Dirichlet-tree Distribution Enhanced Random Forests for Head Pose Estimation

Yuanyuan Liu[1,2,3], Jingying Chen[1,2], Leyuan Liu[1,2], Yujiao Gong[1] and Nan Luo[1]

[1]*National Engineering Research Center for e-Learning, Central China Normal University, Wuhan, China*

[2]*Collaborative & Innovative Center for Educational Technology (CICET), Wuhan, China*

[3]*Huazhong University of Science and Technology Wenhua College, Wuhan, China*

Keywords: Dirichlet-tree Distribution Enhanced Random Forests, Head Pose Estimation, Gaussion Mixture Model, Positive Patch Extraction.

Abstract: Head pose estimation is important in human-machine interfaces. However, illumination variation, occlusion and low image resolution make the estimation task difficult. Hence, a Dirichlet-tree distribution enhanced Random Forests approach (D-RF) is proposed in this paper to estimate head pose efficiently and robustly under various conditions. First, Gabor features of the facial positive patches are extracted to eliminate the influence of occlusion and noise. Then, the D-RF is proposed to estimate the head pose in a coarse-to-fine way. In order to improve the discrimination capability of the approach, an adaptive Gaussian mixture model is introduced in the tree distribution. The proposed method has been evaluated with different data sets spanning from -90° to 90° in vertical and horizontal directions under various conditions.The experimental results demonstrate the approach's robustness and efficiency.

1 INTRODUCTION

Head pose estimation is important in many human machine interfaces such as (Chen and Chen, 2011; McFarlane, 2002). Head orientation is related to a persons direction of attention, it can present useful information about what the person is paying attention to. Different methods have been developed for two types of image data, i.e., 2D images or depth data. Methods on depth data can provide high accuracy, however they require special hardware (e.g. expensive depth sensor) and need more computations. In this study, we focus on 2D images. Lots of work have been done on head pose estimation for 2D images, some are based on local facial features (Shotton and Fitzgibbon, 2011; Sun and Kohli, 2012; McFarlane, 2002), while others are based on the globe image (Dantone and Gall, 2012; Gourier and Hall, 2004; Li and Wang, 2010). However, illumination variation, occlusion and low image resolution make the estimation task difficult. Hence, a Dirichlet-tree distribution enhanced Random Forests approach (D-RF) is proposed in this paper to estimate head pose efficiently and robustly under various conditions.

Random Forest (RF) (Breiman, 2001) is a popular method in computer vision given their capability to handle large training datasets, high generalization power and speed, and easy implementation. Some works showed the power of random forest in mapping image features to votes in a generalized Hough space (Gall and Lempitsky, 2009) or to real-valued functions (Sun and Kohli, 2012). Recently, multiclass RF has been proposed in (Huang and Ding, 2010) for real-time head pose recognition from 2D video data and 3D range images (Fanelli and Gall, 2011; Fanelli and Weise, 2011; Shotton and Fitzgibbon, 2011). Furthermore, Gall *et al.* (McFarlane, 2002) improved the classification rate by modifying the optimization scheme at each node of the trees. Matthias *et al.* (Dantone and Gall, 2012) proposed a conditional random forest to estimate head pose under various conditions only in the horizontal direction. The accuracy rate reaches 72.3% with five yaw angle classes. In order to improve the accuracy and efficiency, a Dirichlet-tree distribution algorithm is introduced into random forest framework to estimate head pose.

The Dirichlet-tree distribution was proposed by Dennis (Minka, 1999). It is the distribution over leaf probabilities that results from the prior on branch probabilities. Minka proved the high accuracy and efficency of the distribution. Some researchers used a Dirichlet-tree distribution in multi-objects track-

Liu Y., Chen J., Liu L., Gong Y. and Luo N..
Dirichlet-tree Distribution Enhanced Random Forests for Head Pose Estimation.
DOI: 10.5220/0004825000870095
In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2014), pages 87-95
ISBN: 978-989-758-018-5

ing (Yan and Han, 2011) and affective computing (Figueiredo and Jain, 2002). In this work, the D-RF is proposed to estimate head poses in vertical and horizontal directions under various conditions (occlusion, different expression, low image resolution and various, illuminations) as shown in Figure 1, where the estimation results are given in the upper left corner in the images.

