ISSN (e): 2250 – 3005 || Vol, 04 || Issue, 12 || December – 2014 ||
International Journal of Computational Engineering Research (IJCER)

Facial Expression Recognition System: A Digital Printing Application

Mahasweta Mandal[1], Somnath Banerjee[2]
[1] Department of Printing Engineering, Jadavpur University, India
[2] Department of Computer Science and Engineering, Jadavpur University, India

ABSTRACT:
Human Computer Interaction (HCI), an emerging field of research in science and engineering, is aimed at providing natural ways for humans to use computers as aids. Humans prefer to interact with each other mainly through speech, but also through facial expressions and body gestures, to emphasize a certain part of that speech and display of emotions. The identity, age, gender, as well as the emotional state of a human being can be acquired from his/her faces. The impression that we receive from a reflected expression on face affects our interpretation of the spoken word and even our attitude towards the speaker himself. Although emotion recognition is seemingly an easy task for humans, it still proves to be a tough task for computers to recognize the user's state of emotion. Progress in this field promises to equip our technical environment with means for more effective interaction with humans and hopefully, in the days ahead, the influence of facial expression on emotion recognition will grow rapidly. The application of digital printing has rapidly grown over the past few years with substantial developments in quality. Digital printing has brought about fast turnaround times and printing on demand in terms of cost. In this paper, we describe the empirical study of the state-of-the-art classifier combination approaches, namely ensemble, stacking and voting. Each of these three approaches was tested with Naïve Bayes (NB), Kernel Naïve Bayes (k-NB), Neural Network (NN), auto Multi-Layer Perceptron (auto MLP) and Decision Tree (DT).The main contributions of this paper is the enhancement of classification accuracy of the emotion recognition task on facial expressions. Our person-dependent and person-independent experiments show that using these classifier combination methodologies provide significantly better results than using individual classifiers. It has been observed from the experiments that overall voting technique with majority voting achieved best classification accuracy.

KEYWORDS: *Human Computer Interaction, facial emotion recognition, facial expressions, facial action coding system, classifier combination, facial features, AU-Coded facial expression, CK+ database, digital printing.*

I. INTRODUCTION

Human beings can express their emotion through voice, body gestures and facial expression. But the most expressive way a human being displays his/her emotional state is through facial expressions. Facial expressions are the facial changes in response to a human being's internal emotional states, intentions, or social communications. Face is the primary signal system to show the emotion of a person. Face recognition and automatic analysis of facial expressions are one of the most challenging research areas in the field of Human-Computer Interaction (HCI) and have received a special importance. In the 1990s, there has been growing interest to construct automatic methods of recognizing emotions from facial expressions in images or video. Emotion as a private state is not open to any objective observation or verification. So, the recognition of the emotional state of a person is really a challenging issue. Relativity, categorization and identification of emotion are three crucial factors in emotion analysis. The relativity factor of emotion depends on the person's facial expression or state of mood whereas other two factors are comprehensible but require high technical affluence of computer intelligence. In recent years, computers and automated processing have had a considerable influence on prepress. The integration of prepress and press, as well as automation in printing and the integration of related processes, have also reached a certain maturity. The use of digital printing applications such as advertising, photos, architectural design etc. and integration of these applications into traditional print markets is rapidly expanding.

Emotion is the realm where thought and physiology are inextricably entwined, and where the self is inseparable from individual perceptions of value and judgment toward others and us. In the last few years, automatic emotion recognition through facial expression analysis has been used in developing various real life applications such as security systems, computer graphics, interactive computer simulations/designs, psychology and computer vision. In psychology, emotion is often defined as a complex state of feeling that results in physical and psychological changes that influence thought and behavior. Our emotions are composed of three critical components: a subjective component (how we experience the emotion), a physiological component (how our bodies react to the emotion), and an expressive component (how we behave in response to the emotion). These different elements can play a role in the function and purpose of our emotional responses.

There are a lot of words for the message persons get from the face(afraid, terrified, horrified, apprehensive, worried, to mention a few of those related to fear), but few to describe the source of those messages. Human beings do have the terms smile, grin, frown, squint, but there are relatively few such words that identify particular facial configurations, distinctive wrinkle patterns, or temporary shapes of the facial features. Without terms to refer to the face human beings are incapable in comparing or correcting their interpretations of facial expression. According to Ekman and Friesen[21] , the face gives more than one kind of signal to convey more than one kind of message. Sometimes the people can't differentiate the emotion messages from the other messages conveyed by the face. The face provides three types of signals: static(such as skin color)),slow(such as permanent wrinkles),and rapid(such as raising the eyebrows).The static signals include many more or less permanent aspects of the face like-skin pigmentation, coloration, the shape of the face, bone structure and the size,shape and location of the facial features(brows, eyes, noise, mouth).The slow signals include changes in the facial appearance which occur gradually with time and in addition to the development of permanent wrinkles, changes in muscle tone, skin texture and even skin, coloration occur with age. The rapid signals are produced by the movements of the facial muscles, resulting in temporary changes in facial appearance, shifts in the location and shape of the facial features and temporary wrinkles. These changes have been occurred on the face for a matter of seconds or fractions of a seconds. All three types of facial signals can be modified or disguised by personal choice, although it is hardest to modify the static and slow signals. So one can be misled, intentionally or accidentally, by rapid, slow or static signals. The face is not just a multi-signal system (rapid,slow,static) but also a multi-message system. In order to describe the emotion, it is referred to transitory feelings such as fear, anger, surprise, happiness etc. When these feelings occur, the facial muscles contract and visible changes are appeared on the face. Scientists have found that accurate judgments of emotion can be made from the rapid facial signals .It is important to note that the emotion messages are not transmitted by either the slow or the static facial signals; however these may affect the implications of an emotion message. Sometimes the facial expression analysis has been confused with emotion analysis in the computer vision domain. For emotion analysis, higher level knowledge is required. For example, although facial expressions can convey emotion, they can also express intention, cognitive processes, physical effort, or other intra- or interpersonal meanings. Computer facial expression analysis systems need to analyze the facial actions regardless of context, culture, gender, and so on. In regard to facial expressions of emotion we should have a higher knowledge on rapid facial signals and their distinctive messages.

