# Geometrical Approaches for Facial Expression Recognition using Support Vector Machines

Jovan de Andrade Fernandes Junior, Leonardo Nogueira Matos, Maria Géssica dos Santos Aragão
Computer Science Department, DCOMP
Federal University of Sergipe,Brazil, UFS
Sergipe, Brasil
Email: jovanoasis@gmail.com, lnmatos@ufs.br, gssica.aragao@gmail.com

***Abstract*—This article presents two facial geometric-based approaches for facial expression recognition using support vector machines. The first method performed an experimental research to identify the relevant geometric features for human point of view and achieved 85% of recognition rate. The second experiment employed the Correlation Feature Selection and achieved 96.11% of recognition rate. All experiments were carried out with Cohn-Kanade database and the results obtained are compatible with the state-of-the-art in this in this research area.**

***Keywords*-Facial Expression Recognition; PDM; CFS ;Correlation Features Selection; Cohn-Kanade Database**

## I. Introduction

Human-machine interaction is a multidisciplinary area that involves engineering, linguistic and psychology, among others. An effective Intelligent Human Computer Interface (IHCI) requires the machine to be able to naturally interact with the user in a way similar to the one that humans do. During an interpersonal communication, information is transmitted through audio, visual and tactile levels. Among them, the visual communication is the richest form of information because involves corporal gestures, facial expressions or a combination of them. Thus, the recognition of facial expressions and a better understand of human emotion is an essential task for a rich and high level machine-user interaction.

Human emotion recognition has been investigated since a seminal book by Charles Darwin ”The expression of the emotions in man and animals”, published in 1872, Ekman [1] has also made great contributions to the area by understanding all displacements of the facial muscles and finding that six basic emotions appear to be universally recognized, even in different cultures. These basic emotions matches with distinct facial expressions, which are: anger, disgust, fear, happiness, sadness and surprise. Furthermore, he built a coding system called Facial Action Coding System - FACS [2], wherewith we can measure the facial movement through Action Units(AU). The description of the six basic emotions plus neutral expression is illustrated in table I.

TABLE I
BASIC FACIAL EXPRESSIONS [1]

| Expression | Description |
|---|---|
| Neutral | All face muscles are relaxed. Eye lids are tangent to there iris. The mouth is closed and lips in contact. |
| Anger | The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are pressed against each other or opened to expose the teeth. |
| Disgust | The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often asymmetrically. |
| Fear | The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are tense and alert. |
| Happiness | The eyebrows are relaxed. The mouth is open and the mouth corners pulled back toward the ears. |
| Sadness | The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is relaxed. |
| Surprise | The eyebrows are raised. The upper eyelids are wide open, he lower relaxed. The jaw is opened. |

As described in [3], the contraction of facial muscles produces changes in appearance and skin color which also can be classified into two types: permanent and transient. Examples of features that are part of the first group are: spatial location of eyes, eyelids, lips and brows. Changes in its geometric shape are taking into account for facial expression recognition. Examples of the second group are the appearance of wrinkles, vertical and horizontal lines in forehead and paranasal sinuses, facial furrows or any other feature that are not present when the face is at rest, but appearing during the manifestation of an expression.

The manifestation of the features in the transient group varies widely among individuals. For example, a person may have facial wrinkles around the face even with facial muscles relaxed, the nasolabial folds may occur even with a neutral face or wrinkles between the eyes in some individuals may not appear. Since transient features does not occur in a similar way in different people, using them to recognize facial expressions may not result in a generic and person independent approach. Thus, this work focuses only on features that belong to the permanent group, placing them as input to calculate facial features distances like: euclidean distance between eyes or euclidean distance between lips and brows, here also called as geometric-based features.

On the past twenty years, a lot has been done to try to recognize human facial expressions [4], [5], but this task is not yet performed by a computer with the same efficiency as the human being. Facial expression recognition systems can also be divided in two main categories: the ones related with static images [6], [7] and those that work with dynamic image sequences [8]. The main difference between them is the coverage of the feature vector, because while the static image-based approaches comprises only information about the current input image, the dynamic image sequence-based methods uses temporal information of images to recognize expressions based on one or more frames. This work will focus on methods based on static images and it will consider the six basic expressions and the neutral expression.

*Contributions:* In this article, we present a method to recognize facial expressions based on two feature extraction procedures: Empirical Normalized Distances and CFS (Correlation Features Selection) [9] Attribute Selection. They are both based on geometric attributes calculated with the PDM's (Point Distribution Model) [10], landmarks positions. A support vector machine using an RBF kernel is used to classify the two kinds of feature vectors. In this paper, all experiments were carried out using the Cohn-Kanade (CK+) database of static images [11].

The remainder of this paper is organized as follows. Section II presents related studies on facial expression recognition. Section III explains the proposed technique and in Section IV the experiments are discussed and the results are presented. In Section V the proposed approaches are compared with some other state-of-the-art methods. Finally, the Section VI concludes this study.

## II. Related work

In recent years, a lot of progress has been made in this research area. A full survey can be found in [5], [12], [13], [14], [4].

Generally, the approaches for recognizing facial expressions vary in how they perform feature extraction and classification. According to Wu *et al.* [12], feature extraction and feature selection methods can be divided into three categories: deformable features extraction methods, motion features extraction methods and statistical feature extraction methods. The first category, which is the object of this work, aims to extract information of facial deformation, such as: geometric deformation (changes in distances between feature points caused by the variety of expression) or texture changes (changes in the local gray level pattern spread during a facial expression manifestation). The second one is mainly used to extract some feature points or feature areas motion information from sequential expression images, such as: the movement distance and direction of feature points. The common methods include: feature point tracking, optical flow approaches and model methods. The last one aims to describes the characteristics of expression images by statistics, such as: histogram or moments. Due to the large diversity of approaches to recognize facial expressions, this section focuses on methods closely related to the one proposed herein.

