Locally Adjusted Robust Regression for Human Age Estimation

Guodong Guo
Computer Science
NCCU
Durham, NC, 27707
gdguo@nccu.edu

Yun Fu
Beckman Institute
UIUC
Urbana, IL 61801
yunfu2@uiuc.edu

Thomas S. Huang
Beckman Institute
UIUC
Urbana, IL 61801
t-huang1@uiuc.edu

Charles R. Dyer
Computer Sciences
UW-Madison
Madison, WI, 53706
dyer@cs.wisc.edu

Abstract

Automatic human age estimation has considerable potential applications in human computer interaction and multimedia communication. However, the age estimation problem is challenging. We design a locally adjusted robust regressor (LARR) for learning and prediction of human ages. The novel approach reduces the age estimation errors significantly over all previous methods. Experiments on two aging databases show the success of the proposed method for human aging estimation.

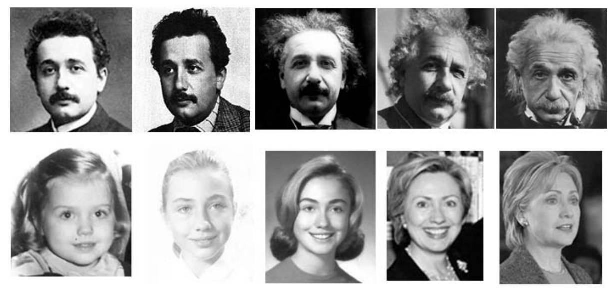


Figure 1. Face aging: each row shows images of the same individual at different ages.

1. Introduction

Human faces convey a significant amount of nonverbal information for human-to-human communication. Our humans have the ability to accurately recognize and interpret faces in real time. Given a captured face image, various attributes can be estimated from it, such as identity, age, gender, expression, and ethnic origin [6]. Facial attributes play a crucial role in real applications including multimedia communication and Human Computer Interaction (HCI). For example, if the user's age is estimated by a computer, an Age Specific Human Computer Interaction (ASHCI) system may be developed. Such a system could be used for secure internet access control in order to ensure young kids have no access to internet pages with adult materials; A vending machine can refuse to sell alcohol or cigarettes to the underage people [19] [9]. In image and video retrieval, users could retrieve their photographs or videos by specifying a required age range [19].

However, automatic age estimation from human face images is a challenging problem. The main difficulty is that different persons 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. Figure 1 shows some face images with different ages.

Surprisingly, there are only a few publications on age estimation in the literature. Existing face-image-based age estimation methods can be divided into three main categories [5]: anthropometric model [17] [22], aging pattern subspace [9], and age regression [18, 19, 24, 23]. The anthropometric model uses the cranio-facial development theory and facial skin wrinkle analysis. The growth related changes of face shape and texture patterns are measured to categorize a face into several age groups. Such methods are suitable for coarse age estimation rather than continuous or refined classification [17] or modelling ages just for young people [22]. The AGing pattErn Subspace (AGES) method [9] models a sequence of personal aging face images by learning a subspace representation, in order to handle incomplete data such as missing ages in the sequence. The age of a testing face is determined by the projection in the subspace that can best reconstruct the face image. For the regression methods, the regression coefficients are estimated from the training data with an assumption of the regression function such as a Quadratic Model (QM) [19]. Yan et al. [24, 23] also dealt with the age uncertainty by formulating a semi-definite programming problem [24] or an EM-based algorithm [23]. Fu et al. showed the existence of an aging manifold [5] visualized in face image subspaces [8]. They suggested the fusion of manifold learning and a quadratic regression model to improve the age estimation performance.

The traditional quadratic model [19] for age regression is based on a Least Square Estimation (LSE) criterion which is not robust to outliers. The outliers could come from some incorrectly labelled ages. In addition, the optimization based on the LSE criterion minimizes the empirical risk which usually cannot generalize well especially when a small number of training data are available.

In this paper, we propose a novel method for automatic age estimation. Our method is called a Locally Adjusted Robust Regressor (LARR) for learning and prediction of the aging patterns. The advantages of the proposed method will be demonstrated with extensive experiments.

The remainder of the paper is organized as follows. In Section 2, the quadratic regression function is introduced which was used in previous approaches. In Section 3, 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 4. Experimental evaluations of the proposed approach and comparisons with previous methods are given in Section 5, and finally the conclusion is provided in Section 6.

