

# FACIAL AGE ESTIMATION BASED ON LABEL-SENSITIVE LEARNING AND AGE-SPECIFIC LOCAL REGRESSION

Wei-Lun Chao, Jun-Zuo Liu, and Jian-Jiun Ding

Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan  
weilunchao760414@gmail.com, jinzuozuo2011@gmail.com, and djjj@ee.cc.ntu.edu.tw

## ABSTRACT

In this paper, a new age estimation framework considering the intrinsic properties of human ages is proposed, which improves the dimensionality reduction techniques to learn the connections between facial features and aging labels. To enhance the performance of dimensionality reduction, a distance metric adjustment step is introduced in advance to achieve a suitable metric in the feature space. In addition, to further exploit the ordinal relationship of human ages, the “label-sensitive” concept is proposed, which regards the label similarity during the learning phase of distance metric and dimensionality reduction. Finally, an age-specific local regression algorithm is proposed to capture the complicated aging process for age determination. From the simulation results, the proposed framework achieves the lowest mean absolute error against the existing methods.

*Index Terms*—Machine learning, Distance learning, Pattern recognition

## 1. INTRODUCTION

Facial age estimation has attracted increasing attention in computer vision and pattern recognition because of its potential usages. An automatic age estimation system can not only facilitate the human-computer interface, but also prevent under ages from accessing pornographic websites, cigarettes, and bears. In addition, the age attribute has also been applied in face verification and retrieval [8].

Estimating human ages is intrinsically a challenging task because of its multi-class nature, where an aging label can be seen as an individual class. This nature makes age estimation easily suffer from over-fitting when the size of database is insufficient. Furthermore, due to the diversity of personal aging processes, it is very difficult to design and determine the type of facial features that can directly represents human ages. To solve these problems, several previous work has been published in the past decade, and the algorithms are generally composed of two parts: Feature extraction [4][6][9][11] and age determination [2][5][10][13].

Besides these two main challenges, three important factors of age estimation should also be considered. At first, there exist the ordinal relationship and correlations among aging labels. For example, age 30 is closer to age 25 than age 10. This relationship makes age estimation more difficult than the traditional multi-class classification problems. Secondly, aging process is rather complicated, which may not be captured by a single classifier and regressor [5]. Finally, inside many aging databases, we found that

the number of images of each age label is highly different, which may result in serious unbalanced learning.

In this paper, a new age estimation framework is proposed, which takes all the above factors and challenges into consideration:

- To avoid over-fitting and explore the connections between facial features and aging labels, locality preserving projection (LPP) [7] is exploited to drastically reduce the dimensionality of features and preserve the most important information for age estimation.
- To better exploit the ordinal relationship, the “label-sensitive” concept is proposed, which regards the label similarity during the learning phase of LPP.
- To capture the complicated aging process, an age-specific local regression algorithm named KNN-SVR is proposed.
- To alleviate the unbalanced problem, several treatments are proposed for each step in the proposed framework.

In addition, to further enhance the performance of LPP, a distance adjustment step is introduced in advance to achieve a suitable metric for neighbor searching, which is an essential step of LPP.

This paper is organized as follows: In Section 2, an overview of the proposed framework is presented, and the concept of “label-sensitive” is introduced in Section 3. In Sections 4 and 5, the algorithms of distance metric adjustment and dimensionality reduction are described, and the proposed age-specific local regression is discussed in Section 6. Finally, the simulation results and conclusion are presented in Sections 7 and 8.

## 2. FRAMEWORK OF THE PROPOSED ALGORITHM

Given a training set  $I = \{i^{(n)}\}_{n=1}^N$  with  $N$  facial images and its corresponding label set  $Y = \{y^{(n)} \in L\}_{n=1}^N$  with  $L = \{l_1, \dots, l_c\}$ , age estimation can be modeled as a supervised learning task. The symbol  $c$  is the total number of aging labels we concerned.

