Proceedings of the Canadian Society for Mechanical Engineering International Congress 2022
CSME Congress 2022
June 5-8, 2022, Edmonton, AB, Canada

Head Pose Estimation and Tracking for Moving Target Active Noise Control Using Convolutional Neural Network

Vishaal Venkatesh[1*], Himanshu Sharma[2], Wintta Ghebreiyesus[3], Fengfeng(Jeff) Xi[4], Reza Faieghi[5]
[1,2,3,4,5]Department of Aerospace Engineering, Ryerson University, Toronto, ON, Canada
*e-mail: vvenkatesh@ryerson.ca

Abstract—This paper introduces a novel approach towards an implementation of head pose estimation techniques for the application of noise control. An existing multi-stage process, which uses convolutional neural network (CNNs), for head pose estimation is modified, implemented, and validated for moving target active noise control. With evaluation and development of application focused methodologies, this paper highlights the modifications necessary alongside the challenges involving real-time tracking. Utilizing such a method to improve noise comfort within aircraft cabins through an integrated real-time tracking system allows for the capability to create a zone of quiet bubble around a passenger's head during their flight time. Through various stand-alone and integrated system tests, the head tracking algorithm proposed shows desirable accuracies, exhibited through low total average % errors. Additionally, the integrated system also shows positive results with low tracking errors alongside effective real-time capabilities. The overall implementation provides a convincing solution for head-tracking system to be integrated with active noise control for moving targets within aircraft cabins.

Keywords: computer vision; head pose estimation; convolutional neural network; active noise control; noise comfort

I. Introduction

Improving passenger comfort has been at the forefront of research for the aviation industry where noise control is one of the largest components to improving passenger comfort within business jets. The noise is primarily generated from the aerodynamic boundary layer flow noise surrounding the aircraft cabin [1]. A method called Moving Target Active Noise Control (ANC) [2, 3, 4, 5, 6, 7] is utilized, where the objective is to track the passengers' ear locations and maintains a desirable dB reduction. With secondary speakers, the primary noise generated from the aerodynamics around the fuselage of the aircraft can be neutralized.

Since the main objective is to improve the noise comfort within an aircraft cabin in real-time, the head tracking system is highly constrained by the weight-requirements. The head-tracking system must utilize a lightweight input sensor while providing the desired accuracy with minimal latency. With the rapid growth in computer vision systems through Convolutional Neural Networks (CNNs), using a camera and adopting a neural network method allows for the implementation of this head-tracking system.

A popular method in computer vision is *head pose estimation*, which is a method that extracts a human head's orientation from imagery primarily attained from cameras [8]. There are various intermediate steps performed internally before attaining the head's directional information some of which include, face detection, facial features' key point extraction, and solving the 2D to 3D rotation transformation which provides the desired Euler angles - yaw and pitch [9, 10].

To attain the pose estimations, the first stage is to detect the face and attain the facial feature key points, one such method that simultaneously performs face detection and facial landmark detection is a cascaded multi-task three-stage framework called MTCNN [11]. This method has proven to have high accuracies while being able to tackle challenges such as large pose variations and unfavourable lighting conditions. This detection model is then adapted to calculate the Euler angles using the facial landmarks detected, intrinsic camera parameters, rotational matrices, and perspective-n-point solver [12]. This paper focuses on adapting and implementing an existing vision-based face detection framework, to attain the head pose estimations primarily the yaw and pitch Euler angles to successfully track the passenger's ear location within an aircraft cabin. Along with the modifications made to the existing MTCNN framework, a novel integration of unique face detection is introduced, this identifier neglects any new face detected outside the initial face detected. This ensures the speaker motion system to track the position of the first passenger detected and reduces any disturbances or unwanted movements.

The head pose estimation method is then validated with a dual axis gimbal and lays the foundation for real-time moving target tracking noise control. The paper is separated into 5 sections, where background is discussed in section II, followed by methodology in section III, implementation in section IV, and lastly conclusion in section V.

