Computing and Informatics, Vol. 40, 2021, 988–1007, doi: 10.31577/cai_2021_5_988

INTEGRATION OF 2D TEXTURAL AND 3D GEOMETRIC FEATURES FOR ROBUST FACIAL EXPRESSION RECOGNITION

Fouzia Adjailia

Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics, Technical University of Košice
Letná 9, 042 00 Košice, Slovak Republic
e-mail: fouzia.adjailia@tuke.sk

Messaoud Ramdani

Laboratory of Automation and Signals (LASA)
Faculty of Engineering, University Badji, Mokhtar of Annaba
P.O. Box 12, Annaba, 23000, Algeria
e-mail: messaoud.ramdani@univ-annaba.org

Andrinandrasana David Rasamoelina

Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics, Technical University of Košice
Letná 9, 042 00 Košice, Slovak Republic
e-mail: andrijdavid@tuke.sk

Peter Sincak

Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics, Technical University of Košice
Letná 9, 042 00 Košice, Slovak Republic
&
Institute of Informatics, Faculty of Mechanical Engineering and Informatics
University of Miskolc
Hungary
e-mail: peter.sincak@tuke.sk

Abstract. Recognition of facial expressions is critical for successful social interactions and relationships. Facial expressions transmit emotional information, which is critical for human-machine interaction; therefore, significant research in computer vision has been conducted, with promising findings in using facial expression detection in both academia and industry. 3D pictures acquired enormous popularity owing to their ability to overcome some of the constraints inherent in 2D imagery, such as lighting and variation. We present a method for recognizing facial expressions in this article by combining features extracted from 2D textured pictures and 3D geometric data using the Local Binary Pattern (LBP) and the 3D Voxel Histogram of Oriented Gradients (3DVHOG), respectively. We performed various pre-processing operations using the MDPA-FACE3D and Bosphorus datasets, then we carried out classification process to classify images into seven universal emotions, namely anger, disgust, fear, happiness, sadness, neutral, and surprise. Using Support Vector Machine classifier, we achieved the accuracy of 88.5 % and 92.9 % on the MDPA-FACE3D and the Bosphorus datasets, respectively.

Keywords: Facial expression recognition, histogram of oriented gradient, local binary pattern, descriptors, feature extraction, voxels

1 INTRODUCTION

As human beings, we use many channels to communicate our thoughts and emotions; verbal, gestures of gait, language of the body or facial expressions. The research conducted by Mehrabian and Ferris [1] showed that the speaker's facial expression contributes 55 % to the spoken message's effect, while the verbal part and the vocal part contribute just 7 % and 38 %, respectively. Facial expressions have been found to be the most reliable and most efficient transmitter of non-verbal communication channels that shape a universal language and allow people to understand each other's emotional states.

Facial expressions are an efficient channel of contact that allows us to recognize human beings' inner state of mind. The human face plays a vital role in understanding the individual's emotional state, regardless of their ethnicity or cultural background.

An observational analysis made by Ekman [2] to figure out whether there are related aspects of feelings in humans. He performed a study on a tribe in New Guinea, and he found that their facial expressions were the same. Eventually, he revealed that we show universal expressions that help us understand feelings.

There are a number of fields that can benefit from this phenomenon:

- Health care: Facial expression recognition can be used to help identify and be aware of the emotional conditions that patients display during their recovery for a clearer assessment of the treatment process by giving greater attention to patients who require it [3].

- Education: Emotion identification is used to provide a clear understanding of learners' adaptation to the study content, and an alteration to the teaching approach is made depending on the study [4].
- User feedback: Analysing the expressions of users and customers while watching a video, playing sports, or shopping may be crucial for the industry to consider the needs of users and customers profoundly and get feedback on their services or goods for greater benefit and improved marketing.
- Safety and security: Monitoring systems have been developed to identify suspicious people based on the identification of facial emotions.

In computer vision, facial expression recognition is the classification of facial features into one of the six basic universal emotions: happiness, sadness, fear, disgust, surprise, and anger, as introduced by Ekman [2]. Facial expressions are those significant movements of the facial muscles that create a visual expression for other people that conveys emotion. The expression displayed on the person's face is a result of components that control the intensity of the expressions. Recognizing an expression from an image can be done through three major steps, as shown in Figure 1:

1. Pre-processing,
2. Feature extraction,
3. Classification.

Face recognition, orientation, normalization, or augmentation are the first steps in facial pre-processing. After the pre-processing phase, several techniques can be adopted to extract meaningful features from the image. Based on the types of features, it is possible to distinguish between global, local, and temporal approaches. The global feature extraction technique attempts to extract features from the whole image by encoding the entire face (this technique is adopted in our research using 3D Histogram of Oriented Gradients). The local feature extraction technique, though, tests local areas of the human face, such as eyes, hair, nose, lips, etc. (our study utilized this approach by using Local Binary Pattern). The temporal approach is also the third approach that collects and extracts facial expression sequences dynamically. Last, using pre-trained classifiers such as Support Vector Machine, and Random Forest, etc., the classification process consists of generating a classification of the extracted features.

