Multi-Feature Ordinal Ranking for Facial Age Estimation

Renliang Weng[1], Jiwen Lu[2], Gao Yang[1] and Yap-Peng Tan[1]
[1]*School of EEE, Nanyang Technological University, Singapore*
[2]*Advanced Digital Sciences Center, Singapore*
Email: weng0017@e.ntu.edu.sg; jiwen.lu@adsc.com.sg; yanggao@ntu.edu.sg; eyptan@ntu.edu.sg

***Abstract*— In this paper, we propose a multi-feature ordinal ranking (MFOR) method for facial age estimation. Different from most existing facial age estimation approaches where age estimation is treated as a classification or a regression problem, we formulate facial age estimation as a group of ordinal ranking subproblems, and each subproblem derives a separating hyperplane to divide face instances into two groups: samples with age larger than k and samples with labels no larger than k. To better extract complementary information from different facial features, we construct multiple ordinal ranking models, each corresponding to a feature set, and aggregate them into an effective age estimator. Experimental results on two public face aging datasets are presented to demonstrate the efficacy of the proposed method.**

I. Introduction

Facial age estimation has been a hot research topic over the past decade. Age is an important biometric trait, and estimating it has many practical applications, such as human-computer interaction, age-specific image retrieval, and age-oriented visual advertisement. Different from other face analysis tasks, facial age estimation has several unique challenges:

- *Group specific*: People of different genders and ethnic groups usually have different aging processes.
- *Individually specific*: Different personal living styles and health conditions make age estimation more specific to individuals.
- *Insufficient data*: There is much difficulty in collecting sufficient data to cover the whole range of ages.
- *Ordinal label*: Age labels are numerical ordinal information. For example, the age label of 10 years old is more closely related to that of 11 than 12.

Recently, there have been a number of facial age estimation methods proposed in the literature [2], [3], [4], [7], [8], [9], [22], [17], [20], and some of them have achieved reasonably good performance. However, most existing methods only utilize a single set of features to represent facial appearance and predict age, which may not comprehensively encapsulate the discriminative information. For example, biologically inspired features (BIF) [10] capture facial saliency, and anthropometric models [13] reflect facial texture information. Each of such features reflects a particular point of view on the age estimation problem and does not comprehensively express the discriminative information encapsulated in features designed from other views. Active Appearance Models (AAM) [5] extract both facial shape and texture information for age estimation. However, these two features are simply concatenated, and their complementary information may not be effectively explored.

Recognizing the value and potential performance gain of combining multiple feature sets, we propose a multi-feature ordinal ranking (MFOR) method for facial age estimation. We perform feature extraction from multiple feature spaces in parallel, and train a binary classifier for each feature space at each age, where any input instance is classified as either older than or no older than the anchor age. The classifiers for each anchor age are then combined to form a stronger classifier, taking into consideration all available features. Different from most existing facial age estimation methods where age estimation is treated as a classification or a regression problem, we order these boosted anchor age classifiers to form an ordinal ranking model, so as to better reflect the ordinal nature of age labels. Experimental results on the widely used FG-NET and MORPH face databases are presented to demonstrate the efficacy of the proposed method.

The rest of this paper is organized as follows. Section II discusses related work. Section III details our proposed approach. Section IV provides the experimental results and Section V concludes the paper.

II. Related Work

Facial Age Estimation: Generally, a facial age estimation system consists of two parts [2], [3], [4], [7], [8], [9], [22], [17], [20]: feature representation and age prediction. For the first stage, discriminative features are extracted to represent the age information of face samples. Representative age-related features include anthropometric models [13], AAM [5], and BIF [10], each of which exploits some age traits for feature representation. For the second stage, age estimation can be cast as a multi-class classification [26], [8], [14] or regression problem [6], [18], [10], [9], [27], [24], [25]. While these methods have achieved some encouraging performance, most of them fail to explicitly utilize the ordinal age information of facial images in the prediction stage. To address this, several recent works formulated age prediction as a ranking problem [15], [19], [23], and ordinal hyperplanes ranker (OHRank) [4] has been the state-of-the-art method for ranking-based age estimation. Different from these age estimation methods, we present a multi-feature ranking model to estimate human ages by jointly learning multiple ordinal ranking models with multiple features, such

