Determining Mood from Facial Expressions
Progress Report

1. Project Description

Nowadays, technology is an everyday part of people's lives and plays a huge role in our development as a human race. Humans spend the majority of their time interacting with various devices such as computers, smartphones, etc. Research shows, most people's work involves working on a computer. However, when compared with human communication, interaction with currently available software interfaces is non-verbal and primitive. If computers have the ability to determine human emotions through facial expression recognition, then human-computer interaction (HCI) would be significantly improved [1]. The communication between a man and machine is becoming one of the fastest developing areas and advancements there would consequently benefit human-robotic interaction (HRI) and many other fields.

Facial expressions play a significant role in human communication [1-5]. Studies have shown human emotions expressed through facial expressions contribute 55% of emotional expression as opposed to 35% for the vocal part and only 7% for the verbal part [2]. This would imply that facial expressions play the major part in communication between people. Before diving in it's essential to make the distinction between facial expression recognition and emotion recognition. Humans express emotions in multiple ways through facial expressions, emotion of voice, body language, whereas facial expression recognition is solely based on visual information which will be the focus of this project [2].

Humans have the ability to interpret facial expressions with little effort. Attempting to develop an automated system that will have that ability is a rather difficult task that presents a few challenges. In order to achieve humanlike results a system must be able to detect an image segment as a face, extract all the information regarding the facial expression that the face is trying to express and finally classify the expressions in categories of emotion [2-5]. With the current advancements in technology and the various existing techniques available for solving this problem, a system that can accurately guess human emotions in laboratory conditions has been developed with accuracy around 97%, however when tested in an unconstrained real-time environment it doesn't perform as well with accuracy as low as 50% [4]. Analysis of facial expressions can be incredibly challenging in a real-time environment due to constantly occurring small transient changes. The main three challenges that arise in an unconstrained real-world environment are illumination variation, head pose and subject dependence [4]. Different illumination levels can directly affect the accuracy of face feature extraction. In a controlled environment under laboratory conditions the head pose of a subject is usually frontal. Yet, in a real-world environment due to a subject's transient movements of the head, sometimes a frontal view might not be available thus presenting a great challenge. Subject dependence has to do with the fact a Facial Expression Recognition (FER) system can only recognize pre-trained faces. To overcome this challenge, a large dataset and a reliable classifier would be required [4].

As mentioned above, developing a FER system has three main parts: pre-processing, feature extraction and image classification [2-5]. Pre-processing is a process which can substantially increase the accuracy of a FER system by applying different cropping and scaling techniques to the images and preparing the data for feature extraction [5]. Feature extraction is a process which dimensionally reduces initially large, raw data to a more manageable dataset by combining different features, while accurately and completely representing the original dataset [6, 7]. After the necessary features are extracted, images can then be classified by emotions at the last stage [5].

The possibilities for application of FER in various areas are endless. As mentioned above, the main areas which could be drastically improved by an effective FER system are HCI and in particular HRI. Having the ability to determine people's emotions would enable an HRI systems to simulate a more natural and friendly interaction with humans, thus improving almost every industry and occupation where computers are used. For example, they can be used in a medical care to detect various mental states through facial expression analysis [8, 9]. By being able to distinguish between the different healthy emotional states such as happiness, satisfaction, and unhealthy emotional states (anger, sadness, frustration) automated systems can improve quality of life. In addition, FER could drastically help mentally ill patients by exploring their behavioural patterns. Researchers [10] and [11] have successfully shown that by investigating facial expressions of patients, they can identify emotional conflicts and as a result, mental disorders such as anxiety or autism can be diagnosed. Furthermore, study has shown people are more likely to trust a machine rather than another person [12]. This opens up a great application of FER in automated counselling systems. Fatigue detection is another area which could significantly benefit by a FER system [13]. According to NHTSA organization statistics, the main factor for road accidents is driver fatigue [13]. With the ability to read human emotions vehicles would be able to prevent the that alerting the drivers. In a similar way, FER can help prevent accidents at the workplace by monitoring employees fatigue levels especially for machine operators. Another area where FER could be applied is teaching [14] - currently automated tutoring systems are becoming increasingly popular allowing anyone with computer access to learn from the comfort of their own home. An automated tutoring system with FER would be able to adapt individually to every student and determine how well they're responding to the material. This would allow the system to make decisions and adjustments, which would make the process seem less artificial, more pleasant, thus making it a more effective learning process on an individual level. Computer graphics is an area which is already largely benefitting from efficient FER systems. The modelling of a human face by precisely parameterizing the geometry of the face and muscle motions is done by FER systems [15]. Newer technologies such as Augmented Reality (AR) and Virtual Reality (VR) also apply effective FER systems to achieve a more natural, effortless communication with humans. Other areas where FER can be applied are automated surveillance systems [48], behaviour prediction [47], lie detection [28], and music/lighting for mood, etc.

2. Related Work

As mentioned above, HCI is an increasingly growing field that attracts a lot of people. Currently existing techniques achieve very high results in lab-controlled environment with accuracy above 95%. However, due to the challenges discussed earlier, in an unconstrained real-world environment, FER systems achieve around 50% accuracy. One of the most studied topics in computer vision is face detection and more recently facial expression recognition. As a result, a large number of algorithms have been developed to overcome various challenges such as illumination variation, pose, occlusion, and others. The following sections describe the most popular and effective techniques used today along with performance comparison for the three stages in a FER system: pre-processing, feature extraction and image classification.

A) Pre-processing

In image processing it's quite common nowadays to work with large datasets, meaning the majority of the images will have different graphical properties. The model trained on such data will most likely suffer in performance [16]. In order to overcome that challenge different pre-processing techniques are applied to the data. Pre-processing is a process which prepares the data for feature extraction and can majorly improve performance of FER [17]. Image pre-processing in this case involves different cropping, scaling, contrast adjustment techniques to improve the data [5]. In almost all cases for an accurate FER system, various pre-processing techniques must be applied to the data as large disparities between image parameters such as illumination levels, contrast, size can drastically decrease the performance, hence why pre-processing is a key stage. The most popular pre-processing methods implemented for this part are Normalization [18], Localization [19, 20], Region of Interest (ROI) [21]. Normalization is the most popular technique, which can be used for reduction of illumination variation in different images and reduction of face image variation, thus providing more clarity to the images. Moreover, normalization is used for extraction of different features such as eyes, mouth, which helps develop a more robust FER system when it comes to personality differences. In addition to normalization, localization is another pre-processing method that can be applied [19,20]. Localization is used for detecting faces within an image and isolating them in face images. Its main purpose is spot the location and size of the face image. ROI are samples of data, isolated for a particular purpose. In image processing, ROI segmentation is an important method for identifying the different parts of a face such as mouth, nose, eyebrows [21, 26]. Another less popular image processing technique is Histogram Equalization [19,20], which is used for adjusting the contrast in images. Histogram the graphical representation of the colour value distribution of an image [22]. This method uses the image's histogram to spread out the most frequent intensity values. This means local areas with low contrast gain a higher contrast. This technique is useful for images with similar tonal distribution in the background and foreground such as x-ray images [22]. In FER pre-processing, ROI is the most popular technique used as it precisely detects all the different face parts that people use to express emotions [21,26].

