2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Cross-Heterogeneous-Database Age Estimation with Co-Representation Among Them

Qing Tian*†‡, Songcan Chen*
*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R.China
†Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, P.R.China
‡School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, P.R.China
Email: {tianqing, s.chen}@nuaa.edu.cn

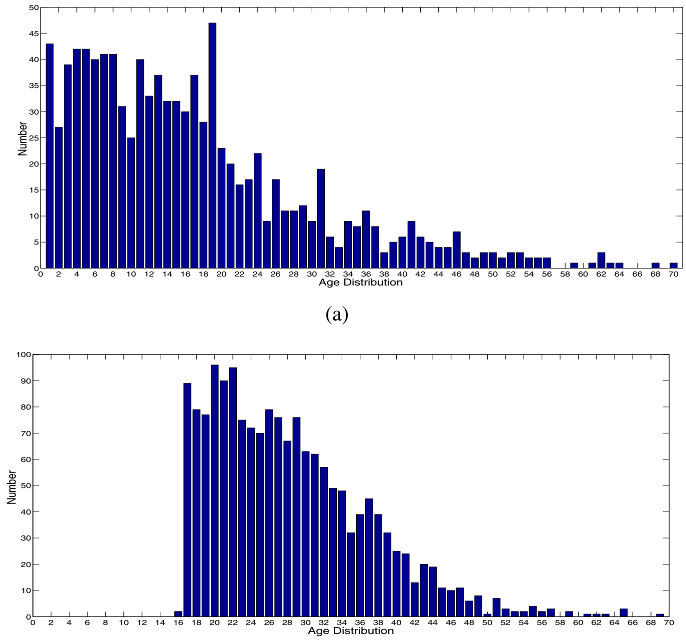
***Abstract*—Human age estimation is an important research topic and can find its applications in such as commodity recommendation and security monitoring. The establishment of existing estimators basically follows a same pipeline, i.e., an estimator is built from a given training dataset like FG-NET and then evaluated on a holdout testing set to determine its effectiveness. In doing so, a usually-followed assumption is that both training and testing sets should share the same age distribution and the same feature representation, implying that 1) once the true age of a human image to be tested is out of the age range of the training set, a mis-estimation is naturally inevitable; 2) estimators built on datasets with different feature representations cannot be directly applied to make predictions on testing sets of each other unless re-trained, because their features are usually different (i.e., the databases are heterogeneous). To the best of our knowledge, the age distributions of different aging databases are usually not consistent and complementary to each other. Motivated by this fact, in order to incorporate such a complementarity in age distributions to improve the generalization ability of the age estimator, in this paper we propose a unified cross-heterogeneous-database age estimation method by first projecting the training samples, usually represented with different features, of different aging databases into a common feature space, and then constructing an age estimator in their mixed sample space. By this way, the age-distribution-incompleteness of the aging datasets can be alleviated by co-representation among them and thus the discriminating ability of the age estimator can be reinforced. Finally, experimental results demonstrate the superiority of the proposed method.**

I. Introduction

Human age estimation is an important research topic and has attracted increasing attention in recent years due to its wide applications in recommendation systems [1], [2], security access control [3], [4], biometrics [5], [6] and entertainment [7], [8], etc.

To perform age estimation according to human facial appearance, a variety of methods have been proposed. Generally, these methods can be grouped into three categories: classification-based (e.g., [9], [10], [11], [12], [13], [14], [15]), regression-based (e.g., [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]), and their hybrid methods (e.g., [3], [28]). When treating each age as a separate class, we can perform age estimation using the existing classification frameworks. According to this principle, artificial neural networks (ANN) [9], conditional probability neural networks (CPNN) [10], Gaussian mixture models [11], and extreme learning machines (ELM) [13] have been successively employed for age classification. More recently, Alnajar et al. [14] proposed an expression-insensitive age estimation method. Dibeklioglu et al. [15] performed age estimation by incorporating facial dynamics together with the appearance information. Actually, age estimation is more of a regression problem rather than ordinary classification due to its characteristics of continuity and monotonicity. Along this line, quadratic function [16], [22], multiple linear regression [17], ξ-SVR [18], SDP regressors [19], [20], aging pattern subspace (AGES) [21], multi-instance regressor [24], and KNN-SVR [27] have been successively proposed for human age regression. Besides the above pure classification or regression based methods, their hybrid methods have also been adopted to train a more powerful estimator. For example, Guo et al. [3] established a so-called locally adjusted robust regression (LARR) to predict human age by combining a series of classifiers and regressors.

