

Article

A Multi-Layer Fusion-Based Facial Expression Recognition Approach with Optimal Weighted AUs

Xibin Jia ^{1,*}, Shuangqiao Liu ¹, David Powers ^{1,2} and Barry Cardiff ³

¹ Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing University of Technology, Beijing 100124, China; liushuangqiao@emails.bjut.edu.cn (S.L.); david.powers@flinders.edu.au (D.P.)

² School of Computer Science, Engineering and Mathematics, Flinders University of South Australia, Adelaide 5001, Australia

³ School of Electrical, electronic & communications Engineering, University College Dublin, Dublin 4, Ireland; barry.Cardiff@ucd.ie

* Correspondence: jiaxibin@bjut.edu.cn; Tel.: +86-10-6739-2370 (ext. 607)

Academic Editor: Antonio Fernández-Caballero

Received: 29 November 2016; Accepted: 17 January 2017; Published: 24 January 2017

Abstract: Affective computing is an increasingly important outgrowth of Artificial Intelligence, which is intended to deal with rich and subjective human communication. In view of the complexity of affective expression, discriminative feature extraction and corresponding high-performance classifier selection are still a big challenge. Specific features/classifiers display different performance in different datasets. There has currently been no consensus in the literature that any expression feature or classifier is always good in all cases. Although the recently updated deep learning algorithm, which uses learning deep feature instead of manual construction, appears in the expression recognition research, the limitation of training samples is still an obstacle of practical application. In this paper, we aim to find an effective solution based on a fusion and association learning strategy with typical manual features and classifiers. Taking these typical features and classifiers in facial expression area as a basis, we fully analyse their fusion performance. Meanwhile, to emphasize the major attributions of affective computing, we select facial expression relative Action Units (AUs) as basic components. In addition, we employ association rules to mine the relationships between AUs and facial expressions. Based on a comprehensive analysis from different perspectives, we propose a novel facial expression recognition approach that uses multiple features and multiple classifiers embedded into a stacking framework based on AUs. Extensive experiments on two public datasets show that our proposed multi-layer fusion system based on optimal AUs weighting has gained dramatic improvements on facial expression recognition in comparison to an individual feature/classifier and some state-of-the-art methods, including the recent deep learning based expression recognition one.

Keywords: feature fusion; multi-layer ensemble; action units (AUs); association rules; facial expression recognition

1. Introduction

Facial expression plays a significant role in human society for interpersonal communication. As indicated by Mehrabian, 55% of human emotional expression is transmitted by facial expressions [1]. Teaching computers to understand the psychology of human beings from their emotions will enhance the intelligence and friendliness of computers, which will bring outstanding benefit for facilitating human-machine interaction and providing services to humans. Increasingly, our computers are becoming like robots, interacting with people in a sensory-motor world. Although current

research has achieved improving results, making computers understand human expressions is still a challenge. Part of the reason is due to the complexity of human expression, which causes difficulty in discriminating feature extraction and classifier determination. Actually, current existing features and classifiers achieve good results in facial expression recognition from a different perspective, so it is natural to consider integrating the functions of single feature and classifier to solve the complexity problem of facial expression, but which one is not an easy problem to solve. In this paper, we propose a multi-feature and classifier fusion-based facial expression and utilize a data-driven method to learn the contributions of each feature and classifier. Furthermore, we also consider action units instead of whole face as the basic processing units to deal with the different roles of facial regions, such as mouth and eyes, in facial expression recognition. This is because each facial action unit contributes differently in expressing understanding. To obtain more ideas for our proposed final solution, we review the literature on relative research of facial expression recognition, mainly in features, classifiers and fusion solution. In addition, we also analyse the related work in facial expression recognition by integrating local region interpretation.

The remainder of this paper is organized as follows. Section 2 presents the review of related work. Section 3 introduces the overview of our proposed fusion scheme. Section 4 addresses our basic idea of affective relative AUs selection and relative feature extraction methods. Section 5 explores our fusion performance analysis of feature pairs based on the Kappa-Error diagram. Section 6 elaborates our AU weighting multi-layer fusion algorithm for facial expression recognition. Then experimental results are given in Section 7 and the discussion and further works are presented in the final section.

2. Related Work

2.1. Features for Facial Expression Recognition

Facial expression analysis usually employs feature extraction methods to describe images in different ways and to facilitate system classification. In the current practice of facial expression recognition analysis, there are many typical feature descriptors, where different extracting strategies have different principles. According to our understanding, there are three categories.

