Facial Expression Detection for video sequences using local feature extraction algorithms

Kennedy Chengeta and Professor Serestina Viriri

University of KwaZulu Natal
School of Computer Science and Mathematics,
Westville Campus, Durban, South Africa
216073421@ukzn.ac.za

Abstract. Facial expression image analysis can either be in the form of static image analysis or dynamic temporal 3D image or video analysis. The former involves static images taken on an individual at a specific point in time and is in 2-dimensional format. The latter involves dynamic textures extraction of video sequences extended in a temporal domain. Dynamic texture analysis involves short term facial expression movements in 3D in a temporal or spatial domain. Two feature extraction algorithms are used in 3D facial expression analysis namely holistic and local algorithms. Holistic algorithms analyze the whole face whilst the local algorithms analyze a facial image in small components namely nose, mouth, cheek and forehead. The paper uses a popular local feature extraction algorithm called LBP-TOP, dynamic image features based on video sequences in a temporal domain. Volume Local Binary Patterns combine texture, motion and appearance. VLBP and LBP-TOP outperformed other approaches by including local facial feature extraction algorithms which are resistant to gray-scale modifications and computation. It is also crucial to note that these emotions being natural reactions, recognition of feature selection and edge detection from the video sequences can increase accuracy and reduce the error rate. This can be achieved by removing unimportant information from the facial images. The results showed better percentage recognition rate by using local facial extraction algorithms like local binary patterns and local directional patterns than holistic algorithms like GLCM and Linear Discriminant Analysis. The study proposes local binary pattern variant LBP-TOP, local directional patterns and support vector machines aided by genetic algorithms for feature selection. The study was based on Facial Expressions and Emotions (FEED) and CK+ image sources.

Keywords: Local binary patterns on TOP · Volume Local Binary Patterns(VLBP)

DOI : 10.5121/sipij.2019.10103

1 Introduction

Video based facial expression analysis has received prominent roles of late, in crowd analysis, security and border control among others[13, 23]. Its also used in image retrieval, clinical research centers and social communication. Facial expressions remain the most effective way of emotion display[14, 15, 11]. Previous work on 2D facial expression recognition focused more into single image frame based analysis than video or image sequence analysis. The former assumes one image is representing the facial expression where as for the video image sequence, each facial image is a temporal dynamic process[14, 15, 11]. Facial action units change in dynamic appearance and dynamic texture and are more difficult to track than static images. A video is classified as a spatial texture mixture in a temporal environment and dynamic features are retrieved[1, 15, 20].

Facial expression algorithms are either holistic or localized feature extraction algorithms. Holistic algorithms cover the whole facial image whereas localized algorithms break down a facial image into smaller units. The study focused on facial motions and locating key facial components namely, nose, eyes, face and mouth. The dynamic features were extracted using key algorithms like local GLBP from three orthogonal planes (LGBP-TOP) which is a LBP variant with Gabor filtering[14, 15, 11]. The LBP-TOP feature descriptor was used for retrieving video dynamic textures for facial expression changes [14, 15, 11]. The facial expression video image sequences were then modeled as a histogram sequence which is a sum of concatenated local facial regions[1]. The experiments used the extended Cohn-Kanade (CK+) as well as the FEED databases. The facial expression sequence was modeled into a histogram by adding local regions histograms for the LGBP-TOP maps. KNN and Support Vector Machines algorithms were used to classify the datasets. Feature selection was enhanced by genetic algorithms. The extended Cohn-Kanade database (CK+) experiments demonstrated that local feature extraction, support vector machines and genetic algorithms achieved the better results compared to holistic feature extraction methods in last couple of years[22, 24].

Video-based facial expression recognition is made of face detection, tracking and recognition[16, 10, 6, 17]. The video sequences picked up depict key universal expressions (surprise, sadness, joy, disgust and anger). Each signal expression is performed by 7 different subjects beginning from the neutral expression. The paper mainly focused on the integration of spatial-temporal motion LBP with Gabor multi-orientation fusion and compared the 3 LBP histograms on three orthogonal planes to find accuracy of facial expression recognition[1, 14]. A support vector machine classifier was used for each plane and the overall LBP-TOP algorithm. The LBP-TOP was also compared against its variants like LBP-MOP [1, 14]. Experiments conducted on the extended Cohn-Kanade (CK+) database and FEED database proved that the 3D LBP TOP algorithm compared better than the 2D image texture local binary pattern variants.

