

# Mapping Forests: A Comprehensive Approach for Nonlinear Mapping Problems

Mahdi Jampour<sup>1,2</sup> · Mohammad-Shahram Moin<sup>2</sup> ·  
Lap-Fai Yu<sup>3</sup> · Horst Bischof<sup>1</sup>

Received: 23 September 2016 / Accepted: 25 July 2017 / Published online: 5 August 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** A new and robust mapping approach is proposed entitled mapping forests (MFs) for computer vision applications based on regression transformations. Mapping forests relies on learning nonlinear mappings deduced from pairs of source and target training data, and improves the performance of mappings by enabling nonlinear transformations using forests. In contrast to previous approaches, it provides automatically selected mappings, which are naturally nonlinear. MF can provide accurate nonlinear transformations to compensate the gap of linear mappings or can generalize the nonlinear mappings with constraint kernels. In our experiments, we demonstrate that the proposed MF approach is not only on a par or better than linear mapping approaches and the state-of-the-art, but also is very time efficient, which makes it an attractive choice for real-time applications. We evaluated the efficiency and performance of the MF approach using the BU3DFE and Multi-PIE datasets for multi-view facial expression recognition application, and Set5, Set14 and SuperTex136 datasets for single image super-resolution application.

**Keywords** Automatic nonlinear mapping · Mapping forests · Multi-view facial expression recognition · Nonfrontal facial expression recognition · Single image super-resolution

## 1 Introduction

The first ideas of decision forests and subsequently random forests (RFs) belongs to more than two decades ago. Due to their fast processing speed, forests have been extensively used to solve computer vision problems especially those requiring real-time processing. An extensive study on decision forests and their applications is provided by Criminisi et al. [4]; where they discussed models for classification, manifold, supervised and semi-supervised learning, regression, density estimation, etc., by means of decision forests. They also discussed the advantages and disadvantages of decision forests and extended the idea of decision forests to solve continuous problems such as regression and density estimation.

We define “mapping” or data (features) mapping as a function that transforms data from a source space to a destination space, where tasks such as classification can be processed more easily. There exist several problems, which can be solved using mapping approaches or by transforming from a source space to a new target space. Linear mappings by means of basic linear regression, ridge regression, logistic regression or other extension are examples which increasingly attracted attention in many computer vision applications [16,21]. Indeed, since most of the real problems are not linear, multiple solutions have been proposed for them, using kernel-based nonlinear methods. However, choosing an appropriate nonlinear kernel is difficult because it needs behavioral systems analysis (BSA) of the problem, which is not always feasible for complicated problems such as those involving

---

✉ Mahdi Jampour  
jampour@icg.tugraz.at

Mohammad-Shahram Moin  
moin@itrc.ac.ir

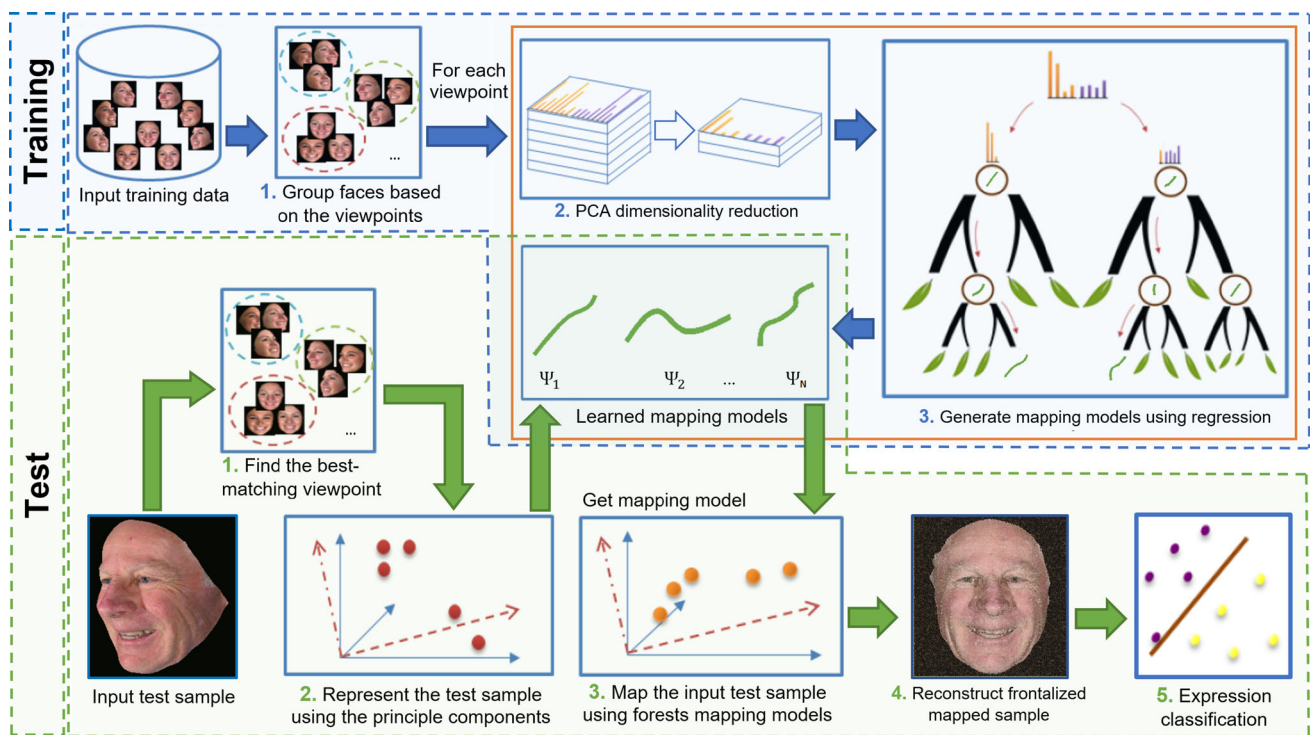
Lap-Fai Yu  
craigyu@cs.umb.edu

Horst Bischof  
bischof@icg.tugraz.at

<sup>1</sup> Graz University of Technology, Graz, Austria

<sup>2</sup> Information and Communication Technology Research Institute, Tehran, Iran

<sup>3</sup> University of Massachusetts, Boston, MA, USA



**Fig. 1** Mapping forests architecture for nonfrontal face analysis. Training: we classify the training samples based on the viewpoints and reduce the dimension using PCA for each viewpoint. MF estimates then the mapping models by means of forests which provides nonlinear mappings from ensemble of linear transformations. Testing: the best subset

is selected based on the viewpoint and represent the test sample in the new space spanned by the principal components (obtained from the training step). We transform then the features to the target space using detected learned models

humans and human faces. Therefore, proposing an automatic approach which is able to provide nonlinear mappings is desirable. On the other hand, there are methods, which provide a set of linear mappings, as a solution for nonlinear mapping. Forests are ensemble learning approaches capable of learning several linear mappings for nonlinear problems.

In this work, we propose an approach for automatic learning of multiple nonlinear mappings using decision forests. In continuation of the definitions provided in [19–21], as an application of nonfrontal face analysis, we define forests for each specific head pose from a discrete set of possible head poses. To determine which forests to use for a new input sample, we rely on another regressor to predict the head pose of the face. Since the mappings are adapted to the pose of the input face, this approach yields significantly better results than using a single mapping [21, 24]. For each specific pose, the input test sample is applied to the corresponding forests to explore the best mapping. The selected mapping among the forests is naturally nonlinear and can produce very accurate transformations, due to the ensemble of linear mappings.

Our main contribution is exploring continuous mapping functions using a highly efficient structure of decision forests

trained with partial data. This model is trained to discriminatively map from a source space to a target space. The main advantage of our approach over (local or global) linear mapping techniques is its ability to provide learned nonlinear mapping using ensemble of linear mappings on each specific subset of data. Moreover, our nonlinear mappings are determined automatically through the mapping forests. As an illustrative example, Fig. 1 shows a general overview of our MF approach applied to the problem of nonfrontal face analysis, which includes both the training and test steps. We have evaluated the performance of our MF approach on two well-known computer vision open problems, namely multi-view facial expression recognition (MFER) and single image super-resolution (SISR).

