A Multi-Task Model for Simultaneous Face Identification and Facial Expression Recognition

Hao Zheng[a,b,c,d], Xin Geng[a,*], Dacheng Tao[e], Zhong Jin[c]

[a] *MOE Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, China*
[b] *Key Laboratory of Trusted Cloud Computing and Big Data Analysis, NanjingXiaoZhuang University, Nanjing, China*
[c] *Jiangsu Key Laboratory of Image and Video Understanding for Social Safety,Nanjing University of Science and Technology, Nanjing,China*
[d] *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China*
[e] *Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology,University of Technology, Sydney, Australia*

Abstract

Regarded as two independent tasks, face identification and facial expression recognition perform both poorly given small size training sets. To address this problem, we propose a multi-task facial inference model (MT-FIM) for simultaneous face identification and facial expression recognition. In particular, face identification and facial expression recognition are learnt simultaneously by extracting and utilizing appropriate shared information across them in the framework of multi-task learning, in which the shared information refers to the parameter controlling the sparsity. MT-FIM simultaneously minimizes the within-class scatter and maximizes the distance between different classes to enable the robust performance of each individual task. We conduct comprehensive experiments on three face image databases. The experimental results show that our algorithm outperforms the state-of-the-art algorithms.

Keywords: face identification, facial expression recognition, multi-task learning

*Corresponding author
Email address: xgeng@seu.edu.cn (Xin Geng)

1. Introduction

Face recognition is still a very active research area in public security, human-computer interaction, financial security, etc. The two primary face recognition tasks are identification and verification. The goal of face identification task is to identify a person based on the face image, i.e., the captured face image needs to be matched to a gallery of known people. Therefore, Addressing face identification task is the key issue for face recognition. Though typical applications of face recognition have been used in many years, many real world situations are still a challenge due to illuminations, pose, and occlusions [1]. At the same time, facial expression recognition is also an important topic due to its wide range of applications [2–5]. Most Facial expression recognition methods aim to recognize a set of prototypic expressions (i.e. surprise, anger, joy, sadness, fear, and disgust). Though much progress has been made [6–10], the performance of facial expression recognition is still unsatisfactory due to the subtlety, complexity and variability of facial expressions.

For face recognition, many algorithms about classifiers have been proposed. The nearest- neighbor (NN) algorithm is simple and it is accurate and applicable to various problems. But the shortcoming of NN algorithm is that only one training sample is used to represent the test face image, so the nearest feature line classifier [11] was proposed through using two training samples for each class to represent the test face image. Then the nearest feature plane classifier [12] was proposed through using three samples to represent the test image. Later, for representing the test image by all the training samples of each class, the local subspace classifier [13] and the nearest subspace classifier [14, 15] were proposed. Because samples from a specific object class are known to lie on a linear subspace, linear regression classification (LRC) [16] algorithm was proposed by formulating the pattern recognition problem in terms of linear regression. Another well-known classifier is support vector machine (SVM) classifier [17], which is solidly based on the theory of structural risk minimization in statistical learning. It is well known that the SVM can efficiently perform a non-linear

classification and map the inputs to a high-dimensional feature space, then find a large margin hyperplane between the two classes which can be solved through the quadratic programming algorithm. However, SVM cannot be applied when the vectors defining out samples have missing entries. It can be seen when occlusions are present for face recognition. Fortunately sparse representation based classification (SRC) was reported by Wright et al.[18] for robust face recognition. In Wright et al.'s pioneer work, the training face images are used as the dictionary of representative samples, and an input test image is coded as a sparse linear combination of these sample images via l_1-norm minimization. The experimental results in [18, 19] of SRC were exciting in FR, which could lead to high classification accuracy, especially well handling the problems of face occlusion and corruption. Furthermore, many extended methods were proposed [20–22], e.g. Gabor SRC [20], Heteroscedastic SRC [21], and SRC for continuous occlusion [22].

For facial expression recognition, many previous works are based on statistical learning. Ekman [23] developed the Facial Action Coding System (FACS) in which the movements on the facial expression are described by action units. Many extension methods [24, 25] were also proposed, and various classifiers were applied for facial expression recognition. In [26] , after the facial features are extracted and represented, an artificial neural network (ANN) was employed to recognize the action units. In order to recognize the naturalistic affective expressions, Meng [27] adopted Hidden Markov Models (HMM) to model this spontaneous process, and finalize the classification process. Then Cohen [28] proposed a new architecture of HMM for automatically segmenting and recognizing human facial expressions. SVM classifier [29] that incorporates statistical information of the classes under examination was also proposed and used to recognize either the six basic facial expressions or a set of Facial Action Units. In [30] , popular machine learning methods including SVM were compared for facial expression recognition, and SVM was proved to be the most effective classifier.

However, in real-world applications for face identification and facial expression recognition, the number of training samples from each class is usually lim-

ited. Unfortunately, the performances of face identification and facial expression recognition by the above methods usually degrade significantly with the decrease of training samples. Therefore the small sample size problem becomes one of the most prominent issues in face identification and facial expression recognition. In fact, the goal of most face identification methods is to find a similarity measure invariant to illumination changes, pose, and facial expressions, so that images of faces can be recognized in spite of existence of these variation. While the goal of expression recognition is to find a model for non-rigid patterns facial expression, so that facial expression can be classified in spite of a wide range of variation. Therefore, Information about facial appearance and expression patterns can be jointed to carry face identification and facial expression at the same time [31, 32] . Both face identification task and facial expression recognition task aim to learn universal model which make the tasks are robust for various situation. In general, multi-task learning can be adopted to join face identification and facial expression recognition by extracting the appropriate shared information. Especially multi-task learning has been shown, both empirically and theoretically, to be able to significantly improve the performance of learning each task separately. In this paper, motivated by the above idea, we propose a multi-task facial inference model (MT-FIM) for simultaneous face identification and facial expression recognition. In MT-FIM, the classifiers of face identification and facial expression recognition are learned at the same time by extracting and utilizing appropriate shared information across them. More importantly, face identification and facial expression recognition can benefit from each other, resulting in better performance. In addition, we minimize the within-class distance and maximize the distance between different classes. Because MT-FIM is convex, optimization of MT-FIM is employed. In conclusion, the contributions of our work are three-fold.

