INDIAN STATISTICAL INSTITUTE

INTEGRATED APPROACH TO RECOGNITION OF MACRO & MICRO EXPRESSIONS FROM FACIAL IMAGES

Dissertation Submitted in Partial Fulfillment of
The Requirements for the Award of The Degree

Of

Master of Technology
in
Computer Science

Submitted by:
Debayan Mukherjee.
Roll No.: CS-1510

Under Supervision of
Prof Dipti Prasad Mukherjee.

M.Tech(CS) DISSERTATION THESIS COMPLETION CERTIFICATE

Student: Debayan Mukherjee (CS-1510)

Topic: Integrated Approach to Recognition of Macro and Micro Expressions From Facial Images

Supervisor: Prof. Dipti Prasad Mukherjee

This is to certify that the thesis titled " **Integrated Approach to Recognition of Macro and Micro Expressions From Facial Images"** submitted by **Debayan Mukherjee** in partial fulfillment for the award of the degree of Master of Technology (Computer Science) is a bonafied record of work carried out by him under my supervision. The thesis has fulfilled all the requirements as per the regulations of this Institute and, in my opinion, has reached the standard needed for submission. The results contained in this thesis have not been submitted to any other university for the award of any degree or diploma.

Date: Prof. Dipti Prasad Mukherjee

Dedication

To my parents and my well wishers, without your help and encouragement it would not have been possible.

Acknowledgements

I would like to thank my dissertation supervisor Prof. Dipti Prasad Mukherjee for agreeing to guide me and for helping me to undertake work in the topic.

Abstract

We present a computer based approach which recognizes emotional facial expressions along with subtle micro expressions. Our system captures live video streams and recognizes both macro and micro expression at the same time. This procedure integrates the task where a set of frames is classified whether it contains micro expressions or macro expressions at first. Afterwards corresponding micro or macro expressions are further classified to its type. We have introduced Motion Magnification using Eulerian Video Magnification method to magnify subtle local motions that helps identifying types of micro expression. We have done extensive experiments on two publicly available data sets CK+ containing videos of posed macro expressions and CASME II containing videos of posed micro expressions. Using variants of LBP features such as Uniform LBP and Rotation Invariant LBP feature dimension is reduced by 80% compared to normal LBP based features that leads to reduction of response time and increases computational efficiency and results in effective classifier performance.

Contents

Chapter 1 :: Introduction

FACIAL expressions (FE) are one of the major ways that humans convey emotions. Aside from ordinary FEs that we see every day, under certain circumstances emotions can also manifest themselves in the special form of micro-expressions(ME). An ME is a very brief, involuntary FE shown on people's face according to experienced emotions. ME may occur in high-stake situations when people try to conceal or mask their true feelings for either gaining advantage or avoiding loss. In contrast to ordinary FEs, MEs are very short (1/25 to 1/3 second, the precise length definition varies[10], [11]), and the intensities of involved muscle movements are subtle.

A major reason for the considerable interest in MEs is that it is an important clue for lie detection [12]. Spontaneous MEs occur fast and involuntarily, and they are difficult to control through one's willpower. In high-stake situation for example when suspects are being interrogated, an ME fleeting across the face could give away a criminal pretending to be innocent, as the face is telling a different story than his statements. Furthermore as has been demonstrated in [20] people who perform better at ME recognition tests are also better lie detectors.

However, detecting and recognizing MEs are very difficult for human beings (in contrast to normal FEs, which we can recognize effortlessly). Study shows that for ME recognition tasks, people without training only perform slightly better than chance on average. This is because MEs are too short and subtle for human eyes to process. The performance can be improved with special training, but it is still far below the 'efficient' level. Moreover, finding and training specialists to analyze these MEs is very time-consuming and expensive.

The objective of this project is to build a system where a person sitting in front of live camera giving some facial expression intentionally or unintentionally is captured and analyzed in real time. The system detects, recognizes and reports Facial Macro Expressions that are given intentionally by one person along with the Micro Expressions that are appearing in ones face together at the same time. This is a challenging task as occurrence of micro expression is very subtle and detecting subtle changes can be affected by some noise of input device or some global head movement or some regular subtle motions like eye blink etc.

Related Works:

Several works have been done in this domain previously where individually facial expressions are analyzed and identified from posed facial expressions. Examples of realtime recognition of emotions from facial expressions include Berretti et al. [13] extracted scale invariant feature transform (SIFT) features from some sample points on the depth map of 3D scan of face. To choose a small set of features from initial 10112 dimensional feature, they employed minimum redundancy maximum relevance (mRMR) feature selection tool. Support vector machine (SVM) was used as the classifier. While Berretti et al. [13] applied SIFT feature on the landmark points, Danelakis et al. [14] proposed a spatio-temporal feature by applying wavelet transformation on the positional information of automatically detected six landmark points on the 3D face scan. Rashid et al [15] showed that combining features from both audio and visual channels of video improves emotion recognition accuracy.

Some researchers investigated ME recognition on posed ME databases. Polikovsky et al. [16], [17] used a 3D gradient descriptor for the recognition. Wu et al.[18] combined Gentleboost and an SVM classifier to recognize synthetic ME samples from the METT training tool. Several recent studies also reported tests on spontaneous ME databases. Pfister et al. [19] were the first to propose a spontaneous ME recognition method. The method combined three steps: first, a temporal interpolation model (TIM) to temporally 'expand' the micro-expression into more frames; second, LBP-TOP feature extraction; and third, using Multiple kernel learning for classification.

This work is focused upon the integrated approach to detect both the facial macro and micro expression in parallel along with finding out the percentage of all predefined expression within an appearance of mixed expression.

Chapter 2 :: Methodology

Finding out and recognizing Facial Macro Expression along with the Micro Expression is carried out by two levels of classification procedure as shown in the diagram.