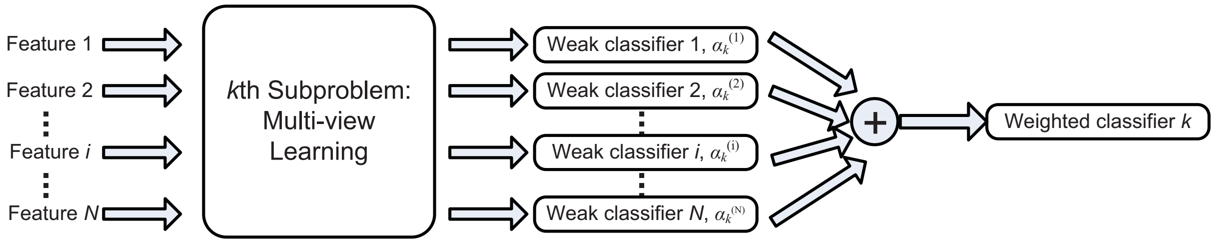

Fig. 1. Learning procedure of our proposed approach.

that discriminative information from different features can be effectively combined and boosted for age estimation.

Multi-View Learning: In machine learning literature, multi-view learning aims to learn a model from multiple feature representations, such that information observed from multiple views can be effectively fused for different classification or regression tasks, hence our approach is a multi-view learning approach. There have been many inspiring multi-view learning works proposed. Xia *et al.* [21] developed a multi-view spectral embedding method to find a low dimensional subspace where the distribution of each view is continuous and the complementary information of different views can be exploit; Lu *et al.* [16] presented a multi-view neighborhood repulsed metric learning method to obtain a discriminative distance metric for kinship verification. Different from these methods, our method explicitly learns a multi-view ranking model for age estimation, which is complementary to existing multi-view learning methods. Kittler *et al.* [12] proposed a method for jointly combining classifiers which utilize different feature sets and concluded that many existing approaches can be traced to decision fusion schemes, such as majority voting, max, min, weighted sum, etc., among which sum rule achieved the best result. Therefore in our proposed method, we fuse classifier outputs by using weighted sums scheme.

III. Proposed Approach

Given a face image, we can extract different feature sets from different views to represent the age information of the person. However, it's still not clear which feature set is the best representation for age estimation. Hence, it is natural to explore multiple feature sets simultaneously to enhance the discriminative power of the input features. Another motivation is that different feature sets can reflect different aging effects. For example, facial shape mainly changes during childhood and adolescence, while skin texture changes gradually during adulthood. Therefore, it's beneficial to combine different feature sets together by learning a set of weighted classifiers to form a stronger classifier for age estimation. Based on these reasons, we propose our multi-feature discriminative model for facial age estimation.

For a human observer, it is easier to distinguish the older one between two people than to estimate their actual ages from their face images. Inspired by this fact, we deem estimation as an ordinal ranking problem. Specifically, we divide the estimation problem into $K-1$ subproblems, where K is the number of age labels in the database, and the kth subproblem is constructed from its anchor age k, by which we separated the database of the ith feature set into two subsets, $P_k^{(i)}$ and $N_k^{(i)}$, as follows:

$$\begin{aligned} &P_k^{(i)} = \{(x_j^{(i)}, +1) | y_j > k\} \\ &N_k^{(i)} = \{(x_j^{(i)}, -1) | y_j \leq k\} \\ &s.t.\ 1 \leq k \leq K-1 \end{aligned} \tag{1}$$

