Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression

Guodong Guo, *Senior Member, IEEE*, Yun Fu, *Student Member, IEEE*, Charles R. Dyer, *Fellow, IEEE*, and Thomas S. Huang, *Life Fellow, IEEE*

Abstract—**Estimating human age automatically via facial image analysis has lots of potential real-world applications, such as human computer interaction and multimedia communication. However, it is still a challenging problem for the existing computer vision systems to automatically and effectively estimate human ages. The aging process is determined by not only the person's gene, but also many external factors, such as health, living style, living location, and weather conditions. Males and females may also age differently. The current age estimation performance is still not good enough for practical use and more effort has to be put into this research direction. In this paper, we introduce the age manifold learning scheme for extracting face aging features and design a locally adjusted robust regressor for learning and prediction of human ages. The novel approach improves the age estimation accuracy significantly over all previous methods. The merit of the proposed approaches for image-based age estimation is shown by extensive experiments on a large internal age database and the public available FG-NET database.**

Index Terms—**Age manifold, human age estimation, locally adjusted robust regression, manifold learning, nonlinear regression, support vector machine (SVM), support vector regression (SVR).**

I. INTRODUCTION

HUMAN faces, as important visual cues, convey a significant amount of nonverbal information to facilitate the real-world human-to-human communication. As a result, the modern intelligent systems are expected to have the capability to accurately recognize and interpret human faces in real time. Facial attributes, such as identity, age, gender, expression, and ethnic origin, play a crucial role in real facial image analysis applications including multimedia communication, human computer interaction (HCI), and security. In such applications, various attributes can be estimated from a captured face image to infer the further system reactions. For example, if the user's age is estimated by a computer, an age specific human computer interaction (ASHCI) system may be developed for secure network/system access control. The ASHCI system ensures young
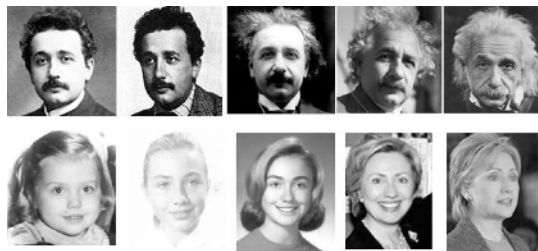
Fig. 1. Face aging of two individuals. Each row shows images of the same individual at different ages.

kids have no access to internet pages with adult materials. A vending machine, secured by the ASHCI system, can refuse to sell alcohol or cigarettes to the underage people [10], [23]. In image and video retrieval, users could retrieve their photographs or videos by specifying a required age range [23]. Ad-agency can find out what kind of scroll advertisements can attract the passengers (potential customers) in what age ranges using a latent computer vision system.

Although automatic image-based age estimation is an important technique involved in many real-world applications, it is still a challenging problem to estimate human ages from face images. Fig. 1 shows some face images of two individuals with different ages. Since different individuals age quite differently, the aging process is determined by not only the person's gene but also many external factors, such as health, living style, living location, and weather conditions. Males and females may also age differently due to the different extent in using makeups and accessories. How to extract general discriminative aging features while reducing the negative influence of individual differences still remains an open problem.

There exits some work on age synthesis and rendering in the last several decades [9], [27], [29], [31], but there are only few publications on age estimation due to the complexity of aging patterns. The age progression displayed on faces is uncontrollable and personalized [10], [26]. Such special characteristics of aging variation cannot be captured accurately due to the prolific and diversified information conveyed by human faces. On the other hand, the age estimation problem is different from the problem of face recognition with age variation [24], where the goal is to estimate facial identities while no ages are estimated from the input faces. The research effort on age estimation may help recognizing faces containing age variations.

There are three main categories [8] that can categorize most existing image-based age estimation methods, such as anthropometric model [21], [27], aging pattern subspace [10], and age

regression [7], [22], [23], [33], [34]. The cranio-facial development theory and facial skin wrinkle analysis are used to create the anthropometric model. The changes of face shape and texture patterns related to growth are measured to categorize a face into several age groups. These methods are suitable for coarse age estimation or modelling ages just for young people [27]. However, they are not designed for continuous or refined age classification [21]. In order to handle incomplete data such as missing ages in the training sequence, the aging pattern subspace (AGES) method [10] models a sequence of individual aging face images by learning a subspace representation. The age of a test face is determined by the projection in the subspace that can best reconstruct the face image. For the regression methods, facial features are extracted by the active appearance models (AAMs) [3] that incorporate the shape and appearance information together. An input face image is then represented by a set of fitted model parameters. The regression coefficients are estimated from the training data with an assumption of the regression function such as a quadratic model (QM) [23]. Yan *et al.* [33], [34] also dealt with the age uncertainty by formulating a semi-definite programming problem [34] or an EM-based algorithm [33].

As Deffenbacher *et al.* [4] argued, the aging factor has essentially sequential patterns which are quite similar to the age morphing [9]. Ages could present a significant trend of underlying sequential patterns when a large number of aging data are provided. This inspires us to explore the refined age estimation technique by age manifold analysis. The age manifold analysis has two advantages to facilitate the age estimation task. First, the manifold analysis is a way to represent the original age data in low dimensionality which is necessary to overcome lack-of-fit of the regression model. Second, the manifold learning captures the underlying face aging structure which is important for accurate modeling and age prediction. To the best of our knowledge, except for our preliminary work in [8], no previous work has investigated manifold learning [6], [28], [30] for age estimation.

On the other hand, the traditional quadratic model [23] for age regression is based on a least square estimation (LSE) criterion. The LSE is not robust to outliers since the outliers could come from some incorrectly labelled ages. Moreover, the optimization based on the LSE criterion minimizes the empirical risk which usually cannot generalize well especially when a large number of training data are not available. The small sample size problem is typical in age estimation because of the difficulty in collecting aging images.

In this paper, we propose a novel scheme for aging feature extraction and automatic age estimation. The basic idea is to learn a low-dimensional embedding of the aging manifold using an appropriate subspace learning method. Then we design a new method, called locally adjusted robust regressor (LARR), for robust learning and prediction of the aging patterns [14]. The effectiveness and advantages of the proposed method will be demonstrated with extensive experiments on a large internal age database and the public available FG-NET database.

II. AGE ESTIMATION FRAMEWORK

As shown in the diagram of Fig. 2, the proposed age estimation framework mainly consists of five modules, which
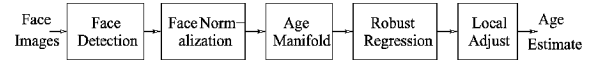

Fig. 2. Age estimation framework based on face image analysis and statistical learning.

are face detection, face normalization, manifold learning, robust regression, and local adjustment. For training, face image patches are automatically detected and cropped from images by face detection. A large number of training images are collected from a broad range of subject ages. The cropped face patches undergo a normalization including geometric alignments and illumination normalization (basically histogram equalization). Then the age manifold is learned to map the original face image data into a low-dimensional subspace. A robust regression function is applied to fit the manifold data. Finally, a local adjustment of the regression results is performed to refine the local fitting of the data. For test, an input face picture goes through the same process of face detection and normalization. Then the normalized face image is projected on the learned manifold which was computed in the learning stage. Finally, the age of the input face image is predicted by the locally adjusted robust regression function. Compared with our previous work in [8], this paper focuses more on the regression part of the whole framework to improve the robustness and effectiveness of age estimation.

