SELECTION AND EFFICIENT USE OF LOCAL FEATURES FOR FACE AND FACIAL EXPRESSION RECOGNITION IN A CORTICAL ARCHITECTURE

[1] P.S.Sumathi,
[1] Research Scholar,
[1] Research Centre in Computer Science,

[2] Dr. D.Christopher Durairaj,
[2] Associate Professor in Computer Science,
[1&2] Research Centre in Computer Science, [1&2] V.H.N.S.N College(Autonomous),
[1&2] Tamilnadu,India.

Abstract

Emotional health plays very vital role to improve people's quality of lives, especially for the elderly. Negative emotional states can lead to social or mental health problems. To cope with emotional health problems caused by negative emotions in daily life, we propose efficient facial expression recognition system to contribute in emotional healthcare system. Thus, facial expressions play a key role in our daily communications, and recent years have witnessed a great amount of research works for reliable facial expressions recognition (fer) systems. Hence, facial expression analysis from video data is considered to be a very challenging task in the research areas of computer vision, image processing, and pattern recognition. In this paper proposed to selection and efficient use of local features for Face and facial expression recognition in a Cortical architecture process.

Keywords: Facial Expression, Recognition, Generic object, Detection, Neural Network,

1. Introduction

There are growing physiological and practical confirmations that demonstrate usefulness of part (e.g., local feature) based approaches in nonexclusive protest recognition which is hearty to variability in appearance because of impediment and to changes in posture, size and illumination. It is no uncertainty clear that low dimension features, for example, edges are important and used in the greater part of visual recognition tasks. In any case, there are just a couple of concentrates that address economical and efficient use of intermediate visual features for larger amount intellectual capacity. In this chapter, inspired by cortical processing, we will address the issue of efficient selection and economical use of visual features for face recognition (FR) as well as facial expression recognition (FER). We demonstrate that via training our recently proposed hierarchical neural system architecture (altered convolutional neural systems: MCoNN) for face recognition (FD), higher request visual capacity, for example, FR and FER can be organized for shared use of such local features. The MCoNN is not quite the same as those recently proposed systems in that training is done layer by layer for intermediate as well as global features with resulting responsive field size of neurons being larger for higher layers. More elevated amount (e.g., more unpredictable) features are defined as far as spatial arrangement of lower level local features in a preceding layer. In the chapter, we will define a typical framework for more elevated amount subjective capacity using the same system architecture (i.e., MCoNN) as substrate as pursues. We will demonstrate two examples of learning local features suitable for FD in our MCoNN (Matsugu and Cardon, 2004). One approach is heuristic, directed training by showing exemplar local features or patches of images, and the other is unsupervised training using SOM (self-organizing map) combined with managed training in MCoNN. In the proposed framework, both FR and FER use regular local features (e.g., corner like end-stop structures) learnt from exemplary image fragments (e.g., mouth corners, eye-corners) for FD. Specifically, in Section 3, spatial arrangement information of such local features is extracted certainly for FR as feature vectors used in SVM classifiers. In the case of FER depicted in Section 4, spatial arrangement of regular local features is used expressly for standard based analysis. We will appear, by simulation, that learnt features for FD end up being useful for FR and FER as well.

2. Learning Local Feature For Generic Object Detection

2.1 Modified convolutional neural network (MCoNN)

Convolutional neural systems (CoNN), with hierarchical feed-forward structure, comprise of feature detecting (FD) layers, each of which pursued with a feature pooling (FP) layer or sub-sampling layer. CoNN as well as Neocognitrons have been used for face detection and recognition.

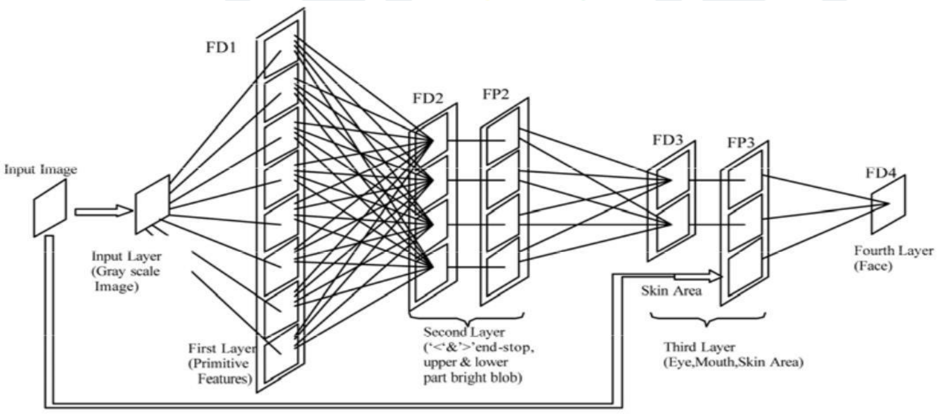


Figure 1. Modified convolutional neural network (MCoNN) architecture for face detection