Most research for face analysis utilizes 2D static images or 2D dynamic videos. Recently, 3D static or dynamic data has received increasing attention due to its explicit representation of geometric structures and its inherent capacity to handle facial pose and illumination variations. Similar to 2D images, 3D modality can also be represented in static space and dynamic space. There are difficulties with 2D imaging that cannot be resolved with current state-of-the-art techniques, rendering face and facial expression identification impossible:

- Changes in illumination are triggered mostly by lighting conditions when an image is taken. Images of the face will look differently if there is a difference in

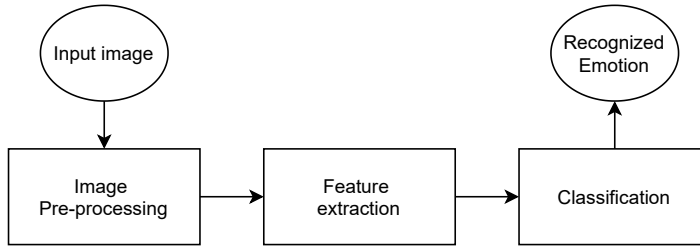

Figure 1. Typical facial expression recognition system

illumination. This influence can cause different forms of image problems, such as creating shadows or rendering half of a face black entirely. On a computer, though, shadows on a face will create higher or lower values of strength relative to the areas unaffected.

- Pose Variations: for any image processing that involves the face, they will trigger major issues. In 2D images, presenting variants such as the rotation of the head will mean that a large amount of the detail of the face will be lost. This would make it impossible for image recognition techniques to benefit from all facial traits, effectively rendering certain algorithms unreliable.
- Occlusions: these are facial image artifacts and differences that allow only half of the mask to be visible. This is a major issue, particularly for face recognition, as it can be unrecognizable to a computer if there is a small difference in facial expression. Glasses, caps, scarves, body hair, and so on may be examples of occlusions.
- Facial aging: The problems it causes are the facial shape, which in the early years varies slightly. In combination with aging, with wrinkles mostly on the forehead, the skin texture grows rougher.

3D images can be computationally intensive to work with in comparison to 2D images. However, they have a more accurate and comprehensive description of the face. They give the parameter of face depth, which is important for the recognition of facial expressions because muscles cannot be easily seen with 2D images such as a concave or convex facial structure. For facial expression analysis, the muscle activity of the face may significantly help to understand what emotion is conveyed, and these muscular movements can be captured in detail by 3D imaging. In order to change all the 2D image processing problems of technology that is shifting facial imaging to other dimensions, 3D mapping will be addressed. For 3D imaging, variations in lighting are not a problem, since the mask is not deformed and the geometric data does not differ based on the lighting. The main contributions to this research paper are as follows:

- Help new researchers understand basic notions of facial emotion recognition and its components, including feature extraction.

- Provide and compile the literature and related work.
- Introduce the standard 3D datasets for facial emotion recognition databases along with their characteristics.
- Establish the boundaries of feature extraction techniques while also identifying core constructs and their relationships and key aspects compared between those approaches.
- Propose a new approach to achieve state of the art results for the classification of facial expression recognition from 2D and 3D data.

This paper is structured as follows: Sections 1 and 2 contain background information about facial expression recognition, research directions are identified and previous work in 3D facial expression recognition is presented (FER). Section 4 discusses the approach proposed. In Sections 5 and 6, the results were presented and discussed, and then we concluded our work in Section 7.

2 RELATED STUDIES

Until around the early 2000s, algorithms integrating effects from 2D and 3D data did not exist. Nowadays, the most direct techniques in this field use the combination of characteristics acquired independently of bi-dimensional or three-dimensional methods, such as texture details, position of landmarks, facial forms, and curvature, to identify expressions and calculate their intensity. Experiments show that integrating features obtained in various modalities helps to capture the general features of facial deformation and increases the precision of identification.

The topic of facial expression recognition was explored by [27]. To this end, an initial method is suggested that measures scale-invariant feature transform (SIFT) descriptors on a range of depth image facial landmarks, and then selects the sub-set of the most important characteristics. An overall identification rate of 77.5 % on the BU-3DFE database was obtained using SVM classification of the chosen features. Comparative assessment of a typical experimental setup illustrates that state-of-the-art outcomes can be obtained from their approach.

In [5], using 3D face details, authors investigated the problem of facial expression recognition. Their methodology was based on a local form analysis of a given face scan of many different regions. These regions were derived from facial surfaces and defined by closed curve sets. In order to derive the form analysis of the derived patches, a Riemannian system is used. The applied structure allowed the similarity (or dissimilarity) of distances between patches to be measured and the optimum deformation between them to be calculated. These measurements were used as inputs for classification techniques such as AdaBoost and Support Vector Machines (SVM).

In [6], using 3D geometry data, we tackled the problem of face expression recognition. To this end, a fully automated approach was suggested that relies on the identification of a collection of facial keypoints, the computation of SIFT

feature descriptors of the face's depth images around sample points identified starting from the facial keypoints, and the selection of the maximum appropriate subset of features. An average identification score of 78.43 % on the BU-3DFE database was obtained by training a Support Vector Machine (SVM) for each facial expression recognition using BU-3DFE database in a common laboratory environment.

In [7], authors used Local Binary Patterns (LBP) for the understanding of facial expression recognition. On several databases, different machine learning approaches were routinely investigated. Extensive studies showed that for facial expression recognition, LBP features were reliable and accurate. They formulated Boosted-LBP using Support Vector Machine classifiers with Boosted-LBP functionality, the most discriminating LBP characteristics were extracted and the best recognition efficiency was achieved. In addition, they investigate LBP features for the identification of low-resolution facial expressions, which were a crucial concern. In their tests, they observed that LBP features behave stably and robustly over a valuable spectrum of low face picture resolutions, and delivered a promising performance in low-resolution compressed video sequences taken in real-world environments.