The remaining of the paper is organized as follows. Section 2 is dedicated to reviewing related works. In Sect. 3, we explain the theory of Random Decision Forests. Section 4 includes our approach of mapping forests. In Sects. 5 and 6, we present mapping forests for multi-view facial expression recognition and image super-resolution, as two computer vision problems solved with our approach. Finally, Sects. 7 and 8 belong to Discussions and Conclusions, respectively.

## 2 Related Work

There are multiple linear and nonlinear approaches for mapping data from source to target space [19, 27, 31]. Most of the approaches use the more discriminant space for classification or segmentation purposes. Moreover, there are attempts that provide robustness by means of mapping approaches [1] or supervised mappings between noisy data and ground truth data [13]. A recent approach proposed by [21] introduced pairwise linear transformation by means of linear ridge regression to map data from source space to destination space. Although they achieved state-of-the-art, most of the real problems are naturally nonlinear that cannot be handled using linear transformations properly. Another approach [19] performed nonlinear mapping that fills the gap of [21], but a challenge still remains, since the employed nonlinear kernel needs to be defined manually, which is a difficult task. Performing a general nonlinear mapping has multiple applications in computer vision such as image denoising, pseudo-coloring, super-resolution and facial analysis. In addition, forests as an efficient structure have been widely used for other related works [7, 8, 13, 32]. For instance, Fanello et al. [13] introduced filter forests and explained its ability for predicting continuous variables. They proposed to learn optimal filtering kernels and to employ such kernels for image denoising, where the kernel size and values are determined based on spatial context. Similarly, Schuster et al. [32] extended decision forests to regression forests for super-resolution applications. Successes in extending decision/random forests for different computer vision applications show that the forests-based approaches can not only achieve high accuracy, but also fast processing speed due to their intrinsic structure of the trees, compared to other techniques. These factors motivated us for extending forests to solve mapping problems. Therefore, we aim to perform efficient nonlinear mappings automatically to achieve better or comparable results as those of the state-of-the-art.

### 2.1 Multi-view Facial Expression Recognition

The problem of facial expression recognition (FER) is one of the hot spots in facial analysis. The first attempt was made by Ekman and Friesen [11]. They addressed the problem of facial expression recognition by proposing facial action coding systems (FACS) to recognize basic expressions. They introduced codes for facial action units in such a way that each facial expression changes the codes. A group of the units refers to an expression that can be found from the detected codes. Their study focused on six basic expressions, including anger (AN), disgust (DI), fear (FE), happiness (HA), sadness (SA) and surprise (SU), in contrast to neutral (NE). Several subsequent approaches attacked the problem with different techniques [28]. They

can be categorized into three categories: (1) appearance-based approaches which use texture information by means of appearance feature descriptors such as LBP, HOG and SIFT; (2) geometric-based approaches that benefit from facial geometric information such as landmarks displacements; and (3) hybrid methods that employ both appearance and geometric information.

The progress in this area became tremendous, while new and more complex challenges like multi-view facial expression recognition (MFER) attracted attentions. The reason is that assuming a frontal view in real-world applications is unrealistic and may limit applicability. Moreover, some of the facial features, which are necessary for expression recognition are not available or only partially available on nonfrontal faces. Therefore, MFER is still an active problem in face analysis with potential applications in human computer interaction (HCI), children education, games, etc.

Currently, the most successful approaches on MFER including Zheng [44]; Rudovic et al. [29] and Jampour et al. [21] map or find the relationship between facial features extracted from nonfrontal views and their corresponding features in the frontal view by mapping all the available training data to a common reference space. Rudovic et al. [30] used the same mapping regardless of the pose of the head; Zheng [44] proposed a complex method that relies on a time-consuming optimization process; Jampour et al. [20] proposed linear mapping between nonfrontal and frontal pairs, while the problems are nonlinear due to the variations in gender, age, ethnicity, skin tone, etc. They recently proposed an extension of their idea as a nonlinear approach with improvement in performance [19]. Their new approach outperformed the state-of-the-art, nevertheless, the problem is that kernel-based approaches are strongly dependent on the kernel definition and only a few well-known kernel functions are shown to work well.

### 2.2 Single Image Super-Resolution

The problem of single image super-resolution is another well-studied problem and hot topic within the computer vision community that aims to provide high-resolution images from corresponding low-resolution images. There are various techniques for image super-resolution. The traditional approaches include bilinear, bicubic or Lanczos interpolation. The modern approaches include statistical image priors [14, 34] and dictionary-based methods [41]. One of the successful directions for image super-resolution is using supervised machine learning approaches, where a model is trained from one or some training samples and used for test images [32]. In this work, we also provide supervised single image super-resolution using our MF approach to show the generality of our method. The details are discussed in Sect. 6.

### 3 Random Decision Forests

We begin with a brief review of random decision forests, which are highly efficient structure for many computer vision applications. It has also been used in different classification and regression tasks [8,22,23,33]. A random decision forest is an ensemble of random decision binary trees, where a tree  $f_i(x) : \mathcal{X} \rightarrow \mathcal{Y}$  classifies sample  $x \in \mathcal{X} \subseteq \mathcal{R}^n$  through the depth of tree to the leaves. In the following, we briefly review both random and regression forests.

#### 3.1 Random Forest

Random forest is an ensemble of randomized trees. Each tree is built, trained and tested independently from other trees. The training data for each tree are generated from subsampling of the original data. During the training, each node splits the training data into subsets using splitting functions, as shown in Eq. 1:

$$\psi(x, \gamma) = \begin{cases} 0 & \text{if } r_\gamma(x) < 0 \\ 1 & \text{otherwise,} \end{cases} \quad (1)$$

where  $\gamma$  defines the response function  $r_\gamma(x)$ . There are several kinds of response functions  $r(\cdot)$ , which have been used in different tasks [23,35]. For instance,  $r_\gamma(x) = x[\gamma_1] - \gamma_{th}$  was used in [32]. They defined operator  $[\cdot]$ , for selecting one dimension of  $x$  such that  $\gamma \in [1..len(x)]$  and  $\gamma_{th}$  is an arbitrary threshold. In addition, there are also adapted splitting functions for image analysis, for example, the distribution of Hough votes which is described in [12]. Each tree of the forest refers to  $f_i(x) : \mathcal{X} \rightarrow \mathcal{Y}$  and the forest is  $F = \{f_1, \dots, f_T\}$ , where  $T$  is the number of trees. The probability of class example  $k$  in the case of classification task is:

$$p(k|x) = \frac{1}{T} \sum_{t=1}^T p_t(k|x), \quad (2)$$

$$C(x) = \arg \max_{k \in \mathcal{Y}} p(k|x), \quad (3)$$

where  $p_t(k|x)$  is the density of class label  $k$  estimated by the  $t$ th tree and  $C(x)$  is the final class label.

#### 3.2 Regression Forest

Regression forest is a set of randomly trained regression. A regression tree splits a complex nonlinear regression problem into several smaller problems, which are easier to solve [4]. Regression forest has been widely used for nonlinear regression of a pair of explanatory and dependent variables. The main difference between random and regression forests

is the continuous nature of regression forest, which can work with continuous data, where both training and test labels are continuous. In applying regression forests, we can use different types of objective functions (e.g., linear, polynomial, probabilistic) for a subspace of input data [7]. Again, similar to the classification case, output of the regression forest is average of all  $T$  trees.

In our approach, we define the mapping function  $f$  to be dependent on continuous input data for mapping from source to target space. In the following, we explain how our MF approach can efficiently provide mapping functions using decision forests.

### 4 Mapping Forests (MFs)

Mapping forests, inspired from decision forests, consist of randomized trees, which are adapted for our specific problem. We split the data into smaller subsets in each node of the trees, and then perform pairwise ridge regression for each subset of data at leaf nodes, to learn the best optimized mapping solution. The optimization problem is formulated such that the objective function is computed by minimizing the mapping error between pairs of training data.

