# Amrita Vishwa Vidyapeetham

# Amrita School of Computing, Chennai

# Computer Science and Engineering-Cyber Security

# 20CYS212 / 20CYS215 – Multimedia Processing / ML in Cybersecurity

**A Literature review on Emotion Recognition for Various Facial Emotional Extraction.**

*UG scholar Amrita School of Computing, Amrita Vishwa Vidyapeetham – Chennai.*

**Introduction:**

Facial emotion recognition is the process of detecting and identifying emotions based on the analysis of facial expressions. This technology involves the use of artificial intelligence and machine learning algorithms to analyse the facial features of individuals and determine their emotional state.

Facial emotion recognition systems typically use cameras or other sensors to capture images of the face, and then analyse those images to determine the presence and intensity of specific facial expressions associated with different emotions. These expressions can include changes in the shape and position of the eyebrows, mouth, eyes, and other facial features.

The technology behind facial emotion recognition is based on the idea that certain facial expressions are associated with specific emotional states. For example, a smile is generally associated with happiness, while a frown is often associated with sadness or anger. By analysing these facial expressions, facial emotion recognition systems can accurately identify a person's emotional state.

Humans interact socially with the help of emotions, which are considered as a universal language. These emotions surpass cultural diversities and ethnicity. Facial expressions are responsible for conveying the information, which was difficult to perceive. It gives the mental state of a person that directly relates to his intentions or the physical efforts that he must be applying for performing tasks. As a result, automatic recognition of emotion with the help of high-quality sensors is quite useful in a variety of areas such as image processing, cybersecurity, robotics, psychological studies, and virtual reality applications to name a few. Efforts in this area are being made to gather information of high-quality to meet the demands of the system so that it can read, process and simulate human emotions. Geometric and machine learning based algorithms for an effective recognition are being refined, emphasizing emotion recognition in real-time and not just ideal laboratory conditions. Hence, building a system that is capable of both face detection and emotion recognition has been a crucial area of research.

It is a well-established fact that human beings are responsible for the depiction of six basic emotions, namely happiness, anger, surprise, sadness, fear, and disgust . These primary emotions form the primary classification of the study of human emotional responses. Apart from these basic emotions, several other emotions have been considered for research. These include contempt, envy, pain, drowsiness and various micro expressions. Facial expression is seen as the primary mode of recognition of human emotion. It works on facial motion and the deformations of facial features to classify them into emotion categories. This classification is based on visual information and may not be the sole indicator of emotion. Other factors also contribute to the recognition of a person’s emotional state such as voice, body language, gestures or even the direction of the gaze. Emotion recognition, therefore, demands a more precise knowledge of all these factors together with contextual information to convey more accurate results.

* **Joy-**The emotion evoked by well-being, success, or good fortune or by the prospect of possessing what one desires (delight: the expression or exhibition of such emotion.)
* **Sadness-**Sadness is an emotional pain associated with, or characterized by, feelings of disadvantage, loss, despair, grief, helplessness, disappointment and sorrow. An individual experiencing sadness may become quiet or lethargic, and withdraw themselves from others.
* **Surprise -**Surprise is defined as to cause of someone to feel in amazing feelings.
* **Anger-**Anger can occur when a person feels their personal boundaries are being or going to be violated.
* **Disgust-**Disgust is a feeling of dislike. Human may feel disgust from any taste, smell, sound or tough.
* **Surprise -**Surprise is defined as to cause of someone to feel in amazing feelings.

Facial emotion recognition is a complex task and the machine learning approach to recognize faces requires several steps to perform it.