(1) We propose a MT-FIM method for face identification and facial expression recognition, which is a multi-task method that joints two related tasks. To the best of our knowledge, this is the first work to simultaneously learning

the face identification and facial expression recognition.

(2) Since MT-FIM adopts the multi-task and increases the samples size for each recognition tasks, the sample size problem can be solved efficiently. MT-FIM achieves better recognition results than the state-of-the-art methods.

(3) In MT-FIM, we introduce the within-class covariance which makes the same class samples easier to cluster. MT-FIM strikes a balance between minimize the within-class distance and maximize the distance between different classes.

The rest of this paper is organized as follows. Section 2.1 reviews the technology of multi-task learning. Section 2.2 presents our multi-task facial inference model. Section 2.3 describes optimization process of MT-FIM, and then complexity of MT-FIM is discussed in Section 2.4. Finally, Section 3 conducts experiments to validate the proposed model and Section 4 concludes the paper.

2. Multi-Task Facial Inference Model (MT-FIM)

In this section, firstly we discuss the related multi-task learning. Then we present the proposed MT-FIM algorithm. Finally optimization of the algorithm is stated.

2.1. Multi-task learning

Multi-task learning has continuously received increasing attention in computer vision, image recognition. It aims to simultaneously learn multiple related tasks and utilizing the shared information among related tasks for improving performance of each task. In the past years many multi-task learning methods have been studied, existing methods can be broadly classified into two categories: implicit structure sharing and explicit parameter sharing. Methods under implicit structure sharing implicitly capture some common structures; for example, the methods in [33, 34] constrain all tasks to share a common low rank subspace and the methods in [35, 36] constraint all tasks to share a common set of features. While methods under explicit parameter sharing explicitly

share some common parameters; examples include hidden units in neural networks [37, 38] , prior in hierarchical Bayesian models [39, 40] , parameters of Gaussian process [41] , classification weight [42] , feature mapping matrix [43] , and similarity metric [44, 45]. In two categories, methods under explicit parameter sharing were often adopted, and multi-task sparse learning was employed by learning multiple classifiers from different tasks to share similar parameter sparsity patterns. For increasing the training samples size, multi-task learning can effectively solve the problem of the small sample size. Especially the Lasso regularized methods [46, 47] are widely used to image recognition for the simplicity and effectiveness. The Lasso method in MTL is a penalized least square method imposing a l_1-norm penalty on the regression coefficients and the parameter controlling the sparsity is shared by all related tasks. A general model is illustrated as below:

$$\min_{W} \sum_{i=1}^{t} \Gamma(W) + \rho \left\| W \right\|_1 , \tag{1}$$

where W is classifier to be estimated from the training samples, $\Gamma(W)$ is the empirical loss on the training set, and $\left\| W \right\|_1$ is the regularization term that encodes task, t is the number of tasks. In the model, ρ is the regularization parameter for controlling sparsity.

2.2. Multi-Task Facial Inference Model (MT-FIM)

But in the multi-task model within-class discriminative information is not considered, so the multi-task learning sometimes wrongly classifies the samples of the same class to the other class. To address this problem, we join the fisher discrimination idea and employ the within-class scatter matrix to minimize the within-class distance. A geometrical explanation of within-class scatter matrix of MT-FIM is shown in Figure 1. From Figure 1, we can see that the samples may be mistakenly classified by the methods without within-class scatter. For considering only the within-class scatter of face identification and facial expres-

sion recognition, within-class covariance S_w of them are defined as:

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T, \tag{2}$$

where x_i^j is the i-th sample of class j, μ_j is the mean of class j, c is the number of classes, and N_j is the number of samples in class j.

When S_w is singular, we can add a diagonal matrix to the within-class covariance matrix, so an identity matrix with a small scalar multiple is used as below:

$$w^T(S_w + \lambda I)w. \tag{3}$$

Following the idea of sharing the common information between the tasks, we introduce the parameter γ to control the sparsity. At the same time, the term $w^T(S_w + \lambda I)w$ provides a measure of the closeness of the projected data, therefore γ can be adjusted for minimizing the within-class distance and maximizing the distance between different classes. As is shown as below:

$$\gamma(w^T(S_w + \lambda I)w). \tag{4}$$

From above analysis, the common shared information γ makes the related tasks benefit from each other. In this paper, we are interested in the multi-task binary classification problem. Finally, we propose to add an elastic term into the model for controlling the l_2-norm penalty. Let $X_i = \{x_1, \cdots, x_n\}'$ be $n \times d$ the training samples of the i-th task, where n is the sample size, d is the dimensionality of the feature space, and denote by $Y_i = \{y_1, \cdots, y_p, \cdots, y_n\}'$ the corresponding label vector of the i-th task, where $y_p \in$ {-1, 1}, the proposed MT-FIM is defined as:

$$\min_W \sum_{i=1}^{t} (\left\|W_i^T X_i - Y_i\right\|_F^2 + \gamma(W_i^T(\lambda I + S_w^i)W_i)) + \eta \left\|W\right\|_F^2, \tag{5}$$

where t is the number of tasks, W_i is classifier of the i- th task, S_w^i is within-class covariance matrix of the i-th task, η is a parameter. In MT-FIM, the number of the tasks is sum of the number of the face identification task and the

number of the facial expression recognition task. For the binary classification, the number of the face identification task is the number of the subject classes, and the number of the facial expression recognition task is often 6. In fact, common parameter γ can be determined via cross validation through the experiments. By adopting the tune parameter γ, MT-FIM simultaneously minimizes the within-class scatter and maximizes the distance between different classes. Meanwhile owing to increasing the number of training samples, performance of each individual recognition task will be improved. In sum, we can learn the face identification classifiers and the facial expression recognition classifiers from the small size training set via MT-FIM, then identify the class and recognize facial expression of the query samples. Figure 2 illustrates the flowchart of MT-FIM. In Figure 2, training images are firstly selected from various expressions for each subject, then face identification tasks and face expression recognition tasks are created according to the training samples matrix and corresponding label vector. Next the regularization of within-class scatter matrix and shared parameters encourage the classifier not only to share common information but also to minimize the within-class distance and maximize the distance between different classes. Finally, face identification and expression recognition of the test images can be classified by the learned classifiers.