#### 4.1 Mapping as Objective Function

In the following, we illustrate how mapping forests can be applied for the problem of nonfrontal face analysis. Let  $X$  be a set of aligned vectorized features between frontal and nonfrontal views, which is extracted by appearance-based descriptors as described in 4.3. The notation  $X_{NF}$  refers to a set of nonfrontal facial features and  $X_{Fr}$  refers to their corresponding frontal facial features in our MFER problem 5. Similarly,  $X_L$  refers to a set of vectorized features of low-resolution images, where their corresponding high resolution is denoted as  $X_H$  in our SISR problem as described in 6. Note that in the rest of the paper, we use the same notations. In MFER problem,  $X_{\theta_i}$  is a subset of facial features in  $X_{NF}$  from viewing angle  $\theta_i$ , where  $X_{\theta_i} = [I_1^{\theta_i}, I_2^{\theta_i}, \dots, I_N^{\theta_i}]$  is a matrix of size  $(q \times N)$ , and refers to the  $N$  vectorized facial features denoted by  $I_k^{\theta_i} \in \mathbb{R}^{(q \times 1)}$ . Note that  $I_k^0$  and  $I_k^{\theta_i}$  are columned feature vectors of the  $k$ th facial image of the training data from the same person in different poses. Based on this, we define pairwise sets of training data,  $X_{Fr}$  and  $X_{\theta_i}$ , where the former is the set of frontal and the latter is a set of corresponding nonfrontal features in viewing angle  $\theta_i$ . With respect to Eq. 4, we need to find function  $f(\cdot)$  using training data and then extend it for test samples, in such a way that the summation in Eq. 4 is minimized. Eq. 5 is a specific case of Eq. 4 in case of linear condition where its closed form solution is provided by ridge regression in Eq. 6.

$$\arg \min_{f(\cdot)} \sum_{k=1}^N \left( x_{Fr}^k - f(x_\theta^k) \right)^2 \tag{4}$$

$$\arg \min_M \|X_{Fr} - MX_\theta\|_2^2 + \lambda \|M\|_2^2 \tag{5}$$

$$M = X_{Fr} \left( X_\theta^T X_\theta + \lambda I \right)^{-1} X_\theta^T \tag{6}$$

We learn mapping function  $M$  through the forest that is dependent on input training data  $X_\theta$  and its correspondence of  $X_{Fr}$ . Our models architecture, which is based on forests, performs an ensemble of linear mappings that is more efficient than a unique linear mapping for nonlinear problems, as it splits a complex problem into several smaller problems. An estimation of the frontalized sample is given by:

$$\hat{x}_{\theta \rightarrow 0} = Mx_\theta \tag{7}$$

where  $x_\theta$  is a vectorized input test sample in an arbitrary viewing angle  $\theta$ ; similarly  $\hat{x}$  is a vector of features mapped into the frontal. In Eq. 6, parameter  $\lambda$  allows us to solve the singularity problem and makes the problem nonsingular even if  $(X_\theta^T X_\theta)$  is not invertible. It can be shown that its total variance is a monotonically decreasing sequence with regard to  $\lambda$ . Furthermore,  $X_\theta$  can also be replaced in Eq. 6 by nonlinear kernels, such as polynomial kernel and radial basis function (Gaussian) kernel. We propose to use MF for minimizing the loss function in a way that selects a branch of each node in the tree with maximum matches with the input vector. Therefore, we use Eq. 7 as the objective function to decrease the mapping error in each node and consequently having ensemble of linear mappings at the leaves. In the following subsections, we provide more details about our model.

### 4.2 Training the Trees

Similar to [4,32] we learn all trees in MF independently and define  $S_j = \{x_{Fr}, x_\theta\} \in \mathcal{X} \times \mathcal{Y}$  as a subset of training data for a given node  $j$ . The goal is to find the splitting function  $\psi(x_\theta, \gamma_i)$  at  $S_j$  such that it splits the data into two branches and minimizes the entropy as in Eqs. 8–12. The splitting starts from the root through a branch of the tree to the leaves. We define  $\gamma_i = x_\theta\{i - \eta, \dots, i + \eta\}$ , where  $\eta = \sqrt{\text{len}(x_\theta)}/2$ . This means that  $2\eta + 1$  members of the feature vectors  $x_\theta$  decide for splitting at each node of the trees.

$$\psi_j = \arg \min_{\psi} I(S_j, \psi) \tag{8}$$

$$I(S_j, \psi) = H(S_j) - \sum_{k \in \{L, R\}} \frac{|S_j^k|}{|S_j|} H(S_j^k) \tag{9}$$

our specific form of regression information gain is as:

$$H(S) = -\frac{1}{|S|} \sum_{x_\theta \in S} \int_{x_{Fr}} p(x_{Fr}|x_\theta) \log p(x_{Fr}|x_\theta) dx_{Fr} \tag{10}$$

where  $H(S)$  is average entropy for subset  $S$ . Inspired from [4], by means of Gaussian distribution for conditional probability  $p(x_{Fr}, x_\theta)$  we rewrite:

$$p(x_{Fr}, x_\theta) = N(x_{Fr} : \bar{x}_{Fr}(x_\theta), \sigma_{x_{Fr}}^2(x_\theta)) \tag{11}$$

$$I(S_j, \psi) \propto \sum_{x_\theta \in S_j} \log(\sigma_{x_{Fr}}(x_\theta)) - \sum_{k \in \{L, R\}} \left( \sum_{x_\theta \in S_j^k} \log(\sigma_{x_{Fr}}(x_\theta)) \right). \tag{12}$$

where  $S_j^L$  and  $S_j^R$  are left and right children at node  $j$  and  $x_\theta, x_{Fr}$  are source and target data, respectively. The output is as Eq. 13, such that we have  $T$  (e.g., 50) output mappings obtained from individual trees and use average of them.

$$M_{\text{Final}} = \frac{1}{T} \sum_{t=1}^T M_t, \tag{13}$$

where  $M_t$  is predicted mapping function of  $t$ th trees.

### 4.3 Features Types

In this work, we employed the basic features for the input face images  $\mathcal{I}$  in all three color channels. The features are extracted by HOG [6] and LBP [26] descriptors with cell size 25 pixels that are concatenated as a feature vector for each train and test sample. The main motivation to employ these two descriptors is that HOG provides gradient information of the images, whereas LBP describes the intensity. The concatenation of these two feature descriptors creates large feature vectors and very large dataset. Therefore, it will be expensive to process such data during the learning through the multiple forests in terms of both time and space complexity, while unnecessary features can be either compressed or ignored. To this end, we employed the well-known dimensionality reduction technique, PCA, and applied it to each cluster. The different datasets that we used have feature vectors with dimensions greater than 14,000 (e.g., BU3DFE >17,800 and Multi-PIE >14,000 dimensions). We select  $d = 500$  in applying the PCA to reduce the dimensionality of both training and test data to a constant number. This means that the first 500 principal components are selected for this purpose, which makes the similarity equal to or greater



**Fig. 2** Visualization of raw-based samples. Nonfrontal faces are transformed to frontal using MF despite different variations present in the input face images

than 98% of the original data. For test samples, we extracted, selected and concatenated features similarly as training samples and then constructed the feature vectors by means of these 500 principal components.

#### 4.4 Forests Setup

We set up our MF approach using 50 trees with four levels of maximum depth on basic features reduced to 500 dimensions as described in Sect. 4.3. The trees' predictions are averaged on the number of all trees and mapping functions at the leaves concatenate subset of data through the depth of the trees. Increasing the number of trees can make the results more stable but does not bring noticeable change, as depicted in Fig. 4. Also we found that increasing the depth further than four levels, will just increase the running time without bringing noticeable improvement, as we experienced. On the other hand, the results may fluctuate with the type of features or number of features used in splitting each node. Nevertheless, the runtime information is described in Sect. 5.3. The proposed framework can be generally used for solving mapping problems. In the following sections, we present the result of using the proposed approach to solve two well-known computer vision problems, i.e., multi-view facial expression recognition and image super-resolution.

