Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.024450
Article

Tech Science Press

Development of Algorithm for Person Re-Identification Using Extended Openface Method

S. Michael Dinesh[1,*] and A. R. Kavitha[2]

[1]Anna University, Chennai, 600025, India
[2]Department of Computer Science & Engineering, SRM Institute of Science & Technology, Chennai, 600026, India
*Corresponding Author: S. Michael Dinesh. Email: ermichaeldinesh@gmail.com
Received: 18 October 2021; Accepted: 14 January 2022

Abstract: Deep learning has risen in popularity as a face recognition technology in recent years. Facenet, a deep convolutional neural network (DCNN) developed by Google, recognizes faces with 128 bytes per face. It also claims to have achieved 99.96% on the reputed Labelled Faces in the Wild (LFW) dataset. However, the accuracy and validation rate of Facenet drops down eventually, there is a gradual decrease in the resolution of the images. This research paper aims at developing a new facial recognition system that can produce a higher accuracy rate and validation rate on low-resolution face images. The proposed system Extended Openface performs facial recognition by using three different features i) facial landmark ii) head pose iii) eye gaze. It extracts facial landmark detection using Scattered Gated Expert Network Constrained Local Model (SGEN-CLM). It also detects the head pose and eye gaze using Enhanced Constrained Local Neural field (ECLNF). Extended openface employs a simple Support Vector Machine (SVM) for training and testing the face images. The system's performance is assessed on low-resolution datasets like LFW, Indian Movie Face Database (IMFDB). The results demonstrated that Extended Openface has a better accuracy rate (12%) and validation rate (22%) than Facenet on low-resolution images.

Keywords: Constrained local model; enhanced constrained local neural field; enhanced hog; scattered gated expert network; support vector machine

1 Introduction

Identifying human faces is something that humans habitually do naturally without much effort, yet it is still a hard issue on the ground of computer vision. A facial recognition framework is a framework that recognizes an individual from a computerized image or a video snippet. Due to its dynamic role, exclusively in video surveillance and forensic tasks, it has increasingly attracted the attention of academia and industry in the last decade [1]. Over the past epoch, there has been huge research and studies evolved over the problem of facial recognition from grayscale to colour images. To meet user demands, a wide range of researches ranging from an easy problem with an image that contains only a person looking straight into the camera, to the colour image that holds multiple persons in the photo, the face bowed a trivial angle or partially hidden and the complication of the background image has already geared up.

However, there are still open challenges due to many uncontrollable changes in lighting, point of view, or occlusion. Facenet, a modern face recognition tool introduced by Google was able to produce 100% results on LFW datasets irrespective of the factors like occlusion, viewpoint, and illumination. Facenet utilised a DCNN model of Zeiler and Fergus, L2 normalisation, and triplet loss for performing face recognition. It also achieves 99.63% results on LFW and 95.12% on youtube faces. But as the input size of the image decreases, the accuracy also decreased [2]. In real-world face recognition applications, the most frequently happening scenario is to recognize face images in a long shot. The long-shot images tend to have very poor resolution. Facenet was tested on 15000 images with low resolution (10 * 10), the accuracy rate and validation rate dropped to 58.1% and 1.6% unexpectedly. Hence it becomes inevitable to develop facial recognition (person re-identification) system that can perform well on low-resolution datasets.

This research paper aims at proposing a new system to enhance the accuracy rate of person re-identification on low-resolution images. Extended openface uses three parameters i) facial landmark detection ii) head pose orientation and eye gaze estimation that decides the accuracy of person re-identification. The followings are the highlights of this paper: 1) facial landmark detection is performed using SGEN-CLM. This convolutional neural network (CNN) models the appearance of all the landmarks individually by deriving response maps for each region of interest in the face image. It nullifies the impact of the occlusion. 2) Head pose orientation and Eye gaze estimation are performed using ECLNF. It utilizes both sparse-holistic landmark detection and dense-local landmark detection. 3) Feature Representation is performed using an Enhanced Histogram of Gradient (EHOG). EHOG feature is obtained through fixed-size blocks and variable-size blocks. It improves the contour description to generate better results during classification. The introduction of the problem statement, the need for this research, and the challenges involved in this research had already been well discussed in segment 1. The remaining part of the paper is laid out as follows: A summary of researches associated with the proposed methods is presented in segment 2. The proposed system is discussed in segment 3. The experimental setup and dataset used for verification are presented in segment 4. The systematic experimental outcomes and discussions are presented in segment 5. A comparative study with the existing methods is presented in segment 6. The conclusion is presented in segment 7.

2 Related Works

Facial Recognition is not a budding field, it has evolved a lot over years. At the initial stage, the research studies were more focused on feature extraction and classification. Recently, there was a paradigm shift in this field with the emergence of machine learning and artificial intelligence. Deep learning is a form of machine learning that is also important in facial recognition research. Even Google and Facebook have come up with their deep learning-based facial recognition system namely Facenet, DeepFace. This section provides an overview of recent advancements in facial recognition.

2.1 DeepFace

DeepFace is a facial recognition system that works based on deep learning. It is formulated by a group of researchers at Facebook [3]. It follows a siamese network. It is a neural network of nine layers with more than 120 million connection weights. Facenet was also trained on a Facebook database that consists of nearly four million images. The system achieved a success rate of 97% in the LFW dataset.

2.2 Facenet

Facenet is a facial recognition system created by the developer's community of Google. Facenet claimed a success rate of 99.96% in the LFW dataset. But for low-resolution facial images, the situation is quite the opposite. This low-resolution challenge arises in various prevailing facial recognition algorithms, and for

these reasons, reasonable performance is difficult to achieve. The researches prove that the accuracy of facenet drops gradually as the resolution of the input image is reduced. Facenet recorded its lowest accuracy of 58.2% for an input image of 10 * 10 resolution. Fig. 1 shows the accuracy.

