International Journal of Pattern Recognition and Artificial Intelligence

An Evolutive Approach for Smile Recognition in Video Sequences

DAVID FREIRE-OBREGÓN
MODESTO CASTRILLÓN-SANTANA
SIANI - Universidad de Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, 35017, Spain
dfreire@siani.es

Facial expression recognition is one of the most challenging research areas in the image recognition field and has been actively studied since the 70's. For instance, smile recognition has been studied due to the fact that it is considered an important facial expression in human communication, it is therefore likely useful for human-machine interaction. Moreover, if a smile can be detected and also its intensity estimated, it will raise the possibility of new applications in the future. We are talking about quantifying the emotion at low computation cost and high accuracy. For this aim, we have used a new support vector machine (SVM) based approach that integrates a weighted combination of local binary patterns (LBPs) and principal component analysis (PCA) based approaches. Furthermore, we construct this smile detector considering the evolution of the emotion along its natural life cycle. As a consequence, we achieved both low computation cost and high performance with video sequences.

Keywords: Facial expression; smile detection; LBP; PCA; SVM

1. Introduction

Since relatively recently, it is known that emotions play a very significant role in decision-making by humans[7]. For example, the content of an email can generate a person to feel confused and consequently puckered, or in addition furrowed, leaning slightly toward the screen. The ability to show, and above all, to interpret emotions is essential in human-machine interaction. There exist authors like Picard[28] that justify the need to develop this new communicative dimension with machines, not only for the possibility of developing new forms of interpretative or gesture-based interaction, but also to help reduce, from a psycho-social perspective, user frustration.

Unlike general facial expression recognition, research in detecting smiles had produced, until recently, much less literature. However, in recent years these type of works have been focused on the development of two critical applications: first being the "smile trigger" of digital cameras[6] that allows for an image to be captured automatically when persons of a scene smile at the lens and, on the other hand, social detectors for detecting human smiles during their daily lives[16]. For

both applications, the smile detection system needs to operate with a wide range of conditions such as lighting, geographical location, ethnicity, gender or age, in addition to dealing with the device hardware acquisition of images (resolution of the camera, zoom, etc).

The contributions of this paper are: 1) an exhaustive analysis of the importance of facial dynamics in order to identify truly smile expressions, 2) a novel smile detector from a dynamic perspective in which the presence (or absence) of a smile is not determined by the facial configuration at a certain frame, but defined by the evolution of the facial action of a person along a temporary interval, 3) and a new dynamic approach for smile detection considering a temporary interval and a weight distribution combining multiple static approaches. Contrary to the second contribution, our particular contribution combines different approaches in order to obtain a behavior response of the life cycle of the facial expression. On the other hand, different intensities of the facial expression are also considered in the study. In fact, this is an interesting way to analyze the facial expressions. Recently, Werner et al.[37] have proposed an approach in order to determine the dynamic intensity variation of the pain expression during a period of time.

The paper is structured as follows: the following section provides a review of the state of the art and of the particularities of the human smile. Section 3 provides the theoretical framework of the research carried out in this article. The features and the classification approach considered can be seen in section 4. The proposed dynamic approach is described in section 5. Hence, results of the proposal are described in section 6. Finally, sections 7 and 8 present the discussion and conclusions respectively.

2. State of the Art

The smile is one of the facial expressions that occurs more often in our daily life and therefore requires special attention. People smile to be nice to each other and generally it is a common expression reflecting pleasure but can also be conceived from contrary feelings, such as anger. In this sense, this expression often happens unintentionally in a state of anxiety. Ekman showed that smiling is a normal reaction and natural to certain stimuli and occurs independently of the culture when it happens[9]. Among the existing types of smiles, the common smile or "Duchenne smile" is the most studied smile and involves the movement of the zygomaticus major and minor muscles close to the mouth region and the orbicular muscle near the eye region. The "Duchenne smile" is defined as a symmetrical smile, produced as an involuntary response to a genuine emotion, what is called a "genuine smile"[9]. Within detection of this type of expression, there exist works aimed at detecting strictly this Duchenne smile. They are based on the Facial Action Coding System (FACS) developed by Ekman and Friesen[10]. The Facial Action Coding System classifies expressions in individual actions of the various facial components. Some authors identify these types of smiles as those in which units of action AU12 and

AU6 happen simultaneously. In Figure 1, one can see a set of subjects classified as genuine or fake smiles, according to Liu[24]. As can be seen, the genuine are those in which there are changes in the already mentioned, eye (orbicularis muscle) and mouth (zygomatic muscle) regions.

Fig. 1. Genuine Smile vs Fake Smile[24].

On the other hand, Ekman[9] claimed that the presence of the subjectivity of a person over the emotions he perceives, determines the decision-making process about the facial expression he ought to show. The same author also suggests that the relativity of the perceived emotion can be extrapolated to the reaction of the subjects, and even with the possibility of generalization when encoding the expression. However, there is always the possibility that this analysis cannot be extrapolated to everyone. In other words, although the smile exists in all cultures, there may be nuances between how a subject has to smile inside of each culture and, depending on the context, it is possible that, see Figure 1, there exist smiles that are genuine marked as fake. This is the reason because this paper is not intended to elucidate the quality of the smile based on FACS, but to improve the existing mechanisms of smile detection through the temporal behavior of facial muscles.