### 5 Mapping Forests for Multi-view Facial Expression Recognition

We performed quantitative and qualitative experiments on MFER using two popular datasets. Also compare our results with the state-of-the-art works. The datasets are used with different protocols, therefore, we first explain the protocols in the following and then discuss the results. We use  $k$ -fold ( $k =$

5) cross-validation on both datasets; therefore, the results in the rest of the paper are averaged by 5 validations. In the training step, we categorize training data into several smaller groups using supervised techniques based on the viewpoints that is explained in Sect. 5.2. We then learn mapping models for each group using our MF approach.

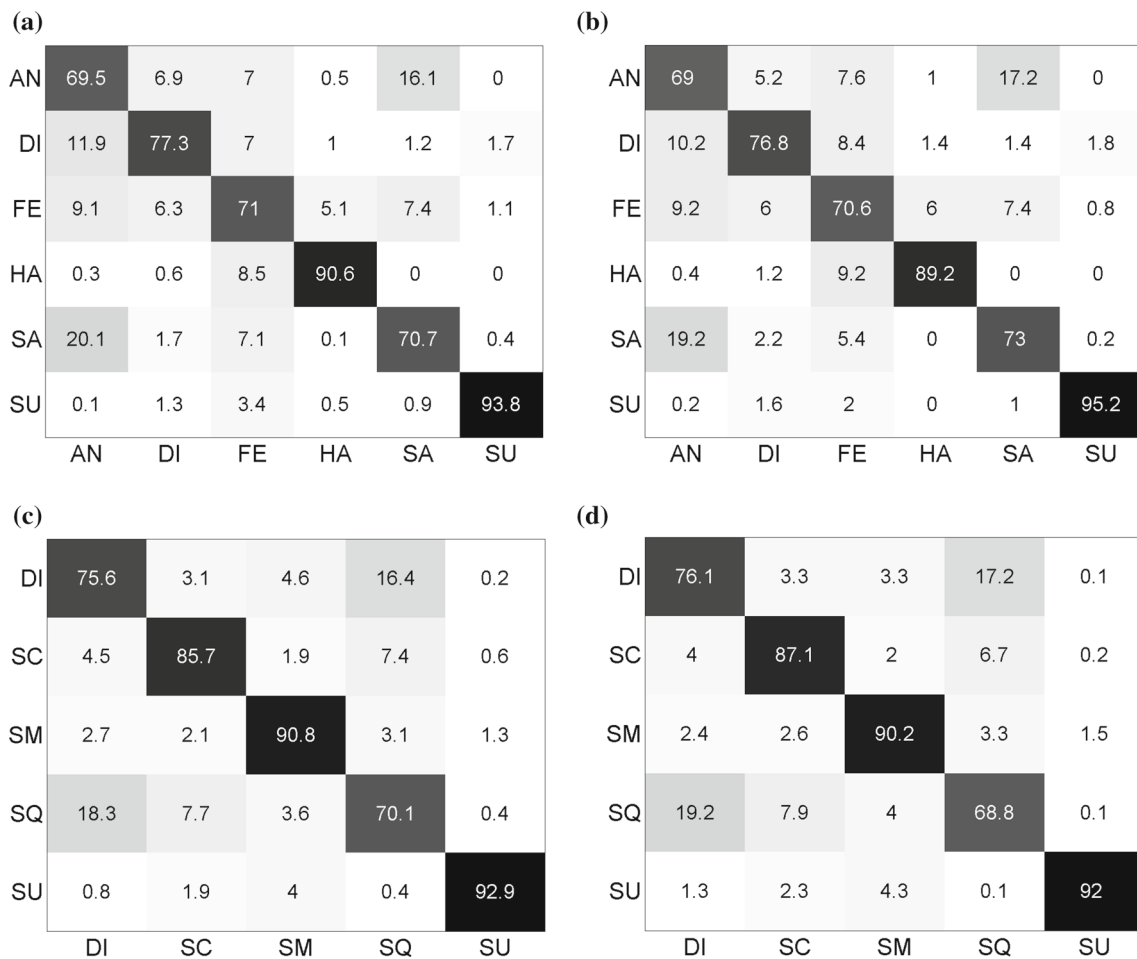
In the test step, we first approximate the class of the test samples based on the viewpoints and then use the corresponding mapping model provided by MF. Finally, the mapped data is classified by linear SVM [3] for expression recognition. The advantage of this pipeline is that, each test sample with various head poses (which are not available in our training data) will be adapted with the closest subset. Therefore, the variation of the head pose will be handled using our MF approach. Consequently, the pipeline can produce frontal faces which are useful for facial analysis, and also provide robustness against head pose. Figure 2 depicts a visualization of several test samples in frontal view generated by our MF approach from a nonfrontal viewpoint. Although the reconstructed frontal faces are blurry, due to high dimensionality reduction, they are still useful for expression recognition. Note that the raw features are used here just for visualization.

#### 5.1 MFER Datasets

In this section, we introduce two well-known and popular facial expression datasets. We used protocols, similar to that of related works, in order to perform equitable comparisons.

##### 5.1.1 BU3DFE

BU3DFE is a publicly available dataset containing 3D scanned faces of 100 subjects with six basic expressions.



**Fig. 3** Confusion matrices for multi-view facial expression recognition using mapping forests. **a** BU3DFE—first protocol, **b** BU3DFE—second protocol, **c** Multi-PIE—first protocol, **d** Multi-PIE—second protocol

**Table 1** Pose-specific MFER using Random Forests and SVM in comparison with Mapping forests

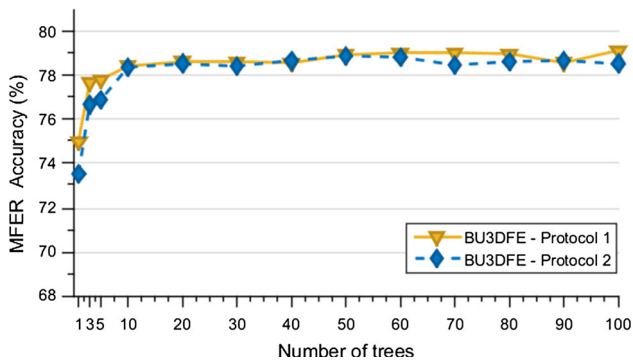
Dataset	BU3DFE		Multi-PIE	
	P1	P2	P1	P2
Random Forests	59.93%	58.10%	68.95%	68.49%
PSC-SVM [21]	77.66%	76.36%	80.94%	81.55%
Mapping forests	78.92%	78.84%	82.89%	82.85%

More details can be found in [42]. We rendered multiple views from the 3D faces using seven pan angles ( $0^\circ$ ,  $\pm 15^\circ$ ,  $\pm 30^\circ$ ,  $\pm 45^\circ$ ) and five tilt angles ( $0^\circ$ ,  $\pm 15^\circ$ ,  $\pm 30^\circ$ ). Altogether there are 35 viewpoints containing 21,000 samples in our first protocol. In addition, we generated the second protocol with views  $0^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$  and  $90^\circ$  with 3000 samples. Each vector has a dimension-

ality of 17,832, which is reduced to 500 by means of PCA.

### 5.1.2 Multi-PIE

CMU Multi-PIE is a multi-purpose dataset containing 337 subjects taken with different poses between  $-90^\circ$  to  $90^\circ$  with an interval of  $15^\circ$ . Altogether there are 13 different viewpoints [15]. The dataset contains five facial expressions, and we selected 145 subjects that have all the expressions available. We cropped facial regions using a semiautomatic algorithm into the dimension of  $175 \times 200$  pixels. In order to evaluate our model, we use two protocols: (1) Protocol 1, similar to [18, 21] containing 13 viewpoints; and (2) Protocol 2, similar to [25, 44] containing 7 viewpoints. Each feature vector has a dimensionality of 14,952, which is reduced to 500 (similar to the BU3DFE features) using PCA.