In MT-FIM, the shared information parameter γ enforces face identification and facial expression recognition to integrate by multi-task learning, and it makes them benefit from each other. For more details, please refer to [37,42]. An obvious advantage fo MT-FIM is that the learned classifier has incorporated both the multi-task regularization and within-class scatter. Equivalently, the learned classifier considers relationship from face identification and facial expression recognition instead of a single task.

Morever MT-FIM can be extended through the union of face image database. Intuitively from MT-FIM we can see that if several face image databases are jointed, more tasks are created and more classifiers are learned. Because the learning of big data can reveal the information representation of the images, the performance of MT-FIM will be improved further, which is verified by the

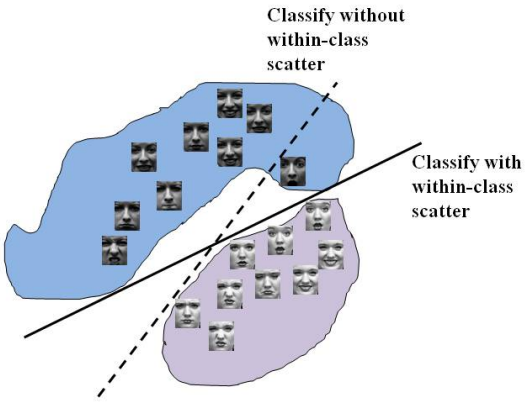

Figure 1: A geometrical explanation of within-class scatter matrix of MT-FIM

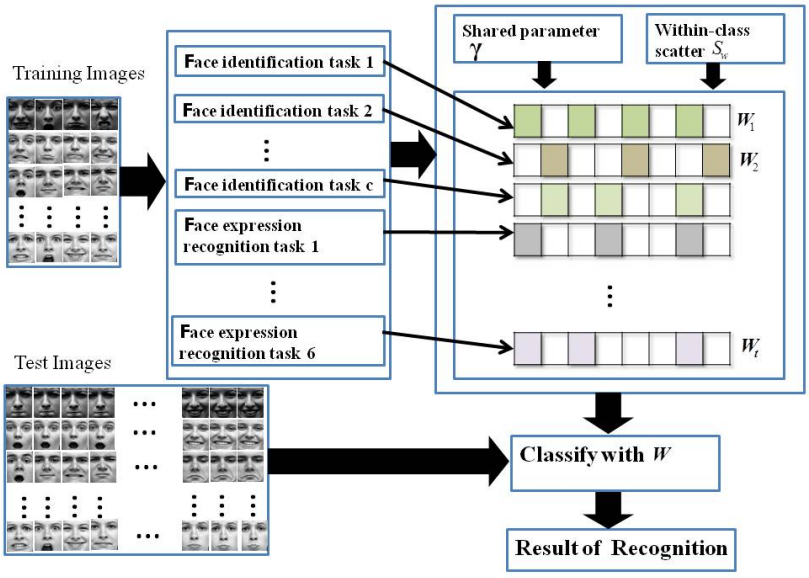

Figure 2: flowchar of MT-FIM

union experiments of Section 3.

2.3. Optimization of MT-FIM

If MT-FIM is a convex problem, we can solve it by adopting the convex optimization method. So we first prove the convexity of Eq. (5) with respect to the variable.

Theorem 1 Eq. (5) is convex with repect to W.

Proof:

It is easy to see that the first and third term in the objective function of Eq. (5) are convex with respect to the variable. We set $\Omega = \lambda I + S_w^i$, the second term in the objective function is rewritten as $\sum_{i=1}^{t} \gamma(W_i^T \Omega W_i)$, where W_i denotes the i-th column of W. $W_i^T \Omega W_i$ is called a matrix fractional function in Example 3.4 (page 76) of [48] and it is proved to be a convex function with respect to W_i when Ω is a positive semidefinite matrix (which is satisfied by S_w). Because the summation operation can preserve convexity according to the analysis on page 79 of [48], $\sum_{i=1}^{t} \gamma(W_i^T \Omega W_i)$ is convex with respect to W. So the objective function is convex with respect to variable and hence Eq. (5) is jointly convex.

Next this section shows how to optimize the proposed model efficiently. Assume $f(x) = (W_i^T X_i - Y_i)^2 + \gamma W_i^T (\lambda I + S_w^i) W_i)$, the optimization of MT-FIM iterates between updating the gradient $f(x)'$ and aggregation matrix V. The optimization progress consists of three steps: (1) within-class scatter matrices generalizing step that computer the original matrices, and (2) an aggregation step that updates V, and (3) a generalized gradient mapping step that updates the gradient $f(x)' = 2\gamma(\lambda I + S_w^i) x_i$.

(1) Within-class scatter matrices generalizing: Given the training samples and labels, we obtain the within-class scatter matrices by solving Eq. (2).