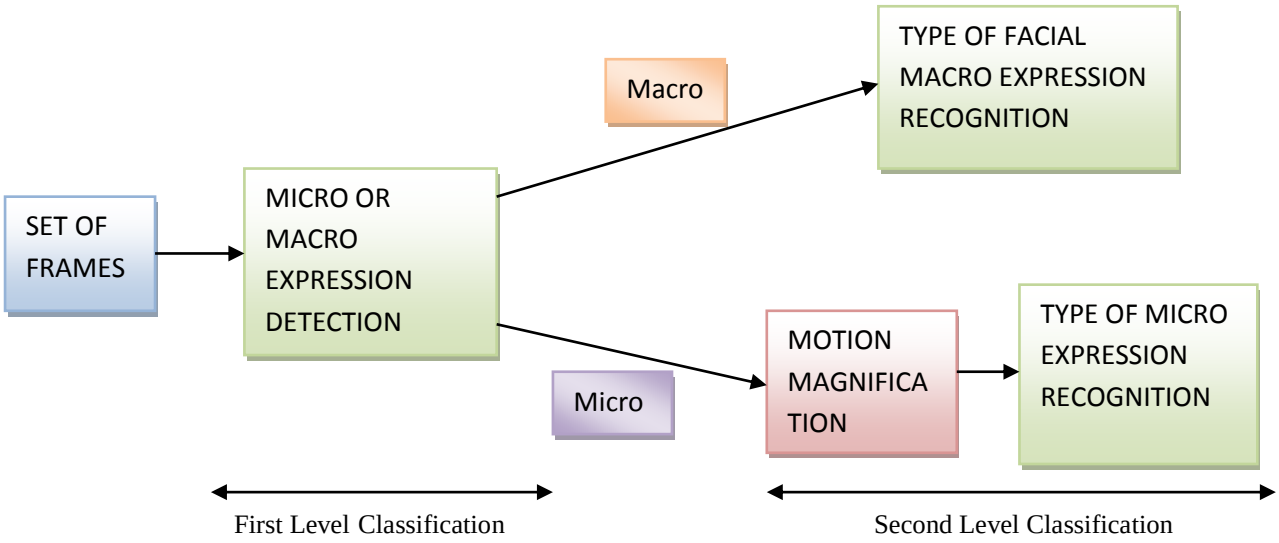

Figure 1: *Block diagram of algorithm comprising Micro and Macro expression detection together*

The first level of classification is to detect whether a video or a set of input frames containing one face is having micro expression or not. If micro expression is not detected within a particular set of frames then the set of frames is further sent to another classifier to recognize the type of facial macro expression. On the other hand if micro expression is detected among a set of frames then the set of frames are carried out for motion magnification and then the magnified video is sent to the classifier to recognize the type of micro expression.

Our motivation is to capture the set of frames using live video camera or web camera. Thus in this project the whole algorithm is experimented over live video stream to detect and recognize both facial macro and micro expression from live data.

The training part for the above algorithm is elaborated in next section

2.1 Steps to Detect Macro and Micro Expressions:

2.1.1. Feature Generation

Two datasets – CK+ [4] for facial macro expression and CASME II [5] for micro expressions are used for this classifier. These databases contain several expressions and for one particular expression there are several videos or set of frames. For each such set of frames containing one particular macro or micro expression one feature vector is generated at the very beginning of the algorithm.

The procedure to generate feature vector from one set of frames is shown below:

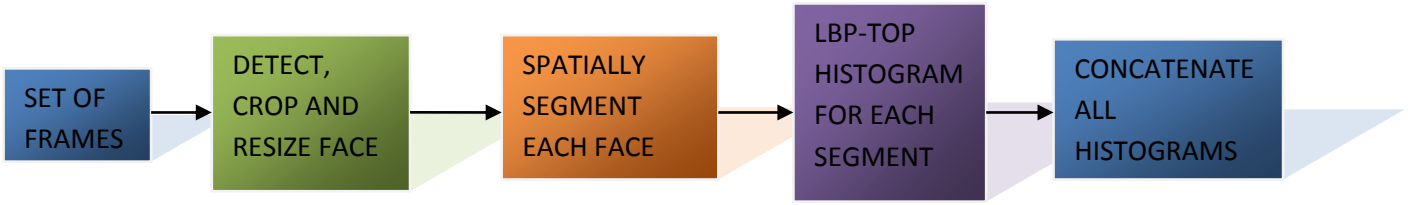

Figure 2: *Block diagram to explaining generation of feature vectors*

To obtain the feature vector, first faces are cropped out from each frame. In this step Viola Jones [6] face detection algorithm is applied to the first and last frame of one video of the training set. Thus we might get two different rectangular regions comprising single face area from first and last frame because of the presence of some global or local head movements. The union of two such regions is used as the standard region that is applied to each frame to crop out faces. After each faces being cropped out, all of them are resized to a fixed size say H×W many pixels.

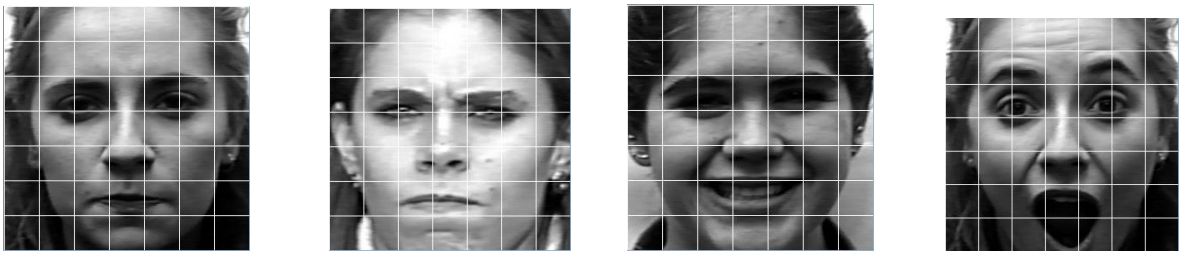

Figure 1: *Some sample frames containing cropped out and segmented face.*

Thus from one set of frames we obtain one set of faces. Each face from that set of faces are then divided into several identical spatial segments. Let us assume M×N numbers of spatial segments are taken out from one face and there are T numbers of faces for one particular training data. Thus considering one particular segment we get H/M × W/N × T number of pixels. LBP-TOP histogram is then calculated over H/M × W/N × T number of pixels for each special segment.

Description of LBP Feature In LBP, as described in [1], [2] each pixel (say, q_i with gray value g(qi)) of an image is assigned a number that represents the texture in the neighborhood of that pixel. Let, pj ∈ Nζ(qi) where Nζ(q_i) denotes the ζ-neighbors of a pixel q_i in an image and let the gray value of pj be represented by g(pj). The gray level values of ζ number of pixels along the circumference of a circle with center $\mathbf{q_i}$ and radius **r** are thresholded with respect to g(qi). Thus we get a ζ−bit binary number (say, $b_{\zeta-1}$ ...$b_3b_2b_1b_0$) such that b_j = 1, if g(p_j)>g(q_i), else b_j = 0,∀j ∈{ 0,1,2,...,ζ−1}. The ζ-bit binary number ($b_{\zeta-1}$ ...$b_3b_2b_1b_0$) is the LBP pattern of the pixel q_i. For representation purpose, this binary pattern may be converted into decimal (say, d(q_i)). For ζ number of neighboring pixels, 2^{ζ} such LBP patterns are possible. Now, LBP histogram is obtained via calculating total number of pixels corresponding to each of such LBP patterns.