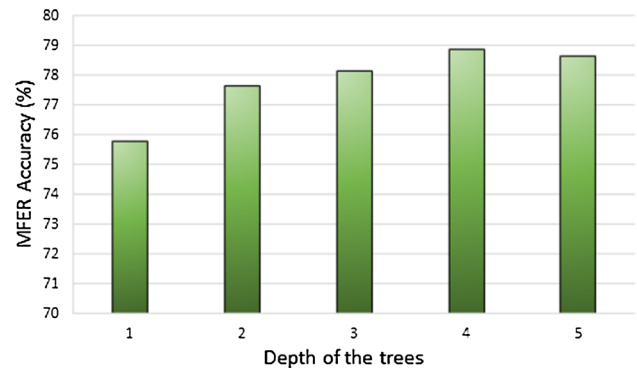
**Fig. 4** Effects of increasing the number of trees on the final MFER accuracy

## 5.2 Supervised Data Assortment

Splitting the data into several smaller subsets is a convenient idea to improve recognition accuracy. Differences in expression, viewpoint, age, skin tone, ethnicity, etc., account for the most important variations among the data. Therefore, if we mitigate the effects imposed by these variations, the final accuracy will increase. The trick is to split the data based on the viewpoints as we have the information of head poses during the training. We could use either supervised or unsupervised methods for data assortment. We used the well-known classification method, linear SVM [3], for our purpose. More specifically, we re-labeled all training data by means of viewpoint information and trained a SVM classifier using this data. In the test phase, a new, unseen test sample is first classified into the correct subset and then we employed the mapping model of that subset, which has been trained by MF, to map the test sample into the frontal view. Table 1 shows the results of the averaged expression recognition with the classified data as baseline, in comparison with the results of the MF.

## 5.3 Experimental Results

The results in Table 1 are produced from each specific viewpoint and then averaged through all of them that compared with our mapping forests approach. The comparison includes all four protocols (two protocols from BU3DFE and two protocols from Multi-PIE). Moreover, Fig. 3 depicts their confusion matrices, where can be seen that most confusion occurs between sadness and anger in both protocols of BU3DFE, and similarly, between disgust and squint in both protocols of Multi-PIE. The best discrimination in both datasets is on surprise and smile (happy) due to their clear variations. Furthermore, the time complexity in both training and test parts are strongly dependent on the feature vectors, number of samples in each protocols and hardware speci-



**Fig. 5** Effects of increasing depth of the trees on the final MFER accuracy (assessed on the BU3DFE-P2)

fications. We used a PC with Intel Core i7-3930 processor and 32 GB RAM. Our first protocol of BU3DFE includes 16,800 training images (80 training samples  $\times$  6 classes of expressions  $\times$  35 viewpoints) with 500 dimensions. We first generated a classification model by means of SVM for estimating the head pose among 35 viewpoints. It took 581 s which could be considered as offline preprocessing. Then, the next time-consuming process was training the forests. As explained earlier, we used 50 trees with a maximum depth of 4. The average training time on our PC was 748 s for each viewpoint. This means that our total training process is about 35 viewpoints  $\times$  748 s  $\times$  fivefold cross-validation + 581 s of head pose estimation, which is about 131,481 s or about 37 h. The running time during the test step is 84 s for estimating the viewpoints of 120 samples, which means 0.7 s per sample for head pose estimation and 1.4 s for predicting mapping of 120 test samples, which means 12 ms for each sample. As we expected, the running time of the forest is very short due to the efficient structure of the trees, through the training step, which could be considered as an offline process, needs several hours. In addition, the effects of increasing the depth and number of trees are shown in Figs. 4 and 5.

## 5.4 Comparison with the Related Work

In this section, we compare our approach with the state-of-the-art on both protocols of BU3DFE and Multi-PIE. Table 2 shows that MF is on a par with the state-of-the-art and related works in both protocols of BU3DFE and Multi-PIE. In addition, [24] proposed an approach similar to PSC in [21] but that is based on a new descriptor called LGBP. They have reported 80.17% accuracy on Multi-PIE dataset with 7 viewpoints similar to Multi-PIE-P2 but using six expressions from 100 subjects. [44], on the same dataset, reported 81.7% for their GSRRR method whereas our MF performs 82.85% for 7 viewpoints (same as those in the previous works) on 5 expressions and 145 subjects. Table 2



**Table 2** Multi-view facial expression recognition comparison between the proposed MF approach and the state-of-the-art methods

Method	Dataset	Accuracy
BDA/GMM [45]	BU3DFE-Protocol 1	68.20
EHMM [36]	BU3DFE-Protocol 1	75.30
GSCF [37]	BU3DFE-Protocol 1	76.10
SSVQ [38]	BU3DFE-Protocol 1	76.34
SSE [39]	BU3DFE-Protocol 1	76.60
PSR [21]	BU3DFE-Protocol 1	78.04
KPSNM [19]	BU3DFE-Protocol 1	79.26
<b>Mapping forest</b>	BU3DFE-Protocol 1	<b>78.92</b>
$LB P^{ms}$ [25]	BU3DFE-Protocol 2	72.43
DNPE [18]	BU3DFE-Protocol 2	72.47
LPP [17]	BU3DFE-Protocol 2 <sup>a</sup>	73.06
LGBP [25]	BU3DFE-Protocol 2	77.67
PSR [21]	BU3DFE-Protocol 2	77.87
KPSNM [19]	BU3DFE-Protocol 2	78.79
<b>Mapping forest</b>	BU3DFE-Protocol 2	<b>78.84</b>
DNPE [18]	Multi-PIE-Protocol 1 <sup>b</sup>	76.83
PSR [21]	Multi-PIE-Protocol 1	81.96
KPSNM [19]	Multi-PIE-Protocol 1	82.43
<b>Mapping forest</b>	Multi-PIE-Protocol 1	<b>82.89</b>
PSR [21]	Multi-PIE-Protocol 2	82.55
KPSNM [19]	Multi-PIE-Protocol 2	83.09
<b>Mapping forest</b>	Multi-PIE-Protocol 2	<b>82.85</b>

Our approach results are shown in bold

<sup>a</sup>With 4 level of intensities

<sup>b</sup>100 subjects instead of 145 in our protocol

also shows that there are different approaches to address the problem of MFER while our MF approach is one of the best techniques for MFER, being very intuitive, fast to apply and accurate, making it an ideal choice for practical applications.

## 6 Mapping Forests for Image Super-Resolution

In the previous section, we showed the efficiency of MF in multi-view facial expression recognition as an application. In this section, we use MF to solve another well-known problem in computer vision; i.e., image super-resolution including multi-images super-resolution and single image super-resolution, where the latter is more complicated since the training data are limited. We then compare our results with those of the state-of-the-art methods.

Image super-resolution (ISR) is an important and hot topic in computer vision that has several applications, such as image manipulation and enhancing medical images, astronomical images, and old videos. There are various techniques for image super-resolution. The traditional approaches

include bilinear, bicubic and Lanczos interpolation. The modern approaches include using statistical image priors [14,34], dictionary-based methods [41], etc. One of the successful directions for ISR is using supervised machine learning approaches, in which a model is trained using training samples and then is evaluated on test images. A training sample is first down-scaled into a small image (e.g., image  $I_O$  of  $512 \times 512$  into image  $I_D$  of  $256 \times 256$  with a down-sampling scale 2). We then up-sample image  $I_D$  into a bigger image with the original scale (e.g., image  $I_D$  of  $256 \times 256$  into image  $I_U$  of  $512 \times 512$ ). All the image resizing uses standard bicubic interpolation. Image  $I_O$  and Image  $I_U$  have the same size; image  $I_U$  is low resolution due to the resizing and missing information.

Image  $I_U$  is a copy of image  $I_O$  but in low resolution. We refer to the low-resolution image as LR and the high-resolution image as HR throughout the rest of the paper. We would like to recover LR to be as similar as possible to HR, which is an ill-posed problem. The solution for such problems is to learn from multiple image patches to find the best mapping function from LR to HR. Unlike the dictionary-based approaches, our proposed approach does not need multiple training images and it can work even with a low-resolution input sample. The details are described in the following.

