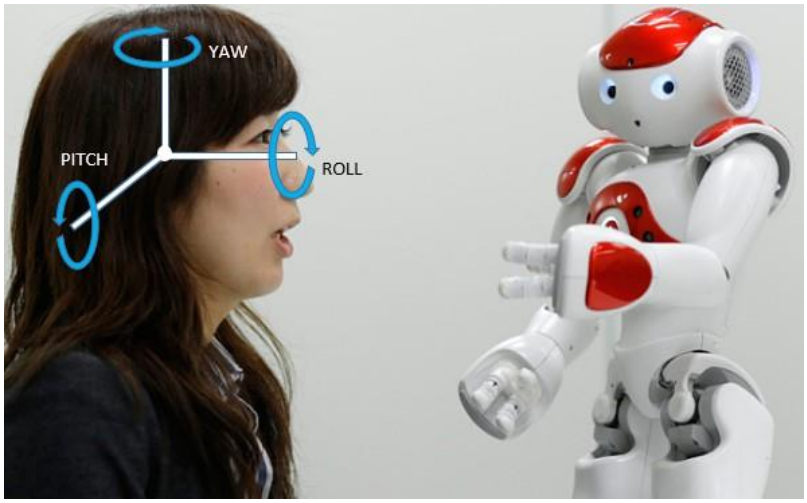
HTI RESEARCH PROJECT
0HM170

The Effect of Improved Face Detection Algorithm on Existing Head Pose Estimation Model



Suvadeep Mukherjee - 0926042
EINDHOVEN UNIVERSITY OF TEHCNOLOGY

Table of Contents

Abstract

Non-verbal ways of communication seems really important these days when we talk about implementation of automated medical assistance in home care settings where aged persons could be inspected by robots continuously. Hence, robots should really be smart to measure the head orientation of any person to get a feeling whether he/she wants to talk to them in a two-way conversations to be happened. But before that it should detect a human face correctly irrespective of the varying room lighting levels or wearing glass or cap etc. The performances of an extensively used face detection algorithm proposed by Viola & Jones (2001) and HOG detection algorithm proposed by Dalal & Triggs (2005) are compared in terms of the range of detected head orientations tested on two publicly available database of face images. For this, an already trained profile face HAAR cascade is used along with the popular frontal face HAAR cascade. Later, Viola-Jones detector is attempted to be used with an improvement in this research where the detector can detect faces maintaining a symmetrical range of orientations as it is found as biased towards right orientation tested on those same face databases. As the head pose algorithm works based on the range of orientations the face detector can detect, it thus can be expected to work for higher head orientations if the detector detects higher face angles. The inclusion of an additional HAAR cascade is also criticized on the basis of the performance comparison with the regular frontal HAAR cascade. Moreover, it is also analysed if the model shows any symmetrical nature in the error estimation in finding yaw angles as the detector works symmetrically with the head angles. Finally, it is attempted to find out the reason behind the large errors occurred in calculating pitch angles as well as for higher yaw angles.

Keywords: robust face detection, head pose estimation, robot interaction.

Introduction

The day is not far when robots would be implemented in a home care system where old-aged people would be assisted as well as cared by them the entire day. For this to be happened successfully, it is crucial that the interaction between the patient and the robot needs to be flawless and clearly interpreted. As we know that humans generally communicate with each other in both verbal and non-verbal ways, it is still a challenge for a robot to communicate with human in the latter one. Speaking about non-verbal ways, eye gaze and head orientation are the most familiar approaches in the era of intelligent autonomous system. Gazing behaviour does not only facilitates conversations (Knapp, Hall & Horgan, 2013) but it also enables us to have a shared focus on an object (Kleinke, 1986). As per the research of Johnson & Cuijpers (2013), gaze direction is linearly related to head pose. Therefore correct estimation of human face orientation in real time is indeed a very crucial task in human-robot interaction. It has been verified and tested that head orientation presumably gives a much better indication of engagement than eye gaze (van der Pol, Cuijpers & Juola, 2011). As Needless to say that the system should work to their variable appearances and wide range of poses that they can adopt along with in different light illuminations. In other words, the head orientation system should be a robust one and a time efficient too.

But before going further, it is always a challenging task for any robots to detect the face first. The problem of finding and analysing faces is a fundamental task in computer vision. Face detection is still dominated by discriminatively-trained scanning window classifiers, most ubiquitous of which is Viola Jones detector due to its open source implementation in the OpenCV library (Zhu & Ramanan, 2012). Apart from this, some other good algorithms are also used such as feature based detection (Yow & Cipolla, 1996), training support vector machines (Osuna, Freund & Girosi, 1997), neural network based detection (Rowley, Baluja & Kanade, 1998) etc. but Viola Jones detector outperforms all of them in terms of processing speed as AdaBoost works very selectively, does not always consider all the outputs of a weak classifier (Viola & Jones, 2001).

Purpose of the research

The purpose of this research is to compare two existing face detection algorithms and based on the results, applying improvements if possible. The improved face detection algorithm is then used to detect faces more efficiently and pass them to the existing robust artificial neural network (based on the work by van der Pol, Cuijpers & Juola (2011)) that would be able to detect the pitch and the yaw angle of a human face more accurately in real time. The reason behind choosing such a non-linear regression method is that the solution is fast and can handle low resolution images efficiently. The neural network works based on training a good range of human faces with varying orientation, illuminations as well as distances. This helps the network to learn about processes and the connections between various nodes which finally helps it to acquire knowledges. According to van der Pol, Cuijpers & Juola (2011) work, to train images, there were two particular databases were used which are freely available for research purposes: Face Pointing04 (Gourier, Hall & Crowley, 2004) and Extended Yale Face Database B (Georghiades, Belhumeur & Kriegman, 2001). The performance of the face detection algorithm in terms of error estimation is then discussed for pitch as well as yaw angles detected. So the research question could be formulated as below:

How to improve a face detection algorithm to detect faces with a larger range of head orientations and how well the improved version of the detector performs in calculating pitch and yaw angles in an existing head pose model.

Huijben (2015) made use of the frontal HAAR like face detector in her research which could detect only faces with yaw angles within [-45°, +90°] and pitch angles within [-15°, +30°] with a detection rate of 85%. Hence it is attempted to increase the range of yaw angles by including a publicly accessible and trained profile face HAAR cascade along with a trained frontal face HAAR cascade which could detect faces beyond ±45° yaw angles. In the process, it is also tested and reported that whether the pitch angles get improved at all from what Huijben (2015) already verified in her research. Another issue which Huijben (2015) encountered in her research was asymmetry in yaw angles detected which was found skewed towards positive yaw angles. Since, human faces are symmetrical in nature horizontally, it is also attempted to flip the faces horizontally if not detected by either of the cascades and then re-apply the cascades again to detect the face. This step has also a positive consequence when faces are trained in the neural network as it works really well with data trained in a

symmetrical fashion because the network may exhibit some chaotic or erroneous behaviour if non-symmetric weights are used (Hopfield, 1982).

Based on the research question that have already been stated, several hypothesises can also be generated which are tested in subsequent sections like below:

Hypothesis 1. The improved face detector shows significant improvement in the range of both yaw and pitch angles in detecting faces.

Hypothesis 2. HOG classifier (Dalal & Triggs, 2005) works better than Viola-Jones detector (HAAR) in terms of range of orientations.

Hypothesis 3. Asymmetry in detecting varying head orientations can be resolved by flipping.

Hypothesis 4. For higher yaw angles such as beyond ±30°, profile face HAAR cascade outperforms frontal face HAAR cascade in terms of correct angle estimation.

Hypothesis 5. Higher oriented faces produce more errors in estimation of yaw angles as the neural network used in this research is not trained with faces of higher yaw angles. Pitch angles should be fine as there is no operation performed to increase pitch angles.

Approach to well tuning

As we know that before the neural network gets fed with the collected images, it is crucial that the system detects the position of faces correctly in the images at first. To do so, it is tested with the accuracy between two specific well recognized object detection methods: HAAR classifier with AdaBoost used by Viola & Jones (2001) and Histogram of Oriented Gradients (HOG) introduced by Dalal & Triggs (2005). It is to mention that Dalal & Triggs (2005) tested their algorithms to detect full bodied humans instead of only faces specifically but their algorithms got improved over time to detect human faces too.

According to Viola & Jones (2001), some predefined and tested HAAR filters can be applied over the images which scans sequentially using a fixed length window and collects as much as features that it could detect to form integral images. They called detected feature sets from each image as weak classifiers which are then processed through Adaboost to fasten the feature selection process. At the end, all the classifiers are cascaded to decide on detection of faces.