B) Feature Extraction

Feature Extraction is the most important, second step in designing a FER system. Large datasets, which are increasingly common, usually contain thousands of features furthermore if the number of features is greater than the number of observations in the dataset, it is highly likely the model would suffer from overfitting [6,7,24]. Overfitting refers to a dataset that has been modelled more than necessary to the point where it actually degrades the performance of the machine learning model. As a result, the model would only learn to perform well on the dataset that it's been trained on. To overcome this obstacle, it's vital to apply regularization or dimensionality reduction techniques –feature extraction [24]. Other than overfitting risk reduction, feature extraction is especially useful where the images are large in size and a more compact feature representation, stripped of any redundant information, is required for further processing. This also necessitates less computing power and respectively improves training times [24]. During this stage, the graphical data of an image is depicted as implicit numerical data describing the texture properties [5], which is then given as input to the classification algorithm. Feature extraction techniques are divided into five types: texture-feature based, edge based, global and local feature-based, patch-based and geometric shape based [5].

With texture feature-based algorithms [20, 22, 33, 34, 35, 37] all features are formed from properties defining the texture of an image. Textures are one of the most important characteristics of an image used to classify and recognize objects and find similarities between images. Scale Invariant Feature Transform (SIFT) [22, 34] has proved to be a very powerful technique for object detection/recognition, however, SIFT might not be optimal for analysing face images [23]. Keypoints-Preserving-SIFT (KPSIFT) includes all the initial key points as features. Partial-Descriptor-SIFT (PDSIFT), includes all key points detected at large scale and uses a partial descriptor for identifying face boundaries. Both prove to be more efficient than the original SIFT [23]. Gabor filters are linear orientation-sensitive filters used for edge and texture analysis that can extract local features in frequency and spatial domain [25-27]. Local Binary Patterns (LBP) is a simple, efficient and robust local descriptor that has proven to do well in various domains such as texture analysis, facial expression recognition, and facial recognition [29]. LBP represents pixels as binary numbers by thresholding the neighbouring pixels and its most important property is its robustness to monotonic gray-scale changes such as illumination variations [30]. Combined with its computational simplicity, LBP is one of the more popular feature extraction techniques used in FER systems [20,22]. For multi resolution approaches, LBP is combined with Three Orthogonal Planes (TOP) method [31]. Weighted Project Based LBP (WPBLBP) [35] is an extended LBP extraction that's formed from instructional domains for which the LBP is extracted. After that, depending on the importance of these instructive regions, the extracted features are weighted [5]. Gaussian Laguerre (GL) wavelets have powerful frequency extraction capabilities for extracting features of facial expressions [16]. In comparison, GL uses a single filter instead rather than multiple ones with Gabor filters [5], which means Gabor Filters require significantly more computational power. In addition, Vertical Time Backward (VTB) method takes out the shape related attributes of facial components. This makes it really effective on spatiotemporal planes. Spatiotemporal derivatives are usually contained in images produced by catadioptric sensors, which contain a significant amount of radial distortion and variation in inherent scale [32]. Weber Local Descriptor (WLD) is another texture-based method for extracting features, which consists of differential excitation component and orientation component, that contains abundant local information [33]. In most cases, WLD uses Supervised Descent Method (SDM) to estimate the distance between various components of the face [5]. WLD performs better than LBP while still being as computationally efficient as LBP. As mentioned, SIFT is a sparse descriptor, whereas WLD is a dense descriptor computed for every pixel and depends on the magnitude of the centre pixel's intensity [34]. Lastly, Discrete Contourlet Transform (DCT) [36] method is a combination of a multi-scaled Laplacian pyramid and multi-directional filter banks. Compared to one-dimensional transforms such as Fourier and wavelet DCT is a two-dimensional feature extraction method that can capture geometrical structures of an image that are key to visual information. Using multidirectional filter banks with DCT multi-resolution and directional image representation contour segments, hence the name contourlet transform [36].

Edge-based feature extraction methods [38 - 43] are used for object recognition where colour or texture cannot be used as a cue for recognition. Instead, the distinctive features of such objects are edges and geometric locations between them. As mentioned above, one of the biggest challenges when it comes to image retrieval is illumination variation. Compared to texture-based feature extraction methods, techniques that use edge information instead are partially illumination invariant and require less memory, which makes them a preferred choice in some cases. Line Edge Map (LEM) [39] can be used to construct a compact face feature for face coding and recognition. The investigation [40] shows that with LEM the extracted features of the face are discriminative and non-discriminative. Active Shape Model using GPU(GASM) [41, 42] is a popular statistical model for object localization based on the famous Snake algorithm. Compared to CPU, the graphics-processing unit allows for substantially bigger facial feature extractions in video or image sequences. In fact, the acceleration improvement is so significant the GPU reported a performance boost 48 times greater than compared to CPU implementation [41]. Main advantage of the ASM compared to other feature extraction algorithms is the model can only deform in ways learnt from the training set, meaning it can deform considerably and maintain specificity to the object intended for representation at the same time. Histogram of Oriented Gradients (HOG) [43] are an efficient descriptor for object detection and recognition. They are generally used in computer vision, pattern recognition and image processing for detecting and recognizing graphic objects such as faces. HOGs are especially efficient in detecting faces with occlusions, pose and illumination variation, because of the robust feature set, in which face features are extracted in a regular grid.

A different approach combines techniques for extracting global and local features. Principal Component Analysis (PCA) [44, 49] is a common method used in statistical pattern recognition and signal processing invented back in 1901. PCA extracts global and low-dimensional features about a pattern in an image. The pattern often contains redundant information. In order to get rid of that redundancy, while still preserving the key descriptive information, the pattern is matched to a feature vector. As a result, the extracted features are used to discriminate between input patterns in an image. In comparison, Independent Component Analysis (ICA) [45, 46, 49] is a novel statistical technique in machine learning and signal processing used to extract local features by using multi-channel observations. ICA aims to find linear projections of data features that maximize their mutual independence. Stepwise Linear Discriminant Analysis (SWLDA) [50] is another efficient technique for extracting localized features. SWLDA employs forward and backward regression models to extract a small set of features. During forward regression, the most correlated features are isolated on the basis of defined class labels or F-test values, while the least significant are removed from the regression model during backward regression [29]. The main advantages of SWLDA over other techniques are its computational simplicity, predictive ability and its performance doesn’t suffer from illumination variation.

The first techniques used for feature extraction in facial expression recognition were geometry-based [19, 51, 52]. This means the extracted features we were based around the shape of the face and its parts – mouth, eyebrows, nose, rather than describing the texture of the face. Most geometric-based techniques employ Active Appearance Model (AAM) or variations of it to localize and track a dense set of facial points [51]. This means that the shape of the face is extracted by monitoring and combining these facial points in different ways as the expression evolves. Mainly because of that, these methods perform better on a dataset comprised of videos or image sequences rather than individual images. Local Curvelet Transform (LCT) [19] is an effective feature extraction method that achieves localization in frequency and time domain. Some of the points and lines on a face can be better extracted than wavelet transform which deals with point singularities [52]. One disadvantage of LCT is that the features extracted are usually quite large and other dimensionality reduction techniques must be applied to overcome this issue [19, 52]. Patch-based feature extraction techniques are less used for facial expression recognition that extract patches of the image. After that, the patches are classified to a specific class and the whole image is classified based on the individual patches. This approach is useful when the image architype is too complex [53].

[20, 22, 33, 34, 35, 37] have shown that when it comes to designing a FER system, for the feature extraction stage texture - based techniques yield best results, since appearance-based extracted features have more significance than others. More recently developed similar techniques are Discrete Wavelet Transform (DWT) [54], Local Directional Number (LDN) Pattern [55], Local Directional Ternary Pattern (LDTP) [56] and KL-transform Extended LBP (KELBP) [57]. It’s also important to note that in recent years, various dimensionality reduction techniques have been applied to features that have high dimensional vectors, in addition, to better determine the significance of the features similarity scores and various algorithms for e.g. Adaptive Boosting (AdaBoost).

C) Image Classification

Image Classification is the last phase in developing a FER system where the output of the feature extraction process is fed as an input to a classification algorithm or classifier and during this stage expressions are categorized as emotions such as happiness, sadness, anger, fear, etc. Similar to feature extraction stage, because of the fastest development and interest of FER in computer vision, a lot of different classifiers have been developed over the years. A distance-based classifier is Euclidean distance metric [58]. The training and datasets for one subject consist of images with different expressions shown. When the model has to classify a certain image, Euclidean distance is calculated between the points on the test image and the points on the training images for the same subject extracted during feature extraction. The expression shown on the test image is classified as the expression shown on the training image, for which the minimum Euclidean distance is found. Euclidean distance is more suitable for static images due to its ambiguity for real-time or robust images [58]. Another distance-based classifier is Minimum Distance Classifier (MDC) [59]. Although its classification accuracy is usually lower than more complex classifiers such as Convolutional Neural Networks (CNN) or Support Vector Machine (SVM), MDC is still used in various areas of pattern recognition due to its computational simplicity and fast execution time. With MDC, an unknown pattern is classified to a category to which the nearest prototype to the pattern belongs [59, 60]. K-Nearest Neighbours (KNN) classifier [16, 61, 62] is another simple, distance-based algorithm that classifies objects based on nearest training examples in the feature space. The object is classified according to adjacent object classes, meaning the objects is assigned to most common class among its k nearest neighbours (k > 0), so if the value of k is one, then the object is classified to the class of that nearest neighbour. After this stage, images are converted to vectors of fixed length with real numbers and a distance-based algorithm such as MDC or Euclidean distance is used to classify the whole image [61,62]. KNN can be defined as lazy learning or instance – based learning in which the function is only calculated locally, and evaluation is postponed until classification. In this way, KNN can be used to determine the significance on the k-nearest neighbours of an object by weighing their contribution. Common factors for why all distance-based algorithms are applied for classification is simplicity, speed, and ease of understanding. A more complex classification technique is Hidden Markov Model (HMM). HMMs are probabilistic models which consist of two random processes – Markov Chain comprised of several countable states, and a second process that defines the output transitions and corresponding emissions [50, 63]. HMMs are often used in speech recognition and more recently for FER in image sequences as well, since they can model temporal dependencies [50]. Support Vector Machine (SVM) [21, 64, 65] is a technique that combines related supervised methods used for classification and regression. SVMs use machine learning techniques which use a hypothesis space of linear functions in high dimensional feature space to maximize prediction accuracy [64]. In fact, they are one of the most efficient classification techniques for dimensionally large data [21]. SVM systems are computationally complex, but perform as well as sophisticated neural networks, delivering high accuracy in FER systems, hence why they're really popular classification technique for pattern recognition-based problems and regression-based applications. For real-world FER systems Extreme Learning Machine (ELM) [66 - 68] is feed-forward neural network technique classification technique with only one hidden layer of nodes originating from the study of single hidden layer feedforward neural networks [66]. Unlike conventional neural networks, with ELM there's no need of tuning the hidden layer for weight adjustment, which makes them substantially faster and greatly reduces processing time for the data. In comparison to SVM, LBP, KNN, Extreme Machine Learning is an efficient classification method for real-world FER systems where the data is noisy and imperfect with a lot of constant transient changes. In addition, Online Sequential Extreme Learning Machine (OSELM) originates from ELM, where the data is split and learned in individual batches, for which the output weights are constantly updated using a Recursive Least Squares (RLS) algorithm. The main advantage of OSELM is that it can provide better generalization performance at a much greater learning speed compared to other classification techniques [68]. A popular rule-based classifier used for FER is ID3 Decision Tree (DT) [69]. ID3 builds a decision tree from a fixed set of examples, which is later used to classify other data. From the decision tree predefined rules are extracted to produce competent rules [5]. Compared to other algorithms, ID3 is robust to noise and has an easily interpretable tree structure (if-then-else), meaning it can be extended to multiple output values. Learning Vector Quantization (LVQ) [70] is a supervised, prototype-based, artificial neutral network (ANN) algorithm that supports both two-class and multi-class classification problems. Compared to KNN, LVQ allows for choosing how many training instances to hang onto and learns exactly what those instances are supposed to look like, rather than having to hold onto all training instances. Finally, probably the most used technique in image processing and in particular FER use neural networks. Other than the ones already discussed above, the most popular ones for FER are Bayesian neural network, Deep Neural (DNN), Network, Artificial Neural Network (ANN) and Convolutional Neural network (CNN).

The main problem with image classification is to find useful features from the feature extraction stage and this exactly what neural networks are great at as they can automatically create and select the most important useful features along with having the ability to learn extremely complex classification models. Some types of neural networks are also able to extract useful features invariant to transformation such as transformation, transposition, scaling, relocation etc.

Table 1. taken from [5]

Performance analysis of FER techniques.

Author name, year	FER method name	Database name	Complexity	Recognition accuracy (%)	No. of expressions recognized	Major contribution	Advantages
Gao et al. (2003)	LEM, dLHD	AR	Less	86.6	3	Oriented structural features are extracted	Suitable for real time applications
Noh et al. (2007)	Action based, ID3 decision tree	JAFFE	Less	75	6	Facial features are discriminative & non discriminative	Cost effective in speed and accuracy
Bashyal et al. (2008)	GF, LVQ	JAFFE	Less	88.86	Not reported	LVQ performs better recognition for fear expressions	Better accuracy for fear expressions
Zhao and Pietikäinen (2009)	GASM, SVM	CK	High	93.85	6	Adaboost learning for multi resolution features	Flexible feature selection
Song et al. (2010)	LBP-TOP, SVM	JAFFE, CK Realtime	Less	86.85	7	Detection of facial features point motion & image ratio features	More robust to lighting variations
Wang et al. (2010)	SVM	JAFFE	Less	87.5	Not reported	DKFER for emotion detection	More efficient emotion detection
Zhang et al. (2011)	Patch based, SVM	JAFFE, CK	Less	82.5	6	Capture facial movement features based on distance features	Effective recognition performance
Poursaberi et al. (2012)	GL Wavelet, KNN	JAFFE, CK, MMI	Medium	91.9	6	Extraction of texture and geometric information	Wealthy capability for texture analysis
Ji and Idrissi (2012)	LBP, VTB, Moments, SVM	CK, MMI	Medium	95.84	6	Extraction of spatial temporal Features	Effective image based recognition
Taylor et al. (2014)	PCA, ICA, HMM	Own	Less	98	6	Multilayer scheme to conquer similarity problems	High accuracy with own dataset
Owusu et al. (2014)	GF, MFFNN	JAFFE, Yale	High	94.16	7	Feature selection based on Adaboost	Lowest computational cost
Demir (2014)	LCT, OSLEM	JAFFE, CK	High	94.41	7	Extraction of statistical features mean, entropy and S.D	Reliable algorithm for recognition
Zhang et al. (2014)	GF, SVM	JAFFE, CK	Less	82.5	7	Template matching for finding similar features	High robustness & fast processing speed
Dahmane and Meunier (2014)	HOG, SVM	JAFFE	High	85	7	SIFT flow algorithm for face Alignment	Robust to rotation, occlusion & clutter
Mahersia and Hamrouni (2015)	Streerable pyramid, Bayesian NN	JAFFE, CK	Less	95.73	7	Statistical features are extracted from the steerable representation	Robust features & achieve good results
Hernandez-matamoros et al. (2015)	Gabor function, SVM	KDEF	Less	99	Not reported	Segmentation of face into two Regions	High performance with low cost
Happy et al. (2015)	LBP, SVM	JAFFE, CK+	Less	93.3	6	Facial landmarks lip and eyebrow corners are detected	Lower computational complexity
Biswas (2015)	DCT, SVM	JAFFE, CK	Less	98.63	6	Each image is decomposed up to fourth level	Very fast & high accuracy
Siddiqi et al. (2015)	SWLDA, HCRF	JAFFE, CK+,MMI, Yale	High	96.37	6	Expressions are categorized into 3 major categories	High accuracy
Cossetin et al. (2016)	LBP, WLD, Pairwise classifier	JAFFE, CK, TFEID	Less	98.91	7	Each pair wise classifier uses a particular subset	High accuracy & less computation power
Salmam et al. (2016)	SDM, CART	JAFFE, CK	Less	89.9	6	Decision tree for training	Improved recognition accuracy
Kumar et al. (2016)	WPLBP, SVM	JAFFE, CK+, MMI	Medium	98.15	7	Extraction of discriminative features from informative face regions	Lower misclassification
Hegde et al. (2016)	GF, ED, SVM	JAFFE, Yale	Less	88.58	6	Projects feature vector space into low dimension space	Improves the recognition efficiency

Literature review summarized at table 1 shows that the datasets used for FER systems are JAFFE, CK/+, MUG, TFEID, Yale, AR, MMI, KDEF, MUG. The facial expressions recognized are usually 6: happiness, sadness, anger, disgust, surprise, neutral – however, with LEM only 3 facial expressions are recognized while some FER systems can recognize 7 expressions successfully with contempt as well. Accuracy performance is outstandingly high in lab – controlled environment with ROI segmentation at pre-processing stage and Gabor function for feature extraction with SVM classifier giving best results 99% [26].

3. Project Specification

The goal of this project is to design, document, build and test a system that does real-time FER. As mentioned above, this includes three stages. Firstly, different techniques will be applied to the training data in order to normalize it. Then I will be developing a CNN model for the feature extraction and image classification stages. The model's accuracy, error rate and complexity will be evaluated with k-fold cross validation. The next part will involve comparing the developed model's performance with the most popular models used for image classification in order to determine which technique yields the best results for FER. Again, k-fold cross validation will be used for this part. After performance comparison and evaluation, I will choose the best technique and embed it in a desktop/mobile application. The app will serve as a wrapper of the FER system. Its main functionality will be to extract the users' faces, which will then be passed to the

FER system for analysis. Once classified, the output will be presented to the user/s. Since facial expression recognition must happen in real time the app will have a webcam on which the user/s faces will be detected and extracted. The detected faces will be surrounded with square boxes. The output from the FER system will be presented as a label attached to the square box surrounding the users' face. For every frame per second the facial expressions will be re-evaluated, and the results presented to the user. From this specification the following functional requirements follow:

FER

- Face detection and emotion classification in real-time.
- Detect and classify all human faces visible on the webcam.
- Developed FER model compared and evaluated against existing models with k-fold cross validation.
- 6 or more emotions recognized - happiness, sadness, anger, disgust, surprise, neutral.

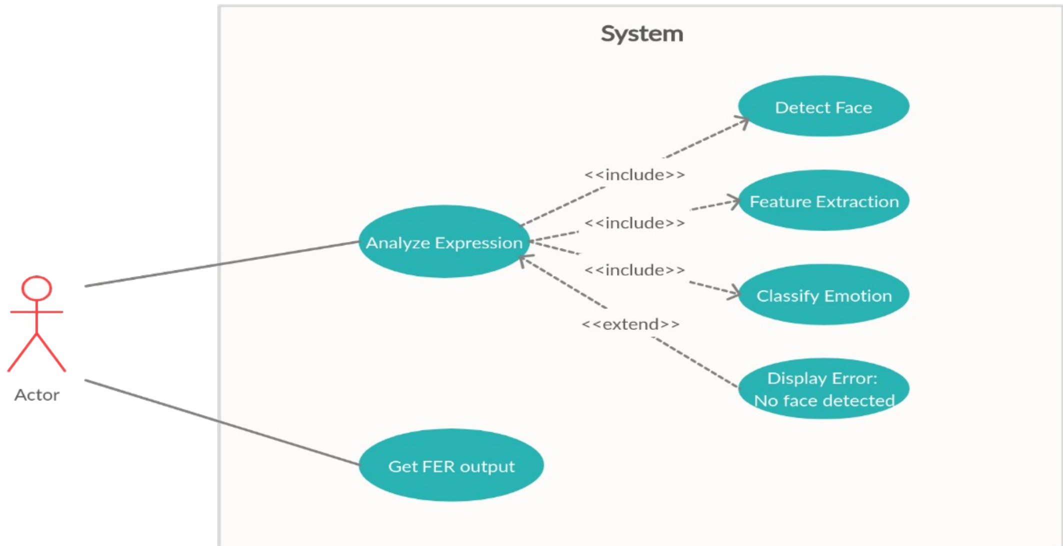
App

- Have a streaming webcam where face detection will happen.
- Re-evaluation of facial expressions for every frame per second.
- The output from FER presented to the user for every frame per second.

Additional requirement (if time allows)

- Allow the user the option to upload an image to be analysed.

The following use case diagram summarises the user/s' interaction with the application:

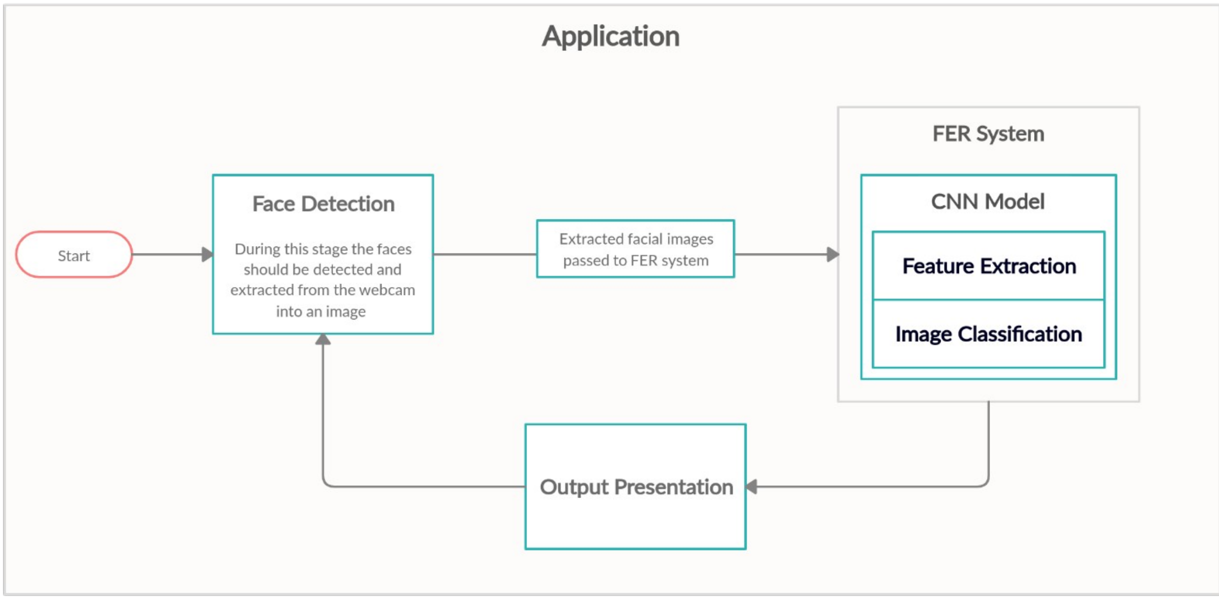


A) Design

As for the FER model design, after literature review, I've decided the best approach would be to develop a deep convolutional neural network model. Because of its popularity in image processing and its ability to learn highly complex models I believe it would be the most appropriate way to go. In addition, using the pre – trained ResNet50 model during prototype development yields 95% accuracy, which proves deep CNN models can give very good results for FER. Moreover, the feature extraction stage, which as mentioned above is the most significant stage of developing a FER system, is in a way performed in an unsupervised, automatic way with CNN and no special algorithms have to be applied in order to extract useful feature. This is called feature learning. CNN extracts features by constructing its own feature map using filters or kernels represented by 3x3 matrices. This is done by the convolution layers of a CNN, in which the network effectively uses adjacent pixel information to efficiently down sample an image, which preserves spatial information about features and that's the key thing about CNN. The preservation of spatial information compared to a pixel vector algorithm for example means CNN will perform much better during the classification stage. A CNN consists of 4 main layers for feature extraction – convolutional layer, Relu-activation layer or also called non-linear, pooling and fully connected layers and a classifier in the end. The convolutional layer of a CNN consists of three pieces: an input tensor, a filter tensor and the output tensor.

An input tensor can be considered as a 2d matrix of the pixels of an image, each pixel being a feature of the input. The input tensor is a really simple 2d matrix for black and white images and for colour images the input is a 2d matrix of size 3 vectors corresponding to the RGB components for each pixel. A filter tensor or also called the kernel as mentioned above is essentially what allows to capture patterns in an image by capturing snapshots of the input matrix via dot-products. The filter tensor is a smaller matrix than the input. The output tensor then stores all the snapshots taken by the filter tensor. The snapshots are a way to summarize the information on a part of the input. The output tensor is essentially a representation of the input that has been summarized by the filter tensor. Once snapshots from the whole image are saved in the output tensor, the convolution layer is finished, and its output is fed as an input to the next convolutional layer. The convolution layer consisting of these 3 pieces is how the network looks for characteristics such as boundaries or curvatures, distinct features of the base level. The whole CNN network will usually consist of several convolutional layers mixed with pooling and non-linear layers. A non-linear layer has an activation function that brings non-linear property applied after the convolution layer in a CNN network. Approximation power does not increase for linear networks by adding more layers or “going deeper”, unlike for non-linear networks. The universal approximation theorem proves that a feed-forward neural network (FNN) with a single hidden layer, can approximate any continuous function for input sets with a fixed size [71]. One of the conditions for that theorem to be valid is that the neural network must be a composition of non-linear activation functions. Therefore, the non-linear layers are added to the convolutional neural network. The pooling layer comes after the non-linear layer and performs down sampling of an image. As a result, the volume of the image is reduced and if some features for example curvatures, have been previously identified in the convolution operation then image is compressed to less detailed pictures. After a series of convolutional, non-linear and pooling layers a fully connected layer is attached to the network. This layer takes the output information from the convolutional networks that can then be classified with a classifier function.

The overall application design can be observed on the following diagram:



B) Implementation and technology proposals

The FER model will be developed with FastAI library built on top of PyTorch. FastAI is a powerful machine learning library for developing deep learning models. The library has different pre-trained models such as Resnet, Densenet Squeezenet, Alexnet that can be easily trained for different classification problems. The model developed will be trained on Google Cloud Platform in order to exploit the GPU on it and expedite the training process. For face detection in real-time OpenCV (Open source computer vison). OpenCV can be used to accurately detect faces from every frame and extract them in image files. The images will then be classified by the best FER implementation model. In addition, I will be testing all available models on the FastAI library and after cross-validating all of them I’ll be comparing their performance with my own model. For app implementation, if I choose mobile app it will most likely be developed with Ionic Framework. The

Ionic Framework allows for the same code base to be deployed as different native applications, so that the app can function both on android and iOS. However, if a desktop app is chosen it will be developed with Java.

C) Testing

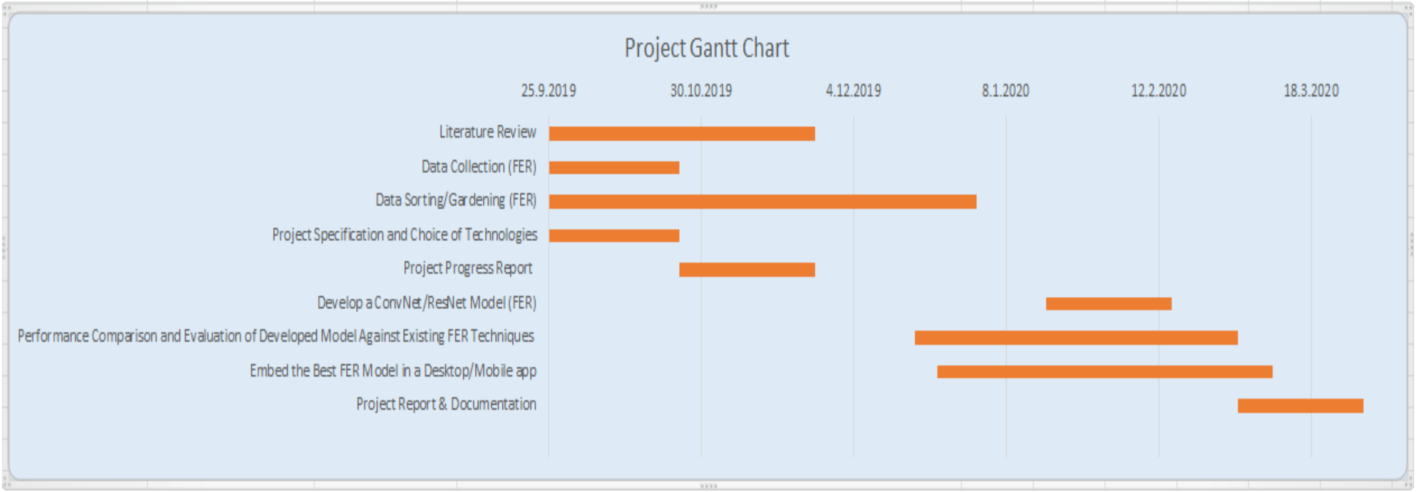
The FER system will be tested with k-fold cross-validation. Cross-validation is a procedure used to evaluate the performance of machine learning models on an independent dataset. K-fold refers to the number of groups that a given data will be split into. The proposed cross-validation is 10-fold cross-validation which is most widely used technique to test the performance of a machine learning technique. K-fold cross validation works as follows: After splitting the data in 10 groups, each individual group is taken as a test set or a hold out. Then the model is trained on the other 9 groups and tested on the group taken as a test set. The performance accuracy is then saved, and the model is discarded. Cross-validation is an especially useful technique when the amount of data is limited, or the model is suffering from overfitting which needs to be mitigated. With cross-validation we can ensure that performance accuracy is not biased or because of chance. The best-case scenario is if using accuracy as a measurement the model's accuracy is similar in all the 10 folds of the data. This means that the algorithm is consistent. However, the model's accuracy varies throughout the 10 folds then it can be assumed that the model suffers from overfitting, is unable to learn, or the data is too complex to be learned.

D) Evaluation

Performance of every model will be evaluated by checking how accurately each image is classified according to one of the 6 emotions. Other than recognition accuracy, the models will also be evaluated and compared on complexity, number of recognized emotions. Complexity will be calculated by measure for a given amount of time and space how well does a model classify images to the different emotion categories.

4. Project Plan

Start Date	End Date	Duration	Task Description
25.9.2019	25.11.2019	61	Literature Review
25.9.2019	25.10.2019	30	Data Collection (FER)
25.9.2019	1.1.2020	98	Data Sorting/Gardening (FER)
25.9.2019	25.10.2019	30	Project Specification and Choice of Technologies
25.10.2019	25.11.2019	31	Project Progress Report
17.1.2020	15.2.2020	29	Develop a ConvNet/ResNet Model (FER)
18.12.2019	1.3.2020	74	Performance Comparison and Evaluation of Developed Model Against Existing FER Techniques
23.12.2019	9.3.2020	77	Embed the Best FER Model in a Desktop/Mobile app
1.3.2020	30.3.2020	29	Project Report & Documentation



FER implementation

1. Data Collection – in order to ensure a robust and reliable FER system data collection is an important step. Main requirements for the data are to have enough of it and in order to ensure robustness the training data should have different head poses, illumination variation, scale, occlusion, etc. There are plenty datasets available specifically for FER as mentioned above. I’ve decided to use and currently have access to JAFFE, TFEID, CK, CK+, Yale, KDEF & AKDEF, and Oulu_Casia.
2. Data Sorting/Gardening – during this step data from all the datasets mentioned above will have to be extracted and sorted according to the emotions shown. During the gardening process the data will be prepared for the next stage - pre-processing.
3. Develop ConvNet/ResNet model and compare its performance with existing models.
    - Pre-processing - includes applying different techniques to the data in order to normalize it and make it more similar. This would ensure substantially better performance of FER.

App

4. Embed the best model in an app. The app will have a streaming webcam for face detection. OpenCV library will be used for face detection and extraction.

Picture 1.

epoch	train_loss	valid_loss	error_rate	time
0	0.258826	0.287585	0.098814	00:26
1	0.188545	0.239541	0.086957	00:26
2	0.134160	0.181967	0.075099	00:26
3	0.100222	0.187592	0.055336	00:25
4	0.075475	0.185064	0.055336	00:26

Prototype development, it can be seen on picture 1 a pre-trained resnet50 model achieving 95% accuracy. The model was developed with FastAI library built on top of PyTorch, trained on Google Cloud Platform (GCP) over only two of the datasets - JAFFE and TFEID with Normalization applied in advance. After testing Normalization increased the accuracy of the trained model by 20%. The ResNet50 is a 50-layer residual network. Residual network is similar to deep convolutional neural networks except the network learns residuals instead of features at each level. A residual can be represented as subtraction of feature learned from input of that layer. ResNet uses skip connections to propagate information over layers allowing for building deeper networks. Skip connections allow the network to understand global features by adding outputs from previous layers to the outputs of stacked layers. The resnet50 model has 5 stages and each one of them has convolution and identity block. Neural networks with that many layers are usually particularly hard to train because of the vanishing gradient problem. The vanishing gradient problem is a notorious problem when adding more layers to a neural network the gradients of the loss function approach zero making training to work ineffectively. It’s also important to note no cross-validation was performed on the model. Instead, 75% of the data was used for training and 25% for testing.

As for development methodology, I suggest the traditional waterfall approach. The waterfall method is a rigid, linear model that consists of sequential phases (requirements, design, implementation, verification, maintenance) each of which focuses on distinct objectives. In order to proceed to the next phase, the previous one must be 100% completed. In comparison to agile development, there’s usually no going back to modify the direction or the project. That would not be a problem for this project, since the project requirements and objectives are clearly outlined, hence why I believe that the waterfall method would be a good fit.

REFERENCES:

[1] F. Abdat, C. Maaoui and A. Pruski, "Human-Computer Interaction Using Emotion Recognition from Facial Expression," *2011 UKSim 5th European Symposium on Computer Modeling and Simulation*, Madrid, 2011, pp. 196-201.
doi: 10.1109/EMS.2011.20

[2] International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463 Vol. 3 Issue 2, February-2014, pp: (108-111), Impact Factor: 1.252, Available online at: www.erpublications.com

[3] Li, Shan and Weihong Deng. “Deep Facial Expression Recognition: A Survey.” *ArXiv* abs/1804.08348 (2018): n. pag.

[4] Samadiani, Najmeh et al. “A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data.” *Sensors (Basel, Switzerland)* vol. 19,8 1863. 18 Apr. 2019, doi:10.3390/s19081863

[5] Revina, I.M., Emmanuel, W.R.S. A Survey on Human Face Expression Recognition Techniques. Journal of King Saud, University – Computer and Information Sciences (2018, https://doi.org/10.1016/j.ksuci.2018.09.002

[6] M. C. Popescu, L. M. Sasu, "Feature extraction feature selection and machine learning for image classification: A case study", Optimization of Electrical and Electronic Equipment (OPTIM) 2014 International Conference on, pp. 968-973, 2014.

