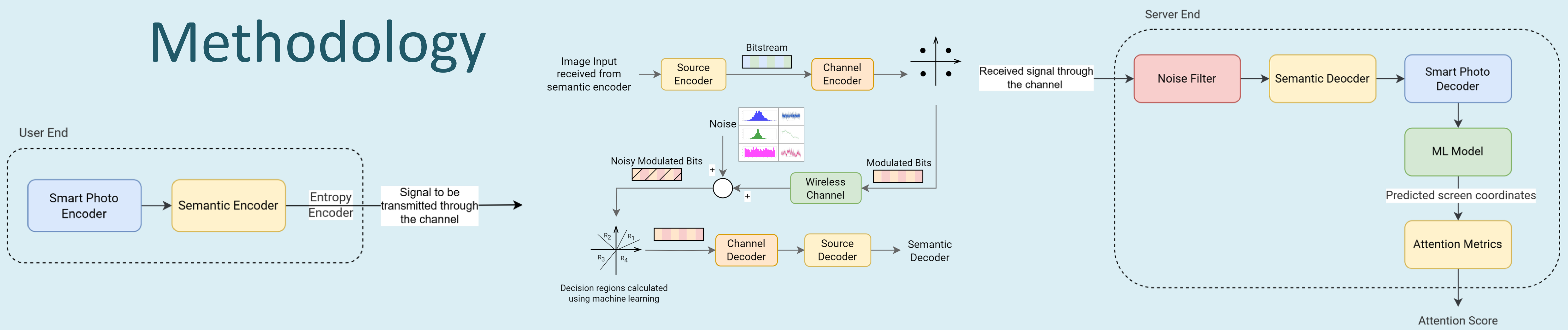
University of Peradeniya

Identification of Attention Level by Gaze Recognition using a Semantic Communication System

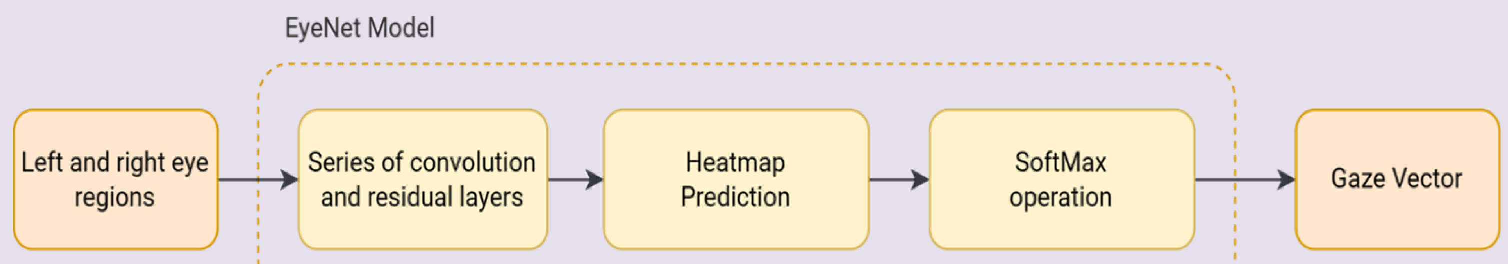
Nethmi S. Hewawiththi[1], M. Mahesha Viduranga[1], Vanodhya G. Warnasooriya[1]
Supervised by: Himal A.Suraweera[1], Maheshi B. Dissanayake[1]
[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya

Abstract- Artificial Intelligence plays a significant role inside classrooms with the rapid development in online education during the COVID-19 pandemic. Our work addresses online education challenges by using gaze estimation and head pose detection to improve attention tracking without any wearable devices. It also employs novel semantic communication methods to reduce bandwidth limitations, enhancing the online learning experience.

