Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 703514, 11 pages
http://dx.doi.org/10.1155/2015/703514

Research Article

Head Pose Estimation with Improved Random Regression Forests

Gaoli Sang, Hu Chen, and Qijun Zhao

State Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University Chengdu, Sichuan 610064, China

Correspondence should be addressed to Hu Chen; huchen@scu.edu.cn

Received 20 May 2015; Accepted 30 September 2015

Academic Editor: Panos Liatsis

Perception of head pose is useful for many face-related tasks such as face recognition, gaze estimation, and emotion analysis. In this paper, we propose a novel random forest based method for estimating head pose angles from single face images. In order to improve the effectiveness of the constructed head pose predictor, we introduce feature weighting and tree screening into the random forest training process. In this way, the features with more discriminative power are more likely to be chosen for constructing trees, and each of the trees in the obtained random forest usually has high pose estimation accuracy, while the diversity or generalization ability of the forest is not deteriorated. The proposed method has been evaluated on four public databases as well as a surveillance dataset collected by ourselves. The results show that the proposed method can achieve state-of-the-art pose estimation accuracy. Moreover, we investigate the impact of pose angle sampling intervals and heterogeneous face images on the effectiveness of trained head pose predictors.

1. Introduction

Research on head pose estimation has been attracting increasing attention of researchers thanks to its importance in a wide variety of applications, such as behavior analysis, gaze estimation, fatigue driving detection, and face recognition. During the past decades, a number of methods have been proposed. In general, these methods can be roughly divided into two major categories depending on whether or not they require facial landmarks in pose estimation. A thorough review on various head pose estimation methods can be found in [1].

The methods using facial landmarks estimate pose angles mostly based on the assumption that the geometrical structure among the facial landmarks is relevant to the pose angles. Some of the methods [2–5] in this category fit a 3D face model to the landmarks detected on 2D face images by changing pose angles, and pick as the final result the angles which make the 3D face model fit the landmarks best. These methods have achieved impressive results especially on the faces with relatively small pose angles. However, their performance highly depends on the accuracy of landmark localization, which is itself a challenging task. Besides, the assumption underlying these methods is also arguable, because the difference between the facial landmark configurations might be due either to different head poses or to different faces (i.e., identities).

The methods [6–9] not using landmarks treat the pose estimation problem as a classification or regression problem. They first train a pose angle predictor using a set of face images whose pose angles are known and then use the predictor to estimate the pose angles in unseen face images. These methods do not need to detect facial landmarks but directly extract appearance features from the face images. They are thus easier to apply. However, there are still some open issues related to these methods. First, the available training face images usually have discrete pose angles (say from −90 to 90 degrees with an increment of 5 degrees), though the head pose angle itself is continuous. In other words, the available training data is merely a sampling of the full head pose space. But it is still unknown how the pose angle sampling interval of the training face images affects the effectiveness of the obtained pose angle predictor. Second, several popular feature extraction methods have been used for head pose estimation, such as Gabor filter responses, Local Binary Patterns or LBP, Histograms of Oriented Gradients

of HOG, and even Grey-scale intensity features. In this paper, we will investigate the performance of different feature extraction methods using our proposed method. Moreover, existing studies usually use the face images captured with same devices and under similar conditions for both training and testing. In real-world applications, however, it might have to apply a trained pose angle predictor to a scenario different from the training phase.

In this paper, we focus on the appearance and training based method, which does not need to detect facial landmarks before pose estimation and is thus easier to use. Specifically, we attempt to estimate the head pose from a single 2D face image by using improved random forests with appearance-based features. To this end, we first extract the features from 2D face images and then use supervised locality preserving projections (SLPP) to reduce the feature dimension. After applying SLPP, each dimension of the features is associated with an importance index (i.e., a weight). Finally, we construct random forests based on the weighted features [10, 11]. When constructing random forests for head pose estimation, unlike previous methods, we consider the importance of different features and conduct tree screening to discard deficient trees [12–14]. Thanks to the weighting and screening processes, our constructed random forests are more effective, as will be demonstrated by our experimental results. The main contribution of this paper includes

(i) employing feature weighting and tree screening to improve the effectiveness of the constructed random forest based head pose predictor,

(ii) investigating the impact of pose angle sampling interval of training face images on the accuracy of trained pose angle predictors,

(iii) studying the performance of pose angle predictors when the training and testing face images are captured under different scenarios (i.e., heterogeneous face images).

The remainder of this paper is organized as follows. Section 2 briefly introduces related work. Section 3 presents the proposed head pose estimation method. Section 4 then reports the experimental results. Finally, conclusions are drawn in Section 5.

2. Related Work

A random forest is an ensemble of tree predictors that can cope with classification and regression problems. It has been successfully applied to many problems [15–21] including head pose estimation. Head pose estimation based on random forests was first proposed by Huang et al. [22] in 2010. The method enhanced the strength of the decision tree by LDA with minimum Euclidean distance classifier as the node tests and had achieved good results. Li et al. [23] used random regression forests to estimate head poses in two steps: (i) half-face and tree structured classifiers with cascaded-Adaboost algorithm are adopted to detect faces with various head poses; (ii) based on the cropped face images, random regression forests are learned and applied to estimate head orientations. Recently, Fanelli et al. [24, 25] exploited random forests for estimating 3D head poses in 3D face images. They utilized depth information as feature vectors when constructing the tree predictors in random forests. However, because depth information is not available in 2D images, this method cannot be applied to 2D face images.

In this paper, we improved the strength of decision trees in the random forest via selecting features according to their importance and screening the trees to retain only those with higher accuracy. Our method differs from previous random forest based head pose estimation methods in the following three ways.

(i) Our method assumes that different features are not equally important and selects important features with higher probabilities.

(ii) Our method evaluates the trees and retains them only if they have sufficiently high accuracy. This screening process improves the quality of the obtained forests.