We organize the remainder of the paper as follows. In Section III, a simple description of three manifold learning methods is given, and manifold visualizations are illustrated for our age data. In Section IV, the quadratic regression model is introduced which was used in previous approaches. In Section V, we introduce the support vector regression method which was adopted as our robust regressor. A local adjustment of the regression results is presented in Section VI. Experimental evaluations of the proposed approach and comparisons with previous methods are presented in Section VII, and, finally, the conclusion and future work are provided in Section VIII.

III. MANIFOLD LEARNING FOR HUMAN AGING

Age is one of the basic attributes in facial images. Suppose the image space is represented by a set of aligned face images $\mathcal{X} = \left\{ \mathbf{x}_i : \mathbf{x}_i \in \mathbb{R}^D \right\}_{i=1}^n$ with image dimension D in the order of subject ages. A ground truth set $\mathcal{L} = \left\{ l_i : l_i \in \mathbb{N} \right\}_{i=1}^n$ associated with the images provides the age labeling. Our goal is to learn a low-dimensional manifold in the embedded subspace as well as its representation $\mathcal{Y} = \left\{ \mathbf{y}_i : \mathbf{y}_i \in \mathbb{R}^d \right\}_{i=1}^n$ with $d \ll D$, which is a one-to-one mapping to $\mathcal{X}$. So, the projection from image space to manifold space can be modelled as $\mathcal{Y} = P(\mathcal{X}, \mathcal{L})$, where $P(\cdot)$ denotes the projection function, which can be either linear or nonlinear. Considering the label information in the modeling, we adopt supervised learning approaches to find $P(\cdot)$. Recent advances in manifold learning suggest several ways of calculating the manifold embedding.

The objective of manifold embedding is to find an $n \times d$ matrix $\mathbf{P}$ satisfying $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$ or directly find $\mathbf{Y}$, where $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_n \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}$, $\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_d \end{bmatrix}$, and $d < n$. In a supervised manner [6], manifold embedding constrains to search nearest neighbors
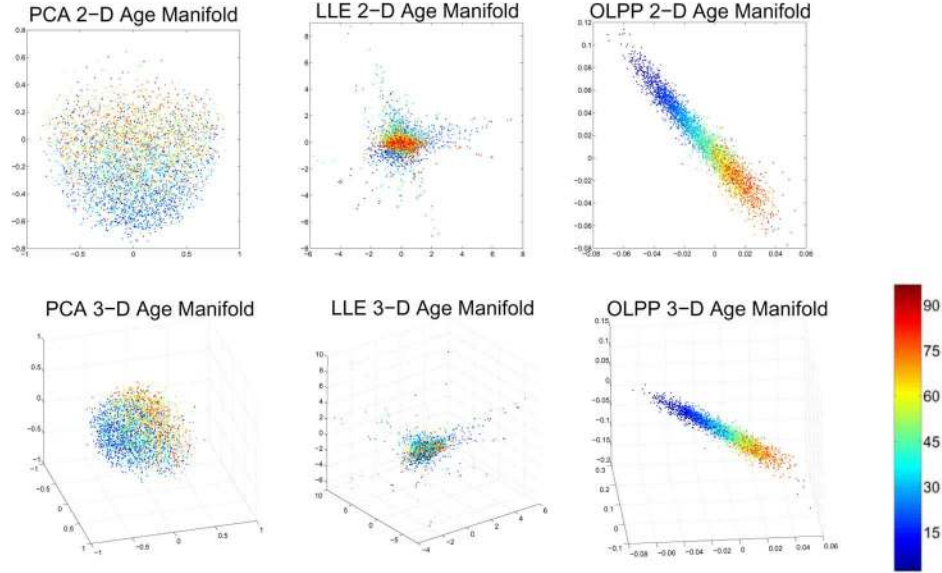
Fig. 3.   Age manifold visualization. The two rows illustrate the 2-D and 3-D manifold of UIUC-IFP-Y age data learned by PCA, LLE, and OLPP algorithms. The datapoints of age from 0 to 93 are colored from blue to red.

in the data space with the same labels of the query. Some typical dimensionality reduction and manifold embedding methods are summarized as follows.

1) *Principal Component Analysis (PCA)* [5]: The PCA method finds the embedding that maximizes the projected variance, $\mathbf{p} = \arg\max_{\|\mathbf{p}\|=1} \mathbf{p}^T \mathbf{S} \mathbf{p}$, where $\mathbf{S} = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$ is the scatter matrix, and $\bar{\mathbf{x}}$ is the mean vector of $\{\mathbf{x}_i\}_{i=1}^{n}$. The PCA method is mentioned here because it is very popular for many tasks such as face recognition [32].

2) *Locally Linear Embedding (LLE)* [28]: The LLE algorithm seeks the nonlinear embedding in a neighborhood-preserving manner by exploiting the local symmetries of linear object class reconstructions, and seeking the optimal weights for local reconstruction.

3) *Orthogonal Locality Preserving Projections (OLPP)* [2]: The OLPP method produces orthogonal basis functions based on the LPP [19] to obtain more discriminating power for embedding. The LPP searches the embedding that preserves essential manifold structure by measuring the local neighborhood distance information. It defines the affinity weight as $s_{ij} = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t\right)$ when $\mathbf{x}_i$ and $\mathbf{x}_j$ are k nearest neighbors of each other, otherwise $s_{ij} = 0$, and $\mathbf{S}(i,j) = s_{ij}$ is a symmetric matrix. It also defines a diagonal matrix $\mathbf{D}(i,i) = \sum_j s_{ji}$, and a Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{S}$, then the optimal projection is

$$\mathbf{p} = \arg \min_{\mathbf{p}^T \mathbf{XDX}^T \mathbf{p}=1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathbf{p}^T \mathbf{x}_i - \mathbf{p}^T \mathbf{x}_j\right)^2 s_{ij}.$$

Define $\mathbf{A}^{(i-1)} = [\mathbf{p}_1 \cdots \mathbf{p}_{i-1}]$, $\mathbf{B}^{(i-1)} = [\mathbf{A}^{(i-1)}]^T (\mathbf{XDX}^T)^{-1} \mathbf{A}^{(i-1)}$ and compute $\mathbf{p}_1$ as the eigenvector of $\mathbf{Q} = (\mathbf{XDX}^T)^{-1} \mathbf{XLX}^T$

associated with the smallest eigenvalue. Then another $\mathbf{p}_i$ is computed as the eigenvector of

$$\mathbf{M} = \mathbf{Q} - (\mathbf{XDX}^T)^{-1} \mathbf{A}^{(i-1)} [\mathbf{B}^{(i-1)}]^{-1} [\mathbf{A}^{(i-1)}]^T \mathbf{Q}$$

associated with another smallest eigenvalue.

The age manifolds of the UIUC-IFP-Y age data learned by the two linear embedding methods, PCA and OLPP, and the nonlinear LLE are visualized in Fig. 3. The datapoints of age from 0 to 93 are colored from blue to red. From the 2-D and 3-D views, the OLPP [2] method characterizes the age manifold better than the PCA and LLE methods with a distinct aging trend. More discussions are provided in Section VII. Furthermore, the OLPP method can give a lower error rate than PCA and a linearization form of the LLE for age estimation in our preliminary experiment [8]. So, we choose to use the OLPP method for our age manifold learning and the OLPP features are extracted for each cropped and normalized face image.

