

Analysis and Recognition of Emotions Using Zernike Moment Based Facial Expression System

Deepak Saini¹, Dipti Bansal²

M.Tech Student, Department of Electronics and Communication Engineering, Punjabi University, Patiala, India¹

Assistant Professor, Department of Electronics and Communication Engineering, Punjabi University, Patiala, India²

ABSTRACT: Automatic human emotion recognition is an important and emerging aspect of human facial expression recognition for human-machine interface. Facial expressions are mostly the reflections of one's emotional state. So to know the true emotional state of a person, accurate evaluation of his/her facial expressions is must. In recent times, major advances were made in understanding and analyzing facial expressions by applications of image processing and machine learning methods. In this paper we have proposed a novel approach which uses neural network to classify and recognize facial expressions. The system is able to recognize six basic emotions. The proposed system works in three steps. First, the face is detected from image, and then features are extracted from it using Zernike moments. In the last step the neural network is trained to classify the emotions. The system so designed was implemented on Cohn-Kanade database and accuracy of around 74 % was achieved.

KEYWORDS: Face detection, feature extraction, Zernike moments, and neural network.

I. INTRODUCTION

An emotion is the physiological and mental state of feelings of individual involving feelings, physiological changes, thinking, triggering of nervous system, and change in facial expressions due to behavioural changes. Emotional state refers to the state of human's emotions for e.g. he was in good mood or the boy is gripped in fear etc. one can quote emotion as mental state which arises spontaneously and is accompanied by physiological changes; a feeling: feeling of love, hate, sorrow, or joy. Human face is the best way to express emotions to the outer world. Though human face can express 21 distinct emotions but in 1969, Friesen and Ekman classified six basic emotional expressions to be universal: anger, sad, surprise, happiness, disgust, and fear [1]. Human beings can very easily recognize these expressions just by seeing a human face but it is very exigent and tedious for a computer system to recognize human facial expressions with better accuracy. But now days there have been significant advances in such fields viz. pattern recognition, image processing which have facilitated researchers to take the studies of this phenomenon from the field of human psychology to automatic analysis, classification, synthesis and even expressive animation.

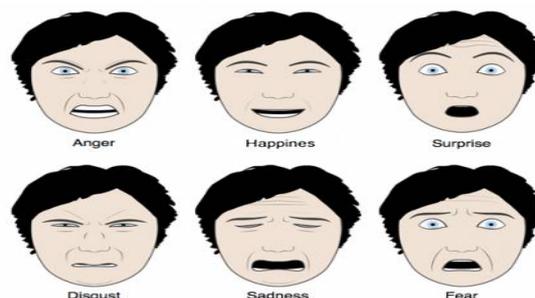


Fig.1 Basic Six Emotions

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

The basic idea to implement automatic emotion recognition system is as follows:-

1. Face Detection.
2. Feature Extraction.
3. Neural Network training and testing.
4. Facial Expression Classification.

In first part of the system face part is detected from a still image or from a sequence of images (video). Face detection is the most fundamental step in automatic facial expression system. The objective of face detection is to find out whether there are any faces present in the image or not and if face/s are present then to return or to localize the face in it. While face detection may be a trivial task for human vision but for computer vision it is a challenging task. Here in this particular work we have implemented face detection using Viola-Jones algorithm.

In the second stage, the system extracts features of the detected face and makes a feature vector. The feature vector thus formed is later used to train the neural network so that the neural network can classify various emotions based on this feature vector. Among different feature extraction algorithms available, here Zernike moments are used due to their inherit property to discard redundant information.

In third step of the system the feature vector obtained in second step is used to train the neural network. To furnish the process of classification using neural network a new top of the line software "STATISTICA" is used. *STATISTICA Automated Neural Networks (SANN)* is one of the most advanced and best performing neural network applications available in the market. It offers numerous unique advantages and is quite appealing to neural network experts and new users. The total feature vector is supplied to the SANN, and then SANN itself decides about the percentage of samples to be used for training and testing. A general rule is to use 70% of the samples for training and 30% of them for testing. Once the neural network is trained, it can be used to classify and recognize facial expressions of random test images.

