Age Estimation Using a Min-Max Modular Support Vector Machine

Age Estimation Using a Min-Max Modular Support Vector Machine

Hui-Cheng Lian and Bao-Liang Lu
Department of Computer Science and Engineering, Shanghai Jiao Tong University
1954 Hua Shan Rd., Shanghai 200030, China
Email: {lianhc,blu}@cs.sjtu.edu.cn

***Abstract*— We introduce min-max modular support vector machine (M^3-SVM) into the age estimation problem in this paper. We extract features from the face images by using a facial point detection and Gabor wavelet transform method firstly. Then we divide the training data sets into several subsets with a 'part-versus-part' task decomposition method. The most important advantage of this task decomposition method over existing random method is that the explicit prior knowledge about genders contained in the face images is used in task decomposition. The classification task is then solved by using M^3-SVM. The experimental results indicate that M^3-SVMs with our new task decomposition method has better performance than traditional SVMs and M^3-SVMs with random task decomposition method. This work has three contributions to age estimation: (1) M^3-SVM is firstly introduced into the age estimation problem; (2) A new task decomposition method using the explicit gender information contained in the face images; and (3) A training data shrunk strategy is proposed to improve the accuracy of age estimation.**

I. Introduction

Face recognition has attracted much attention due to its potential values for applications as well as theoretical challenges. However there are still much information including emotional state, ethnic origin, gender, and age in a face image needed to be recognized and interpreted for contact-less human-computer interaction (HCI) based on facial images currently for the improvement of the interaction between humans and machines. Despite the fact that the age of a person plays an important role during interaction, so far no researcher has been involved in estimating personal age based on face images.

The work of Young *et al.* [15] may be the first published contribution for age classification. By locating the facial points including eyes, mouth, nose, chin and virtual top of the head, facial feature ratios were computed that permitted the distinguishing of babies from others. Next, wrinkles were analyzed that permitted the distinguishing of seniors from younger. Young's work is relatively elementary and the evaluation is not quantified and the details of age were experientially divided into three kinds: babies, younger, and seniors. Hayashi *et al.* [13], [14] researched about the age estimation based on wrinkle texture and color of facial images by using a special Hough transform called Digital Template Hough Transform to extract wrinkle on the face. A look up table was then used to evaluate the ages. Lanitis [17] performs Principal Component Analysis on generating facial models, and he aimed to establish the relationship between the models and the age of the subjects by a quadratic function. In his work, Lanitis also found that the area around the eyes be the most significant for the task of automatic estimation. Ryotatsu *et al.* [12] had developed a system which estimated gender and age with classifiers using support vector machines (SVMs) and voting based on facial features. Gabor Wavelet Transformation (GWT), texture, geometric arrangement, color and hair were used for age estimation. The age estimation rate they achieved is 58.4%.

Although face recognition have been widely studied, age estimation is still one of the most challenging problems for face recognition researchers. Similar to any pattern recognition problems, two key points for age estimation are feature extraction and pattern classification. From the view of feature extraction, the most simple method is to use gray-scale or color pixel vectors as features [5]. The second kind of methods come from the theory of subspace transformation such as PCA, ICA and LDA, which project faces into a low-dimensional space and then recognize them [4]. This kind of method has been shown not to be very robust to variations of face orientation. The third kind of methods is using texture information such as winkle and complexion [14]. Some of existing age estimation systems are reported based on these kinds of information[14], [12]. The last kind of methods is combining the facial feature detection with the Gabor Wavelet Transformation (GWT) to extract the local facial feature for recognition [6], [7]. This facial feature detection makes face recognition system more accurate to catch the personal features, and the wavelet transformation can model the fine characteristics of these facial points, like the analysis of face's wrinkles and shapes.

Traditional pattern classifiers such as k-nearest-neighbor, Fisher linear discriminant, neural networks and SVMs are often used to pattern recognition. Among these classifiers, SVMs seem to be superior to all other classifiers [5]. The advantage of SVMs is to find the optimal linear hyper-plane such that the expected classification error for unseen samples is minimized. However, similar to almost all traditional classifiers, SVMs treat all data in one class as a whole in training phase, and will perform coarsely than the method of further dividing the training data set in each class into a number of subsets.

