Chapter

Facial Emotion Recognition Feature Extraction: A Survey

Michele Mukeshimana, Abraham Niyongere and Jérémie Ndikumagenge

Abstract

Facial emotion recognition is a process based on facial expression to automatically recognize individual emotion expression. Automatic recognition refers to creating computer systems that are able to simulate human natural ability of detection, analysis, and determination of emotion by facial expression. Human natural recognition uses various points of observation to make decision or conclusion on emotion expressed by the present person in front. Facial features efficiently extracted aid in improving the classifier performance and application efficiency. Many feature extraction methods based on shape, texture, and other local features are proposed in the literature, and this chapter will review them. This chapter will survey some recent and formal feature expression methods from video and image products and classify them according to their efficiency and application.

Keywords: facial emotion recognition (FER), feature extraction, human computer interaction, automatic emotion recognition, machine learning

1. Introduction

Recent research in Computer Science is more driven by constructing a solution smarter product (hardware and software). Computing is becoming ubiquitous and pervasive, with human at the center. Devices are attaining more ability to average human intelligent actions and interactions. Human beings are basically emotional and affective. They express their emotion in many ways and they require emotion expression in their natural interaction no matter what they interact with (human, machine, or nature) [1]. Together with the objective of having human-centered digital solutions, affective computing aims to endow computers with the ability of sensing, recognition, and expressing emotion [2, 3].

Automatic emotion recognition is one of the recent research trends in Artificial Intelligence, especially in the field of Machine Learning. Based on scientific ground, emotion recognition is about a mapping from feature space to emotion descriptors or label space. This feature space is built from different identified cues extracted from an original element, which is the subject of study [4]. These cues seem to help distinguish two different situations or cases during a classification task and minimize differences within elements of the same class.

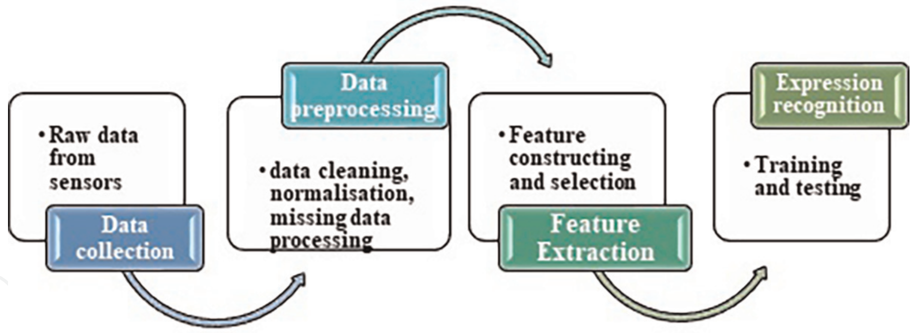


Figure 1.
Emotion recognition processes.

In order to recognize human affect state automatically, some of the steps studied and worked on consist of data collection, data preprocessing, feature extraction, and emotion recognition, as represented in **Figure 1**.

Data collection as the first step in automatic recognition consists of reassembling raw data from different sensors according to the work at hand, that is, the modality to study or its application [5]. This chapter is about acquiring a video with a recognizable human face and expressing emotion. Collected data are tarnished with many noises and unwanted details that need to be removed [6, 7]. Data preprocessing generally involves data cleaning, normalization (or standardization), and missing data processing. Cleaned data serve as basic space for extraction of main features, which convey more information for an expected pattern. The feature extraction step consists of representing data in a digital form to present to a filter. It draws out the values, which are more informative and nonredundant for a future easy learning process and quick generalization.

It is very important to extract an effective facial representation from all considered facial images for any effective facial expression recognition system. The resulting representation should preserve indispensable information possessing distinguished contrast power and stability, which lessens within-class variations of expressions whereas expands between-class variations [8]. Extracted features aid in emotion classification [9]. At this level, two procedures are done: training a classifier and testing it. Emotion classification is the last step, resulting in the process of classification of a new case into its category using the trained classifier. Classification performance is greatly subjective to the quality of information contained in the expression representations [10]. Thus, the step of feature extraction has a great influence on the classification outcome.

This chapter contains a global point of view on feature extraction, and different types of facial expression recognition feature extraction methods are detailed in the following chapters.

2. Feature extraction

Features are also called attributes or input variables. Feature extraction consists in draw out the feature relating to the modality. The precision of the most relevant feature for extraction in emotion recognition research is still an open topic [11–13]. However, the often-studied modalities are face expression, speech, body motion, hand gestures, and physiological signals. They are the representation of the data and

DOI: http://dx.doi.org/10.5772/intechopen.110597

can be in binary form, categorical, discrete, or continuous. Feature extraction is subdivided into two processes, that is, features construction and feature selection.

2.1 Feature construction

The feature construction consists of determining the good data representation, according to the domain specifications and measurements availability [13]. The extracted features are proper to modalities and an interesting task. In emotion recognition, feature extraction focuses on cues that convey better the affect expression. Actually, referring to human natural emotion or intention expression and perception, there are many studies that have proved some frequently observed units to convey useful information for emotion categorization.

Table 1 represents a summary of the frequently observed and studied units for feature extraction, according to the recording methods or the study of interest.

Table 1 presents a summary of the list of the combinations and cues considered according to modality in the study. Modality means any human body parts that can be used to express emotion. In affect detection, some basic units encompass other intermediate units. This list relates to the most cited elements in the literature. The modalities are defined as the main objectively observed entities, which convey most information about emotion expression. Basic units are the small elements of the whole modality and can stand for an independent study [14–18]. Intermediate units are more detailed than the basic units. These unity measurements produce multiple feature values, which constitute the vector feature of the modality [19]. The features

Modality		**Units**	
		Basic	**Intermediates**
Face expression		Eyes, eyebrows, nose, mouth	Action Units, pupil
Speech		Linguistic	Word, multi-word, phrases, sentences, documents
		Paralinguistic	Pitch intensity of utterances, bandwidth, duration, voice quality, Mel frequency Cepstral coefficients (MFCC)
Body	Head gestures	Head position	
		Head movement	
	Hand Gestures	Shape	
		Motion; keystrokes	
	Body motion	Spinal column	Neck, chest and abdomen
		DOF body	Symmetrical arms
		Body center mass	Movement of body center of mass
		Joints	Degree of joint rotation
Physiologic		Hearth Brain Limbs Blood	Electrocardiogram (ECG); breath rate; electro-dermal activity (EDA); electro-myogram (EMG)

Table 1.
Modality and extracted features.