In this paper, a new age estimation framework is proposed, which consists of four steps: **Feature extraction, distance metric adjustment, dimensionality reduction, and age determination**. Suggested by previous work, the active appearance model (AAM) [3] is adopted for feature extraction, which jointly considers the appearance and shape information from human faces and results in a feature vector  $\mathbf{x} \in R^d$  for each image  $i$ . Then, a distance metric adjustment step is introduced to learn a suitable metric in this feature space, which can enhance the performance of the following dimensionality reduction step. The resulting features after these two steps are denoted as  $\mathbf{x}_{adjust} \in R^d$  and  $\mathbf{z} \in R^p$ . Finally, according

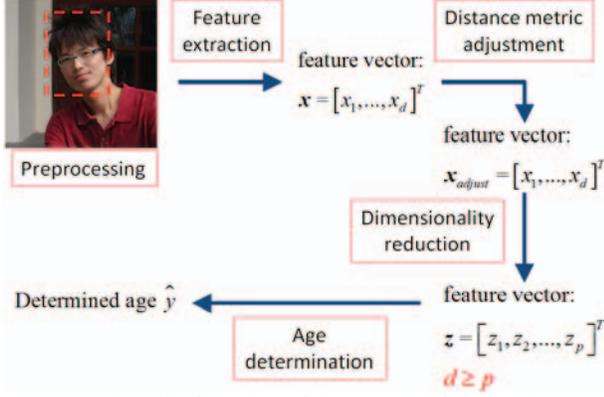


Fig. 1 The flowchart of the proposed framework

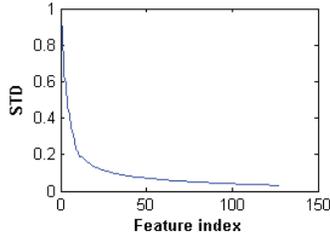


Fig. 2 The standard deviation of each AAM feature.

to  $z \in R^p$ , an age determination function is trained to estimate the aging label  $\hat{y}$ . The flowchart of this framework is plotted in Fig. 1.

### 3. THE LABEL-SENSITIVE CONCEPT

Before describing in detailed about the proposed framework, we first introduce the “label-sensitive” concept. In the traditional multi-class classification, a class is treated independently of other classes, and a uniform penalty is given when a sample is misclassified into any other classes. While in the task of age estimation, there intrinsically exists the ordinal relationship among human ages, and different penalties should be given for different misclassified cases. To better exploit this ordinal relationship in our work, the “label-sensitive” concept is proposed.

During the learning phase of distance metric and dimensionality reduction, several statistical measures are required to compute for each class. Instead of treating each class individually, the “label-sensitive” concept claims that samples with similar class labels should also be considered in this process, and the weights of these samples are assigned based on the label similarity. For example, when computing the scatter matrix for age 30, samples with ages around 30 are also regarded. In the following two sections, we will show how to embed this concept into distance metric learning and dimensionality reduction.

### 4. DISTANCE METRIC ADJUSTMENT

AAM [3] extracts the shape and appearance information from human faces, which may not directly correspond to aging labels. Besides, the dimensionality of AAM features is usually too high to train a robust age classifier or regressor. To overcome these problems, the popular locality preserving projection (LPP) [7] is exploited to learn the connections between features and labels and drastically reduce the feature dimensionality. LPP is a manifold

Table 1: The algorithm of RCA

#### Presetting:

- Training set:  $X = \{\mathbf{x}^{(n)} \in R^d\}_{n=1}^N$ ,  $Y = \{y^{(n)} \in L\}_{n=1}^N$
- Define  $X_i$  as the feature set containing all feature samples with label  $l_i$ . The number of samples in  $X_i$  is denoted as  $N_i$ .
- RCA finds  $W_{RCA} \in R^{d \times d}$ , then  $\mathbf{x}_{adjust} = W_{RCA}^T \mathbf{x} \in R^d$ .

#### Algorithm:

- For each  $l_i$ , compute the mean  $\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{\mathbf{x}^{(n)} \in X_i} \mathbf{x}^{(n)}$  and the intra-class scatter  $S_i = \frac{1}{N_i} \sum_{\mathbf{x}^{(n)} \in X_i} (\mathbf{x}^{(n)} - \boldsymbol{\mu}_i)(\mathbf{x}^{(n)} - \boldsymbol{\mu}_i)^T$ .
- Compute the total scatter matrix  $S = \frac{1}{N} \sum_{i=1}^c N_i \times S_i$ .
- Perform eigendecomposition  $S = V \Lambda V^T$ , and  $W_{RCA} = V \Lambda^{-1/2}$ .