In our previous work, we have proposed a 'part-versus-part' task decomposition method [3] and developed a new modular SVM called M^3-SVM for solving large-scale pattern classification problems [3]. Our studies show that M^3-

SVMs have three main advantages over traditional SVMs: (1) Massively parallel training of SVMs can be easily implemented in parallel computing systems; (2) Large-scale pattern classification problems can be solved efficiently; and (3) The generalization accuracy can be obviously improved. Hereto we have succeeded in applying M³-SVM to several pattern recognition problems including large scale text categorization [8], multi-view face recognition [10] and gender recognition problem [9].

The remainder of this paper is organized as follows: In the next section, we present an overview of the feature extraction for age estimation. We give a brief introduction to the min-max modular support vector machine in Section III. Section IV presents the task decomposition strategies and preparations for age estimation. Experimental results and analysis are presented in Section V. The final section provides conclusions.

II. Feature Extraction for Age Estimation

The well-performed face feature extraction method [6], [7] will be used to generate feature vectors for our M³-SVM classifiers. The main idea of this face feature extraction is to detect the face in an image firstly and then locate the facial points including eyes, nose and mouth. Gabor wavelet transformation plus retina sampling is then used to extract the facial point characteristics which are combined to form a feature vector. The extracted feature vectors are processed as the inputs to our estimation system.

For facial point location, a method called retina sampling [7] is set to the facial feature points search. With retina sampling, the feature point is more precise around the important facial features like eyes and mouth, and less precise as one moves away from these areas. When paired with GWT, it is possible to evenly extract localized features as well as more wide-area features. And by pairing GWT and retina sampling, it is possible to create a feature extraction algorithm that approximates the human visual sense, and it is also possible to extract features, which are valuable to estimate age. Furthermore, it has the added benefit of being highly adaptive to facial variations such as orientation, expression, hairstyle, lighting, eyeglasses and so on [7].

Gabor filter can be expressed as follows:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \cdot \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (1)$$

where μ and ν define the orientation and the scale of the Gabor kernels, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu}, \quad (2)$$

where $k_\nu = k_{max}/f^\mu$ and $\phi_\mu = \pi\mu/8$, k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. Fig.1 shows Gabor wavelets of five different scales, $\nu \in \{0, ..., 4\}$ and eight orientations, $\mu \in \{0, ..., 7\}$.

For more details about this feature extraction, one can refer to [6], [7]. Here we will simply describe the scales of the