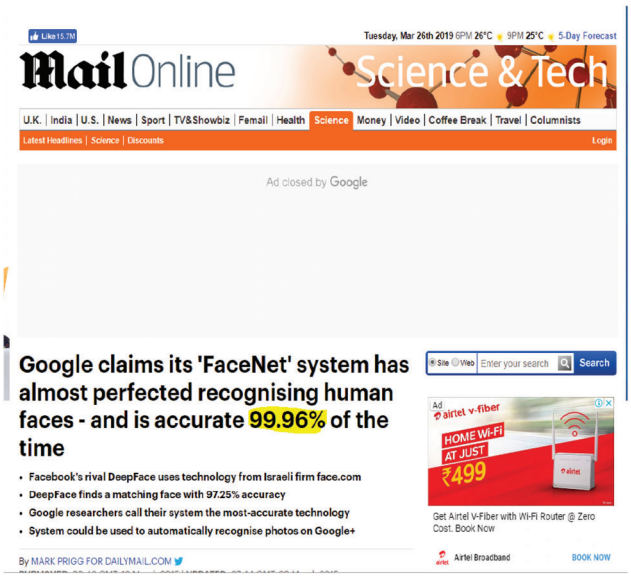


Figure 1: Facenet shows 99.96% of accuracy (courtesy: mail online, march 2019)

2.3 Open Face

Besides DeepFace, Facenet being owned by Facebook and Google, there also exists a free open-source face recognition system termed as Openface [3]. The fundamental working of the open face can be separated into two sections i) training ii) recognition. The training phase aims at producing a trained neural network. The recognition phase aims at classifying the image-based using the trained model.

It is an open-source free face recognition that works based on deep neural networks. It works based on i) facial landmarks ii) head pose and iii) eye-gaze features. It works by projecting the face on a 128-dimensional unit hypersphere and forming an embedding. An embedding is capable of differentiating two faces clearly, thus it simplifies clustering, similarity detection, and classification tasks. It is well suited at the point where Euclidean distance between features is not effective. On evaluating the performance of the openface [4] system against the datasets like LFW, Scface, Forenface, the researchers suggest that it is still not suited for forensic application where higher accuracies are expected. Santoso et al. proposed Face Recognition using modified openface [5] using modified triplet loss and Adaptive Moment Estimation. Modified Openface produced better results than the original openface.

2.4 Image-Based Facial Recognition

Zhang et al. proposed [6] a light weighted multitasked cascading CNN for joint face recognition and alignment. It is used for online hard sampling to advance the precision of face detection and alignment. Zeng et al. demonstrated hierarchical deep learning facial recognition [7] in which feature learning and metric learning play a critical role. This approach holds three stages i) Identity Generalised Embedding for feature representation learning using DCNN ii) efficient Cross-view Quadratic Discriminant Analysis (XQDA) is used for metric learning. Wei Li et al. projected Deep Joint Learning of Multi-Loss Classification [8] showed the advantages of mutually learning the local and global features. A structure sparsity-based feature selective learning mechanism to improve the efficiency of deep joint learning of

multi-loss classification. Zepp Uibo on his thesis termed as Comparison of Face Recognition Neural Networks. The famous facial recognition neural networks like Visual Geometry Group (VGG), CASIA, and Openface were tested on LFW and FOTIS datasets. The author concluded VGG performed better among all three networks.

2.5 Video-Based Person Re-Identification

Shuang Li et al. proposed Diversity Regularized Spatiotemporal Methodology [9] that involves Encoding Discriminative Image Region than the whole image using Multiple Spatial Attention Model. But Diversity Regularization is to be considered to be operational overhead. Wei Zhang et al. [10] proposed a new methodology Learning Compact Appearance Representation where several reference frames were selected using flow energy profile (FEP). Features are pooled using CNN and distance metric learning is employed for identification. But feature pooling is very fragile to ambiguity caused by the background. Mclaughlin, Rincon et al. proposed Recurrent Convolutional Network (RCN) [11] that follows Siamese architecture. The convolutional network, recurrent layer, and temporal pooling layer are prepared together and play a role as a feature extractor for video-based search. The efficiency of this method is unstable when it comes to real multitarget tracking options. Zhen Zhou et al. proposed the recurrent neural networks [12] that involve two-step processes namely i) feature learning ii) metric learning. Feature learning was implemented using the temporal attention model whereas Metric learning was implemented using the spatial attention model. This method suffers due to serious occlusions, heavy illumination changes, and different clothing of the same person. Deqiang Ouyang et al. proposed a self-paced learning framework [13] for identifying discriminative features and deep reinforcement for recognition. Yu Wu et al. proposed a stepwise learning approach [14] that improves the discriminating capability of CNN. It engaged a progressive method to exploit the unlabelled tracks efficiently. The strategy additionally utilizes a distance-based testing basis for label estimation and candidate selection to upgrade the accuracy of label estimation. Shuangjie Xu introduced Spatial-Temporal Pooling Networks (ASTPN) [15]; it performs by feeding a pair of videos to the Siamese network & producing Euclidean distance between them. Mang Ye et al. proposed Robust Anchor Embedding for Unsupervised Video Face Recognition in the Wild [16]. It uses an unsupervised deep feature rendering learning framework to achieve similarity between unlabeled sequences and anchor points to perform the label estimation. The framework also utilises both visual and intrinsic similarity to reduce the false positives. Dapeng Chen et al. proposed Competitive Snippet-similarity Aggregation and Co-attentive Snippet Embedding [17]. This method splits a long video into many brief video segments and sums the high-ranked snippet similarities for sequence-similarity estimation.

The following conclusion is drawn based on the various related works discussed above.

1) Deepface and Facenet have gradually increased the accuracy of face recognition. Deepface achieved 97% and Facenet achieved 99% on LFW datasets. But the accuracy rate gradually drops with the decrease in the resolution of an image.
2) Openface also yielded an impressive performance on LFW datasets. However, it is not suitable for forensic applications in its current form due to its performance at the low quality of the images collected from CCTV.

The performance degradation of the facial recognition systems remains an unsolved problem, as extracting facial features from a low-resolution image is an extremely tedious job for people and technologies. Hence, there is a huge need for a facial recognition system that can work on low-resolution images. This research paper goal at bringing out the better performance of facial recognition systems on low-resolution images.

3 Proposed System

To enhance the face recognition system's results in low-resolution frames, the Openface system is improved using the techniques SGEN-CLM, ECLF, and EHOG. Extended Openface works in five steps namely i) capture videos from the source ii) convert video into the frames iii) Feature Extraction and Representation Formation Using SGEN-CLM, ECLNF, and EHOG iv) Predict the face in the video using SVM Classifier. Fig. 2. shows the architecture diagram of Extended openface.

