Real Time Customer Satisfaction Analysis using Facial Expressions and Headpose Estimation

Nethravathi P. S[1]
Faculty, Institute of Engineering & Technology, Srinivas University, Mangalore, India

Manjula Sanjay Koti[2]
Faculty, Dept. of MCA, Dayananda Sagar Academy of Technology & Management, Karnataka, India

Taramol. K.G[3]
Faculty, School of Business, Manipal Academy of Higher Education, Dubai International Academic City, Dubai

Soofi Anwar[4]
Faculty of Business with Accounting, University of Stirling, Al Dhait South, Ras Al Khaimah, United Arab Emirates

Gayathri Babu J[5]
Faculty, Department of Business Administration, Shree Devi Institute of Technology, Mangalore, India

Rajermani Thinakaran[6]
Faculty of Data Science and Information Technology, INTI International University, Nilai, Negeri Sembilan, Malaysia

***Abstract*—One of the most exciting, innovative, and promising topics in marketing research is the quantification of customer interest. This work focuses on interest detection and provides a deep learning-based system that monitors client behaviour. By assessing head position, the recommended method assesses customer attentiveness. Customers whose heads are directed toward the promotion or the item of curiosity are identified by the system, which analyses facial expressions and records client interest. An exclusive method is recommended to recognize frontal face postures first, then splits facial components that are critical for detecting facial expressions into iconized face pictures. Mainly consumer interest monitoring will be executed. Finally, the raw facial images are combined with the iconized face image's confidence ratings to estimate facial emotions. This technique combines local part-based characteristics through holistic face data for precise facial emotion identification. The new method provides the dimension of required marketing and product findings indicate that the suggested architecture has the potential to be implemented because it is efficient and operates in real time.**

Keywords*—Customer monitoring; convolutional neural network; facial expression recognition; facial analysis; head pose estimations component; CNN Model; object localization; face boosting

I. Introduction

The usual way is for a salesperson to study client behavior during the shopping phase or advertisement watching and then recall customer interest. However, every salesperson needs special talents for this job, and each spectator may interpret consumer behavior differently. In this aspect, only a few extraordinarily tactful and competent salespeople can have good salesperson-customer interactions [1]. According to [2], subjective emotional perception-based approaches may not always represent the human emotional state appropriately. On the other hand, automatic measurements provide a more exact and dependable result. As a result, developing non-invasive, objective, and quantifiable measures for tracking client interest is crucial.

Human choices can be analyzed in various ways such as brain images [3]; an electroencephalogram also known as EEG [4], [5]; eye tracking [6], [7]; heart rate registration [8]; and other approaches have been a recent topic in the existing literature. Customer behavior classification [9 - 11] and customer face analysis studies have also been used in several studies [12], [13].

Estimating a client's visual focus of attention is one approach of evaluating their interest. Head posture is quantified in research on visual center of attention [14 - 16]. Recognizing consumer sentiments for advertising purposes is also a difficult and quickly growing academic area [17]. The intuitive decision-making process is substantially influenced by one's mood [17]. People who are in a good mood assume that everything is well and that they are safe in their surroundings. When they're in a poor mood, though, they believe things aren't going well and that an incident is approaching, and needing their attention [18]. Marketers must consider their customers' emotions and moods [19]. Knowing mental status of the buyer helps marketer in creating good business [19]. According to [20], annoyance, anxiety, sorrow, and disgust are all bad feelings, but happiness is a good emotion. Because it's difficult to discriminate between good and negative emotions, surprise isn't mentioned. According to [21], happy clients are positive, confident, passionate, stimulated, and thrilled.

The goal of this study is to develop a deep learning-based method for tracking client interest that relies on head-pose alignment and facial expression identification. The recommended method identifies the visual center of attention by first recognizing the human face and then analyzing the head posture orientation because the camera is an important item, the frontal faces suggest the visual center of focus. If the observed face is concentrated on the advertisement or product of interest, the algorithm begins to recognize the facial expression.

Customers' facial expressions are captured by the equipment over a period of time that may be assessed by specialists. Based on the gathered facial expressions, the system decides if the customer is in a good or poor mood. Based on the gathered data the system concludes he/she is in good or poor mood. The image can be captured by the camera and used to estimate face expression detection. As a result, determining client interest may be done non-invasively, quantitatively, and at a minimal cost. The suggested approach might be beneficial for identifying marketing campaigns and other corporate initiatives that clients would be interested in. Salespeople can also make changes to their marketing materials based on client feedback. This software could also help salespeople respond more effectively to customer emotions, leading to higher satisfaction.

The suggested study's main contributions can be stated in four points:

1) A system is proposed for measuring customer interest that is non-invasive, objective, and quantifiable.

2) Because the proposed system does not have the feature to save customer facial photos, personal privacy is protected. The system processes current client photos and does not require them to be saved for later processing. If there is a need to keep track of a customer's facial expressions, de-identified iconized face photos can be saved. Face expression data is included in iconized face photos while maintaining personal privacy.

3) A three-cascade Convolutional Neural Network (CNN) model is proposed which performs multiple tasks. In the third part CNN combines raw face images with iconized image confidence values. Part-based information is included in the confidence values of iconized images, whereas holistic details are included in raw face images. Improved facial expression detection is possible thanks to the combination of part-based and holistic data.

