

Fast and Accurate Head Pose Estimation via Random Projection Forests: Supplementary Material

Donghoon Lee¹, Ming-Hsuan Yang², and Songhwai Oh¹

¹Electrical and Computer Engineering, Seoul National University, Korea

²Electrical Engineering and Computer Science, University of California at Merced

donghoon.lee@cpslab.snu.ac.kr, mhyang@ucmerced.edu, songhwai@snu.ac.kr

In this supplementary material, we describe experimental results which were not included in the paper due to the page limitation. In addition, we also explain label corrections made to mislabeled samples in the HIIT and QMUL datasets [4]. Experimental results on the QMUL dataset with noisy, occluded, and blurred inputs are reported in Section 1. The results show that the proposed algorithm is more robust than other methods. Graphs for analysis of the proposed algorithm are described in Section 2. Examples of the classification and regression on head pose images are described in Section 3 and 4. Section 5 shows experimental results from the original datasets without label corrections. The result shows a similar tendency with the corrected datasets used in the paper. Section 6 describes changes made to the head pose labels in the HIIT and QMUL datasets and shows some examples.

1. Experiments on the QMUL Dataset with Noisy, Occluded, and Blurred Inputs

Figure 1 shows results of the QMUL dataset with noisy, occluded, and blurred inputs. Similar to the HIIT dataset results in Figure 10, 12, and 13 of the paper, it shows that the proposed algorithm is fairly robust against above corruptions. An example of five settings of occluded images are shown in 2.

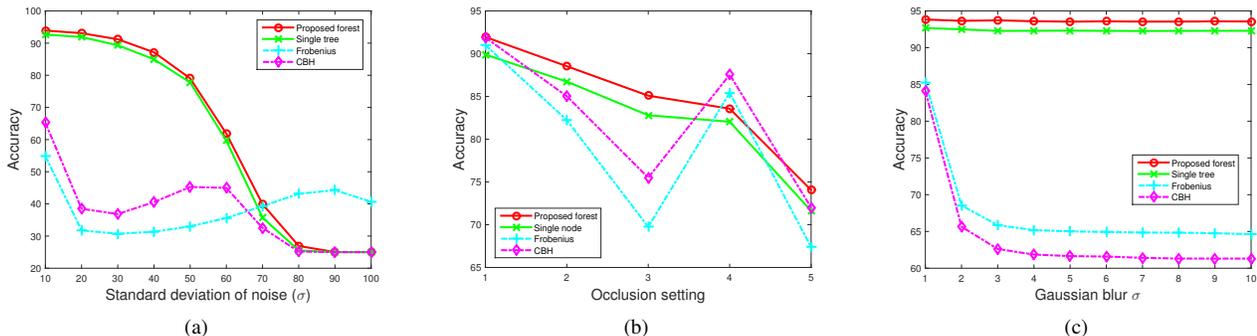


Figure 1. (a) Head pose estimation accuracy at different noise levels. (b) Head pose estimation accuracy at different occlusion settings. The proposed algorithm is robust to occlusions. (c) Head pose estimation accuracy with blurred images. Gaussian blur with filter size 5×5 is applied (using the QMUL dataset).

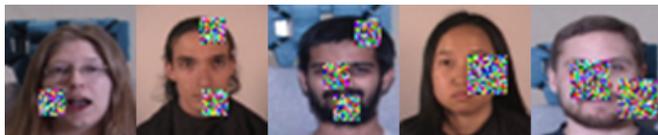


Figure 2. Five settings of occlusion. Occluded regions are randomly generated with random RGB pixel values $[0, 255]$.

2. Analysis of the Proposed Algorithm

In this section, we report the results of effects of random projection matrices and random projection forest parameters in Figure 3(a) and Figure 3(b), respectively.