(2) Aggregation: we adopt the Nesterov's method[49]. In contrast to traditional optimization methods, the Nesterov's method has a much faster convergence rate. The Nesterov's method has as convergence rate of $O(1/d^2)$ which is higher than convergence rates of $O(1/d)$ from other gradient methods

descent. We need to computer two matrices W and V in which W is the matrix of approximate solutions, and V is the matrix of search points. The k-th iterations of W and V are described as W^k and V^k respectively. V^k is the affine combination of W^k and W^{k-1} as

$$V^k = (1 + \alpha^k)W^k - \alpha^k W^{k-1}, \tag{6}$$

where α^k is the combination coefficient.

(3) Gradient Mapping: The approximate solution W^{k+1} is computed as

$$W^{k+1} = \varphi_G(V^k - \frac{1}{\mu^K} f'(W^k)), \tag{7}$$

where $\varphi_G(u)$ is the Euclidean projection [48] problem

$$\varphi_G(u) = \min_{x \in G} \frac{1}{2} \|x - u\|^2 \tag{8}$$

The value of μ^k will increase with the iteration progress. When $f(W^{k+1}) \leq f_{\mu,V^k}(W^{k+1})$, μ^k is obtain to appropriate for V^i, where $f_{\mu,V^k}(W^{k+1}) = f(V^k) + \langle f'(V^k), W^{k+1} - V^k \rangle + \frac{\mu}{2} \left\| W^{k+1} - V^k \right\|^2$.

The proposed multi-task facial inference model is summarized in Algorithm 1.

2.4. complexity of MT-FIM

In this section, time complexity of MT-FIM is discussed. Time complexity of MT-FIM is mainly caused by solving the convex optimization Eq. (5) and Eq. (8), respectively. Suppose that we denote the number of training samples, the dimension of the samples, the number of face identification classes, the number of facial expression classes, and the desired accuracy as n, d, p, q, and e respectively. According to [49], time complexity of solving Eq. (5) is $O(nd)$. For the binary classifier, the number of tasks is just the number of classes, therefore time complexity of Eq. (8) is $O(d(p+q))$. As a summary time complexity of MT-FIM is $O(\frac{1}{\sqrt{\varepsilon}}(nd + d(p+q))$.

Algorithm 1: Proposed Multi-task Facial Inference Model algorithm

1. Input: $\{X_i, Y_i\}$, $i = 1, 2, \cdots, t$, where $X_i \in \Re^{n \times d}$ is the training samples of the i- th task, Y_i is the corresponding label vector, and max iteration number q, parameter μ^0 for iteration.
2. Output: W.
3. Set $W^1 = W^0$, $t_{-1} = 0$, $t_0 = 1$.
4. Compute S_w^i by Eq. (2).
5. for k= 1 to q do
6. Set $\alpha^k = \frac{t_{k-2}-1}{t_{k-1}}$, $V^k = (1 + \alpha^k)W^k - \alpha^k W^{k-1}$
7. while(true)
8. Set $\mu^k = \mu^{k-1} \times 2$
9. Computer $W^{k+1} = \varphi_G(V^k - \frac{1}{\mu^k} f'(W^k))$,
10. if $f(W^{k+1}) \leq f_{\mu,V^k}(W^{k+1})$ then
11. break
12. end-if
13. end-while
14. Set $W = W^k$
15. If stopping criteria satisfied then break for loop.
16. Set $t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$
17. end-if
18. end-for

3. Experiments

The proposed model is evaluated on the Cohn-Kanade [50], BU-3DFE Facial Expression [51], and Oulu-CASIA VIS [52] facial expression databases. To evaluate more comprehensively the performance of MT-FIM, in section 3.2 we test face identification and facial expression recognition in three databases respectively, then in section 3.3 we test the performance in the union of three databases. In all the methods, we used PCA to reduce the feature dimension such as 150, 300, 400, and selected the small training sample size such as 3 or 4 for each subject. Before evaluating the recognition performance of the proposed model, we want to select the common parameter of the tasks. Here the parameters are tuned by ten cross validation, the parameters are set as γ=0.001, η=0.01. For statistical stability, we generate ten different training and test dataset pairs by randomly permuting 10 times, and average classification accuracies over these splits are reported. We compare the performance of the proposed MT-FIM with the state-of-the-art classifiers, such as nearest neighbor (NN), linear regression classification (LRC)[16], linear support vector machine (SVM)[17], and sparse representation based classifier (SRC)[18]. For simple description, 'FI' and 'FE' are denoted as face identification and facial expression recognition.

3.1. Comparison methods and configurations

To verify the performance of MT-FIM, we selected the following the methods to compare.

(1) NN. Here, choice of the distance metric is Euclidean and rule of used to classify the sample is nearest.

(2) LRC. LRC uses the PCA method to reduce the dimensionality.

(3) SVM. Here, we use the one-versus-all strategy, and select RBF kernel. The radius parameter is tuned to their optimal values through cross validation.

(4) SRC. SRC is an effective classifier which codes a testing sample as a sparse linear combination of all the training samples. In order to obtain the better

performance, we select the basic pursuit algorithm. The parameter value 0.001 of the SRC is selected by cross validation.

3.2. Experiment on the facial expression databases

We first validate the performance of MT-FIM in the Cohn-Kanade, the BU-3DFE Facial Expression, and Oulu-CASIA VIS facial expression databases respectively.