Toolkit	Modality	Feature extracted/functionality	Brief description
PRAAT [22]	Audio	Duration, F0, Range, Movement, Slope, Energy features	PRAAT (a system for doing phonetics)
FEELTRACE [23, 24]	Audio	Labeling	Allowing the emotional dynamics of speech episodes to be examined.
OpenEAR [25]	Audio	Signal Energy, Loudness, Mel-/Bark-/Octave-Spectra, MFCC, PLP-CC	openEAR provides efficient (audio) feature extraction.
OpenSMILE [26]	Audio	Signal Energy, Loudness, Formants, Mel-/Bark-/Octave-Spectra, MFCC, PLP-CC, Pitch, Voice quality (Jitter, Shimmer), LPC, Line Spectral, Pairs(LSP), Spectral, Shape description	It is an open source toolkit, for feature extraction in machine learning and data mining [27]
EyesWeb [28]	Body	Quantity of motion, cue, Contraction index of the body, velocity, Acceleration, fluidity of the hand's, barycenter	Open software for extended Multimodal Interaction.
Luxand FSDK 1.7	Face	Action units	Facial recognition software [29]
ANVIL [30]	Audio	Annotation tool in a multimodal dialog	Free for research purposes [31]

Table 2.
Some automatic feature extraction tools.

construction can be manually processed and/or complemented by automatic feature construction methods [20, 21].

Recently, the research in feature extraction techniques has ended up by proposing some automatic feature extraction tools and algorithms. Some examples are given in **Table 2**.

In **Table 2**, the toolkit column corresponds to the name given to the tool or algorithm in the literature. The modality column means the channel conveying needed information. The listed tools are mostly available online and free of charge, and are compatible with the most popular platforms, such as Windows, Linux, and Macintosh. The references within the table are the work that has utilized the tool or the reports of the authors.

The step of feature construction builds a feature set which is full of some unnecessary or superfluous data. In order to clean that feature set, a feature selection is necessary to prepare a proper dataset useful in the learning process.

2.2 Feature selection

The step of features selection mainly aims to select some features, which are more relevant and explanatory to the study in view. The feature construction creates thousands of features that require an important amount of storage and slows down the training process, the curse of dimensionality. The feature selection uses a data reduction method to eliminate irrelevant and redundant information to a sufficient minimum dimension. The main objective is to get attributes with a large distance between classes and small variance in the same class [7].

The step of feature extraction success affects the training process, recognition accuracy, and application efficiency. It constitutes a subject of study on its own, and it is the subject of the present work, extensions are limited to facial feature extraction.

3. Facial expression feature extraction

Facial feature extraction is all about exactly localizing different features on the face, which include the detection of eyes, brows, mouth, nose, chin, etc. [32]. Facial features are often subdivided into appearance or transient features and geometric or intransient features [10, 33, 34]. Local appearance-based methods extract appearance changes of the face or a region of the face, while geometric features express the shape of the facial components (eyebrows, eyes, mouth, etc.) and the location of prominent points of the face (comers of the eyes, mouth, etc.).

3.1 Geometric feature extraction

3.1.1 Facial feature points (FFP)

The shape and location-related features could be achieved using Active Appearance Methods (AAM) [35]. It has been used to label 68 facial feature points (FFPs) as related in the work of Wu et al. [36]. Facial feature points are visible marks in facial images or points that constitute interesting part of images, such as eye centers, nose tip, mouth corners, and other salient facial points. They are often used as a reference or for measurement. **Figure 2** represents an example of the FFPs extracted based on the AAM alignment and the corresponding animation parameters, and this figure is extracted from the work of Wu et al. [36].

Facial feature points are also referred to as facial points, fiducial facial points, or facial landmarks [37]. The points shown in **Figure 2** can be concatenated to represent a shape $x = (x_1, \cdots, x_N, y_1, \cdots, y_N)^T$, where (x_i, y_i) denotes the location of the i-th point and N is the number of points (here **Figure 2**, N equals 68). The FFPs are grouped into Facial Animation Parameters (FAPs), to facilitate the normalization among people. Every FAP limits a segment of a key distance on the face. The AAM was initially developed in the work of Cootes and Taylor [35], and has presented strong promise in multiple technologies of facial recognition technologies, including in recognizing emotions by its ability to both aid in beginning face-search algorithms and feature extraction based on texture and shape [38].

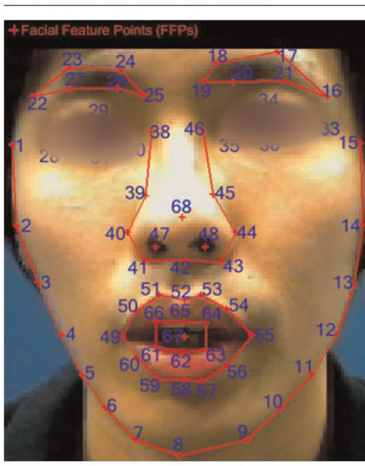
Extracted facial feature points (FFPs)	Facial regions	FAPs Num.	Euclidean distance between FFPs	Comparing FFPs displacement with neutral frame
	Eyebrows	1, 2	$D_{vertical,1}(22, 30)$, $D_{vertical,2}(16, 35)$	$D_{v,1\_Neutral}-D_{v,1}$, $D_{v,2\_Neutral}-D_{v,2}$
		3, 4	$D_{vertical,3}(25, 30)$, $D_{vertical,4}(19, 35)$	$D_{v,3\_Neutral}-D_{v,3}$, $D_{v,4\_Neutral}-D_{v,4}$
		5, 6	$D_{vertical,5}(22, 28)$, $D_{vertical,6}(16, 33)$	$D_{v,5\_Neutral}-D_{v,5}$, $D_{v,6\_Neutral}-D_{v,6}$
		7, 8	$D_{vertical,7}(23, 28)$, $D_{vertical,8}(17, 33)$	$D_{v,7\_Neutral}-D_{v,7}$, $D_{v,8\_Neutral}-D_{v,8}$
		9, 10	$D_{vertical,9}(25, 28)$, $D_{vertical,10}(19, 33)$	$D_{v,9\_Neutral}-D_{v,9}$, $D_{v,10\_Neutral}-D_{v,10}$
		11, 12	$D_{vertical,11}(23, 30)$, $D_{vertical,12}(17, 35)$	$D_{v,11\_Neutral}-D_{v,11}$, $D_{v,12\_Neutral}-D_{v,12}$
		13	$D_{m,13}(19, 25)$	$D_{h,13\_Neutral}-D_{h,13}$
	Eyes	14, 15	$D_{vertical,14}(29, 31)$, $D_{vertical,15}(34, 36)$	$D_{v,14\_Neutral}-D_{v,14}$, $D_{v,15\_Neutral}-D_{v,15}$
		16, 17	$D_{vertical,16}(28, 49)$,$D_{vertical,17}(33, 55)$	$D_{v,16\_Neutral}-D_{v,16}$, $D_{v,17\_Neutral}-D_{v,17}$
		18, 19	$D_{horizontal,18}(28, 30)$,$D_{horizontal,19}(33, 35)$	$D_{h,1\_18Neutral}-D_{h,18}$, $D_{h,19\_Neutral}-D_{h,19}$
	Nose	20, 21	$D_{vertical,20}(52, 68)$, $D_{vertical,21}(58, 68)$	$D_{v,20\_Neutral}-D_{v,20}$, $D_{v,21\_Neutral}-D_{v,21}$
		22, 23	$D_{vertical,22}(49, 68)$, $D_{vertical,23}(55, 68)$	$D_{v,22\_Neutral}-D_{v,22}$, $D_{v,23\_Neutral}-D_{v,23}$
	Mouth	24, 25	$D_{vertical,24}(52, 58)$, $D_{horizontal,25}(49, 55)$	$D_{v,24\_Neutral}-D_{v,24}$, $D_{h,25\_Neutral}-D_{h,25}$
	Facial Contours	26, 27	$D_{horizontal,26}(5, 58)$, $D_{horizontal,27}(11, 58)$	$D_{h,26\_Neutral}-D_{h,26}$, $D_{h,27\_Neutral}-D_{h,27}$
		28, 29	$D_{horizontal,28}(2, 68)$, $D_{horizontal,29}(14, 68)$	$D_{h,28\_Neutral}-D_{h,28}$, $D_{h,29\_Neutral}-D_{h,29}$
		30	$D_{vertical,30}(8, 68)$	$D_{v,30\_Neutral}-D_{v,30}$