Among the most notable works in the past about the smile recognition, various techniques have been developed. Ito[18] used lip edges and a perceptron. This denotes that lips are the most important region as it involves the movement of the, previously mentioned, zygomatic muscles to raise the ends of the mouth. In fact, in the model based approaches facial key points like eyes, eyebrows, nose, lips etc. have to be located and tracked in order to estimate the facial expression according to the relative position of these facial key points[21]. However, the edges of the lips may be insufficient features for recognizing whether a person smiles or not. Hence, Kowalik[22] made use of a set of points located around the mouth in order to train a

neural classifier. In 2011, Zhang[40] proposed a Facial Expression Recognition (FER) system by using "salient" distance features in order to consider facial element and muscle movements. More recently, Shan[29] introduced an efficient approach to smile detection, in which the intensity differences between pixels in the grayscale face images are used as features. Their experiments show that our approach has similar accuracy to the state-of-the-art method but faster. On the other hand, Déniz et al.[8] proposed an approach for real time interaction. They suggested that the eye, nose and mouth regions provide enough information to decide whether a person is smiling or not. Meanwhile, the work of Shinohara[31] offers a correlation system of higher order for smile detection, reaching very high rates of accuracy in controlled environments. Finally, the conclusions of the work of Whitehill[38] argue that there is a big difference between typical testing literature and the results obtained in actual situations. These authors contend that the training set plays almost as important role as the detection method applied, especially in terms of variability and size, and asserts that a normal training phase may handle thousands of images. More recently, works using public databases have been proposed by Huang[17], Yadappanavar[39] and Shimada[30]. These works as well as two commercial systems are discussed in section 7.

3. Theoretical Background

A facial expression is not generated discretely at any given moment but it does so continuously over a period of time. Koelstra[20] labeled this phenomenon of formation of expressions in the human face as a temporal sequence that goes through these segments: neutral, activation, splendor, and deactivation. What this theory states is that expressions are not actuated through a switch, but that the evolution of the facial map determines the intensity of the expression and the phases this expression go throughout its life cycle.

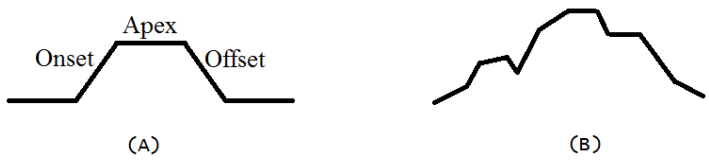

Fig. 2. Natural life cycle of a facial expression. On the left image, the life cycle of a genuine facial expression can be appreciated. On the other hand, the irregular behavior of the life cycle leads a fake facial expression shown on the right image graph[20].

In this context, Figure 2 presents two timelines of behavior that facial expressions can experience over their natural life cycle; the natural behavior described above can be seen in the image on the left. The activation stage begins when the facial muscles start to contract and increase the intensity of their movements

while the splendor stage comes when intensity reaches a stable state. Finally, the off stage arises from the relaxation of the muscles until a neutral facial expression is reached. On the other hand, the image on the right shows a set of steps in which the muscular movement of the face is not continuous or spontaneous, resulting in a mock facial expression. This includes a hesitant activation step, a very short duration (or too long) on the stage of splendor and a very sharp deactivation.

Other authors[15] have stated that the existence of motion video allows the analysis of relevant aspects of facial expressions such as specific muscle actions, intensity levels, expression asymmetry or other characteristics.

Relying on that theory of temporal facial expressions, this research aims at analyzing in a meticulous way whether, once found an interesting approach for the detection of smiles, it is possible to apply it in a dynamic approach for smile detection. In other words, the proposed dynamic approach takes advantage of the temporal coherence that video provides and image does not.

Indeed, the proposed work is intended to exploit the life cycle derivative of the smile. The objective is not only to determine whether a person smiles or not, but also to quantify the intensity of this facial expression. Finally, it will be appreciated how Koestra theory can be applied on real cases. Hence, in the next section a set of techniques and the classification approach used in the proposed work in Section 4 are explained.

4. A static approach

This article extends our earlier publication[11] in this topic where we made an extensive study on the problem of smile detection. The idea followed in that work was to analyze both the full face as well as its most important regions according to Ekman; eyes and mouth. For that reason, several techniques were combined such as principal component analysis, simplified LBP or uniform LBP with two types of classifiers; the support vector machine (SVM) and the k nearest neighbors algorithm (k-NN). The best results arose from the following approaches:

- PCA space obtained as a result of the original grayscale images.
- The concatenation of histograms obtained from the original image encoded with Uniform LBP (ULBP) considering the eyes and mouth areas.
- The concatenation of pixel values from the image encoded with Simplified LBP (SLBP), considering only the mouth region.

Furthermore, that research showed that an approach based on a SVM classifier outperformed k-NN for most features spaces.