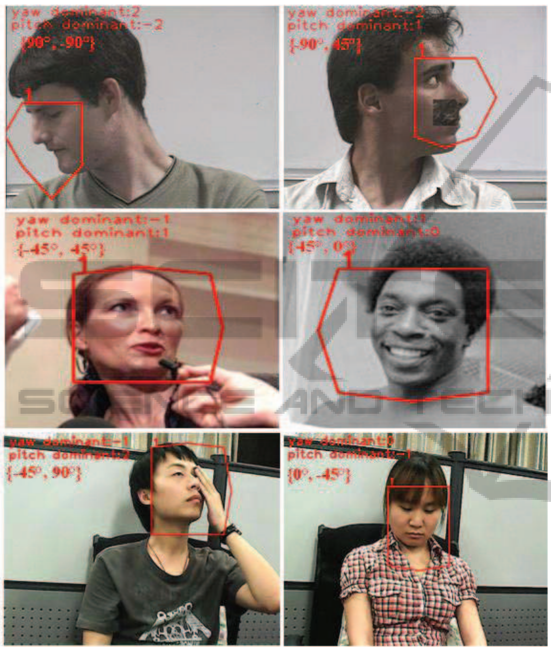


Figure 1: Examples of head pose estimation in the horizontal and vertical direction under various conditions.

The main contributions of this paper are as follows. First, in order to eliminate the influence of occlusion and noise, histogram distributions of facial squares and Gabor features based PCA are extracted for positive and negative patch classification, where PCA is used to reduce Gabor feature dimensions. Then, a D-RF approach is proposed to estimate head poses in a coarse-to-fine way. Meanwhile, an adaptive Gaussian mixture model is introduced in the classification framework to improve the accuracy. Details are discussed in the following sections.

The rest of the paper is organized as follows: Sections 2 details about Dirichlet-tree distribution enhanced random forests for multiclass head pose estimation; Section 3 presents the experiment results and discussions; Section 4 gives the conclusions.

2 D-RF FOR HEAD POSE ESTIMATION

The flowchart of the proposed approach is given in Figure 2. In the first stage, facial patches are extracted and classified. In the second stage, a Dirichlet-tree distribution is introduced into the random forests framework to estimate head pose in the horizontal and vertical directions. The proposed D-RF consists of four layers. D-L1 and D-L2 are two layers in the horizontal direction, D-L1 represents the coarse classification while D-L2 is the refined classification. D-L3 and D-L4 are two layers in the vertical direction. D-L3 represents the vertical coarse classification based on the refined classification in the horizontal direction, while D-L4 the represents final refined classification in the vertical and horizontal directions. Details are given in the following.

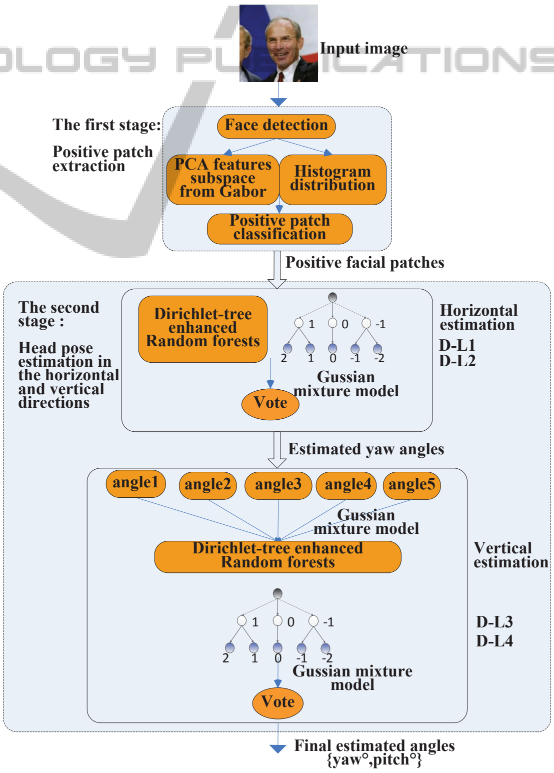
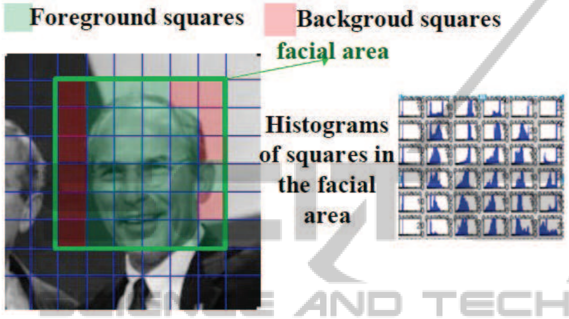


Figure 2: The flowchart of the proposed approach.

2.1 Facial Patch Extraction and Classification

The extracted facial area using Jones&Viola detector (Jones and Viola, 2003) usually includes some noise for head pose estimation, such as hair, neck and occlusion. In order to eliminate noise, the facial area is segmented into foreground and background areas. The foreground areas include positive patches and negative patches, where the positive patches are contribute to estimate head pose while the negative pathces including occlusion or noise may introduce errors for the task. In the work, we segment background areas based on histogram distributions firstly(see Figure 3). The process of positive facial patch extraction is given in Figure 4.