II. BACKGROUND

A. Face Detection - Mutli-Task Cascaded Convolutional Neural Network (MTCNN)

The overall framework of MTCNN can be split into three stages namely, the proposal network (P-Net), the refine network (R-Net), and the output network (O-Net). The training process for these CNN detectors is based on three specific tasks, face classification, bounding box regression, and facial landmark localization.

Face Classification: The face classification is defined as a two-class classification problem, which pertains to classifying the image as a face or not a face. Cross-entropy loss is used for each sample x_i, as shown in (1).

$$L_i^{det}= -\left(\left(1-y_i^{det}\right)\left(1-\log\left(p_i\right)\right)-\left(y_i^{det}\log\left(p_i\right)\right)\right) \quad (1)$$

Where p_i represents the probability of the sample being a face, determined by the network, and y_i^{det} denotes the ground-truth label which is either 0 or 1.

Bounding Box Regression: The bounding boxes are based on four main labeled locations left, top, height and width, the regression problem utilizes the Euclidean loss for each sample x_i, as shown in (2).

$$L_i^{box}=\left\|\hat{y}_i^{box}-y_i^{box}\right\|_2^2 \quad (2)$$

Where $\hat{y}_i^{box}$ is the regression value attained from the network and y_i^{box} represents the labeled ground-truth coordinates. All truth-values and predicted values belong to either of the four coordinates mentioned above.

Facial Landmark Localization: The facial landmark detection takes a similar approach to that of the bounding boxes as seen earlier, Euclidean loss function is shown in (3).

$$L_i^{box}=\left\|\hat{y}_i^{landmark}-y_i^{landmark}\right\|_2^2 \quad (3)$$

Where $\hat{y}_i^{landmark}$ is the predicted coordinates attained from the network and $y_i^{landmark}$ is the true labeled coordinates.

Since there are three different networks involved to perform multiple tasks within each CNN, an overall learning target takes each learning objective and balances it within each network architecture. The overall learning target is shown in (4).

$$\min \sum_{i=1}^{N} \sum_{j\in\{det,box,landmark\}} \alpha_j \beta_i^j L_i^j \quad (4)$$

Where N is the total number of training samples, α_j represents the importance of task for that particular network which is categorized as the following: $\alpha_{det} = 1$, $\alpha_{box} = 0.5$, and $\alpha_{landmark} = 0.5$ in the P-Net and R-Net, while $\alpha_{det} = 1$, $\alpha_{box} = 0.5$, and $\alpha_{landmark} = 1$ in the O-Net. β_i^j is the sample type indicator which can be interpreted as $\beta_i^j \in \{0,1\}$. Lastly, L_i^j is the loss functions shown in equations (1)-(3). Figure 1 outlines the structure for the P-Net, R-Net, and O-Net [11].

B. Perspective-n-Point Problem

One of the problems that have been predominant in the computer vision research field is the motion estimation of an object when the camera is fixed at a certain position, this motion estimation has been termed to a Perspective-n-Point (PnP) problem as seen in figure 2. Where the objective is to estimate the position and rotation of a calibrated camera based on an identifiable 3D-2D point mapping between a 3D model and its corresponding image projections. The PnP problem has proven to be able to achieve accurate results while maintaining robustness with the estimations, there are various components within this problem definition which would require the analysis of an appropriate camera model.

