# Web Enabled Spontaneous Facial Expression Database (WESFED): Challenges and Design

MAHESH M. GOYANI
Assistant Professor, Department of Computer Engineering,
Government Engineering College, Modasa
Gujarat, India.

Accurate Facial Expression Recognition (FER) can lead to improved man-machine interaction. The success of pattern recognition depends heavily on the amount and the quality of training samples. Over the time, many expression datasets have been published, but most of them are designed under controlled environment. Expressions are faked by subjects with peak intensity in front of the high-resolution camera. For the realistic use, the system should be able to cope with the spontaneous expressions, even in low-resolution environment. Although significant work has been done in the field of FER, a generalization of investigated methods is still unknown. Due to the multidimensional diversities in race, ethnicity and surrounding environment, it becomes difficult to design facial expression image dataset, which can serve as a benchmark tool for the research community. Existing dataset provides the baseline for the research, but they do not guarantee the real time use of systems trained with those datasets. In this paper, we present a 2D static Web Enabled Spontaneous Facial Expression Database (WESFED), which covers almost every type of diversities necessary to build real life expression recognition system. In this dataset, we overcome most of the limitations of existing dataset by incorporating variations in ethnicity, age, gender, illumination, inplane and out of plane rotation of the face. The work addresses the issues that needs to be considered while creating database. In addition, state-of-the-art datasets are also reviewed with their merits and demerits.

Keywords: Affective Computing, Facial Expression Recognition, Action Units, Facial Expression Dataset

## 1. INTRODUCTION

Affective computing is the next generation computing, in which an automated response is generated from human behaviour and expressions. It focuses on recording, interpreting and modelling the users mental state into appropriate computer actions (Pantic et al. [2006]). Facial expression recognition is the process of finding the emotional state of the person from the facial image. Facial expression is not just a physical change on the face but is a physio-psychological process. It emerges from the mind and gets reflected on the face in the form of contraction of muscles. This adjustment of muscles keeps going for a limited period, roughly from 250 ms to 5 sec (Fasel and Luettin [2003]), so tracking of exact expression is difficult.

Progressively, future of computation is being humanistic instead of computer-centric. Automatic facial expression recognition can act as a component of many human-machine interfaces (HCI) applications. Apart from socially sensitive Human- Computer Interaction (Cid et al. [2013]), AFER can be utilized in many applications such as detection of mental disorders (Gaebel and Wlwer [1992]), safety against road rage (Nasoz et al. [2004]), security (Butalia et al. [2012]), animations and video games (Bartlett et al. [2003]), automation applications, etc. It can also be used to include emotion related information in automatic image captioning systems (Cowie et al. [2001]). Multi-modality provides a further clue for the precise action. Synthetic speech with expressions seems more pleasing and convincing than a monotonous voice. Talking heads, avatars and computer agents can be trained to learn user preferences through expressions (Fasel and Luettin [2003]), (T. Kanade and Tian [2000]). Traditional HCI frameworks do not account the mental condition of the individual. Facial expressions are the most grounded segments, which mirrors the mental state of the person. As well as, they also provide an important behavioural measure for the study of emotions, cognitive processes and social interactions (Donato et al. [1999]). Ekman and Friesen [1971] categories range of expressions into basic six classes. Those six prototype expressions are anger, disgust, fear, happy, sad and surprise. However, machines are expected to recognize expressions falling outside this cluster. Detailed survey on facial expression recognition can be found in the literature Tian et al. [2000], Pantic and Rothkrantz [2000], Pantic and Rothkrantz [2003], Fasel and Luettin [2003], Zeng et al. [2009], Corneanu et al. [2016], Goyani and Patel [2017].

Success of such system depends on the large training and testing samples. Static FER system extracts the facial features from the single image, whereas configurations of muscles can be better understood on image sequences. Dynamics provides the temporal deformation of facial muscles influenced by a particular expression. Expression database is essential component for training and validating FER systems. Various facial expression datasets such as CK (T. Kanade and Tian [2000]), JAFFE (Lyons [1999]), BND (Douglas-Cowie et al. [2003]), 3D-BFU (Yin et al. [2006]), MMI (Pantic et al. [2005]), Bosphorus (Savran et al. [2010]) etc. are designed with different modalities and goals. Most of these databases are designed under controlled environment with limited subjects and images. Due to no or little diversity, and with faked expression, these datasets do not provide an ideal ground for expression recognition for real life scenario. Systems designed using such database suffers from performance degradation with a change in head pose, illumination, occlusion, age, and ethnicity. The objective of posed expression database is to provide the foundation for the design and evaluation of novel algorithms. Once the algorithms are tested they advance for spontaneous expression database. Due to growing research in the field of FER, there is a need of comprehensive wide scale dataset, which highlights images with the more realistic scenario.