4) The system recognises and localises the face components that are crucial for facial expression detection in the facial component segmentation step, resulting in an ionised image. The CNN permits directed training at the facial emotion recognition stage by compelling earlier layers of the architecture to learn to identify and locate the essential face components using the confidence values of that ionized picture as input.

Consumer interest monitoring using CNN is performed for all mentioned three steps. Detecting frontal faces, classifying facial expressions, head posture estimation and facial expression detection are performed.

The next parts, which are organized as follows, give further information. The literature evaluation of current state-of-the-art methods that are relevant to the proposed methodology is summarized in Section II. Section III delves into the suggested technique, while Section IV covers the results, analysis, arguments, and inferences, while Section V wraps up the paper by summarizing some of the work's future directions.

II. Literature Survey

Several publications in the literature describe how to use images or videos to solve real-world problems. For video processing, speed is critical, and many frameworks are being developed to improve it [22 - 24]. Real-world challenges include object identification [25], text detection [26], [27], facial expression recognition [28], head position estimation [15], and so on. This section looks into the work's two key components: facial expression recognition and head posture estimation.

A. Facial Expression Recognition

Avatar animation [29]; smart environments [30 - 33]; robotics [34]; medical [35]; traffic [36], [37]; and human–computer interaction [38 - 44] are some of the applications of automated facial expression recognition. Ekman and Friesen's six universal facial expressions, namely disgust, happiness, fear, wrath, sadness, and surprise, were commonly utilised in automated facial expression detection experiments in their early research [45].

Geometric and appearance-based algorithms for facial expression recognition have been identified [46]. The features derived from positional correlations between facial components focus on geometric-based approaches [46]. Appearance-based features determine face texture [47 - 49]. Histogram of oriented gradients (HOG) [50], principal component analysis (PCA) [51], local binary pattern (LBP) operator [52], and other appearance-based methods have been used for facial emotion identification.

In the field of face expression analysis, the use of machine learning approaches specifically the deep learning is a recent trend. To extract just particular properties for expressions and examine the six essential expressions, Lopes et al. [53] used a CNN network with picture pre-processing techniques such as image rotation, face cropping, and intensity normalization. Pitaloka et al. [54] analyzed six essential sentences using a CNN. They used many data normalization techniques, scaling, face detection, cropping, and resizing algorithms. Matsugu et al. [55] recommended that CNNs be used to give a rule-based technique for detecting smiles and faces.

B. Head Pose Estimation

Visual surveillance [56], [57]; driver attention [58], [59]; the visual focus of attention [15], [60]; and robotics [61] have all been investigated using head posture estimation appearance-based, model-based, manifold embedding, and nonlinear regression techniques are used to create head position prediction systems [62]. Appearance-based strategies compare a new head picture to a set of head posture templates to determine which viewpoint is the most related Appearance-based approaches have the drawback of only being able to predict discrete posture locations [63].

Furthermore, certain templates [63] necessitate long picture comparisons. In model-based strategies, geometric information or facial landmark locations are employed to estimate head position [63]. The amount and quality of geometric signals produced from the image determine the accuracy of model-based techniques. In manifold embedding techniques like PCA [63], dimensionality reduction strategies are used. [64] Estimated head location by projecting images into a PCA subspace and comparing the results to a collection of embedded templates. The problem with manifold embedding is that it can be modified by factors other than location and identity, such as lighting [63]. A labelled training set is used in nonlinear regression algorithms to construct a nonlinear mapping from pictures to postures [62]. According to [62], a dataset with consistency is required to train the parameters a nonlinear regression.

CNNs [62] are a nonlinear regression technique. CNNs give better accuracy in performance for difficult head pose orientations. In [65] described a CNN technique for estimating head posture. Three CNNs make up their network, each of which corresponds to one of three head posture types: yaw, pitch, or roll. To estimate head posture, [66] used a combination of regression models and CNN-based classification. [67] estimated head posture and located landmarks using local and global data collected from a CNN. For head posture estimation, [62] employed CNN and adaptive gradient methods. The literature shows that none of the existing methods used a cascade of the best deep learning architecture in each element and explored how it works. As we already cited, there is substantial evidence that the use of cascaded CNNs has the potential to provide the most optimized and robust results. Hence, in this study, we proposed a three-cascaded CNN architecture to analyze customer satisfaction in real-time.

III. Methodology

Each face picture is trimmed using the Viola and Jones Algorithm [68] to eliminate background information and leave just expression and head pose-specific data. The proposed technique makes use of the Viola and Jones algorithm, which allows for quick feature evaluations while also reducing the complexity of feature detection for each frame [68]. When this coordinate is included, as shown in (1), the integral image at the point x, y comprises the total of the pixels above and to the left of x, y [68].

$$jj\,(x, y) = \sum_{x`<=x,y`<=y} j(x`, y`) \tag{1}$$

A raw picture is j(x,y), while an integral image is jj(x,y). The integral image may be computed in one pass over the raw picture using (2) and (3) [68]. Where a(x, y) stands for the cumulative row total, a(x, -1) stands for zero, and jj(-1, y) stands for zero.

$$a\,(x, y) = a\,(s, y\text{-}1) + j\,(x, y) \tag{2}$$

$$jj\,(x, y) = jj\,(x\text{-}1, y) + s\,(x, y) \tag{3}$$

The recommended system's initial stage determines if the client is looking at the correct advertisement or product. The suggested approach that classifies a head in frontal versus non-frontal profile can use the coarsest level head position estimation. Non-frontal faces are ignored when frontal faces are provided to CNN-2 for facial component segmentation. At 00, 450, 900, 1300, and 1800, the CNN-1 has been taught to estimate head position (Fig. 1).