Dalal & Triggs (2005) came up with a different approach such that detecting object appearance and shape within an image which can be described by the distribution of intensity gradients or edge directions. The image is first divided into small connected cells and for the pixels within each cell a histogram of gradient directions is compiled. The HOG descriptor then concatenates these histograms. The key advantage of using HOG approach is that it is invariant to geometric and photometric transformations as it operates on local cells.

Although there are extensive research had already conducted based on HAAR face detection as it is fast and more robust in compared to other available approaches currently, it is decided

to test the performance of the HOG approach upon the already mentioned face databases for the face detection in terms of accuracy comparing with the HAAR detection.

In the next section, it is discussed about the criteria and efficiency to decide which among HAAR and HOG to be used for the final face detection method in the real time system based on their performances on some widely used datasets. To investigate on the further effectiveness of HAAR like classifier in face detection, flipping faces horizontally are also tested in case the face is not detected in normal condition as the human face is symmetrical horizontally not vertically. This step is considered because the bigger range of face angles detected the more numbers of yaw and pitch angles the head pose model could calculate in real time.

Databases used for detection

The Face Pointing04 database contains varying appearances of people but with static lighting and white backgrounds. On the other hand, the Yale B database includes face images with varying lighting conditions and backgrounds. The differences are that the Yale B database contains either frontal or left oriented faces only and the amount of faces are almost 6 times of Face Pointing04 database. The Face Pointing04 database contains faces of 15 people with varying yaw and pitch angles. Each person has two series of 93 images in which each series differs sometimes slightly among each other e.g. glasses put on faces. So, there are total of 2790 faces tested with both the above mentioned detection algorithms. On the other hand, the extended Yale B database contains faces of 28 people with 9 different poses and 64 different lighting illumination conditions for each pose. So, total of 16128 faces are tested with the two detection algorithms accordingly.

Algorithms used for face detection

As already discussed earlier that the approach proposed by Viola & Jones (2001) using several HAAR like features and approach proposed by Dalal & Triggs (2005) calculating the histograms of feature gradients to detect a face successfully. For detection using HAAR like features, an extensively used library OpenCV (*ver. 2.4.9*) is used and for HOG detection, an open source library called dlib (*ver 18.6*) is used. In both the cases, publicly accessed already trained classifiers are used. For normal HAAR like detection, two efficient cascades *haarcasecade_frontal_alt2.xml* for detecting frontal faces and *haarcascade_profileface.xml* for detecting profile faces are used provided the frontal face cascade does not detect that face. As we will see later that the detection results are biased towards the right orientation (positive yaw angles) and does not show any symmetry even though human faces are symmetrical in nature horizontally, it is decided to create a mirror image (flipping horizontally) in case both the above mentioned cascades do not detect any face and then apply one after other on the flipped face to detect it successfully. This HAAR like detection with flipping is also discussed in terms on its performance in the next section.

Analysis of different detection methods

This section is divided with methods, results and discussion for each of the detection methods on each of the databases used along with a conclusion on the findings at the end of each database processing.

HAAR like detection (with and without flipping) on FacePointing04 database

Methods

It is found that some of the faces in Face Pointing04 database have yaw or pitch angles too large and therefore the face detection algorithms could not find the face correctly. To systematically remove those highly oriented faces, a threshold detection ratio is set at 85%. In figure 1, angles with a detection ratio above the threshold are drawn with dark blue points and those are less than 85% but greater than 80% have been indicated using red colour. The second threshold of 80% was considered as it is found that the points drawn in the plot are not symmetrical over at least one axis.

Result

HAAR detection without flipping overall detects (yaw, pitch) = ([-15°, +90°], [-15°, +30°]) with detection rate as 80%. But a symmetrical dataset is intended to have as training such dataset in the neural network would make the network unbiased of any left or right orientation. Such second threshold helps to draw a black coloured rectangle as depicted in the figure 1 which is symmetric at least over yaw angle. Hence, the range of the symmetric threshold function is found as (yaw, pitch) = ([-15°, +15°], [-15°, +30°]) with detection rate as 80%.

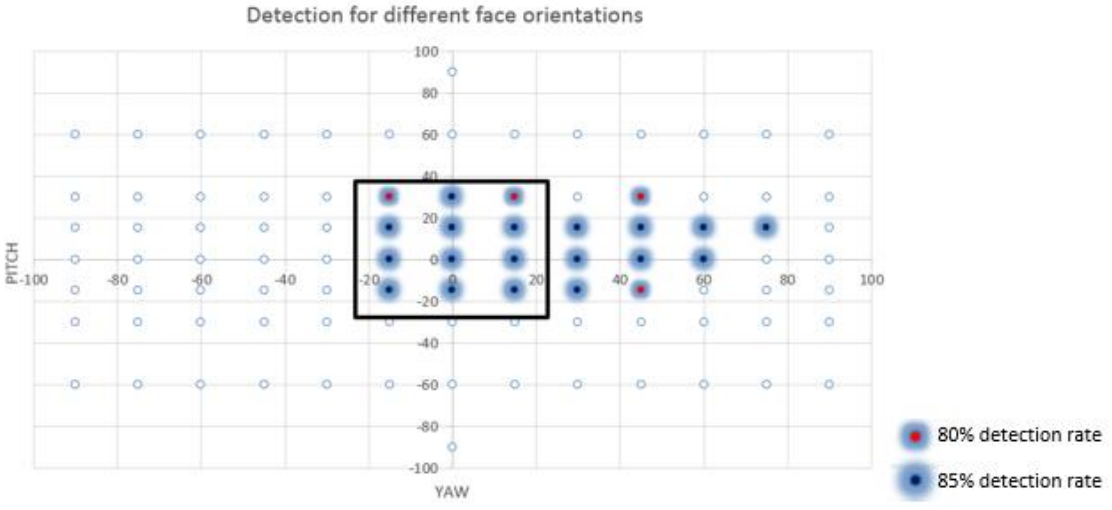


Figure 1. HAAR detection (without flipping) for Face Pointing04 database; Detected: -15≤ yaw ≤+90, -15≤ pitch ≤+30; black coloured boundary is the threshold function

But as it is also found that the plot is skewed much more towards the positive yaw angles (right faces), it is further decided to flip the undetected left oriented faces over y axis and then test again with the same HAAR filters applied. Figure 2 shows the improved version of the detection window.

HAAR detection with flipping overall detects (yaw, pitch) = ([-90°, +75°], [-15°, +30°]) with detection rate as 80%. And the range of the symmetric threshold function (black coloured bounded region) is found as (yaw, pitch) = ([-60°, +60°], [-15°, +30°]) with detection rate as 80%.

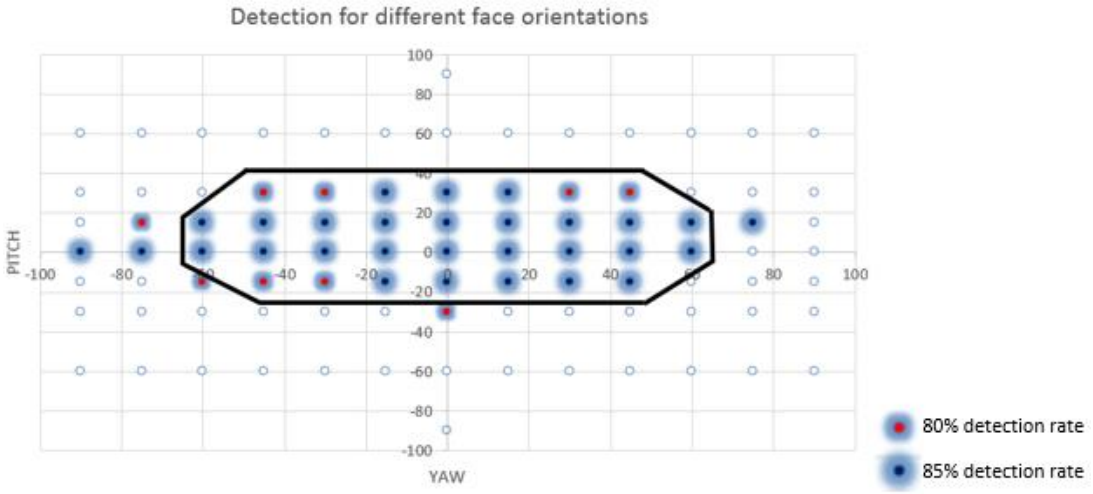


Figure 2: HAAR detection with flipping for Face Pointing04 database; Detected: -90≤ yaw ≤+75, -15≤ pitch ≤+30; black coloured boundary is the threshold function

However, flipping the image indeed takes some additional processing time over simply HAAR detection. As calculated, normal HAAR detection takes 1.5439 secs on average (with standard deviation = 0.7383) to process 30 images i.e. each image takes around 0.05146 secs to get processed and detects faces from those for each of the 93 varying angles on 80% detection threshold whereas flipping and then applying HAAR detection takes 1.5642 secs on average (with standard deviation = 0.9473) to process 30 images i.e. each image takes around 0.05214 secs to get processed and detects faces for each of the 93 varying angles on the same detection threshold.