2 Literature Review

In recent researches use of spatio-temporal representations has been successfully used to address limitations of static image analysis[8, 3–5, 2]. Successful research has been done by fusing PCA, Gabor Wavelets, local binary patterns and local directional patterns[23]. Classification has included support vector machines(SVM), AdaBoost, k-nearest neighbor and neural networks[7, 8, 14]. The permanent features like eyes, mouth or lips' feature vectors were generated from the facial appearances in the spatial and frequency domains.

2.1 Static Facial Expression Analysis Background

Clinical facial expression research has been widely studied with 2D images either as a combination of facial expressions in 2D images or universal global facial [7, 8, 3, 5]. The Facial Action Coding System (FACS) describes facial expressions based on a mix of action units (AU)[7, 8, 3, 5]. Each facial action is a representation of muscular facial actions resulting in sudden facial expression variations. Global facial expressions represent the wholesome facial changes. The key expression changes are being happy, sad, anger as well as fear. The FACS solution improved its 2D image success to 3D to analyze video image changes based on time, content quality and valence [7, 8, 3, 5].

Challenges with static 2D images Based Methods Major clinical research in facial expression analysis includes subjective and qualitative scenarios in the 2D image family[8, 3]. The 2D static images lack temporary dynamics[14, 4, 13, 15, 16, 18]. They are also prone to subjective judgements and poor qualitative features. The 2D static images do not capture temporary dynamics and expression changes [14, 4, 13, 15, 16, 18]. Over recent times there has been need to capture and measure quantitatively facial expressions in video and motion scenarios. Various frameworks have used dynamic analysis to deduce emotions giving accurate results.[14, 4, 13, 15, 16, 18]. Successful studies have chosen frameworks that included video face detection and tracking incorporating shape variability where features are extracted and classifiers applied on the given histograms[18, 1].

2.2 Facial Recognition with video image sequences

Facial expressions are subdivided into 3 segment groups namely the beginning, apex and end[18, 14, 15, 2, 1]. Facial expressions are also denoted as magnitudes of facial motions or motion units moments[18, 1]. Hidden Markov models and naive bayes classifiers have also been successful in video sequence facial expression recognition[14, 16]. Successful video sequence in facial expressions applied a two stage approach to classify images in 3D by measuring the video intensity using optical flows [1, 16]. Several probabilistic methods like particle filtering and condensation can also track facial expression in video sequences [18, 1]. Separate manifold substances have also been applied in video based facial expression analysis. To track video sequences models like 3d wireframe models, facial mesh models, net models and ASM models were successfully used[14, 20, 1, 10, 12]. Videos subtle changes of facial expression can be measured on video facial expression recognition than on static image analysis[14, 16, 1, 10, 12].

3 Local Based Facial Expression Feature Extraction

Facial expression analysis influences wide areas in human computer interaction. Local binary patterns and their 2D and 3D variants have been used in this field[14, 16, 1, 10, 12]. Holistic and local based feature extractors have also been used successfully. PCA feature extractors are prominent holistic algorithms and local binary patterns, Gabor filters and Gabor wavelets and local directional patterns have been successfully applied as local feature extractors[14, 16, 1, 10, 12].

3.1 Local Binary Patterns (LBP) for static image feature extraction

Local binary patterns are based on facial images being split into local sub regions. The challenges of facial occlusion and rigidness are plenty though grey scale image conversion is used to reduce illumination[8, 7, 3]. Local binary patterns are invariant to grey level images. Localized feature vectors derived are then used to form the histogram which is used by machine learning classifiers or deep learning methods. The local features are position dependent [8, 7, 3]. For local binary patterns, the facial region is divided into small blockers like mouth, eyes, ears, nose and forehead[4]. The aggregate histograms are then grouped to form one feature vector or histogram for the facial image. The popular local binary pattern variants include uniform local binary patterns, central symmetric local binary patterns, elongated local binary patterns, multi block local binary patterns, ternary local binary patterns and rotational local binary patterns[8, 7, 3, 12, 23, 20, 1]. The key parameters include the radius of the local binary pattern and the given number of neighbors[8, 9, 3, 12, 23, 20, 1].

The basic local binary pattern non center pixels use the central pixel as the threshold value taking binary values [8, 7, 3]. Uniform binary patterns are characterized by a uniformity measure corresponding to the bitwise transition changes. The local binary pattern has 256 texture patterns. The local binary LBP r,n operator is represented mathematically in the following equation where the radius is given as r and the number of neighborhoods as n.

$$LBP_{(n,r)} = \sum_{n-1}^{n=0} s(p_n - p_c)2^n. \tag{1}$$

The neighborhood is depicted as an m-bit binary string leading to n unique values for the local binary pattern code. The grey level is represented by 2^n-bin distinct codes. The gray scale of the middle pixel is given as p_c whilst p_n represents the neighboring pixels.