Fig. 1. Sample Image Snapshots with Different Head Pose Positions.

CNN 1: To modify the backbone network design for the pose estimation issue, we use differentiable neural architecture search (NAS). Differentiable Neural Architecture Search is what we employ (NAS). NAS is formulated as a nested optimization problem. Fig. 2 shows the complete information, including the Differentiable neural architecture search, efficient backbone, efficient head, and Cost optimization.

CNN 2: Faces such as the mouth, eye, and brow areas are separated from the rest of the image by the CNN-2. Face component segmentation is a binary classification problem using the face component and the backdrop. The original raw images and corresponding training masks are partitioned into 16 * 16 non-overlapping blocks before moving to the training step. The majority class is assigned to blocks that contain more than 80% of the face component or backdrop pixels. The remaining mixed-class blocks are omitted during training. According to our tests, the threshold value of 80% was calculated. Fig. 3 depicts the image's construction block steps. Two channels are the output of the fully connected layer, one of which provides confidence values for facial components and the other of which contains confidence values for the background. The higher component confidence value in the fully linked layer creates iconized facial images. Additionally, two channels' confidence values are passed to the CNN-3's input for guided image training and more powerful face expression identification. A movable window is used for testing, as stated in [69].

Stage	Output size	Layer	
		Small	Large
Input	256 × 256	—	
Conv3 × 3	128 × 128	[32, s2]	
SepDepth3 × 3	128 × 128	[16, s1]	[24, s1]
NAS Stage1	64 × 64	[24, s2] [24, s1] × 3	[32, s2] [32, s1] × 5
NAS Stage2	32 × 32	[32, s2] [32, s1] × 5	[64, s2] [64, s1] × 7
NAS Stage3	16 × 16	[64, s2] [64, s1] × 9	[96, s2] [96, s1] × 9
NAS Stage4	16 × 16	[96, s1] × 8	[160, s2] × 10
TransConv	32 × 32	[64, s2]	
TransConv	64 × 64	[32, s2]	

Fig. 2. Neural Architecture Search Architecture.

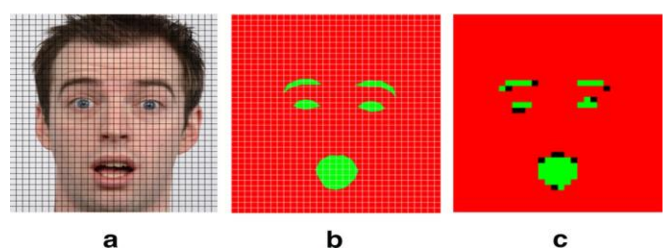

Fig. 3. Sample Image which Shows the Non Overlapping Parts. Black is the Ignored Part, Green is the Facial Component and Red is the Background in the Image C which Shows the Built and Labeled Part. Ground Truth is B and A is the Sample Divided Input Image.

CNN 3: The standard Xception model inspired this architecture. SeparableConv is the modified depthwise separable convolution, as shown in Fig. 4. SeparableConvs are considered as Inception Modules and used throughout the deep learning architecture, as can be established. All flows have residual (or shortcut/skip) connections, which ResNet first proposed.

The proposed method for consumer interest monitoring uses CNNs to complete the three learning steps. The method begins with CNN-1 detecting frontal faces. The frontal pictures are then split using CNN-2 in order to maintain facial components such as critical and important trains of facial expression. Finally, CNN-3 uses CNN-2's fully connected layer confidence values together with raw facial photos to classify facial expressions. The proposed CNN design is shown in Fig. 5.

Fig. 5. Sample Image which shows the non-overlapping parts. Black is the ignored part, Green is the facial component and red is the background in the image C which shows the built and labeled part. Ground truth is B and A is the sample divided input image.