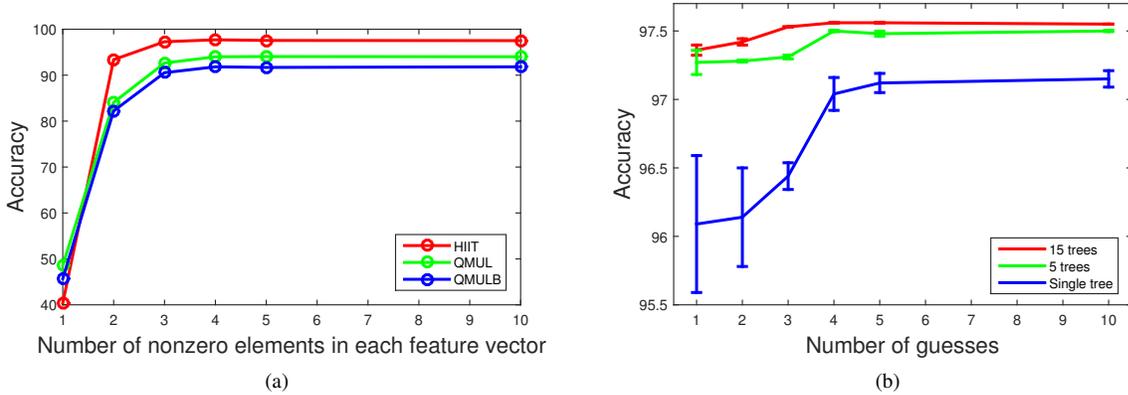


Figure 3. (a) Estimation accuracy with respect to the number of nonzero elements in each feature vector. (b) Estimation accuracy with respect to the number of guesses at each random forest node (HIIT dataset).

3. Classification Examples

We report all misclassified images in the HIIT dataset in Table 1-6. As shown in tables below, failed cases of the proposed algorithm are mainly occurred at the boundary of each class. For example, in Table 2, the proposed algorithm misclassifies fourteen ‘Front left’ images as ‘Front’ class images and misclassified images are actually similar to the ‘Front’ class.

Table 1: Images that are wrongly classified as ‘Front left’ in the HIIT dataset.

Correct label	Method	Images
Front	Proposed	

[5]-CBH		
[5]-Frobenius		
Front right	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Left	Proposed	
	[5]-CBH	

[5]-Frobenius



Table 2: Images that are wrongly classified as 'Front' in the HIIT dataset.

Correct label	Method	Images
Front left	Proposed	
	[5]-CBH	

[5]-Frobenius



Front right

Proposed



[5]-CBH



[5]-Frobenius

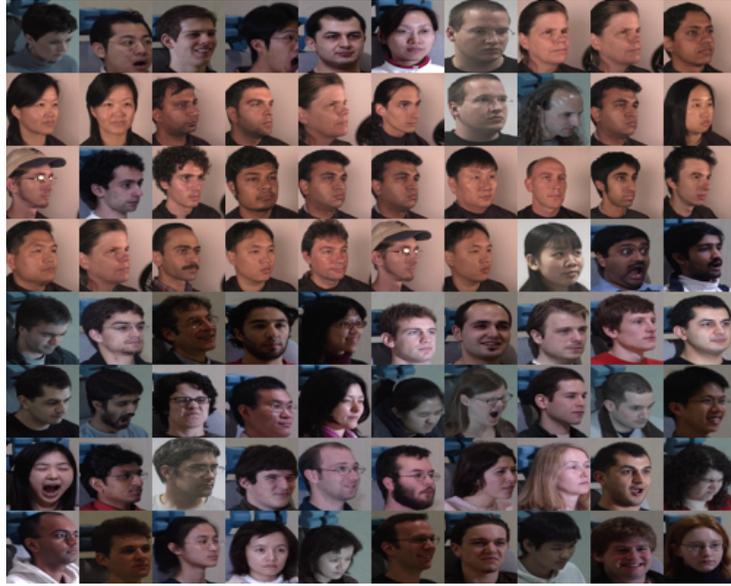
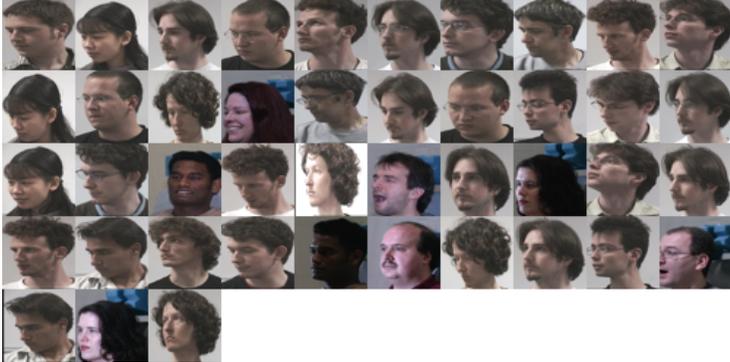