1) Cohn-Kanade database: The Cohn-Kanade facial expression database consists of 100 university students with ages from 18 to 30 years. Among the students, sixty-five percent were female, fifteen percent were African-American, and three percent were Asian or Latino. Six of the displays were based on descriptions of prototypic basic emotions (i.e., joy, surprise, anger, fear, disgust, and sadness). Image sequences from neutral to target display were digitized into 640 by 480 or 490 pixel arrays with 8-bit precision for grayscale values. In our experiments, 374 sequences were selected from the database. This selection criteria was that each selected sequence to be labeled can be judge one of the six emotions. So 97 subjects were selected from the sequences and each sequence includes one labeled emotion. All these images were aligned according to the center points of the eyes. In the experiments, we select the last four frames which contain the most expressive expressions for each sequence. The images are cropped to a size of 50×50. Some sample images are shown in Figure 3. For each subject, the four images were randomly used for training, while the remaining images were used for testing. Note the training samples are labeled by the classes and expressions. Table 1 shows the FI accuracy and FE accuracy versus feature dimension by NN, LRC, SVM, SRC, and MT-FIM. It can be seen that both FI accuracy and FE accuracy of MT-FIM outperform than other methods, especially FI accuracy is at least 10% higher than other methods when dimension is 400. We can see that FI accuracy of all methods except LRC achieve their maximal accuracies at the dimension of 300, with 84.35% for NN, 90.44% for SVM, 91.12% for SRC, 96.11% for MT-FIM. From Table 1, one can see that MT-FIM still achieve high accuracy of 92.57% even if the

Figure 3: Sample images of one person on Cohn-Kanade facial expression database

training sample size is small and feature dimension is low.This is because MT-FIM makes full use of the common information of the face identification task and facial expression recognition task, and MI-FIM is stable for dimensional change. Why FI accuracy improves greater than FE accuracy in MT-FIM? The reason is that the number of the classes in face identification is larger that in facial expression recognition, therefore the number of the tasks in face identification is much more than that in facial expression recognition. As a result, face identification obtains more benefit than facial expression.

Table 2 shows the confusion matrix of facial expression obtained by MT-FIM when the dimension is 300. In the six different expressions, accuracy of anger expression is lower than that of other expressions. This is because anger expression is similar appearance with sadness expression, and anger is often misclassified as sadness.

In addition, for illustrating the convergence of MI-FIM, Figure 4 is plotted to show the change in value of the objective function in Eq. (5). We find that the objective function value decreases rapidly and then levels off, showing the fast convergence of the MI-FIM which takes no more than 10 iterations.

2) BU-3DFE Facial Expression database: BU-3DFE database contain images depicting facial expressions of Anger, Surprise, Disgust, Joy, Sadness, Fear and Neutral. The facial expressions in the BU-3DFE database are acted at four different levels of intensity. The database presently contains 100 subjects (56% female, 44% male), ranging from 18 to 70 years of age, with a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino. In our experiments, we rendered 2D facial images of 44 subjects from the database. The images are cropped and normalized face

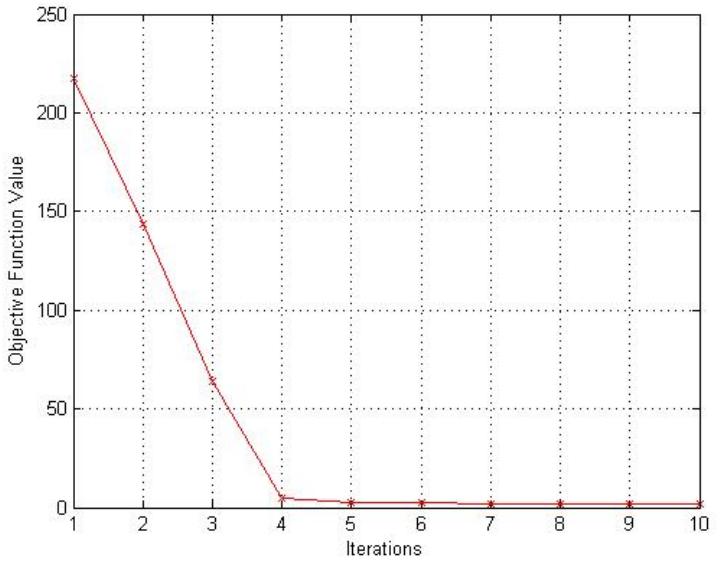

Figure 4: Convergence of objective function value

Table 1: Accuracy on Cohn-Kanade face expression database

	NN	LRC	SVM	SRC	MT-FIM
Dimensions(d=150)					
FI Accuracy	0.7782	0.9131	0.8232	0.9058	**0.9257**
FE Accuracy	0.7803	0.8713	0.9396	0.9479	**0.9434**
Dimensions(d=300)					
FI Accuracy	0.8435	0.8824	0.9044	0.9112	**0.9611**
FE Accuracy	0.8057	0.8637	0.9265	0.9223	**0.9396**
Dimensions(d=400)					
FI Accuracy	0.7866	0.8162	0.8399	0.8471	**0.9571**
FE Accuracy	0.8169	0.8726	0.9220	0.9264	**0.9353**

Table 2: Confusion matrix of facial expression obtained by MT-FIM on Cohn-Kanade facial expression database

	Anger(%)	Disgust(%)	Fear(%)	Joy(%)	Sadness(%)	Surprise(%)
Anger	**91.3**	0	2.5	0	6.2	0
Disgust	1.9	**95.2**	0	2.9	0	0
Fear	0	0	**92.4**	5.4	2.2	0
Joy	0	0	3.9	**96.1**	0	0
Sadness	5.1	0	2.1	0	**92.8**	0
Surprise	0	0	0	0	4.1	**95.9**

images of size 50×50. Some sample images are shown in Figure 5. For each subject, the four images were randomly used for training, while the remaining images were used for testing. Table 3 lists the accuracy of different methods. MT-FIM still achieves the best results in three dimensions. Influenced by the background of facial images, accuracies of all the methods decline comparing those in the above database. Particularly, FE accuracy has a great decrease for NN, LRC, and SVM. But FE accuracy of MT-FIM remains high recognition rate which is at least 10% higher than other methods. This illustrates MT-FIM method is robust and stable. We can see that facial expression recognition is easily affected by the original images, and background of face images make the expression recognition more difficult. The MT-FIM introduces the discriminative information and adopts the shared parameter to control sparsity, therefore the model is not sensitive for the background of face images. In Table 4, accuracies of six expressions are above 60%, which illustrate MT-FIM is stable in spite of a wide range of variation. This is the reason that face identification benefits facial expression recognition, resulting in better performance.