In order to test discriminative potential for human emotion processing, [8] analyzed a number of different multimodal attributes from video and audio. The authors could extract scale-invariant feature transform (SIFT), Local Binary Patterns from Three Orthogonal Planes (LBP-TOP), Pyramid of Histograms of Orientation Gradients (PHOG), Local Phase Quantization from Three Orthogonal Planes (LPQ-TOP) and audio features for each clip. For any form of feature on the EmotiW 2014 Challenge dataset, they trained various classifiers and suggested a new hierarchical classifier fusion approach for all extracted features. 47.17 % is the accuracy they achieved on the test series.

Authors in [9] suggested a fully automated approach to 3D facial expression recognition. To capture both global and local facial surface deformations that usually occur during facial expressions, a novel facial representation, namely Differential Mean Curvature Maps (DMCMs), was suggested. By measuring the mean curvatures thanks to an integral computation, the DMCMs were directly derived from 3D depth images. They were further normalized into an aspect ratio deformation to allow for facial morphology variations. Finally, a histogram of oriented gradients (HOG) was added to regions of these structured DMCMs to create facial features that could be fed to the Multiclass-SVM classification algorithm. The proposed methodology demonstrated competitive efficiency while being fully automated, using the protocol proposed by [10] on the BU-3DFE dataset.

A new feature descriptor, local normal binary patterns (LNBPs), which was used for detecting facial action units, was proposed by [11]. Once LNBPs have been used to form descriptor vectors that captured the detailed shape of the action, a GentleBoost (GB) algorithm was used to pick the function, and support vector machines (SVMs) were applied to detect each AU. Along with the same test using 3D local binary pattern (3DLBP) descriptors that add the LBP operator to

the depth map of the face, this method was validated on the Bosphorus database. In the identification of several individual Action Units (AUs), LNBP descriptors have been demonstrated to outperform 3DLBPs. Finally, to merge the advantages of the 3DLBPs and both of the LNBP descriptors, feature fusion was used, with the best outcome reaching a mean receiver operating characteristic (ROC AuC) of 96.35.

3 DATASETS DESCRIPTION

The main goal of this section is to provide an overview of the datasets used in our experiment. The primary reason for selecting the datasets are:

- Primarily used in literature
- Free of charge
- Open source
- Provide different formats of data (both 2D and 3D data) that aligns with the goal of this research.

3.1 Bosphorus

The Bosphorus database is a dataset used for scientific purposes, it was presented by Savran et al. [12]. A total of 4 666 scans obtained from 105 subjects are included in the collection, 61 of whom are male and 44 are female. Included are some facial expressions that are expressed in two ways, first being the fundamental expressions of happiness, surprise, fear, sadness, anger, disgust. The other are Action Units-based expressions. For each word, each subject will preferably include a single frontal face picture and a 3D landmark file, see Figure 2.

a) b) c) d) e) f)

Figure 2. 3D samples from happiness expression captured from actors/actresses. Emotional expressions: a) happiness, b) surprise, c) fear, d) sadness, e) anger, f) disgust

3.2 IMPA-FACE3D

To assess the analysis of facial expression recognition, in 2008, the database IMPA-FACE3D was developed, in particular. The neutral face (face with the location of the front camera -0 degrees and without facial expression) and the six universal expressions suggested by Ekman amongst human races are the basis for this purpose: happy, sad, surprise, anger, disgust, and fear. A record of geometric details with color is the key feature of this dataset.

It entails the acquisition of 38 individuals with a neutral face sample, samples relating to six universal facial expressions, and other expressions related to five samples containing open and/or closed mouth and eyes. Two samples matching the lateral profiles of people were also considered. Overall, the data collection consists of 22 men and 16 women, the bulk of which are between 20 and 50 years of age. For all entities, 14 samples were acquired, summarizing 532 samples, see Figure 3.

3D images were captures using 3D sensor called non-contact Konica Minolta Vivid 910 [13].

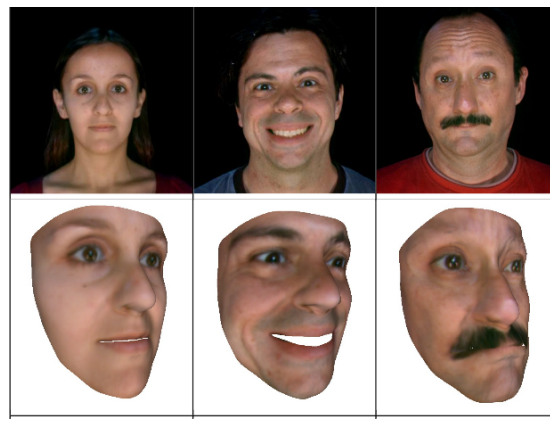

Figure 3. Samples from IMPA-FACE3D dataset

4 PROPOSED APPROACH

In this section, we include the details of our proposed approach. We first illustrate the pre-processing methods we applied on both 2D and 3D images as an input. Then, we offer a detailed description of the method used for feature extraction using the Local Binary Pattern and 3D voxel histogram of oriented gradients for the 2D and 3D images, respectively. This is followed by the process of classification.