Figure 2.
Example of facial feature points labeled using AAM alignment [36].

3.1.2 Facial affective coding systems (FACS)

Other works consider the Facial Affective Coding System (FACS) and define the Active Unities (AUs) as the facial muscle action [39, 40]. Facial action unit research studies the movement of facial muscles [41] and describes facial movement changes. Based on the work of Ekman Paul and Friesen [42], Facial Action Coding System (FACS) contributes as one of the most representative methods for facial expression application in measurement technology. Action units can precisely extract facial expressions, but they are less applied in facial expression recognition because of their exact positioning. **Figure 3** represents some examples of Action Unities.

In **Figure 3**, the examples display the considered action unities detected on facial images. Those action units are randomly chosen for illustration. The description is about facial muscle movement or portrayal. Muscles indicate the action done on the facial muscles or the whole head. The emotion expression corresponds to an ascertained combination of some specific action unities, and **Table 3** represents some examples of possible combinations, their description in facial muscles, and the corresponding emotion expression.

In **Table 3**, the combinations of Action Unities are referred to the work of the visual book of group iMOTIONS. For more details, we refer to the above-mentioned review [36] and the work in Refs. [39–42] and references therein.

3.2 Appearance-based features

Local appearance descriptors in the literature are mostly the LBPs (Local Binary Pattern) and its derived, the Local Direction Number pattern (LDN) and the Edge-Oriented Histogram. Local appearance-feature-based methods are used because of their close descriptor of the appearance.

3.2.1 Local binary pattern (LBP)

Local Binary Pattern (LBP) [43] method is a texture operator mostly used in computer vision and image processing applications, such as in object detection, object

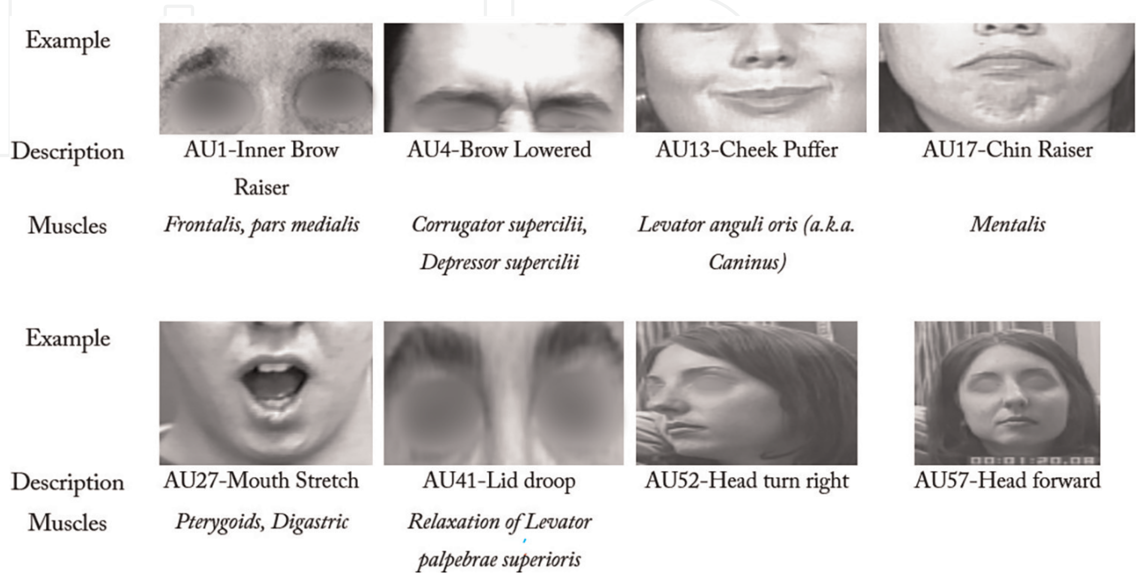


Figure 3.
Examples of action Unity description and muscles involved.

Action Unities combination	Description	Emotion
4 + 5 + 7 + 23	Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener	Anger
9 + 15 + 16	Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor	Disgust
1 + 2 + 4 + 5 + 7 + 20 + 26	Inner Brow Raiser, Outer Brow Raiser, Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop	Fear
6 + 12	Cheek Raiser, Lip Corner Puller	Happiness/Joy
1 + 4 + 15	Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor	Sadness
1 + 2 + 5 + 26	Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop	Surprise

Table 3.
Example of action unities combination for emotion analysis.

tracking, face recognition, and fingerprint matching [44–46]. It is a good operator for real time and very high frame rate applications. The LBP computes features for each image pixel; therefore, real-time extraction of LBP features requires considerable computational performance. It was proposed for a texture analysis [29], and it is insensitive to illumination changes and has an extension to rotation invariant [31].

An LBP feature is a binary vector obtained from a neighborhood around the current image pixel. The basic LBP operator is the 3*3 neighborhood pixels, which is called LBP 8, 1, that is, there are nine pixels with one center and eight neighborhood pixels. The value of the LBP feature is the result of the thresholding of every pixel's luminance against the center pixel's luminance. It is equal to 1 if the difference is positive and to 0 otherwise. The resultant binary number is computed by concatenating all the above binary codes in a clockwise direction, beginning from the top-left one, as shown in **Figure 4**, and the corresponding decimal value is used for labeling [45]. The obtained numbers are known as Local Binary Patterns or LBP codes.

