International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

Survey on Emotion Recognition Using Facial Expression

Moe Moe Htay[a]*, Zin Mar Win[b]

[a,b]*University of Computer Studies, Mandalay, Myanmar*

[a]*Email: moemoehtay@ucsm.edu.mm*

[b]*Email: zinmarwin@ucsm.edu.mm*

Abstract

Automatic recognition of human affects has become more interesting and challenging problem in artificial intelligence, human-computer interaction and computer vision fields. Facial Expression (FE) is the one of the most significant features to recognize the emotion of human in daily human interaction. FE Recognition (FER) has received important interest from psychologists and computer scientists for the applications of health care assessment, human affect analysis, and human computer interaction. Human express their emotions in a number of ways including body gesture, word, vocal and facial expressions. Expression is the important channel to convey emotion information of different people because face can express mainly human emotion. This paper surveys the current research works related to facial expression recognition. The study attends to explored details of the facial datasets, feature extraction methods, the comparison results and futures studies of the facial emotion system.

Keywords: facial expression; facial features; feature extraction; emotion classification.

1. Introduction

In Artificial Intelligent era, Facial Expression Recognition is interesting and challenging task with the problems of limited dataset, different environments, pose, occlusion, person variation etc. FER systems have been applied many systems such as human-computer–interaction (HCI), games, animation of data-driven, surveillance, clinical monitoring etc., [2]. Ekman and Friesen, psychologists from America defined six universal facial expressions: fear, happiness, anger, disgust, surprise, and sadness and also explored Action Units based facial action coding system (FACS) to describe facial features of expressions [24].

----

* Corresponding author.

Some literatures work adding on other emotions neutral, contempt, and many compound facial emotions. Some researchers employed on handcrafted features extracted using algorithms and others employed on complicated features extracted using deep learning methods.

2. Current Research Literature Review

Yang and his colleagues (2018) presented a FER model using Haar Cascades face components detection and Neural Network (NN) to train the eye and adding mouth features on JAFFE Japanese database. Comparison of the result of proposed method with Sobel Edge detection methods is that the system has achieved more good accuracy. The problem of illumination and pose of the image and to make fully meet theory and practical requirements by integrating other biometric authentication methods and HCI perception methods is still existed [1].

Kalsum and his colleagues (2018) examined emotion recognition system using hybrid feature descriptors combining spatial Bag of features and spatial scale-invariant feature transform (SBoF-SSIFT) and classifiers of K-nearest neighbor. Codebook construction is applied after features extraction to represent large feature sets by grouping similar features into a specified cluster numbers. The experimentation accuracy has showed 98.33% and 98.5% on JAFFE and extended cohn-canade (CK+) dataset respectively. However, the recognition performance depends on the number of clusters for codebook generation, number of detected features, levels for image segmentation, and size of training dataset [2].

Qi and his colleagues (2018) implemented cognition and mapped binary pattern based FER using basic emotion model and circumplex model on CK+ with 100 images for training and 50 images for testing. In the preprocessing step, unwanted information such as hair, ear, and background are removed from the facial image. LBP and pseudo 3D model are used to extract the facial contours and to segment face area into sub-regions. To reduce the dimension of the features mapped local binary pattern is employed and then used two classifiers of SVM and softmax. The result found that local features and expressions are correlated. Moreover, the two classifiers have a little difference in performance. The existence of occlusion, complex conditions, and micro-expression recognition will be conducted in future FER system [5].

Shabat and his colleagues (2018) proposed a method Angled Local Directional Pattern (ALDP) for texture analysis of facial expression with six classifiers k-NN, SVM, DT, RF, Gaussian NB and Perceptron on CK+ dataset. Firstly, facial image was detected using Haar-like as [1] and then cropped and normalized the detected image. The accuracy improved 99% with ALDP method with no preprocessing [6]. Sreedharan and his colleagues (2018) also proposed Grey Wolf Optimization for feature selection and GWO-Neural Network (GWO-NN) for feature classification. The parts of face eyes, nose, mouth and ears are detected using Viola-John algorithm and then SIFT feature extraction is used feature points. The accuracy 89.79% on CK+ is less than [6] and achieved 91.22% [7].

Majumder and his colleagues (2016) used geometric feature extraction, regional local binary pattern (LBP) features extraction, fusion of both the features using auto-encoders and self-organizing map (SOM)-based classifier. The average accuracy 97.55% of MMI and 98.95% of CK+ database. The accuracy of SOM-based classifier is significant improvement over SVM with 3.94% increase for CK+ and 4.36% for MMI dataset respectively [9]. Chen and his colleagues (2018) explored multiple feature fusion applying Histogram of oriented gradients from three orthogonal planes (HOG-TOP) with experimentation of three datasets CK+, GEMEP-FERA 2011, and Acted Facial Expression in the Wild (AFEW) 4.0[13].

Zeng and his colleagues (2018) proposed a framework with high-dimensional features combination of appearance and geometric features. The system used deep sparse auto-encoders (DSAE) to learn robust discriminative feature and active appearance model (AAM) to locate the facial landmarks 51 points. Three feature descriptors HoG, gray value and LBP are utilized to describe the local features. Linear dimension reduction method of PCA is used to compress the features and then give the map as the input of DASE. The accuracy of the proposed framework achieved 95.79% of CK+ dataset by using leave on subject out cross-validation method [14].

Tang and his colleagues (2018) presented three models of differential geometric fusion network (DGFN) with extraction of handcrafted features, deep facial sequential network (DFSN) based on CNN with auto-extracted features, and DFSN-1 combination of the advantages of DGFN and DFSN by mapping and concatenation of handcrafted and auto-extracted features. DFSN-1 achieved the best performance among the three models on all of CK+, Oulu-CASIA and MMI dataset [19]. Mayya and his colleagues (2016) used deep convolutional neural network (DCNN) using caffe framework and Telsa K20Xm GPU. The frontal face is detected and cropped applied by openCV in facial images preprocessing from CK+ and JAFFE. The accuracy of experiment achieved 97% with leave-one-subject-out cross validation on CK+ and 98.12% with 10-folds cross validation on JAFFE [22].

Li and his colleagues (2018) released facial images database of real world affective facial database (RAF-DB). In the study, authors explored deep locality preserving convolutional neural network (DLP-CNN) to classify 7 classes and 11 class compound emotion expressions. The proposed network is more preferable for dataset under uncontrolled environment [16]. Tautkute and his colleagues (2018) classified 7-class and 3 class of facial expression using emotional deep alignment network (EmotionalDAN) model. AffectNet dataset (seven emotions, 68 facial landmarks) is used for training and CK+, JAFFE and the Indian spontaneous expression database (ISED) are used for testing. The result shows that 0.736% for CK+, 0.465% for JAFFE and 0.62% for ISED on seven classes and also achieves 0.921% for CK+, 63.45% for JAFFE and 0.896% for ISED on three classes respectively [11]. Slimani and his colleagues (2018) reviewed analysis of 22 Local Binary Pattern variances on JAFFE and CK databases using the simple parameter-free Nearest Neighbor classifier (1-NN). For JAFFEE database, the highest recognition accuracy achieved 97.14% by using dLBPα, ELGS and LTP, while CK database, the highest recognition rate of 100% by using AELTP, BGC3, CSALTP, dLBPα, nLBPd, STS, and WLD descriptors. The basic LBP descriptor achieved the acceptable performance of 95.71% on JAFFE and 99.28% of CK database. The study can be extended including other problems and other datasets. Sang and his colleagues used DCNN adding data augmentation, cross entropy and L2 multi-class SVM[3]. In [4], weighted center regression adaptive feature mapping (W-CR-AFM) for feature distribution and CNN for feature training on CK+, Radbound Faces database (RaFD), Amsterdam dynamic facial expression set (ADFES) and proprietary database. Different of other papers, spatial normalization and feature enhancement preprocessing methods are used. The recognition obtained 89.84%, 96.27%, 92.70% for CK+, RaFD and ADFES respectively. Munir and his colleagues (2018) address illumination problem of real world facial images using fast Fourier transform and contrast limited adaptive histogram equalization (FFT+CLAHE) for poor illumination and then applied merged binary pattern code (MBPC). PCA is used as a method of feature dimension reduction and k-NN as a classifier on SFEW dataset [8]. Guo and his colleagues (2017) released a new database iCV-MEFED at FG work-shop. Multi-modality CNN is compared with CNN for micro emotion recognition in the paper. The proposed network extracted firstly visual and geometrical information of features then concatenated these into a long vector. The feature vector is fed to the hinge loss layer. The framework is better performance than CNN with the misclassification of 80.212137 using caffe [17]. Guo and his colleagues (2018) also proposed another three works of the work-shop. The first winner method using CNN with geometric representation of landmark displacement leading better results compared with texture-only information. The recognition accuracy achieves 51.84% for seven expressions and 13.7% for compound emotion with the performance of average time 1.57ms using GPU or 30ms using CPU [10]. Barros and his colleagues (2017) employed deep emotional attention model using cross channel CNN by adding attention modulator on the bimodal face and body (FABO) benchmark database. The system applied CNN to learn the location of emotion expressions in a cluttered scene. The study has shown that the experimentation of one expression attention mechanism and two expression attention mechanism. The accuracy of the framework with attention is better than that of without attention [15]. Zhang and his colleagues (2018) proposed a robust facial landmark extraction method by combining data-driven of fully convolution network (FCN) and model-driven of pre-trained point distribution model (PDM) with three steps Estimation-Correction-Tuning (ECT). The computation of response maps of global landmark estimation is trained by FCN and then the maximum points of the maps are fitted with PDM to generate initial facial shape. In the final, a weighted version of regularized landmark mean-shift (RLMS) is applied to fine-tune the facial shape iteratively [18]. Fereira and his colleagues (2018) designed to learn NN architecture with three loss functions fully supervised, weekly supervised and hybrid regularization. The experimentation of the proposed model has achieved promising results on CK+, JAFFE under lab-environment and SFEW in the wild [20]. Yan and his colleagues (2018) proposed transductive deep transfer learning (TDTL) architecture to address the problem of cross-database non-frontal facial expression recognition applying VGGface 16-Net on BU-3DEF and Multi-PIE datasets. The study found that feature representation with VGG network is better than traditional handcrafted features such like SIFT and LBP to represent complicated features [21]. In the [23] paper, Zheng and his colleagues (2018) also used the two datasets for the experimentation to address the problem of cross-domain and cross-view of facial expressions using transductive transfer regularized least-square regression (TTRLSR) model, color SIFT (CSIFT) features with 49 landmarks and SVM classifiers. The two databases have only four identical categories neutral, surprise, happy and disgust. The experimentation of the study conducted two kinds cross-domain and same view and cross-view and same domain. PCA algorithm also used to reduce the dimension of features. The studies in references [1, 2, 5, 9] classified six basic emotions as Happiness, Angry, Sadness, Surprise, Fear, Disgust. In [3,4,7,8,11] have classified one more class as Neutral and [6,10,11] have done Contempt class. All of eight classes have been classified by the studies in [14,17,19]. However, [21] and [23] have worked on Neutral, Happiness, Surprise and Disgust expressions. The paper [13] employed with 5 classes of GEMEP-FERA 2011 database and 7 classes of CK+ and AFEW. Li and his colleagues [16] explained seven basic emotions and 11 compound emotions sadly fearful, sadly angry, sadly surprised, happily surprised, happily disgusted, sadly disgusted, fearfully angry, fearfully surprised, angrily surprised, angrily disgusted and disgustedly surprised. The paper [20] has worked classification 6 universal classes of JAFFE, SFEW with classes of 6 basic and neutral, and CK+ with 8 classes including contempt.

3. State of the Problem

FER system is need to be taken up as research problems of occlusion, illumination, poses, lighting, viewing angle for the real world environment. The major challenges of the research include:

- The research works on micro-expression and compound emotion are limited.
- Most of researches classify basic emotion but fine-grain emotion is relatively small.
- Mathematical model is needed to be created for extracting more discriminant features facial images in the wild.
- Real time emotion recognition systems are needed to be developed to meet practical applications.

4. Comparison of Research Works

Based on the several literatures [1-24], Table I describes the comparison of the most prominent research papers.

Table 1: Comparison of Major Research Work

Authors (year)	Findings and Limitation
Yang and his colleagues (2018)	• The eye and mouth are important parts for facial expression recognition. • Non-frontal faces were not considered. • Illumination and pose of the real-life images are not included. • FER system needs to fully meet practical requirements.
Qi and his colleagues (2018)	• Local facial features and expression are correlated • Circumplex model works well than discrete emotion model. • Occlusion, complex condition and micro expression recognition were conducted in future.
Majumder and his colleagues (2016)	• Fusion of features • SOM-based classifier is significant improvement over SVM. • Real-life application would be developed. • Facial occlusion need to address for researchers.
Munir and his colleagues (2018)	• Use FFT-CLAHE • Real world images were implemented. • Increase number of features tends to decrease accuracy because of redundant features. • Illumination problem is considered in the paper. • This could be tried to handle noise in the preprocessing stage.

5. Typical Facial Expression Recognition

Typical FER system is showed in the following system flow Fig.1. In the detection of face consists of three works: locate the face, crop the face, scale the face. Haar-cascade method is mostly used in the face detection stage. Suitable methods for features extraction, dimension reduction and classification could be selected.

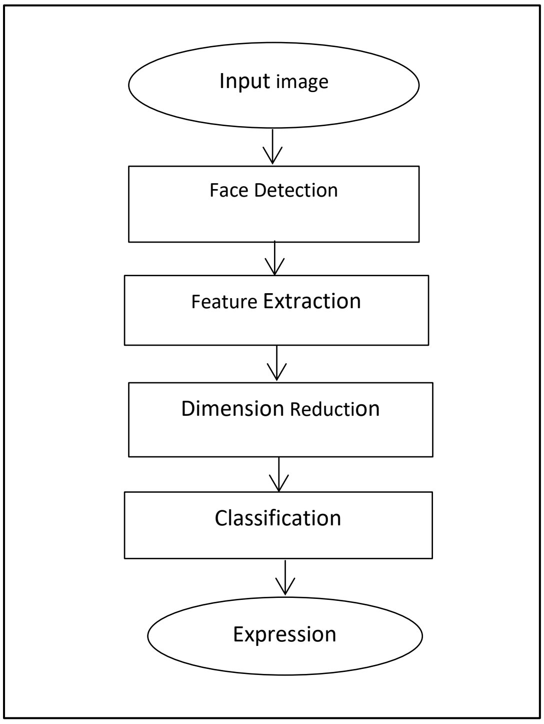

Figure 1: Typical FER System

6. Feature of Facial Images

Most of the FER system used geometrical features or visual features or both of these features to extract the features from the images of face

6.1. Geometrical methods

Geometrical methods can estimate facial landmarks location or some components of facial images such as the eyebrows, the mouth, and the nose and these features can be measured by distances, curvatures, deformations, and other geometric properties to represent the geometric facial features as they are sensitive to noise [5, 7, 9 ,10 ,13 ,17].

6.2. Appearance methods

Appearance methods such as Scale Invariant Feature Transform (SIFT), Gabor appearance, Local Phase Quantization can detect the multi-scale, multi-direction of the local texture changes on either specific regions or the whole face to encode changes on either specific regions or the whole face to encode the texture [6, 7, 9, 13, 17]. The effects of the approached are time-consuming, and the characteristic dimension is huge, so the dimensionality reduction methods are used to affect the accuracy of facial expression recognition.

7. Datasets

Facial expression datasets have two types of creation of images: posed expressions images and spontaneous expressions images datasets. Researchers acquire facial images in three ways such as peak expression images only, image sequences portraying an emotion from neutral to its peak, and video clips with emotional annotations. The two widely used datasets are CK+ and JAFFE [25-28]. The real world facial databases are FER-2013, FERG-DB, SFEW2.0 (static facial expression in the wild), RAF-DB (real world affective face database) and AffectNet database.

7.1. Extend Cohn-Canade Dataset (CK+)

CK+ data set has been widely used in many years in facial expression system. This data set comprises of 593 sequences of image vary in duration from 10 to 60 frames collected from 123 subjects. The age range of subjects is 18-50 years, where 31% are men and 69% are women. The images express seven categories of expressions: happy, sad, surprise, anger, fear, disgust, and neutral that cover the basic emotions. Each image has 640 * 640 or 490 pixels resolution.

7.2. Japanese Female Facial Expression Dataset (JAFFE)

JAFFE data set is also widely used in expression recognition of human emotion. This dataset consists of 213 images of 10 Japanese females including seven expressions: six basic (happy, surprise, sad, anger, fear and disgust) and neutral. Each image has the resolution of 256 * 256 pixels.

7.3. FER-2013 Dataset

FER-2013 2013 data set contains 28,000 images that are labeled. The dataset is created in 2013 for learning focused on three challenges: the black box learning, the facial expression recognition challenges and the multimodal learning challenges. The images are 48 * 48 pixels grayscale of faces in seven expressions: six basic expression and neutral.

7.4. FERG-DB Dataset

FERG-DB stands for facial expression research group database that consists of face images of six stylized characters grouped into seven types of expressions: six basic expressions and neutral. The dataset includes 555767 images.

7.5. Static Facial Expression in the Wild Dataset (SFEW)

The images in the SFEW are extracted from a temporal facial expressions databases Acted Facial Expressions in the Wild (AFEW) which has been extracted from movies. The database contains 700 images that have been labeled into six basic expressions.

7.6. Real-World Affective Face Database (RAF-DB)

RAF-DB database is a large-scale facial expression database that includes facial images downloaded from internet. The dataset is annotated seven dimensional expression distribution vectors for each image.

7.7. AffectNet Dataset

AffectNet is a largest database of facial expression in the real-world and contains more than 1,000,000 facial images downloaded from the internet search by six different languages with 1250 emotion related keywords. The database defined eleven categories of expression: six basic expressions, neutral, contempt, none, uncertain, and non-face.

8. Conclusion and Recommendations

Facial expression recognition is an active research area and more interesting for researcher under the problem of occlusion, brightness, viewing angle, pose, and background in the real-life images, sequence of images and videos. This review paper has presented methods of preprocessing, feature extraction and classification scheme. Most of the studies are experimentation of under controlled environment facial images. In the paper, the methods of face detection, feature extraction, and expression recognition are presented mainly. The FER research goes on to meet real-life applications for driver drowsiness recognition, assistant of distance learning, clinical patient monitoring and teaching robot, health care system for autism children. In the future, FER system will be developed for fined grained facial expressions recognition and compound emotions recognition by using facial images.

Acknowledgements

The authors wish to thank Professor Dr. G.R. Sinha for teaching how to read literature of research and how to write survey paper.

References

[1]. Yang, D., Abeer Alsadone, P.W.C., Dingh, A. K., and Elchouemi, A. An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environmet. Procedia Computer Science 125 (2018) 2-10.

[2]. Kalsum, T., Anwar, S. M., Majid, M., Khan, B., & Ali, S. M.(2018) Emotion recognition from facial expressions using hybrid feature descriptors. 12(6), IET Image Processing, 1004-1012.

[3]. Sang, D. V., and Van Dat, N. (2017, October) "Facial expression recognition using deep convolutional neural networks" In Knowledge and Systems Engineering (KSE), 2017 9th International Conference on, pp. 130-135). IEEE.

[4]. Wu, B.F., and Lin, C.H. "Adaptive Feature Mapping for Customizing Deep Learning Based Facial Expression Recognition model. IEEE Access 6, 2018; 12451-12461.

[5]. Qi, Chao, et al. "Facial Expressions Recognition Based on Cognition and Mapped Binary Patterns" (2018):18759-18803, IEEE Access 6.

[6]. Shabat, A.M., and Tapamo, J.R. "Angled local directional pattern for texture analysis with an application to facial expression recognition", ISSN 1751-9632, IET Computer Vision (February 2018).

[7]. Sreedharan, N. P. N., Ganesan, B., Raveendran, R., Sarala, P., and Dennis, B. (September, 2018) "Grey Wolf optimization-based feature selection and classification for facial emotion recognition" IET Biometrics, doi: 10.1049/iet-bmt.2017.0160.

[8]. Munir, A., Hussain, A., Khan, S. A., Nadeem, M., & Arshid, S. (2018)"Illumination invariant facial expression recognition using selected merged binary patterns for real world images"Optic 158: 1016-1025.

[9]. Majumder, A., Behera, L., & Subramanian, V. K. "Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion". IEEE Transactions on cybernetics. (2016).

[10]. Guo, Jianzhu, et al. "Dominant and Complementary Emotion Recognition From Still Images of Faces" IEEE Access (April 2018), Volume 6, 26391-26403.

[11]. Tautkute, , I., Trzcinski, T., and Bielski, A. (2018). I Know How You Feel: Emotion with Facial Landmarks, arXiv: preprint arXiv:1805.00326.

[12]. Slimani, K., Kas, M., El Merabet, Y., Messoussi, R., & Ruichek, Y. (2018, March), "Facial emotion recognition : A comparative analysis using 22 LBP variants" In Proceedings of the 2nd Mediterranean Conference on Pattern Recognition and Artificial Intelligence. (pp. 88-94). ACM.

[13]. Chen, J., Chen, Z., Chi, Z., & Fu, H.(2018). Facial expression recognition in video with multiple feature fusion. IEEE Transactions on Affective Computing, 9(1), 38-50.

[14]. Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., & Dobaie, A. M. (2018) "Facial expression recognition via learning deep sparse autoencoders" Neurocomputing, 273, 643-649.

[15]. Barros, P., Parisi, G.I., Weber, C., & Wermter, S. (2017). Emotion-modulated attention improves expression recognition: A deep learning model. Neurocomputing, 253, 104-114.

[16]. Li, S., and Deng, W. "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition", IEEE trans. On Image Processing, 28(1), 356-370, 2019.

[17]. Guo, J., Zhou, S., Wu, J., Wan, J., Zhu, X., Lei, Z., & Li, S. Z. (2017, May). Multi-modality network with visual and geometrical information for micro emotion recognition. In Automatic Face and Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on (pp.814-819). IEEE.

[18]. Zhang, H., Li, Q., Sun, Z., and Liu, Y. (2018). Combining data-driven and model-driven methods for robust facial landmark detection. IEEE Transactions on Information Forensics and Security, 13(10), 2409-2422.

[19]. Tang, Y., Zhang, X. M., & Wang, H. (2018). Geometric-Convolutional Feature Fusion Based on Learning Propagation for Facial Expression Recognition. IEEE Access, 6, 42532-42540.

[20]. Ferreira, P. M., Marques, F., Cardoso, J. S., & Rebelo, A. (2018). Physiological Inspired Deep Neural Networks for Emotion Recognition. IEEE Access, 6, 53930-53943.

[21]. Yan, K., Zheng, W., Zhang, T., Zong, Y., & Cui, Z. (2018). Cross-database non-frontal facial

expression recognition based on transductive deep transfer learning. arXiv preprint arXiv:1811.12774.

[22]. Mayya, V., Pai, R. M., & Pai, M. M. (2016). Automatic facial expression recognition using DCNN. Procedia Computer Science, 93, 453-461.

[23]. Zheng, W., Zong, Y., Zhou, X., & Xin, M. (2018). Cross-domain color facial expression recognition using transductive transfer subspace learning. IEEE transactions on Affective Computing, 9(1), 21-37.

[24]. Ekman, P., Friesen, W.V., Facial Action Coding System a technique for the measurement of facial movement. Palo AltO: Consulting Psychologists Press, pp. 271-302, 1978.

[25]. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on (pp. 94-101). IEEE.

[26]. Lyons, M. J., Akamatsu, S., Kamachi, M., Gyoba, J., & Budynek, J. (1998, April). The Japanese female facial expression (JAFFE) database. In Proceedings of third international conference on automatic face and gesture recognition (pp. 14-16).

[27]. Loob, C., Rasti, P., Lüsi I., Junior, J.C.J., Baró, X., Escalera, S., Sapinski, T., Kaminska, D. and Anbarjafari, G., 2017, May. Dominant and complementary multi-emotional facial expression recognition using c-support vector classification. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on (pp. 833-838). IEEE.

[28]. Dhall, A., Goecke, R, Lucey, S., & Gedeon, T.Static facial expression in the wild: data experiment protocol. CVHCI Google Scholar.