SenTion: A framework for Sensing Facial Expressions

Rahul Islam*, Karan Ahuja*, Sandip Karmakar*, Ferdous Barbhuiya* *IIIT Guwahati
{rahul.islam, karan.ahuja, sandip, ferdous}@iiitg.ac.in

Abstract—Facial expressions are an integral part of human cognition and communication, and can be applied in various real life applications. A vital precursor to accurate expression recognition is feature extraction. In this paper, we propose SenTion: A framework for sensing facial expressions. We propose a novel person independent and scale invariant method of extracting Inter Vector Angles (IVA) as geometric features, which proves to be robust and reliable across databases. SenTion employs a novel framework of combining geometric (IVA's) and appearance based features (Histogram of Gradients) to create a hybrid model, that achieves state of the art recognition accuracy. We evaluate the performance of SenTion on two famous face expression data set, namely: CK+ and JAFFE; and subsequently evaluate the viability of facial expression systems by a user study. Extensive experiments showed that SenTion framework yielded dramatic improvements in facial expression recognition and could be employed in real-world applications with low resolution imaging and minimal computational resources in real-time, achieving 15-18 fps on a 2.4 GHz CPU with no GPU.

Index Terms—Inter Vector Angles, HoG, Facial Expression, CK+, JAFFE.

✦

1 INTRODUCTION

Facial expressions play a very important role for sensing the perceived human emotion and intention in people. Driven by advancements in machine learning, image processing and human cognition they have found a wide variety of applications in interpreting user reactions, human computer interaction systems, emotion based interactive games, creation of virtual characters, surveillance, driver-less car, etc.

Facial expression analysis consist of three major stages: feature extraction, feature selection and classifier construction. In the first stage, features based on geometry or appearance textures of the face are extracted from a image or from a sequence of images. These facial features are extracted from face by using various feature extraction method like HOG, LBP, Gabor-Wavelet, AMM models etc. From these extracted set feature a subset feature is selected in second stage which best fitted to distinguish between different expression. Then in third stage classifier is constructed using these subset of features. Although various methods has been proposed for facial expression recognition, robust recognition that performs well across databases and subjects is still a challenging task, as most of the methods are person dependent, intolerant to scale changes and have low rate of accuracy across emotions and high computational overhead.

In this paper, we propose SenTion: a person independent and scale invariant framework for sensing emotions from facial expressions. Accuracy of the proposed system has been tested on the Cohn-Kanade Plus and JAFFE database, where it shows considerable improvement with low computational overhead making is suitable for realtime application. We validate our results with a user study, that compares the recognition accuracy of our system across different expressions and overlays it with the corresponding human perception associated with it. The specific contributions of our work is threefold. We first describe in detail our novel techniques of fusing geometrical and appearance based features. We then demonstrate the goodness of our system by evaluating it

on benchmark databases. Finally, we validate the practicality of facial expression recognition systems via a first of it's kind user study that measures the ease of expressing and understanding of various facial expressions by humans.

2 RELATED WORK

Among the various communications mechanism human being uses such as speech, hand gestures and emotions etc, emotions and the encoded feelings prove to be an important factor for an effective communication. One easy of way of understanding emotion is to analyze facial expression which have significance in human communication, interaction and can be also applied in facial biometric-based intelligent systems.

Among the notable contribution for determining facial expression Yacoob et al. [1] fitted local parametric motion models on face and then classified the parameters by a nearest neighbor classier to recognize expression. Hoey et al. [2] used a multilevel Bayesian network to learn in a weakly supervised manner the dynamics of facial expression. De la Torre et al. [3] proposed a geometric-invariant clustering algorithm to decompose a stream of one person's facial behavior into facial gestures. Shan et al. [4] explored the use of local binary patterns (LBP) for the given task. Bartlett et al. [5] proposed a method based on gabor filters and AdaSVMs for the same purpose.

For facial expression detection mainly two types of feature are used: appearance based and geometry based. Appearance features describe the texture of the face caused by expression, such as wrinkles and furrows. Geometric features describe the shape of the face and its components such as the mouth or the eyebrows [6]. Among the appearance based feature, Gabor wavelet representation is widely adopted in facial expression analysis [5] [7]. These gabor features are largely affected by scaling and computation of gabor features is time intensive. Local Binary Patterns(LBP) is also an effective appearance features for facial expression analysis [8] [9] [4] . Geometric feature based approaches use active appearance models (AAMs) or similar kind technique to track a dense set

of facial points on face. The locations of these points are then used to infer the shape of facial features such as the mouth or the eyebrows and thus to classify the facial expression [6]. Sebe et al. [10] proposed a geometric feature based system where he used Piecewise Bzier volume deformation tracking after manually locating a number of facial points.