3) Oulu-CASIA VIS Facial Expression database: The experiments of Oulu-CASIA VIS Facial Expression database is more challenge. In this database, six basic emotions including anger, disgust, fear, happiness, sadness and surprise were taken from 80 subjects between 23 and 58 years old, with 73.8% of the

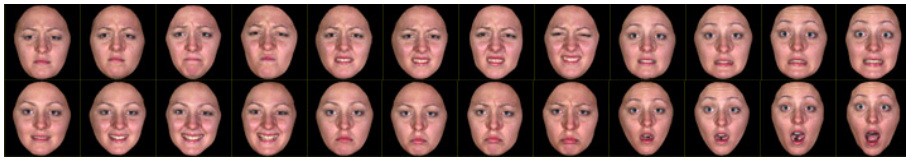

Figure 5: Sample images of one person on BU-3DFE facial expression database

Table 3: Accuracy on BU-3DFE facial expression database

	NN	LRC	SVM	SRC	MT-FIM
Dimensions(d=150)					
FI Accuracy	0.6319	0.6578	0.6566	0.6581	**0.6691**
FE Accuracy	0.2876	0.3058	0.4436	0.5641	**0.6699**
Dimensions(d=300)					
FI Accuracy	0.6289	0.6255	0.6254	0.6311	**0.6615**
FE Accuracy	0.2881	0.3532	0.4382	0.5548	**0.6572**
Dimensions(d=400)					
FI Accuracy	0.6258	0.6437	0.6011	0.6122	**0.6661**
FE Accuracy	0.2934	0.4012	0.4650	0.5478	**0.6557**

Table 4: Confusion matrix of facial expression obtained by MT-FIM on BU-3DFE facial expression database

	Anger(%)	Disgust(%)	Fear(%)	Joy(%)	Sadness(%)	Surprise(%)
Anger	**62.3**	6.3	11.7	0	19.7	0
Disgust	9.3	**64.7**	15.6	10.4	0	0
Fear	0	0	**64.2**	20.9	14.9	0
Joy	0	15.6	18.2	**66.2**	0	0
Sadness	25.3	0	10.6	0	**64.1**	0
Surprise	0	0	14.8	0	16.8	**68.4**

subjects are males. All the images were taken under the challenging uncontrolled light condition. All expressions are captured in three different illumination conditions: normal, weak and dark. Normal illumination means that good normal lighting is used. Weak illumination means that only computer display is on and subject sits on the chair in front of the computer. Dark illumination means near darkness. We used PCA to reduce the feature dimension to 300. The images are cropped to a size of 50×50. Some sample images are shown in Figure 6. For each subject, the three images were randomly used for training, while the remaining images were used for testing. Table 5 shows the performance of five methods under the dark, strong, weak illumination conditions, respectively. Clearly, SRC achieves better results than NN, LRC, SVM, while MT-FIM works the best. In the strong illumination condition, FI accuracy of MT-FIM is at least more than 2.5% improvement than other methods, and FE accuracy of MT-FIM is slightly more that SVM but much more than NN, LRC. Both FI accuracy and FE accuracy of all the methods are the best in the strong illumination, and those in the dark illumination are the worst. It implies that the illumination condition is important for face identification and facial expression recognition. Because the sparse constraint is also robust for the illumination, MT-FIM achieves the best performance in three different illumination conditions. From Table 6, we can see that accuracy of surprise is still higher than those of other expressions. This is because appearance of surprise is more different than those of other expressions, which is not easy to misclassify.

3.3. Experiment on the union of the facial expression databases

To evaluate more comprehensively the performance of MT-FIM, we joint three facial expression databases of the Cohn-Kanade, BU-3DFE Facial Expression, and Oulu-CASIA VIS, then learn their classifiers based on more samples. For convenient comparison, we conduct the same experiments setting with above experiments. We used PCA to reduce the feature dimension to 300. The images are cropped to a size of 50×50. In the Cohn-Kanade and BU-3DFE facial Expression database, four images were randomly used for training, while in the

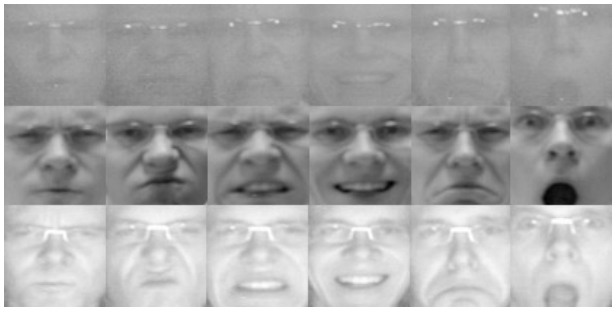

Figure 6: Sample images of one person from three illumination conditions on Oulu-CASIA VIS facial expression database. The first to third rows are images including six expressions under the dark, strong, weak illumination conditions, respectively.