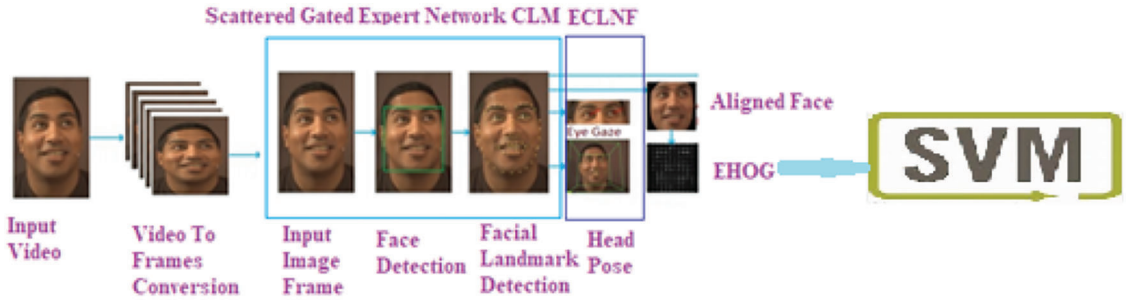


Figure 2: Architecture diagram of extended openface

3.1 Obtaining Video & Video to Image Conversion

First and foremost, Extended Open face will accept the input in the form of the videos. Videos can be obtained from offline video (stored on hard disk) or real-time video that can be captured from a web camera. Then, the video clips are partitioned into the keyframes or images using short boundary detection.

3.2 Feature Extraction & Representation Formation

In the second step, the feature extraction and representation are carried out. The feature that is extracted are i) facial landmarks ii) head pose iii) eye gaze. Facial landmark features are extracted using SGEN-CLM whereas head pose and eye gaze features are extracted using ECLNF. Facial landmark detection using SGEN-CLM is discussed in Section 3.2.1. Head pose and eye gaze detection are explained in Section 3.2.2 and 3.2.3 respectively. After feature extraction, the aligned face is represented using EHOG.EHOG is discussed in 3.2.4

3.2.1 Facial Landmark Detection

Facial landmark detection is performed with the help of combined usage of constrained local model and mixture of expert layer [18]. SGEN-CLM model consists of five steps namely i) computing region of interest ii) contrast normalizing iii) convolution iii) expert layer v) deriving the output in the form of response map. The novelty of SGEN-CLM is it treats every landmark individually, thus eliminating the effect of occlusion. The diagrammatic representation of SGEN-CLM is shown in Fig. 3.

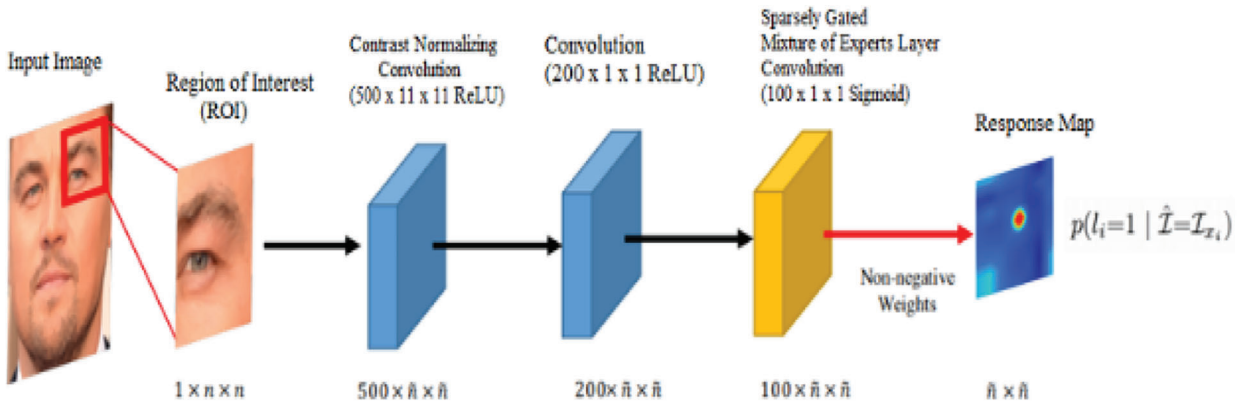


Figure 3: Working of scattered gated expert network-CLM

An algorithmic description of SGEN-CLM is described below.

Algorithm: SGEN-CLM

Step 1: Compute Region of Interest (ROI)

- Collect the n × n ROI on the existing estimate of the landmark location as input and generate a response map that evaluates the likelihood of landmark alignment at each pixel location. (Note: Localization of individual landmarks is performed by assessing the landmark alignment probability at a specific pixel location. The region around the estimated landmark is treated as ROI (n × n pixel))

Step 2: Perform Contrast Normalization

- Using z-score normalization, the region of interest is convolved with a 500 × 11 × 11(Kernel Shape) contrast normalized convolution layer. The mean and standard deviation of contrast in kernel shape is computed. The contrast value in ROI is subtracted with a mean value of kernel shape and then divided by the standard deviation of kernel shape to obtain the normalized contrast value of ROI.
- Correlation between input and kernel is computed.
- Output of this layer is forwarded to the convolution layer.

Step 3: Perform Sharpening using Convolution Layer

- Output response of contrast normalizing layer is convoluted in this layer
- The convolution layer is of capacity 200 × 1 × 1 ReLU neurons.

Step 4: Model Facial Landmark Alignment using Scattered Gated Mixture of Expert (SGMoE)

- SGMoE is the third layer of this model consisting of i) experts ii) gating network. An expert is a feed-forward neural network.
- Gating network can be prepared to choose a scattered blend of the experts to process each input. SGMoE comprises of a set of 'n' expert networks E1, E2…. En and a gating network G, whose output is a sparse n-dimensional vector. In this process of facial landmark alignment, each expert can present its own probability alignment, and the Gating network decides the best outcome based on the maximum votes of experts.
- Compute y

$$y = \sum_{i=1}^{n} G(x)IEi(x) \tag{1}$$

where G(x) → output of the gated network

Ei(x) → output of the ith expert network with input x

E1, E2…. En → Expert is a feed-forward neural network that can make facial landmark alignment.