According to these observations, we selected SVM as classification framework. The next sections introduce briefly the features used by the three best approaches, and formalize the classification problem.

4.1. *Methods and classification*

The local binary pattern (LBP) is an image descriptor which is usually used in classification tasks. It was initially introduced by Ojala et al.[26] for texture classification, but its popularity grew because of the invariance presented under changes in illumination and the low cost processing. This characteristics allowed them to be applied relatively successful in solving classification problems, especially when a certain robustness to changes in the processing image luminance is needed [33].

The computation of this type of pattern is relatively simple. Given a grayscale image, a pixel is selected and the LBP operator is applied to its neighbors in order to obtain the LBP pattern for the selected central pixel. Therefore, a threshold operation is performed in a circular way involving the surrounding pixels and the central pixel previously selected. Then, the result of each of these threshold operations leads to the computation of the binary pattern. Ojala[27] considered for his basic version a 3×3 pixel window and built the local binary pattern based on the central pixel information and their eight neighbors. However, the own definition LBP allows an easily adaptation to other ratios (R) considering P neighbors:

$$\begin{aligned} LBP_{P,R}(x_c, y_c) &= \sum_{k=0}^{P-1} s(g_p - g_c)2^k \ , \\ s(x) &= \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \end{aligned} \tag{1}$$

On the other hand, the LBP is a circular pattern, therefore it is invariant under rotations. The experience gained in the experiments conducted by Ojala et al. suggested that only a particular subset of local binary patterns use to appear in most of the analyzed pixels. This author defines this subset of patterns as uniform patterns (ULPB) and they are characterized by the fact that they contain, as most, two bitwise transitions from 0 to 1 or viceversa. In other words, this mean that a uniform pattern is a pattern with no transitions between 0 and 1 (for example, 00000000 or 11111111) or it has, at most, two bitwise transitions (00011100 or 100000011).

More recently, it has been established that LBP has been found useful in describing the facial appearance. In fact, once the pattern LBP is computed, most authors apply a data representation based on histograms[25,1]. However, as shown in recent works which have made use of this LBP approach based calculate on histograms, there is a loss of the relative location of the image data[25,34]. These authors suggested that using LBP as preprocessing method does not favor certain classification strategies because LBP has the effect of emphasizing edges and noise. To reduce this noise, Tao Qian[34] proposed a modification of Eq. (1). In order to compute this new approach, the weight distribution among all the pixels is exactly the same. This operation generates the descriptor known as simplified local binary pattern (SLBP). Eq. (2) describes de SLBP computation.

$$SLBP_{P,R}(x_c, y_c) = \sum_{k=0}^{P-1} s(g_p - g_c) \ , \qquad s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2}$$

According to Tao Qian, this new approach shows some benefits applied to facial verification due to the fact that by simplifying the weights, the image is more robust to changes in lighting (providing a maximum of nine possible pixel values). Thus, the number local patterns is dramatically reduced and the image is represented in a more restricted domain.

The principal component analysis (PCA)[19] is a very useful statistical technique that has been remarkable applied in the area of automatic facial analysis during the last decades. In fact, it is a quite common technique to find patterns in sets of data which offer multiple dimensions and express this data in a way to highlight both; similarities between data and differences without losing much information during the process. For this reason, a covariance analysis between the different factors is performed. Hence, given a target image, the PCA decomposition projects it into the space (see Figure 3) and the appearance of different individuals is represented in lower dimensionality through average latent variables v_i[35].

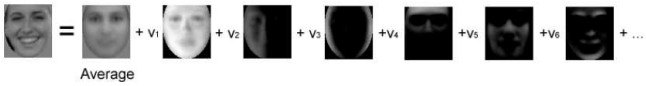

Fig. 3. PCA descomposition of a human face.

After extracting the features from face images, the support vector machines[2] (SVM) are used as strong classifiers. Moreover, the SVM are intended to find a separating hyperplane which fits properly according to the features provided by the input data, maximizing the separation between both classes. Furthermore, the main property of support vector machines is the ability to minimize the classification empirical error as well as maximize the margin between the different classes involved in the classification process.

5. A dynamic approach

In this work, a combination of classifiers considering two dimensions is proposed. Firstly, authors consider, for a given classifier, a temporal window using both current and previous frame decisions to take a decision based on the dynamic behavior of the face. And secondly, authors fuse for the final decision the information provided by different classifiers for the current frame (see Figure 4). Our aim is to

reduce classification errors. Moreover, it is expected a reduction of classification errors by means of the resulting fusion of classifiers in time.

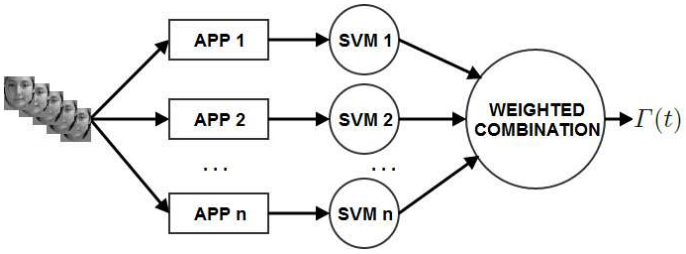

Fig. 4. Flowchart of the dynamic approach.

