wayne.edu/~mdong/cvpr17.pdf>.
- [31] X. Wang, R. Guo and C. Kambhamettu, “Deeply-Learned Feature for Age Estimation”, *Proceedings of IEEE Winter Conference on In Applications of Computer Vision*, pp. 534-541, 2015.
- [32] Tianyue Zheng, Weihong Deng and Jiani Hu, “Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 12-16, 2017.
- [33] Peng Hou, Xin Geng, Zeng-Wei Huo and Jia-Qi Lv, “Semi-Supervised Adaptive Label Distribution Learning for Facial Age Estimation”, *Proceedings of 31st AAAI Conference on Artificial Intelligence*, pp. 2015-2021, 2017.