After analyzing the above methods, we can find that they basically follow a same pipeline, i.e., an estimator is built from a given training dataset like FG-NET and then evaluated on a testing set to determine its effectiveness. In doing so, a usually-followed assumption is that both training and testing sets should share the same age distribution and the same feature representation, which implies that 1) once the true age of a human image to be tested is out of the age range of the training set, a mis-estimation is naturally inevitable; 2) estimators built on datasets with different feature representations cannot be directly applied to make predictions on testing sets of each other unless re-trained, because their required dimensions of features are usually not matched. As a result, when faced with cross-heterogeneous-database age estimation scenarios, all the aforementioned methods become not applicable because the feature representations and age distributions of the databases are typically not compatible. Without loss of generality, let us take the FG-NET (commonly with AAM feature representation) and Morph (typically with BIF feature representation) databases as examples, two widely used aging datasets for age estimation as shown in Figure 1. The samples of FG-NET mainly distribute at the age range of 0 to 36 years old. By contrast, the age distribution of Morph mostly lies between 16 and 50 years old. It can be intuitively



(a)

(b)

Fig. 1: The age distributions of FG-NET (a) and Morph (b) aging databases. The x-axis and y-axis respectively denote the age distribution and the number of samples at corresponding age.

found that the age distributions of FG-NET and Morph are quite inconsistent, but complementary to some extent. For example, there are no samples at the ages from 0 to 15 years old in Morph itself, but if it is mixed with FG-NET, there will be samples from FG-NET to compensate for the sample space corresponding to 0 to 15 years old of Morph. Therefore, it is preferred to train an age estimator across databases. However, it is a challenging learning problem because the feature representations and age distributions of these databases are not compatible. Although there has been a related work performing cross-database age estimation [29], it requires the databases share the same feature representation.

Remark

At first glance the existing multi-task [30] and/or zero-shot [31] learning strategies seem able to be extended to handle such a cross-database age estimation problem, however it is quite not feasible due to that 1) in multi-task framework, the tasks involved require the same feature representation and output distribution; 2) for zero-shot learning, although it can train an estimator to recognize patterns that have not appeared in training phase, it requires the intermediate semantic attributes to bridge the feature representation and the output label. Unluckily, no semantic attributes are currently provided in the case of cross-database age estimation, implying which cannot be implemented through zero-shot learning.

Although the cross-database age estimation is so challenging and there are no existing methods able to be referred, the complementarity-in-age-distributions of different databases inspires us to present a unified cross-heterogeneous-database age estimation method by first projecting the aging datasets into a common feature space, and then making age prediction in their mixed sample space. By this way, the age-distribution-incompleteness of the aging datasets can be alleviated by co-representation among them and thus the age discriminating ability of the estimator can be reinforced. More importantly, our method just need train a single unified estimator on all the involved aging databases, while all the existing methods must train a respective estimator for each of the databases and the resulting estimators can not make estimations across each other. Finally, we experimentally demonstrate the superiority of the proposed method.

The rest of this paper is organized as follows. In Section II, we briefly review a related work. In Section III, we propose our method. In Section IV, we conduct experiments to evaluate the proposed method. Finally, Section V concludes this paper.