Proposed architecture in Figure 1 accompanies the property of vigor in object recognition, for example, translation and deformation invariance as in understood neocognitrons, which also have similar architecture. The MCoNN contains the same three properties as the original CoNN as well as Neocognitrons: local open fields, shared weights, and alternating feature detection/pooling mechanism to recognize some intermediate (as in local however not very basic) local features. Those properties are can be broadly found in cortical structures. Feature pooling (FP) neurons perform either maximum value detection as in Riesenhuber and Poggio and Serre et al or local averaging in their responsive fields of appropriate size. Our model (MCoNN) for face detection as appeared in Figure 1 is not quite the same as traditional ones in many aspects. To begin with, it has just FD modules in the base and best layers. The intermediate features recognized in FD2 comprise an arrangement of figural alphabets. Local features in FD1 are used as bases of figural alphabets, which are used for eye or mouth detection. Face detecting module in the best layer is sustained with an arrangement of yields from facial part (e.g, for example, eye, mouth) finders as spatially requested arrangement of local features of intermediate unpredictability. Second, we don't train FP (or sub-sampling) layers (FP neurons perform either maximum value detection or local averaging in their responsive fields). Third, we use a detection aftereffect of skin shading area as input to the face detection module in FD4. The skin area is obtained basically by thresholding of tone data of input image in the range of [-0.078,0.255] for the full range of [-0.5,0.5], which is very broad indicating that skin shading feature plays just auxiliary part in the proposed framework. Third, in our MCoNN show, in contrast to the original CoNN, local features to be distinguished in particular layers are pre-defined, and trained module by module (i.e., for each local feature class) for specifi category of local features; edge-like features in the principal layer, and then in the second layer, corner-like structures (i.e., '<' and '>' end-stop), elongated masses (i.e., upper part splendid mass, and lower part brilliant mass) are identified. The second and third layers are made out of feature detecting layer and feature pooling layer as in original CoNN and Neocognitrons. Local features identified in the second layer establish some alphabetical local features in our framework, and details will be explained in the following segment. Eye and mouth features are recognized in the third layer. Finally, a face is identified in the forth layer using yields from the third layer and skin area data defined by some confined range of shade and saturation values.

The training continues as pursues. As in, training of the MCoNN is performed module by module using fragment images as constructive data extracted from freely available database (e.g., Softpia Japan) of in excess of 100 people. Other irrelevant fragment images extracted from background images are used as negative samples. In the initial step, two FD layers from the base, namely FD1 with 8 modules and FD2 with 4 modules, are trained using standard back-propagation with intermediate local features (e.g., eye corners) as positive training data sets. Negative examples that don't comprise the corresponding feature category are also used as false data. Specifically, we trained the FD2 layer, the second from the base FD layer to form finders of intermediate features, for example, end-stop structures or masses (i.e., end-stop structures for left and right side and two sorts of horizontally elongated masses (e.g., upper part brilliant, bring down part splendid) with varying sizes, rotation (up to 30 deg. with rotation in-plane axis as well as head axis). These features for training are fragments extracted from face images. More mind boggling local feature finders (e.g., eye, mouth locators, yet not limited to these) are trained in the third or fourth FD layer using the patterns extracted from transforms as in the FD2 layer. Because of these training successions, the best FD layer, FD4, learns to locate faces in complex scenes. The span of partial images for the training is set with the goal that just a single class of explicit local feature is contained. The quantity of training data set is 14847 including face images and background image for FD4 module, 5290 for FD3, and 2900 for FD2.

3. Proposed Component Based Face Recognition

Proposed face recognition framework uses intermediate features extracted from face detection framework using MCoNN, which are bolstered to SVM for classification. This combination of MCoNN with SVM is similar in soul to late works by Serre et al. (2007) and Mutch and Lowe (2006). Figure 2 indicates detailed structure of the MCoNN for face detection as well as face recognition. Here, we depict feature vectors and the technique for their generation in face recognition. A feature vector, F, used in SVM for face recognition is a N dimensional vector, combined from an arrangement of local yield circulations, F1 (as appeared in Figure 2, in a module detecting edge-like feature in FD1 layer in addition to yield conveyances, F2, (as appeared in Figure 2(2)) of two intermediate-level modules detecting eye and mouth in FD2 layer. Along these lines, F = (F1, F2) where F1 = (F11, … , F1m) and F2 = (F21, … , 2n) are combined vectors formed by component vectors, F1k (k=1, ..., m) and F2k (k=1, ..., n), individually.

Each component vector speaks to plausibility or nearness of explicit class of local feature in an assigned local area. Measurement of a component vector is the area of a rectangular locale as in Figure 3. Subsequently measurement of feature vector, N, is the total summation of individual elements of component vectors. In particular, F1 =(F11, F12, … , F1,15), and local areas, total number of assigned areas being 15 as in Figure 9 (1), for component vectors are set around eye, nose, and mouth, using the identified eye location from the MCoNN. F1 reflects shape information of eye, mouth, and nose. F2 = (F21, F22, F23), and each component vector reflects spatial arrangement of eye or eye and nose, and so forth., depending on how local areas in FD2 (e.g., positions and size) are set. The methodology for feature vector generation is summarized as pursues. To begin with, we define an arrangement of local areas for FD1 as well as FD3 modules based on the CNN yield in FD3 modules for eye and mouth detection. Places of local areas in FD1 module are set around explicit facial components (i.e., eyes, mouth) as illustrated in Figure 3 (1).

