

# Regression From Uncertain Labels and Its Applications to Soft Biometrics

Shuicheng Yan, *Member, IEEE*, Huan Wang, Xiaou Tang, *Senior Member, IEEE*, Jianzhuang Liu, *Senior Member, IEEE*, and Thomas S. Huang

**Abstract**—In this paper, we investigate two soft-biometric problems: 1) age estimation and 2) pose estimation, within the scenario where uncertainties exist for the available labels of the training samples. These two tasks are generally formulated as the automatic design of a regressor from training samples with uncertain nonnegative labels. First, the nonnegative label is predicted as the Frobenius norm of a matrix, which is bilinearly transformed from the nonlinear mappings of a set of candidate kernels. Two transformation matrices are then learned for deriving such a matrix by solving two semidefinite programming (SDP) problems, in which the uncertain label of each sample is expressed as two inequality constraints. The objective function of SDP controls the ranks of these two matrices and, consequently, automatically determines the structure of the regressor. The whole framework for the automatic design of a regressor from samples with uncertain nonnegative labels has the following characteristics: 1) the SDP formulation makes full use of the uncertain labels, instead of using conventional fixed labels; 2) regression with the Frobenius norm of matrix naturally guarantees the nonnegativity of the labels, and greater prediction capability is achieved by integrating the squares of the matrix elements, which to some extent act as weak regressors; and 3) the regressor structure is automatically determined by the pursuit of simplicity, which potentially promotes the algorithmic generalization capability. Extensive experiments on two human age databases: 1) FG-NET and 2) Yamaha, and the Pointing'04 head pose database, demonstrate encouraging estimation accuracy improvements over conventional regression algorithms without taking the uncertainties within the labels into account.

**Index Terms**—Semidefinite programming, soft biometrics, uncertain labels.

## I. INTRODUCTION

**H**ARD BIOMETRICS, such as the face/human identification from face, iris, fingerprint, and gait, have attracted a lot of attention over the past decades. Many commercial sys-

Manuscript received November 01, 2007; revised January 30, 2008. Current version published November 19, 2008. This work was supported by the U.S. Government VACE program. The views and conclusions are those of the authors and not of the U.S. Government or its agencies. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anil Jain.

S. Yan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: eleyans@nus.edu.sg).

H. Wang is with the Department of Computer Science, Yale University, New Haven, CT 06520-8234 USA (e-mail: huan.wang@yale.edu).

X. Tang and J. Liu are with the Department of Information Engineering, Chinese University of Hong Kong, Hong Kong, China (e-mail: jzliu,xtang@ie.cuhk.edu).

T. S. Huang is with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: huang@ifp.uiuc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2008.2006585

tems have been developed based on different biometric features. Compared with the hard-biometric tasks, soft-biometric tasks, such as age estimation and pose estimation which do not predict deterministic identity but output certain human characteristics, were much less studied, and the current techniques for these tasks are still far from real application. These two tasks, especially the pose estimation, for the labels of the data can be considered continuous quantities; hence, they are essentially regression problems.

A face image encodes different types of useful information, such as identity, expression, gender, age, and pose. It is commonly believed that humans can provide satisfying and consistent ground truths about the identity, expression, and gender, by clues from the speech, hair style, and costume. But for age and pose, the ground truths labeled by different individuals are often far from consistent due to the complex effects of living conditions, cosmetics, personal specialties, gender differences, facial geometry, and so on. Three examples of pose labeling from 13 different participants are shown in Fig. 1, and large variations are observed among the labels from different participants [22]. The average standard deviation for all of the evaluated images is  $5.73^\circ$  as reported in [22].

Despite the fact that age is an important characteristic in humans, only a few works have been dedicated to the problem of age estimation [10], [11]. The latest work is from Geng *et al.* [7], based on the statistical modeling of aging patterns. All of these algorithms require the age label to be a fixed value. However, it is often difficult to acquire the accurate ages in real applications; instead, only rough age ranges are often obtained. Moreover, even if the age of a certain person is labeled as a fixed value, such as 30, the actual age can be any real value within the interval [30 31]. Hence, it makes more sense for the age to be finally expressed as an interval instead of a fixed value.

Head pose estimation has many useful applications, such as gaze detection, safe driving, automouse in a large screen, etc. Much research [6], [19] has been dedicated to this specific problem, and the latest work was proposed by Nallure *et al.* [16], based on biased manifold learning. All of these algorithms cannot directly handle the scenario with uncertain labels. Although the pose labels can be obtained from hardware devices in certain experimental scenarios, it is still desirable and of great application importance to propose a new formulation for the automatic design of a regressor based on training samples with uncertain labels.

A natural question is whether we can gain more from the interval labels by designing a new learning framework. Our answer to this question is positive. In this paper, the estimation of age or pose information is considered as a nonlinear regres-

TABLE I  
A SUMMARY OF THE DATABASE SIZE, AGE PRECISION, AND MAIN FEATURES USED FOR DIFFERENT AGE ESTIMATION ALGORITHMS.

Algorithm	Total Images	Subjects	Precision	Features
Kwon [11]	47	-	Three groups	Geometric and wrinkle features
Hayashi [10]	300	-	5y/group	Geometric and wrinkle features
Lanitis [12]	400	40	0-35y	Active Appearance Models [5]
Geng [8]	1002	82	0-69y	Active Appearance Models [5]
This work	8,000	1,600	0-93y	Gray-Level Intensities

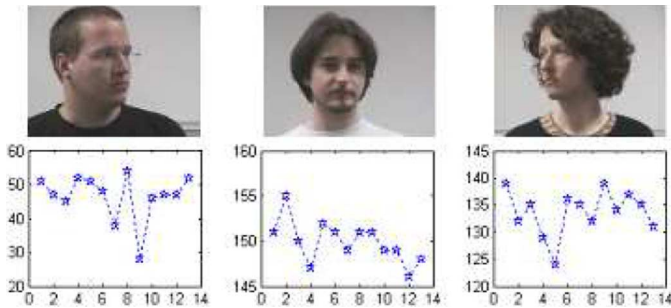


Fig. 1. Estimated pose labels of the three images in [8] from 13 different observers by rotating a 3-D head model. We can see that large standard deviations exist for these labeled ground truths [22].

sion problem based on training samples with uncertain nonnegative labels. The nonlinearity from the input image features to the label is embodied with the kernel trick [15], and a set of kernels is used for designing a regressor. The age and pose values can be considered nonnegative and with a lower bound of zero, and we compute the age as the Frobenius norm of a dimension-flexible matrix. Its advantage over direct linear combination of features or kernels for regression is that it provides a flexible way to integrate a set of weak regressors, namely the square outputs of the matrix elements, for a better approximation. This flexible matrix is bilinearly transformed from the mappings of the candidate kernels. The learning of these two transformation matrices is formulated as a semidefinite programming (SDP) problem [3]. More specifically speaking, the uncertain label, namely the age or pose interval, of each sample is expressed as two inequality constraints in the SDP formulation, and its objectives function is used to pursue a regressor with a simple structure, which potentially promotes the algorithmic generalization capability.

The rest of this paper is organized as follows. We present the backgrounds and motivations of this work in Section II. Section III introduces the details of the SDP formulation for automatic designing of a regressor based on training samples with uncertain nonnegative labels. Its relationship with other SDP problems as well as traditional regression algorithms is discussed in Section IV. Section V provides comparison experiments on two human age databases and one pose database, and the concluding remarks are given in Section VI.