Saeed *et al.* [15] infer the facial expressions using features based on the location of eight facial points, calculated by a PDM technique, representing the shape and location of three facial components (eye, eyebrow, and mouth). Then, the authors derived six geometrical features and performed two experiments. They achieved $83\%$ accuracy when did not take into account the neutral expression. When adding neutral expression, the recognition rate dropped to $73.6\%$ using an SVM classifier.

Shan *et al.* [7] investigated the impact of image resolution on the accuracy of the result of facial expression recognition and concluded that methods based on geometric features do not handle low resolution very well, while those based on appearance, like Gabor Wavelets and LBP (Local Binary Patterns), that are not so sensible to the image resolution. Furthermore, the authors also performed a deep study that, in the best scenario, they achieved an accuracy rate of $95.1\%$ using SVM and LBP as feature extractor. It is currently the state-of-the-art method on the Cohn-kanade database [11].

Due to the small size of target set, SVM is quite suitable for facial expression recognition. Hsieh *et al.* [16] used Adaboost and ASM (Active Shape Model), [10], to identify human faces and locate facial components and subsequently the authors used Gabor filters and Laplace of Gaussian (LoG) edge detection to propose "semantic facial features". Finally, SVM were used to classify the user facial expressions into one of the six class of expression (excludes sad expression), achieving in average $94,5\%$ recognition rate on the Cohn-Kanade database.

Michel *et al.* [17] defined 22 features points for automatic tracking. The motions of all the feature points from neutral to peak expression were measured as a feature vector. Chen *et al.* [18] used feature points displacements and local texture differences between the normalized neutral and expressive face images for recognition. The combined feature vector contains a $42$ dimensional geometric feature vector and a 21 dimensional texture feature vector. The average accuracy was $95\%$ using an SVM on the Cohn-Kanade database (excluding "Contempt" from database).

The system proposed in [19] achieved $99.7\%$ recognition rate with the drawback that requires the Candide grid to be manually placed upon the facial area and moreover it requires the manual detection of the neutral state in a video sequence, requiring the whole video of facial expression development from the neutral state to the fully expressive image.

## III. The Geometrical Approach

In this section, a geometrical based approach for facial expressions recognition is described. An overview of the proposed method is illustrated in Fig. 1.

The facial expression recognition system is composed of three main phases: PDM tracker, features extraction and selection and feature classification. In PDM tracker phase, the system tracks the defined landmarks. In the next step, the system extracts the Euclidean distances of all points and then selects the most relevant features. Lastly, the system classifies the expression with a basic emotion label using a support vector machine with a RBF kernel. The remaining of this section explains each phase on details.

### A. Point Distribution Model

As a non-rigid object, the human face requires a robust method to deal with several problems, such as: translation,

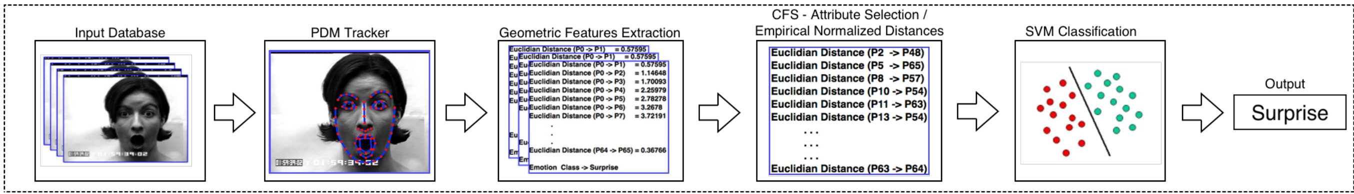



Fig. 1. Overview of the proposed method.

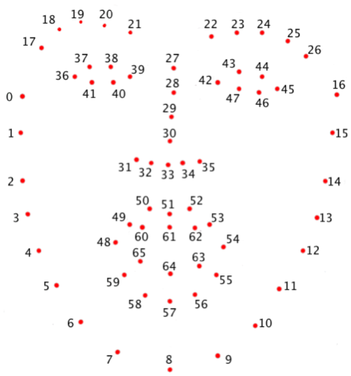


Fig. 2. Set of the points used to represent facial features.

rotation, scale, orientation (pose), among others. Shape and appearance changes of the facial components impose additional difficulties to the usual image conditions like lightness and shadow. Furthermore, we need to deal with partial occlusion of the face on the presence of non-structural components such as beard, mustache, hat, glasses and others. The Point Distribution Model (PDM) tries to model a deformable shape of an object. In this approach, the shape is described with a set of $n$ points or landmarks, obtained through a database containing examples of the different shapes that the object can be observed, ordinarily manually marked. These set of points can represent the internal and external features of an object, for example, it can represent the facial contours and internal elements of a human face like eye, mouth, nose, lips, eyebrows. In this paper, it was used a set of 66 $landmarks$ that are enough to describe the movement of all the facial permanent features, suggested from MUCT Database [20]. The arrangement of the used points can be seen in Fig. 2.