One category is texture-based feature. In this category, one of the common features is a global one, using transforming domain methods, such as Gabor wavelet and Principal Component Analysis (PCA), to extract facial appearance information [2–6]. In the last decades, the most efficient facial description method utilizes histograms to record feature vectors such as Histograms of Oriented Gradients (HOG) [7,8], Histograms of Local Binary Patterns (HLBP) [9–11] and Weber Local Descriptor (WLD) [12,13], which calculated the change in orientation or a stimulus (such as sound, lighting) with facial expression. The other popular feature extraction methods include Haar-features [14,15] and scale-invariant feature transform (SIFT) descriptors [16].

One category is a geometric feature derived from landmark extraction. This is normally based on an established two dimensional or three dimensional (2D/3D) facial model to locate feature points based on shape parameters and various derived facial detection models, such as Active Appearance Mode (AAM) [17,18]. Some fast landmark extraction methods have been proposed in recent years by using the combining method with face and eye detection with the Vilas–Jones algorithm and together with a local descriptor to detect the interested region and extract the extreme points as landmarks [19]. Other research extracts the accurate localization of landmarks, relying on the geometrical properties of facial shape with the differential geometry as the theoretical substratum [20]. The derived geometric feature could be simple, with geometric features such as lip width, and with lengths describing the expression states such as how open the mouth is. Some research derived more complex geometric features, for example, calculating a set of geodesic and Euclidean distances, together with nose volume and ratios between geodesic and Euclidean distances; specifically, on the mouth and eyebrows as a separate study for facial expression analysis in paper [21].

One category is temporal feature, calculated from the image sequence, which has the benefit of revealing the dynamic attribution of expression and improves the discriminative performance in expression recognition. Optical flow is one popular and traditional dynamic information extraction method for generating features [16,22]. Face tracking is another common approach to locate and track the landmarks and derive the temporal features. As demonstrated in the research by Deepak Ghimire et al., they use Elastic Bunch Graph together with eye detection to establish the landmarks, then the tracking is done among two adjacent frames by iteratively processing and calculating displacements with Gabor jet. The feature is generated with a derived geometric representation of a single frame image and a set of the transformed landmark coordinate positions of succeeding frame images. It needs to be mentioned here that the ensemble classification scheme achieves high recognition accuracy using multi-class AdaBoost and support vector machines, respectively [23].

With development of the deep learning algorithm, the learning based feature extraction, instead of the above manual construction ones, is applied in the facial expression recognition fields [24]. However, the requirement of the large training sample of deep learning based feature extraction is not practical in most facial expression applications, with the obstacle of capturing and labelling enough expression samples. Furthermore, the traditional facial expression feature becomes mature after decades of research. Therefore, we still pursue an effective solution based on the typical facial features.

All in all, the above three categories of features have diverse formula rooted indifferent understandings of facial properties and lead to the different discriminate performance in facial expression recognition. In this paper, we explore a fusion solution to achieve overall performance improvement.

2.2. Fusion Strategy for Facial Expression Recognition

Although the existing feature extraction approaches provide the discriminate features to foster the facial expression recognition performance, none has proved to be capable of achieving satisfactory results in all application cases. As we illustrated in Section 2.1, part of the reason is that a certain feature or learning approach has an inherent scheme and usually fits to a certain category of problem. Therefore, a fusion-based solution has been proposed to improve the overall performance by taking advantage of feature extraction approaches' complementary performance. To get a better understanding of feature extraction approaches' fusion performance, this paper proposes an analysis method based on a diversity and distinction evaluation metric by utilizing the Kappa-error diagram [25].

In current research, many fusion-based expression recognition approaches have appeared. Bartlett et al. employ an Adaboost and Support Vector Machine (SVM) fusion method together with the Gabor feature for expression recognition. They gained a 93% recognition rate on the Cohn–Kanade database. Additionally, based on the Facial Action Coding System (FACS) technique, the recognition rate increases by 1.8% [2]. Thiago et al. utilize Gabor and LBP as feature descriptors. Their experiments on the Cohn–Kanade and Japanese Female Facial Expression (JAFPE) datasets showed that the recognition rate with multiple features and classifiers is 10% and 5% above that with a single feature and classifier respectively [26]. Koelstra et al. propose a dynamic texture-based approach based on facial Action Units (AUs). They use a Gentle Boost ensemble algorithm with the Hidden Markov Model, which obtained an 89.8% classification rate for expression recognition on the Cohn–Kanade dataset [27]. Valstar and Pantic exploited a fully automated facial AU detecting system with Gabor wavelet features and a SVM classifier which classified the 15 AUs occurring along or in combination with other AUs with a mean recognition rate of 90.2% in the Cohn–Kanade dataset [28].