Given the ith feature set $X^{(i)} = [x_1^{(i)}, \ldots, x_n^{(i)}] \in \mathbb{R}^{r_i \times n}$, wherein r_i is the feature dimension of the ith set, and $x_j^{(i)}$ is the jth instance. Let $\alpha_k^{(i)}$ be the weight for ith set for the kth ranking problem, we formulate our approach as the following optimization problem:

$$\begin{aligned} \min_{w_k^{(i)}, b_k^{(i)}, \xi^{(i)}} & \sum_{i=1}^{N} \alpha_k^{(i)} \left(\frac{1}{2} \left\langle w_k^{(i)}, w_k^{(i)} \right\rangle + C \sum_j \xi_j^{(i)} \right) \\ s.t.\ \ & z_k[j](\left\langle w_k^{(i)}, \phi_k(x_j^{(i)}) \right\rangle + b_k^{(i)}) \geq 1 - \xi_j^{(i)} \\ & \xi_j \geq 0, \sum_{i=1}^{N} \alpha_k^{(i)} = 1,\ \forall i, j \end{aligned} \tag{2}$$

where N is the number of feature sets, ϕ_k is a mapping into high dimensional space, accompanied with a kernel function as its inner product calculation, and $(w_k^{(i)}, b_k^{(i)})$ are the hyperplane parameters for ith feature set to solve the kth subproblem. $z_k[j] = 1$ if $x_j^{(i)} \in P_k^{(i)}$ and $z_k[j] = -1$ if $x_j^{(i)} \in N_k^{(i)}$. Fig. 1 illustrates the basic idea of our learning procedure.

The solution to Eq. (2) is $\alpha_k^{(i)} = 1$ corresponding to the minimal $\left(\frac{1}{2}\left\langle w_k^{(i)}, w_k^{(i)} \right\rangle + C \sum_j \xi_j^{(i)}\right)$ over different feature sets, and $\alpha_k^{(i)} = 0$ otherwise, which means only the best feature set would be chosen for classification, hence this scheme could be considered as max rule [12]. Therefore, we term this approach as Multi-Feature Max-fusion Ranking (MFMaxR). Note that MFMaxR is not the same as OHRank, as the former could utilize different feature sets for different subproblems, while in the latter case only one feature set is exploited throughout the problem. To fully utilize multiple feature sets, we modify $\alpha_k^{(i)}$ to be $\left(\alpha_k^{(i)}\right)^p$ for Eq. (2), the new objective function is defined as

$$\min_{w_k^{(i)},b_k^{(i)},\xi^{(i)}} \sum_{i=1}^{N} \left(\alpha_k^{(i)}\right)^p \left(\frac{1}{2}\left\langle w_k^{(i)}, w_k^{(i)}\right\rangle + C\sum_j \xi_j^{(i)}\right)$$
$$s.t.\ \ z_k[j]\left(\left\langle w_k^{(i)}, \phi_k(x_j^{(i)})\right\rangle + b_k^{(i)}\right) \geq 1-\xi_j^{(i)}$$
$$\xi_j \geq 0, \sum_{i=1}^{N}\alpha_k^{(i)} = 1,\ \forall i,j \tag{3}$$

Hence, MFMaxR is a special case of our MFOR approach when $p = 1$. Transforming the equation above by using lagrangians we could obtain

$$\begin{aligned} J = \min\ &\sum_{i=1}^{N}\left(\alpha_k^{(i)}\right)^p\left(\frac{1}{2}\left\langle w_k^{(i)}, w_k^{(i)}\right\rangle + C\sum_j \xi_j^{(i)}\right) \\ &- \sum_{i,j}\mu_{ij}\left(z_k[j]\left(\left\langle w_k^{(i)}, \phi_k(x_j^{(i)})\right\rangle + b_k^{(i)}\right) - 1 + \xi_j^{(i)}\right) \\ &- \lambda\left(\sum_{i=1}^{N}\alpha_k^{(i)} - 1\right) - \sum_{i,j}\tau_{ij}\xi_j^{(i)} \\ s.t.\ &\mu_{ij}\geq 0,\ \ \tau_{ij}\geq 0,\ \ \xi_j^{(i)} \geq 0 \end{aligned} \tag{4}$$