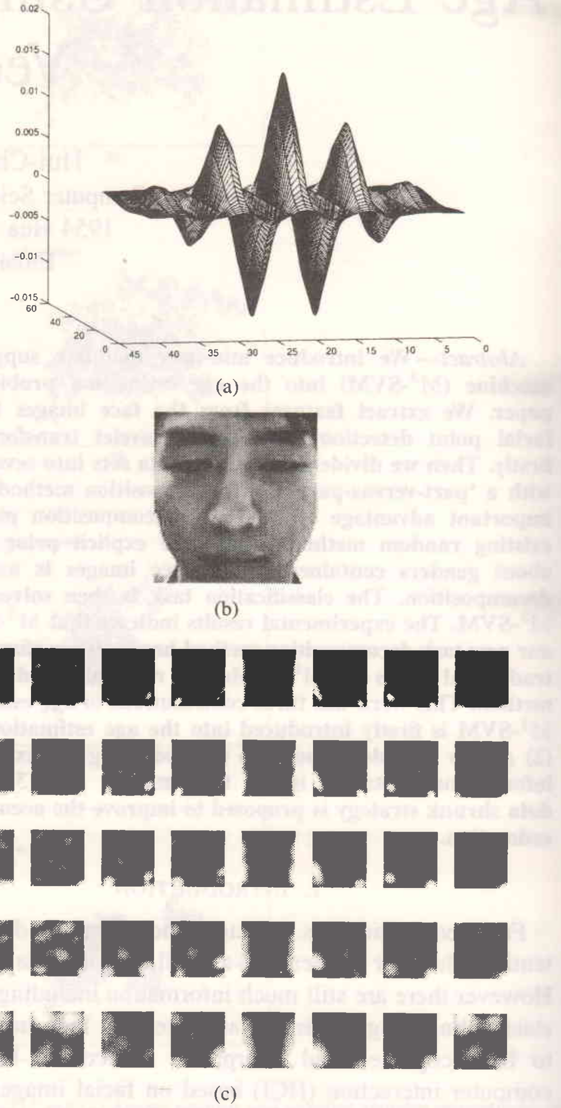

Fig. 1. Gabor filter and Gabor representations of a facial image. (a) A Gabor filter; (b) Facial image; and (c) Eight orientations and five scales of Gabor representations (the magnitudes) on the facial image

gallery sets and probe sets for our M³-SVM age estimation system in section IV.

III. Min-Max Modular Support Vector Machine

The min-max modular network was firstly introduced in [1]. In this framework, a K-class classification problem is decomposed into a series of $K(K-1)/2$ two-class problems. These two-class problems are to discriminate class C_i from C_j for $i = 1, ..., K-1$ and $j = i+1, ..., K$, while the existence of the training data belonging to the other K-2 classes is ignored. If the two-class problems are still hard to be learned, they can be divided into a set of two-class subproblems as small as needed. Consequently, a large-scale and complex K-class classification problem can be solved effortlessly and efficiently by learning a series of smaller and simpler two-class problems in a parallel way.

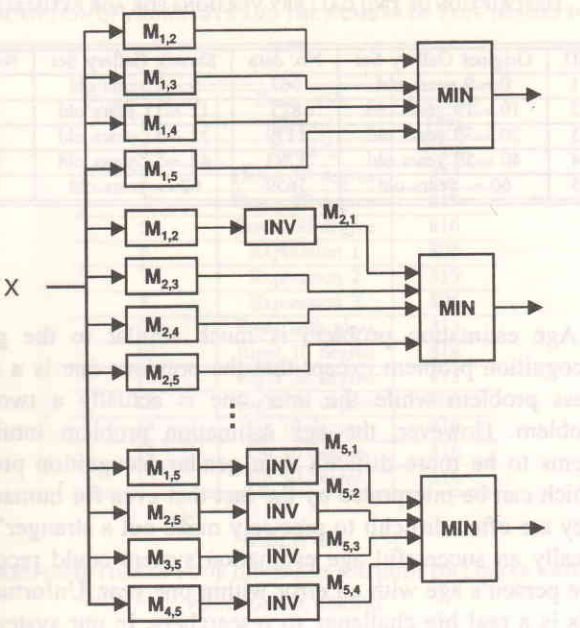

Fig. 2. Illustration of module combination of a five-class problem. Here $\mathbf{M}_{i,j}$ means a two-class problem of class i and class j and **MIN**, **INV** means a **MIN** unit and an **INV** unit respectively.

A. *Task Decomposition and Module Combination*

Let $\mathcal{T}$ be the training set of a K-class classification problem and the K classes are represented by $C_1, C_2, \ldots, C_K$, respectively.

$$\mathcal{T} = \{(X_l, Y_l)\}_{l=1}^{L} \tag{3}$$

where $X_l \in \mathbf{R}^{\mathrm{d}}$ is the input vector, $Y_l \in \mathbf{R}^{\mathrm{K}}$ is the desired output, and L is the number of training data.

Suppose the K training input sets, $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_K$, are expressed as

$$\mathcal{X}_i = \left\{ (X_l^{(i)} \right\}_{l=1}^{L_i} \quad \text{for} \quad i = 1, \ldots, K \tag{4}$$

where L_i is the number of training inputs in class C_i, $X_l^{(i)}$ is the lth sample belonging to class C_i and all of $X_l^{(i)} \in \mathcal{X}_i$ have the same desired outputs, and $\sum_{i=1}^{K} L_i = L$.