(a) (b)

Figure 3: Foreground and background squares.

Step 1. Segment Background Squares: The detected facial area is divided into 6*6 non-overlapping squares as shown in Figure 3(a), histogram distributions of the squares are computed as shown in Figure 3(b). We utilize the uniformity of histogram distributions of the squares to segment most of the background areas.

Step 2. Classify the Positive and Negative Patches: 200 patches are randomly extracted from the rest of patches with background removed, which includes positive and negative facial patches. The positive and negative patches are classified using RF (Fanelli and Gall, 2011; Breiman, 2001; Dantone and Gall, 2012). In order to model the random tree, the train set of positive facial patches are labelled as 1 and the negative facial patches are labelled as 0. A tree T grows based on Gabor features and gray histograms distribution of the labelled patches. The training and testing similar to the RF (Fanelli and Gall, 2011; Breiman, 2001; Dantone and Gall, 2012). When all test patches P arrive at leaves of trees in the forest, we use the probability $p(c=k|l_t(P))$ stored at the leaf to judge whether the test patch belongs to a class k , where $k=1$ represents the positive patches while $k=0$ represents the negative patches. The probability of the forest is obtained by averaging over all trees' leaves:

$$p(c_i|P)=\frac{1}{T}\sum_t p(c=k|l_t(P)) \quad (1)$$

where l_t is the corresponding leaf for the tree T_t . The algorithm diagram is shown in Figure 4.

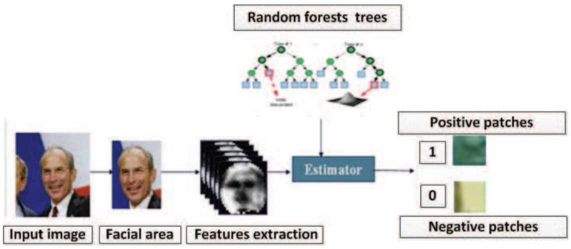


Figure 4: Positive patches extraction and classification.

2.2 Head Pose Estimation using D-RF

Methods using RF to estimate head pose in the horizontal direction are presented in (Dantone and Gall, 2012; Murphy-Chutorian and Trivedi, 2009) . Actually, head pose in the vertical direction is also useful to indicate a person's attention. Hence, D-RF is proposed in this work to estimate head pose in both horizontal and vertical directions under various conditions.

2.2.1 D-RF

The Dirichlet-tree is the distribution over leaf probabilities $[p_1...p_i]$ that results from this prior node probabilities $[a_1...a_k]$ on branch probabilities b_{ji} (Minka, 1999), where i is the number of a leaf , k is the number of a prior node, j is the layer of a branch as shown in Figure 5. Because of this distribution's high accuracy and efficiency (Minka, 1999), it is introduced into the random forests framework to estimate head pose in a coarse-to-fine way in the paper.

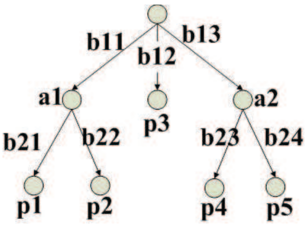


Figure 5: A general Dirichlet-tree distribution.

From the Dirichlet-tree distribution (see Figure 5), it is noted that each child layer in the forest is related to his parent. Hence, the D-RF only computes the probability of a tree in the child layer instead of all trees' probabilities in the forest. While the original random forest estimates the head pose using leaves' probabilities of all training trees to vote in the horizontal and vertical directions. Therefore, D-RF can provide high accuracy and efficiency. The training and testing of the D-RF are given below.

Training. Each tree T in the forest $T = \{T_t\}$ is built and selected randomly from a different set of the training images. From each image, we extract a set of facial patches $P_i = \{I_i, C_i\}$. Where I_i represents the appearance and C_i represents the set of annotation angles of different head poses in the Dirichlet-tree.

In our case, the patch appearance I_i is defined by multiple channels $I = (I_i^1, I_i^2, I_i^3)$. I_i^1 contains the gray values of the raw facial patch with dimension as 31*31. I_i^2 represents the Gabor features based PCA of positive facial patches with dimensions as 35*12. I_i^3 is the histogram distributions of the patches. The set of $C_i^n = (c_i^1, (c_i^2|c_i^1), (c_i^3|c_i^2, c_i^1), (c_i^4|c_i^3, c_i^2, c_i^1))$ contains the annotated discrete angles in different layers of the Dirichlet-tree, where c_i^1 are 3 yaw rotation angles in the first layer of the Dirichlet-tree distribution , $c_i^2|c_i^1$ are 5 yaw angles refined from c_i^1 in the second layer, $c_i^3|c_i^2, c_i^1$ are 15 pitch angles under condition of each yaw angle c_i^2 in the third layer, $c_i^4|c_i^3, c_i^2, c_i^1$ are 25 refined angles based on the above annotated angles at leaves of the Dirichlet-tree in the fourth layer.