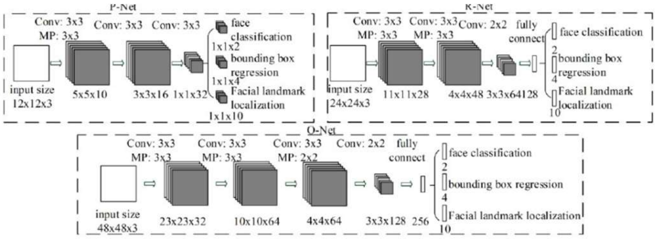

Figure 1: MTCNN Overall Architecture [11]

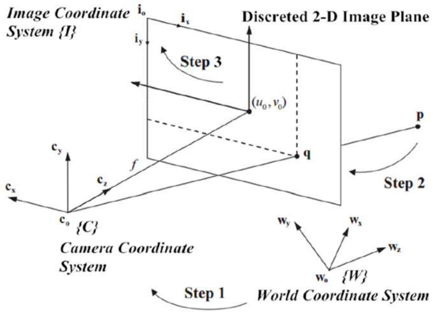

Figure 2: General Procedure of Pinhole Camera Model [13]

Based on the research performed on various camera models, the pinhole model is the most popular which defines an explainable projection of a point in the 3-D world coordinate system to the 2-D image plane. Consider point *p* in the 3-D world coordinate system defined as: $\{W\colon w_o, w_x, w_y, w_z\}$, and the relative position of *p* in this system is defined to be $X_w=(x_w,y_w,z_w,1)^T$. The pinhole camera model for a particular point *p* can be defined as shown in (5).

$$m^{'}=DK_0MX_w \quad (5)$$

Where,

$$D=\begin{bmatrix}\frac{1}{dx} & -\frac{cot\theta}{dx} & 0\\ 0 & \frac{1}{sin\theta dy} & 0\\ 0 & 0 & 1\end{bmatrix}, K_0=\begin{bmatrix}f & 0 & x_0 & 0\\ 0 & f & y_0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix} \quad (6)$$

And

$$M=\begin{bmatrix}\mathbf{R} & \mathbf{t}\\ 0 & 1\end{bmatrix} \quad (7)$$

The general outlined procedure of the pinhole model, where the first step utilizes matrix M defined in (7) to transform the coordinates of *p* from its world coordinates into the camera coordinates, where **R** is a 3 x 3 matrix that corresponds to rotation of the camera and **t** is a 3 x 1 translation vector. Then, K_0 projects point *p* onto the image plane which utilizes the camera focal length f and the optic center of the plan (x_0, y_0). Finally, matrix D which contains the individual pixel size

(dx, dy) in a CCD/CMOS image sensor with an intersection angle of θ, discretizes 2-D coordinates in the Image Coordinate System $\{I\}$. Based on these steps, the multiplication of matrices D and K_0 contains the intrinsic properties while matrix M contains the extrinsic properties. With a calibrated camera, the extrinsic parameters can be estimated using a set of 3-D points in the world coordinates and their comparable 2-D projections in the image coordinates [13].

III. METHODOLOGY

To achieve a real-time head tracking mechanism using vision reference, accurate estimations of the head rotation angle, and position are required. Additionally, it is necessary to avoid disturbances which are inherent within an aircraft cabin environment. Two methodologies are established individually, *Euler Angle Estimation* and *Unique Face Identification*, which can be integrated into a single head pose estimation. The individual systems' architecture and their working principles are discussed below.

A. Head Pose Estimation

The head tracking system is primarily based on head pose estimation, which is an active computer vision research field that utilizes cameras to attain the angular rotation and position of the head. This work leverages this concept to estimate the angular rotation of a passenger's head to initiate a real-time noise control strategy. A facial landmark-based approach has been utilized, where the first step is to detect the face and identify the critical facial landmarks, which are simultaneously performed by the utilized MTCNN method [11] with a high accuracy. Once this is complete, the following adaptations are applied to attain the head's angular rotations with minimized disturbances and accurate estimations; unique face identification, 2D to 3D rotation, and Euler angle estimations. Figure 3 outlines the overall framework of this head tracking system.

