Real-Time Customer Satisfaction Analysis using Facial Expressions and Head Pose Estimation

Nethravathi P. S. [1] & P. S. Aithal [2]

[1] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore, India.
OrcidID: 0000-0001-5447-8673; Email: nethrakumar590@gamil.com

[2] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore, India.
OrcidID: 0000-0002-4691-8736; Email: psaithal@gmail.com

Subject Area: Computer Science.
Type of the Paper: Conceptual Research.
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed In: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.6592194
Google Scholar Citation: IJAEML

How to Cite this Paper:
Nethravathi, P. S., & Aithal, P. S., (2022). Real-Time Customer Satisfaction Analysis using Facial Expressions and Head Pose Estimation. *International Journal of Applied Engineering and Management Letters (IJAEML)*, *6*(1), 301-312. DOI: https://doi.org/10.5281/zenodo.6592194

International Journal of Applied Engineering and Management Letters (IJAEML)
A Refereed International Journal of Srinivas University, India.

Crossref DOI : https://doi.org/10.47992/IJAEML.2581.7000.0138

Received on: 04/04/2022
Published on: 31/05/2022

Real-Time Customer Satisfaction Analysis using Facial Expressions and Head Pose Estimation

Nethravathi P. S. [1] & P. S. Aithal [2]

[1] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore, India.
OrcidID: 0000-0001-5447-8673; Email: nethrakumar590@gamil.com

[2] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore, India.
OrcidID: 0000-0002-4691-8736; Email: psaithal@gmail.com

ABSTRACT

Background/Purpose: *Quantification of consumer interest is an interesting, innovative, and promising trend in marketing research. For example, an approach for a salesperson is to observe consumer behaviour during the shopping phase and then recall his interest. However, the salesperson needs unique skills because every person may interpret their behaviour in a different manner. The purpose of this research is to track client interest based on head pose positioning and facial expression recognition.*

Objective: *We are going to develop a quantifiable system for measuring customer interest. This system recognizes the important facial expression and then processes current client photos and does not save them for later processing.*

Design/Methodology/Approach: *The work describes a deep learning-based system for observing customer actions, focusing on interest identification. The suggested approach determines client attention by estimating head posture. The system monitors facial expressions and reports customer interest. The Viola and Jones algorithms are utilized to trim the facial image.*

Findings/Results: *The proposed method identifies frontal face postures, then segments facial mechanisms that are critical for facial expression identification and creating an iconized face image. Finally, the obtained values of the resulting image are merged with the original one to analyze facial emotions.*

Conclusion: *This method combines local part-based features with holistic facial information. The obtained results demonstrate the potential to use the proposed architecture as it is efficient and works in real-time.*

Paper Type: *Conceptual Research.*

Keywords: Customer monitoring, Convolutional neural network, Head pose estimation, Facial expression recognition, Facial analysis.

1. INTRODUCTION :

The usual way is for a salesperson to study client attitude when he/she watching or shopping phase and then recall customer interest. However, every salesperson needs special talents for this job, and each spectator may interpret consumer behaviour 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.

Knowing the choices using various tools such as brain images [3], 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]; [10] and customer face analysis studies have also been used in several studies [11].

Estimating a client's graphic focus of attention is one approach to gauging their interest. Assessment of head posture is used in research on the visual focus of attention [12]. Moreover, identifying client opinions for selling a product is a quickly growing and difficult research field. The intuitive decision-making process is substantially influenced by one's mood [13]. 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 needs their attention. Customers' emotions and moods must be considered by marketers [14]. According to [15], identify as positive or negative feelings, the surprise isn't included and difficult as well [16].

Our goal is to design a deep learning system for tracking client interests. The proposed method detects the visual centre of attention by first recognizing the human face and then estimating head posture orientation. This work could be useful for identifying things like marketing campaigns and other company efforts that clients could 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.

2. OBJECTIVES OF THE STUDY :

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

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

(2) Because the suggested work will not save 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 CNN (Convolutional Neural Network) model with multiple tasks is suggested. The third 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 recognizes and localizes the facial apparatuses that are crucial for detection in the facial component segmentation stage, which results in an ionized image. The CNN permits guided training at the facial expression recognition stage by compelling the earlier layers of the architecture.