We define a patch comparison feature as our binary tests ϕ, similar to (Fanelli and Gall, 2011; Breiman, 2001; Dantone and Gall, 2012):

$$\phi = |R_1|^{-1} \sum_{j \in R_1} I^f(j) - |R_2|^{-1} \sum_{j \in R_2} I^f(j) > \tau \qquad (2)$$

where R_1 and R_2 are two random rectangles within the positive facial patches, $I^f(j)$ is the feature channel $f \in \{1, 2...\}$ and τ is a threshold.

The training of a sub-forest in the D-RF is given below:

1. Divide the set of patches P into two subset P_L and P_R for each ϕ.

$$P_L = \{P|\varphi < \tau\}, P_R = \{P|\varphi > \tau\} \qquad (3)$$

where ϕ is the patch comparison feature (Eq.(2)) and τ is a threshold that has been predefined.

2. Select the splitting candidate ϕ which maximizes the evaluation function Information Gain(IG).

$$IG = \arg\max_{\phi}(H(P|a_j) - (\omega_L H(P_L|a_j) + \omega_R H(P_R|a_j))) \qquad (4)$$

where ω_L, ω_R are the ratio between the number of samples in set P_L (arriving to left subset using the binary tests), set P_R (arriving to right subset using the binary tests) and set P (total node samples). $H(P|a_j)$ is the defined class uncertainty measure and the entropy of the continuous patch labels,

$$H(P|a_j) = -\sum_{i=1}^{N} \frac{\sum_i p(c_i|a_j, P_n)}{|P|} \log(\frac{\sum_i p(c_i|a_j, P_n)}{|P|}) \qquad (5)$$

where $p(c_i|a_j, P_n)$ indicates the probability that the patch P_n belongs to the head pose class c_i in the subfores a_j of the j-th layer in the D-RF.

3. Create a leaf l when IG is below a predefined threshold or when a maximum depth is reached. Otherwise continue recursively for the two subsets P_L and P_R at the first step.

Testing. We initially run a positive facial patch extraction algorithm (Sec.2.1) to find the position and the size of the positive patch. Each positive facial patch is then fed to the trees in the D-RF. At each node of a tree, the patches are evaluated according to the stored binary test and passed either to the right or left child until a leaf node is reached. By passing all the positive patches down all the trees in the D-RF for head pose estimation, each positive patch P_n ends in a set of leaves L of the different sub-forest of D-RF instead of ending all leaves of the random forest. In each leaf l, there are classification probabilities of head pose and the distributions of the continuous head pose parameter by a multivariate Gaussian as in (Breiman, 2001; Dantone and Gall, 2012):

$$p(c_i^m|l_{a_j} = N(c_i^m|a_j; \overline{c_i^m|a_j}, \sum\nolimits_{a_j}) \qquad (6)$$

where $c_i^m|a_j$ and $\sum_{a\,j}$ are the mean and covariance matrix of the head pose classification probabilities of the sub-forest a_j of the j-th layer in the D-RF.

When the patch reaches to the leaves of the sub-forest, the next sub-forest from D-RF should be loaded based on the prior class decision $C(P)$.

The class decision function of the sub-forest is defined as

$$C(P) = \arg\max_{a_j \in C_i^n} p(c_i|a_j, p) \qquad (7)$$

where $p(c_i|a_j, p)$ is the estimation probability of D-RF in condition of sub-forest a_j of the j-th layer and is computed by an adaptive Gaussion mixture model described in the following. The final head pose is then obtained by performing adaptive Gaussian mixture model for voting.

2.2.2 Adaptive Gaussian Mixture Model for Voting

Because the D-RF is a distribution of multi-layer random forests, so an adaptive Gaussian mixture model is introduced into classify final head pose probability in this study. We can improve the Eq.(6) as

$$\begin{gathered} p(c_i|a_j, l_{ji}) = N(c_i|a_j; \overline{c_i|a_j}, \sum\nolimits_{l_{ji}}), \\ c_i|a_j = \{\delta_{ji}(k) \cdot c_i^j(k)\}, k = 1, 2, 3, 4 \\ c_i \in \{-90^\circ, -45^\circ, 0^\circ, 45^\circ, 90^\circ\} \end{gathered} \qquad (8)$$

where j is the sub-forest number in the layer of the D-RF, i is the child node of the sub-forest j, k is the layer number in the D-RF, $c_i|a_j$ and $\sum_{l_{ji}}$ are the mean and covariance matrix of the*ii*-th head pose class under the j-th layer in the D-RF.