Table 5: Accuracy on Oulu-CASIA VIS face expression database

	NN	LRC	SVM	SRC	MT-FIM
Dimensions(d=150)					
FI Accuracy	0.6493	0.6809	0.7179	0.7233	**0.7482**
FE Accuracy	0.4286	0.4958	0.5863	**0.5998**	0.5993
Dimensions(d=300)					
FI Accuracy	0.6810	0.6998	0.7337	0.7411	**0.7927**
FE Accuracy	0.5172	0.5512	0.6929	0.6931	**0.7013**
Dimensions(d=400)					
FI Accuracy	0.6691	0.6820	0.7242	0.7299	**0.7515**
FE Accuracy	0.4619	0.5321	0.6030	0.6083	**0.6153**

Table 6: Confusion matrix of facial expression obtained by MT-FIM on Oulu-CASIA VIS facial expression database

	Anger(%)	Disgust(%)	Fear(%)	Joy(%)	Sadness(%)	Surprise(%)
Anger	**61.4**	24.1	0	0	14.5	0
Disgust	14.9	**69.7**	0	0	15.4	0
Fear	0	0	**62.3**	20	17.7	0
Joy	0	0	11.3	**77.2**	11.5	0
Sadness	19.5	18.1	0	0	**62.4**	0
Surprise	0	0	8.3	0	8.6	**83.1**

Oulu-CASIA VIS facial expression database, three images were randomly used for training. In the description below, MT-FIM (Union) is denoted as MT-FIM method in the union of three databases. In Figure 7, the first and second column show FI accuracy and FE accuracy respectively. The first to third row show accuracy in the Cohn-Kanade, BU-3DFE, Oulu-CASIA VIS database respectively. We can see that both FI accuracy and FE accuracy achieve the improvement by MT-FIM (Union). Due to their features of images in three facial expression databases are similar, and the tasks are related, therefore MT-FIM(Union) can adopts more training samples and learning more tasks. This demonstrates that more related tasks can improve their performances, which more validate our idea. In Figure 7, we notice that the FE accuracy of MT-FIM (Union) is slightly worse than that of MT-FIM in the Cohn-Kanade database when the dimension is 150. Because union of different databases has influence on feature space, MT-FIM may misclassify. In spite of this, MT-FIM (Union) brings the improvement of the performance in face identification and facial expression recognition.

4. Conclusion

In this paper, we proposed a multi-task model for simultaneous face identification and facial expression recognition (MT-FIM) to handle the small sample

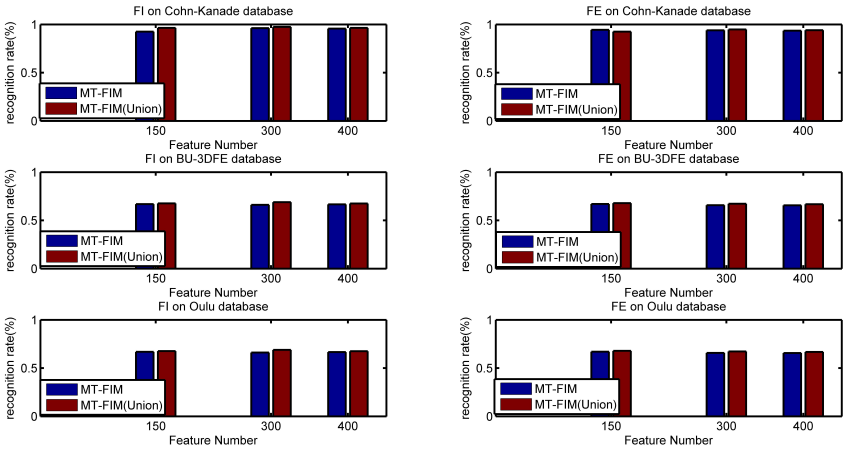

Figure 7: Comparison of performance on Union facial expression database. The first and second column show FI accuracy and FE accuracy respectively. The first to third row show accuracy in the Cohn-Kanade, BU-3DFE, Oulu-CASIA VIS database respectively

size problem of face identification and facial expression recognition. Apart from the improved accuracy, one important advantage of MT-FIM is that it can simultaneously learn the classifiers of face identification and facial expression recognition. In order to improve the discrimination of the model, MT-FIM introduces the within-class scatter matrix which is adjusted to minimize the within-class distance and maximize the distance between different classes. We evaluated the proposed model on three different facial expression databases. The experimental results clearly demonstrated that the proposed MT-FIM has much better performance than the state-of-the-art methods. Especially MT-FIM is more competitive for small training samples. Nonetheless, performance of MT-FIM will be increased slightly when more facial expression databases are jointed, and experiments in more face databases need to be conducted, which is considered in the future work.

Acknowledgement

This work is partially supported by the Project funded by China Postdoctoral Science Foundation Under grant No. 2014M5615556, supported by the National Science Foundation of China (61273300, 61232007) and Jiangsu Natural Science Funds for Distinguished Young Scholar (BK20140022). And, it is also partially supported by grants 15KJB520024 from Jiangsu University Natural Science Funds, supported by grants KFKT2014B18 from the State Key Laboratory for Novel Software Technology from Nanjing University, supported by grants 30920140122007 from Project supported by Jiangsu Key Laboratory of Image and Video Understanding for Social Safety. Finally, the authors would like to thank the anonymous reviewers for their constructive advice.

References

[1] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Computer Vision and Pattern Recognition (CVPR), 2005.

[2] M. Pantic, L. Rothkrantz, Automatic analysis of facial expressions: the state of art, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1424–1445.

[3] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognition 36 (2003) 259–275.

[4] M. Pantic, L. Rothkrantz, Toward an affect-sensitive multimodal human computer interaction, in: Proceeding of the IEEE, Vol. 91, 2003, pp. 1370–1390.

[5] C. Shan, S. Gong, P. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, Image and Vision Computing 27 (2009) 803–816.

[6] Y. Yacoob, L. Davis, Recognizing human facial expression from long image sequences using optical flow, IEEE Trans. Pattern Anal. Mach. Intell. 18 (6) (1996) 636–642.

[7] I. Essa, A. Pentland, Coding analysis interpretation and recognition of facial expressions, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1999) 1357–1362.

[8] M. J. Lyons, M. Lyons, S. Akamatsu, Automatic classification of single facial images, IEEE Trans. Pattern Anal. Mach. Intell. 12 (6) (1999) 1357–1362.

[9] G. Donato, M. Bartlett, J. Hager, P. Ekman, T. Sejnowski, Classifying facial actions, IEEE Trans. Pattern Anal. Mach. Intell. 21 (10) (1999) 974–989.