Following, we consider uniform binary patterns [1],[2]. Uniform binary patterns (u2) are those binary patterns where the transition from 0to1 or from 1to0 is not more than 2 times while considering circular pattern. Considering 8-neighbors, 256 binary patterns are possible, whereas there are only 58 uniform patterns (u2) for 8-bit positions. It is seen that among different LBP patterns, the uniform patterns are mainly used to represent the texture of facial expressions. The uniform LBP patterns significantly reduce the number of feature dimensions but still represent reliable expression-related information.

Another significant variation of LBP is Rotation Invariant LBP [1], [2] where for one particular bit pattern it is rotated circularly and among all such rotated bit patterns the smallest valued one is considered. This again reduces the number of binary patterns significantly. For example if we again consider 8-neighbors then Rotation Invariant LBP reduces number of bit patterns from 256 to 36 uniform patterns.

Both the Uniform Binary Pattern and Rotation Invariant LBP are used in this project. In order to extract the LBP-TOP histogram the LBP feature is calculated over three orthogonal planes X-Y, X-T and Y-T and concatenated, where T signifies the time axis and X, Y signifies the spatial axis i.e. rows and columns of each frame. Thus we get LBP-TOP feature for one particular spatial segment of the whole face.

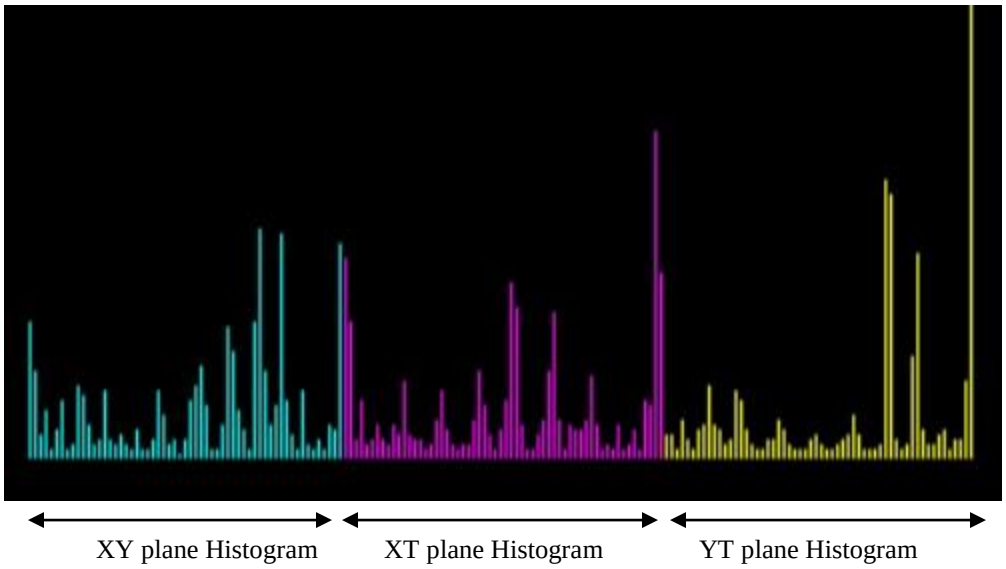

Figure 3: *Sample LBP-TOP histogram from one special segment of one face*

Finally LBP-TOP histogram of all M×N no of segments are concatenated one after another to get the final feature vector from one set of frames.

2.1.2. Classification

Using the above methodology features are generated from each set of frames or videos belonging to CK+ [4] or CASME II [5] dataset. After that all the features corresponding to facial macro expression obtained from CK+ dataset are treated as the one class of data and features corresponding to micro expressions obtained from CASME II dataset are treated as another class of data. Then a binary classifier is applied to classify the data. To get the probabilistic measurement of whether a strong micro expression is occurring or not it is better to use random forest as a classifier to classify among micro and macro expression.

CASME II [5] dataset contains videos where each of them starts with Neutral Facial expression and after occurrence of some subtle micro expression again it comes back to the Neutral Facial expression. Thus compared to the Facial Macro expressions of CK+ [4] data sets, CASME II [5] dataset contains videos that contain neutral facial expression. As a result if in this classifier a set of frames is classified as micro expression then it can be inferred that the set of frames is more likely to have neutral facial expression rather than any macro expression. Thus along with the micro or macro expression classification, neutral facial expression can also be detected in this step.

2.2. Steps to Recognize Type of Facial Macro Expression

This step is carried out when the first level classifier classifies a set of frames contains macro expression. This is because the first level classifier classifies a set of frames containing micro expression when neutral facial expression is maintained throughout all the frames as mentioned in section 1.2. Steps to recognize facial macro expressions are as follows:

2.2.1. Feature Generation

Feature generation in this step is identical to the steps of 1.1. Only CK+ [4] dataset along with the type of expressions as their class labels is used in this step in order to classify only the type of facial macro expressions. To recognize different types of facial macro expressions it is better to divide in special segments in such a way that each segment caries some significant part of our face that usually changes its shape or location during occurrence of one facial expression.

Concatenated LBP-TOP [1], [2] histogram is used in this step as the feature vector. For computational efficiency and to get rid from the rotational noise of one face both the rotational invariant LBP and uniform LBP is used. Thus all the LBP patterns consisting among both of the Uniform LBP and Rotation Invariant LBP are considered to be taken as histogram bucket. This LBP-TOP measurement is again applied for different segment of one set of face, followed by a concatenation of each LBP-TOP feature for different segments. This suffices the feature generation task.

2.2.2. Classification

This is a multi class classification process where different expressions comprise different class labels. Random Forest classifier is trained for this purpose to get a probabilistic or fuzzy measurement from the test stream of data comprising the amount of presence of each expressions or probability of each expression in a set of faces.

2.3. Steps to Recognize Type of Micro Expression:

This process is carried out when a strong micro expression is detected in the first level of classification. Strong micro expression detection signifies that the first level classifier classifies a set of frames as containing micro expression with a higher probability. This is important because if a set of frame is having no micro expression but only set of frames containing neutral expression then compared to macro expression of CK+ [4] data set it will classify this set of frame as containing micro expression as mentioned in section 1.2. Only CASME II [5] dataset is used to train this model where different kind of expressions are different class labels.

2.3.1. Motion Magnification

As Micro Expressions are very subtle in nature and lasts for very small amount of time in ones face, Motion Magnification plays important role for recognition of micro expression. Motion Magnification is done via using Eulerian Video Magnification method [7].

