ABSTRACT-
This paper evaluates facial recognition based on Local Binary Patterns on three orthogonal planes, Pyramid of Histogram of Gradients, and a geometrical feature based on distance between fiducial points for person-independent facial expression recognition. Different machine learning methods are systematically examined on two databases (posed and spontaneous). For posed database Cohn-Kanade has been used and for spontaneous database FeedTUM has been used. Extensive experiments illustrate that Local Binary Patterns on three orthogonal planes is effective and efficient for facial expression recognition. The best recognition performance is obtained by using Support Vector Machine classifier.

1. INTRODUCTION
The computer is a tool that humankind has been using for only a short time, and the way we use it is constantly and rapidly changing. The functioning of the computers performing these tasks is hidden to the general public. What is missing at this moment is a breakthrough in the way machines sense human actions to understand their intentions. Only then can a natural, human-centered way of interaction with our computerized environment be created. The machine will read human motion, identity, posture, gait, hand gestures, facial expression etc. Facial expression is one of the most powerful, natural and immediate means for human beings to communicate emotions and intentions. Often emotions are expressed through the face before they are verbalized. In the past decade, much progress has been made in building computer systems to understand and use this natural form of human communication. Humans detect and interpret faces and facial expressions in a scene with little or no effort. Still, development of an automated system that accomplishes this task is rather difficult. There are several related problems: detection of an image segment as a face, extraction of the facial expression information and classification of the expression (e.g., in emotion categories). A system that performs these operations accurately and in real time would form a big step in achieving a human-like interaction between man and machine. It is meant to serve as an ultimate goal and a guide for determining recommendations for development of an automatic facial expression analyzer.

Two procedures are necessary for an automatic expression analysis system: facial feature extraction and facial expression recognition. In facial feature extraction, there are mainly two types of approaches: geometric feature-based methods and appearance-based methods. In geometric feature-based methods, the facial components or facial feature points are extracted to form a feature vector that represents the face geometry. In appearance based methods, image filters are applied to either whole-face or specific regions in a face image to extract a feature vector. Geometric feature extraction can be more computationally expensive, but is more robust to variation in face position, scale, size, and head orientation. In facial expression recognition, most automatic expression analysis systems attempt to recognize a small set of prototypic expressions (i.e. joy, surprise, anger, sadness, fear, and disgust). The aim of this paper is to compare geometric and texture feature descriptors for facial expression recognition and to find the more suitable feature descriptor for facial expression recognition on posed and spontaneous datasets.

1.1 ORIGIN OF THE RESEARCH PROBLEM
Automatic facial expression analysis has been an active research area in the fields of computer vision and pattern recognition. Since 1978, the focus of attention has shifted a couple of times. First it shifted from the analysis of still images to analysis of face video. Later some researchers moved away from detecting a small number of prototypic expressions, such as emotions, to the detection of automatic facial actions. Recently the focus has shifted to two new aspects of facial expression analysis: the analysis of spontaneously displayed expressions and the analysis of the temporal dynamics of facial expressions. Most automatic facial expression analysis systems developed so far target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger [6, 22]. Even though automatic recognition of the six basic emotions from face images and image sequences recorded under controlled conditions is considered largely solved, reports on novel approaches are still being published even to this date. While message judgment studies distinguish only a very limited number of facial expressions, human experts can manually code nearly any anatomically possible facial expression using FACS. This is achieved by decomposing an expression into the specific AUs and their temporal segments. As AUs are independent of any (cultural) interpretation, they can in turn be used for any higher order decision making process. It is not surprising, therefore, that automatic AU coding in face images and face image sequences attracted the interest of computer vision researchers. Historically, the first attempts to encode AUs in images of faces in an automatic way were reported by Bartlett et al, Lien et al and Pantic et al. These three research groups are still the forerunners in this research field.