[10] Y. Tian, T. Kanade, J. Cohn, Recognizing action units for facial expression analysis, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 97–115.

[11] S. Li, J. Lu, Face recognition using nearest feature line method, IEEE Trans. Neural Network 10 (1999) 439–443.

[12] J. Chien, C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 1644–1649.

[13] J. Laaksonen, Local subspace classifier, in: Int'l Conf. Artificial Neural Networks, 1997.

[14] K. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 684–698.

[15] S. Li, Face recognition based on nearest linear combinations, in: Computer Vision and Pattern Recognition (CVPR), 1998.

[16] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 2106–2112.

[17] B. Heisele, P. Ho, T. Poggio, Face recognition with support vector machine: Global versus component-based approach, in: IEEE Int'l Conf. Computer Vision (ICCV), 2001.

[18] J. Wright, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.

[19] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.

[20] M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with gabor occlusion dictionary, in: European Conf. Computer Vision(ECCV), 2010.

[21] H. Zheng, J. Xie, Z. Jin, Heteroscedastic sparse representation classification for face recognition, Neural Processing Letters 35 (3) (2012) 233–244.

[22] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, Y. Ma, Face recognition with contiguous occlusion using markov random fields, in: IEEE Int'l Conf. Computer Vision (ICCV), 2009.

[23] P. Ekman, W. Friesen, Facial action coding system: Investigator's guide, consulting Psychologists Press (1993).

[24] M. Bartlett, G. Littlewort, L. Fasel, J. Movellan, Real time face detection and expression recognition: Development and application to human-computer interaction, in: Computer Vision and Pattern Recognition (CVPR), 2003.

[25] F. Bourel, C. Chibelushi, A. Low, Robust facial expression recognition using a state-based model of spatiallylocalised facial dynamic, In: Proc. of Automatic Face and Gesture Recognition (2002) 113–118.

[26] Y. Tian, T. Kanade, J. Cohn, Recognizing action units for facial expression analysis, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 97–115.

[27] H. Meng, N. Bianchi-Berthouze, Naturalistic affective expression classification by a multi-stage approach based on hidden markov models, Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science 6975 (2011) 378–387.

[28] I. Cohen, N. Sebe, A. Garg, L. Chen, T. S. Huang, Facial expression recognition from video sequences: Temporal and static modeling, Computer Vision and Image Understanding 91 (2003) 160–187.

[29] I. Kotsia, I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines, IEEE Trans. Image Process. 16 (2007) 172–187.

[30] N. Sebe, M. S. Lew, I. Cohen, Y. Sun, T. Gevers, T. Huang, Authentic facial expression analysis, Image and Vision Computing 25 (2007) 1856–1863.

[31] C. Antonioand, F. Brendanand, H. Thomas, Face recognition with contiguous occlusion using markov random fields, in: Computer Vision and Pattern Recognition (CVPR), 1999.

[32] D. Singh, P. Gupta, U. Tiwary, Face recognition using facial expression: A novel approach, in: Proceedings of SPIE, Mobile Multimedia/Image Processing, Security, and Applications, Vol. 6982, 2008.

[33] S. Negahban, M. Wainwright, Estimation of (near) low-rank matrices with noise and high-dimensional scaling, The Annals of Statistics 39 (2) (2011) 1069–1097.

[34] T. Pong, P. Tseng, S. Ji, J. Ye, Trace norm regularization: Reformulations, algorithms, and multi-task learning, SIAM Journal on Optimization 20 (6) (2010) 3465–3489.

[35] S. Negahban, M. Wainwright, Joint support recovery under high-dimensional scaling: Benefits and perils of l_1 regularization, in: Advances in Neural Information Processing Systems(NIPS), 2008.

[36] A. Argyrious, T. Evgenious, M. Pontil, Convex multi-task feature learning, Machine Learning 73 (2008) 243–272.

[37] R. Caruana, Multi-task learning, Machine Learning 28 (1) (1997) 41–75.

[38] J. Baxter, A model of inductive bias learning, Journal of Articial Intelligence Research 12 (2000) 149–198.

[39] B. Bakker, T. Heskes, Task clustering and gating for bayesian multitask learning, JMLR 4 (2003) 83–99.

[40] A. Schwaighofer, V. Tresp, K. Yu, Learning gaussian process kernels via hierarchical bayes, in: Advances in Neural Information Processing Systems(NIPS), 2005.

[41] N. Lawrence, J. Platt, Learning to learn with the informative vector machine, ICML (2004).

[42] T. Evgeniou, M. Pontil, Regularized multi-task learning, SIGKDD (2004) 109–117.

[43] R. Ando, T. Zhang, A framework for learning predictive from multiple tasks and unlabeled data, JMLR 6 (2005) 1817–1853.

[44] S. Parameswaran, K. Weinberger, Large margin multi-task metric learning, NIPS 23 (2010) 1867–1875.

[45] Y. Zhang, D. Yeung, Transfer metric learning by learning task relationships, SIGKDD (2010) 1199–1208.

[46] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1996) 267–288.

[47] T. T. Wu, K. Lange, Coordinate descent algorithms for lasso penalized regression, The Annals of Applied Statistics 2 (2008) 224–244.

[48] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press (2004).

[49] Y. Nesterov, Introductory lectures on convex optimization: A basic course, Kluwer Academic Publishers (2003).

[50] T. Kanade, J. Cohn, Y. Tian, Comprehensive database for facial expression analysis, In: Proc. of Automatic Face and Gesture Recognition 2 (2000) 46–53.

[51] J. Wang, L. Yin, X. Wei, Y. Sun, 3d facial expression recognition based on primitive surface feature distribution, In: Proc. of CVPR (2006) 1399–1406.

[52] G. Zhao, X. Huang, M. Taini, S. Li, M. Pietikainen, Facial expression recognition from near-infrared videos, Image and Vision Computing 29 (2011) 607–619.