Table 3: Images that are wrongly classified as 'Front right' in the HIIT dataset.

Correct label	Method	Images
Front left	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Front	Proposed	



Table 4: Images that are wrongly classified as 'Left' in the HIIT dataset.

Correct label	Method	Images
Front left	Proposed	
	[5]-CBH	

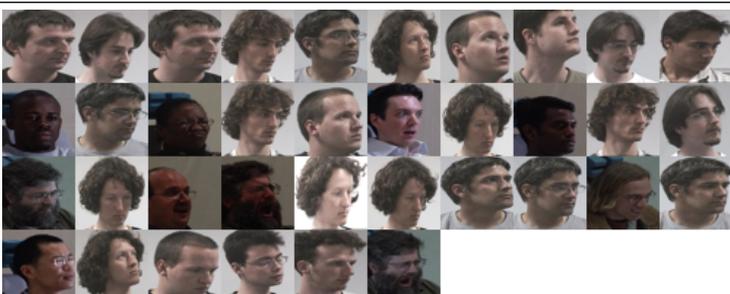
	[5]-Frobenius	
Front	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Front right	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Rear	Proposed	
	[5]-CBH	
	[5]-Frobenius	

Table 5: Images that are wrongly classified as ‘Rear’ in the HIIT dataset.

Correct label	Method	Images
Front left	Proposed	
	[5]-CBH	
	[5]-Frobenius	

Left	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Right	Proposed	
	[5]-CBH	
	[5]-Frobenius	

Table 6: Images that are wrongly classified as ‘Right’ in the HIIT dataset.

Correct label	Method	Images
Front left	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Front	Proposed	
	[5]-CBH	
	[5]-Frobenius	
Front right	Proposed	

[5]-CBH



	[5]-Frobenius	
Left	Proposed	
	[5]-CBH	
	[5]-Frobenius	

Head pose estimation results on the PETS2009 dataset [2] and the Towncentre dataset [1] are described in Figure 4 and Figure 5, respectively. The results show that false positives are effectively removed and head poses are fairly well estimated by the proposed algorithm.

4. Regression Examples

We plot the estimated pose and the ground truth pose of the FacePix dataset [3] in Figure 6. The results show that the proposed algorithm can estimate head poses finely. The errors are mainly generated around the boundary, i.e., when the head pose is at -90° or 90° .

5. Experiments on the Original Datasets

As mentioned in the paper, the original datasets contain mislabeled data. For a fair comparison with other methods, we report results with the original datasets in Table 7. The result shows a similar tendency compared to the corrected datasets as shown in Table 3 of the paper. In addition, the proposed algorithm using a random projection forest is more robust against the mislabeled data. We plan to make the corrected dataset available publicly.



(a)



(b)

Figure 4. Head detection and pose estimation results on Towncentre dataset. For the detection results, the score of each detection box is provided. For the pose estimation results, we use a red arrow to visualize the estimated direction. Its confidence score is written near the box. Dashed-line rectangles are the detections that are estimated as the background (i.e., false positive) by the proposed algorithm.

6. Dataset Corrections

We give an overview of changes made to the head pose labels in the HIIT and QMUL datasets and show some examples. However, we cannot submit the entire corrected datasets due to the file size limitation for supplementary materials. Table 8 shows an overview of changes in the HIIT dataset. Table 9 shows an overview of changes in the QMUL dataset. Examples of corrected HIIT and QMUL dataset images are shown in Table 10 and 11, respectively.