Provides more details are mentioned in the subsequent sections and are organized as shown below. Section II gives a brief idea about the literature survey on the existing state-of-the-art methods related to the defined methodology. Section III demonstrates the proposed method. Section IV contains the results, analysis, discussions, and inferences. Section V concludes the paper with some details regarding the future directions.

3. 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 [17], [18]. Real-world challenges include object identification [19], text detection ([20],[21]), facial expression recognition [22], head position estimation [23], and so on. The two main elements of this work, facial expression recognition, and head pose estimation studied in this section.

A. Facial expression recognition:

It is a growing area with applications in avatar animation [24], smart environments [25], robotic [26], medical [27], traffic [28], and human–computer interaction [29]. In their early research, the six universal facial expressions described by Ekman and Friesen, namely disgust, happiness, fear, rage, sadness, and surprise, are typically used [30].

Geometric-based and appearance-based algorithms for facial expression recognition have been identified [31]. The features derived from positional correlations between facial components focus on geometric-based approaches. Appearance-based features determine face texture [32]. Local binary pattern (LBP) operator [33], histogram of oriented gradients (HOG) [34], principal component analysis (PCA) [35] etc.

B. Head pose estimation:

This has been studied for a variety of applications, including visual surveillance [36], driver attention [37], the visual focus of attention [38], and robotics. Appearance-based approaches compare a fresh

head image to a set of head posture templates to determine which perspective is the most similar. Appearance-based approaches have the drawback of being limited to predicting discrete pose positions [39].
Furthermore, certain templates require lengthy image comparisons. Geometric information or facial landmark locations are used for head posture estimation in model-based techniques. The accuracy of model-based approaches is determined by the number and quality of geometric signals derived from the image. Dimensionality reduction strategies are used in manifold embedding techniques like PCA used picture projection into a PCA subspace and compared the results [40]. According to [39], a consistent dataset to train the parameters is required for nonlinear regression.
CNNs [39] are a nonlinear regression technique. CNNs give high-accuracy results for head pose orientation difficulties described as 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, used a combination of both models estimated head posture and located landmarks using global and local data collected from a CNN. For head posture estimation, employed CNN and adaptive gradient methods.

4. RESEARCH GAP :

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.

5. METHODOLOGY :

The Viola and Jones Algorithm [41] is used to trim each facial image to remove background information and leave only expression and head pose-specific data. The suggested method employs the Viola and Jones algorithm, which allows for rapid feature assessments while simultaneously lowering the complexity of every border. As demonstrated in Eq. 1, the integral picture at the location x, y contains the total pixels above and left of x, y when this coordinate is included.

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

Where j(x,y) is a raw image and jj(x,y) is an integral image. Using Eqs. 2 and 3, the integral image may be considered [82]. Where a(x, y) denotes the cumulative row total, a(x, -1) denotes zero, and jj(-1, y) denotes zero.

$$a\ (x, y) = a\ (s, y-1) + j\ (x, y) \qquad (2)$$

$$jj\ (x, y) = jj\ (x-1, y) + s\ (x, y) \qquad (3)$$

The proposed system's initial stage determines whether the consumer looks correct item. For the suggested system that groups ahead versus non-frontal profile, the coarsest level head position estimation is sufficient. Non-frontal faces are ignored when frontal faces are presented to CNN-2 for segmentation. The CNN-1 has been trained to estimate 0 °, 45 °, 90 °, 130 °, and 180 ° of head posture (Figure 1).

Fig. 1: Sample Image snapshots with different head pose positions (Compiled by the Authors)

(Note: All the colored images can be used only for the Srinivasa University journals. For the rest of the journals, either you can replace the same with your students' image or you can delete the image)
CNN 1: We apply differentiable Neural architecture search (NAS) to adapt the backbone network design for the pose estimation problem. Figure 1 shows the complete information, including the Differentiable neural architecture search, Efficient backbone, Efficient head, and Cost optimization.

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 (Compiled by the Authors)

CNN 2: Faces such as the mouth, eye, and brow areas are separated from the rest of the image by the CNN-2. The problem of face section clustering is framed as a binary classification problem of facial component vs. backdrop. According to our tests, the threshold value of 80% was calculated. Figure 5 depicts the image's construction block steps. The output is two channels, 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 71.