LBP Variants Various LBP variants were successfully proposed and used. These include TLBP for Ternary Local Binary Pattern as well as Central Symmetric Local Binary Patterns [8, 9, 4]. Over-Complete Local Binary Patterns (OCLBP) is another key variant that takes into account overlapping into adjacent image blocks. The rotation invariant LBP is designed to remove the effect of rotation by shifting the binary structure[8, 4]. Other variants include the monogenic and central symmetric (MCS-LBP).

3.2 Volume Local Directional Binary Patterns (VLDBP)

For video sequencing facial image analysis, Volume Local Directional binary patterns (VLDBP) and Local Gabor Binary Patterns from Three Orthogonal Planes have been successful compared to other [1, 14, 15]..

Volume local directional binary patterns (VLDBP) are used as an extension of LBP in the dynamic texture field. Dynamic texture extends the temporal domain and is used in video image analysis. The face regions of the video sequence images are modeled with VLDBP which incorporates movement and appearance [1, 14, 15]. It uses 3 planes where the central plan includes the central pixel used to the LBP algorithm. VLBP considers co-occurrences of surrounding points from three planes and generates binary representatives [1, 14, 15].. The extraction considers local volumetric neighborhoods against the pixels. The center pixel grey values and the surrounding pixels are then compared against each other.

$$VLBP_{LPR} = \sum_{q=0}^{3P+1} v_q 2^q \tag{2}$$

3.3 Local Gabor Binary Patterns based on Three Orthogonal Planes

3D dynamic texture recognition which concatenates three histograms from LBP on three orthogonal planes has been widely used (LBP-TOP) [1, 14, 17]. LBP-TOP extracts features from the local neighborhoods over the 3 planes. The spatial-temporal information in the 3 dimensional (X, Y, T) space is given where X,Y are spatial coordinates and the T axis represents temporal time[20]. [1, 14, 15]. LBP-TOP derives local binary patterns from a central pixel by threshholding the neighboring pixels [1, 14, 17][20]. The algorithm decomposes the 3 dimensional volume into 3 orthogonal planes[1, 2]. The XY plane indicates appearances features in the spatial domain and XT, visual against time whilst the YT plane is for motion in the temporal space domain[20].

The spatial plane, XY is similar to the regular LBP in static image analysis. The vertical spatio-temporal YT plane and horizontal XT plane are the other 2 planes in the 3 dimensional space[20]. The resulting descriptor enables encoding of spatio-temporal information in video images. The performance and accuracy of the latter was also comparable to the LBP-TOP. The LBP STCLQP or spatio-temporal completed local quantized patterns (STCLQP) was also used to consider the pixel sign, orientation and size or magnitude[20]. Local gabor binary patterns from Three Orthogonal Planes (LGBP-TOP) add gabor filtering to improve accuracy. With the added filtering algorithm rotational misalignment of consecutive facial images is mitigated[1, 16, 10]. To avoid LBP-TOP statistical instability, a re-parameterization algorithm based on another local Gaussian jet was suggested [1, 14, 17].

$$k = (H_L, X_Y, H_L, XT, H_L, Y_T, H_C, X_Y, HC, XT, HC, YT) \tag{3}$$

(LBP/C)TOP feature is denoted in vector form where Hv, m (v= LBP/C, and m = XY, XT, YT where m represents 6 LBP sub-histograms with contrast features in 3 planes [1, 14, 17]. The LBP-TOP algorithm describes video sequence changes in both spatial and temporal domains hence captures structural information of the former domain and longitudinal data of the latter [16, 18, 1]. LBP histogram features encode spatial data in the XY plane and the histograms from XT, YT planes include the temporal and spatial data[20]. With facial actions causing local and expression changes over time, the dynamic descriptors have an edge in facial expression analysis over the static descriptors [16, 18, 1].

$$H_q, x = \sum q, v, tf_j(q, v, t) = q \tag{4}$$

Contrasts in 3 orthogonal planes are denoted as Cm represented as (m= XY, XT and YT) [1] and defined as 3 sub-histograms Hx ,y (x= C and y= XY, XT , YT) [16, 18, 1]. Image device quality of the facial expression videos also impacts frame rates and spatial resolution quality as well [16, 18, 1].