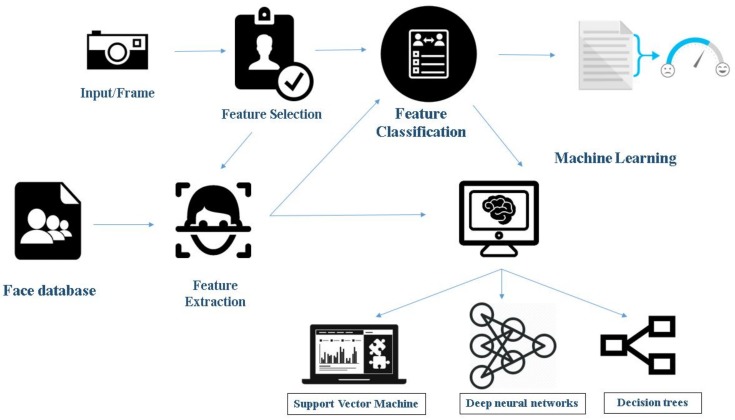
**Feature selection:** This stage refers to attribute selection for the training of the machine learning algorithm. The process includes the selection of predictors for construction of the learning system. It helps in improving prediction rate, efficiency, and cost-effectiveness. Many tools such as Weka and sci-kit-learn have inbuilt tools for automated feature selection.

**Feature classification:** When it comes to supervised learning algorithms, classification consists of two stages. Training and classification, where training helps in discovering which features are helpful in classification. Classification is where one comes up with new examples and, hence, assigning them to the classes that are already made through training the features.

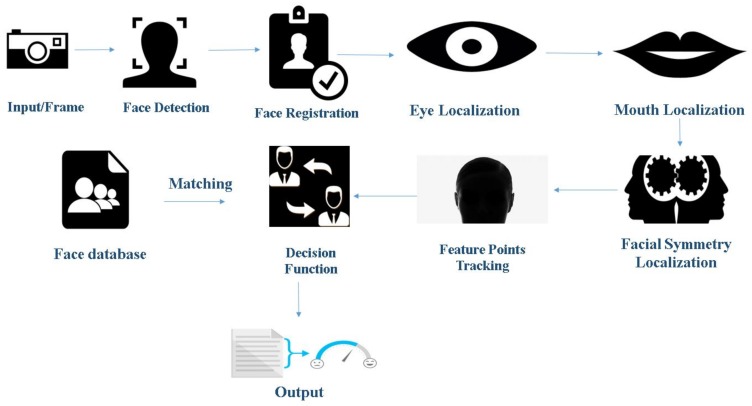
**Feature extraction:** Machine learning requires numerical data for learning and training. During feature extraction, processing is done to transform arbitrary data, text or images, to gather the numerical data. Algorithms used in this step include principal component analysis, local binary patterns, linear discriminant analysis, independent component analysis, etc.

**Classifiers:** This is the final step in this process. Based on the inference from the features, the algorithm performs data classification. It comprises classifying the emotions into a set of predefined emotion categories or mapping to a continuous space where each point corresponds to an expressive trait. It uses various algorithms such as Support Vector Machine (SVM), Neural Networks, and Random Forest Search.

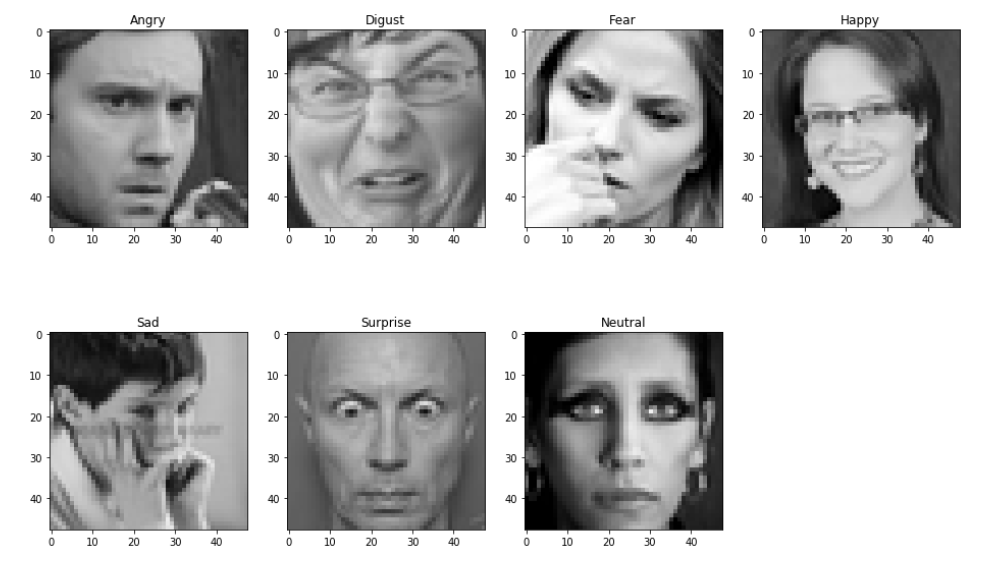
Pictorial Representation:



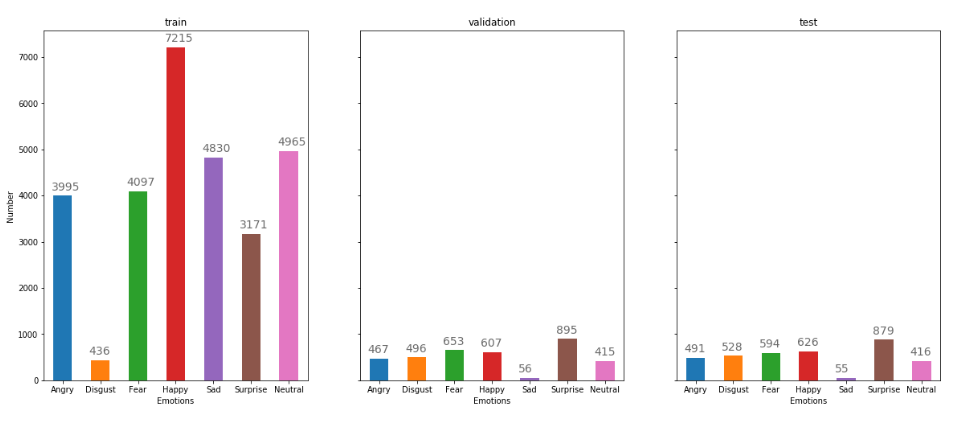
**Figure 1: Face detection and emotion recognition using machine learning.**



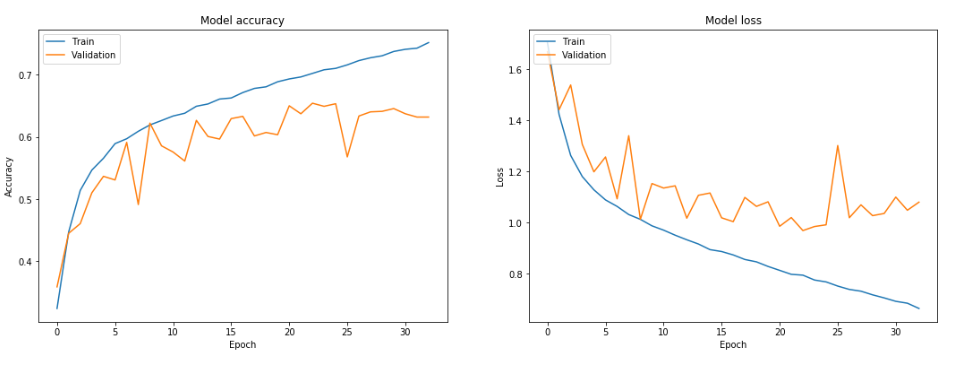
**Figure 2: Face detection and emotion recognition using geometric feature-based process.**



#### **Figure 3: Let's look at some images...**

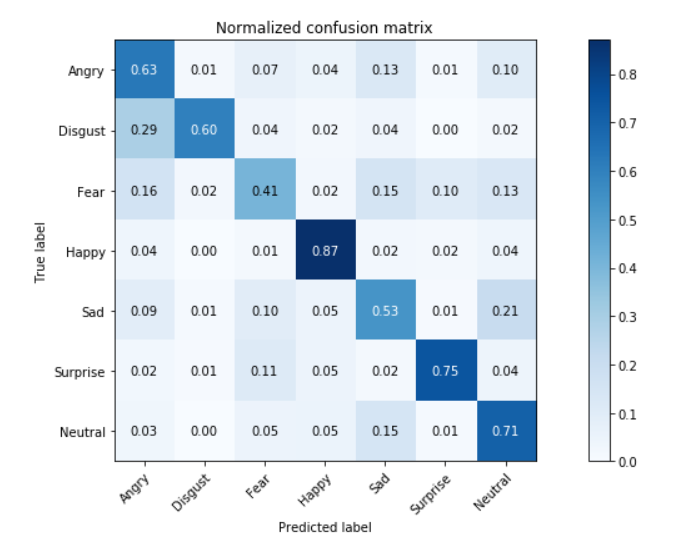


**Figure 4: The size of train, validation, test are 80%, 10% and 10%, respectively.**

****

## **Figure 5: Visualize Training Performance.**

## **More Analysis using Confusion Matrix**



**Tables:**

Table 1: Emotion based on attributes.