The problem above is not semidefinite quadratic, we use alternative optimization to obtain the optimal parameters. Firstly, we fix $\alpha_k^{(i)}$, then J could be minimized as below:

$$\begin{aligned} \frac{\partial J}{\partial w_k^{(i)}} &= \left(\alpha_k^{(i)}\right)^p w_k^{(i)} - \sum_j \mu_{ij} z_k[j]\phi_k(x_j^{(i)}) = 0; \\ \frac{\partial J}{\partial b_k^{(i)}} &= \sum_j \mu_{ij} z_k[j] = 0; \\ \frac{\partial J}{\partial \xi_j^{(i)}} &= \left(\alpha_k^{(i)}\right)^p C - \mu_{ij} - \tau_{ij} = 0 \end{aligned} \tag{5}$$

Combining Eq. (4) and Eq. (5), we could obtain dual function

$$\begin{aligned} \max &- \sum_{i=1}^{N}\frac{1}{2(\alpha_k^{(i)})^p}\sum_{m,n}\mu_{im}\mu_{in}z_k[m]z_k[n]\phi_k^T(x_m^{(i)})\phi_k(x_n^{(i)}) \\ &+ \sum_{i,j}\mu_{ij} \\ s.t.\ &0 \leq \mu_{ij} \leq \left(\alpha_k^{(i)}\right)^p C \end{aligned} \tag{6}$$

Now the problem is in the form of standard semi-definite quadratic programming problem, which could be solved by QP solver robustly. Having obtained $\mu_{ij}, w_k^{(i)}, \xi_j^{(i)}$ and $b_k^{(i)}$, J could be simplified as

$$\begin{aligned} J = \min &\sum_{i=1}^{N}\left(\alpha_k^{(i)}\right)^p\left(\frac{1}{2}\left\langle w_k^{(i)}, w_k^{(i)}\right\rangle + C\sum_j \xi_j^{(i)}\right) \\ &- \lambda\left(\sum_{i=1}^{N}\alpha_k^{(i)} - 1\right) \end{aligned} \tag{7}$$

TABLE I

MFOR TRAINING ALGORITHM FOR THE kTH SUBPROBLEM

Input: Training samples with multiple feature sets. $P_k^{(i)}, N_k^{(i)}$.
Parameters: iteration number T, and convergence error ϵ

Output: Weak classifiers and their weights $(w_k^{(i)}, b_k^{(i)}), \alpha_k^{(i)}$.

Step 1 (Initialization):
Let $\alpha_k^{(i)} = \frac{1}{N}$, for all $i = 1, 2, ..., N$.

Step 2 (Local Optimization):
For $r = 1, 2, ...T$, repeat
2.1 Optimize Eq. (6) to update $(w_k^{(i)}, b_k^{(i)})$ and $A_k^{(i)}$.
2.2 Update $\alpha_k^{(i)}$ through calculating Eq. (9).
2.3 Calculate J^r via Eq. (8).
2.4 If $r > 2$ and $|J^r - J^{r-1}| < \epsilon$, go to Step 3.

Step 3 (Output Results):
Output $(w_k^{(i)}, b_k^{(i)})$ and $\alpha_k^{(i)}$.

let's denote $A_k^{(i)} = \frac{1}{2}\left\langle w_k^{(i)}, w_k^{(i)}\right\rangle + C\sum_j \xi_j^{(i)}$, then

$$J = \min\sum_{i=1}^{N}\left(\alpha_k^{(i)}\right)^p A_k^{(i)} - \lambda\left(\sum_{i=1}^{N}\alpha_k^{(i)} - 1\right) \tag{8}$$

J could be optimized as below:

$$\frac{\partial J}{\partial \alpha_k^{(i)}} = p(\alpha_k^{(i)})^{p-1}A_k^{(i)} - \lambda = 0 \Rightarrow \alpha_k^{(i)} = \left(\frac{\lambda}{p * A_k^{(i)}}\right)^{\frac{1}{p-1}} \tag{9}$$

Since $\sum_{i=1}^{N}\alpha_k^{(i)} = 1$, if $p = 2$, we have

$$\alpha_k^{(i)} = \frac{\frac{1}{A_k^{(i)}}}{\left(\sum_j (A_k^{(i)})^{-1}\right)} \tag{10}$$

Having obtained $\alpha_k^{(i)}$, we can update $(w_k^{(i)}, b_k^{(i)})$ and $A_k^{(i)}$ by optimizing Eq. (6). These obtained updated parameters are then used to solve Eq. (8) to update $\alpha_k^{(i)}$ in return. The proposed MFOR training algorithm for the kth subproblem is summarized in Table I.