The ability to recognize human emotions can be used in various fields viz. effective marketing of consumer products, developing video games that are able to react dynamically with current player by recognizing his/her emotions. Facial expressions due to emotional change in a human being can help individuals in law enforcement agencies to detect a person who is speaking lie or whose expressions don't comply with his words as it is seen that people often get emotional when they lie. There may be many factors which can provoke emotions in a human being like the negative emotions can occur because of fear, anger, disgust etc. Reading facial expressions and micro-expression of emotion can aid the development of rapport, trust, and collegiality; they can be useful in making credibility assessments, evaluating truthfulness of convicts and detecting deception; and better information about emotional states provides the basis for better cooperation, negotiation, or sales.

Recent improvements in this particular field of facial expression recognition have encouraged researchers around the globe to extend the applicability of facial emotion recognition to areas like chat room avatars and video conferencing avatars. The ability to recognize emotion can be very valuable in suspect detection systems and intelligence improvement systems for children with brain development disorders.

II. FACE DETECTION

Starting step of any facial expression recognition system is to acquire image or to collect a uniform dataset (particularly in developing a new system). In our approach, we have used Cohn-Kanade Dataset [2] as input dataset. The goal of face detection is to determine whether or not there are any face present in image, if present, returns the image location and extent of each face [3]. There are numerous approaches and algorithms that have been employed for detection of faces [4].

Local Binary Patterns (**LBP**) first described in 1994 [5] [6]. It is a very powerful feature for pattern classification. LBP uses local binary patterns based on each pixel which compares the eight neighbouring pixels making it very efficient texture operator and describes texture features of image very effectively. The main disadvantage of LBP is that it can only be used for binary/gray images with low accuracy and is not recommended when small changes in face location.

- AdaBoost Algorithm is the first practical boosting algorithm which gives considerable accurate prediction output based on relatively weak and incorrect input. It combines simple classifiers linearly to produce a strong

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

classifier. Its detection rate is quite high as it uses multiple iterations to generate a strong classifier. Although it is very simple to implement but it is sensitive to noisy data and outline.

- Neural Network Based algorithm is used to detect frontal views of the face image. It examines small windows of an image and chooses whether each window contains face or not. The system intercedes between several neural networks to improve the overall performance which eliminates the complex task of manually selecting non face training examples [6]. The only drawback of such a system is that it is very complex and has low detection rate due to training of non-face images.
- Viola Jones object detection is the first framework to provide competitive object detection rates in real-time. It is proposed by Paul Viola and Michael Jones in 2001. It is robust algorithm and has very high detection rate and very low false positive rate [7]. It performs well in aspects of both accuracy and speed needed for real time face detection applications.

The feature extractor described in Viola Jones algorithm for object detection also performs the work of face detection very effectively and efficiently [10]. It also uses a part of AdaBoost Algorithm which combines classifiers in a cascade. So due to above said properties we in this particular work have used it to perform face detection part. Once the face is detected, it is cropped from the input image to avoid unwanted data/features using Bounding Box (rectangle) in this case using “`insertObjectAnnotation`” function of matlab.

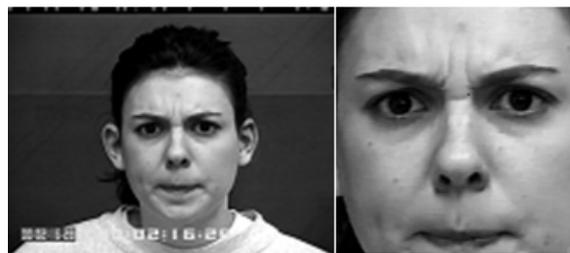


Fig.2 Example of Face Detection

III. FEATURE EXTRACTION

Once the face is detected from the input image the next step is to extract features from the face image. The features thus obtained are used to construct a feature vector which is then used to train the neural network to perform the complicated task of classification of emotions. The various feature extraction techniques are being classified into following categories by Zeng et al [11]:

A. Geometric based feature extraction

In this approach the features of the face are extracted using the properties of face, for example the relationship between the mouth and nose, relationship between areas of eyes, edge structures of the different areas of the face. These may include the shape and location of the corners of the mouth, nose, eyes, which are also known as fiducial points of face [12-16].

B. Modal based methods

This method is also known as template matching method as it attempts to find out the areas in the image whose characteristics match with a defined template of the face. This template may be based on the texture of image or on pixel intensity, edges or moments of the face. This technique gives better results over other available techniques but it is bit time consuming. In 2005 and early 2007 researchers used Gabor wavelets and haar features [17-19] to extract features, some also used temporal templates etc. [20]. Recently some researchers have also tried to combine both geometric and modal based features in order to get better results [21, 22].