### 6.1 Image Super-Resolution Learning Model

We define two sets of vectorized patches from LR and HR images as  $X_L$  and  $X_H$ , respectively, where LR is down-sampled and then upsampled using basic bicubic interpolation. LR and HR images have the same size and the patch size in our work is  $4 \times 4$  pixels. Therefore,  $X_L$  corresponds to  $X_H$  with the same size of  $N \times 16$ . We would like to find a mapping function  $f$ :

$$\arg \min_{f(\cdot)} \sum_{i=1}^N \|x_H^i - f(x_L^i)\|^2 \quad (14)$$

This is similar to Eq. 4 and the solution is:

$$M = X_H(X_L^T X_L + \lambda I)^{-1} X_L^T \quad (15)$$

$$\hat{x}_H = M x_L \quad (16)$$

We discussed, in Sect. 4, that our mapping forests solves this ridge regression by means of decision forests. MF provides mappings for each patch from LR to HR. We note that our learning part is based on only one input sample  $I$ , where we generate  $I^{1/s}$  in such a way that  $I^i$  is a resized image of  $I$  with scale factor  $i$ . During the training, we learn the relation between low and high resolution of the images using our mapping model. In test phase, we estimate an original image

**Table 3** Average PSNR on Set5, Set14 and SuperTex136 datasets with different scale factors

Dataset	Factor	Bicubic interpolation (%)	Zeyde [43] (%)	ANR [40] (%)	SRCNN [9] (%)	SRCNN (SoTA) [10] (%)	Our MF approach (%)
Set5	2×	33.66	35.78	35.83	36.34	36.66	36.03
Set5	3×	30.39	31.90	31.92	32.39	32.75	32.26
Set5	4×	28.42	29.69	29.69	30.09	30.49	29.86
Set14	2×	30.10	31.81	31.80	32.00	32.45	32.05
Set14	3×	27.42	28.67	28.65	28.86	29.30	28.84
Set14	4×	25.90	26.88	26.85	27.07	27.50	27.05
SuperTex136	2×	27.98	28.10	28.26	29.44	NA	30.32
SuperTex136	3×	25.40	25.33	25.38	26.37	NA	26.62
SuperTex136	4×	23.97	24.52	24.54	24.70	NA	24.48

We used only one current low-resolution sample to train our model (single image super-resolution)

**Table 4** Average PSNR on Set5, Set14 and SuperTex136 (SuperTex) datasets with different scale factors

Dataset	Scale factor	Learned by	Bicubic interpo. (%)	Our MF approach (%)
Set5	2×	Set14	33.66	35.82
Set5	3×	Set14	30.39	32.73
Set5	4×	Set14	28.42	30.81
Set5	2×	SuperTex	33.66	36.80
Set5	3×	SuperTex	30.39	30.69
Set5	4×	SuperTex	28.42	30.21
Set14	2×	Set5	30.10	31.28
Set14	3×	Set5	27.42	28.29
Set14	4×	Set5	25.90	26.89
Set14	2×	SuperTex	30.10	31.38
Set14	3×	SuperTex	27.42	26.80
Set14	4×	SuperTex	25.90	26.52
SuperTex	2×	Set5	27.98	30.47
SuperTex	3×	Set5	25.40	27.44
SuperTex	4×	Set5	23.97	26.05
SuperTex	2×	Set14	27.98	30.54
SuperTex	3×	Set14	25.40	27.49
SuperTex	4×	Set14	23.97	26.09

We used cross-dataset to train our mapping model

$I$  from  $I^{1/s}$  where the unavailable pixels are estimated by our MF approach using the available pixels. This is our single image super-resolution setup, which is valid as we do not use the ground truth. On the other hand, low-resolution input image information is always available for any new test sample by down-sampling it.

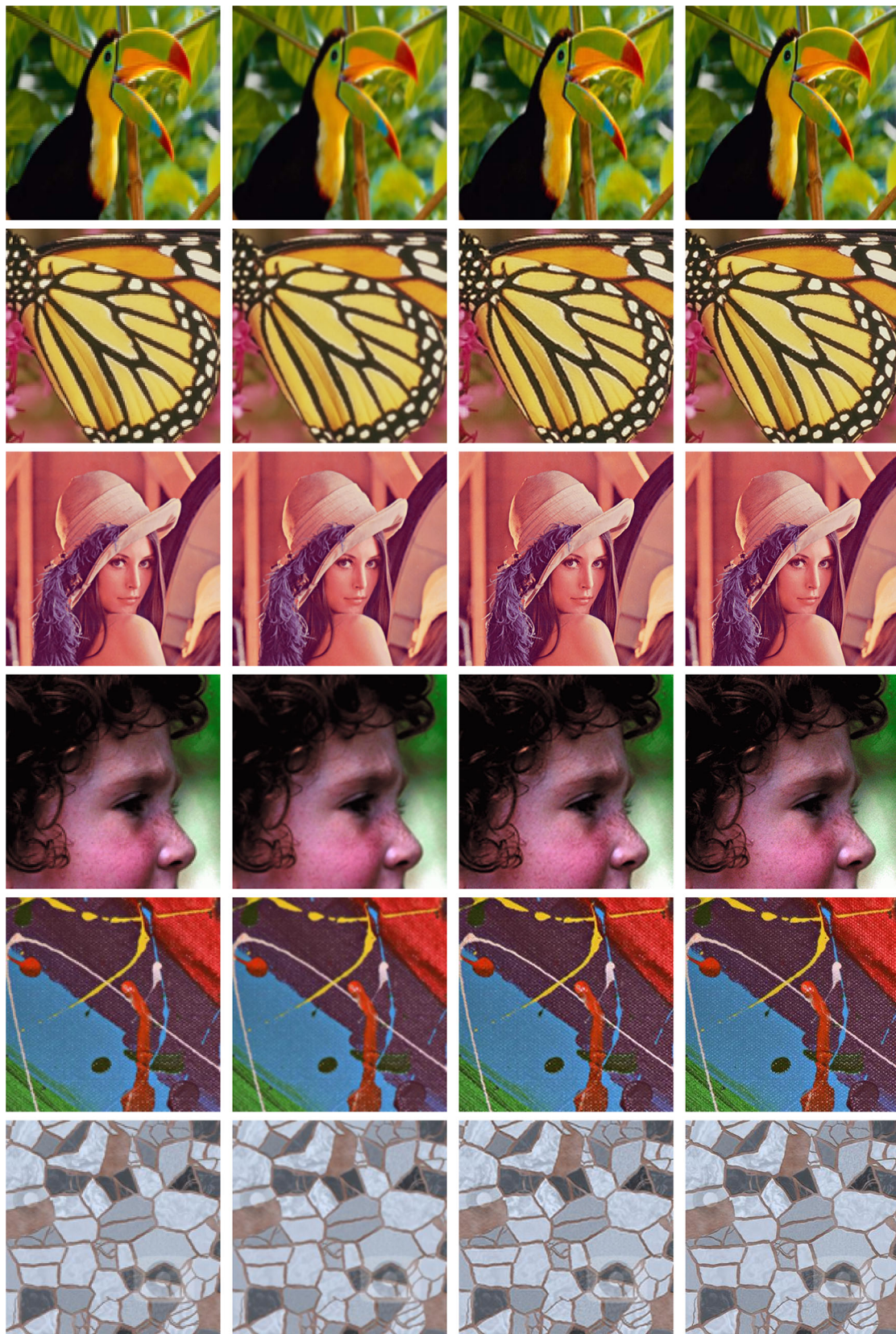
### 6.2 Image Super-Resolution Datasets

In this section, we introduce three popular datasets used for image super-resolution application and describe our protocols, which is intended to be similar with the related works, as much as possible, for comparison purposes. We have provided results by means of MF with two validation protocols:

(1) single image super-resolution (SISR), where we use only the input test sample in low resolution for training our model in Sect. 6.1; and (2) cross-dataset, where we use a dataset to train our model and tested it on another dataset.