We summarize the whole process of our MFOR approach as follows:

1. For each anchor age k, where $1 \leq k < K$,
   a) For each feature set, divide the training data into two sets: $P_k^{(i)}$ and $N_k^{(i)}$.
   b) Use algorithm tabulated in Table I to establish a weighted classifier k with $f_k(x)$ as its decision function:

$$f_k(x) = sign\left(\sum_{i=1}^{N}\alpha_k^{(i)}\left(\left\langle w_k^{(i)}, \phi_k(x_j^{(i)})\right\rangle + b_k^{(i)}\right)\right) \tag{11}$$

2. Construct an age estimation rule $E(x)$ by collecting preferences information from all subproblems:

$$E(x) = 1 + \sum_{k=1}^{K-1}\frac{1}{2}\left(f_k(x) + 1\right) \tag{12}$$

Hence, our multi-feature approach is a decision-level fusion approach and our fusion scheme could be traced to the sum rule [12].

Fig. 2. Several example facial images of one subject with different age values in the FG-NET database.

Fig. 3. Several example facial images with different age values in the MORPH database.

IV. Experiments

A. Data Sets

We have evaluated our proposed MFOR algorithm by conducting a number of age estimation experiments on two popular databases: FG-NET [1] and MORPH Album 2 [11]. There are 1002 color or grayscale facial images of 82 identities, covering large ranges in pose, expression and lighting in FG-NET. And the subjects' age values range from 0 to 69. In terms of MORPH Album 2, it is a large-scale database containing 55,608 facial images with two to four images per person from 16 to 77 years old. To reduce the influence of group variation, we select 3952 images of males of Caucasian descent. Before performing feature extraction, all the input images of both datasets have been converted to grayscale, aligned at the eye positions as well as normalized to the same size. Histogram equalization was undertaken to reduce the impact of illumination. Fig. 2 and 3 show some examples together with their age labels drawn from these two databases respectively.

B. Experimental Settings

For FG-NET, we extracted two feature sets from raw facial images, namely AAM [5] features and bio-inspired features (BIF) [10]. AAM was selected because it could extract both shape and appearance features from raw images, and a number of methods also used AAM for feature extraction. BIF was selected for feature extraction for its reported high age estimation accuracy on FG-NET and it could provide complementary information for AAM. The dimension of features of AAM was set to preserve 95 percent of the variability. For BIF features, the number of bands was set to be 8 (thus 16 scales in total) with 4 orientations each. Its final dimension was set to be 100 after PCA dimension reduction. In terms of the training stage, radial basis function (RBF) kernel was selected and all its correspondent parameters were sought via five-fold cross validation. We compared our results with AGES [8], WAS [14], RUN1 [23], RUN2 [24], OHRank [4] and Multi-Task Warped Gaussian Process (MTWGP) [27] by using leave-one-person-out (LOPO), a popular testing procedure deployed by most existing estimation works on FG-NET database. We also investigated age-inferring power of shape model and texture model on FG-NET with respect to threshold age's variation. We split AAM features into two feature sets corresponding to the shape model and the texture model respectively, each with a dimension of 50. By inspecting their weights' variation, we could compare their age-inferring power for various threshold ages.