(iii) A machine learning technique, that is, SLPP, is applied to enhance the discriminant power of the features extracted from 2D face images. Moreover, SLPP provides a natural measurement of the importance of different features.

3. The Proposed Head Pose Estimation Method

Figure 1 shows the block diagram of the proposed method. Details of each component are given below.

3.1. Feature Extraction. Faces are first detected in the input images by using the method in [26]. They are then cropped from the original images and normalized to grey-scale images of 64×64 pixels. Then certain texture features (e.g., Gabor [8], LBP [21], or HOG [22]) are extracted from the normalized face images. The features are further processed by using the SLPP method for two purposes: (i) reducing and weighting feature dimensions and (ii) enhancing the discriminant power of features.

SLPP is a method for supervised linear dimensionality reduction [27]. It seeks a set of projection axes $\{\alpha_i\}$, which can preserve the local structure of samples that are from the same classes while maximizing the separation margin between the samples of different classes. Let $x_1, x_2, \ldots, x_N \in R^n$ be N n-dimensional features extracted from a given set of N training samples belonging to K classes $c_1, c_2, \ldots, c_K$. The objective function of SLPP is defined as follows:

$$\alpha_{\text{opt}} = \arg\max \frac{\alpha^T X L_b X^T \alpha}{\alpha^T X L_w X^T \alpha}, \tag{1}$$

$$L_b = D^B - B, \tag{2}$$

$$L_w = D^W - W, \tag{3}$$

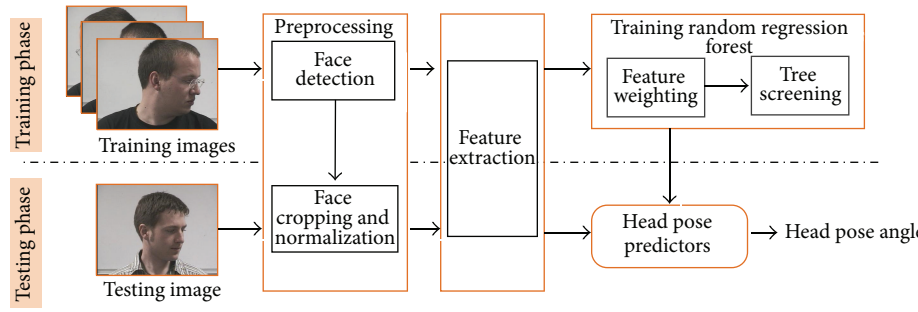

Figure 1: Block diagram of the proposed head pose estimation method.

$$D^B = \begin{cases} \sum_j B_{ji}, & i = j \\ 0, & i \neq j, \end{cases} \tag{4}$$

$$D^W = \begin{cases} \sum_j W_{ji}, & i = j \\ 0, & i \neq j, \end{cases} \tag{5}$$

where $X = \{x_1, x_2, \ldots, x_N\}$, L_b and L_w are called Laplacian matrix, B and W are the within-class scatter matrix and the between-class scatter matrix, and T denotes the transpose operator. The optimal projection axes of SLPP are obtained by maximizing the between-class separation and simultaneously minimizing the within-class scatter. It can be mathematically shown that the optimal projection axes of SLPP are the eigenvectors corresponding to the nonzero eigenvalues of the following general Eigen-decomposition problem:

$$XL_bX^T\alpha = \lambda XL_wX^T\alpha. \tag{6}$$

In the SLPP generated feature space, each dimension corresponds to one of the projection axes (i.e., the eigenvectors of (6)), and its associated eigenvalue as well. The larger the eigenvalue is, the better the feature along the corresponding dimension fulfills the objective function in (1). This motivates us to measure the importance of different features based on the eigenvalues. Specifically, in the lower dimensional feature space, the ith feature is assigned with a weight, which is defined as follows:

$$w_i = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j}, \tag{7}$$

where λ_i is the eigenvalue associated with the ith eigenvector and m is the dimension of the reduced SLPP feature space. The weights of different features will be used to guide the selection of features in constructing tree predictors in random regression forests.

3.2. Training Random Regression Forests. Based on the N training samples, a random regression forest is constructed. Let $Y = \{y_i \in R^m \mid i = 1, 2, \ldots, N\}$ denote the extracted features, and let $L = \{l_1, l_2, \ldots, l_N\}$ denote the annotated class labels of the training samples (i.e., yaw or pitch angles in this paper). For each of the S trees in the random forest, a subset of N_s samples is randomly chosen by bootstrap from the N training samples. To train a tree using the chosen training samples, the parameters involved in the binary tests at their nonleaf nodes have to be determined.

The binary test at a node, denoted by Φ, determines which branch the sample arriving at the node should be forwarded to. In this paper, the binary test is defined by a chosen feature and its associated threshold. Given a sample, the binary test compares its feature with the threshold and forwards the sample to the left branch if the evaluation result is lower than the threshold or forwards it to the right branch otherwise. In the training phase, a pool of binary tests $\{\Phi_i\}$ is generated at every nonleaf node by randomly choosing a feature from a subset of M features and a threshold for it. The subset of M features is randomly chosen at each nonleaf node out of the m features. In order to choose the best binary test from these randomly generated binary tests, we use the information gain to evaluate the effectiveness of each binary test. Specifically, all the training samples (denoted by $\widehat{Y}$) arriving at the node are first split into the left and right branches by the binary test under evaluation. The two subsets are denoted by $\widehat{Y}_L$ and $\widehat{Y}_R$, respectively. The information gain of this splitting is then computed as follows:

$$\begin{aligned} \mathrm{IG}(\Phi_i) &= E(\widehat{Y}) - E(X_{\widehat{Y}_L}) - E(X_{\widehat{Y}_R}), \\ E(\widehat{Y}) &= -\sum_{i=1}^{m} P(y_i) \log P(y_i), \end{aligned} \tag{8}$$

where $E(\cdot)$ is information entropy and $P(y_i)$ is the probability of ith feature occurring at the node. The binary test with the highest information gain is taken as the best one for the node.