Table 2: The proposed lsRCA algorithm

- Define the sample-class weight ( $\sigma$  and  $\varepsilon$  are tunable):

$$e_i^{(n)} = \begin{cases} \exp(-(y^{(n)} - l_i)^2 / \sigma), & \text{if } |y^{(n)} - l_i| \leq \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

- Modify the computation of  $\boldsymbol{\mu}_i$ ,  $S_i$ , and  $S$  as:

$$\boldsymbol{\mu}_i = \frac{1}{\omega_i} \sum_{n=1}^N e_i^{(n)} \mathbf{x}^{(n)}, \quad S_i = \frac{1}{\omega_i} \sum_{n=1}^N e_i^{(n)} (\mathbf{x}^{(n)} - \boldsymbol{\mu}_i)(\mathbf{x}^{(n)} - \boldsymbol{\mu}_i)^T, \quad \text{and}$$

$$S = \sum_{i=1}^c \omega_i S_i / \sum_{i=1}^c \omega_i, \quad \text{where } \omega_i = \sum_{n=1}^N e_i^{(n)}$$

- Follow the algorithm of RCA to compute  $W_{RCA} \in R^{d \times d}$

learning algorithm and aims to minimize the average neighbor distance after projection. Generally, manifold learning assumes that the input space is locally Euclidean and utilizes the Euclidean metric for neighbors searching, while this assumption may not be held. To ensure this, we simply check the standard deviation (STD) of each AAM feature. The result is shown in Fig. 2, which shows a strong variation of STD among these features. This observation destroys the assumption and may distort the computation of neighbor similarities as well as degrade the overall performance.

To deal with this issue, general unsupervised treatments such as normalizing or scaling the STD can be used. While in our work, relevant component analysis (RCA) [1] is adopted because of its supervised nature and efficiency. RCA is a supervised distance metric learning algorithm, which aims at whitening each intra-class scatter matrix and conducts a globally Euclidean metric inside each class. Although this metric cannot directly guarantee the locally Euclidean assumption, it does improve the overall performance in the experiments. The algorithm of RCA is summarized in Table 1.

To further consider the ordinal nature, a label-sensitive form of RCA called lsRCA is proposed and summarized in Table 2, where the computation of  $\boldsymbol{\mu}_i$ ,  $S_i$ , and  $S$  are modified. Finally, to balance the influence of each label in lsRCA, an unbalance-compensated version called C-lsRCA is proposed, where the computation of  $S$  is replaced by

$$S = \frac{1}{c} \sum_{i=1}^c S_i. \quad (1)$$

**Table 3: The proposed *lsLPP* algorithm****Presetting:**

- Training set:  $X_{adjust} = \{\mathbf{x}_{adjust}^{(n)} \in R^d\}_{n=1}^N, Y = \{y^{(n)} \in L\}_{n=1}^N$
- Define the similar-label set for each sample  $\mathbf{x}_{adjust}^{(i)}$  :

$$N^{(+)}(i) = \{\mathbf{x}_{adjust}^{(j)} \mid |y^{(j)} - y^{(i)}| \leq \varepsilon, j \neq i\}$$

- Create an  $N \times N$  sample similarity matrix  $B = [b_{ij} = 0]_{1 \leq i, j \leq N}$
- *lsLPP* finds  $W_{LPP} \in R^{d \times p}$ , then  $\mathbf{z} = W_{LPP}^T \mathbf{x}_{adjust} \in R^p$ .