Still facial detection is an open problem in computer vision. However, in the last few years the research in facial part detection has reached the level of advancement which makes it usable for use in lab setting. In 2007, Guoqing Zhao (senior member IEEE) proposed a spatial-temporal facial feature descriptor based on local binary pattern [12]. In 2011, Roland Goecke proposed a method for automatic emotion recognition. The system extracts pyramid of histogram of gradients (PHOG) and local phase quantization (LPQ) features for encoding the shape and appearance information [20]. Shiqing Zhang in 2012 presented a new method of facial expression
recognition based on local binary patterns (LBP) and local Fisher discriminated analysis (LFDA) [21].

Peter W. Mc Owan in 2013 [24] presented a survey on automatic facial expression analysis, impacts, important applications in many areas such as human–computer interaction and data-driven animation. He evaluated facial representation based on statistical local features, Local Binary Patterns, for person-independent facial expression recognition. The research done by Guoqing Zhao and by Roland Goecke was based on feature descriptors only. They did not use geometric descriptor, so in this work we will look after the performance based on both geometric and appearance descriptors. Further another very commonly used feature PHOG can be experimented.

1.2 Basic Structure of Facial Expression Analysis Systems

Facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to automatic facial expression analysis (AFE/A) consists of three steps (Fig.1): face acquisition, facial data extraction and representation, and facial expression recognition.

![Fig1. Basic structure of facial expression analysis systems [36]](image)

Face acquisition is a processing stage to automatically find the face region for the input images or sequences. It can be a detector to detect face for each frame or just detect face in the first frame and then track the face in the remainder of the video sequence. To handle large head motion, the head finder, head tracking, and pose estimation can be evaluated facial representation based on statistical local features, Local Binary Patterns, for person-independent facial expression recognition. The research done by Guoqing Zhao and by Roland Goecke was based on feature descriptors only. They did not use geometric descriptor, so in this work we will look after the performance based on both geometric and appearance descriptors. Further another very commonly used feature PHOG can be experimented.

1.3 Deliberate Versus Spontaneous Expression

Most face expression data have been collected by asking subjects to perform a series of expressions. These directed facial action tasks may differ in appearance and timing from spontaneously occurring behavior. Deliberate and spontaneous facial behavior are mediated by separate motor pathways, the pyramidal and extra pyramidal motor tracks, respectively [5, 23, 24]. As a consequence, fine-motor control of deliberate facial actions is often inferior and less symmetrical than what occurs spontaneously. Many people, for instance, are able to raise their outer brows spontaneously while leaving their inner brows at rest; few can perform this action voluntarily [8, 14]. Spontaneous depression of the lip corners and raising and narrowing the inner corners of the brow are common signs of sadness. Without training, few people can perform these actions deliberately which incidentally is an aid to the detection [22]. Differences in the temporal organization of spontaneous and deliberate facial actions are particularly important in that many pattern recognition approaches and are highly dependent on the timing of the appearance change [23, 24]. Unless a database includes both deliberate and spontaneous facial actions, it will likely prove inadequate for developing face expression methods that are robust to these differences.

1.4 DATABASE

Two databases have been used in this research work to meet the objectives. FG-NET database for spontaneous facial expressions and COHN-KANADE dataset for posed facial expressions.

1.4.1 The FeedTUM Database

The FeedTUM Database with Facial Expressions and Emotions from the Technical University Munich is an image database containing face images showing a number of subjects performing the six different basic emotions defined by Eckman & Friesen. The database has been generated as part of the European Union project FG-NET (Face and Gesture Recognition Research Network). One of the dominating problems when gathering a database with emotions and facial expressions is the phenomena that played emotions differ from the natural ones. Therefore one of the underlying paradigms of this database is to let the observed people react as natural as possible. As a consequence, it was tried to wake real emotions by playing video clips or still images after a short introduction phase instead of telling the person to play a role. After extraction the images are separately stored in subdirectories as follows: {anger, disgust, fear, happy, neutral, sadness, surprise}. It contains 385 sequences.