## II. BACKGROUNDS AND MOTIVATIONS

In this section, we discuss the backgrounds of age estimation and pose estimation, and also the motivations for our work of learning autostructured regressors from uncertain nonnegative labels.

### A. Background and Motivation From Age Estimation

Kwon and Lobo [10] presented a method for age classification based on the craniofacial development theory and skin wrinkle analysis, with human faces finally classified into three groups: 1) babies, 2) young adults, and 3) senior adults. Hayashi *et al.* [9] proposed classifying age and gender based on the wrinkle and geometry relationships between different parts of a face, with the human ages divided into multiple groups at the five-year intervals. Lanitis *et al.* [11] utilized active appearance models [4] to extract the combined shape and intensity parameters, and then compared the age estimation accuracies of the algorithms, including simple quadratic fitting, shortest distance classifier, and the neural network. Recently, Geng *et al.* [7] proposed to conduct age estimation by modeling the statistical properties of aging patterns, namely, a sequence of personal facial age images, based on the assumption that multiple images of different ages are available for each person. Besides the age estimation problem, many other problems related to human age were also studied in the computer vision literature. Lanitis *et al.* [12] and Ramanathan and Chellappa [18] studied the problem of aging simulation on faces. Ramanathan and Chellappa [17] proposed a Bayesian age-difference classifier built on a probabilistic eigenspace framework, and face verification was conducted in an age-invariant manner.

Research on age estimation is still at a preliminary stage, and most works are based on small databases with coarse age precision. A detailed comparison of the previous algorithms along with our solution is summarized in Table I. It shows that the largest database ever used consists of 1 002 face images from only 82 subjects, which is far from statistical importance given the high variety of face images. In this paper, we tackle the age estimation problem on a more challenging face database with 8 000 images, 1 600 subjects, and ages ranging from 0 to 93.

The motivation of this work is from the observation on the age labels of face images. The age labels are nonnegative and ordinal, and are often uncertain for some databases, while these characteristics are rarely exploited by previous algorithms. Traditional algorithms for age estimation can be roughly classified into two categories. The first category divides all possible ages into multiple groups, and then considers the age estimation problem as a general multiclass classification problem. The methods in [9] and [10] belong to this category. They ignore the ordinal information of the age labels and can only predict the age at a coarse precision. The second category considers age estimation as a regression problem and current techniques require the age label to be a fixed value in the model training stage, not an interval. However, when we obtain the facial images, especially multiple images of different ages from the same person, it is often difficult to acquire the accurate ages, instead only rough

age ranges are available. On the other hand, as mentioned before, the age is commonly recorded as an integer value, but the truth is that when we say a person is age 30, we mean that the age is possibly within the interval  $[30\ 31)$ . Hence, it is more reliable for the age to be finally expressed as an interval instead of a fixed value. Traditional regression algorithms cannot directly solve the problem with uncertain labels, and they may also output negative ages; hence, it is desirable to propose a new formulation for the age estimation problem.

### B. Backgrounds and Motivations From Pose Estimation

Similar to age estimation, current research on appearance-based head pose estimation can also be roughly divided into two categories. The first category [13], [14] formulates pose estimation as a conventional multiclass pattern recognition problem, and only rough pose information is inferred from these algorithms. The second category takes pose estimation as a regression problem, and nonlinear regression algorithms (e.g., neural networks [20]) are used for learning the mapping from the original appearance features to the head pose label.

The motivation of this paper is from the observations on the methods to obtain the pose ground truths of the samples. For some databases, the ground truths are obtained from the estimations of several observers, and as shown in Fig. 1, a large variation often exists among the estimated labels from different observers. Another popular way to obtain the pose ground truths is displayed in Fig. 2 used for Pointing'04 [8]. All images of this database were taken using the FAME Platform of the PRIMA Team in INRIA Rhone-Alpes [8]. To obtain different poses, markers are put in the whole room, and each marker corresponds to a pose in two directions. Post-its are used as markers. The whole set of post-its covers a half-sphere in front of the person. In order to obtain the face in the center of the image, the person is asked to adjust the chair to see the device in front of him or her. After this initialization phase, the person is asked to stare successively at 93 Post-it marks, without moving his or her eyes [8]. This kind of system requires the precise cooperation of the users and, hence, the obtained pose labels are not so exact, which again motivates us to propose a new learning framework for pose estimation by taking the uncertainties of the pose labels into account.

### III. NONLINEAR REGRESSION WITH UNCERTAIN NONNEGATIVE LABELS

For the human age or pose estimation problem, the image set for model training is denoted here as a matrix  $X = [x_1, x_2, \dots, x_N]$ ,  $x_i \in \mathbb{R}^m$ , where  $N$  is the image number and  $m$  is the feature dimension. The uncertain nonnegative label for the image  $x_i$  is denoted as  $[l_i, L_i]$ , where  $l_i \geq 0$ . The task is to predict the nonnegative label of the new image  $x$ , and our solution to this general problem is as follows.

#### A. Problem Formulation

1) *Nonlinear Regression From Image Features to Nonnegative Label:* Generally, rough age or pose estimation can be conducted by separating all possible ages or poses into several groups, and then formulating them as a general multiclass classification problem. In this way, for a more precise estimate, a

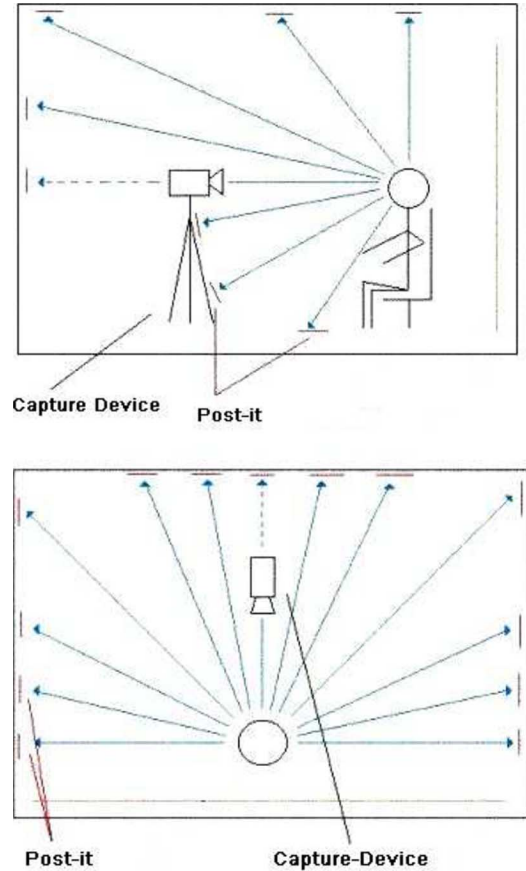


Fig. 2. Pose label capture system used in [8] (with some modifications on labels).

greater number of classes will be used and, consequently, more samples will be required for learning a reliable estimator. In this paper, we take the age/pose estimation problem as a nonlinear regression problem, and the nonlinear formulation is based on the kernel trick [15].

Assume that we have a set of kernel mapping functions, denoted as  $\{\phi^1(x), \phi^2(x), \dots, \phi^n(x)\}$ , where  $n$  is the number of kernels, and  $\phi^o(x) : \mathbb{R}^m \rightarrow \mathfrak{F}^o$ ,  $o = 1, 2, \dots, n$ , is the kernel mapping function with  $\mathfrak{F}^o$  as the higher or infinite dimensional Hilbert space and the corresponding kernel function  $k^o(x, y) = \langle \phi^o(x), \phi^o(y) \rangle$ . Meanwhile, let the combined kernel mapping function be  $\phi(x) = [\phi^1(x)^T, \phi^2(x)^T, \dots, \phi^n(x)^T]^T$ .