The basic operator of 3*3 neighborhoods is small to capture dominant features with large-scale structures. Later on, Ojala et al. [47] proposed an advanced operator, which is proficient to deal with texture at different scales by using neighborhoods of different sizes. A set of sampling points is evenly spaced on a circle centered at the current pixel to label and define a local neighborhood. A bilinear interpolation permits interpolation of the points that do not fall within the pixels, thus allowing to use a radius of any size and to have any number of sampling points in the neighborhood. Some examples are illustrated in **Figure 5**.

Figure 5 represents an example of LBP extended operator with the circular (8, 1), (16, 2), and (24, 3) neighborhoods.

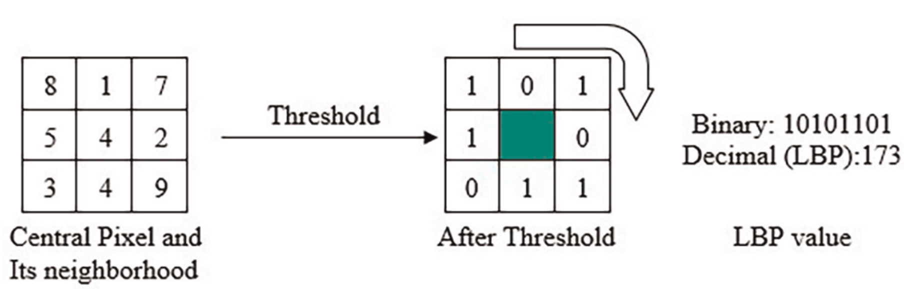


Figure 4.
LBP operator.

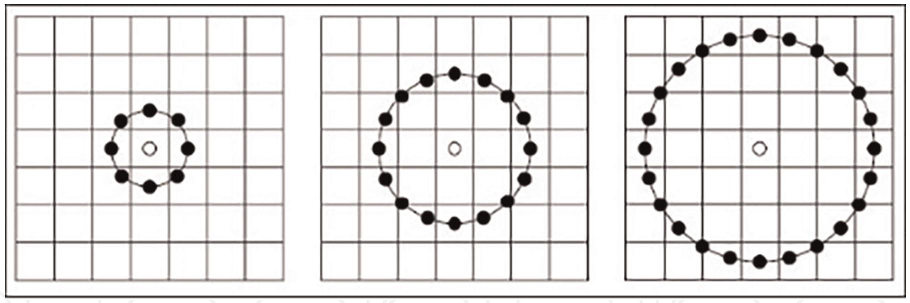


Figure 5.
Examples of the extended LBP operator.

Given a pixel at (x_c, y_c), for an extended LBP $_{(P,\ R)}$ operator with P sampling points neighborhood on a circle of radius R, the LBP can be computed as follows in decimal form:

$$LBP_{(P,R)}\left(x_c, y_c\right) = \sum_{P=0}^{P-1} s(i_P - i_c)2^P \tag{1}$$

where i_c and i_p are gray-level values of the central pixel and its neighborhood, the P is the number of surrounding pixels in the circle neighborhood with a radius R, and the function is defined as follows:

$$s(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases} \tag{2}$$

The LBP $_{(P,\ R)}$ operator produces 2^P different output values, corresponding to 2^P different binary patterns formed by P pixels in the neighborhood. That makes the extended LBP sensitive to image rotation, and in order to deal with it, a rotation invariant LBP was proposed and is computed as follows:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, i) | i = 0, 1, \dots, P - 1\} \tag{3}$$

where $ROR(u, i)$ executes a circular by bit right shift on the P-bit number u*i times. This operator computes occurrence statistics of individual rotation invariant patterns corresponding to certain micro-features in the image. It is a good operator for real time and very high frame rate applications.

LBP is invariant against monotonic gray-scale variations and has extensions to rotation invariant texture analysis. In the work of Ojala et al. [47], it was shown that there are patterns containing more information than others do and they were called "uniform patterns" denoted $LBP_{(P,R)}^{U2}$. In fact, it is possible to use a subset of 2^P binary pattern to represent the image's texture. Uniform local binary patterns are the patterns containing at most two bitwise transitions from 0 to 1 or *vice versa* when the corresponding bit string is considered circular. For example,

•00000000 (0 transitions).
•01110000 (2 transitions).
•11,001,111 (2 transitions).
•11,001,001 (4 transitions).
•01010011 (6 transitions).

In natural images, LBP is uniform. The uniform value can be found using the equation below:

$$LBP_{P,R}^{u2} = \begin{cases} \sum_{p=0}^{P-1} s\left(i_p - i_0\right), & U(LBP_{P,R}) \leq 2 \\ P(P-1)+2 & otherwise \end{cases} \quad (4)$$

where

$$U(LBP_{P,r}) = |s(i_{P-1} - i_c) - s(i_0 - i_c)| + \sum_{p=1}^{P} \left|s(i_P - i_c) - s\left(i_{p-1} - i_c\right)\right| \quad (5)$$

If U ≤ 2, it is a uniform LBP otherwise it is nonuniform LBP. The LBP space dimension is reduced from 2^P to P*(P-1) +2 output values. **Figure 6** represents an example of uniform and nonuniform patterns.

However, there have been different improvements in the LBP operator performance, such as improvement of its discriminative capability [48–53], enhancement of its robustness [54, 55], selection of its neighborhood [56–58], extension to 3D data [59–61], and combination with other approaches [62–65]. For more details, we refer to the survey done by Huang et al. [29].

3.2.2 Local directional numbers pattern (LDN)

A Local Directional Numbers Pattern (LDN) is proposed in the work of Rivera et al. [66]. It is a face descriptor that enables to acquire structural information and the intensity variations of the face texture. LDN descriptor extracts features by analysis of all eight (08) directions at every pixel position with a compass mask and generates a code from the analysis of its directional information. From all directions, the top positive and top negative directions are chosen to return a significant descriptor for different textures with similar structural patterns.

Local Directional Number Pattern (LDN) is a six-bit binary code. The resulting feature describes the local primitives, including different types of curves, corners, and junctions, more stably and more informative. They allow making differences in intensity changes in the texture. **Figure 7** represents an example of LDN code computation, and it is proposed in the work of Rivera et al. [66].

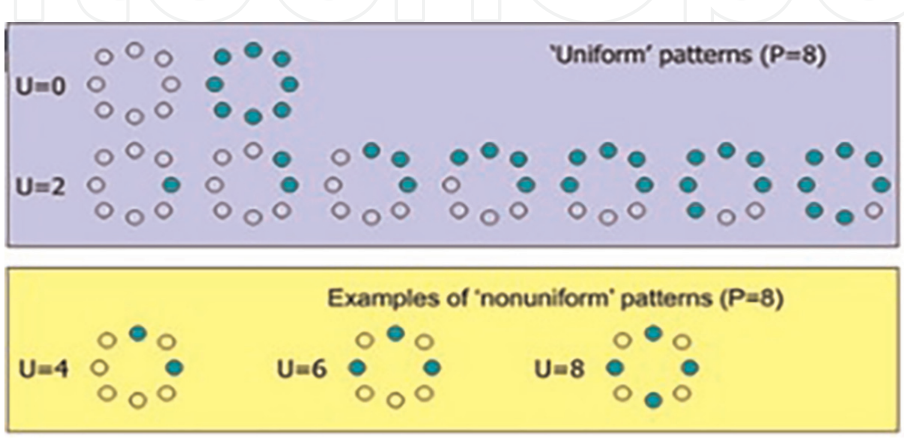


Figure 6.
Uniform and nonuniform patterns.