![Fig: 2 Sample face expression images from the FeedTUM database [25]](image)
1.4.2 The Cohn-Kanade Database

In its current distribution, the CK contains 586 sequences across 97 subjects. Each of the sequences contains images from onset (neutral frame) to peak expression (last frame). Facial behavior of 210 adults was recorded using two hardware synchronized cameras. Participants were 18 to 50 years of age, 69% female, 81% Euro-American, 13% Afro-American, and 6% other groups. Participants were instructed by an experimenter to perform a series of 23 facial displays; these included single action units and combinations of action units. Each display began and ended in a neutral face with any exceptions noted. Image sequences for frontal views and 30-degree views were digitized into either 640x490 or 640x480 pixel arrays with 8-bit gray-scale or 24-bit color values.

![Sample face expression images from the Cohn–Kanade database](image)

2. FEATURE DESCRIPTORS

Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. Caifeng shan et al. in 2009 examined different machine learning methods, including template matching, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and the linear programming technique, to perform facial expression recognition using LBP features [32]. Their study demonstrated that, compared to Gabor wavelets, LBP features can be derived very fast in a single scan through the raw image and lie in low-dimensional feature space, while still retaining discriminative facial information in a compact representation.

There are two common approaches to extract facial features: geometric feature-based methods and appearance-based methods [22]. Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry. However, the geometric feature-based methods usually require accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations [32].

2.1 GEOMETRIC DESCRIPTOR

Geometric facial features describe the shapes and locations of facial components such as the eyebrows, eyes, nose, mouth and chin. Based on these facial components, which for example could be described by facial points, feature vectors that represent the face geometry are computed. The first geometric feature based approach was proposed in 1978 by Suwa et al. They proposed a system that attempts to automatically recognize facial expressions by analyzing the positions of 20 characteristic facial points. The first step in our fully automated facial expression recognition system is to find the coordinates of those facial points in each frame of a video [14].

![Geometric points tracked on a face](image)

The locations of these points are found in the first frame using a point detector which locates 49 facial points. In all frames after the first we will find the positions of these 49 points using tracker that uses the coordinates of the points found in the first frame to initialize the tracker.

![Distance between points](image)

After tracking the points, the distance between points will be found.

2.2 APPEARANCE DESCRIPTORS

2.2.1 Local binary patterns on three orthogonal plane (LBP-TOP)

Dynamic or temporal textures are textures with motion. The goal of facial expression recognition is to determine the emotional state of the face, for example, happiness, sadness, surprise, neutral, anger, fear and disgust, regardless of the identity of the face. The operator labels the pixels of an image by thresholding a 3 × 3 neighborhood of each pixel with the center value and considering the results as a binary number and the 256-bin histogram of the LBP labels computed over a region is used as a texture descriptor. The derived binary numbers (called Local Binary Patterns or LBP codes) codify local primitives including different types of curved edges, spots, flat areas, etc.

![Procedure of LBP](image)

We begin by sampling neighboring points in the volume, and then thresholding every point in the neighborhood with the value of the center pixel to get a binary value. Finally, we produce the LBP code by multiplying the thresholded binary values with weights given to the corresponding pixel and we sum up the result. After labeling a image with the LBP operator, a histogram of the labeled image can be defined as

\[ H_i = \sum I(x,y)=i, i=0,\ldots,n-1 \]

where \( n \) is the number of different labels produced by the LBP operator. The LBP-TOP histograms in each block are computed and concatenated into a
single histogram. All features extracted from each block volume are connected to represent the appearance and motion of the facial expression sequence.

In this way, we effectively have a description of the facial expression on three different levels of locality. The labels (bins) in the histogram contain information from three orthogonal planes describing appearance and temporal information at the pixel level. The labels are summed over a small block to produce information on a regional level expressing the characteristics for the appearance and motion in specific locations, and all information from the regional level is concatenated to build a global description of the face and expression motion.