Discussion

Talking about the effectiveness as a whole, simply HAAR detection detects 1078 faces with only one false positives whereas flipping HAAR detection detects 1569 faces with two false positives. So HAAR with flipping does indeed have an improvement in the detection model.

HOG detection on FacePoint04 database

Methods

Now the same database is tested with HOG filter with taking 85% as threshold detection ratio (dark blue coloured points). Similarly, to draw a symmetrical detection window, a second detection threshold of 80% (red coloured points) is considered too.

Result

Figure 3 depicts the findings below. Unlike HAAR detector, the points are found symmetrically distributed over both yaw and pitch angles. Therefore, no flipping operation is needed additionally for HOG detection. Hence, the range of the symmetric threshold function (black coloured bounded region) is found as (yaw, pitch) = ([-45°, +45°], [-30°, +30°]) with detection rate as 80%.

However, HOG detection takes much more processing time 2.6569 secs on average (with standard deviation = 0.1516) i.e. each image takes around 0.0886 secs to get processed and detects faces from those 30 images for each of the 93 varying angles on 80% detection threshold.

Discussion

Talking about effectiveness, it detects 1428 faces with only one false positive. However the points for HOG detection are found to be symmetrically distributed over both yaw and pitch angle whereas they are found to be symmetrically distributed over only yaw angle for HAAR detection. The processing time is too way higher for HOG detection compared to the HAAR one (with or without flipping). The accuracy of finding faces is found somewhat comparable though between them.

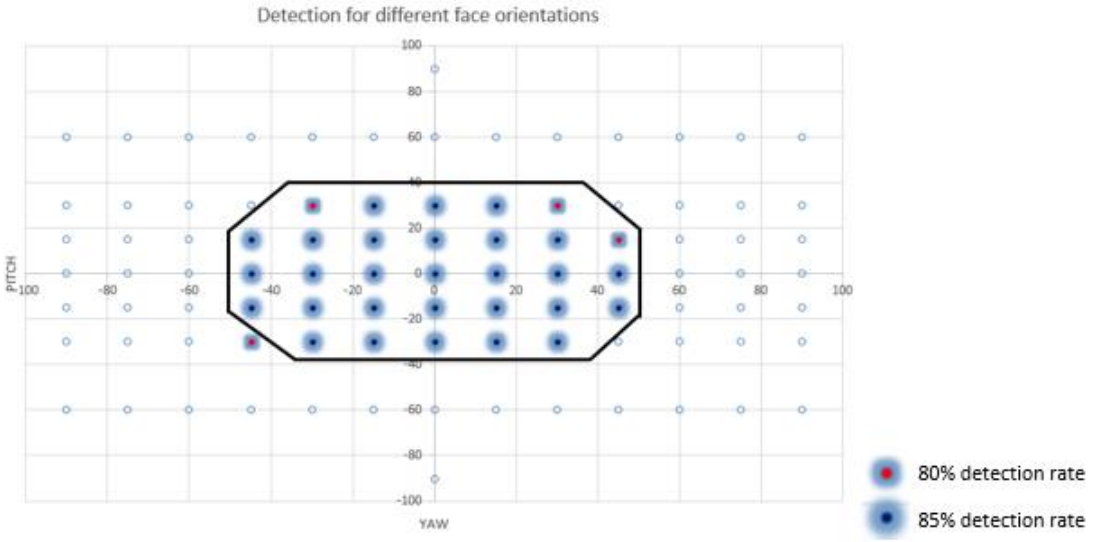


Figure 3: HOG detection for Face Pointing04 database; -45≤ yaw ≤+45, Detected: -30≤ pitch ≤+30; black coloured boundary is the threshold function

Conclusion on detection method based on FacePointing04 database

To compare how much data each of the above discussed methods removes from the database, it is found that around 87% images would be removed if HAAR (without flipping) detection is applied, around 66% images would be removed in case HAAR with flipping is applied and around 67% images would be removed if HOG detection is applied. Since, HAAR with flipping has a wider range of yaw angles detected than the other two, has processing speed (tested on machine with 2.40 GHz processor) almost similar to HAAR without flipping and much lower than that of HOG and retains almost as many images as HOG detector, it indeed outperforms the others in terms of accuracy as well as efficiency for FacePointing04 database images.

HAAR like detection (with and without flipping) on Extended Yale B database

Methods

Next, the faces from Extended Yale B database are tested with each of the detection methods discussed above one at a time. As this database contains images almost 6 times the Face Pointing04 database and there are no such extreme variations of face orientation like Face Pointing04 images, the threshold for detection ratio is therefore set at 95% (dark blue coloured points). However to make the data points more symmetrical, the second threshold is decided to be set at 85% (red coloured points).

Result

Figure 4 depicts the findings for HAAR normal detection. Overall, HAAR detection without flipping detects (azimuth, elevation) = ([-70°, +60°], [-20°, +35°]) with detection rate as 85%. In order to get a symmetrical distribution, the black coloured diamond shaped region is considered as the collection of detected faces under varying azimuth and elevation ranges with detection threshold of 85%. So, the range of the symmetric threshold function (black coloured bounded region) is found as (azimuth, elevation) = ([-10°, +10°], [-20°, +20°]) with detection rate as 85%.

Overall, it takes 40.5715 secs on average (with standard deviation = 3.0921) to process 252 images i.e. each image takes around 0.1609 secs to get processed and detects faces out of those for each of the 64 illumination conditions on 85% detection threshold.

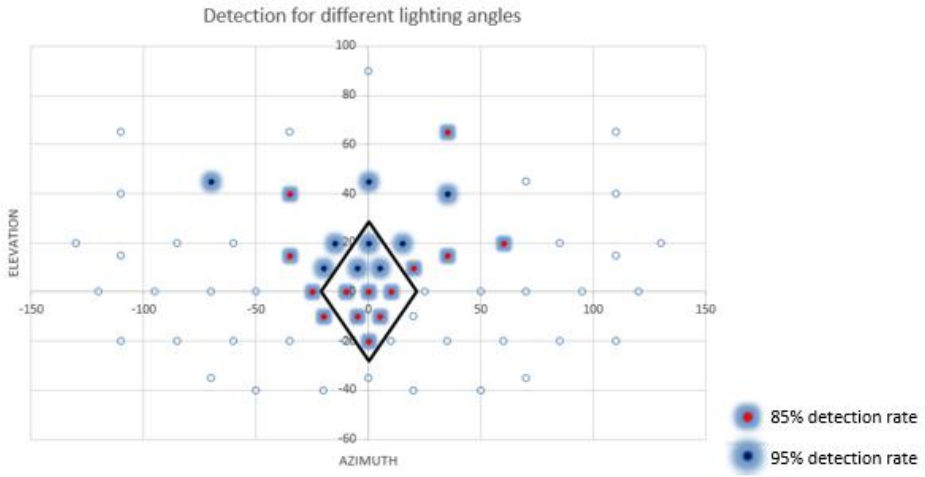


Figure 4. HAAR detection (without flipping) for Yale database; Detected: -70≤ Azimuth ≤+60, -20≤ Elevation ≤+35; black coloured boundary is the threshold function

But as the Extended Yale B database contains 9 different poses with either frontal or left oriented faces, it is further decided to flip the non-detected faces to the other side similarly done for Face Pointing04 database stated above and then test with it again. Overall, HAAR detection with flipping detects (azimuth, elevation) = ([-85°, +70°], [-20°, +65°]) with detection rate as 85%. Figure 5 shows detection window which is symmetrical over azimuth angle only. The range of the symmetric threshold function (black coloured bounded region) is found as (azimuth, elevation) = ([-70°, +70°], [-20°, +65°]) with detection rate as 85%.