During the construction of a tree, the tree keeps growing until the number of samples arriving at the node is sufficiently small, or the samples are all of the same class, or the tree reaches the maximum number of layers. A leaf is created when the tree stops growing. The leaf stores a real-valued, multivariate distribution computed from the pose parameters ($\Theta_i = \{\Theta_{\mathrm{yaw}}, \Theta_{\mathrm{pitch}}\}$) of the samples arriving at it. Figure 2 illustrates a random regression forest for head pose estimation.

As proven by Breiman [28], the generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Motivated by this fact, we choose features according to

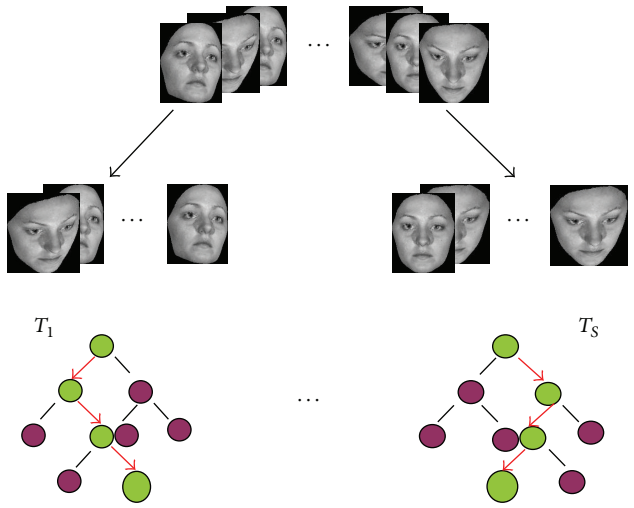

Figure 2: Illustration of a random regression forest for head poses estimation. For each tree, the binary tests at the nonleaf nodes direct an input sample towards a leaf, where a real-valued, multivariate distribution of the output parameters (i.e., head pose angles) is stored. The forest combines the results of all its trees to produce a probabilistic prediction in the real-valued output space.

their weights. That is, the feature with larger weight has higher probability to be chosen. At each nonleaf node, we randomly select M features from m features without repeat according to their weights. This feature weighting method improves the effectiveness of the constructed trees, as will be demonstrated by our experiments.

3.3. Tree Evaluation and Screening. Furthermore, once a tree is built, we evaluate its quality to determine whether the tree should be retained or discarded. Only the trees which have sufficiently high accuracy will be kept. As mentioned above, the bootstrap method is used to randomly choose a subset of N_s samples from the given N training samples to construct a tree. These chosen samples are called in-of-bag (IOB) data, while the remainder is called out-of-bag (OOB) data. The OOB data are utilized to evaluate the tree constructed based on the IOB data.

Given a tree regressor $h_k(x)$ built from the kth training data subset, we define the mean absolute error (denoted by MAE) of the tree $h_k(x)$ as

$$\text{MAE} = \frac{\sum_{i=1}^{N-N_s} \left(h_k\left(x_i\right) - l_i\right)}{N - N_s}, \tag{9}$$

where x_i is a sample in the OOB data and l_i is the class label of the sample x_i. It is obvious that the tree with low MAE has high accuracy. Hence, we accept a tree whose MAE on the OOB data is below the prespecified threshold. In this way, the trees constructed in the forest are all with lower MAE, and a better random forest can be obtained.

3.4. Predicting Head Poses. Once the random regression forest is constructed, it can be used to predict the head pose in a new unseen 2D face image. At each node of a tree in the forest, the stored binary test evaluates a feature and sends the testing sample either to the right or to the left branch, all the way down until a leaf. When the sample arrives at a leaf, the tree gives an estimate for the pose parameters in terms of the stored distribution. The multivariate distributions predicted by different trees in the forest are averaged to give the final estimation of the head pose of the testing sample.

4. Experiments

In this section, we evaluate the proposed method on four public databases: Pointing'04 [29], CAL-PEAL [30], NCKU [31], and CUbiC FacePix [32, 33]. For all the images in these databases, the face detector in [26] is applied to locate the face regions (when the method fails to detect the right face region, we crop the face region manually), and the face regions are cropped and normalized to the same size of 64 × 64 pixels.

We compare our method (denoted by WSRF) with some existing random forests based methods, including the method (denoted by VRF + LDA) in [22], the method (denoted by SRF) in [23] and the method (denoted by WRF) in [7], and some other state-of-the-art methods, including the methods proposed in [34] (denoted by LPLS and KPLS), [35] (denoted by LBIF and SLBIF), [36] (denoted by MGD), [2] (denoted by 3DMM), and [9] (denoted by MLD). In the experiment, we investigate how different features (e.g., Gabor, LBP, HOG, and Grey-scale intensity) work with our proposed method. We use Gabor kernels with five spatial frequencies and eight orientations. To extract LBP features, face images are divided into cells of 16 × 16 pixels. For each pixel in a cell, we consider a circular neighborhood with radius of eight pixels. When extracting the HOG features, each scanning window over the face images is divided into cells with a fixed size of 6 × 6 pixels, and each group of 3 × 3 cells is integrated into a block in a sliding fashion. The extracted Gabor, LBP, and HOG features are reduced to 256 dimensions by SLPP. For each method, we report the mean absolute errors (MAE) and the accuracy (the percentage of the samples whose pose estimation error is less than 5 degrees) of head pose estimation. Furthermore, in order to evaluate the generalization ability of the proposed method, we evaluate it on a set of samples captured by surveillance cameras.

In our experiment, when constructing random regression forests, the number of trees is empirically determined as being done in many random forest based methods [37, 38]. Specifically, the number of trees is gradually increased until the prediction error becomes stable or increasing.