References

- [1] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *Proc. of the IEEE Computer Vision and Pattern Recognition*, pages 3457–3464. IEEE, 2011. 12

Table 7. Estimation accuracy on the uncorrected HIIT, QMUL, and QMULB datasets at different image sizes.

Original Dataset	Size	[5]-Frobenius	[5]-CBH	Proposed
HIIT	15 × 15	82.6%	84.2%	96.48%
	25 × 25	88%	90%	96.86%
	50 × 50	96%	96%	96.93%
QMUL	15 × 15	57.9%	60%	91.33%
	25 × 25	78%	80%	91.49%
	50 × 50	91%	92%	91.55%
QMULB	15 × 15	51.9%	54.2%	88.34%
	25 × 25	74%	76%	88.38%
	50 × 50	90%	91%	88.40%

Table 8. Comparison between the original HIIT dataset and corrected HIIT dataset.

# of images	Original training set	Corrected training set	Original test set	Corrected test set
Front	2000	2003	2000	2009
Rear	2000	2000	2000	2000
Right	2000	2000	2000	2000
Left	2000	1995	2000	2002
Front right	2000	1999	2000	1998
Front left	2000	1997	2000	1988

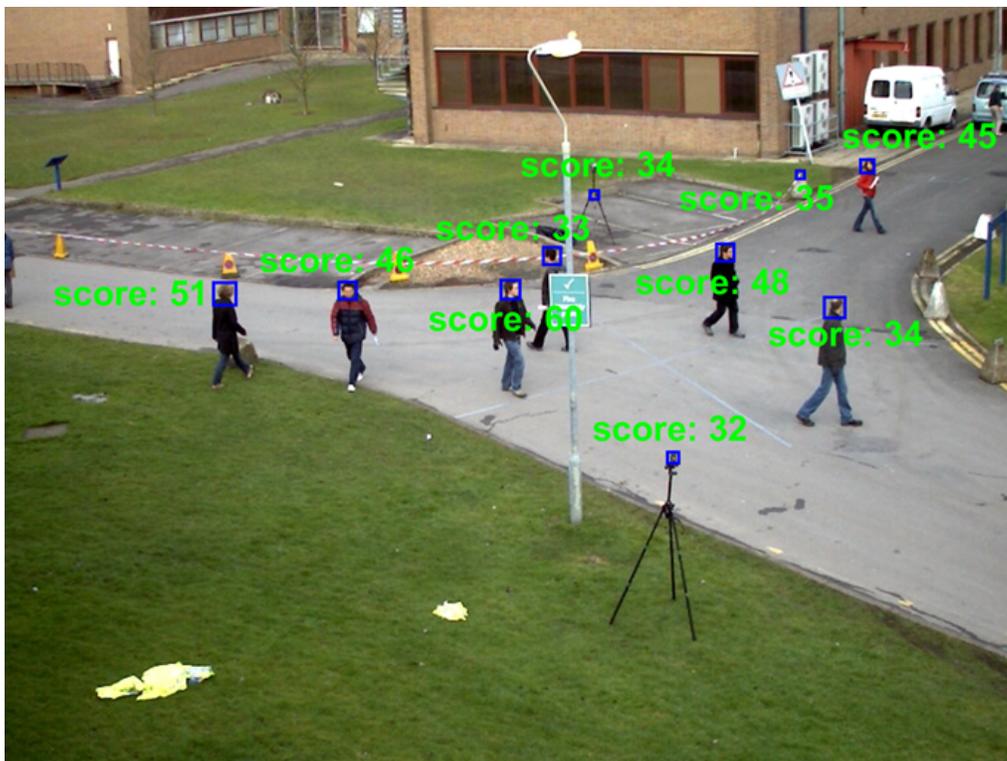
Table 9. Comparison between the original QMUL dataset and corrected QMUL dataset.