For the temporal window combination, as expressed in Eq. (3), before deciding whether a detected face in a frame is smiling or not, the response to the set of previous frames is integrated. A weighted combination is defined giving larger weights to more recent frames (see Figure 5). For a given approach, *app*, the temporal based decision within a temporal window size of w_s frames is defined as:

$$\begin{aligned} Dt_{app} &= \frac{\partial w(t) \partial \tau_{app}(t)}{\partial t} \\ &= \sum_{k=t-w_s}^{t} \left(\frac{1}{1+sqrt(t-k)}\right) * (\tau_{app}(k)) \end{aligned} \tag{3}$$

Where $\tau_{app}(k)$ is the decision given by a single classifier on the frame at a given instant k. On the other hand, each classifier can only generate outputs in the range [-1; +1]. Hence, a positive score indicates a smiling face, a negative its opposite. In order to define a larger weight to the closer frames, each frame decision is weighted by the term $(\frac{1}{1+sqrt(t-k)})$. Attending to the temporal window size, a larger window size can provide a greater confidence in classifying a stable emotion. However, it may also affect negatively the system reactivity during the activation or deactivation phase of the emotion (see Figure 2).

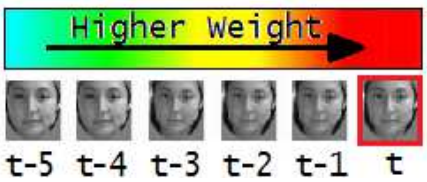

Fig. 5. Weight influence during the processing of a keyframe at a given instant t (framed in red) considering a five frames window size ($w_s = 5$).

A second fusion process is performed considering a set of classifiers. After evaluating for each classifier or approach the recent temporal window, the resulting scores are combined within a frame making use of a set of weights. These weights are defined attending to the individual approaches considered. In our case, the approaches studied in section 4 are used. Their combination is defined as follows:

$$\Gamma(t) = \sum_{app=1}^{n} \psi_{app} * Dt_{app} \tag{4}$$

Where ψ_{app} defines the weight of each classifier. A greater weight is assigned to those approaches that report better rates. With this aim, an experimental study is carried out in section 6 in order to provide these weights to the considered approaches. In any case, this assigment is done avoiding that a single classifier weigth may beat the other decisions unilaterally. Hence, weights are assigned considering that there is no classifier with an absolute majority on the decision process. The decision of the classifier with the highest weight is not guaranteed to succeed; the others classifiers can revoke the proposed weak decision. Thus, given n approaches, the weight $\psi_i \in \mathbb{R}$ of an approach i is defined as:

$$\forall i, j \in \mathbb{N} \implies \psi_i \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} \psi_j \tag{5}$$

6. Experiments

6.1. *Datasets*

Different independent datasets are chosen for training and testing. For training we have used the same dataset considered in Freire et al.[11] that combines images randomly downloaded from internet with images taken from facial databases such as the BIOID dataset[13]. All these images were manually annotated indicating whether or not the subject smiles. Finally, these images have served as input to train the different classifiers that are combined in the proposed approach.

For testing, we firstly made use of the non public SIANI video database to find the most suitable temporal window configuration for each approach. This dataset contains spontaneous recordings of 106 people. This collection was not created specifically for facial expression analysis, but for face detection[3]. Although this database cannot be distributed due to license agreement, the results achieved are included to illustrate the reader the performance with a less restrictive dataset. Moreover, this database contains 59386 frames, where 17406 of them show smiling people and the rest of them (41980 frames) show not smiling people.

Once the best configuration for each approach has been studied, the DaFEx database provides an interesting framework because it has been widely used by the community for this problem. Furthermore, this database contains eight professional

Fig. 6. Different intensity levels of emotions at the DaFEx database. A total number of 12988 frames where extracted from video sequences of subjects playing a smile. From this number of frames, 6783 correspond to smiling, and 6205 to non smiling face.

actors that perform seven facial expressions (six basic facial expressions and one neutral) with three intensity levels (low, medium and high). Thus, each actor presented a facial expression twice along with an intensity level (see Figure 6). Hence, the DaFEx database is directly analyzed considering the previously collected best window configuration.

Considering that the system is processing video sequences, we have also made use of a face detector for video sequences[3] that combining several cues outperforms single cue detectors[36] when applied to video. The resulting detected faces are cropped and normalized to 59×65 pixels.

6.2. *Results*

Before addressing the combined approach, a number of individual tests have been made for each of the three proposed approaches in section 4;

- **App1** - The PCA space obtained as a result of the original grayscale images.
- **App2** - The concatenation of histograms obtained from the original image encoded with ULBP considering eyes and mouth.
- **App3** - The concatenation of pixel values from the image encoded with SLBP, considering only the mouth.

