African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 4, Issue 3, 2025
Page No: 552-574
Website: https://aaasjournals.com/index.php/ajapas/index
ISI 2025: 1.126 **SJIFactor 2024: 6.752** **معامل التأثير العربي: 1.62**

Novel Geometric Key Frame & Feature Extraction for Four-dimensional Dynamic Facial Expression Recognition

Ali Ali Milad [1*], Elhadi Elfitory Algarai [2]
[1] Computer Science Department, Faculty of Science, Elmergib University, Alkhoms, Libya
[2] Software Engineering Department, Faculty of Information Technology, Elmergib University, Alkhoms, Libya

إطار رئيسي هندسي جديد واستخراج الميزات للتعرف على تعبيرات الوجه الديناميكية رباعية

علي علي ميلاد [1*] ، الهادي الفيتوري الجراي [2]
[1] قسم علوم الحاسوب، كلية العلوم، جامعة المرقب، الخمس، ليبيا
[2] قسم هندسة البرمجيات، كلية تقنية المعلومات، جامعة المرقب، الخمس، ليبيا

[*]Corresponding author: a.amilad@elmergib.edu.ly

Received: July 25, 2025	Accepted: September 08, 2025	Published: September 20, 2025

Abstract
This paper presents facial expression synthesis traditionally used in Face recognition systems which leads to good performance under the variations that occurred in poses or expressions face. There are several challenges faced by the 4D FER, and they are not limited to imbalanced, high computational complexity, occlusion, large data requirements, especially for deep learning frameworks. Facial expression synthesis is one of the popular researches used in varying applications. In this paper, we initiate a model to synthesize a natural facial image from given various facial expressions while maintaining the initial facial features. The ability to produce both synthetic images of subjects in the training set which could be employed to models requiring larger training data is one of the novelties of our approach. The feature engineering techniques proposed in this paper may be adapted for real-time embedded systems due to the strategies implemented to reduce computational complexity, memory, and maintain a relatively high degree of accuracy. Also, geometric PSNR, IMMSE and Entropy information is used to detect apex or key frames, and a novel alpha axes-angular feature extracted from geometric facial landmarks data. This leads to the generation of input feature vector from ZY axes for alpha angles with respect to origin in three-dimensional Euclidean space. The NCA feature selections are applied to obtain optimal feature subset and trained using multiclass ECOC-SVM on MATLAB 2020a. The problem of 3D facial expression recognition is modeled as an imbalanced problem and average recognition accuracy are used as a performance metric. The results showed a highly informative alpha angular feature on the BU4DFE dataset and demonstrate the effectiveness of the proposed landmark-based approach to classifying emotions.

Keywords: 4D dynamic facial expression, Facial synthesis, texture mapping, image processing face, modeling, Axes-angle feature, facial expression synthesis.

الملخص
تقدم هذه الورقة البحثية تركيب تعبيرات الوجه المستخدم تقليديًا في أنظمة التعرف على الوجوه والذي يؤدي إلى أداء جيد في ظل الاختلافات التي حدثت في أوضاع أو تعبيرات الوجه. هناك العديد من التحديات التي تواجه D FER4، وهي لا تقتصر على عدم التوازن والتعقيد الحسابي العالي والانسداد ومتطلبات البيانات الكبيرة، وخاصة لأطر التعلم العميق. يعد تركيب تعبيرات الوجه أحد الأبحاث الشائعة المستخدمة في تطبيقات مختلفة. في هذه الورقة البحثية، نبدأ نموذجًا لتجميع صورة وجه طبيعية من تعبيرات وجه مختلفة معينة مع الحفاظ على ملامح الوجه الأولية. إن القدرة على إنتاج كل من الصور التركيبية للموضوعات في مجموعة التدريب والتي يمكن استخدامها للنماذج التي تتطلب بيانات تدريب أكبر هي إحدى المستجدات في نهجنا. يمكن تكييف تقنيات هندسة الميزات المقترحة في هذه الورقة البحثية للأنظمة المضمنة في الوقت الفعلي بسبب الاستراتيجيات المطبقة لتقليل التعقيد الحسابي والذاكرة والحفاظ على درجة عالية نسبيًا من الدقة. كما تُستخدم معلومات PSNR الهندسية، وIMMSE، والإنتروبيا للكشف عن الإطارات الرئيسية أو الإطارات الرئيسية، بالإضافة إلى خاصية

زاوية ألفا جديدة مُستخرجة من بيانات معالم الوجه الهندسية. يؤدي هذا إلى توليد متجهات ميزات إدخال من محاور ZY لزوايا ألفا بالنسبة إلى الأصل في الفضاء الإقليدي ثلاثي الأبعاد. تُطبق اختيارات خصائص NCA للحصول على مجموعة فرعية مثالية من الميزات، وتُدرَّب باستخدام ECOC-SVM متعدد الفئات على MATLAB 2020a. تُنمذج مشكلة التعرف على تعبيرات الوجه ثلاثية الأبعاد كمشكلة غير متوازنة، ويُستخدم متوسط دقة التعرف كمقياس للأداء. أظهرت النتائج خاصية زاوية ألفا غنية بالمعلومات على مجموعة بيانات BU4DFE، وتُظهر فعالية النهج المُقترح القائم على المعالم في تصنيف المشاعر.

الكلمات المفتاحية: تعبير الوجه الديناميكي رباعي الأبعاد، تركيب الوجه، رسم الملمس، معالجة صور الوجه، النمذجة، ميزة المحاور والزاوية، تركيب تعبير الوجه.

Introduction

The research of facial expressions synthesizing has had its major focus in the area of the primary emotions of humans, there are very few studies that focused on non-primary human emotions such as Chan et al. (2006)[1]. Although the 3D data used in this work unlike 2D data that has very low cost, has its cons which include; greater number of dimensions, which consequently raises storage and processing costs. Hence, the motivation to provide a solution. Also, the 3D data has 3D data inherently exhibits robustness to these variations, being unaffected by illumination and, to some extent, resilient to pose changes [2]. Researchers have also proposed many tools for analyzing and reconstructing motions based on Igarashi's idealized concept of spatial key frames. [3]. Also, derived from the human skeletal structure using 3D articulated joint motion trajectories, [4] offer an innovative and a computationally efficient approach for recognizing actions from 3D data. Geometric-based algorithms determine the location of landmarks such as (eyes - nose) and extract the shape of faces [5]. They use an active appearance model (ASM) to Identify landmarks from the facial region and perform tracking of facial points, which is a dependable technique to addressing the lighting issues faced by appearance-based systems.

A Kinect sensor is used in a study published in [6] to recognize emotions. It uses an active appearance model to track the face area (AAM). In AAM, fuzzy logic aids in the observation of variation in essential properties. It uses previous knowledge based on the FACS (Facial Action Coding System) to determine emotions. This project is limited to a single subject and three expressions. These appearances and geometry-based techniques both have the drawback of requiring the user to choose a good characteristic to represent face expression system. In geometry-based features, the feature vector is linked to landmarks, and poor identification of the landmark points may lead to reduced recognition performance. Appearance-based traits are less likely to be affected by backdrop variations and misalignment [7]. These descriptors can be incorporated into color, gray value, texture, and statistically deformable shape aspects to create a robust input for architectural performance [8]. Handcrafted features are sensitive to changes in position, aging, and the appearance of the face in general. Traditional approaches, on the other hand, use less memory than neural network-based alternatives. As a result, the aforementioned methodologies for real-time embedded applications are still used in research [9]. For facial expression recognition, [10] proposes a multitask global-local network (MGLN), the approach integrates two components: a Global Face Module (GFM), which extracting spatial features from the frame exhibiting peak expression, and a Part-Based Module (PBM), that models temporal dynamics across the nose , mouth, and eyes regions. GFM and PBM characteristics are extracted based on a CNN network and a LSTM (long short-term memory), with both models combined to effectively capture robust variations in facial expressions. This approach demonstrates significant effectiveness in in-the-wild facial expression recognition (FER), this study suggests a novel deep learning strategy for the merging of 2D and 3D modalities [11]. In [12], the author used nonlinear autoregressive, multilayer perceptron MLP with eXogenous inputs NARX, ra-dial basis function RBF network, and CNN to introduce distance and texture signature characteristics in face emotion identification.

In [13], they introduce a novel hybrid feature representation for facial expression identification from a single image frame which combines deep-learning and SIFT features of various levels extracted from the CNN model, then the hybrid features are subsequently employed for facial expression classification through Support Vector Machines (SVM). Spectral approaches were investigated as shape descriptors for 3D face expression analysis using 3D data [14]. For FER, handcrafted elements are also essential. Deep and handmade characteristics, normally taken from 2D depth pictures and 3D scans for 3D emotion identification, were expressed in combinatorial form by the author in [15]. In [16], the author described this approach utilizes a sparsity-aware deep network, where convolutional features are employed to generate sparse deep features, which are subsequently used to train an LSTM network for phrase detection in conjunction with TOP-landmarks.

Motivation

The face recognition problems from video has been forced, where datasets often lack sufficient data and vary greatly due to differences in expressions, accessories, lighting, and other factors. Expanding a training dataset to include synthetic examples in addition to real samples can increase the model's capabilities by minimizing the domain gap, ac-cording to Nerea Aranjuelo et al, 2021 [17]. This paper develops a technique for creating synthetic data, which is very useful for deep learning applications requiring a lot of data.

Though the proposed dataset in this study comprises a total of roughly 60,000 observations, the Multi-SVM classifier is used and does not necessarily require the use of synthetic data. The synthetic data generation approach in this study could be employed for the deep learning framework, which is outside the scope of this study, according to section **1.2.** Video sequences introduces other challenges which include high computational cost, low accuracy due to redundant data in the datasets, and so forth. Some of the limitations of 2D facial recognition, is handled by 3D or 4D geometric data, though computationally expensive. In this light, information from all three-dimensional data is used to extract features, which we refer to as alpha axes-angle features in this case. While doing so, it is pertinent to reduce the feature space, since three-dimensional data necessitates more computations. The strategy used to solve this problem in two ways, namely, key frame, and feature selection. When an emotion is expressed, the transition from neutral to peak expression, known as the apex [18], is critical for emotion recognition. Using all frames for training may lead to suboptimal classification performance.
Consequently, we prioritize frames featuring the apex of the expression, as they provide the most valuable information, while excluding frames that contribute less to the emotional context [19]. To acquire the peak intensity frames with respect to frame difference between the reference neutral frame and all others, the geometric PSNR, IMMSE, and entropy approaches are utilized. In addition, to minimize the dimension of calculations to only that of relevant predictors, a non-parametric and filter-based feature selection method is used. Finally, the selected features are normalized and input into a multiclass SVM.

For the first time in 4D FER, geometric PSNR, IMMSE, and mean entropy are used for rapid and reliable key-frame selection. Some of the problems associated with 3D or 4D FER problems for use in some frameworks, especially deep learning is the need for larger datasets. Though this study is not on deep learning models application in 4DFER, this paper attempts to initiate a solution to the 3D or 4D FER problem, by Synthetic data generation models for data augmentation in frameworks requiring large datasets, and to solve highly imbalanced dataset problems. Synthetic data that is closely similar to the real data and has a low MSE with respect to original image is generated successfully. The application of synthetic data is sufficiently founded for deep learning applications and necessary for use in conventional 4D FER approaches, since the dataset class distribution is slightly skewed and effects of imbalance maybe negligible. Also, we propose a novel alpha axes angle feature extraction technique.

For the first time in 4D FER, the fairly optimized NCFS is used to make tradeoff decisions between accuracy and feature size. Before applying it to expression recognition, this algorithm investigates the impact of feature selection on the model in terms of accuracy and amount of features. Standardization of features is used to improve classifier performance in 4D FER in the proposed model.

Methodology

The database has been executed on FER using the 3D facial landmarks identified, analyzing performance using the performance metrics specified in latter part of this paper, and the results of this research are compared with other traditional Euclidean distance-based geometric methods and modern hybrid methods and hybrid approaches. The latter part of this study contains a description of the data employed, as well as the experimental framework or design, outcomes, and analysis.

Facial expression databases

To conduct these experiments, this research used two 4D facial databases which state-of-the-art. One of the databases that used is the BU-4DFE [20]; which consists of 101 subjects displaying 6 prototypic facial expressions (anger, happiness, fear, disgust, sadness, and surprise). There are 58 ladies and 43 males, ranging in age from 18 to 45, with a variety of ethnic and racial ancestries among the participants. The subjects in the data total 101. Based on variable intensity levels, an algorithm can be used to extract key-frames from each sequence (neutral, onset, apex, offset). The movie was recorded at a frame rate of 25 frames per second, with each expression sequence reaching about 100, totaling 60,402 frames. Each 3D model has approximately 35,000 vertices in its resolution. There are six classes in this multiclass classification system. The distribution of classes is skewed slightly and its effect in this is considered negligible.

Experimental design:

Based on the data model, 83 facial landmarks are detected and an 83-dimensional vector is constructed for each face and expression, where each landmark is represented by its 3D coordinates (x, y, z). The vector of model is an 83-dimension vector. Furthermore, different feature dimensionality reduction or feature selection algorithms are employed such as Neighborhood component analysis NCA. The NCFS feature selection algorithm is applied to reduce the feature space by only selecting optimal features with respect to feature weights and importance.

Eighty-three (83) facial landmarks on 60,402 observations in the dataset, and 15,150 observations when processed in real-time mode (i.e, for the first one second or 25 frames). The resulting feature vector is then used to train a separate SVM [21]. We employed a 10-fold cross-validation approach, in which the data is divided into ten subsets at random, nine of which are used for training and the other for testing. This is repeated ten times, with each subset being tested separately. To help reduce overfitting, the average error of all iterations is used.

Data Visualization:

By projecting feature points from one frame on a scatter diagram, the data can be viewed in 2D or 3D. A 3D scatter plot with a view angle of [60, -70] is shown in Figure 1. When data is represented in this fashion, it becomes more meaningful since it allows one to see how distinct feature points are scattered in 3D space.

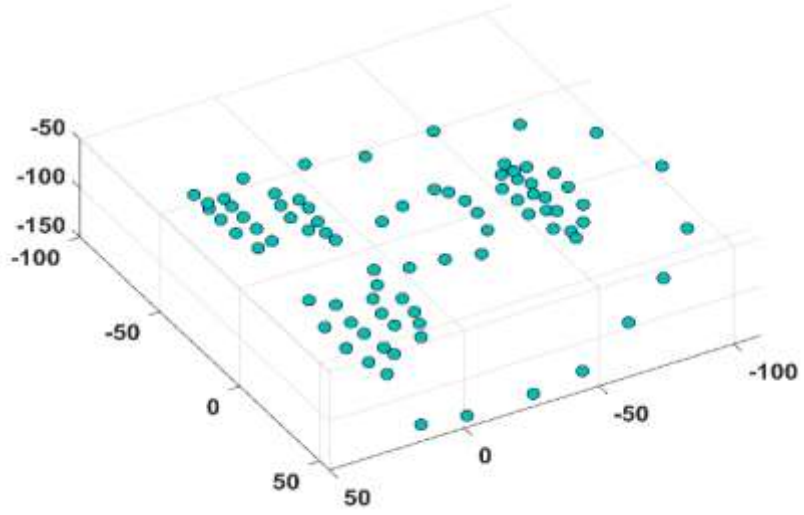

Figure 1. Visualize the 2D scatter plot for subject 43, Male Surprise expression.

Proposed Method

In this study, we take a different approach by creating synthetic data for emotion recognition. A block diagram showing the structure of the proposed network is shown in figure 2.0 below.

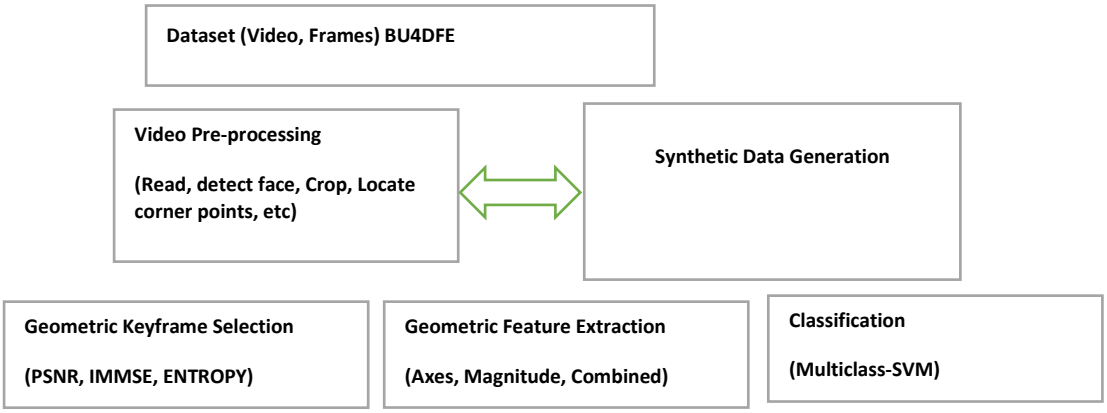

Figure 3. The proposed network Structure

Pre-processing

Preprocessing is a technique for improving the performance of the FER system, it can be applied before performing feature extraction [22]. To improve the expression frames, image preparation involves a variety of operations such as scaling, image clarity and contrast correction, and further enhancement techniques [23].
Cropping and scaling are applied to the face image, while localization employs the Viola-Jones algorithm for detecting facial regions within the input image as part of preprocessing [24, 25]. In real-time, each frame in the video sequence is quantitatively analyzed across all facial expressions to derive corresponding landmark features. Landmark features used is for the full key frame sequence, and This approach aims to improve dynamic expression recognition, but it increases computational complexity. For landmark processing, min-max normalization is applied to the combined axes-angle and magnitude features after concatenation. The landmarks are stored in a matrix, where rows represent observations and columns represent feature points. Each subject's data consists of folders with video sequences for the six expression classes. The proposed algorithm is designed to traverse each sequence for all 101 subjects and organize them in rows in preparation for feature extraction. For full sequence processing, the feature matrix is of m*n dimension, where m is 60,402 and n is 83. Also, the labels have processed all observations, its dimension is k*s, where k is 60,402 and s is equal to 1.

Feature extraction
The FER system's next stage is the feature extraction process. Finding and presenting positive characteristics of concern within an image for further analysis. This process is vital in image processing as it converts raw graphical data into a more abstract form. Techniques employed in feature extraction include texture-based, edge-based, global/local feature-based, and geometric feature-based methods, and patch-based method are some of the types of feature extraction methods. This framework's feature extraction is geometric-based, and such derives its 3D landmark points from 3D texture images in the XY, or XZ, or YZ axes.

3D Model Construction
The 3D face model construction requires a database of faces which contain on a set of training with different face expression in order to testing model, the model used the BU_3DFE database which includes 100 people (44 male, 56 female) with 2500 facial expression. [26].
The 3D model and 3D face mask extraction have been implemented by previous researches [27],

Image quality Measures
The model performance analysis was measured using Peak signal-to-noise ratio (PSNR) to measure the noise of the virtual image which obtained from synthesized process [28, 29]. The measurement unit of PSNR is decibels (db). The PSNR is a tool used to measure the difference between the quality of original images and reconstructed images. When PSNR ratio is high, the quality of reconstructed image is better, the acceptable value in PSNR start from 30 db.
The MSE (Mean Square Error) is the cumulative squared error between the original and the reconstructed image, while PSNR used as a measure of the peak error. When the MSE value is low, the error will be low. The PSNR and MSE can be calculated by:

$$MSE = \frac{\sum_{R,C}[J_1(r,c)-J_2(r,c)]^2}{R*C} \qquad (7)$$

The MSE calculates the mean-squared error, while R represented the number of rows and C represented the number of columns in the input image.

$$PSNR = 10log_{10}\left(\frac{R^2}{MSE}\right) \qquad (8)$$

The PSNR was calculated be previous equation, R = 255. The PSNR applied on the sample included 7 people from BU-3DFE database with 6 different facial expression for each one. The model synthesized the natural image from different facial expression which can be classified to anger, happiness, disgust, surprise, fear, sadness face as shown in Figure 4.

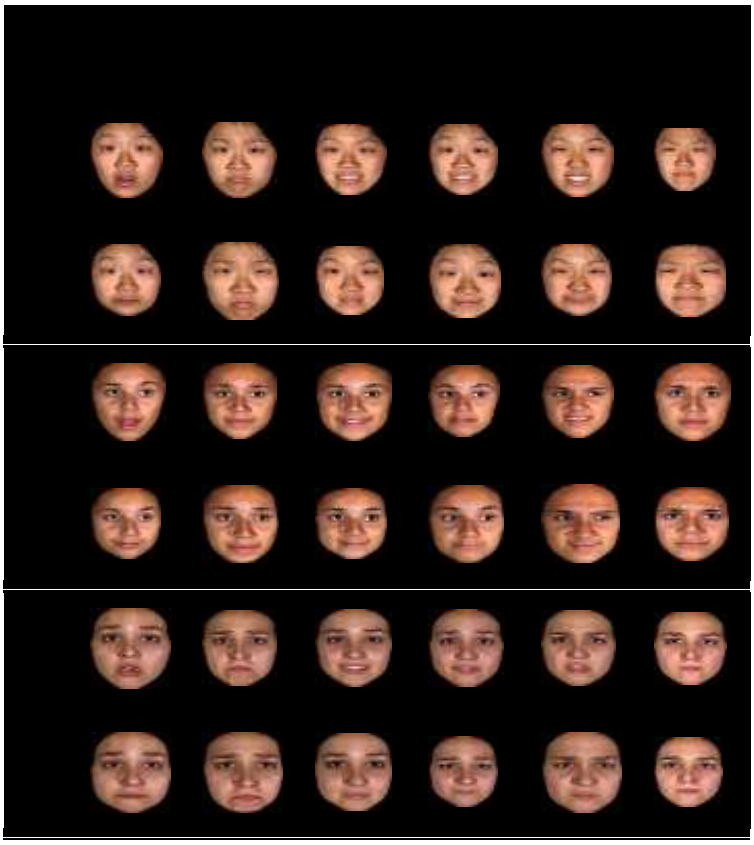

Figure 4. shows the subjects with different intensity levels of the six facial expression (BU3DFE)

Synthetic data generation results

The result of testing model in the table 1 was calculated using PSNR analysis. The table illustrate the similarity ratio between the natural images which generated from different facial expressions.

Table 1. PSNR analysis between original natural face and the synthesized face.

Sample / PSNR	HAPPINESS	FEAR	SADNESS	DISGUST	ANGER	SURPRISE
Sample 1	35.57	34.59	34.37	36.47	33.79	35.69
Sample 2	36.24	35.77	36.51	34.15	34.87	34.57
Sample 3	34.65	34.63	35.36	35.22	35.65	36.65
Sample 4	35.42	34.48	33.47	33.87	36.54	37.23
Sample 5	34.21	36.62	34.59	34.91	33.65	35.31
Sample 6	36.56	35.02	33.95	35.34	35.55	34.11
Sample 7	35.38	34.32	35.88	34.11	34.98	36.52
Avrage	**35.43**	**35.06**	**34.87**	**34.86**	**35.00**	**35.72**

Image quality Measures

This section describes key frame extraction methods, along with the associated issues and challenges. Various approaches for extracting key frames have been proposed in the literature. Studies \[30] and \[31] categorize these methods into different types, including sequential frame comparison, global frame comparison, and methods based on minimizing correlations between frames, minimum reconstruction error in frames, temporal variance between frames, maximum coverage of video frames, reference key frame, curve simplification, key frame extraction using clustering, object- and event-based key frame extraction, and panoramic key frames.

Towards this goal, we propose a novel PSNR-based model to learn effectively the relationships (or differences) between image pairs (Reference & other frames) associated with different facial expressions. The result is a categorization based on SNR intensity levels, and the peaks are recorded and registered for 4D FER. To describe but a few;

Predetermined reference frame:

A key frame is created using this method by comparing a predefined, each frame in the shot is referenced to a reference frame, as discussed in [32]. The main advantage of this approach lies in its low computational overhead and straightforward implementation. Its disadvantage is that it does not effectively portray the global or overall context of an image. The comparison is made in terms of PSNR, MSE, or structural similarity, and entropy.

Entropy Cluster-based key frame selection:

For entropy, the class cluster mean entropy or local mean class entropy is used as a threshold per expression. The frames in each cluster with Frames with the highest entropy are identified as key frames. The advantage of cluster-based approaches is that they capture the global characteristics of the entire expression. However, the drawback of these methods is that they require significant computational resources for cluster generation and subsequent processing key frame selection [40]. Some of the approaches used in this work is described in brief as follows;

- PSNR based
- IMMSE based
- Local mean class Entropy geometric-based
- Mean Peak Entropy texture-based

The difference between frames in a sequence is first used to create a new feature matrix using the input vector (geometric coordinates). This is accomplished by utilizing the matlab peakfinder function to do basic peak analysis. Here all of the frames with high expression intensities are located [33].

The Geometric PSNR approach is applied, and any frames with a low peak-signal-to-noise ratio indicate redundancy. It indicates that the picture intensity value is closer to the reference (neutral) image value. The peakfinder function searches for all local maxima or peaks in the sequence, since the goal is to find key frames that have more discriminative features from the neutral expression. The frames with low PSNR decibels in the distribution is extracted as high intensity frame or relevant key frames. In this context,

a local peak is a data sample that's larger than its neighbors or equal to infinity (Inf). Endpoints that are not Inf are ignored. For flat peaks, only the point with the lowest index is returned, and in cases of multiple flat peaks, the function will return the lowest-indexed point and the indices or location at which the peaks occur. The proposed framework assumes that the PSNR of image data is in the range [0, 1] for images of data type double. Peakval is set to 1 by default in MATLAB.

$$\text{Geometric } PSNR = 10\log_{10}(peakval^2/MSE) \qquad (9)$$

Where the *MSE* is the mean square error between all other frames and the reference frame [34]. This technique helps to reduce the redundancy, yet retaining about 80% of the entire dataset as shown in the table below.

Table 2. Geometric PSNR between reference image (neutral) and all other frames in the sequence.

S/N	Class Size	Key-frames	Redundancy
AN	10124	8299	1825
DI	10171	8503	1668
FE	10044	8422	1622
H	9973	8190	1783
Sa	10142	8177	1965
Su	9948	8424	1524

Table 3. Geometric IMMSE between reference image (neutral) and all other frames in the sequence.

S/N	Class Size	Key-frames	Redundancy
AN	10124	8314	1810
DI	10171	8512	1659
FE	10044	8446	1598
H	9973	8209	1764
Sa	10142	8203	1939
Su	9948	8435	1513

Table 4. Geometric ENTROPY between reference image (neutral) and all other frames in the sequence.

S/N	Class Size	Key-frames	Redundancy
AN	10124	5138	4986
DI	10171	5087	5084
FE	10044	4902	5142
H	9973	5009	4964
Sa	10142	5154	4988
Su	9948	4828	5120

Data split threshold is 7000 per class, which is approximately 80 percent of key frame dataset.

Table 5. General table showing the overall testing and training data size.

	Full dataset size	Key-frames	Training	Test
PSNR based	60402	50015	42000	8015
IMMSE	60402	50119	42000	8119
ENTROPY	60402	30118	24000	6118

Normalization / Feature selection:

Data normalization, a common preprocessing approach, involves scaling or transforming the data to ensure equal contribution from each feature. The effectiveness of machine learning algorithms is heavily influenced by the quality of the data, which is critical for constructing a generalized predictive model for classification problems. Several studies have underscored the role of data normalization in improving both data quality and the performance of machine learning models. Normalization helps to handle outliers. Visualization of the feature space using scattered plot helps us appreciate this.

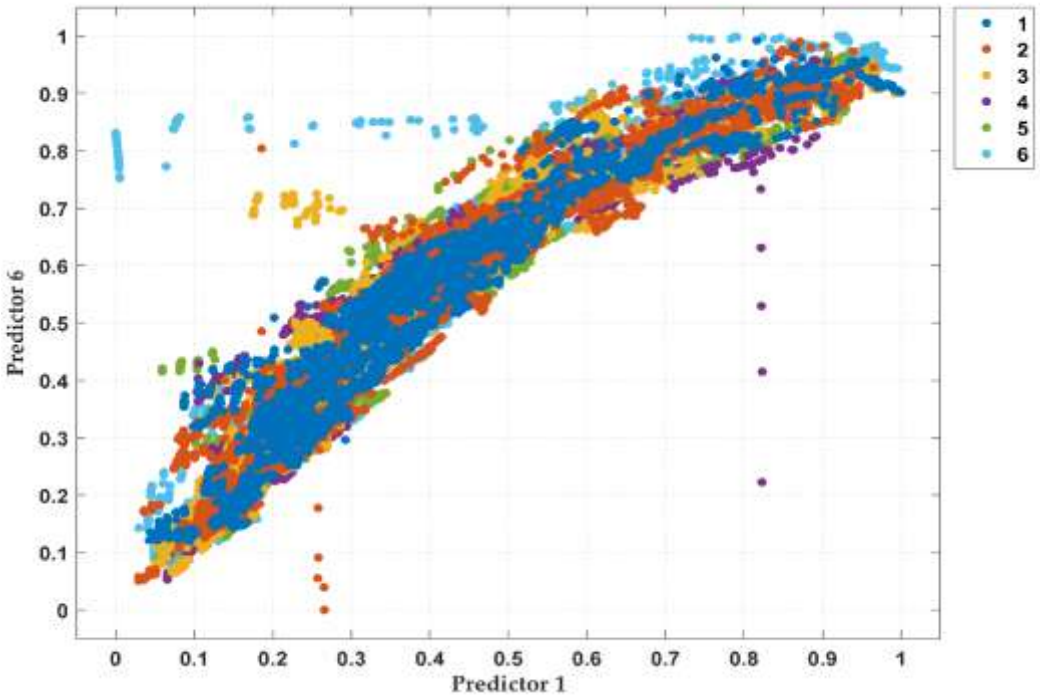

Figure 5. Scattered diagram showing 1st predictor against the 6th predictor

But, while recent work on 3D and 4D data often overlooks feature selection and weighting approaches—critical trends in machine learning for boosting performance—this study examines the effect of fourteen different data normalization techniques on classification performance, considering full feature sets, feature selection, and feature weighting. Furthermore, we introduce a modified Nearest Component Analysis (NCA) that performs an exhaustive search for feature subsets and the optimal feature weights, alongside fine-tuning the Nearest Neighbor Classifier's parameters. Experiments are carried out on the BU4DFE dataset, and results are analyzed in terms of classification accuracy, feature reduction percentage, and runtime.

Key-frame and Feature Selection:

Key-Frame Selection
The video is recorded at 25 frames per second. In this experiment, we develop an algorithm that reduces the video length by selecting only key frames with relevant features. The video sequence is approximately 100 frames, we implement peak entropy key frame selection that implements feature extraction at second of the video sequence for each facial pose. In this paper, our contributions are as follows: We show that (i) ecoc- kernel svm classification methods are highly precise for emotional expression estimation using landmark data only and (ii) they enable early and reliable estimation of peak expression per second and this represents full range of expression intensities, i.e, neutral, onset, apex, offset is consistent with real time recognition systems. This novel idea is in contrast to conventional methods, where feature extraction provides all possible combination of this feature points, hence the feature vector could be in thousands hence increasing computational complexity.

Peak-Entropy Key frame Extraction:
The entropy algorithm is implemented to select key frames with high intensity values. The video AVI files are read and entropy of each image is computed by detection of early expression in each video sequence at the frame rate of 25 fps. Only fames with peak expressions (high entropy) are considered to reduce complexity.
The facial action units change position during emotion expression, different facial points may be activated at different frame in the sequence. The mouth points may peak in frame 'a', while the nose points may peak in frame 'b' and the eye corner points may peak in frame 'c'. This complex dynamics presents a careful observation of multiple peaks across the sequence. Therefore, the peak finder algorithm is useful to detect all peaks in the sequence, and the frames with intensities above mean entropy value are also selected. This provides us with the broad range of summarized frame to represent the expression.

Proposed Peak-Entropy Key frame Selection Algorithm.
EKFE Algorithm: The algorithm for key frame extraction is as follows:

Start

- *Detect face and crop relevant face ROI*
- *Computer PSNR of each frame to the reference frame and count number of frames*
- *Compute peak of PSNR of frames per expression*
- *find index of peaksexpressions in the PSNR vector*
- *Merge results to form key frame index per expression*
- *Repeat for next batch of frames in the next sequence*
- *Use index to select relevant columns from matrix (key-frames)*
- *Extract Axes-angle or distance features from key frames only.*

End

Start

- *Detect face and crop relevant face ROI*
- *Computer entropy of each frame and count number of frames*
- *Compute local mean entropy of frames per expression*
- *Set mean entropy as threshold*
- *find index of frames greater than threshold value*
- *Merge results to form key frame index per expression*
- *Repeat for next batch of frames*
- *Use index to select relevant columns from matrix (key-frames)*
- *Extract Axes-angle or distance features from key frames only.*

End

The expression intensity is harnessed using entropy technique to determine the frames relevant to achieve aim of this paper. Figure 6.0 below shows in block the procedures for proposed entropy key frame selection.

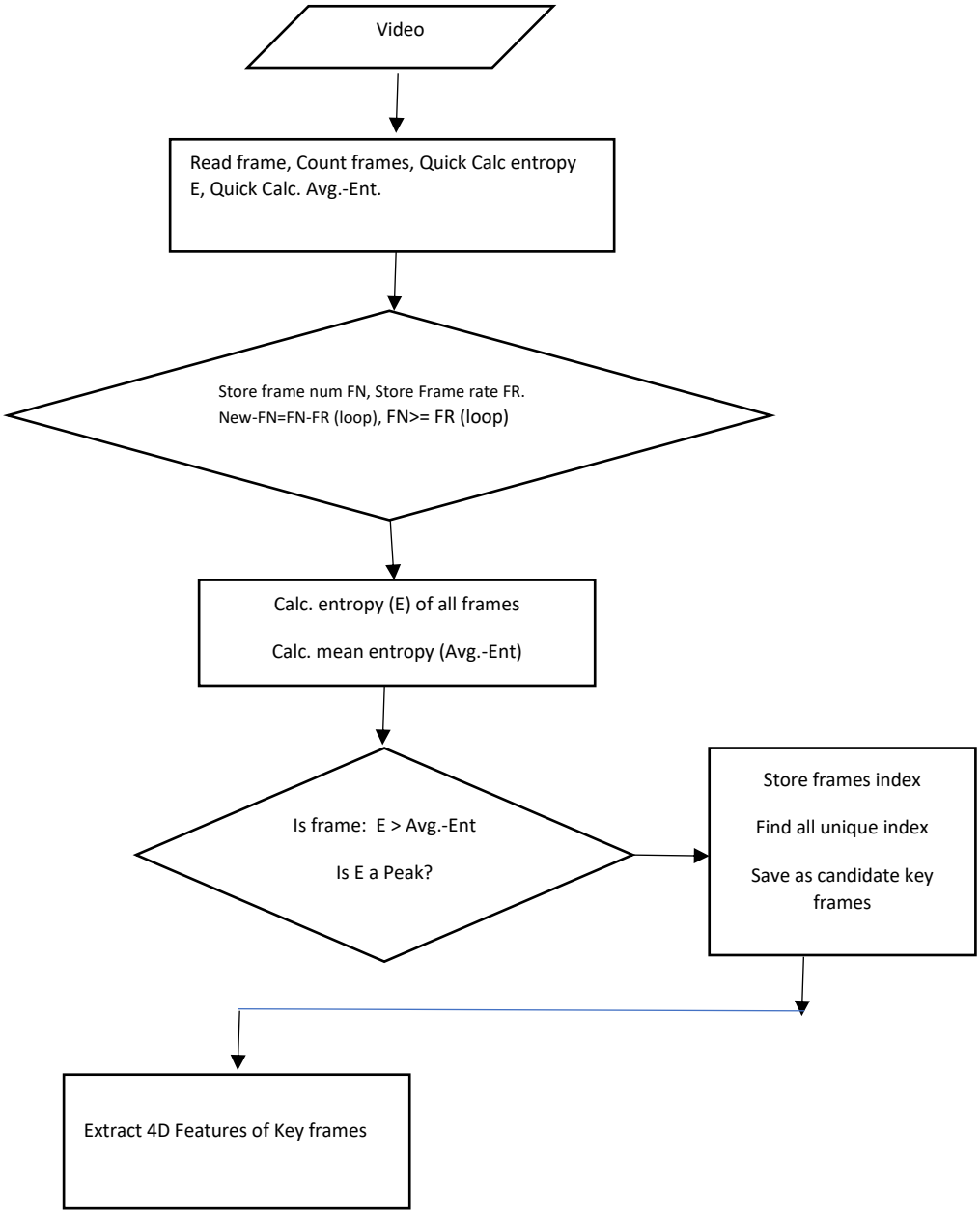

Figure 6. proposed key frame extraction flow chart

Nca Feature Selection Ncfs

Improving 4d Fer Performance:

The objective is to maximize accuracy or performance and probably minimize number of features to reduce time complexity. When the number of features producing this accuracy makes the model computationally expensive, there could be a trade-off between number of features and accuracy. The feature selection algorithm provides feature weights of different predictors and the selection is done by tuning the feature weight starting with initial tol of 0.01 and up to the mean weight by an equal interval 0.1. The limit or threshold that provides the maximum or best performance is assumed to be have the optimal subset of features for the model [35]. The figure below shows a block diagram that describes the NCA feature selection design process.

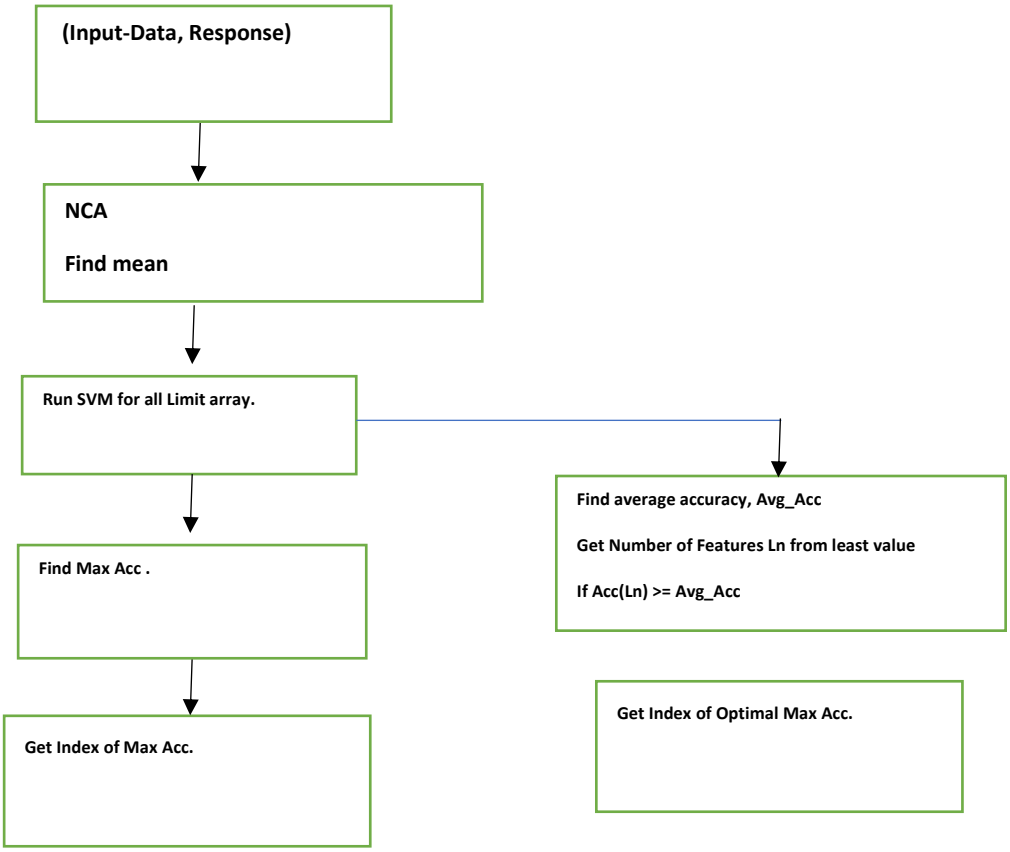

Figure 7. Proposed block diagram for optimized NCA

Feature Extraction

Angle deformations exhibit invariance to both scale and rotation, thus normalization is not necessary. Fig. 8 presents the 3D facial landmarks.

Implementation:

Axes Angular Extraction Equation:

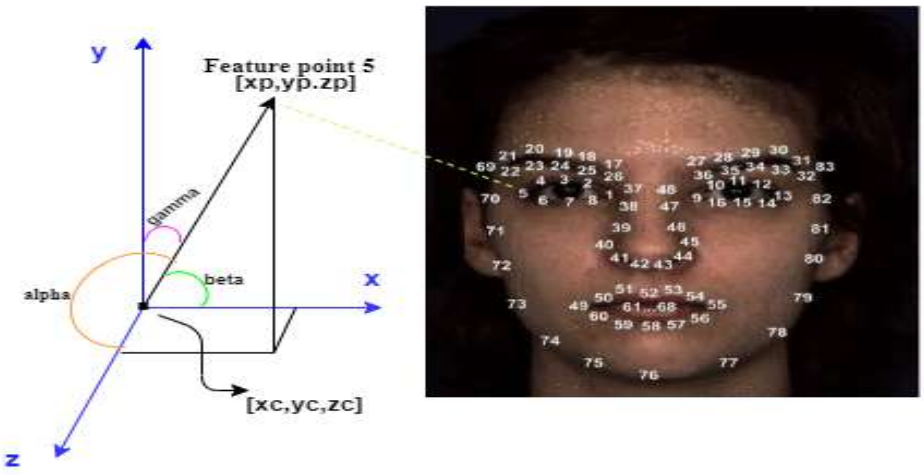

Figure 8. A 3D representation of a face in the BU-4DFE dataset, with the Z-axis oriented towards the observer (Frontal view).

The line segment (also referred to as a vector) connecting the origin Pc (0, 0, 0) to each feature point P(x, y, z), in space which used to label each feature point in the 3D space. The angles between these line segments and the coordinate axes can be calculated with respect to the difference axes in 3D Euclidean space.

F_p = Points of Feature.

Q = Feature points Number.

X_p, Y_p, Z_p = Landmark Coordinates for geometric landmark points.

X_c = origin or center in three-dimensional (3D) space $\mathbb{R}^3$.

The normal vector of the plane, when viewed from the front, is along the Z-axis. In this study, we use combined features and normalize the vectors, as the angles are measured in radians, and the magnitudes/distances represent the coordinates. The features used are as follows: 83 distances from the origin to the feature points, and 83 angles between each feature point and the origin. The angles are calculated with respect to the ZY plane and are measured in radians. This can be mathematically expressed as

zy axes;

$$\textbf{Alpha}\ (\alpha_{p,c}) = \text{atan}\Big(\sum_{p=1}^{Q} \Big(\sqrt{(Z_p - Z_C)^2 + (Y_P - Y_C)^2}\ \ / \ (X_P - X_C)\Big)\Big) \tag{10}$$

Where p=1, 2, 3… Q.

$$\textbf{Distance}\ (\gamma_{p,c}) = \sum_{p=1} (\sqrt{((X_p - X_c)^2 + (Y_p + Y_c)^2 + (Z_p - Z_c)^2)}) \tag{11}$$

Standardization / Z-score normalization:

This is performed so that the features are rescaled to inherit the properties of a standard normal distribution with mean μ=0 and standard deviation from the mean σ=1.

$$Z = (x-\mu)/\sigma \tag{12}$$

Standardizing the features to be centered on 0 with a standard deviation of 1 is not only important for comparing quantity's that have distinct units. It is important to standardize in this study to allow results obtained from both angle-based and distance-based features to be comparable but it is also a general prerequisite for many robust machine learning algorithms [36]. The model suffers from some outlier effect and to contain this a min-max normalization is applied.

Min-Max Normalization technique makes the data to be scaled to a fixed range - usually zero (0) to one (1) as seen in figure above. The significance of having this bounded range in contrast to z-score normalization is an outcome that presents smaller standard deviations, which in turn suppress the effect of outliers present in the data. It is expressed as follows:

$$\boldsymbol{X_{norm}} = \frac{\mathbf{X} - \boldsymbol{X_{min}}}{\boldsymbol{X_{max}} - \boldsymbol{X_{min}}} \tag{13}$$

4D FER FULL SEQUENCE DATA

The whole sequence data contains all frames in the dataset, and it is processed using the pipeline configuration illustrated in the table 5 below. With entire sequence data, performance may be obtained before feature selection and compared to performance after feature selection. The description and parameters of this block is summarized in the table 6.

Table 6. shows the parameters of previous block

Class	Label	No. of Observation	Gamma Feature Size	Magnitude Feature Size
1	Anger	10124	83	83
2	Disgust	10171	83	83
3	Fear	10044	83	83
4	Happy	9973	83	83
5	Sad	10142	83	83
6	Surprise	9948	83	83
Total		60,402		

Classification:
Using support vector machine (SVM) binary learners, we train a multiclass error-correcting output codes (ECOC) model. The available hyper-parameter settings of multi-SVM is shown in table 7 below. The Fine Gaussian SVM has a kernel scale of sqrt. (P)/4, allowing it to distinguish between classes in fine detail. The function fitcecoc uses $K(K-1)/2$ binary support vector machine (SVM) models and the one-versus-one (1vs1) coding design to create a Classification-ECOC model Mdl., where K is the number of unique class labels (levels). A one-versus-one strategy for six classes yields 15 binary learners [37]. For multi-class SVM, the one-versus-one technique uses a max-wins voting strategy, in which each classifier assigns the instance to one of two classes, then increases the vote for that class by one vote, and finally the class with the most votes determines the instance classification. If 'auto-scale' is set to false, the SVM model is more sensitive to variables with large variance and less sensitive to variables with small variance. To treat all variables on equal footing, regardless of variance values, the 'auto-scale' is set to true. When variables are standardized, the optimal kernel width is often close to 1. To optimize the box constraint, a uniformly spaced logarithmic scale, from 1e-6 to 1e+6 is used [38] as shown on table 7 below.

The performance of the classification network is dependent on the on the various model hyper-parameter settings such as kernel function, kernel scale, multiclass strategy, standardization state, box constraint. The grid settings are depicted in table below.

Table 7. SVM hyper-parameters.

S/N	Type	Variables
1	Model preset	Quadratic, Cubic, Fine, Median, Coarse
2	Kernel function	Linear, Gaussian, Polynomial
3	Kernel scale	Positive values log-scaled in the range [1e-3,1e3]
4	Box constraint	Positive values log-scaled in the range [1e-3,1e3]
5	Multiclass method	One vs One, One vs All
6	Standardization	True, False
7	Polynomial order	Integers in the range [2,4]

Performance Analysis
Using K-fold cross-validation with Name-Value pair arguments is a well-established policy for evaluating machine learning models [39]. To assess the model's generalization ability, 10-fold cross-validation is performed in this research. The data is randomly divided into ten equal-sized folds, with one-fold used for validation while the model is trained on the remaining nine. This process is iterated 10 times, ensuring each fold is used once for validation.
The confusion matrix serves as a key tool for evaluating classification models in machine learning. It is a table that summarizes a model's performance, helping to identify top-performing models. To compare the angle and combined methods, we present the confusion matrix in percentage format to assess the proportion of correctly and incorrectly classified features. Furthermore, recognition accuracy, derived from the matrix, is used to evaluate the models' performance and determine which one has the strongest predictors or highest accuracy [40].

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{Total Population.}) \tag{14}$$

Implementation of NCFS feature selection:

CASE 1: Distance + IMMSE key + NCFS

The number of MSE key-frames include; 8314, 8512, 8446, 8209, 8203, 8435 for expression Angry (AN) to Surprise (SU) respectively (see table 3). The mean feature weight obtained is 1.1764, and the search grid range is built in steps of 0.3 in reverse order from the mean value to a predetermined minimum of 0.01 [0.01: 0.3: Mean]. To facilitate the establishment of a larger search space, the step size chosen is minimal (0.3). This provides a three step range; 0.010, 0.310, 0.610, 1.010. The training data samples are 42,000 and test data samples are 8119. This provides four (4) steps in this range: 0.010, 0.310, 0.610, and 1.010. There are 42,000 training data samples and 8119 test data samples.

Table 8. Results of NCFS Feature selection for Distance + IMMSE + NCFS.

Weight	Features	Performance	Exec. Time (sec)
0.0100	78	0.991009	220.8262
0.3100	56	0.983741	154.5812
0.6100	50	0.980416	149.0264
0.9100	47	0.978322	147.6487

Table 8 shows that the parameters on the top row have the best performance (0.991009), the most features (78), and the longest execution time (220.8262 seconds), which is less desired for real-time applications. The system selects the first row (feature weight 0.01, feature size 78) for final emotion categorization since one of the goals is to employ only predictors with strong discriminative power to achieve maximum recognition accuracy. Figure 10 shows a 3D plot depicting the relationship between the amount of features, execution time, and recognition accuracy.

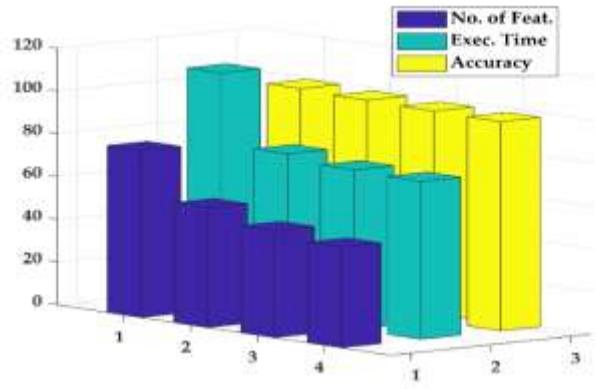

Figure 9. 3D plot of Feature size (dark blue) versus Execution time (Light blue scaled *2) versus Accuracy (Yellow), at a viewing angle [-39.9, 9].

The NCFS algorithm eliminates all feature indexes that are considered redundant features, namely 9, 37, 48, 64, and 66, according to the chosen setting. As a result, the size of the subset feature space is 78.

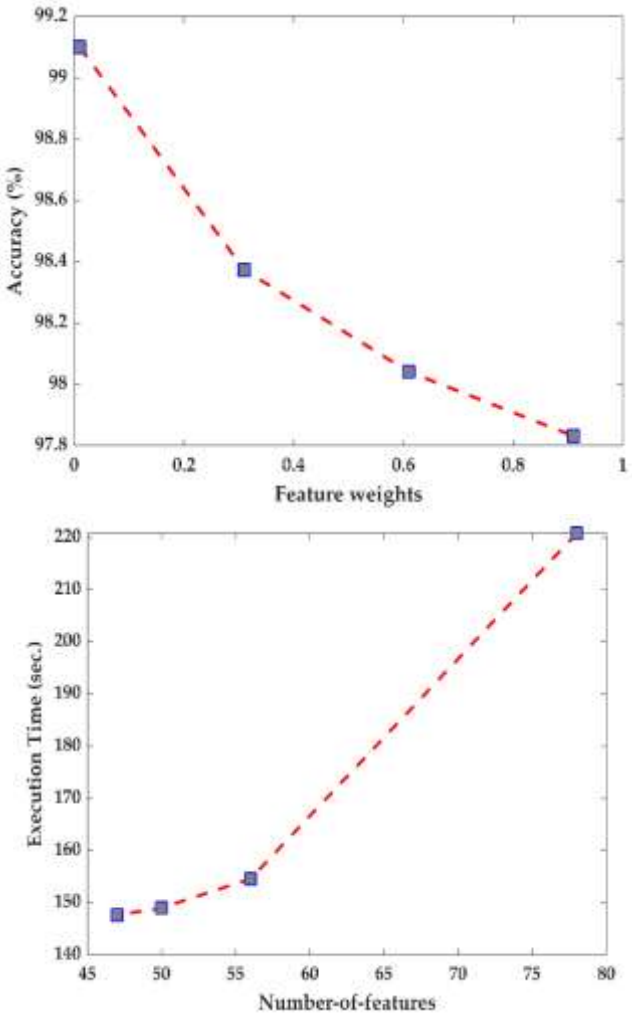

Figure 10 (a.) Line plot of Accuracy versus Feature Weight, (b). Exec. time (sec) versus Number of features

Figure a shows a line plot of performance or accuracy against feature weight. Feature weights closer to zero, according to the NCA algorithm, are less relevant. As feature weights approach zero, accuracy improves, as shown in the graph. As a result, the smallest feature weight (0.01) yields a performance of 99.1009%. (See table 8). Accuracy and performance are used interchangeably in this study. With fewer features, performance drops. The feature weights W, and Accuracy Acc. appear to have a negative or inverse relationship; as W grows, Acc.

declines. This is the same with the relationship between features weight W and Feature Size Sz; an increase in feature weight, also means decrease in overall Sz.
The figure b. above shows that as features increases, the execution time also increases accordingly. The least feature size (47) would be processed at the least execution time of 80 seconds.
As seen in figure b., as the number of features increases, so does the execution time. The least feature size (47) would be processed at the least execution time of 80 seconds. Consider a tradeoff between accuracy and amount of features: if the goal is to reduce computational complexity, the algorithm chooses the model with the shortest execution time, even if it has the lowest accuracy (97.83 percent). Alternatively, a threshold with the average performance value can be set so that all models above it are evaluated when choosing the least accurate model. The relationship between the number of features or Sz and execution time is positive as shown on the figure (10b).

Table 9. Confusion matrix for Validation and Test for Distance + IMMSE key

Val.	**AN**	**DI**	**FE**	**H**	**Sa**	**Su**		**Test**	**AN**	**DI**	**FE**	**H**	**Sa**	**Su**
AN	6963	30	4	1	2	0		AN	1310	0	2	1	1	0
DI	66	6842	58	11	19	4		DI	11	1478	6	1	3	1
FE	20	40	6852	24	57	7		FE	3	4	1411	10	0	6
H	10	15	51	6808	96	20		H	1	5	0	1160	6	1
Sa	9	3	10	12	6947	19		Sa	1	1	1	2	1197	1
Su	17	21	11	17	148	6786		Su	2	0	5	0	0	1490

The Distance + IMMSE key + NCFS + SVM had an average validation recognition accuracy of **99.3190476190476** percent, and the test accuracy is **99.10.**

CASE 2: Angle + IMMSE key + NCFS

The number of MSE key-frames is the same as those of distance because the key-frames are selected using the landmark coordinates. The mean of the feature weights is 0.6944, and the search grid range is 0.01 to the mean value [0.01: 0.3: Mean]. This gives you a three-step range of 0.010, 0.310, and 0.610. There are 42,000 training data samples and 8119 test data samples.

Table 10. Results of NCFS Feature selection for Angle + IMMSE key + NCFS.

Weight	Feature size	Performance	Elapse time (sec)
0.0100	41	0.971794555979801	257.400578200000
0.3100	30	0.955905899741347	144.271639600000
0.6100	27	0.948269491316665	155.790281000000

Maximum performance recorded is 0.971794 percent, which has also the most number of features and the higher computational cost in terms of execution time (257.4 sec.). The model with the highest accuracy is chosen as the subset feature subset selected with size of 41. The following features are eliminate as redundant;
The highest reported performance is 0.971794 percent, which also has the most features and a greater computational cost in terms of execution time (257.4 sec.). The subset feature with size of 41 is chosen as the model that produce the highest performance and is hence adequate for facial expression classification. The following 42 feature indexes are eliminated because they are considered redundant in this experiment: 1, 2, 5. 8, 11, 12, 13, 14, 15, 16, 17, 20, 21, 23, 24, 25, 30, 31, 32, 33, 35, 37, 38, 42, 50, 51, 52, 54, 55, 56, 57, 58, 59, 62, 63, 64, 67, 68, 71, 76, 77, 81

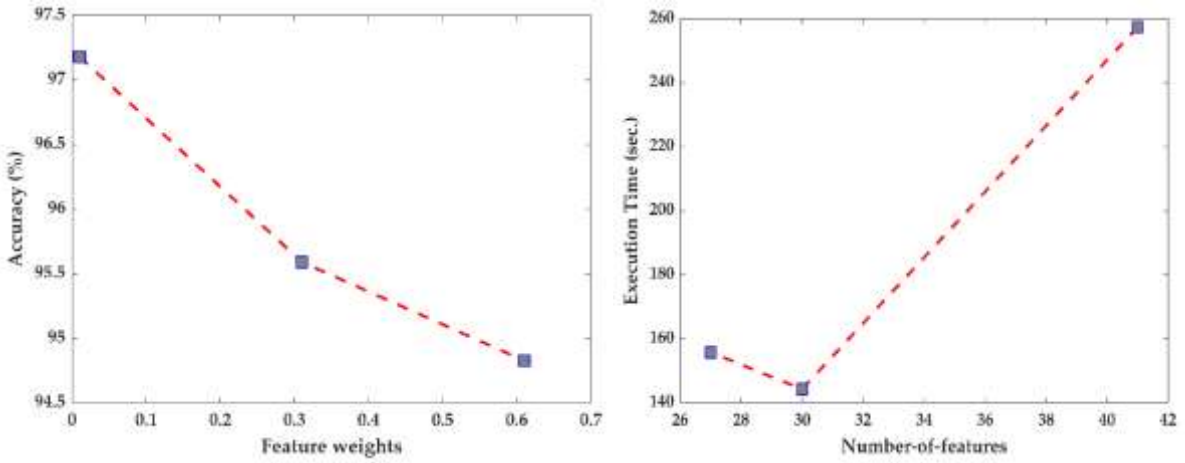

Figure 11. (a.) Line plot of Accuracy versus Weight, (b.) Elapse time versus feature size

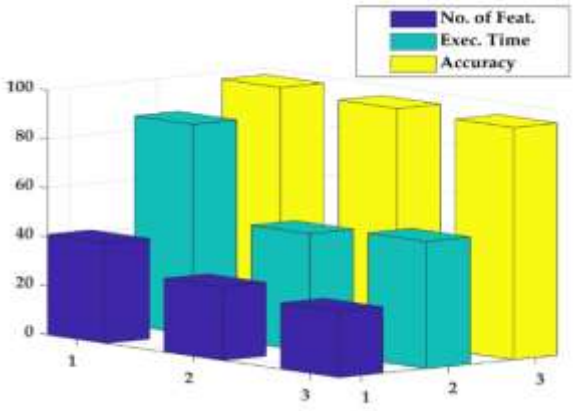

Figure 12. 3D plot of Feature size (dark blue) versus Time taken (Light blue scaled *3) versus Accuracy (Yellow), viewing angle [-39.9, 8].

Table 11. Confusion matrix for Validation and Test for Angle + IMMSE key + NCFS

Val.	AN	DI	FE	H	Sa	Su		Test	AN	DI	FE	H	Sa	Su
AN	6786	82	38	17	55	22		**AN**	1278	14	8	2	11	1
DI	123	6663	83	69	27	35		**DI**	24	1426	30	19	3	10
FE	119	80	6618	76	79	28		**FE**	21	21	1344	22	26	12
H	108	30	174	6440	177	71		**H**	19	6	36	1102	40	6
Sa	42	37	119	57	6637	108		**Sa**	10	9	20	11	1138	15
Su	75	35	127	84	135	6544		**Su**	17	5	29	22	25	1337

The Angle + IMMSE key + NCFS + SVM had an average validation recognition accuracy of 98.2714 percent, and the test accuracy is 97.8076**.**

CASE 3. Angle + Pointwise PSNR key + NCFS

The mean of the feature weights is 1.3651, and the search grid range is 0.01 to the mean value [0.01: 0.3: Mean]. There are five steps in this range: 0.010 0.310 0.610 0.910 1.210. There are 42,000 training data samples and 8015 test data samples.

Table 12. Results of NCFS Feature selection for Angle + PSNR key + NCFS

Weight	Feature size	Performance	Elapse time (sec)
0.0100	51	0.972925764192140	201.048504000000
0.3100	36	0.963942607610730	139.359831200000
0.6100	30	0.953087960074860	142.157317200000
0.9100	27	0.947473487211478	146.546642100000
1.2100	24	0.935870243293824	146.700644100000

The feature indexes that were chosen by the algorithm are: 1, 2, 3, 4, 5, 7, 9, 10, 16, 17, 18, 19, 22, 25, 26, 27, 29, 30, 34, 35, 36, 39, 40, 41, 43, 44, 45, 46, 48, 49, 50, 53, 55, 57, 60, 61, 65, 66, 67, 69, 71, 72, 73, 74, 77, 78, 79, 80, 81, 82, 83.

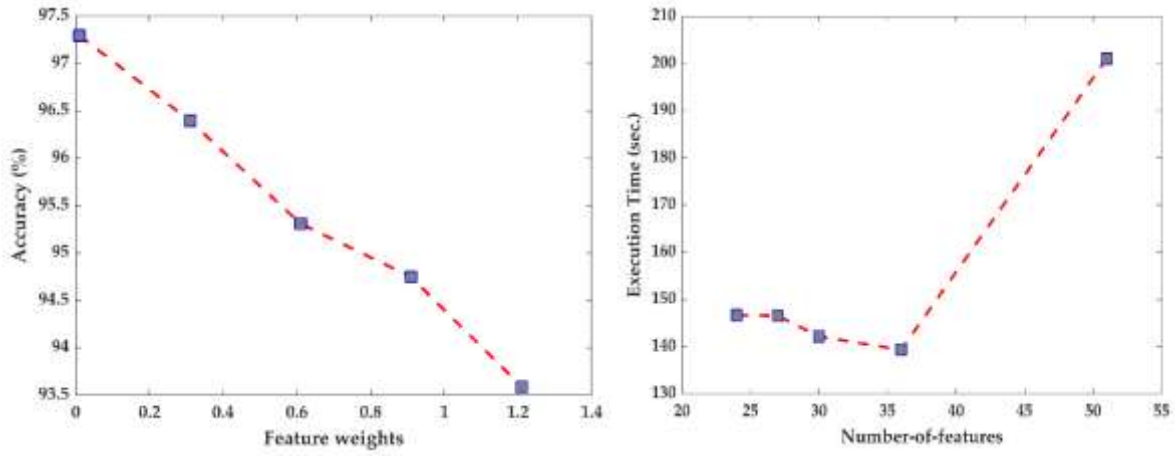

Figure 13. (a.) Line plot of Accuracy versus Weight, (b.) Elapse time versus feature size.

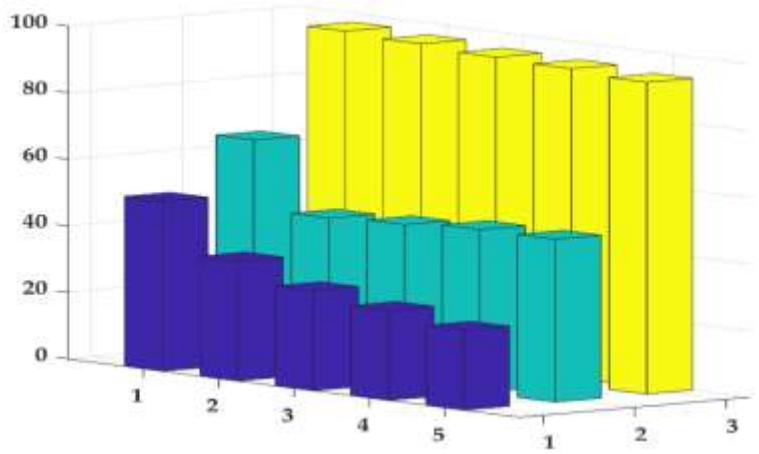

Figure 14. 3D plot of Feature size (dark blue) versus Time taken (Light blue scaled *3) versus Accuracy (Yellow), viewing angle [-39.5, 9].

Table 13. Confusion matrix for Validation and Test for Angle + PSNR key + NCFS.

Val.	AN	DI	FE	H	Sa	Su		Test	AN	DI	FE	H	Sa	Su
AN	6924	25	14	6	16	15		AN	1282	6	7	0	3	1
DI	53	6835	66	23	4	19		DI	16	1453	14	9	2	9
FE	37	28	6845	55	26	9		FE	4	3	1388	12	6	9
H	51	9	65	6809	37	29		H	6	4	12	1147	10	11
Sa	15	1	68	50	6835	31		Sa	3	0	14	14	1141	5
Su	10	8	47	13	58	6864		Su	1	7	9	4	25	1378

The Angle + PSNR key + NCFS + SVM had an average validation recognition accuracy of 98.6809 percent, and the test accuracy is 97.9663**.**

CASE 4. Distance + Pointwise PSNR key + NCFS

Fine Gaussian SVM, Gaussian kernel function, Kernel scale of 2.3, Box constraint level of 1, and One-vs-One multiclass technique are among the classification models settings used in this study. The feature weight average is 1.0502, and the search grid range is from 0.01 to the mean value [0.01: 0.3: Mean]. This provides a four-step range: 0.010, 0.310, 0.610, and 0.910. As indicated in table 2, the training data samples are 42,000 and the test data samples are 8015.

Table 14. Result of NCFS Feature selection for Distance + PSNR key + NCFS

Weight	Feature size	Performance	Elapse time (sec)
0.01000	64	98.45290	637.338022400000
0.31000	49	97.41734	270.291037700000
0.61000	45	97.01809	198.144943200000
0.91000	40	96.29444	191.718121800000

The highest recorded accuracy is 98.4529 percent, which likewise contains the most features, but its execution time is more than double that of the other feature sizes in the table. The acceptable feature size for the 4D FER is 64 for expression classification. Using the hyper-parameters and 10-fold cross validation, this parameter is used to construct a new vector for Multi-SVM classification.

The index of the optimal feature subset is as follows; 3, 5, 7, 11, 12, 13, 14, 15, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 55, 56, 58, 59, 60, 61, 64, 65, 66, 67, 69, 70, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83.

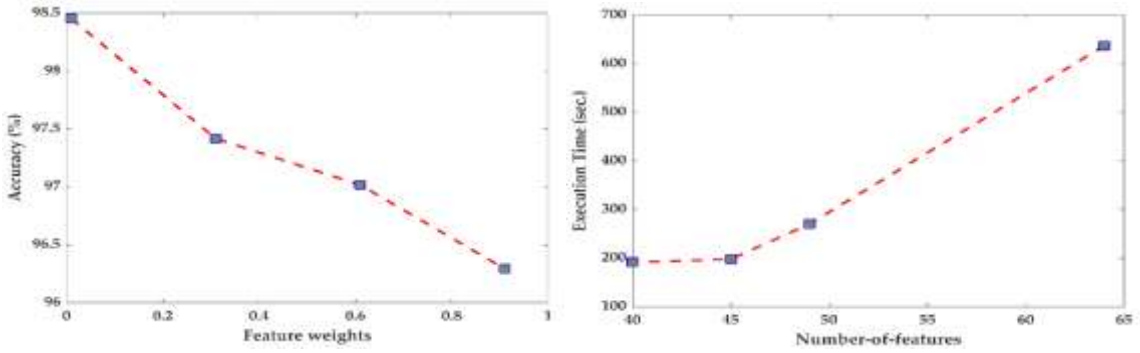

Figure 15. (a.) Line plot of Accuracy versus Weight, (b.) Elapse time versus feature size

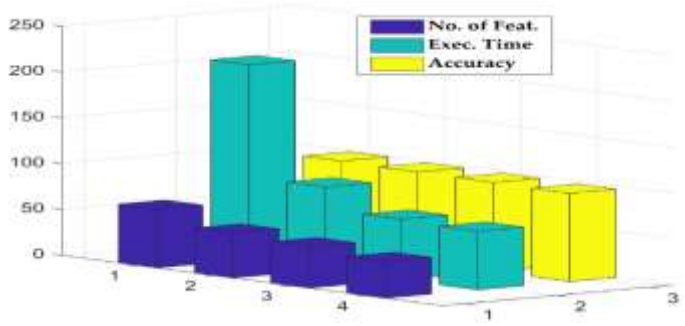

Figure 16. 3D plot of Feature size (dark blue) versus Time taken (Light blue scaled *3) versus Accuracy (Yellow), viewing angle [-39.9, 10].

Table 15. Confusion matrix for Validation and Test for Distance + PSNR + NCFS

Val.	AN	DI	FE	H	Sa	Su		Test	AN	DI	FE	H	Sa	Su
AN	6860	44	32	7	53	4		**AN**	1269	10	12	1	7	0
DI	147	6647	130	27	42	7		**DI**	32	1412	40	11	8	0
FE	48	51	6721	51	113	16		**FE**	8	15	1355	12	24	8
H	17	32	120	6635	166	30		**H**	4	6	33	1099	42	6
Sa	26	10	39	37	6867	21		**Sa**	4	4	7	5	1153	4
Su	22	49	37	34	238	6620		**Su**	6	22	14	5	62	1315

The Distance + PSNR key + NCFS + SVM had an average validation recognition accuracy of 98.9642 percent, and the test accuracy is 98.1659**.**

CASE 5. Distance + Entropy key + NCFS

The mean of the feature weights is 0.9889, and the search grid range is 0.01 to the mean value [0.01: 0.3: Mean]. This gives a three-step range: 0.010, 0.310, 0.610, and 0.910. As indicated in table 3, the training data samples are 24,000 and the test data samples are 6118.

Table 16. Result of NCFS Feature selection for Distance + Entropy key + NCFS

Weight	Feature size	Performance	Elapse time (sec)
0.0100	68	0.9797319	61.8312
0.3100	54	0.9717227	59.6325
0.6100	52	0.9691075	68.0869
0.9100	42	0.9576659	69.0746

Max accuracy recorded is 97.973 percent, which has also the maximum number of features but its execution time is more than double the time table the other feature sizes recorded in the table. The acceptable feature size for the 4D FER is 42. This parameter is used to create a new vector for Multi-SVM classification using the obtained optimized hyper-parameters and 10-fold cross validation.

The maximum accuracy reported is 97.973%, and it also contains the most features, but its execution time is slightly more than that of the row 1, but definitely less than rows 3 and 4 in the table. The acceptable feature size for the 4D FER is 42 for expression classification. Using the chosen SVM hyper-parameters settings and K-fold cross validation, this parameter is used to construct a new vector for Multi-SVM classification.

The chosen indexes of the feature subset is as follows; 5, 7, 8, 12, 13, 14, 20, 21, 22, 25, 31, 32, 33, 35, 40, 42, 43, 46, 47, 49, 50, 51, 52, 53, 56, 58, 59, 65, 67, 69, 70, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83.

Table 17. Confusion matrix for Validation and Test for Distance + Entropy key + NCFS

Val.	AN	DI	FE	H	Sa	Su		Test	AN	DI	FE	H	Sa	Su
AN	3891	52	22	13	13	9		**AN**	1077	25	15	4	6	11
DI	60	3814	76	20	13	17		**DI**	19	1031	26	8	2	1
FE	28	79	3742	61	58	32		**FE**	7	23	828	18	14	12
H	25	16	108	3747	74	30		**H**	5	3	41	923	20	17
Sa	1	14	52	27	3891	15		**Sa**	0	5	7	12	1125	5
Su	19	29	47	36	147	3722		**Su**	2	7	19	7	27	766

The Distance +Entropy key + NCFS + SVM had an average validation recognition accuracy of 97.9208 percent, and the test accuracy is 97.3030**.**

CASE 6. Angle + Entropy key + NCFS

The mean of the feature weights is 0.7522, and the search grid range is 0.01 to the mean value [0.01: 0.3: Mean]. This provides a four-step range: 0.010, 0.310, 0.610, and 0.910. As indicated in table 3, the training data samples are 30,000 and the test data samples are 5439.

Table 18. Result of NCFS Feature selection for Angle + Entropy key + NCFS.

Weight	Feature size	Performance	Elapse time (sec)
0.0100	55	0.980143	85.7919
0.3100	38	0.979040	79.2816
0.6100	33	0.976833	66.5122

The highest observed accuracy is 98.014 percent, which likewise has the most features, but its execution time is slightly more than that of the other feature sizes in the table. The maximum feature size for the 4D FER is 55. The indexes of the selected feature subset is as follows; 3, 4, 5, 6, 7, 8, 9, 10, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 31, 34, 36, 37, 39, 40, 44, 45, 46, 47, 48, 49, 50, 54, 55, 56, 60, 61, 64, 65, 67, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 83. Also, the key frames are as thus; AN (6235), DI (5824), FE (5988), H (5877), SA (5795), SU (5720)

Table 19. Confusion matrix for Validation and Test for Angle + Entropy key + NCFS.

Val.	AN	DI	FE	H	Sa	Su		Test	AN	DI	FE	H	Sa	Su
AN	4813	94	43	8	38	4		**AN**	1183	25	10	1	15	1
DI	78	4773	78	4	22	45		**DI**	15	774	18	2	6	9
FE	79	73	4745	35	42	26		**FE**	18	13	931	12	10	4
H	55	40	107	4648	127	23		**H**	11	7	21	812	15	11
Sa	38	25	84	19	4797	37		**Sa**	10	6	12	0	759	8
Su	19	28	64	13	102	4774		**Su**	5	2	8	3	19	683

The Angle + Entropy key + NCFS + SVM had an average validation recognition accuracy of 98.5066 percent, and the test accuracy is 97.775 percent**.**

Table 20. General table for all methods.

S/N	Method	Validation	Test
1	Distance + IMMSE key + NCFS + SVM	99.31904	**99.06392**
2	Angle + IMMSE key + NCFS + SVM	98.27142	97.80761
3	Angle + Pointwise PSNR key + NCFS	98.68095	97.96631
4	Distance + Pointwise PSNR key + NCFS	98.96428	98.1659
5	Distance + Entropy key + NCFS	97.92083	97.30304
6	Angle + Entropy key + NCFS	98.50666	97.77532

Table 21. Results as per maximization of accuracy objective.

S/N	Model	Weight	Feature size	Performance
1	Angle + PSNR	0.0100	51	**97.29257**
2	Angle + IMMSE	0.0100	41	**97.17945**
3	Angle + Entropy	0.0100	55	98.01434
4	Distance + PSNR	0.0100	64	98.45290
5	Distance + IMMSE	0.0100	78	**99.10087**
6	Distance + Entropy	0.0100	68	**97.97319**
	Average		59.5	98.00222

Table 22. Results as per Minimization of feature size objective.

S/N	Model	Weight	Feature size	Performance
1	Angle + PSNR	1.2100	24	**93.58702**
2	Angle + IMMSE	0.6100	27	**94.82694**
3	Angle + Entropy	0.6100	33	97.68339
4	Distance + PSNR	0.9100	40	**89.29444**
5	Distance + IMMSE	1.0100	47	**97.83224**
6	Distance + Entropy	0.9100	42	95.7665
	Average		35.5	94.83176

Table 23. Results as per least Time complexity objective – minimize time and feature size.

S/N	Model	Max Feature Size / Time	Feature size	Time (sec.)
1	Angle + PSNR	51(201.05)	24	146.70
2	Angle + IMMSE	41(257.01)	27	155.79
3	**Angle + Entropy**	55(85.79)	33	**66.51**
4	Distance + PSNR	64(637.34)	40	191.72
5	Distance + IMMSE	78 (220.82)	47	147.68
6	Distance + Entropy	68(61.83)	42	69.07
	Average		35.5	129.58

From the foregoing, the algorithm has complexity of *O(Mc)*, and *Mc* represents the number of features. The experiment was conducted on MATLAB 2020a, Intel-R Core-i5 (3.60 GHz) with RAM capacity 16GB. The proposed framework has a complexity of *O(Te)*, where *Te* is the number of input expression images.

Top performing model:

The Distance + IMMSE approach was the top performer in this experiment in terms of maximum accuracy and least execution time. Its confusion matrix (see Table 9) displays its class performance in terms of the number of samples properly and erroneously classified, and is expressed in percentages per class for the six fundamental facial emotions (see Table 24).

. Confusion matrix for Validation and Test for Distance + IMMSE key

Val.	AN	DI	FE	H	Sa	Su		Test	AN	DI	FE	H	Sa	Su
AN	6963	30	4	1	2	0		AN	1310	0	2	1	1	0
DI	66	6842	58	11	19	4		DI	11	1478	6	1	3	1
FE	20	40	6852	24	57	7		FE	3	4	1411	10	0	6
H	10	15	51	6808	96	20		H	1	5	0	1160	6	1
Sa	9	3	10	12	6947	19		Sa	1	1	1	2	1197	1
Su	17	21	11	17	148	6786		Su	2	0	5	0	0	1490

Table 24. Class performance for top performing model for testing.

Class	Recognition Accuracy
Angry	99.695
Disgust	98.533
Fear	98.396
Happy	98.891
Sad	99.501
Surprise	99.532
Avg.	**99.091**

Average recognition accuracy is **99.10** for Distance + IMMSE which means 8046 samples out of 8119 samples are recognized correctly by the system. Surprise is one of the most difficult emotions to recognize, according to Maria Guarnera et al 2015 [41]. However, as indicated in table 24, the surprise class did well (99.1096).

Experimental Results and Discussions

The aim of this survey is to investigate the general nature of the facial information from 101 subjects to recognize expressions of anger, happiness, fear, surprise, sadness, disgust emotion by using the proposed techniques in this paper. We first discuss our findings in existing studies from the perspective of model type, and the methodology. We evaluate the synthetic generation model based on MSE separately. We perform geometric Pointwise PSNR and MSE key frame selection.

We analyze the results of the key frame selection in terms of accuracy, execution time, and feature size. Also the result of support vector machine classification is analyzed for the facial expression classes. Then, we perform a unified evaluation in terms of recognition accuracy and identify the top performing model(s).

Based on the results of our tests, it is clear that our proposed method can effectively cope with 4D face expression recognition. Using an axes-angle feature-based method, we proposed a new and concise feature engineering framework for expression recognition from 3D dynamic images (video) in this research. The Axes-angle and magnitude predictors extracted separately to feature vectors. In this study, a 4-D facial expression network is configured to process the video data, and then the data is split into test and train set, by 10-fold cross validation (10-CV) for all experiments.

According to its capacity to test multiple feature weights, a trade-off between the quantity of features and the accuracy acquired from the selection process, the proposed NCFS greatly surpasses conventional NCA. More exactly, NCFS identifies a more reliable feature size value for classifying emotions. Although the tradeoff scenarios include both circumstances where the goal is to minimize execution time or enhance accuracy, the focus of this research is on the latter (See Table 21-23).

The table 23 provides the general result for all methods the SVM model's performance is assessed with and without feature selection. Scaling the data using z-score normalization leads to a significant improvement in classification accuracy. This enhancement occurs because SVM is more sensitive to variables with larger variance and less sensitive to those with smaller variance. When the data is normalized, the ideal kernel width for SVM is generally found to be close to 1. For fair comparison, on table 25 we compared system performance of all proposed methods and also included results from other recent published work or state-of-the-art approaches conducted on the BU-4DFE data set.

Table 25. Performance (%) comparison of the state-of-the-art methods with proposed methods based on the BU-4DFE dataset.

Method	Experimental Setting	Acc %
Abd El Rahman et al 2020 [42]	3D Deformation Signature	99.98
Zhen et al. [43]	Spatial facial deformation+temporal filtering 60S, 10CV Keyframe, HMM	95.13
Zarbakhsh and Demirel [44]	4D FER, multimodal time series-geometric landmark-based deformations + NCFS 100S, 10CV AC-DTW	92.50
Li et al. [45]	Geometric images (DPI, NCI, SII) Multimodal 60S, 10CV	92.22
Yao et al. [46]	Texture and geometric scattering. 60S, 10CV Keyframe, MKL	90.12
Konstantinos Papadopoulos, et al 2021 [47]	Spatio-temporal graphs from facial landmarks	88.45
Yurtkan and Demirel [48]	3D geometrical facial feature point positions + SVM	87.50
Proposed	Geometric PSNR key+Angle Features + NCFS + Multi-SVM	**97.96**
Proposed	Geometric IMMSE key + Distance Features + NCFS + Multi-SVM	**99.10**

The top performing model is the distance + IMMSE method as shown in table 20, but when compared to other state of the art approaches, it performed very well. The recognition rate of our work is higher than the results given in [43-48]. A very study by Abd El Rahman et al 2020 [42] outperforms our method by 0.89 percent. Our proposed solution, on the other hand, relies on a set of landmarks rather than frontal view faces and has been designed to reduce complexity hence may be applicable in real time FER systems.

Conclusion and Future works

An integrated 3D face model has been introduced in this paper to generate the natural face from different facial expression. The model used 20 points of facial features which represent all sensitive points in the human face to adapt the face model on the input object which done by dividing the face into several parts and each part is individually modified, then calculate the width and height of each face region separately, the parts include the silhouette of the face model, left eye, right eye, nose and mouth of the face model. The measure that used to

evaluate the quality of model is Peak signal-to-noise ratio (PSNR), the result show that synthesis natural facial from different facial expression can be an alternative solution to many problems which face recognition systems have suffered, especially in database training. The ZY axis is used to extract Angle and Distance or magnitude features. The NCFS was utilized for feature selection, features were standardized, and Multi-SVM was used for classification, with the highest performing Distance + IMMSE model having an acceptable recognition rate of 99.10 percent. The future works could implement all three axis in 3D space at the same time, with real time entropy key frame selection.

Author Contributions: H.U did the literature survey, and the write-up. H.U and K.Y handled the revision. K.Y supervision. All authors contributed to the article and approved the submitted version of the manuscript.

Acknowledgments: This paper was supported by the Cyprus International University whose Library resources proved relevant to achieving this research objectives.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1- Chan, C. H., Kittler, J., & Messer, K. (2007, August). Multi-scale local binary pattern histograms for face recognition. In International conference on biometrics (pp. 809-818). Berlin, Heidelberg: Springer Berlin Heidelberg.
2- Malassiotis, S., & Strintzis, M. G. (2005). Robust face recognition using 2D and 3D data: Pose and illumination compensation. Pattern Recognition, 38(12), 2537-2548.
3- Igarashi, T., Moscovich, T., Hughes, J.F.: Spatial keyframing for performance-driven animation. In: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2005, pp. 107–115. ACM, New York (2005). https://doi.org/10.1145/1073368.1073383
4- H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Santiago, Chile, December 2015.
5- Zarbakhsh, P.; Demirel, H. 4D facial expression recognition using multimodal time series analysis of geometric landmark-based deformations. Vis Comput36, 951-965(**2020**). https://doi.org/10/1007/s00371-019-01705-7.
6- Sujono and A. A. S. Gunawan, "Face expression detection on Kinect using active appearance model and fuzzy logic," *Procedia Computer Science*, vol. 59, pp. 268–274, 2015.
7- McCarthy, P. A., Meyer, T., Back, M. D., & Morina, N. (2023). How we compare: A new approach to assess aspects of the comparison process for appearance-based standards and their associations with individual differences in wellbeing and personality measures. *PLoS One*, *18*(1), e0280072.
8- Banerji, S., Sinha, A., & Liu, C. (2013). New image descriptors based on color, texture, shape, and wavelets for object and scene image classification. *Neurocomputing*, *117*, 173-185.
9- K. Li, Y. Jin, M. W. Akram, R. Han, and J. Chen, "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy," *The Visual Computer*, vol. 36, no. 2, pp. 391–404, 2020.
10- M. Yu, H. Zheng, Z. Peng, J. Dong, and H. Du, "Facial expression recognition based on a multi-task global-local network," *Pattern Recognition Letters*, vol. 131, pp. 166–171, 2020.
11- Thai Son Ly, Nhu-Tai Do, Soo-Hyung Kim, Hyung-Jeong Yang, Guee-Sang Lee, A novel 2D and 3D multimodal approach for in-the-wild facial expression recognition, Image and Vision Computing, Volume 92, 2019, 103817, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2019.10.003.
12- Asit Barman, Paramartha Dutta, Facial expression recognition using distance and texture signature relevant features, Applied Soft Computing, Volume 77, 2019, Pages 88-105, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2019.01.011.
13- Fengyuan Wang, Jianhua Lv, Guode Ying, Shenghui Chen, Chi Zhang, Facial expression recognition from image based on hybrid features understanding, Journal of Visual Communication and Image Representation, Volume 59, 2019, Pages 84-88, ISSN 1047-3203, https://doi.org/10.1016/j.jvcir.2018.11.010.
14- Dmytro Derkach, Federico M. Sukno, Automatic local shape spectrum analysis for 3D facial expression recognition, Image and Vision Computing, Volume 79, 2018, Pages 86-98, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2018.09.007.
15- Walid Hariri, Nadir Farah, Recognition of 3D emotional facial expression based on handcrafted and deep feature combination, Pattern Recognition Letters, Volume 148, 2021, Pages 84-91, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2021.04.030.
16- Muzammil Behzad, Nhat Vo, Xiaobai Li, Guoying Zhao, Towards Reading Beyond Faces for Sparsity-aware 3D/4D Affect Recognition, Neurocomputing, Volume 458, 2021, Pages 297-307, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.06.023.
17- Nerea Aranjuelo, Sara García, Estíbaliz Loyo, Luis Unzueta, Oihana Otaegui, Key strategies for synthetic data generation for training intelligent systems based on people detection from omnidirectional cameras, Computers & Electrical Engineering, Volume 92, 2021, 107105, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2021.107105.
18- Costa B.F., Esperança C. (2020) Motion Capture Analysis and Reconstruction Using Spatial Keyframes. In: Cláudio A. et al. (eds) Computer Vision, Imaging and Computer Graphics Theory and Applications. VISIGRAPP 2019. Communications in Computer and Information Science, vol 1182. Springer, Cham. https://doi.org/10.1007/978-3-030-41590-7_3
19- Terra, S. C. L., & Metoyer, R. A. (2007). A performance-based technique for timing keyframe animations. *Graphical Models*, *69*(2), 89-105.
20- L. Yin, X. Chen, Y. Sun, T. Worm and M. Reale,"A high-resolution 3D dynamic facial expression database,"in Proc. 8th IEEE Int. Conf. on Automatic Face & Gesture Recognition, Amsterdam, Netherlands, pp. 1–6, 2008.
21- S. R. Jannat, D. Fabiano, S. J. Canavan and T. J. Neal,"Subject identification across large expression variations using 3D facial landmarks,"in Int. Conf. on Pattern Recognition Workshops, Milan, Italy, pp. 2122, 2020.
22- John, A., Abhishek, M. C., Ajayan, A. S., Sanoop, S., & Kumar, V. R. (2020, August). Real-time facial emotion recognition system with improved preprocessing and feature extraction. In *2020 Third international conference on smart systems and inventive technology (ICSSIT)* (pp. 1328-1333). IEEE.
23- Qi, Y., Yang, Z., Sun, W., Lou, M., Lian, J., Zhao, W., ... & Ma, Y. (2022). A comprehensive overview of image enhancement techniques. *Archives of Computational Methods in Engineering*, *29*(1), 583-607.

24- Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; pp. 511–518.
25- Smriti Tikoo, Nitin Malik. Detection of Face using Viola Jones and Recognition using Back Propagation Neural Network. https://arxiv.org/abs/1701.08257v1
26- Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang and M. J. Rosato, "A 3D facial expression database for facial behavior research," 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006, pp. 211-216, doi: 10.1109/FGR.2006.6
27- Milad, A., & Yurtkan, K. (2023). RETRACTED ARTICLE: An integrated 3D model based face recognition method using synthesized facial expressions and poses for single image applications. *Applied Nanoscience*, *13*(3), 1991-2001.
28- A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," 2010 20th International Conference on Pattern Recognition, 2010, pp. 2366-2369, doi: 10.1109/ICPR.2010.579.
29- Bommisetty, R.M., Prakash, O. & Khare, A. Keyframe extraction using Pearson correlation coefficient and color moments. Multimedia Systems 26, 267–299 (2020). https://doi.org/10.1007/s00530-019-00642-8
30- Dang C, Radha H. RPCA-KFE: Key frame extraction for video using robust principal component analysis. IEEE Transactions on Image Processing (TIP). 2015;24(11):3742-3753
31- Mentzelopoulos M, Psarrou A. Key frame extraction algorithm using entropy difference. In: Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, MIR; 15–16 October, 2004; New York, NY, USA. pp. 39-45
32- Ujwalla Gawande, Kamal Hajari and Yogesh Golhar (February 12th 2020). Deep Learning Approach to Key Frame Detection in Human Action Videos, Recent Trends in Computational Intelligence, Ali Sadollah and Tilendra Shishir Sinha, IntechOpen, DOI: 10.5772/intechopen.91188. Available from: https://www.intechopen.com/chapters/71081
33- Natan (2021). Fast 2D peak finder (https://www.mathworks.com/matlabcentral/fileexchange/37388-fast-2d-peak-finder), MATLAB Central File Exchange. Retrieved May 26, 2021.
34- Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing. Addison-Wesley, New York, 1992.
35- Yang, W., K. Wang, W. Zuo. "Neighborhood Component Feature Selection for High-Dimensional Data." Journal of Computers. Vol. 7, Number 1, January, 2012.
36- Raju, V.N.G.; Lakshmi, K.P.; Jain, V.M.; Kalidindi, A.; Padma, V. Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 729-735, doi: 10.1109/ICSSIT48917.2020.9214160.
37- Scholkopf, B.; Smola. *A. Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond, Adaptive Computation and Machine Learning.* Cambridge, MA: The MIT Press, **2002**.
38- Christianini, N.; Shawe-Taylor, J.C. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press, **2000**.
39- Pal, K.; Patel, B.V. Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques. *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020, pp. 83-87, DOI: 10.1109/ICCMC48092.2020.ICCMC-00016.
40- Susmaga, R. Confusion Matrix Visualization. In: Kłopotek M.A., Wierzchoń S.T., Trojanowski K. (eds) Intelligent Information Processing and Web Mining. Advances in Soft Computing, 2004, vol 25. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39985-8_12.
41- Guarnera, M., Hichy, Z., Cascio, M. I., & Carrubba, S. (2015). Facial Expressions and Ability to Recognize Emotions From Eyes or Mouth in Children. Europe's journal of psychology, 11(2), 183–196. https://doi.org/10.5964/ejop.v11i2.890
42- Abd El Rahman Shabayek, Djamila Aouada, Kseniya Cherenkova, Gleb Gusev, and Bj¨orn Ottersten. 3d deformation signature for dynamic face recognition. In 45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020), Barcelona 4-8 May 2020, 2020.
43- Zhen, Q., Huang, D., Drira, H., Amor, B.B., Wang, Y., Daoudi, M.: Magnifying subtle facial motions for effective 4D expression recognition. IEEE Trans. Affect. Comput. (2017). https://doi.org/ 10.1109/TAFFC.2017.2747553
44- Zarbakhsh, P; Demirel, H. 4D facial expression recognition using multimodal time series analysis of geometric landmark-based deformations.The Visual Computer (2020) 36:951–965. https://doi.org/10.1007/s00371-019-01705-7
45- Li,W., Huang, D., Li, H.,Wang,Y.: Automatic4D facial expression recognition using dynamic geometrical image network. In: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), pp. 24–30. IEEE (2018)
46- Yao, Y., Huang, D., Yang, X., Wang, Y., Chen, L.: Texture and geometry scattering representation-based facial expression recognition in 2D+3D videos. ACM Trans. Multimed. Comput. Commun. Appl. 14(1s), 18:1–18:23 (2018)
47- Konstantinos Papadopoulos, Anis Kacem, Abdelrahman Shabayek and Djamila Aouada. Face-GCN: A Graph Convolutional Network for 3D Dynamic Face Identification/Recognition. arXiv:2104.09145v2 [cs.CV] 20 Apr 2021.
48- Yurtkan, K.; Demirel, H. Feature selection for improved 3D facial expression recognition. Pattern Recognit. Lett. 2014, 38, 26–33.