II. MOTIVATION AND RELATED WORK

Facial expression is one of the most significant ways for human beings to communicate their emotions and intentions. The face can express sooner than people verbalize or even realize their feelings. Since the mid of 1980s a remarkable novelty has been brought to build computer system to understand and use the natural form of human communication. There are lot of praiseworthy researches have been carried out in this Human-Computer Interaction (HCI) field during last decades. In recent years, the developments in HCI have abetted the user to interact with the computer in novel ways beyond the traditional boundaries of the keyboard and mouse. This emerging field includes areas, such as computer science, engineering, psychology and neuroscience. As facial expressions provide important clues about emotions, several approaches have been *envisaged* to classify human affective states. Facial expressions are visually observable, conversational, and interactive signals that regulate our interactions with the environment and other human beings in our vicinity [3]. Therefore, behavioural research, bimodal speech processing, videoconferencing, face/visual speech synthesis, affective computing and perceptual man-machine interfaces are those principle driving applications that have lent a special impetus to the research problem of automatic facial expression analysis and produced a great number of interests in this research topic. In 1872, Darwin [1] took attention by firstly demonstrating the universality of facial expressions and their continuity in human beings and animals. He claimed that there are specific inborn emotions, which originated in serviceable associated habits. He also explained that emotional expressions are closely related to survival. In 1971, Ekman and Friesen [4] classified human emotion into six archetypal emotions: surprise, fear, disgust, anger,

Happiness, and sadness. These prototypic emotional displays are also referred to as basic emotions. Ekman and Friesen [2,13] also developed the *Facial Action Coding System* (FACS) for describing facial expressions by *action unit*s (AUs). Their work helped to attract the attention of many researchers to analyze facial expressions by means of image and video processing. By tracking facial features and measuring the amount of facial movement, they tried to categorize different facial expressions. Recent works on facial expression analysis and recognition [9, 15, 17, 18, 19,20] used these universal "*basic expressions*" or a subset of them. Suwa et al. [5] presented a preliminary investigation on automatic facial expression analysis from an image sequence.Being motivated by some psychological studies,

De Silva et al.[6] carried out experiments on 18 people to recognize emotion using visual and acoustic information separately from an audio-visual database recorded from two subjects. They claimed that sadness and fear emotions are being better identified with audio, and some emotions, e.g., anger and happiness, are better identified with video. Furthermore, Chen et al.[7] cited the use of audio visual information in a multimodal HCI scenario for computers to recognize the user's emotional expressions. He concluded that the performance of the system increased when both modalities were considered together. The features are taken typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio, which use global statistics of the acoustic features. Pantic [8] explored the complete review of recent emotion recognition systems based on facial expressions. He strongly argued to include the essence of emotional intelligence into HCI design in the near future so that the system could be able to recognize a user's affective states —in order to become more human-like, more effective, and more efficient. Moreover, he provided new techniques for developing the initial phase of an intelligent multimodal HCI—an automatic personalized analyser of a user's nonverbal affective feedback.

Mase [9] developed an emotion recognition system based on the major directions of specific facial muscles . He was one of the first to introduce image processing techniques to recognize facial expressions. He used optical flow to estimate facial muscle actions which can then be recognized as facial expressions in both a top-down and bottom-up approach. In both approaches, the objective was on computing the motion of facial muscles rather than of facial features. Facial expressions are the result of facial muscle actions which are triggered by the nerve impulses generated by emotions. The muscle actions cause the movement and deformation of facial skin and facial features such as eyes, mouth and nose. As the texture of a fine-grained organ of facial skin has helped in extracting the optical flow, muscle actions has been extracted from external appearance. By the use of K-nearest neighbor for classification, four emotions, namely: happiness, anger, disgust and surprise have been recognized with an accuracy of 80% . Yacoob et al. [10] proposed a similar approach for analyzing and classifying facial expressions from optical flow based on qualitative tracking of principle regions of the face and flow computation at high intensity gradients points. Instead of using facial muscle actions, they constructed a dictionary to convert local directional motions associated with edge of the mouth, eyes and eyebrows into a linguistic, per-frame, mid-level representation. The main goal of this approach was to develop computational methods that relate such motions as cues for action recovery. Face region motion refers to the changes in images of facial features caused by facial actions corresponding to physical feature deformations on the 3-D surface of the face.