CNN 3: The standard Xception model inspired this architecture. SeparableConv is the modified depthwise separable convolution, as shown in Figure 2. SeparableConvs are taken as Inception Modules.

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 essential and critical facial expression traits. 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 Figure 4.

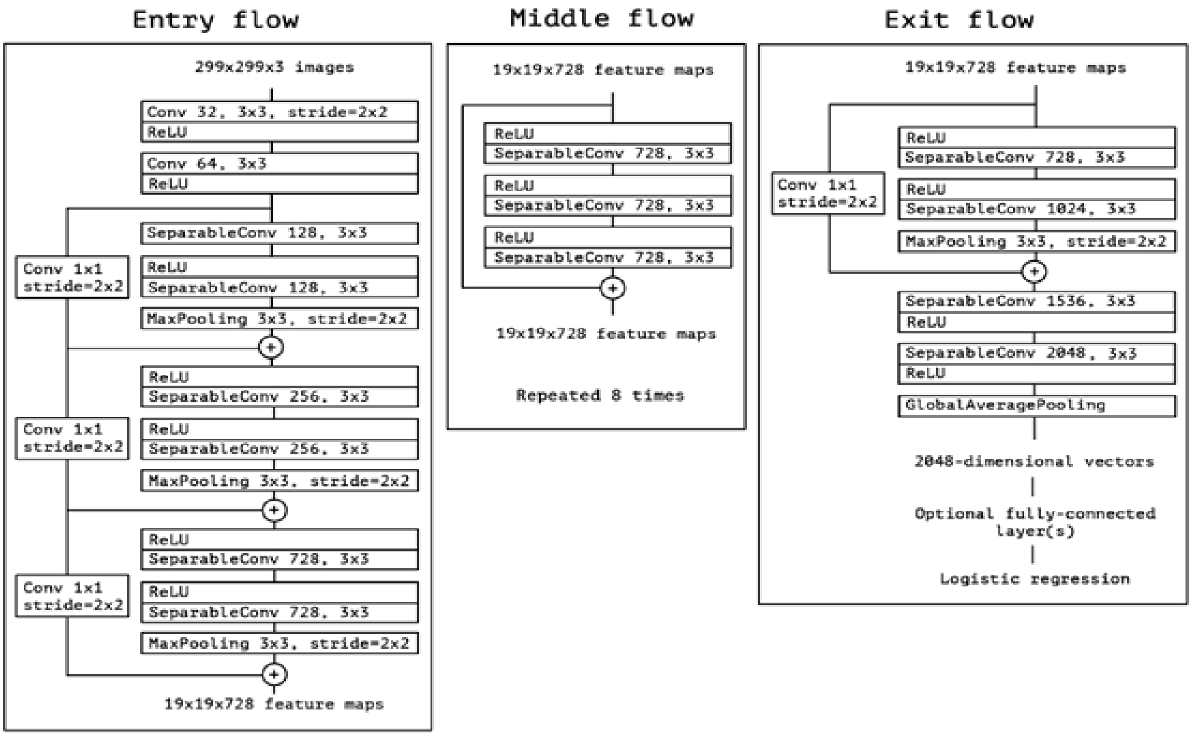

Fig. 3: Standard Exception Architecture (Compiled by the Authors)