When presenting with a test image, the adaptive Gaussian mixture model can adaptively select T_t trees from D-RF based on the estimated probability $p(\alpha|P)$ in different layer nodes of the D-RF. Similarly, the probability $p(\alpha|P)$ can be learned by a RF on the prior training set α. To this end,

$$p(c_i|\alpha, P) = \frac{1}{T_t}\sum_{j}\sum_{t=1}^{k_j} p(c_i|l_{t,a_j}(P)) \qquad (9)$$

where l_t, a_j is the corresponding leaf for patch P of the tree in D-RF. The discrete values k_j are computed such that $\Sigma_j k_j = T_t$ and

$$k_j \approx T_t \cdot \int_{\alpha \in a_j} P(\alpha|P)da \qquad (10)$$

2.2.3 Head Pose Estimation in the Horizontal and Vertical Direction

In order to obtain head pose estimation in the horizontal and vertical directions under various conditions, D-RF is trained as described in Sec.2.2.1. Since it is difficult to obtain continuous ground truth head pose data from 2D images, we annotate rotation angles as ”1, 0, 1” and ”2, -1, 0, 1, 2” in two layers. ”1, 0, 1” represent yaw rotation angles as ”$90°, 0°, 90°$” and ”2, -1, 0, 1, 2” represent refined yaw rotation angles as ”$90°, -45°, 0°, 45°, 90°$”. We store the multivariate adaptive Gaussian distribution in the leaf as define in Eq.(8). The Dirichlet-tree distribution (Figure 5) is introduced to RF as D-RF (see Figure 6 and 7). Figure 6 shows the framework of head pose estimation using D-RF in the horizontal direction, where a is the estimation result in the horizontal direction, and D-L1 and D-L2 are two layers in the horizontal direction in D-RF. Then, five yaw angles can be estimated in the second layer in the D-RF.

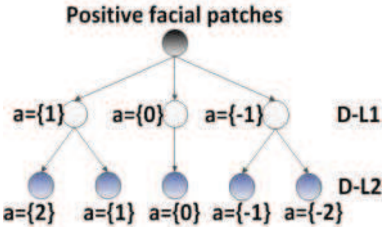


Figure 6: Head pose estimation in the horizontal direction.

After the yaw angles have been classified, pitch angles are estimated under the condition of the classified yaw angles a. Figure 7 shows the framework of estimation using D-RF in the vertical direction, and D-L3 and D-4 are two layers in the vertical direction in D-RF. And the angle annotation in the vertical direction are similar to horizontal rotation angles. When the patches are sent down through all vertical layers in D-RF, sub-trees are selected from sub-forests in D-L3 and D-L4 layers of the D-RF using Eq.(10) and Eq.(8). Finally, we can estimate 25 discrete yaw and pitch angles that are stored at leaves of the D-RF, i.e. $\{90°, 90°\}, \{90°, 45°\}...\{0°, 0°\}...\{-90°, -90°\}$.

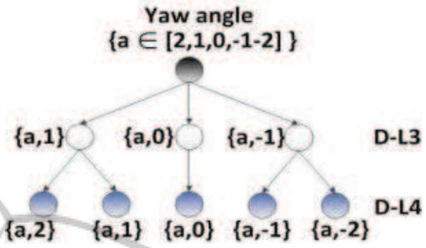


Figure 7: Head pose estimation in the vertical direction.

3 EXPERIMENTS

The proposed approach have been tested with Pointing’04 head pose database (Gourier and Hall, 2004), LFW database (Huang and Ramesh, 2007) and the database of our laboratory (see Figure 8). The Pointing’04 database consists of 2940 images with different poses and expressions. The LFW database consists of 5749 individual facial images. The images have been collected ’in the wild’ and vary in poses, lighting conditions, resolutions, races, occlusions, and make-up. Our laboratory database has been collected using 20 different persons with different poses, expressions and occlusions, and the reference angles have been annotated using the method similar to LFW (Huang and Ramesh, 2007).

For evaluation, we divided the datasets into a training set and a testing set. The training set consists of 2100 images from Pointing’04 database. The testing set includes the rest of 840 images from Pointing’04 database, 1500 images from LFW database and 200 images from our lab database.

3.1 Training

For training the trees in the Pointing’04 database, we fixed some parameters on the basis of empirical observations, e.g., the trees have a maximum depth of 15 and at each node we randomly generate 2000 splitting candidates and 25 thresholds. Each tree grows based on a randomly selected subset of 186 images. Sub-trees in different layers of the Dirichlet-tree have been trained independently.