As purposed by Cootes et al.[10], the statistical properties of the shape are observed to study its the variation modes and to apply global restrictions in order to prevent the generation of invalid forms. They proposed the Active Shape Models - ASM, one type of PDM, which initially aligns the shapes described by landmarks on a database by applying the Procrustes alignment algorithm [21] and then builds a linear shape model of the object by applying a Principal Component Analise - PCA. This means that the shape $s$ can be expressed as a base shape $s_0$ plus a linear combination of $m$ shape variations $s_i$, illustrated in equation (1), where the coefficients $p = (p_1, ..., p_m)^T$ are the shape parameters. Furthermore, the authors also proposed to extract each landmark pattern, in this case based on the gray levels, by searching the neighborhood of the current landmark position estimation before appling the global shape restrictions obtained through PCA.

$$s = s_0 + \sum_{i=1}^{m} p_i s_i \qquad (1)$$

Over the past decade, several different PDM based approaches were proposed to improve the model fit to a target image like Active Appearance Models - AAM [22] and the Constrained Local Models - CLM [23].

*Active Appearance Models - AAM:* In addition to the linear shape model, the AAM approach builds an $2D$ triangulated mesh appearance model within the base mesh $s_0$. Let $s_0$ also denote the set of pixels $u = (u, v)^T$ that lie inside the base mesh $s_0$. The appearance of the AAM is then an image $A(u)$ defined over the pixels $u \in s_0$. Therefore, the appearance $A(u)$ can be expressed as a base appearance $A_0(u)$ plus a linear combination of $l$ appearance images $A_i(u)$, as expressed in equation (2), where the coefficients $\lambda_i$ are the appearance parameters. The base (mean) appearance $A_0$ and appearance images $A_i$ are usually computed by applying PCA to the (shape normalized) training images.

$$A(u) = A_0(u) + \sum_{i=1}^{l} \lambda_i A_i(u) \qquad (2)$$

Thus, to generate a model instance with shape parameters $p$ and appearance parameters $\lambda_i$ we warp the appearance $A$ from the base mesh $s_0$ onto the model shape mesh $s$. In particular, the pair of meshes $s0$ and $s$ defines a piecewise affine warp from $s_0$ to $s$ denoted $W(u; p)$. More details can be seen in [22].

*Constrained Local Model - CLM:* The CLM is quite similar to the AAM approach. The only difference between them is that instead of considering all textures over the whole face, the CLM approach consider the texture information around each landmark. In this case, the approach uses an ensemble of local detectors to determine $s$. All of these methods have the following two goals: the first one is to perform an exhaustive local search for each PDM landmark around their current estimate using some kind of feature detector; the second goal is to optimize the PDM parameters in a way that the landmarks detection responses are jointly maximized. More details can be seen in [23].

### B. Feature Extraction and Selection

With the shape of the face described in a set of points, that represents the facial geometry, we have to select the features that maximize the decision boundary between classes when the distances between landmarks changes, by identifying important attributes which enhance the performance of the classifier and eliminates irrelevant attributes.

The feature selection techniques can be classified in different ways, one of which relates to their tie with induction algorithm that ultimately learns from the reduced data. To evaluate the quality and the performance of a subset of attributes, three approaches are commonly used: *Built-in*, *Filter-based* and *Wrapper-based*.

In the first case, the selection of the subset is built-in or integrated with induction algorithms, like decision trees. The filter based approach operates independently of any induction algorithm, unnecessary attributes are filtered out of the data before induction takes place. The techniques on this group, for example, verifies the correlation between the attributes and are usually faster than the latter group. The wrapper based approach argues that the bias of a particular induction algorithm should be taken into account when selecting features. Thus, for each possible subset, the induction algorithm is consulted and the subset that has the highest reduction of the error rate is selected in general.

In this work, to achieve the better subset of features, firstly all Euclidean Distances $D$ between the landmarks that describes the shape of the human face (illustrated in Fig. 2) have been calculated through equation (3). Since we have 66 landmarks, we found $\binom{66}{2} = 2145$ distances. To reduce the dependence of different face sizes, scale and translation variations, every point $P_i$ is normalized with equation (4) where $Dn$ is a normalization coefficient calculate through the Euclidean Distance $D$ between the left corner of the right eye and the right corner of the left eye, Eq. (5). This distance was selected because it remains unchanged during deformation of the facial muscles.

$$D(P_\imath, P_\jmath) = \|P_{\imath_{normalized}} - P_{\jmath_{normalized}}\|_2 \tag{3}$$

$$P_{\imath_{normalized}} = P_\imath / Dn \tag{4}$$

$$Dn = D(P_{42}, P_{39}) \tag{5}$$

In this work, two features selection approaches were used. The first was based on a human point of view obtained through an empirical experiment and reveals promising results. The second uses the Correlation Features Selection - CFS. Both belongs to the group of filter-based approaches. These two proposed features selection methods are described in details in section IV.

### C. Feature Classification

We used a Support Vector Machine, a well-know linear binary classifier, and formulated the facial expression recognition task as a multiclass learning process, where one class was assigned to each expression. Since we have a linear binary classification task with training data $x_i (i = 1, ..., N)$, having corresponding classes $y_i = \pm 1$, the decision function can be formulated as equation (6), where $\mathbf{w}^T\mathbf{x} + b = 0$ denotes a separating hyperplane, $b$ is the bias and $\mathbf{w}$, called weight vector, is ortogonal to the separating hyperplane. The values $(\mathbf{w}, b)$ can be obtained by solving a constrained optimization problem, where (7) is minimized subject to the constraints in (8).

$$g(\mathbf{x}) = sign(\mathbf{w}^T\mathbf{x} + b) \tag{6}$$

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{7}$$

$$y_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1, \qquad \forall_i \tag{8}$$

Originally SVM is formulated as a linear classifier. In order to deal with non linear problems, a kernel function must be applied to obtain an effective classification. Here, we employed LIBSVM [24] with RBF (Radial Basis Function) kernel.