IV. QUADRATIC REGRESSION

Given the extracted features for each face image, a regression function is often used to characterize the relationship between the extracted features, $\mathbf{y}$, and the age labels, L

$$L = f(\mathbf{y}).$$

A typical choice of the regression function, f, is the quadratic model (QM). For example, Lanitis *et al.* proposed to use the QM function for age regression [22], namely

$$\hat{L} = w_0 + \mathbf{w}_1^T \mathbf{y} + \mathbf{w}_2^T \mathbf{y}^2 \tag{1}$$

where $\hat{L}$ is the estimate of the age, w_0 is the offset, $\mathbf{y}$ and $\mathbf{y}^2$ are the extracted feature vector and its square, and $\mathbf{w}_1$ and $\mathbf{w}_2$ are weight vectors.
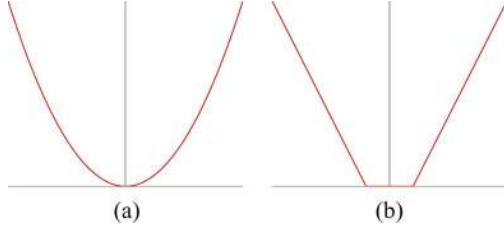
Fig. 4. Regression criteria. (a) Quadratic regression loss function. (b) ϵ-insensitive loss function, which is less sensitive to outliers than the quadratic loss function. Another benefit from this function is a sparse set of support vectors to represent the regression function, i.e., only points outside the ϵ zone contribute to the regression function. The horizontal and vertical axes are $\mathbf{y}$ and $f(\mathbf{y})$, respectively.

The QM method has been used for age regression in previous approaches [22], [23]. The model parameters are optimized by minimizing the difference between the actual ages of the individuals, L, and the ages estimated using (1), i.e., $\|L - \hat{L}\|$. The loss function usually corresponds to a least square estimation (LSE) criteria. However, there are some disadvantages for the QM method: 1) the aging is a complex nonlinear regression problem, especially for a large span of years, e.g., 0–90. The simple quadratic function may not model properly the complex aging process; 2) the least square estimation is sensitive to outliers that come from incorrect labels in collecting a large image database; and 3) the least square estimate criterion only minimizes the empirical risk which may not generalize well for unseen examples, especially with a small number of training examples. This is typical in age estimation because of the difficulty in collecting age images and the diversity of age patterns due to different living conditions, cosmetics, gender differences, and facial shapes. In sum, we need to seek a robust model for modelling the aging patterns.

For the purpose of robust regression of the aging process, we adopt the support vector regression (SVR) method [35]. The SVR might attack the three limitations of the traditional quadratic regression model to facilitate the age estimation task.

V. SUPPORT VECTOR REGRESSION

The basic idea of SVR is to find a function $f(\mathbf{y})$ that has most ϵ deviation from the actually obtained target z_i for the training data $\mathbf{y}_i$, and, at the same time, is as flat as possible. In other words, we do not care errors as long as they are less than ϵ. This property determines the SVR to be less sensitive to outliers than the quadratic loss function. In comparison with the conventional quadratic loss function shown in Fig. 4(a), the ϵ-insensitive loss function of SVR is shown in Fig. 4(b). Given the same input, the ϵ-insensitive loss function is more robust than the quadratic function in dealing with outliers.

A. Linear SVR

Consider the problem of approximating the set of data $\mathcal{D} = \{(\mathbf{y}_1, z_1), \ldots, (\mathbf{y}_n, z_n)\}$, $\mathbf{y}_i \in \mathbb{R}^d$, $z_i \in \mathbb{R}$, with a linear function

$$f(\mathbf{y}) = \langle \mathbf{w}, \mathbf{y} \rangle + b. \tag{2}$$

The optimal regression function [35] is given by

$$\min_{\mathbf{w}, \xi} \quad \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{n} (\xi_i^+ + \xi_i^-)$$

$$\text{subject to} \quad \begin{aligned} z_i - \langle \mathbf{w}, \mathbf{y}_i \rangle - b &\leq \epsilon + \xi_i^+ \\ \langle \mathbf{w}, \mathbf{y}_i \rangle + b - z_i &\leq \epsilon + \xi_i^- \\ \xi_i^+, \xi_i^- &\leq 0 \end{aligned} \tag{3}$$

where constant $C > 0$ determines the tradeoff between the flatness of f and data deviations, and ξ_i^+, ξ_i^- are slack variables to cope with otherwise infeasible constraints on the optimization problem of (3). The ϵ-insensitive loss function as shown in Fig. 4(b) is

$$L_\epsilon(\mathbf{y}, z) = \begin{cases} 0, & \text{if } |f(\mathbf{y}) - z| < \epsilon \\ |f(\mathbf{y}) - z| - \epsilon, & \text{otherwise.} \end{cases} \tag{4}$$

The *primal* problem of (3) can be solved more efficiently in its *dual* formulation [35] resulting in the final solution given by

$$\mathbf{w} = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \mathbf{y}_i \tag{5}$$

and

$$f(\mathbf{y}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \langle \mathbf{y}_i, \mathbf{y} \rangle + b \tag{6}$$

where α_i, α_i^* are Lagrange multipliers. The value of b in (2) can be determined by plugging (5) into (2) [11].

B. Toy Example

To illustrate the SVR idea and see the importance of proper setting of the parameter ϵ, we use a toy example that contains 30 points in 2-D with 10 in a line and the remaining 20 being outliers distributed on both sides of the line [13]. Hence, the data contains 67% outliers. Using the SVR algorithm implemented by Gunn [11] (which provides a user interface) and a linear kernel with $\epsilon = 0.02$, the result is shown in Fig. 5(a). Observe that the line was correctly estimated despite the high percentage of outliers.

On the other hand, observe that SVR returns 27 support vectors (90% of the input data) and seven of them are very close to the boundaries (two dashed lines), but there are actually 20 outliers in the original data. So, we cannot simply classify the support vectors (SVs) as outliers. Increasing the ϵ value might "drag" the seven closest support vectors inside the dashed boundaries, and then only the outliers in the data would be returned as support vectors. However, when we increase ϵ gradually up to 0.09, there are still 26 SVs returned which are still not the true outliers, as shown in Fig. 5(b), and even worse, the slope of the line has changed significantly. This demonstrates that using a large ϵ is not a good idea because it may degrade the model structure.