It is known that a K-class problem defined by (3) can be divided into $K(K-1)/2$ two-class subproblems, each of which is given by

$$\mathcal{T}_{ij} = \left\{ (X_l^{(i)}, +1) \right\}_{l=1}^{L_i} \cup \left\{ (X_l^{(j)}, -1) \right\}_{l=1}^{L_j}$$
$$\text{for} \quad i = 1, \ldots, K-1 \quad \text{and} \quad j = i+1, \ldots, K \tag{5}$$

Suppose each of the two-class problems have been learned, all of the individual trained modules which were assigned to learn associated two-class problems can be easily integrated into a M^3-network by using the **MIN**, or/and **INV** units according to the proposed module combination principles [1].

Let y denote the actual output vector of the whole M^3-network for a K-class classification problem, let $g(x)$ denote the transfer function of the M^3 network. We may then write

$$y = g(x) = [g_1(x), \ldots, g_K(x)]^T \tag{6}$$

where $y \in \mathbf{R}^K$, and $g_i(x)$ is the *discriminant function*, which discriminates the patterns of C_i from those of the rest classes, The M^3-network is said to assign an input x to class C_i if

$$|g_i(x) - (1-\epsilon)| \leq \delta \quad \text{and} \quad |g_j(x) - \epsilon| < \delta \quad \text{for} \quad j \neq i \tag{7}$$

where $1-\epsilon$ and ϵ denote the high and low desired outputs, respectively, δ is a real number, which denotes the error tolerance. For example, ϵ and δ can be set to 0.01 and 0.5.

The discriminant function $g_i(x)$ of the M^3 network which is constructed to learn the $K(K-1)/2$ two-class problems can be given by

$$g_i(x) = \min \left[\min_{j=i+1}^{K} (h_{ij}(x)), \min_{r=1}^{i-1} \mathrm{INV}(h_{ri}(x)) \right] \tag{8}$$

where $h_{ij}(x)$ is the activation function of the module $\mathbf{M}_{ij}$ trained on $\mathcal{T}_{ij}$ (5) and INV is the inverse operator defined in [1].

Fig. 2 shows the module combination of a five-class problem. In this figure $\mathrm{M}_{i,j}$ means a two-class problem of class i and class j and **MIN**, **INV** means a **MIN** unit which is to find a minimum value from its multiple inputs and an **INV** unit which is to invert its single input, respectively.

B. *Fine Task Decomposition and Module Combination*

Even though a K-class problem is broken into $K(K-1)/2$ relatively smaller two-class problems, some of these two-class problems may be still hard to be learned. We have suggested that $\mathcal{T}_{ij}$ defined by (5) can be further decomposed into a number of two-class subproblems as small as needed according to the class relations among training data [1].

Assume that the input set $\mathcal{X}_i$ defined by (4) is further partitioned into $N_i (1 \leq N_i \leq L_i)$ subsets in the form of

$$\mathcal{X}_{ij} = \left\{ (X_l^{(ij)} \right\}_{l=1}^{L_i^{(j)}} \quad \text{for} \quad j = 1, \ldots, N_i \tag{9}$$

where $L_i^{(j)}$ is the number of training inputs included in $\mathcal{X}_{ij}$, and $\cup_{j=1}^{N_i} \mathcal{X}_{ij} = \mathcal{X}_i$, the training set for each of the smaller and simpler two-class problems can be given by

$$\mathcal{T}_{ij}^{(u,v)} = \left\{ (X_l^{(iu)}, +1) \right\}_{l=1}^{L_i^{(u)}} \cup \left\{ (X_l^{(jv)}, -1) \right\}_{l=1}^{L_j^{(v)}}$$
$$\text{for} \quad u = 1, \ldots, N_i, v = 1, \ldots, N_j$$
$$i = 1, \ldots, K-1 \quad \text{and} \quad j = i+1, \ldots, K \tag{10}$$

where $X_l^{(iu)} \in \mathcal{X}_{iu}$ and $X_l^{(jv)} \in \mathcal{X}_{jv}$ are the input vectors belonging to class C_i and class C_j, respectively. Here $\sum_{u=1}^{N_i} L_i^{(u)} = L_i$, and $\sum_{v=1}^{N_j} L_j^{(v)} = L_j$.