II. Related Work

In [29], Su et al. proposed a cross-database age estimation method based on transfer learning. They combined the discriminant subspace learning and transfer learning to perform transfer age estimation between different databases. Concretely, let W denote the discriminant projection matrix, S_b and S_w denote the between-class and within-class scatter matrices of the source database, respectively. Besides, assume μ_k and Σ_k describe the mean and covariance of the k-th Gaussian component in the target database, and their new mean and covariance in the discriminant projection space are $\mu_k^{'} \triangleq W^T\mu_k$ and $\Sigma_k^{'} \triangleq W^T\Sigma_k W$. The objective function of their cross-databases age estimation is formulated as below

$$W_{opt} = arg\max_{W} \left((F(W) + \lambda H(W)\right), \tag{1}$$

where $F(W) \triangleq \frac{tr(W^TS_bW)}{tr(W^TS_wW)}$ stands for the discriminant subspace learning term on the source database, $H(W) \triangleq \sum_{i=1}^{N}\sum_{k=1}^{K}\varpi\{\mathtt{ln}\pi_k + \mathtt{ln}N(y_i|\mu_k^{'},\Sigma_k^{'})\}$ describes the likelihood on the target database with ϖ being the normalization factor, and λ is the nonnegative tradeoff parameter. It is through the regularization term $H(W)$, they achieved transfer discriminant subspace learning to cater for cross-database age estimation.

Although Su et al. achieved the goal of cross-database age estimation by employing the transfer learning strategy, the prior knowledge of inconsistence-in-age-distribution of different aging databases and the complementarity-in-age-distribution between the datasets has not been taken into account yet. To address such drawbacks, in the next section, we propose a novel co-representation based cross-database age estimation method.

III. The Proposed Method

In the scenario of cross-heterogeneous-database age estimation, two key issues are involved. The first is the feature dimension issue: the feature representations of different databases are usually not compatible, e.g., FG-NET with AAM feature representation and Morph with BIF representation, and consequently their feature dimensions are different. The second issue is how to conduct collaborative learning between the databases. To perform cross-heterogeneous-database age

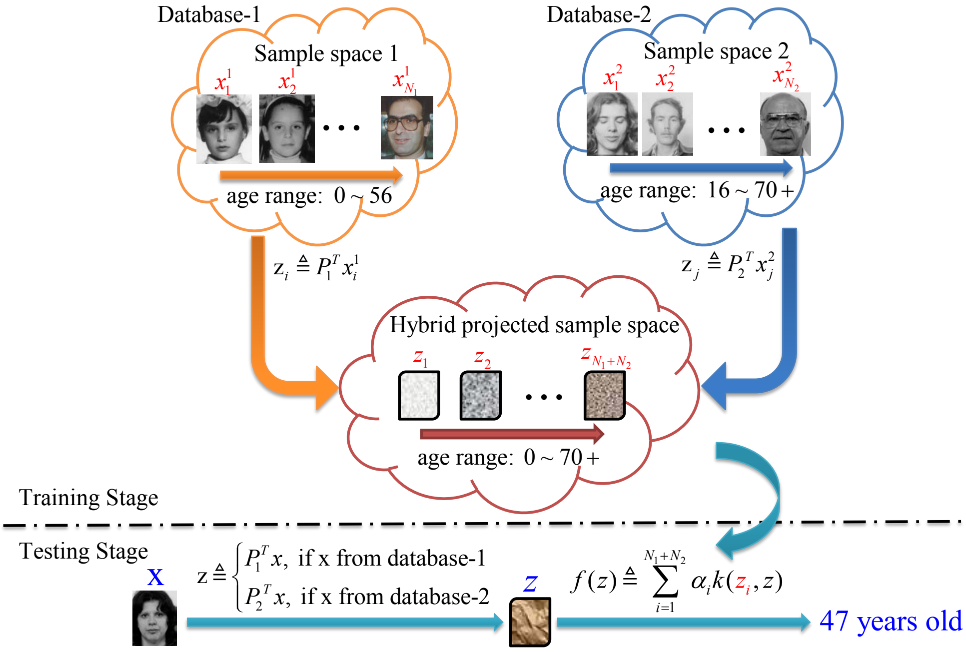


Fig. 2: Illustration of the proposed co-representation based cross-heterogeneous-database age estimation.