Based on the above analysis, fusion provides a solution to overcome the shortage of single feature and improves the overall performance. Therefore, our paper adopts the fusion strategy to perform expression recognition. Considering that stack fusion provides a solution for the different source integration, we select the stack fusion based approach, with which we have achieved good results in our previous work on multi-classifier fusion for facial expression recognition [29].

2.3. Facial Partition Strategy for Expression Recognition

In research by Cunningham et al., they prove that most expressions rely primarily on a single facial area to convey meaning with different expressions using different areas [30]. From the technology points, there is much research on making facial expressions with separate facial areas as a basis. Hsieh et al. locate facial components by an active shape model to extract seven dynamic face regions and they trained a multi-class support vector machine (SVM) to classify six facial expressions [31]. This method provides a typical idea of applying dynamic face regions to facial expression recognition.

3. Overview of Our Proposed Multi-Layer Fusing Facial Recognition Approach

Previous work has demonstrated that fusion based facial expression recognition approaches including multiple features and classifiers outperform the single techniques. These approaches have the advantage of integrating the complementarity of each source. Additionally, applying FACS, developed by American psychologists Ekman and Friesen [32], helps to improve the performance of facial expression recognition by integrating the action prediction of affective related action units. Motivated by the good performance of the fusion strategy, we propose a multi-view fusion approach that deploys stacking architecture to integrate AU blocks with multiple typical features and classifiers. Furthermore, to facilitate counting the contribution of AUs for the different expressions, an AU-weighted matrix is utilized at the top level of facial expression recognition. The associated rule algorithm is employed in mining each AU's contribution to each basic expression. From the global point of view, the flow chart diagram of the overall method is illustrated in Figure 1.

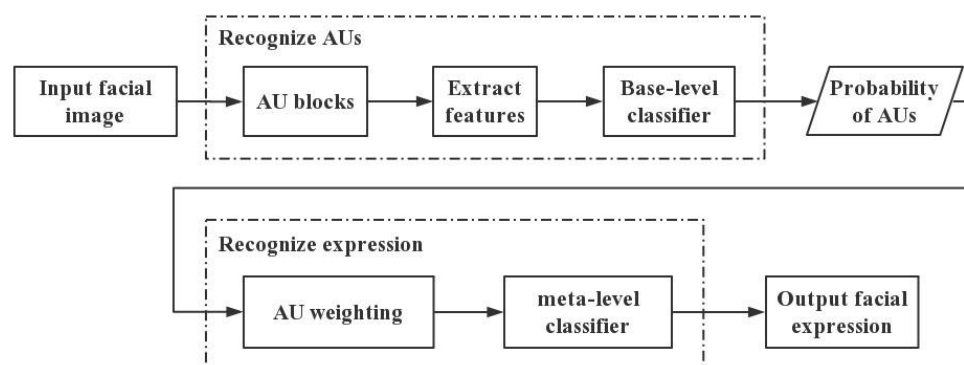


Figure 1. Flow chart diagram of the fusion-based expression recognition approach.

As shown in Figure 1, there are two layers to our facial expression recognition methods. The first layer, called base-level classifier, works on the AU prediction. In this part, each AU block is represented with multiple features. Then, each feature is input to several classifiers separately and obtains the relative AU prediction probability. The second layer, viz. meta-level classifier, works on facial recognition. At this layer, the obtained AU prediction probability pool from first layer is weighted. Then, the weighted probability pool is input to the meta-level classifier which outputs the expression determination results.

The architecture of our proposed approach is shown in Figure 2. There are two learning functional modules. One is the two-layer fusion for recognition with the stacking framework. The other is AU contribution learning, which utilizes association rules to explore AUs' weight matrix for expression estimation. The details of the processing procedure are illustrated as follows.