1) Euler Angle Estimation

As explained in section II, a predominant issue with head pose estimation is the PnP problem, with the MTCNN method the 2D coordinates (x, y) of the facial landmarks can be attained, and the 3D coordinates of those same points can be selected arbitrarily. For the application the 3D points selected are shown in table 1, they correlate to the five facial landmarks detected by the MTCNN model. The location of a point *P* in (X, Y, Z) camera coordinates can be attained based on these 3D locations of a point *P* in (U, V, W) world coordinates, with the rotation matrix **R** and translation vector **t** of the world coordinates with respect to the camera coordinates. This is defined in (8) where the rotation matrix and translation vector are unknown.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \tag{8}$$

The next relationship to consider is the representation of point *P* in the image coordinates with the intrinsic camera parameters which is shown in (9) below. Where u and v are the 2D coordinates of point *P*, s is the scaling factor, which is ignored for this use case, f_x and f_y are the focal lengths in the x and y directions and lastly, (c_x, c_y) is the optical center. By combining the intrinsic and extrinsic parameters in (8) and (9), the overall non-linear equation to solve is shown in (10).

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{9}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \tag{10}$$

To solve this equation, a method known as Direct Linear Transform (DLT) [14] which approximately estimates the rotation matrix **R** and translation vector **t**. Since the DLT method lacks the desired accuracy as the reprojection error requires iterative minimization, the Levenberg-Marquardt optimization method [15] is applied after the DLT. This optimization method iteratively solves this problem by perturbing the **R** and **t** matrix and vectors, respectively until the reprojection error decreases, allowing for accurate estimations of the rotation matrix and translation vector. Once the accurate **R** and **t** are attained, the 2D locations of the 3D facial points can be predicted on the image through a projection matrix that projects the 3D points onto the 2D image. The pitch (θ), yaw (ψ), and roll (ϕ) Euler angles are calculated by decomposing the projection matrix into the rotation matrices $\mathbf{R_x}$, $\mathbf{R_y}$, and $\mathbf{R_z}$ which are rotations about the x, y, and z axis respectively. The projection matrix **P** and the combined rotational Directional Cosine Matrix (DCM) is shown in (11) and (12), respectively.

$$P = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \tag{11}$$

$$\mathbf{R_{DCM}} = \mathbf{R_z}(\phi)\mathbf{R_y}(\theta)\mathbf{R_x}(\psi) \tag{12}$$

The Euler angles can then be calculated from the $\mathbf{R_{DCM}}$ matrix, for the purpose of this application only the pitch and yaw angles are considered and shown in (13) and (14), respectively.

$$\theta = \mathrm{atan2}\left(-\mathbf{R_{DCM}}_{31}, \sqrt{\left(\mathbf{R_{DCM}}_{11}\right)^2 + \left(\mathbf{R_{DCM}}_{21}\right)^2}\right) \tag{13}$$

$$\psi = \mathrm{atan2}\left(\mathbf{R_{DCM}}_{21}, \mathbf{R_{DCM}}_{11}\right) \tag{14}$$

Lastly, the passenger and the camera coordinates vary slightly as the images attained from the camera are inverted which impacts the Euler angles calculated, to tackle this, the yaw angle ψ is inverted to accommodate the inversion between the world and camera coordinates. The final yaw angle utilized is shown in (15), while no change is made to the pitch angle calculated.

$$\psi = -\mathrm{atan2}\left(\mathbf{R_{DCM}}_{21}, \mathbf{R_{DCM}}_{11}\right) \tag{15}$$

This method allows to accurately estimate the head pose at any given time and provides the ability to track the head with varying head poses. This establishes a framework to perform various experiments that tests the accuracies of the head pose predictions.