Based on this experiment, we observe: 1) the SVR technique can potentially deal with data containing a high percentage of outliers; 2) classifying support vectors as outliers is not workable; 3) using a large value for ϵ is not a good idea for SVR; and
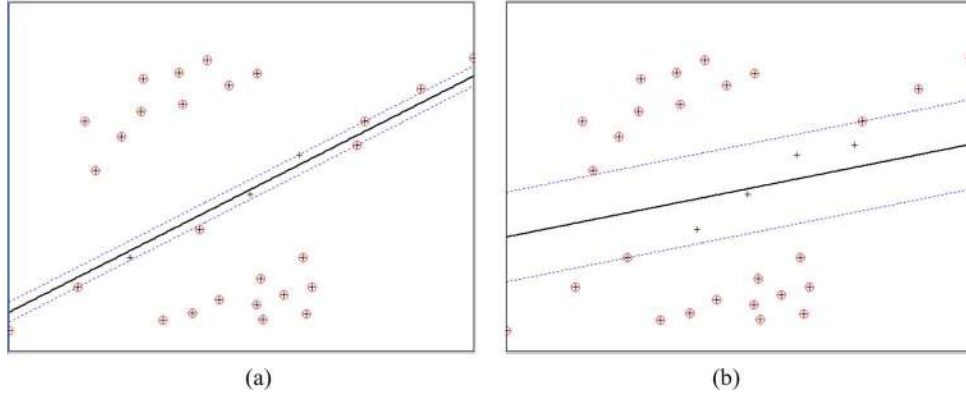
Fig. 5. SVR on real 2-D data with $\epsilon = 0.02$ in (a) and $\epsilon = 0.09$ in (b). Note that the support vectors (marked by circles) are not the true outliers in either case.

4) using small ϵ is preferable, especially when a large number of outliers are present.

This toy example and the above observations were first presented by Guo *et al.* in [13]. The robust regressor, SVR, was applied successfully for outlier detection and removal in affine motion tracking with the setting of a small ϵ. Here we adopt the same idea but use it for another application—robust age regression. Instead of using the simple linear regression, we need a nonlinear SVR for the complex aging patterns.

C. Nonlinear SVR

A nonlinear regression function may be required in practice to adequately model the data. It can be obtained by using kernels, in the same manner as a nonlinear support vector machine (SVM) for classification [35]. A nonlinear mapping can be used to map the data into a high-dimensional feature space where a linear regression is performed. Different kernels, such as polynomials, sigmoid, or Gaussian radial basis functions, can be used depending on the tasks. For our robust age regression, we found that the Gaussian radial basis function kernel performs much better than the linear regression. The reason is that the linear regression cannot model the complex aging process. A radial basis function is of the form

$$k(\mathbf{y}, \mathbf{y}') = e^{-\gamma \|\mathbf{y} - \mathbf{y}'\|^2} \tag{7}$$

where γ is a constant to adjust the width of the Gaussian function. Given the kernel mapping, the solution of the nonlinear SVR is obtained as [35]

$$\langle \mathbf{w}, \mathbf{y} \rangle = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \, k(\mathbf{y}_i, \mathbf{y}) \tag{8}$$

and

$$f(\mathbf{y}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \, k(\mathbf{y}_i, \mathbf{y}) + b. \tag{9}$$

The difference to the linear regression is that $\mathbf{w}$ is no longer given explicitly. Also note that in the nonlinear case, the optimization problem corresponds to finding the *flattest*, or linear
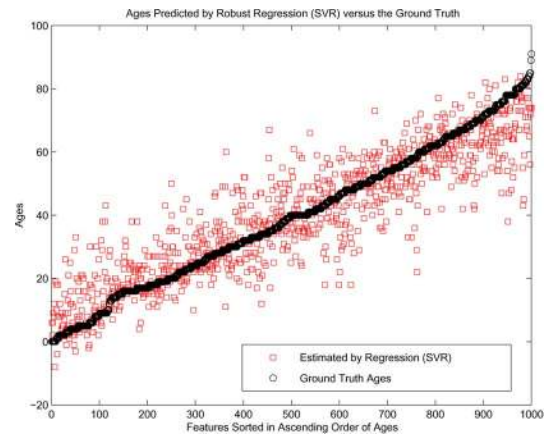

Fig. 6. Plot of the true ages (black circles) versus the estimated ages (red squares) for one thousand female face images. The ages are predicted by the nonlinear SVR with a Gaussian kernel.

regression function in the higher dimensional *feature* space,[1] not in the input space.

VI. LOCALLY ADJUSTED ROBUST REGRESSION

Now, a question may be asked, is it "good" enough using the SVR as a robust regressor for human age prediction? To answer this question, let us look at an estimation result using the SVR. Fig. 6 shows the predicted ages (red squares) with respect to the ground truth ages (black circles). Note that this is not a regression curve. One thousand data points are sorted in ascending order of the ground truth ages, i.e., from 0 to 91 years for females. The predicted ages are obtained from the SVR method. From this figure, we observe that the SVR method can estimate the global age trend, but cannot predict the ages precisely. By inspecting the result carefully, we find that the SVR predictions give bigger age values for many younger people, and smaller age values for some older people. In some cases, the estimated age values could be far away from the true ages, e.g., more than 40 years.

Why the SVR method cannot show better performance than we expect for age prediction? The reason can be in two aspects:

[1]Note that the feature space means a higher dimensional space in support vector regression, which is different from the feature extracted from data in image processing. Actually, the extracted features from images are the input data for SVR in our age modelling.
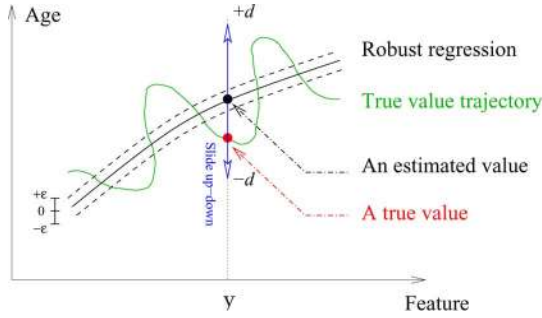
Fig. 7. Illustration of the idea of locally adjusted robust regression.

First, the problem of age prediction is really challenging because of the diversity of aging variation. Each individual may age in his/her own way and be affected by external factors, such as health, living condition, and exposure to weather conditions. Second, the SVR method attempts to find a flat curve to approximate the data in order to obtain good generalization capability. As shown in Fig. 7, the SVR computes a flat curve within a small ϵ tube. However, the age data may distribute like the (green) irregular curve. One cannot expect the SVR to estimate an irregular curve like this because of the over-fitting problem. Further, one cannot assign a large ϵ to enclose all true data points inside the ϵ tube, as demonstrated in the toy example in Section V-B. So how do we model the aging function by allowing the irregular distribution of true ages?

A. Local Adjustment of the Regression Result

One feasible solution is to adjust the age regression values locally so that the estimated age values can be "dragged" towards the true ages. We call it a locally adjusted robust regressor (LARR). The idea of LARR is illustrated in Fig. 7. Suppose the predicted age value by SVR is $f(\mathbf{y})$, corresponding to the input data $\mathbf{y}$. The point $f(\mathbf{y})$ is displayed by the black dot on the regression curve. The estimated age, $f(\mathbf{x})$, may be far away from the true age value, L, shown as the red dot on the true age trajectory curve. The idea of the LARR method is to slide the estimated value, $f(\mathbf{y})$, up and down (corresponding to greater and smaller age values) by checking different age values, $t \in [f(\mathbf{y}) - d, f(\mathbf{y}) + d]$, to see if it can come up with a better age estimation. The value d indicates the range of ages for local search. Hopefully the true age value, L, is also within this range, i.e., $L \in [f(\mathbf{y}) - d, f(\mathbf{y}) + d]$.