In the learning phase, all of the two-class subproblems are independent from each other and can be efficiently learned

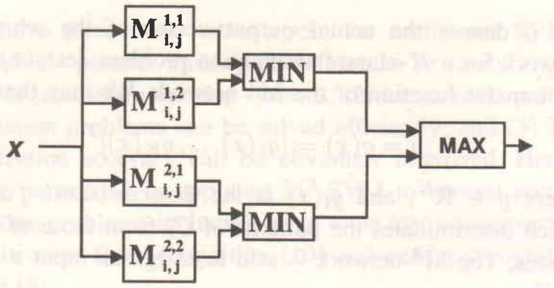

Fig. 3. Illustration of fine decomposition and module combination of a two-class problem. Here the two-class problem of class i and class j is further decomposed into 2×2 sub-problems. Therefore the M^3-network consists of 2×2 individual network modules, 2 MIN units, and one MAX unit.

in a massively parallel way. These individual modules which were assigned to learn associated subproblems can be trained by any existing classifier, such as KNN, MLP and SVMs [1], [3], [8], [9]. SVMs, due to their powerful learning ability and good generalization performance, are chosen as the classifiers of subproblems in this paper. Therefore we called it as min-max modular support vector machine [3].

After training all the individual subproblems of class C_i and class C_j, the $N_i \times N_j$ smaller SVMs are integrated into a M^3-SVM with N_i MIN units and one MAX unit according to the two combination principles [1], [2], [3] as follows,

$$h_{i,j}(x) = \max_{u=1}^{N_i} \left[\min_{v=1}^{N_j}(h_{i,j}^{u,v}(x)) \right] \tag{11}$$

where $h_{i,j}^{u,v}(x)$ denotes the transfer function of the trained SVMs on individual module $\mathrm{M}_{i,j}^{u,v}$ corresponding to the two-class subproblem of $\mathcal{T}_{i,j}^{(u,v)}$ defined by (10).

Fig. 3 shows this idea of fine decomposition and module combination for two-class problem. Here the two-class problem of class i and class j is further decomposed into 2×2 sub-problems. Therefore the M^3 network consists of 2×2 individual network modules, 2 MIN units, and one MAX unit.

IV. Task Decomposition Strategies and Preparations for Age Estimation

A M^3-SVM needs to divide the training data sets into several subsets in its first step. So how to divide the training data set effectively is an important issue. Although dividing the data set randomly is a simple and straightforward approach [2], the geometric relation among the original training data may be damaged [8]. The data belonging to a reasonable cluster may be randomly separated into other clusters. From the viewpoint of SVMs, random task decomposition might lead the boundaries of subproblems complex. In our precious work [9], we have proposed a prior knowledge based strategy (PK) for the decomposition of gender recognition and have shown that M^3-SVM with PK task decomposition has better performance than traditional SVMs and M^3-SVM with random task decomposition method.

TABLE I

Description of Two Gallery Versions for Age Estimation

ID	Original Gallery Set	No. data	Shrunk Gallery Set	No. data
1	0 ~9 years old	967	0 ~7 years old	703
2	10 ~19 years old	823	12 ~17 years old	502
3	20 ~39 years old	1139	22 ~37 years old	825
4	40 ~59 years old	1793	42 ~57 years old	1414
5	60 ~ years old	1659	62 ~ years old	1453

Age estimation problem is much similar to the gender recognition problem except that the previous one is a multi-class problem while the later one is actually a two-class problem. However, the age estimation problem intuitively seems to be more difficult than gender recognition problem which can be interpreted by the fact that even for humankind, they are often difficult to precisely make out a stranger's age. Ideally an successful age estimation system could recognize one person's age with an error within one year. Unfortunately this is a real big challenge to researchers. In our system, we simply divide the age samples into 5 categories which range from 0~9, 10~19, 20~39, 40~59, and over 60 years old, implying that our estimation problem is translated into a five-class pattern recognition problem.