4.1. Results on Pointing'04. The Pointing'04 database [29] contains 2D face images of 15 individuals. Each subject has 2 series of 93 images at different poses. The images display variations in expressions, skin colors, and occlusions (e.g., wearing glasses). The pose is described by rotation of yaw and pitch. Yaw angles range between −90 and 90 with increments of 15 degrees, and pitch angles range between −90 and 90 with increments of 30 degrees. Figure 3 shows some example face images of a subject in the Pointing'04 database. It is worth mentioning that the subjects in this database were

Table 1: The accuracy and MAE for different methods on the Pointing'04 database.

Method	MAE		Accuracy	
	Yaw	Pitch	Yaw	Pitch
WRF [7]	7.5°	7.8°	80.15%	78.23%
VRF + LDA [17]	11.05°	N/A	66.95%	N/A
SRF [18]	9.6°	13.9°	N/A	N/A
LPLS [29]	11.29°	10.52°	45.57%	58.70%
KPLS [29]	6.56°	6.61°	67.36%	80.36%
MGD [31]	6.9°	8.0°	64.51%	62.72%
MLD [9]	**4.24°**	**2.69°**	N/A	N/A
WSRF (Gabor)	6.32°	5.92°	84.14%	85.41%
WSRF (LBP)	6.23°	5.65°	83.01%	85.78%
WSRF (HOG)	**5.21°**	**4.06°**	**87.32%**	**89.41%**
WSRF (intensity)	6.54°	6.41°	83.68%	84.86%

N/A denotes that the measure is not available in the paper where the method was proposed (same for the other tables in this paper). For the method SRF, 14 subjects are used to train the random regression forests while the rest subjects are applied for test.

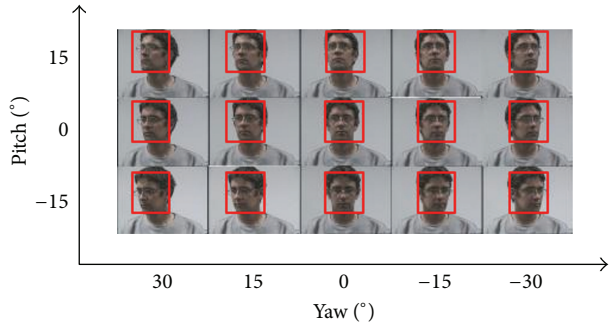

Figure 3: Some example face images of a subject in the Pointing'04 database. The red bounding boxes in the figure denote the face regions provided in the database.

asked to rotate their heads and the only-one camera in front of them captured their face images. Consequently, the head pose angles in this database are not very precise (see Figure 4).

In the experiments, we set $S = 30$ and $M = 14$. For a fair comparison with the methods in [9, 17, 29, 31], we use 80% of Pointing'04 images for training, 10% for cross-evaluation, and 10% for testing.

Table 1 shows the accuracy and MAE of yaw and pitch angles by different methods on the Pointing'04 database and by our proposed improved random regression forest based method with different features. The proposed method with HOG feature achieves MAE of 5.21 and 4.06 degrees for yaw and pitch, respectively, which are comparable to the MLD method and better than the other methods. Considering that the MLD method is specially designed for handling the uncertainty of head pose angles in this database, our proposed method could be further improved by incorporating the idea in MLD into the training process (say using face images with yaw angles of 40 to 50 degrees as the training samples for the yaw angle of 45 degrees. This is among our future work).

Table 2: The accuracy and MAE for different methods on the CAS-PEAL database.

Method	MAE	Accuracy
WRF [7]	6.9°	92.72%
VRF + LDA [17]	N/A	97.23%
LBIF [30]	N/A	94.57%
SLBIF [30]	N/A	94.55%
WSRF (Gabor)	5.35°	97.29%
WSRF (LBP)	5.54°	97.18%
WSRF (HOG)	**4.53°**	**98.54%**
WSRF (intensity)	5.83°	97.02%

4.2. Results on CAS-PEAL. The CAS-PEAL face database [24] contains 99,594 images of 1,040 individuals (595 males and 445 females). Each subject displays twenty-one poses combining seven yaw angles (−45, −30, −15, 0, 15, 30 and 45) and three pitch angles (−30, 0 and 30). In this experiment, we consider only yaw angles as in [6].

For a fair comparison with [6, 22], in our experiments, we use a subset containing totally 4,200 images of 200 subjects whose IDs range from 401 through 600. The images are ranked by their subject IDs and then divided into three subsets. Two subsets are taken as the training set and the other subset is taken as the testing set. Figure 5 shows some example cropped face regions in this database. In the experiments, the number of trees and the number of randomly selected features at each node are set as 40 and 14, respectively.

Table 2 summarizes the accuracy and MAE achieved by different methods as well as by our proposed method with different features on the CAS-PEAL database. Obviously, the proposed method is much better than the method in [30]. In particular, the proposed method with HOG feature achieves MAE of 4.53 degrees for yaw, which is the best results among the methods under consideration. In addition, the method in [2] achieved the MAE of 3.78 degrees, but testing on yaw angles that range from −15° to 15°. By comparing the results of WRF and the proposed method with different features in the list, it can be seen that screening trees is effective in improving the pose estimation accuracy of random forests.

4.3. Results on NCKU. The NCKU face database [31] contains 6,660 images of 90 subjects (78 males and 12 females). Each subject has 74 images, where 37 images were taken every 5 degrees from right profile (defined as +90) to left profile (defined as −90) in the yaw rotation. The remaining 37 images are generated (synthesized) by the existing 37 images using commercial image processing software in the way of flipping them horizontally. In our experiments, we use a subset containing the first 37 images of all subjects. Figure 6 shows some example face images of a subject in the NCKU face database.

In the experiment, the images are ranked by their subject IDs and divided into three subsets. Two subsets are taken as the training set and the other subset is used for testing. The number of trees and the number of randomly selected features at each node are set as 30 and 14, respectively.

(a) (b)

Figure 4: Different head poses in the Pointing'04 database were acted by the subjects who rotated their heads in front of the camera. (a) Two images of the same subject who is displaying the same pose (pitch: −60; yaw: −90) in both images. (b) Another two images, each from one of two subjects who are displaying the same pose (pitch: −30; yaw: −60). Obviously, the specified pose angles are not precise.