It would be efficient and easier if all researchers test their system on same dataset. This will make comparison of state of the art methods straightforward and simple. However, designing common cross application dataset for all future application is really a challenging task. FERET has been proved a de-facto standard for face recognition systems, but standardized facial expression dataset is still an open problem. Since the surveys Fasel and Luettin [2003], Pantic and Rothkrantz [2000] published on facial expression datasets, significant progress has been made in the development of standard dataset.

In this paper, we have proposed novel Web Enabled Spontaneous Facial Expression Dataset (WESFED), which accounts almost every type of diversities necessary to design real life facial expression system. WESFED contains seven expressions (six basic expressions and a neutral one). Framework for dataset creation is proposed and discussed.

Rest of the paper is organized as follows: State of the art facial expression datasets are reviewed in section 2. Challenges and design of WESFED are discussed in section 3. Section 4 covers conclusions and discussions followed by references.

## 2. BACKGROUND AND RELATED WORK

Automatic facial expression recognition is essential for affective computing, but due to the limited datasets or limitations of datasets, the problem is not well studied. The expression can be modeled using descriptive or judgmental coding scheme (Corneanu et al. [2016]). Descriptive coding defines the expression in terms of position and relation between facial muscles. The face is described by the set of facial muscles configuration, called Action Units (AUs). Ekman and Friesen [1978] reported more than 7000 such combinations. CK and MMI are descriptive coding based datasets. Judgmental coding describes expression using subjective parameters such as wrinkles, texture etc. JAFFE, AR, CASIA, TFEID etc. are judgmental coding based datasets. Besides basic six prototype expressions, few datasets also include some other common expressions such as neutral (Lyons and Akamatsu [1998]), (Martinez and Benavente [1998]), (Sim et al. [2002]), wink (Sim et al. [2002]), scream (Martinez and Benavente [1998]), talk (Sim et al. [2002]) and smile (Martinez and Benavente [1998]). Sociologists have classified expressions into 18 different classes:

despair, grief, surprise, flurry, laugh, horror, disgust, fury, fear, doubt, disparagement, impatience, contempt, sneer, smile, hate, plea, and worry (Zhou [2003]). Xue et al (Xue et al. [2006]) proposed a dataset which contains 25 expressions (18 pure, 3 mixed and 4 complex expressions). Most of the studied dataset contains six or seven expressions.

MMI has proven as a major turnaround in facial expression recognition. It was designed by Pantic et al. (Pantic et al. [2005]), (Maat et al. [2005]). It consists of both posed and spontaneous expressions including profile view. MMI is freely available and was made available online on February 2009 (mmi []). Few other researchers have created own datasets like Ekman et al. (P.Ekman and W.Irwin []), Donato et al. (Donato et al. [1999]), Chen-Huang (Chen [2000]) etc. Another important database is Ekmans datasets (ekm []) which is not freely available. It contains peoples form diverse ethnicity, including Japanese, Caucasians and stone-age tribes.

Most of the research work is centered around recognition of expressions from 2D data. In recent time, few 3D facial expression datasets also designed to address this issue (Yin et al. [2006]). Apart from the mentioned databases, there are many other databases like the RU-FACS-1, USC-IMSC, UA-UIUC, QU, PICS and others. Comparative study of these datasets can be found in literature Pantic et al. [2005]. Characteristics of few facial expression datasets are compared in Table II. Comparative characteristics of these datasets are stated in Table I.

Table I: Comparative characteristics of expression datasets

| **Label** | **Interpretation** |
|---|---|
| 1 | Image sequences |
| 2 | Diverse ethnicity |
| 3 | Uniform light |
| 4 | Occlusion |
| 5 | AU-Coded |
| 6 | Profile view |
| 7 | Multiple images |
| 8 | Color images |
| 9 | Only front view |
| 10 | Spontaneous |

Table II: General overview of different datasets

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| CK | ○ | ○ | ○ | ● | ○ | ● | ○ | ● | ○ | ● |
| PICS | ● | × | × | ● | ● | ○ | ● | × | ● | ● |
| JAFFE | ● | ● | × | ● | ● | ● | ● | ● | ○ | ● |
| AR | ● | × | ● | ○ | ● | ● | ● | ○ | ○ | ● |
| PIE | ● | × | ● | ○ | ● | ○ | ● | × | ● | ○ |
| MMI | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ | ○ | ○ |
| CASIA | ○ | × | × | ● | ● | ● | ● | ● | ○ | ● |
| TFEID | ● | ● | × | ● | ● | ● | ○ | ○ | ○ | ● |
| BU-3DFE | ● | × | × | ● | ● | ● | ○ | ○ | ○ | ● |
| Bosphorus | ● | ● | × | ● | ● | ● | ○ | ○ | ● | ● |
| BND | ○ | ● | ○ | ● | ● | ○ | ● | ○ | ● | ○ |
| NVIE | ● | ● | ● | ● | ● | ○ | ● | ○ | ● | ○ |