However when using such a translation, samples near the boundaries such as the side between 9 and 10 years old may be unfair to the system and they might effect the final results of evaluation if not properly mended. Considering the fact that some of the age samples do not help to intensify the common feature of one category but fuzz up the boundaries between translated categories. For example, a 9 years old sample is more possible to be estimated as 10 years old rather than be estimated as 5 years old. Based on this consideration, we propose an approach to discarding the boundary samples. So another substituted strategy is to similarly divide the age samples into new 5 categories but range from 0~7, 12~17, 22~37, 42~57, and over 62 years old respectively. We called these 5 categories as shrunk data sets. We will compare the evaluation results on the original 5 categories and the shrunk 5 categories in the later parts. Table I shows these two versions of data sets.

In our precious work [9], we have used the age information for gender data decomposition, i.e. we sort the samples using age information from young to old, and then divide them into different subsets. However, to an age estimation problem, considering that we have had the whole personal gender information of data sets and the fact that faces between man and woman have much different features even in the same age phase, we employ the gender information in task decomposition for M^3-SVM to solve the age estimation problem. According to these gender information we divide each of five categories of age samples into two subsets respectively. Therefore there are totally 40 (i.e. $\frac{5\times4}{2} \times 4 = 40$) two-class problems to learn the original age estimation problem.

TABLE II

DESCRIPTION OF PROBE SETS AND THE NUMBER OF FACE IMAGES FOR TEST

Probe Set ID	Description	No. data
1	Front 1	1278
2	Front 2	1066
3	Down 10 degree	820
4	Down 20 degree	819
5	Down 30 degree	816
6	Expression 1	805
7	Expression 2	815
8	Expression 3	805
9	Front with glass	813
10	Right 10 degree	814
11	Right 20 degree	815
12	Right 30 degree	805
13	Up 10 degree	819
14	Up 20 degree	816
15	Up 30 degree	816

TABLE III

RESPONSE TIME OF FOUR DIFFERENT METHODS ON LINEAR AND POLYNOMIAL KERNEL

Methods	Training Time (s)		Time A[a] / Time B[b] (ms/p)	
	Linear	Poly	Linear	Poly
SVM	553.7	667.7	97.3 / –	120.4 / –
M3-SVM-RND	455.7	463.3	520.3 / 42.1	571.8 / 44.5
M3-SVM-PK	439.5	442.7	512.8 / 40.9	531.8 / 43.3
M3-SVM-S-PK	411.6	436.3	450.3 / 37.5	424.4 / 35.8

[a] response time in serial

[b] response time of largest individual module, i.e. response time in parallel

V. EXPERIMENTS

In this section, we present experimental results on the age data sets to compare the traditional SVMs with M^3-SVM using our proposed task decomposition methods. We use linear SVMs and polynomial SVMs to evaluate the performance of the proposed method. The reason why RBF SVMs are not selected in this paper is that the optimal parameters of RBF kernel are hard to be obtained. All SVMs and SVMs in M^3 network are trained by LibSVM [11] and the parameters are set to the default parameters of LibSVM, i.e. C=1 for linear SVM, and C=1 and Degree=3 for polynomial SVMs.

Two versions of gallery sets with their samples extracted from a large face images database using the feature extraction method mentioned in Section II are shown in Table I. Here, the number of dimensions of each feature is 4350. The 15 kinds of probe sets including frontal images, various degree view faces, faces with glasses and different expressions and so on, are shown in Table II. And all of the performance evaluations are running on these fifteen probe sets.

In the experiment I, all of the SVMs are linear kernel. We use traditional SVM, M^3-SVM-RND, M^3-SVM-PK and M^3-SVM-S-PK to evaluate the performance, where M^3-SVM-RND, M^3-SVM-PK and M^3-SVM-S-PK denote M^3-SVM with random task decomposition on original gallery sets, M^3-SVM with prior knowledge (PK) task decomposition on original gallery sets and M^3-SVM with PK task decomposition on the shrunk gallery sets, respectively. In experiment II, all the classifiers are the same as the experiment I except that the kernel of SVMs is changed to polynomial kernel.