- This layer possesses the capacity to model the final probability alignment by using the output received from the previous layer.
- It is activated by a sigmoid function, which generates votes from various experts based on the likelihood of alignment.

Step 5: Generate Response Map

- SGEN-CLM uses the non-negative weights from the last layer to combine the response graphs. The response graph can be tracked to get a facial landmark.

Once facial landmark detection is performed through SGEN-CLM, the system proceeds to the next step head pose estimation.

3.2.2 Head Pose Estimation

Estimating head posture has never attracted a similar level of attention as detecting facial landmarks. Watson's system implemented based on the Generalized Adaptive View-based Appearance Model is also used in head pose estimation. Other than Watson, many other frameworks also use depth data to perform head pose estimation, but it does not suit webcam applications. The proposed model can extract a head pose that includes translation and orientation. Head pose estimation using ECLNF includes two major steps i) Precompute ii) Procedure Fit. Precompute is mainly focused on camera calibration and face model generation, whereas procedure fit concentrates on the deriving pose and deformation parameters. The algorithmic description of head pose estimation using ECLNF is described below:

Algorithm: ECLF

Precompute

- Calibrate camera intrinsic matrix K

$$K = \begin{pmatrix} fx & \alpha & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

Where fx, fy → focal length of a camera

cx, cy → principal point of a camera

α → skew

- Develop 3D mean shape s0 and eigenvectors Φ
- Prepare 2D texture filters Hk around n PDM points

End

Procedure Fit (deformation d, pose p, the image I)

Using p to reduce and rotate I

separate head Iface

if the First frame or tracking is lost then

Use Max-Margin Object Detection to identify faces.

Create a filter of global correlation

else

Produce global response map

Look for tracking failure

Compute global shift among frames

Update global filter

repeat

for k ← 1,…, n points in the head model

do

(Continued)

Algorithm: (continued)

Create Γk via local regions of Iface and Hk
Find mean-shift estimates vk from Γk
Project the 3D point Xk into the image,W(Xk, q)
Compute deformable derivatives ∂Xk/∂d
Evaluate derivatives of pose ∂ ˜ Xk/∂˜ r and ∂ ˜ Xk/∂t
Evaluate projection derivatives ∂W(Xk, q)/∂q
Combine derivatives for final Jacobian matrix J
Compute the parameter updates Δq =Δd Δp
Update deformation parameters d ← d + Δd
Update pose p ← p + Δp, with rot. Adjustment
Find the new 3D points X
until (kΔqk <) or (iter = maxIter)

ECLNF generates pose and deformation parameters at the end of the processing. The parameter pose (p) helps in the generation of yaw, pitch, and roll that decides the quality of head pose attributes.

3.2.3 Eye Gaze Estimation

Given the limited computing resources, the poor quality of the integrated RGB front camera, and the small screen to which the gaze is assigned, the gaze has been deemed a challenge on portable devices [19]. ECLNF is a deformable shape registration method that detects eye-region landmarks. Eye region landmark contains eyelids, iris, and the pupil. SynthesEyes dataset is used to train ECLNF patch experts. Once the positions of the eyes and pupils are detected using the ECLNF model, the same information is used to calculate the gaze vector of each eye separately. To find the position of the pupil in the coordinates of the 3D camera, a light beam is emitted from the origin of the camera to the pupil's center in the image plane, and the intersection point with the eyeball is calculated. The vector from the 3D eyeball center to the pupil location provides an estimated gaze vector.

3.2.4 Extended Histogram of Gradient Representation

EHOG works based on unidirectional gradient, it can produce better unique features for each pattern thus making the classification process easier. All three features i) facial landmark ii) head pose iii) eye gaze is represented in the form of EHOG. EHOG is preferred as HOG maps gradients of different directions in a cell into the same bin of the histogram, it is unable to distinguish between some patterns and produces the same feature for two different patterns. EHOG that works based on unidirectional gradient produces more accurate unique for each pattern thus making the classification easier. EHOG is fed into the SVM classifier for training and testing. EHOG feature is collected through the fixed-size blocks and variable size blocks, which can improve the description of the contour. EHOG function is based on each group of unidirectional gradient images. It calculates the vector in each uni-orientation gradient image of the group and then combined it into the final EHOG features. The diagrammatic representation of EHOG is shown in Fig. 4.

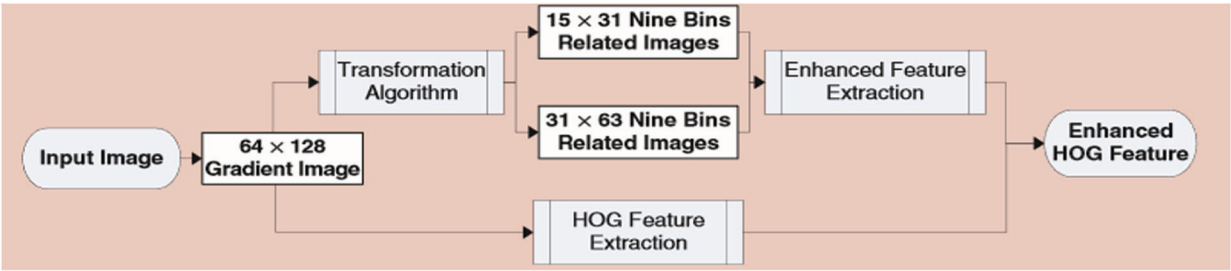


Figure 4: Diagrammatic representation of EHOG

Algorithm: EHOG

Input: 15 * 31-pixel uni-orientation gradient image

Output: EHOG feature in form of dimensional vector

Procedure:

- Partition each 15 * 31-pixel uni-orientation gradient image of the group into 7 * 3 = 21 blocks, each block should be of 8 * 8 pixels.
- Fragment each block of 8 * 8 pixels into 4 * 4-pixel cells.
- Histogram of all 8 * 8 blocks is developed by four bins through bilinear interpolation. The usage of the corresponding weight, pixel's value not only contributes to the cell to which the pixel belongs but also to that of neighbouring cells.
- Every pixel(i, j) in the block has weight WN(i, j) where N refers to the number of bins.
- Weights related to the spatial location of the pixels are computed as below.

$$W1(i,\ j) = (1 - ti)(1 - tj) \tag{3}$$

$$W2(i,\ j) = (1 - ti)(tj) \tag{4}$$

$$W3(i,\ j) = (ti)(1 - tj) \tag{5}$$

$$W4(i,\ j) = (ti)(tj) \tag{6}$$

where W1, W2, W3, W4 → weights are related to the spatial location of the respective pixels.