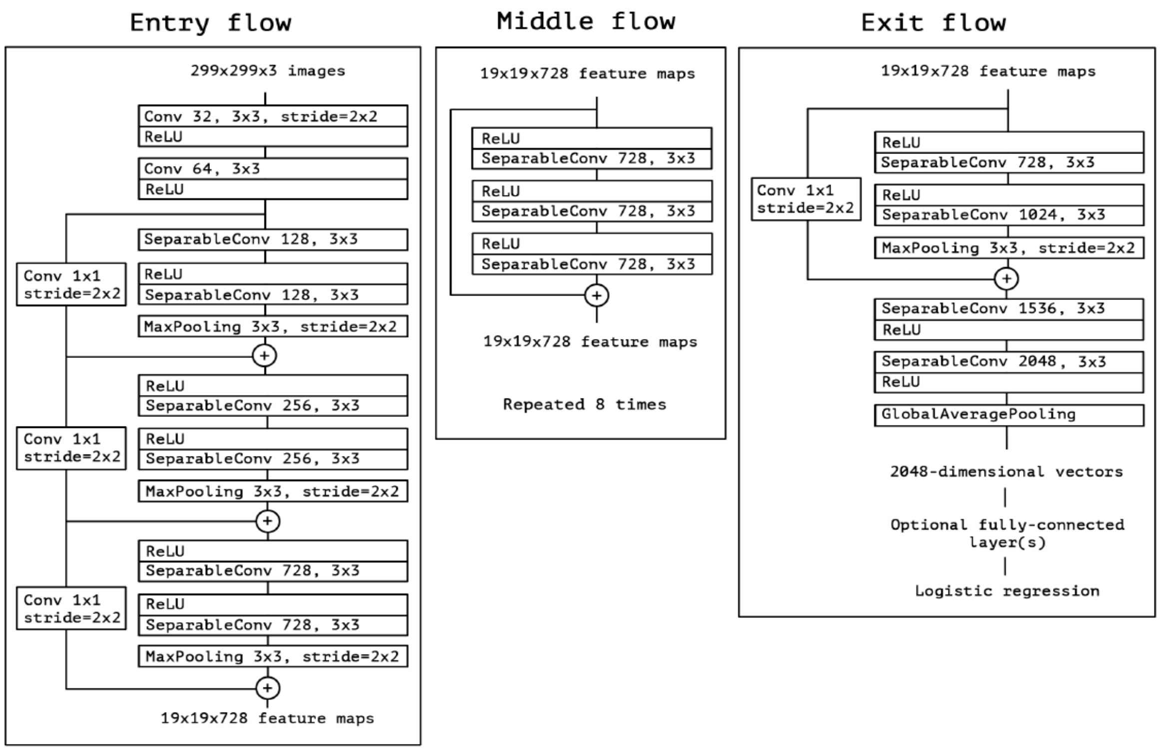

Fig. 4. Standard Exception Architecture.

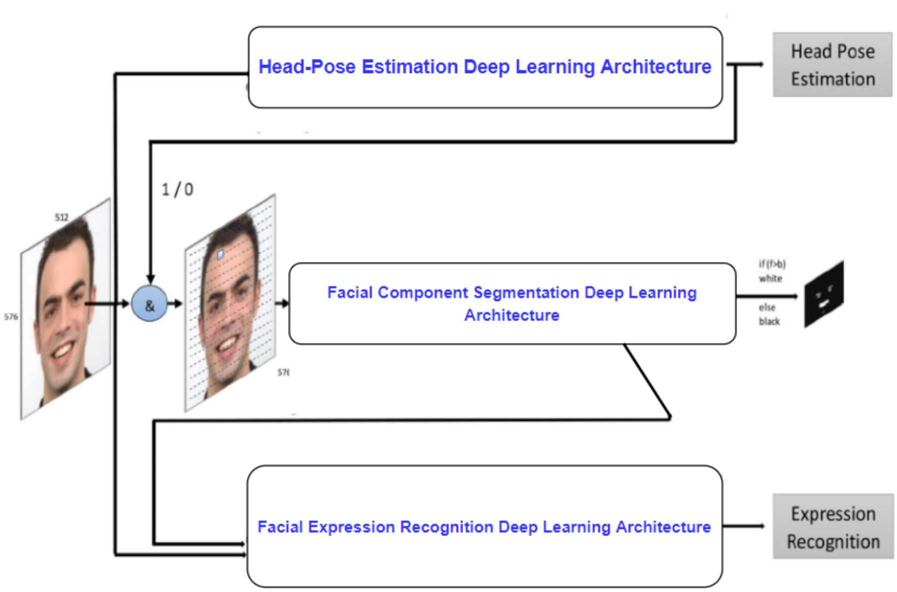

Fig. 5. Complete Work Flow of the Proposed Architecture.

IV. Results and Analysis

The proposed deep learning architecture was built using the Keras library with Tensor Flow as the back-end. The Radboud Face Database (RaFD) was used to train and test the network [70].

In this work, 8040 photographs were utilised to estimate head posture, while 1206 frontal images were used to detect face expressions from RaFD (for angry, pleased, disgust, surprise, sad, and fear). The training and testing sets for head posture estimation and facial expression detection were divided into two equal halves. There are no photographs in the testing and training sets that belong to the same person. Raw facial images were used to train the algorithm in the head posture assessment phase. Although the suggested approach can distinguish between frontal and non-frontal profile head positions, it was trained to recognise five separate yaw movement angles. Table I shows the head posture estimate confusion matrix for RaFD.

Using the Karolinska Directed Emotional Face (KDEF) database, the proposed technique was also evaluated [71]. KDEF is ideal for evaluating proposed system performance since it incorporates both head posture angle and facial expression labels of the face photos. KDEF consists of 4900 pictures taken from five different angles and with seven different face expressions. 4900 shots were utilised to estimate head position, while 840 frontal images were used to recognise facial expressions. Because of the small number of images in these collections, they were divided into two groups for training (90%) and testing (10%). Table II shows the KDEF's head posture estimate confusion matrix.

Each picture was clipped to contain facial areas of the brows, eyes, and mouth in order to build learning masks for the face element segmentation approach. The Face++ toolbox (Face++ 2017) was then used to detect critical facial points. On a human face, the toolkit can recognise 83 important regions. Training masks were made at 45 key locations. Each essential point was connected to make a polygon. These polygons were used to create training masks. After that, the pictures were thresholded to create a final training set. Fig. 6 depicts the producing process. The system was trained on a 5-channel input for facial emotion recognition (comprising a 3-channel raw facial picture and 2-channel ionized face confidence values).