We show the experimental results in Fig. 4 and Table III. Fig. 4(a) shows the total number of correct outputs produced by four methods on fifteen probe sets using linear kernel. Fig. 4(b) plots the detailed accuracy of the four methods on fifteen probe sets. From Figs 4(a) and 4(b) we can see that: (1) M^3-SVM with random task decomposition outperforms traditional SVMs; (2) M^3-SVM with PK task decomposition outperforms M^3-SVM with random task decomposition; and (3) training data shrunk strategy can help to improve the performance of age estimation. Similar results can also be seen from Fig. 4(c) for polynomial kernel.

Table III also shows the response time of these four methods on different kernel functions. From this table we can see that the training time of M^3-SVM is less than that of traditional SVMs, but the response time of M^3-SVM is longer than that of the traditional SVMs. However there is a large space to improve the response time of M^3-SVM since there are much redundant computing inside M^3-SVM. We can speed up the response time by pruning redundant module or using selection algorithm to speed up response time in serial [18].

VI. CONCLUSIONS

We have proposed a new task decomposition approach using M^3-SVM to deal with the age estimation problem. We have compared four methods on age estimation problem. From experimental results, we can draw the following conclusions: (a) M^3-SVM outperforms traditional SVMs on age estimation problem in the estimation accuracy and training time. (b) The proposed task decomposition method that using the gender information inside age data samples can help to improve the performance of M^3-SVM in the classification accuracy. (c) A training data shrunk strategy can help to improve both estimation accuracy and response time of M^3-SVM on the age estimation problem. There are several issues should be further exploited, such as pruning the redundant computation inside M^3-SVM and using unsupervised task decomposition for M^3-SVM.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China via the grants NSFC 60375022 and NSFC 60473040.

REFERENCES

[1] B. L. Lu and M. Ito, "Task decomposition based on class relations: a modular neural network architecture for pattern classification", in *Biological and Artificial Computation: From Neuroscience to Technology*, Lecture Notes in Computer Science, Springer Vol.1240 (1997), pp. 330-339.

[2] B. L. Lu and M. Ito, "Task Decomposition and Module Combination Based on Class Relations: a Modular Neural Network for Pattern Classification," *IEEE Transactions on Neural Networks*, 10(1999) pp. 1244 -1256.

[3] B. L. Lu, K.A. Wang, M. Utiyama and H. Isahara, "A Part-versus-part Method for Massively Parallel Training of Support Vector Machines," in *Proc. of IJCNN'04*, Budapast, July 25-29 (2004) pp. 735-740.

[4] K. Balci and V. Atalay, "PCA for Gender Estimation: Which Eigenvectors Contribute?" in *16 th International Conference on Pattern Recognition* (ICPR'02) Volume 3 August 11 - 15, 2002 Quebec City, QC, Canada, pp. 303-363.

[5] B. Moghaddam and M.H. Yang, "Gender Classification with Support Vector Machines," in: *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, 2000, pp.306-311.

[6] M. Kawade, "Vision-Based Face Understanding Technologies and Applications," in *2002 International symposium on micromechatronics and human science*, 2002, pp.27-32.

[7] S. Hosoi, E. Takikawa and M. Kawade, "Ethnicity Estimation with Facial Images," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, May 17-19, 2004 Seoul, Korea, pp.195-200.

[8] K. A. Wang, Z. Hai and B. L. Lu, "Task Decomposition Using Geometric Relation for Min-Max Modular SVM," in *Proc. Advances in Neural Networks-ISNN 2005, Lecture Notes in Computer Science*, Vol. 3496, 2005, pp. 887-892.

[9] H. C. Lian and B. L. Lu, "Gender Recognition Using a Min-Max Modular Support Vector Machine," in *Proc. ICNC'05-FSKD'05, Lecture Notes in Computer Science*, 2005, pp. 433-436.