Recently researcher have employed deep learning for the problem of facial expression analysis. Liu et al. [11] used hybrid deep learning model i.e. CNN + RBM for automatic detection basic facial expression. Kim et al. [12] compared different models of Deep Belief Network(DBN) for unsupervised feature learning in audio-visual emotion recognition. Liu et al. [13] proposed a Boosted Deep Belief Network(BDBN) to perform three task namely feature learning, feature selection and classifier construction on CK+ and JAFFE Dataset. In contrast to these computationally intensive models, our work aims to integrate the simplicity of geometric models and the thoroughness for salient feature extraction of appearance based models, by combining the two to form a hybrid model which achieves state of the art accuracy while being robust and reliable across databases.

3 MODEL CONSIDERATIONS

In this section, we present the primary modules of our proposed facial expression recognition algorithm. Figure 1 illustrates our proposed framework of feature extraction, training and subsequent testing.
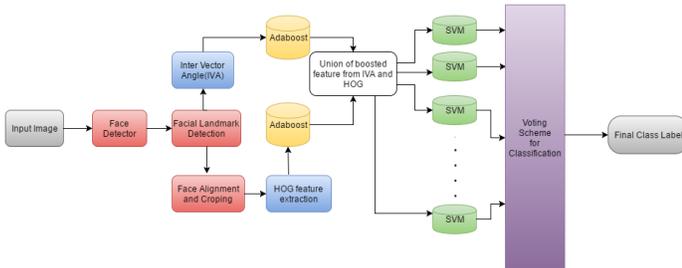

Fig. 1. Schematic representation of proposed model.

It has been observed that facial expressions are usually manifested by contraction and expansion of facial muscles which alters the position of facial landmarks. Along with facial muscles the texture of the area also changes. [14] Our proposed work attempts to exploit the changes by making use of geometric features to capture changes in facial landmarks and appearance based features for capturing more salient features such as texture and contour.

For each image, the area of the face is localized first, followed by the detection of various facial landmarks. We adapt the concept of scale invariant Inter Vector Angles (IVA) to act as geometric features, which relies upon prior facial landmarks for feature computation. We use this approach in junction with appearance based features, which are represented by Histogram of Gradients proposed by [15]. Prior to HOG feature extraction, we crop the face from the image using face detection and then align and normalize the Region of Interest for uniform extraction and training of features across different datasets. Once the feature vectors are obtained, they are passed through AdaBoost for feature selection and then Support Vector Machines for learning and classification.

3.1 Facial Landmark Detection

Happy et al. [14] states that the importance of accurate facial land marking is undeniable to achieve better performance for facial expression recognition. A pre-requisite to facial landmark extraction is face detection, for which we make use of the dlib library [16], which provides a powerful frontal face detector that is robust to pose changes. We then detect 68 facial landmarks from the above detected image using Vahid et al [17] state of the art real time face pose estimation algorithm. Their algorithm uses a general framework based on gradient boosting for learning an ensemble of regression trees for estimating the face's landmark positions directly from a sparse subset of pixel intensities.

3.2 Geometry Based Feature

Many existing literature's such as [18] use the 2D locations of facial landmark in the image space as geometric features. Similar angle based representation have also been explored by [19], [20] and [21]. However, their approaches are only defined in the discrete image space and lack a mathematical formulation. Hence, while such models can adapt to different scales, they fail to capture the individual differences of faces. To overcome these shortcomings, we propose the concept of Inter Vector Angle(IVA). In this work, we not only study the holistic spatial analysis of geometric features, but derive a novel mathematical formulation of angle based representations. The generality of our foundation helps it to span over various feature landmarks, without being custom designed for a specific feature region such as lips, eyes, brow, etc. as is the case in [19]. IVA's are scale invariant geometric features computed from the previously extracted facial landmark as follows:

Let P be the set of the location of all 2D landmarks such that $(P_1, P_2, P_3, ...P_{68}) \in P$ where P_i denoted the (x,y) coordinate of the facial landmark in image space where i=1:68. Let V_{ij} denote the vector between any two facial landmark P_i and P_j such that $i \neq j$. Therefore $\overrightarrow{V_{ij}} = P_i - P_j$. The set of all such vector $\overrightarrow{V}$ can be denoted by $\overrightarrow{V} \Leftarrow \forall \overrightarrow{V_{ij}}$, where $i \neq j$. To compute the we select any two vectors $\overrightarrow{V_{ij}}, \overrightarrow{V_{mn}} \in \overrightarrow{V}$ such that $i = j$ and $m \neq n$. Now the angle between any two such vector is given by:

$$\theta_{ijn} = \cos^{-1}(\frac{\overrightarrow{V_{ij}}.\overrightarrow{V_{mn}}}{|\overrightarrow{V_{ij}}||\overrightarrow{V_{mn}}|}) \qquad (1)$$

where . denotes the dot product and $|\overrightarrow{x}|$ signifies the magnitude of the vector $\overrightarrow{x}$. We define $\theta \Leftarrow \forall \theta_{ijn}$ such that $j \neq n$. Here θ denotes the set of all angles between any two vectors joining a set of facial landmarks, and having a common point.

We observe that the cardinality of θ to be 150348. This is because we have $|P| = 68$. The total number of triangle that can be formed, has the maximum possibility of $^{68}C_3$ considering no three points are co-linear. Each triangle has three angles therefore leading to $|\theta| = 3 \times {}^{68}C_3 = 150348$. Out of these, the points which are co-linear do not contribute to feature weightage, and their angles are assigned to zero.

The use of θ as IVA geometric feature imbibes the scale invariant property of angles in our system. Hence, the face Region of Interest no longer needs to be taken into normalized space leading to lossy down-sampling as is the case with most appearance and geometric features which depend upon the size of the image such as those implemented in state of the art algorithms such as in [14] and [22]. We also observe in section 4.1.1 that IVA out performs traditional geometric features such as those which consider only the vector length or landmark locations.

3.3 Face Alignment

In order to make our feature model scale space and rotation invariant, we need to map it to a common reference frame. This also helps us to better capture and analyze the texture of the face. We use the eye centers (averaged from the detected eye corner) to act as the reference points when computing the similarity transform to obtain a cropped, scaled and rotated face. The resultant is a 192X192 pixel image of the face.

3.4 Appearance Based Features

Appearance based feature are extracted on aligned facial images. We extract Histogram of Gradient(HOGs) of aligned face based on the configuration proposed by Felzenswalb et al. [15], prior to which we use Gaussian Pyramids to upscale the image [23]. HOG features are defined using 9 contrast insensitive gradient orientations, a cell size of 8 and truncation 0:2. This leads to a 36-dimensional feature vector. We have used Scikit-learn [24] implementation of HOG to extract HOG features.

3.5 Feature Boosting

Literature suggest that Adaboost and SVM works in harmony [25] [26] [27] [7]. Using SVM as a weak learner for Adaboost, it performs not only as a fast classifier, but also as a feature selection technique. An advantage of feature selection by Adaboost is that features are selected contingent on the features that have already been selected [28]. In our work, we use the Decision Tree Classifier as the base estimator and use SAMME.R as the boosting algorithm [29]. We select the best feature obtained by applying Adaboost for every class using One-Against-All strategy. We do this by binning the best weak-learners obtained in all the iterations for every one-against-all sub-problem and picking the top 'k' features. The final feature set is taken to be the union of best features obtained for every sub-problem.

3.6 Classification and Validation

For face expression detection, we use Support Vector Machine(SVM) as our learning model. SVM [30] is a popular machine learning algorithm which maps the feature vector to a different plane, usually to a higher dimensional plane, by a non-linear mapping, and finds a linear decision hyper plane for classification of two classes. Since SVM is binary classifier multiclass problem can be extended using two methods: One-Against-One (OAO) or One-Against-All (OAA).

In our work, we have used OAO method to detect multiple expression using SVM. SVMs are trained for each combination of classes and the final class label is decided using majority voting strategy. Hence, $^{K}C_2$ number of classifiers need to be constructed where K is the number of classes (facial expressions to be classified).