Figure 4 represents the overall architecture of the proposed approach, it consists of three parts, image pre-processing of both 2D and 3D input images, as well as the feature extraction phase, where Histogram of oriented gradients (HOG3D) and Local

Binary Patterns (LBP) used on 3D and 2D images respectively to extract features, the two features vectors are concatenated, the integration of the two feature vectors are fed to different classifiers to assess the facial expression classification process.

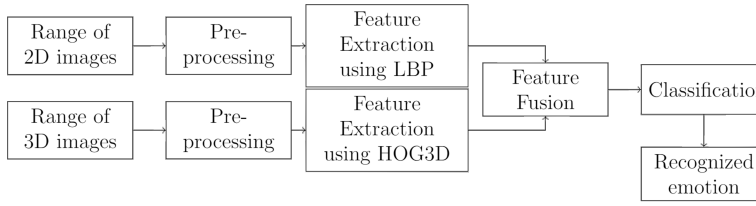

Figure 4. Pipeline of the proposed approach

4.1 Data Pre-Processing

In this section, we present the pre-processing and several transformations methods we used in our experiment as follows:

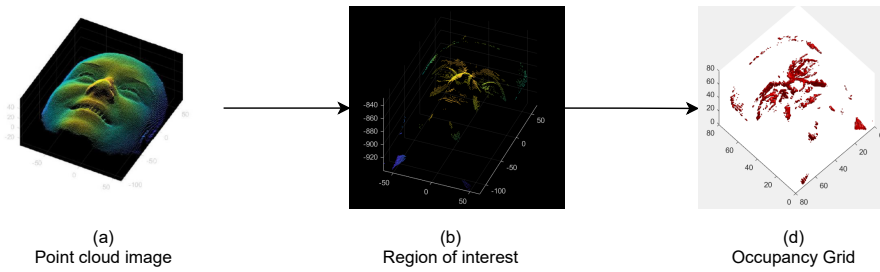

Figure 5. Pre-processing methods applied in our experiment on IMPA-FACE3D dataset

- Cropping: For each 2D image we apply face cropping transformation.
- Write Point cloud files: create an object for storing point cloud, in order to create the point cloud files, we need 3D coordinates. For IMPA-FACE3D dataset, point cloud files are provided, however for Bosphorus dataset, the files provided are in the format of (.bnt) these files contains:
  - zmin: minimum depth value denoting the background,
  - nrows: subsampled number of rows,
  - ncols: subsampled number of columns,
  - imfile: image file name,
  - data: $N \times 5$ matrix where columns are 3D coordinates and 2D.

- Uniform down-sampling: This approach [14] gives the chosen points a homogeneous distribution. For this, from the point cloud, a voxel configuration is derived and the points nearest to each voxel core are chosen. The mean sampling distance corresponds, then, to the edge length of the voxels.
- Region of interest: Region of interests are samples defined for specific use within a data set, it returns the points in the input point cloud within an area of interest. The points inside the Region of interest defined are obtained using a search algorithm based on Kd-tree, the cuboid parameter in our case is shallow point on tip of the nose.
- Occupancy grid: As described in Figure 5, we next apply a simple voxelization transformation to each input point cloud to transform our input point cloud into a pseudo-image. This is done by a) reading point cloud input data, b) spatially bin the point cloud into a 125-by-125-by-125 grid, and c) Build an occupancy grid.

Table 1 presents the transformation approaches applied to the Bosphorus and IMPA-FACE3D datasets, we applied face cropping for the 2D texture images from IMPA-FACE3D, however, images from Bosphorus were cropped.

Figure 5 represents an example of the pre-processing phases applied for the 3D images from IMPA-FACE dataset, input images were of type point cloud, we applied uniform down-sampling for the point clouds, then applied a function of region of interest to eliminate any extra information, convert the point cloud to pseudo-images in the shape of occupancy grids.

Dataset	**Bosphorus**		**IMPA-FACE3D**	
	2D	**3D**	**2D**	**3D**
Write Point cloud files		×		
face cropping			×	
grey scale	×		×	
uniform down-sampling		×		×
region of interest		×		×
occupancy grid		×		×

Table 1. Transformation methods applied to Bosphorus and IMPA-FACE3D

4.2 Feature Extraction

4.2.1 Histogram of Oriented Gradients

3D Voxel HOG (3D VHOG) is based on the original Histogram of Oriented Gradients by [15]. It extends the method by using voxels over pixels and expands the original histograms into 2 dimensions. The multiple steps for 3D gradient orientation descriptor calculation are shown in Figure 6.

The cell grid is split into the service area of the STIP (spatio-temporal interest points). Likewise, each cell is divided into a grid of blocks. An integral video-based rapid measurement then calculates the 3D mean gradient in each row, the direction of which is quantified using a regular polyhedron to form a block histogram. After that a cell histogram is generated by summing up all the block histograms within that cell. Finally, within the support zone of the STIP, the HOG3D descriptor is a concatenation of all cell histograms.