First of all, a temporal window analysis on the SIANI database has been carried out to define the most suitable window size for each approach. Hence, the results considering the most suitable temporal window configuration for each approach are; App1 (w_s=5) with an error rate of 17.4% (15.3% FNR and 22.3% FPR), App2 (w_s=15) with an error rate of 23.8% (22.3% FNR and 27.2% FPR) and App3 (w_s=20) with an error rate of 26.7% (25% FNR and 30.6% FPR). As can be appreciated, the temporal window study on the SIANI database shows that the most suitable window size for each approach varies depending in the characteristics of the approach. Furthermore, the approach that provides the best rates (App1) requires a small temporal window size. On the other hand, those approaches that consider the image texture (App2 and App3) require a larger window size to achieve their best results. Therefore, the correlation between the temporal window size and

Table 1. Error rates for each single descriptor approach using the three DaFEx database intensity levels for smiling video sequences. Different window size for each approach was considered according to Eq. (3). FNR stands for "False Negative Rate" while FPR stands for "False Positive Rate". Error rates are calculated over the total number of frames. For each video sequence, a smile was detected in at least one frame.

DaFEx	Approach	Window	Error	FNR	FPR
Complete Database	App1	5	11.6%	16.2%	07.3%
	App2	15	18.8%	20.1%	17.7%
	App3	20	20.1%	21.1%	19.6%
Low intensity set	App1	5	13.4%	14.4%	12.0%
	App2	15	20.9%	20.0%	21.6%
	App3	20	22.2%	23.3%	21.3%
Medium intensity set	App1	5	13.3%	18.1%	08.7%
	App2	15	19.9%	21.5%	18.0%
	App3	20	20.4%	20.4%	20.6%
Hight intensity set	App1	5	06.4%	11.5%	01.7%
	App2	15	16.2%	19.5%	13.3%
	App3	20	17.8%	19.7%	16.6%

the accuracy of the analyzed approach suggests that the higher the accuracy is, the higher the confidence will be.

The results for each classifier after processing the smiling videos sequences of the DaFEx database can be observed in Table 1. As can be appreciated, the error rates for the SIANI video dataset are quite different from the DaFEx database due to the head rotations and the pose variations during the acquisition process of the SIANI dataset. Accordingly to the SIANI database results, the best performance on the DaFEx database is obtained using directly the grayscale image, i.e. App1. On the other hand, the App2 presents really interesting rates because it models in an appropriate way smile textures through the use of normalized histograms. Finally and contrary to the App2, the App3 gains a better balance between the rate of false positives and negatives, but its error rate is quite similar to the App2.

Once different conclusions in this first experiment have been extracted, we propose a new approach according to Eq. (4). As it has happened before, the SIANI database provides an interesting framework to work under less restricted conditions. Our experiments have shown that a 5 frames window size has provided the best error rate for each test (considering image detection at 30 fps). Figure 7 shows two different graphs of the classifier applied on two sample video sequences recorded for the SIANI dataset. The graphs include; 1) the manual annotation and the results achieved for different approaches (+1 smiling and -1 non smiling) indicated by thick blue lines; and 2) the results achieved with five different temporal window size (red, green, blue, sky blue and black for respectively 5, 8, 10, 15 and 18 window size). In addition, the horizontal line is the boundary between the state smiling (above the line) and the no smiling (below the line). The handmade labeling does not allow to clearly discern stages of activation (onset) or deactivation (offset), the slope is

maximal without any natural evolution of the facial expression according to what is disclosed in Figure 2. On the left graph, the classifier response presents a more *natural* behavior of the evolutionary cycle, ensuring a relative stability of the apex stage and providing a progressive behavior of the onset and offset stages. An opposite behavior can be appreciated on the right graph, there is no stability on the smile detection process. According to Koelstra[20], this second behavior suggests a not genuine smile. This suggestion is not based on the smile features of the frame, but on the life cycle of the facial expression.

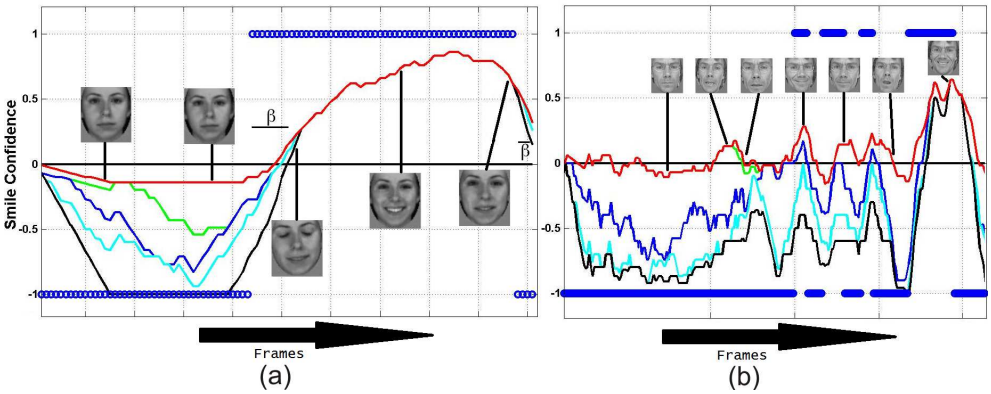

Fig. 7. Temporal evolution response of the proposed approach trough two video sequences. As it can be appreciated, the X-axis line marks the boundary between smiling (above the line) or not smiling (under the line) depending on the response signal. For instance, blue marks at the +1 y-axis and -1 y-axis stand for the manual annotations, while the several color lines stand for the result of the classifier for different window size based on Eq. (4). The β range shows the small delay due to the temporal window use.