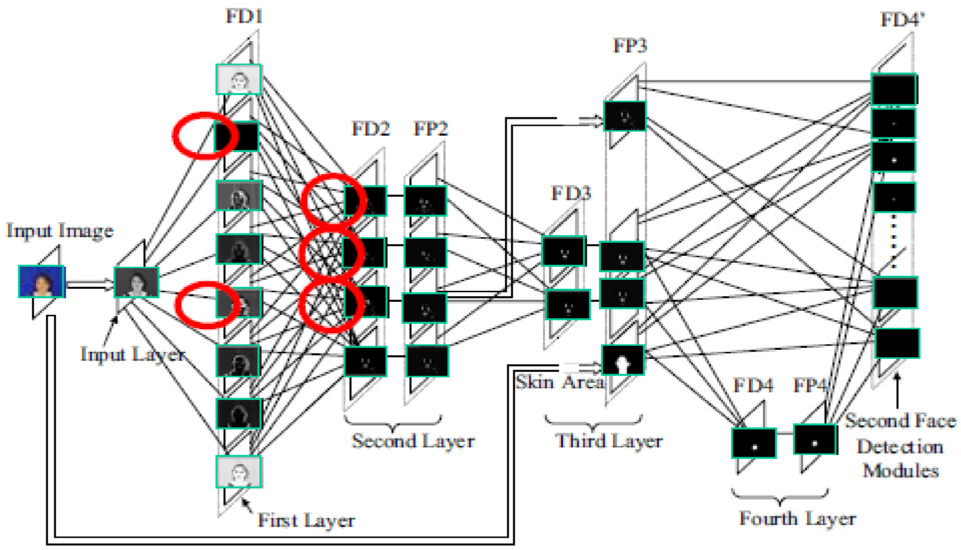


Figure 2: MCoNN for face recognition and facial expression recognition. Outputs from encircled modules in FD1 and FD2 layers are used for face recognition

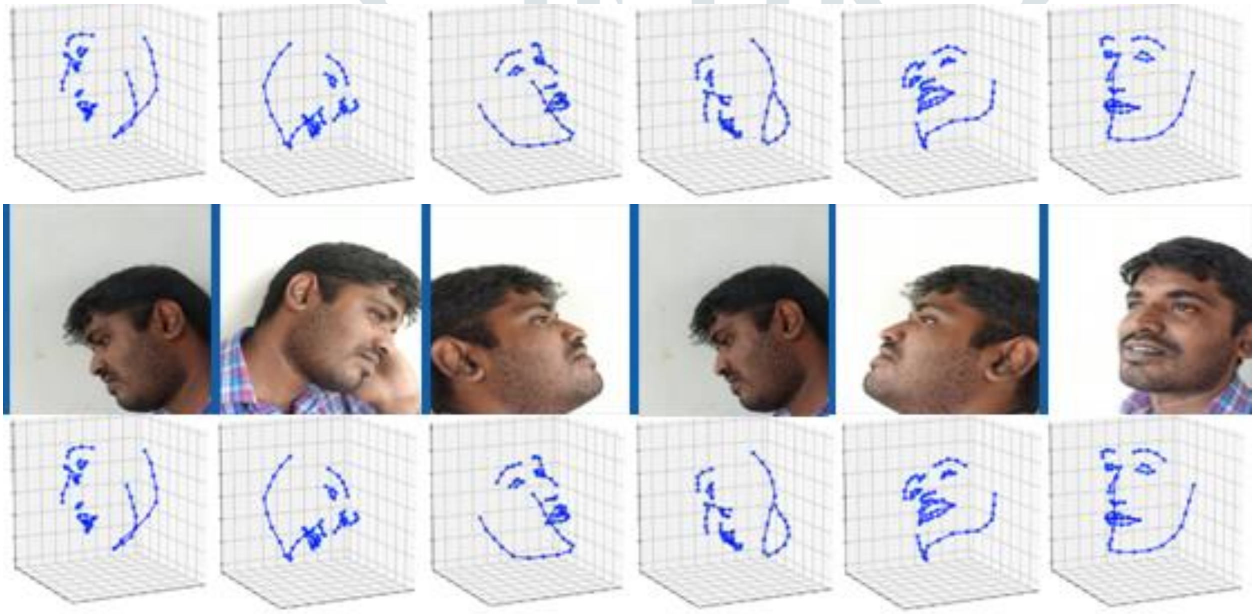


Figure 3: Local image fragments for training the second and third layers of MCoNN

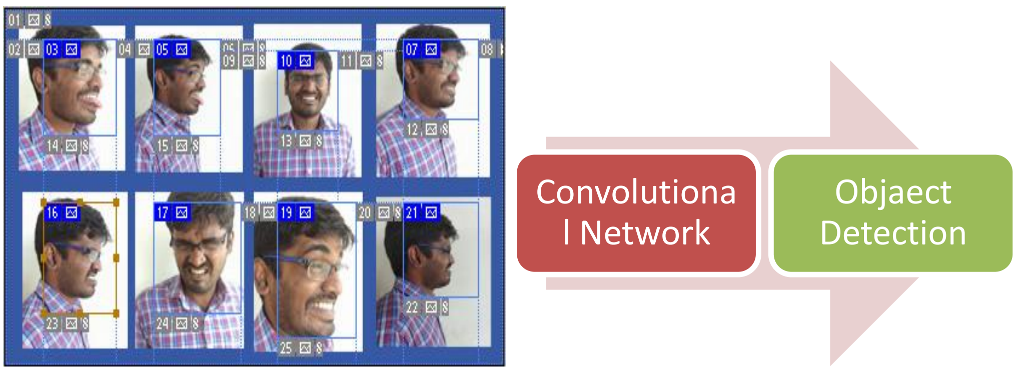


Figure 4: Schematic diagram of learning system for generic object recognition

The measure of input image is of VGA, and the extent of local areas for FVs is 15 x15, 125 x 65, or 45 x 65 depending on the class of local features. As indicated in Figure 4, the quantity of local areas for FD1 feature and FD2 feature is fourteen and two, separately. The quantity of FVs for one individual is 30, which are obtained under varying image capturing conditions with the goal that measure, present, facial expression, and lightning states of particular faces are marginally extraordinary.