In this work we have used another kind of statistical features called moments of image. Moments of an image are well-established shape descriptors and they have been used as features for various pattern recognition tasks [21].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

The potential of using moments of image as features was first proposed by Hu [22], in which he used moment invariants for pattern recognition which were invariant to effects of scaling, translation and rotation. The orthogonal moments proposed by Teague [23] were very effective in image recovery from different moments of the image. Some of the orthogonal moments are Zernike moments, Legendre moments and pseudo-Zernike moments. Lajevardi et.al used orthogonal moments [24]. This set of orthogonal moments is called Zernike moments and it is used for facial expression classification and recognition tasks.

IV. ZERNIKE MOMENTS

Zernike moments are geometric moments which are two dimensional functions of orthogonal polynomials on the unit disk. These are the projections of the image on an orthogonal set of polynomials (Zernike polynomials). They were first introduced by Mr. Zernike in 1934 [25]. The main advantage of such moments is that they are invariant to shift, rotation, and scale and are robust in the presence of noise which makes them suitable for pattern recognition tasks. Zernike moments differ from normal geometric moments in a way that the redundant moments are eliminated and the recovery of image from the moments is simple and straightforward.

The Zernike polynomial is defined as:

$$z_{nm}(x, y) = Z_{nm}(r \cos \theta, r \sin \theta) = R_{nm}(r) e^{im\theta} \quad (1)$$

Where n is the order of the polynomial and m represents the repetition.

The Zernike Polynomials are a series of complex-valued functions (polynomials) which are orthogonal on the unit disk:

$$x^2 + y^2 \leq 1 \quad (2)$$

Now the radial polynomial R_{nm} is given by:

$$R_{nm}(r) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+|m|}{2}-s\right)! \left(\frac{n-|m|}{2}-s\right)!} (r)^{n-2s} \quad (3)$$

Where $n \geq 0, n \geq |m|$ and $n - |m|$ is even. The Zernike moments A_{nm} of order n and repetition m for a function $f(r \cos \theta, r \sin \theta)$ or $f(r, \theta)$ is defined as:

$$A_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(r \cos \theta, r \sin \theta) Z_{nm}^*(r, \theta) r dr d\theta \quad (4)$$

Value of m can vary from $-\infty$ to ∞ . For image $f(x, y)$ the continuous Zernike moment defined in Eq. (3) is given in its discrete approximation in the following manner:

$$A_{nm} = \frac{n+1}{\pi} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) z_{nm}^*(x, y) \quad (5)$$

$$A_{nm} = \frac{n+1}{\pi} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) R_{nm}^*(r_{xy}) e^{-jm\theta_{xy}} \quad (6)$$

M, N define size of the image, $j = \sqrt{-1}$, and $*$ denotes the complex conjugate.

So, the Zernike moment of an image is the inner product between the image function and Zernike polynomial and is computed by using Eq. (4).

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Coordinates of image are normalized to 0 and 1. Once normalization is done, the discrete polar coordinates at the pixel of (x, y) are given by the following equations where NxN is the dimension of image:

$$r_{xy} = \frac{\sqrt{(2x-N+1)^2 + (N-1-2y)^2}}{N}, \theta_{xy} = \tan^{-1} \left(\frac{N-1-2y}{2x-N+1} \right) \quad (7)$$