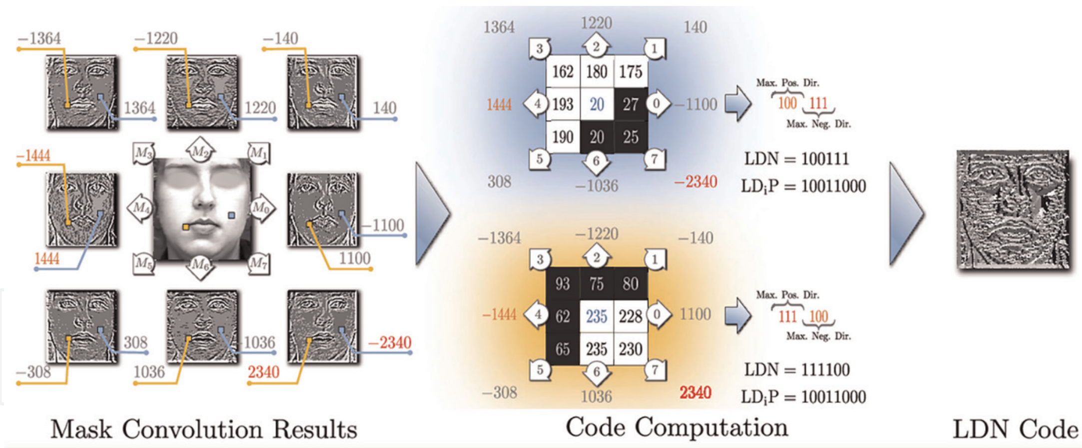


Figure 7.
LDN coding.

The produced code represents information on the texture structure and intensity transitions of each pixel of the input images. The LDN descriptor permits to use of the information of the entire neighborhood, instead of using sparse points. In the coding scheme, LDN code is generated by analyzing the edge response of each mask, representing edge significance in its respective direction, and combining the dominant directional numbers.

Edge responses are not equally important; a high negative or high positive value signals a prominent dark or bright area. The encoding of these outstanding areas is based on the sign information, the top positive directional number represents the three most significant bits in the code and the top negative the three least significant bits. The masks are shown in **Figure 8**; they take names of basic and secondary directions. The code is defined as:

$$LDN_{(x,y)} = 8i_{x,y} + j_{x,y} \tag{6}$$

where (x, y) is the central pixel of the neighborhood to encode, and ix,y and jx,y are directional number maximum positive and minimum negative responses, respectively, which are defined by:

$$\begin{aligned} i_{x,y} &= \arg\max_{i} \{\Pi^{i}(x,y) | 0 \leq i \leq 7\} \\ j_{x,y} &= \arg\min_{j} \{\Pi^{j}(x,y) | 0 \leq j \leq 7\} \end{aligned} \tag{7}$$

$$\underset{M_0}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}} \underset{M_1}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \underset{M_2}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \underset{M_3}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}}$$

$$\underset{M_4}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}} \underset{M_5}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \underset{M_6}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \underset{M_7}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}}$$

Figure 8.
Kirsch edge response masks.

where Π^i is the convolution of the original image, I, and the i^{th} mask, defined by $\Pi^i = I * M^i$.

This approach allows us to distinguish intensity changes (e.g., from bright to dark and *vice versa*) in the texture that otherwise will be missed most evident directions descriptor uses the information of the whole neighborhood, it does not use sparse points for its computation as it is for LBP. LDN translates the directional information of the face's textures (i.e., the texture's structure) in a compact way, producing a more discriminative code.

3.2.3 Edge orientation histogram

Edge Orientation Histogram (EOH) engenders a feature set extracted based on the gradient of the pixels that correspond to edges of an image. It is used as a descriptor in classification or detection tasks. These descriptors rely on the abundance of the information of edge and are invariant to global illumination [67–69]. The edge is computed by filtering the gray-scale image using the Sobel operator. Five operators provide information about the strength of the gradient in five particular directions, as represented in **Figure 9**.

Figure 9 represents the Sobel mask for five directions; in (a) it is the vertical direction, (b) it is the horizontal direction, (c) and (d) are the diagonals directions, and (e) it is the non-direction case. The gradient pixels are classified into β images corresponding to β orientation ranges; they are also designated as bins. Therefore, a pixel in bin $k_n \in \beta$ contains its gradient magnitude if its orientation is inside β's range, otherwise it is null. Integral images are now used to store the accumulation image of each of the edge bins. **Figure 10** represents the Edge Orientation Histogram.

$$\underset{(a)}{\begin{matrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{matrix}} \quad \underset{(b)}{\begin{matrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{matrix}} \quad \underset{(c)}{\begin{matrix} 2 & 2 & -1 \\ 2 & -1 & -1 \\ -1 & -1 & -1 \end{matrix}} \quad \underset{(d)}{\begin{matrix} -1 & 2 & 2 \\ -1 & -1 & 2 \\ -1 & -1 & -1 \end{matrix}} \quad \underset{(e)}{\begin{matrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{matrix}}$$

Figure 9.
Sobel mask for five directions [70].

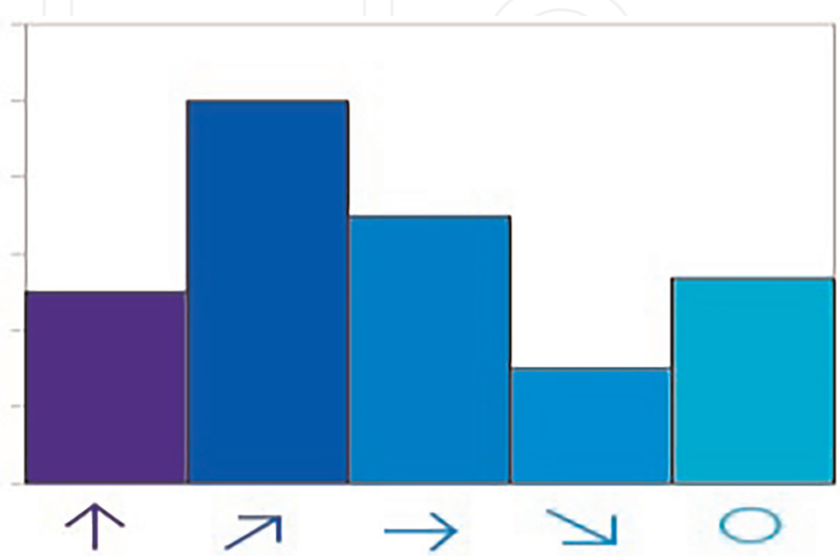


Figure 10.
Edge orientation histogram [70].