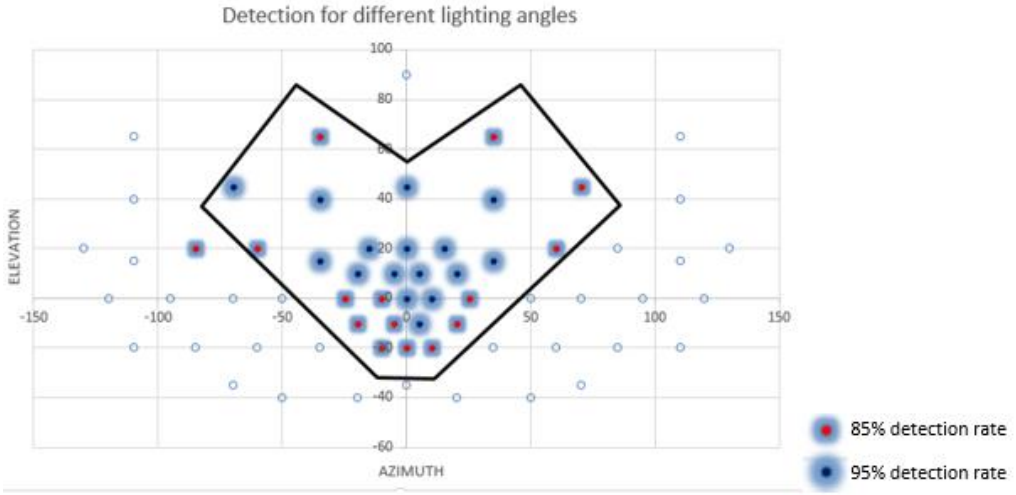


Figure 5. HAAR detection with flipping for Yale database; Detected: -85≤ Azimuth ≤+70, -20≤ Elevation ≤+65; black coloured boundary is the threshold function

Now, it takes 50.8488 secs on average (with standard deviation = 8.9011) to process 252 images i.e. each image takes around 0.2018 secs to get processed and detects faces out of those for each of the 64 illumination conditions on 85% detection threshold. The time taken

is much higher than the simple HAAR detection because of spending addition processing time for flipping.

Discussion

Talking about the effectiveness as a whole, it detects 10975 faces correctly with 79 false positives whereas HAAR detection with flipping detects total of 12476 faces with 94 false positives. This time too, HAAR with flipping does indeed show an improvement in the detection model.

HOG detection on Extended Yale B database

Method

Now the same database is tested with HOG filter with taking 95% as threshold detection ratio (dark blue coloured points). Similarly, to draw a symmetrical detection window, a second detection threshold of 85% (red coloured points) is considered too. So it can now detect faces with symmetrical lighting angles i.e. both for azimuth and elevation angles.

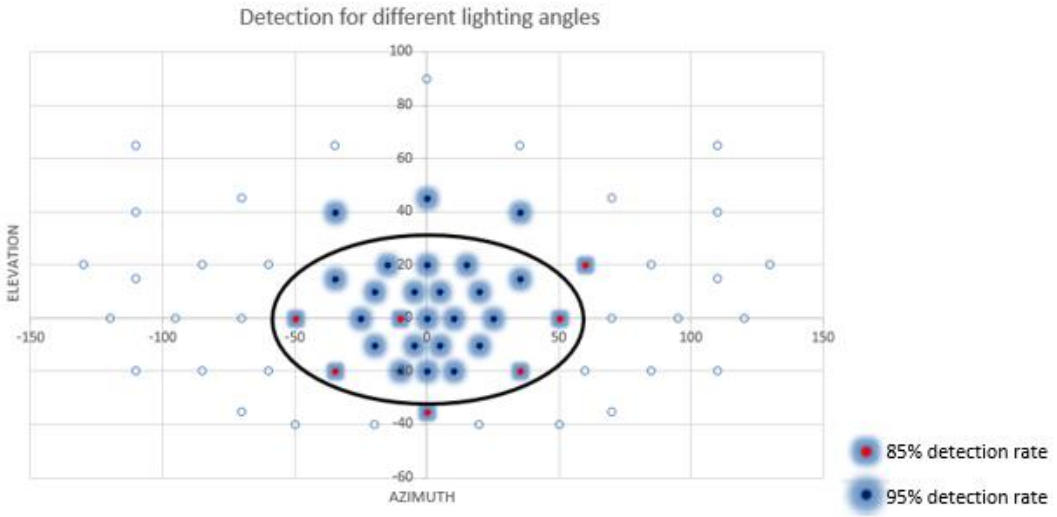


Figure 6. HOG detection for Yale database; Detected: -50≤ Azimuth ≤+60, -35≤ Elevation ≤+45; black coloured boundary is the threshold function

Result

Figure 6 depicts the findings for HOG detection. Overall, HOG detects (azimuth, elevation) = ([-50°, +60°], [-35°, +45°]) with detection rate as 85%. In order to get a symmetrical distribution, a black coloured oval shaped region is considered as the collection of detected faces under varying azimuth and elevation ranges with detection threshold of 85%. So, the

range of the symmetric threshold function (black coloured bounded region) is found as (azimuth, elevation) = ([-50°, +50°], [-20°, +20°]) with detection rate as 85%.

It takes around 60.7903 secs on average (with standard deviation = 5.5073) to process 252 images i.e. each image takes around 0.2412 secs to get processed and detects faces out of those for each of the 64 illumination conditions on 85% detection threshold. So, this kind of detector is much slower than the HAAR detector (with or without flipping).

Discussion

Talking about effectiveness, it detects total of 9343 faces with only 5 false positives. This could snatch the attention from HAAR detector in terms of accuracy in varying lighting conditions. Similar to the FacePointing04 database, the points for HOG detection are found to be symmetrically distributed over both azimuth and elevation angles whereas they are found to be symmetrically distributed over only azimuth angle for HAAR with flipping detection. The processing time is too way higher for HOG detection compared to the HAAR one (with or without flipping). So, the accuracy of finding faces is found somewhat comparable though between them.

Conclusion on detection method based on Extended Yale B database

Finally, to compare how much data each of the above discussed methods removes from the Yale database, it is found that around 86% images would be removed if HAAR (without flipping) detection is applied, around 53% images would be removed in case HAAR with flipping is applied and around 61% images would be removed if HOG detection is applied. Since, HAAR with flipping has a wider range of azimuth as well as elevation angles detected than other two, has processing speed (tested on machine with 2.40 GHz processor) lies in between HAAR without flipping and HOG and retains maximum images than that of others, it indeed outperforms the others in terms of efficiency. But it can still be debatable as HOG detector causes only 5 false positives whereas HAAR with flipping results in 94 false positives. Since, we consider a miss is more expensive than identifying a false face, HAAR with flipping successfully identifies almost 1.5 times faces than HOG detector which in turn supports its better performance than the later.

Discussion on selection of the detection method

As we look at the comparison between different detection approaches tested above, it is evident that HAAR like detection method with flipping images cover the maximum amount of yaw angle of a human face and it's more robust and faster than other two methods. Besides, it allows less exclusion of the faces from the database as the number of faces it detects much higher than others but causes higher number of false alarms too than HOG detector. So there is a trade-off between accuracy and range of possible angles detected. As false alarms could

be considered than being missed a correct face, HAAR detection with flipping is taken as the selected method for further processing.

Database used for testing images

To test and evaluate the performance of the detector, 34% of the remaining FacePointing04 database images are used to calculate the pitch and yaw angles with the help of existing head pose neural network. Those images come with actual yaw and pitch angles of the faces during their creation which in turn are made use of calculating the deviation and error that are measured in subsequent sections.

Evaluation of HAAR detection method (with flipping) using existing head pose model

The head pose estimation model based on the work by van der Pol, Cuijpers & Juola (2011) is considered to be used in evaluating the performance of the selected face detection method discussed in the previous section. The face detection algorithm outputs the region of faces in an image depicted in a rectangular fashion and the centre of that rectangle which is then fed to the head pose model for further processing. But before we start calculating the pitch and yaw angles of the face, it is indeed an important step to pre-process the face in order to feed it into the neural network. The next sections describe some important steps of pre-processing of faces and finally the evaluation of the detection method with the help of head pose model accordingly.

Pre processing

The rectangular shaped cut out face detected is first transformed into a grayscale image. The noise in the image is then removed by applying a Gaussian blur. The histogram equalization is also applied to increase the contrast of the image. Now, the face is cropped in such a way that the output image would consist only 40x90 pixels which would basically contain the most important features in a face surrounded by the centre of the rectangle. Post cropping, the edges of that face are determined as we would only need some important features of the face to feed into the neural network while discarding the redundant information simultaneously. For that, Laplacian filter, a popular edge detection method is used since it calculates the second derivatives to find the edges effectively. Figure 7 illustrates the procedure below.

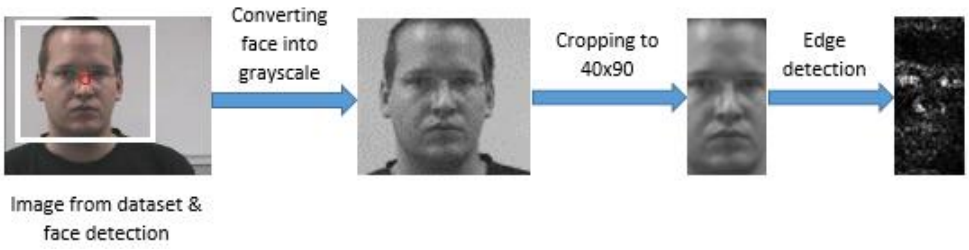


Figure 7. Pre-processing of the detected face

Comparison of performances between frontal & profile HAAR cascade

Now, the pre-processed image is fed into the neural network which returns the calculated pitch and yaw angles based on the already trained .mat file. Some of the faces are detected using frontal HAAR cascade and some others using profile HAAR cascade on which yaw and pitch angles are calculated accordingly. But some of the yaw angles are detected by both of them in different images. Table 1 depicts the comparison between estimated yaw angles along with the standard deviation of the faces detected by either frontal or profile face HAAR cascade. Some missing values are mentioned as *NULL* or '-' in the table as well.