4 EVALUATION

In this section, we discuss the performance of our methodology. For experiments, we use a hardware configuration of Intel Pentium CPU 2020M @ 2.40 GHz and 4 GB RAM. We run our experiments on the two widely used facial expression datasets, that is, Japanese Female Facial Expression (JAFFE) [31] and Cohn-Kanade (CK+) [32]. We use 10 fold cross validation to evaluate the performance of our proposed methodology. We train our method using both geometric (IVA) and appearance based (HOG) features, and construct $^{6}C_2$ number of SVM classifiers for evaluating the performance on the test set.

4.1 Experiments on the Cohn-Kanade Database

The Cohn-Kanade+ database consists of 593 posed sequences from 123 subjects, with all images having dimensions 640x490 pixels. Each sequence in the database goes from neutral to target display, with the last frame being AU encoded. The images contain six different facial expressions: anger, disgust, fear, happiness, sadness and surprise. [33] For our experiments, we use the images whose sequences could be labeled as one of these six basic facial expressions. We showcase the performance of different feature selection techniques, namely: Inter Vector Angles (IVA), Histogram of Gradients (HOG) and a hybrid model using HOG + IVA. We use AdaBoost for feature selection as the ensemble method, employing a Decision Tree Classifier as the base estimator and maximum number of estimators as 100. This generates a feature vector of size 531 and 508 for IVA and HOG respectively, giving us a total value of k (as described in section 3.5) as 1039.

4.1.1 Using Geometric Features:

We test the accuracy of our system using Geometric Features, namely Inter Vector Angles described in section 3.2 with AdaBoost described in section 3.5. As this method is scale invariant we do not need to re-size the image, hence making it applicable on the fly. A critique of most geometric feature and shape based models is that the distance between facial landmarks vary from person to person, thereby making the system person dependent and less reliable [14]. However, the use of IVA's make the features independent of people as they are found to be reliable across subjects. If we choose only vector lengths (i.e. distance between landmarks) as features, we achieve a limited accuracy of 87.9%, as compared to the current system of IVA features which achieves an accuracy of 97.42%. This establishes the robustness of our system.

As compared to traditional methods that make use of Gabor Wavelets, such as those proposed by [5] and [7] are affected by the resolution of the image, our method is more computationally sensitive and is scale invariant.The method also beats the state of the art accuracies reported by appearance based models such as the 94.09% accuracy of [14], which employ appearance based salient feature patches and 93.14% of [22] which employ both appearance based features (in the form of salient patches) and geometric features (in the form of salient distances) in their method using Facial Movement Features. The proposed system achieves an average F1 Score of 96.8%, precision of 97.29% and recall of 97.12%. Figure 2 depicts the confusion matrix of this method.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	93.5	4	0	0	2.5	0
Disgust	0	100	0	0	0	0
Fear	0	0	100	0	0	0
Happy	2.9	0	0	97.1	0	0
Sad	3.4	0	3.4	0	93.2	0
Surprise	0	0	0	1.3	0	98.7

Fig. 2. Confusion Matrix of IVA on CK+ database

4.1.2 Using Appearance Features:

Histogram of Gradients (HOGs) are widely used for automatic facial action unit detection and have showcased state of the art performance in the same. [18]. We make use of HOGs with AdaBoost as proposed in section 3.4 and 3.5 respectively to test it's individual performance and contribution to the system.

Using the HOG features extracted from the aligned face images, we achieve an accuracy of 96.73%, precision of 96.87%, recall of 95% and F1 Score of 95.42%. This depicts that HOG are very robust at capturing the texture and salient patches of the face, without prior knowledge of discriminative patches. Figure 3 depicts the confusion matrix of this method.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	95	2.5	0	0	2.5	0
Disgust	3.3	96.7	0	0	0	0
Fear	3.3	0	88.4	0	0	8.3
Happy	0	0	0	100	0	0
Sad	9.9	0	0	0	90.1	0
Surprise	0	0	0	0	0	100