Figure 5: Example cropped face images in the CAS-PEAL database.

Table 3 summarizes the accuracy and MAE achieved by different methods on the NCKU face database. It can be seen that our proposed method with HOG feature achieves the best results. This again proves the effectiveness of our proposed method in improving the pose estimation accuracy of random forests.

4.4. Results on CUbiC FacePix. The FacePix database [32] contains 16,290 images of 30 individuals (25 males and 5 females). Each subject has two sets of face images: a set with pose angle variations and a set with illumination angle variations. In the experiment, we only use the images of the pose variations. Each subject displays 181 poses from right profile (denoted by +90) to left profile (denoted by −90) with increments of 1 degree in the yaw rotation. Figure 7 shows some images of a subject displaying yaw angles from −90 to 90 with an interval of 10.

In the experiments, we set $S = 45$ and $M = 14$, respectively. The images are ranked by their subject IDs and divided into three subsets. Two subsets are used for training and the rest one is used for testing.

Table 4 summarizes the accuracy and MAE achieved by different methods on the CUbiC FacePix face database. The proposed method with HOG features obtains an MAE of 4.51 degrees and accuracy of 89.15% for yaw, which are significantly better than the existing method.

4.5. Results on Surveillance Data. In order to evaluate the generalization ability of our method in more realistic environment, the forests trained by samples from Pointing'04, CAS-PEAL, NCKU, and FacePix are directly applied to a set of samples captured by surveillance cameras. There are totally 30 participating volunteers. To capture images with pose variation of a subject, the subject is asked to look upwards, look right into the camera, and look downwards, when he/she walks through under the surveillance cameras.

Table 3: The accuracy and MAE for different methods on the NCKU database.

Method	MAE	Accuracy
WRF [7]	7.60°	80.42%
WSRF (Gabor)	5.97°	86.37%
WSRF (LBP)	6.04°	85.89%
WSRF (HOG)	**4.83°**	**88.32%**
WSRF (intensity)	6.28°	85.56%

Table 4: The accuracy and MAE for different methods on the CubiC FacePix database.

Method	MAE	Accuracy
WRF [7]	7.95°	76.33%
3DMM [2]	4.63°	N/A
WSRF (Gabor)	5.76°	87.56%
WSRF (LBP)	5.88°	87.42%
WSRF (HOG)	**4.51°**	**89.15%**
WSRF (intensity)	6.05°	87.07%

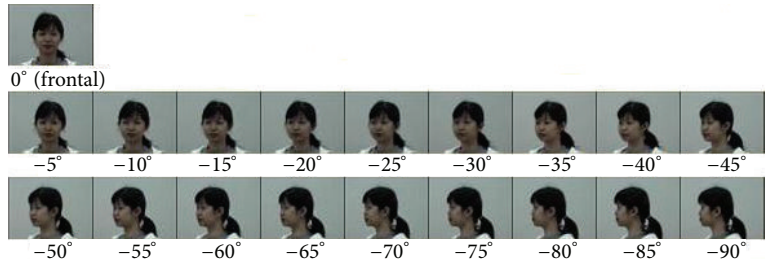

Figure 6: Some example face images of a subject in the NCKU face database.

Figure 7: Some example face images of a subject in the FacePix Database.

Table 5: The accuracy and MAE on the surveillance dataset achieved by the random regression forests trained on different databases.

Database	MAE		Accuracy	
	Yaw	Pitch	Yaw	Pitch
Pointing'04	8.23°	7.08°	80.17%	82.13%
CAS-PEAL	7.28°	6.40°	81.35%	85.48%
NCKU	**6.39°**	N/A	**84.39%**	N/A
CUbiC FacePix	6.92°	N/A	82.36%	N/A
Pointing'04 + CAS-PEAL + NCKU + CUbiC FacePix	**5.86°**	**5.68°**	**86.26%**	**87.93%**

As a result, 15 images are captured for each subject, showing yaw angles of (−30, −15, 0, 15, and 30) and pitch angles of (−30, 0, and 30). Figure 8 shows the 15 images of a subject in this surveillance dataset.

The face regions in the captured images are located by using the face detector in [26] and then cropped and normalized to the same size of 64 × 64 pixels. HOG features are extracted from the normalized cropped face regions and projected to the SLPP-reduced feature space. Table 5 shows the MAE and accuracy of yaw and pitch angles achieved on this surveillance dataset by, respectively, using the random forests trained on the Pointing'04, CASPEAL, NCKU, CUbiC and FacePix databases. In order to examine whether using face images from multiple sources to train the random forest helps, we combine the four databases (i.e., Pointing'04, CAS-PEAL, NCKU, and CUbiC FacePix) and use all the face images together to train another random forest based head pose predictor. The results are shown in Table 5.

It can be observed from Table 5 that when using a single training database, the smallest pose estimation error is achieved by the NCKU-trained predictor, for which the training data have an interval of 5 degrees in pose angles. In contrast, the other three predictors trained using data with 1-degree interval (i.e., FacePix) or 15-degree interval (i.e., Pointing'04 and CAS-PEAL) all have larger errors. A possible reason is that too coarse sampling of pose angles cannot well interpolate all possible pose variations, whereas too dense sampling of pose angles might in fact introduce more confusing information, rather than more discriminative information. On the other hand, by combining multiple databases for training, the generalization ability of the resulted head pose predictor can be enhanced, thanks to the improved diversity of the training data.

Figure 9 shows some face images for which the trained predictors fail to accurately estimate the pose angles. The MAE for these images is larger than 10 degrees. As can be seen, the cluttered background and intense facial expressions could generate distracting features for the predictors trained using the face images collected under laboratory conditions. How to cope with such poor quality images is among the topics of our future work.

