Computers, Materials & Continua
DOI:10.32604/cmc.2022.023328
Article

Tech Science Press

Automated Facial Expression Recognition and Age Estimation Using Deep Learning

Syeda Amna Rizwan[1], Yazeed Yasin Ghadi[2], Ahmad Jalal[1] and Kibum Kim[3,*]

[1]Department of Computer Science, Air University, Islamabad, 44000, Pakistan
[2]Department of Computer Science and Software Engineering, Al Ain University, Abu Dhabi, 122612, UAE
[3]Department of Human-Computer Interaction, Hanyang University, Ansan, 15588, Korea
*Corresponding Author: Kibum Kim. Email: kibum@hanyang.ac.kr
Received: 03 September 2021; Accepted: 08 November 2021

Abstract: With the advancement of computer vision techniques in surveillance systems, the need for more proficient, intelligent, and sustainable facial expressions and age recognition is necessary. The main purpose of this study is to develop accurate facial expressions and an age recognition system that is capable of error-free recognition of human expression and age in both indoor and outdoor environments. The proposed system first takes an input image pre-process it and then detects faces in the entire image. After that landmarks localization helps in the formation of synthetic face mask prediction. A novel set of features are extracted and passed to a classifier for the accurate classification of expressions and age group. The proposed system is tested over two benchmark datasets, namely, the Gallagher collection person dataset and the Images of Groups dataset. The system achieved remarkable results over these benchmark datasets about recognition accuracy and computational time. The proposed system would also be applicable in different consumer application domains such as online business negotiations, consumer behavior analysis, E-learning environments, and emotion robotics.

Keywords: Feature extraction; face expression model; local transform features and recurrent neural network (RNN)

1 Introduction

Recognition of human age and expressions has engaged many researchers in various fields including sustainable security [1], forensics [2], biometrics [2], and cognitive psychology. Interest in this field is spreading fast and is fuelled by scientific advances that provide a better understanding of personal identity, attitudes, and intentions based on facial expressions and age. Facial expressions have a great impact on interpersonal communication. Human emotional responses are very complex and are most directly expressed in facial expressions. In the Mehrabian oral communication effects model, it is stated that 7% intonation, 38% expressions account when the people speak, and 55% body language accounts along with the facial expressions. Over the past few decades, researchers have conducted studies for human facial expressions recognition and age estimation (FERAE) systems that use advanced sensors such as video cameras, eye trackers, thermal cameras, human vision component

sensors [3–5], and stereo-cam [6,7] to intelligently recognize the human behaviours, gestures [8–10], emotions and to predict the age of an individual. Problems that arise in automatic FERAE systems are pose variations, uncontrolled lightning, complex backgrounds, partial occlusions, etc. Researchers face many challenges in attempting to overcome these problems.

Human subjects normally present various expressions all the time in daily life. To develop a sustainable expression recognition and age estimation system, we need to determine whether age estimation is influenced by changes in facial expression, how significant the influence is, and if a solution can be developed to solve the problem caused by facial expressions. Existing works on age estimation are mostly founded on expressionless faces. Most age estimation and expression recognition systems contain mainly frontal-view, neutral expressions, although some used variations in illumination, pose, and expression. To perform a systematic study on age estimation with various expressions, we need to use databases with clear ground truth labels for both age and expression.

The main distribution of our proposed model is as follows; First, face detection is done using the YCbCr skin color segmentation model. Second, landmark points are plotted on the face based on the connected components technique. Third, Synthetic face a mask is mapped on the face, based on landmark points localization. Fourth, features are extracted and subdivided into two categories. For age estimation, Anthropometric model, energy-based point clouds, and wrinkles are used for feature extraction. For expression recognition, HOG-based symmetry identification, energy-based point clouds, and geodesic distances between landmark points are extracted. Finally, Recurrent Neural Network (RNN) is used for the correct recognition of facial expressions and age.

The main contributions of the proposed system are:

- Synthetic face mask mapping increases the multi-face expressions and age recognition accuracy.
- Our local transform features of both age and expression recognition provide far better accuracy than other state-of-the-art methods.
- Recurrent Neural Network (RNN) classifier for the accurate age prediction and expressions recognition of individuals.

Our proposed sustainable FERAE model is evaluated using different performance measures over two multi-face benchmark datasets, namely, the Gallagher collection person dataset and an images of groups dataset which fully validated our system's efficacy showing that it outperforms other state-of-the-art methods.

This article is structured as follows: Section 2 describes related work for both facial expression and age recognition. Section 3 gives a detailed overview of the proposed model that intelligently recognizes multi facial expressions and age. In Section 4, the proposed model performance is experimentally assessed on two publicly available benchmark datasets. Lastly, in Section 5 we sum up the paper and future directions are outlined.

2 Related Work

Over the past few years, many researchers have done remarkable work on both single and multi-facial expressions recognition and age estimation. In this section, a comprehensive review of recent related studies of both facial expressions recognition and age estimation models are given in Section 2.1 and 2.2 respectively.

2.1 Multi-facial Expressions Recognition Systems

In recent years, many RGB-based facial expressions recognition systems have been proposed. In [11], the authors first detected facial features using Multi-task Cascaded Convolution Neural Network. After that, CNN and a VGG-16 model were used for the classification of facial expressions as Neutral, Positive, or Negative. The facial expression recognition accuracy on the Public dataset was 74%. In [12], the authors developed a system to recognize facial expressions in a variety of social events. Seetaface was used to detect the faces and align them. Visual facial features, i.e., PHOG, CENTRIST, DCNN features, and VGG features using VGGFace-LSTM and DCNN-LSTM, were then extracted. The system was tested on Group Affect Database 2.0 and achieved recognition accuracy of 79.78%. In [13], a hybrid network was developed in which CNN was used to pretrain the faces and extract scene features, skeleton features, and local features. These fused features were used to predict emotions. The system was tested on a public dataset and achieved system validation and testing accuracies 80.05% and 80.61% respectively. In [14], the authors developed a mood recognition system by first capturing the images from the web cam and trained two machine learning algorithm i-e., Gradient boosting classifier and K-Nearest Neighbor (KNN). The recognition accuracies achieved are 81% and 73% respectively.