Table 1. Yaw_Estimated vs Yaw_Actual on faces detected by different cascades

Yaw_Actual	**Yaw_Estimated (median)**			
	Frontal HAAR cascade		Profile HAAR cascade	
	Median	Std. deviation	Median	Std. deviation
-60°	-28.78°	-	-33.77°	5.79°
-45°	-30.65°	5.22°	-31.43°	9.11°
-30°	-25.04°	6.19°	-25.03°	12.16°
-15°	-12.17°	6.32°	-	-
+0°	-1.76°	4.57°	-	-
+15°	+14.05°	6.11°	+23.13°	-
+30°	+25.37°	5.93°	+29.06°	5.31°
+45°	+26.61°	5.51°	+30.53°	5.17°
+60°	+27.96°	4.10°	+34.04°	4.12°

Figure 8 presents actual yaw vs estimated yaw for different HAAR cascades. It is found that the differences in estimated yaw angles detected with different HAAR cascades are significant when yaw = +30°, +45° and +60° ($p<0.01$). Looking at the table and figure 8, it can also be inferred that for yaw ≥ +30°, profile HAAR cascade works better than the frontal one considering their median values. It is also found that profile cascade overestimates for yaw = +15°. For yaw = -15° and +0°, all faces are detected using frontal cascade only. For yaw = -60°, profile cascade again outperforms the frontal one as the median values found bigger than the latter. The standard deviation at yaw = -30° is found as largest (= +12.16°) among all which is caused by profile HAAR cascade.

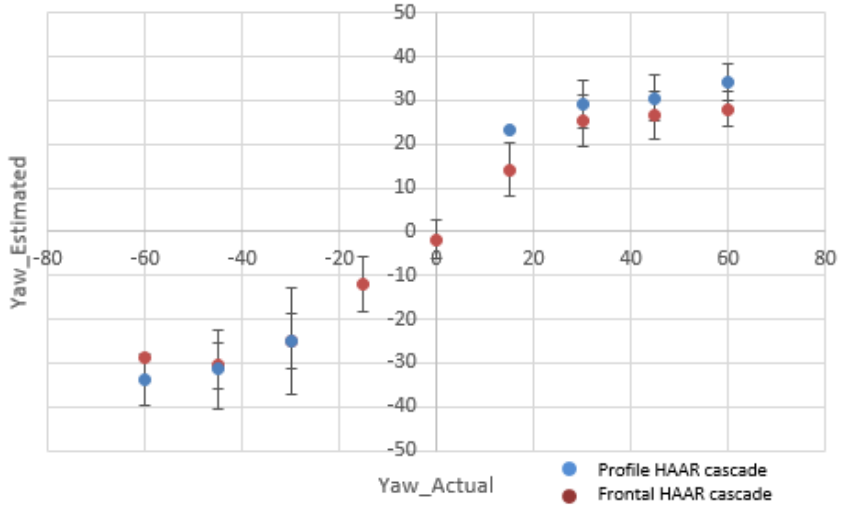


Figure 8. Actual Yaw vs Estimated Yaw for different HAAR cascades

Similarly, table 2 and figure 9 show the comparison between estimated pitch angles along with the standard deviation of the faces detected by either frontal or profile face HAAR cascade.

Table 2. Pitch_Estimated vs Pitch_Actual on faces detected by different cascades

Pitch_Actual	Pitch_Estimated (median)			
	Frontal HAAR cascade		Profile HAAR cascade	
	Median	Std. deviation	Median	Std. deviation
-15°	-7.42°	5.39°	-1.66°	8.67°
+0°	-2.64°	5.61°	-0.79°	6.64°
+15°	+5.74°	5.91°	+1.22°	6.59°
+30°	+13.33°	5.14°	+4.92°	7.35°

Unlike the previous case, the differences in estimated pitch angles detected with different HAAR cascades are found significant for all pitch angles ($p<0.01$). For only pitch = +0° profile HAAR cascade works better ($p<0.01$) but for other pitch angles frontal cascade outperforms ($p<0.01$) the profile one. Variations around bias values are shown using bars in figure 9.

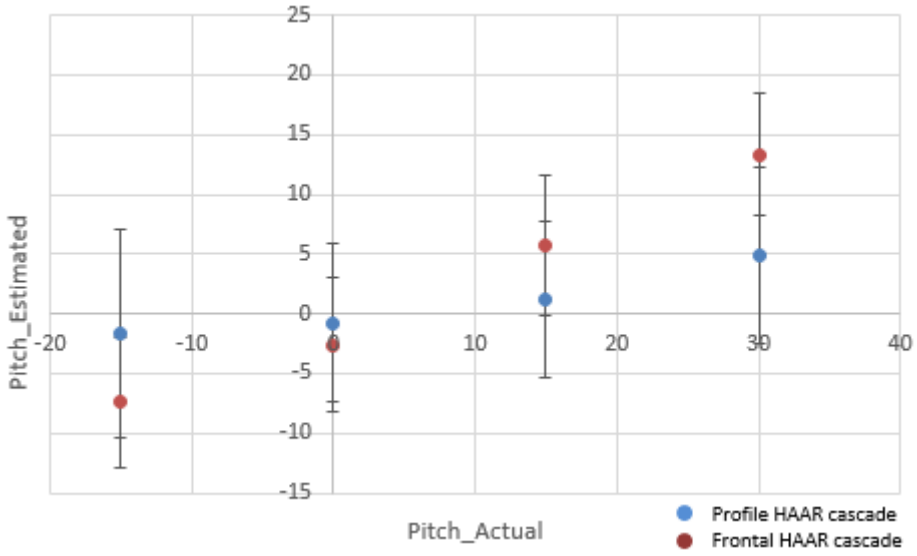


Figure 9. Actual Pitch vs Estimated Pitch for different HAAR cascades

Error measurement in calculating pitch and yaw angle

As the pre-processed image is fed into the neural network with the help of already trained .mat file, it calculates pitch and yaw angles and returns them accordingly. But the model looks still not accurate for higher angles possibly for reasons such as limited range of the face orientations were trained etc. It is decided to measure the median of signed errors of both the angles based on the test images outcomes to interpret how much deviation the model still causes from the expected values.

Table 3 below presents the calculated median of signed error (MSE) and the standard deviation accordingly for each and every pitch and yaw angle detected. It can be seen that the model does not perform well in calculating higher values of both yaw and pitch angles in terms of accuracy.

Table 3. MSE and Std. dev for all pitch and yaw angles calculated

Pitch	Pitch = -15°		Pitch = +0°		Pitch = +15°		Pitch = +30°		
Pitch MSE	9.222°		-1.976°		-10.875°		-18.409°		
Std. dev	6.8235°		6.1387°		6.6749°		6.9011°		
Yaw	Yaw= -60°	Yaw= -45°	Yaw= -30°	Yaw= -15°	Yaw= +0°	Yaw= +15°	Yaw= +30°	Yaw= +45°	Yaw= +60°
Yaw MSE	26.35°	13.84°	4.97°	2.83°	-1.76°	-0.95°	-3.65°	-15.59°	-26.70°
Std. dev	5.76°	8.24°	8.01°	6.32°	4.57°	6.14°	6.04°	5.68°	4.68°

Figure 10 plots the median of the estimated yaw angles calculated by the model for each and every actual yaw angle. A linear regression line is attempted to fit the graph. The slope of the fitted line is found as +32.73° which is bit lesser than +45°. Looking at the table 3, it is evident that from +0° to up to ±30°, the error is comparatively lesser than for other angles which can also be seen from the plot that those points are very close to the expected yaw angles causing their slopes very close to +45° as well. So the effect of other angles such as ±45° and ±60° might have decreased down the slope value overall. The standard deviation bars give an indication of the variation around the bias values. The standard deviation around +0° and ±60° yaw angles are found lesser than others.