The system was trained on a 5-channel input for facial emotion recognition (comprising a 3- channel raw facial picture and 2-channel iconized face confidence values). Table III shows the RaFD confusion matrix for emotion recognition. When the confusion matrix is used to recalculate the accuracy of positive and negative expressions, positive expression (happy) has a 95.11 percent accuracy, while negative expression (anger, disgust, fear, and sadness) has a 92.88 percent accuracy. Table IV also includes the KDEF confusion matrix for face expression recognition.

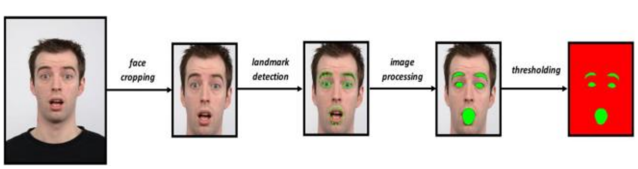

Fig. 6. Sample Image form the Database with Generating Flow.

TABLE I. HEAD POSE ESTIMATION ON RAFD: CONFUSION MATRIX

Actual	**Predicted (in %)**				
	0	***45***	***90***	***135***	***180***
0	98.66	1.34	0	0	0
45	0	99.11	0.89	0	0
90	0	0	100	0	0
135	0	0.82	0	99.18	0
180	0	0	0	0	100
Average:	99.43				

TABLE II. HEAD POSE ESTIMATION ON KDEF: CONFUSION MATRIX

Actual	**Predicted (in %)**				
	0	***45***	***90***	***135***	***180***
0	99.02	0.98	0	0	0
45	0	100	0	0	0
90	0	0	100	0	0
135	0	0	0	100	0
180	0	0	0	0.84	99.16
Average:	99.64				

TABLE III. CLASSIFICATION CONFUSION MATRIX ON KDEF DATABASE: EMOTION RECOGNITION (%)

Actual\ Predicted	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprised**
Anger	95.92	1.11	0	0	2.97	0
Disgust	2.01	95.66	0	0	0	2.33
Fear	0	0	88.23	0	10.00	1.77
Happy	0	3.11	0	95.11	1.78	0
Sad	3.85	4.55	0	0	92.34	0
Surprised	0	0	0	0.77	0	99.23

TABLE IV. CLASSIFICATION CONFUSION MATRIX ON KDEF DATABASE: FACIAL EXPRESSIONS (%)

Actual\ Predicted	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprised**
Anger	89.12	3.14	0	0	7.74	0
Disgust	0	91.66	0	0	8.34	0
Fear	0	0	84.23	0	11.23	4.54
Happy	0	0	0	98.11	0.34	1.55
Sad	5.66	0	0	0	94.34	0
Surprised	0	0	0	1.77	0	98.23

Because the RaFD database contains strange expressions, our system may experience certain mistakes. Fig. 7 depicts a visual representation of some of our mistake instances. The proposed system's performance was compared to that of previous face emotion recognition studies. The comparison's findings are shown in Table V. The results of facial expression recognition utilising 3-channel raw image, 1-channel iconized image, and 5-channel composite image for RaFD are shown in the table below. Table VI illustrates the execution times for three different types of CNNs, as well as the overall execution time, with estimated values for one picture. The segmentation procedure takes up a significant amount of time.

The proposed pipeline is intended for photos with a resolution of 576 * 512 pixels and discrete facial components. CNN-1 and CNN-3 use 64x64 images as inputs, while CNN-2 uses 576x512 images. As a result, it takes longer for CNN-2 to process the input pictures. Performance declines by 13% if the segmentation phase is bypassed in order to save time (85.12 percent for raw input image data vs. 95.11 percent for the data from the five channel).

The details of hyperparameters and their values are Threshold 0.005, Learning rate 0.10, Input channels 3 5 5, regularization strength of 0.0005and the Batch-size 3. The train and test of the aligned face are regular and saturated well, but the unaligned looks are quite different. But they still saturate at the same number of epochs, but the variations can be observed in Fig. 8.

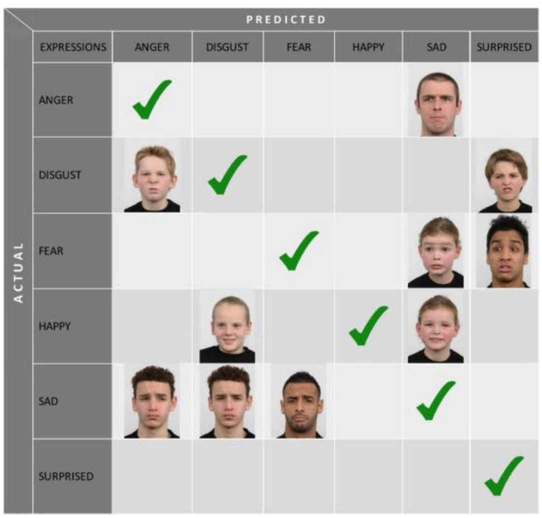

Fig. 7. Misclassification of Results and its Visualizations.