One common method for representing and analyzing facial emotions is through the use of tables. These tables typically contain columns for different facial features, such as eyebrow position, mouth shape, and eye gaze direction, as well as rows for different emotions, such as happiness, sadness, anger, and surprise

|  |  |  |  |
| --- | --- | --- | --- |
| Emotion | Eyebrow Position | Mouth Shape | Eye Gaze |
| Happy | Raised | Smiling | Direct |
| Sad | Lowered | Frowning | Downward |
| Angry | Lowered | Frowning | Direct |
| Surprise | Raised | Open | Upward |

Table 2: Emotion Intensity Levels

Another approach to facial emotion recognition is to analyze the intensity levels of various emotions. This method involves categorizing emotions based on their intensity, rather than their specific facial movements. Here's an example of a table that categorizes emotions based on their intensity:

|  |  |
| --- | --- |
| Emotion | Intensity Level |
| Happy | Low |
| Content | Low-Medium |
| Amused | Medium |
| Excited | Medium-High |
| Angry | Low-High |
| Disgusted | High |
| Afraid | High |
| Sad | Low-High |

In this table, emotions are categorized based on their intensity level, ranging from low to high. For example, a low-intensity happy expression might involve a slight smile or raised eyebrows, while a high-intensity expression might involve a broad smile or laughter. By measuring the intensity of various emotions in a person's facial expression, an algorithm can determine their emotional state.

Table 3: Facial Action Coding System (FACS)

The Facial Action Coding System (FACS) is a widely used method for analyzing facial expressions. It involves breaking down facial movements into specific action units (AUs) and determining the intensity and duration of each AU. Here's an example table of AUs and their corresponding emotions:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Emotion | AU 1 | AU 2 | AU 4 | AU 6 | AU 7 | AU 9 | AU 12 |
| Happy | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| Sad | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| Angry | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| Surprise | 1 | 0 | 1 | 0 | 1 | 0 | 0 |

In this table, each emotion is associated with a specific combination of AUs. For example, a happy expression is typically associated with the activation of AU 2 (raising of the outer corners of the eyebrows) and AU 7 (raising of the cheeks). By detecting and measuring the intensity of these AUs in a person's facial expression, an algorithm can determine if they are happy or not.

The AUs in Table 3 refer to specific facial muscle movements that are associated with particular emotions. For example, AU 1 (inner brow raiser) is associated with sadness, while AU 2 (outer brow raiser) is associated with happiness. By detecting and measuring the intensity of these AUs in a person's facial expression, an algorithm can determine the likely emotional state of the individual.

**Literature Review:**

The main aim of this research work is to classify the emotional expression from the mouth region of the human face. As the initial task is to extract the mouth region from the facial image, a survey on various existing research works to segment the face expression images is reviewed and discussed.

The worth of Emotion Recognition was recognized when “The Expression of Emotion in Man and Animals” was written by Charles Darwin .This book greatly inspired the study of emotions. Emotion Recognition due to its various applications gained immense importance like a drowsy driver could be spotted using emotion recognition systems .Corneanu et al., Matusugu et al., and Viola et al. gave a primary classification for the emotion recognition using multimodal approaches. They primarily talked about techniques and the parameters used for emotion recognition. The methods mentioned consisted of localization of the face using detection and segmentation, which made use of Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) algorithms. Along with all these techniques, Corneanu et al. concentrated on the categorization for emotion recognition considering the two principal components, one was parametrization, and the other was recognition of facial expressions. In his research, parametrization was relating the emotions detected while recognition of facial expressions was accomplished by using the algorithms such as Viola and Jones. This study also experiments with other algorithms like CNN and SVM , and this study concludes by proving that CNN demonstrates to have comparatively better accuracy on Viola and Jones algorithms.