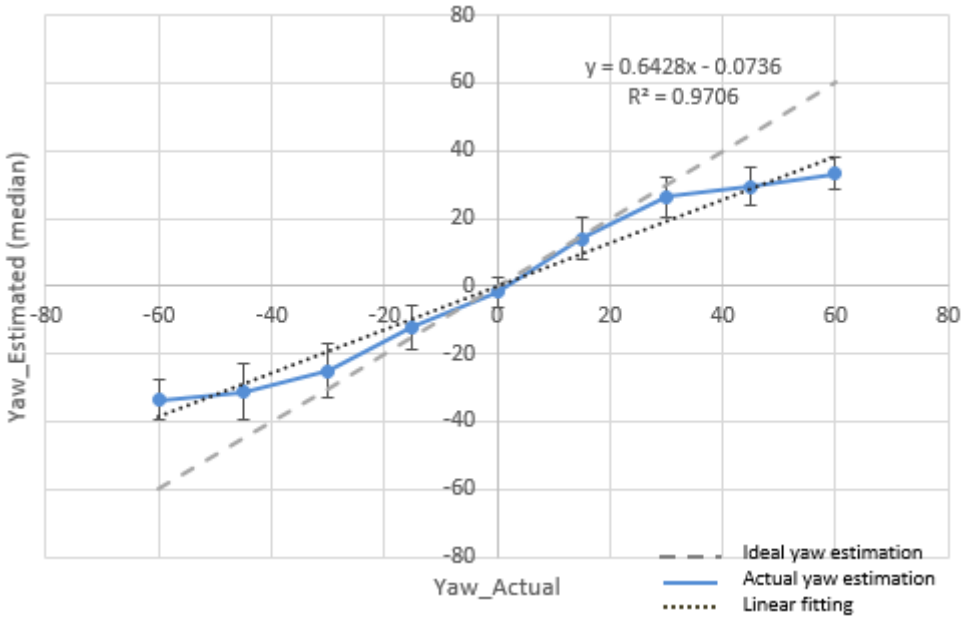


Figure 10. Actual Yaw vs Estimated Yaw

Similarly, figure 11 plots the median of the estimated pitch angles calculated by the model for each and every actual pitch angle. A linear regression line is also attempted to fit the graph. The slope of the fitted line is found as only +21.22° which is far less than +45°. Looking at the table 3, it is evident that only for +0°, the error is almost close to zero which can also be seen

from the plot. But for other higher angles, the model performs not that convincing causing higher deviations from the expected values. Besides, the standard deviation bars give an indication of the variation around the bias values here too. The variations are found somewhat comparable among them.

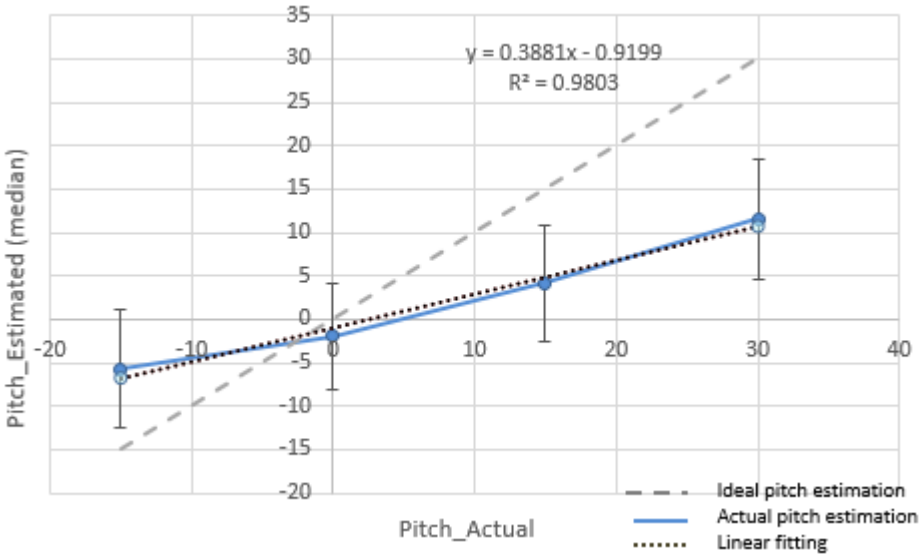


Figure 11. Actual Pitch vs Estimated Pitch

To get a much clear picture of the performance of the model, errors are too plotted in graphs for yaw and pitch angles. In both the plots, the standard deviation bars give an indication of the variation around the bias values. Figure 12 shows the plot with median values of the signed error drawn with for each and every yaw angle. The MSE for yaw = +15° is found the lowest and for yaw = -15° and +0°, the error is found also very low. So for those angles the model performs the best. As the yaw angle increases, MSE also gets increased. After ±30°, the increment looks almost like monotonically. The standard deviation for yaw = -45° gets the highest value (=8.24°) among all yaw angles. It is further investigated whether the MSE for yaw angles calculated are significantly symmetrical as face detection is done in a symmetrical fashion which is already described. None of the symmetrical yaw angles (±15°, ±30°, ±45°, ±60°) are found significantly different (*p>0.1*) from each other in terms of performance. As expected, the main effect of yaw angles on the yaw MSE is found as significant (*F(8)=430.61, p<0.01*). But the main effect of pitch angles on the yaw MSE is not found as significant (*F(3)=2.10, p=0.091*). So the pitch angles do not have any effect on the variations of estimated yaw angles. On the other hand, the interaction effect of pitch and yaw angles on yaw MSE is found as significant (*F(31)=117.27, p<0.01*) which signifies that the effect of pitch angles has an impact on the effect of yaw angles on yaw MSE. The effect size of the interaction is recorded as high (partial $\eta^2 = 0.81$). The slope of the fitted linear regression line is found as -19.66°. As it can be seen from the plot that the value of slope increases more when yaw > ±30°, the slope found till ±30° is -7.98° which is much closer to +0° than overall slope of the fitted linear line. And for yaw > +30°, the slope found is -36.53° and yaw < -30°, the slope found is -39.83°, which undoubtedly increase the overall slope.

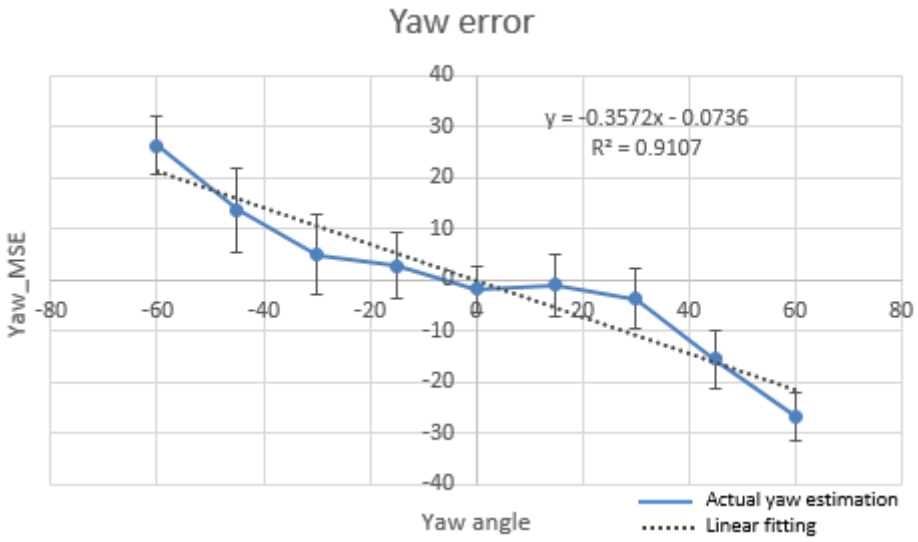


Figure 12. MSE values for different yaw angles

Next, figure 13 shows the plot with median values of the signed error drawn with for each and every pitch angle. It can be seen from the figure that for pitch = ±15° and +30°, the MSE (=+9.22°, -10.88° and -18.41° respectively) values are higher than to be expected. The main effect of pitch angles on the pitch MSE is found as significant (*F(3)=696.03, p<0.01*). Unlike the previous case, the main effect of yaw angles on the pitch MSE is found as significant (*F(8)=3.13, p<0.01*). So the yaw angles have an effect on the variations of estimated pitch angles. On the other hand, the interaction effect of pitch and yaw angles on pitch MSE is also found as significant (*F(31)=87.10, p<0.01*) which signifies that the effect of yaw angles has an impact on the effect of pitch angles on pitch MSE. The effect size of the interaction is recorded as high (partial $\eta^2 = 0.76$). The slope of the fitted linear regression line is found as -31.46°.