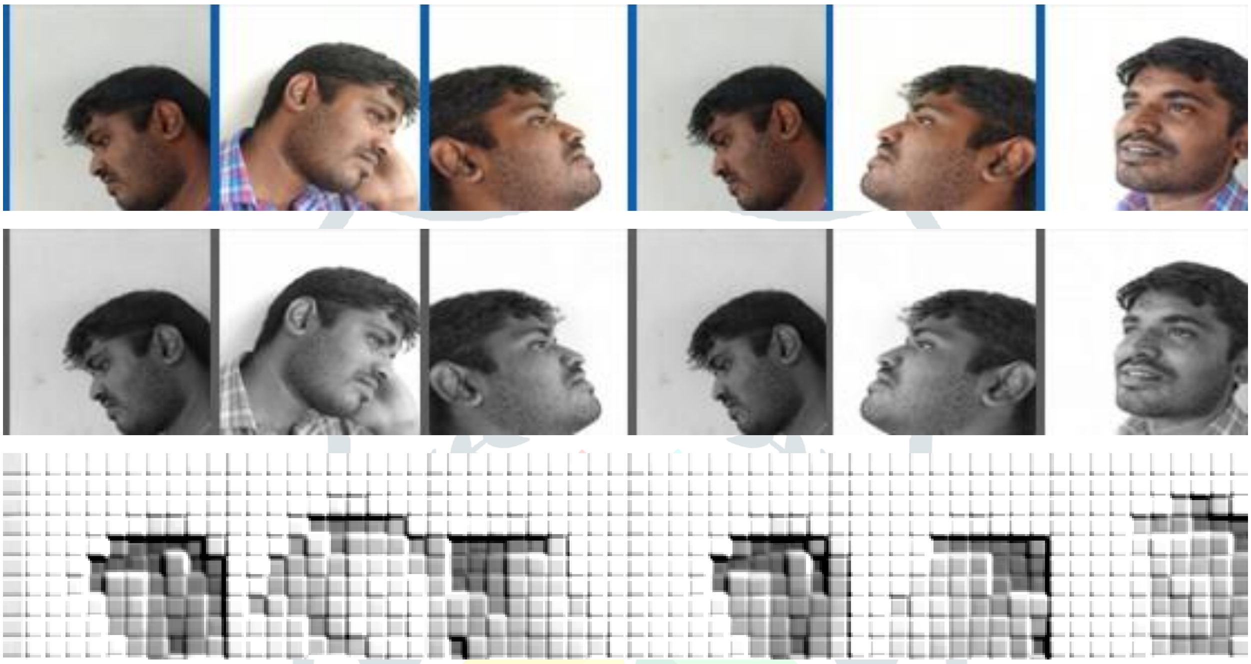


Figure 5. Intermediate output from MCoNN (1):input image, (2) output example from FD1, (3) intermediate outputs from encircled modules of FD2

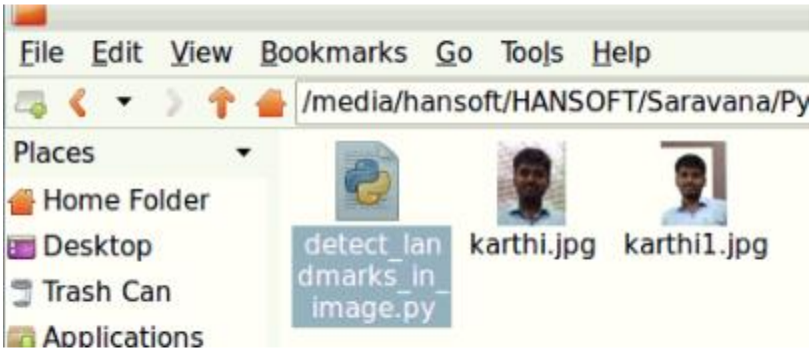
Face image database used ofr training and testing is in-house DB (10 subjects, 1500 images) and PIE database (we used part of the DB: 15 subjects 60 images) by CMU. We compared results obtained from McoNN's intermediate yields with those obtained from raw data using the same local area as in Figure 3. ROC bends in Figure 5 obtained for in-house face database demonstrate that using intermediate yields rather than raw data give better performance. Using the same dataset, we compared our model with commercially available software which is based on DLM (Wiskott et al., 1997). The recognition rate ended up being almost the same for the relative size of 0.8 to 1.2, while F.A.R. is marginally inferior to our model (i.e., F.A.R. isn't flawlessly zero), suggesting that our model involving a lot easier operations equals to the performance of outstanding amongst other models (Matsugu et al., 2004).