In this paper, we model the nonnegative label as the Frobenius norm of a matrix, and specifically, we have

$$a = \|W^T(\phi(x) \odot V)\|^2 \quad (1)$$

where  $V \in \mathbb{R}^{n \times n'}$  is the matrix to give different weights for different kernels, and the operator  $\odot$  is defined as

$$\phi(x) \odot V = \begin{bmatrix} \phi^1(x)V_{11} & \phi^1(x)V_{12} & \cdots & \phi^1(x)V_{1n'} \\ \phi^2(x)V_{21} & \phi^2(x)V_{22} & \cdots & \phi^2(x)V_{2n'} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^n(x)V_{n1} & \phi^n(x)V_{n2} & \cdots & \phi^n(x)V_{nn'} \end{bmatrix}.$$

The symbol  $W$  is the transformation matrix that transforms the feature from the higher dimensional feature space

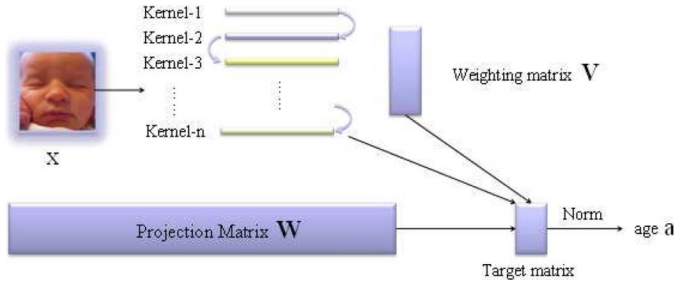


Fig. 3. Illustration of the process from the original image datum to the final label.

into a lower dimensional one, and it is represented as the linear combination of  $\{\phi(x_i), i = 1, 2, \dots, N\}$ , that is,  $W = [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]U$  where  $U \in \mathbb{R}^{N \times N'}$ . The number  $n'$  and  $N'$  are the expected column numbers of the matrix  $V$  and  $W$ , and are automatically determined as described afterward. The whole flowchart from the original image data to final labels is illustrated in Fig. 3.

Here, for each sample  $x_i$ , we define a data-specific kernel matrix  $K_{x_i} \in \mathbb{R}^{N \times n}$  as

$$K_{x_i}(j, o) = k^o(x_i, x_j) \quad (2)$$

and for an image  $x$ , this kernel matrix is written as  $K_x$ . Then, the function in (1) can be rewritten as (3), shown at the bottom of the page, that is

$$a = \|U^T K_x V\|^2. \quad (4)$$

This formulation of the prediction function is much simpler than that in (1).

2) *Discussion*: Why do we use the Frobenius norm of a flexible matrix for approximating a label? The reasons are twofold. On the one hand, the labels commonly have lower bounds and can be assumed to be nonnegative after a translation, say  $[0, +\infty)$  for age and  $[0, 180]$  for pose in this work, and the norm operator can naturally guarantee the nonnegative property. On the other hand, (4) is equal to  $\sum_{i=1}^{N'} \sum_{j=1}^{n'} \|U_i^T K_x V_j\|^2$ , where  $\|U_i^T K_x V_j\|^2$  can be considered weak regressors and, hence, the sum of these weak regressors may bring greater approximation capability than a single weak regressor.

3) *Uncertain Labels to Inequality Constraints*: According to (4), the uncertain nonnegative label of the sample  $x_i$  (i.e.,  $[l_i L_i]$ ) can be expressed as the two inequalities

$$\|U^T K_{x_i} V\|^2 \leq L_i, \quad (5)$$

$$-\|U^T K_{x_i} V\|^2 \leq -l_i. \quad (6)$$

4) *Objective Function: Avoiding Overfitting and Pursuing Simplicity*: As described before,  $\|U_i^T K_x V_j\|^2$  can be considered a weaker regressor. A reasonable way to reduce the possibility of overfitting and promote the algorithmic generalization capability is to control the rank of the transformation matrices  $U$  and  $V$  and, consequently, reduce the number of weak regressors. Moreover, there may be infinite solutions that satisfy all of the constraints in (5) and (6). It is desirable to provide a criterion for guiding the selection of the optimal solution and pursuing lower ranks of the parameter matrices is a feasible strategy.

We will present a method for controlling the ranks of the transformation matrices based on sparsity property. Similar to the sparse support vector machine [2] method, our algorithm employs the  $L_1$  norm to control the sparsity of the parameters.

As we expect to control the ranks of the transformation matrices, we do not compute the matrix  $U$  and  $V$  directly; instead, we compute the matrices

$$S^u = UU^T, \quad S^v = VV^T. \quad (7)$$

Here, we take  $S^u$  as an example to demonstrate how to control the matrix rank. Take the singular value decomposition of  $S^u$  as

$$S^u = U_r \Lambda^u U_r^T \quad (8)$$

where  $U_r \in \mathbb{R}^{N \times N}$  is a square orthogonal matrix and  $\Lambda^u = \text{diag}\{\lambda_1^u, \lambda_2^u, \dots, \lambda_N^u\}$ . Then, controlling the rank of the matrix  $U$  is equivalent to controlling the sparsity of the diagonal elements of matrix  $\Lambda^u$ . Enlightened by the Sparse SVM, we simply minimize the  $L_1$  norm of the diagonal elements of  $\Lambda^u$ , that is

$$\min_{S^u} \sum_{i=1}^N |\lambda_i^u| = \min_{S^u} \text{Tr}(S^u) \quad (9)$$

where  $\text{Tr}(\cdot)$  is the trace of a square matrix. The equality is satisfied owing to the fact that all  $\lambda_i^u$ 's from the positive semidefinite matrix  $S^u$  are nonnegative.

*Remark*:  $\text{Tr}(S^u)$  can also be considered as the  $L_2$  norm of  $(\tau_1^u, \tau_2^u, \dots, \tau_N^u)$ , which are the singular values of  $U$  with  $\lambda_i^u = \tau_i^{u2}$ . But as discussed later, the optimization problem takes  $S^u$  and  $S^v$  as variables directly, and the constraints are all linear inequalities. Hence, it is reasonable to control the sparsity of  $(\lambda_1^u, \lambda_2^u, \dots, \lambda_N^u)$  directly. Similarly, we can control the rank of the matrix  $V$  by minimizing  $\text{Tr}(S^v)$ .

According to the definition in (7), the lefthand items in constraints (5) and (6) can be expressed by  $S^u$  and  $S^v$  as

$$\begin{aligned} \|U^T K_{x_i} V\|^2 &= \text{Tr}(U^T K_{x_i} V V^T K_{x_i}^T U) \\ &= \text{Tr}(K_{x_i} V V^T K_{x_i}^T U U^T) \\ &= \text{Tr}(K_{x_i} S^v K_{x_i}^T S^u) \end{aligned} \quad (10)$$

$$\left\| U^T [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]^T \begin{bmatrix} \phi^1(x) V_{11} & \phi^1(x) V_{12} & \cdots & \phi^1(x) V_{1n'} \\ \phi^2(x) V_{21} & \phi^2(x) V_{22} & \cdots & \phi^2(x) V_{2n'} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^n(x) V_{n1} & \phi^n(x) V_{n2} & \cdots & \phi^n(x) V_{nn'} \end{bmatrix} \right\|^2 = \|U^T K_x V\|^2 \quad (3)$$

where the first equality is obtained from the fact that  $\|A\|^2 = \text{Tr}(AA^T)$  for any matrix  $A$ , and the second equality is obtained from the fact that  $\text{Tr}(AB) = \text{Tr}(BA)$  for any matrices  $A$  and  $B$  with proper dimensions.