2.2.2 Pyramid of histogram of gradients (PHOG)
We consider the problem of image classification where our main goal is to explore how the spatial distribution of shape can benefit recognition. A descriptor is used which has the advantages of both: it captures the spatial distribution of edges, but is formulated as a vector representation. The objective is to represent an image by its local shape and the spatial layout of the shape. Here local shape is captured by the distribution over edge orientations within a region, and spatial layout by tiling the image into regions at multiple resolutions [28-41]. The descriptor consists of a histogram of orientation gradients over each image sub region at each resolution level a Pyramid of Histograms of Orientation Gradients (PHOG). The distance between two PHOG image descriptors then reflects the extent to which the images contain similar shapes and correspond in their spatial layout. The flexibility of the spatial histogram level weighting means that a spectrum of spatial correspondences between two images can be represented. If only the coarsest level is used, then the descriptor reduces to a global edge or orientation histogram. If only the finest level is used, then the descriptor enforces correspondence for tiles (spatial bins) over the image. This extreme is what is captured by where histograms are computed over local image regions [41].

Each image is divided into a sequence of increasingly finer spatial grids by repeatedly doubling the number of divisions in each axis direction (like a quad tree). The number of points in each grid cell is then recorded. This is a pyramid representation because the number of points in a cell at one level is simply the sum of those contained in the four cells. It is divided into at the next level. The cell counts at each level of resolution are the bin counts for the histogram representing that level. The soft correspondence between the two point sets can then be computed as a weighted sum over the histogram intersections at each level. Similarly, the lack of correspondence between the point sets can be measured as a weighted sum over histogram differences at each level. A HOG vector is computed for each grid cell at each pyramid resolution level. The final PHOG descriptor for the image is a concatenation of all the HOG vectors. In forming the pyramid the grid at level 1 has 2L cells along each dimension. Consequently, level 0 is represented by a K-v vector corresponding to the K bins of the histogram, level 1 by a 4K-vector etc, and the PHOG descriptor of the entire image is a vector with dimensionality K= \( \sum_{l=0}^{L-1} 4^l \)

For example, for levels up to \( L = 1 \) and \( K = 20 \) bins it will be a 100-vector. In the implementation we limit the number of levels to \( L = 3 \) to prevent over fitting. The PHOG is normalized to sum to unity. This normalization ensures that images with more edges, for example those that are texture rich or are larger, are not weighted more strongly than others [28]. PHOG is not the same as a scale space pyramid representation of edges as there is no smoothing between levels of the pyramid, and all edges are computed on the high resolution image. Similarity between a pair of PHOGs is computed using a distance function, with appropriate weightings for each level of the pyramid. The advantages of Phog are its insensitivity to small rotation. PHOG is a compact vector descriptor suitable for use in standard learning algorithms with kernels. PHOG is flexible, since it builds in spatial pyramid matching, and is able to cope with varying degrees of spatial correspondence by design.

2.3 Support Vector Machines (SVM)
SVMs have proven to be extremely efficient classifiers, achieving classification rates unparalleled by any other classifier in domains as diverse as marine biology, face detection and speech recognition. They are sparse, non linear and generalize very well given only a small training set. But probably the most important aspect is the well founded mathematical theory on which the classifier is based. It is inspired by statistical learning theory that performs structural risk minimization on a nested set structure of separating hyper planes. SVMs (Support Vector Machines) are a useful technique for data classification. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value (i.e. the class labels) and several attributes (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes [12].
The procedure for using an SVM package is:

- Transform data to the format of an SVM package
- Conduct simple scaling on the data
- Consider the RBF kernel
- Use cross-validation to find the best parameter C and gamma
- Use the best parameter C and gamma to train the whole training set
- Test

There are two parameters for an RBF kernel: C and gamma. It is not known beforehand which C and gamma are best for a given problem; consequently some kind of model selection (parameter search) must be done. The goal is to identify good (C; gamma) so that the classifier can accurately predict unknown data (i.e. testing data). A common strategy is to separate the data set into two parts, of which one is considered unknown [12]. The prediction accuracy obtained from the “unknown” set more precisely reflects the performance on classifying an independent data set.