Description of Motion Magnification: Motion magnification using Eulerian Video Magnification [7] is carried out using the following steps:

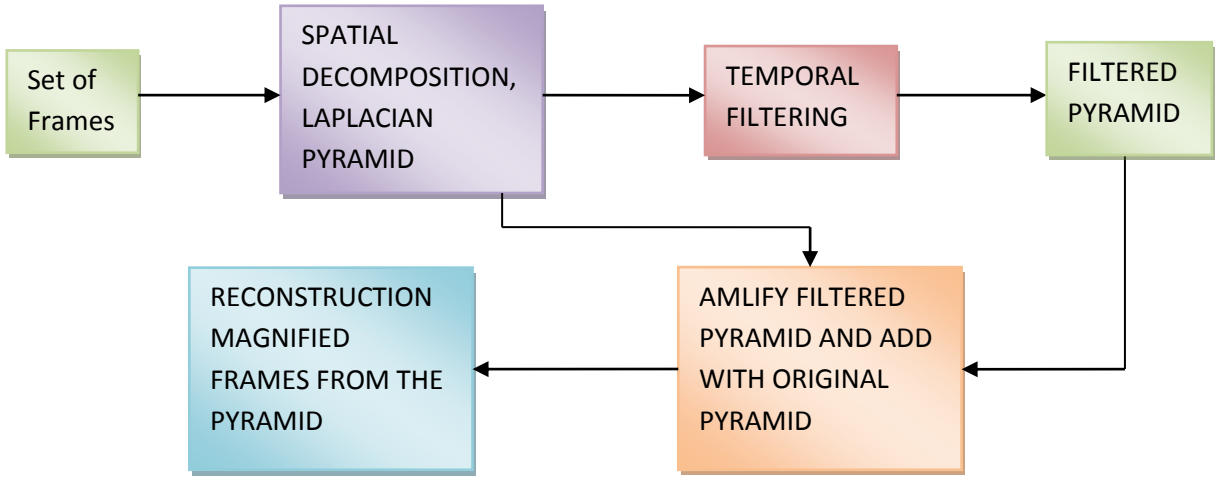

Figure 4: *Block diagram of motion magnification procedure*

Spatial Decomposition: The following diagram represents the special decomposition phase.

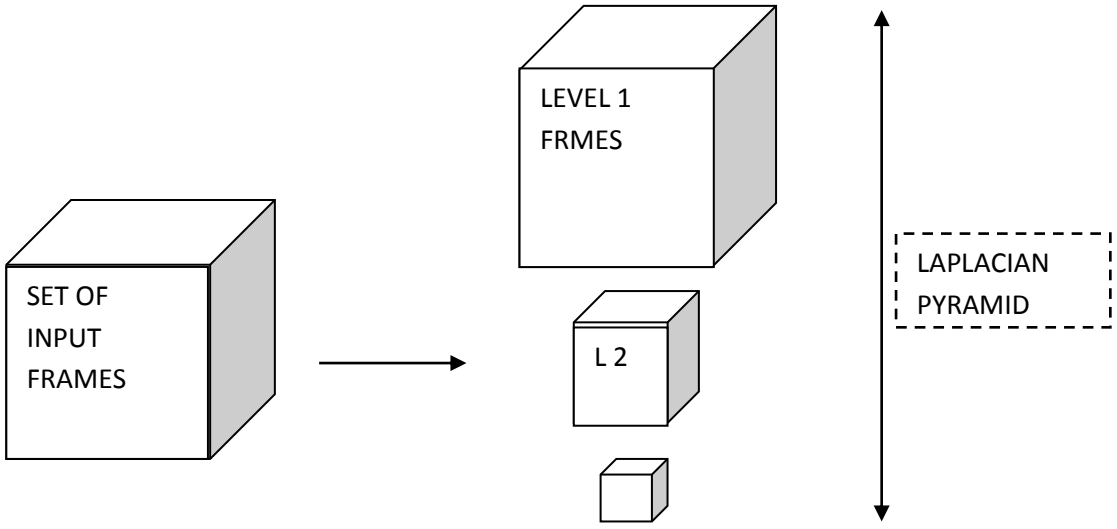

Figure 5: *Construction of Laplacian pyramid from a set of input frames*

This step basically creates standard Laplacian pyramid for each frame among the set of input frames. Laplacian pyramid is a kind of pyramid where each level except the bottom most stored high frequency information. This pyramid is generated by successively applying pyramid down and pyramid up mechanism.

In pyramid down procedure one frame is convolved with some Gaussian low pass mask. Then even rows and columns are deleted from the frame. Thus from one frame of size M × N we get another frame of size M/2 × N/2. On the other hand in pyramid up procedure blank rows and columns (rows and columns with all 0 values) are inserted between every successive rows and columns causing size of one frame increase from M × N to 2M × 2N. Afterwards the gray value of blank rows and columns are predicted using Gaussian convolution.

One level of pyramid is achieved by applying one pyramid down operation followed by one pyramid up operation. Thus applying these two operations successively will generate a frame F' with equal size but contain low pass information of the input frame F. Thus frame for one level is obtained by taking difference between the input and low pass frame (F – F'). The frame generated after pyramid down process is treated as the input frame for the next level and level by level the same procedure is carried out. For N number of input frames each level will contain exactly N number of frames as for each frame the Laplacian pyramid generation is carried out identically.

The bottom most level of this Laplacian pyramid contains low pass information and rest of the levels contains high pass information. For more depth of one level, the amount of noise will be less. Thus actual motion information excluding the noise is prominent at lower levels. Thus this step is one of the most significant steps for motion magnification.

Temporal Filtering:

This step is important to capture motion information in temporal domain. A motion is nothing but the change of pixel values throughout time. Thus this step concentrates upon finding out change of grey values among successive frames. Frequency information over successive frames plays a significant role to capture that gray value difference information.

An IIR band pass filter is applied over successive frames to extract motion information. For the range of frequency of bandpass filter f_h to f_l, the temporally filtered frames are generated using the formula

Lowpass1 = $(1 - f_l)$ * Lowpass1 + f_l * (i^{th} Input frame)

Lowpass2 = $(1 – f_h)$ * Lowpass2 + f_h * (i^{th} Input frame)

i^{th} filtered frame = Lowpass2 – Lowpass1

-initializing Lowpass1 and Lowpass2 to be the 1^{st} input frame at the beginning.