- Then we perform normalisation

$$W\ = \sum I(i,\ j)2 \tag{7}$$

$$vi\ =\ vi/\left(€+ \sqrt{w}\right) \tag{8}$$

where € → small error rate

vi → the dimensional vector for enhanced HOG

W → normalised weight.

EHOG contains an additional feature that makes the representation accurate and easier. EHOG is fed to SVM for facial recognition in the next step.

3.3 Testing & Training of SVM

As a final step, the representation obtained from ECLNF and EHOG was used for training and testing SVM. SVM is preferred over the neural network as it can work well with less amount of dataset and because of its compatibility to operate well with HOG kind of representation.

SVM receives a bunch of inputs and forecasts two of the potential classes [20]. The classifier is trained from scratch and the trained classifier is allowed to infer the testing dataset. In the training phase, SVM is fed with EHOG Representation and training label leading to the formation of the trained model. The trained model classifies the EHOG representation provided.

4 Experimental Setup and Dataset for Verification

Extended Open Face is a deep neural network-based face recognition system that is implemented in Python and Torch. Open Face depends on some external libraries like dlib, OpenCV, a torch that is freely available opensource. The newly developed system was tested against the LFW, IMFDB, MPIIGaze datasets.

4.1 LFW

LFW is expanded as Labelled Faces in the Wild. LFW contains nearly 13233 images and 5749 people. LFW Dataset is formed and maintained by scholars at the University of Massachusetts. Images are in the format of jpg and it is of the dimension of 250 * 250. The entire Labelled Faces in the Wild database can be downloaded as a zipped tar file from the official website of LFW. LFW dataset is used to evaluate i) the performance of SGEN-CLM for facial landmark detection and ii) the performance of Extended Openface on low-resolution images.

4.2 IMFDB

IMFDB contains 34512 2D-frames of multiple Indian film celebrities composed from various videos. It was released by the Centre for Visual Information Technology at IIIT, Hyderabad [21,22]. The below guidelines were considered while forming the dataset i) selection of frames ii) cropping of faces. Only one frame is chosen unless there is an alternative frame with noteworthy change with the previously chosen frame. If there are numerous differences presented in a shot, a face with occlusion and pose difference were preferred. The dataset can be freely downloaded from the official website of IMFDB. This dataset contains a huge number of low-resolution images. IMFDB is used to assess the performance of Extended Openface on low-resolution images.

4.3 MPIIGaze

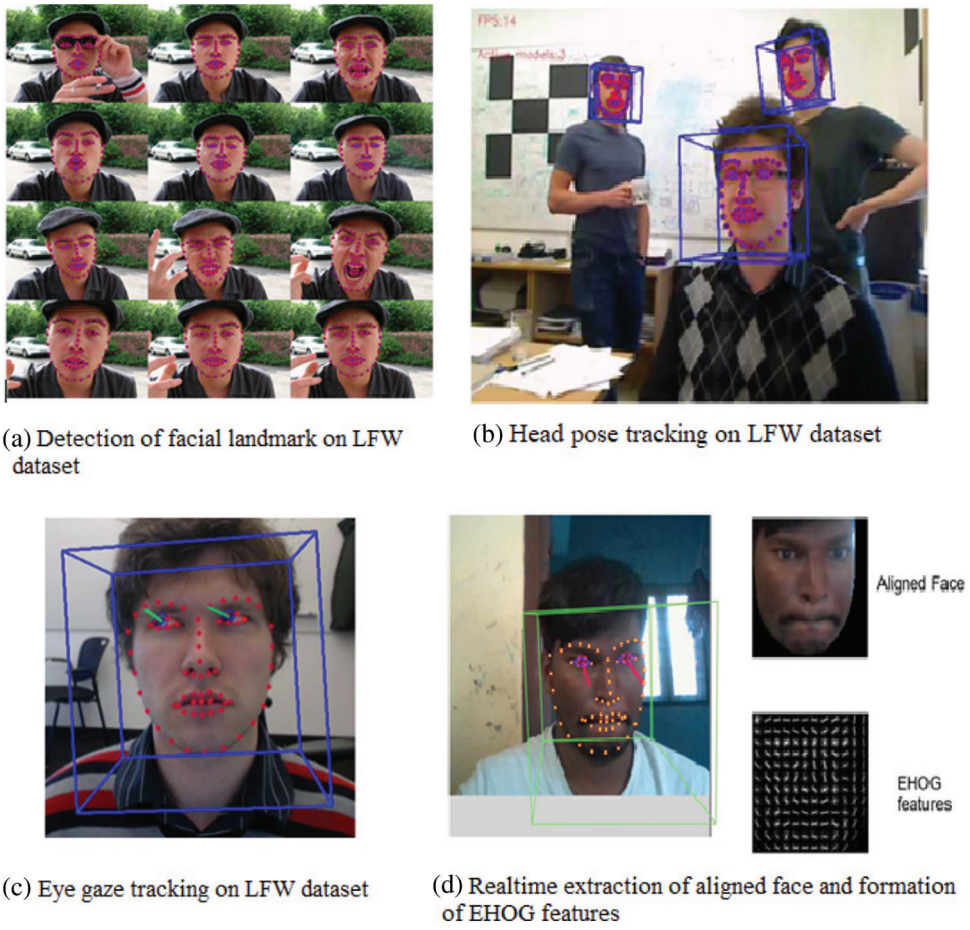
MPIIGaze dataset is a dataset that comprises a huge number of images from a different set of people, capturing the different moments of their daily life. All pictures in the dataset are marked with 3D annotations of gaze target and spotted eye/head positions. This dataset also offers manual facial landmark annotations on a subset of images, which allows a comprehensive assessment of gaze estimation results. This dataset is used to evaluate the performance of ECLNF for eye gaze estimation.