Though these two feature extraction approaches are mostly present in the literature but there are other works that considered hybrid approach [71]. A hybrid method means to at the same time use of appearance features and other shape features to make them complementary. Mixing these two types of features will improve classifier performance.

3.3 Feature extraction method classification

The facial expression recognition rate is more influenced by the basic features used for classifier training. From different works on facial feature extraction research, there are two main categories of feature extraction methods as mentioned above: geometric based and appearance based. In this work, we propose **Table 4** for a classification.

From this classification in **Table 4**, there are mainly two categories of feature extraction: appearance based and geometric based. Different cited methods

Feature category	Feature details	Techniques	References	Applications
Geometric based	FFP	Active Appearance Model (AAM)	Ratliff & Patterson [38]	Texture and Shape
		Active Shape Model (ASM)	Iqtait et al. [72]	Shape
	FACS	Holistic spatial analysis based on PCA, Feature-based approach and Facial motion analysis	Tian et al. [39]	Action Units
		Convolutional Experts Constrained Local Model (CE-CLM) and Histograms of Oriented Gradients (HOG	Yang et al. [40]	Geometric and appearance features
Appearance based	LBP	Improved LBP (Mean LBP)	Jin et al. [48], Bai et al. [49]	Effects of central pixels
		Hamming LBP	Yang and Wang [50]	Decrease of error rate caused by noise disturbances
		Extended LBP	Huang et al. [51]	Deals with variations of illumination
		Completed LBP	Guo et al. [53]	Better texture classification for rotation invariant
		Local Ternary Patterns	Tan and Triggs [54]	Discriminant and less sensitive to noise in uniform regions
		Soft LBP	Ahonen and Pietikäinen [55]	Robust to noise and output continuous according to input
		Elongated LBP	Liao and Chung [56]	New feature Average Maximum Distance Gradient Magnitude (AMDGM)

Feature category	Feature details	Techniques	References	Applications
		FMulti-Block LBP	Liao and Si [57]	More robust and consider integral image
		Three/Four Patch LBP	Wolf et al. [58]	Improves multi-option identification and same/not-same classification
		3D LBP	Fehr [59]	Texture analysis in 3D
		Volume LBP	Zhao and Pietikäinen [61]	Combines motion and appearance
		LBP and SIFT	Heikkilä et al. [63]	Tolerance to lighting changes, robustness on even image areas, and computational efficiency
		LBP and Gabor wavelet	Zhang et al. [73]	No need training procedure to build the face model
			He et al. [62]	
		LBP Histogram Fourier	Ahonen et al. [65]	Rotation invariant image descriptor
	EOH	Haar wavelet and EOH	Gerónimo et al. [67]	Object change in cluttered environments
		EOH for smile	Timotius and Setyawan [69]	Discriminate lip to depict a smile
	LDN	LDN basic	Rivera et al. [66]	Directional information of the face textures
		LDPv	Kabir et al. [10]	texture and contrast information of facial components

Table 4.
Classification of different feature extraction methods.

or techniques are used for facial expression feature extraction as well as other feature extraction-related work [68]. Among geometric feature extraction, the active appearance model is mostly used combined with the principal component analysis method to reduce the vector dimension for efficient application in real time. Among the appearance-based feature extraction, the local-based pattern algorithm is the mostly found in the literature and highly expended.

In recent work, in view of collecting enough features to enhance facial expression recognition rate by including more details, researchers propose hybrid methods [71, 72].

4. Conclusions

Automatic facial emotion recognition is a recent research trend that is applied in many areas, such as security, health, education, and social interaction. Facial feature extraction is one of the crucial steps in order to get a good and quick classifier at the end. In view of getting a performant classifier firstly, facial feature representation has to distinguish different individuals well and at the same time tolerate that there can be minor variation within-class members. It should be easy to be extracted from the basic facial images to speed up further processing; all that demands is that the final sample space must stay in a low dimensional space to reduce classification complexity.

This work pictures different methods used in facial feature extraction and their best usage. It can serve as a reference and guide to researchers in facial expression recognition. Hereby, cited methods are mainly applied to 2D images and but works considering 3D mage are also related. Actually, as devices are getting smarter and averaging natural perception, it is a judiciary that the corresponding software development follows.

Author details

Michele Mukeshimana*, Abraham Niyongere and Jérémie Ndikumagenge
Research Center in Infrastructure, Environment and Technology (CRIET), University of Burundi, Bujumbura city, Republic of Burundi

*Address all correspondence to: michele.mukeshimana@ub.edu.bi

References

[1] Cheng X, Zhan Q, Wang J, Ma R. A high recognition rate of feature extraction algorithm without segmentation. In: Proceeding of IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA). Tokyo, Japan: IEEE; 12-15 April, 2019. pp. 923-927. DOI: 10.1109/IEA.2019.8714943

[2] Picard RW. Affective Computing. United States of America (USA): MIT Press, MIT Media Laboratory Perceptual Computing Section Technical Report No 321; 1997

[3] Picard RW. Affective computing: From laughter to IEEE. IEEE Transactions on Affective Computing. Jan. 2010, vol. 1, no. 1, pp. 11-17. DOI: 10.1109/T-AFFC.2010.10

[4] Kamarol SKA, Jaward MH, Parkkinen J, Parthiban R. Spatiotemporal feature extraction for facial expression recognition. IET Image Processing. 2016; **10**(7):534-541

[5] Li SZ, Jain AK, editors. Handbook of Face Recognition. London Limited: Springer-Verlag; 2011. DOI: 10.1007/978-0-85729-932-1_4

[6] Han J, Kamber M, Pei J. Data preprocessing. In: Data Mining: Concepts and Techniques. Waltham, MA, USA: Elsevier Inc.; 2012. pp. 84-124

[7] Theodoridis S, Koutroumbas K. Feature selection. Pattern Recognition. 4th ed. Boston: Academic Press; 2009

[8] Shan C, Gong S, McOwan P. Robust facial expression recognition using local binary patterns. In: Proceedings of IEEE International Conference Image Processing. Genova Italy: IEEE; 14-14 September 2005. pp. 914-917

[9] Keche J-K, Dhore MP. Facial feature expression based approach for human face recognition: A review. International Journal of Innovative Science Engineering & Technology. 2014;**1**(3):1-5

[10] Kabir H, Jabid T, Chae O. Local directional pattern variance (LDPv): A robust feature descriptor for facial expression recognition. The International Arab Journal of Information Technology. 2012;**9**(4): 1-10