Rosenblum et al. [19] also computed optical flow of regions on the face, then applied a radial basis function network architecture that learned the correlation between facial feature motion patterns and human emotions. This architecture was specially invented to classify expressions. Besides, Lanitis et al. [20] invented a flexible shape and appearance model for image coding, person identification, pose recovery and facial expression recognition. Black et al. [15] developed a mid-level and high-level representation of facial actions using parametric models to extract the shape and movements of the mouse, eye and eyebrows . This approach is also recommended in [10] with 89% of accuracy. Tian et al. [16] obtained 96% accuracy using permanent and transient facial features such as lip, nasolabial furrow and wrinkles. They also used geometrical models to locate the shapes and appearances of those features . In this study the objective was to recognize the AU , developed by Ekman and Friesen [2]. Essa et al. [17] introduced a system that quantified facial movements based on parametric models of independent facial muscle groups and achieved 98% accuracy. They invented spatial-temporal templates that were used for emotion recognition. In this study, the face was being modelled by the use of an optical flow method coupled with geometric, physical and motion-based dynamic models. Matsuno et al. [22] described an approach for recognizing facial expressions from static images focusing on a pre-computed parameterization of facial expressions. Their approach plotted a grid over the face and warped it based on the gradient magnitude using a physical model. In that model, the amount of wrapping was represented by a multi-variate vector which was different than a learned vector of four facial expressions (i.e., happiness, sadness, anger and surprise)

Sebe et al. [11] recommended a method introducing the Cauchy Naive Bayes classifier which used the Cauchy distribution as the model distribution for recognizing emotion through facial expressions displayed in video sequence. Their person-dependent and person-independent experiments showed that the Cauchy distribution assumption typically produced better results than Gaussian distribution assumption. They used the simplified model proposed by Tao and Huang [12] which took an explicit 3D wireframe model of the face. The face model consisted of 16 surface patches embedded in Bezier volumes. The wireframe model and the 12 facial motion measurements were being measured for facial expression recognition. The 12 features have been used to measure the facial motion in the face model and using these features the 7 basic classes of facial expression have been defined for classification.

(a) Anger (b) Disgust (c) Fear (d) Happiness (e) Sadness (f) Surprise

Figure 1.1: Examples of images from the video sequences used in the experiment

Sebe et al. [11] also demonstrated that reduction of the emotion recognition problem to a mood recognition problem might increased the classification results significantly higher. After proposing Cauchy Naive Bayes classifier, Ira Cohen et al. [14] approached a Tree-Augmented-Naïve Bayes (TAN) classifier for recognizing emotions through facial expressions displayed in video sequences .The reason behind the introduction of TAN classifier was to learn the dependencies between the facial features It has been observed from the person-dependent and person-independent experiments that TAN structure provided significantly better outcome than that of simpler NB-classifier.

In the last few years, automatic emotion recognition through facial expression analysis has been used in developing various real life applications such as security systems, computer graphics, interactive computer simulations/designs, psychology and computer vision.Though many researchers employed machine learning classifiers (e.g., Cauchy-Gaussian assumption and support vector machine) independently to recognise emotion state on facial expression analysis, but the proposed method adopt classifier combination methodology. To the best of our knowledge, classifier combination methods were not used yet by any researcher in emotion classification task. So, we employed the classifier combination methodology to enhance the accuracy of emotion classification task.

III. FEATURES SET

Features are basically function or properties of some variable. Feature classification is to classify the features into classes for our purpose that may be our target class or noisy class. Efficient measurement of facial expression is necessary to understand the functionality of face-to-face communication. Most of the studies on automated expression analysis focus on classification of basic expressions (i.e., joy, fear, sadness, disgust, surprise, and anger) [23]. Suitable features selection is necessary to accomplish this classification task. We have used action unit, landmark , intensity and their combination as features in our experiment.

Action Unit

Action Units (AUs) are the fundamental actions of individual muscles or groups of muscles. AUs represent the muscular activity that produces facial appearance changes defined in Facial Coding System by Ekman and Friesen [2]. The reason behind in using the term AU is that more than one action have been separated from what most anatomists described as one muscle.

Landmark and Intensity : Landmark is one of the important features used in this experiment.It has been used in CK+ database invented by Cohn et al.[25]. The similarity normalized shape, denoted by s_n , refers to the 68 vertex points for both the x and y coordinates. The 68 vertices produces a raw 136 dimensional feature vector. These points are the vertex locations after removing all the rigid geometric variation (translation, rotation and scale), relative to the base shape.The similarity normalized shape s_n can be obtained by synthesizing a shape instance of s, that ignores the similarity parameters p. FACS provides a description of all possible and visually detectable facial variations in terms of 44 Action Units (AUs). Usually, a trained human FACS coder identifies the occurrence of an action unit and codes its intensity in a given facial image. Although the FACS coding is a precise tool for studying facial expressions, it is labor intensive. Therefore, automating the FACS coding and measuring the intensity of AUs would make it easier and widely accessible as a research tool in behavioral science. Genearally intensities of AUs are measured from absent to maximal appearance using a six-point intensity metric (i.e., 0 to 5).

IV. DATASET

The *Cohn-Kanade*[1]AU-Coded Facial Expression Database is publicly available from Carnegie Mellon University. CK database was invented for the purpose of promoting research into automatically detecting individual facial expressions in 2000. Since then, the CK database has become one of the most widely preferred test-beds for algorithm development and evaluation. During this period, three following limitations have been clearly seen:

- While AU codes are well validated, emotion labels are not, as they refer to what was requested rather than what was actually performed,
- The lack of a common performance metric against which to evaluate new algorithms, and
- Standard protocols for common databases have not emerged.

The cumulative effect of these factors has made benchmarking various systems very difficult or impossible. This is highlighted in the use of the CK database [24], which is among the most widely preferred corpus for developing and evaluating algorithms for facial expression analysis. It contains image sequences of facial expressions from men and women of varying ethnic backgrounds. Each subject (resolution 640 X 480) was instructed to perform a series of 23 facial displays that include single action units and combinations of action units. A total of 504 sequences are available for distribution in this database. After development of *Extended Cohn-Kanade database* (CK+) the number of sequences and the number of subjects are increased by 22% and 27% respectively. Emotion expressions included happy, surprise, anger, disgust, fear, and sadness. Examples of the expressions are shown in Fig. 4.3.