Furthermore, by comparing the results in Tables 1, 2, 3, and 5, we can see that the predictors can obtain higher accuracy when the testing images are captured by the same devices as the training images. This is essentially a problem of heterogeneous image analysis. To better understand the impact of heterogeneous images, we evaluate our method by using the surveillance dataset for both training and testing. Specifically, the images of 20 subjects are used for training, while the images of the other 10 subjects are used for testing. The number of trees and the number of randomly selected features at each node are set as 10 and 14. Table 6 gives the accuracy and MAE of yaw and pitch angles achieved by the predictor trained on the surveillance dataset. The best MAE of the estimated yaw and pitch angles are, respectively, 5.43 and 5.21 degrees, which are obviously lower than that obtained by the predictors trained using other databases and even lower than that achieved by the predictor trained by

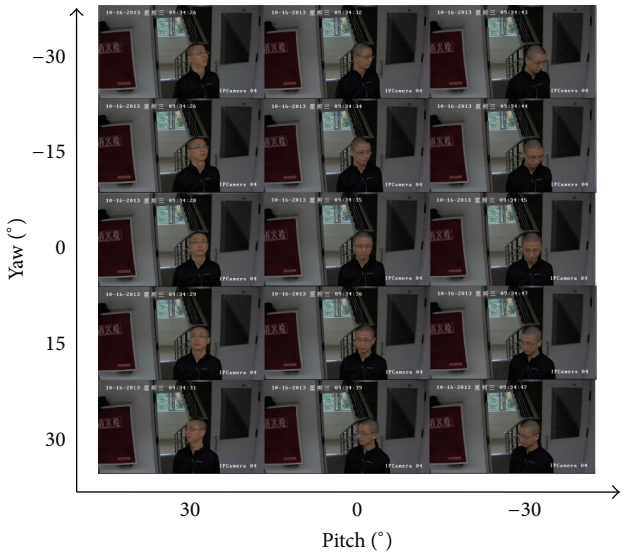

Figure 8: Example face images of a subject in the surveillance dataset captured by surveillance cameras.

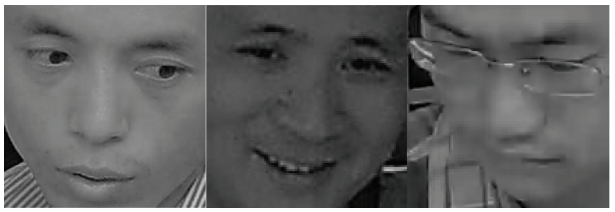

Figure 9: Some images from the surveillance dataset whose MAE are more than 10 degrees.

Table 6: The accuracy and MAE achieved by the proposed method when part of the surveillance dataset is used for training and the remaining in the dataset is used for testing.

Database	MAE		Accuracy	
	Yaw	Pitch	Yaw	Pitch
WSRF (Gabor)	6.13°	5.89°	86.35%	88.01%
WSRF (LBP)	6.57°	6.26°	85.81%	87.65%
WSRF (HOG)	**5.43°**	**5.21°**	**89.36%**	**89.79%**
WSRF (intensity)	6.90°	6.76°	84.29%	85.94%

combined databases (refer to Table 5). This again demonstrates the difficulty in handling heterogeneous data.

4.6. The Threshold in Tree Screening. In tree screening, we evaluate the performance of a tree in terms of its MAE on the OOB data and impose a threshold on its MAE. In the above experiments, the threshold of OOB MAE is set to 8. That is, we accept the trees whose OOB MAE is below 8. Figures 10 and 11, respectively, show the MAE and accuracy of yaw and pitch angles on testing data when different MAE thresholds are used for tree screening in training. Figure 12 plots the required training time versus MAE thresholds.

Table 7: The mean ($\bar{S}$) and variance (σ_s) of the similarity of all pairs of trees in the random regression forests constructed on different databases.

Databases	Similarity mean ($\bar{S}$)	Similarity variance (σ_s)
Pointing'04	0.169	0.067
CAS-PEAL	0.190	0.013
NUKU	0.173	0.015
CUbiC FacePix	0.206	0.025
Surveillance	0.120	0.012

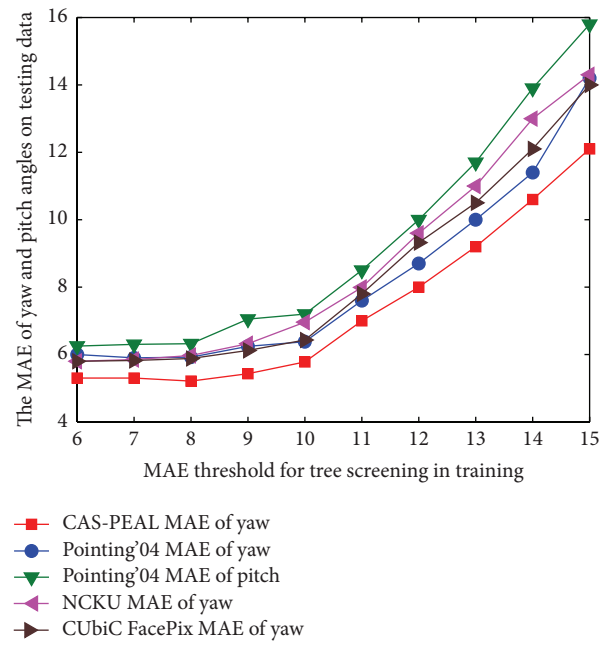

Figure 10: The MAE of yaw and pitch angles on testing data when different MAE thresholds are used for tree screening in training.

From these curves, we can see that when the MAE threshold increases, it takes a shorter time to train a random forest, but its accuracy becomes lower. On the other hand, as the MAE threshold decreases to 8 and even smaller, the resulting random forests' accuracy becomes relatively stable. This suggests that using an MAE threshold smaller than 8 cannot significantly improve the accuracy while the training time increases substantially. Therefore, we report above the performance of our proposed method when the MAE threshold is set as 8.