○: Yes, ●: No, ×: Not available/applicable

CK (T. Kanade and Tian [2000]) database sets up the modern research in FER. Images and expressions in CK database are more controlled with categorical labels. CK+ database was published later, covering more number of images with spontaneous expression and AU coded images. JAFFE (Japanese Female Facial Expression Database) (Lyons and Akamatsu [1998]) is another widely used FE database, representing 10 Japanese female models. Images in JAFFE are captured under uniform lighting condition and has a limited number of subjects with a frontal view only. TFEID (Taiwanese Facial Expression Image Database) (Chen and Yen [2007]) contains 40 subjects, with a male-female ratio of 1:1. Subjects of JAFFE and TFEID belong to a single ethnicity. Yin et al. [2006] built 3D facial expression database of static images and image sequences. These databases are known as BU-3DFE (Binghamton University 3D Facial Expression) database and BU-4DFE (3D + time) database, respectively. At present, BU-3DFE contains 100 subjects (56% female, 44% male), age ranging from 18 - 70 years, with a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic-Latino. Aleix Martinez and Robert Benavente (AR) (Martinez and Benavente [1998]) database consists of only four expressions - angry, neutral, scream and smile. It consists of 4000+ static color images of 70 males and 56 females. MMI is image/video database, designed by (Pantic et al. [2005]). This database contains more than 1500 images plus video sequences, with frontal and profile view. Age group of subjects covers a wide range of 19 - 62 years. BND (Belfast Naturalistic Database) (Douglas-Cowie et al. [2003]) was created by Queen's University of Belfast, which covers 250 videos of a wide range of expression. Each expression considers neutral as ground truth and reaches to the apex and comes back to neutral. Bosporus (Savran et al. [2010]) is also 3D dataset with a wide range of subjects, head pose and a number of expressions. NVIE (Natural Visible and Infrared facial Expression) database (Wang [2010]) contains visible and infrared images with 412 static images of 215 subjects having age group of 17 - 31 years. A comprehensive survey on facial expression datasets based on the subject, content and modality is described in Table III, Table IV and Table V, respectively.

Table III: Subject based Survey on Facial Expression Recognition Datasets

| | #Subject | #Male | #Female | Age Group | Ethnic Diversity |
|---|---|---|---|---|---|
| CK | 97 | 35% | 65% | 18 30 | Y |
| JAFFE | 10 | 0 | 10 | N/A | N |
| TFEID | 40 | 20 | 20 | N/A | N |
| AR | 126 | 70 | 56 | N/A | N/A |
| BU-3DFE | 100 | 44 | 56 | 18 70 | Y |
| Bosphorus | 105 | 60 | 45 | 25 35 | N |
| MMI | 19 | 56% | 44% | 19 62 | Y |
| BND | 125 | 31 | 94 | N/A | N |
| NVIE | 215 | 157 | 58 | 17 31 | N |

Table IV: Content based Survey on Facial Expression Recognition Datasets

| | #Images/Videos | Static/Videos | Frame Rate | Single / Multiple faces | coding scheme | 3D Data | Resolution |
|---|---|---|---|---|---|---|---|
| CK | 486 | Videos | 12 fps | Multiple | Gray | N | 640 × 490, 640 × 480 |
| JAFFE | 213 | Static | N/A | Single | Gray | N | 256 × 256 |
| TFEID | 1975 | Static | N/A | Multiple | Gray + color | N | 600 × 480 |
| AR | 4000+ | Static | N/A | Single | Color | N | 768 × 576 |
| BU-3DFE | 2500 | Static | N/A | Multiple | Color | Y | 512 × 512 |
| Bosphorus | 4666 | Static | N/A | Multiple | Color | Y | N/A |
| MMI | V: 848, I: 740 | Both | 24 fps | Single | Color | N | 720 × 576 |
| BND | 250 | Video | N/A | Single | Color | N | N/A |
| NVIE | 412 | Static | N/A | Single | Color | N | 1920 × 1080 |

N/A: Not Available or Not Applicable, Y: Yes, N: No, I: Image, V: Video, A: Anger, AP: Apex, C: Contempt, D: Disgust, F: Fear, H: Happy, N: Neutral, SA: Sad, SC: Scream, SM: Smile, SU: Surprise, O: Onset, FP: Frontal & Profile, P: Posed, S: Spontaneous, FV: Front view

Table V: Modality based Survey on Facial Expression Recognition Datasets

| | Face pose | Expressions | Intensity | Posed / Spontaneous | Illumination |
|---|---|---|---|---|---|
| CK | FV | 23 facial displays + single or combinations of AU(s) | Neutral to apex | P | Uniform |
| JAFFE | FV | A, D, F, H, N, SA, SU | N/A | P | N/A |
| TFEID | FV captured at 0o and 45o | A, C, D, F, H, N, SA, SU | slight & high intensity | P | N/A |
| AR | FV | A, N, SC, SM | with and without occlusions | P | Left, right and all lights on |
| BU-3DFE | FV | A, D, F, H, N, SA, SU | Neutral to apex | P | N/A |
| Bosphorus | Yaw, Pitch and Cross rotation | A,D, F, H, SA, SU + 25 pure and 3 combination of AUs | Head poses, Occlusion | P | N/A |
| MMI | FV + FP captured by two cameras | 79 facial displays including single / combinations of AUs | Neutral-apex-neutral | P | Uniform |
| BND | Various | Wide range of facial expressions | Neutral-apex-neutral | S | Indoor |
| NVIE | Front, left, right | A, AP, D, F, H, O, SA, SU | N/A | P + S | Left and right illuminated images |

These databases can be categorized into two categories: 1) 2D image or sequence based database, and 2) 3D static images or 3D sequence based database. JAFFE, CK, CMU PIE, MMI etc. belong to the first category, whereas BU-3DFE and Bosphorus belongs to second. Although extensive research has been done on 3D face recognition and face animation, 3D facial expression is not explored enough (Yin et al. [2006]). The Same fact applies to spontaneous images. Lack of the work is mainly due to the unavailability of the suitable datasets. Recently, researchers have explored and used the 3D information to improve the recognition rate. They have used either partial 3D information, such as multiple views (Pantic and Rothkrantz [2004]) or 3D models (Huang and Huang [1997]). Interpreting human face as a bumpy surface would have more practical applications rather than considering it flat 2D surface (Yin et al. [2006]).

A number of 2D face and facial expression recognition datasets are designed by the research community, but very few 3D databases are published. Lack of common evaluation strategy makes the comparison even difficult. 3D and even 2D datasets need careful consideration. Some of the 2D and 3D face and facial expression datasets are listed in Table VI.

Table VI: Existing 2D and 3D Face and Facial Expression Datasets

| Type | Face Recognition | Facial Expression Recognition |
|---|---|---|
| 2D | AR (Martinez and Benavente [1998]),<br>BioID (bio []),<br>FERET (Phillips et al. [2000]),<br>CMU-PIE (Sim et al. [2002]),<br>FRGC (Phillips [2005]),<br>XM2VTSDB (Messer et al. []),<br>UT-Dallas (OToole [2005]) | CK (T. Kanade and Tian [2000]),<br>JAFFE (Lyons and Akamatsu [1998]),<br>TFEID (Chen and Yen [2007]),<br>MMI (Pantic et al. [2005]),<br>BND (Douglas-Cowie et al. [2003]),<br>NVIE (Wang [2010]) |
| 3D | Xm2vtsdb (Messer et al. []),<br>3D FRGC (Chang et al. [2005]),<br>PRISM-ASU (pri []) | BU-3DFE (Yin et al. [2006])<br>Bosphorus (Savran et al. [2010]) |

Next big task ahead for dataset designer was to label the data and augment the data with useful metadata. Expert observers or AU coders were labeling the data manually. This process is very much time consuming and tedious. Sometimes subjects themselves were asked to rate the expression they have created. Labeling data requires ample amount of training and careful observation of individual image. Labeling of AU requires special skill. To become a certified FACS coder, it takes 100 hours of training to learn reliable judgments. Even a good coder takes one to three hours of time to label one minute video on a frame by frame basis Bartlett et al. [2006], Cohen et al. [2003]. Cohen et al. [2003] used Bayesian Network Classifier like Nave Bayes (NB), Tree Augmented nave Bayes (TAN) and Stochastic Structure search (SSS) for semi-supervised learning for auto labeling of data.

## 3. CHALLENGES AND CREATION OF WESFED

Sebe et al. [2007] listed the major problems associated with spontaneous expression image acquisition. This has proved important step in that direction. Their observations were as follow:

—Different subjects express the same emotions at different intensities

—Expression loses its intensity if subject becomes aware that he is being captured

—Laboratory conditions may not encourage the subject to display spontaneous expressions.

Authors developed a video kiosk for real time recording of subject expressions. Expressions are recorded using hidden cameras. After recording procedure is done, each subject is notified and asked for their permission to use their video for research purpose. 50% subjects turned up with positive response. Subjects themselves were asked to rate their own videos for the true emption labeling, and their reviews were documented for further analysis (Sebe et al. [2007]). Study shows that it was very difficult to cover wide range of expression, in particular fear and sad were found to be more difficult to induce. The study helps in understanding the problems appearing in spontaneous expression acquisition, and the mechanisms that can be used to elicit the issues.

In CK dataset, each video sequence ends in peak facial expression, and hence the sequence was incomplete in terms of its temporal pattern. Cohen et al. [2003] reported that due to this limitation, they couldnt use their own dataset to train and test a Multi-Level HMM classifier. From this fact, dataset designer should infer that the video sequence must have complete temporal pattern including onset, apex and offset images.

Wang and Yin [2007] reported another specif problem of topographical modeling of facial expression. Topographic modeling is pixel based approach and hence it fails against illumination changes. And authors were unable to find the dataset having expressive faces with different intensity modeling. Occlusion related issues were addressed by Kotsia et al. [2008]. Authors couldnt find the dataset that had expressive faces with occlusions. Authors themselves manually preprocessed the CK and JAFFE dataset graphically add the occlusion by introducing the white patches on face.

Proposed work is based on the design of 2D spontaneous facial expression dataset. This is the rare attempt to foster the research in the field of facial expression analysis with such a large and dynamic dataset. The dataset can be useful to generalize FER systems as it covers a wide range of diversity including age, ethnicity, gender, occlusion, illumination, pose etc. The proposed dataset consists of seven basic expressions. Expressions are coded using judgmental features. Rigorous analysis of various datasets led us to the implementation of more realistic and spontaneous facial expression dataset. It would not be much difficult to design a system for expressions with peak intensity, however, in real life, the formation of expression is not ideal. Schmidt and Cohn [2001] noted 18 unique classes of a smile. Also, other expressions may have a number of variants - gentle to peak. Thus, training the classifier and recognition of facial expression is not an easy task. The intensity of expression varies over the time for an individual, as well as for two different persons at a time. Subsequently, it is difficult to precisely determine the intensity of expression without referring to the neutral face of a given subject.

Although lots of expression databases were built, none of these databases collected spontaneous facial expression. The subjects were instructed to perform specific facial expressions which were faked in front of camera. The fact is that it is very difficult to collect all six basic spontaneous expressions from daily life incidents, and ground truth labeling is also not an easy task. Recognition of the expressions with low intensity is much more difficult than recognizing exaggerated expressions.

The design of FERS is challenging due to variations encountered in age, gender, race, ethnicity, head pose, illumination, and many other factors. For effective machine learning, it is necessary to collect a large number of training facial images acquired from numerous subjects. Generalization of FERS is very important because the subject specific system has little or no practical use. Hence, it becomes a crucial task to acquire a rich database representing natural expressions with certain variability.

In WESFED, we tried to capture all possible diversities. Raw images were acquired from google search engine using text-based search criteria. Due to limitations of the search engine, google showed only limited images. Out of which, some were dead linked and some were not relevant. Different keywords with identical meaning are used to collect the large pool of raw images. For example, Surprised faces are collected using surprised faces, astonished characters, amazed human etc. keywords. List of keywords used to search raw images for all seven expressions are listed in Table VII.

Table VII: Keywords used for text-based image search

| Expression | Keywords used for text-based search |
|---|---|
| Anger (AN) | Angry, danger, displeasure, irritate, hostile |
| Disgust (DI) | Disgust, abhorrence, aversion, shame, dislike |
| Fear (FE) | Fear, terror, panic, scare, alarming |
| Happy (HA) | Happy, cheerful, joyful, jolly, gleeful, delighted, pleased, |
| Sad (SA) | Sad, unhappy, sorrow, depressed, down, miserable, gloomy |
| Surprise (SU) | Surprise, astonished, amazed, wondered |
| Neutral (NE) | Neutral, disinterested, unbiased, unemotional, impartial |

Data collected in this way contains many irrelevant images also. At first, color images are converted to gray scale and faces are cropped using viola jones cascade object detection algorithm. We manually removed irrelevant images from cropped faces. The framework of the implementation of WESFED is portrayed in Figure 1.

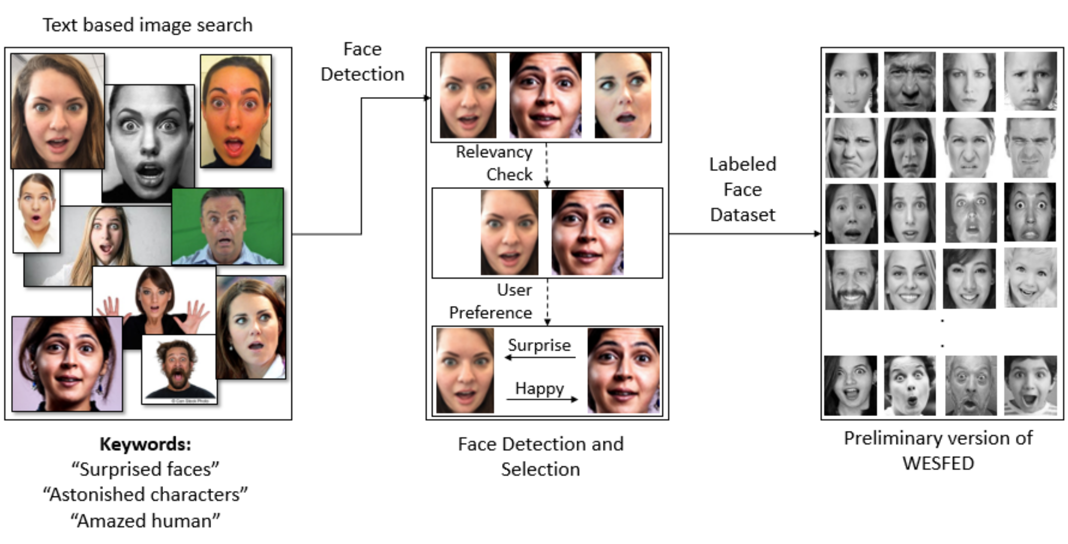


Figure 1. Framework for design of WESFED

Sometimes expressions are mixture of multiple emotions and it is difficult to accurately label them. For the robust class label, user feedback based scheme was employed. Each cropped face is presented to 10 different persons and they are asked to vote them from one of the seven predefined classes. Class with maximum votes is assigned to the face. Cropped images are normalized to 160 × 140 pixel and stored as an 8-bit grayscale image. Confusion matrix for the subjective voting scheme is shown in Table VIII. As observed in other datasets, sad-fear and disgust-angry are the most confusing pair of expressions.

Table VIII: Subjective measurement of each class

| | AN | DI | FE | HA | SA | SU | NE |
|---|---|---|---|---|---|---|---|
| AN | **95.3** | 2.7 | 1.2 | 0 | 0.8 | 0 | 0 |
| DI | 2.8 | **94.6** | 2.1 | 0 | 0.5 | 0 | 0 |
| FE | 1.2 | 1.7 | **93.8** | 0 | 2.6 | 0.7 | 0 |
| HA | 0 | 0 | 0.5 | **97.9** | 0 | 1.2 | 0.4 |
| SA | 1.1 | 1.8 | 2.7 | 0 | **94.3** | 0 | 0.1 |
| SU | 0.3 | 0 | 0.8 | 1.3 | 0.2 | **97.4** | 0 |
| NE | 0 | 0.3 | 0 | 0.8 | 0.4 | 0 | **98.5** |

WESFED consists of 885 images, having 56:44 male female ratio. Statistical analysis of WESFED is shown in Table IX.

Table IX: Statistical detail of WESFED

| Expression | AN | DI | FE | HA | SA | SU | NE | Total | Percentage |
|---|---|---|---|---|---|---|---|---|---|
| #Samples | 81 | 65 | 44 | 200 | 143 | 137 | 215 | 885 | 100.0 % |
| #Male | 40 | 41 | 16 | 78 | 75 | 66 | 176 | 492 | 55.59 % |
| #Female | 41 | 24 | 28 | 122 | 68 | 71 | 39 | 393 | 44.41 % |

Existing datasets rarely addresses the issues of spontaneous expressions. Most of the time, images are acquired under a static environment with fixed illumination source with the front pose. On the other hand, real life scenarios are very different. We addressed all possible issues by considering images of different ethnicity, age, pose, illumination, and occlusion. None of the existing datasets contains child subject, which is the source of most natural expressions. In WESFED, we tried to emulate expression of children too. For pose variation, we considered in-plane and out of the plane rotation of the face. Illumination varies from brightest to darkest. Faces with glasses, beard, mustache, hair occlusion are present in the dataset for each of the seven expressions. Variety in ethnicity is obvious as images returned by search engine come from the web. Diversities in WESFED is presented in Figure 2.

People rarely express their feeling with fixed posture and gesture. Most of the past and current research is focused on expression recognition from posed faces. In WESFED, we tried to incorporate variations as much as possible from different dimensions. Most of the datasets has featured young male and female subjects. Age group of JAFFE dataset is 18-30, NVIE is 17-31. BU-3DFE covers wide range of 18-70 and MMI covers the subjects of age group 19-62. But none of the dataset includes the child subjects. In WESFED, age group varies from few month kid to 80+ old man. Thats another unique feature of WESFED compared to other datasets. Rarely search engine returns a subject with all seven expressions. In most of the cases, we could find a single expression from a subject. So unlike other person dependent database, WESFED does not require cross database testing as most of the subjects in database are different for all expression classes. Few samples from WESFED are shown in Figure 3.

## 4. CONCLUSION AND FUTURE WORK

This experiment may provide motivation to spontaneous facial expression recognition. Apart from stated diversities, the intensity of expressions varies from mild to peak. Next version of the database is under consideration. Idea is to add more images covering even higher degree of diversities. This database is person independent and hence does not need cross database validation. Like other datasets, WESFED also suffers from certain limitations. The database

Figure 2. Diversity in WESFED (Ethnicity, Age, Pose, Illumination Occlusion - from top to bottom)

Figure 3. Snapshot design of WESFED

is lacking statistical validation and certain expressions like disgust has a fairly lower number of samples compared to some other expression like happy and neutral. Feature enhancement of this issue is under consideration. Expressions and poses have many versatilities but the number of expressions are still limited to basic prototypic expression, which can be extended. Despite all this, WESFED is the first dataset of its own kind. It would serve as one of the motivational factors to gear up the research for spontaneous expression recognition.

## References

Bioid. `http://www.humanscan.de/support/downloads/facedb.php`.

Ekman dataset. `http://www.paulekman.com/product-category/facs/`.

Mmi. http://emotion-research.net/toolbox/toolboxdatabase.2006-09-28.5469431043.

Prism-asu. http://prism.asu.edu/3dface/default.asp.

Bartlett, M., G. Littlewort, M. F., Lainscsek, C., Fasel, I., and Movellan, J. 2006. Fully automatic facial action recognition in spontaneous behavior. In *IEEE 7th International Conference on Automatic Face and Gesture Recognition*. 223–230.

Bartlett, M., Littlewort, G., Fasel, I., and Movellan, J. 2003. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*. Vol. 5. 53–53.

Butalia, A., Ingle, M., and Kulkarni, P. 2012. Facial expression recognition for security. *International Journal of Modern Engineering Research 2,* 4, 1449–1453.

Chang, K., Bowyer, K., and Flynn, P. 2005. An evaluation of multimodal 2d+3d face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27,* 4, 619–624.

Chen, L. 2000. Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction. *University of Illinoisat UrbanaChampaign*.

Chen, L.-F. and Yen, Y.-S. 2007. Taiwanese facial expression image database. *Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan*.

Cid, F., Prado, J., Bustos, P., and Nunez, P. 2013. A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using gabor filtering. In *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 2188–2193.

Cohen, I., Sebe, N., Cozman, F. G., Cirelo, M. C., and Huang, T. S. 2003. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 595–601.

Corneanu, C., Simn, M., J.Cohn, and Guerrero, S. 2016. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38,* 8, 1548–1568.

Cowie, R., Douglas-Cowie, E., and Tsapatsoulis, N. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine 18,* 1, 32–80.

Donato, G., Bartlett, M., Hager, J., Ekman, P., and Sejnowski, T. 1999. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21,* 10, 974–989.

Douglas-Cowie, E., Campbell, N., Cowie, R., and Roach, P. 2003. Emotional speech:towards a new generation of databases. *Speech Communication 40,* 1, 33–60.

Ekman, P. and Friesen, W. V. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology 17,* 2, 124–129.

Ekman, P. and Friesen, W. V. 1978. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press*.

Fasel, B. and Luettin, J. 2003. Automatic facial expression analysis: A survey. *Pattern Recognition 36,* 1, 259–275.

Gaebel, W. and Wlwer, W. 1992. Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience 242,* 1, 46–52.

Goyani, M. and Patel, N. 2017. Judgmental feature based facial expression recognition systems and fer datasets-a comprehensive study. *International Journal of Next-Generation Computing 8,* 1, 62–81.

Huang, C. and Huang, Y. 1997. Facial expression recognition using model-based feature extraction and action parameters classification. *Journal of Visual Communication and*

*Image Representation 8,* 3, 278–290.

Kotsia, I., Buciu, I., and Pitas, I. 2008. An analysis of facial expression recognition under partial facial image occlusion. *Image and Vision Computing 26,* 7, 1052–1067.