2. Quadratic Regression

Given the extracted features for each face image, people usually took a regression function, $L = f(x)$, to characterize the relationship between the extracted features, x, and the age labels, L. A typical choice of the regression function, f, is the Quadratic Model (QM) [18], $\hat{L} = w_0 + w_1^T x + w_2^T x^2$, where $\hat{L}$ is the estimate of the age, w_0 is the offset, x and x^2 are the extracted feature vector and its square, and w_1 and w_2 are weight vectors.

The QM method has been used for age regression in previous approaches [18] [19]. The model parameters are optimized by minimizing the difference between the actual ages of the individuals and the estimated ages. The loss function usually corresponds to a Least Squares 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. Therefore, we need to seek a robust model for modelling the aging patterns.

For the purpose of robust aging regression, we adopt the Support Vector Regression (SVR) method [25]. The SVR might attack the three limitations of the traditional quadratic regression model.

3. Support Vector Regression

The basic idea of SVR is to find a function $f(x)$ that has most ϵ deviation from the actually obtained target y_i for the training data x_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. In comparison with the conventional quadratic loss function shown in Figure 2(a), the ϵ-insensitive loss function of SVR is shown in Figure 2(b).

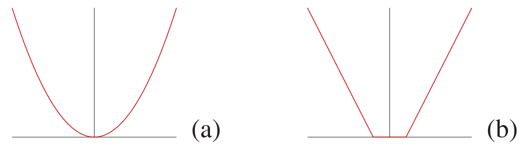


Figure 2. (a) quadratic regression loss function; (b) ϵ-insensitive loss function.

Consider the problem of approximating the set of data $\mathcal{D} = \{(x_1, y_1), \ldots, (x_l, y_l)\}, x \in R^n, y \in R$, with a linear function,

$$f(x) = \langle w, x \rangle + b. \tag{1}$$

The optimal regression function [25] is given by

$$\begin{array}{ll} \min\limits_{w,\xi} & \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} (\xi_i^+ + \xi_i^-) \\ & y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i^+ \\ \text{subject to} & \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^- \\ & \xi_i^+, \xi_i^- \leq 0 \end{array} \tag{2}$$

where constant $C > 0$ determines the trade-off 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 (2). The ϵ-insensitive loss function as shown in Figure 2(b) is

$$L_\epsilon(x, y) = \begin{cases} 0 & \text{if } |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon & \text{otherwise} \end{cases} \tag{3}$$

The *primal* problem of (2) can be solved more efficiently in its *dual* formulation [25] resulting in the final solution given by

$$w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i, \tag{4}$$

and

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b, \tag{5}$$

where α_i, α_i^* are Lagrange multipliers. The value of b in Eq. (1) can be determined by plugging Eq. (4) into Eq. (1) [10].

A nonlinear regression function can be obtained by using kernels, in the same manner as a nonlinear Support Vector Machine (SVM) for classification [25]. Different kernels, such as polynomials, sigmoid, or Gaussian radial basis functions, can be used depending on the tasks. For our robust age regression, the Gaussian radial basis function kernel was adopted. A radial basis function is

$$k(x, x') = e^{-\gamma \|x - x'\|^2}, \tag{6}$$

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 [25],

$$\langle w, x \rangle = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x), \tag{7}$$

and

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x) + b. \tag{8}$$

4. Locally Adjusted Robust Regression

Is it “good” enough for human age prediction using the SVR? To answer this question, let us look at an estimation result using the SVR. Figure 3 shows the predicted ages (squares) with respect to the ground truth ages (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. For example, 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.

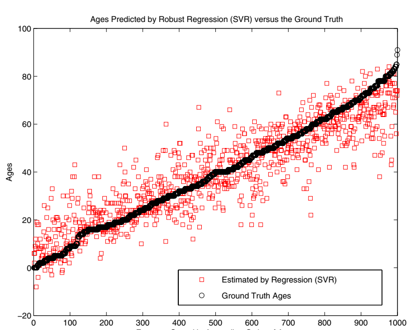


Figure 3. The true ages (circles) versus the estimated (squares) for 1,000 females.

Why the SVR method cannot show better performance than we expect for age prediction? The reason can be in two aspects: First, the problem of age prediction is really challenging because of the diversity of aging variation. Second, the SVR method attempts to find a flat curve within a small ϵ tube to approximate the data in order to obtain good generalization capability. But the age data may distribute like the (green) irregular curve shown in Figure 4. 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 motion estimation [12]. So how to model the aging function by allowing the irregular distribution of true ages?

4.1. Local Adjustment

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 Figure 4. Suppose the predicted age value by SVR is $f(x)$, corresponding to the input data x. The point $f(x)$ is displayed by the black dot on the regression curve. The estimated age, $f(x)$, 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, $f(x)$, up and down (corresponding to greater and smaller age values) by checking different age values, $t \in [f(x) - d, f(x) + d]$, 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., $L \in [f(x) - d, f(x) + d]$.