Six Intersection Points (SIP) The LBP-SIP or Local Binary Pattern— Six Interception Points (LBP-SIP) considered 6 unique points along the intersecting lines of the 3 orthogonal planes to derive the binary pattern histograms [16, 18, 1].

$$AB, DF, EG = L_A \cap L_B \cap L_C \tag{5}$$

where AB, DF and EG are intersection points. Six neighbor points carry enough data to describe spatio-temporal textures around point C[16, 18, 1]. LBP-SIP gives a concentrated group of high dimensional features spaces where there is sparse data [16, 18, 1].

LBP-Three Mean Orthogonal Planes (MOP) LBP-MOP or mean orthogonal plane is another variant to have been successfully used by concatenating mean images from image stacks derived along the 3 orthogonal planes[16, 18, 1]. It also preserves essential image patterns and reduces redundancy which affects encoded features.

4 Feature Selection and Edge Detection

Feature selection and images are enhanced for classification through edge detection and selection of the fittest images by genetic algorithms. Prominent detection algorithms include, Sobel, Canny, Kirsch edge detector which forms the base for local directional patterns and Hewitt edge detector. The study uses the Kirsch edge detector.

4.1 Local directional patterns

LDP includes compass mask which allows for information extraction based on prominent edges or directions. The focus is on facial image edges on prominent facial regions[9, 3]. Convolution is applied based on the base images to get the edge detected images[9, 3]. For local directional patterns or LDP a key edge detection local feature extractor, the images were divided into LDPx histograms, retrieved and then combined into one descriptor[9, 3, 5, 1, 11]. The local directionary pattern, includes edge detection using the Kirsch algorithm. The operator takes one kernel which is rotated in 8 directions at forty five degrees based on one kernel marsk[9, 3]. The Kirsch operator's edge size is calculated as the maximum size for all directions and this is shown in the local directional pattern Kirsch convolutionary equation with the associated example M_0:

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

Fig. 1. Local Directional Patterns (LDP)

$$LDP_x(\sigma) = \sum_{K}^{r=0} \sum_{L}^{r=0} f(LDP_q(o, u), \sigma). \tag{6}$$

$$M_0 = (85x - 3) + (32x - 3) + (26x5) + (10x5) + (45x5) + (38x - 3) + (60x - 3) + (53x - 3) = -399 \tag{7}$$

4.2 Genetic Algorithm

Genetic algorithms were introduced in the 1970s as a class of evolutionary algorithms[22, 24]. These are heuristic approaches that find solutions based on evolutionary biology concepts. Genetic algorithms select a subset of features by removing unimportant features[22]. In this study the GA algorithm is also used to select best SVM kernel function. The algorithm uses techniques like mutation, crossover, and selection to regenerate the population. The algorithm starts with a randomly generated set of facial expression images and this evolves as a new generation where the fittest images are selected for the next iteration[22]. The fitness function optimizes the objective function. The population selected once its satisfies the fitness, undergoes mutation or crossover for them to be selected for next iteration[24]. The convergence is reached when certain given generations has been achieved or when a given fitness value has been achieved. In facial expression recognition the algorithm is used to also optimize computational cost and video temporal correlation[22, 24].

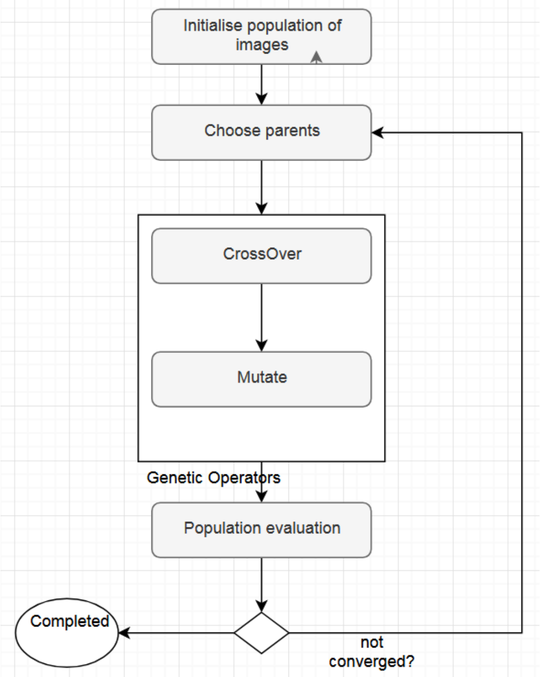

Fig. 2.