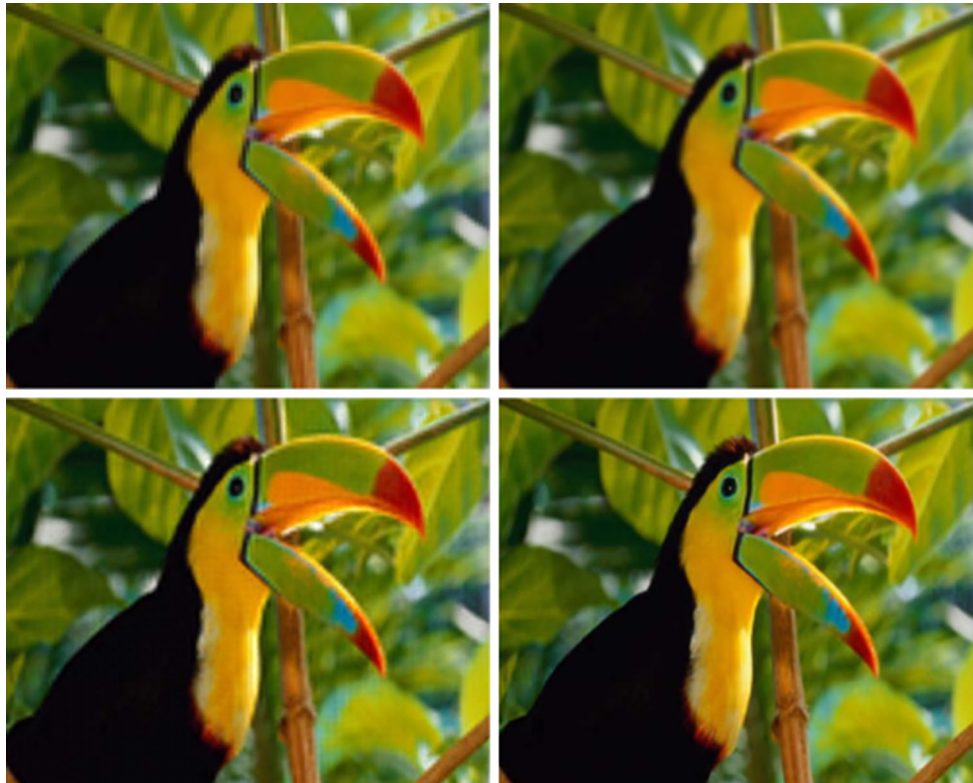
#### 6.2.1 Dataset5, Set14 and SuperTex136

Set5 [2] and Set14 [43] include, respectively, 5 and 14 images in different conditions, including grayscale or color images taken from nature, human, animal, and city building. Images in Set5 are square in different size, but Set14 images are without any constraint. In contrast, SuperTex136 [5] is a dataset with 136 grayscale or color texture images with a size of



**Fig. 6** Single image super-resolution results by means of mapping forests. *First column* is the input samples from various datasets (Set5, Set14 and Tex136); *second column* is the results by means of bicubic

interpolation; *third column* is the results using our MF approach; and the *last column* shows the original images



**Fig. 7** Our MF results on single image super-resolution with  $4\times$  zoom for qualitative comparison. *Top-left* input low resolution sample. *Top-right* result of Bicubic interpolation. *Bottom-left* our MF result. *Bottom-right* the ground truth

$300 \times 300$ . Two samples from all datasets are shown in Fig. 6 (first column).

### 6.3 Experimental Results

The results of the first protocol are reported in Table 3. Table 4 shows the results using the same datasets but the training process involves using several images from the cross-dataset (protocol 2). Note that in case of Set5, we only used 5 images to learn our model. A qualitative comparison is also shown in Fig. 6, which shows that our MF approach is successful for image super-resolution problem too. Figure 7 provides illustrative and comparison results in  $4\times$  zoom.

## 7 Discussions

Our approach is a comprehensive technique for mapping problems and can automatically consider an ensemble of mappings that provide a nonlinear solution for the input problem. We believe that most computer vision problems are naturally nonlinear and finding an appropriate kernel for them is not straightforward. Therefore, automatic nonlinear approaches are desirable for this kind of problems, and the proposed mapping forests is a promising approach toward

this direction. In our experiments, we showed that each part of the source space can be transformed to a favorable destination space by means of the proposed MF approach. MF defines a set of mapping functions that provide appropriate, continuous nonlinear transformations. Following to our previous study [21], we showed that our automatic MF approach can efficiently provide nonlinear solution instead of kernel-based approach. In that work, we found also that small pose errors cannot significantly affect the results. As we use similar technique of mappings, we expect the same robustness while our current approach is nonlinear. Although in SISR problem our MF approach does not outperform the state-of-the-art CNN-based method (SRCNN) [10], its advantage is that it can be automatically and efficiently applied to different applications. Moreover, our approach does not need a very large dataset for training, as CNN-based method does.

## 8 Conclusion

In this paper, we proposed mapping forests (MFs) as a comprehensive approach for solving nonlinear mapping problems. Our MF approach relies on decision forests that exploits continuous mapping functions for transforming from a source space to a destination space. It can automatically deal

with varying conditions in the input data. We demonstrated the efficacy of our approach on two well-known computer vision problems, multi-view facial expression recognition (MFER) and image super-resolution (ISR). We compared the results of our approach with that of the state-of-the-art and related work approaches in these problems, and found that our approach outperforms or is on a par with them.