An improved version of this procedure is known as cross-validation. In cross-validation, we first divide the training set into v subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining v - 1 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified [41]. The cross-validation procedure can prevent the over fitting problem (choosing the best classifier). A grid-search on C and gamma using cross-validation is done. Various pairs of (C; gamma) values are tried and the one with the best cross-validation accuracy is picked. Since doing a complete grid-search may still be time-consuming, it is recommended to use a coarse grid first. After identifying a “better” region on the grid, a finer grid search on that region can be conducted [29]. The SVM predicts the accuracy of the methods used when an optimum value of c and gamma is used.
Confusion matrix for COHN KANNADE database for seven emotions:

Table No. 1 Confusion matrix for Geometric Points

<table>
<thead>
<tr>
<th>Emotion/Inputs</th>
<th>Angry</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Neutral</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry (9)</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Disgust (21)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Fear (7)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Happy (28)</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Neutral (12)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad (15)</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Surprise (24)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

The results of confusion matrix using Geometric Points show that out of the seven emotions, happy (15/28) and surprise (7/24) is most recognizable. The resultant accuracy is 25/112.
The results of confusion matrix using LBP-TOP show that out of the seven emotions, happy (18/28) and surprise (14/24) is most recognizable. The resultant accuracy is 52/112.

The results of confusion matrix using Max Phog show that out of the seven emotions, happy (24/28) and surprise (12/23) is most recognizable. The resultant accuracy is 59/11.

The results of confusion matrix using Avg Phog show that out of the seven emotions, happy (23/28), sad (12/15) and surprise (16/24) is most recognizable. The resultant accuracy is 58/112.
FeedTUM DATASET

The gnu plot graph for finding values of c and gamma for FeedTUM database:

**Geometric points**

```
log[10](C)  log[10](gamma)
-1  -6
-1  -2
1  -6
1  -2
```

**LBP-TOP**

```
log[10](C)  log[10](gamma)
-1  -6
-1  -2
1  -6
1  -2
```

**Max Phog**

```
log[10](C)  log[10](gamma)
-1  -6
-1  -2
1  -6
1  -2
```

**Avg Phog**

```
log[10](C)  log[10](gamma)
-1  -6
-1  -2
1  -6
1  -2
```
Confusion matrix for FEEDTUM database for seven emotions:

Table No. 1 Confusion matrix for Geometric Points

<table>
<thead>
<tr>
<th>Emotion/inputs</th>
<th>Angry (15)</th>
<th>Disgust (15)</th>
<th>Fear (9)</th>
<th>Happy (15)</th>
<th>Neutral (15)</th>
<th>Sad (9)</th>
<th>Surprise (15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry (15)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Disgust (15)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear (9)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Happy(15)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Neutral (15)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Sad (9)</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Surprise (15)</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

The results of Confusion matrix using Geometric Points show that out of the seven emotions, neutral (9/15) and surprise (4/15) is most recognizable. The resultant accuracy is 22/112.

Table No. 2 Confusion matrix for LBP-TOP

<table>
<thead>
<tr>
<th>Emotion/inputs</th>
<th>Angry (15)</th>
<th>Disgust (16)</th>
<th>Fear (9)</th>
<th>Happy (15)</th>
<th>Neutral (15)</th>
<th>Sad (9)</th>
<th>Surprise (15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry (15)</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disgust (16)</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear (9)</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Happy (15)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Neutral (15)</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad (9)</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Surprise (15)</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

The results of Confusion matrix using LBP-TOP show that out of the seven emotions, angry (6/15), neutral (5/15) and surprise (6/15) is most recognizable. The resultant accuracy is 29/112.