[10] Z. G. Fan and B. L. Lu, "Multi-View Face Recognition with Min-Max Modular SVMs," in *Proc. ICNC'05-FSKD'05, Lecture Notes in Computer Science*, LNCS 3611, 2005, pp.396-399.

[11] C. C. Chang and C. J. Lin, "LIBSVM : a library for support vector machines," http://www.csie.ntu.edu.tw/~ cjlin/papers/libsvm.ps.gz

[12] R. Iga and K. I. A, "Gender and Age Estimation System from Face Images," in *SICE Annual Conference in Fukui*, August 4-6, 2003, pp.756-761.

[13] J. Hayashi, "A Method for Estimating and Modeling Age and Gender using Facial Image Processing," in *Proc. of the Seventh International Conference on Virtual Systems and Multimedia* (VSMM'01), 2001, pp.439-448.

[14] J. Hayashi, M. Yasumoto, H. Ito and H. Koshimizu, "Age and gender estimation based on wrinkle texture and color of facial images," in *16th International Conference on Pattern Recognition*, 11-15 Aug. 2002, pp.405-408.

[15] H. K. Young and V. L. Niels, "Age Classification from Facial Images," in *Proc. IEEE Conf. on Computer Vision.and Pattern Recognition*, Seattle, Washington, U. S. A., 1994, pp.762-767.

[16] A. Lanitis, "Comparing Different Classifiers for Automatic Age Estimation," *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, vol. 34, no. 1, Mar. 2004, pp. 621-628.

[17] A. Lanitis, "On The Significance Of Different Facial Parts For Automatic Age Estimation," in *14th International Conference on Digital Signal Processing 2002*, pp. 1027-1030.

[18] H. Zhao and B. L. Lu, "On Efficient Selection of Binary Classifiers for Multi-class Classification", in *Proc. of IJCNN'05*, Montreal, Canada, July 31-August 4, 2005, pp. 3186-3191.

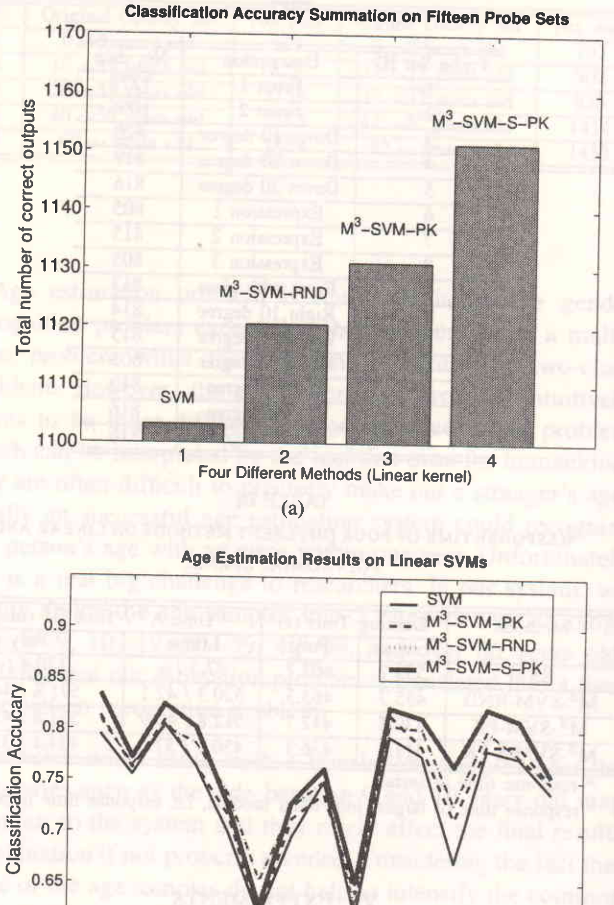

Fig. 4. The comparative results of experiment I and II. (a) Classification accuracy summation on fifteen probe sets using linear kernel. (b) Plot results of four methods using linear kernel. (c) Classification accuracy summation on fifteen probe sets using polynomial kernel.