Based on the aforementioned constraints and the objective function, the regression problem with uncertain nonnegative labels can be formally defined as

$$\begin{aligned} (S^u, S^v)^* &= \arg \min_{S^u, S^v} \text{Tr}(S^u) + \text{Tr}(S^v), \text{ s.t.} \\ 1: S^u &\succeq 0, S^v \succeq 0; \\ 2: \text{Tr}(K_{x_i} S^u K_{x_i}^T) &\leq L_i, \quad i = 1, 2, \dots, N; \\ 3: \text{Tr}(-K_{x_i} S^v K_{x_i}^T) &\leq -l_i, \quad i = 1, 2, \dots, N \end{aligned} \quad (11)$$

where  $S^u \succeq 0$  and  $S^v \succeq 0$  mean that  $S^u$  and  $S^v$  are positive semidefinite.

In this problem, the objective function is convex, yet the feasible solution set is possibly nonconvex; hence, it is essentially a nonconvex optimization problem and, consequently, a closed-form solution does not exist. Naturally, we present a procedure to optimize  $S^u$  and  $S^v$  iteratively, and in each step, the problem is then converted into a convex optimization problem. The semidefinite programming toolbox can be applied for the step-wise optimization.

### B. Iterative Parameter Optimization

Iterative optimization along different axes is very common in nonconvex optimization literature. Here, we solve the optimization problem with respect to  $(S^u, S^v)$  by iteratively optimizing one parameter matrix while fixing the other one.

For the given  $S^v$ , the constraints 2–3 in the optimization problem (11) are changed to

$$\begin{aligned} \text{Tr}(C_i^u S^u) &\leq L_i, \quad i = 1, 2, \dots, N \quad (12) \\ \text{Tr}(-C_i^u S^u) &\leq -l_i, \quad i = 1, 2, \dots, N \quad (13) \end{aligned}$$

where  $C_i^u = K_{x_i} S^v K_{x_i}^T$ .

With the aforementioned constraints and the objective function in (11), the matrix  $S^u$  can be obtained by optimizing a semidefinite programming problem as shown in Algorithm 1. The objective function in Algorithm 1 is convex, and the optimization does not suffer from the local optimum issue [21]. There are several general-purpose toolboxes and polynomial-time solvers available for solving the semidefinite programming problem. In this paper, we utilize the solver SeDuMi and the CSDP 4.9 toolbox in MATLAB [3]. The computational complexity of semidefinite programming depends on the implementation, and more details are referred to in [3].

---

#### Algorithm 1 Procedure to learn matrix $S^u$

---

- 1) Inputs: image set  $X$  along with the uncertain nonnegative labels  $[l_i, L_i], i = 1, 2, \dots, N$ , and the matrix  $S^v$ .
- 2) Solve the semidefinite programming problem defined below. Minimize  $\text{Tr}(S^u)$ .
  - 1:  $S^u \succeq 0$ ;
  - 2:  $\text{Tr}(C_i^u S^u) \leq L_i, i = 1, 2, \dots, N$ ;

$$3: \text{Tr}(-C_i^v S^u) \leq -l_i, \quad i = 1, 2, \dots, N.$$

3) Output the matrix  $S^u$ .

Similarly, for the given  $S^u$ , the constraints 2–3 in the optimization problem (11) are changed to

$$\begin{aligned} \text{Tr}(C_i^v S^v) &\leq L_i, \quad i = 1, 2, \dots, N \quad (14) \\ \text{Tr}(-C_i^v S^v) &\leq -l_i, \quad i = 1, 2, \dots, N \quad (15) \end{aligned}$$

where  $C_i^v = K_{x_i}^T S^u K_{x_i}$ .

Then, the optimization problem in (11) is converted into a standard semidefinite programming problem as listed in Algorithm 2. Similarly, it can be solved with the general-purpose toolbox for the SDP problem.

Algorithms 1 and 2 are iteratively conducted to obtain the stepwise result  $(S_t^u, S_t^v)$  until the following stop criteria are satisfied:

$$\begin{cases} \|S_t^u - S_{t+1}^u\| < \varepsilon N^2 \\ \|S_t^v - S_{t+1}^v\| < \varepsilon n^2 \end{cases} \quad (16)$$

where  $\varepsilon$  is a manually defined threshold and is empirically set to be  $10^{-6}$  in this paper. Smaller  $\varepsilon$  may result in smaller rank for  $S^u$  and  $S^v$ , but it has little affection on the final regression results since the weights for the extra components will be very small.

*1) Convergency Discussion:* The optimization problem in (11) is nonconvex due to the nonconvexity of the feasible solution set and, hence, we cannot guarantee that the solution will be globally optimal. Here, instead we prove that the iterative algorithm will converge to a local optimum. Denote the objective function as  $F(S^u, S^v) = \text{Tr}(S^u) + \text{Tr}(S^v)$ , then we have

$$F(S_t^u, S_t^v) \geq F(S_{t+1}^u, S_t^v) \geq F(S_{t+1}^u, S_{t+1}^v). \quad (17)$$

Therefore, the objective function is nonincreasing, and we have  $f(S^u, S^v) \geq 0$ , which means that the objective function has a lower bound. Then, we can conclude that the objective function will converge to a local optimum.

After the convergence of the iterative procedure, the transformation matrices  $U$  and  $V$  can be obtained from the singular value decomposition of the obtained matrices  $S^u$  and  $S^v$

$$S^u = U_r \text{diag}\{\lambda_1^u, \lambda_2^u, \dots, \lambda_{N'}^u, 0, \dots, 0\} U_r^T \quad (18)$$

$$S^v = V_r \text{diag}\{\lambda_1^v, \lambda_2^v, \dots, \lambda_{n'}^v, 0, \dots, 0\} V_r^T \quad (19)$$

where  $\lambda_i^u > 0$  and  $\lambda_i^v > 0$ . We then have

$$U = U_r(:, 1 : N') \text{diag}\{\lambda_1^u, \lambda_2^u, \dots, \lambda_{N'}^u\}^{1/2} \quad (20)$$

$$V = V_r(:, 1 : n') \text{diag}\{\lambda_1^v, \lambda_2^v, \dots, \lambda_{n'}^v\}^{1/2} \quad (21)$$

where  $U_r(:, 1 : N')$  means the submatrix consisting of the left  $N'$  column vectors of the matrix  $U_r$  and similarly  $V_r(:, 1 : n')$  is the submatrix consisting of the left  $n'$  column vectors of the matrix  $V_r$ .

*Remark:* The column numbers of the matrix  $U$  and  $V$  are automatically determined and, consequently, the number of weak regressors as well as the structure of the regressor are determined in an automatic manner.

### ALGORITHMIC ANALYSIS

In this section, we analyze the relationship between the traditional regression formulation and that defined by SDP in this paper. We then compare our algorithm with other SDP based algorithms.