On the other hand, there is a straight relation between the confidence of the decision made by the approach and the height of the output curves shown in Figure 7. In other words, there is a higher confidence if the approach's response generates a higher curve because it means that each individual classifier agrees with the others.

A last consideration, a small delay can be seen at the beginning and the end of the facial expression, just after the stages of onset and offset (see β in Figure 7-(a)). This occurs due to the use of a temporal window because, even if it provides more stability and supports a progressive response to the activation and deactivation phases, it also generates a small gap before the beginning of changes in the facial expression of a given individual.

Besides, an empirical study considering the Eq. (5) has been carried out to define the weight assigned to each approach. Those approaches that offered a lower error rate during the initial study have been assigned a greater weight in this study. It may seem that there is a correlation between the weight distribution (ψ) and the temporal window size (w_s). Therefore, the weight distribution shows an

improvement of the App2 as far as the smile intensity increases. On the other hand, the App3 seems to behave better in the opposite case. Moreover, the appearance of low intensity values benefits the accuracy of this approach. However, the higher weight is assigned to App1 almost always due to the performance achieved during the initial test shown in Table 1. The improvement for the SIANI database is remarkable, using the same temporal window size and weight configuration shown on Table 2, the error rate is 14.7% (17.7% FPR and 13.44% FNR). More in detail, the results for the new approach after computing the DaFEx smile dataset are presented in Table 2. Furthermore, in Table 2 it is observed that the new approach combines a low error rate and a balanced relationship between the rate of false positives and negatives.

Table 2. Error rates of the combined approaches using the three DaFEx database smiling intensity subsets. FNR stands for "False Negative Rate" while FPR stands for "False Positive Rate". Error rates are calculated over the total number of frames. For each video sequence, a smile was detected in at least one frame.

DaFEx	Window	Error	FNR	FPR	ψ_i
Complete Database	5	10.5%	11.7%	09.0%	App1-2.0 App2-1.3 App3-0.7
Low intensity set	5	13.8%	17.1%	11.2%	App1-2.0 App2-1.3 App3-0.7
Medium intensity set	5	11.3%	14.4%	07.8%	App1-2.0 App2-1.5 App3-1.0
Hight intensity set	5	05.7%	06.2%	05.4%	App1-1.8 App2-1.7 App3-0.5

7. Discussion

The experiments presented in this paper explore the temporal coherence which provides a video sequence, but also to explore the life cycle of the facial expressions to determine accurately the activation (and deactivation) of a user's smile. Furthermore, a face detector and a smile classifier have been combined in order to work as a real-time system; the face detector provides the facial cues and the smile classifier detects the smile and determines its intensity. Contrary to Werner et al. proposal [37], our approach does not specify frame pairs for which the first frame has a lower intensity than the second. In our case, frames are computed no matter the intensities they exhibit in order to find the optimal response.

Table 3 provides a comparison among the different methods used for the smile detection in recent years. Two commercial smile recognition systems are addressed;

Table 3. Comparison between different approaches for the smile detection. The symbol "+++" stands for unknown condition due to the fact that we are dealing with a comercial system and the required specifications are not available. Error rates are calculated over the total number of frames. For each video sequence, a smile was detected in at least one frame.

Approach	Classifier	Training Data	Test Data	Error
Optical Flow[17]	Adaboost	BIOID	FGnet	11.5 %
Sony T300[17]	Adaboost	+++	FGnet	27.3 %
3D Face mapping[5]	+++	+++	FGnet	05.9 %
Gabor Filters[38]	SVM	GENKI	GENKI	03.7 %
Pixel Comparisons[29]	Adaboost	GENKI4K	GENKI4K	10.3 %
Pixel's value[39]	SVM	Local Images	Local Images	17.8 %
LIH + CS-LBP[30]	SVM	FEED	FEED	02.1 %
PCA[11]	SVM	BIOID	DaFEx	13.5 %
ULBP Histograms[11]	SVM	BIOID	DaFEx	21.4 %
SLBP[11]	SVM	BIOID	DaFEx	20.7 %
Digital Signature[12]	SVM	BIOID	DaFEx	11.5 %
Proposed Dynamic Approach	SVM	BIOID	SIANI dataset	14.7 %
	SVM	BIOID	DaFEx	10.5 %

the Sony T300 and the smile detection of Omron's technology[5]. On the other hand, Huang et al. make use of the FGnet database ("Facial Expression and Emotion Database") which consists of a set of recorded videos with different emotions through a front camera, very similar to the image recorded on a webcam built into the screen[4]. Huang et al.[17] propose a method of smile detection based on the application of optical flow techniques on the mouth region exclusively. Their theory is based on the intuition that, around the mouth region and in the event of a smile, the optical flow vectors point outward and upward. However, this system, ignores extreme facial rotations or image scaling, for example when the subject is approaching or moving away from the camera. Nonetheless, Huang et al. experiments provide rates around 88.5% accuracy, achieving an improvement of up to 15% over the Sony T300 camera rates and approaching enough to the Omron's smile detection rates. Precisely, this detector developed by Omron[5] uses a 3D mapping technique to determine whether the subject smiles or not.