## IV. Experiments

This study employed the Extended Cohn-kanade (Ck+) database [11] to verify the facial expression classification performance. This database is one of the most comprehensive in the current facial expression research community, facilitating the comparison of this work with others. It comprises 327 image sequences of university students (18-30 years old) including the universal emotions described in table I and the contempt expression (disregarded in this work). Each sequence begins with a neutral expression and ends with the target facial expression. In this work, in every image sequence, to enhance the amount of samples of each emotion, the first (when considering the neutral expression) and the last 3 images of the sequences were selected to train and test the SVM classifier in a 10-fold cross-validation approach. Furthermore, in this work, we consider both 6-class expression recognition and 7-class expression recognition by including the neutral expression.

As presented by some authors [17], [7], [25], geometric features-based methods provide similar or better performance than appearance-based approaches in facial expression recognition. However, the geometric feature-based methods usually requires accurate and reliable facial feature detection and tracking, directly impacting on recognition performance. Therefore, in this work, all experiments were carried out using two PDMs. The first is an AAM provided with the database, which is expected to have a better performance than others PDMs because is person-dependent, that is, the PDM model that was trained specifically for the Cohn-kanade database samples. The last one is a Constrained Local Model - CLM

[23], a type of PDM, trained using annotated data from the MUCT Database [20] with multiple identities, expressions, lighting conditions and other sources of variability. Thus, the results obtained through this more generic or person-independent face alignment is expected to be less accurate than the first since it is not specialized on the database. The implementation of the CLM used in this article was the proposed by Saragih [23].

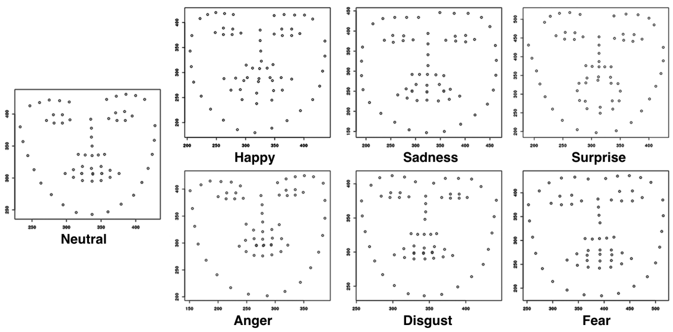



Fig. 3. Examples of images used in an experimental research that aims to identify the human perception of important landmarks for facial expression classification.

*Empirical Normalized Distances Approach:* To identify which distances should be used in the feature classification step, initially, an experimental research was performed asking a group of ten people to look at the Cohn-Kanade database images containing only the data of the landmarks corresponding to people images expressing a basic emotion. Examples of images used in this research can be seen in Fig. 3. Empirically, when participants hit the expression shown, it was asked to inform which characteristics were taken into consideration to arrive at their conclusions. The purpose of this study was to identify, from a human point of view, which landmarks or distances between them are taken into consideration to classify facial expressions.

Based on the experiment described above and in the method proposed by Soyel [26], we propose seven geometric-features that are described in table II. Here, every obtained distance is also normalized with the equation (3). In this approach, to reduce measurement problems, when possible, we take into account the symmetrical properties of the human face by adding the equivalent distance and then we consider the average value of the distances.

Using a person-dependent model of a PDM to the 6-class problem, this approach obtains 99.67% of recognition rate. However, when the quality of the feature detection fall down while using a person-independent PDM, the recognition rate decays to 89.75%. The obtained confusion matrix using the empirical normalized distances to the 6-class recognition is shown in table III.

Despite the high recognition rate obtained in 6-class problem, on the 7-class problem the recognition rate with the person-dependent facial alignment decays to 96.27%. Hence, the recognition rate using a person-independent model decays to 85.03%. The obtained confusion matrix using the empirical normalized distances adding the neutral expression is presented in table IV, which shows that the recognition of the additional expression causes a strong impact on the recognition rate especially in anger, fear and sadness expressions, suffering considerable confusion using this approach.

Although the adopted approach of using the first (in 7-class experiment) and the three latest images of the sequences provided by Cohn-Kanade [11] to increase the number of samples,

TABLE II
THE SELECTED EMPIRICAL NORMALIZED DISTANCES USED IN THE FIRST EXPERIMENT.

| **Distance Name** | **Description** |
|---|---|
| Eye Opening | $\frac{Dreal(P_{37},P_{41})+Dreal(P_{38},P_{40})}{2}$ |
| Eyebrow Height | $\frac{Dreal(P_{20},P_{38})+Dreal(P_{18},P_{37})}{2}$ |
| Eyebrow Distance | $Dreal(P_{21}, P_{22})$ |
| Mouth Height | $\frac{Dreal(P_{60},P_{65})+Dreal(P_{61},P_{64})+Dreal(P_{62},P_{63})}{3}$ |
| Mouth Width | $Dreal(P_{54}, P_{48})$ |
| Chin Height | $Dreal(P_{30}, P_{8})$ |
| Lip Stretching | $Dreal(P_{36}, P_{48})$ |

TABLE III
CONFUSION MATRIX USING THE EMPIRICAL DISTANCES APPROACH. FOR EACH EXPRESSION, THE FIRST LINE PRESENTS THE RESULTS OBTAINED WITH A PERSON-DEPENDENT PDM AND THE SECOND ONE WAS OBTAINED WITH A PERSON-INDEPENDENT PDM.