Fig.4.3. Cohn-Kanade AU-Coded Facial Expression database. Examples of emotion-specified expressions from image sequences.

In the present work, CK+ image database used as corpus. In CK+ database each set consists of sequences of image of a subject (i.e., man/woman). During person dependent experiment, each of the sets was used separately. However in person independent experiment, the image database was divided into two parts: the first part, which was used as training, contains the 70% of the entire dataset and the second part, which was used as testing, contains 30% of the entire dataset.

V. PREPARING TRAINING MODELS

Many researchers investigated the technique of combining the predictions of multiple classifiers to produce a single classifier (Breiman, 1996c; Clemen, 1989; Perrone, 1993; Wolpert, 1992). The resulting classifier is generally more accurate than any of the individual classifiers making up the ensemble. Both theoretical (Hansen and Salamon, 1990; Krogh and Vedelsby, 1995) and empirical (Hashem, 1997; Opitz and Shavlik, 1996a, 1996b) researches have been carried out successfully. Two sets of experimentation were carried out: person dependent and person independent facial emotion recognition. Thus two different sets of training models were prepared and tested. Altogether sixty three classifier combination models, namely ensemble,

[1] http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html

stacking and voting were prepared for the identification of person independent facial emotions. During experimentations, the features were added in incremental fashion. First AU feature was used to train model. Then intensity and landmark features were added gradually to train. The training models preparation for person dependent and person independent task is same except the size of the training data. So the following four sub-sections describe the training models preparation for both cases.

Models using Individual Classifiers : Initially the five classifiers- NB, k-NB, NN, auto MLP and DT were trained individually with the gradually incremented feature set, i.e., first they were trained with AU feature only; then they were trained with AU and intensity features; and finally they were trained with AU, intensity and landmark feature set.

Models using Classifiers Combination : This section describes the empirical study of the state-of-the-art classifier combination approaches, namely ensemble, stacking and voting. Each of these three approaches was tested with Naïve Bayes (NB), Kernel Naïve Bayes (k-NB), Neural Network (NN), auto Multi-Layer Perceptron (auto MLP) and Decision Tree (DT). The previous work [11][14][19] on facial emotion identification used these classifiers independently. In this work, the five classifiers were used to establish the effect of combining models. The following sections report the experimentations and obtained results.

Training Ensemble Models : Two popular methods for creating accurate ensembles are- bagging (Breiman, 1996c) and boosting (Freund and Schapire, 1996; Schapire, 1990). These methods rely on resampling techniques to obtain different training sets for each of the classifiers. Bagging approach was applied separately to five base learners, namely NB, k-NB, NN, auto MLP and DT. Initially the size (number of iteration) of the each base learner is set to 2. Then the experiments were performed with gradually increased size (size > 2).
It has been observed that initially the classification accuracy is increased with increasing in size parameter, but after a certain size value, the accuracy is almost stable. Cross validation technique was used to set the suitable size for base learners. Table-5.1 reported the suitable size of the base learners, i.e., NB, k-NB, NN, auto MLP and DT.It can be noticed from Table 5.1 that bagging using NN and auto MLP classifiers require less iterations than that of other classifiers, i.e., NB, kNB and DT.

Table –5.1: Size variation in Bagging

	Size		
Base Learners	**AU**	**AU+INT**	**AU+INT+LM**
NB	18	15	13
kNB	20	17	15
DT	15	12	10
Auto MLP	12	10	8
NN	12	9	7

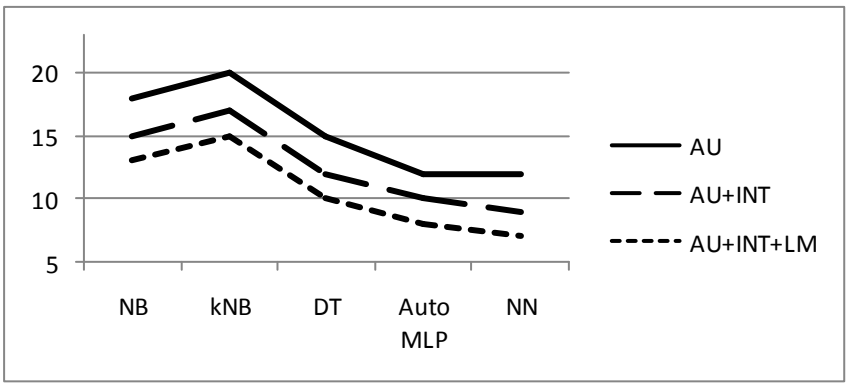

Fig. 5.1: Size variation in Bagging

Like Bagging, Boosting (AdaBoost.M1) approach was also applied separately to five base learners, namely NB, k-NB, NN, auto MLP and DT. The size (i.e., number of iterations) parameter of boosting was fixed empirically for each base learner because if the value of $1/\beta_t$ in boosting is less than 1 for some iteration value, then the weight of the classifier model may be less than zero for that iteration which is not valid. Table-5.2 reports the obtained size parameter of the base learners for boosting.

Table –5.2: Size variation in Boosting

	Size		
Base Learners	**AU**	**AU+INT**	**AU+INT+LM**
NB	21	18	14
kNB	24	20	17
DT	18	15	14
Auto MLP	16	14	11
NN	15	13	9

If we compare the size requirement of two state-of-the-art ensemble technique, it is clearly observed from Table-5.1 and Table-5.2 that boosting ensemble requires little extra iteration than that of bagging ensemble for each of the classifiers. Like bagging, NN classifiers needs less iterations to give maximum performance among the five classifiers.