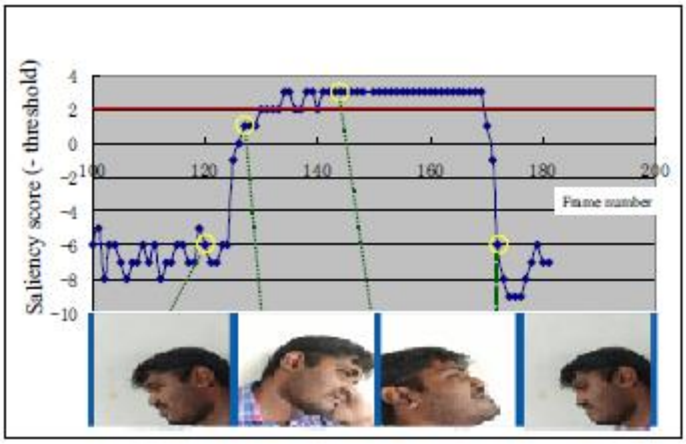
Facial expressions as manifestations of emotional states, in general, will in general be diverse among individuals. For example, smiling face as it appears may have changed emotional implications for various people in that 'smiling face', seen by others, for some individual does not necessarily speak to really smiling state for that individual. Just a couple of algorithms (e.g., Ebine and Nakamura, 1999) have addressed heartiness to such individuality in facial expression recognition. Moreover, in request for facial expression recognition (FER) to be used for human-PC interaction, for example, that algorithm must have great ability in dealing with variability of facial appearance (e.g., posture, size, and translation invariance). Most algorithms, up until this point, have addressed just a part of these issues (Wallis and Rolls, 1997). In this examination, we propose a framework for facial expression recognition that is strong to variability that originates from individuality and viewing conditions. Recognizing facial expression under unbending head developments was addressed by (Black and Yacoob, 1995). Neural system display that learns to perceive facial expressions from an optical stream field was accounted for in (Rosenblum et al., 1996). Guideline based framework was accounted for in (Yacoob and Davi s, 1996) and (Black and Yacoob, 1997), in which primary facial features were tracked all through the image succession. As of late, Fasel (2002) has proposed a model with two independent convolutional neural systems, one for facial expression and the other for face character recognition, which are combined by a MLP.

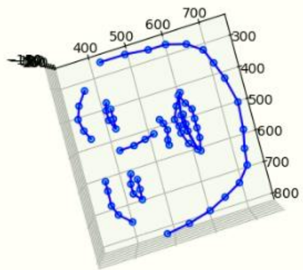
3.1 Facial expression recognition using local features extracted by MCoNN

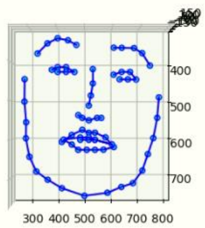
We appear, in this segment, proposed rule-based processing plan to enhance subject independence in facial expression recognition. We found that some of lower level features extracted by the first FD layer of MCoNN for face detection as well as face recognition are also useful for facial expression recognition. Primary features used in our model are horizontal line sections made up of edge-like structures similar to step and roof edges (extracted by two modules in FD1 layer, hovered in Figure 3 representing parts of eyes, mouth, and eyebrows. For example, changes in distance between end-stops (e.g., left-corner of left eye and left side end-stop of mouth) within facial components and changes in width of line sections in lower part of eyes or cheeks are identified to obtain saliency scores of an explicit facial expression. Primary prompts related to facial actions adopted in our facial analysis for the detection of smiling/laughing faces are as per the following.

1. Distance between endpoints of eye and mouth gets shorter (lip being raised)
2. Length of horizontal line fragment in mouth gets longer (lip being extended)
3. Length of line fragments in eye gets longer (wrinkle around the tail of eye gets longer)
4. Gradient of line fragment connecting the mid point and endpoint of mouth gets more extreme (lip being raised)
5. Step-edge or splendor inside mouth area gets increased (teeth being appeared)
6. Quality of edges in cheeks increased (wrinkle around cheeks being developed)


File Edit View Bookmarks Go Tools Help
/media/hansoft/HANSOFT/Saravana/Py
Places
Home Folder
Desktop
Trash Can
Applications
detect_landmarks_in_image.py
karthi.jpg
karthi1.jpg


Saliency score (- threshold)
Frame number
4
2
0
-2
-4
-6
-8
-10
100
120
140
160
180
200


400
500
600
700
300
400
500
600
700
800


400
500
600
700
300 400 500 600 700 800

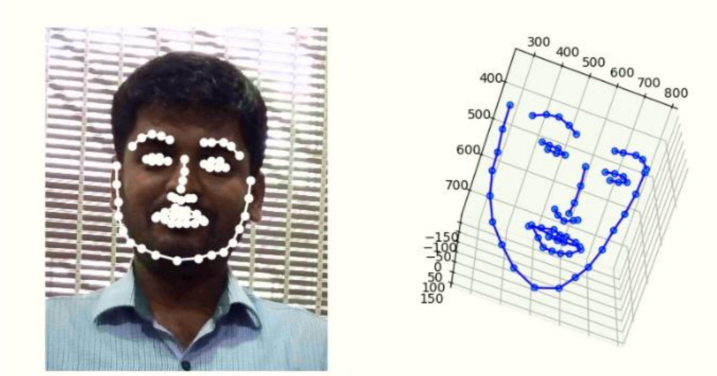


Figure 6: Normalized saliency score subtracted by constant value for smiling face detection