2.2 Multi-facial Age Estimation Systems

In recent years, different methodologies have been adopted by researchers for the estimation of age or age group. In [15], The authors developed the system to estimate the age of real-life persons. features were extracted via Local Binary Pattern (LBP) and Gabor techniques. For classification, SVM was used. The system was tested on the Images of Group dataset and achieved an accuracy of 87.7%. In [16] extracted the features using LBP and FPLBP techniques. SVM was used for accurate age group classification and achieved an accuracy of 66.6%. In [17] developed a system for automatic classification of age and gender. Features were extracted via Multi-Level Local Binary Pattern (ML-LBP) whereas SVM with non-linear RBF kernel was used to classify according to correct age groups and gender. The system was tested on the Images of Group dataset and achieved an accuracy of 43.4%. In [18] proposed a system, the authors extracted the features and classified the correct age group using Convolution Neural Network (CNN). The system was tested on the OUI Adience dataset and achieved an accuracy of 62.34%.

3 Material and Methods

This section describes the proposed framework for facial expressions recognition and age estimation. Fig. 1 shows the general architecture of the proposed system.

3.1 Pre-processing and Face Detection

Our sustainable FERAE system starts with the preprocessing step, which involves two steps; 1) background subtraction, and 2) aligning the faces of both datasets at an angle of 180°. First, complex backgrounds are removed from the images to detect the faces more accurately. This is done using the median filtering technique using a 5×5 window to remove the noise and suppress the undesirable distortion in the images. Then, the K means clustering is used for background subtraction. Secondly, if the positions of most of the faces in both datasets are not aligned properly this can be problematic for the detection of faces in the images. Thus, we set the face alignment of both the Gallagher collection person dataset and the Images of Groups dataset using the code available on GitHub [19].

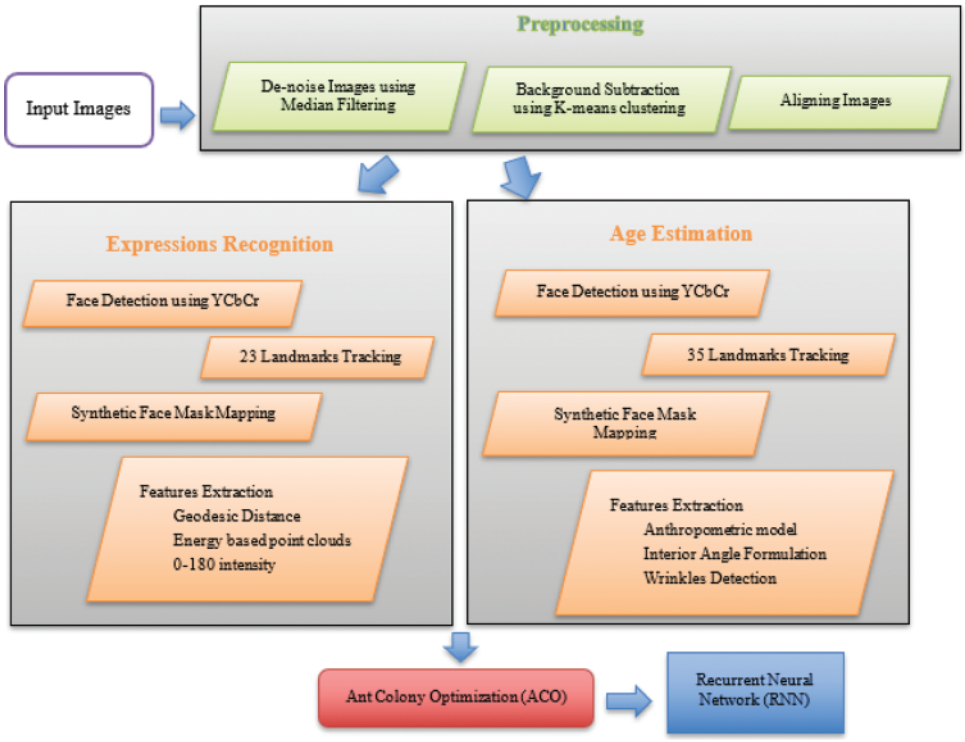


Figure 1: System architecture of the proposed FERAE system

For face detection, a skin color segmentation technique YCbCr is used. This skin color segmentation model provides remarkable results to detect faces in a scene using YCbCr color space. The skin color of each individual varies, so to get full coverage of each skin pixel the RGB images are converted to YCbCr color space to easily distinguish skin from non-skin pixels. Fig. 2 shows the examples of face detection in the Images of Groups dataset. This technique is not affected by the illumination condition Y(luma) factor. Skin representation is based on two components *Cb* (blue difference) and *Cr* (red difference). The skin color model is formulated as in Eqs. (1) and (2) [2].

$$Y_{lum} = 0.299\mathrm{R} + 0.287\mathrm{G} + 0.11\mathrm{B} \tag{1}$$

$$Cr = R - Y_{lum},\ Cb = B - Y_{lum} \tag{2}$$

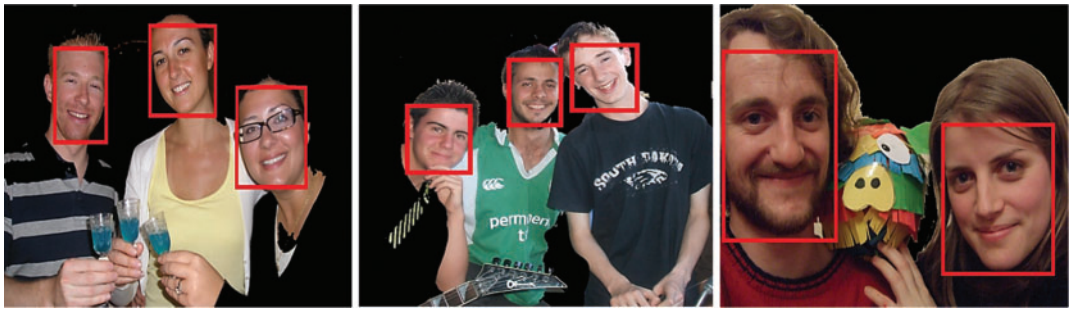


Figure 2: Some results of pre-processing and face detection over the Images of the group dataset

3.2 Landmarks Tracking

Landmarks tracking is the primary step towards face mask mapping. The landmarks are plotted on the facial features to track the pixel positions. They will help us extract different point-based features for the accurate classification of multi-face expressions and age. This section is divided into two subsections; Section 3.2.1 explains landmarks tracking over the Gallagher benchmark dataset for multi-face expressions recognition and Section 3.2.2 describes landmarks tracking over the Images of groups dataset for multi-face age estimation.