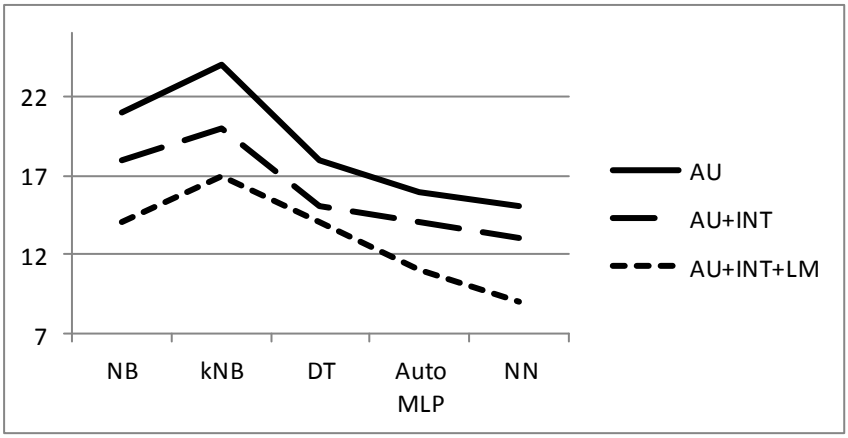

Fig. 5.2: Size variation in Boosting

Training Stacking Models : In stacking, one classifier is used as model learner whereas the others as base learners. In this work, out of five classifiers four classifiers were used as the base learners and the remaining classifier was used as model learner. So, five training models were prepared where each classifier got a chance to be the model learner.

Table –5.3: Model setup in Stacking

Base Learners	Model Learner
kNB, NN, Auto MLP, DT	NB
NB, NN, Auto MLP, DT	kNB
NB, kNB, Auto MLP, DT	NN
NB, kNB, NN, DT	Auto MLP
NB, kNB, NN, Auto MLP	DT

Training Voting Models : Though in stacking five different models were obtained using five classifiers (i.e., NB, k-NB, NN, auto MLP and DT), but in voting the scenario is different. In voting, only one model was obtained. In this case, four classifiers altogether were used as the base learners and majority vote was used as voting approach.

VI. MODEL TESTING AND RESULTS

Evaluation : The actual impacts of the proposals or the predicted results are interpreted by the evaluation. The proposed facial emotion model was evaluated using standard *precision*, *recall* and *F-measure* matrices.

In a classification task, the precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class).

Precision (P) = $\frac{TP}{TP+FP}$

where, TP = *true positive*, FP = *false positive*

Recall in the classification context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labeled as belonging to the positive class but should have been).

Recall (R) = $\frac{TP}{TP+FN}$

where, TP = *true positive*, FN = *false negative*

In a classification task, a precision score of 1.0 for a class C means that every item labeled as belonging to class C does indeed belong to class C but it states nothing about the number of items from class C that were not labeled correctly; whereas, a recall of 1.0 means that every item from class C was labeled as belonging to class C but it states nothing about how many other items were incorrectly also labeled as belonging to class C).

Often, there is an inverse relationship between precision and recall, where it is possible to increase one at the cost of reducing the other. In this context, F-measure is the harmonic mean of precision and recall that combines precision and recall.

Precision (P) = $2\ x\ \frac{Precision\ .\ Recall}{Precision\ +\ Recall}$ s

Testing Models : As discussed in the training section, the training of models is categorized into two sets of experiments, namely person dependent test and person independent test. The following subsections details the outcome obtained for those experiments.

Person dependent Testing : In Person dependent testing, we used the leave-one-out strategy, i.e., one sequence of each testemotion was left out, and the rest were used as the training sequences. Table-5.4 shows the recognition rate for each person for the five classifiers, and the total recognition rate averaged over the five people. Notice that the fifth person has the worst recognition rate. The fact that subject-5 was poorly classified can be attributed to the inaccurate tracking result and lack of sufficient variability in displaying the emotions. It can be noticed that the any of the five classifiers does not significantly decrease the recognition rate (and improves it in some cases), even though the input is unsegmented continuous sequence. Among five classifiers, auto MLP and NN performs almost the same with high recognition rate. The performance of DT is slightly better than that of kNB.

Table –5.4: Results applying individual classifiers

		Subject→					
		1	**2**	**3**	**4**	**5**	**Average**
NB	Precision	81.65%	87.46%	85.47%	82.46%	55.59%	78.53%
	Recall	78.94%	83.87%	82.21%	80.53%	53.32%	75.77%
	F-measure	80.27%	85.63%	83.81%	81.48%	54.43%	77.12%
kNB	Precision	82.63%	88.56%	86.31%	83.79%	57.32%	79.72%
	Recall	79.89%	84.76%	83.82%	81.59%	54.41%	76.89%
	F-measure	81.24%	86.62%	85.05%	82.68%	55.83%	78.28%
DT	Precision	83.25%	90.01%	87.43%	85.05%	58.64%	80.88%
	Recall	79.82%	85.11%	83.71%	81.82%	55.28%	77.15%
	F-measure	81.50%	87.49%	85.53%	83.40%	56.91%	78.97%
auto MLP	Precision	84.62%	91.32%	89.03%	85.98%	60.21%	82.23%
	Recall	80.74%	86.27%	84.64%	82.59%	56.89%	78.23%
	F-measure	82.63%	88.72%	86.78%	84.25%	58.50%	80.18%
NN	Precision	84.79%	91.36%	89.22%	86.03%	60.53%	82.39%
	Recall	80.81%	86.21%	84.52%	82.51%	56.97%	78.20%
	F-measure	82.75%	88.71%	86.81%	84.23%	58.70%	**80.24%**

The results of person dependent testing using classifier combination methods are given in the following tables. It is noticed from the Table-5.5 and Table-5.6 that bagging and boosting performs almost same in facial emotion classification. Again, ensemble with NN classifier outperforms the auto MLP with slight margin and others three (i.e., NB, kNB and DT) with satisfactory margin. It is also noted that ensemble technique improves the classification accuracy than that obtained using individual classifiers.