We use these numerous prompts as supporting proof of explicit facial expression (i.e., grin). Each sign was scored based on the level of positive changes (i.e., designated changes as offered above) to the emotional state (e.g., happiness). Saliency score of explicit emotional state is calculated with weighted summation of particular scores, which is then thresholded for judging whether the subject is smiling/laughing or not. Greater weighting factors are given to signs of less individuality (i.e., more typical prompts across individuals): (I), (ii), and (v). Figure 6 demonstrates an arrangement of normalized saliency scores indicating effective detection of smiling faces with an appropriate edge level. The system demonstrated the ability to discriminate smiling from talking based on the duration of saliency score above edge (longer duration infers greater plausibility of smiling; Matsugu et al., 2004). We obtained results demonstrating reliable detection of grins with recognition rate of 97.6% for 5600 still images of in excess of 10 subjects. In contrast to various approaches (Donato et al., 1999), invariance properties as far as translation, scale, and posture, inherent in our non-spiking adaptation of MCoNN (Matsugu et al., 2002), brings heartiness to dynamical changes both in head developments and in facial expressions without requiring unequivocal estimation of movement parameters. Because of the topographic property of our system which saves the position information of facial features from base to top layers, the translation invariance in facial expression recognition is in this manner inherently incorporated with our convolutional architecture with feedback mechanism for locating facial features. Specifically, intermediate facial features, for example, eyes and mouth are identified and used for tracking useful crude local features extracted by the base layer FD1 of MCoNN. Certain location information of eyes and mouth recognized in the MCoNN are used, through the feedback circle from the intermediate layer FP3, to confine the processing area of standard based facial feature analysis, which analyzes contrasts as far as at least six prompts. It worked out that the framework is very insensitive to individuality of facial expressions with the assistance of the proposed standard based processing using single however individual normal face. Because of the voting of scores for various signals as far as contrasts of facial features in neutral and emotional states, individuality is averaged out to obtain subject independence.