1. Initialization of the parent population from the image database and choosing population size.
2. Evaluating the population
3. Selecting fit facial images and use fitness function to find the criterion. [24]
4. Crossover of facial images. The individuals which have frame numbers nearer to each other have higher probability of crossover[24].
5. The mutations are done and the fittest individual is computed. In this computation of fittest, the older generation does not participate. Though it does die outs only after reproducing 10 fit individuals[24].

6. This process continues until there is at least 10 fix facial images within 100 frames.

Genetic algorithms has proven to be a successful algorithm to derive optimal support vector machine kernels[24]. The kernel functions are crucial to support vector machines as they affect classification. A Genetic classification algorithm approach was widely suggested to choose the best kernel and its parameters[24].

5 Facial Expression Implementation Approach

The implementation involves analyzing video streams to track facial feature points over time. The feature vectors are then calculated and emotions detected from the trained models. Training and classification of the models is done using the popular algorithms namely support vector machines and genetic algorithm to remove unwanted facial images and reduce the error rate. Recognition of the model and new expressions on new images is then done on the selected annotated databases which includes CK+ database and FEED database. The section describes the approach, databases selected and then classification algorithm chosen and implemented.

The study's objective was to recognize facial expressions from video sequences. The approach involved locating and tracking the faces and expressions during the video segmentation and sequential modeling phase. The video sequence detection involved landmark detection and tracking, which define the facial shapes[15]. Viola Jones OpenCV detection tools are used. The features were then extracted using various 3D video feature extraction variants of the LBP-TOP algorithm. Gabor filters [16, 18, 1, 12, 5] were then applied during preprocessing. Geometric features were normalized and they were immune from skin color and illumination changes. [16, 18, 1].

```
Data:  Copy and preprocess video image datasets
Result: Facial expression classification results for the image datasets
while For each image I inside the CK+ and FEED database do
    1. divide the database into training and test sets;
    2. for each image inside the given datasets;
    3. apply Viola Jones algorithm for extraction and preprocess the image
      using Principal Component Analysis;
    3b Genetic algorithm and PCA algorithm are applied to remove unwanted
      images which have no bearing to the facial expression classification.
    4. The study then finds features using LBP-TOP, LBP-XY,LBP-XT and
      LBP-YT algorithms;
    5. extract the features using the LBP-MOP, LBP-SIP algorithm
    5b Local directional patterns are also applied to remove unwanted edges
      from the images 6. apply Gabor Filters to get the LGBP-TOP and
      LGBP-MOP features
    7. calculate the Euclidian distance matrix;
    8. apply the classification on each with support vector machine and
      genetic algorithm;Use genetic algorithms to select the optimal support
      vector machine kernel.
    9. the best classification results is then labelled the best algorithm;
    End For'
end
```

Algorithm 1: Local Gabor Binary Patterns from Three Orthogonal Planes to analyze video sequences[21, 1]

The machine learning classifiers namely support vector machines and genetic algorithm were used for the classification and feature selection. The algorithm used is shown in Algorithm 1. The study analyzed a sequence of frames that change from one form to another to detect faces from a live video based on the CK+ dataset and the FEED dataset.

5.1 Facial Expression Preprocessing

Preprocessing is achieved using the Principal Component Analysis algorithm. With this the number of parameters is streamlined to include only relevant and important parameters only[21, 1]. PCA is used to eliminate other less important feature vectors dimensions. Gabor filters (a linear filter) were then used to detect edges in texture analysis. In the spatial domain, a given gabor filter acts like a Gaussian kernel function as shown in the equation below [21, 16, 18, 1] . The Gabor filters extract expression-invariant features.

$$G\_c[t,w] = Te^{-\frac{(t^2+w^2)}{2\sigma^2}} \cos(2\pi f(t\cos\theta + w\sin\theta)); \tag{8}$$

$$G\_s[t,w] = We^{-\frac{(t^2+w^2)}{2\sigma^2}} \sin(2\pi f(t\cos\theta + w\sin\theta)); \tag{9}$$

where T and W are normalizing factors that will be derived[21].

5.2 Facial Expression Databases

The study used video sequences lasting around 10 seconds with 15 second frames per second. The facial expressions are dynamic and evolve over time from the start, when reaching the apex and offsets. The video sequence datasets that could have been used included the CK+, YouTube Faces Database, Acted Facial Expressions(AFEW) in the Wild as well as the BU-3DFE, MMI+ dataset and the Facial Expressions and Emotions Database (FEED) [16, 18, 1]. The FEED dataset included 400 webcam video extracts from 18 voluntary participants in mpg format of sizes 480 times 640. There were labeled as 6 facial expression classes [16, 18, 1]. YouTube Faces Database data set contains over 3 000 videos from about a thousand and five hundred video sequence images. The videos average around 2 to 3 seconds and clips frame sizes from 48 to 6000 frames with a mean of 180. The BU-4DFE database has 101 subjects for identifying the emotion and it also has 83 feature points to recognize the emotion. In the total 101 subjects, 58 subjects are female and remaining 43 subjects are male. The study chose the FEED and CK+ dataset for implementation [16, 18, 1]. For static image analysis the study used the CK+ dataset and Google set dataset. The static image analysis was then compared to the video sequence databases.