Table –5.5: Results applying ensemble-Bagging

		Subject→					
		1	**2**	**3**	**4**	**5**	**Average**
NB	Precision	82.61%	88.56%	86.42%	83.61%	56.63%	79.57%
	Recall	80.01%	84.79%	83.26%	81.63%	54.29%	76.80%
	F-measure	81.29%	86.63%	84.81%	82.61%	55.44%	78.16%
kNB	Precision	83.53%	89.61%	87.43%	84.72%	58.35%	80.73%
	Recall	80.84%	85.68%	84.78%	82.63%	55.34%	77.85%
	F-measure	82.16%	87.60%	86.08%	83.66%	56.81%	79.26%
DT	Precision	84.21%	90.98%	88.51%	85.99%	59.62%	81.86%
	Recall	80.79%	86.23%	84.79%	82.87%	56.26%	78.19%
	F-measure	82.46%	88.54%	86.61%	84.40%	57.89%	79.98%
auto MLP	Precision	85.57%	92.53%	90.08%	87.01%	61.31%	83.30%
	Recall	81.69%	87.21%	85.56%	83.61%	57.91%	79.20%
	F-measure	83.58%	89.79%	87.76%	85.28%	59.56%	81.20%
NN	Precision	85.82%	92.51%	90.15%	87.05%	61.43%	83.39%
	Recall	81.73%	87.33%	85.49%	83.69%	58.02%	79.25%
	F-measure	83.73%	89.85%	87.76%	85.34%	59.68%	**81.27%**

Table –5.6: Results applying ensemble-Boosting

		Subject→					
		1	**2**	**3**	**4**	**5**	**Average**
NB	Precision	82.73%	88.61%	86.54%	83.58%	56.71%	79.63%
	Recall	79.98%	84.72%	83.32%	81.68%	54.21%	76.78%
	F-measure	81.33%	86.62%	84.90%	82.62%	55.43%	78.18%
kNB	Precision	83.49%	89.68%	87.51%	84.68%	58.41%	80.75%
	Recall	80.97%	85.70%	84.72%	82.67%	55.31%	77.87%
	F-measure	82.21%	87.64%	86.09%	83.66%	56.82%	79.29%
DT	Precision	84.25%	90.91%	88.61%	86.00%	59.72%	81.90%
	Recall	80.83%	86.28%	84.71%	82.84%	56.27%	78.19%
	F-measure	82.50%	88.53%	86.62%	84.39%	57.94%	80.00%
auto MLP	Precision	85.61%	92.49%	90.05%	86.98%	61.38%	83.30%
	Recall	81.70%	87.23%	85.59%	83.67%	57.96%	79.23%
	F-measure	83.61%	89.78%	87.76%	85.29%	59.62%	81.21%
NN	Precision	85.86%	92.48%	90.13%	87.10%	61.48%	83.41%
	Recall	81.75%	87.32%	85.53%	83.65%	58.04%	79.26%
	F-measure	83.75%	89.83%	87.77%	85.34%	59.71%	**81.28%**

The results obtained using stacking and voting are depicted in Table-5.7 and Table-5.8. In stacking, maximum performance is achieved using NB, kNB, DT and auto MLP as base learner (BL) and NN as model learner (ML). However, from Table-5.7 it is clear that voting technique with majority voting achieved maximum classification accuracy.

Table –5.7: Results applying Stacking

			Subject→					
BL	ML		**1**	**2**	**3**	**4**	**5**	**Average**
kNB, DT, auto MLP, NN	**NB**	Precision	83.81%	89.69%	87.61%	84.52%	58.08%	80.74%
		Recall	81.02%	85.69%	84.47%	82.86%	55.29%	77.87%
		F-measure	82.39%	87.64%	86.01%	83.68%	56.65%	79.28%
NB, DT, auto MLP, NN	**kNB**	Precision	84.66%	90.73%	88.65%	85.81%	59.57%	81.88%
		Recall	81.94%	86.78%	85.84%	83.72%	56.28%	78.91%
		F-measure	83.28%	88.71%	87.22%	84.75%	57.88%	80.37%
NB, kNB, auto MLP, NN	**DT**	Precision	85.19%	92.08%	89.53%	87.13%	61.02%	82.99%
		Recall	81.89%	87.31%	85.76%	83.97%	57.32%	79.25%
		F-measure	83.51%	89.63%	87.60%	85.52%	59.11%	81.08%
NB, kNB, DT, NN	**auto MLP**	Precision	86.71%	93.51%	90.98%	88.07%	62.45%	84.34%
		Recall	82.70%	88.26%	86.63%	84.70%	59.05%	80.27%
		F-measure	84.66%	90.81%	88.75%	86.35%	60.70%	82.25%
NB, kNB, DT, auto MLP	**NN**	Precision	86.78%	93.53%	91.24%	88.05%	62.78%	84.48%
		Recall	82.74%	88.42%	86.49%	84.71%	59.14%	80.30%
		F-measure	84.71%	90.90%	88.80%	86.35%	60.91%	82.33%