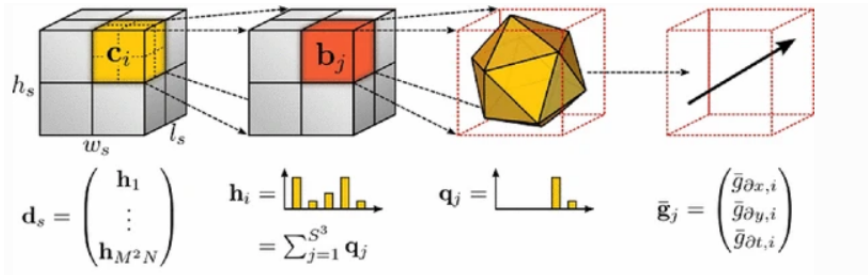

Figure 6. Overview of HOG3D descriptor computation. a) The support region of an STIP is divided into a grid of cells and the final descriptor is a concatenation of all cell histograms. b) A cell is divided into a grid of blocks and a cell histogram is the sum of all block histograms within the cell. c) A block histogram is obtained by regular polyhedron-based orientation quantization of mean block gradient. d) Mean gradient in a block [18].

3D Voxel HOG [16] is based on Dalal and Triggs' original Histogram of Oriented Gradients [17]. By using voxels over pixels, it extends the approach and stretches the original histograms into 2 dimensions. This implementation was developed for the identification of local object structures, for use in a context for risk analysis in which it is used to define the object's risk-related properties (sharp edges, points, etc.).

In order to extract 3D Histogram of Oriented Gradients features, the function has the following parameters:

Voxel volume: a $[n \times n \times n]$ matrix defining voxels with values in the range of 0–1.

Cell size: spacial size of a 3D cell (integer).

Block size: spacial size of a 3D block defined in cells (integer).

Theta histogram bins: the number of bins to break the angles in the xy plane (180 degrees).

Phi histogram bins: the number of bins to break the angles in the xy plane (360 degrees).

Step size: optional integer defining number of cells.

Features: a structure containing the position of a block and a 3D matrix that holds the theta and phi information for each cell in that block. Additionally holders

are created to define if that feature represents a part of an object and defining marker.

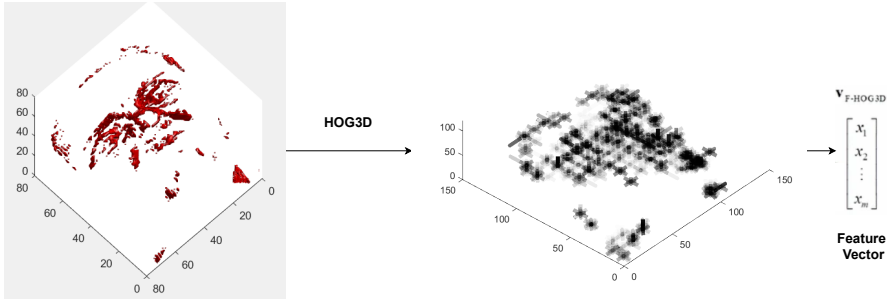

Figure 7. Feature extraction from the 3D images

Table 2 represents the parameters used for the implementation of 3D HOG.

Parameters	**Value**
Voxel volume	$125 \times 125 \times 125$
Cell size	16
Block size	2
Theta histogram bins	9
Phi histogram bins	18
Step size	2
Features	32

Table 2. Parameters used in our experiment

4.2.2 Local Binary Pattern

The Local Binary Pattern (LBP) is a non-parametric operator used to define a pixel's local environment by generating a pattern from the pixel's binary derivatives. When it comes to numerical calculations, the algorithm itself is simple and resilient to monotonic gray change; which is why the operator is typically applied to gray scale images; making it a common and successful tool for analyzing texture.

Figure 8 is an example of the operator of a simple LBP [19]. The initial neighborhood 3×3 on the left is thresholded by the middle pixel value and the center pixel value. There is a binary pattern code generated. The LBP code of the neighborhood's middle pixel is obtained by translating the binary code into a decimal code.

Figure 9 presents the local binary pattern extraction from Bosphorus dataset where the number of neighbors is 8.

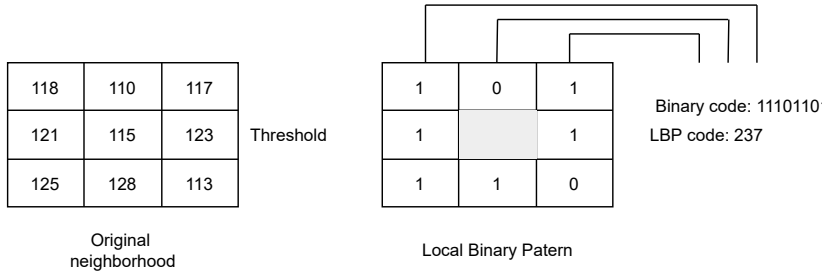

Figure 8. The basic LBP operator

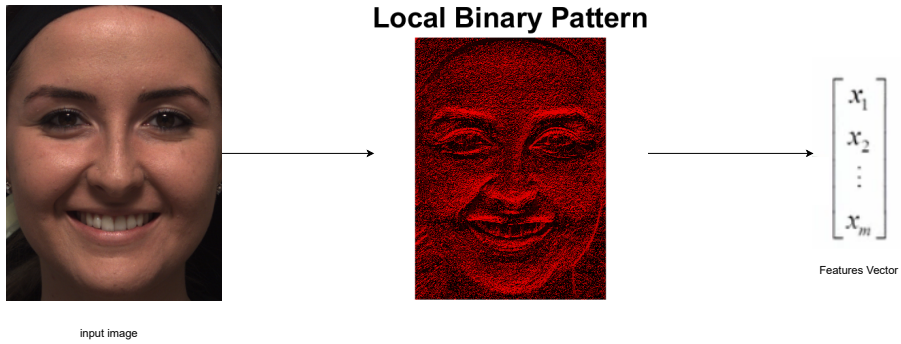

Figure 9. Feature extraction from the 2D images using LBP

4.3 Classification

For the classification purpose we used Support Vector Machine (SVM); Support Vector Machine is a common classification technique [20]. SVM performs implicit mapping of data to a higher dimensional feature space, where linear algebra and geometry can be used to isolate data that is only available separable with non-linear rules in the input space.