estimation while handling the above issues, we propose a novel co-representation based cross-database age estimation method as illustrated in Figure 2. In our method, we first perform dimension reduction for the databases to dimension-identical feature spaces. By this way, the sample spaces of different databases can be projected to a hybrid sample space. More importantly, the age distribution incompleteness of individual databases can be alleviated by complementarity from each other. Then, in the projected hybrid sample space, we can construct a unified age predictor over all the databases involved. Concretely, without loss of generality, we just consider two databases and let $\{x_i^1\}_{i=1}^{N_1} \in \mathbb{R}^{d_1}$ and $\{y_i^1\}_{i=1}^{N_1} \in \mathbb{R}$ respectively denote the N_1 training samples and corresponding age labels from the first aging database, and $\{x_j^2\}_{j=1}^{N_2} \in \mathbb{R}^{d_2}$ and $\{y_j^2\}_{j=1}^{N_2} \in \mathbb{R}$ the N_2 training samples and corresponding age labels from the second aging database[1]. In order to reduce $\{x_i^1\}_{i=1}^{N_1}$ and $\{x_j^2\}_{j=1}^{N_2}$ to a d-dimensional ($d \leq min\{d_1, d_2\}$) common feature space, we seek two projection matrices, denoted as $P_1 \in \mathbb{R}^{d_1 \times d}$ and $P_2 \in \mathbb{R}^{d_2 \times d}$, such that after projection the dimensions of their new representations $z_i \triangleq P_1^T x_i^1 \in \mathbb{R}^d$ and $z_j \triangleq P_2^T x_j^2 \in \mathbb{R}^d$ are identical. In practice, such P_1 and P_2 can be obtained by performing dimension reduction such as Canonical Correlation Analysis (CCA) [32] on the data sets. By this way, we can merge the sample space of database 1 and the one of database 2 into a hybrid common space $\{z_p\}_{p=1}^{N_1+N_2}$. Then, in the hybrid sample space, we can construct an age prediction function $f(z) \triangleq \sum_{p=1}^{N_1+N_2} \alpha_p \cdot k(z_p, z)$ with $k(z_p, z)$ being similarity metric function and representation coefficients α obtained by optimizing the following objective function:

$$\min_{\alpha} \sum_{i=1}^{N_1+N_2} \| \sum_{p=1}^{N_1+N_2} \alpha_p \cdot k(z_p, z_i) - y_i \|_2^2 + \lambda_1 \|\alpha\|_2^2 + \lambda_2 \|\alpha\|_1, \tag{2}$$

[1]Actually, our modelling strategy can be similarly extended to three or more databases.

where $\alpha \triangleq [\alpha_1, \alpha_2, \cdots, \alpha_{N_1+N_2}]^T$, and λ_1 and λ_2 are two nonnegative trade-off parameters. As shown in Eq. (2), the first term penalizes the loss on training data, meanwhile the second term, i.e., L_2 norm, controls the smoothness of α. Unlike the L_2 term, the L_1 term (i.e. the third term) plays a role in inducing the sparseness of α, which essentially encourages each age to be represented by a subset rather than the whole sample sets while automatically dropping the outline samples (i.e., the samples whose age labels are inaccurate) at each age. It is motivated by the fact that an age can be represented by the samples at ages close to it [21], [33].

For the sake of simplifying the optimization, we equivalently rewrite Eq. (2) as

$$\min_{\alpha} \|\mathbf{K}\alpha - y\|_2^2 + \lambda_1 \|\alpha\|_2^2 + \lambda_2 \|\alpha\|_1, \tag{3}$$

in which $\mathbf{K} \in \mathcal{R}^{(N_1+N_2)\times(N_1+N_2)}$ is the kernel matrix defined on training data with $\mathbf{K}_{i,j} \triangleq k(z_i, z_j)$, and $y \triangleq [y_1, y_2, \cdots, y_{N_1+N_2}]^T$. In order to optimize Eq. (3), we first respectively augment the kernel matrix $\mathbf{K}$ and class labels y to $\mathbf{K}^*$ and y^*, defined as below:

$$\mathbf{K}^* \triangleq (1+\lambda_2)^{-\frac{1}{2}} (\mathbf{K}, \sqrt{\lambda_2}\mathbf{I})^T \quad and \quad y^* \triangleq (y^T, \mathbf{0})^T,$$

where $\mathbf{I}$ and $\mathbf{0}$ denote an (N_1+N_2)-order identity matrix and (N_1+N_2)-dimensional zero vector, respectively. Let $\gamma \triangleq \frac{\lambda_1}{\sqrt{1+\lambda_2}}$ and $\alpha^* \triangleq \sqrt{1+\lambda_2}\alpha$. Then, Eq. (3) equivalently becomes

$$\min_{\alpha^*} \|\mathbf{K}^*\alpha^* - y^*\|_2^2 + \gamma \|\alpha^*\|_1. \tag{4}$$