Methodology



Gaze Estimation



In the approach of estimating the gaze, the eye images are sent through a series of convolution and residual layers to predict the heatmap for the eye which represents the likelihood of each pixel being the gaze point. The probability of each pixel being the gaze point can be calculated from

$$P[i,j] = \frac{\exp(H[i,j])}{\sum \exp(H[k,l])} \forall (i,j); \text{ in the eye image}$$

Next, the softmax operation is applied to obtain a continuous probability density function over the image plane. Finally, the gaze vector is calculated by applying the softargmax operation on the normalized probability distribution. The normalized horizontal and vertical gaze coordinates are calculated as mentioned.

$$\text{Vertical Gaze} = \sum (j \times P[i,j]) \forall (i,j); \text{ in the eye image}$$

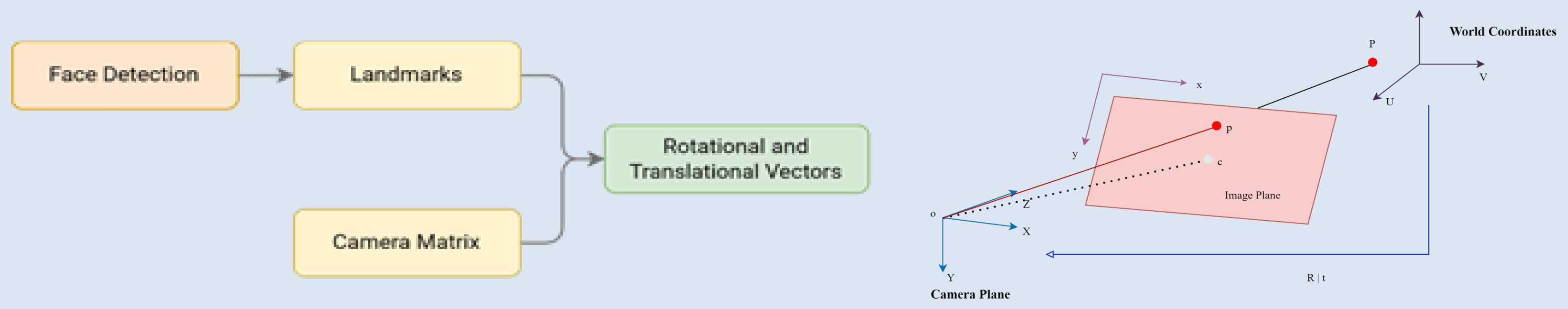
$$\text{Horizontal Gaze} = \sum (i \times P[i,j]) \forall (i,j); \text{ in the eye image}$$

$$\text{Normalized vertical gaze} = \frac{2 \times \text{Vertical gaze}}{\text{frame height}} - 1$$

$$\text{Normalized horizontal gaze} = \frac{2 \times \text{Horizontal gaze}}{\text{frame height}} - 1$$