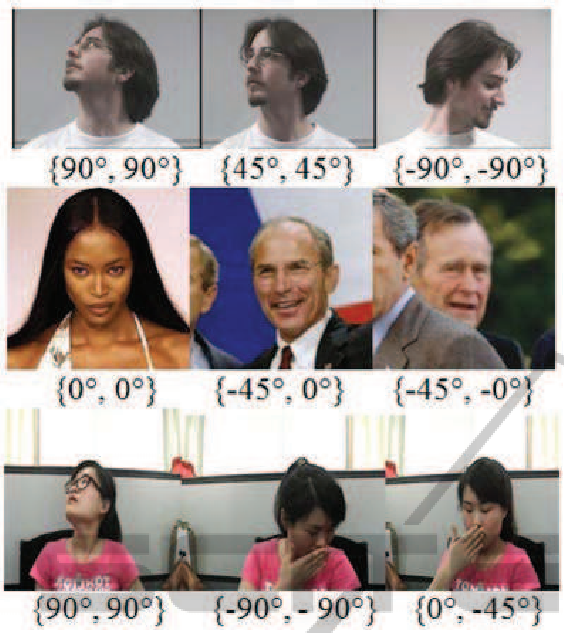


Figure 8: Examples of images from the databases, Pointing'04 database (the first row), LFW database (the second row), and our lab database (the third row).

3.2 Testing

In order to evaluate the proposed approach, estimation accuracy is defined in Eq.(11), where *Num* is the number of correct estimation samples in the testing set and *Total* is the number of testing images. Let Y_0, Y_1, Y_2, Y_3, Y_4 be the estimation accuracies of 5 yaw angles and P_0, P_1, P_2... be the estimation accuracies of the pitch angles under the correspondent yaw angle. $Q(P_i|Y_i)$ denotes the final estimation accuracy in leaves of the last layer, which is defined as:

$$Accuracy = \frac{Num}{Total} \tag{11}$$

$$Q(P_i|Y_i) = \frac{<P_i, Y_i> \cdot P_i}{\sum_{j=1}^{n} <P_j, Y_i> \cdot P_j} \tag{12}$$

D-RF consist of four layers, layer 1(D-L1), layer 2(D-L2), layer 3(D-L3), layer 4(D-L4). Figure 9 shows final estimation accuracies using different layers of D-RF for 25 head pose classes estimation. None represents the average accuracy of 25 head pose classes using the original random forest. While L1 to L4 represent the accuracies of 25 head pose classes using 1 to 4 layers of the D-RF. L1 and L2 represent the estimated average accuracies of 25 head pose classes using only one layer (i.e.D-L1) and two layers(i.e. D-L1 and D-L2) in D-RF, respectively. L3 and L4 represent the estimated average accuracies of 25 head pose classes using three layers (i.e. D-L1, D-L2 and D-L3) and four layers (i.e. D-L1,D-L2, D-L3 and D-L4) in D-RF, respectively. As shown in Figure 9, the final accuracy of original random forest (RF) reaches to 63.23%, and the proposed approach improves the accuracy with the introduction of the different layers of the Dirichlet-tree. The optimal estimation accuracy is 71.83% using 4 layers of the D-RF.

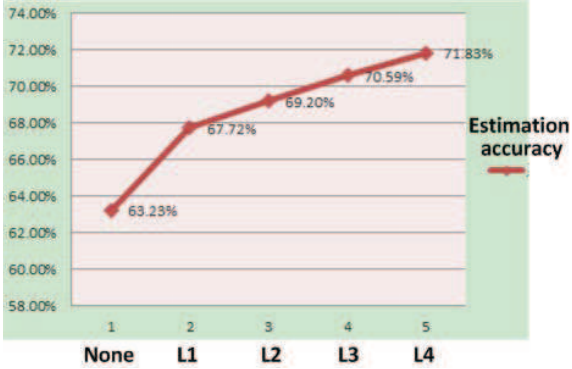


Figure 9: Accuracy comparison in different layers of the D-RF.

3.2.1 Comparison between the D-RF and RF

In order to compare the proposed D-RF with the RF, the same features are used in the comparison experiments.

1) Head pose estimation in the horizontal direction: The estimation results for different yaw rotation angles are presented in Table 1. The first row (RF) is the estimation accuracy using RF and the second row (D-RF) is the estimation accuracy using D-RF. The average accuracy of the D-RF and RF are 83.52% and 78.40% respectively. D-RF provides higher average accuracy than RF in the horizontal direction.

Table 1: The Yaw estimation accuracies (%) comparison.

	90°	45°	0°	−45°	−90°
RF%	80.59	76.79	78.66	75.62	80.29
D-RF%	82.63	83.87	82.2	83.25	84.14

2) Head pose estimation in vertical and horizontal directions under various conditions: The experiments results are shown in Table 2, where columns D-RF and RF describe the accuracies obtained using the proposed D-RF and RF respectively.

As shown in Table2, the average accuracy of the D-RF and RF are 71.83% and 62.23% respectively. D-RF provides higher average accuracy than RF in the horizontal and vertical directions.