---

#### Algorithm 2 Procedure to learn matrix $S^v$

---

- 1) Inputs: image set  $X$  along with the uncertain nonnegative labels  $[l_i, L_i]$ ,  $i = 1, 2, \dots, N$ , and the matrix  $S^u$ .
- 2) Solve the semidefinite programming problem defined below.  
Minimize  $Tr(S^v)$ .  
1:  $S^v \succeq 0$ .  
2:  $Tr(C_i^v S^v) \leq L_i$ ,  $i = 1, 2, \dots, N$ .  
3:  $Tr(-C_i^v S^v) \leq -l_i$ ,  $i = 1, 2, \dots, N$ .
- 3) Output the matrix  $S^u$ .

#### C. Traditional Regressor versus SDP Formulation

A direct approach for age or pose estimation is to design the proper regressor. In [11], the relationship between the label and the features is modeled with quadratic regression, namely

$$a = c + w_1^T x + w_2^T (x * x) \quad (22)$$

where  $x$  and  $(x * x)$  are the vectors containing the image features and the squares of the features, respectively;  $c$ ,  $w_1$ , and  $w_2$  are parameters to learn. This simple regression method works well for simple quadratic fitting problems, but for a complex nonlinear regression problems, such as age and pose estimation, it cannot produce a satisfactory fit.

Another popular regression algorithm is multilayer perceptrons (MLP) with back propagation learning [20]. It has been widely applied in various applications, such as face detection and recognition. Yet, the structure of MLP need be manually determined; the tuning of the structure is often time consuming and needs an extra validation data set.

Although simple quadratic regression and MLP can be used for age or pose estimation, they cannot directly handle regression problems with uncertain labels similar to what our proposed SDP-based formulation does. Another disadvantage of quadratic regression and MLP is that the output from them can be negative, which is inconsistent with the nonnegative assumption in this paper.

#### D. Other Problems Formulated With SDP

In the past few years, the optimization tool SDP has been used for problem formulation in several works. Weinberger and Saul [21] formulated the manifold embedding task as a semidefinite programming problem and provided a new perspective beyond spectral analysis for manifold learning. The sparse principal component analysis problem was also formulated with SDP as demonstrated in [1]. Though our proposed algorithm is also formulated as an SDP problem, the objective function and the constraints in our algorithm are unique. Our proposed algorithm

utilizes SDP as a tool for formulating the nonlinear regression problem with uncertain nonnegative labels.

## IV. EXPERIMENTS

In this section, we first introduce the implementation details of our algorithm for nonlinear regression with uncertain nonnegative labels, referred to as RUN hereafter. Then, the superiority of uncertain labels over fixed labels is justified with a toy problem; and the algorithmic convergency property is verified by the experiments on the FG-NET<sup>1</sup> database. Finally, the human age databases—FG-NET and Yamaha databases—and Pointing'04 pose database are used to systematically evaluate the effectiveness of the RUN algorithm in estimation accuracy in comparison with the state-of-the-art algorithms [7].

### A. Implementation Details

---

#### Algorithm 3 Procedure to learn matrix $S^u$ with Relaxation.

---

- 1) Inputs: image set  $X$  along with the uncertain nonnegative labels  $[l_i, L_i]$ ,  $i = 1, 2, \dots, N$ , and the matrix  $S^v$ .
- 2) Solve the semidefinite programming problem defined below.  
Minimize  $Tr(S^u) + \gamma \sum_{i=1}^N (\epsilon_i^1 + \epsilon_i^2)$ .  
1:  $S^u \succeq 0$ ,  $\epsilon_i^1 \geq 0$ ,  $\epsilon_i^2 \geq 0$ ,  $i = 1, 2, \dots, N$ .  
2:  $Tr(C_i^u S^u) - \epsilon_i^1 \leq L_i^2$ ,  $i = 1, 2, \dots, N$ .  
3:  $Tr(-C_i^u S^u) - \epsilon_i^2 \leq -l_i^2$ ,  $i = 1, 2, \dots, N$ .
- 3) Output the matrix  $S^u$ .

In our implementation, two strategies are applied to facilitate the RUN algorithm. It is possible that not all of the constraints in Algorithm 1 and 2 can be satisfied in real applications and, hence, we add relaxation parameters to ensure that the feasible solution set is not empty. Following [3], we take Algorithm 1 as an example to introduce how to add relaxation parameters, and the details are listed in Algorithm 3.

As described in Section II-A, the column vectors of the matrix  $W$  lie within the space spanned by the kernel mappings of all training samples. Thus, the size of  $U$  increases along with the growth of the training set and, consequently, the size of  $S^u$  will be very large. To improve the scaling capability, we constrain the column vectors of  $W$  to be the combination of certain prototypes from the training set, namely  $W = [\phi(x_1^p), \phi(x_2^p), \dots, \phi(x_M^p)]U$ , where  $x_i^p$  is the selected prototype,  $U \in \mathbb{R}^{M \times N'}$ , and  $M$  is the prototype number. We conduct the  $K$ -means algorithm for clustering the training samples, and then the samples close to the cluster centers are selected as prototypes. For all of the age estimation experiments, the prototype number is set as 400, and the parameter  $\gamma$  in Algorithm 3 is set to 1. Note that generally more prototypes are used and better performance can be achieved; and greater  $\gamma$  means more strict inequality constraints. The number 400 is selected for the tradeoff between accuracy and computational cost. The Gaussian kernels  $k_o(x, y) = \exp\{-\|x - y\|^2 / \delta_o^2\}$

<sup>1</sup>Online. Available: <http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm>.

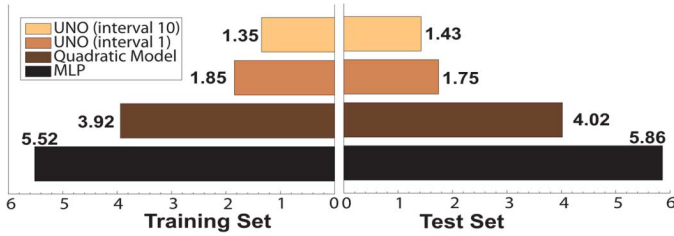


Fig. 4. MAEs of the RUN algorithms with an uncertain label interval of width 10 and width 1, QM, and MLPs on the toy data.

are applied as the kernel candidates and we use four kernels with parameters  $\delta_o = 2^{(o-10)/2.5\delta}$ ,  $o = 0, 1, 2, 3$  in all of the experiments, where  $\delta$  is the standard deviation of the sample data. For the pose estimation experiments, the prototype number is set as 40, and the other parameters are the same as those for age estimation.

In this work, RUN is compared with the traditional regression algorithms quadratic models (QM) and supervised neural networks [11]. According to [7], the least squares fit (LSF) optimization algorithm commonly gives better performance than the genetic algorithm; thus, the least-square fit optimization is utilized for the quadratic models. For the neural networks, we adopt the same configuration as in [12] (i.e., MLP with the back propagation learning algorithm) and the network architecture and parameters are also set the same as in [12].

1) *Estimation Output*: For an image  $x$ , its estimated age is output as  $\hat{a} = \|U^T K_x V\|$ .

2) *Accuracy Measurements*: Two measures are used to evaluate algorithmic performance. The first one is the mean absolute error (MAE) criterion used in [7], [11], and [12]. MAE is defined as the average of the absolute errors between the estimated labels and the ground truth labels, i.e.,

$$\text{MAE} = \frac{\sum_{i=1}^{N_t} |\hat{a}_i - a_i|}{N_t}$$

where  $\hat{a}_i$  is the estimated age for the  $i$ th testing sample,  $a_i$  is the ground truth age, and  $N_t$  is the number of testing images. Another popular measure is the cumulative score [7]

$$\text{CumScore}(\theta) = \frac{N_{e \leq \theta}}{N_t} \times 100\%$$

where  $N_{e \leq \theta}$  is the number of samples in which the estimator makes an absolute error of not higher than  $\theta$ .