Therefore, the LARR method is a two-step procedure: 1) a robust regression over all ages of the training data by using the SVR method. This step can be considered as a global regression process; 2) a local adjustment within a limited range of ages centered at the regression result.

Now the key issue is how to verify different age values within a specified range for the purpose of local adjustment. Remember, our goal is to "drag" the initially estimated age value, $f(\mathbf{y})$, by the global regressor, toward the true age, L, as close as possible. We take a classification approach to locally adjust or verify different ages, considering each age label as one class. Because only a small number of age labels are used for each local adjustment, regression methods cannot work properly. For our classification-based local adjustment, there

are many possible choices of classifiers, but here we adopt a linear SVM for our local age adjustment. The main reason is that the SVM can learn a classifier given a small number of training examples. This has been demonstrated by the author previously for learning in the small sample case, such as face recognition [17], [18], image retrieval [15], audio classification and retrieval [16], and face expression recognition [12]. The capability of learning a classifier in the small sample case is also important for human age prediction. Usually the number of training examples, e.g., 50, is smaller than the feature dimension, e.g., 150, in age estimation, even though we perform experiments on a large database (see Section VII for details).

B. Linear SVM

Given a set of training vectors belong to two separate classes, $(\mathbf{y}_1, z_1), \ldots, (\mathbf{y}_n, z_n)$, where $\mathbf{y}_i \in \mathbb{R}^d$, $z_i \in \{-1, +1\}$, the linear SVM learns an optimal separating hyperplane, $\mathbf{w}\mathbf{y} + b = 0$, that maximizes the margin [35]. The SVM learning is to find the saddle point of the Lagrange functional

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \parallel \mathbf{w} \parallel^2 - \sum_{i=1}^{n} \alpha_i \left\{ z_i \left[(\mathbf{w} \cdot \mathbf{y}_i) + b \right] - 1 \right\}$$
(10)

where α_i are the Lagrange multipliers. The Lagrangian has to be minimized with respect to $\mathbf{w}$, b and maximized with respect to $\alpha_i \geq 0$. The optimization is usually transformed to its *dual* problem

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \left\{ \min_{\mathbf{w}, b} L(\mathbf{w}, b, \alpha) \right\}$$
(11)

and the optimal hyperplane is represented by the dual solution, α

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i z_i \mathbf{y}_i.$$
(12)

The value of b can be estimated by plugging $\mathbf{w}$ into the original equation, $\mathbf{w}\mathbf{y} + b = 0$.

In testing, the classification is given by

$$f(\mathbf{y}) = \text{sign} \left(\mathbf{w} \cdot \mathbf{y} + b \right)$$
(13)

for any new data point $\mathbf{y}$. If the training data are nonseparable, slack variables ξ_i can be introduced. See [35] for more details.

C. Binary Tree Search With Limited Range

The classical SVMs are designed to deal with the two-class classification problems. There are three typical ways to extend it to a multiclass classification application: 1) learning classifiers for each pair of classes, and taking a binary tree search in testing; 2) training SVMs for each class against all the remaining classes; and 3) training SVMs for all classes simultaneously. The last two schemes are not appropriate for our purpose, because in the local adjustment only partial classes of age data are involved. If the last two schemes are used, the SVMs have to be re-trained dynamically for each adjustment, which is computationally expensive. The first scheme is feasible to fulfill our task since there is no need to re-train the SVMs online. Therefore, all pair-wise SVM classifiers can be trained offline. Only a limited number of classes are involved in the binary tree search for test.

The binary tree structure for multiclass SVM classification has been used successfully for 3-D object recognition [25] and face recognition [18]. In general, the number of pair-wise comparisons is $n_c - 1$ for each test in an n_c-class classification problem. Here, the number of pair-wise comparisons is limited to $m_c - 1$ when only m_c classes are involved in each local adjustment, and $m_c < n_c$. Each age corresponds to one class label.

D. Local Search Range Determination

The local search range, m_c, is determined by several factors, such as the scale of the data (large versus small scale) and the performance of the robust regressor (here the kernel SVR).

It is difficult to determine the local search range theoretically. However, there are some guidelines for choosing local search ranges. The larger the search range, the bigger the chance to contain the true ages within that range. If the search range is too small, the true age label might not be reached and the local search may find an arbitrary age label. On the other hand, if the search range is too big, it also increases the possibility to obtain an adjusted age that is far away from the true age, because the local classification is just a locally optimal search.

In our experiments, we specify different ranges and demonstrate the effects of different local search ranges on the results. The main goal is to show that the local adjustment can really improve the performance over the robust regressor for human age estimation.

VII. EXPERIMENTS

A. Data Sets

Extensive age estimation experiments are performed on the UIUC-IFP-Y Internal Aging Database,[2] a large size age database used in a few existing papers [7], [8], [33], [34]. The database contains 8000 high-resolution RGB color face images captured from 1600 different voluntary Asian subjects in an outdoor environment, 800 females and 800 males, in the age range from 0 to 93 years. Each subject has five near frontal images with provided ground truth ages. To our knowledge, this human age database is the largest one ever reported.

The FG-NET Aging Database [1] is a public available age database that we adopt for the experiments. The database contains 1002 high-resolution color or grey scale face images with large variation of lighting, pose, and expression. There are 82 subjects (multiple races) in total with the age ranges from 0 to 69 years, and each face image has 68 labelled points characterizing shape features. The shape features can be combined with appearance features to form a face representation, called active appearance models (AAMs) [3]. The AAMs use 200 parameters to model each face for the purpose of age estimation [10], [33], [34].

B. Results

To evaluate the age estimation performance, we first use a face detector to find the face area in each image, and label the eye corner locations of each face subject. Based on the face and

[2]The images in the database cannot be shown here due to the IP and confidential issues.
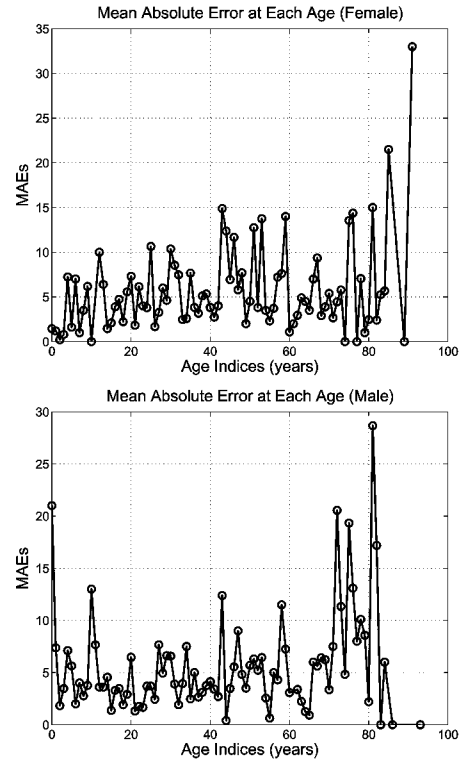

Fig. 8. MAEs at each age for females and males on the UIUC-IFP-Y database, given by our LARR method.

eye corner locations, the face images are cropped, scaled, and transformed to 60×60 gray-level patches. The images have significant variances in illumination since the photographs were taken in the outdoor environment. The grey-level values of each face image are normalized to a normal distribution with zero-mean and one standard deviation in order to reduce the effect of out-door illumination changes. The database also contains some facial expression variations and makeup.