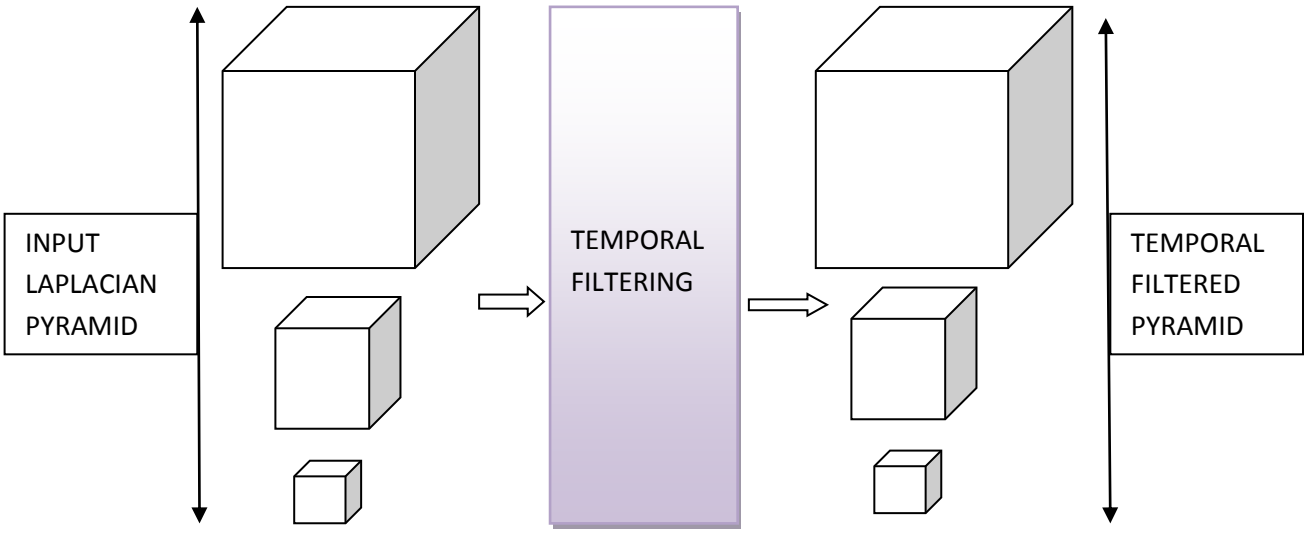

Figure 6: *Block diagram of temporal filtering of motion magnification.*

Amplification: Amplification of temporal filtered frames is a delicate process as amplifying the filtered frames with a high amplification factor may cause the noise gets amplified at a stake. Thus the amplification factor for deeper levels are kept higher compared to the amplification factor for the higher levels. This is because the deeper the level the less amount of noise will be there. Thus for one level the amplification factor is set as $\alpha * 2^i$ where i corresponds to one level.

Another important aspect is that the frequency for noise in temporal domain is usually higher than the usual motion in temporal domain. Thus to get rid from noise be amplified the amplification factor is gradually degraded beyond certain frequency or below certain wave length of temporal domain. This wave length is called cutoff wavelength. The following graph represents the change of amplification factor with respect to temporal wave length.

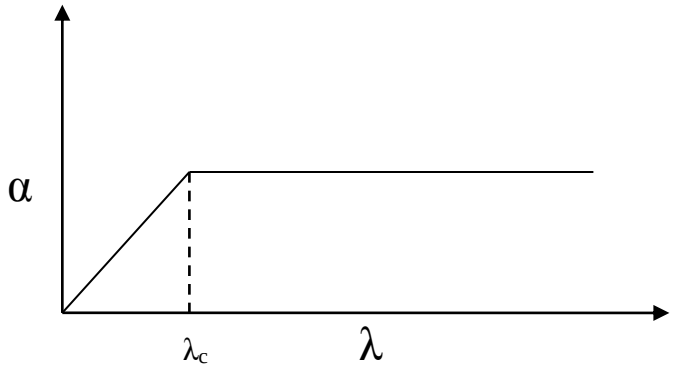

Figure 7: *Graph showing rate of change of amplification factor with respect to temporal wavelength*

where, α is the amplification factor, λ is the temporal wave length and λ_c is the cut off wave length.

The amplification along with addition procedure is shown in the following diagram:

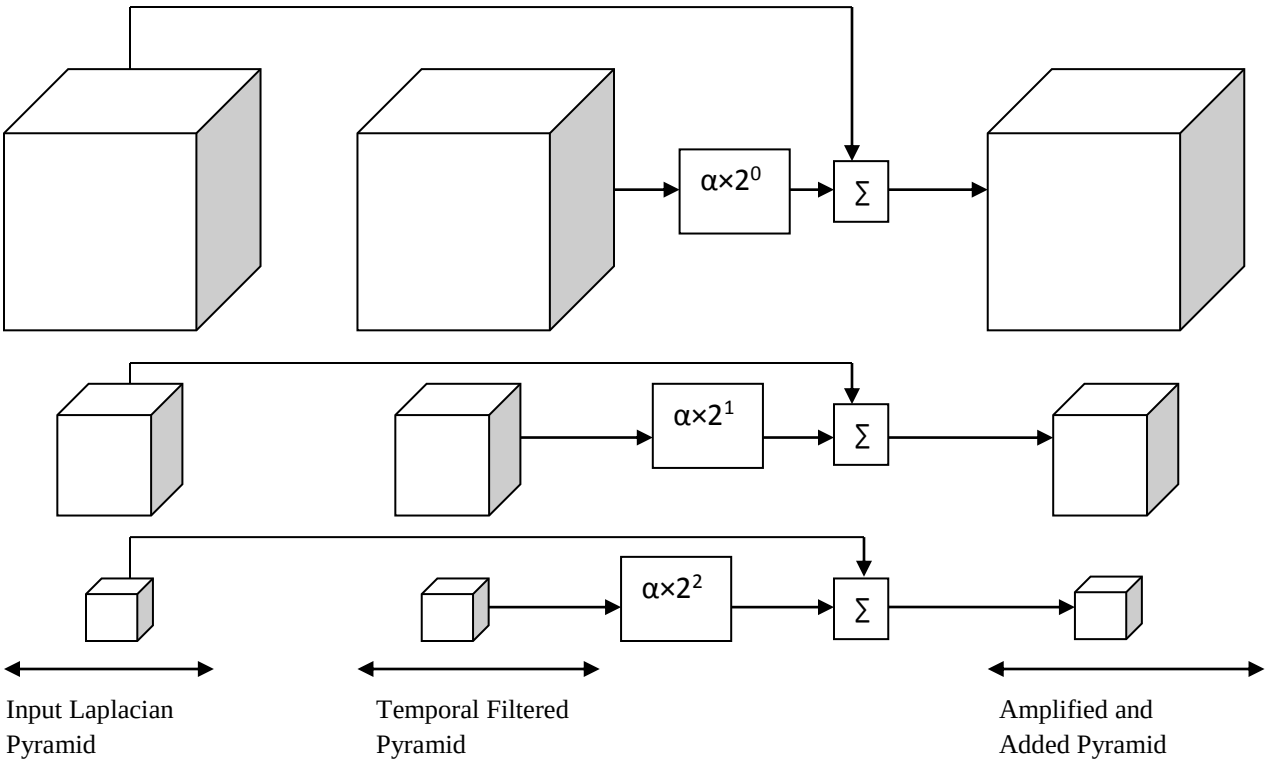

Figure 8: *Block diagram of amplification and addition procedure of motion magnification*