2) Unique Face Identification

Within an aircraft cabin setting, there are high possibilities for various personnel to interfere with the head tracking system,

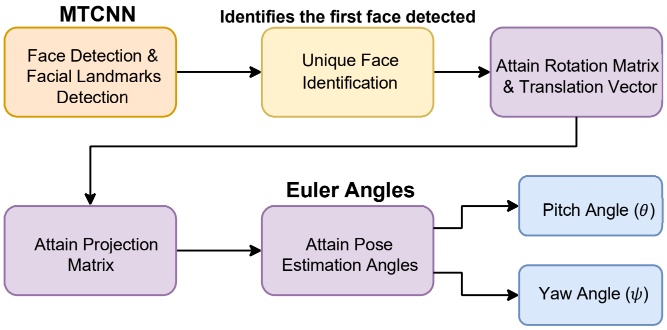
MTCNN
Identifies the first face detected
Face Detection & Facial Landmarks Detection
Unique Face Identification
Attain Rotation Matrix & Translation Vector
Euler Angles
Attain Projection Matrix
Attain Pose Estimation Angles
Pitch Angle (θ)
Yaw Angle (ψ)

Figure 3: Head Tracking System Architecture

Table 1: Arbitrary 3D Points Selected

Facial Landmarks	**Global Coordinates in (U, V, W)**
Left Eye	(-165, 170, -115)
Right Eye	(165, 170, -115)
Nose	(0, 0, 0)
Left Mouth Corner	(-150, -150, -125)
Right Mouth Corner	(150, -150, -125)

to tackle this problem a novel method called unique face identification has been developed which creates a sphere bounding box around the first face detected and ensures that any other faces that enter this sphere will be tagged as a "*New Face Detected*" and will prevent further angular calculations. Figure 4 showcases the working mechanism of the unique face identifier, where the yellow bounding box attained from the MTCNN model is used as a reference to then create a dynamic sphere bounding box with an established d_i and d_{i_max} which are the edge distances taken from the right side between the face bounding box, and the sphere bounding box.

These distance measures dynamically change as the head moves, to reduce noisy tracking data during real-time unique face identification a distance ratio $\mathrm{d}_i/\mathrm{d}_{i\_\max\,(initial)}$ is used as the metric to detect new faces. Where d_i dynamically changes with head rotations and $\mathrm{d}_{i\_\max\,(initial)}$ is the initial distance value established with the first detected face. Based on experimental analysis the valid head tracking distance ratio is $0.1 < \mathrm{d}_i/\mathrm{d}_{i\_\max\,(initial)} < 0.85$, anything outside this range would stop angular calculations and flag the system with a "*New Face Detected*" tag. This overall mechanism ensures angular changes from only one face is being calculated and processed for downstream control tasks, which reduces the likelihood of unnecessary data being transmitted.

IV. Implementation

Based on the methodology developed and explained in the previous section, the implementation is split into two components experimental setup and experimental results. The head tracking system is tested and evaluated as a stand-alone system initially and later integrated with a speaker motion system for real-time testing. The setup and results are separated in such a manner to understand the system's behaviour at a finer level. Since the focus of the development of the head tracking system is to orient the speakers to follow the passenger's ear, the dual-axis gimbal system and the head tracking system create a continuous loop in which the inputs from the head tracking mechanism are treated as the reference points for the gimbal system to adjust the position of the speakers accordingly. The motor control is determined by the encoders which aim to correct the motors' position to match the reference input angle from the head tracking system.

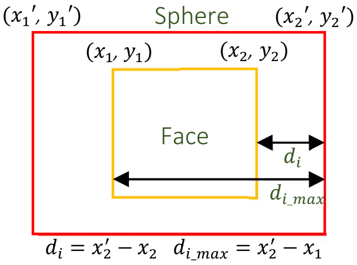

Figure 4: Unique Face Identification

A. *Experimental Setup*

1) *Head Tracking System*

As seen in figure 6, the experimental setup for the stand-alone system is done by placing a mannequin head above a turn table at 30 in. away from the source camera, the mannequin head is used for the tests as the MTCNN algorithm would have to recognize a human face in order to perform the detection. A turn table is used to accurately verify the angles predicted by the head tracking system, and a distance of 30 in. from the source camera is selected to simulate the realistic distance between the passenger and the camera within the aircraft cabin. Lastly, the source camera used for these experiments is a built-in laptop camera.