CK+ dataset The CK+ dataset includes 593 video sequences and 7 expression types from 123 participants. The participants included African-Americans and Euro-Americans and other races accounted for 6 percent [16, 18, 1]. The video sequences were 640 by 490 by 640 by 480 pixels. The grey images made with 8-bit precision made up the frames dataset[13]. The study used 90 participants and considered the 6 expressions namely anger, disgust, fear, happiness, sadness, and surprise[9].

5.3 Facial Expression Video Sequences Classification

The study used genetic algorithms and support vector machines[19, 8, 5] for classification and Kirsch based local directional pattern gave the edge detection advantage to remove unwanted features and reduce the error rate.

Support Vector Machine Support vector machines consider points close the given class boundaries[10]. A hyperplane is chosen to separate 2 classes which are initially given as linearly separable. The hyperplane separating the two classes is represented by the given equation[19, 9, 12, 15]:

$$w^T x_n + b = 0, \tag{10}$$

such that:

$$w^T, x_n + b1 \qquad y_n = +1, \tag{11}$$

The genetic algorithm is also used to select the optimal kernel type to be used in the algorithm from one the following options namely ‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’ and lastly callable[19, 9, 12, 15]. The kernel parameter was designed to take in linear, poly or rbf parameters. Alternatively, the gamma parameter was also tuned and the cost parameter of the support vector machine needed an optimal value[19, 9, 12, 15].

```
model = svm.svc(kernel=’linear’, c=2, gamma=3)
```

6 Experiment and Results

6.1 Static image analysis experimental results

For the 2D experiments, the CK+ dataset was tested against support vector machines aided by genetic algorithm for feature selection and local binary and directional patterns for feature selection. Whilst the 2D classification results showed greater accuracy they lacked the 3D and dynamic spatial properties. The best classification was found on a combined LBP+ELBP and Gabor Filters combination with a 16, 2 radius combination that resulted in a classification rate of 98.76 percent for the support vector machine plus genetic algorithm classifier.

6.2 Experimental Results on CK+ and FEED 3D Datasets

Three experiment types were executed on the CK+ dataset’s facial motions. Recognition rates were executed for the LBP-XY plane, LBP-XT plane as well as the LBP-YT planes. The combined recognition rate for the LBP-TOP was also calculated. The static tests used a support vector machine and genetic algorithm classifier. The temporal dynamic classification of the CK+ and FEED datasets was based on the XY, YT and XT plane dimensions with an average length of 0.9 seconds. The minimum length was 0.78 seconds and the highest length was 0.934 seconds. The CK+ dataset the combined LGBP-TOP with Gabor Filtering and support vector machine and genetic algorithm classifier achieved an accuracy of 98.03 percent from a sequence of 593 video sequences. This was the modal accuracy based on a run of twenty experiments. For the FEED database with 400 video image sequences, the corresponding modal accuracy was 98.44 percent from a run of around twenty experiments.

The combined feature extractor LBP-TOP achieved higher classification rates as compared to the specific dimension LBP-XT, LBP-XT and LBP-YT accuracy rates. Better variation was experienced for the LBP-XT based plane. The second experiments evaluated the efficiency of using Gabor Filters to enable multi-orientation fusion to the spatial temporal advantages of the LBP-TOP algorithm. Support vector machines and genetic algorithm classifiers were also used in this scenario The classifier with Gabor-Filters and LBP-TOP feature extractor showed greater accuracy to the normal LBP-TOP algorithm. The other LBP-TOP variants like SIP and MOP also achieved greater accuracy but the LGBP-TOP with parameters of 8,3 on each dimension achieved better accuracy to all the LGBP-TOP variants.

6.3 Video Sequence Confusion Matrices

The confusion matrix obtained from the video datasets showed an overall success of 98.03 percent and 98.44% when Gabor Filtered on the CK+ and FEED database respectively. The following confusion matrix gives detail of the precision recall accuracy for the CK+ dataset which included 593 video datasets with video lengths of less than 1 second. For classification, support vector machine algorithm with kernel based selection using genetic algorithm is used.