4.4 Biwi

Biwi is a dataset that contains over 15 K pictures of 20 individuals (6 women and 14 men). The head pose range covers about +−75 degrees yaw and +−60 degrees pitch. All the images in the datasets are annotated. This dataset is used to evaluate the performance of ECLNF for head pose estimation.

5 Experimental Results

In this segment, the test outcomes of the proposed system are shown as below i) test results on LFW shown in Fig. 5 v) test result on IMFDB shown in Fig. 6 vi) test results on MPII gaze datasets shown in Fig. 7. Sections 5.1 and 5.2 show the results and accuracy/validation rate calculation of the proposed system on LFW and IMFDB datasets. (Note: dotted representation points to the facial landmark detection, cubic representation points to head pose detection and straight line coming out of eye represents the gaze.)



(a) Detection of facial landmark on LFW dataset

(b) Head pose tracking on LFW dataset

(c) Eye gaze tracking on LFW dataset

(d) Realtime extraction of aligned face and formation of EHOG features

Figure 5: Results of extended openface on LFW dataset

5.1 Experimental Results on Low-Resolution LFW Dataset

LFW dataset has an image of dimension 250 * 250. To obtain a low-resolution image data set, a large variation was performed to the image's size used for assessment. Using nearest neighbour interpolation, the photos were first resized to 10 * 10, 20 * 20, 40 * 40, 60 * 60, 80 * 80, and 100 * 100, respectively.

$$\text{Accuracy Rate} = (\text{T.P} + \text{T.N})/(\text{T.P} + \text{T.N} + \text{F.P} + \text{F.N}) \quad (9)$$

$$\text{Validation Rate} = (\text{T.P})/(\text{T.P} + \text{F.N}) \quad (10)$$

where T.P stands for True Positive, T.N stands for True Negative, F.P stands for False Positive, F.N Stands False Negative. Tab. 1 shows the results of the proposed system on low resolution images of LFW.

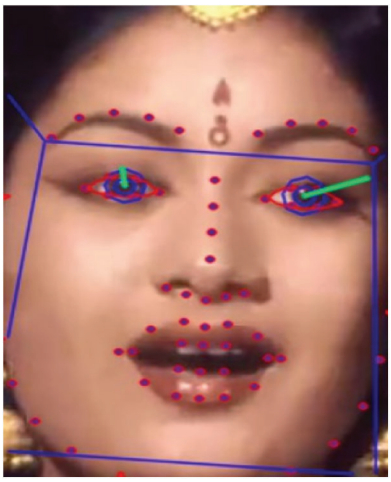


Figure 6: Facial landmark, Eye gaze, and Head pose on IMFDB

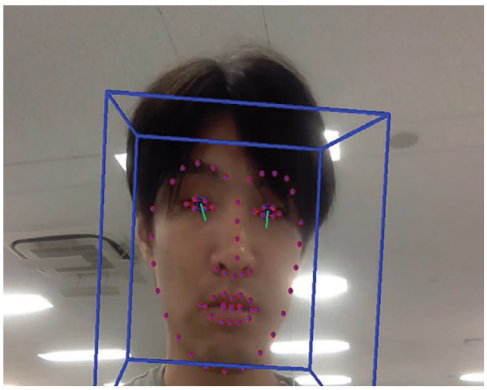


Figure 7: Facial landmark, Eye gaze, and Head pose on MPII Gaze

Table 1: The results of the proposed system on low resolution images of LFW

Input image size	Accuracy rate (%)	Validation rate (%)
160 * 160	97.2	96.8
100 * 100	96.9	96.9
80 * 80	96.1	95.7
60 * 60	95.5	93.9
40 * 40	94.1	89.5
20 * 20	93.7	11.2
10 * 10	90.2	7.4

5.2 Experimental Results on IMFDB Dataset

IMFDB dataset contains images ranging from the resolution of 110 * 110 to 40 * 40 pixels. To obtain a low-resolution image dataset, the alterations were made in the size of the images used for evaluation. The images in the size of 40 * 40 were reduced to 10 * 10, 20 * 20 using nearest-neighbour interpolation.

The images on the scale of 110 * 110 were enlarged to 160 * 160 using zooming interpolation. Tab. 2 shows the performance of the proposed system on low-resolution images of IMFDB.

Table 2: The performance of the proposed system on low-resolution images of IMFDB

Size of input image	Accuracy (%)	Validation (%)
160 * 160	95.2	95.3
100 * 100	96.4	94.7
80 * 80	96.3	94.1
60 * 60	95.1	93.8
40 * 40	94.6	89.1
20 * 20	92.7	12.1
10 * 10	79.2	8.4

6 Comparison with the Existing Methods

This segment aims at comparing the proposed methodology against various existing methodologies. Comparative analysis was done with three research questions i) How well the proposed technique SGEM-CLM for facial landmark detection performs in comparison with the other state of art methods? ii) How well does the proposed technique ECLNF for head pose and eye gaze detection perform in comparison with the other state of art methods? iii) Does Extended Openface perform better than Facenet on low-resolution datasets. The three proposed techniques are compared against various existing techniques respectively i) Section 6.1 shows a comparison of SGEN-CLM facial landmark detection with the other related works ii) Section 6.2 shows a comparison of ECLNF head pose detection with other works iii) Section 6.3 shows a comparison of ECLNF eye gaze estimation with other works iv) Section 6.4 compares the performance of the proposed system on the low-resolution images.

6.1 Comparative Analysis of Proposed Facial Landmark Detection Using SGEN-CLM

The detection of facial landmarks was evaluated on the LFW and IMFDB datasets. At the first, the proposed system has been evaluated. Then the proposed technique SGEN-CLM is equated against the other algorithms whose executions are found on the web and that have been trained to identify identical facial landmarks. The existing systems used for the comparison were: LNF, SVR, CEN. Two parameters were used for squared correlation (r^2) and root mean square error (RMSE). r^2 is a value that shows how well it fits your training data. RMSE compares the predicted value and the actual value. RMSE was multiplied with 10^3 as its value was in the precision of three decimal points. For making it easier for the comparison purpose, it was multiplied by 10^3. Tab. 3 shows the comparison of facial landmark detection with other methods.