The stand-alone system test is comprised of collecting data from a series of 9 different yaw positions set at 2 pitch positions, accumulating to a total of 18 tests. The yaw positions ranged from -40° to 40° and the two fixed pitch positions are 10° and 20°, at each pitch position the turntable is set to the 9 different yaw positions where approximately 350 datapoints are collected at each position. Based on the data collected at each test, the average percent error is calculated for further analysis.

2) *Speaker Motion System*

The objective of the gimbal system is to orient the speaker in a way that the zone of quiet bubble is constantly maintained around the passenger's head even with position and angular changes of the head over time. The speakers' motion is mainly defined in the pitch and yaw motion, figure 5 also showcases the coordinate frames where the yaw motion corresponds to the speakers moving left and right with rotation about the y-axis, while the pitch motion is defined when the speakers move up and down with rotation about the x-axis. The angular information attained from the head tracking system is directly utilized for the motion of the speakers as they both share the same coordinate frames. The integrated system tests the dual-axis gimbal motion with the head tracking system using a passenger's realistic head movements, during testing the passenger moves the head dynamically in both pitch and yaw motion and the gimbal system replicates the motion of the head. As seen in figures 7 and 8, the participant moves their head in the pitch and yaw directions respectively. The test cases are classified into two scenarios: "*Head Left*" and "*Head Right*" depending on the location of the passenger's head from the neutral position. The results for these tests are primarily reported based on the motion along the yaw axis since this direction has a higher range of motion than that along the pitch axis.

B. *Experimental Results*

1) *Head Tracking System Results:*

Based on the 18 test cases defined earlier, the main metric used to measure the accuracy of the head tracking system is the average % error, as it would consider all ranges of the dataset and is described in (16).

$$\text{Avg. \%Error} = \frac{\Sigma_{i=1}^{n}\left|\frac{v_{A(i)}-v_E}{v_E}\right|.100\%}{n} \tag{16}$$

Where v_A is the actual observed value, v_E is the static expected values, and n is the total number of data points within each particular test case. Based on the average % error calculated at each test, a total average is taken for better generalization. As seen in table 2 the total averages for both yaw and pitch motions were maintained predominantly under 5%, while the total average % errors at the 10^o pitch angle is higher than those of the 20^o pitch angle. Figures 9 a) and b) showcase the average % errors at each test for both yaw and pitch motions at the 10^o and 20^o pitch angles, respectively. As it can be seen in both figures, the system is able to generalize the predictions in the yaw direction better than the pitch, where the errors are higher at the extreme yaw angles of $\pm 40^o$.

The model has an overall better performance at the 20^o pitch angle, when compared to the 10^o with very similar yaw errors but much higher pitch errors. Factors such as lighting and initial camera calibration contribute to the discrepancies in pitch performance between the two test categories. Lastly, the distribution of yaw % error at each test is also studied through boxplots, figures 10 a) and b) showcase the data distribution at each yaw test at the 10^o and 20^o pitch angles, respectively. It can be seen that the Interquartile Range (IQR) is predominantly within 5% for both tests, with an outlier in the 20^o pitch at -40^o . Despite the single outlier, the system has shown promising consistencies with its accuracy by maintaining low % errors at each test. Based on the defined test conditions, the vision system is able to accurately predict the head poses while maintaining minimal errors.

2) *Integrated System Results:*

The real-time capabilities of the head tracking system are tested with the integrated system test, the head tracking is expected to relay the data to the speaker motion system to effectively move the speakers. Figures 11 a) and b) shows the response of the speaker motion system with the real-time head tracking system, it can be seen that the head tracking system is able to provide head pose predictions consistently. The encoders are then able to follow the predicted head rotations very closely with some errors observed. The integrated test validates the performances of the head tracking and the speaker motion system with respect to achieving the desired accuracy and low latency, which meets the overall objective of real-time noise comfort within an aircraft cabin.