4.7. Diversity Analysis. This experiment evaluates the diversity of the trees in the obtained random forests. A random forest with higher diversity is more preferred because such a forest is believed to be more generative. To assess the diversity of a forest, we define that two trees share a common node if there are two nodes which are at the same position in them and use the same feature. The similarity $S \in [0, 1]$ of two trees can be then computed as

$$S = \frac{2 \times N_{cd}}{N_d^1 + N_d^2}, \tag{10}$$

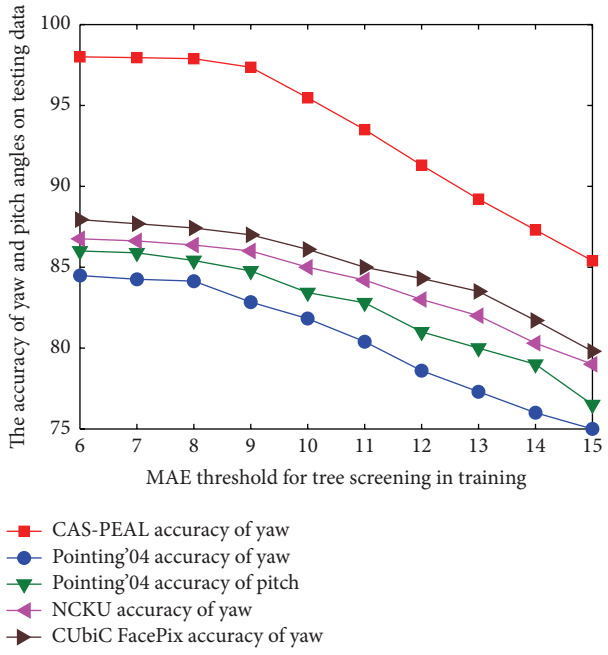

Figure 11: The accuracy of yaw and pitch angles on testing data when different MAE thresholds are used for tree screening in training.

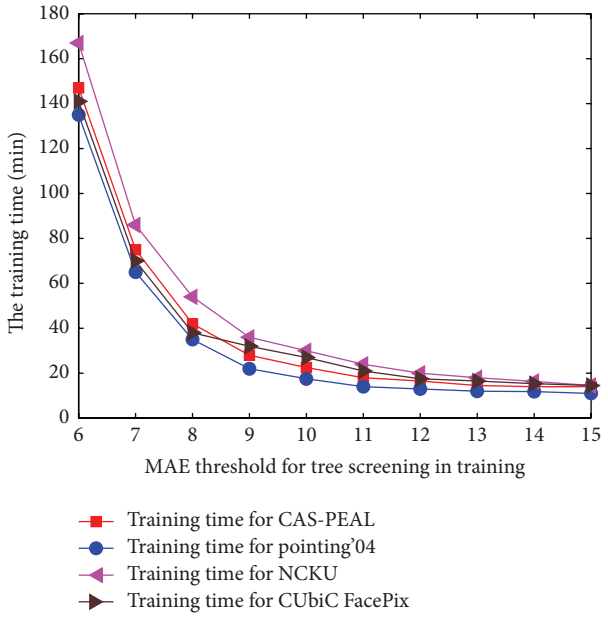

Figure 12: The training time when different MAE thresholds are used for tree screening in training.

where N_d^1 and N_d^2 are, respectively, the total number of nodes in the two trees and N_{cd} denotes the number of common nodes between them. The diversity of the forest is measured by the mean and variance of the similarity of all pairs of trees in it. A forest with low similarity mean and variance is deemed to have high diversity.

Table 7 shows the similarity mean and variance of the random forests obtained on the five databases. As can be clearly seen, all are small, which indicates high diversity of these forests. It is worth mentioning that if we further consider the thresholds associated with the nodes, the diversity could be even higher. Therefore, although the trees in a forest are established based on the same set of weighted features, they still have high diversity thanks to the random division of both training data and candidate features (i.e., for each node, a random subset of training data and a random subset of candidate features are used to learn its parameters).

5. Conclusions

An improved random forest based method has been presented in this paper for estimating the head pose in 2D face images. The method weights the features when constructing tree predictors in the random forest and screens the trees to retain only the trees with high accuracy. A series of evaluation experiments have been done on four public databases and a set of surveillance images captured by ourselves. The experimental results demonstrated that the random forests constructed with our proposed method indeed improve the head pose estimation accuracy.

We have systematically evaluated the accuracy of random forest based head pose predictors when training face images with different pose angle sampling intervals are used. Our results suggest that neither too dense nor too coarse sampling intervals are preferred, but a sampling interval of median (say five) degrees generates the best pose angle prediction accuracy.

We have also investigated the performance of trained pose angle predictors on a heterogeneous testing dataset. The results demonstrate that the accuracy decreases when the testing data is captured in a different scenario from that of training data. This is obvious when the predictor is trained with laboratory data and tested with real-world surveillance data. Such heterogeneous problem is not aware by more and more researchers.

In our future work, we are going to focus more on the heterogeneous problem in head pose estimation and further improve the pose estimation accuracy for real-world face images by exploring more advanced face image processing techniques and more effective feature representations.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Key Scientific Instrument and Equipment Development Project of China (no. 2013YQ49087903) and National Natural Science Foundation of China (Grant no. 61202160).

References

[1] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: a survey," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 31, no. 4, pp. 607–626, 2009.

[2] Y. Cai, M. Yang, and Z. Li, "Robust head pose estimation using a 3D morphable model," *Mathematical Problems in Engineering*, vol. 2015, Article ID 678973, 10 pages, 2015.

[3] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[4] J. Aghajanian and S. J. D. Prince, "Face pose estimation in uncontrolled environments," in *Proceedings of the 20th British Machine Vision Conference (BMVC '09)*, vol. 1, pp. 1–11, September 2009.

[5] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2879–2886, IEEE, Providence, RI, USA, June 2012.