TABLE V. FACIAL EMOTION RECOGNITION OF PROPOSED METHOD WITH OTHER STUDIES ON RAFD / KDEF DATABASES

Methods	Database	Accuracy (in %)
HoG + NNE [1]	RaFD	93.75
Surf Boosting [64]	RaFD	90.64
Gabor F. + GLCM [36]	RaFD	88.41
LSiBP + SVM[68]	KDEF	84.07
HoG + AdaBoost [38]	KDEF	87.20

TABLE VI. EXECUTION TIME FOR CNN STUCTURES

CNN Structures	Task	45	90	135	Time
1	Estimation of Head Pose				~0.02
2	Segmentation of Facial Component				~0.13
3	Recognition of Facial Expression				~0.80
0.31 Sec. is the total execution time for one image					

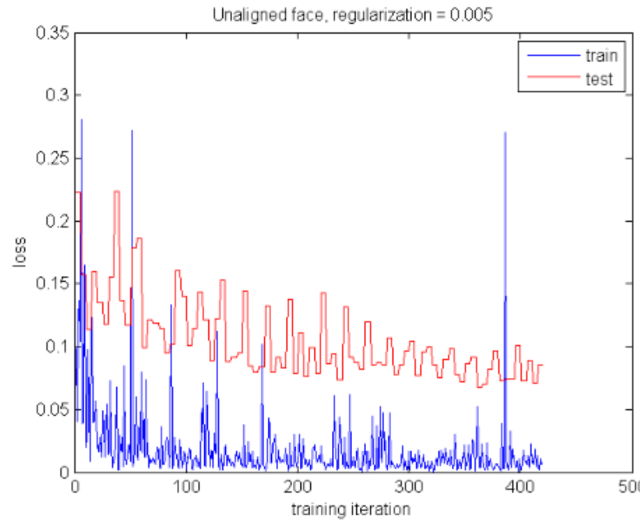

Fig. 8. Train and Test Curves of Unaligned Face Images.

V. CONCLUSION AND FUTUTRE DIRECTIONS

A unique deep learning system for automatic head posture estimation and face emotion identification is proposed. The proposed method started with developing a noninvasive, quantifiable approach for tracking client interest. The suggested system is made up of three CNN structures in a cascade. The first CNN's task is to estimate head posture. Face segmentation was taught to the second CNN structure. Face recognition and classification were encoded into the third CNN. The latter two steps (CNN-2 and CNN-3) allow for guided picture classification and the integration of part-based and holistic data. The RaFD dataset demonstrated that head posture estimation was 99.43 percent accurate and face expression detection was 93.20 percent accurate in experimental tests. The average accuracy of positive and negative emotions is 93.99.

The proposed approach can aid in the measurement of relevant marketing and product likability as well as the quantification of the client interest. It can also be used to identify sales-boosting company initiatives. According to customer feedback, marketing efforts can alter their methods. Future direction includes the temporal analysis of the data and the optimization space used in the architecture. Track human faces and use object localization so that the system can watch and index an individual's facial expressions over time. Exploring the proposed architecture with people of different regions and cultural background and testing it on various other domains.

REFERENCES

[1] Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R., & Akan, A. (2019, October). Real time emotion recognition from facial expressions using CNN architecture. In *2019 medical technologies congress (tiptekno)* (pp. 1-4). IEEE.

[2] Missaglia, A. L., Oppo, A., Mauri, M., Ghiringhelli, B., Ciceri, A., & Russo, V. (2017). The impact of emotions on recall: An empirical study on social ads. Journal of Consumer Behaviour, 16(5), 424-433.

[3] Dockès, J., Poldrack, R. A., Primet, R., Gözükan, H., Yarkoni, T., Suchanek, F., ... & Varoquaux, G. (2020). NeuroQuery, comprehensive meta-analysis of human brain mapping. *Elife*, *9*, e53385.

[4] Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, *16*(3), 031001.

[5] Cook, I.A., Warren, C., Pajot, S.K., Schairer, D. and Leuchter, A.F., 2011. Regional brain activation with advertising images. Journal of Neuroscience, Psychology, and Economics, 4(3), p.147.

[6] Wedel, M. and Pieters, R., 2000. Eye fixations on advertisements and memory for brands: A model and findings. Marketing science, 19(4), pp.297-312.

[7] Ungureanu, F., Lupu, R.G., Cadar, A. and Prodan, A., 2017, October. Neuromarketing and visual attention study using eye tracking techniques. In 2017 21st international conference on system theory, control and computing (ICSTCC) (pp. 553-557). IEEE.

[8] Micu, A.C. and Plummer, J.T., 2010. Measurable emotions: How television ads really work: Patterns of reactions to commercials can demonstrate advertising effectiveness. Journal of Advertising Research, 50(2), pp.137-153.

[9] Popa, M., Rothkrantz, L., Yang, Z., Wiggers, P., Braspenning, R. and Shan, C., 2010, October. Analysis of shopping behavior based on surveillance system. In 2010 IEEE International Conference on Systems, Man and Cybernetics (pp. 2512-2519). IEEE.

[10] Liu, J., Gu, Y., & Kamijo, S. (2017). Customer behavior classification using surveillance camera for marketing. Multimedia Tools and Applications, 76(5), 6595-6622.

[11] Liu, J., Gu, Y. and Kamijo, S., 2018. Integral customer pose estimation using body orientation and visibility mask. Multimedia Tools and Applications, 77(19), pp.26107-26134.

[12] Kasiran Z, Yahya S (2007) Facial expression as an implicit customers' feedback and the challenges. In: Computer graphics, imaging and visualisation, IEEE, pp 377–381

[13] Teixeira, T., Wedel, M. and Pieters, R., 2012. Emotion-induced engagement in internet video advertisements. Journal of marketing research, 49(2), pp.144-159.