### B. Toy Data: Effectiveness of Uncertain Labels

To examine the effectiveness of uncertain labels, we introduce a toy example where we know the exact labels. In this toy problem, the feature dimension is 2, and the underlying relationship between the data and the label is

$$l(x) = l(x_1, x_2) = 100 \times \cos(\|x\|) + 100 \times \exp(-\|x\|^2)$$

where  $x_1$  and  $x_2$  independently follow the uniform distribution on the unit interval  $[0, 1]$ , and  $l(x)$  is the exact label from data  $x$  with  $l(x) > 0$ .

To obtain the uncertain labels, the observed label is assumed to be  $l^n(x) = l(x) + \varepsilon$ , where  $\varepsilon$  is random noise evenly dis-



Fig. 5. Sample aging images from one person in the FG-NET aging database.

tributed within the interval  $[-5, 5]$  such that  $l^n(x) > 0$ . Two-hundred data samples are randomly sampled as the training set and the labels with noise are used for model training; also, another 200 data are sampled for testing, and the exact labels are used for computing training and testing accuracy. For comparison, the results from QM and MLP are also reported.

For our proposed RUN algorithm, we compare two types of uncertainty scales: one is the uncertain labels with an interval of width 10, namely,  $[l^n(x) - 5, l^n(x) + 5]$ , and the other is the uncertain labels with a smaller interval of width 1, namely  $[l^n(x) - 0.5, l^n(x) + 0.5]$ , which is very close to the fixed labels. In these experiments, 100 prototypes are used, the kernel number and  $\gamma$  are set as for age estimation experiments. Typically, it takes about 70 s for one iteration on a computer with a 2.8-GHz CPU and 2-GB memory. The comparison results are displayed in Fig. 4, from which we have two observations: 1) RUN is better than the other two regression algorithms in label estimation accuracy and 2) the uncertain labels which consider the noise of the labels can further promote estimation accuracy of the RUN algorithm.

### C. Databases for Age and Pose Estimation

Two aging face databases are used in our experiments. One is the FG-NET aging database<sup>2</sup>, which contains 1002 face images of 82 persons with ages ranging from 0 to 69. Some sample images of a person are displayed in Fig. 5. The evaluation method for the FG-NET database is the leave-one person out (LOPO).

The other age database, Yamaha<sup>3</sup>, contains 8000 Japanese facial images of 1600 people in age ranging from 0 to 93. Each person has five images and the Yamaha database is divided into two parts with 4000 images from 800 males and another 4000 images for 800 females. Our experiments are carried out separately on female and male subsets. For each subset, 1000 images are randomly selected for model training while the remaining 3000 samples are used for testing. To the best of our knowledge, Yamaha is the largest aging database ever reported.

For the FG-NET database, each person has multiple images of different ages and, hence, many other algorithms such as aging patterns subspace (AGES) [7] and weighted appearance specific (WAS) [12] are applicable. For comparison, we use the same feature set as in [7] for the FG-NET database. First, the first 200 appearance parameters [5] based on the 68 key facial points are used as input for age estimation. For detailed information on

<sup>2</sup>Online. Available: <http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm>.

<sup>3</sup>To protect the portrait rights of the participants, sample images of the Yamaha face database are not shown here.

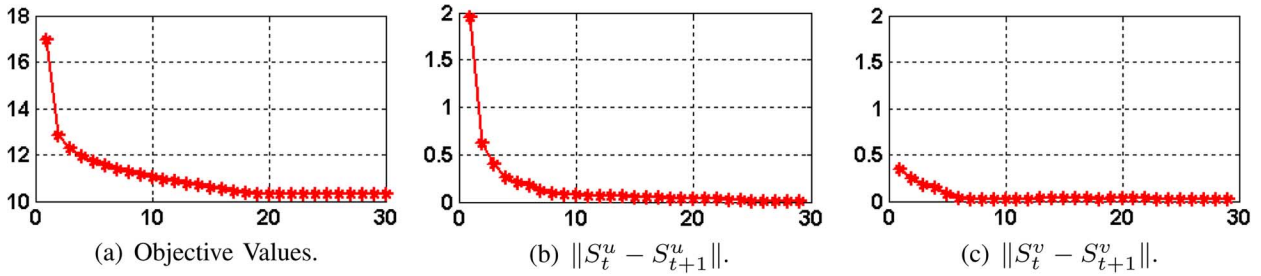


Fig. 6. Convergence of the objective function  $S^u$  and  $S^v$ . (a) Objective value versus iteration number. (b)  $\|S_t^u - S_{t+1}^u\|$  versus iteration number. (c)  $\|S_t^v - S_{t+1}^v\|$  versus the iteration number.

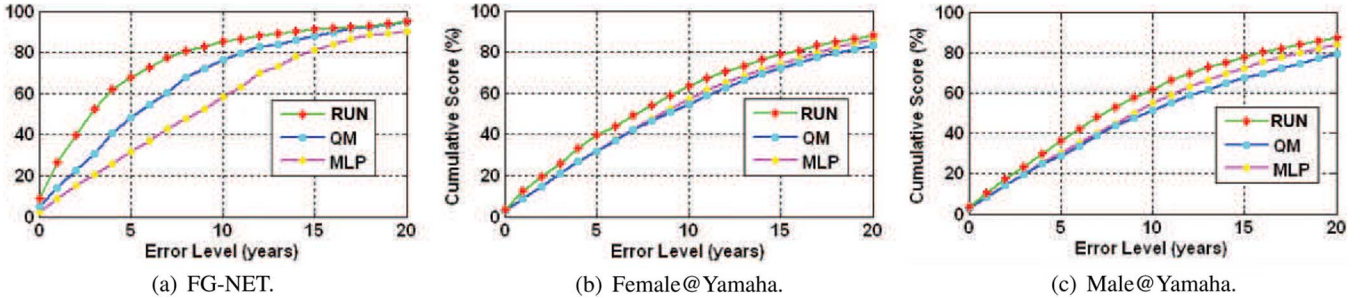


Fig. 7. Cumulative scores of the age estimation for QM, MLP, and RUN at error levels from 0 to 20 years for different data sets.

shape, texture, and appearance parameters, please refer to [4]. For the Yamaha database, the positions of the key facial points are not provided and, hence, the original gray-level values are used instead as features by normalizing the images to size  $64 \times 64$  pixels and fixing the locations of the two eyes.

The Pointing'04 head pose database [8] consists of 15 subjects. The nose tips are manually marked, and we crop the faces to the size of  $64 \times 64$  pixels. In our experiments, 52 images from four subjects are used, and 13 observers are invited for labeling the poses by rotating a 3-D face head[22].<sup>4</sup> As shown in Fig. 1, uncertainty exists in the labels from different observers.

#### D. Convergency Justification

In this subsection, we systematically evaluate the convergency property of the RUN algorithm from three aspects: 1) the convergency of the objective function value; 2) the convergency of the transformation matrix  $S^u$ , which is characterized by  $\|S_t^u - S_{t+1}^u\|$ ; and 3) the convergency of the transformation matrix  $S^v$ , which is characterized by  $\|S_t^v - S_{t+1}^v\|$ .