3.2.1 Landmarks Tracking for Multi-face Expressions Recognition

To plot the landmarks over the Gallagher collection person dataset, the same procedure is used for marking the landmarks on eyebrows, eyes, and lips as mentioned in Section 3.3.1. For the localization of landmarks on the nose. First, the nose is detected using a cascade algorithm. The two nostril points are obtained by applying the concept of connected components inside the bounding box. Then, 3 points are obtained, one on the nose tip and two are on the nostrils. Therefore, a total of 23 landmarks are plotted on the entire face. Figs. 3a and 3b show the landmark points symmetry over both benchmark datasets, respectively.

(a) (b)

Figure 3: Landmark points symmetry over the (a) Gallagher collection person dataset and (b) the Images of groups dataset, respectively

3.2.2 Landmarks Tracking for Multi-face Age Estimation

After detection of the face, 35 landmarks are plotted on the face, on the eyebrows, eyes, and lips, by converting the RGB image into a binary image and detecting the facial features using blob detection. The edges of each facial feature blob are marked with landmarks by taking the central point of each edge using Eq. (3). The nose is detected using the ridge contour method and a total of seven landmark points are marked on the nose. To plot the area of the chin, jawline, and forehead, the midpoints of the face blob or bounding box edges are marked and these are calculated using Eq. (4) [2];

$$e_1 = \frac{a}{2};\ e_2 = \frac{b}{2};\ e_3 = \frac{c}{2};\ e_4 = \frac{d}{2} \tag{3}$$

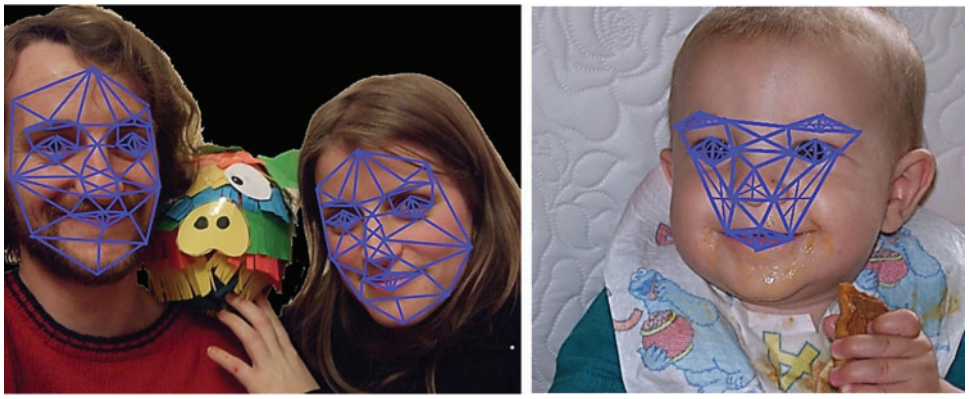
$$e_n = \frac{(e_i + e_j)}{2} \tag{4}$$

where a, b, c, d denotes the edges length and the $e_1, e_2, e_3,$ and e_4 are the midpoints of the blob edges.

3.3 Synthetic Face Mask Prediction

Synthetic mask prediction is a robust technique for the accurate prediction of the multi-face age of an individual and to recognize the expressions or emotions of a person. This technique is widely used for face detection, face recognition, face aging estimation, etc. For the generation of synthetic masks on the face, we utilized the 35 landmark points for age estimation and 23 landmarks for multi-face expressions recognition. The technique used for both the masks is the same, i.e., three-sided polygon meshes and perpendicular bisection of a triangle are applied [15]. However, the synthetic mask is only generated on facial features for multi-face expression recognition using the sub-triangle formation. The main variations appear on the facial features during changes in facial expressions. Algorithm 1 describes the overall procedure of synthetic face mask prediction over the Gallagher collection person dataset for multi-face expressions recognition.

Given a face image with 35 and 23 landmarks points over the Images of the Group dataset and the Gallagher collection person dataset images respectively, a multivariate shape model is generated using the landmarks points via polygon meshes and the perpendicular bisection of triangles for age estimation and expression recognition, the large triangles, and sub-triangle formation rule is used. The perpendicular bisections help us distinguish the changes occurring from infancy till adulthood where triangular meshes will further help to extract features for both multi facial expressions and age estimation. Figs. 4a and 4b show the synthetic mask prediction over the Images of the Group dataset and the Gallagher collection person dataset, respectively.



(a) (b)

Figure 4: Synthetic face mask prediction over (a) the Images of Groups dataset for age estimation and over (b) the Gallagher collection person dataset for expression recognition respectively

Algorithm 1: Multi-face expression recognition synthetic mask prediction

Input: Input X=Position of 23 landmarks localization points;
Output: Mesh of triangles of Y:TM(Y);
//initiating feature descriptors matrix//
begin
1 Calculate the pixel positions of three outer corners of a triangle (c1, c2, c3);
2 **for**
3 TM(Y):= (c1, c2, c3);

(Continued)

Algorithm 1: Continued

```
/* Initialize TM(Y) a large triangle*/
The sub-triangles formed inside the TM(Y) is S:=(s1, s2, s3);
do
c1A bisects ∠c_1;
c2B bisects ∠c_2;
c3C bisects ∠c_3;
 end for
 TM(Y)←S;
 return TM(Y);
 end
```

3.4 Feature Descriptors

For the estimation of age and the accurate recognition of facial expressions, we have extracted the age and expression features individually. For age group prediction, the features extraction methods include; 1) Anthropometric model, 2) Interior Angles formulation and, 3) Wrinkles detection (See Section 3.4.1). For expressions recognition, the extracted features are, 1) Geodesic distance, 2) Energy-based point clouds, and 3) 0–180° intensity (See Section 3.4.2).

3.4.1 Feature Extraction for Age Group Classification

The anthropometric model is the study of the human face and facial features by dimensions and sizes [20]. The landmark points marked on the facial features are known by anatomical names e.g., the lip corners are known as the left and right cheilion and are denoted by lch and rch, likewise, the inner corners of the eyebrows are known as Nasion and are denoted by n. By using this model, we have taken several distances between the facial features which are calculated using the Euclidean distance using Eq. (5) [21].

$$dis = \sqrt{p_1 - q_1)^2 - (p_2 - q_2)^2} \tag{5}$$

where p_1, p_2, q_1 *and* q_2 are the pixel locations along x and y coordinates, respectively. Fig. 5 shows the anatomical names and calculated dimensions.