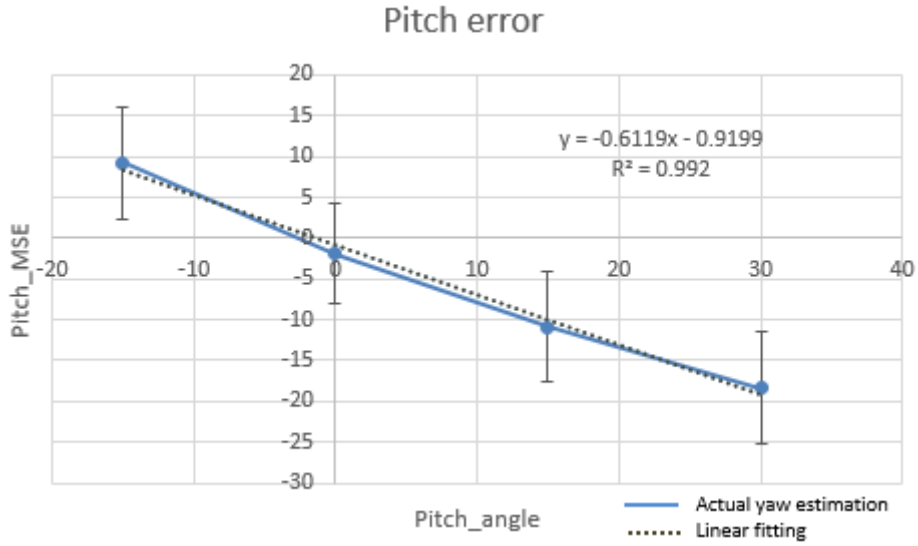


Figure 13. MSE values for different pitch angles

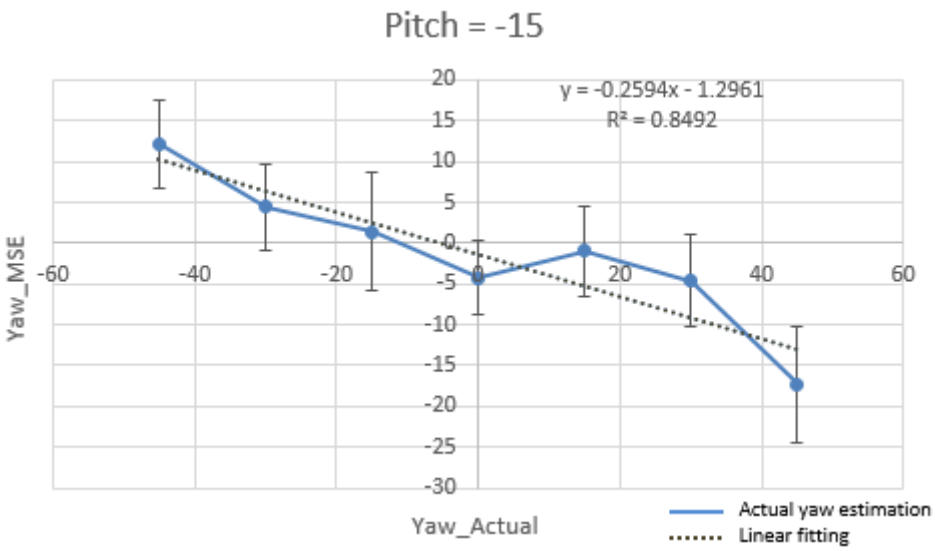
In the next section, it is investigated and discussed about whether the yaw estimations show any significant differences over different yaw angles for each and every pitch angle tested. They are also analysed through figures plotting for each pitch angle.

Table 4 depicts yaw MSE values and standard deviations for each of the pitch and yaw angles below. As it is already been discussed that the detection model does not recognize face orientation at yaw = ±60° for pitch = -15°and +30°, they are mentioned as *NULL* or '-' in the table accordingly.

Table 4. Pitch vs Yaw angles (in terms of MSE & SD)

	Pitch = -15°		Pitch = +0°		Pitch = +15°		Pitch = +30°	
	MSE	Std. dev	MSE	Std. dev	MSE	Std. dev	MSE	Std. dev
Yaw=-60°	-	-	24.85	4.30	27.73	6.57	-	-
Yaw=-45°	12.21	5.43	12.97	13.54	14.41	3.67	14.19	6.11
Yaw=-30°	4.44	5.27	6.61	6.31	3.05	6.69	6.07	11.55
Yaw=-15°	1.40	7.18	0.91	5.28	1.75	5.64	5.84	5.63
Yaw=+0°	-4.30	4.55	-3.19	3.49	-1.02	4.85	1.81	4.26
Yaw=+15°	-0.93	5.49	-3.65	6.47	-1.92	6.26	2.36	5.11
Yaw=+30°	-4.58	5.64	-4.58	5.64	-3.12	6.34	-1.83	5.19
Yaw=+45°	-17.32	7.19	-15.03	5.07	-15.45	5.14	-15.64	5.42
Yaw=+60°	-	-	-27.00	3.96	-26.38	5.38	-	-

Figure 14 shows all the plots of yaw MSE vs yaw angles for each pitch angle. The main effect of yaw angles on yaw MSE is still found as significant for all pitch angles. However, the overall slope of the fitted linear regression line is found as the lowest (≈ -14.55°) for pitch = -15° and +30°. For pitch = +0°, the slope from yaw = +0° up to +30° is found as -2.65° which is very close to zero. For pitch = +15°, the slope up to yaw = ±30° is found as -6.09° which is too close to zero. None of the symmetrical yaw angles (±15°, ±30°, ±45°, ±60°) are found significantly different (*p>0.1*) from each other in terms of performance for each and every pitch angle.



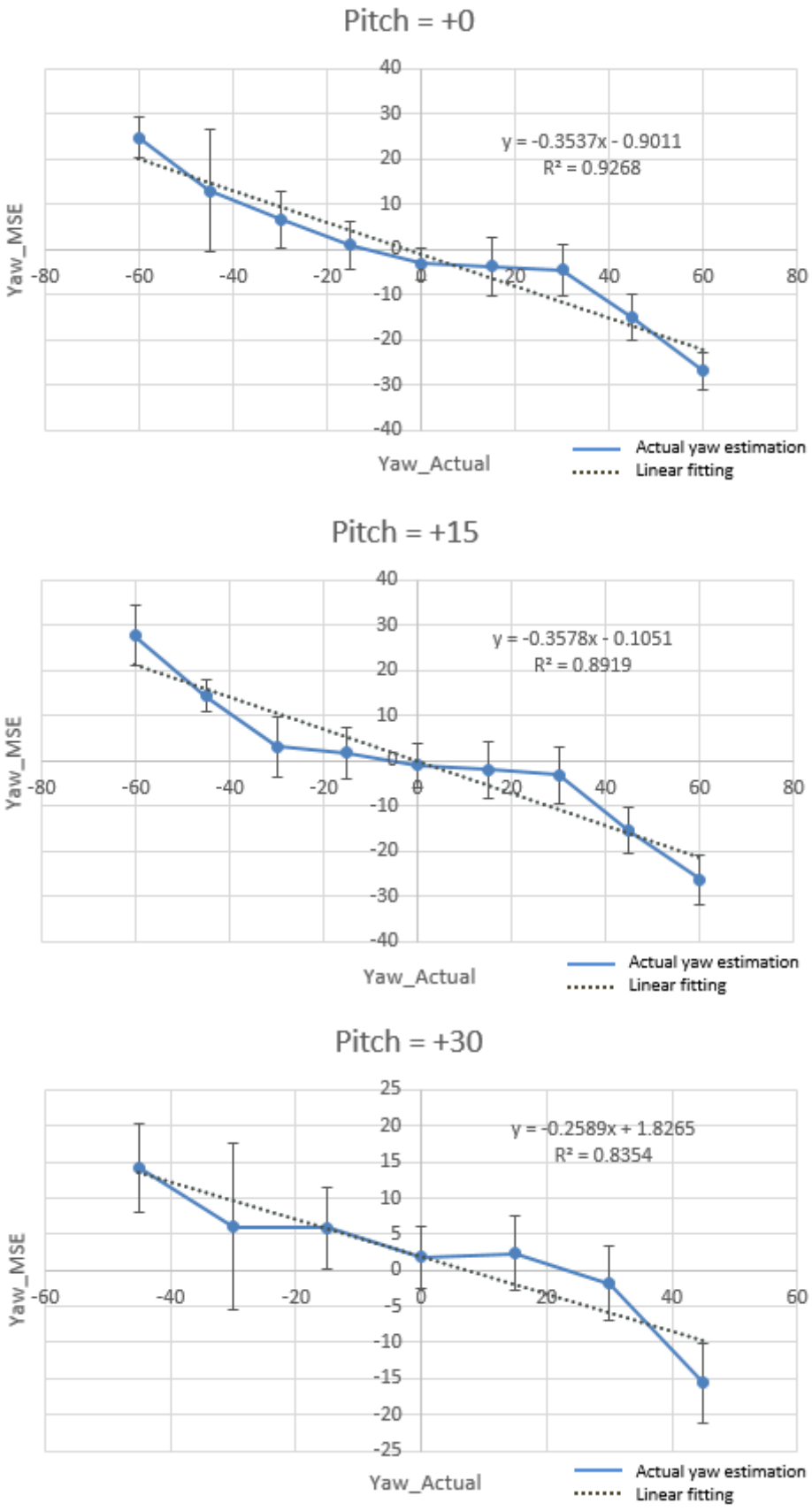


Figure 14. Yaw_MSE vs Yaw_Actual for different pitch angles

Now, it is also investigated whether the yaw estimations differ over pitch angles for each yaw angle. There is a significant difference (*F(3)=5.74, p<0.01*) found in the yaw MSE over different pitch angles for yaw = -15°. The effect size has a large size (partial $\eta^2 = 0.14$). So it does have an impact on the calculation of this particular yaw angle. A further investigation reveals that the difference in calculated yaw values for between pitch = +30° and rest of the pitch angles are found significant (*p<0.01*) in which pitch = +30° causes more errors than other pitch angles.