Table 3: Comparison of facial landmark detection with other methods

Detector	r^2	RMSE * 10^3
SVR [23]	21.31	66.8
LNF [24]	36.57	59.2
CEN	64.22	37.9
CEN (with no ME layer)	23.81	65.11
SGEN-CLM	66.74	36.5

The proposed technique SGEN-CLM network showed higher r^2 and lower RMSE than the other facial landmark detection techniques.

6.2 Comparative Analysis of Proposed Head Pose Estimation Using ECLNF

The publicly available datasets with current ground truth head pose estimation: Biwi is used to measure the adequacy of the proposed system on the head pose estimation [25]. For comparison, the detailed analysis of the results of the Chehra framework, CLM, CLM-Z, and Regression Forests is presented in Tab. 4. The proposed model Extended openface is evaluated based on four parameters Yaw, Pitch, Roll, Mean.

The proposed technique ECLNF for head pose detection had better results than the existing techniques in terms of parameters of yaw, pitch, and roll.

Table 4: Comparison of head pose estimation results with various methods

Method	Yaw	Pitch	Roll	Mean
Reg.Forests [26]	9.2	8.5	8.0	8.6
CLM	8.2	8.2	6.5	7.7
CLM-Z	8.0	6.1	6.0	6.7
Chehra [27]	13.9	14.7	10.2	12.9
Openface	7.9	5.6	4.5	6.0
ECLNF	8.01	5.8	4.7	6.1

6.3 Comparative Analysis on Proposed Eye Gaze Estimation Using ECLNF

Eye gaze estimation is performed on the MPIIGaze dataset. Nearly 750 images of the MPIIGaze dataset were used for calculating gaze error. The performance comparison of the various system is displayed in Tab. 5.

Table 5: Comparison of gaze error with the various methods

Model	Gaze error
EyeTab [28]	47.1
Conv.NN on UT	13.91
Conv.NN on SynthesEyes	13.55
Conv.NN on SynthesEyes + UT	11.12
OpenFace	9.96
ECLNF	9.91

The proposed technique ECLNF for eye-gaze detection had low gaze error than the existing methodologies [29].

6.4 Comparative Analysis of Extended Open Face with Facenet

On comparing the outcome of Facenet and the proposed model, though the outcome of the proposed model is comparatively less in the high-resolution images between the resolution range 160 * 120

between 40 * 40. But the outcome of the proposed model was much better than Facenet in low-resolution images of range between 20 * 20 and 10 * 10. The performance of Facenet and Extended Openface based on accuracy rate is shown in Fig. 8.

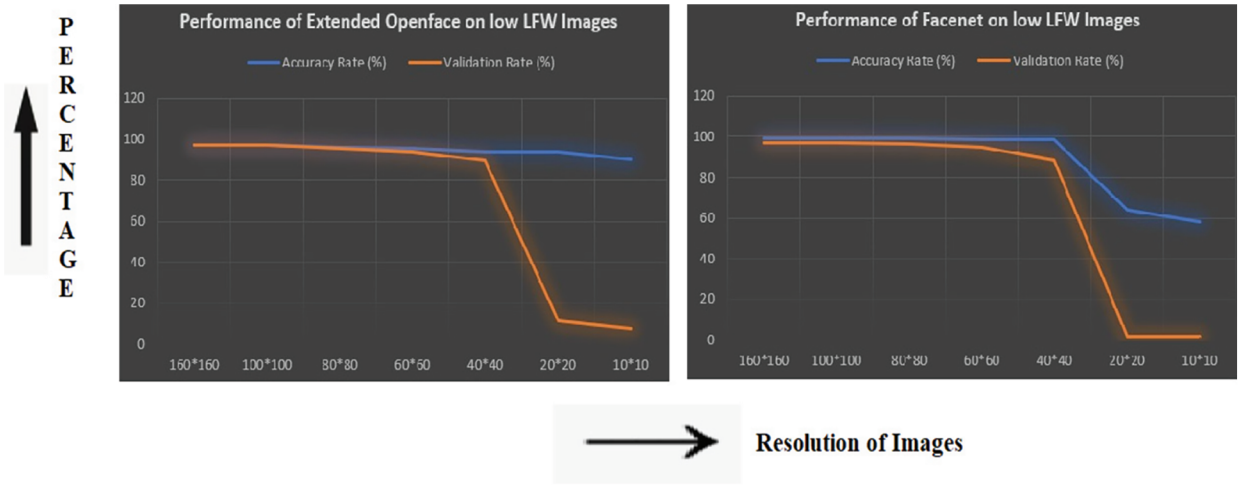


Figure 8: Performance comparison of Facenet & Extended Openface on low-resolution images

The following findings are clear based on the proposed system's results and comparisons with other state of art methods i) the accuracy and validation rate of Extended Openface is much better than Facenet on low-resolution datasets ii) the results of individual proposed techniques used for facial landmark detection (SGEN-CLM), head pose estimation (ECLNF) and eye gaze estimation are also better than the other state of art methods.

7 Conclusion

This research paper describes an effective technique "Extended Openface" for Person Re-identification. The three major contributions of this work are i) Scattered Gated Expert of Network for facial landmark detection ii) Enhanced Constrained Local Neural Field for head pose and eye gaze estimation and iii) Enhanced HOG for feature representation. SGEN-CLM network showed higher r^2 (66.74) and lower RMSE (36.5) than the other facial landmark detection techniques. ECLNF for head pose detection had better results than the existing techniques in terms of the parameter of yaw (8.01), pitch (5.8), and roll (4.7). ECLNF for eye-gaze detection had low gaze error (9.91) than the existing methodologies. Holistically, Extended openface recorded an accuracy rate of 90.2% and a validation rate of 7.4% on the LFW dataset with the image resolution of 10 * 10. It also recorded the accuracy rate of 79.2% and validation rate of 8.4% on the IMFDB dataset with the image resolution of 10 * 10. The proposed system showed an improved accuracy rate and validation rate than Facenet by 30% and 6% respectively.

Therefore, it is evident that Extended Openface improved the results of person re-identification on the low-resolution images. In our future works, this research can be applied to the surveillance system where the low-resolution image steps in as an input.