Kahou et al. worked on a hybrid CNN—RNN (Recurrent Neural Network) architecture for emotion recognition. The hybrid approach was done by performing aggregation of CNN and RNN. The paper explored three CNN structures:

* a very deep one with 3 × 3 frame size;
* a three-layer with 5 × 5 filter size; and
* finally in the third one increased the filter size to 9 × 9.

**Integrating Face Detection and Pose Estimation**

To exploit the posited synergy between face detection and pose estimation, we must design a system that integrates the solutions to the two problems. Merely cascading two systems where the answer to one problem is used to assist in solving the other will not optimally take advantage of the synergy. Therefore, both answers must be derived from one underlying analysis of the input, and both tasks must be trained together.

**Robust face detection using CNN**

Internal representation of face is provided by a hierarchically ordered set of convolutional kernels defined by the local receptive field of FD neurons. Face model is represented as a spatially ordered set of local features of intermediate complexity, such as eyes, mouth, nose, eyebrow, cheek, or else, and all of these features are represented in terms of a fixed set of lower and intermediate features.

The work done by them used RNN to aggregate the frame features. The main reason to do this was that RNN could learn from an event irrespective of the time at which it must have occurred in a sequence. They performed a feature level and a decision level fusion, which, in turn, provided a significant improvement. The fusion with feature level was done using an MLP that had different hidden layers for each modality. In the fusion of decision level, the weighted sum of the class probabilities that were estimated was used. This architecture outperformed all the other methods like the aggregation of CNN-RNN performed and averaged of per frame based classifications.

**Problem Description:**

1. **Inter- and Intra-Individual Variability:** There is significant variability in the way people express emotions, even for the same emotion. This makes it challenging to develop models that can accurately recognize emotions across different individuals and under different circumstances. Inter-individual variability in FER can be caused by a variety of factors, including age, gender, culture, and personality. For example, some cultures may express emotions more subtly than others, making it more difficult for a FER system trained on one culture to accurately detect emotions in individuals from a different culture. Similarly, individuals with certain personality traits may exhibit more or less expressive facial expressions, which can affect the performance of FER systems. Intra-individual variability in FER can be caused by factors such as changes in lighting conditions, facial hair, or facial expression intensity. For example, changes in lighting conditions can affect the visibility of certain facial features, making it more difficult for FER systems to accurately detect emotions. Similarly, subtle changes in facial expression intensity can result in different AU activations, which can affect the accuracy of FER systems.
2. **Integration with Real-World Applications**: Integrating facial emotion recognition technology into real-world applications, such as human-computer interaction or security and surveillance, requires addressing technical, ethical, and legal challenges. Here are some examples:

* **Healthcare**: FER can be used in healthcare settings to monitor and diagnose patients with mental health disorders, such as depression and anxiety, by analyzing their facial expressions and emotional responses. FER can also be used to improve the quality of life of people with autism spectrum disorders by assisting with social interaction and communication.
* **Security**: FER can be integrated with security systems to improve surveillance and threat detection in public places such as airports and train stations. FER can also be used in law enforcement settings to identify suspects and detect deception.
* **Education**: FER can be used in educational settings to monitor student engagement and emotional responses to learning materials, providing valuable feedback for teachers to improve their teaching strategies.
* **Entertainment**: FER can be used in the gaming and entertainment industries to create more immersive and interactive experiences for users, such as personalized avatars that respond to users' emotions in real-time.

However, there are also potential ethical and privacy concerns with the integration of FER technology in real-world applications.

**Methodology:**

Facial emotion recognition is a complex process that involves multiple steps, including data collection, pre-processing, feature extraction, and classification. The following is a general overview of the methodology for facial emotion recognition:

**Data collection**: The first step in facial emotion recognition is to collect data in the form of facial images or videos. This can be done using a variety of methods, such as capturing video footage of participants during an experiment or using publicly available datasets.

**Pre-processing:** The collected data is then pre-processed to remove any noise or artifacts and to normalize the images for size and orientation. This can involve techniques such as image cropping, resizing, and grayscale conversion.

**Feature extraction:** The next step is to extract features from the pre-processed images that are relevant to facial emotion recognition. Commonly used features include AUs, facial landmarks, and texture descriptors.