[14] Ba, S. O., & Odobez, J. M. (2010). Multiperson visual focus of attention from head pose and meeting contextual cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(1), 101-116.

[15] Riener, A. and Sippl, A., 2014. Head-pose-based attention recognition on large public displays. IEEE computer graphics and applications, 34(1), pp.32-41.

[16] Sheikhi, S. and Odobez, J.M., 2015. Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions. Pattern Recognition Letters, 66, pp.81-90.

[17] Karu-Salo, I., 2013. The effect of universal emotions on customer behaviour. Estonian Business School, pp.1-72.

[18] Kahneman D. Thinking, fast and slow. Macmillan; 2011 Oct 25.

[19] Sangeeta, S., 2013. Module–6 Consumer Behavior. Vinod Gupta School of Management, Indian Institute of Technology: India. Article. Available on: http://nptel. ac. in/courses/110105029/pdf, 20, pp.6-23.

[20] Ekman P, Friesen WV (1975) Unmasking the face: a guide to recognising emotions from facial clues. Consulting Psychologists Press, New Jersey (Prentice Hall [Palo Alto, CA]).

[21] Laros, F.J. and Steenkamp, J.B.E., 2005. Emotions in consumer behavior: a hierarchical approach. Journal of business Research, 58(10), pp.1437-1445.

[22] Yan, C., Zhang, Y., Xu, J., Dai, F., Li, L., Dai, Q. and Wu, F., 2014. A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors. IEEE Signal Processing Letters, 21(5), pp.573-576.

[23] Yan, C., Zhang, Y., Xu, J., Dai, F., Zhang, J., Dai, Q. and Wu, F., 2014. Efficient parallel framework for HEVC motion estimation on many-core processors. IEEE Transactions on Circuits and Systems for Video Technology, 24(12), pp.2077-2089.

[24] Bayrakdar, S., Akgün, D. and Yücedağ, İ., 2017. An accelerated approach for facial expression analysis on video files. Pamukkale University Journal of Engineering Sciences, 23(5), pp.602-613.

[25] Yan, C., Li, L., Zhang, C., Liu, B., Zhang, Y. and Dai, Q., 2019. Cross-modality bridging and knowledge transferring for image understanding. IEEE Transactions on Multimedia, 21(10), pp.2675-2685.

[26] Yan, C., Xie, H., Chen, J., Zha, Z., Hao, X., Zhang, Y. and Dai, Q., 2018. A fast uyghur text detector for complex background images. IEEE Transactions on Multimedia, 20(12), pp.3389-3398.

[27] Yan, C., Xie, H., Liu, S., Yin, J., Zhang, Y. and Dai, Q., 2017. Effective Uyghur language text detection in complex background images for traffic prompt identification. IEEE transactions on intelligent transportation systems, 19(1), pp.220-229.

[28] Oztel, I., Yolcu, G., Öz, C., Kazan, S. and Bunyak, F., 2018. iFER: facial expression recognition using automatically selected geometric eye and eyebrow features. Journal of Electronic Imaging, 27(2), p.023003.

[29] Zalewski, L. and Gong, S., 2005, June. 2d statistical models of facial expressions for realistic 3d avatar animation. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 2, pp. 217-222). IEEE.

[30] Kim, J. B., Hwang, Y., Bang, W. C., Lee, H., Kim, J. D., & Kim, C. (2013, January). Real-time realistic 3D facial expression cloning for smart TV. In 2013 IEEE International Conference on Consumer Electronics (ICCE) (pp. 240-241). IEEE.

[31] Takahashi, M., Clippingdale, S., Naemura, M. and Shibata, M., 2015. Estimation of viewers' ratings of TV programs based on behaviors in home environments. Multimedia Tools and Applications, 74(19), pp.8669-8684.

[32] Teyeb, I., Jemai, O., Zaied, M., & Amar, C. B. (2015, December). Vigilance measurement system through analysis of visual and emotional driver's signs using wavelet networks. In 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA) (pp. 140-147). IEEE.

[33] Lozano-Monasor, E., López, M. T., Vigo-Bustos, F., & Fernández-Caballero, A. (2017). Facial expression recognition in ageing adults: from lab to ambient assisted living. Journal of Ambient Intelligence and Humanized Computing, 8(4), 567-578.

[34] De Carolis, B., Ferilli, S., & Palestra, G. (2017). Simulating empathic behavior in a social assistive robot. Multimedia Tools and Applications, 76(4), 5073-5094.

[35] Yolcu, G., Oztel, I., Kazan, S., Oz, C., Palaniappan, K., Lever, T. E., & Bunyak, F. (2017, November). Deep learning-based facial expression recognition for monitoring neurological disorders. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1652-1657). IEEE.

[36] Shaykha, I., Menkara, A., Nahas, M., & Ghantous, M. (2015, September). FEER: Non-intrusive facial expression and emotional recognition for driver's vigilance monitoring. In 2015 57th International Symposium ELMAR (ELMAR) (pp. 233-237). IEEE.

[37] Zhang, Y., & Hua, C. (2015). Driver fatigue recognition based on facial expression analysis using local binary patterns. Optik, 126(23), 4501-4505.

[38] Terzis, V., Moridis, C. N., & Economides, A. A. (2013). Measuring instant emotions based on facial expressions during computer-based assessment. Personal and ubiquitous computing, 17(1), 43-52.