The above properties are evaluated on the FG-NET database, and detailed results are shown in Fig. 6. In this experiment, we can see that the objective value monotonically decreases and converges together with the parameter matrix  $S^u$  and  $S^v$  after about 20 iterations.

#### E. Age Estimation Results

In this subsection, we conduct detailed age estimation experiments on the aforementioned age databases. According to [7], AGES achieves the best performance for age estimation, followed by the WAS [12]. Thus, we compare RUN with AGES and WAS for the experiments on the FG-NET aging database. For the RUN algorithm, the fixed age label is replaced with the

corresponding uncertain one. As mentioned before, when we say that the label is  $a_i$  (integer) for the sample  $x_i$ , his or her exact age can be any real value within the interval  $[a_i, a_i + 1)$  and, hence, for RUN, the uncertain label for sample  $x_i$  is set as  $[a_i, a_i + 1 - \epsilon]$ , where  $\epsilon$  is the minimal positive number that a computer can encode. Of course, a more reliable estimator can be achieved if we are able to obtain specific uncertainty information for each training image.

Fig. 7 displays the cumulative scores of different algorithms. Fig. 8 shows the comparison results of age estimation on FG-NET with only shape, texture, and the combined shape with texture. Fig. 9 displays the MAEs of different age groups for RUN with uncertain labels and RUN with fixed labels, and Table II lists the detailed MAEs of different algorithms. In our implementation, the vectors in the matrix  $U$  and  $V$  corresponding to the singular values of less than  $10^{-6}$  are removed. For the FG-NET database, the rank of the derived matrix  $U$  ranges from 9 to 33 with respect to different training sets in the LOPO strategy, and the derived matrix  $U$  takes the rank of 10 and 11 for the Yamaha female and male databases, respectively. The final rank of matrix  $V$  turns out to be 1 or 2 for these two aging databases.

From these results, several interesting observations are derived.

- 1) RUN reaches the lowest MAEs across both databases. On the FG-NET database, RUN brings an approximate 15% deduction of MAE compared with the state-of-the-art algorithm AGES [7]. Also, the uncertain labels bring extra accuracy improvement over fixed labels for the RUN algorithm as reported in Fig. 9 on the FG-NET database.
- 2) The Yamaha database is more challenging for age estimation than the FG-NET database, and the estimation accuracies from the evaluated algorithms are relatively lower than those in the FG-NET database.

<sup>4</sup>These data labels are collected for comparison research of man and machine in pose estimation capability.



TABLE II  
MAES OF DIFFERENT ALGORITHMS ON TWO DIFFERENT DATABASES AND OVER DIFFERENT AGE RANGES.  
NOTE THAT #SAMPLES MEANS THE NUMBER OF SAMPLES FOR EACH AGE GROUP OF THE FG-NET DATABASE

FG-NET					Female@Yamaha				Male@Yamaha			
Range	#Samples	RUN	QM	MLP	Range	RUN	QM	MLP	Range	RUN	QM	MLP
0-9	371	2.51	6.26	11.63	0-9	11.21	11.97	14.33	0-9	9.86	13.42	14.08
10-19	339	3.76	5.85	3.33	10-19	6.23	9.58	8.85	10-19	7.52	10.33	9.46
20-29	144	6.38	7.10	8.81	20-29	7.95	9.29	9.70	20-29	8.85	10.21	9.35
30-39	79	12.51	11.56	18.46	30-39	8.17	9.85	9.66	30-39	7.76	9.35	8.60
40-49	46	20.09	14.80	27.98	40-49	8.64	10.45	8.78	40-49	8.67	11.71	9.10
50-59	15	28.07	24.27	37.20	50-59	9.43	10.15	9.53	50-59	11.10	13.38	10.08
60-69	8	42.50	37.38	49.13	60-69	11.12	13.49	10.88	60-69	12.49	15.99	13.44
Average		<b>5.78</b>	7.57	10.39	70-93	15.56	19.66	16.52	70-93	16.60	20.44	19.69
Average	AGES: 6.77 [8]		WAS : 8.06 [8]		Average	<b>9.79</b>	11.80	11.03	Average	<b>10.36</b>	13.10	11.72

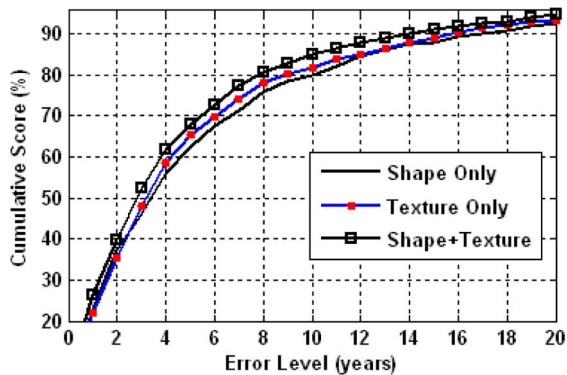


Fig. 8. Cumulative scores of the age estimation from RUN at error levels from 0 to 20 years on the FG-NET database for three cases: shape only, texture only, and combining shape with texture.

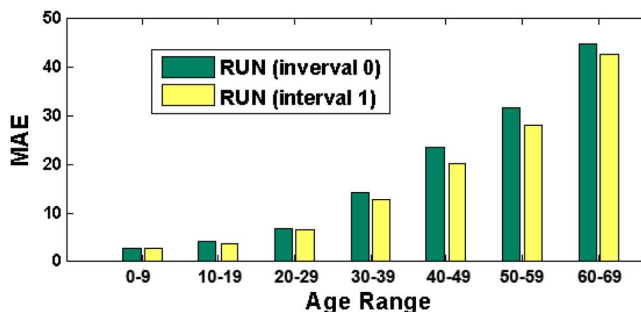


Fig. 9. MAEs for different age groups of the FG-NET database compared between RUN with an age interval of width 1 and RUN with interval of width 0, namely, fixed labels.

- 3) The average MAE from the QM-based estimator is about 27% lower than that of the average MLP-based estimator in the FG-NET database; while the MLP generally performs better than the QM in the Yamaha database.
- 4) The results in Fig. 8 show that shape information is also very important for age estimation, and its results are even almost as good as the results from texture only. One evidence for this observation may be that the facial contours of children are often more likely round than seniors. For the experiments on the FG-NET database, the facial positions are manually labeled and, hence, the results are reasonably good; but our offline experiments on the Yamaha database show that the current face alignment algorithm

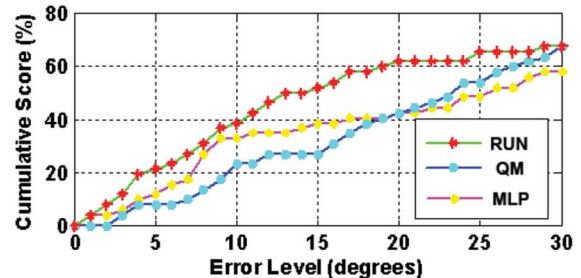


Fig. 10. Cumulative scores of the pose estimation for quadratic model (QM), MLPs, and RUN at error levels from 0 to 30°.

is still not good enough to provide accurate positions for shape-only-based age estimation.

#### F. Pose Estimation Result

For pose estimation, all algorithms are conducted 13 times, and each time, the labels from one observer are used for training QM and MLP, and the labels with the interval length as the standard deviation of the labels from all 13 observers are used for training the RUN algorithm. The leave-one-image-out experimental results are shown in Fig. 10, which illustrates that RUN is better than QM and MLP in accumulative scores for almost all error levels. The average MAE for RUN is 24.7°, which is much smaller than 27.1° for QR and 33.2° for MLP.