Firstly, facial images are separated into several AU blocks. Then, four kinds of feature extraction methods are used separately in describing every AU set. After that, the occurrence probability of each AU action is predicted with the base-level classifier. Finally, the facial expression classification is done, with the obtained AU probability as an input at the meta-level. Note that the probability of AU occurring is weighted before classification in the paper. The weights refer to the contribution

of each AU in determining each basic expression. We describe it as a weight matrix, as shown at the bottom right of Figure 2. This is obtained by sample learning at the second process, where we utilize association rules to explore AUs' weight matrix. As manual labelling of AU in each expression is difficult, we still use the automatic AU labelling on expression instances. In this paper, we use only a simple feature and classifier to train and obtain the corresponding AU label of each expression instance in the dataset. By excavating a strong relationship between AUs and expressions from the dataset, AUs' weight matrix is set.

In our proposed method, we take full consideration of a variety of each AU block in facial expression interpretation and make full use of the complementarity of different features and classifiers on facial expression recognition. The stacking ensemble architecture ensures the contribution of each source view derived from the data learning. In this way, the effectiveness and robustness of facial expression recognition is supposed to be improved.

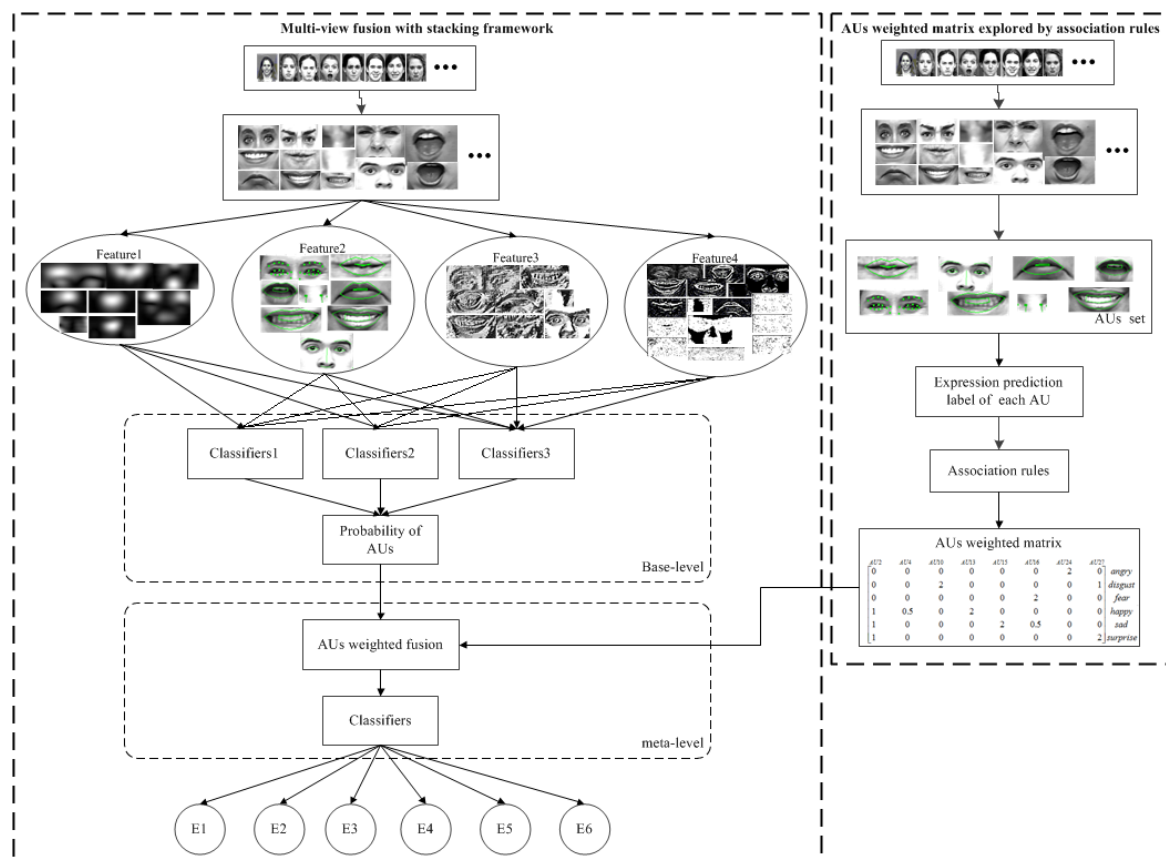


Figure 2. The architecture for our proposed fusion-based expression recognition approach (Note: we denote the outputs, viz. the six basic expressions—angry, disgust, fear, happy, sad, surprise—as E1, E2, E3, E4, E5, E6, separately).