FG-NET is a relatively small database, to further validate the efficacy of our algorithm, we conducted experiments on MORPH Album 2. For the feature representation stage, there were three feature sets extracted from MORPH Album 2 database. The first feature set was obtained by PCA dimension reduction from raw images to a dimension of 100. The second feature set was extracted by LBP histogram descriptor with dimension set to be 128, and the third one was BIF features, whose configuration was the same as the one on FG-NET database. For the training stage, we randomly split the database into five parts, where four of them were used for training and the remaining one for testing. With this configuration, 30 trials were performed. In the experiment, we first compared our MFOR algorithm with OHRank on various feature sets. For MFOR, we conducted it on four multi-feature sets, which were one tri-feature set (PCA+BIF+LBP), and three bi-feature sets (PCA+BIF, BIF+LBP and PCA+LBP) respectively. In terms of OHRank, it was evaluated on these three mono-feature sets separately. We then compared our MFOR approach with multi-view data fusion approaches. The first one is a feature-level fusion approach, which simply concatenated three mono-feature vectors into an extended feature vector, after which it would be fed into OHRank classifiers for age estimation. We denote this method as Multi-Feature Concatenation Ranking method (MFConR). We also compared our approach with MFMaxR.

C. Results and Analysis

In our experiment, two metrics are used to evaluate the estimation performance. The first one is the Mean Absolute Error (MAE) criterion [8], [14], [4], [27], which is defined

TABLE II

MAEs OF COMPARED AGE ESTIMATION ALGORITHMS ON THE FG-NET DATABASE

Method (Feature sets)	MAE
MFOR (AAM+BIF)	**4.25**
OHRank (AAM)	4.60
OHRank (BIF)	4.92
MTWGP (AAM)	5.05
RUN1 (AAM)	5.78
RUN2 (AAM)	5.33
AGES (AAM)	6.82
WAS (AAM)	7.46

TABLE III

MAEs OF COMPARED AGE ESTIMATION ALGORITHMS ON THE MORPH ALBUM 2 DATABASE

Method (Feature sets)	MAE
MFOR (PCA+LBP+BIF)	**4.20**±0.03
MFOR (PCA+BIF)	4.30±0.07
MFOR (LBP+PCA)	4.37±0.10
MFOR (BIF+LBP)	4.50±0.08
MFConR (PCA+LBP+BIF)	4.46±0.05
MFMaxR (PCA+LBP+BIF)	4.59±0.12
OHRank (PCA)	4.82±0.04
OHRank (BIF)	4.95±0.14
OHRank (LBP)	5.53±0.22

as the average value of the absolute errors between the estimated age labels and the correspondent true age values:

$$MAE = \sum_{j=1}^{N_t} |\hat{y_j} - y_j|/N_t, \tag{13}$$

where N_t is the number of testing instances, $\hat{y_j}$ and y_j are estimated age label and ground truth age value respectively.

The second measure is cumulative score (CS) proposed by Geng *et al.* [8] , defined as

$$CS = N_{e\leq L}/N_t \times 100\%, \tag{14}$$

where $N_{e\leq L}$ is the number of test images with absolute error e less than L.

Table II and Fig. 4 display the MAE results and CS curves derived on the FG-NET database respectively, both of which demonstrate that our MFOR method consistently outperforms all other mono-feature algorithms.

Table III and Fig. 5 further show supreme efficacy of our MFOR method, from which we made two observations:

- Multi-feature approaches produce better results than mono-feature algorithms. MFOR, MFConR and MF-MaxR have lower MAEs and higher CS than OHRank, indicating higher estimation accuracy could be achieved by multi-feature algorithms, which validates our claim that exploiting multiple feature sets could enhance the estimation result.
- Our MFOR approach performs better than MFConR and MFMaxR. Note that MFOR on tri-feature set achieve the highest estimation accuracy and results of MFOR on bi-feature sets are even better than MFConR on tri-feature sets. This interesting result indicates that although MFConR encompassed three feature sets by concatenation, this simple concatenation might hamper estimation performance as it ignores distinct statistical property of each feature. Similarly, although MFMaxR could choose the classifier with best classification capacity for each subproblem, it fails to discover complementary information among different feature sets.