3) Computation time: The experiments have been

Table 2: The final estimation accuracies (%) using D-RF and RF.

Yaw / Pitch	90°		45°		0°		−45°		−90°	
	D-RF (%)	RF (%)	D-RF (%)	RF (%)	D-RF (%)	RF (%)	D-RF (%)	RF (%)	DRF (%)	RF (%)
90°	72.1	61	69.7	62.5	72.3	71.6	68.3	55.6	70.6	65.9
45°	73	52.3	72.6	73.1	73.2	69.4	71.9	43.9	71.5	66.2
0°	79.3	75.2	75.9	64.1	78.7	70.6	74	69.8	80	75.7
−45°	72.4	66	70.5	72.8	70.1	73.2	68.8	50.7	67.9	49.4
−90°	67.2	58.8	70.3	60.3	70.7	67	69.4	45.2	65.3	60.4

Table 3: Computing time in D-RF and in RF.

Time $\bar{u}$(s) /algorithm	Positive patches extraction	Yaw estimation	Pitch estimation	Total running Time
Our approach	0.206914	0.430245	0.352829	0.98995
RF	—	0.696799	0.671794	1.36859

conducted on a PC with Intel(R)Core(TM) i5-2400 CPU@ 3.10GHz. The computation time of D-RF and RF is given in Table 3. From the table, one can see that the D-RF is faster than the RF.

3.2.2 Results with Different Composition Methods

Experiment results using different composition methods are given in Table 4. The training and testing are based on the proposed D-RF. First, Table 4 describes whether the proposed positive patch extraction benefits the head pose estimation or not. The results show that positive patches extraction can increase the estimation accuracy by 3.31%. Then it gives the results using different image features, one is the feature combination of Gabor features, gray values and Histogram distributions of facial patches and the other is only gray values of raw image patches. Finally, it shows different estimation results with different voting models, i.e. adaptive Gaussian mixture model and fixture Gaussian model. From this table, one can see that the composition method of positive patches extraction, features combination and the adaptive Gaussian mixture model can give optimal estimation accuracy.

Table 4: Comparison of estimation accuracies of different methods.

Different Methods	Accuracy (%)
1.Using Positive patches extraction or not	
Positive areas extracting	71.83
Non Positive areas extracting	67.14
2.Using different features	
Features combination (Gabor+Gray+Histogram)	71.83
Image pixels gray	68.37
3.Using different voting models	
Adaptive Gaussian mixture model	71.83
Fixed Gaussian model	53

3.2.3 Results of the Occluded Face Images

We randomly add black blocks on the images from the databases. Some example results on the occluded test images using the proposed approach are shown in Figure 10, where the estimation results are given in the upper left corner in the images. Comparison results of the same occluded test images using the D-RF and RF respectively are shown in Figure 11. From this figure, one can see that the D-RF performs better than the RF, the D-RF classifies the poses correctly while the RF fails to do it.

4 CONCLUSIONS

In this paper, we propose a robust and efficient approach for head pose estimation in the vertical and horizontal directions under various conditions. First, in order to eliminate the influence of occlusion and noise, Gabor features and gray histogram distributions of facial areas are extracted for positive and negative patch classification. Then, a Dirichlet-tree distribution enhanced random forests approach is proposed to estimate head poses in a coarse-to-fine way. Meanwhile, an adaptive Gaussian mixture model is introduced in the classification framework to improve the accuracy. Experiment results show that the positive feature patch extraction benefits the head pose estimation and the D-RF performs more accurate and

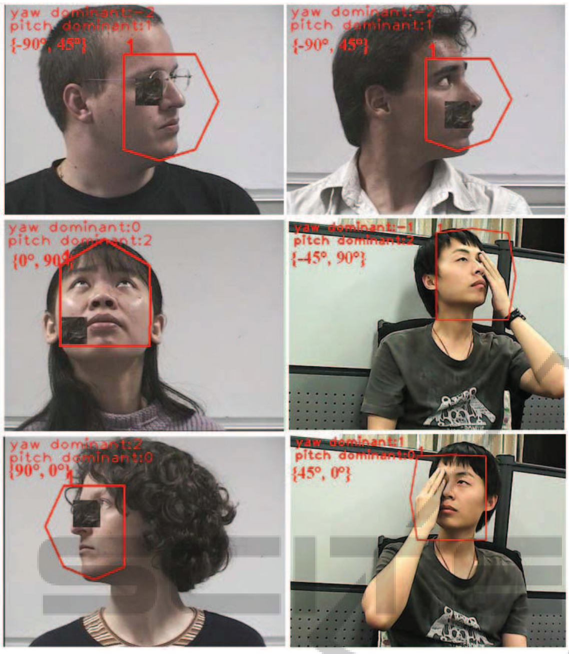


Figure 10: Example results of the occluded test images using D-RF.