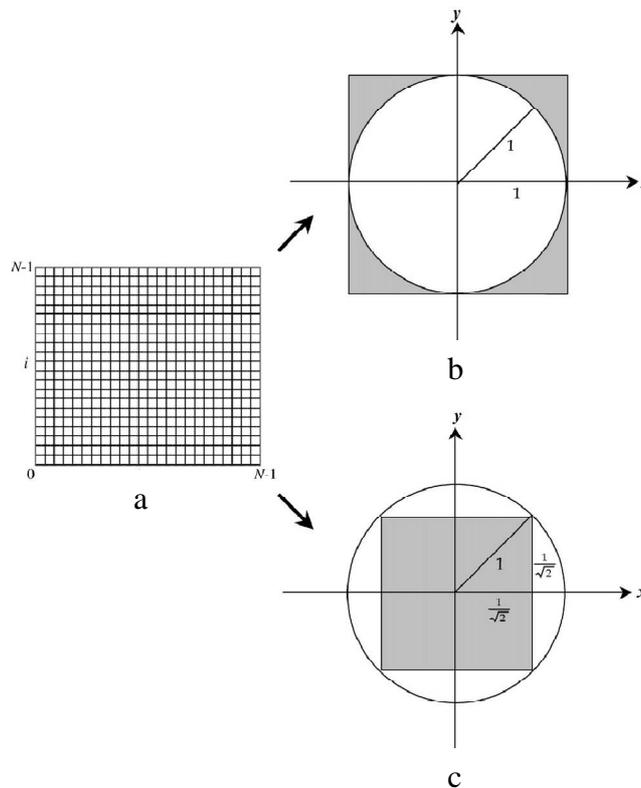


Fig.3 (a) Image plane N×N, (b) mapping of image plane N×N over a unit circle and (c) mapping of image plane N×N inside a unit circle.

To calculate Zernike moments of an image, the image is first written in polar coordinates on a unit circle, so that the centre of image is the centre of the circle and the pixels which are only inside the circle are used for calculations but the pixels outside the circle are not used in calculations [26].

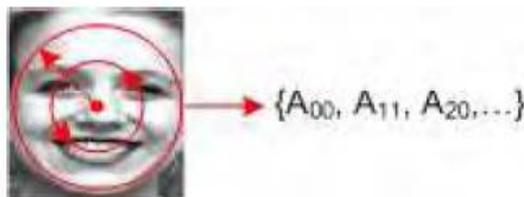


Fig.4 Example of ZM for feature extraction with different orders and repetition

While calculating moments of signature images, moments of higher order upto 10 are calculated. Table 1 lists the various Zernike moments according to their order and dimensionality.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Table I: the first 10order Zernike moments

Order	Dimensionality	Zernike moments
0	1	A_{00}
1	2	A_{11}
2	4	A_{20}, A_{22}
3	6	A_{31}, A_{33}
4	9	A_{40}, A_{42}, A_{44}
5	12	A_{51}, A_{53}, A_{55}
6	16	$A_{60}, A_{62}, A_{64}, A_{66}$
7	20	$A_{71}, A_{73}, A_{75}, A_{77}$
8	25	$A_{80}, A_{82}, A_{84}, A_{86}, A_{88}$
9	30	$A_{91}, A_{93}, A_{95}, A_{97}, A_{99}$
10	36	$A_{10,0}, A_{10,2}, A_{10,4}, A_{10,6}, A_{10,8}, A_{10,10}$

V. METHODOLOGY

A. Dataset

To perform analysis we have used Cohn-Kanade dataset. This dataset is mostly used by the researches for training and testing of facial expression recognition system. The dataset comprises of approximately 500 images of 100 different persons. In our study we have taken one image of each person with particular facial expression for all six expressions in this way we have total of 36 images. The selected images constitute six basic emotions viz. anger, disgust, fear, sad, happy, and surprise of each of the individual.

B. Image Pre-processing

For system to work efficiently and flawlessly the input image data should be normalized, to do so we perform image pre-processing using following steps: frame selection from sequence of images, face detection, intensity normalization and at last making all face images of uniform size.

C. Face Detection

Face detection part is done by using HAAR like feature base cascade classifier called Viola-Jones algorithm. In Viola-Jones algorithm the cascade function is trained with machine learning algorithm using a set of images. Once system is trained, the classifier is used to detect face from the input image. After that the size of the face image is normalized to 60x60 pixels.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

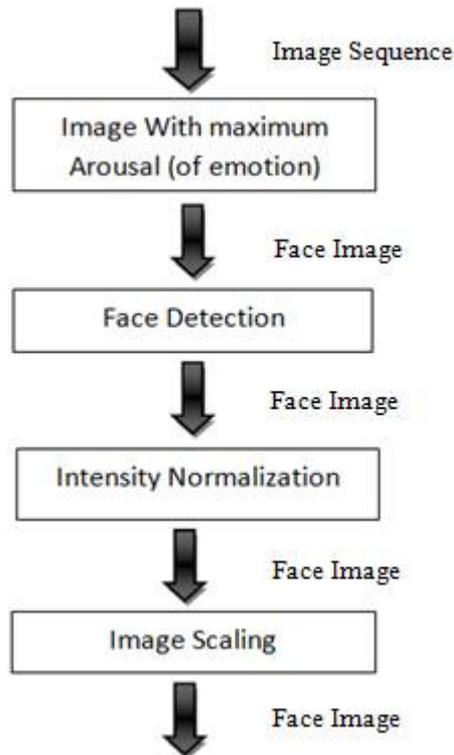


Fig. 5 Image Pre-processing Flowchart

D. Feature Extraction

After pre-processing of image the next step is to extract features of the image and to construct the feature vector of the image. The features are extracted by finding Zernike moments of the image. It was found out that higher order Zernike moments lead to higher efficiency in facial expression classification. For each class of emotion we have six images of six different individuals and feature vector is constructed using higher order Zernike moments. The features thus obtained are stored in the form of a table.