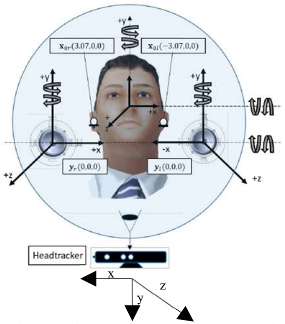

Figure 5: Speaker and Camera System Coordinate Frames

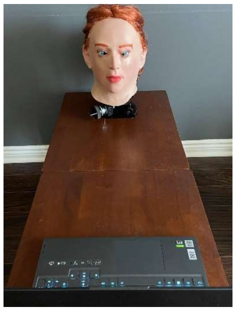

Figure 6: Standalone System Test Setup

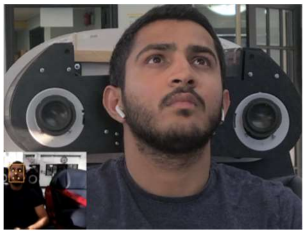

Figure 7: Integrated Pitch Test Setup

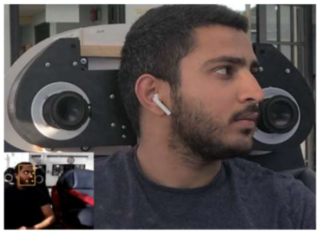

Figure 8: Integrated Yaw Test Setup

V. Conclusion

In this study, a modified head pose estimation approach has been examined and integrated with a real time tracking mechanism for a moving target within aircraft cabins. The stand-alone tests prove the head tracking system's accuracy at varied yaw and pitch angles where the dual axis prediction capability makes this a desirable approach for pose estimations of generic passenger head movements. The integrated system has provided a framework to accurately track the passenger's head, while maintaining high accuracies and balancing effective real-time responses. The real-time capability of the overall system, while minimizing individual errors is very promising for active noise

Table 2: Total Average % Error Results

Pitch Angle	Total Average % Error (Pitch)	Total Average % Error (Yaw)
10°	5.22	3.41
20°	2.71	3.23

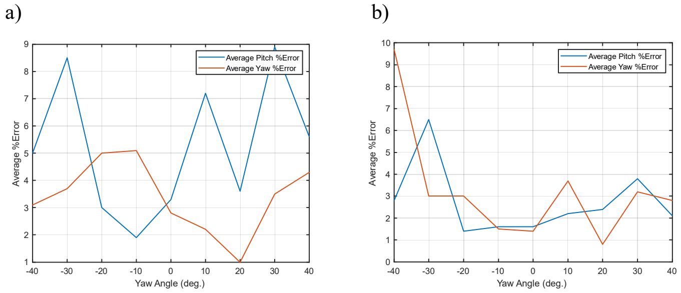

Figure 9: Yaw & Pitch Average % Error. a) At 10° Pitch, b) At 20°Pitch

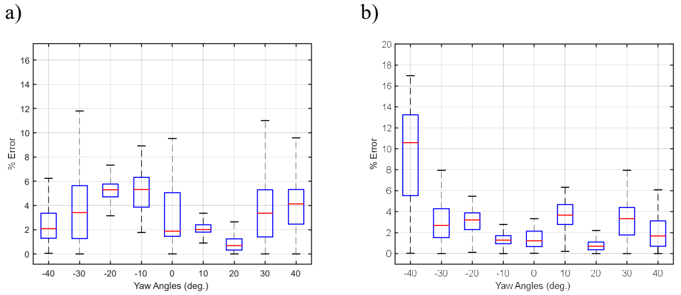

Figure 10: Yaw % Error Boxplot. a) At 10° Pitch, b) At 20°Pitch

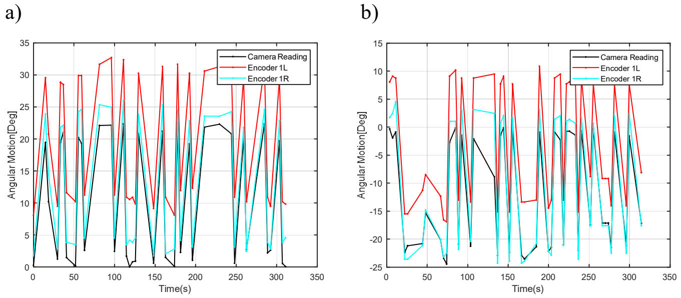

Figure 11: Real-time Integrated Yaw Test. a) Head Right, b) Head Left

control tasks that can improve the passenger's noise comfort experience within the aircraft cabin. This head tracking approach has the potential to be extended to other pose estimation tasks such as driver assistance, motion capturing, and gaze estimation to name a few.

ACKNOWLEDGMENT

The authors would like to thank Bombardier Aerospace and NSERC for their generous support sponsoring this project.

REFERENCES

[1] Grewal, A. (A.), F Nitzsche, Zimcik, D.G. (D. G.), and Leigh, B. (B.), "Active control of aircraft cabin noise using collocated structural actuators and sensors", Journal of Aircraft, vol. 35, no. 2, pp. 324–331, Mar. 1998.

[2] Y. J. Liao, A. Slade, H. Luo, en L. Chen, "The study of active noise control method for moving target in noisy space", in Applied Mechanics and Materials, vol 52–54, 2011, pp. 1592–1597.

[3] T. Xiao, X. Qiu, and B. Halkon, "Ultra-broadband local active noise control with remote acoustic sensing," *Scientific Reports*, vol. 10, no. 1, 2020.

[4] I. IKEDA, S. KIJIMOTO, K. MATSUDA, and Y. KOBA, "Active noise control with a moving evaluation point," Journal of System Design and Dynamics, vol. 2, no. 1, pp. 362–369, 2008.

[5] T. OKUYAMA, H. MATSUHISA, H. UTSUNO, and J. G. PARK, "Active noise control for a moving evaluation point using transfer function interpolation," *JSME International Journal Series C*, vol. 49, no. 3, pp. 865–872, 2006.

[6] S. M. Kuo, K. Kuo, en W. S. Gan, "Active noise control: Open problems and challenges", in The 2010 International Conference on Green Circuits and Systems, pp. 164–169, 2010.

[7] S. M. Kuo en D. R. Morgan, "Active noise control: a tutorial review", Proceedings of the IEEE, vol 87, no 6, pp. 943–973, 1999.

[8] E. Murphy-Chutorian en M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 31, no 4, pp. 607–626, 2009.

[9] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. "Joint 3d face reconstruction and dense alignment with position map regression network," In Proceedings of the European Conference on Computer Vision (ECCV), pp. 534–551, 2018.

[10] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. "Face alignment across large poses: A 3d solution," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.146–155, 2016.

[11] K. Zhang, Z. Zhang, Z. Li, en Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks", IEEE Signal Processing Letters, vol 23, no 10, pp.1499–1503, 2016.

[12] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPNP: An accurate o(n) solution to the PNP problem," International Journal of Computer Vision, vol. 81, no. 2, pp. 155–166, 2008.

[13] X. X. Lu, "A Review of Solutions for Perspective-n-Point Problem in Camera Pose Estimation", Journal of Physics: Conference Series, vol 1087, bl 052009, Sep 2018.

[14] Y. I. Abdel-Aziz, H. M. Karara, en M. Hauck, "Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry*", Photogrammetric Engineering & Remote Sensing, vol 81, no 2, pp. 103–107, 2015.

[15] B. M. Wilamowski en H. Yu, "Improved Computation for Levenberg–Marquardt Training", IEEE Transactions on Neural Networks, vol 21, no 6, pp. 930–937, 2010.