[11] Vertegaal R, Slagter R, van der Veer G, et al. Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Seattle, WA, USA: ACM Press; 2001. pp. 301-308

[12] Karpouzis, K. Editorial: “Signals to Signs” – Feature Extraction, Recognition, and Multimodal Fusion. In: Cowie R, Pelachaud C, Petta P, editors. Emotion-Oriented Systems. Cognitive Technologies. Berlin, Heidelberg: Springer; 2011. pp. 65-70 DOI: 10.1007/978-3-642-15184-2_5

[13] Guyon I, Elisseeff A. An introduction to feature extraction. Guyon I et al, editor. Feature Extraction, Studies in Fuzziness and Soft Computing. vol. 207. New York: Springer; 2006. pp. 1-25

[14] Cruz A, Garcia D, Pires G, et al. Facial Expression Recognition Based on EOG toward Emotion Detection for Human-Robot Interaction. In:

Proceedings of BIOSIGNALS. Lisbon, Portugal: SciTePress; 2015. pp. 31-37

[15] Sadr J, Jarudi I, Sinha P. The role of eyebrows in face recognition. Perception. 2003;**32**:285-293

[16] Rani PI, Muneeswaran K. Facial emotion recognition based on eye and mouth regions. International Journal Pattern Recognition and Artificial Intelligence. 2015;**30**(2016):1655020. DOI: 10.1142/S021800141655020X

[17] Li ZJ, Duan XD, Wang CR. Automatic expression recognition based on mouth shape analysis. Applied Mechanics and Materials. 2014;**644-650**: 4018-4022

[18] Hasan HS, Kareem SBA. Gesture feature extraction for static gesture recognition. Arabian Journal for Science and Engineering. 2013;**38**:3349. DOI: 10.1007/s13369-013-0654-6

[19] Graves A, Schmidhuber J, Mayer C, et al. Facial expression recognition with recurrent neural networks. In: International Workshop on Cognition for Technical Systems. Munich, Germany: coTeSys; 2008

[20] Vukadinovic D, Pantic M. Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics Waikoloa, Hawaii. October 10-12, 2005

[21] Bhatta LK, Rana D. Facial feature extraction of color image using gray scale intensity value. International Journal of Engineering Research & Technology (IJERT). 2014;**3**(3):1177-1180

[22] Huang ZQ, Chen L, Harper M. An open source prosodic feature extraction tool. In: Proceedings of the Language Resources and Evaluation Conference (LREC'2006). Genoa, Italy: European Language Resources Association (ELRA); 2006. pp. 2116-2121

[23] Pantic M, Caridakis G, André E, et al. Multimodal emotion recognition from low-level cues. In: P. Petta Pelachaud C, Cowie R, editors. Emotion-Oriented Systems, Cognitive Technologies, C©. Berlin, Heidelberg: Springer-Verlag; 2011. DOI 10.1007/978-3-642-15184-2_8

[24] Cowie R, Douglas-Cowie E, Savvidou S, et al. FEELTRACE: An instrument for recording perceived emotion in real time. In: Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research. Belfast, Ireland: ISCA; 2000. pp. 19-24

[25] Eyben F, Wöllmer M, Schuller B. Open EAR-introducing the Munich open-source emotion and affect recognition toolkit. In: Proceedings of the Third International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009). Amsterdam, Netherlands: De Rode Hoed; 2009. pp. 1-6

[26] Eyben F, Wöllmer M, Schuller B. openSMILE-the Munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM International Conference on Multimedia (MM'10). Firenze, Italy: Association for Computing Machinery (ACM); 2010. pp. 1459-1462

[27] Grimm M, Kroschel K, Narayanan S. The Vera am Mittag German audio-visual emotional speech database. In: Proceeding of the IEEE International Conference on Multimedia and Expo (ICME). Hannover, Germany: Institute of Electrical and Electronics Engineers (IEEE); 2008. pp. 865-868

[28] Castellano G, Kessous L, Caridakis G. Emotion recognition through multiple modalities: Face, body gesture, and speech. In: Christian P, Beale R, editors. Affect and Emotion in HCI. Vol. 4868. Springer: Berlin, of the series LNCS; 2008. pp. 92-103

[29] Huang D, Shan C-F, Ardebilian M, et al. Local binary patterns and its application to facial image analysis: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2011;**41**(6):765-781

[30] Kipp M. Anvil - a generic annotation tool for multimodal dialogue. In: Proceedings of Eurospeech. Aalborg, Denmark: Citeseer; 2001. pp. 1367-1370

[31] Mohamed E, ElGamal A, Ghoneim R, et al. Local binary patterns as texture descriptors for user attitude recognition. International Journal of Computer Science and Network Security (IJCSNS). 2010;**10**(6):222-229

[32] Wu YM, Wang HW, Lu YL, Yen S, Hsiao YT, Pan JS, et al. Intelligent Information and Database Systems. ACIIDS 2012. Lecture Notes in Computer Science, vol. 7196. Berlin, Heidelberg: Springer. pp. 228-238. DOI: 10.1007/978-3-642-28487-8_23

[33] Tian YL, Kanade T, Cohn JF. Facial Expression Analysis: Handbook of Face Recognition. New York, NY: Springer; 2005. DOI: 10.1007/0-387-27257-7_12

[34] Sumathi CP, Santhanam T, Mahadevi M. Automatic facial expression analysis a survey. International Journal of Computer Science & Engineering Survey (IJCSES). 2012;**3**(6):259-275

[35] Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. In: Proceedings of the Fifth European Conference on Computer Vision (ECCV'98). Vol. 1407. Freiburg, Germany: LNCS; 1998. pp. 484-498

[36] Wu C-H, Lin J-C, Wei W-L. Survey on audiovisual emotion recognition: Databases, features, and data fusion strategies. APSIPA Transactions on Signal and Information Processing. 2014; **2014**(3):e12

[37] Ahdid R, Taifi K, Safi S, Manaut B. A survey on facial feature points detection techniques and approaches. International Journal of Computer and Information Engineering. 2016;**10**(8):1-8

[38] Ratliff MS, Patterson E. Emotion recognition using facial expressions with active appearance models. In: Proceedings of HCI 2008 (HCI). Liverpool, UK: Citeseer; 1-5 September 2008. pp. 89-98

[39] Tian Y-L, Kanade T, Cohn J-F. Recognizing AUs for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; **23**(2):91-116

[40] Yang J, Zhang F, Chen B, Khan SU. Facial Expression Recognition Based on Facial Action Unit. In :Proceedings of the Tenth International Green and Sustainable Computing Conference (IGSC). Alexandria, VA, USA: IEEE, October 21-24, 2019. pp. 1-6. DOI: 10.1109/IGSC48788.2019.8957163.2019:1-6

[41] Hager JC. A comparison of units for visually measuring facial actions. Behavior Research Methods Instruments & Computers. 1985;**17**(4):450-468