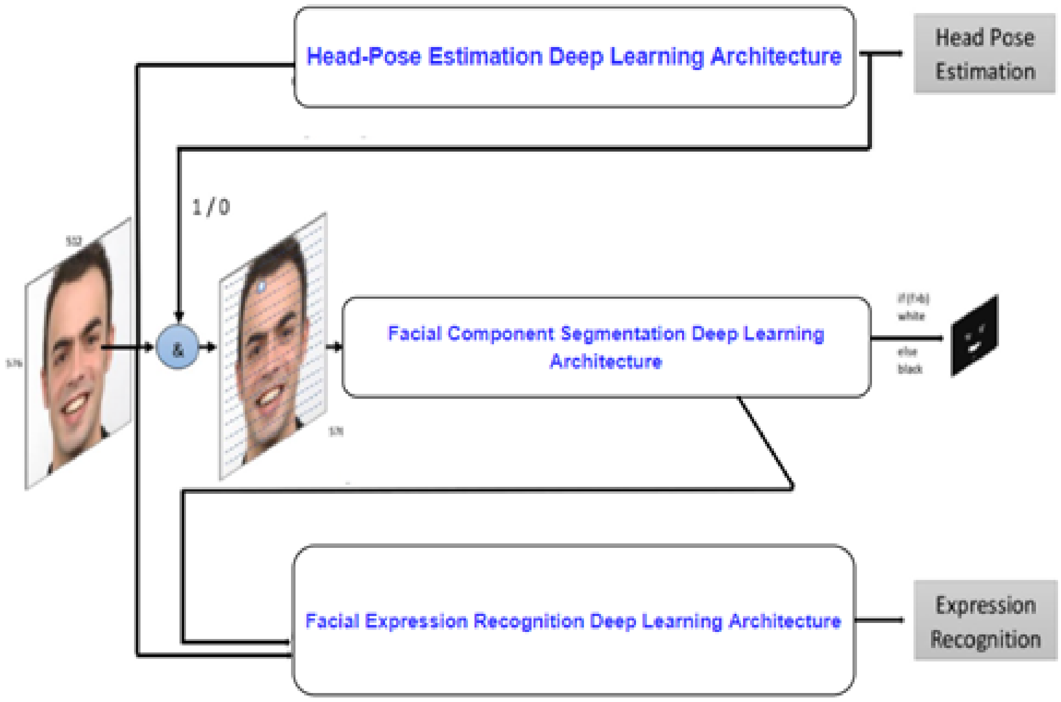

Fig. 4: Complete workflow of the proposed architecture (Compiled by the Authors)

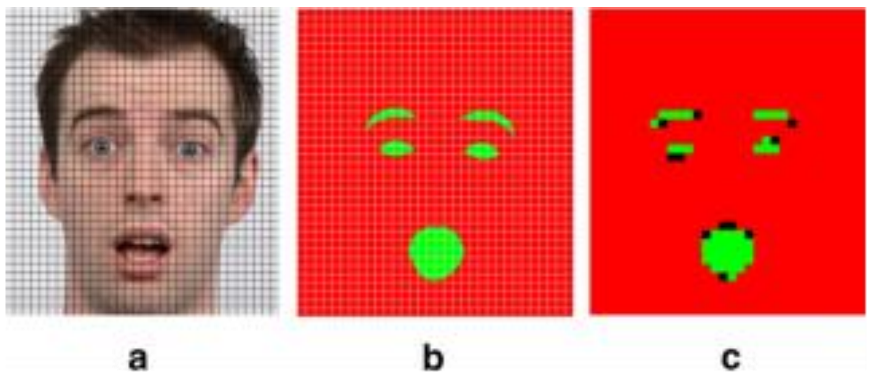

Fig. 5: Sample Image which shows the non-overlapping parts. Black is the ignored part, Green is the facial component and red is the background in image C which shows the built and labeled part. B is the real one and A is the sample divided image.

6. RESULTS AND ANALYSIS :

The planned deep learning architecture was built by the Keras library with Tensor Flow as the back-end. The Radboud Face Database (RaFD) was used to train and test the network. Totally 8040 photos were utilized, while 1206 frontal pictures were used for detection from RaFD (for angry, pleased, etc). The identification sets for training and testing were separated into two equal halves. Any image of the same person is not included in the testing or training sets. Raw facial photos are used to train algorithms in the head pose estimation portion. Although the suggested system can categorize a head posture as a frontal or non-frontal profile. Table 2 shows the head posture estimate confusion matrix for RaFD.

Table 1: Head pose estimation on RaFD: Confusion Matrix (Compiled by the Authors)

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				

The Karolinska Directed Emotional Face (KDEF) was also used here. KDEF is ideal for assessing proposed system performance since it incorporates both head posture angle and facial emotions. KDEF consists of 4900 photos with five different viewing angles and seven different facial expressions. Four thousand nine hundred photos utilized to estimate, while 840 frontal images were used for facial expression identification. Because there are few photos in these sets, they were divided into two groups to train (90%) and testing (10%). Table 3 shows the KDEF's head posture estimate confusion matrix.