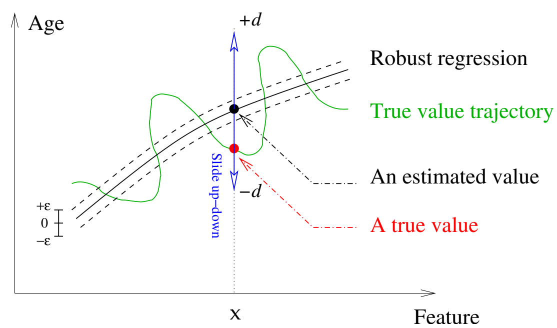


Figure 4. Illustration of the LARR idea.

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, $f(x)$, by the global regressor, towards 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. For our classification-based local adjustment, there are many possible choices of classifiers, but here we adopt a linear Support Vector Machine (SVM) [25] for our local age adjustment. The main reason is that the

SVM can learn a classifier given a small number of training examples, which has been demonstrated successfully in problems such as face expression recognition [11], audio classification [14], and image retrieval [13].

4.2. Limited Binary Tree Search

The classical SVMs deal with the two-class classification problem. To extend to a multi-class classification, there are three typical ways: (1) learning classifiers for each pair of classes, and taking a binary tree search in testing [20] [15]; (2) training SVMs for each class against all the remaining classes; and (3) training SVMs for all classes simultaneously. However, the last two schemes are not appropriate for our purpose here. Because in our 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. While in the first scheme there is no need to re-train the SVMs. All pair-wise SVM classifiers can be trained off-line, and only a limited number of classes are involved in the binary tree search in testing.

In general, the number of pair-wise comparisons is $n-1$ for each test in an n-class classification problem [20] [15]. Here in our age adjustment, the number of pair-wise comparisons is limited to $m-1$ when only m classes are involved in each local adjustment, and $m < n$.

5. Experiments

We perform age estimation experiments on a large database, the UIUC-IFP-Y Aging Database. The database contains 8,000 high-resolution RGB color face images captured from 1,600 different voluntary Asian subjects, 800 females and 800 males, in the age range from 0 to 93 years. Each subject has 5 near frontal images with provided ground truth ages. To our knowledge, this human age database is the largest one ever reported.

For age estimation, 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 eye corner locations, the face images are cropped, scaled, and transformed to 60×60 gray-level patches. The grey-level values of each face image are normalized to reduce the effect of out-door illumination changes. The face image patches are fed into a manifold learning module. We use the orthogonal locality preserving projections [2] method for age manifold learning, similar to the approach in [5]. 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. The system then learns a robust regression function using the kernel SVR method for females and males separately. A small ϵ value was chosen, $\epsilon = 0.02$, for the loss function in Eqn. (3). In SVR learning, parameters C and γ are determined on a validation set. Different ranges such as 4, 8, 16, 32, and 64 were tried for the local adjustment of the global regression results, and compared to see the effect of local adjustment. The purpose of choosing the powers of two is to simplify the binary search structure. The pair-wise linear SVM classifiers were used for the local adjustment, centered at the age value (or label) obtained from the global regressor.

Table 1. MAEs of our approaches.

Various Setup	U-I-Y/F	U-I-Y/M	FG-NET
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	–
SVM	5.55	5.52	7.16

To evaluate the accuracy of our algorithms for age estimation on the UIUC-IFP-Y age database, we perform a standard 4-fold cross validation test. 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. The performance of age estimation can be measured by two different measures: the Mean Absolute Error (MAE) and the Cumulative Score (CS). The MAE is defined as the average of the absolute errors between the estimated ages and the ground truth ages, $MAE = \sum_{k=1}^{N} |\hat{l_k} - l_k|/N$, where l_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 [19] [18] [9] [24]. The cumulative score [9] is defined as $CS(j) = N_{e \leq j}/N \times 100\%$, where $N_{e \leq j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years.

Experimental results are shown in Table 1 and Figure 5 (a) and (b). The first and second columns in Table 1 show the MAEs for females 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. Different ranges of adjustment do have different MAEs. For comparison, we also show the results using purely the SVM classifiers in the last row. The best LARR result in terms of MAE is 5.29 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.

The CS measures are shown in Figure 5 (a) and (b) for females and males, separately. One can observe that

Table 2. MAE comparisons with others.