| In/Out | Anger | Disgust | Fear | Happy | Sadness | Surprise |
|---|---|---|---|---|---|---|
| Anger | **100%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| | **85.19%** | 9.63% | 0.00% | 2.22% | 0.74% | 2.22% |
| Disgust | 0.00% | **100%** | 0.00% | 0.00% | 0.00% | 0.00% |
| | 3.39% | **89.27%** | 0.00% | 6.21% | 0.56% | 0.56% |
| Fear | 0.00% | 0.00% | **96.00%** | 0.00% | 0.00% | 4.00% |
| | 1.33% | 0% | **88.00%** | 6.67% | 4.00% | 0.00% |
| Happy | 0.00% | 0.00% | 0.00% | **100%** | 0.00% | 0.00% |
| | 2.42% | 4.83% | 1.93% | **87.92%** | 0.00% | 2.90% |
| Sadness | 0.00% | 0.00% | 0.00% | 0.00% | **100%** | 0.00% |
| | 2.38% | 1.19% | 5.95% | 0.00% | **85.71%** | 4.76% |
| Surprise | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | **100%** |
| | 1.61% | 0.00% | 0.80% | 0.80% | 0.80% | **95.98%** |

TABLE IV
CONFUSION MATRIX ADDING NEUTRAL EXPRESSION RECOGNITION TO THE THE EMPIRICAL DISTANCES APPROACH. FOR EACH EXPRESSION, THE FIRST LINE PRESENTS THE RESULTS OBTAINED WITH A PERSON-DEPENDENT PDM AND THE SECOND ONE WAS OBTAINED WITH A PERSON-INDEPENDENT PDM.

| | Neutral | Anger | Disgust | Fear | Happy | Sad. | Surp. |
|---|---|---|---|---|---|---|---|
| Neut. | **94.82%** | 0.97% | 1.29% | 0.00% | 0.00% | 1.62% | 1.29% |
| | **84.79%** | 2.91% | 2.59% | 1.29% | 0.97% | 2.91% | 4.53% |
| Ang. | 1.48% | **98.52%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| | 25.93% | **71.85%** | 1.48% | 0.00% | 0.00% | 0.00% | 0.74% |
| Disg. | 4.52% | 0.00% | **95.48%** | 0.00% | 0.00% | 0.00% | 0.00% |
| | 5.08% | 2.82% | **89.27%** | 0.00% | 2.26% | 0.00% | 0.56% |
| Fear | 4.00% | 0.00% | 0.00% | **96.00%** | 0.00% | 4.16% | 0.00% |
| | 24.00% | 0.00% | 0.00% | **73.33%** | 1.33% | 1.33% | 0.00% |
| Hap. | 0.00% | 0.00% | 0.00% | 0.00% | **100%** | 0.00% | 0.00% |
| | 3.86% | 0.00% | 1.93% | 1.45% | **92.75%** | 0.00% | 0.00% |
| Sad. | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | **100%** | 0.00% |
| | 34.72% | 2.38% | 1.19% | 0.00% | 0% | **59.52%** | 2.38% |
| Surp. | 6.83% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | **93.17%** |
| | 4.42% | 0.40% | 0.00% | 0.00% | 0.00% | 0.00% | **95.18%** |

TABLE V

RECALL, PRECISION AND F-MEASURE FOR 6-CLASS EMPIRICAL DISTANCES APPROACH. FIRST LINE OF EACH EXPRESSION PRESENTS RESULTS WITH PERSON-DEPENDENT FACIAL ALIGNMENT AND THE SECOND ONE WITH THE PERSON-INDEPENDENT APPROACH.

| **Expression** | **Recall** | **Precision** | **F-Measure** |
|---|---|---|---|
| Anger | 100%<br>85.18% | 100%<br>86.46% | 1.00<br>0.8582 |
| Disgust | 100%<br>89.26% | 100%<br>86.81% | 1.00<br>0.8802 |
| Fear | 96.00%<br>88.00% | 100%<br>85.71% | 0.9795<br>0.8684 |
| Happy | 100%<br>87.92% | 100%<br>89.65% | 1.00<br>0.8878 |
| Sadness | 100%<br>85.71% | 100%<br>91.13% | 1.00<br>0.8834 |
| Surprise | 100%<br>95.98% | 98.80%<br>94.46% | 0.9940<br>0.9521 |

TABLE VI

RECALL, PRECISION AND F-MEASURE FOR 7-CLASS EMPIRICAL DISTANCES APPROACH. FIRST LINE OF EACH EXPRESSION PRESENTS RESULTS WITH PERSON-DEPENDENT FACIAL ALIGNMENT AND THE SECOND ONE WITH PERSON-INDEPENDENT APPROACH.

| **Expression** | **Recall** | **Precision** | **F-Measure** |
|---|---|---|---|
| Neutral | 94.82%<br>84.78 | 90.71%<br>70.43% | 0.9272<br>0.7694 |
| Anger | 98.51%<br>71.85% | 97.79%<br>85.08% | 0.9815<br>0.7791 |
| Disgust | 95.48%<br>89.26% | 97.68%<br>91.32% | 0.9657<br>0.9028 |
| Fear | 96.00%<br>73.33% | 100%<br>88.70% | 0.9795<br>0.8029 |
| Happy | 100%<br>92.75% | 100%<br>96.00% | 1.00<br>0.9434 |
| Sadness | 100%<br>59.52% | 94.28%<br>83.33% | 0.9710<br>0.6944 |
| Surprise | 93.17%<br>95.18% | 98.30%<br>92.94% | 0.9567<br>0.9404 |

the number of instances of each class was unbalanced, favoring neutral expression, which biases the classifier and affects its performance for the classes with lower recall rate as sadness, anger and fear. Therefore, for a better analysis of the results, in tables V and VI are illustrated the recall, precision and f-measure rates for the 6-class and the 7-class experiments respectively.

*CFS Attribute Selection Approach:* Despite the good results of the previous experiment using a person-dependent model, the recognition rate for the generic model drops significantly. Thus, we tried to apply a formalized feature selection approach.

Like the majority of feature selection approaches, Correlation Features Selection - CFS uses a search algorithm along with a function to evaluate the merit of the feature subset. The heuristic by which CFS measures the goodness of feature subset takes into account the usefulness of individual features for predicting the class label along with the level of intercorrelation among them. The hypothesis on which the heuristic is based can be stated: *A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other* [9].

TABLE VII

CONFUSION MATRIX OBTAINED USING THE FEATURES SELECTED BY CORRELATION FEATURES SELECTION. FOR EACH EXPRESSION, THE FIRST LINE PRESENTS RESULTS OBTAINED WITH A PERSON-DEPENDENT PDM AND THE SECOND ONE WAS OBTAINED WITH A PERSON-INDEPENDENT PDM.

| In/Out | Anger | Disgust | Fear | Happy | Sadness | Surprise |
|---|---|---|---|---|---|---|
| Anger | **100%**<br>**93.33%** | 0.00%<br>2.22% | 0.00%<br>0.00% | 0.00%<br>0.74% | 0.00%<br>0.00% | 0.00%<br>3.70% |
| Disgust | 0.00%<br>1.13% | **100%**<br>**89.27%** | 0.00%<br>0.00% | 0.00%<br>7.91% | 0.00%<br>0.00% | 0.00%<br>1.69% |
| Fear | 0.00%<br>2.67% | 0.00%<br>0.00% | **96.00%**<br>**94.67%** | 0.00%<br>0% | 0.00%<br>0.00% | 4.00%<br>2.67% |
| Happy | 0.00%<br>0.48% | 0.00%<br>3.86% | 0.00%<br>0.48% | **100%**<br>**93.72%** | 0.00%<br>0.48% | 0.00%<br>0.97% |
| Sadness | 0.00%<br>0.00% | 0.00%<br>0.00% | 0.00%<br>1.19% | 0.00%<br>1.19% | **100%**<br>**96.43%** | 0.00%<br>1.19% |
| Surprise | 0.00%<br>0.80% | 0.00%<br>0.40% | 0.00%<br>0.40% | 0.00%<br>0.40% | 0.00%<br>0.80% | **100%**<br>**97.19%** |

The equation (9) formalizes this heuristic.

$$G_s = \frac{k\bar{r}_{ci}}{\sqrt{k + k(k-1)\bar{r}_{ii}}} \tag{9}$$

Equation (9) is, in fact, Pearson's correlation coefficient, where all variables have been standardized. The numerator can be thought of as giving an indication of how predictive of the class a group of features are; the denominator of how much redundancy there is among them. The heuristic goodness measure should filter out irrelevant features as they will be poor predictors of the class. Redundant features should be ignored as they will be highly correlated with one or more of the other features.

After the extraction of the $2145$ Euclidean distances, the CFS algorithm was run and selected $44$ features which achieves a better performance than the previous experiment. In fact, using a person-dependent model of a PDM this approach also obtained 99.67% of recognition rate. However, when the quality of the feature detection fall down while using a person-independent PDM, the recognition rate decay to 94.06%, improving the results of the previous method. The confusion matrix using the CFS distances is described in table VII.

In addition to detecting the six basic expressions, this method also recognize the neutral expression. Nonetheless, the recognition rate obtained with the person-dependent decays to 97.97%. Hence, the recognition rate using a person-independent model decays to 96.11%. The obtained confusion matrix using the CFS distances adding the neutral expression is shown in table VIII.

As explained in the previous experiment, the tables IX and X illustrate the recall, precision and f-measure rates for the 6-class and the 7-class experiments for CFS attribute selection distances respectively.

## V. COMPARISON AND DISCUSSION

Although the CFS Attribute Selection method achieved better results than the Empirical Normalized Distances approach, the second one has the advantage that the considered distances are already defined whereas the other has to evaluate the

TABLE VIII
CONFUSION MATRIX ADDING NEUTRAL EXPRESSION RECOGNITION TO THE FEATURES SELECTED BY CORRELATION FEATURES SELECTION APPROACH. FOR EACH EXPRESSION, THE FIRST LINE PRESENTS THE RESULTS OBTAINED WITH A PERSON-DEPENDENT PDM AND THE SECOND ONE WAS OBTAINED WITH A PERSON-INDEPENDENT PDM.

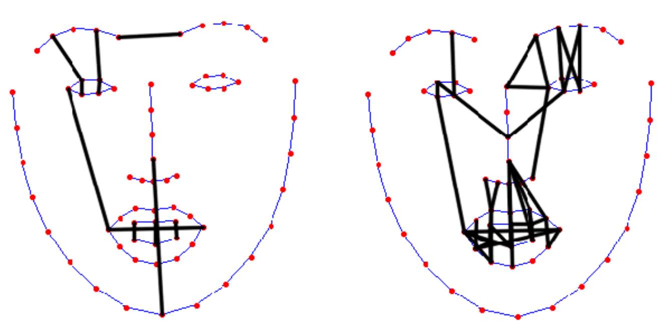
| | Neutral | Anger | Disgust | Fear | Happy | Sadness | Surprise |
|---|---|---|---|---|---|---|---|
| Neut. | **97.09%** | 0.32% | 0.97% | 0.00% | 0.00% | 0.32% | 1.29% |
| | **92.23%** | 0.32% | 0.97% | 0.65% | 0.00% | 1.29% | 4.53% |
| Ang. | 0.00% | **100%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| | 0.00% | **100%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Disg. | 1.69% | 0.00% | **98.31%** | 0.00% | 0.00% | 0.00% | 0.00% |
| | 0.56% | 0.56% | **98.31%** | 0.00% | 0.00% | 0.00% | 0.56% |
| Fear | 4.00% | 0.00% | 0.00% | **96.00%** | 0.00% | 0.00% | 0.00% |
| | 5.33% | 0.00% | 0.00% | **90.67%** | 0.00% | 0.00% | 4.00% |
| Hap. | 0.00% | 0.00% | 0.00% | 0.00% | **100%** | 0.00% | 0.00% |
| | 0.00% | 0.00% | 0.00% | 0.00% | **98.55%** | 0% | 1.45% |
| Sad. | 2.38% | 0.00% | 0.00% | 0.00% | 0.00% | **97.62%** | 0.00% |
| | 3.57% | 0.00% | 0.00% | 0.00% | 0.00% | **94.05%** | 2.38% |
| Surp. | 3.21% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | **96.79%** |
| | 2.41% | 0.0% | 0.00% | 0.00% | 0.00% | 0.00% | **97.59%** |

TABLE IX
RECALL, PRECISION AND F-MEASURE FOR 6-CLASS CFS SELECTION ATTRIBUTES APPROACH. FIRST LINE OF EACH EXPRESSION PRESENTS RESULTS WITH PERSON-DEPENDENT FACIAL ALIGNMENT AND SECOND ONE WITH PERSON-INDEPENDENT APPROACH.

| Expression | Recall | Precision | F-Measure |
|---|---|---|---|
| Anger | 100%<br>93.33% | 100%<br>94.73% | 1.00<br>0.9402 |
| Disgust | 100%<br>89.26% | 100%<br>92.94% | 1.00<br>0.9106 |
| Fear | 96.00%<br>94.66% | 100%<br>95.94% | 0.9795<br>0.9530 |
| Happy | 100%<br>93.71% | 100%<br>91.94% | 1.00<br>0.9282 |
| Sadness | 100%<br>96.42% | 100%<br>96.42% | 1.00<br>0.9642 |
| Surprise | 100%<br>97.18% | 98.8%<br>94.90% | 0.9940<br>0.9603 |

database to obtain the feature vector. The Fig. 4 illustrates the distances considered by the Empirical Normalized Distances approach ((a)) and the CFS Attribute Selection approach ((b)) applied on the Cohn-Kanade database.

The Cohn-Kanade facial expression database for facial expression recognition has attracted the attention of numerous research groups. Some recent work using geometric-based



(a) Empirical Normalized Distances (b) CFS distances

Fig. 4. Considered distances in the two proposed approaches.

TABLE X
RECALL, PRECISION AND F-MEASURE FOR 7-CLASS CFS SELECTION ATTRIBUTES APPROACH. FIRST LINE OF EACH EXPRESSION PRESENTS RESULTS WITH PERSON-DEPENDENT FACIAL ALIGNMENT AND THE SECOND ONE WITH PERSON-INDEPENDENT APPROACH.

| Expression | Recall | Precision | F-Measure |
|---|---|---|---|
| Neutral | 97.08%<br>92.23 | 94.93%<br>95.31% | 0.9600<br>0.9375 |
| Anger | 100%<br>100% | 99.26%<br>98.54% | 0.9963<br>0.9926 |
| Disgust | 98.30%<br>98.30% | 98.30%<br>98.30% | 0.9830<br>0.9030 |
| Fear | 96.00%<br>90.66% | 100%<br>97.14% | 0.9795<br>0.9379 |
| Happy | 100%<br>98.55% | 100%<br>100% | 1.00<br>0.9927 |
| Sadness | 97.61%<br>94.04% | 98.79%<br>95.18% | 0.982<br>0.9461 |
| Surprise | 96.78%<br>97.50% | 98.36%<br>91.35% | 0.9757<br>0.9436 |

features, static images and SVM are summarized in Table XI. As we can observe, the recognition rate achieved by the proposed methods matches the best accuracy founded. Here, we list of some investigative papers closely related to the one introduced in this article.

TABLE XI
PERFORMANCE COMPARISON OF THE PROPOSED CFS SELECTED ATTRIBUTES APPROACH USING A PERSON-INDEPENDENT PDM MODEL WITH THE STATE-OF-THE-ART METHODS THAT RECOGNIZE BASIC EMOTIONAL EXPRESSIONS FROM STATIC IMAGES USING COHN-KANADE DATABASE AND GEOMETRIC FEATURES.

| Method | Year | N-Class | Accuracy |
|---|---|---|---|
| Hsieh et al. [16] | 2015 | *6-class | 94.7% |
| Hsu et al. [27] | 2014 | 7-class | 89.6% |
| Lajevardi and Hussain [28] | 2012 | 7-class | 91.9% |
| Xiao et al. [29] | 2011 | 6-class | 96.57% |
| Zhang et al. [30] | 2011 | 6-class | 94.48% |
| ***Kotsia et al. [19] | 2007 | 6-class | 99.7 |
| Chen et al. [18] | 2012 | **7-class | 95% |
| Tsai et al. [31] | 2010 | 7-class | 98.59% |
| Barlett et al. [32] | 2005 | 7-class | 90.9% |
| Saeed et al. [15] | 2014<br>2014 | 6-class<br>7-class | 83.01%<br>73.06% |
| Shan et al. [7] | 2009 | 7-class<br>6-class | 91.40%<br>95.10% |
| Proposed Empirical Distances Method | 2016 | 7-class<br>6-class | 85.03%<br>89.75% |
| Proposed CFS Attribute Selection Method | 2016 | 7-class<br>6-class | 96.11%<br>94.60% |

*The Sadness expression was not considered.
**The Contempt expression was considered and Neutral was disregarded.
***Method is not fully automatic, requiring manual labeling.

Despite our best results were obtained using the person-dependent facial alignment, in this section, every comparison with another works were made when testing it with the person-independent approach, because it is more suitable for real world cases. We achieved $96.11\%$ accuracy rate using $44$ attributes selected from the CFS feature selection and SVM with RBK kernel form feature classification, and $85\%$ of recognition rate using empirical geometric distances using also an SVM classifier. So far, the system in [19] has shown superior performance, and has achieved a 99.7% recognition rate. In their method, the landmark initialization was a manual process, and the number of landmarks were also larger than the number the one in the proposed method. But, on the other hand, the proposed method is fully automatic. To improve the performance of their system, Hsieh et al. [16] have included some appearances features and they did not consider the Sadness expression. Chen *et al.* [18] also achieved good results but the neutral expression was not considered. For other listed works excepted Tsai *et al.* [31], Xiao *et al.* [29] and Shan et al. [7] our CFS-based method outperforms, demonstrating the contribution of this work.

An interesting fact is that the CFS Attribute Selection method achieves greater results on 7-class problem than in 6-class problem for person- independent face alignment. This may have occurred because the CFS selection was applied considering the neutral expression which is not considered in this case. Thus, another subset of features could be more appropriate for the 6-class problem.

## VI. Conclusion

In this paper, we introduced two geometric-based techniques for facial expression recognition. The results on the Cohn-kanade database showed that both produced relevant scores, which are comparable with the current state-of-the-art of the research area.

In future work, others machine learning algorithms could be evaluated and we are particularly confident on the adequacy of CNN (Convolutional Neural Network) in this area. We are also motivated to perform further experiments using different databases and verify the performance of the method in a wider scenario.

## References


[1] P. Ekman, "Basic emotions," in *The Handbook of Cognition and Emotion*, T. Dalgleish and T. Power, Eds., 1999.

[2] P. Ekman and W. Friesen, *Facial Action Coding System (FACS): Manual.*, Consulting Psychologists Press, Palo Alto, CA, USA, 1978.

[3] Y. li Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 97–115, 1999.

[4] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, 2009.

[5] A. Danelakis, T. Theoharis, and I. Pratikakis, "A survey on facial expression recognition in 3d video sequences," *Multimedia Tools and Applications*, vol. 74, no. 15, pp. 5577–5615, 2015.

[6] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1805–1812.

[7] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, 2009.

[8] Y.-H. Byeon and K.-C. Kwak, "Facial expression recognition using 3d convolutional neural network," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 12, 2014.

[9] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29 - July 2, 2000*, 2000, pp. 359–366.

[10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Comput. Vis. Image Underst.*, pp. 38–59, 1995.

[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[12] T. Wu, S. Fu, and G. Yang, *Advances in Brain Inspired Cognitive Systems: 5th International Conference, BICS 2012, Shenyang, China, July 11-14, 2012. Proceedings*. Springer Berlin Heidelberg, 2012, ch. Survey of the Facial Expression Recognition Research.

[13] C. Sumathi, T. Santhanam, and M. Mahadevi, "Automatic facial expression analysis a survey," *International Journal of Computer Science and Engineering Survey*, 2012.

[14] V. Bettadapura, "Face expression recognition and analysis: The state of the art," *arXiv: Tech Report*, pp. 1–27, 2012.

[15] A. Saeed, A. Al-Hamadi, R. Niese, and M. Elzobi, "Frame-based facial expression recognition using geometrical features," *Adv. in Hum.-Comp. Int.*, 2014.

[16] C.-C. Hsieh, M.-H. Hsih, M.-K. Jiang, Y.-M. Cheng, and E.-H. Liang, "Effective semantic features for facial expressions recognition using svm," *Multimedia Tools and Applications*, pp. 1–20, 2015.

[17] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," in *Proceedings of the 5th International Conference on Multimodal Interfaces*. ACM, 2003.

[18] J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, and L. Wang, "Facial expression recognition using geometric and appearance features," in *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service*, ser. ICIMCS '12. ACM, 2012.

[19] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, 2007.

[20] S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT Landmarked Face Database," *Pattern Recognition Association of South Africa*, 2010.

[21] G. B. Dijksterhuis and J. C. Gower, "The interpretation of generalized procrustes analysis and allied methods," *Food Quality and Preference*, pp. 67–87, 1991/2.

[22] I. Matthews and S. Baker, "Active appearance models revisited," *Int. J. Comput. Vision*, 2004.

[23] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *Int. J. Comput. Vision*, 2011.

[24] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[25] M. F. Valstar and M. Pantic, "Combined support vector machines and hidden markov models for modeling facial action temporal dynamics," in *Proceedings of the 2007 IEEE International Conference on Human-computer Interaction*, ser. HCI'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 118–127.

[26] H. Soyel and H. Demirel, *Image Analysis and Recognition: 4th International Conference, ICIAR 2007, Montreal, Canada, August 22-24, 2007. Proceedings*. Springer Berlin Heidelberg, 2007, ch. Facial Expression Recognition Using 3D Facial Feature Distances, pp. 831–838.

[27] F.-S. Hsu, W.-Y. Lin, and T.-W. Tsai, "Facial expression recognition using bag of distances," *Multimedia Tools and Applications*, 2014.

[28] S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: feature extraction and selection," *Signal, Image and Video Processing*, 2012.

[29] R. X. Q. Z. D. Z. P. Shi, "Facial expression recognition on multiple manifolds," *Pattern Recognition*, 2011.

[30] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," *IEEE Transactions on Affective Computing*, 2011.

[31] H. H. Tsai, Y. S. Lai, and A. Y. C. Zhang, "Using svm to design facial expression recognition for shape and texture features," in *2010 International Conference on Machine Learning and Cybernetics*, 2010.

[32] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.