For the calculation of the interior angles, the above-mentioned face mask is used. From infancy to adulthood the shape of the face mask changes and this results in the variations of angles. We calculated the interior angles θ_1, θ_2, θ_3 using the law of cosine in Eqs. (6), (7) and (8) [21];

$$\theta_1 = cos^{-1}\left(\frac{p^2 + q^2 - r^2}{2pq}\right) \tag{6}$$

$$\theta_1 = cos^{-1}\left(\frac{p^2 + r^2 - q^2}{2pr}\right) \tag{7}$$

$$\theta_1 = cos^{-1}\left(\frac{r^2 + q^2 - p^2}{2rq}\right) \tag{8}$$

where p, q and r are the sides of the triangles formed by the face mask. Different measurements of interior angles on two different age groups are shown in Fig. 6.

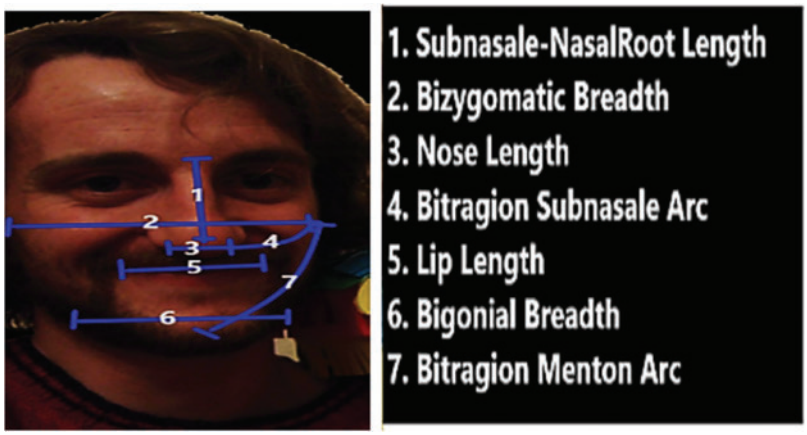


Figure 5: Anatomical names of the given dimensions

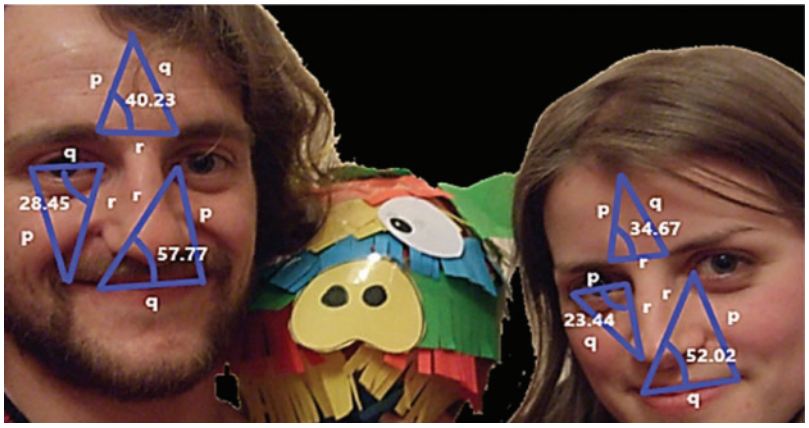


Figure 6: Interior angle formulations over the Images of groups dataset

With time, human skin texture changes due to environment, stress, health issues, and many other factors. This texture variation appears in the form of wrinkles, under-eye bags, sagging skin, etc. For wrinkle detection, the Canny edge detection method is used. In Fig. 7, the winkles are displayed in the form of edges, i.e., the white pixels in the binary image over the Images of groups dataset. The quantity of the edges is equal to the number of wrinkles on the face which exhibits the age of the person. These wrinkles are calculated using the Eq. (9) [22];

$$WF = \left(\frac{F}{T1} + \frac{LE}{T2} + \frac{RE}{T3} + \frac{UE}{T4} + \frac{AL}{T5}\right) \tag{9}$$

where *F, LE, RE, UE* and *AL* are the white pixels, i.e., *T1, T2, T3, T4* and *T5* are the total number of pixels on the forehead, left-eyelid, right-eyelid, under-eyes and around the lips.

Figure 7: The results of wrinkles formation over the images of groups dataset

3.4.2 Feature Extraction for Facial Expressions Recognition

The geodesic distance on the surface of the face is the shortest distance between two points. To calculate the geodesic distance, Kimmel and Sethian proposed a method known as fast marching using the Eikonal equation as in Eq. (10) [23];

$$|\nabla\_x(u)| = F(u); \quad u \in \Omega \tag{10}$$

where ∇ denotes the gradient and $|.|$ is the Euclidean norm.

The fast-marching algorithm is based on Dijkstra algorithm which computes the shortest distances between two points. In this work, we calculate geodesic distances on the surface of the face using the values of gradient only. *Img* is an image having multiple landmark points. To calculate the geodesic distance between two landmark points the distance is $(D = d1, d2)$. The geodesic distance is taken as the parametric manifold which can be represented by mapping $F: R^2 \rightarrow R^3$ from the parameterization P to the manifold which is given as in Eq. (11);

$$F(D) = F(d_1, d_2) = (d_1, \ d_2, d_3(d_1, d_2)) \tag{11}$$

The metric tensor g_{ij} of the manifold is given as in Eq. (12);

$$g_{ij} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} X.X & X.Y \\ Y.X & Y.Y \end{bmatrix} \tag{12}$$

The geodesic distance is calculated as the shortest distance between the two points. We can calculate the geodesic distance on the surface of the face between 15 landmark points. The geodesic distance $\delta_{(A,B)}$ between two points A and B is calculated as in Eq. (13);

$$\delta_{A,B} = min\gamma(\beta(A, B)) \tag{13}$$

The distance element on the manifold is given as in Eq. (14);

$$\delta_{c,d}\sqrt{g_{c,d^{A,B}}} \tag{14}$$

where the values of c and d are 1 and 2. We can compute the geodesic distance between 15 landmark points and select the most significant distance that helps in expression recognition. As a result, we obtained a total of 15 distances.

Energy-based point clouds are the techniques that work on the principle of the Dijkstra algorithm. According to our best knowledge, this technique is used for the very first time for age estimation and expression recognition, simultaneously. This technique is efficient, robust, and quite simple to implement. Using this technique, a central landmark point labeled $f \epsilon F$ is marked at the center of

the face. Its distance is fixed to zero, i.e., $d(f) = 0$. After that, this value is inserted into a priority queue Q, where the priority is based on the smallest distance between the landmark points. The remaining points are marked as $d(q) = \infty$. In the priority queue, one point f is selected then the shortest distances between that point to other points are calculated based on the Dijkstra algorithm. Based on those distances, energy-based point clouds are displayed on the face. The alignment of these point clouds changes with variations in the distances from the central point to the other landmark points. The distances from the central point to other varying landmark points are known as optimal distances [24]. Fig. 8 shows the hierarchical steps for energy-based point clouds extraction.