[6] B. Ma, R. Huang, and L. Qin, "VoD: a novel image representation for head yaw estimation," *Neurocomputing*, vol. 148, pp. 455–466, 2015.

[7] R. Zhu, G. Sang, Y. Cai et al., "Head pose estimation with improved random regression forests," in *Proceedings of the 8th Chinese Conference on Biometric Recognition (CCBR '13)*, pp. 457–465, Jinan, China, November 2013.

[8] M. Fenzi, L. Leal-Taixe, B. Rosenhahn, and J. Ostermann, "Class generative models based on feature regression for pose estimation of object categories," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 755–762, Portland, Ore, USA, June 2013.

[9] X. Geng and Y. Xia, "Head pose estimation based on multivariate label distribution," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 1837–1842, IEEE, Columbus, Ohio, USA, June 2014.

[10] D. Amaratunga, J. Cabrera, and Y.-S. Lee, "Enriched random forests," *Bioinformatics*, vol. 24, no. 18, pp. 2010–2014, 2008.

[11] B. Xu, J. Z. Huang, G. Williams, Q. Wang, and Y. Ye, "Classifying very high-dimensional data with random forests built from small subspaces," *International Journal of Data Warehousing and Mining*, vol. 8, no. 2, pp. 44–63, 2012.

[12] H. Kim, H. Kim, H. Moon, and H. Ahn, "A weight-adjusted voting algorithm for ensembles of classifiers," *Journal of the Korean Statistical Society*, vol. 40, no. 4, pp. 437–449, 2011.

[13] M. Robnik-Šikonja, "Improving random forests," in *Machine Learning: ECML 2004: 15th European Conference on Machine Learning, Pisa, Italy, September 20–24, 2004. Proceedings*, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds., vol. 3201 of *Lecture Notes in Computer Science*, pp. 359–370, Springer, Berlin, Germany, 2014.

[14] H. B. Li, W. Wang, H. W. Ding, and J. Dong, "Trees Weighting Random Forest method for classifying high-dimensional noisy data," in *Proceedings of the IEEE International Conference on E-Business Engineering (ICEBE '10)*, pp. 160–163, Shanghai, China, November 2010.

[15] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[16] A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[17] H. Pang, A. Lin, M. Holford et al., "Pathway analysis using random forests classification and regression," *Bioinformatics*, vol. 22, no. 16, pp. 2028–2036, 2006.

[18] A. Bosch, A. Zisserman, and X. Muñoz, "Image classification using random forests and ferns," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, October 2007.

[19] S. Bernard, S. Adam, and L. Heutte, "Using random forests for handwritten digit recognition," in *Proceedings of the 9th IEEE International Conference on Document Analysis and Recognition (ICDAR '07)*, pp. 1043–1047, Curitiba, Brazil, September 2007.

[20] R. Khan, A. Hanbury, and J. Stoettinger, "Skin detection: a random forest approach," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 4613–4616, IEEE, Hong Kong, China, September 2010.

[21] L. Zhang and P. N. Suganthan, "Random forests with ensemble of feature spaces," *Pattern Recognition*, vol. 47, no. 10, pp. 3429–3437, 2014.

[22] C. Huang, X. Ding, and C. Fang, "Head pose estimation based on random forests for multiclass classification," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 934–937, Istanbul, Turkey, August 2010.

[23] Y. Li, S. Wang, and X. Ding, "Person-independent head pose estimation based on random forest regression," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 1521–1524, Hong Kong, September 2010.

[24] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, "Random forests for real time 3D face analysis," *International Journal of Computer Vision*, vol. 101, no. 3, pp. 437–458, 2013.

[25] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 617–624, IEEE, Providence, RI, USA, June 2011.

[26] S. Liao, A. K. Jain, and S. Z. Li, "Unconstrained face detection," Tech. Rep. MSU-CSE-12-15, Department of Computer Science, Michigan State University, East Lansing, Mich, USA, 2012.

[27] Z.-H. Shen, Y.-H. Pan, and S.-T. Wang, "A supervised locality preserving projection algorithm for dimensionality reduction," *Pattern Recognition and Artificial Intelligence*, vol. 21, no. 2, pp. 233–239, 2008.

[28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[29] Pointing'04 database, http://www-prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html.

[30] W. Gao, B. Cao, S. Shan et al., "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 38, no. 1, pp. 149–161, 2008.

[31] The National Cheng Kung University face database, http://www.datatang.com/data/14866.

[32] J. A. Black Jr., M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan, "A framework for performance evaluation of face recognition algorithms," in *Proceedings of the International Conference on Information Technologies and Communications (ICITC '02)*, pp. 163–174, 2002.

[33] G. Little, S. Krishna, J. Black, and S. Panchanathan, "A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. II89–II92, March 2005.

[34] M. A. Haj, J. Gonzalez, and L. S. Davis, "On partial least squares in head pose estimation: how to simultaneously deal with misalignment," in *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR '12), pp. 2602–2609, IEEE, Providence, RI, USA, June 2012.

[35] B. Ma, X. Chai, and T. Wang, "A novel feature descriptor based on biologically inspired feature for head pose estimation," *Neurocomputing*, vol. 115, pp. 1–10, 2013.

[36] V. Jain and J. L. Crowley, "Head pose estimation using multi-scale gaussian derivatives," in *Image Analysis: 18th Scandinavian Conference, SCIA 2013, Espoo, Finland, June 17–20, 2013. Proceedings*, vol. 7944 of *Lecture Notes in Computer Science*, pp. 319–328, Springer, Berlin, Germany, 2013.

[37] K. L. Lunetta, L. B. Hayward, J. Segal, and P. van Eerdewegh, "Screening large-scale association study data: exploiting interactions using random forests," *BMC Genetics*, vol. 5, no. 1, article 32, 2004.

[38] C. Strobl and A. Zeileis, "Danger: high power!—exploring the statistical properties of a test for random forest variable importance," Tech. Rep. 017, University of Munich, Munich, Germany, 2008.