Reconstruct Frames from Laplacian Pyramid:

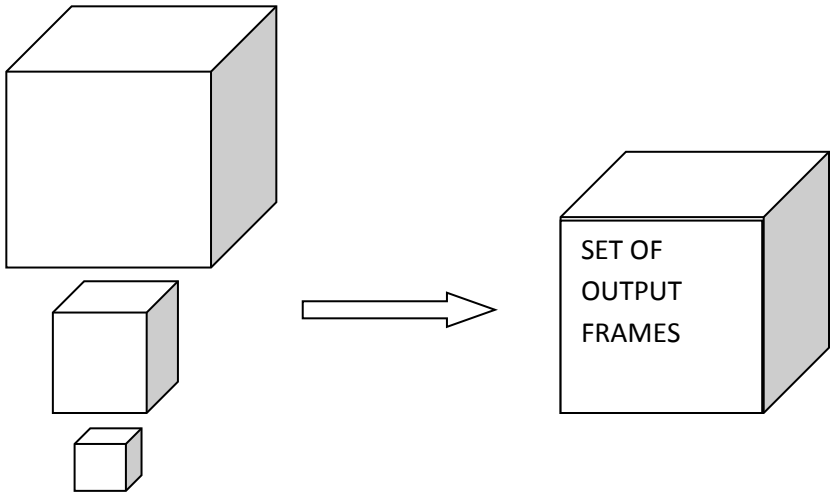

Figure 9: *Block diagram of reconstruction of output frames from amplified and added pyramid*

In this procedure from the bottom most level of the Amplified and Added Laplacian pyramid, Pyramid Up is carried out and added with the next upper level of the pyramid successively to get the final output frames. This output frames comprises the magnified video.

2.3.2. Feature Generation

Feature generation in this step is identical to the steps of 1.1. Only CASME II[5] dataset along with the type of expressions as their class labels is used in this step in order to classify only the micro expressions. To recognize different types of micro expressions it is better to divide in special segments in such a way that each segment caries some significant part of our face that usually changes its shape or location during occurrence of one facial expression. Motion magnified videos gives better classification result compared to raw frames.

Concatenated LBP-TOP histogram is used in this step as the feature vector. For computational efficiency and to get rid from the rotational noise of one face both the rotational invariant LBP and uniform LBP is used. Thus all the LBP patterns consisting among both of the Uniform LBP and Rotation Invariant LBP are considered to be taken as histogram bucket. This LBP-TOP measurement is again applied for different segment of one set of face, followed by a concatenation of each LBP-TOP feature for different segments. This suffices the feature generation task.

2.3.3. Classification

This is a multi class classification process where different micro expressions comprise different class labels. Random Forest classifier is trained for this purpose to get a probabilistic or fuzzy measurement from the test stream of data comprising the amount of presence or probability of all type micro expression in a set of faces.

Chapter 3 :: Results

Online Test Setup:

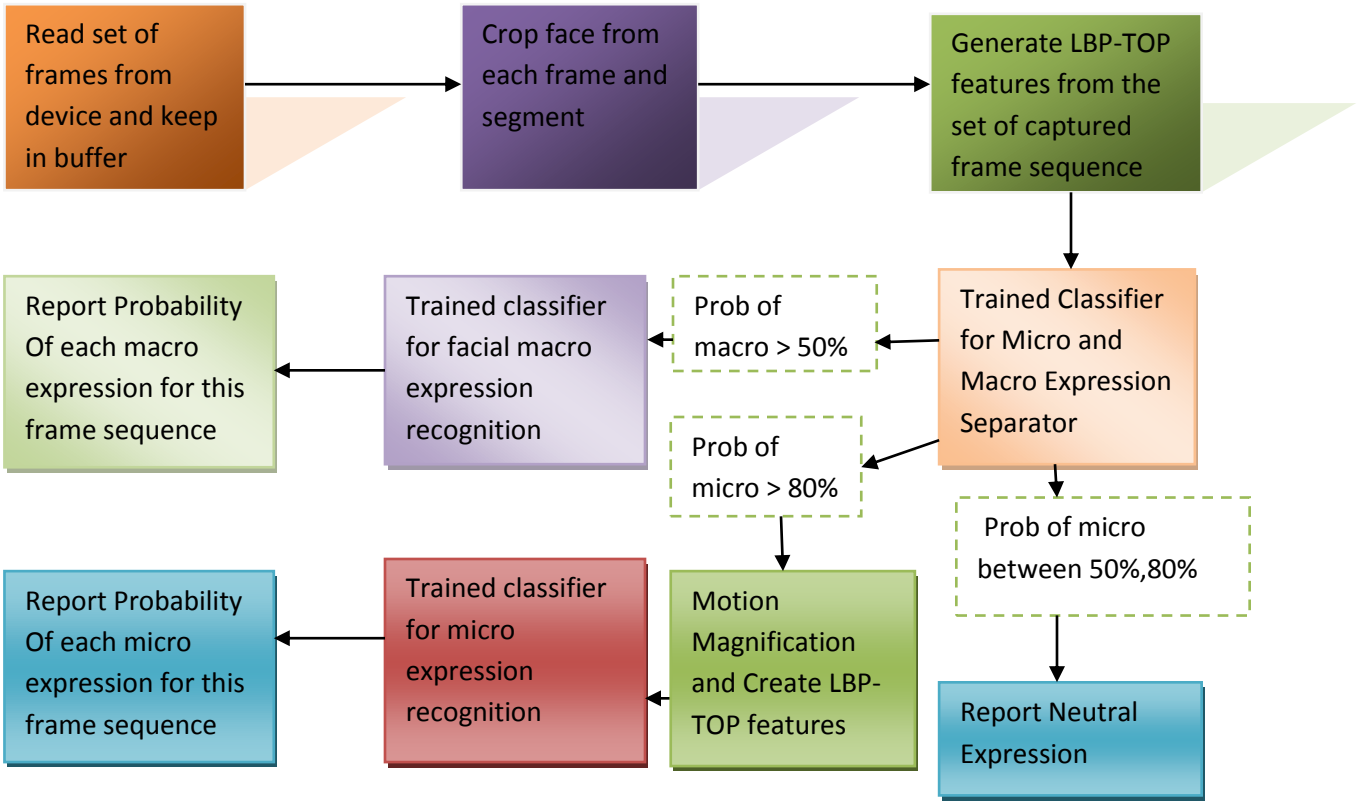

Figure 10: *Steps to process online video stream.*

Description of raining data set: As mentioned in the methodology part two data sets – (a)CK+ [4] and (b)CASME II[5] are used for testing the above algorithm. CK+ data sets includes 300 image sequence having six basic emotional expression (Anger, Disgust, Fear, Joy, Sorrow, Surprise) and CASME II data sets includes 253 videos having seven different type of micro expressions (Disgust, Fear, Happiness, Repression, Sadness, Surprise, Others).

Experimental setup: In the first level of classification all 300 frame sequence of CK+ data set generating 300 data points is considered as one class of data and similarly 253 data points of CASME II data set are considered as another class of data.

The second level of classification is performed over some specific micro and macro expressions. The facial macro expression classifier is trained using four different expressions Anger, Disgust,

Joy and Surprise. The second level micro expression classifier is trained using four micro expressions – Disgust, Happiness, Repression and Surprise. This is because these expressions contains numerous samples and well distinguishing in nature.

Experimental Setup for Face Cropping and Segmentation: For better performance over online videos for each level of classification each cropped face are resized to 280 by 280 pixels and divided into 7 partitions row wise and 7 partition column wise. Thus each segment contains 40 by 40 pixels. This gives better result compared to other because the 7 by 7 segment - partitions one face in such a way that significant spatial locations of a face are kept in different partition.

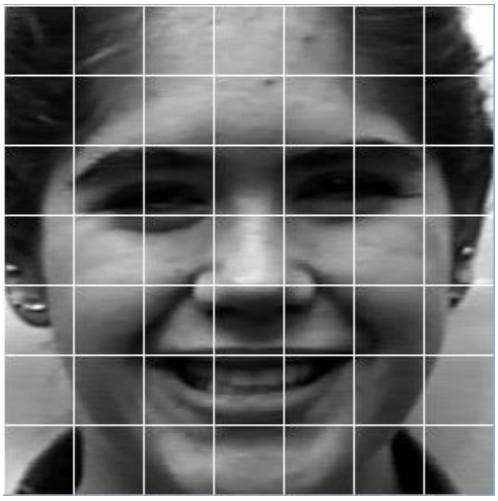

Figure 11: *Sample frame representing cropped out and segmented frame*

In the above sample it can be noticed that inner eye corner and the outer eye corners, nose tip, both the edges of lip, different portions of eye brows, and portion between two eyes are represented with separate spatial blocks. Thus temporal changes of all those significant face portions are taken into account differently for detection of one particular emotion.

Experimental Setup to Generate LBP Feature: LBP-TOP histogram bins are constructed using both Uniform and Rotational Invariant LBP. To obtain this first a dictionary is constructed that contains 58 unique values of uniform LBP. Then for each pixel the Rotational Invariant LBP value is measured. If that particular value belongs to any of that 58 Uniform LBP pattern then the corresponding histogram bin is used. Otherwise an extra bin is allocated that contains all the non uniform LBP pixel counts. Thus length of LBP feature corresponding to one orthogonal plane of one image segment is 59. For three orthogonal planes we have 59*3 numbers of bins. Thus from one frame sequence the length of feature or the LBP-TOP histogram is 59*3*7*7. This is the actual feature length extracted from one frame sequence. These standards are maintained for all three of the classification process.

Training Result for the First Level of Classification:

Random forest is used to train this level of classification. Several parameters are checked to obtain good performance. Compared to other parameters, keeping maximum depth of the random forest 5 and applying the constraint that no leaf node should contain more than 10 data points having more than one class and total number of tress 150, gives us the best result for all the classification problems.

Using the above parameters the first micro and macro expression classifier gives 100% accuracy to discriminate datasets from CASME II and CK+ using 50% of the whole data as training and rest 50% as test data. Applying the trained for live video stream results good performance classifying micro and macro expression. Generally, giving no facial posed expression results in classifying the set of online frames as containing micro expression according to as mentioned before.

Training Result for Recognition of Facial Macro Expression:

Data Separation: For classification among four facial macro expression – Anger, Disgust, Joy, Surprise, from 249 frame sequence containing those four expressions, 249 data points are generated and used for training purpose. Training is done by separating 80% of the data as training sample and 20% of the data as test sample. For that 80% dataset 10 fold cross validation is performed and the test data is kept completely unknown to the trained model.

Classifier Parameters: This training algorithm is executed several times with several parameters of Random Forest Classifier. For computational efficiency maximum depth of each tree is kept 5. The spitting condition of a node is decided as if a node contains 10 or more data points having different class label then it is split. For several number of trees the algorithm is executed and results for these are listed in the table below:

Number of trees	**Accuracy of Anger**	**Accuracy of Disgust**	**Accuracy of Joy**	**Accuracy of Surprise**	**Overall Accuracy**
50	82%	75%	85%	89%	84%
100	**90%**	78%	**90%**	86%	85%
125	85%	77%	**90%**	91%	85%
150	87%	81%	**90%**	**92%**	**88%**
200	86%	**82%**	**90%**	**92%**	87%

Table 1: *Training result for facial macro expression recognition*

Thus we get an overall accuracy of 90% for this four class classification problem.

For applying the trained classifier over online video the prediction for joy and surprise is better compared to the prediction of anger and disgust. For posed expression starting from neutral to an expression with high intensity is predicted correctly almost in all situations. The output of online video stream is given in terms of the probability of all four trained expressions.

Training Result for Recognition of Micro Expression:

Data separation: For classification among four facial macro expression – Disgust, Joy, Repression, Surprise from 146 frame sequence containing those four expressions, 146 data points are generated and used for training purpose. Training is done by separating 80% of the data as training sample and 20% of the data as test sample. For that 80% dataset 10 fold cross validation is performed and the test data is kept completely unknown to the trained model.

