International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 1, January - February 2017 **ISSN 2278-6856**

Facial Semantics Recognition Method for Content based Video Retrieval Systems

B. S. Daga[1], A. A. Ghatol[2] ,V.M.Thakare[3]

[1]Associate Professor, Computer Engineering Department,
Fr. Conceicao Rodrigues College of Engineering, Mumbai,India

[2]Former VC, Dr.BabasahebAmbedkar Technological University, Lonere, India

[3] Professor and Head in Computer Science, Faculty of Engineering & Technology, Post Graduate Department of Computer Science, SGBAmravati University, Amravati,India

Abstract

With growing video databases, the accurate facial expression identifier systems are proving their importance. State of the art literature suggests the use of facial texture pattern while identifying the facial expression from a image frame or photograph, and exercise with moving geometric patterns while dealing with dynamics expression identification from a video. Accordingly they make use of local binary patterns (LBP) from a image frame or facial landmark tacking (FLT) to extract out the semantics from a video database. Actual classification and identification of expression is then performed by support vector machine (SVM) based classifier. This work primarily presents a comparative assessment of LBP and FLT methods as semantic means for video retrieval systems. Moreover herein introduces possible implication of probabilistic neural network (PNN) as more effective alternative to conventionally practiced SVM for classification problem. The modular system so developed has been tested with well-established database and comparative results of the methods are presented. Results presented indicate that the implication of order of magnitude faster PNN can be efficient replacement of SVM. Moreover, considering the near future technologies, use of facial landmark tracking technique is most viable solution to yield accurate and meaningful results. The facial expression identification based on different facial semantics has been one of well-researched areas for quite some time. However, expecting a proven system implication into real practice needs to focus on systems performance and processing speed. The usefulness of current work lies in its contribution towards presenting a comparative study of classifiers for content based video retrieval systems.

Keywords: Facial Landmark Tracking, Support Vector Machine, Neural Network, Semantic, Facial Expressions, Content Based Video Retrieval System

1. INTRODUCTION

Automatic identifying the human facial expressions is of prime importance in many computer related applications. A feature is a characteristic that can capture a certain visual property of a face. This feature extraction is the key function in content based video retrieval systems[3]. The CBVR (Content Based Video Retrieval) have received intensive attention in the literature of video information retrieval since this area was started couple of years ago, and consequently a broad range of techniques has been proposed. An effective search system would include to betterment of computers and robots to serve better to humans. Not only that but at a day to day level, where humans are involved in auto chats with computing systems, would find it very useful . Today's age of online market, where sellers contact to human customers through computer interface over the internet, it would be of huge gain for sellers to know human customer emotions by identifying the facial expressions.

Basically, when a facial expression forms there are many facial body elements and muscles contribute to generate it. Such as the chic bones, eye lines, etc. Even the differential eye movement or the blinking rate of eyes conveys viable information about the persons feeling and the emotional state of the subject. Different movement of muscles can contribute to different emotion expressions totally altogether. Such as the lip movement, as even used in text emoticons, an upward movement of lip corners expresses the smile and happy face. On the other hand downward movements of lip corner points denote the unhappy or sad facial expression. Whereas the differential movement of lip corners suggests contempt expressions look. Similar to eyes, the brows, lips, and chic muscles show key movements during an expression. As computers technology is getting mature day over day, its percolation in human life is advancing as well. Machine learning is one of important section where, researchers striving their best to improve on human-computer interface. The budding ability of computers to detect human facial expression and there behind emotion is one of the primary importance. And so is its improvisation attracts researchers and programmers.