**Algorithm:**

- Find the  $k_1$ -nearest samples of  $\mathbf{x}_{adjust}^{(i)}$  in  $N^{(+)}(i)$  and denote these samples as  $KNN^+(i)$ , where  $k_1$  is adjustable.
- For each sample pair  $\{\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\}$ , if  $\mathbf{x}^{(j)} \in KNN^+(i)$  or  $\mathbf{x}^{(i)} \in KNN^+(j)$ , set:

$$b_{ij} = \exp\left(-\frac{\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2}{t}\right) \times \exp\left(-\frac{(y^{(i)} - y^{(j)})^2}{\sigma}\right).$$

- Compute  $L = D - B$ , where  $D$  is diagonal with  $d_{ii} = \sum_j b_{ij}$ .
- Compute the generalized eigendecomposition:  $(XLX^T)\mathbf{v}^{(i)} = \lambda^{(i)}(XDX^T)\mathbf{v}^{(i)} \rightarrow (XLX^T)V = (XDX^T)VA$ , where  $A$  is arranged in the descending order.
- $W_{LPP} = [\mathbf{v}^{(N-p+1)}, \mathbf{v}^{(N-p+2)}, \dots, \mathbf{v}^{(N)}] = V \begin{bmatrix} O_{p \times (N-p)} & I_{p \times p} \end{bmatrix}^T$ .

**5. DIMENSIONALITY REDUCTION**

The usage of RCA or its modified versions results in a suitable metric, where the Euclidean distance now can be applied for neighbor searching in LPP [7]. Originally, LPP is an unsupervised dimensionality reduction technique, which can be modified into a supervised formulation by searching neighbors with the same class label. To take the ordinal nature into consideration, the label-sensitive concept is applied in LPP and achieves an improved version named *lsLPP*, where the same-label constraint is replaced by a similar-label one. In addition, *lsLPP* defines a new neighbor weighting function, which regards both the feature and label similarity between neighbors. Table 3 summarizes this algorithm, where  $\varepsilon$ ,  $\sigma$ , and  $t$  are tunable for feature and label similarity.

To balance the influence of each label, we modify the neighbor size  $k_1$  and the similarity range  $\varepsilon$  for each sample based on the number of samples of the corresponding label. These two treatments may result in asymmetric  $B$  and  $L$ , which disobeys the definition of graph Laplacian. To overcome this situation,  $B$  is simply replaced by  $(B + B^T) / 2$  before computing  $L$ .

**6. AGE DETERMINATION**

After *lsLPP*, the resulting  $p$ -dimensional vector  $\mathbf{z}$  ( $p$  is usually much smaller than  $d$ ) now can be used to train an age determination function. To capture the complicated aging process, the proposed framework utilizes local regression instead of global regression for age determination. Inspired by the work in [12] and the  $L_1$  loss of support vector regression (SVR) emphasized in [5], an age-specific local regression algorithm named KNN-SVR is proposed and summarized in Table 4. Now given a new image,

**Table 4: The proposed KNN-SVR algorithm****Presetting:**

- Training set:  $Z = \{\mathbf{z}^{(n)} \in R^p\}_{n=1}^N, Y = \{y^{(n)} \in L\}_{n=1}^N$

**Algorithm:** ( $k$  is tunable)

- For a query  $\mathbf{z}$ , find its  $k$ -nearest neighbors  $\{\mathbf{z}_{KNN}^{(i)}, y_{KNN}^{(i)}\}_{n=1}^k$  in  $Z$ .
- Train an RBF-kernel SVR regressor based on  $\{\mathbf{z}_{KNN}^{(i)}, y_{KNN}^{(i)}\}_{n=1}^k$ , and use it to predict the age for  $\mathbf{z}$ .

**Table 5: The definitions of MAE and CS**

Test set	$\mathcal{D}_{test} : I = \{\mathbf{i}^{(n)}\}_{n=1}^{N_t}, Y = \{y^{(n)} \in L\}_{n=1}^{N_t}$
MAE	$MAE = \frac{1}{N_t} \sum_{n=1}^{N_t}  y^{(n)} - \hat{y}^{(n)} $
CS	$CS(j) = \frac{1}{N_t} \sum_{n=1}^{N_t} \mathbb{I} \left[  y^{(n)} - \hat{y}^{(n)}  \leq j \right]$ , where $\begin{cases} \mathbb{I}[\text{true}] = 1 \\ \mathbb{I}[\text{false}] = 0 \end{cases}$

the extracted AAM features are processed by  $W_{RCA}$  and  $W_{LPP}$  in order. Finally, the age is estimated by the proposed KNN-SVR.