Table –5.8: Results applying Voting

		Subject→					
		1	**2**	**3**	**4**	**5**	**Average**
NB, kNB, DT, auto MLP, NN	Precision	87.82%	95.08%	92.31%	89.13%	63.86%	85.64%
	Recall	83.81%	89.47%	87.53%	85.74%	61.23%	81.56%
	F-measure	85.77%	92.19%	89.86%	87.40%	62.52%	83.55%

Person independent Testing : In the previous section, it is seen that a good recognition has achieved when the training sequences are taken from the same subject as the test sequences. However, the main challenge is to see if this can be generalized to a person independent recognition. Like person dependent test, first five classifiers were tested independently. The results of the test are shown in Table-5.9.It is observed that if classifiers are applied individually then NN achieved best performance for all cases (80.88%, 83.89% 86.33% with AU, AU + Intensity and AU + Intensity + Landmark respectively).

Table –5.9: Results applying individual classifiers

Features→

Classifier→	**AU**			**AU + INT**			**AU+ INT+ Landmark**		
	P (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)
NB	75.35	73.53	74.43	78.35	76.33	77.33	80.13	78.47	79.29
kNB	77.36	75.77	76.56	80.36	78.87	79.61	82.45	81.32	81.88
DT	78.98	75.82	77.37	81.98	78.52	80.21	84.02	80.38	82.16
Auto MLP	81.03	78.93	79.97	84.03	81.73	82.86	86.23	84.76	85.49
NN	82.11	79.69	**80.88**	85.11	82.71	**83.89**	87.49	85.21	**86.33**

The results obtained using classifier ensemble methods are given in Table-5.10 and Table-5.11. It is noticed from the Table-5.10 and Table-5.11 that bagging and boosting performs almost same (the difference between classification accuracy is almost 0.02% for all cases) in facial emotion classification for person independent testing. Again, ensemble with NN classifier outperforms the auto MLP with slight margin (0.01% and 0.03% for bagging and boosting respectively) and others three (i.e., NB, kNB and DT) with satisfactory margin. It is also noted that ensemble technique improves the classification accuracy than that obtained using individual classifiers.

Table –5.10: Results applying Ensemble-Bagging

Features→

Classifier→	**AU**			**AU + INT**			**AU+ INT+ Landmark**		
	P (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)
NB	77.21	74.97	76.07	80.23	78.33	79.27	81.98	80.31	81.14
kNB	79.04	77.48	78.25	82.04	80.57	81.30	84.24	83.47	83.85
DT	80.87	77.79	79.30	83.59	80.65	82.09	85.89	82.29	84.05
Auto MLP	82.97	80.76	81.85	85.89	83.71	84.79	88.03	86.79	87.41
NN	83.95	81.82	**82.87**	87.01	84.59	**85.78**	89.51	87.36	**88.42**

Table –5.11: Results applying Ensemble-Boosting

Features→

Classifier→	**AU**			**AU + INT**			**AU+ INT+ Landmark**		
	P (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)
NB	77.19	75.02	76.09	80.29	78.27	79.27	82.01	80.32	81.16
kNB	78.98	77.58	78.27	81.97	80.65	81.30	84.17	83.61	83.89
DT	80.95	77.68	79.28	83.68	80.59	82.11	85.97	82.21	84.05
Auto MLP	83.02	80.84	81.92	85.93	83.74	84.82	88.06	86.77	87.41
NN	83.87	81.93	**82.89**	86.93	84.67	**85.79**	89.45	87.46	**88.44**

Table-5.12 and Table-5.13 show the results obtained using stacking and voting techniques. In person dependent testing, stacking technique achieved maximum performance when NB, kNB, DT and auto MLP were used as BL and NN was used as ML. However, in person independent testing the stacking technique performed best when NB, kNB, DT and NN were used as BL and auto MLP was used as ML. ML with auto MLP outperforms the ML with NB, kNB, DT and NN with 8.21%, 5,47%, 5.33% and 0.37% respectively. Table-X confirms that voting technique with majority voting achieved maximum classification accuracy (90.04%).

Table –5.12: Results applying Stacking

Features→

		AU			**AU + INT**			**AU+ INT+ Landmark**		
BL	**ML**	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)
kNB, DT, auto MLP, NN	NB	78.25	76.09	77.15	81.29	78.27	79.75	83.01	80.32	81.64
NB,DT, auto MLP, NN	kNB	80.01	78.41	79.20	82.89	81.58	82.23	85.17	83.61	84.38
NB, kNB, auto MLP, NN	DT	82.05	78.51	80.24	84.51	81.43	82.94	86.97	82.21	84.52
NB, kNB, DT, NN	Auto MLP	85.02	82.69	**83.84**	88.11	85.61	**86.84**	90.68	89.03	**89.85**
NB, kNB, DT, auto MLP	NN	84.59	82.84	83.71	88.03	85.59	86.79	90.61	88.37	89.48

Table –5.13: Results applying Voting

Features→

	AU			**AU + INT**			**AU+ INT+ Landmark**		
BL	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)	**P** (%)	**R** (%)	**F** (%)
NB, kNB, DT, auto MLP, NN	85.23	82.74	83.97	88.23	85.65	86.92	91.03	89.07	90.04

Conclusion and future work

In this work, the proposed methodology for recognizing emotions through facial expressions displayed in video sequences using the state-of-the-art classifier combination approaches, namely ensemble, stacking and voting.