4. Experimental Results

4.1 Dataset 1

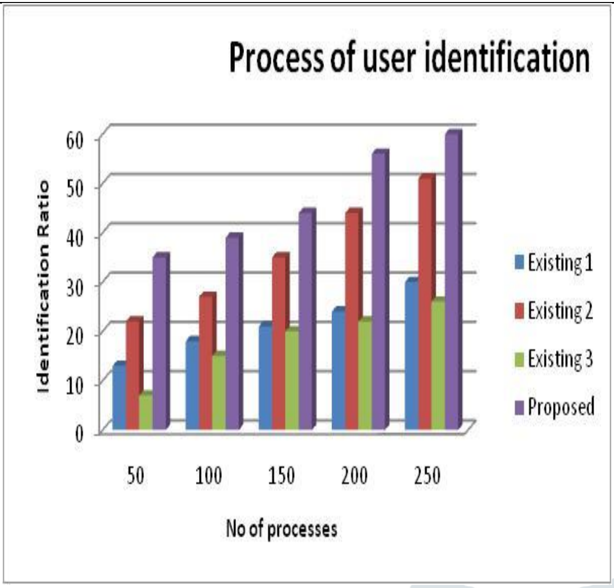


Figure 7: Comparison chart of process of user identification

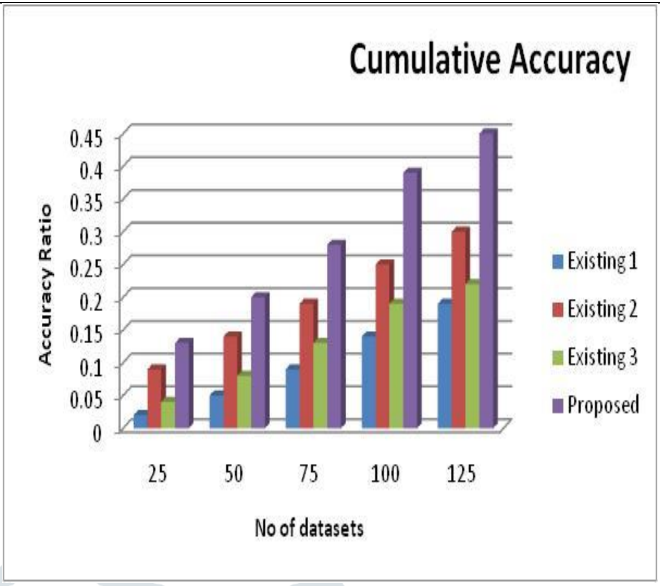


Figure 8: Comparison chart of cumulative accuracy

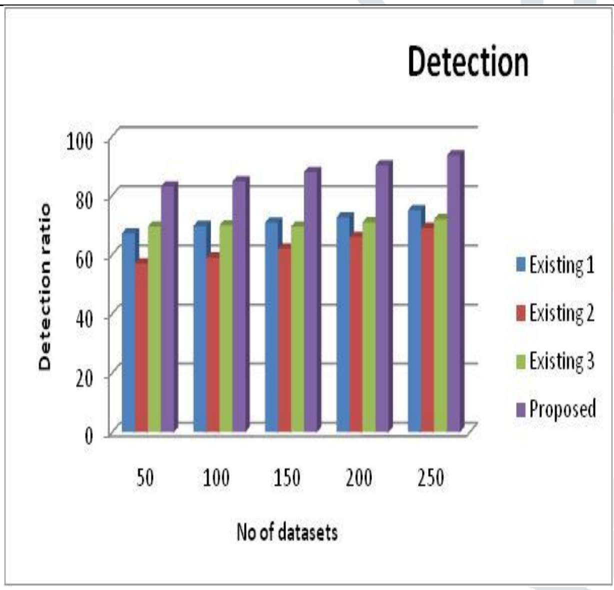


Figure 9: Comparison chart of detection

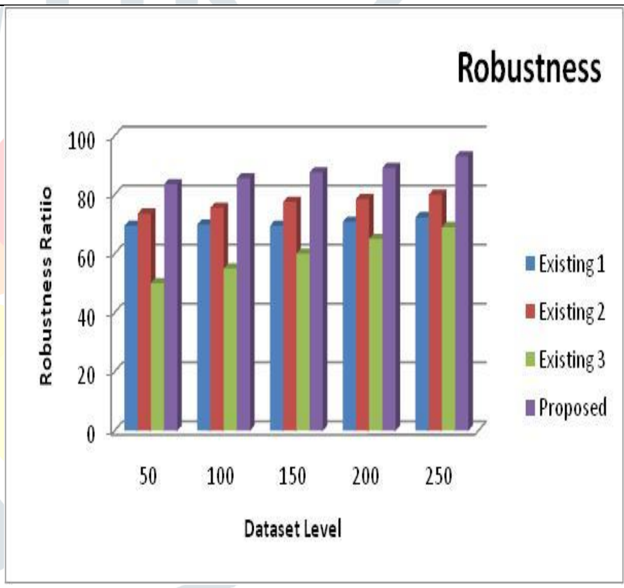


Figure 10: Comparison chart of robustnesss

Figure 7 the comparison chart of process of user identification shows the different values of existing and proposed method. No of processes in x axis and identification ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 13 to 30 existing 2 values start from 22 to 51 existing 3 values start from 7 to 26 and proposed method values start from 35 to 60. Figure 8 the comparison chart of cumulative accuracy shows the different values of existing and proposed method. No of datasets in x axis and accuracy ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 0.02 to 0.19 existing 2 values start from 0.09 to 0.3 existing 3 values start from 0.04 to 0.22 and proposed method values start from 0.13 to 0.45. Figure 9 the comparison chart of detection ratio shows the different values of existing and proposed method. No of datasets in x axis and detection ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 67.2 to 75 existing 2 values start from 57 to 69 existing 3 values start from 69.5 to 72 and proposed method values start from 83 to 93.6. Figure 10 the comparison chart of robustness shows the different values of existing and proposed method. No of datasets in x axis and robustness ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 69.5 to 72.4 existing 2 values start from 73.6 to 80.01 existing 3 values start from 50 to 69 and proposed method values start from 83.6 to 93.