**7. SIMULATION RESULTS**

The age estimation experiments are performed on the most widely-used FG-NET aging database [14], which contains 1002 facial images from 82 individuals and provides 68 landmarks on each face. These images are ranging from age 0 to age 69, while more than 700 of them are under age 20, which makes the FG-NET database highly unbalanced.

Suggested by the experimental setup in previous work, the leave-one-person-out (LOPO) testing strategy is adopted, where the estimation algorithm is repeatedly trained on images from 81 people and tested on images of the remaining person. To evaluate the performance, two popular measures, mean absolute error (MAE) and cumulative score (CS), proposed in [4] are computed for each age estimation algorithm. MAE computes the average  $L_1$  loss during testing, which fits the loss function of SVR and that is why we adopt SVR in the proposed local regression algorithm. The formulations of MAE and CS are defined in Table 5.

The experiments in this paper are conducted in two stages. At first, we test different algorithm combinations in the proposed framework to demonstrate the improvements achieved by the usage of RCA and the proposed C-*lsRCA*, *lsLPP*, and KNN-SVR. Then, the combination with the lowest MAE is further compared with existing methods. In our implementation, the 68 landmarks are used for AAM training, and 127 features are extracted to maintain 98% shape and appearance variation. Besides, the tunable parameters of C-*lsRCA*, *lsLPP*, and KNN-SVR are selected through cross validation. The optimal dimensionality  $p$  of *lsLPP* and the parameter  $k$  of KNN-SVR are reached around 10 and 15 (increasing or reducing them will degrade the overall performance).

Tables 6 and 7 list the MAE results of the two experiments, and Fig. 3 depicts the comparison of CS with existing algorithms. From these results, we show that the proposed algorithms and modifications (C-*lsRCA*, *lsLPP*, and KNN-SVR) not only improve the performance of the proposed framework, but also achieve the lowest MAE and outperform the state-of-art algorithms. In addition, the proposed framework can be efficiently trained and tested. It takes only 6 seconds for training the C-*lsRCA* and *lsLPP*

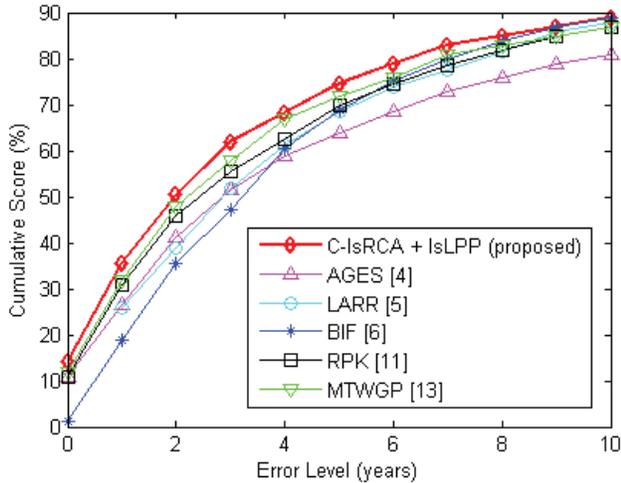


Fig. 3 Cumulative scores of the proposed and existing algorithms

**Table 6: The comparison of different algorithm combinations in the proposed framework (STDN means STD normalization)**

		No change	STDN	RCA	C-IsRCA
SVR	LPP	7.35	5.79	5.43	5.55
	lsLPP	7.28	5.85	5.44	5.34
KNN-SVR	LPP	4.84	5.10	4.67	4.74
	lsLPP	4.81	4.85	4.49	<b>4.38</b>

**Table 7: Comparing the MAE of the proposed algorithm with those of the existing algorithms**

Algorithm	MAE	Algorithm	MAE
AGES [4]	6.77	RPK [11]	4.95
RUN1 [10]	5.78	MTWGP [13]	4.83
RUN2 [10]	5.33	BIF [6]	4.77
RED-SVM [2]	5.24	OHR [2]	4.48
LARR [5]	5.07	<b>Proposed</b>	<b>4.38</b>

matrices using Matlab on a duo-core PC. Although KNN-SVR is an on-line algorithm, it requires only 0.002 second for searching neighbors in the low-dimensional space and training SVR with only the  $k$  nearest neighbors. To further ensure the effectiveness of lsLPP for learning the feature-label connection, we depict the first two lsLPP features of the whole database in Fig. 4. As shown, obvious feature-label dependences have been reached by lsLPP.