On the other hand, Whitehill[38] suggested an approach based on the application of Gabor filters all over the face in order to determine whether a person is smiling or not smiling. The rates offered by this author are remarkable and similar to those offered by other authors like Shimada[30]. He proposed a method that combines a new LBP based approach and local intensity histograms. Shimada also combines several SVM creating a cascade detector in order to keep a low computational cost as well as a low error rate. For the experiments, this author used the Facial Expression and Emotion Database (FEED)[14] as well as several images obtained from TV programs. As in Whitehill's research, Shimada considers the entire face for smile detection. Shan[29] analyzes the intensity differences between pixels in the grayscale face images and use the Adaboost to combine intensity differences. Finally, Yadappanavar[39],

made an extensive study on local images. He also combines several SVM classifiers trained by a set of descriptors obtained through the pixel's intensity value. This author ony considers the mouth region.

Once that the Table 3 techniques have been addressed, it is easier to understand the context of the developed research in this paper. The rates provided in this paper show a progressive enhancement of results as the different approaches presented in section 4 were combined. Using a standard and publicly accessible database, DaFEx, a more validity to the results has been provided. On the other hand, our rates can be compared with those obtained in the past by the approach known as Digital Signature. Indeed, the proposed dynamic approach outperforms the rates achieved by the Digital Signature. However, the Digital Signature outperforms both LBP based approaches almost always.

The Digital Signature approach is based on dividing the mouth region in a set of blocks. Each of these block is compared with a central patch of this region and the computed sum of squared differences is stored in a vector, which is known as Digital Signature. This technique cannot be included in the proposed approach due to the fact that it creates a bottleneck. The time needed to compute the Digital Signature is excesive compared to the considered approaches. For instance, the execution time of processing a set of faces by the Digital Signature is almost 50% slower than processing them with any of the three considered techniques for the proposed dynamic approach. Finally, another important advantage of our dynamic approach is that it has collected the inherent strengths of each individual approach such as robustness under lighting conditions changes.

8. Conclusions

There has been a huge development of perceptual user interfaces during the last decade. The affective computing has played a significant role in this development. Indeed, the capability of making peripherals more "friendly" and "adaptable" pretends to introduce a new generation of intelligent devices in the technological market. The main purpose of this new generation of devices is to reduce the "intelligence gap" by facing the problem trough an affective perspective. Nonetheless, a new dimension of communication between humans and machines requires the machines to understand the life like humans do; the symbolism, the people, the culture, etc. Exploring biological alternatives have shown pretty good results in the past. Thus, in this paper authors have explored the human behavior in order to better understand how machines can possibly be aware of our "human signals". Moreover, this paper provides an intensive research on the smile dynamics and proposes a new approach based on these dynamics. Recently, an interesting approach for human expressions recognition considering a set of facial points was presented by Sohail and Bhattacharya[32]. They observed the "discrete features" of these points responsible for the facial expression. Therefore, the proposed approach in this paper provides an interesting framework for considering not only the discrete features at a given

moment, but also to analyze the dynamic of those features in order to improve their results. Furthermore, another possible research line is to combine our temporal coherence approach with a three-dimension (3D) model of the face. Following this line, Lee et al.[23] have not just considered the two-dimensional (2D) facial motions, but also the three-dimensional facial motions which can fit with our proposal.

The experiments proposed in this paper exploit the temporal coherence provided by a video sequence and the life cycle of the face expressions, in order to accurately determine the activation of a user's smile. Furthermore, a face detector has been combined with a smile classifier to work as a real-time system; the detector provides facial cues and the smile classifier detects smile as well as the intensity of this smile. The authors have proved that the results achieved of processing sets of consecutives frames have shown a higher real-time performance than our previous work where a static image based approach was considered.