Fig. 3. Confusion Matrix of HOG on CK+ database

4.1.3 Hybrid Model: IVA + HOG:

We combine the scale invariant property of Inter Vector Angles with the texture and salient patches capturing property of Histogram of Gradients, to create a hybrid model which is an amalgam of both geometric and appearance based features. This model, achieved an accuracy of 97.38%, with a precision of 97.24%, recall of 96.94% and F1 Score of 96.71%. The confusion matrix for the same can be found in Figure 4. Table 1 compares the performance of this method from the results obtained by other approaches reported in literature. Subsequently Table 2 showcases the ranks of these methods. In both the tables we compare the accuracy across 6 facial expressions: Angry, Disgust, Fear, Happy, Sad and Surprise. We attain the best rank out of all the methods, with an average rank of 1.66. Our approach is the best for classifying images belonging to Anger, Fear and Disgust. Due to the simplicity of computation, scale in-variance and robustness of our model, it can be used in real life applications and on devices with constrained resources such as tablets and mobiles. Our model is stable, efficient and does not require any GPU's for training and even outperforms current Deep Learning techniques in terms of accuracy on the CK+ database. Our method beats the average accuracy of 92.22% achieved by Liu et al. [11] using AU-Aware Deep Networks(AUDN's) and 99.6% achieved by [34] which employs deep convolutional neural networks. It also exceeds the 8 fold cross validation accuracy of 96.7% achieved by [13] using Boosted Deep Belief Networks (BDBN). While other deep neural network architectures such as [35] and [36] exist, they predict Facial Action Units rather than categorizing facial expressions and hence our approach cannot be compared to their's.

Therefore, being able to perform significantly better on the standard CK+ database with low computational requirement, high speed of 15 frames per second and stability, makes our work the most applicable among the known state of the art.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	91	6.5	0	0	2.5	0
Disgust	0	100	0	0	0	0
Fear	0	0	100	0	0	0
Happy	1.4	0	0	98.6	0	0
Sad	3.3	0	3.3	0	93.3	0
Surprise	0	0	0	1.2	0	98.8

Fig. 4. Confusion Matrix of Hybrid Model on CK+ database

Method	Anger	Disgust	Fear	Happy	Sad	Surprise
Uddin [37]	82.5	97.5	95	100	92.5	92.5
Poursaberi [38]	87.03	91.58	90.98	96.92	84.58	91.23
Zhong [33]	71.39	95.33	81.11	95.42	88.01	98.27
Song [39]	90.56	86.04	84.61	93.61	90.24	92.3
Zhang [22]	87.1	90.2	92	98.07	91.47	100
Liu [11]	81.48	95.48	82.67	99.52	71.43	97.59
Happy [14]	87.8	93.33	94.33	94.2	96.42	98.46
Our Method	91	100	100	98.5	93.2	98.7

TABLE 1
Performance comparison (in percentage) of different Facial Expression Recognition methods on the CK+ Database

Method	An	Di	Fe	Ha	Sa	Su	Avg. Rank
[37]	6	2	2	1	3	6	3.33
[38]	5	6	5	5	7	8	6
[33]	8	4	8	6	6	4	6
[39]	2	8	6	8	5	7	6
[22]	4	7	4	4	4	1	4
[11]	7	3	7	2	8	5	5.33
[14]	3	5	3	7	1	3	3.66
Our's	1	1	1	3	2	2	1.66

TABLE 2
Performance-based ranks and average ranks of State-of-the art methods on the CK+ Database. A lower value indicates a better rank.

4.2 Experiments on JAFFE Database

The JAFFE database contains 213 gray images of seven facial expressions (six basic + neutral) taken from 10 Japanese Female models, with each image having a resolution of 256X256 pixels. In this paper, all the images of six basic expressions from the JAFFE database are used, namely: anger, disgust, fear, sad, happy and surprise.

To evaluate the robustness and accuracy of our system on different databases, we test the accuracy of our system on the JAFFE database. While testing on the JAFFE database, we use the same parameters as those used in the CK+ database; so as to evaluate the reliability of our system with images of different face sizes and resolutions without additional adaptation. We found the JAFFE database to be more challenging than the CK+ database and hence set the maximum number of estimators in AdaBoost to 400, to avoid loss of important features. Figure 5 denotes the performance (F1 score, precision and recall) of our method on the various facial expressions. We test our model using both geometric (IVA), appearance based (HOG) and a hybrid model of features as proposed in section 3. The confusion matrix for the same can be found in Figure 6. It is interesting to note that using IVA, we obtain an accuracy of 78.89% and using HOG we obtain an accuracy of 88.83%. Using our hybrid model (IVA+HOG) we get an accuracy of 88.2%. As opposed to the previous case (in CK+ dataset), appearance based features have a greater impact than geometric features on the JAFFE dataset. Our proposed hybrid model correctly captures this, and remains stable and adapts well to the changes across databases.