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i \gamma_i K(x_i.x) + b\right) \tag{1}$$

where α_i is a dual optimization problem multiplier, and $K(x_i, x)$ is a kernel function. Provided a non-linear mapping Φ that embeds input data into a function space, the kernels have the form $K(x_i, x_j) = (\Phi(x_i) * \Phi(x_j))$. SVM finds a linear hyperplane dividing the maximum margin to distinguish training data in a function space. b is the ideal hyperplane parameter.

5 EXPERIMENTAL RESULTS

For our experiments, 906 scans (instances) were taken from the Bosphorus dataset, while 497 scans were taken from the IMPA-FACE3D dataset. Specifically, Tables 3 and 4 show the number of instances obtained per facial expression in the Bosphorus and the IMPA-FACE3D datasets, respectively. The selected scans correspond to faces showing the universal facial expressions. All these images are from frontal faces, with no pose variation, and without occlusions.

Facial Expressions	**2D Instances**	**3D Instances**	**Total**
Surprise	71	71	142
Sadness	66	66	132
happy	106	106	212
Anger	71	71	142
Disgust	69	69	138
Fear	70	70	140
Total	**453**	**453**	**906**

Table 3. Number of instances per facial expression in the Bosphorus dataset

Facial Expressions	**2D Instances**	**3D Instances**	**Total**
Neutral	36	36	72
Surprise	35	71	71
Sadness	36	35	71
Joy	36	34	70
Anger	36	35	71
Disgust	36	35	71
Fear	36	35	71
Total	**252**	**245**	**497**

Table 4. Number of instances per facial expression in the IMPA-FACE3D dataset

5.1 Quantitative Evaluations of the Proposed Approach

The experiment were carried out with 8.00 GB RAM using an Intel core i7 7th Gen CPU 2.70 GHZ system. The analyses were carried out using Matlab. Using the MDPA-FACE3D and the Bophorus datasets, the methods were applied and evaluated.

Table 5 reports the confusion matrix and shows the result obtained on the IMPA-FACE 3D dataset. We achieved a model accuracy of 88.5 %.

- Kernel scale: Automatic.
- Box constraint level: 1.
- Mutliclass method : one-vs-one.

- Standardize data: true.
- Total misclassification cost: 29.
- Prediction speed: ~ 960 obs/sec.

PC/AC	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
anger	29	7	0	0	0	0	0
disgust	0	28	8	0	0	0	0
fear	0	0	35	1	0	0	0
joy	0	0	1	34	1	0	0
neutral	0	0	0	1	35	0	0
sadness	0	0	0	0	0	36	0
surprise	0	0	0	0	0	10	26

Table 5. Confusion Matrix; PC: predicted classes; AC: actual classes

Table 6 presents the confusion matrix and shows the result obtained on the Bosphorus dataset. We achieved a model accuracy of 92.9 %:

- Kernel scale: Automatic
- Box constraint level: 1.
- Mutliclass method : one-vs-one
- Standardize data: true
- Total misclassification cost: 32
- Prediction speed: ~ 690 obs/sec
- Training time: 9.274 sec

PC/AC	Anger	Disgust	Fear	Joy	Sadness	Surprise
anger	103	3	0	0	0	0
disgust	0	62	4	0	0	0
fear	0	1	68	2	0	0
joy	0	0	3	68	1	0
sadness	0	0	0	7	62	0
surprise	0	0	0	0	12	58

Table 6. Confusion Matrix; PC: predicted classes; AC: actual classes

6 EVALUATION AND DISCUSSION

In comparison with state-of-the-art models, the findings demonstrate that the proposed approach outperforms the results obtained in recent research, see Table 7.

Reference	Database	Methodology	Expressions	Accuracy
[21]	Bosphorus	Zernike moments + SVM	6	60.00
[22]	Bosphorus	LBP + SVM	6	76.98 %
[23]	Bosphorus	LBP + SVM	6	76.56 %
[24]	Bosphorus	SURF+SVM with PE	**7**	79.00 %
[24]	Bosphorus	SURF + SVM with EPE	7	84.00 %
[25]	BU–3DFE/ Bosphorus	Multi-modal 2D and 3D descriptors	6/7	79.72 %
[26]	BU-3DFE	Geometric scattering representations	6	82.73 %
[27]	BU-3DFE	scale-invariant feature transform (SIFT)	6	77.5 %
Our Approach	**Bosphorus**	**HOG3D + LBP + SVM**	**6**	**92.9 %**
Our Approach	**MPA-FACE 3D**	**HOG3D + LBP + SVM**	**7**	**88.5 %**

Table 7. Comparison in terms of classification rate

While the emotion class is the ultimate result of our suggested method, we performed comprehensive tests to assess the intermediate stages of our strategy. To begin, we quantified the accuracy of the 3D expression parameters estimation. We evaluate its performance using a split of two datasets, namely the Bosphorus and the IMPA-FACE. Tables 5 and 6 compare the proposed technique to the state of the art. We obtained accuracy of 92.9 % and 88.5 %, respectively, which surpassed the state of the art for 3D facial expression recognition, and can be considered as a benchmark for the IMPA-FACE dataset.

7 CONCLUSION AND FUTURE WORK

In this paper, we proposed a new approach to effectively recognize human facial expressions. We applied cutting edge techniques for 3D image pre-processing. We used Local Binary Patterns and 3D voxel Histogram of Oriented Gradients to extract features from 2D and 3D images, respectively. Future work should focus on the the computation of regional statistics of vHOG histograms (eyes, mouth, etc.), the use of automatic landmark point detection to compute useful expression-dependent features (better discrimination) and the combination of classifiers.

REFERENCES

[1] Mehrabian, A.—Ferris, S. R.: Inference of Attitudes from Nonverbal Communication in Two Channels. Journal of Consulting Psychology, Vol. 31, 1967, No. 3, pp. 248–252. doi: 10.1037/h0024648.

[2] Ekman, P.: Universal Facial Expressions of Emotions. California Mental Health Research Digest, Vol. 8, 1970, No. 4, pp. 151–158.

[3] Leo, M.—Carcagnì, P.—Mazzeo, P. L.—Spagnolo, P.—Cazzato, D.—Distante, C.: Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches. Information, Vol. 11, 2020, No. 3, Art. No. 128, doi: 10.3390/info11030128.

[4] Linnenbrink-Garcia, L.—Patall, E. A.—Pekrun, R.: Adaptive Motivation and Emotion in Education: Research and Principles for Instructional Design. Policy Insights from the Behavioral and Brain Sciences, Vol. 3, 2016, No. 2, pp. 228–236, doi: 10.1177/2372732216644450.

[5] Maalej, A.—Ben Amor, B.—Daoudi, M.—Srivastava, A.—Berretti, S.: Local 3D Shape Analysis for Facial Expression Recognition. 2010 20th International Conference on Pattern Recognition, IEEE, 2010, pp. 4129–4132, doi: 10.1109/icpr.2010.1003.

[6] Berretti, S.—Ben Amor, B.—Daoudi, M.—del Bimbo, A.: 3D Facial Expression Recognition Using SIFT Descriptors of Automatically Detected Keypoints. The Visual Computer, Vol. 27, 2011, No. 11, Art. No. 1021, doi: 10.1007/s00371-011-0611-x.

[7] Shan, C.—Gong, S.—McOwan, P. W.: Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study. Image and Vision Computing, Vol. 27, 2009, No. 6, pp. 803–816, doi: 10.1016/j.imavis.2008.08.005.

[8] Sun, B.—Li, L.—Zuo, T.—Chen, Y.—Zhou, G.—Wu, X.: Combining Multimodal Features with Hierarchical Classifier Fusion for Emotion Recognition in the Wild. Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14), 2014, pp. 481–486, doi: 10.1145/2663204.2666272.

[9] Lemaire, P.—Ardabilian, M.—Chen, L.—Daoudi, M.: Fully Automatic 3D Facial Expression Recognition Using Differential Mean Curvature Maps and Histograms of Oriented Gradients. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1–7, doi: 10.1109/FG.2013.6553821.

[10] Gong, B.—Wang, Y.—Liu, J.—Tang, X.: Automatic Facial Expression Recognition on a Single 3D Face by Exploring Shape Deformation. Proceedings of the 17th ACM International Conference on Multimedia (MM '09), 2009, pp. 569–572, doi: 10.1145/1631272.1631358.

[11] Sandbach, G.—Zafeiriou, S.—Pantic, M.: Local Normal Binary Patterns for 3D Facial Action Unit Detection. 2012 19th IEEE International Conference on Image Processing, 2012, pp. 1813–1816, doi: 10.1109/ICIP.2012.6467234.

[12] Savran, A.—Alyüz, N.—Dibeklioğlu, H.—Çeliktutan, O.—Gökberk, B.—Sankur, B.—Akarun, L.: Bosphorus Database for 3D Face Analysis.

In: Schouten, B., Juul, N. C., Drygajlo, A., Tistarelli, M. (Eds.): Biometrics and Identity Management (BioID 2008). Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 5372, 2008, pp. 47–56, doi: 10.1007/978-3-540-89991-4_6.

[13] Mena-Chalco, J.: Reconstruçao de Faces 3D Através de Espaços de Componentes Principais (Reconstruction of 3D Faces Through Principal Component Spaces). Ph.D. Thesis, IME-USP, 2010, doi: 10.13140/RG.2.2.24366.15685 (in Portugal).

[14] Glira, P.—Pfeifer, N.—Briese, C.—Ressl, C.: A Correspondence Framework for ALS Strip Adjustments Based on Variants of the ICP Algorithm. Photogrammetrie – Fernerkundung – Geoinformation, Vol. 2015, 2015, No. 4, pp. 275–289, doi: 10.1127/pfg/2015/0270.

[15] Dalal, N.—Triggs, B.: Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), Vol. 1, 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[16] Dupre, R.—Argyriou, V.—Greenhill, D.—Tzimiropoulos, G.: A 3D Scene Analysis Framework and Descriptors for Risk Evaluation. 2015 International Conference on 3D Vision, 2015, pp. 100–108, doi: 10.1109/3DV.2015.19.

[17] Nguyen, T. Q.—Kim, S. H.—Na, I. S.: Fast Pedestrian Detection Using Histogram of Oriented Gradients and Principal Components Analysis. International Journal of Contents, Vol. 9, 2013, No. 3, pp. 1–9, doi: 10.5392/IJoC.2013.9.3.001.

[18] Klaeser, A.—Marszałek, M.—Schmid, C.: A Spatio-Temporal Descriptor Based on 3D-Gradients. In: Everingham, M., Needham, C. (Eds.): Proceedings of the British Machine Vision Conference (BMVC 2008), 2008, pp. 99.1–99.10, doi: 10.5244/c.22.99.

[19] Ojala, T.—Pietikainen, M.—Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002, No. 7, pp. 971–987, doi: 10.1109/TPAMI.2002.1017623.

[20] Vapnik, V.: Statistical Learning Theory. First Edition. Wiley, New York, 1998, pp. 62.

[21] Vretos, N.—Nikolaidis, N.—Pitas, I.: 3D Facial Expression Recognition Using Zernike Moments on Depth Images. 2011 18th IEEE International Conference on Image Processing, 2011, pp. 773–776, doi: 10.1109/ICIP.2011.6116669.

[22] Chun, S. Y.—Lee, C. S.—Lee, S. H.: Facial Expression Recognition Using Extended Local Binary Patterns of 3D Curvature. In: Park, J., Ng, J. Y., Jeong, H. Y., Waluyo, B. (Eds.): Multimedia and Ubiquitous Engineering. Springer, Dordrecht, Lecture Notes in Electrical Engineering, Vol. 240, 2013, pp. 1005–1012, doi: 10.1007/978-94-007-6738-6_124.

[23] Wang, Y.—Meng, M.—Zhen, Q.: Learning Encoded Facial Curvature Information for 3D Facial Emotion Recognition. 2013 Seventh International Conference on Image and Graphics, 2013, pp. 529–532, doi: 10.1109/ICIG.2013.112.

[24] Azazi, A.—Lutfi, Syaheerah L.—Venkat, I.—Fernández-Martínez, F.: Towards a Robust Affect Recognition. Expert Systems with Applications: An International Journal, Vol. 42, 2015, No. 6, pp. 3056–3066, doi: 10.1016/j.eswa.2014.10.042.

[25] Li, H.—Ding, H.—Huang, D.—Wang, Y.—Zhao, X.—Morvan, J.—Chen, L.: An Efficient Multimodal 2D + 3D Feature-Based Approach to Automatic Facial Expression Recognition. Computer Vision and Image Understanding, Vol. 140, 2015, pp. 83–92, doi: 10.1016/j.cviu.2015.07.005.

[26] Yang, X.—Huang, D.—Wang, Y.—Chen, L.: Automatic 3D Facial Expression Recognition Using Geometric Scattering Representation. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2010, pp. 1–6, doi: 10.1109/FG.2015.7163090.

[27] Berretti, S.—Del Bimbo, A.—Pala, P.—Ben Amor, B.—Daoudi, M.: A Set of Selected SIFT Features for 3D Facial Expression Recognition. 2010 20th International Conference on Pattern Recognition, 2010, pp. 4125–4128, doi: 10.1109/ICPR.2010.1002.

Fouzia Adjailia obtained her Ph.D. at the Department of Cybernetics and Artificial Intelligence at the Technical University of Košice, Slovakia in 2021. Her primary research interest is computer vision, the intersection between facial expression recognition from 2D and 3D images, and human-robot interaction. She has productively collaborated with training related to artificial intelligence. She had a visit to Japan to attend JST-CREST/IEEE-RAS Spring School. In Germany, she conferred research related to artificial intelligence and urban living in BMW Summer School. In addition to her academic work, she provides online courses related to artificial intelligence.

Messaoud Ramdani received his state doctorate (doctorat d'Etat, Ph.D. degree) in automatic control from the University of Annaba, Algeria, in 2006. From 2004 to 2005, he has held student research visiting position at the Research Centre for Automatic Control of Nancy (CRAN), University of Lorraine, Nancy, France. He is currently Professor in the Department of Electronics, Faculty of Engineering, University Badji-Mokhtar of Annaba, Algeria. His current research interests include fuzzy systems, model predictive control, variable structure control, environmental and renewable energy engineering, statistical process monitoring, machine learning, intelligent data analysis and the application of computational intelligence.

Andrinandrasana David Rasamoelina received his M.Sc. degree in software engineering and database management in 2019 at Ecole Nationale d'Informatique, University of Fianarantsoa, Madagascar. Currently, he is pursuing his Ph.D. degree in artificial intelligence at the Department of Cybernetics and Artificial Intelligence at the Technical University of Košice, Slovakia. His research focuses on few-shot learning, the use of artificial intelligence with scarce data.

Peter Sincak is Professor of artificial intelligence and he is responsible for artificial intelligence domain at the Technical University of Košice. Currently he is serving as Head of Department of Cybernetics and Artificial Intelligence and Head of Scientific Board of Artificial Intelligence Slovakia National Slovak Platform. His main interests are deep learning, AI computing, social robotics and explainable AI. He is coauthor of 100+ publications and also editor of number of books published in Springer-Verlag, IOS Press and World Scientific. He gave number of invited lectures in Japan, China, Italy and many other countries. He is also Visiting Professor in the University of Miskolc, Hungary.