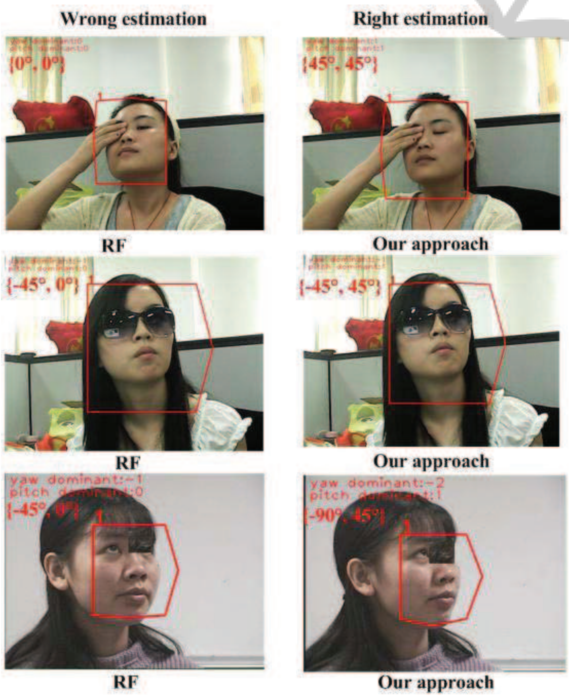


Figure 11: Results of the same occluded test images using the D-RF and RF respectively, where the estimation results are given in the upper left corner in the images.

efficient than the RF . In future work, more experiments will be conducted to evaluate the methods performance under different noise. Also, this method could be used to estimate the head pose in a wide scene, e.g. the attention of students in a classroom.

ACKNOWLEDGEMENTS

This research was supported by the National Key Technology Research and Development Program (No.2013BAH72B01) and MCM20121061, the Natural Science Foundation of Hubei Province (No. 2011CDB159), Research Funds of CCNU from the Colleges Basic Research and Operation of MOE (Grant No.: CCNU13B001), Wuhan Chenguang Project (2013070104010019), Central China Normal University Research Start-up funding (Grant No.: 120005030223) , the Scientific Research Foundation for the Returned Overseas Chinese Scholars (Grant No.: (2013) 693), Hubei province natural science foundationNo. 2013CFB209),New Century Excellent Talents in University (NCET-11-0654),and Young foundation of Huazhong university of science and tecknology, WenHua collоge(J0200540102).

REFERENCES

Breiman, L. (2001). Random forests. In *Machine Learning*.

Chen, J. and Chen, D. (2011). A feature-based detection and tracking system for gaze and smiling behaviours. In *International Journal of Computer Systems Science Engineering. 3: 207214*.

Dantone, M. and Gall, J. (2012). Real time facial feature detection using conditional regression forests. In *CVPR*.

Fanelli, G. and Gall, J. (2011). Real time head pose estimation with random regression forests. In *CVPR*.

Fanelli, G. and Weise, T. (2011). Real time head pose estimation from consumer depth cameras. In *DAGM*.

Figueiredo, M. and Jain, A. (2002). Unsupervised learning of finite mixture models. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*.

Gall, J. and Lempitsky, V. (2009). Class-specic hough forests for object detection. In *CVPR*.

Gourier, N. and Hall, D. (2004). Estimating face orientation from robust detection of salient facial features in pointing 2004. In *ICPR international Workshop on Visual Observation of Deictic Gestures*.

Huang, C. and Ding, X. (2010). Head pose estimation based on random forests for multiclass classification. In *ICPR*.

Huang, G. and Ramesh, T. (2007). Learned-miller. labeled faces in the wild:a database for studying face recognition in unconstrained environments. In *Technical report, University of Massachusetts*.

Li, Y. and Wang, S. (2010). Person-independent head pose estimation based on random forest regression. In *ICIP*.

McFarlane, D. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. In *Human-Computer Interaction*.

Minka, T. (1999). The dirichlet-tree distribution. In *http://research.microsoft.com/minka/papers/dirichlet/minkadirtree.pdf*.

Murphy-Chutorian, E. and Trivedi, M. (2009). Head pose estimation in computer vision: A survey. In *Transactions on Pattern Analysis and Machine Intelligence*.

Shotton, J. and Fitzgibbon, A. (2011). Real-time human pose recognition in parts from single depth images. In *CVPR*.

Sun, M. and Kohli, P. (2012). Conditional regression forests for human pose estimation. In *CVPR*.

Yan, X. and Han, C. (2011). Mutiple target tracking by probability hypothesis density based on dirichlet distribution. In *Journal of XiAn JiaoTong University*.