Eq. (4) is a classical LASSO problem [34] and can be solved using coordinate descent [35] with the p-th element of α^*

updated as below

$$\alpha_p^* = sgn(\alpha_p^{ols})(|\alpha_p^{ols}| - \gamma)_+ = \begin{cases} \alpha_p^{ols} - \gamma, & \alpha_p^{ols} > \gamma \\ 0, & -\gamma \leq \alpha_p^{ols} \leq \gamma \\ \alpha_p^{ols} + \gamma, & \alpha_p^{ols} < \gamma \end{cases} \tag{5}$$

where $sgn(\cdot)$ is a sign function, α^{ols} stands for the analytical solution of $\min_\alpha \|\mathbf{K}^*\alpha^* - y^*\|_2^2$, and $(|\alpha_p^{ols}| - \gamma)_+$ equals to $|\alpha_p^{ols}| - \gamma$ if $|\alpha_p^{ols}| - \gamma > 0$, and 0 otherwise. Based on Eq. (5), we conduct cyclic iterations between all elements of α^* until they converge. With the obtained optimal α^*, we can recover the optimal value of α through $\alpha \triangleq \frac{1}{\sqrt{1+\lambda_2}}\alpha^*$. The complete optimization algorithm for Eq. (3) is summarized in Algorithm 1.

Algorithm 1 The optimization algorithm of Eq. (3).

Input:
Training data: $\{x_i^1, y_i^1\}_{i=1}^{N_1}$, $\{x_j^2, y_j^2\}_{j=1}^{N_2}$.
Parameters: λ_1, λ_2.
Output:
The optimal representation coefficients α of Eq. (3).

Training Stage

Project the cross-database instances $\{x_i^1\}_{i=1}^{N_1}$ and $\{x_j^2\}_{j=1}^{N_2}$, through transformation $z_i \triangleq P_1^T x_i^1$ and $z_j \triangleq P_2^T x_j^2$, into common space $\{z_p\}_{p=1}^{N_1+N_2}$;
Centralize the $\{z_p\}_{p=1}^{N_1+N_2}$ and then construct the kernel matrix $\mathbf{K}$ on them;
Augment the kernel matrix $\mathbf{K}$ and corresponding labels y to $\mathbf{K}^*$ and y^*, and then convert Eq. (3) to Eq. (4);
do
  for $p = 1, 2, \cdots, N_1{+}N_2$ **do**
    Update α_p^* based on Eq. (5);
  end for
until α^* converges
Compute $\alpha = \frac{1}{\sqrt{1+\lambda_2}}\alpha^*$.

IV. Experiments

To evaluate the proposed method, in this section we conduct experiments on two aging databases, i.e., FG-NET and Morph album I. For FG-NET dataset, it consists of 1002 facial images captured from 82 persons aged from 0 to 36 years old. For Morph Album I, there are about 1690 facial images from 631 persons aged from 16 to about 77 years old. Their age distributions and image examples are shown in Figure 1 and Figure 3, respectively.