Fig. 3. LBP-TOP , CK+ Dataset Facial Expression Recognition dataset from 593 video sequences

	precision	**recall**	**f1-score**	**support**
anger	0.969	1	0.976	122
disgust	0.962	0.964	0.970	61
fear	0.967	0.973	0.974	132
happy	0.977	0.971	0.974	117
neutral	1	0.979	0.986	95
sadness	0.962	0.970	0.979	66
avg/total	0.981	0.981	0.9803	**593**

	Confusion Matrix						
anger	[[**115,**	0,	0,	3,	3,	1],	
disgust	[	0,	**58,**	0,	1,	1,	1],
fear	[	1,	0,	**127,**	1,	2,	1],
happy	[	2,	3,	1,	**111,**	0,	0],
neutral	[	1,	0,	1,	0,	**93,**	0],
sadness	[	0,	0,	0,	2,	2,	**62**]]

The FEED Dataset had 400 video sequence images analyzed over the 5 expression types namely anger, disgust, fear, happy, sadness and neutral. For the FEED dataset, the anger expression type showed modal frequency in the confusion matrix and for the CK+ video datasets, the fear expression type was highest.

Fig. 4. LBP-TOP , FEED Dataset Facial Expression Recognition dataset of 400 web-cam videos image sequences

	precision	**recall**	**f1-score**	**support**
anger	0.974	0.969	0.966	99
disgust	0.984	0.985	0.976	80
fear	0.983	1	0.985	46
happy	0.978	0.995	0.977	74
neutral	0.977	0.997	1	62
sadness	0.965	0.979	0.964	39
avg/total	1	1	0.984	400

	Confusion Matrix						
anger	[**90,**	1,	2,	3,	2,	2]	
disgust	[	2,	**71,**	1,	3,	2,	1]
fear	[	1,	2,	**43,**	0,	0,	0]
happy	[	2,	2,	1,	**68,**	1,	0]
neutral	[	1,	0,	1,	0,	**60,**	0]
sadness	[	1,	0,	0,	1,	0,	**37**]

7 Conclusion

The feature extraction methods of LBP-TOP variants applied to major facial components used by the research to analyze facial expressions in video datasets showed marked improvements compared to traditional holistic methods used before. For each facial component angle namely XY, XT and YT-3D dimension, the classification rate was a combination of support vector machine with genetic algorithm to optimize the kernel function and enhanced feature selection. The Gabor filters improved the accuracy and the LBP-TOP variants also showed great accuracy. The study also proved that using genetic algorithms improves feature selection and error rate during image selection as well reducing computation time by twenty percent. The edge detection capability of local directional patterns also allowed the removal of less important information on the facial images that does not influence the facial images. Local feature extraction algorithms, in this scenario, local binary pattern from three orthogonal planes with local directional patterns (for enhanced edge detection) showed better accuracy than holistic algorithms in terms of accuracy and performance.

8 Future Work

Future work in video facial expression recognition and classification includes investigating applicability of analyzing video media like video conferencing, video streamed data, skype and other forms of media. The research also recommends analyzing the expressions of people in a group conversation and determining if their expressions are correlated based on the conversation at hand. The research also recommends analyzing expressions in an African context where there are different cultures with each culture having different ways of expressing themselves. Some of the key cultures suggested include the Zulu culture in South Africa, Swahili culture in Eastern Kenya and Tanzania, as well as Shona culture in Zimbabwe.

References

1. Y.Wang,J.See,R.C.-W.Phan,Y.-H.Oh,Lbp with six intersection points:Reducing redundant information in lbp-top for micro-expression recognition, in: Computer Vision—ACCV 2014, Springer, Singapore, 2014, pp. 525–537.
2. Y. Wang, J. See, R.C.-W. Phan, Y.-H. Oh, Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition, PloS One 10 (5) (2015).
3. M. S. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, et al.: "The auto- matic detection of chronic pain-related expression: requirements, challenges and a multimodal dataset," Transactions on Affective Computing, 2015.
4. P. Pavithra and A. B. Ganesh: "Detection of human facial behavioral ex- pression using image processing,"
5. K. Nurzynska and B. Smolka, "Smiling and neutral facial display recognition with the local binary patterns operator:" Journal of Medical Imaging and Health Informatics, vol. 5, no. 6, pp. 1374–1382, 2015-11-01T00:00:00.
6. Rupali S Chavan et al, International Journal of Computer Science and Mobile Computing Vol.2 Issue. 6, June- 2013, pg. 233-238

7. P. Lemaire, B. Ben Amor, M. Ardabilian, L. Chen, and M. Daoudi, “Fully automatic 3d facial expression recognition using a region-based approach,” in Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, J-HGBU ’11, (New York, NY, USA), pp. 53–58, ACM, 2011.
8. C. Padgett and G. W. Cottrell, “Representing face images for emotion clas- sification,” Advances in neural information processing systems, pp. 894–900, 1997.
9. P. Viola and M. J. Jones: “Robust real-time face detection,” Int. J. Comput. Vision, vol. 57, pp. 137–154, May 2004.
10. Yandan Wang , John See, Raphael C.-W. Phan, Yee-Hui Oh, Spatio-Temporal Local Binary Patterns for Spontaneous Facial Micro-Expression Recognition, May 19, 2015, https://doi.org/10.1371/journal.pone.0124674
11. A. Sanin, C. Sanderson, M. T. Harandi, and B. C. Lovell, “Spatio-temporal covariance descriptors for action and gesture recognition,” in Proc. IEEE Workshop on Applications of Computer Vision (Clearwater, 2013), pp. 103–110.
12. K. Chengeta and S. Viriri, ”A survey on facial recognition based on local directional and local binary patterns,” 2018 Conference on Information Communications Technology and Society (ICTAS), Durban, 2018, pp. 1-6.
13. S. Jain, C. Hu, and J. K. Aggarwal, “Facial expression recognition with temporal modeling of shapes,” in Proc. IEEE Int. Computer Vision Workshops (ICCV Workshops) (Barcelona, 2011), pp. 1642–1649.
14. X. Huang, G. Zhao, M. Pietikainen, and W. Zheng, “Dynamic facial expression recognition using boosted component-based spatiotemporal features and multiclassifier fusion,” in Advanced Concepts for Intelligent Vision Systems (Springer, 2010), pp. 312–322.
15. R. Mattivi and L. Shao, “Human action recognition using LBP-TOP as sparse spatio-temporal feature descriptor,” in Computer Analysis of Images and Patterns (Springer, 2009), pp. 740–747.
16. A. S. Spizhevoy, Robust dynamic facial expressions recognition using Lbp-Top descriptors and Bag-of-Words classification model
17. B. Jiang, M. Valstar, B. Martinez, M. Pantic, ”A dynamic appearance descriptor approach to facial actions temporal modelling”, IEEE Transaction on Cybernetics, vol. 44, no. 2, pp. 161-174, 2014.
18. Y. Wang, Hui Yu, B. Stevens and Honghai Liu, ”Dynamic facial expression recognition using local patch and LBP-TOP,” 2015 8th International Conference on Human System Interaction (HSI), Warsaw, 2015, pp. 362-367. doi: 10.1109/HSI.2015.7170694
19. Aggarwal, Charu C., Data Mining Concepts, ISBN 978-3-319-14141-1, 2015, XXIX, 734 p. 180 illus., 173 illus. in color.
20. Pietikäinen M, Hadid A, Zhao G, Ahonen T (2011) Computer vision using local binary patterns. Springer, New York. https://doi.org/10.1007/978-0-85729-748-8
21. Ravi Kumar Y B and C. N. Ravi Kumar, ”Local binary pattern: An improved LBP to extract nonuniform LBP patterns with Gabor filter to increase the rate of face similarity,” 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP), Mysore, 2016, pp. 1-5.
22. Arana-Daniel N, Gallegos AA, López-Franco C, Alanís AY, Morales J, López-Franco A. Support Vector Machines Trained with Evolutionary Algorithms Employing Kernel Adatron for Large Scale Classification of Protein Structures. Evol Bioinform Online. 2016;12:285-302. Published 2016 Dec 4. doi:10.4137/EBO.S40912
23. K. Chengeta and S. Viriri, ”A survey on facial recognition based on local directional and local binary patterns,” 2018 Conference on Information Communica-

tions Technology and Society (ICTAS), Durban, 2018, pp. 1-6. doi: 10.1109/ICTAS.2018.8368757

24. İlhan İlhan, Gülay Tezel,
A genetic algorithm–support vector machine method with parameter optimization for selecting the tag SNPs,
Journal of Biomedical Informatics,
Volume 46, Issue 2,
2013,
Pages 328-340,
ISSN 1532-0464,
https://doi.org/10.1016/j.jbi.2012.12.002.
(http://www.sciencedirect.com/science/article/pii/S1532046412001852)