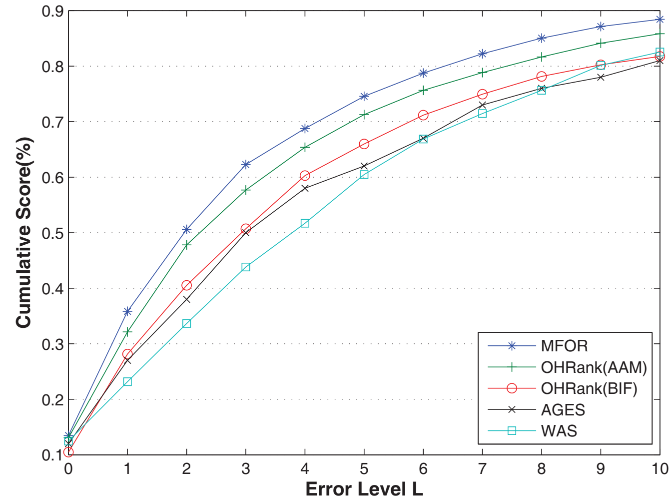

Fig. 4. Comparisons of CS curves of different age estimation algorithms on FG-NET database.

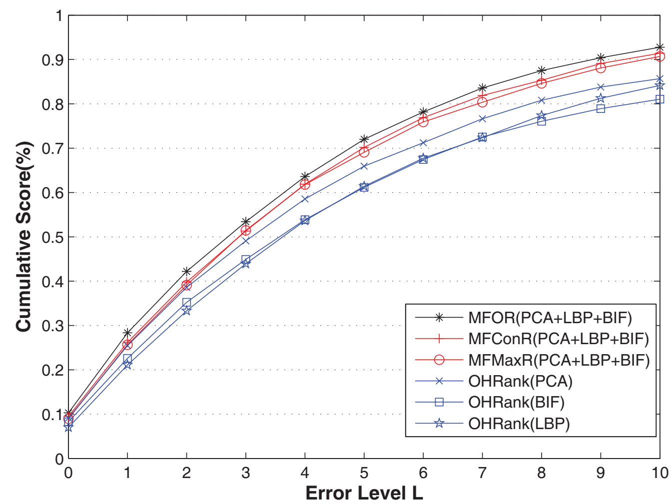

Fig. 5. Comparisons of CS curves of different age estimation algorithms on MORPH Album 2 database.

All these results lend a hand to prove that our MFOR approach could effectively exploit multiple features to improve the final estimation result.

We have also investigated the age-inferring power of shape model and texture model on FG-NET. Fig. 6 shows that when the anchor age is below 10, the shape model has much larger weights than the texture model, and when the anchor age is above 25, texture model has consistently larger weights. This result corresponds to the human aging process. Facial aging pattern appears as skeleton variations during childhood

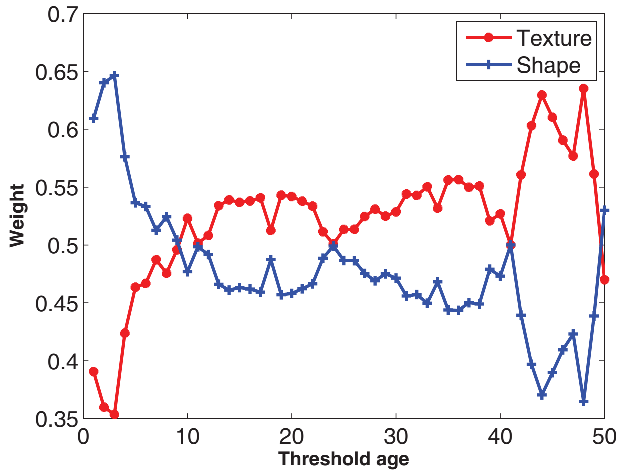

Fig. 6. Weights variation of texture model and shape model on FG-NET with respect to change of threshold age.

and facial texture change during adulthood. Note that we only display the result up to 50 years old, since there are insufficient instances from FG-NET database with age labels above 50 to render a meaningful result.

V. Conclusion and Future Work

In this paper, we have proposed a multi-feature ordinal ranking method for age estimation by learning weighted classifiers on multiple feature sets. Complementary information between different feature sets are explored by assigning weights to their correspondent classifiers through joint learning. Moreover, the age estimation problem is split into a group of $K-1$ subproblems of binary classifications according to the ordinal property of age labels. Our experimental results demonstrate that our proposed MFOR method outperforms state-of-the-art approaches and other multi-view data fusion approaches. In the future, we are interested to explore age-inferring power of more feature sets, and endeavor to extend our work to be a multi-feature framework for the age estimation problem.