Parameters for Motion Magnification: It has been experimentally tested that 5 levels of special decomposition to construct Laplacian pyramid gives us good result and computationally efficient. Thus in Laplacian pyramid 5 level of high pass frames and one level of low pass frames are maintained as standard. For temporal filtering the experimental range of the bandpass filter is kept between 0.2 and 0.5. Greater range includes more noise causing the classifier performance worse. The cutoff wavelength (λ_c) is assigned to 600. The classification result varies with base amplification factor. It can be noticed that for low amplification factor the motion magnification is inefficient as it cannot magnify motions effectively and for high amplification factor the noise gets amplified worsening the classification result. Classification result with respect to amplification factor is shown in the table below:

Classifier Parameters: This training algorithm is executed several times with several parameters of Random Forest Classifier. For computational efficiency maximum depth of each tree is kept 5. The spitting condition of a node is decided as if a node contains 10 or more data points having different class label then it is split. For several number of trees the algorithm is executed and results for these are listed in the table below:

Number of trees	**Accuracy of Disgust**	**Accuracy of Joy**	**Accuracy of Repression**	**Accuracy of Surprise**	**Overall Accuracy**
50	39%	42%	44%	38%	40%
100	41%	37%	42%	42%	42%
125	40%	**43%**	**45%**	39%	42%
150	**44%**	41%	43%	**46%**	**43%**
200	42%	**43%**	40%	**46%**	**43%**

Table 2: *Training result for micro expression recognition without motion magnification*

Base Amplification Factor	Accuracy of Disgust	Accuracy of Joy	Accuracy of Repression	Accuracy of Surprise	Overall Accuracy
0.5	45%	42%	44%	46%	45%
1	44%	40%	47%	47%	45%
2	**48%**	**46%**	48%	**55%**	**52%**
5	**48%**	41%	**50%**	51%	49%
10	43%	39%	43%	47%	42%

Table 3: *Training result for micro expression recognition applying motion magnification*

In this table, number of trees are maintained 150 for all experiments.

Thus we get an overall accuracy of 52% for this four class classification problem after applying motion magnification and 40% without applying motion magnification.

For applying the trained classifier over online video the prediction for joy and surprise is better compared to the prediction of anger and disgust. For posed expression starting from neutral to an expression with high intensity is predicted correctly almost in all situations. The output of online video stream is given in terms of the probability of all four trained expressions.

Chapter 4 :: Conclusion and future direction

There are several contributions of this project over the facial macro and micro expression detection. First of all in several previous research works Micro expression has been detected via an unsupervised spotting algorithm that tracks difference measurement among frames via tracking movements in temporal domain. Some larger difference measure over smaller interval of time is reported afterwards as spotting of micro expression. This particular method is replaced by a supervised classification task of micro and macro expression detection. This also gives benefit for the detection of Neutral expressions considering data sets containing micro expressions are having Neutral expression in most of the frames.

Secondly, all the rotational invariant LBP values that are uniform are taken as histogram bucket, to generate features, which reduces length of the feature vector drastically and performs live emotion detection technique in real time. One last contribution is to modify the amplification process in motion magnification algorithm. This helps us reduce the outlier to get magnified and achieve a less noisy magnified video that leads to better performance.

The alignment of face in this project is not implemented explicitly. It is assumed that in live video of high fps capturing device, for some number of frames as the input set of frames the global motion is negligible. Though a very stable device with fewer amounts of head movements are required for this algorithm to give better performance.

We are working on extending the same framework for recognition of expressions for mood estimation of TV viewers or internet users of social networking sites. Note that our system is designed to recognize posed emotional expressions from near-frontal face videos. Posed expressions are expected to be exaggerated as compared to spontaneous ones. We may need more dense features to represent spontaneous facial expressions and micro expressions.

Chapter 5 :: References

1) Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. 27, 803–816 (2009). doi:10.1016/j.imavis.2008.08.005
2) Zhao, G., Pietikinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 915–928 (2007)
3) Kanade,T.,Cohn,J.,Tian,Y.L.:Comprehensive database for facial expression analysis. In: International Conference on Automatic Face and Gesture Recognition, pp. 46–53. IEEE (2000). doi:10. 1109/AFGR.2000.840611
4) Lucey, P., Cohn, J.F., Kanade, T., Saragih, J.: The extended cohnkanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: Computer Vision and Pattern Recognition, pp. 94–101 (2010)
5) W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," PloS one, vol. 9, no. 1, p. e86041, 2014.
6) Viola,P.,Jones,M.J.:Robust real-time face detection .Int.J.Comput. Vis. 57(2), 137–154 (2004)
7) H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world." ACM Trans. Graph., vol. 31, no. 4, p. 65, 2012.
8) Swapna Agarwal, Bikash Santra, Dipti Prasad Mukherjee : Anubhav: recognizing emotions through facial expression
9) Xiaobai Li, Xiaopeng Hong, Antti Moilanen, Xiaohua Huang, Tomas Pfister, Guoying Zhao, and Matti Pietik¨ainen : Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-expression Spotting and Recognition Methods
10) W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," Journal of Nonverbal Behavior, vol. 37, no. 4, pp. 217–230, 2013.
11) D. Matsumoto and H. Hwang, "Evidence for training the ability to read microexpressions of emotion," Motivation and Emotion, pp. 1–11, 2011.
12) ——, "Lie catching and microexpressions," The Philosophy of Deception, pp. 118–133, 2009.
13) Berretti, S., Amor, B.B., Daoudi, M., Del Bimbo, A.: 3d facial expression recognition using sift descriptors of automatically detected keypoints. Vis. Comput. 27(11), 1021–1036 (2011)
14) Danelakis, A., Theoharis, T., Pratikakis, I.: A spatio-temporal wavelet-based descriptor for dynamic 3d facial expression retrieval and recognition. Vis. Comput. 1–11 (2016). doi:10.1007/ s00371-016-1243-y
15) Rashid,M.,Abu-Bakar,S.,Mokji,M.:Human emotion recognition from videos using spatio-temporal and audio features.Vis.Comput. 29(12), 1269–1275 (2013)
16) S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in ICDP. IET, 2009, pp. 1–6.
17) S. Polikovsky and Y. Kameda, "Facial micro-expression detection in high-speed video based on facial action coding system (FACS)," IEICE Transactionson In formation and Systems, vol.96,no.1,pp.81–92,2013.

18) Q. Wu, X. Shen, and X. Fu, “The machine knows what you are hiding: an automatic micro-expression recognition system,” Affective Computing and Intelligent Interaction, pp. 152–162, 2011.
19) T. Pfister, X. Li, G. Zhao, and M. Pietikainen, “Recognising spontaneous facial micro-expressions,” in ICCV. IEEE, 2011.