#### V. CONCLUSIONS AND FUTURE WORKS

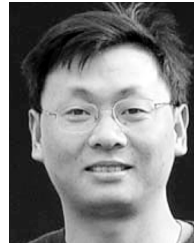
We have proposed in this paper a semidefinite programming formulation for automatically designing regressor based on training samples with uncertain nonnegative labels. Encouraging experimental results were achieved on two human aging databases, one of which is the largest one ever reported, and one pose database compared with state-of-the-art regression algorithms which did not take into account the uncertainty property of the labels.

Our proposed algorithm is general for regression problems with uncertain nonnegative labels, and we are planning to further investigate this algorithm from three aspects: 1) how to select the parameter  $\gamma$  in a rational way such that it acts as the weighting parameter in support vector machines (SVMs) [15]; 2) how to efficiently solve the SDP problem when the number

of prototypes is over 2000; 3) the experiments on the effectiveness of the shape + texture-based age estimation motivate us to design a better algorithm for more accurately locating the key feature points on the face and further improving the age estimation accuracy; and 4) person-independent age estimation may be achieved if we consider the identity information in designing regressors.

## REFERENCES

- [1] A. Aspremont, L. Ghaoui, M. Jordan, and G. Lanckriet, "A direct formulation for sparse pca using semidefinite programming," *Adv. Neural Inf. Process. Syst.* 17, pp. 41–48, 2005.
- [2] J. Bi, K. Bennett, M. Embrechts, C. Breneman, and M. Song, "Dimensionality reduction via sparse support vector machines," *J. Mach. Learning Res.*, vol. 3, pp. 1229–1243, 2003.
- [3] B. Borchers, "Csdp, a C library for semidefinite programming," *Optim. Meth. Softw.*, vol. 11, no. 1, pp. 613–623, 1999.
- [4] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
- [5] G. Edwards and A. Lanitis, "Statistical face models: Improving specificity," *Image Vis. Comput.*, vol. 16, no. 3, pp. 203–211, 1998.
- [6] Y. Fu and T. Huang, "Graph embedded analysis for head pose estimation," in *Proc. Int. Conf. Automatic Face Gesture Recognition*, 2006, pp. 3–8.
- [7] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.
- [8] N. Gourier, D. Hall, and J. Crowley, "Estimating face orientation from robust detection of salient facial features," in *Proc. Pointing, Int. Workshop on Visual Observation of Deictic Gestures*, Cambridge, U.K., 2004.
- [9] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu, "A method for estimating and modeling age and gender using facial image processing," in *Proc. 7th Int. Conf. Virtual Systems and Multimedia*, 2001, pp. 439–448.
- [10] Y. Kwon and N. Lobo, "Age classification from facial images," *Comput. Vis. Image Understand.*, vol. 74, no. 1, pp. 1–21, 1999.
- [11] 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.
- [12] 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.
- [13] S. Li, Q. Fu, L. Gu, B. Scholkopf, Y. Cheng, and H. Zhang, "Kernel machine based learning for multi-view face detection and pose estimation," in *Proc. Int. Conf. Computer Vision*, 2001, vol. 2, pp. 674–679.
- [14] S. Li, X. Lu, X. Hou, X. Peng, and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 705–712, Jun. 2005.
- [15] K. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, Mar. 2001.
- [16] V. Nallure, J. Ye, and S. Panchanathan, "Biased manifold embedding: A framework for person-independent head pose estimation," presented at the IEEE Conf. Computer Vision and Pattern Recognition, 2007.
- [17] N. Ramanathan and R. Chellappa, "Face verification across age progression," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2005, pp. 462–469.
- [18] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006, vol. 1, pp. 387–394.
- [19] B. Raychev, I. Yoda, and K. Sakaue, "Head pose estimation by nonlinear manifold learning," in *Proc. 17th Int. Conf. Pattern Recognition*, 2004, pp. 23–26.
- [20] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
- [21] K. Weinberger and L. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004, pp. 988–995.
- [22] Z. Zhang, Y. Hu, and T. Huang, "Pose estimation comparison for man and machine," Internal Rep. IFP, UIUC.



**Shuicheng Yan** (M'06) received the B.S. and Ph.D. degrees from the Applied Mathematics Department, School of Mathematical Sciences from Peking University, Peking, China, in 1999 and 2004, respectively. His research interests include computer vision and machine learning.

Currently he is an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore.



**Huan Wang** received the B.Eng. degree in information engineering from Zhejiang University, Hangzhou, China, the M.Phil. degree in information engineering from The Chinese University of Hong Kong, Hong Kong, China, and is currently pursuing the Ph.D. degree in computer science at Yale University, New Haven, CT.

His research interests include artificial intelligence, machine learning, and computer vision.



**Xiaoou Tang** (S'93–M'96–SM'02) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1990, the M.S. degree from the University of Rochester, Rochester, NY, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

Currently, he is a Professor and the Director of the Multimedia Lab in the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong, China. He is also the Group Manager

of the Visual Computing Group with Microsoft Research Asia, Beijing, China. His research interests include computer vision, pattern recognition, and video processing.

Dr. Tang is a Local Chair of the IEEE International Conference on Computer Vision (ICCV) 2005, an Area Chair of ICCV'07, a Program Chair of ICCV'09, and a General Chair of the ICCV International Workshop on Analysis and Modeling of Faces and Gestures 2005. He is a Guest Editor of the Special Issue on Underwater Image and Video Processing for the *IEEE Journal of Oceanic Engineering* and the Special Issue on Image- and Video-based Biometrics for the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*. He is an associate editor of the *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (PAMI)*.



**Jianzhuang Liu** (M'02–SM'02) received the B.E. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 1983, the M.E. degree from Beijing University of Posts and Telecommunications, Beijing, in 1987, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 1997.

From 1987 to 1994, he was a faculty member in the Department of Electronic Engineering, Xidian University, Xi'an, China. From 1998 to 2000, he was a Research Fellow at the School of Mechanical and

Production Engineering, Nanyang Technological University, Singapore. He then was a Postdoctoral Fellow with the Chinese University of Hong Kong, for several years. Currently, he is an Assistant Professor in the Department of Information Engineering at the Chinese University of Hong Kong. His research interests include image processing, computer vision, pattern recognition, and graphics.



**Thomas S. Huang** received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, China, in 1956, and the M.S. and D.Sc. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1960 and 1963, respectively.

He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and on the Faculty of 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, IL, where he is currently William L. Everitt Distinguished Professor of Electrical and Computer Engineering, 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, University of Illinois at Urbana-Champaign, and Co-Chair of the Institute's major research theme human computer intelligent interaction. His research interests are in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 20 books and more than 500 papers in network theory, digital filtering, image processing, and computer vision.

Dr. Huang received the IEEE Jack S. Kilby Medal in 2001. In 2002, he received the King-Sun Fu Prize, International Association of Pattern Recognition, and the Pan Wen-Yuan Outstanding Research Award. In 2005, he received the Okawa Prize, and in 2006, he was named the Electronic Imaging Scientist of the year by IS&T and SPIE. He 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 IEEE International Association of Pattern Recognition and the Optical Society of America. He 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 and the Honda Lifetime Achievement Award for "contributions to motion analysis" the same 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*.