E. Classification using Artificial Neural Network (ANN)

The most effective and efficient way to perform pattern classification and recognition is ANN. The neural network consists of neurons which are interconnected, where each unit takes an input, applies a function to it and then passes its output to next layer.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

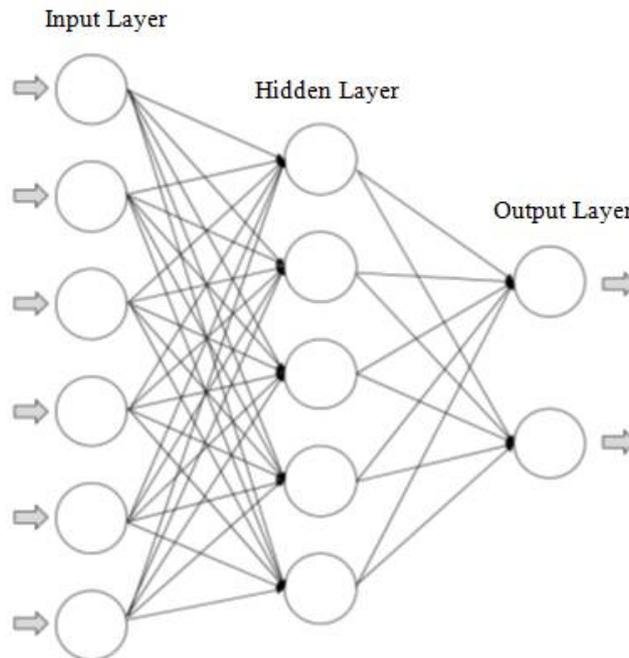


Fig.5 3-layer Neural Network

Once the Zernike moments of face images are found out, the classification process is carried out by using ANN. The feature vector thus obtained is fed to the neural network to train it and then to identify the individual emotion for given data. To classify the emotions through facial expressions we have made use of a statistical tool “STATISTICA” version 12. The data to be classified is loaded to the software, the tedious task of defining type of training algorithm and number of layers, neurons is done by the software itself. It automatically finds out the best learning algorithm. In our case we got best recognition rate for Multilayer Perceptron model. The model is of 3 layers with input and output layers comprising of 6 neurons and 5 hidden layer neurons.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

To test our system we take 36 input images from Cohn-Kanade dataset. Then these images are preprocessed and are saved to a different directory. After this the images are used to compute Zernike moments and a feature vector comprising these is constructed for ANN based classification system. The complete feature vector is fed to STATISTICA, which then itself divides the available dataset into training data and testing data. Once the network is trained it can be used for testing purposes. During testing phase the system was able to achieve accuracy of about 74%. The primary objective of this study was to develop a system which could automatically detect emotional state of a human being by using his/her facial expressions.

VII. CONCLUSION

In this work we have used Zernike moments as features of images for recognizing emotions of human beings. After pre-processing higher order Zernike moments are used to construct feature vector which are then fed to the classifier. Experimental results show that the system was able to recognize emotions with an accuracy of 74%. The accuracy can be increased if we increase the number of images of input dataset. The other way to increase accuracy is to combine various feature extraction techniques.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

VIII. ACKNOWLEDGEMENT

The authors would like to express sincere gratitude to Mr. Amitava Das, Principal Scientist, CSIR-CSIO Sector 30 Chandigarh for much needed assistance and technical advice in conducting the research.