References

1. T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Machine Intelligence*, 28:2037–2041, 2006.
2. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
3. M. Castrillón-Santana, O. Déniz-Suárez, M. Hernández-Tejera, and C. Guerra-Artal. ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation*, 18(2):130–140, April 2007.
4. Company. Internet source: FGNET, Face and Gesture Recognition Working Group, 2009.
5. Company. Internet source: OKAO Vision, 2009.
6. Company. Internet source: SONY DSC T200, 2012.
7. A. R. Damasio. *Descarte's Error: Emotion, Reason and the Human Brain*. Picador, 1994.
8. O. Déniz, M. Castrillón, J. Lorenzo, L. Antón, and G. Bueno. Smile detection for user interfaces. *Advances in Visual Computing. Lecture Notes in Computer Science*, 5359:602–611, 2008.
9. P. Ekman, R. Davidson, and W. Friesen. *The duchenne smile: emotional expression and brain physiology*. 1990.
10. P. Ekman and W. Friesen. *Manual for the Facial Action Coding System*. Consulting Psychologists Press, 1977.
11. D. Freire-Obregón, M. Castrillón-Santana, and O. Déniz-Suárez. Smile detection using local binary patterns and support vector machines. In *Proceedings of Computer Vision Theory and Applications (VISAPP'09)*, 2009.
12. D. Freire-Obregón, M. Castrillón-Santana, and O. Déniz-Suárez. New techniques applied for facial expression detection. *Journal on Biometrics Systems, Design and Applications*, 1:93–108, 2011.
13. R. W. Frischholz and U. Dieckmann. Bioid: A multimodal biometric identification system. *IEEE Computer*, 33(2), February 2000.
14. W. H. Internet source: Facial expressions and emotion database, 2006.
15. J. Hager. *The inner and outer meanings of facial expression*. Guilford, 1983.

16. J. Hernandez, M. Hoque, and R. Picard. Internet source: MIT Mood Meter, 2012.
17. Y. H. Huang and C. S. Fuh. Face detection and smile detection. *Proceedings of IPPR*, (1), 2009.
18. A. Ito, X. Wang, M. Suzuki, and S. Makino. Smile and laughter recognition using speech processing and face recognition from conversation video. In *Procs. of the 2005 IEEE Int. Conf. on Cyberworlds (CW'05)*, 2005.
19. Y. Kirby and L. Sirovich. Application of the Karhunen-Loéve procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1), July 1990.
20. S. Koelstra, M. Pantic, and I. Patras. A dynamic texture-based approach to recognition of facial actions and their temporal models. In *Proceedings on Pattern Analysis and Machine Intelligence*, volume 32, pages 1940–1954, 2010.
21. I. Kotsia and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Processing, IEEE Transactions on*, 16(1):172–187, Jan 2007.
22. U. Kowalik, T. Aoki, and H. Yasuda. Broaference - a next generation multimedia terminal providing direct feedback on audience's satisfaction level. In *INTERACT*, pages 974–977, 2005.
23. C. Lee and D. Samaras. Analysis and control of facial expressions using decomposable nonlinear generative models. *IJPRAI*, 28(5), 2014.
24. H. Liu and P. Wu. Comparison of methods for smile deceit detection by training au6 and au12 simultaneously. In *Proceedings of ICIP*, number 3, pages 1805–1808, 2012.
25. S. Marcel, Y. Rodriguez, and G. Heusch. On the recent use of local binary patterns for face authentication. *International Journal of Image and Video Preprocessing*, 1:1–9, 2007.
26. T. Ojala, M. Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29:51–59, 1996.
27. T. Ojala, M. Pietikinen, and T. Menp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
28. R. W. Picard. *Affective Computing*. MIT Press, 1997.
29. C. Shan. Smile detection by boosting pixel differences. *Image Processing, IEEE Transactions on*, 21(1):431–436, Jan 2012.
30. K. Shimada, T. Matsukawa, Y. Noguchi, and T. Kurita. Appearance-based smile intensity estimation by cascaded support vector machines. In *Proceedings of the 2010 international conference on Computer vision*, pages 277–286, 2010.
31. Y. Shinohara and N. Otsu. Facial expression recognition using Fisher weight maps. In *Procs. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2004.
32. A. S. M. Sohail and P. Bhattacharya. Classifying facial expressions using level set method based lip contour detection and multi-class support vector machines. *IJPRAI*, 25(6):835–862, 2011.
33. X. Tan. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on Biometrics Compendium*, 19:1635 – 1650, 2010.
34. Q. Tao and R. Veldhuis. Illumination normalization based on simplified local binary patterns for a face verification system. In *Proc. of the Biometrics Symposium*, pages 1–6, 2007.
35. M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.

36. P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):151–173, May 2004.
37. P. Werner, A. Al-Hamadi, and R. Niese. Comparative learning applied to intensity rating of facial expressions of pain. *IJPRAI*, 28(5), 2014.
38. J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Towards practical smile detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(1):2106–2111, 2009.
39. H. Yadappanavar. Machine learning approach for smile detection in real time images. In *International Journal of Image Processing and Vision Sciences (IJIPVS)*, pages 32–37, 2012.
40. L. Zhang and D. Tjondronegoro. Facial expression recognition using facial movement features. *Affective Computing, IEEE Transactions on*, 2(4):219–229, Oct 2011.

Biographical Sketch and Photo

David Freire-Obregón received the M.Sc. and Ph.D. degrees in computer science from Las Palmas de Gran Canaria University, in 2010 and 2014 respectively. His research interests include: computer vision for human interaction, artificial intelligence and machine learning.

Modesto Castrillón-Santana received the M.Sc. and Ph.D. degrees in computer science from Las Palmas de Gran Canaria University, in 1992 and 2003 respectively. His research activities focus particularly on the automatic facial analysis problem, but covering also different topics related to image processing, perceptual interaction, human-machine interaction and computer graphics. Currently, he is an associate professor at ULPGC.