A. Experimental Setup

In the experiments, we extract 200-dimensional AAM parameters from FG-NET as its original feature representations, and draw BIF features from Morph database. The $k(z_i, z_j)$ in Eq. (2) is uniformly taken the Gaussian function $\exp^{(-\|z_i - z_j\|^2/\delta)}$ with δ being the kernel bandwidth. Without loss of generality, we adopt the CCA algorithm to reduce the AAM and BIF features of FG-NET and Morph to a lower dimension-identical feature space where the optimal dimension, as well as all the other hyper-parameters involved

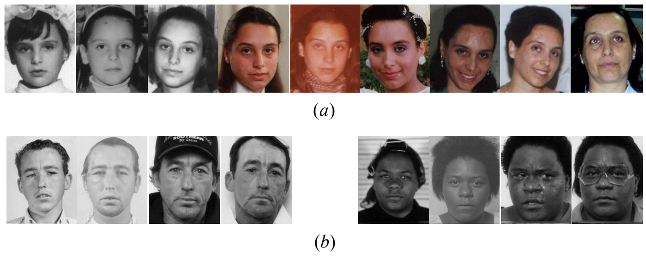


Fig. 3: Image examples of FG-NET (a) and Morph (b).

in the experiments, is determined by *cross-validation*. Besides, we uniformly adopt the Mean Absolute Errors (MAE) as the performance measure, where $MAE := \frac{1}{N}\sum_{i=1}^{N}|\widehat{l_i} - l_i|$ with l_i and $\widehat{l_i}$ denoting the true and predicted age values, respectively. All the reported results are averaged over 10 runs, each with the same experimental setup. Before reporting the results, we first introduce the methods compared in the experiments as below:

- **TLDA**: the transfer learning based method of [29][2].
- **on individual DB**: the method of Eq. (3) learned on a single database (DB).
- **cross DBs with L_2 norm (ours)**: the method of Eq. (3) learned across aging databases penalized with only the L_2 norm on the representation coefficients α (i.e., removing the L_1 norm from the objective function).
- **cross DBs with L_2 and L_1 norms (ours)**: the method of Eq. (3) learned across aging databases penalized with L_2 and L_1 norms on the representation coefficients α.

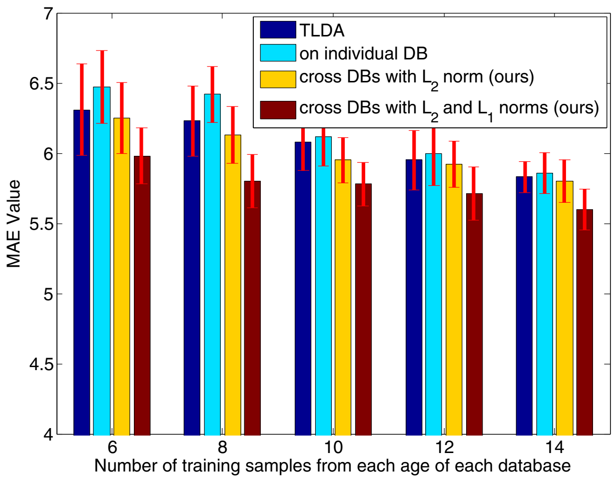
B. Experimental Results

On the FG-NET and Morph aging databases, we conduct comparative age estimation experiments. The results on the two databases are plotted in Figure 4. From it, we can see that

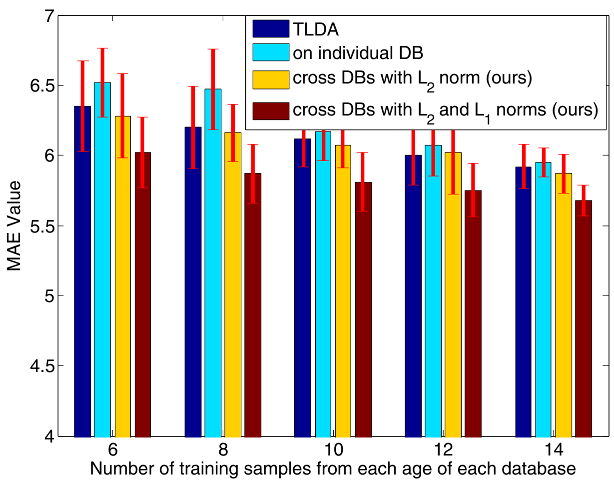
- Compared with the method learning just on individual database, the transfer learning or cross-database based methods can yield better age estimations with lower MAEs. It shows that making age estimation by using the age classes from two age-distribution-inconsistent databases can improve the generalization ability of an age estimator than using the ones just from a single database.
- The MAEs yielded cross DBs with L_2 norm penalization[3] are consistently lower than those by TLDA. It demonstrates the superiority of the proposed methods over the transfer learning based method.