## 8. CONCLUSION

A new age estimation framework considering the intrinsic factors of human ages is proposed in this paper. After feature extraction, RCA is utilized to achieve a suitable metric for neighbor searching. Then based on this metric, LPP is trained to reduce the feature dimensionality and learn the connections between features and aging labels. To further consider the ordinal nature of human ages as well as the unbalanced learning problem in RCA and LPP, the “label-sensitive” concept and several unbalance treatments are proposed and results in new algorithms called C-IsRCA and lsLPP. In addition, an age-specific local regression algorithm called KNN-SVR is proposed to capture the complicated human aging process. The simulation results performed on the widely-used FG-NET aging database show that the proposed algorithms and framework achieve the lowest MAE against the state-of-art algorithms.

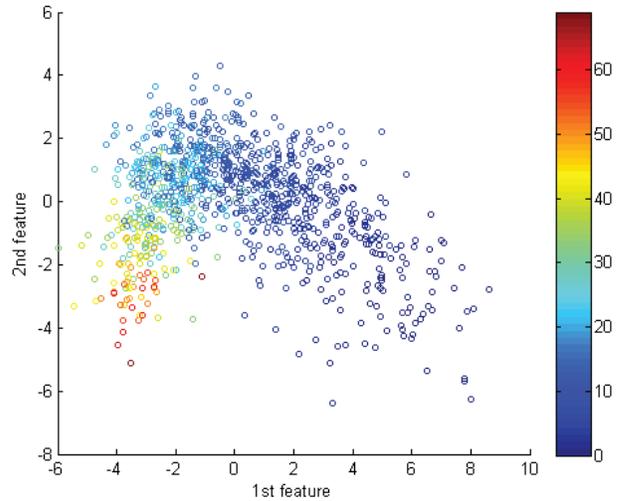


Fig. 4 The dimensionality reduction result after C-IsRCA + lsLPP. The distribution of the 1<sup>st</sup> and the 2<sup>nd</sup> features of all FG-NET images is shown for visualization purpose.

## 9. REFERENCES

- [1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, “Learning Distance Functions Using Equivalence Relations,” *ICML*, 2003.
- [2] K. Y. Chang, C. S. Chen, and Y. P. Hung, “Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation,” *CVPR*, 2011.
- [3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active Appearance Models,” *ECCV*, 1998.
- [4] X. Geng, Z. H. Zhou, and K. Smith-Miles, “Automatic Age Estimation Based on Facial Aging Patterns,” *IEEE Trans. PAMI*, vol. 29, no. 12, pp. 2234-2240, 2007.
- [5] G. Guo, Y. Fu, C. Dyer, and T.S. Huang, “Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression,” *IEEE Trans. Image Processing*, vol. 17, no. 7, pp. 1178-1188, 2008.
- [6] G. Guo, G. Mu, Y. Fu, and T. S. Huang, “Human Age Estimation Using Bio Inspired Features,” *CVPR*, 2009.
- [7] X. He and P. Niyogi, “Locality Preserving Projections,” *NIPS*, vol. 16, 2003.
- [8] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar, “Describable Visual Attributes for Face Verification and Image Search,” *IEEE Trans. PAMI*, vol. 33, no. 10, pp. 1962-1977, 2011.
- [9] A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing Different Classifiers for Automatic Age Estimation,” *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 34, no. 1, pp. 621-628, 2004.
- [10] S. Yan, H. Wang, T. S. Huang, and X. Tang, “Ranking with Uncertain Labels,” *ICME*, 2007.
- [11] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, “Regression from Patch-Kernel,” *CVPR*, 2008.
- [12] H. Zhang, A. Berg, M. Maire, and J. Malik, “SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition,” *CVPR*, 2006.
- [13] Y. Zhang and D. Yeung, “Multi-Task Warped Gaussian Process for Personalized Age Estimation,” *CVPR*, 2010.
- [14] FG-NET Aging Database, <http://www.fgnet.rsunit.com/>.