4. Selection of Affective Relative AUs and Feature Extraction

4.1. Choosing Affective Relative AUs Based on FACS

There are 43 muscles in the human face, which can be combined into more than 10,000 kinds of facial expression. At least 3000 of these have specific affective meaning. Ekman and Friesen established the popular Facial Action Coding System (FACS) [32], which divided the human face into several independent “action units” (AUs). Based on our understanding of human facial anatomy and physiological structure, these AUs are considered to be critical for representing and recognizing facial expressions. Although the individual appearance differences in the human face are considerable,

the structure of faces is physically similar. In addition, some researchers had verified that most expressions rely primarily on a single facial area to convey meaning, and the combination of rigid head, eyes, eyebrows, and mouth motions is sufficient to produce expressions that are as easy to recognize as the original recordings [30]. Thus, based on the above two standpoints and cognitive experience, we chose eight AUs as a basis for predicting six basic emotions. The selected AUs are shown in Table 1.

These AUs are selected as representatives of the muscle action in particular facial regions, and are initially represented as image regions manually marked up based on sequences of video frames (peak movement vs. rest state). Figure 3 illustrates eight kinds of AUs which are contained in six basic facial expressions.

Table 1. Facial expression result of each feature and fusion feature performance with AUs.

AU	FACS Name
AU2	Outer Brow Raiser
AU4	Brow Lowerer
AU10	Upper Lip Raiser
AU13	Sharp Lip Puller
AU15	Lip Corner Depressor
AU16	Lower Lip Depressor
AU24	Lip Pressor
AU27	Mouth Stretch

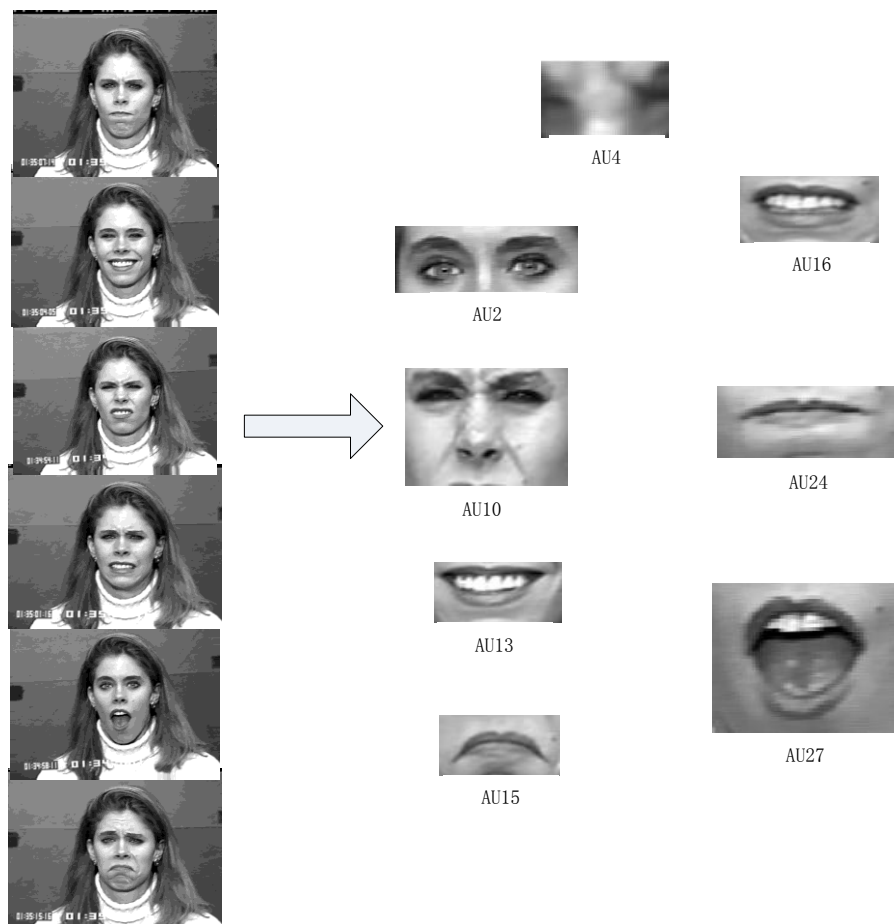


Figure 3. AU blocks from six basic facial expressions.