Acknowledgement: I am honoured and grateful to thank my guide, Dr. A.R. Kavitha M.E., Ph.D. for her invaluable advice and unwavering support. I'd like to thank my professor, Dr. A. Packialatha M.E., Ph.D. for her untiring support and encouragement during my doctoral path. Last but not least, I'd want to thank my family (Late Simon–My Father, Mrs. Shanthi Simon–My Mother, Mr.Aloseous Kingsly–My Brother-in-law, Mrs. Gnana Rita M.A., B.Ed. – My Sibling, Mrs. Pearlin M.E–My Wife, Baby. Farren Fernley & Baby. Natalie Nashley–My nieces) and friend (Mr. K. Saravanan M.Com.,M.B.A) for their boundless honest support and consistent motivation.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] T. Baltrušaitis, P. Robinson and L. P. Morency, "3D constrained local model for rigid and non-rigid facial tracking," in *IEEE Conf. on Computer Vision and Pattern Recognition*, USA, pp. 2610–2617, 2012.

[2] M. R. Golla and P. Sharma, "Performance evaluation of facenet on low-resolution face images," in *Int. Conf. on Communication, Networks and Computing*, Gwalior, India, pp. 317–325, 2018.

[3] T. Baltrušaitis, P. Robinson and L. P. Morency, "Openface: An open source facial behavior analysis toolkit," in *IEEE Winter Conf. on Applications of Computer Vision (WACV)*, NY, USA, pp. 1–10, 2016.

[4] A. Fydanaki and Z. Geradts, "Evaluating OpenFace: An open-source automatic facial comparison algorithm for forensics," *Forensic Sciences Research*, vol. 3, no. 3, pp. 202–209, 2018.

[5] K. Santoso and G. P. Kusuma, "Face recognition using modified open face," *Procedia Computer Science*, vol. 135, pp. 510–517, 2018.

[6] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[7] M. Zeng, C. Tian and Z. Wu, "Person re-identification with hierarchical deep learning feature and efficient xqda metric," in *Proc. of the 26th ACM Int. Conf. on Multimedia*, Seoul Republic of Korea, pp. 1838–1846, 2018.

[8] W. Li, X. Zhu and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *Proc. IJCAI-26*, Melbourne, Australia, pp. 2194–2200, 2017.

[9] S. Li, S. Bak, P. Carr and X. Wang, "Diversity regularized spatiotemporal attention for video-based person re-identification," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, PR, USA, pp. 369–378, 2018.

[10] W. Zhang, S. Hu, K. Liu and Z. Zha, "Learning compact appearance representation for video-based person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 8, pp. 2442–2452, 2018.

[11] N. McLaughlin, J. M. Del Rincon and P. Miller, "Recurrent convolutional network for video-based person re-identification," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 1325–1334, 2016.

[12] Z. Zhou, Y. Huang, W. Wang, L. Wang and T. Tan, "See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 4747–4756, 2017.

[13] D. Ouyang, J. Shao, Y. Zhang, Y. Yang and H. T. Shen, "Video-based person re-identification via self-paced learning and deep reinforcement learning framework," in *Proc. of the 26th ACM Int. Conf. on Multimedia*, Seoul Republic of Korea, pp. 1562–1570, 2018.

[14] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Ouyang *et al.,* "Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 5177–5186, 2018.

[15] S. Xu, Y. Cheng, K. Gu, Y. Yang, S. Chang *et al.,* "Jointly attentive spatial-temporal pooling networks for video-based person re-identification," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 4733–4742, 2017.

[16] M. Ye, X. Lan and P. C. Yuen, "Robust anchor embedding for unsupervised video person re-identification in the wild," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 170–186, 2018.

[17] D. Chen, H. Li, T. Xiao, S. Yi and X. Wang, "Video person re-identification with competitive snippet-similarity aggregation and co-attentive snippet embedding," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 1169–1178, 2018.

[18] A. Zadeh, Y. Chong Lim, T. Baltrusaitis and L. P. Morency, "Convolutional experts constrained local model for 3d facial landmark detection," in *Proc. of the IEEE Int. Conf. on Computer Vision Workshops*, Venice, Italy, pp. 2519–2528, 2017.

[19] E. Wood and A. Bulling, "Eyetab: Model-based gaze estimation on unmodified tablet computers," in *Proc. of the Symposium on Eye Tracking Research and Applications*, Safety Harbor Florida, pp. 207–210, 2014.

[20] A. R. Kavitha and C. Chellamuthu, "Online medical healthcare application for cancer disease prediction based on pattern matching and SVM classifier with MVC (PMSMVC) framework," *International Journal of Biomedical Engineering and Technology*, vol. 12, no. 2, pp. 177–188, 2013.

[21] A. W. Senior and R. M. Bolle, "Face recognition and its application," in *Biometric Solutions*, Boston, MA: Springer, pp. 83–97, 2002. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4615-1053-6_4.

[22] S. Setty, M. Husain, P. Beham, J. Gudavalli, M. Kandasamy *et al.,* "Indian movie face database: A benchmark for face recognition under wide variations," in *Fourth National Conf. on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Jodhpur, India, pp. 1–5, 2013.

[23] J. M. Saragih, S. Lucey and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.

[24] T. Baltrusaitis, P. Robinson and L. P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *Proc. of the IEEE Int. Conf. on Computer Vision Workshops*, Sydney, NSW, Australia, pp. 354–361, 2013.

[25] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 815–823, 2015.

[26] G. Fanelli, T. Weise, J. Gall and L. Van Gool, "Real time head pose estimation from consumer depth cameras," in *Joint Pattern Recognition Symposium*, Germany, pp. 101–110, 2011.

[27] A. Asthana, S. Zafeiriou, S. Cheng and M. Pantic, "Incremental face alignment in the wild," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, OH, USA, pp. 1859–1866, 2014.

[28] E. Wood, T. Baltrusaitis, X. Zhang, Y. Sugano, P. Robinson *et al.,* "Rendering of eyes for eye-shape registration and gaze estimation," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Santiago, Chile, pp. 3756–3764, 2015.

[29] X. Zhang, Y. Sugano, M. Fritz and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 4511–4520, 2015.