The face image patches with the same size of 60×60 are fed into the manifold learning module. To visualize the embedded age manifold using different techniques, we choose 4000 face images of female subjects [8]. Fig. 3 displays the 2-D and 3-D manifold visualization of the PCA, LLE, and OLPP algorithms. The datapoints of ages from 0 to 93 are colored from blue to red. We can obtain three observations. 1) The PCA method does not shown clear manifold trend of ages. The reason is that the PCA is purely unsupervised without using any age label information, which seems to be important for learning the embedded manifold from the complex aging patterns. 2) The manifold learned by the LLE (a nonlinear embedding method) is approximately an ellipsoid with higher age (red) in the center and lower age (blue) at periphery. 3) The OLPP algorithm achieves good visualization of the age manifold with a distinct aging trend. Therefore, we choose the OLPP method in our age manifold learning module in the age estimation framework.

After the age manifold was learned, each face image can be projected onto the age manifold to extract a feature vector. We use the first 150 features for each face image by following our previous work in [8]. The system then learns a robust regression function using the kernel SVR method for females and males separately. Actually the manifold was learned for the female
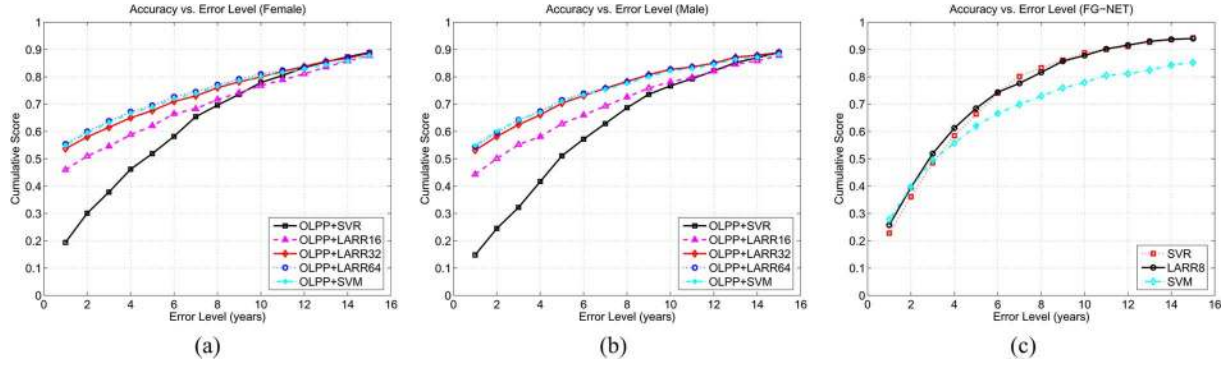
Fig. 9.   Cumulative scores of the algorithms with different settings for (a) female age estimation, (b) male age estimation on the UIUC-IFP-Y database, and (c) age estimation on the FG-NET database, at error levels from 1 to 15 years.

and male independently. As demonstrated in the toy example in Section V-B, a small ϵ value should be chosen for the ϵ-insensitive loss function in (4). We set $\epsilon = 0.02$ for our age estimation task. In SVR learning, parameters C and γ are determined on a validation set. Experimentally, we found that a good choice is $C = 40$ and $\gamma = 12$, separately. To locally adjust the global regression results, we tried different local search ranges as powers of two, e.g., 4, 8, 16, 32, and 64 classes, and the results from different search ranges are compared to see the effect of local adjustment. The purpose of choosing the powers of two is to simplify the binary search structure. One can observe that the local search range does influence the age estimation results. Currently, we cannot set the local search range automatically in a data-driven manner, and we will investigate this issue in the future. The pair-wise linear SVM classifiers were used for the local adjustment, centered at the age value (or label) obtained from the global regressor.

We perform a standard four-fold cross validation test to evaluate the accuracy of our algorithms for age estimation on the UIUC-IFP-Y age database. The test was executed on the female and male subsets separately. The reason is that we found females and males age quite differently in the database. So, our current system has a latent module to first estimate gender before age estimation in the UIUC-IFP-Y database. For each experiment, about 1/3 of the training data are used as a validation subset to determine the optimal parameter setting such as C and γ. Then the parameters are fixed and the whole training data set is used to learn the robust regression function. The pair-wise linear SVM classifiers are learned using the same training data and used for local adjustment in testing. Finally, all performance measures are reported on the unseen test data.

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, $\text{MAE} = \sum_{k=1}^{N} |\hat{l_k} - l_k|/N$, where l_k is the ground truth age for the test image k, $\hat{l_k}$ is the estimated age, and N is the total number of test images. The MAE measure was used previously in [10], [22], [23], [34]. The cumulative score [10] is defined as $\text{CS}(j) = N_{e \le j}/N \times 100\%$, where $N_{e \le j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years.

Table I and Figs. 8 and 9 show the experimental results. The first and second columns in Table I show the MAEs for females

TABLE I
MAEs OF OUR METHOD WITH DIFFERENT SETTINGS

Various Setup	UIUC-IFP-Y/F	UIUC-IFP-Y/M	FG-NET
SVM	5.55	5.52	7.16
SVR	7.00	7.47	5.16
LARR4	6.83	7.21	**5.07**
LARR8	6.48	6.81	**5.07**
LARR16	5.86	5.95	5.12
LARR32	5.29	**5.30**	6.03
LARR64	**5.25**	5.38	—

and males in the UIUC-IFP-Y age database, separately. Different ranges, e.g., 4, 8, 16, 32, and 64, were tried for local adjustment of the global regression results. One can see that the local adjustment truly reduces the errors of the global regression. For example, the MAE of the SVR is seven years for the female (column 1 in Table I), but is reduced to 5.86 (column 1, row 5) when 16 local classes are used for the LARR method, and so on. Different ranges of adjustment do have different MAEs. For comparison, we also show the results using purely the SVM classifiers in the first row. One can see that the classification scheme has lower errors than the pure regression method for both females and males, but it has higher error rates than some of the locally adjusted results. The best LARR result in terms of MAE is 5.25 years for females when the local search range is 64 classes, while it is 5.30 years for males when the adjust range is 32 classes. The ranges of local adjustment depend on the data and the global regression results. To illustrate the MAEs at each age, two pictures for female and male results are displayed in Fig. 8, respectively.