Table No. 3 Confusion matrix for max PHOG

<table>
<thead>
<tr>
<th>Emotion/inputs</th>
<th>Angry (10)</th>
<th>Disgust (16)</th>
<th>Fear (9)</th>
<th>Happy (15)</th>
<th>Neutral (15)</th>
<th>Sad (9)</th>
<th>Surprise (15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry (10)</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Disgust (16)</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Fear (9)</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Happy (15)</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Neutral (15)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad (9)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Surprise (15)</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

The results of confusion matrix using max PHOG out of the seven emotions, angry (8/10), happy 8/15), neutral (15/15) and is most recognizable. The resultant accuracy is 46/112.
A support vector machine (SVM) classifier was selected since it is well founded in statistical learning theory and has been successfully applied to various object detection tasks in computer vision. Since SVM is only used for separating two sets of points, the seven expression classification problem is decomposed into 15 two class problems (happiness-surprise, anger-fear, sadness-disgust etc.) then a voting scheme is used to accomplish recognition. For evaluation, we separated the subjects in ratio of 70:30. The SVM classifier was trained on 70 subjects and tested on 30. The same subjects did not always appear in both training and testing. The testing was therefore done with “novel faces” and was person independent. The best recognition performance is found to be 45.45% for Cohn Kanade database using LBP-TOP which is quite more than that for geometric points (23.14%). This proves that appearance-based feature provide better performance than geometrical approaches in Action Unit recognition. Further, it is observed that, Disgust, Joy, Surprise and Neutral can be recognized with high accuracy while the recognition rates for Fear and Sadness are much lower.

### CONCLUSIONS
In this work empirically evaluated Geometric points, Local binary patterns on three orthogonal plane and Pyramid of histogram of gradients to describe appearance changes of expression images. Different techniques are examined on both spontaneous (FeedTUM) and posed (Cohn kanade) databases. Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. Experiments on two DT databases with a comparison to the state-of-the-art results showed that our method is effective for DT recognition. Our approach is computationally simple and robust in terms of grayscale and rotation variations, making it very promising for real time application problems.

Extensive experiments illustrate that LBP-TOP features are effective and efficient for facial expression recognition. One limitation of the existing facial expression recognition methods is that they attempt to recognize facial expressions from data collected in a highly controlled environment given high resolution frontal faces [26]. However, in real-world applications such as smart meeting and visual surveillance, the input face images are often at low resolutions. Obviously low-resolution images in real world environments make real-life expression recognition much more difficult. In this work, we investigated LBP-TOP features for low-resolution facial expression recognition. Experiments on different image resolutions show that LBP-TOP features perform stably and robustly over a useful range of low resolutions of face images. The encouraging performance on real-world compressed video sequences illustrated their promising applications in real-world environments.

### REFERENCES

---

**Table No. 4 Confusion matrix for Avg PHOG**

<table>
<thead>
<tr>
<th>Emotion/inputs</th>
<th>Angry</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Neutral</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry (15)</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Disgust (15)</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Fear (9)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Happy (15)</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Neutral (10)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Sad (9)</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Surprise (15)</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

The results of confusion matrix using Avg PHOG show that out of the seven emotions, angry (7/15), happy (7/15) and surprise (6/15) is most recognizable. The resultant accuracy is 35/112.

The accuracy achieved by the SVM classifier for the two classifiers is given in the table:

<table>
<thead>
<tr>
<th></th>
<th>Geometric points</th>
<th>LBP-TOP</th>
<th>Max Phog</th>
<th>Avg Phog</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEDTUM</td>
<td>23.49%</td>
<td>33.31%</td>
<td>35.40%</td>
<td>34.10%</td>
</tr>
<tr>
<td>COHN KANADE</td>
<td>23.15%</td>
<td>45.45%</td>
<td>43.85%</td>
<td>42.14%</td>
</tr>
</tbody>
</table>