Table 2: Head pose estimation on KDEF: Confusion Matrix (Compiled by the Authors)

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				

For the construction of training masks, and face component segmentation procedure, all pictures are trimmed to contain facial regions of the brows, eyes, and mouth. After that, the Face++ toolkit (Face++ 2017) was utilized to identify facial important points. The toolbox can recognize 83 critical areas on a human face. 45 important spots make training masks and a polygon was formed by connecting each critical point.

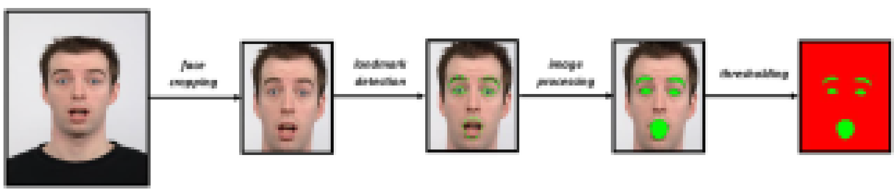

Fig. 6: Sample image from the database with Generating Flow

The system was trained on a 5-channel input for facial emotion recognition (Table 3 shows the RaFD confusion matrix for face look identification and when it 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. The KDEF step output is also given in Table 4.

Table 3: Classification Confusion Matrix on KDEF Database: Facial Expressions (%) (Compiled by the Authors)

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 4: Classification Confusion Matrix On KDEF Database: Facial Expressions (%) (Compiled by the Authors)

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. Figure 7 depicts a visual representation of some of our mistake instances. This work equated to previous studies on facial expression recognition. The comparison findings are shown in Table 5. Table 6 shows the time stages of CNN and the overall execution time, with approximate values for one picture. The segmentation procedure takes up a significant amount of time, as shown in the table.

Fig. 7: Misclassification of results and its visualizations (Compiled by the Authors)

All the image used in the paper are taken from the same databases used in the study. 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. If the segmentation step is skipped in order to save time, performance drops by 13% (85.12 percent for raw image input vs. 95.11 percent for 5-channel data).

Table 5: Facial emotion recognition of proposed method with other studies on RaFD /KDEF databases (Compiled by the Authors)

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 6: Execution time for CNN Structures (Compiled by the Authors)

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
Total Exen., Time = 0.31 Sec. for one image					

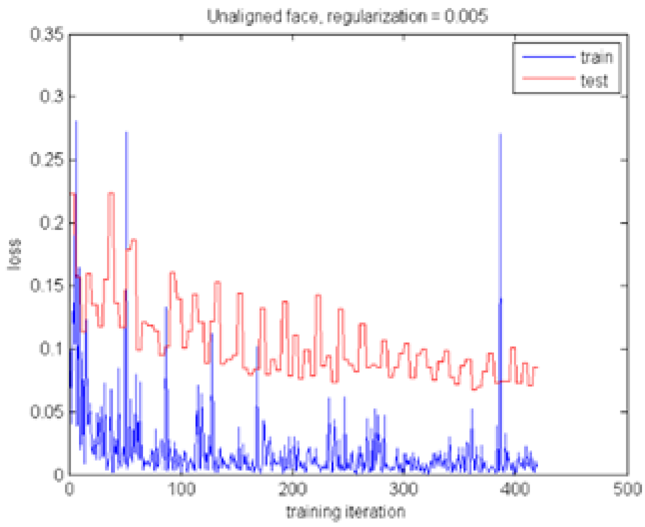

Fig. 8: Train and test curves of unaligned face images (Compiled by the Authors)

The details of hyper parameters 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 Figure 6.

7. RESEARCH AGENDA :

(i) A unique deep learning system for automatic head posture estimation and face emotion identification is presented in this article.
(ii) This research is the first step in developing a noninvasive, quantifiable approach for tracking client interests.

8. CONCLUSION AND FUTURE DIRECTIONS :

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. The third CNN was programmed to recognize and classify facial expressions. The last two steps allow for guided picture grouping and the integration of part-based and holistic data. On the RaFD dataset, testing test results showed 99.43 percent precision posture prediction and 93.20 percent accuracy for face appearance detection. 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's interest. It can also be used to identify sales-boosting company initiatives. According to customer feedback, marketing efforts can alter their methods. The 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 backgrounds and testing it on various other domains.