# of images	Original training set	Corrected training set	Original test set	Corrected test set
Front	2256	2250	1772	1529
Rear	2256	2244	2096	2064
Right	2256	2105	2248	1966
Left	2256	2157	1502	1345
Background	2256	2246	1107	1053

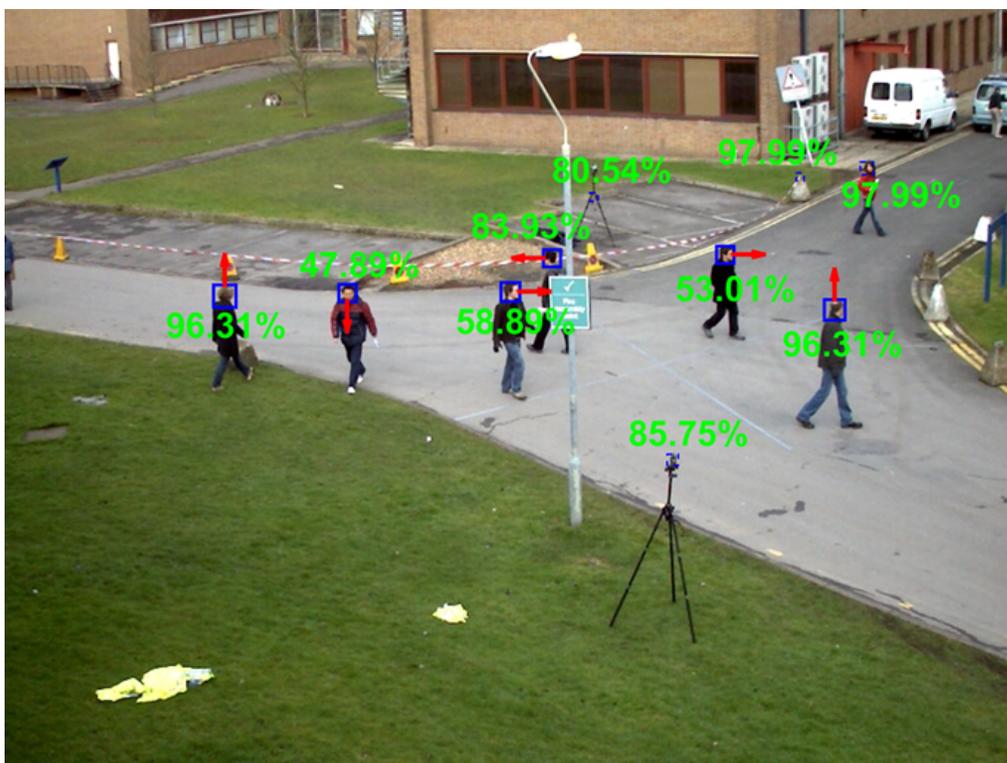
- [2] J. Ferryman and A. Shahrokni. PETS2009: Dataset and challenge. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pages 1–6. IEEE, 2009. 12
- [3] D. Little, S. Krishna, J. Black, and S. Panchanathan. A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005. 12
- [4] D. Tosato. "ARCO (array of covariance matrices), code and datasets,". <https://sites.google.com/site/diegotosato/ARCO>. 1
- [5] D. Tosato, M. Spera, M. Cristani, and V. Murino. Characterizing humans on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1972–1984, 2013. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14

Table 10. Example images that are wrongly labeled in the original HIIT dataset.

File name	Image	Label before correction	Label after correction
head (22).png		Front left (test)	Front (test)
head (65).png		Front right (train)	Front (train)
head (83).png		Front left (test)	Front (test)
head (97).png		Front right (train)	Front (train)
head (99).png		Front left (test)	Front (test)
head (130).png		Front left (test)	Front (test)
head (198).png		Front right (train)	Front (train)
head (304).png		Left (train)	Front (train)
head (318).png		Left (train)	Front (train)
head (321).png		Left (train)	Front (train)
head (341).png		Left (train)	Front (train)
head (354).png		Left (train)	Front (train)