[39] Zhao, Y., Wang, X., Goubran, M., Whalen, T., & Petriu, E. M. (2013). Human emotion and cognition recognition from body language of the head using soft computing techniques. Journal of Ambient Intelligence and Humanized Computing, 4(1), 121-140.

[40] Niforatos, E., & Karapanos, E. (2015). EmoSnaps: a mobile application for emotion recall from facial expressions. Personal and Ubiquitous Computing, 19(2), 425-444.

[41] Baldassarri, S., Hupont, I., Abadía, D., & Cerezo, E. (2015). Affective-aware tutoring platform for interactive digital television. Multimedia Tools and Applications, 74(9), 3183-3206.

[42] Benmohamed, A., Neji, M., Ramdani, M., Wali, A., & Alimi, A. M. (2015). Feast: face and emotion analysis system for smart tablets. Multimedia Tools and Applications, 74(21), 9297-9322.

[43] Arigbabu, O. A., Mahmood, S., Ahmad, S. M. S., & Arigbabu, A. A. (2016). Smile detection using hybrid face representation. Journal of Ambient Intelligence and Humanized Computing, 7(3), 415-426.

[44] Samara, A., Galway, L., Bond, R., & Wang, H. (2019). Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), 2175-2184.

[45] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124.

[46] Pons, G., & Masip, D. (2017). Supervised committee of convolutional neural networks in automated facial expression analysis. IEEE Transactions on Affective Computing, 9(3), 343-350.

[47] Lee, C. C., Shih, C. Y., Lai, W. P., & Lin, P. C. (2012). An improved boosting algorithm and its application to facial emotion recognition. Journal of Ambient Intelligence and Humanized Computing, 3(1), 11-17.

[48] Ghimire, D., & Lee, J. (2013). Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors, 13(6), 7714-7734.

[49] Perumal Ramalingam, S., & Chandra Mouli, P. V. S. S. R. (2018). Modified dimensionality reduced local directional pattern for facial analysis. Journal of Ambient Intelligence and Humanized Computing, 9(3), 725-737.

[50] Ghimire, D., & Lee, J. (2012). Histogram of orientation gradient feature-based facial expression classification using bagging with extreme learning machine. Advanced Science Letters, 17(1), 156-161.

[51] Sobia, M. C., Brindha, V., & Abudhahir, A. (2014, February). Facial expression recognition using PCA based interface for wheelchair. In 2014 International Conference on Electronics and Communication Systems (ICECS) (pp. 1-6). IEEE.

[52] Zhao, G., & Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE transactions on pattern analysis and machine intelligence, 29(6), 915-928.

[53] Lopes, A. T., De Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern recognition, 61, 610-628.

[54] Pitaloka, D. A., Wulandari, A., Basaruddin, T., & Liliana, D. Y. (2017). Enhancing CNN with preprocessing stage in automatic emotion recognition. Procedia computer science, 116, 523-529.

[55] Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks, 16(5-6), 555-559.

[56] Chen, C., & Odobez, J. M. (2012, June). We are not contortionists: Coupled adaptive learning for head and body orientation estimation in surveillance video. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1544-1551). IEEE.

[57] Kang, S. K., Chung, K. Y., & Lee, J. H. (2014). Development of head detection and tracking systems for visual surveillance. Personal and ubiquitous computing, 18(3), 515-522.

[58] Murphy-Chutorian, E., Doshi, A., & Trivedi, M. M. (2007, September). Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In 2007 IEEE intelligent transportation systems conference (pp. 709-714). IEEE.

[59] Alioua, N., Amine, A., Rogozan, A., Bensrhair, A., & Rziza, M. (2016). Driver head pose estimation using efficient descriptor fusion. EURASIP Journal on Image and Video Processing, 2016(1), 1-14.

[60] Ba, S. O., & Odobez, J. M. (2008). Recognizing visual focus of attention from head pose in natural meetings. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 16-33.

[61] Gaschler, A., Jentzsch, S., Giuliani, M., Huth, K., de Ruiter, J., & Knoll, A. (2012, October). Social behavior recognition using body posture and head pose for human-robot interaction. In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2128-2133). IEEE.

[62] Patacchiola, M., & Cangelosi, A. (2017). Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. Pattern Recognition, 71, 132-143.

[63] Murphy-Chutorian, E., & Trivedi, M. M. (2008). Head pose estimation in computer vision: A survey. IEEE transactions on pattern analysis and machine intelligence, 31(4), 607-626.

[64] McKenna, S. J., & Gong, S. (1998). Real-time face pose estimation. Real-Time Imaging, 4(5), 333-347.

[65] Tran BH, Kim Y-G (2017) Deep head pose estimation for faces in the wile and its transfer learning. In: Seventh international conference on information science and technology (ICIST). IEEE, pp 187–193.

[66] Mukherjee, S. S., & Robertson, N. M. (2015). Deep head pose: Gaze-direction estimation in multimodal video. IEEE Transactions on Multimedia, 17(11), 2094-2107.

[67] Xu, X., & Kakadiaris, I. A. (2017, May). Joint head pose estimation and face alignment framework using global and local CNN features. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 642-649). IEEE.

[68] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154.

[69] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[70] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. Cognition and emotion, 24(8), 1377-1388.

[71] Lundqvist, D., Flykt, A., & Öhman, A. (1998). Karolinska directed emotional faces. Cognition and Emotion.