There had been many of facial expression identification systems reported in literature so far. They all work based

on different principles. Some try to measure key facial landmarks movement during expression build-up. Some try to extracts facial information in terms of texture change under certain image processing techniques. There could be a multiple combinations of holistic and component based approaches possible. Facial landmark tracking (FLT) and Local binary pattern (LBP) are two most reliable techniques as observed from literature. FLT technique works on basis of evolution of facial landmarks movements as an expression build up. Thereby, it is mostly expected to have better dynamic response as compared to LBP, where LBP makes use of static images only. In current paper, an effort has been made to do a comparative assessment of FLT and LBP methods for semantic expression identification. Moreover, most design pattern common classification algorithms used in literature are based on support vector machines (SVM). In current paper, probabilistic neural networks (PNN) had been tried parallel to see its performance in comparison with conventional SVM technique.

Famous American psychologist [1] was the pioneer behind identifying and classifying various human facial expressions. His contribution to creating the humongous database of human expressions proved to be an asset for many modern machine learning researchers and scientists. As per his historical reports, human facial expressions have been classified in 7 key semantics features. This semantic based content identification helps in video data extraction. Facial expressions can be considered important cues which can be used in Video Retrieval [9]. He termed these set of seven facial features as "Seven Universal Facial Expressions of Emotion". Thereby, any automatic computer software system for facial expression classifier is expected to clearly distinguish between these seven features for a given data over a neutral face. These seven expressions are listed as displayed in referred in Fig. 1

2. RELATED WORK

The Facial Action Coding System (FACS) [2] is a method for measuring facial expressions in terms of activity in the underlying facial muscles. FACS is a system to taxonomies human facial movements by their appearance on the face, based on a system as adopted by the same psychologist. It is an index of facial expressions, but does not actually provide any bio-mechanical information about the degree of muscle activation. One has to understand that muscle activation is not part of FACS. Movements of nostrils, eyes, cheeks, lips, lip corners, head, neck, tongue, dimples, winks etc., in positive and negative directions are encoded to specific numbers and mapped to high level features. The retrieval framework can be accomplished through the development of philosophy based on semantic content model and semantic content extraction algorithms[13] Also, sometimes, their intensity of movements is accounted. Thus each of the expression would obtain a specific code or a summation of codes, as illustrated in this article.

Figure 1 Seven basic emotions and their universal expressions

Researchers [16] have demonstrated a system for facial expression detection. Moreover their developed system can do a generic application job. It could identify the face; do facial restructuring, coding of facial landmarks movements etc. The system is basically based on facial landmarks movement tracking. The performance of their system was exceptionally good in classifying the expressions. However, their results showed some flaws in confusing between the few expressions identification. This they had attributed to the fact that the facial landmarks sometime are not getting detected correctly. This may cause a wrong identification.

Support vector machine (SVM) are the most commonly popular machine learning algorithm used for classification problems. Basically, it supports a good mapping mechanism. The SVM primarily works on the basis of vector space. Before using for actual classification problem the SVMs are trained to classification categories with training data sets. For each training data SVM constructs a vector. These vectors when plotted in vector space demark the boundaries to certain categories. Thus once training is done every new data point vector would lie to certain category of classification problem. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. For expression identification problems SVM had always been the tool of choice for classification process among the researchers. Most of the researchers had demonstrated its use in literature [4,5,6].

Facial landmarks detection is not the only method in the field of expression detection. Part of literature is full of

works which do use facial texture information for expression detection. One of such method is use of local binary pattern (LBP) detection. The first evidence of LBP operator was introduced by researchers working in early era of machine learning [7], and was proved a powerful means of texture description. It is a simple image processing technique however proven very effective. In LBP operation, the human face, as builds the expression different pixels are illuminated to different brightness levels. This mainly happens due to the face exposed to constant camera light while muscles performing the movement. In such situation, if the facial extract from a video frame is fragmented, the pixels can be assigned to differential binary patterns. This on multiple combinatorial use of facial fragmentation, results in to a binary code. This is a rather simplistic approach due to its computational simplicity.

Researchers [8] proposed expression recognition system which was fully automatic and did achieve superior performance. They combined Facial texture method, LBP with geometric information, facial landmarks detection (FLT) technique. Combination of these two methods provided the benefit of static and dynamic pattern recognition.