The main contributions of this work are-

- introduction of classifier combination methodologies in the emotion recognition on facial expressions task.
- enhancement of classification accuracy of emotion recognition on facial expressions task.
- Classification accuracy is increased for both the person dependent and independent emotion identification.

Overall voting technique with majority voting achieved best classification accuracy.

The emotion recognition from just the facial expressions is probably not accurate enough. Therefore, other measurements probably have to be employed to interact the emotional state of a human with a computer properly. This work is just another step on the way toward achieving the goal of building more effective computers that can serve us better. For future research, we shall focus on facial expressions and body gestures in individual framework as well as multimodal framework because body movements and gestures have recently started attracting the attention of the HCI community. We will have an approach of skin color segmentation on HSV (Hue, saturation and value) space for facial and body feature extraction to recognise emotion.

One of the future directions of this work may be incorporated the color model (e.g., HSV) into emotion recognition on facial expressions task. The image may be represented to a color model. The color values may be used to compute AU, intensity etc. This may be applied to process facial expression of images in digital printing.

Moreover, the integration of multiple modalities such as voice analysis and context would be expected to improve the recognition rates and eventually improve the computer's understanding of human emotional states. This research will be continued to find better methods to fuse audio-visual information that model the dynamics of facial expressions and speech.

REFERENCES

[1] C. Darwin, The Expression of the Emotions in Man and Animals, J. Murray, London, 1872.

[2] P. Ekman and W. Friesen. The Facial Action Coding System: A Technique For The Measurement of Facial Movement. Consulting Psychologists Press, Inc., San Francisco, CA, 1978.

[3] J. Russell and J. Fernandez-Dols, The Psychology of Facial Expression. New York: Cambridge Univ. Press, 1997.

[4] P. Ekman, W.V. Friesen, Constants across cultures in the face and emotion, J. Personality Social Psychol. 17 (2) (1971) 124–129.

[5] M. Suwa, N. Sugie, K. Fujimora, A preliminary note on pattern recognition of human emotional expression, Proceedings ofthe Fourth International Joint Conference on Pattern Recognition, Kyoto, Japan, 1978, pp. 408– 410

[6] De Silva, L. C., Miyasato, T., and Nakatsu, R. Facial Emotion Recognition Using Multimodal Information. In Proc. IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS'97), Singapore, pp. 397-401, Sept. 1997.

[7] Chen, L.S., Huang, T. S., Miyasato T., and Nakatsu R. Multimodal human emotion / expression recognition, in Proc. of Int. Conf. on Automatic Face and Gesture Recognition, (Nara, Japan), IEEE Computer Soc., April 1998

[8] Pantic, M., Rothkrantz, L.J.M. Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE , Volume: 91 Issue: 9 Sept. 2003. Page(s): 1370 –1390.

[9] Mase K. Recognition of facial expression from optical flow. IEICE Transc., E. 74(10):3474–3483, October 1991.

[10] Yacoob, Y., Davis, L. Computing spatio-temporal representations of human faces. Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on , 21-23 June 1994 Page(s): 70–75.

[11] N. Sebe, I. Cohen, A. Garg, M. Lew, and T. Huang. Emotion recognition using a Cauchy Naive Bayes classifier.In ICPR, 2002, to appear.

[12] H. Tao and T. S. Huang. Connected vibrations: A modal analysis approach to non-rigid motion tracking.In CVPR.pages 735-750, 1998.

[13] Ekman, P., Friesen, W. V., & Tomkins, S. S. Facial affect scoring technique (FAST): A first validity study. Semiotica~ 1971,3 (1), 37-58.

[14] N. Sebe, I. Cohen, A. Garg, M. Lew, and T. Huang. Emotion recognition using a Cauchy Naive Bayes classifier.

[15] Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the International Conference on Computer Vision, pages 374–381. IEEE Computer Society, Cambridge, MA, 1995.

[16] Tian, Ying-li, Kanade, T. and Cohn, J. Recognizing Lower Face Action Units for Facial Expression Analysis.Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), March, 2000, pp. 484 – 490.

[17] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 757–763, July 1997.

[18] G. Faigin, The Artist's Complete Guide To Facial Expression. New York, NY: Watson-Guptill Publications, 1990.

[19] M. Rosenblum, Y. Yacoob, and L. Davis, "Human expression recognition from motion using a radial basis function network architecture," IEEE Transactions on Neural Network, vol. 7, pp. 1121–1138, September 1996.

[20] R. Aris. Discrete Dynamic Programming. Blaisdel, 1964. A. Lanitis, C. J. Taylor, and T. F. Cootes, "A unified approach to coding and interpreting face images," in Proc. 5th International Conference on Computer Vision (ICCV), Cambridge, USA, 1995, pp. 368–373.

[21] Hager, J., Ekman~ P., & Friesen, W. V. A comparison of facial measurement procedures.In P. Ekman & W. V. Friesen (Eds.), An atlas of Facial action. Book in preparation.

[22] K.Matsuno,C.Lee,andS.Tsuji,"Recognition of human facial expressions without feature extraction", Proceedings of the European Conference on Computer Vision, 513-520,1994.

[23] Y.-L. Tian, T. Kanade, and J. Cohn.Facial expression analysis. In S. L. . A. Jain, editor, Handbook offace recognition, pages 247–276. Springer, New York, New York, 2005.

[24] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In Proceedings of the International Conference on Automatic Face and Gesture Recognition, pages 46–53, 2000. 1,2[15] S.

[25] Lucey et al. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression