Machine Learning Techniques for Automatic Depression Assessment

Anna Maridaki[*], Anastasia Pampouchidou[1†], Kostas Marias[*‡], and Manolis Tsiknakis[*‡]
[*]Department of Informatics Engineering, Technological Educational Institute of Crete (TEI), Heraklion, Crete Greece
[†]Laboratoire Electronique, Informatique et Image, Université de Bourgogne, 27011 Le Creusot, Bourgogne France
[‡]Institute of Computer Science (ICS), Foundation for Research and Technology (FORTH), Heraklion, Crete Greece
Email: anna.maridaki@gmail.com

***Abstract*—Depression is one of the most common mood disorder that is inherently related to emotions, involving bad mood, low self-esteem and loss of interest in normal pleasurable activities. The aim of this work is to develop a framework based on the dataset provided by AVEC'14 for depression assessment. The proposed work presents two different motion representation methods: a) Gabor Motion History Image (GMHI), and b) Motion History Image (MHI). Several combinations of appearance-based low level features are extracted from both motion representations. These features were further combined with statistically derived features, and used for training and testing with several machine learning techniques. The proposed approach reached an F1 score of 81.93%, both for MHI and GMHI, with SVM classifier. The achieved performance is comparable to state-of-the-art approaches, while manages to outperform several others. Apart from accomplishing a competitive performance, the proposed work provides an exhaustive exploration of different combinations of the investigated motion representations, descriptors, and classifiers.**

***Keywords*—Affective computing; Depression assessment; Gabor motion history image; Machine learning; Motion history image**

I. INTRODUCTION

Depressive Disorder is classified in the category of mood disorders causing distressing symptoms, affecting feelings and difficulty to handle daily activities. Depression is enlisted as the fourth most significant cause of suffering worldwide and is predicted as the leading cause in 2020. Currently, assessment relies on judgment of experienced professionals. Thus, there is no objective method for depression diagnosis, and as a result misdiagnosing affected individual is a common obstacle, running the risk of subjective biases. According to the World Health Organization (WHO) [1] depression detection failure can be due to the lack of resources and trained care providers. Despite the scientific studies that have already been conducted, remarkably little innovation has occurred in the clinical care of depressive disorder. The most widespread diagnostic method is based on diagnostic and statistical manual of mental disorder (DSM) criteria [2]. A clinical diagnosis can further be supported by self-report instruments, such as the Beck Depression Inventory (BDI) [3].

[1] Anastasia Pampouchidou was funded by the Greek State Scholarship Foundation, under the scholarship program instituted in memory of Maria Zaousi.

Automated depression diagnosis systems have been developed in order to support clinicians' decision and avoid mistaken diagnosis. Such systems could also be helpful to overcome the problem of subjective bias associated with self-reports. Those approaches are based on a variety of non-verbal manifestation of depression. Facial expressions play an integral role in order to detect symptoms of Depression as depressed individuals exhibit restricted facial motion than non-depressed [4].

The "Audio/Visual Emotion Challenge" AVEC 14 [5] provides a dataset which could assist the development of an automated depression detection system. The provided dataset is consisted by video recordings, annotated with a depression index. In the present work we developed an automated framework for categorical depression assessment based on low level features. Two motion representation methods are implemented: a) Motion History Image, and b) Gabor Motion History Image, where moving parts of a video sequence can be engraved in a single image. The main contribution of the proposed work is the comparison of several machine learning algorithms.

The rest part of the paper is organized as follows. Section II briefly reviews the related work in this field while Section III describes the proposed methodology. Experimental results and recommendation for future work are given in Section IV and V respectively.

II. RELATED WORK

Several approaches, for categorical depression assessment based on AVEC 14 dataset, can found in literature. Senoussaoui et al. [6] employed Local Gabor Binary Patterns in Three Orthogonal Planes and achieved 82% accuracy. Another approach utilizing Local Curvelet Binary Patterns in Pairwise Orthogonal Planes presented 74.5% accuracy [7]. A result of 72.8% was attained by Pampouchidou et al. [8] where a multimodal approach was presented. Further, Alghowinem et al. [9] in their cross-corpus approach used eye gaze and head pose

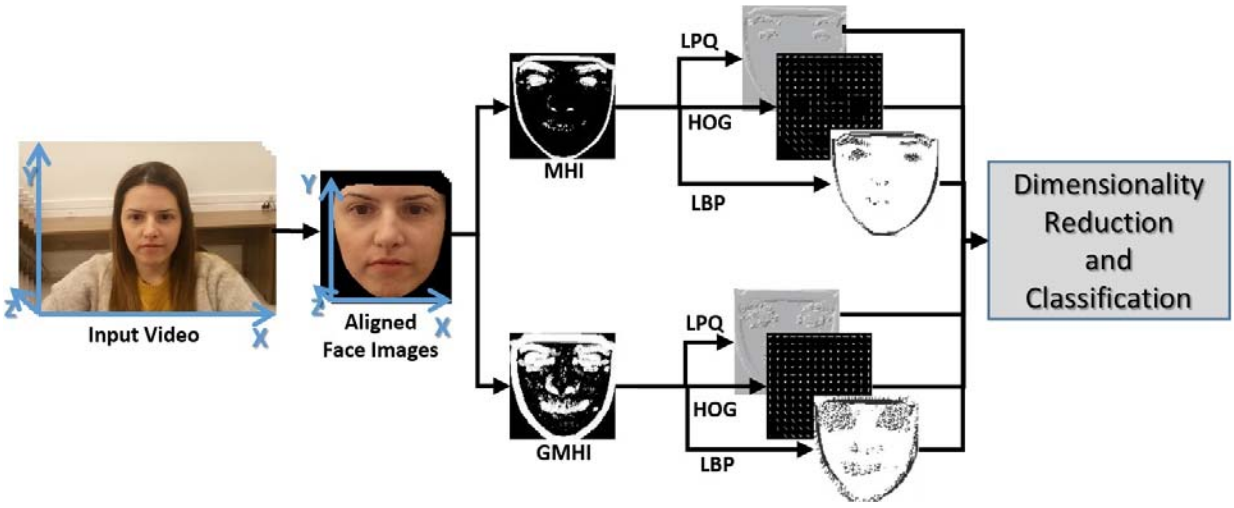


Fig. 1. Pipeline of the Proposed Algorithm

features and achieved recall of 81.3%. Motion History Image (MHI) represents a motion sequence in compact manner, in a scalar-valued image. Valstar et al. [10] implemented Motion History Image for facial action recognition. Another approach of motion representation, Landmark Motion History Image, was introduced, where motion history image derive from sequences of facial lanmarks [11]. In [12] different motion reperesentation methods were implemented succeding 87.4% F1 score with the use of Convolution Neural Networks.

III. METHODOLOGY

This section includes the research strategy that was followed during the development of the current approach. The proposed workflow is illustrated in Fig.1. The first step of the method is the preprocessing, where the Region of Interest (ROI) is detected. In this work this region is the whole face. For the video preprocessing, open source software OpenFace was used [13], where aligned facial images sized 112 × 112 are extracted. OpenFace provides with a binary value of "success" for each aligned image, which specifies the successful and the unsuccessful facial landmarks detection. Score "1" represents successful detection, and "0" the failure. In the current work only the successfully detected aligned images were used for the present framework development.

The Audiovisual Emotion Challenge (AVEC) dataset has been collected for measuring and monitoring depression severity. Depression severity is estimated as a continuous variable by using Beck Depression Inventory score. A subject is classified in a category according to the depression severity level. The standard BDI cut-offs are:

0–13: Minimal depression

14–18: Mild depression

19–28: Moderate depression

30–63: Severe depression

For the categorical depression assessment the cutoff was set in 13/14 points. With such a cutoff 96 subjects were characterized as depressed, and 104 as non-depressed.

A. Motion Representation

In the present work the implemented methods derived from two different motion representation algorithms: (1) Motion History Image (MHI), and (2) Gabor Motion History Image (GMHI).

1) Motion History Image

A view-based approach for action representation and recognition has been developed. Motion History Image algorithm computes a static, scalar-valued image where intensity is a function of recency of motion. Motion is condensed into gray scale images in which the most recent action in a video sequence is illustrated with white pixels. Gray scalar values represent movement that happened less recent. The MHI algorithm is applied at the aligned face image sequence. The MHI can be generated using difference of frames (DOF)

$$\Psi(x,y,t) = \begin{cases} 1 & if\ D(x,y,t) > \xi \\ 0 & otherwise \end{cases} \quad (1)$$

where Ψ(χ,y,t) is the binarization of the difference of frames by considering a threshold ξ, and ξ is the minimal intensity difference between two images. ξ value was empirically set 25, so that motion appearing in the static background due to illumination changes is not represented in the MHI.
D(x,y,t) is defined with difference distance Δ as:

$$D(x,y,t) = |I(x,y,t) - I(x,y,t \pm \Delta)| \quad (2)$$

The MHI $H_T(x,y,t)$ is computed based on:

$$H_T(x,y,t) = \begin{cases} \tau & if\ \psi = 1 \\ max(0, H_T(x,y,t-1) - 1) & otherwise \end{cases} \quad (3)$$

2) Gabor Motion History Image

The second motion representation algorithm implemented in the present work is Gabor Motion History Image. The algorithm has the same operation as the described MHI. For the GMHI computation, Gabor inhibited images were utilized instead of the actual video frames, as they depict the most relevant and important image information reducing the effect of background

texture [14]. The GMHI algorithm is applied at the aligned face image sequence of each video.

In each aligned image a bank of Gabor filters with multiple orientations and wavelengths are applied.

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = e^{-\frac{\tilde{x}^2+\gamma^2\tilde{y}^2}{2\sigma^2}} \cos\left(2\pi\frac{\tilde{x}}{\lambda}+\varphi\right) \quad (4)$$

with:

$$\tilde{x} = x\cos\theta + y\sin\theta, \tilde{y} = -x\sin\theta + y\cos\theta \quad (5)$$

where x,y is Pixel location, λ wavelength, θ orientation, φ phase, σ Standard deviation of the Gaussian γ Spatial aspect ratio

In order to reduce the effect of background texture, anisotropic inhibition was developed [14]. Background texture is estimated and then it is subtracted from the Gabor filter image:

$$w(x,y) = \frac{1}{\|g(DoG)\|_1} g\big(DoG(x,y)\big) \quad (6)$$

Gabor Motion History Image method was developed defining empirically threshold ξ to 8. This configuration was set so that the static background is not represented in the GMHI. Applying Gabor filters and the Gabor inhibited algorithm to the aligned face image of a video compute motion representation GMHI. An example of GMHI computation is illustrated in Fig.2.

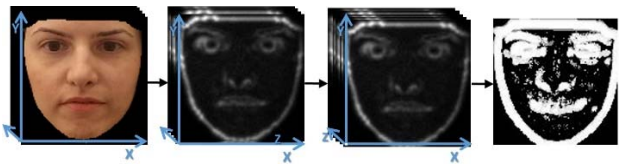


Fig. 1. Process Flow of Gabor Motion History Image

B. Feature Extraction

The Histogram of Oriented Gradients (HOG) algorithm is based on edge information which is described by the distribution of intensity gradients or edge directions. The source picture is divided into a dense grid of small region called cells (8×8 pixels). For each pixel within the cell the vertical and horizontal gradients are computed using 1-D Sobel operators. HOG descriptor results a 1×6084 feature vector.

Local Binary Pattern descriptor (LBP) is an efficient texture operator which aims to encode the local structures around each pixel. For each pixel a binary vector is computed by the comparison of the pixel's intensity with those of its neighbors. The image is divided into overlapping cells (neighborhoods) usually consisted by 3×3 pixels. The center pixel value is subtracted with its eight neighbor's value following the pixels along a circle (clockwise). This procedure end up to an 8-digit binary number. Two set of {radius, neighborhood} were tested hereby. Set {1,8} results in an 1×59, and {2,16} results in an 1×243 feature vectors respectively.

The Local Phase Quantization (LPQ) is based on the blur invariance property of the Fourier phase spectrum. The image is divide into blocks where a short-time Fourier transform (STFT) is applied to extract local phase information. LPQ descriptor results in an 1×256 feature vector.

It was not only appearance-based descriptors that were implemented, but also same statistical features were utilized. Mean (average value), standard deviation (mean deviation of the signal compared to the average) and combined histogram are extracted from motion representation images (MHI and GMHI) and regarded as a single descriptor [HIST-MEAN-STD]. A visualization of all the proposed features, in combination with both MHI and GMHI, is illustrated by Fig. 3.

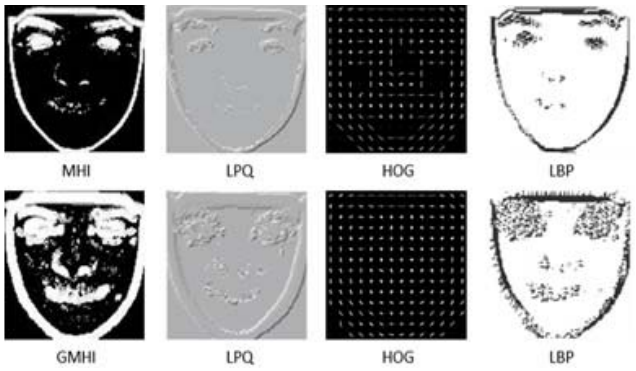


Fig. 2. Visualization of the proposed features for both MHI and GMHI

C. Dimensionality Reduction and Classification

In order to reduce the computational time of classification and the irrelevant features we use a Principal Component Analysis (PCA) model. New variables are linear combinations of the original feature, chosen to capture as much on the original variance as possible. For a feature matrix m×n, with m rows (samples) and n columns (features), the principal components vectors are the eigenvectors of the n×n coefficient matrix, ordered by decreasing magnitude of the corresponding eigenvalue. PCA method reduces feature dimensionality of the original data by projecting the data into a lower dimensional space. The smaller feature set is used as the extracted features for the classification step.

In the classification step, the feature vector that is provided by the feature extraction step is used by a selected classifier in order to evaluate the proposed system. The machine learning algorithms tested in the present work is Naïve Bayes, k-Nearest Neighbors, Random Forest, and Support Vector Machine.

IV. Experimental Results

Several configurations of the parameters involved in the proposed algorithms were tested trying to optimize the overall results, succeeding the maximum F1 score, which is computed based on the confusion matrix C, precision, and recall, defined as follows:

$$C = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (7)$$

$$precision = \frac{TP}{TP+FP} \quad (8)$$

$$recall = \frac{TP}{TP+FN} \quad (9)$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (10)$$

Leave-One-Subject-Out cross validation method was used as more than one recording for each subject exists, which could introduce a bias to the classification process. The combined extracted features HOG, LBP{1,8}, LBP{2,16}, LPQ, and HIST-MEAN-STD result in 31 different combinations. Those combinations are used as different feature vector inputs. The selected values for Principal Component Analysis are 10, 50, 100 and 150; hereby only the best performing (10 and 100) have been reported.

Having tested several configuration set of parameters and comparing the classifier's performance, we noticed that feature combinations results do not outperform results of single descriptors, which is obvious in both Fig.4 and Fig.5. Thus in the following comparison tables (Table I and Table II) feature combination results have not been included. Presented results include F1 score of LBP{1,8}, LBP{2,16}, HOG, LPQ, HIST-MEAN-STD, and feature fusion of all the above together.

For the MHI approach Table I presents the best performance for PCA 10 and 100 for the different classifiers. The best performance is 81.93% achieved with the appearance-based descriptor HOG as well as with the HIST-MEAN-STD, both using SVM classifier for 100 selected features.The best results for Gabor Motion History Image approach are presented in Table II. We chose the F1 Scores for PCA both 10 and 100 for all implemented classifiers. As for GMHI approach the maximum F1 score is 81.93% achieved by the combination of statistical features HIST-MEAN-STD for 10 selected features.

In summary, both MHI and GMHI perform 81.93% F1 score with SVM classifier, which outperforms the rest of the tested classifiers. More specifically, the best F1 score achieved among kNN, Naïve Bayes, and Random Forest was 64.34%, which is much lower than the one achieved by SVM. HOG descriptor does not perform well for PCA 100 for kNN classifier in both MHI and GMHI, but when tested with SVM classifier it reaches up to 80% F1 score in both approaches. LBP descriptor did not perform well at all within the proposed framework. LBP{1,8} is a vector 1×59 thus for PCA 100 and 150 the algorithm cannot be executed. Testing different {radius, neighborhood} parameters could potentially improve LBP descriptor performance.

V. Conclusions

The proposed work presented with an exhaustive test of several variants in terms of motion representation, feature extraction, as well as different classification techniques. The achieved performance outperforms several previous works, while performing comparably to most by reaching F1 score of 81.93%. However, there is still room for further exploitation. A multimodal approach of the implemented framework would be desirable, by combining visual with audio-based features. Another expansion could be more tests with additional classifiers as well as more extracted features could improve the overall performance of the algorithm. Finally, deep learning techniques, which have been proven to perform competitively, could improve the overall high performance presented herein.

TABLE I. F1 Score Results of the Implemented Classifiers for MHI

	PCA 10				PCA 100			
Features	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**
LBP{1,8}	*45.50%*	*35.50%*	*42.71%*	-	-	-	-	-
LBP{2,16}	*49.74%*	*29.09%*	*47.24%*	-	*64.34%*	*27.21%*	*42.11%*	-
HOG	*32.50%*	*38.51%*	*42.62%*	*79.80%*	-	*1.98%*	*57.58%*	***81.93%***
LPQ	*18.90%*	*43.31%*	*50.00%*	*60.20%*	*52.80%*	*31.08%*	*49.29%*	*59.30%*
HIST-MEAN-STD	*64.08%*	*37.57%*	*41.88%*	*80.00%*	*55.02%*	*42.53%*	*54.11%*	***81.93%***
FEATURE FUSION	*64.08%*	*38.37%*	*54.73%*	*54.55%*	*58.56%*	*36.84%*	*61.14%*	*36.59%*

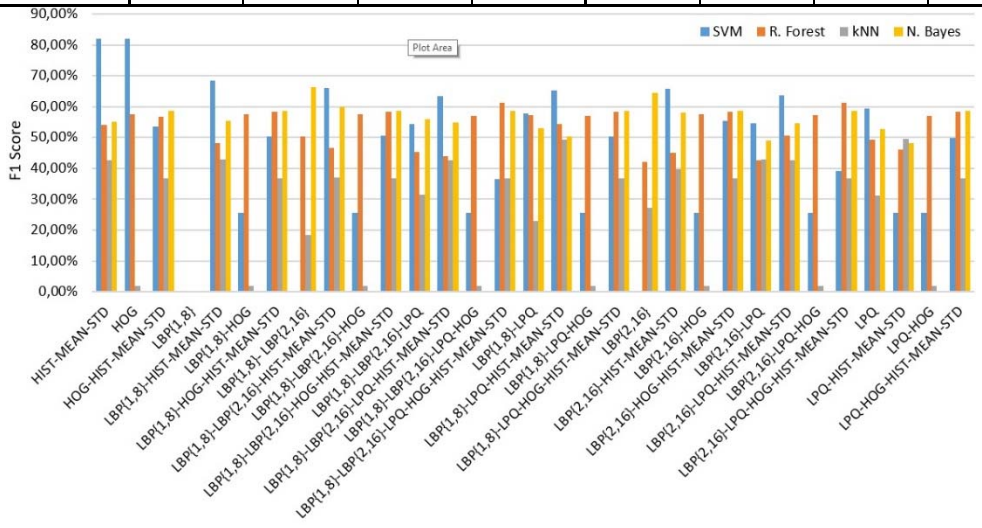
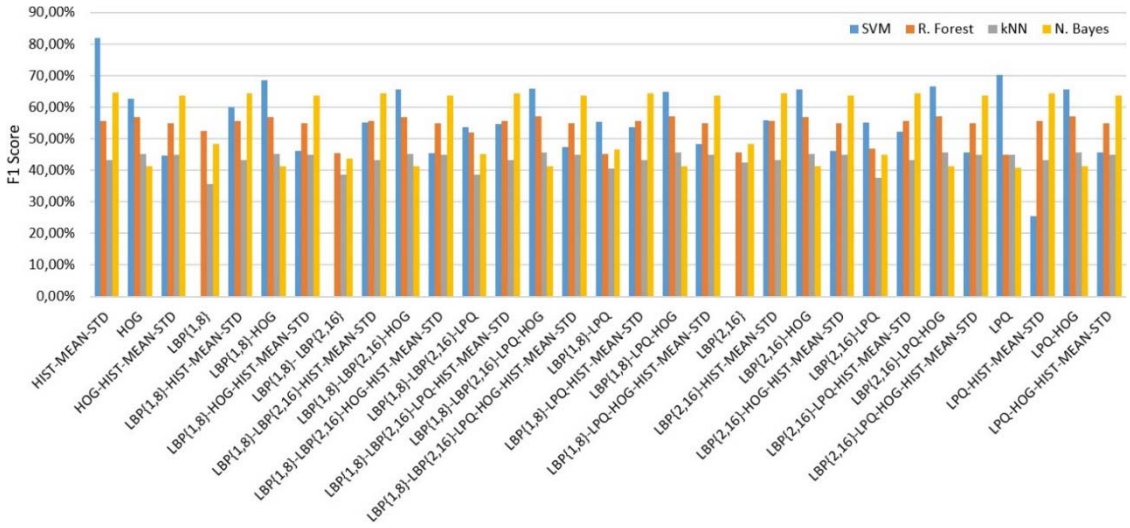