4.3 User Study

In order to find test the feasibility of facial expression systems for various applications such as those described in section 1, we perform a user study. Here we ask 5 people to showcase various near-frontal facial expressions as depicted in Table 3. It has to be kept in mind that these people were not trained actors/actresses and were not provided any guidance on how to showcase their expressions. They were then asked to give their rating on the ease of expressing the respective facial expressions on the well
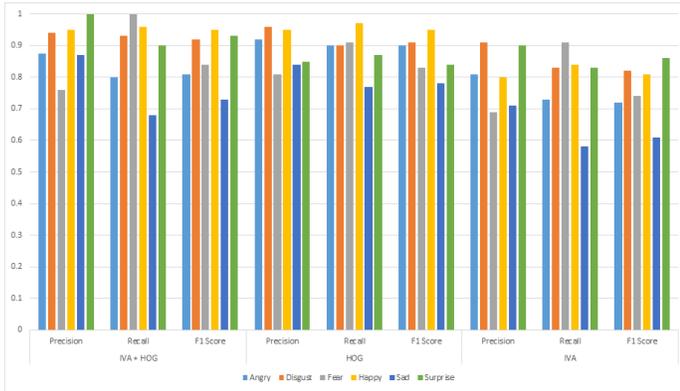
Fig. 5. Performance of our method on the JAFFE database for different facial expressions.

(a) Inter Vector Angles

(b) Histogram of Gradients

(c) Hybrid Model: IVA + HOG

Fig. 6. Confusion matrix of proposed methods on the JAFFE database. Here: AN → Anger, DI → Disgust, FE → Fear, HA → Happy, SA → Sad, SU → Surprise

Emotion	Ease of Expression	Ease of Understanding
Happy	4.41	4.20
Surprise	3.78	3.08
Sad	3.1	2.02
Angry	3.3	2.76
Disgust	2.6	2.42
Fear	2.1	0.665

TABLE 3
Overall ratings of the user Study

5 DISCUSSION

We have trained and cross validated the performance of our system on both CK+ and JAFFE dataset. The accuracy of 97.38% (hybrid model) on the CK+ dataset, shows the proposed approach is robust and out performs other state of the art methods.

Analyzing the confusion matrix of our hybrid model on the CK+ dataset in Figure 4 and JAFFE dataset in Figure 6(c), we find that the proposed method sometimes confuses anger with sadness and disgust. This is due to the similar geometric and appearance features between the facial expressions, especially in the JAFFE database, where the difference is less stark. Other than this, no major problem could be detected.

From the experiment on CK+ dataset we have seen that IVA features gives more accuracy than combined HOG and IVA features (by 0.04%). In JAFFE dataset we have seen that HOG features gives more accuracy than combined HOG and IVA (by 0.63%). From this we can can conclude that IVA geometric features more dominant in CK+ and HOG appearance features are more dominant in JAFFE database. Therefore a system comprising both feature sets (IVA + HOG) proves to be more stable, reliable and accurate to perform facial expression recognition across databases without additional adaptation.

The user study and it's subjective evaluation also showcases the viability of facial expressions acting as a precursor for emotions in an induced environment. While a claim can be made that person-dependent and naturally occurring facial expressions can provide a better insight, the study presents vital information such as which facial expression to use for practical applications in an induced setting where the system cues the user for depicting an emotion such as that for human-computer interaction or creation of virtual characters. Therefore, while the expressions depicted in the user study lacked peak intensity it provided a good sample of facial expressions in everyday life.

6 CONCLUSION AND FUTURE WORK

This paper has presented a computationally efficient facial expression recognition system (SenTion) for accurate and robust classification of the six universal expressions. We proposed a novel approach of taking Inter Vector Angles (IVA) as geometric features, which proved to be scale invariant and person independent. It's combination with appearance based Histogram of Gradients, provided a hybrid model which was reliable and stable across databases without additional adaptation. The effectiveness of the proposed approach of IVA's and appearance HOG based features is validated by it's performance on the CK+ and JAFFE database and comparison with other state-of-the art methods. We outperformed the state of the art algorithm on the CK+ database, achieving an accuracy of 97.38%.

The proposed approach being simplistic, consistent, computationally sensitive and real time can be potentially applied into many applications, such as human emotional state detection, patient monitoring, fatigue detection and tutoring systems. In our

established Likert scale [40], on a scale of 1 to 5. Subsequently, the ease of understanding their facial expressions was measured by a set of 3 human evaluators by means of a blind testing, wherein each evaluator was showed a 5 second clip of the respective expression and was asked to identify the facial expression and then rate it on a Likert Scale form 1 to 5. A miss-classification of expression was rated 0. Every evaluator measured the ease of interpreting each facial expression of each person. The average ratings of the user study are depicted in Table 3. According to the table; happy, surprise and angry are the top 3 facial expressions that can be expressed and understood easily by humans. Also, there is a direct correlation between ease of expression and ease of understanding a facial expression, both being directly proportional to one another.

future work, we will extend our approach to in the wild settings and also test it's performance on video based facial expression recognition systems.

REFERENCES

[1] F. De la Torre, Y. Yacoob, and L. Davis, "A probabilistic framework for rigid and non-rigid appearance based tracking and recognition," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 491–498.

[2] J. Hoey, "Hierarchical unsupervised learning of facial expression categories," in *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on*. IEEE, 2001, pp. 99–106.

[3] F. De la Torre, J. Campoy, Z. Ambadar, and J. F. Cohn, "Temporal segmentation of facial behavior," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[4] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[5] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 568–573.

[6] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer, "Meta-analysis of the first facial expression recognition challenge," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 4, pp. 966–979, 2012.

[7] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, April 2006, pp. 223–230.

[8] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[9] S. Happy, A. George, and A. Routray, "A real time facial expression classification system using local binary patterns," in *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*. IEEE, 2012, pp. 1–5.

[10] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang, "Authentic facial expression analysis," *Image and Vision Computing*, vol. 25, no. 12, pp. 1856–1863, 2007.

[11] M. Liu, S. Li, S. Shan, and X. Chen, "Au-aware deep networks for facial expression recognition," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.

[12] Y. Kim, H. Lee, and E. M. Provost, "Deep learning for robust feature generation in audiovisual emotion recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3687–3691.

[13] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1805–1812.

[14] S. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *Affective Computing, IEEE Transactions on*, vol. 6, no. 1, pp. 1–12, 2015.

[15] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, 2010.

[16] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Dec. 2009. [Online]. Available: http://dl.acm.org/citation.cfm?id=1577069.1755843

[17] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[18] T. Baltrusaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, vol. 6. IEEE, 2015, pp. 1–6.

[19] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. 97–115, 2001.

[20] M. F. Valstar, H. Gunes, and M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features," in *Proceedings of the 9th international conference on Multimodal interfaces*. ACM, 2007, pp. 38–45.

[21] J. F. Cohn and F. De la Torre, "Automated face analysis for affective," *The Oxford handbook of affective computing*, p. 131, 2014.

[22] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," *Affective Computing, IEEE Transactions on*, vol. 2, no. 4, pp. 219–229, 2011.

[23] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=1953048.2078195

[25] E. García and F. Lozano, "Boosting support vector machines." in *MLDM Posters*, 2007, pp. 153–167.

[26] X. Li, L. Wang, and E. Sung, "Adaboost with svm-based component classifiers," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 5, pp. 785–795, 2008.

[27] ——, "A study of adaboost with svm based weak learners," in *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, vol. 1. IEEE, 2005, pp. 196–201.

[28] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, 2013.

[29] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[30] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[31] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 200–205.

[32] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[33] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2562–2569.

[34] P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, "Dexpression: Deep convolutional neural network for expression recognition," *arXiv preprint arXiv:1509.05371*, 2015.

[35] P. Khorrami, T. Paine, and T. Huang, "Do deep neural networks learn facial action units when doing expression recognition?" in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 19–27.

[36] L. A. Jeni, A. Lőrincz, Z. Szabó, J. F. Cohn, and T. Kanade, "Spatio-temporal event classification using time-series kernel based structured sparsity," in *European Conference on Computer Vision*. Springer, 2014, pp. 135–150.

[37] M. Z. Uddin, J. Lee, and T.-S. Kim, "An enhanced independent component-based human facial expression recognition from video," *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 4, pp. 2216–2224, 2009.

[38] A. Poursaberi, H. A. Noubari, M. Gavrilova, and S. N. Yanushkevich, "Gauss–laguerre wavelet textural feature fusion with geometrical information for facial expression identification," *EURASIP Journal on Image and Video Processing*, vol. 2012, no. 1, pp. 1–13, 2012.

[39] M. Song, D. Tao, Z. Liu, X. Li, and M. Zhou, "Image ratio features for facial expression recognition application," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, no. 3, pp. 779–788, 2010.

[40] R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, 1932.