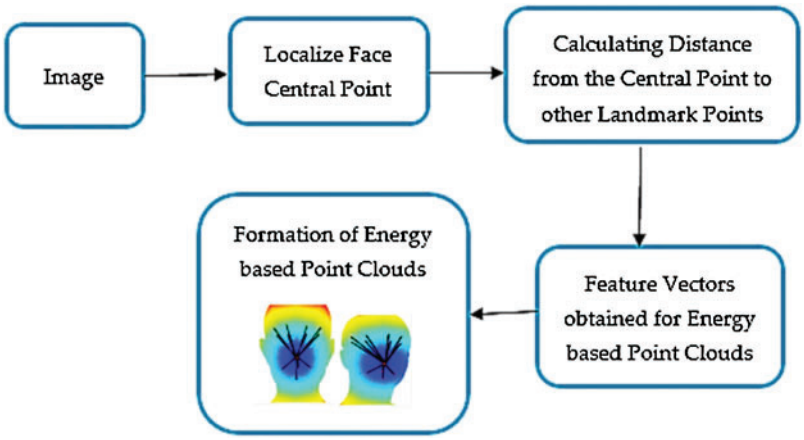


Figure 8: The hierarchical steps for energy-based point clouds extraction

In the 0–180° intensity feature extraction technique, radon transform calculates the projection of an image matrix with some specific axis. The specific axis is used to predict the 2D approximation of the facial expression through different parts of the face using the intensity estimation q along with the specific set of radial line angles θ defined as in Eq. (15) [25];

$$I(q,\ \theta) = \int_{-\infty}^{\infty} f(a, b) \tag{15}$$

where $I(q,\theta)$ is the line integral of the image intensity and $f(a,\ b)$ is the distance from the origin at angle θ of the line junction. All these points on a line satisfy Eq. (8) and the projection function can be rewritten as Eqs. (16) and (17) [25];

$$s = a * sin\theta - b * cos\theta \tag{16}$$

$$I(q, \theta) = \int_{-\infty}^{\infty}\int f(a,\ b)\delta(s - acos\theta - bsin\theta)dadb \tag{17}$$

Finally, we extracted the top 180 levels of each pixel's intensity and combined them into a unified vector for different facial expressions. Fig. 9 shows the different expression intensity levels (0–180°).

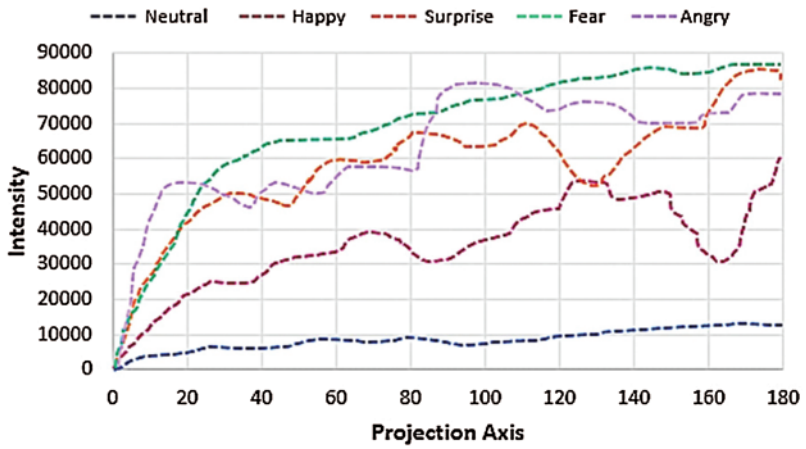


Figure 9: 0–180° intensity levels for different expressions over the Gallagher collection person dataset

3.5 Long Short-Term Memory Based Recurrent Neural Network (RNN-LSTM)

Variations in the facial features, while expressions are changing, can exhibit various positions of the facial features. For instance, in the state of sadness, an individual has drooping eyelids, the outer corners of the lips are pulled in a downward direction and very slow eye blinking occurs. In a state of happiness, fast eye blinking and movement of the cheek muscles around the eyes occur and puffiness occurs under the eyes. By comparison with the state of anger, the eyes open widely, eyebrows are drawn together and the lips are tightly closed and become narrower and thin, or the lips are opened to form a rectangle. Similarly, for the prediction of accurate age group classification, changes in facial textures and features occur. In childhood, an individual has more tight skin, no wrinkles on the face, and no under-eye puffiness, whereas, in adulthood, more wrinkles are formed around the eyes, lips, and cheeks, sagging skin and skin color varies. These feature and texture variations are extracted in the form of feature vectors and the Recurrent Neural Network (RNN) takes advantage of them for accurate classification of multi-face expressions and age.

The feature vectors of expressions and age are fed to the RNN classifier after the features extraction and optimization stage. Our RNN uses one hidden layer along with the 210 unidirectional LSTM fully interconnected cells. The input layer is comprised of 5080 images of the Images of Groups dataset and 589 images of the Gallagher collection person dataset. The vectors size of the Images of groups dataset is $28{,}231 \times 550$ and for the Gallagher collection person dataset it is 931×623 Each features vector is the depiction of the participant facial expressions and age. At the output layer, a SoftMax function, which is responsible for a 1 out K classification task, is used. The SoftMax function output range lies between 0 and 1 where the sum is equal to 1 at every time step. The RNN is trained using the Adam Optimizer with a learning rate of 0.001 [26]. Fig. 12 depicts the hierarchy of RNN for age and expression classification. Algorithm 2 defines the RNN_LSTM training for age estimation and expression recognition.

Algorithm 2: RNN-LSTM Training

Input: Classes ← { “7”,“6”};
Features ←{“Age Estimation”, “Expression Recognition”} ;
Output: A ← dataset {n}. Values;
B ← dataset {Features}. Values;

```
Train_Data, Test_Data,Valid_Data ← Split_Data_Train_Test (A, B, 0.33, 0.25);
Size_of_Batch← 4;
RNN_LSTM← Sequential_Model({
Embedded_Layer (Train_Data.Length, Output_Data_Length, Train_Data.Columns),
RNN_LSTM_Layer (Output_Data_Length),
Dense_Layer (Output_Data_Length, activation_Function=‘Sigmoid’)});
Optimizer←Adam, Epochs← 20;
RNN_LSTM.Compile (Optimizer);
RNN_LSTM.train(Train_Data, Epochs, Size_of_Batch, Valid_Data);
```

4 Performance Evaluation

This section gives a brief description of two datasets used for facial expressions recognition and age estimation, results of experiments conducted to evaluate the proposed FERAE system and comparison with other systems.

4.1 Datasets Description

The description of each dataset used in FERAE system is given in Sections 4.1.1, 4.1.2 and 4.1.3.

4.1.1 The Gallagher Collection Person Dataset for Expression Recognition

The first dataset used for multi-face expression recognition is the Gallagher Collection Person dataset [27]. The images in this dataset were shot in real life at real events of real people with real expressions. The dataset comprises 589 images with 931 faces. Each face is labeled in the image with an expression of Neutral, Happy, Sad, FERAE, Angry and Surprise. The dataset is publicly available. Some examples from this dataset are shown in Fig. 10.

Figure 10: Some examples from the Gallagher collection person dataset

4.1.2 Images of Groups Dataset for Age Estimation

The second dataset is the Images of Groups dataset which is used for multi-face age group classification [28]. The dataset is the largest dataset comprising 5080 images containing 28,231 faces that are labeled with age and gender. The seven-age group labels of this dataset are 0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+. This dataset is publicly available. Some examples of this dataset are shown in Fig. 11.

Figure 11: Some examples from the images of groups dataset

4.2 Experimental Settings and Results

All the processing and experimentation are being performed on MATLAB (R2019). The hardware system used is Intel Core i5 with 64-bit Windows-10. The system has 16 GB and 5 (GHz) CPUs. To evaluate the performance of the proposed system, we used a Leave One Person Out (LOPO) [29] cross-validation method. Experiment 1 determined the facial features detection accuracy rates over both benchmark datasets. Experiment 2 determined the multi-face expressions recognition accuracy rates as shown in the form of a confusion matrix. Experiment 3 determined the multi-face age estimation accuracy rates over the Images of groups dataset. Experiments 4 reveal comparisons in a ROC curve graph of the proposed model with another state-of-the-art models for both multi-face expression recognition and age estimation, respectively.

4.2.1 Experiment 1: Facial Features Detection Accuracies

In this experiment, facial features detection accuracies over the Images of Groups dataset and Gallagher collection person dataset were determined as shown Fig. 12.

4.2.2 Experiment 2: Multi-face Expressions Recognition Accuracy

For multi-face expression recognition, the RNN model is used for the accurate classification of expression and age. The Leave One Subject Out (LOSO) cross validation technique is used for the evaluation of the proposed system. Tab. 1 shows the confusion matrix of multi-face expressions recognition.

4.2.3 Experiment 3: Multi-face Age Estimation Accuracy

For multi-face age estimation, the RNN model was used for the accurate classification of age. The Leave One Subject Out (LOSO) cross validation technique was used for the evaluation of proposed system. Tab. 2 shows the confusion matrix for multi-face age estimation.

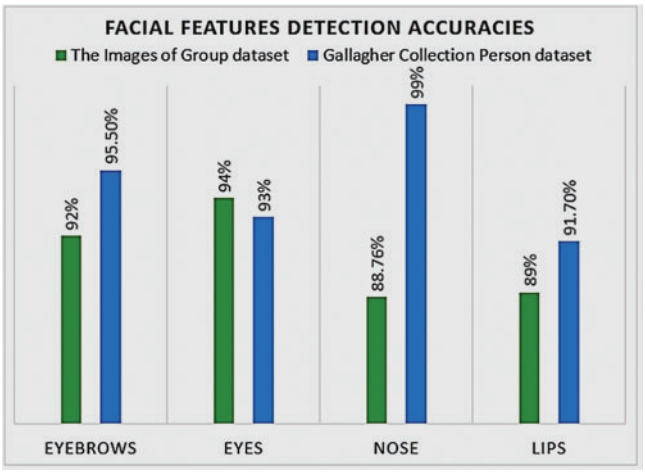


Figure 12: Facial features detection accuracies over both benchmark datasets

Table 1: Confusion matrix for multi-face expressions recognition over the Gallagher person collection dataset

Classes	Neutral	Happy	Sad	Fear	Angry	Surprise
Neutral	**0.894**	0.048	0.033	0.016	0.009	0
Happy	0.064	**0.924**	0	0	0	0.012
Sad	0.059	0.002	**0.80**	0.086	0.053	0
Fear	0.058	0	0	**0.76**	0.135	0.047
Angry	0.017	0.061	0	0.027	**0.803**	0.092
Surprise	0	0.005	0.001	0.015	0.025	**0.954**
Mean Accuracy= 85.5%						

Table 2: Confusion matrix for multi-face age estimation over the images of groups dataset

Classes	0–2	3–7	8–12	13–19	20–36	37–65	66+
0–2	**0.92**	0.08	0	0	0	0	0
3–7	0.04	**0.95**	0.01	0	0	0	0
8–12	0	0.04	**0.89**	0.07	0	0	0
13–19	0	0	0.05	**0.90**	0.05	0	0
20–36	0	0	0.01	0.10	**0.88**	0.01	0
37–65	0	0	0	0	0	**0.93**	0.07
66+	0	0	0	0	0	0.07	**0.93**
Mean Accuracy= 91.4%							

4.2.4 Experiment 4: Results for Comparison of the Proposed Multi-expressions Recognition and Age Estimation Model with Other State of the Art Models.

Figs. 13a–13f, 14a–14f shows the ROC curve graph for all multi-facial expressions and age estimation. The ROC curve is the relationship between the true positive rate and the false positive rate. The true positive is basically showing the sensitivity and false positive rate is showing the 1-specificity. Both the true positive and false positive can be calculated in Eqs. (18) and (19) respectively;

$$True\_Positive = \frac{True\_Positive}{True\_Positive + \ False\_Negative} \tag{18}$$

$$False\_Positive = \frac{False\_Positive}{False\_Positive \ + True\_Negative} \tag{19}$$

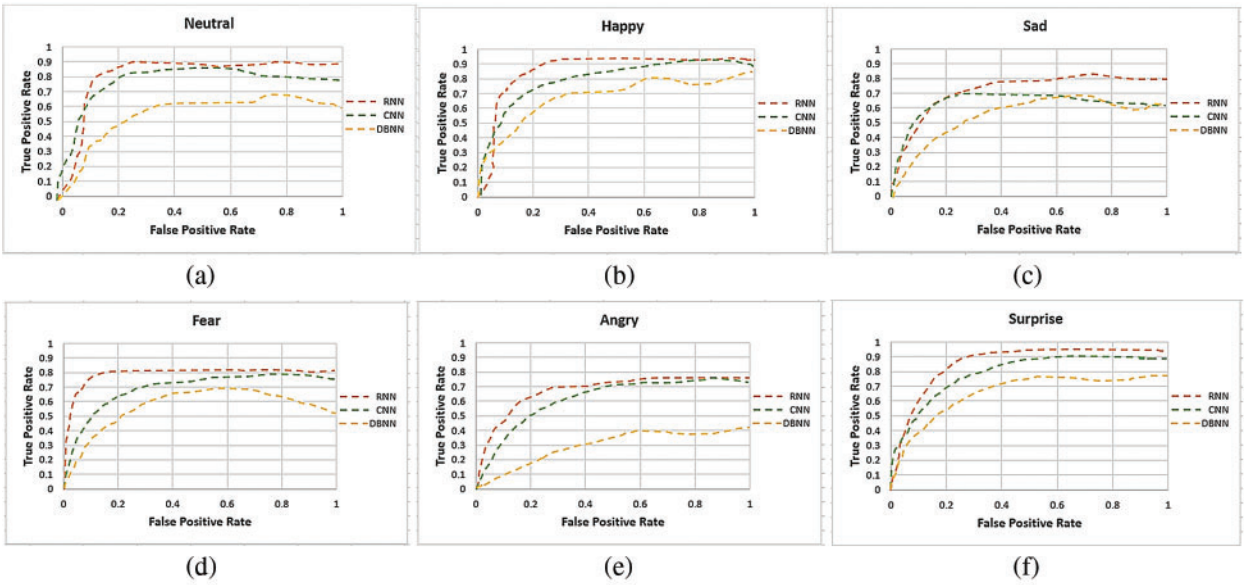


Figure 13: The ROC curve graphs for all multi-facial expressions over the Gallagher collection person dataset. The lowest and highest values in the expressions ROC curve graphs of both the true positive and false positive using RNN are; **Neutral:** (0.03, 0.00) and (0.80, 1.00), **Happy:** (0.02, 0.02) and (0.92, 0.98), **Sad:** (0.14, 0.027) and (0.80, 1.00), **Fear:** (0.10, 0.00) and (0.81, 0.98), **Angry:** (0.09, 0.01) and (0.77 and 1.00) and **Surprise:** (0.12, 0.03) and (0.93, 0.98)

We have tested our multi-facial expression recognition and age estimation system (FERAE) model using the state-of the art methods i-e., Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and Deep Belief Neural Network (DBNN). Experimental Results 4 shows that the RNN along with the other salient feature descriptors of both expression and age provides better results against CNN and DBNN.

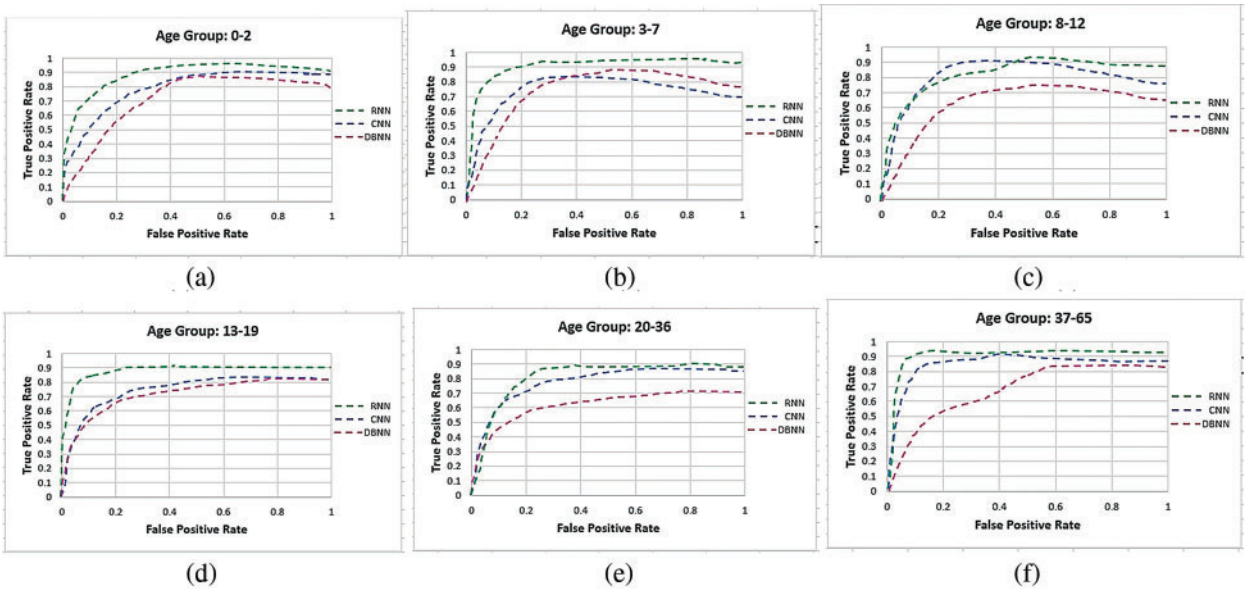


Figure 14: The ROC curve graphs for all the age groups over the Images of groups dataset. The lowest values and highest values in the age groups ROC curve graphs of both the true positive and false positive using RNN are; **0–2**: (0.00, 0.00) and (0.90, 1.00), **3–7**: (0.18, 0.01) and (0.92, 1.00), **8–12**: (0.00, 0.00) and (0.89, 0.99), **13–19**: (0.10, 0.00) and (0.90, 1.00), **20–36**: (0.00, 0.00) and (0.98 and 0.99) and **37–65**: (0.09, 0.01) and (0.91, 0.97)

5 Conclusion

In this paper, a fused model of multi facial expressions recognition and age estimation is proposed. A synthetic face mask is mapped on the face formed by the localization of the landmarks points. The novel point-based and texture-based features obtained using different feature extraction techniques are passed to the RNN classifier for the classification of expressions and age groups The proposed system is tested using the Gallagher collection person dataset for expression recognition and the Images of groups dataset for age estimation. Experimental results show that our approach produced superior classification accuracies i-e., 85.5% over the Gallagher collection person dataset and 91.4% over the images of groups dataset. The proposed system applies to surveillance systems, video gaming, consumer applications, e-learning, audience analysis, and emotion robots. As for limitations, the system fails to detect the detailed facial features of persons from images that are captured too far from the cameras. In the future, we will work on the computational time complexity of the system and also evaluate our system on RGB-D datasets.

Acknowledgement: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2018R1D1A1A02085645). Also, this work was supported by the Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 202012D05-02).

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] N. Kauser and J. Sharma, "Facial expression recognition using LBP template of facial parts and multilayer neural network," in *Proc. Int. Conf. on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, pp. 445–449, 2017.

[2] S. Rizwan, A. Jalal and K. Kim, "An accurate facial expression detector using multi-landmarks selection and local transform features," in *Proc. Int. Conf. on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, pp. 1–6, 2020.

[3] H. Basavegowda and G. Dagnew, "Deep learning approach for microarray cancer data classification," *CAAI Transactions on Intelligence Technology*, vol. 5, pp. 22–33, 2020.

[4] S. Tahir, A. Jalal and K. Kim, "Wearable inertial sensors for daily activity analysis based on adam optimization and the maximum entropy markov model," *Entropy*, vol. 22, pp. 579, 2020.

[5] A. Jalal, N. Khalid and K. Kim, "Automatic recognition of human interaction via hybrid descriptors and maximum entropy markov model using depth sensors," *Entropy*, vol. 22, pp. 817, 2020.

[6] M. Javeed, A. Jalal and K. Kim, "Wearable sensors-based exertion recognition using statistical features and random forest for physical healthcare monitoring," in *Proc. Int. Bhurban Conf. on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, pp. 512–517, 2021.

[7] R. Jiang, X. Mou, S. Shi, Y. Zhou, Q. Wang *et al.,* "Object tracking on event cameras with offline–online learning," *CAAI Transactions on Intelligence Technology*, vol. 5, pp. 165–171, 2020.

[8] A. Ahmed, A. Jalal and K. Kim, "A novel statistical method for scene classification based on multi-object categorization and logistic regression," *Sensors*, vol. 20, pp. 3871, 2020.

[9] A. Jalal, I. Akhtar and K. Kim, "Human posture estimation and sustainable events classification via pseudo-2D stick model and K-ary tree hashing," *Sustainability*, vol. 12, pp. 9814, 2020.

[10] N. Khalid, M. Gochoo, A. Jalal and K. Kim, "Modeling Two-person segmentation and locomotion for stereoscopic action identification: A sustainable video surveillance system," *Sustainability*, vol. 13, pp. 970, 2021.

[11] K. Fujii, D. Sugimura and T. Hamamoto, "Hierarchical group-level emotion recognition in the wild," in *Proc. Int. Conf. on Automatic Face & Gesture Recognition (FG 2019)*, Lille, IEEE, pp. 1–5, 2019.

[12] Q. Wei, Y. Zhao, Q. Xu, L. Li, J. He *et al.,* "A new deep-learning framework for group emotion recognition," in *Proc. Int. Conf. on Multimodal Interaction*, Glasgow, ACM, pp. 587–592, 2017.

[13] X. Guo, L. Polanía and K. Barner, "Group-level emotion recognition using deep models on image scene, faces, and skeletons," in *Proc. Int. Conf. on Multimodal Interaction*, Glasgow, ACM, pp. 603–608, 2017.

[14] A. Dey and K. Dasgupta, "Mood recognition in online sessions using machine learning in realtime," in *Proc. Int. Conf. on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, India, IEEE, pp. 1–6, 2021.

[15] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan *et al.,* "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp. 22492–22503, 2017.

[16] S. Bekhouche, A. Ouafi, A. Benlamoudi and A. T. Ahmed, "Automatic Age estimation and gender classification in the wild," in *Proc. Int. Conf. on Automatic Control, Telecommunications and Signals (ICATS15)*, Annaba, Algeria, pp. 16–18, 2015.

[17] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Boston, MA, USA, pp. 7–12, 2015.

[18] E. Eidinger, R. Enbar and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 2170–2179, 2017.

[19] H. Ansar, A. Jalal, M. Gochoo and K. Kim, "Hand gesture recognition based on auto-landmark localization and reweighted genetic algorithm for healthcare muscle activities," *Sustainability*, vol. 13, pp. 2961, 2021.
[20] T. Zhou, W. Wang, S. Qi, H. Ling and J. Shen, "Cascaded human-object interaction recognition," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 4262–4271, 2020.
[21] W. Lee, B. Lee, X. Yang, H. Jung, I. Bok *et al.,* "A 3D anthropometric sizing analysis system based on north American CAESAR 3D scan data for design of head wearable products," *Computers and Industrial Engineering*, vol. 117, pp. 121–130, 2018.
[22] R. Jana and A. Basu, "Automatic Age estimation from face image," in *Proc. Int. Conf. on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bangalore, India, pp. 87–90, 2017.
[23] R. Ahdid, S. Said, M. Fakir, B. Manaut, Y. Ouadid *et al.,* "Three-dimensional face surface recognition by geodesic distance using jacobi iterations," in *Proc. Int. Conf. on Computer Graphics, Imaging and Visualization*, Marrakesh, pp. 44–48, 2017.
[24] E. Treister and E. Haber, "A Fast-marching algorithm for the factored eikonal equation," *Journal of Computational Physics*, vol. 324, pp. 210–225, 2016.
[25] M. Singh, M. Mandal and A. Basu, "Pose recognition using the radon transform," in *Proc. on Int. Midwest Symp. on Circuits and Systems*, Covington, KY, pp. 1091–1094, 2005.
[26] A. Mostafa, M. I. Khalil and H. Abbas, "Emotion recognition by facial features using recurrent neural networks," in *Proc. Int. Conf. on Computer Engineering and Systems (ICCES)*, Cairo, IEEE, pp. 417–422, 2018.
[27] A. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," in *2008 IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE: Anchorage, AK, pp. 1–8, 2008.
[28] A. Gallagher and T. Chen, "Understanding images of groups of people," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, IEEE, pp. 256–263, 2009.
[29] S. Rizwan, A. Jalal, M. Gochoo and K. Kim, "Robust active shape model via hierarchical feature extraction with sfs-optimized convolution neural network for invariant human age classification," *Electronics*, vol. 10, pp. 465, 2021.