Fig. 3. Performance of the Implemented Classifiers for MHI for all combinations of the feature vectors

TABLE II. F1 SCORE RESULTS OF THE IMPLEMENTED CLASSIFIERS FOR GMHI

	PCA 10				PCA 100			
Features	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**
LBP{1.8}	*48.31%*	*35.58%*	*52.53%*	-	-	-	-	-
LBP{2,16}	*48.35%*	*42.50%*	*45.69%*	-	*62.41%*	*40.52%*	*49.76%*	-
HOG	*41.18%*	*45.24%*	*56.84%*	*62.72%*	*2.06%*	*2.02%*	*46.39%*	*80.00%*
LPQ	*40.76%*	*44.97%*	*45.03%*	*70.39%*	*52.54%*	*22.90%*	*48.70%*	*69.83%*
HIST-MEAN-STD	*64.69%*	*43.24%*	*55.67%*	***81.93%***	*60.40%*	*57.73%*	*53.14%*	*80.00%*
FEATURE FUSION	*63.77%*	*44.92%*	*55.00%*	*47.42%*	*50.98%*	*6.56%*	*51.65%*	*63.95%*



Performance of the Implemented Classifiers for GMHI for all combinations of the feature vectors

REFERENCES

[1] W. H. Organization, The global burden of disease: 2004 update, Geneva: World Health Organization, 2008.

[2] Association American Psychiatric (APA), Diagnostic and Statistical Manual of Mental Disorders , Fifth Edition, USA, 2013.

[3] A.T. Beck et al., "An Inventory for Measuring Depression," *Arch Gen Psychiatry,* pp. 561-571, 1961.

[4] A. Pampouchidou et al., "Automatic Assessment of Depression Based on Visual Cues: A Systematic Review," *IEEE Transactions on Affective Computing,* vol. PP, no. 99, pp. 1-1, 2017.

[5] M. Valstar et al., "AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge," in *Proceedings of the 4th Intern. Workshop on Audio/Visual Emotion Challenge*, Orlando, Florida, USA, 2014.

[6] Mohammed Senoussaoui, et al., "Model Fusion for Multimodal Depression Classification and Level Detection," in *4th Intern. Workshop on Audio/Visual Emotion Challenge*, Orlando, Florida, USA, 2014.

[7] A. Pampouchidou et al., "Video-based depression detection using local Curvelet binary patterns in pairwise orthogonal planes," in *38th IEEE EMBC*, 2016.

[8] A. Pampouchidou et al., "Facial geometry and speech analysis for depression detection," in *39th IEEE*-EMBC, 2017.

[9] S. Alghowinem et al., "Cross-cultural detection of depression from nonverbal behaviour," in *11th IEEE Automatic Face and Gesture Recognition* , 2015.

[10] M. Valstar et al., "Motion history for facial action detection in video," in *2004 IEEE Intern. Conf. on Systems, Man and Cybernetics*, 2004.

[11] A. Pampouchidou, et al., "Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text," in *6th Intern Workshop on Audio/Visual Emotion Challenge*, Amsterdam, The Netherlands, 2016.

[12] A. Pampouchidou et al., "Quantitative comparison of motion history image variants for video-based depression assessment," *EURASIP Journal on Image and Video Processing,* p. 64, 1 December 2017.

[13] T. Baltrusaitis et al., "OpenFace: an open source facial behaivior analysis toolkit," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 2016.

[14] A. Cruz et al., "Facial emotion recognition with anisotropic inhibited Gabor energy histograms," in *IEEE Intern.Conf. on Image Processing*, Melbourne, 2013.