Method	U-I-Y/F	U-I-Y/M	FG-NET
WAS [9]	–	–	8.06
AGES [9]	–	–	6.77
QM [19]	9.96	10.51	6.55
MLPs [19]	10.99	12.00	6.98
RUN 1 [24]	9.79	10.36	5.78
RUN 2 [23]	6.95	6.95	5.33
LARR(Ours)	**5.25**	**5.30**	**5.07**

the LARR methods (with different ranges for local adjustment) improve the score significantly over the pure regression method especially for lower error levels, e.g., $m < 10$ years. We do not show the cumulative scores for 4 and 8 classes here in order to not mess up the figures. Those two CS curves are even lower than 16 classes.

We also compare our results with all previous methods reported on the UIUC-IFP-Y age database. As shown in Table 2, 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. Our method brings about 24% deduction of MAEs over the best result of previous approaches. The comparisons of cumulative scores are shown in Figure 6 (a) and (b).

There is a public available age database, the FG-NET [1]. In this age database, 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 [9] [24] [23]. To evaluate our LARR method on FG-NET, we use the same AAM features as in [9] [24] [23].

A test strategy, called Leave-One-Person-Out (LOPO), was usually taken for the FG-NET age database [9] [24] [23]. We follow the same strategy and compare our results with previous ones. The experimental results are shown in the third column of Tables 1 and 2. One can see that our LARR method has an MAE of 5.07 years which is lower than all previous methods listed in Table 2. Our best MAE was obtained using either 4 or 8 classes for local adjustment as shown in Table 1. Increasing the local search ranges for the LARR method will make the errors bigger.

The cumulative scores of our LARR method on the FG-NET database are shown in Figure 5 (c). LARR8 means using 8 classes for local adjustment. 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 Figure 6 (c). Our LARR method performs much better than the QM an MLP methods. The method of RUN1 [24] is close to our LARR in low age error levels, but worse than LARR in high levels. In contrast, the method of RUN2 [23] 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.

From the experimental evaluations, we summarize 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 has lower errors than all the state-of-the-art approaches to age estimation.

6. Conclusion

We have presented a new approach to automatic age estimation. The Support Vector Machine (SVM) and Support Vector Regression (SVR) methods are investigated for age prediction for the first time. A Locally Adjusted Robust Regressor (LARR) was designed to estimate ages with high performance. Experimental evaluations on a large age database and the public available FG-NET database show that our LARR method performs better than all the state-of-the-art approaches. We expect to see more applications of the LARR method to other challenging real problems.

References

[1] The FG-NET Aging database, http://sting.cycollege.ac.cy/∼alanitis/ fgnetaging/index.htm.

[2] D. Cai, X.F. He, J.W. Han, and H.-J. Zhang, "Orthogonal Laplacianfaces for Face Recognition," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3608-3614, 2006.

[3] T.F.Cootes, G.J. Edwards and C.J.Taylor. "Active Appearance Models", in *Proc. European Conference on Computer Vision* Vol. 2, pp. 484-498, 1998.

[4] R.O. Duda, E.H. Peter, and G.S. David, *Pattern Classification* (2^{nd} Edition), Wiley Interscience, 2000.

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

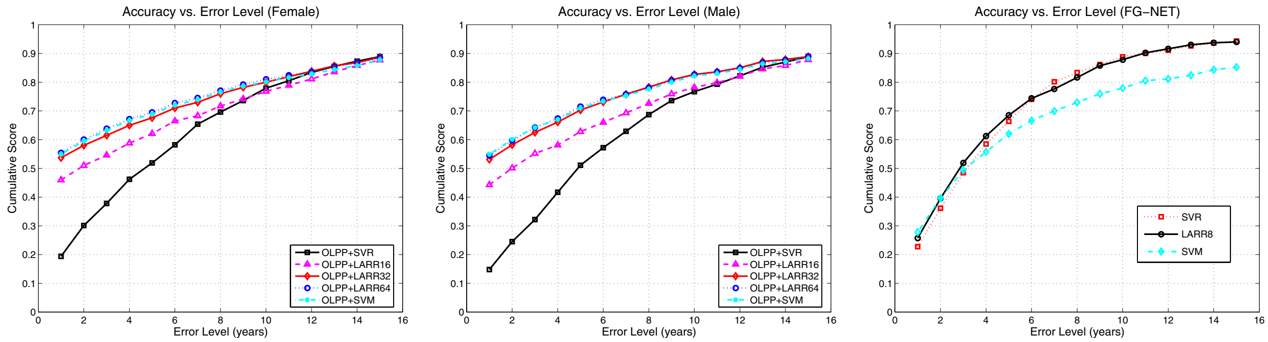
[6] Y. Fu and N.N. Zheng, "M-Face: An Appearance-Based Photorealistic Model for Multiple Facial Attributes Rendering," *IEEE Trans. on CSVT*, vol. 16, no. 7, pp. 830-842, 2006.