[7] https://arxiv.org/abs/1905.02845

[8] G. Muhammad, M. Alsulaiman, S. U. Amin, A. Ghoneim and M. F. Alhamid, "A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities," in *IEEE Access*, vol. 5, pp. 10871-10881, 2017.
doi: 10.1109/ACCESS.2017.2712788

[9]Kojima, Yuriko et al. “Characteristics of facial expression recognition ability in patients with Lewy body disease.” *Environmental health and preventive medicine* vol. 23,1 32. 18 Jul. 2018, doi:10.1186/s12199-018-0723-2

[10]McClure E.B., Pope K., Hoberman A.J., Pine D.S., Leibenluft E. Facial expression recognition in adolescents with mood and anxiety disorders. Am. J. Psychiatry. 2003;160:1172–1174. doi: 10.1176/appi.ajp.160.6.1172. [PubMed] [CrossRef] [Google Scholar]

[11] Wallace S., Coleman M., Bailey A. An investigation of basic facial expression recognition in autism spectrum disorders. Cogn. Emot. 2008;22:1353–1380. doi: 10.1080/02699930701782153. [CrossRef] [Google Scholar]

[12] Swayne, M. (2019). *People more likely to trust machines than humans with their private information*. [online] Phys.org. Available at: https://phys.org/news/2019-05-people-machines-humans-private.html [Accessed 5 Nov. 2019].

[13] M. A. Assari and M. Rahmati, "Driver drowsiness detection using face expression recognition," *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, 2011, pp. 337-341.
doi: 10.1109/ICSIPA.2011.6144162

[14] D. Yang, Abeer Alsadoon, P.W.C. Prasad, A.K. Singh, A. Elchouemi, An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment, Procedia Computer Science, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.12.003

[15] Technologies, V. (2019). *From emotion to animation - Visage Technologies*. [online] Visage Technologies. Available at: https://visagetechnologies.com/emotion-animation/ [Accessed 5 Nov. 2019].

[16] Poursaberi, A., Noubari, H.A., Gavrilova, M. *et al.* Gauss–Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. *J Image Video Proc* **2012,** 17 (2012) doi:10.1186/1687-5281-2012-17

[17] Diah Anggraeni Pitaloka, Ajeng Wulandari, T. Basaruddin, Dewi Yanti Liliana, Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition, Procedia Computer Science, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.10.038.

[18] Yi Ji, Khalid Idrissi, Automatic facial expression recognition based on spatiotemporal descriptors, Pattern Recognition Letters, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2012.03.006.

[19] Uçar, Ayşegül & Demir, Yakup & Güzeliş, Cüneyt. (2014). A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. Neural Computing and Applications. 27. 10.1007/s00521-014-1569-1.

[20] Cossetin, M.J., Nievola, J.C., Koerich, A.L., 2016. Facial expression recognition using a pairwise feature selection and classification approach. IEEE Int. Jt. Conf. Neural Networks, pp. 5149–5155.

[21] Dahmane, Mohamed & Meunier, Jean. (2014). Prototype-Based Modeling for Facial Expression Analysis. Multimedia, IEEE Transactions on. 16. 1574-1584. 10.1109/TMM.2014.2321113.

[22] Happy, S L & Routray, Aurobinda. (2015). Automatic Facial Expression Recognition Using Features of Salient Facial Patches. IEEE Transactions on Affective Computing. 6. 10.1109/TAFFC.2014.2386334.

[23] Cong Geng and X. Jiang, "SIFT features for face recognition," *2009 2nd IEEE International Conference on Computer Science and Information Technology*, Beijing, 2009, pp. 598-602.
doi: 10.1109/ICCSIT.2009.5234877

[24] Ippolito, P. (2019). *Feature Extraction Techniques*. [online] Medium. Available at: https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be [Accessed 6 Nov. 2019].

[25] T.M. Abhishree, J. Latha, K. Manikantan, S. Ramachandran, Face Recognition Using Gabor Filter Based Feature Extraction with Anisotropic Diffusion as a Pre-processing Technique, Procedia Computer Science, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.03.149.

[26] Hernandez-Matamoros, Andres & Bonarini, Andrea & Escamilla-Hernandez, Enrique & Nakano-Miyatake, Mariko & Perez-Meana, Hector. (2015). A Facial Expression Recognition with Automatic Segmentation of Face Regions. 529-540. 10.1007/978-3-319-22689-7_41.

[27] Ebenezer Owusu, Yongzhao Zhan, Qi Rong Mao, A neural-AdaBoost based facial expression recognition system, Expert Systems with Applications, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2013.11.041.

[28] Zeng Xuemei, Wu Qi, Zhang Siwei, Liu Zheying, Zhou Qing, Zhang Meishan, A False Trail to Follow: Differential Effects of the Facial Feedback Signals From the Upper and Lower Face on the Recognition of Micro-Expressions, Frontiers in Psychology, 2018, ISSN=1664-1078, 10.3389/fpsyg.2018.02015

[29] Wang, L., Li, RF., Wang, K. et al. Int. J. Autom. Comput. (2014) 11: 459. https://doi.org/10.1007/s11633-014-0835-0

[30] Matti Pietikäinen (2010) Local Binary Patterns. Scholarpedia, 5(3):9775.

[31] Xiaoming Zhao & Shiqing Zhang (2016) A Review on Facial Expression Recognition: Feature Extraction and Classification, IETE Technical Review, 33:5, 505-517, DOI: 10.1080/02564602.2015.1117403

[32] K. Daniilidis, A. Makadia and T. Bulow, "Image processing in catadioptric planes: spatiotemporal derivatives and optical flow computation," *Proceedings of the IEEE Workshop on Omnidirectional Vision 2002. Held in conjunction with ECCV'02*, Copenhagen, Denmark, 2002, pp. 3-10. doi: 10.1109/OMNVIS.2002.1044483

[33] Dayi Gong, Shutao Li and Yin Xiang, "Face recognition using the Weber Local Descriptor," *The First Asian Conference on Pattern Recognition*, Beijing, 2011, pp. 589-592.
doi: 10.1109/ACPR.2011.6166675

[34] D.G.Agrawal et al. Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 3(Version 1), March 2014, pp.502-506

[35] S. Kumar, M. K. Bhuyan and B. K. Chakraborty, "Extraction of informative regions of a face for facial expression recognition," in *IET Computer Vision*, vol. 10, no. 6, pp. 567-576, 9 2016. doi: 10.1049/iet-cvi.2015.0273

[36] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," in *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091-2106, Dec. 2005.
doi: 10.1109/TIP.2005.859376

[37] Islam, Md & Ahmed, Arif & Kundu, Krishau. (2014). Texture Feature based Image Retrieval Algorithms. International Journal of Engineering and Technical Research. 2.

[38] Ohashi, G. & Shimodaira, Y.. (2003). Edge-Based Feature Extraction Method and Its Application to Image Retrieval. Journal of Systemics, Cybernetics and Informatics. 1.

[39] Yongsheng Gao and M. K. H. Leung, "Face recognition using line edge map," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764-779, June 2002.
doi: 10.1109/TPAMI.2002.1008383

[40] Noh S., Park H., Jin Y., Park JI. (2007) Feature-Adaptive Motion Energy Analysis for Facial Expression Recognition. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2007. Lecture Notes in Computer Science, vol 4841. Springer, Berlin, Heidelberg

[41] Li, Jian & Lu, Yuqiang & Pu, Bo & Xie, Yongming & Qin, Jing & Pang, Wai-Man & Heng, Pheng-Ann. (2009). Accelerating Active Shape Model using GPU for facial extraction in video. Proceedings - 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2009. 4. 522 - 526. 10.1109/ICICISYS.2009.5357636.