[42] Friesen E, Ekman P. Measuring facial movement. Journal of Nonverbal Behavior 1. 1976;**1976**:56–75

[43] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured

distributions [J]. Pattern Recognition. 1996;**29**(1):51-59

[44] Mukeshimana M, Ban X-J, Karani N. Toward instantaneous facial expression recognition using privileged information. International Journal of Computer Techniques. 2016;**3**(6):23-29

[45] López MB, Nieto A, Boutellier J, et al. Evaluation of real-time LBP computing in multiple architectures. Journal of Real-Time Image Processing. 2017;**13**(2):375-396

[46] Pietikäinen M, Hadid A, Zhao G, et al. Lbp in different applications. Computer Vision Using Local Binary Patterns. 2011;**40**:193-204

[47] Ojala T, Pietikäinen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Analysis and Machine Intelligence. 2002;**24**(7):971-987

[48] Jin H, Liu Q, Lu H, et al. Face detection using improved LBP under Bayesian framework. In: Proceeding of International Conference on Image and Graphics (ICIG). Hong Kong, China: Institute of Electrical and Electronics Engineers (IEEE); 2004. pp. 306-309

[49] Bai G, Zhu Y, Ding Z. A hierarchical face recognition method based on local binary pattern. In: Proc. Congress on Image and Signal Processing. Sanya, Hainan, China: Institute of Electrical and Electronics Engineers (IEEE); 2008. pp. 610-614

[50] Yang H, Wang Y. A LBP-based Face Recognition Method with Hamming Distance Constraint. In: Proceedings of the International Conference on Image and Graphics (ICIG 2007), Chengdu, China. 2007. pp. 645-649. DOI: 10.1109/ICIG.2007.144

[51] Huang D, Wang Y, Wang Y. A robust method for near infrared face recognition based on extended local binary pattern. In: Proc. Int. Symposium on Visual Computing (ISVC). Las Vegas, Nevada, USA: Springer Berlin Heidelberg; 2007. pp. 437-446

[52] Huang Y, Wang Y, Tan T. Combining statistics of geometrical and correlative features for 3D face recognition. In: Proc. British Machine Vision Conference (BMVC). Edinburgh, UK: Citeseer; 2006. pp. 879-888

[53] Guo Z, Zhang L, Zhang D. A completed modeling of local binary pattern operator for texture classification. IEEE Transactions on Image Processing (TIP). 2010;**19**(6):1657-1663

[54] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Proc. Analysis and Modeling of Faces and Gestures (AMFG). Rio de Janeiro, Brazil: Springer Berlin Heidelberg; 2007. pp. 168-182

[55] Ahonen T, Pietikäinen M. Soft histograms for local binary patterns. In: Proc. Finnish Signal Processing Symposium (FINSIG). Finland: Citeseer; 2007. pp. 1-4

[56] Liao S, Chung ACS. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In: Proc. Asian Conf. Computer Vision (ACCV). Tokyo, Japan: Springer Berlin Heidelberg; 2007. pp. 672-679

[57] Liao S, Li SZ. Learning multi-scale block local binary patterns for face recognition. In: Proceedings of the International Conference on Biometrics. Seoul, Korea: Springer Berlin Heidelberg; 2007. pp. 828-837

[58] Wolf L, Hassner T, Taigman Y. Descriptor based methods in the wild. In: Proc. ECCV Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. Marseille, France: INRIA; 2008

[59] Fehr J. Rotational invariant uniform local binary patterns for full 3D volume texture analysis. In: Proc. Finnish Signal Processing Symposium (FINSIG). Finland: Citeseer; 2007

[60] Paulhac L, Makris P, Ramel J-Y. Comparison between 2D and 3D local binary pattern methods for characterization of three-dimensional textures. In: Proc. Int. Conf. Image Analysis and Recognition. Póvoa de Varzim, Portugal: Springer Berlin Heidelberg; 2008

[61] Zhao G, Pietikäinen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007;**29**(6):915-928

[62] He L, Zou C, Zhao L, et al. An enhanced LBP feature based on facial expression recognition. In: IEEE Engineering in Medicine and Biology Society 27th Annual Conference. IEEE; 2005. pp. 3300-3303

[63] Heikkilä M, Pietikäinen M, Schmid C. Description of interest regions with local binary patterns. Pattern Recognition. 2009;**42**(3):425-436

[64] Huang D, Zhang G, Ardabilian M, et al. 3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching. In: Proc. IEEE International Conference on Bio-Metrics: Theory, Applications and Systems. Washington DC, USA: Institute of Electrical and Electronics Engineers (IEEE); 2010. pp. 1-7

[65] Ahonen T, Matas J, He C, et al. Rotation invariant image description with local binary pattern histogram fourier features. In: Proc. Scandinavian Conference on Image Analysis (SCIA). Oslo, Norway: Springer Berlin Heidelberg; 2009. pp. 61-70

[66] Rivera AR, Castillo RJ, Chae OO. Local directional number pattern for face analysis: Face and expression recognition. IEEE Transactions on Image Processing. 2013;**22**(5):1740-1752. DOI: 10.1109/TIP.2012.2235848

[67] Gerónimo D, López A, Ponsa D, et al. Haar wavelets and edge orientation histograms for on-board pedestrian detection. Martí J, et al, editors. IbPRIA. Berlin, Heidelberg: Springer-Verlag; 2007. pp. 418-425

[68] Alefs B, Eschemann G, Ramoser H, Beleznai C. Road Sign Detection from Edge Orientation Histograms. IEEE Intelligent Vehicles Symposium. 2007. pp. 993-998. DOI: 10.1109/IVS.2007.4290246

[69] Timotius IK, Setyawan I. Evaluation of edge orientation histograms in smile detection. In: Proceedings of the 6th International Conference on Information Technology and Electrical Engineering (ICITEE). Yogyakarta, Indonesia: IEEE; 2014. pp. 1-5. DOI: 10.1109/ICITEED.2014.7007905

[70] Edge orientation histograms in global and local features (0ctave_ Matlab) - that doesn't make any sense [Internet]. Available from: http://robertour.com/2012/01/26/edge-orientation-histograms-in-global-and-local-features/ revisited on December 9th, 2022

[71] Kaul A, Chauhan S, Arora AS. Hybrid approach for facial feature extraction. International Journal of

Engineering Research & Technology (IJERT). 2016;**4**(15):1-3

[72] Iqtait M, Mohamad FS, Mamat M. Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM). In proceedings of IOP Conf. Series: Materials Science and Engineering. 2018; **332**:012032. DOI: 10.1088/1757-899X/332/1/012032

[73] Zhang W, Shan S, Gao W, Chen X, Zhang H. Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Proceedings of the IEEE International Conference of Computer Vision (ICCV). Beijing. PR China: Institute of Electrical and Electronics Engineers (IEEE); 17-21 October 2005. pp. 786-791