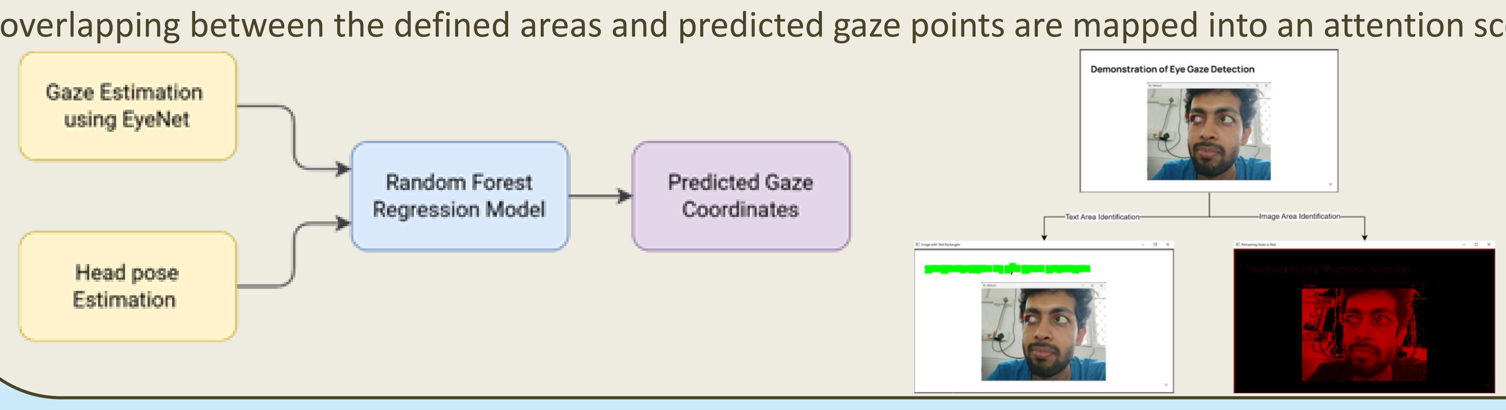
Head Pose Estimation

The gaze direction does not only depend on the eye regions which are extracted, also depends on the head pose orientation which determines the orientation of participants' heads in 3D space.