[42] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. 1995. Active shape models—their training and application. Comput. Vis. Image Underst. 61, 1 (January 1995), 38-59. DOI=http://dx.doi.org/10.1006/cviu.1995.1004

[43] Černá, L., Cámara-Chávez, G., & Menotti, D. (2013). Face Detection : Histogram of Oriented Gradients and Bag of Feature Method.

[44] International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013

[45] Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE transactions on neural networks*, *13*(6), 1450–1464. doi:10.1109/TNN.2002.804287

[46] Z. Lihong, W. Ye and T. Hongfeng, "Face recognition based on independent component analysis," *2011 Chinese Control and Decision Conference (CCDC)*, Mianyang, 2011, pp. 426-429.
doi: 10.1109/CCDC.2011.5968217

[47] Shakya, Subarna & Sharma, Suman & Basnet, Abinash. (2016). Human behavior prediction using facial expression analysis. 399-404. 10.1109/CCAA.2016.7813754.

[48] Al-Modwahi, A.A., Sebetela, O., Batleng, L.N., Parhizkar, B., & Lashkari, A.H. (2012). FACIAL EXPRESSION RECOGNITION INTELLIGENT SECURITY SYSTEM FOR REAL TIME SURVEILLANCE.

[49] Muhammad Hameed Siddiqi, Rahman Ali, Abdul Sattar, Adil Mehmood Khan & Sungyoung Lee (2014) Depth Camera-Based Facial Expression Recognition System Using Multilayer Scheme, IETE Technical Review, 31:4, 277-286, DOI: 10.1080/02564602.2014.944588

[50] M. H. Siddiqi, R. Ali, A. M. Khan, Y. Park and S. Lee, "Human Facial Expression Recognition Using Stepwise Linear Discriminant Analysis and Hidden Conditional Random Fields," in *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1386-1398, April 2015. doi: 10.1109/TIP.2015.2405346

[51] Ghimire, D., & Lee, J. (2013). Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. *Sensors (Basel, Switzerland), 13*(6), 7714–7734. doi:10.3390/s130607714

[52] C, Emmanuel & Donoho, David. (2000). Curvelets - A Surprisingly Effective Nonadaptive Representation For Objects with Edges. Curves and Surfaces.

[53] Hou, Le & Samaras, Dimitris & Kurc, Tahsin & Gao, Yi & Davis, James & Saltz, Joel. (2016). Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016. 10.1109/CVPR.2016.266.

[54] Ridha Ilyas, Bendjillali & Beladgham, Moh & Merit, Khaled & taleb-ahmed, Abdelmalik. (2019). Improved Facial Expression Recognition Based on DWT Feature for Deep CNN. Electronics. 8. 324. 10.3390/electronics8030324.

[55] A. Ramirez Rivera, J. Rojas Castillo and O. Oksam Chae, "Local Directional Number Pattern for Face Analysis: Face and Expression Recognition," in *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1740-1752, May 2013. doi: 10.1109/TIP.2012.2235848

[56] B. Ryu, A. R. Rivera, J. Kim and O. Chae, "Local Directional Ternary Pattern for Facial Expression Recognition," in *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 6006-6018, Dec. 2017. doi: 10.1109/TIP.2017.2726010

[57] Guo, M., Hou, X., Ma, Y., & Wu, X. (2016). Facial expression recognition using ELBP based on covariance matrix transform in KLT. *Multimedia Tools and Applications, 76*, 2995-3010.

[58] Dhavalikar, Anagha & Kulkarni, Ramesh. (2014). Facial Expression Recognition Using Euclidean Distance Method. Journal of Telematics and Informatics. 2. 10.12928/jti.v2i1.1-6.

[59] Bag, Soumen & Sanyal, Prof(Dr.) Goutam. (2011). An efficient face recognition approach using PCA and minimum distance classifier. IEEE 2011 International Conference on Image Information Processing. 10.1109/ICIIP.2011.6108906.

[60] Islam D.I., Anal S.R.N., Datta A. (2018) Facial Expression Recognition Using 2DPCA on Segmented Images. In: Bhattacharyya S., Chaki N., Konar D., Chakraborty U., Singh C. (eds) Advanced Computational and Communication Paradigms. Advances in Intelligent Systems and Computing, vol 706. Springer, Singapore

[61] Sohail A.S.M., Bhattacharya P. (2007) Classification of Facial Expressions Using K-Nearest Neighbor Classifier. In: Gagalowicz A., Philips W. (eds) Computer Vision/Computer Graphics Collaboration Techniques. MIRAGE 2007. Lecture Notes in Computer Science, vol 4418. Springer, Berlin, Heidelberg

[62] Thakare, Prashant and Pravin S. Patil. “Facial Expression Recognition Algorithm Based On KNN Classifier.” (2016).

[63] Schmidt M., Schels M., Schwenker F. (2010) A Hidden Markov Model Based Approach for Facial Expression Recognition in Image Sequences. In: Schwenker F., El Gayar N. (eds) Artificial Neural Networks in Pattern Recognition. ANNPR 2010. Lecture Notes in Computer Science, vol 5998. Springer, Berlin, Heidelberg

[64] Liyuan Chen, Changjun Zhou, Liping Shen, Facial Expression Recognition Based on SVM in E-learning, IERI Procedia, ISSN 2212-6678, https://doi.org/10.1016/j.ieri.2012.06.171.

[65] P.C., Vasanth & K.R., Nataraj. (2015). Facial Expression Recognition Using SVM Classifier. Indonesian Journal of Electrical Engineering and Informatics (IJEEI). 3. 10.11591/ijeei.v3i1.126.

[66] Mahmud, Firoz & Mamun, Md. Al. (2017). Facial Expression Recognition System Using Extreme Learning Machine. International Journal of Scientific and Engineering Research. 8. 26-30.

[67] Mr.R.Sathish Kumar, G.Mohanraj, M.Srivathsan, M.Vishnu Prashanna (2016) Recognition of Facial Emotions Structures Using Extreme Learning Machine Algorithm. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056

[68] Wei Guo, Tao Xu, Keming Tang, Jianjiang Yu, and Shuangshuang Chen, “Online Sequential Extreme Learning Machine with Generalized Regularization and Adaptive Forgetting Factor for Time-Varying System Prediction,” Mathematical Problems in Engineering, vol. 2018, Article ID 6195387, 22 pages, 2018. https://doi.org/10.1155/2018/6195387.

[69] Kankal, Sonal S. and Asst. Prof. Madhavi Mane. “Comparing various techniques of Detection of facial expression with the algorithm ID3 (decision tree based).” (2019).

[70] De Vries, Gert-Jan & Pauws, Steffen & Biehl, Michael. (2015). Facial Expression Recognition Using Learning Vector Quantization. 760-771. 10.1007/978-3-319-23117-4_65.

[71] Anastasis Kratsios, “Universal Approximation Theorems”. (2019), arXiv:1910.03344.