**Acknowledgements** Open access funding provided by Graz University of Technology. We would like to thank Iranian National Elites Foundation for its financial support and Iranian ICT Research Institute (ITRC) for providing complementary supports for this work.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Åström, F., Felsberg, M., Baravdish, G.: Mapping-based image diffusion. *J Math Imaging Vis* **57**(3), 293–323 (2017)
- Bevilacqua, M., Roumy, A., Guillemot, C., line Alberi Morel, M.: Low-complexity single-image super-resolution based on non-negative neighbor embedding. In: Proceedings of BMVC, pp. 135.1–135.10. BMVA (2012)
- Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 27:1–27:27 (2011)
- Criminisi, A., Shotton, J.: *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, London (2013)
- Dai, D., Timofte, R., Gool, L.V.: Jointly optimized regressors for image super-resolution. In: Eurographics, p. 1 (2015)
- Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
- Dollar, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: BMVC, pp. 1–11 (2009)
- Dollar, P., Zitnick, C.: Structured forests for fast edge detection. In: ICCV, pp. 1841–1848 (2013)
- Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV, pp. 184–199. Springer, Berlin (2014)
- Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016)
- Ekman, P., Friesen, W.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto (1978)
- Fanelli, G., Yao, A., Noel, P., Gall, J., Gool, L.: Hough forest-based facial expression recognition from video sequences. In: ECCV Workshops, pp. 195–206 (2010)
- Fanello, S., Keskin, C., Kohli, P., Izadi, S., Shotton, J., Criminisi, A., Pattacini, U., Paek, T.: Filter forests for learning data-dependent convolutional kernels. In: CVPR, pp. 1709–1716 (2014)
- Fattal, R.: Image upsampling via imposed edge statistics. *ACM Trans. Graph.* **26**(3), 95–102 (2007)
- Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. *Image Vis. Comput.* **28**, 807–813 (2010)
- Han, H., Gan, L., Liu, S., Guo, Y.: A novel measurement matrix based on regression model for block compressed sensing. *J. Math. Imaging Vis.* **51**(1), 161–170 (2015)
- Hu, Y., Zeng, Z., Yin, L., Wei, X., Tu, J., Huang, T.: Multi-view facial expression recognition. In: FG, pp. 1–6 (2008)
- Huang, X., Zhao, G., Pietikainen, M.: Emotion recognition from facial images with arbitrary views. In: BMVC, pp. 76.1–76.11 (2013)
- Jampour, M., Lepetit, V., Mauthner, T., Bischof, H.: Piecewise non-linear regression for multi-view facial expression recognition. *Image Vis. Comput.* **58**, 38–46 (2017)
- Jampour, M., Mauthner, T., Bischof, H.: Multi-view facial expressions recognition using local linear regression of sparse codes. In: Computer Vision Winter Workshop (CVWW), pp. 1–8 (2015)
- Jampour, M., Mauthner, T., Bischof, H.: Pairwise linear regression: an efficient and fast multi-view facial expression recognition. In: FG, pp. 1–8 (2015)
- Kazemi, V., Burenius, M., Azizpour, H., Sullivan, J.: Multi-view body part recognition with random forests. In: BMVC, pp. 1–11 (2013)
- Kotschieder, P., Bulo, S., Pelillo, M., Bischof, H.: Structured labels in random forests for semantic labelling and object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 2104–2116 (2014)
- Moore, S., Bowden, R.: Multi-view pose and facial expression recognition. In: BMVC, pp. 1–11 (2010)
- Moore, S., Bowden, R.: Local binary patterns for multi-view facial expression recognition. *Comput. Vis. Image Underst.* **115**, 541 (2011)
- Ojala, T., Pietikinen, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: ICPR, pp. 582–585 (1994)
- Raftopoulos, K.A., Kollias, S.D.: The globallocal transformation for noise resistant shape representation. *Comput. Vis. Image Underst.* **115**(8), 1170–1186 (2011)
- Roychowdhury, S., Emmons, M.: A survey of the trends in facial and expression recognition databases and methods. [arXiv:1511.02407](https://arxiv.org/abs/1511.02407) (2015)
- Rudovic, O., Pantic, M., Patras, I.: Coupled Gaussian processes for pose-invariant facial expression recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1357–1369 (2013)
- Rudovic, O., Patras, I., Pantic, M.: Regression-based multi-view facial expression recognition. In: ICPR, pp. 4121–4124 (2010)
- Sanchez-Lozano, E., Martinez, B., Valstar, M.F.: Cascaded regression with sparsified feature covariance matrix for facial landmark detection. *Pattern Recogn. Lett.* **73**, 19–25 (2016)
- Schulter, S., Leistner, C., Bischof, H.: Fast and accurate image upscaling with super-resolution forests. In: CVPR, pp. 3791–3799 (2015)
- Schulter, S., Leistner, C., Wohlhart, P., Roth, P., Bischof, H.: Accurate object detection with joint classification-regression random forests. In: CVPR, pp. 923–930 (2014)
- Shao, W.Z., Ge, Q., Deng, H.S., Wei, Z.H., Li, H.B.: A unified optimization perspective to single/multi-observation blur-kernel estimation with applications to camera-shake deblurring and non-parametric blind super-resolution. *J. Math. Imaging Vis.* **54**(2), 216–239 (2016)
- Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: CVPR, pp. 1–8 (2008)
- Tang, H., Hasegawa-Johnson, M., Huang, T.: Non-frontal view facial expression recognition based on ergodic hidden markov model supervectors. In: ICME, pp. 1202–1207 (2010)
- Tariq, U., Yang, J., Huang, T.: Multi-view facial expression recognition analysis with generic sparse coding feature. In: ECCV, pp. 578–588 (2012)
- Tariq, U., Yang, J., Huang, T.: T: Maximum margin GMM learning for facial expression recognition. In: FG, pp. 1–6 (2013)

39. Tariq, U., Yang, J., Huang, T.: Supervised super-vector encoding for facial expression recognition. *Pattern Recogn. Lett.* **46**, 89–95 (2014)
40. Timofte, R., De, V., Gool, L.V.: Anchored neighborhood regression for fast example-based super-resolution. In: *ICCV*, pp. 1920–1927 (2013)
41. Wang, S., Zhang, L., Liang, Y., Pan, Q.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: *CVPR*, pp. 2216–2223 (2012)
42. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.: A 3d facial expression database for facial behavior research. In: *FG*, pp. 211–216 (2006)
43. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: *7th International Conference on Curves and Surfaces*, Avignon, France. Revised Selected Papers, pp. 711–730. Springer, Berlin (2012)
44. Zheng, W.: Multi-view facial expression recognition based on group sparse reduced-rank regression. *IEEE Trans. Affect. Comput. S.* **5**, 71 (2014)
45. Zheng, W., Tang, H., Lin, Z., Huang, T.: Emotion recognition from arbitrary view facial images. In: *ECCV*, pp. 490–503 (2010)



**Mahdi Jampour** received his B.Sc. degree in Computer Science from the University of Kerman, Iran, in 2006. He received his M.Sc. degree in Artificial Intelligence from the Islamic Azad University (IAU), Mashhad, Iran in 2009 and Ph.D. in Computer Vision from Graz University of Technology, Graz, Austria in 2016. He received the Best M.Sc. Thesis Award from the IAU Mashhad and full Ph.D. scholarship from the Iran Ministry of Science, Research and

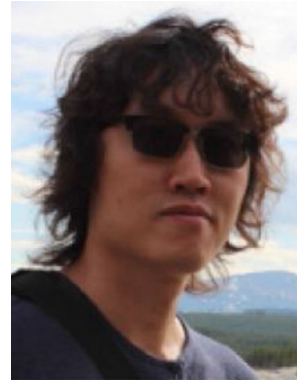
Technology (MSRT). He was also awarded a Marshal Plan Fellowship from the Austrian Marshall Plan Foundation to visit the University of Massachusetts at Boston in 2015. He is currently postdoc researcher at Information and Communication Technology Research Institute. His research interests are in the areas of computer vision including recognition, detection, verification and identification.



**Mohammad-Shahram Moin** received his B.Sc. degree from AmirKabir University of Technology, Iran, in 1988; M.Sc. degree from University of Tehran, Iran, in 1991; and Ph.D. degree from École Polytechnique de Montréal, Montréal, Canada, in 2000, all degrees in electrical engineering. He worked as research associate in Scribens Laboratory, École Polytechnique de Montréal, Canada, and as project manager in biometric and big data topics in ICT Research

Institute (Iran Telecommunication Research Center), Tehran, Iran. Moin is IEEE senior member, associate professor and lecturer of artificial neural networks, digital signal processing, pattern recognition and stochastic processes courses in graduate levels. He has published more

than 100 peer reviewed scientific papers. His current research interests are pattern recognition, image processing, biometrics and big data analytics.



**Lap-Fai Yu** is an assistant professor at the University of Massachusetts Boston, where he directs the Graphics and Virtual Environments Lab. He received his B.Eng. (first class honors) and M.Phil. degrees in computer science from the Hong Kong University of Science and Technology (HKUST) in 2007 and 2009, respectively, and his Ph.D. degree in computer science from the University of California, Los Angeles (UCLA), in 2013, where he received the Cisco Outstanding Graduate Research Award. He was a visiting scholar at Stanford University and a visiting scientist at the Massachusetts Institute of Technology. His research interests include computer graphics and computer vision.



**Horst Bischof** received his M.S. and Ph.D. degree in computer science from the Vienna University of Technology in 1990 and 1993, respectively. In 1998, he got his Habilitation (venia docendi) for applied computer science. Currently, he is vice rector for Research at Graz University of Technology and professor at the Institute for Computer Graphics and Vision at the Graz University of Technology, Austria. Bischof is member of the scientific board of Joanneum

Research and board member of the Fraunhofer-Institut für Graphische Datenverarbeitung (IGD). His research interests include object recognition, visual learning, motion and tracking, visual surveillance and biometrics, medical computer vision, and adaptive methods for computer vision where he has published more than 580 peer-reviewed scientific papers. Horst Bischof was General Chair of CVPR 2015 and was chairman of the DAGM/AGM conference 2012 and co-chairman of international conferences (ICANN, DAGM), and local organizer for ICPR96. He was program co-chair of ECCV 2006 and several times Area chair of all major vision conferences. Currently, he is associate editor for IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition, Computer and Informatics and the Journal of Universal Computer Science. Horst Bischof is member of the European academy of sciences and has received several (19) awards among them the 29th Pattern Recognition award in 2002; the main prize of the German Association for Pattern Recognition DAGM in 2007 and 2012, the Best Scientific Paper Award at BMCV 2007, and the Best Demo Award at 2012 and the Best Scientific Paper Awards at the ICPR 2008, ICPR 2010, PCV2010, AAPR 2010 and ACCV 2012.