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

[8] Y. Fu and T. S. Huang. "Locally linear embedded eigenspace analysis", IFP-TR, UIUC, January, 2005.

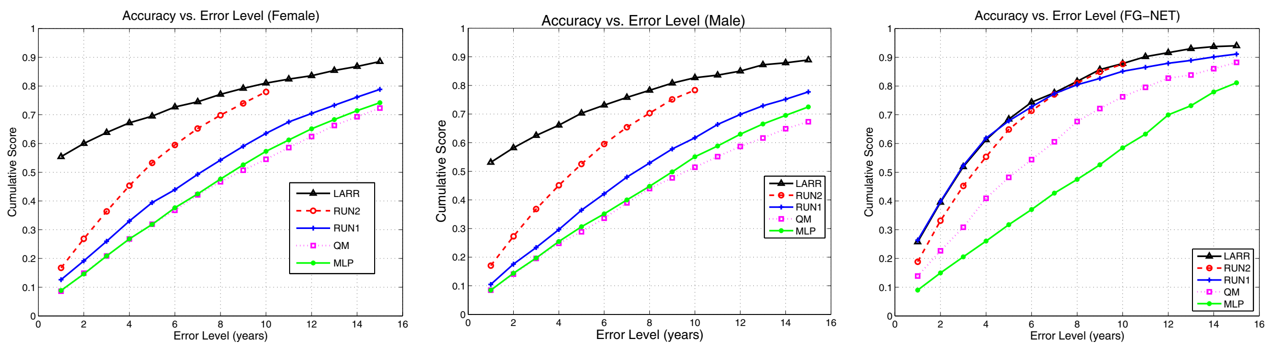
[9] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from Facial Aging Patterns for Automatic Age Estimation," *ACM Conf. on Multimedia (ACM MM'06)*, pp. 307-316, 2006.

[10] S. R. Gunn, "Support vector machines for classification and regression," technical report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.



(a) (b) (c)

Figure 5. 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.



(a) (b) (c)

Figure 6. 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.

[11] G. Guo and C. Dyer, "Learning from Examples in the Small Sample Case: Face Expression Recognition," *IEEE Trans. on SMC-B*, Vol. 35, No. 3, 477-488, 2005.

[12] G. Guo, C. Dyer, and Z. Zhang, "Linear Combination Representation for Outlier Detection in Motion Tracking," *IEEE Conference on CVPR*, Vol. 2, pp. 274-281, 2005.

[13] G. Guo, A. Jain, W. Ma, and H. Zhang, "Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback," *IEEE Trans. on Neural Networks*, Vol. 13, No. 4, 811-820, 2002.

[14] G. Guo and S. Li, "Content-based Audio Classification and Retrieval by Support Vector Machines," *IEEE Trans. on Neural Networks*, Vol. 14, No. 1, 209-215, 2003.

[15] G. Guo, S. Li, and K. Chan, "Face Recognition by Support Vector Machines," *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 196-201, 2000.

[16] T. Joachims, "Making large-scale SVM learning practical," *Advances in Kernel Methods - Support Vector Learning*, B. Schlkopf and C. Burges and A. Smola (ed.), MIT Press, 1999.

[17] Y. Kwon and N. Lobo, "Age Classification from Facial Images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1-21, 1999.

[18] A. Lanitis, C. Taylor, and T. Cootes, "Toward Automatic Simulation of Aging Effects on Face Images," *IEEE Trans. on PAMI*, vol. 24, no. 4, pp. 442-455, 2002.

[19] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing Different Classifiers for Automatic Age Estimation," *IEEE Trans. on SMC-B*, vol. 34, no. 1, pp. 621-628, 2004.

[20] M. Pontil and A. Verri, "Support Vector Machines for 3-D Object Recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, 637-646, 1998.

[21] N. Ramanathan and R. Chellappa, "Face Verification Across Age Progression," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3349-3361, 2006.

[22] N. Ramanathan and R. Chellappa, "Modeling Age Progression in Young Faces," *IEEE Conf. on CVPR'06*, pp. 387-394, 2006.

[23] S. Yan, H. Wang, T. S. Huang, and X. Tang, "Ranking with Uncertain Labels", *IEEE Conference on Multimedia and Expo*, 2007.

[24] S. Yan et al., "Learning Auto-Structured Regressor from Uncertain Nonnegative Labels", *ICCV*, 2007.

[25] V. N. Vapnik, *Statistical Learning Theory*, John Wiley, New York, 1998.