Very recently in 2013, a group of researchers [15], tried to include a test dataset of seven distinct facial expressions and observed a quite successful recognition rate of their component based expression classifier algorithm. The datasets used by Hong et al. are generated using the JACFEE and JACNEUF dataset, which is certified by the Paul Ekman Group LLC [10]. Their objective was to test the system in identifying 7 universal facial semantics, which included the Ekman's expression of contempt along with the neutral face. The data set was consisting large enough data points with nearly 280 colorful images. Out of which 140 were the neutral faces. And remaining 140 were from 7 facial expressions. So as to count, 20 images per expression. They have employed four different image pre-processing techniques are used & listed as –

a. Grayscale Transformation
b. Local Binary Patterns (LBP)
c. Edges using Canny and Sobel
d. Feature points

The facial expressions were classified using a standard multiclass Support Vector Machine (SVM) and a pairwise adaptive multiclass SVM (pa-SVM) – which used pairwise adaptive model parameters. They primarily focused on the performance of SVM classification technique particularly when it comes to classifying the expression contempt with other expressions. Moreover they looked at the factors which could have impact on the performance of the classifier. The overall system performance as they had presented was applauding. They obtained a best correct classification rate of 98.57% with the contemptuous expression included.

While there had been several classification algorithm are available, namely, support vector machine (SVM), extreme learning machine (ELM), sparse representation of image (SRC), Probabilistic neural networks (PNN) that find applications in design pattern recognition of image classifications. In 2008, researchers [11] demonstrated application of PNN method for solving distinct pattern classification problems. In their article, they stressed out on the fact that PNN is predominantly a classifier since it can map any input pattern to a number of classifications. And compared to other state of the art classifiers, PNN provides advantage primarily with fast training process, an inherently parallel structure, guaranteed to converge to an optimal classifier as the size of the representative training set increases and training samples can be added or removed without extensive retraining. Probabilistic neural networks can be used for classification problems. When an input is presented, the first layer computes distances from the input vector to the training input vectors and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes.

Very recently a group of researchers had demonstrated the use of bio-dimensional empirical mode decomposition (BEMD)12 technique for facial expression identification. Their proposed methodology basically works by subtracting the facial image of a subject pertaining to specific expression from that of the same subject's neutral facial image. Moreover, they mainly focus on the selective facial parts of eyes and mouth mainly, since these are the parts of human face where the facial muscles indicate maximum movement or deformations while an expression is being developed on the face. The idea in their work for image subtraction basically works on the principle of image pixel subtractions. Most of the works as recorded in literature stress on different techniques being used for expression encoding which mainly can be classified under component based approaches and holistic approaches. However, there had been publications which also focused on performance improvements of the facial expression identification system. Again, very recently another group of researchers [17] had tried to exercise the implementation of scale invariant feature transform (SIFT) of the local facial features. They claimed that using SIFT kind of approaches the overall system complexity can be brought down with remarkable improvement in system accuracy and overall computational processing speed. Also, such a system, like SIFT helps in kind of

non-dimensional feature extraction method which would free the system from its performance dependency of image resolution or other image related issues which often bug the researchers and developers to drag away from the mainline work of expression identification.

3. PROPOSED SYSTEM

The system architecture of the modular facial expression recognition is illustrated in Fig. 2. As can be seen from the architectural block diagram, the system accepts input video sequence in frame by frame chunked format.

For each frame has to go through a pre-processing block which reorients the frame based on camera observed angle to normal form. Thereafter, the facial landmark tracking (FLT) and local binary pattern (LBP) feature extraction happens in frame-wise manner. The information thus extracted is further parsed through trained SVMs and PNN modules so as to produce final comparative results. Both, LBP and FLT feature extraction are well described in recent literature8, thereby only a brief of it has been included here in algorithm.

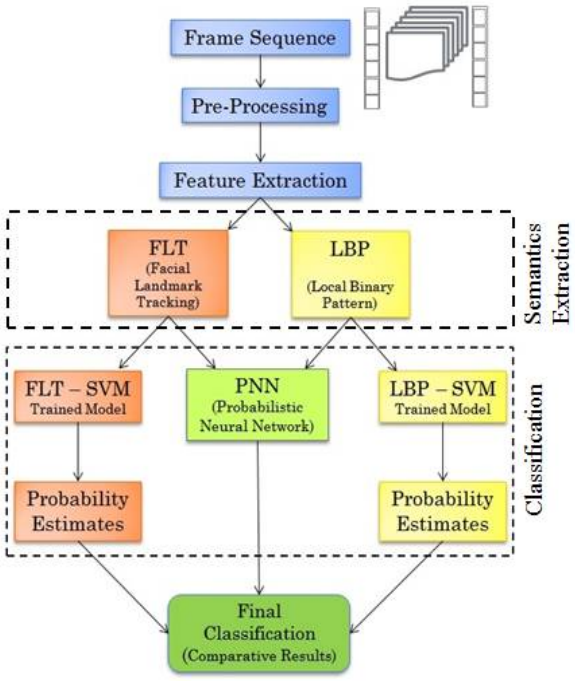


Figure 2 System Architecture

3.1 Algorithm

1. Pre-Processing: At very first, the system needs to have a normalized face of the captured frame. This is achieved by transforming each face as captured by camera frame to a constant coordinate frame. Then as per the facial landmarks detected, with help of it the face is mapped to a normalized neutral face.
2. FLT Detection: Detection and tracking of 49 facial landmarks as shown in Fig. 3.
3. FLT Landmark Trajectory extraction: As the facial expression builds up and captured by the camera, the movements of facial landmarks are sequentially recorded. Each subsequent movement of the facial landmarks is progressively summed up in order to construct the landmarks movement vector. This can be represent

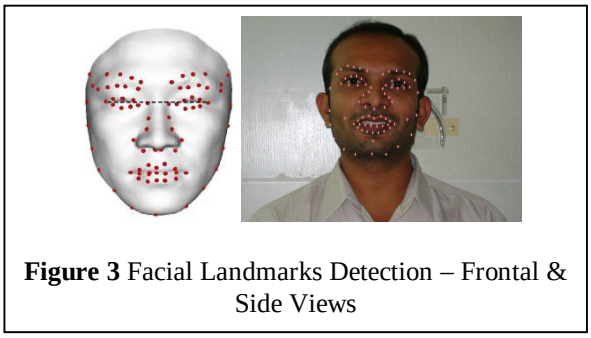


Figure 3 Facial Landmarks Detection – Frontal & Side Views

$$v_j^{(i)} = P_j^{(i+1)} - P_j^{(i)} \qquad (1)$$

Here, $v_j^{(i)}$ is the displacement vector between two sequential frames. $P_j^{(i)}$ is the spatial frame position of a feature landmark, j, at any i^{th} frame.

If within a frame sequence there are N numbers of frames, from starting of expression build up, to its apex, the progressive movements of facial landmarks are summed up as said. Thereafter, the frame sequence is fragmented into, say k equal parts. Each one matches a segment containing a number of frames. The number of segments depends on the accumulated distance V_j^{total}. Thereby, the accumulated distance V_j^{total} of P_j from the first frame to the last frame is:

$$V_j^{total} = \sum_{l=1}^{N-1} \| v_j^{(i)} \| \qquad (2)$$

The larger the distance is, the less frames would be taken. Assume the s^{th} segment contains frame n_{s-1} to frame $n_s - 1$. Then the motion vector w_j^s in the s^{th} segment is constructed as

$$w_j^s = \sum_{l=n_{s-1}}^{n_s-1} . v_j^{(i)} \qquad (3)$$

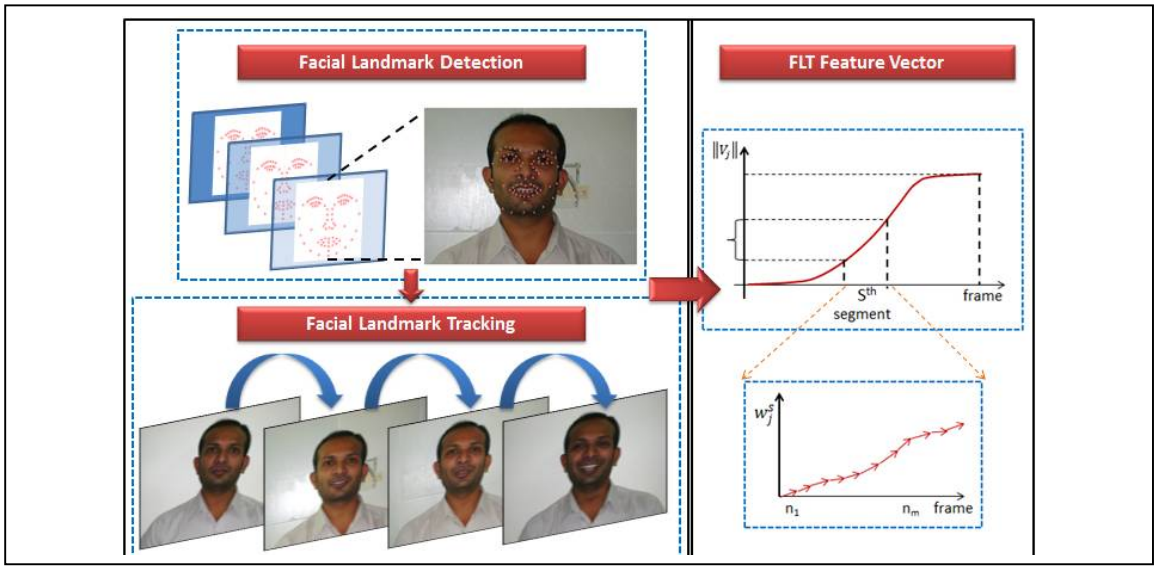
We consider altogether 49 such facial landmarks for tracking. Thus finally the overall motion vector of all facial landmarks for the s^{th} segment is constructed as

$$W_j^s = [w_1^s w_2^s \ldots\ldots w_{49}^s] \qquad (4)$$

As a results of concatenation of all the motion vectors represents the facial expression, which is the single feature vector to be used with support vector machine for further classification of expression. This feature vector is represented by

$$F = [W^1 W^1 \dots \dots W^k] \quad (5)$$

In present work, FLT_k^{adp} signifies the FLT operator. The symbolization of this operator can be understood as the superscript 'adp' denotes the adaptive segmentation. Whereas, the subscript, k tells that each frame sequence has been fragmented into k segments for the purpose. The concept described is depicted in Fig. 4.



4. Feature extraction using Local Binary Pattern (LBP): The multi-resolution uniform LBP descriptor is proposed to be used to construct a binary code for each pixel in the face image. The LBP code is defined as a function of the number of neighbors P and radius R around each pixel, i.e. $LBP^{u2}_{(P,R)}$. The superscript u2 stands for uniform-LBP since only the valid uniform codes are extracted as suggested in literature. Then, subdivide each face into 8 × 8 sub-blocks and compute the histogram of the obtained codes. The concatenation of these histograms of each of these blocks results in the final descriptor of the facial image as shown in Fig. 5.
5. Expression Classification: In current approach, along with the FLT and LBP feature based expression identification, a permutation of

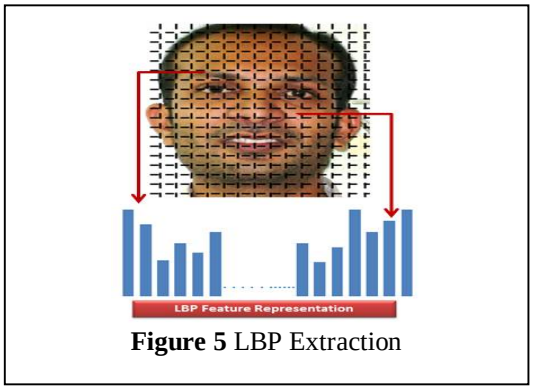


Figure 5 LBP Extraction

6. classification techniques are also employed, namely, support vector machine (SVM) and probabilistic neural network (PNN).
7. Whenever an expression is predicted, its probability of belonging to a certain class is also determined. To decide on a facial expression, we sum up the probability estimates generated from LBP and FLT trained models for each expression individually. We then choose the expression which maximizes the probability.
8. PNN is basically based on Baysian classification and classical estimators for probability density function. As also pointed out literature [11], the basic operation performed by the PNN is an estimation of the probability density function of features of each class from the provided training samples using Gussian Kernel. One can refer to literature [11] for detailed formulation and algorithm. These estimated densities are then used in a Bayes decision rule to perform the classification.

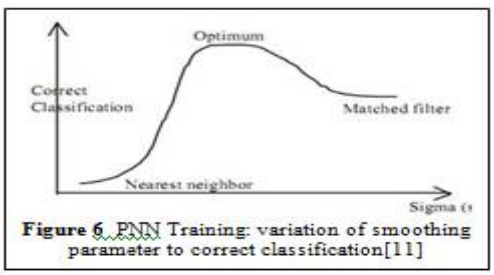


Figure 6 PNN Training: variation of smoothing parameter to correct classification[11]

9. Like a neural network functions, PNN has to be trained so as to arrive at the most efficiency classifier. The effectiveness of the classifier depends on the convergence of the PNN training. Thereby it is important to include as many cases as possible for the PNN training. A full trained PNN would converge to the well-defined classifier as the training set size increases. Basically training of PNN module means arriving at the most optimum value of sigma (σ) which is nothing but the smoothing parameter. Prior research work [11] had illustrated the variation of sigma and its correspondence to the correct classification, i.e. at the optimum value, as shown here in Fig. 6.

4.TEST RESULTS

The modular facial expression identification system thus developed had been tested with one of the most literature cited database, i.e. Cohn-Kanade AU-Coded facial expression or 'CK+' database [14]. It consists of expressions from subjects range in age from 18 to 30years, in total out of which 65% were female, 15% African-American and 3% Asian or Latino. Each subject has contributed with a video sequence of 23 continuous facial expression displays. This included a must of, angry, disgust, happy, sad, surprise and fear expressions.
Every set of video sequences, as available from database, had been tested, for FLT and LBP facial expression extraction method. Moreover, for both methods, FLT and LBP expression identification, a permutation of SVM and PNN classification technique was employed. Here Fig. 7 shows the system snapshot while testing on a random video.

Figure 7 System demonstration

Thus obtained results of confusion matrix bring out the effect of the FLT and LBP method effectiveness with SVM and PNN techniques. Figure 8 demonstrates their confusion matrix results. Primary inference of the results, which is comparing FLT and LBP methods, is that the FLT being more accurate than LBP. This is because as we are testing with frames from dynamic videos, FLT can better capture an expression from its building up to apex mode. This can be attributed to FLT extraction which do form feature vector from successive frame information. Whereas LBP extracts information frame by frame, treating them independently. The secondary observation is for comparing SVM and PNN. Results from Fig. 8 reveal that PNN having an edge over SVM in sense of accuracy however, it can be well confirmed with larger datasets. Average computation time for LBP+SVM was found to be more than 1 hour, which is highly undesirable for expression identification system. It was noticed that LBP methods are more time consuming. LBP+PNN model takes ~20 minutes. On the other hand FLT methods was order of magnitude faster, clocking < 1 min. Moreover, FLT+PNN had proved the quickest to respond in 15 sec.

5.CONCLUSION

The paper has presented the techniques of making use of local binary patterns (LBP) and facial landmark tracking (FLT) as the semantic means for facial expression identification. When it comes to identify facial expression from online videos, facial landmark tracking (FLT) has been found to be more precise, quicker than local binary pattern (LBP).

Moreover, the comparative assessment of performance of commonly known machine learning classification methods, Probabilistic Neural Network (PNN) and Support Vector Machine (SVM) has been presented as the additional vital combinatorial study of methods. Probabilistic neural network (PNN) is certainly a desired replacement to support vector machine (SVM) for expression identification system. Facial landmark tracking being demonstrated the more useful semantic builder for facial expression on video as input and Probabilistic neural networks (PNN) being the faster classifier, the combination of FLT and PNN forms the best modular expression identification system.

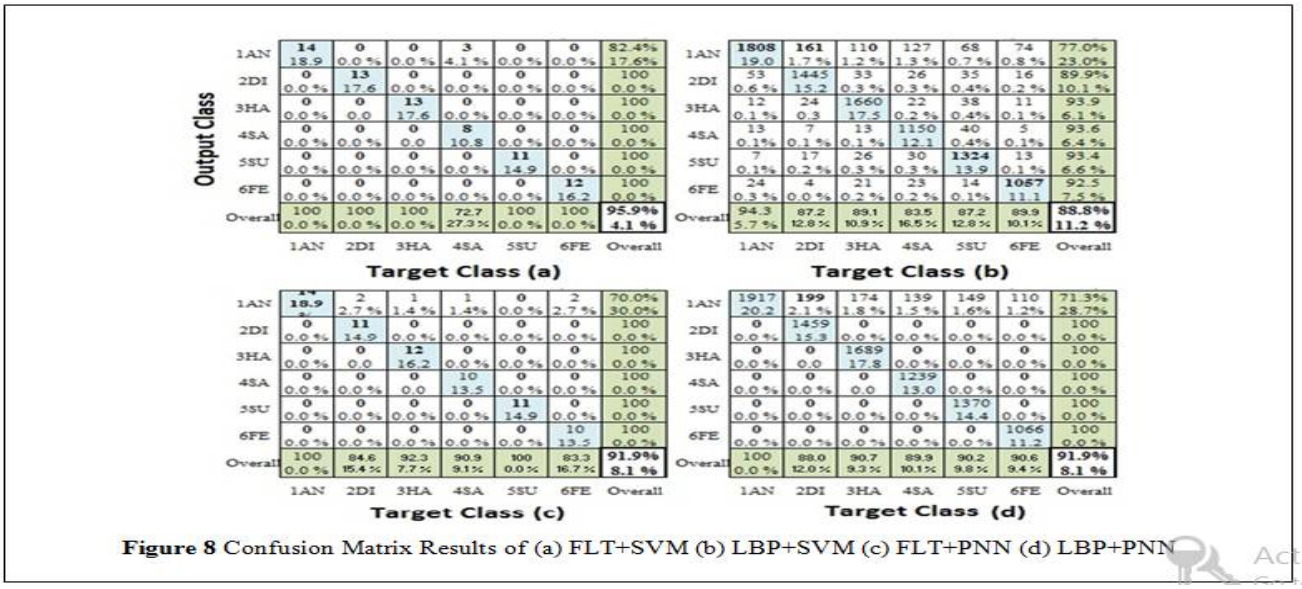


Figure 8 Confusion Matrix Results of (a) FLT+SVM (b) LBP+SVM (c) FLT+PNN (d) LBP+PNN

References

[1] Ekman P, Expression and the nature of emotion. Approaches to Emotion. Hillsdale, NJ: Erlbaum.1984. pp. 319-344.

[2] Ekman P, Friesen W V. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, 1978.

[3] Daga, Brijmohan. "Content Based Video Retrieval Using Color Feature: An Integration Approach." Advances in Computing, Communication, and Control. Springer Berlin Heidelberg, 2013. 609-625.

[4] ChangC, Lin C J. LIBSVM: a Library for Support Vector Machines, 2001.

[5] ShanC, Gong S, McOwan P W. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 2009, 27(6), pp. 803–816.

[6] ZhaoG, PietikainenM. Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern Recognition Letters, 2009, 30(12), pp.1117–1127.

[7] Ojala T, Pietikäinen, M, Harwood D, A Comparative Study of Texture Measures with Classification Based on Feature Distributions. Pattern Recognition, 1996, 19(3), pp.51-59.

[8] Weng C H, Lai S H, Online Facial Expression Recognition Based on Combining Texture and Geometric Information. IEEE International Conference on Image Processing (ICIP) Paris, 2014, pp. 5976-5980.

[9] Daga, B. S., and A. A. Ghatol. "Multicue Optimized Object Detection for Automatic Video Event Extraction." Indian Journal of Science and Technology 9.47 (2016).

[10] Paul Ekman Group LLC. JACFEE and JACNEUF dataset. http://www.humintell.com/for-use-in-research/.

[11] Emary I M, Ramakrishnan S, On the application of various aprobabilistic neural networks in solving different pattern classification problems. World Applied Sciences Journal, 2008,4(6), pp. 772 –780.

[12] SahaA, PukhrambamM, Pradhan S N, Facial Image Analysis for Expression Recognition by Bidimensional Empirical Mode Decomposition. Indian Journal of Science and Technology, August 2016, 9(31).

[13] Daga, B. S., and A. A. Ghatol. "Detection of objects and activities in Videos using Spatial Relations and Ontology based approach in Video Database System." International Journal of Advances in Engineering & Technology 9.6 (2016): 640

[14] CK+ Facial Expression Database. http://www.pitt.edu/~emotion/ck-spread.htmDate accessed: 26/11/2016

[15] Hong K, ChalupS, King R, A Component Based Approach for Classifying the Seven Universal Facial Expressions of Emotion. IEEE Symposium on Computational Intelligence for Creativity and Affective Computing, 2013, 4(1).

[16] Lanitis A, Taylor C J, Cootes T F. Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7), pp. 743–756.

[17] Kim S, Kim Y, Lee S, An improved Face Recognition Based on Scale Invariant Feature Transform (SIFT): Training for Integrating Multiple Images and Matching by Key Point's Descriptor-Geometry. Indian Journal of Science and Technology, September 2016, 9(35)

.

AUTHOR

B. S. Daga was born at Amravati, India in 1967. He received the Bachelor in Computer Engineering from SGB, Amravati University in 1990 and Master in Computer Science & Engineering from National Institute of Technology, Allahabad. He is currently pursuing the PhD degree with Department of Computer Engineering, SGB, Amravati University. He is also associated with Dept. of Science & Technology's(Government of India)Entrepreneurship Development Programs. His research interest includes Multimedia Systems, Data Mining, Artificial Intelligence, Computer Simulation and Modeling.

Dr.A.A.Ghatol was born at Amravati, India in 1949. He received the Bachelor in Electrical Engineering from Nagpur University in 1971 and Master and PhD in Electrical from IIT Mumbai in 1973 and 1984 respectively. He was the Vice Chancellor of Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra. He is the Best Teacher awardee of Govt. of Maharashtra for the year 1999. He has worked in various capacities at Central Govt, professional bodies and at social platform. He is known for his keen interest in interdisciplinary subjects in Engineering.

Dr.V.M.Thakare has received M.E. (Advance Electronics from Amravati University), P.G. DCM, from IICM, Ahamadabad and Ph.D. in Computer Science. He has been invited as a Keynote Speaker, Invited Speaker, Session Chair and Reviewer for more than 30 International & National Conferences & has number of publications in International Journals. He has been actively involved in the research in the area of Robotics and AI, Computer Architectures, ICT, SE.