REFERENCES :

[1] Menon, K., & Dubé, L. (2000). Ensuring greater satisfaction by engineering salesperson response to customer emotions. *Journal of retailing*, *76*(3), 285-307. Google Scholar↗

[2] Menon, K., & Dubé, L. (2000). Ensuring greater satisfaction by engineering salesperson response to customer emotions. *Journal of retailing*, *76*(3), 285-307. Google Scholar↗

[3] Langleben, D. D., Loughead, J. W., Ruparel, K., Hakun, J. G., Busch-Winokur, S., Holloway, M. B., ... & Lerman, C. (2009). Reduced prefrontal and temporal processing and recall of high "sensation value" ads. *Neuroimage*, *46*(1), 219-225. Google Scholar↗

[4] Ohme, R., Reykowska, D., Wiener, D., & Choromanska, A. (2009). Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures. *Journal of Neuroscience, Psychology, and Economics*, *2*(1), 21-30. Google Scholar↗

[5] Cook, I. A., Warren, C., Pajot, S. K., Schairer, D., & Leuchter, A. F. (2011). Regional brain activation with advertising images. *Journal of Neuroscience, Psychology, and Economics*, *4*(3), 147-160. Google Scholar↗

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

[7] Ungureanu, F., Lupu, R. G., Cadar, A., & 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)*, 553-557. IEEE. Google Scholar↗

[8] Micu, A. C., & Plummer, J. T. (2010). Measurable emotions: How television ads really work. *Journal of Advertising Research*, *50*(2), 137-153. Google Scholar↗

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

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

[11] Kasiran, Z., & Yahya, S. (2007). *Facial expression as an implicit customers' feedback and the challenges*, 377-381. IEEE. Google Scholar↗

[12] Sheikhi, S., & 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*(1), 81-90. Google Scholar↗

[13] Karu-Salo, I. (2013). The effect of universal emotions on customer behaviour. *Estonian Business School*. 1-72. Google Scholar↗

[14] Sangeeta, S. (2013). Module–6 Consumer Behavior. *Vinod Gupta School of Management, Indian Institute of Technology: India. 20*(1), 6-23. Google Scholar↗

[15] Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*, Ishk. *10*(1). 1-197. Google Scholar↗

[16] Laros, F. J., & Steenkamp, J. B. E. (2005). Emotions in consumer behavior: a hierarchical approach. *Journal of business Research*, *58*(10), 1437-1445. Google Scholar↗

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

[18] Yan, C., Zhang, Y., Xu, J., Dai, F., Zhang, J., Dai, Q., & 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), 2077-2089. Google Scholar↗

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

[20] Yan, C., Xie, H., Chen, J., Zha, Z., Hao, X., Zhang, Y., & Dai, Q. (2018). A fast Uyghu text detector for complex background images. *IEEE Transactions on Multimedia*, *20*(12), 3389-3398. Google Scholar↗

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

[22] Oztel, I., Yolcu, G., Öz, C., Kazan, S., & Bunyak, F. (2018). iFER: facial expression recognition using automatically selected geometric eye and eyebrow features. *Journal of Electronic Imaging*, *27*(2), 23-30. Google Scholar↗

[23] Riener, A., & Sippl, A. (2014). Head-pose-based attention recognition on large public displays. *IEEE computer graphics and applications*, *34*(1), 32-41. Google Scholar↗

[24] Zalewski, L., & 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)*, *2*(1), 217-222. IEEE. Google Scholar↗

[25] 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)*, 240-241. IEEE. Google Scholar↗

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

[27] 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)*, 1652-1657. IEEE. Google Scholar↗

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

[29] 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. Google Scholar↗

[30] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124-136. Google Scholar↗

[31] 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. Google Scholar↗

[32] 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. Google Scholar

[33] 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. Google Scholar

[34] 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. Google Scholar

[35] 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)*, 1-6. IEEE. Google Scholar

[36] 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*, 1544-1551. IEEE. Google Scholar

[37] 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. Google Scholar

[38] 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. Google Scholar

[39] Patacchiola, M., & Cangelosi, A. (2017). Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognition*, *71*(1), 132-143. Google Scholar

[40] 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. Google Scholar

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

****