Lyons, M. 1999. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21,* 12, 1357–1362.

Lyons, M. and Akamatsu, S. 1998. Coding facial expressions with gaborwavelets. In *3rd IEEE Conference on Automatic Face and Gesture Recognition*. 200–205.

Maat, L., Sondak, R., Valstar, M., Pantic, M., and Gaia, P. 2005. Mmi face database.

Martinez, A. and Benavente, R. 1998. The ar face database. *CVC Technical Report*.

Messer, K., Matas, J., J. Kittler, a. J. L., and Maitre, G. Xm2vtsdb: The extended m2vts database. In *2nd Internation Conference on Audio and Video based Biometric Person Authentication*.

Nasoz, F., Alvarez, K., Lisetti, C., and Finkelstein, N. 2004. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology and Work 6,* 1, 4–14.

OToole, A. 2005. A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27,* 5, 812–818.

Pantic, M., Pentland, A., Nijholt, A., and Huang, T. 2006. Human computing and machine understanding of human behavior: A survey. In *International Conference on Multimodal Interfaces*. 47–71.

Pantic, M. and Rothkrantz, L. 2000. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22,* 12, 1424–1445.

Pantic, M. and Rothkrantz, L. 2003. Toward an affect-sensitive multimodal human computer interaction. *Proceeding IEEE 91,* 9.

Pantic, M. and Rothkrantz, L. 2004. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34,* 3, 1449–1461.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. 2005. Web-based database for facial expression analysis. In *IEEE International Conference on Multimedia and Expo*.

P.Ekman, J.C.Hanger, C. and W.Irwin. Ekman-hanger facial action exeplars. *Human Interaction Laboratory, University of California*.

Phillips, P. 2005. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Phillips, P., Moon, H., Rizvi, S., and Rauss, P. 2000. The feret evaluation methodology for face recognition algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22,* 10, 1090–1104.

Savran, A., Sankur, B., and Bilge, M. T. 2010. Facial action unit detection: 3d versus 2d modality. In *IEEE Conference on Computer Vision and Pattern Recognition*. 71–78.

Schmidt, K. L. and Cohn, J. F. 2001. Dynamics of facial expression: Normative characteristics and individual differences. In *IEEE International Conference on Multimedia Expo*. Vol. 728. 547–550.

Sebe, N., Lew, M., Sun, Y., Cohen, I., Gevers, T., and Huang, T. 2007. Authentic facial expression analysis. *Image and Vision Computing 25,* 12, 1856–1863.

Sim, T., Baker, S., and Bsat, M. 2002. The cmu pose, illumination, and expression database. In *IEEE International Conference on Automatic Face and Gesture Recognition*.

T. Kanade, J. F. C. and Tian, Y. 2000. Comprehensive database for facial expression analysis. In *4th IEEE International Conference on Automatic Face and Gesture Recognition*. 46–53.

Tian, Y., Kanade, T., and Cohn, J. 2000. Recognizing upper face action units for facial expression analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 1. 1–19.

Wang, J. and Yin, L. 2007. Static topographic modeling for facial expression recognition and analysis. *Computer Vision and Image Understanding 108,* 1, 19–34.

Wang, S. 2010. A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Multimedia 12,* 7, 682–691.

Xue, Y. L., Mao, X., and Zhang, F. 2006. Beihang university facial expression database and multiple facial expression recognition. In *International Conference on Machine Learning and Cybernetics*. 3282–3287.

Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. J. 2006. A 3d facial expression database for facial behavior research. In *IEEE International Conference on Automatic Face and Gesture Recognition*. 211–216.

Zeng, Z., Pantic, M., Roisman, G., and Huang, T. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31,* 1, 39–58.

Zhou, C. 2003. The reviews of mental computation.

**Dr. Mahesh Goyani** has obtained Bachelors Degree in field of Computer Engineering from Veer Narmad South Gujarat University, Surat, India in 2005, and did post-graduation from Sardar Patel University, V.V.Nagar, India in 2007. He earned the Ph.D. Degree from at Charotar University of Science and Technology, Changa, India. At present, he is serving as an Assistant Professor at Department of Computer Engineering, Government Engineering College, Modasa, India. He has published number of books in area of Computer Graphics and Analysis of Algorithm. He also has published many research paper in reputed journals and conferences. His area of interest includes Pattern Recognition, Machine Learning, Data Mining, Data Analytics and Image Processing. Prof. Goyani has also served as a reviewer in many international journals. He was also invited as a session chair in International Conference on Computer Science, Engineering and Information Technology, Tirunveli, Tamilnadu, India in 2011. He has been a committee member in many national and international conferences.