[2] Since TLDA requires the source and target databases share a same feature representation (implying with identical feature dimension), we project the involved aging databases to dimension-identical feature space and then perform cross-database age estimation using TLDA.

[3] Actually, the experimental setting is adverse to our method because compared to the other methods, the superiority of our method in discriminating the age classes which have not appeared in the training phase can not be fully demonstrated.



(a)



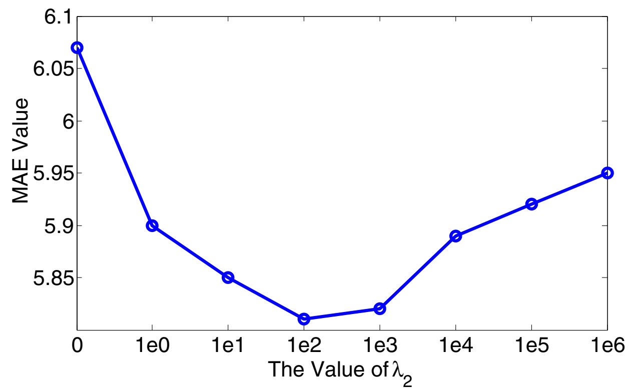
(b)

Fig. 4: Comparison of individual or cross-database age estimation tested on FG-NET (a) and Morph (b).

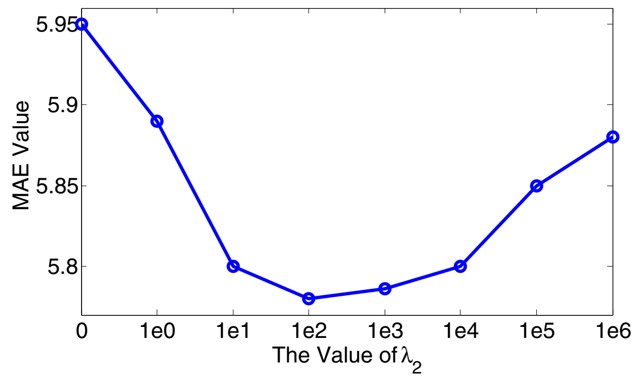
- The MAEs yielded cross DBs penalized with both L_2 and L_1 norms are significantly lower than those just with L_2 norm. It witnesses that a sparse solution of the samples representation coefficients α (see Eq. (3)) help to improve the discriminating ability of the age estimator. It is consistent with the prior knowledge that an age can be well represented by its neighboring ages rather than the whole age range, implying the soundness of imposing L_1 norm on the α.

Besides, in order to figure out the effect of the sparsity of α in Eq. (3), we randomly select 10 samples from each age of the databases[4] and conduct experiments plot the age estimation performance with respect to varying λ_2 in Figure 5. It demonstrates that with increasing value of λ_2, age estimation MAEs of the proposed method (see Eq. (3)) decrease but then tend to increase, which shows although a sparse representation is preferred, an age pattern can not be sufficiently expressed by too few samples.

[4]If the number of samples at some ages is less than 10, all of their samples will be selected for training.



(a)



(b)

Fig. 5: Age estimation performance with respect to varying λ_2 tested on FG-NET (a) and Morph (b).

V. Conclusion

In this paper, we proposed a unified co-representation based cross-heterogeneous-database age estimation method by first projecting different aging datasets into a dimension-identical feature space, and then making age prediction in their sample space with co-representation learning. Finally, we experimentally demonstrated the superiority of the proposed method. It should be noted that although we did not specifically consider the ordinal relationship of the age range [33], [36] for fair comparison with the work of [29], it is trivial to incorporate such a relationship in our method. For better age estimation, label distribution learning [10] as well as database-decentralized age estimation will also be considered in our future work.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China under Grant 61472186, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20133218110032, and the Fundamental Research Funds for the Central Universities and Jiangsu Qing-Lan Project.

References

[1] G. S. Linoff and M. J. Berry, *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2011.