Acknowledgement

Jiwen Lu is supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sciences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore.

References

[1] T. F.-N. aging Database.

[2] V. Balasubramanian, J. Ye, and S. Panchanathan. Biased manifold embedding: A framework for person-independent head pose estimation. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–7, june 2007.

[3] D. Cai, X. He, J. Han, and H.-J. Zhang. Orthogonal laplacianfaces for face recognition. *Image Processing, IEEE Transactions on*, 15(11):3608 –3614, nov. 2006.

[4] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 585 –592, june 2011.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 681–685, June 2001.

[6] Y. Fu and T. S. Huang. Human age estimation with regression on discriminative aging manifold. In *Multimedia, IEEE Transactions on*, volume 10, pages 578–584, 2008.

[7] X. Geng, K. Smith-Miles, and Z. Zhou. Facial age estimation by learning from label distributions. In *24th AAAI Conf. on Artificial Intelligence*, pages 585–592, 2010.

[8] X. Geng, Z. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(12):2234–2240, 2007.

[9] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *Image Processing, IEEE Transactions on*, 17(7):1178–1188, 2008.

[10] G. Guo, G. Mu, Y. Fu, and T. Huang. Human age estimation using bio-inspired features. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 112 –119, june 2009.

[11] K. R. Jr and T. Tesafaye. A longitudinal image database of normal adult age-progression. In *IEEE 7th International Conference on Automatic Face and Gesture Recognition*, pages 341–345, April 2006.

[12] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.

[13] Y. H. Kwon and N. D. V. Lobo. Age classification from facial images. In *In Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 762–767, 1999.

[14] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. Systems, Man and Cybernetics*, 34(1):621–628, 2004.

[15] C. Li, Q. Liu, J. Liu, and H. Lu. Learning ordinal discriminative features for age estimation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2570–2577. IEEE, 2012.

[16] J. Lu, J. Hu, X. Zhou, Y. Shang, Y. Tan, and G. Wang. Neighborhood repulsed metric learning for kinship verification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2594–2601. IEEE, 2012.

[17] X. Luo, X. Pang, B. Ma, and F. Liu. Age estimation using multi-label learning. In *Proceedings of the 6th Chinese conference on Biometric recognition*, CCBR'11, pages 221–228, Berlin, Heidelberg, 2011. Springer-Verlag.

[18] B. Ni, Z. Song, and S. Yan. Web image mining towards universal age estimator. In *Proceedings of the 17th ACM international conference on Multimedia*, MM '09, pages 85–94, 2009.

[19] T. Qin, X.-D. Zhang, D.-S. Wang, T.-Y. Liu, W. Lai, and H. Li. Ranking with multiple hyperplanes. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 279–286, 2007.

[20] N. Ramanathan, R. Chellappa, and S. Biswas. Age progression in human faces: A survey. In *Visual Languages and Computing*, volume 15, pages 3349 – 3361, 2009.

[21] T. Xia, D. Tao, T. Mei, and Y. Zhang. Multiview spectral embedding. In *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, volume 40, pages 1438–1446. IEEE, 2010.

[22] B. Xiao, X. Yang, H. Zha, Y. Xu, and T. S. Huang. Metric learning for regression problems and human age estimation. In *Proceedings of the 10th Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, PCM '09, pages 88–99, 2009.

[23] S. Yan, H. Wang, T. Huang, Q. Yang, and X. Tang. Ranking with uncertain labels. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 96–99. IEEE, 2007.

[24] S. Yan, H. Wang, X. Tang, and T. Huang. Learning auto-structured regressor from uncertain nonnegative labels. In *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007*, pages 1 –8, oct. 2007.

[25] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. Huang. Regression from patch-kernel. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.*, pages 1–8, june 2008.

[26] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *Advances in Biometrics*, pages 464–473. Springer, 2007.

[27] D.-Y. Zhang, Yu Yeung. Multi-task warped gaussian process for personalized age estimation. In *Proceedings 2010 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2622–2629, 2010.