There is also a significant difference (*F(3)=5.71, p<0.01*) found in the yaw MSE over different pitch angles for yaw = +0°. It is also found that the pitch angles have almost a large effect size (partial $\eta^2 = 0.13$). So it does have an impact on the calculation of this particular yaw angle too. A further investigation reveals that the difference in calculated yaw values for between pitch = -15° and +30° is found significant (*t(3)=3.83, p<0.01*) in which pitch = +30° causes more errors than for pitch = -15°. Similarly, the difference in calculated yaw values for between pitch = +0° and +30° is also found significant (*t(3)=3.19, p=0.01)* in which pitch = +30° also causes more errors than for pitch = +0°.

For yaw = +15°, a significant difference (*F(3)=4.19, p<0.01*) is also found in the yaw MSE over different pitch angles. It is also found that the pitch angles have a moderate to large effect size (partial $\eta^2 = 0.11$). So it does have an impact on the calculation of this particular yaw angle too. A further investigation reveals that the difference in calculated yaw values for between pitch = +0° and +30° is found significant (*t(3)=3.37, p<0.01*) in which pitch = +30° causes more errors than for pitch = +0°. Similarly, the difference in calculated yaw values for between pitch = +15° and +30° is also found significant (*t(3)=2.71, p=0.04)* in which pitch = +30° also causes more errors than for pitch = +15°.

Discussion

This research does an improvement over detecting faces with greater yaw angles indeed compared to Huijben's (2015) research. Unfortunately there's no improvement in pitch angles found at all as the HAAR cascade that both of them use usually takes care of yaw angles mainly. In this research, a profile face HAAR cascade is used additionally to the frontal face HAAR cascade to attempt to increase the yaw angles of a face overall supporting hypothesis 1 partially that states the improvement increases detection of both pitch and yaw angles but no steps are actually taken to increase the pitch angles explicitly.

But it can also be seen that using HOG detector does improve the detection of higher pitch angles [-30°, +30°] but at the same time it performs (yaw = [-45°, +45°]) not better than HAAR cascade with flipping (yaw = [-90°, +75°]) for yaw angles. Besides, HOG detector takes more processing time than Viola-Jones detector and maintains always a high criterion than the latter one as it detects lesser amount of faces in FacePointing04 database with a detection rate of 80%. Because of using high criterion, HOG detector also gives lower false positives than Viola-Jones detector which is already been found in the earlier sections. As it is already discussed that missing a face seems to be more expensive than detecting an object falsely as face, Viola-Jones detector (HAAR cascade with flipping) is considered to be used over HOG

detector. For extended Yale B database too, HAAR detection with flipping detects more variation in azimuth and elevation angles of the varying illumination on faces than HOG detector along with less processing time but creating more false positives. Hence, hypothesis 2 which claims that HOG detector performs better that Viola-Jones detector, is proved incorrect at least in this context.

While HOG detector is found detecting face orientation symmetrically over both pitch and yaw angles by default in FacePointing04 database, HAAR detection does not find as such neither on pitch nor on yaw angles. HAAR detection is found as biased towards positive yaw angles (figure 1). In order to get a symmetrical head orientation over yaw angles, a flip of images horizontally is performed as human face is symmetrical horizontally only. This step undoubtedly increases the range of yaw angles which can be supported using figure 2. Hypothesis 3 which claims that asymmetry in detection can be resolved using flipping horizontally seems also get proved. While analysing errors, it is also found that the mirrored yaw angles (±15°, ±30°, ±45°, ±60°) perform similarly with each other as the MSEs are not found as significantly different for each of the pairs. But the standard deviations for each of pairs found somewhat varying.

An additional profile face HAAR cascade is made use of in detecting faces with higher head orientations as the frontal face HAAR cascade was expected either unable to detect higher yaw and pitch angles or to put the bounding boxes incorrectly on the detected faces (not covering the entire face) and thus impacting on the measurement of yaw and pitch angles accordingly. It has already been seen that some of yaw angles within the already specified yaw range (-60°, +60°) are failed to get detected by frontal HAAR cascade but detected by profile HAAR cascade and all the pitch angles within the already specified pitch range [-15°, +30°] are detected by both the cascades. Higher positive yaw angles (≥+30°) are found to be measured by profile HAAR cascade better than frontal HAAR cascade on average. For negative yaw angles, profile HAAR cascade performs almost similar to the frontal cascade. There is also a difference found in calculating pitch angles over using two different cascades. Expect pitch = +0°, frontal face HAAR cascade outperforms in estimating pitch angles for all other pitch values. Hence, hypothesis 4 is found as partially correct that profile HAAR cascade does not always perform better than frontal cascade for higher yaw angles at least not for negative higher yaw angles.

Figure 10 says that the head pose model performed really well up to ±30° causing the slope almost incident on the ideal graph which is the range of yaw angles that was trained in the neural network from the work of van der Pol, Cuijpers & Juola (2011). Other higher yaw angles produce significant errors. But we see that all pitch angles except +0° deviate a lot from the ideal value which is not expected which partially proves hypothesis 5 as correct. However, the main effect of yaw angles are found having significant impact on estimating pitch angles but not the vice versa. Their interaction effects are found significant in estimating both yaw and pitch angles. Finally, the estimation of yaw angles -15°, +0° and +15° in particular are found significantly different over different pitch angles respectively.

Nevertheless, it is also found that the detector creates varying size of bounding boxes based on the result of the cascade used. Some bounding boxes cover all the prime criteria (eyes,

nose, mouth etc.) to decide on faces but some contain only a few decreasing the area of the box. This is to be verified further on why the resultant faces output from the cascades give varying sizes of bounding box around faces. This is an important improvement to be considered as the neural network works best for the optimal cut of faces from the image that is decided by the bounding box.

Talking about the future work, it would be a good idea to consider to re-train the neural network with improved range of database images that can be detected by the improved face detector. As it is already proved that yaw angles show a significant impact on estimating pitch angles for this model which may indicate that the previous training did not include good amount of faces with higher pitch angles as the range of yaw angles detected was lower earlier. The problem with large pitch variation could thus be resolved if the model is re-trained with faces of higher yaw-pitch combination.

References

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 1,* 886-893. doi:10.1109/CVPR.2005.177

Georghiades, A. S., Belhumeur, P. N., & Kriegman, D. J. (2001). From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6)*, 643–660. doi:10.1109/34.927464

Gourier, N., Hall, D., & Crowley, J. L. (2004). Estimating face orientation from robust detection of salient facial structures. *In FG Net Workshop on Visual Observation of Deictic Gestures* (pp. 1-9). Cambridge, UK: FGnet (IST–2000–26434)

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA, 79(8)*, 2554–2558.

Huijben, I. (2015). Effects of Network Structures and Training Databases on the Robustness of Head Pose Estimation Networks.

Johnson, D. O., & Cuijpers, R. H. (2013). Predicting Gaze Direction from Head Pose Yaw and Pitch.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin, 100(1)*, 78-100. doi:http://dx.doi.org/10.1037/0033-2909.100.1.78

Knapp, M., Hall, J., & Horgan, T. (2013). Nonverbal communication in human interaction. *Cengage Learning*.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: an application to face detection. *Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference on*, 130-136. doi:10.1109/CVPR.1997.609310

Rowley, H.A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(1)*, 23-38. doi:10.1109/34.655647

Van der Pol, D., Cuijpers, R. H., & Juola, J. F. (2011). Head Pose Estimation for a Domestic Robot. *Proceedings of the 6th international conference on Human-robot interaction*, 277-278. doi:10.1145/1957656.1957769

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, *1*, 1-511.

Yow, K.C., & Cipolla, R. (1996). Detection of human faces under scale, orientation and viewpoint variations. *Automatic Face and Gesture Recognition, 1996, Proceedings of the Second International Conference on*, 295-300. doi:10.1109/AFGR.1996.557280

Zhu, X., & Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2879-2886. doi: 10.1109/CVPR.2012.6248014

Appendix

Source code and required files: https://github.com/Rctue/headpose/