REFERENCES

- [1] Paul Ekman, "universal facial expressions of emotion," California mental health research digest, pp. 151–158, 1970.
- [2] Kanade, Takeo, Jeffrey F. Cohn, and Yingli Tian. "Comprehensive database for facial expression analysis." In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pp. 46-53. IEEE, 2000.
- [3] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. on PAMI*, 24(1):34–58, 2002.
- [4] T. Ojala, M. Pietikäinen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", *Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994)*, vol. 1, pp. 582 - 585.
- [5] Jump up^ T. Ojala, M. Pietikäinen, and D. Harwood (1996), "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", *Pattern Recognition*, vol. 29, pp. 51-59.
- [6] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, —Neural Network-Based Face Detection, *PAMI* January 1998.
- [7] P. Viola and M. Jones, —Rapid object detection using a boosted cascade of simple features, In *Proc. of CVPR 2001*.
- [8] Markan Lopar and Slobodan Ribaric, "An Overview and Evaluation of Various Face and Eyes Detection Algorithms For Driver Fatigue Monitoring Systems", September 19, 2013.
- [9] Zeng, Zhihong, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. "A survey of affect recognition methods: Audio, visual, and spontaneous expressions." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31, no. 1 (2009): 39-58.
- [10] Pantic, Maja, and Leon JM Rothkrantz. "Facial action recognition for facial expression analysis from static face images." *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 34, no. 3 (2004): 1449-1461.
- [11] Valstar, Michel F., Hatice Gunes, and Maja Pantic. "How to distinguish posed from spontaneous smiles using geometric features." In *Proceedings of the 9th international conference on Multimodal interfaces*, pp.38-45. ACM, 2007.
- [12] Valstar, Michel F., Maja Pantic, Zara Ambadar, and Jeffrey F. Cohn. "Spontaneous vs. posed facial behavior: automatic analysis of brow actions." In *Proceedings of the 8th international conference on Multimodal interfaces*, pp. 162-170. ACM, 2006.
- [13] Pantic, Maja, and Ioannis Patras. "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences." *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 36, no. 2 (2006): 433-449.
- [14] Chang, Ya, Changbo Hu, Rogerio Feris, and Matthew Turk. "Manifold based analysis of facial expression." *Image and Vision Computing* 24, no. 6 (2006): 605-614.
- [15] Littlewort, Gwen C., Marian Stewart Bartlett, and Kang Lee. "Faces of pain: automated measurement of spontaneous all facial expressions of genuine and posed pain." In *Proceedings of the 9th international conference on Multimodal interfaces*, pp. 15-21. ACM, 2007.
- [16] Bartlett, Marian Stewart, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier 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, pp.568-573. IEEE 2005.
- [17] Whitehill, Jacob, and Christian W. Omlin. "Haar features for FACS AU recognition." In *7th International Conference on Automatic face and gesture recognition, FGR 2006*, pp. 99-101. IEEE, 2006.
- [18] Valstar, Michel, Maja Pantic, and Ioannis Patras. "Motion history for facial action detection in video." In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 1, pp. 635-640. IEEE, 2004.
- [19] Zhang, Yongmian, and Qiang Ji. "Active and dynamic information fusion for facial expression understanding from image sequences." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, no. 5 (2005): pp. 699-714.
- [20] Tian, Ying-Li, Takeo Kanade, and Jeffrey F. Cohn. "Facial expression analysis." In *Handbook of face recognition*, pp. 247-275. Springer New York, 2005.
- [21] Kim, Hyung Shin, and Heung-Kyu Lee. "Invariant image watermark using Zernike moments." *Circuits and Systems for Video Technology, IEEE Transactions on* 13, no. 8 (2003): 766-775.
- [22] Ming-Kuei Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol.8, no. 2, pp. 179–187, 1962.
- [23] Teague, Michael Reed. "Image analysis via the general theory of moments*." *JOSA* 70, no. 8 (1980): 920-930.
- [24] Lajevardi, Seyed Mehdi, and Zahir M. Hussain. "Higher order orthogonal moments for invariant facial expression recognition." *Digital Signal Processing* 20, no. 6 (2010): 1771-1779..
- [25] Zernike, F. (1934). "Beugungstheorie des Schneidenverfahrens und Seiner Verbesserten Form, der Phasenkontrast methode". *Physica* 1 (8): 689–704.
- [26] Lajevardi, Seyed Mehdi, and Zahir M. Hussain. "Higher order orthogonal moments for invariant facial expression recognition." *Digital Signal Processing* 20, no. 6 (2010): 1771-1779
- [27] Murari Mandal, Shashi Poddar, Amitava Das. "Comparison of Human and Machine based Facial Expression Classification." *International Conference on Computing, Communication and Automation 2015*. pp. 1198-1203.