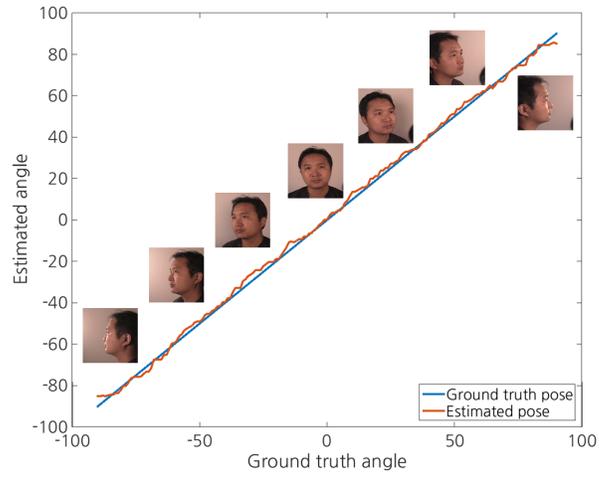


(a)

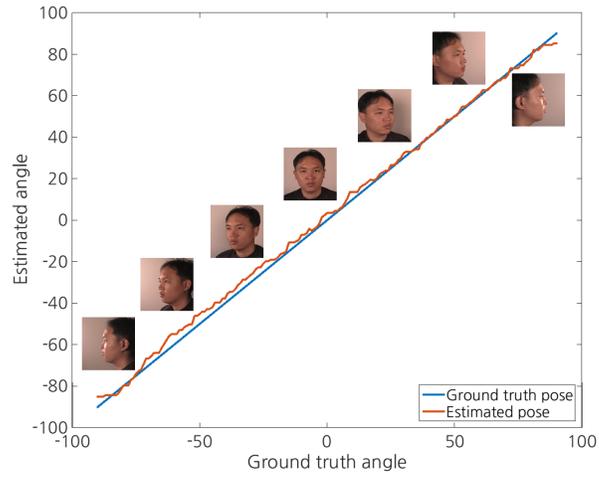


(b)

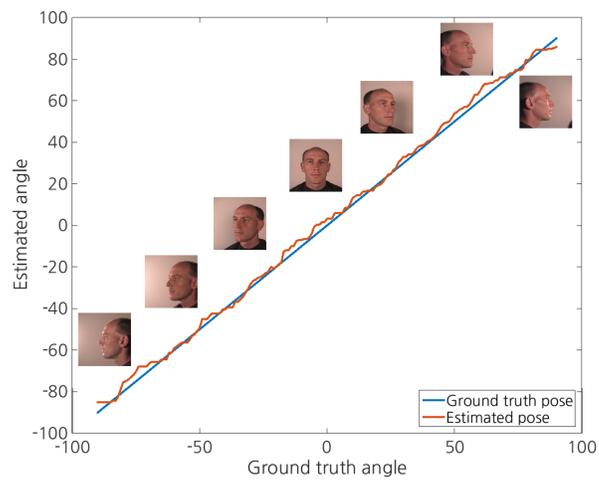
Figure 5. Head detection and pose estimation results on PETS2009 dataset.



(a)



(b)



(c)

Figure 6. Head pose regression results for three different subjects of the FacePix dataset.

Table 11. Example images that are wrongly labeled in the original QMUL dataset.

File name	Image	Label before correction	Label after correction
000099.jpg		Front (test)	Left (test)
001369.jpg		Rear (test)	Front (test)
001663.jpg		Left (test)	Front (test)
001937.jpg		Rear (train)	Left (train)
002604.jpg		Front (train)	Background (train)
003073.jpg		Right (train)	Front (train)
003311.jpg		Right (test)	Front (test)
003944.jpg		Rear (train)	Background (train)
005745.jpg		Left (test)	Front (test)
005801.jpg		Right (test)	Rear (test)
005877.jpg		Front (train)	Background (train)
006248.jpg		Left (train)	Background (train)
neg_000719.jpg		Background (train)	Left (train)
neg_000719_f.png		Background (train)	Right (train)