[2] G. Raab, M. G. J. Goddard, R. A. Ajami, and V. B. Gargeya, *Customer relationship management: A global perspective*. Gower Publishing, Ltd., 2012.

[3] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1178–1188, 2008.

[4] Y. Wu, Z. Wei, H. DENG *et al.*, "Attribute-based access to scalable media in cloud-assisted content sharing," *Multimedia, IEEE Transactions on*, vol. 15, no. 4, pp. 778–788, 2013.

[5] P. Van Leeuwen, S. Lange, A. Klein, D. Geue, and D. H. Grönemeyer, "Dependency of magnetocardiographically determined fetal cardiac time intervals on gestational age, gender and postnatal biometrics in healthy pregnancies," *BMC pregnancy and childbirth*, vol. 4, no. 1, p. 6, 2004.

[6] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King, "Aspects of age variation in facial morphology affecting biometrics," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*. IEEE, 2007, pp. 1–6.

[7] Y.-Y. Chen, A.-J. Cheng, and W. H. Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos," *Multimedia, IEEE Transactions on*, vol. 15, no. 6, pp. 1283–1295, 2013.

[8] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Recognition of genuine smiles," *Multimedia, IEEE Transactions on*, vol. 17, no. 3, pp. 279–294, 2015.

[9] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 1, pp. 621–628, 2004.

[10] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 10, pp. 2401–2412, 2013.

[11] K. Ueki, T. Hayashida, and T. Kobayashi, "Subspace-based age-group classification using facial images under various lighting conditions," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. IEEE, 2006, pp. 6–pp.

[12] F. Alnajar, C. Shan, T. Gevers, and J.-M. Geusebroek, "Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions," *Image and Vision Computing*, vol. 30, no. 12, pp. 946–953, 2012.

[13] 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.

[14] F. Alnajar, Z. Lou, J. Alvarez, and T. Gevers, "Expression-invariant age estimation," in *Proceedings of the British Machine Vision Conference. BMVA Press*, 2014.

[15] H. Dibeklioglu, F. Alnajar, A. Ali Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *Image Processing, IEEE Transactions on*, vol. 24, no. 6, pp. 1928–1943, 2015.

[16] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 442–455, 2002.

[17] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 1383–1386.

[18] K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–5.

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

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

[21] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 12, pp. 2234–2240, 2007.

[22] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *Multimedia, IEEE Transactions on*, vol. 10, no. 4, pp. 578–584, 2008.

[23] 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*. IEEE, 2011, pp. 585–592.

[24] B. Ni, Z. Song, and S. Yan, "Web image and video mining towards universal and robust age estimator," *Multimedia, IEEE Transactions on*, vol. 13, no. 6, pp. 1217–1229, 2011.

[25] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning distance metric regression for facial age estimation," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 2327–2330.

[26] C. Li, Q. Liu, and J. Liu, "Learning ordinal discriminative features for age estimation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2570–2577.

[27] W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognition*, vol. 46, no. 3, pp. 628–641, 2013.

[28] S. Kohli, S. Prakash, and P. Gupta, "Hierarchical age estimation with dissimilarity-based classification," *Neurocomputing*, vol. 120, pp. 164–176, 2013.

[29] Y. Su, Y. Fu, Q. Tian, and X. Gao, "Cross-database age estimation based on transfer learning," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1270–1273.

[30] A. Evgeniou and M. Pontil, "Multi-task feature learning," *Advances in neural information processing systems*, vol. 19, p. 41, 2007.

[31] M. Palatucci, D. Pomerleau, G. E. Hinton, and T. M. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in neural information processing systems*, 2009, pp. 1410–1418.

[32] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[33] Q. Tian, H. Xue, and L. Qiao, "Human age estimation by considering both the ordinality and similarity of ages," *Neural Processing Letters*, vol. 43, no. 2, pp. 505–521, 2016.

[34] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[35] T. T. Wu and K. Lange, "Coordinate descent algorithms for lasso penalized regression," *The Annals of Applied Statistics*, pp. 224–244, 2008.

[36] K. Chen, S. Gong, T. Xiang, and C. Loy, "Cumulative attribute space for age and crowd density estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2467–2474.