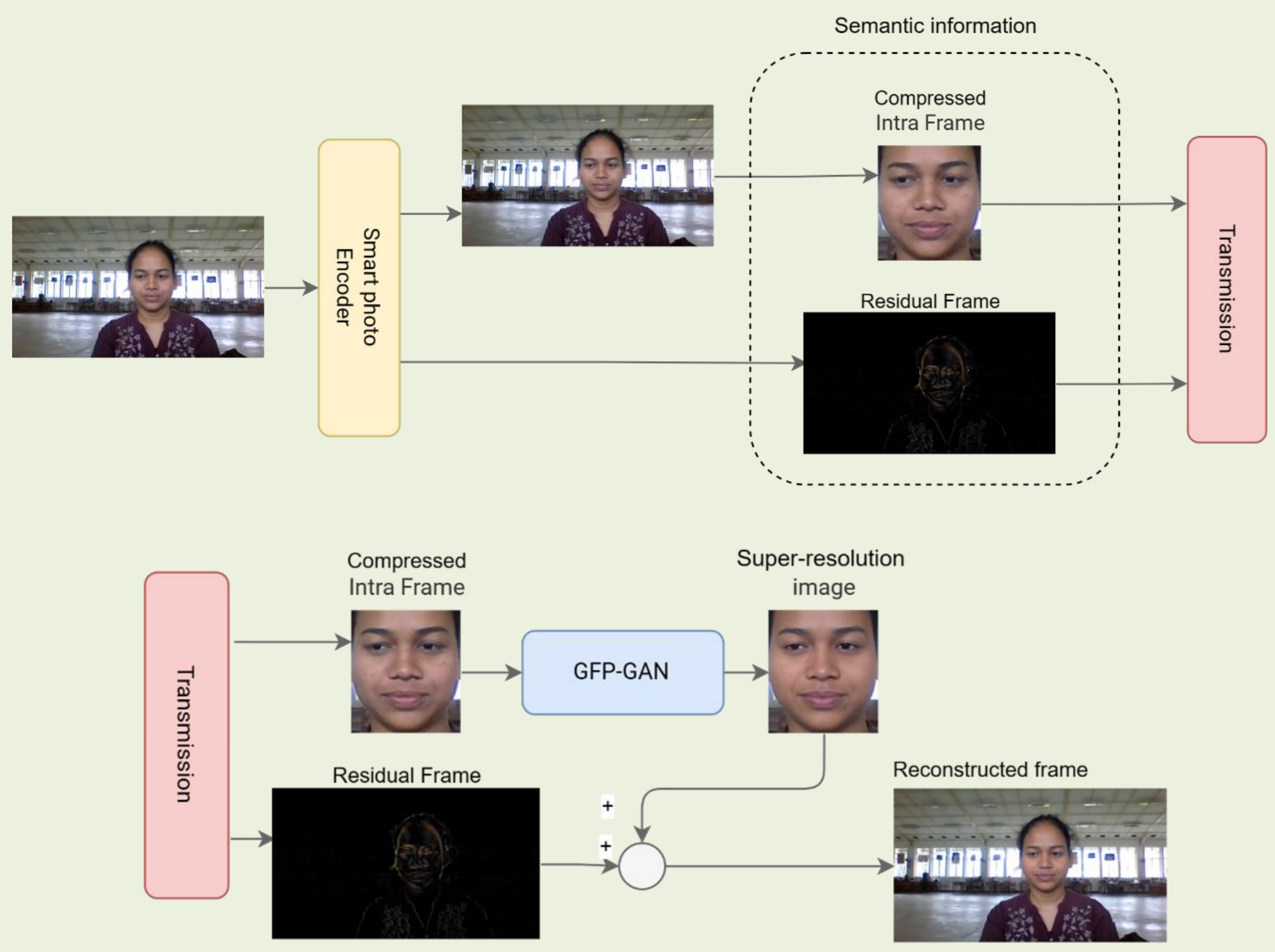
Attention Metric

Gaze estimation and head pose estimation are combined together to predict the point user is looking at the screen. Different areas of the presentation have been assigned a score based on their criticality and overlapping between the defined areas and predicted gaze points are mapped into an attention score.



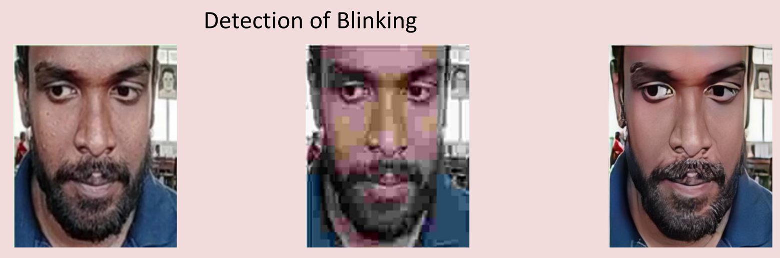
Semantic Framework

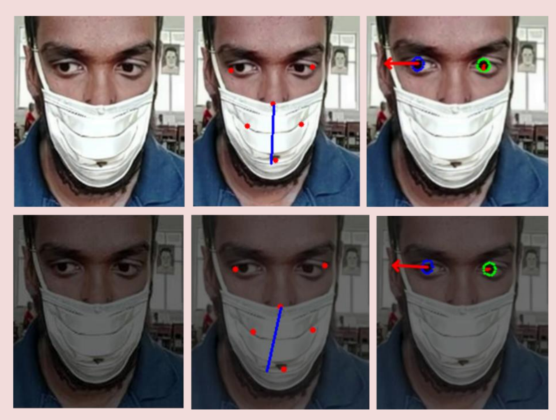
The video frames are extracted from the user and the Structural Similarity Index (SSIM), i.e., metric to evaluate similarity of two images, is calculated. If SSIM is greater than a given threshold, it is sent as an intra-frame and if not, it is sent as a residual frame. The frames selected for Intra Frame transmission are subjected to compression after identifying the face area using the Shape predictor 68 facial landmarks in dlib and OpenCV libraries.



Results

Detection of Blinking


Identified face area of size 17.5 kB. Compressed face Image of size 2 kB. Reconstructed face Image using GFPGAN.


Model results under subjects with masks and low light conditions.

Achieved compression ratio is 0.019. Hence, the new bandwidth required is approximately only 0.2 times of the original bandwidth leading to a substantial reduction of the required bandwidth for video transmission.

Contact details

Name : Himal A. Suraweera
Tel. No.: +94772166088
Email : himal@eng.pdn.ac.lk

Multidisciplinary AI Research Centre (MARC)
University Research Council
University of Peradeniya
Peradeniya, 20400, Sri Lanka