4.2 Dataset 2

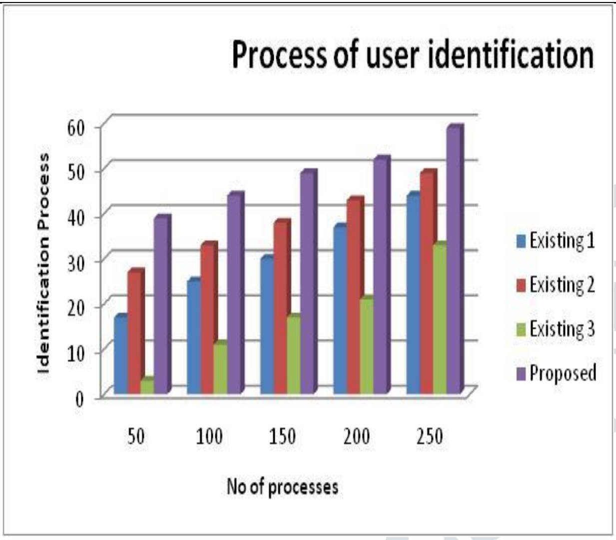


Figure 11: Comparison chart of User Identification

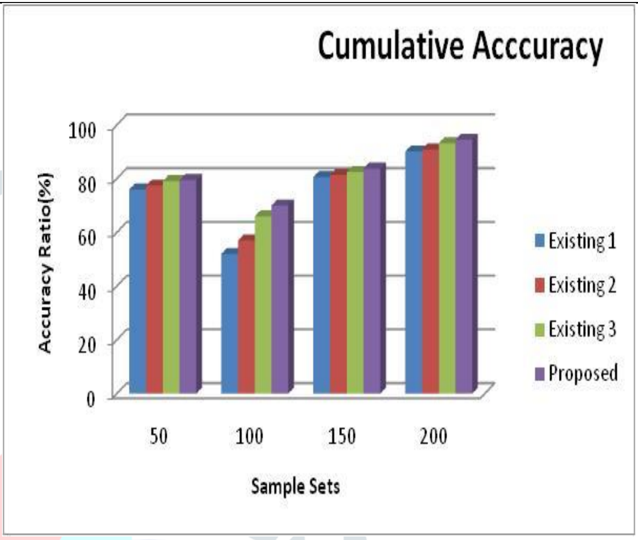


Figure 12: Comparison chart of cumulative Accuracy

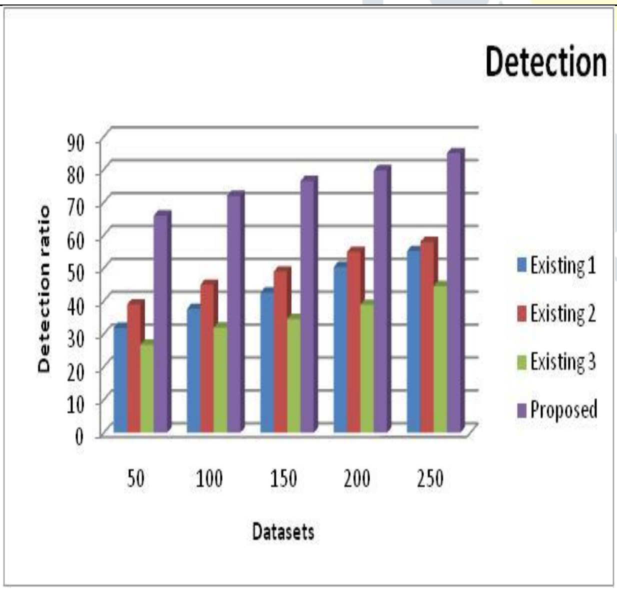


Figure 13: Comparison chart of detection ratio

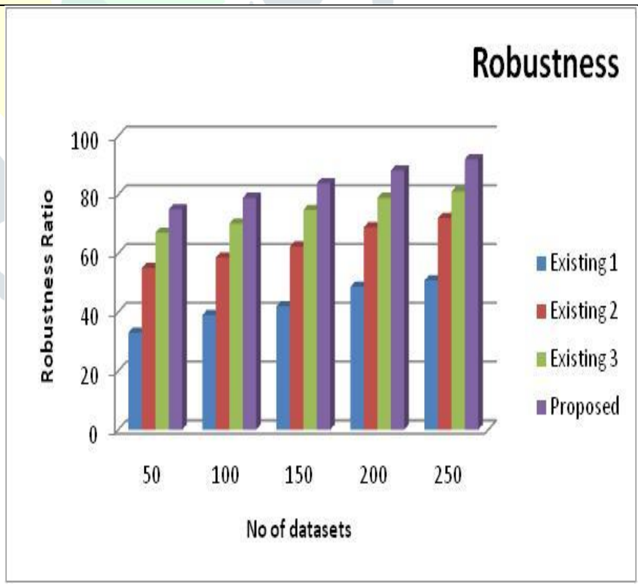


Figure 14: Comparison chart of robustness ratio

Figure 11 the comparison chart of user identification shows the different values of existing and proposed method. No of processes in x axis and identification process in y axis. The proposed method is better than the existing method. Existing 1 values start from 17 to 44 existing 2 values start from 27 to 49 existing 3 values start from 3 to 33 and proposed method values start from 39 to 59. Figure 12 the comparison chart of cumulative accuracy shows the different values of existing and proposed method. Sample sets in x axis and Accuracy ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 76 to 82 existing 2 values start from 52 to 74 existing 3 values start from 80.6 to 85 and proposed method values start from 90.1 to 97. Figure 13 the comparison chart of detection ratio shows the different values of existing and proposed method. Datasets in x axis and Detection ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 31.9 to 55.23 existing 2 values start from 39 to 58 existing 3 values start from 26.77 to 44.56 and proposed method values start from 66 to 85. Figure 14 the comparison chart of robustness ratio shows the different values of existing and proposed method. No of Datasets in x axis and Robustness ratio in y axis. The proposed method is better than the existing method. Existing 1 values start from 33 to 50.76 existing 2 values start from 55 to 72 existing 3 values start from 67 to 81 and proposed method values start from 75 to 92.06.

5. Conclusion

We looked into our recently proposed leaning techniques (unsupervised and supervised) for appropriate and shared (economical) local feature selection and extraction for generic face related recognition. In particular, we demonstrated feasibility of our hierarchical, component based visual pattern recognition show, MCoNN, as a certain constellation display as far as convolutional operation of local feature, providing a substrate for generic object detection/recognition. Detailed simulation think about demonstrated that we can realize face recognition as well as facial expression recognition efficiently and economically with satisfactory performances by using the same arrangement of local features extracted from the MCoNN for face detection.

References:

[1] T. Ojala and M. Pietikainen. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Trans on Pattern Analysis and Machine Intelligence, Vol. 24. No.7, July, 2002.

[2] Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2037–2041 (2006).

[3] AnaghaV.Malkapurkar, Rupali Patil, Prof. Sachin Murarka, A New Technique for LBP Method to Improve Face Recognition, International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, Volume 1, Issue 1, November 2011.

[4] Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Computer Vision, ECCV 2004 Proceedings, Lecture Notes in Computer Science 3021 (2004) 469-481.

[5] Caifeng Shan, Shaogang Gong , Peter W. McOwan Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing 27 (2009) 803–816 [6] C. Shan, S. Gong, P.W. McOwan, Robust facial expression recognition using local binary patterns, in: IEEE International Conference on Image Processing (ICIP), Genoa, vol. 2, 2005, pp. 370–373.

[7] Li Liu, Lingjun Zhao, Yunli Long, Gangyao Kuang, Paul Fieguth, Extended local binary patterns for texture classification : Image and Vision Computing 30 (2012) 86–99.

[8] Md. Abdur Rahim, Md. Najmul Hossain, Tanzillah Wahid & Md. Shafiul Azam: Face Recognition using Local Binary Patterns (LBP), Global Journal of Computer Science and Technology Graphics & Vision, Volume 13 Issue 4 Version 1.0 Year 2013.

[9] Bo Yang, SongcanChen, A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image: Neurocomputing120(2013)365–379.

[10] Timo Ahonen, Abdenour Hadid, and Matti Pietik¨ainen, Face Recognition with Local Binary Patterns, Machine Vision Group, Pattern Analysis and Machine Intelligence 24 (2002) 971–987.

[11] Chetan Ballur & Shylaja's "Application of Local Binary Pattern and Principal Component Analysis for Face Recognition"International Journal of Electrical, Electronics and Data communication, issn (p): 2320-2084, volume-1, issue-, july-2013.

[12] Fabrice Bourel & C.C. Chibelushi's "Facial Expression Recognition:A Brief tutorial Overview" , School of Computing, Staffordshire University, Beaconside, Stafford ST18 0DG, ORSYP, 101 quartier Boieldieu, La Défense 8, F-92042 Paris La Défense Cedex, France.