Fig. 9(a) and (b) shows the CS measures for females and males separately. We can observe that the LARR methods (with different ranges for local adjustment) improve the score significantly over the pure regression method for lower error levels, e.g., $m_c < 10$ years. For example, in one year error level, most LARRs with proper ranges of local adjustment could improve the accuracy by 175% and 267% for females and males separately. This improvement is significant. We also notice that large ranges are required for local adjustment on the UIUC-IFP-Y age database. For instance, when 16 classes are used for local adjustment, the CS curve is explicitly lower than 32 or 64 classes. We do not show the cumulative scores for four and eight classes here in order to not mess up the figures. Those two CS curves are even lower than 16 classes. One may also notice that the CS curve of SVM classifiers is close to the LARR32 and LARR64 for both
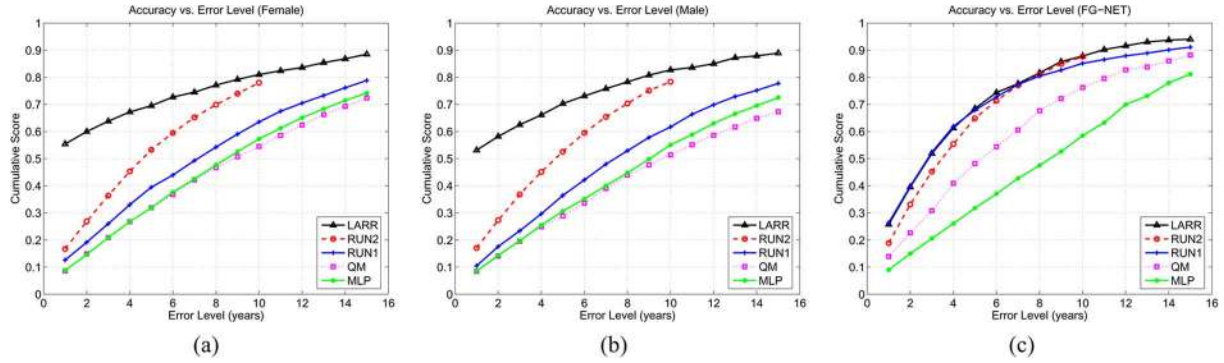
Fig. 10.   Comparisons between our LARR method and other state-of-the-art methods in terms of the cumulative scores for (a) female age estimation, (b) male age estimation on the UIUC-IFP-Y database, and (c) age estimation on the FG-NET database.

TABLE II
MAE COMPARISONS OF DIFFERENT ALGORITHMS

Method	UIUC-IFP-Y/F	UIUC-IFP-Y/M	FG-NET
WAS [10]	–	–	8.06
AGES [10]	–	–	6.77
QM [23]	9.96	10.51	6.55
MLPs [23]	10.99	12.00	6.98
RUN1 [34]	9.79	10.36	5.78
RUN2 [33]	6.95	6.95	5.33
LARR(Ours)	**5.25**	**5.30**	**5.07**

females and males, but the MAEs of the SVM are higher than the LARR16 or LARR32 as shown in Table I. This indicates that we need both MAE and CS measures complementarily to measure the performance of an algorithm in age estimation.

As shown in Table II, we also compare our results with all previous methods reported on the UIUC-IFP-Y age database. It turns out our LARR method has the MAEs of 5.25 and 5.30 years for females and males separately, which are explicitly smaller than all previous results under the same experimental protocol. Our method brings about 24% deduction of MAEs over the best result of previous approaches. The comparisons of cumulative scores are shown in Fig. 10(a) and (b). The method of the RUN2 only has the age error levels up to ten years in [33]. The results were provided by the authors in [33] and [34].

We use the same AAM features as in [10], [33], [34] to evaluate our LARR method on FG-NET database. Since the FG-NET database has small size, we do not learn any age manifold but use the AAM features directly. Our focus is then to evaluate the performance of the new LARR method for age estimation on the FG-NET database. The popular test strategy, namely leave-one-person-out (LOPO), was usually taken for the FG-NET age database, as suggested by the existing work [10], [33], [34]. We follow the same strategy and compare our results with the state-of-the-art methods. The experimental results are shown in the third column of Tables I and II. One can see that our LARR method has an MAE of 5.07 years which is lower than all previous methods listed in Table II. Our best MAE was obtained using either four or eight classes for local adjustment as shown in Table I. Increasing the local search ranges for the LARR method will make the errors larger. For example, the MAE will be 6.03 years when 32 classes are used for local adjustment. We cannot get the result for 64 classes since there are at most 63 or 61 age labels in the LOPO test. In other words, there are missing ages in the FG-NET database. When the pure classifiers, SVMs, are used, the MAE is 7.16,

which is much higher that the 5.16 years of the pure regression. One possible reason is that there is not sufficient data for pair-wise SVM training, while the global SVR uses all the data in the model. Another observation is that the robust regression itself (without local adjustment) has an MAE of 5.16 years, which is still lower than all previous methods shown in Table II. Our LARR method further reduces the MAE to 5.07 years.

Fig. 9(c) shows the cumulative scores of our LARR method on the FG-NET database. LARR8 means using eight classes for local adjustment. We do not show LARR4, LARR16, and LARR32 in order to avoid messing up the display. The cumulative scores of those ranges are close to LARR8 with slight differences. LARR8 has higher accuracy than the pure regression by SVR at lower error levels (1–6), but close to it at higher error levels. The cumulative scores of the pure SVM are much lower than the pure SVR for most error levels, which indirectly indicates the significance of constraining the SVM search in a local range. The cumulative score comparisons are shown in Fig. 10(c). Our LARR method performs much better than the QM and MLP methods. The method of RUN1 [34] is close to our LARR in low age error levels, but worse than LARR in high levels. In contrast, the method of RUN2 [33] is close to our LARR in high age error levels, but worse than the LARR in low error levels. Overall, our LARR method has higher accuracy than both the RUN1 and RUN2 on the FG-NET database.

VIII.   CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new framework for face-image-based automatic age estimation. A manifold learning method was introduced for learning the low-dimensional age manifold. The support vector machine and support vector regression methods are investigated for age prediction based on the learned manifolds. To improve the age estimation performance and robustness, a locally adjusted robust regressor (LARR) was designed. From the experimental evaluations on our internal UIUC-IFP-Y database and the public available FG-NET database, we conclude that 1) the LARR method gives better age estimation than the purely robust regression by SVR or purely classification by SVM and 2) the LARR method outperforms many state-of-the-art approaches to age estimation. We expect to see more applications of the LARR method to other challenging real-world problems. An interesting topic to investigate in the future is to determine the local adjust ranges

in a data-driven manner instead of specified values, which may further improve the accuracy.

ACKNOWLEDGMENT

REFERENCES

[1] The FG-NET Aging Database [Online]. Available: http://www.fgnet.rsunit.com/

[2] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. ECCV*, 1998, vol. 2, pp. 484–498.

[4] K. A. Deffenbacher, T. Vetter, J. Johanson, and A. J. O'Toole, "Facial aging, attractiveness, and aistinctiveness," *Perception*, vol. 27, pp. 1233–1243, 1998.

[5] R. O. Duda, E. H. Peter, and G. S. David, *Pattern Classification*, 2nd ed. New York: Wiley Interscience, 2000.

[6] Y. Fu and T. S. Huang, *Locally Linear Embedded Eigenspace Analysis, IFP-TR*. Urbana, IL: Univ. Illinois, 2005.

[7] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Trans. Multimedia*, to be published.

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

[9] Y. Fu and N. Zheng, "M-face: An appearance-based photorealistic model for multiple facial attributes rendering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 830–842, 2006.

[10] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proc. ACM Conf. Multimedia*, 2006, pp. 307–316.

[11] S. R. Gunn, "Support vector machines for classification and regression," Tech. Rep. Image Speech and Intelligent Systems Research Group, Univ. Southampton, U.K., 1997.

[12] G. Guo and C. Dyer, "Learning from examples in the small sample case: Face expression recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 3, pp. 477–488, Jun. 2005.

[13] G. Guo, C. Dyer, and Z. Zhang, "Linear combination representation for outlier detection in motion tracking," in *Proc. IEEE Conf. CVPR*, 2005, vol. 2, pp. 274–281.

[14] G. Guo, Y. Fu, T. S. Huang, and C. Dyer, "Locally adjusted robust regression for human age estimation," presented at the IEEE Workshop on Applications of Computer Vision, 2008.

[15] G. Guo, A. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 811–820, Apr. 2002.

[16] G. Guo and S. Li, "Content-based audio classification and retrieval by support vector machines," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 209–215, Jan. 2003.

[17] G. Guo, S. Li, and K. Chan, "Support vector machines for face recognition," *Image Vis. Comput.*, vol. 19, no. 9-10, pp. 631–638, 2001.

[18] G. Guo, S. Li, and K. Chan, "Face recognition by support vector machines," in *Proc. IEEE Conf. FG*, 2000, pp. 196–201.

[19] X. F. He and P. Niyogi, "Locality preserving projections," presented at the NIPS, 2003.

[20] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999.

[21] Y. Kwon and N. Lobo, "Age classification from facial images," *Comput. Vis. Image Understand.*, vol. 74, no. 1, pp. 1–21, 1999.

[22] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.

[23] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 621–628, Feb. 2004.

[24] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs, "A study of face recognition as people age," presented at the IEEE Conf. ICCV, 2007.

[25] M. Pontil and A. Verri, "Support vector machines for 3-D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 637–646, Jun. 1998.

[26] N. Ramanathan and R. Chellappa, "Face verification across age progression," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3349–3361, Nov. 2006.

[27] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. IEEE Conf. CVPR*, 2006, pp. 387–394.

[28] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[29] C. M. Scandrett, C. J. Solomona, and S. J. Gibsona, "A person-specific, rigorous aging model of the human face," *Pattern Recognit. Lett.*, vol. 27, no. 15, pp. 1776–1787, 2006.

[30] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.

[31] J. Suo, F. Min, S. Zhu, S. Shan, and X. Chen, "A multi-resolution dynamic model for face aging simulation," presented at the IEEE Conf. CVPR, 2007.

[32] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[33] S. Yan, H. Wang, T. S. Huang, and X. Tang, "Ranking with uncertain labels," in *Proc. IEEE Conf. Multimedia and Expo*, 2007, pp. 96–99.

[34] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," presented at the IEEE Conf. ICCV, 2007.

[35] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

Guodong Guo (M'07–SM'07) received the B.E. degree in automation from Tsinghua University, Beijing, China, in 1991, the Ph.D. degree in pattern recognition and intelligent control from the Chinese Academy of Sciences in 1998, and the Ph.D. degree in computer science from the University of Wisconsin, Madison, in 2006.

He is currently an Assistant Professor in the Department of Computer Science, North Carolina Central University, Durham. He visited and worked at INRIA, Sophia Antipolis, France in 1997, Ritsumeikan University, Japan, in 1998, Nanyang Technological University, Singapore, from 1999 to 2000, and Microsoft Research, China, from 2000 to 2001. His research areas include computer vision, machine learning, and multimedia. He has authored the book *Face, Expression, and Iris Recognition Using Learning-based Approaches* (VDM Verlag, 2008), published over 40 technical papers in face, expression, age, and iris recognition, multimedia information retrieval, and image analysis, and filed three patents on iris and texture image analysis.

Yun Fu (S'07) received the B.E. degree in information and communication engineering from the School of Electronic and Information Engineering, Xi'an Jiaotong University (XJTU), China, in 2001, the M.S. degree in pattern recognition and intelligence systems from the Institute of Artificial Intelligence and Robotics, XJTU, in 2004, and the M.S. degree in statistics from the Department of Statistics, University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2007. He is currently pursuing the Ph.D. degree in the Electrical and Computer Engineering Department, UIUC.

From 2001 to 2004, he was a Research Assistant at the Institute of Artificial Intelligence and Robotics (AI&R), XJTU. From 2004 to 2007, he was a Research Assistant at the Beckman Institute for Advanced Science and Technology, UIUC. He was a Research Intern with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, Summer 2005, and with the Multimedia Research Lab of Motorola Labs, Schaumburg, IL, in Summer 2006. His research interests include machine learning, human computer interaction, image processing, multimedia, and computer vision.

Dr. Fu is the recipient of the 2002 Rockwell Automation Master of Science Award, two Edison Cups of the 2002 GE Fund " *Edison Cup*" Technology Innovation Competition, the 2003 HP Silver Medal and Science Scholarship for Excellent Chinese Student, the 2007 Chinese Government Award for Outstanding Self-financed Students Abroad, the 2007 DoCoMo USA Labs Innovative Paper Award (IEEE ICIP 2007 best paper award), the 2007–2008 Beckman Graduate Fellowship, and the 2008 M. E. Van Valkenburg Graduate Research Award. He is a student member of Institute of Mathematical Statistics (IMS) and a Beckman Graduate Fellow.

Charles R. Dyer (M'74–SM'85–F'98) received the B.S. degree in mathematical sciences from Stanford University, Stanford, CA, in 1973, the M.S. degree in computer science from the University of California, Los Angeles, in 1974, and the Ph.D. degree in computer science from the University of Maryland, College Park, in 1979.

From 1979 to 1982, he was an Assistant Professor in the Department of Electrical Engineering and Computer Science, University of Illinois at Chicago. He has been on the faculty at the University of Wisconsin, Madison, since 1982, where he is currently Professor in the Departments of Computer Sciences and Biostatistics and Medical Informatics. He was the Chair of the Department of Computer Sciences from 1990 to 1993. He has worked in the field of computer vision for over 30 years and has published over 130 technical papers. His research has focused mainly on shape and image representations for computer vision, with contributions on topics such as quadtrees, aspect graphs, pyramid algorithms, computational photography, object recognition, and motion analysis.

Dr. Dyer is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and the International Association for Pattern Recognition (IAPR). He was awarded the Visiting Fellowship at Exeter College, University of Oxford, Oxford, U.K., for 1999–2000. He was General Co-Chair of the 2003 Computer Vision and Pattern Recognition Conference. He served as an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and is currently on the Editorial Board of the journal *Machine Vision and Applications*. He is currently a member of the Governing Board of the IAPR.

Thomas S. Huang (S'61–M'63–SM'76–F'79–LF'01) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., and the M.S. and D.Sc. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge.

He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University, West Lafayette, IN, from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, Urbana, where he is now the William L. Everitt Distinguished Professor of Electlrical and Computer Engineering, a Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology and Co-Chair of the Institute's major research theme Human Computer Intelligent Interaction. He has published 20 books, and over 500 papers in network theory, digital filtering, image processing, and computer vision. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals.

Dr. Huang is a Member of the National Academy of Engineering, a Foreign Member of the Chinese Academies of Engineering and Sciences, and a Fellow of the International Association of Pattern Recognition and the Optical Society of America, and has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. In 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize from the International Association of Pattern Recognition and the Pan Wen-Yuan Outstanding Research Award. In 2005, he received the Okawa Prize. In 2006, he was named by IS&T and SPIE as the Electronic Imaging Scientist of the Year. He is a Founding Editor of the *International Journal Computer Vision, Graphics, and Image Processing* and Editor of the Springer Series in Information Sciences, published by Springer Verlag.