**Classification:** Once the features have been extracted, a machine learning algorithm is used to classify the images into one or more emotional categories. This can involve techniques such as support vector machines, decision trees, or neural networks.

**Evaluation**: The final step is to evaluate the performance of the facial emotion recognition system. This is typically done by comparing the predicted emotions to ground truth labels and calculating metrics such as accuracy, precision, and recall.

There are many variations to this general methodology, and different approaches may be used depending on the specific application and data available. For example, deep learning methods such as convolutional neural networks (CNNs) have shown promising results for facial emotion recognition, and some systems may use additional sources of data such as audio or physiological signals.

**References:**

Main Reference[[Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality - PMC (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5856132/)]

Corneanu C.A., Simón M.O., Cohn J.F., Guerrero S.E. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Trans. Pattern Anal. Mach. Intell.*2016;**38**:1548–1568. doi: 10.1109/TPAMI.2016.2515606. [[PMC free article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7426891/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/26761193)] [[CrossRef](https://doi.org/10.1109%2FTPAMI.2016.2515606" \t "_blank)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=IEEE+Trans.+Pattern+Anal.+Mach.+Intell.&title=Survey+on+RGB,+3D,+thermal,+and+multimodal+approaches+for+facial+expression+recognition:+History,+trends,+and+affect-related+applications&author=C.A.+Corneanu&author=M.O.+Sim%C3%B3n&author=J.F.+Cohn&author=S.E.+Guerrero&volume=38&publication_year=2016&pages=1548-1568&pmid=26761193&doi=10.1109/TPAMI.2016.2515606&)]

 Matsugu M., Mori K., Mitari Y., Kaneda Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Netw.*2003;**16**:555–559. doi: 10.1016/S0893-6080(03)00115-1. [[PubMed](https://pubmed.ncbi.nlm.nih.gov/12850007)] [[CrossRef](https://doi.org/10.1016%2FS0893-6080(03)00115-1" \t "_blank)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Neural+Netw.&title=Subject+independent+facial+expression+recognition+with+robust+face+detection+using+a+convolutional+neural+network&author=M.+Matsugu&author=K.+Mori&author=Y.+Mitari&author=Y.+Kaneda&volume=16&publication_year=2003&pages=555-559&pmid=12850007&doi=10.1016/S0893-6080(03)00115-1&)]

Viola P., Jones M. Rapid object detection using a boosted cascade of simple features; Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001); Kauai, HI, USA. 8–14 December 2001;; Piscataway, NJ, USA: IEEE; 2001. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Proceedings+of+the+2001+IEEE+Computer+Society+Conference+on+Computer+Vision+and+Pattern+Recognition+(CVPR+2001)&title=Rapid+object+detection+using+a+boosted+cascade+of+simple+features&author=P.+Viola&author=M.+Jones&volume=Volume+1&publication_year=2001&)]

 Osadchy M., Cun Y.L., Miller M.L. Synergistic face detection and pose estimation with energy-based models. *J. Mach. Learn. Res.*2007;**8**:1197–1215. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=J.+Mach.+Learn.+Res.&title=Synergistic+face+detection+and+pose+estimation+with+energy-based+models&author=M.+Osadchy&author=Y.L.+Cun&author=M.L.+Miller&volume=8&publication_year=2007&pages=1197-1215&)]

Dalal N., Triggs B. Histograms of oriented gradients for human detection; Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR 2005); San Diego, CA, USA. 20–25 June 2005; pp. 886–893. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Proceedings+of+the+IEEE+Computer+Society+Conference+on+Computer+Vision+and+Pattern+Recognition,+(CVPR+2005)&title=Histograms+of+oriented+gradients+for+human+detection&author=N.+Dalal&author=B.+Triggs&volume=Volume+1&pages=886-893&)]

<https://openaccess.thecvf.com/content_cvpr_2017_workshops/w33/papers/Kowalski_Deep_Alignment_Network_CVPR_2017_paper.pdf>

Aleksic PS and Katsaggelos AK, “Automatic facial expression recognition using facial animation parameters and multistream hmms,” TIFS, vol. 1, no. 1, pp. 3–11, 2006. [Google Scholar]