4.2. AU Feature Extracting Solution

In present research, a wide variety of feature extraction algorithms are used in facial analysis: Gabor wavelet, PCA (Principal Component Analysis), AAM (Active Apperance Model), HOG (Histogram of Oriented Gradient), SIFT (Scale-invariant feature transform), LBP (Local Binary Pattern), WLD (Weber Local Descriptor), etc. They reveal the discriminating property of objects for a certain classification target with difference principles. As each AU has a different texture structure, we prefer to choose several diverse features and fuse their complementary contribution in representing AUs. This paper employs Gabor, AAM, LBP, and WLD feature descriptors, which have different feature extraction mechanisms. The quantitative fusion performance between these selected features is examined in the next part. The details of the employed features are briefly illustrated as follows:

- Gabor feature:** Selecting a two-dimension Gabor wavelet with the scale and direction $v = 2$ and $u = 4\pi/8$ respectively. Together with PCA, dimension reduction Gabor is extracted to represent each AU.
- Geometry feature:** Utilizing the AAM shape model to calibrate feature points and generate corresponding geometry features for AU regions, and making the coordinate of feature points the geometry feature vector.
- LBP feature:** Choosing a flexible circle operator to extract the local texture of each AU, and using a histogram to record the LBP feature vectors of AU regions.
- WLD feature:** Using an emerging method that synthesizes multiple orientation and excitation operators of sub-image sets, and forming a histogram to record WLD feature vectors of each AU region.

Some results of Gabor/geometry/LBP/WLD feature extraction for AUs region are shown in Table 2.

Table 2. Four kinds of feature extraction results for AU blocks.

Image	Gabor	Geometry	LBP	WLD

5. Analysing Feature Fusion Performance According to Kappa-Error Diagrams

To evaluate fusion performance between the above selected typical features, we utilize kappa-error diagrams. The kappa-error diagram is a popular tool for analysing ensemble methods proposed by Margineantu and Dietterich [33], which is used to gain insight into ensemble performance. Kuncheva points out that there are two main factors reflecting the fusion performance: accuracy and pairwise diversity [25]. Higher accuracy and higher diversity (evaluated with kappa) indicates the better ensemble performance of a pair of methods. However, these two performance factors normally conflict and are unable to increase at the same time. Therefore, a bound in the diagram is defined, indicating ensemble performance of pairs with trade-off of two factors. In the kappa-error diagram, the pairwise point closing the bound curve indicates the better ensemble performance of these two methods. Referring to the concluded judge rules, this paper adopts the kappa-error diagram and bound in AU feature fusion analysis. The diversity and complementarity of every feature pair are evaluated for AU recognition. This paper applies this strategy to feature ensemble evaluation.

5.1. Theory of Kappa-Error Diagram Evaluation

Suppose that F_1 and F_2 are a pair of features that are under-estimated. For each AU from one dataset, each feature is extracted respectively. We employ several typical classifiers—KNN (K nearest neighbour), SMO (sequential minimal optimization), MLP (multilayer perception)—in AU recognition. To use various classifiers is to ensure the general and authority of AU judgment. The corresponding pairwise contingency table is counted with each classifier as in Table 3.

Table 3. Contingency table with two classification results.

Predict Result	F_1 Correct	F_1 Wrong
F_2 Correct	a	b
F_2 Wrong	c	d

In the table, parameter a represents the number of samples for both extraction methods doing the right classification with the same classifier; b and c are the number of samples for which one extraction method is right and another is wrong; and d is the number of samples for which both extraction methods are wrong. Then, error “ e ” and kappa “ $kappa$ ” values are computed as follows [29]:

$$e = \frac{1}{2} \left(\frac{c+d}{N} + \frac{b+d}{N} \right) \quad (1)$$

$$kappa = \frac{OA - AC}{1 - AC} \quad (2)$$

where N is the number of samples in the dataset: here $N = a + b + c + d$. OA and AC are computed as follows:

$$OA = \frac{a+d}{N} \quad (3)$$

$$AC = \frac{(a+d)(a+c) + (b+d)(c+d)}{N^2} \quad (4)$$

where OA represents the average sample number for which both extraction methods have the same results with one classifier. Thus, it describes the coherence of two extraction methods. The higher the coherence, the closer the functions of two extraction methods will be. It means no improvement of performance with fusion. On the contrary, it indicates that the functions of two extraction methods are various. Fusing of these two features will help to provide complementary contribution and improve the whole ensemble’s results. To provide more objective criteria than accuracy directly, kappa is used, which removes the chance factor by minus AC . The diversity of two technologies being evaluated is based on the rule that the lower the kappa is, the higher the diversity is.

Error in Equation (1) is the average of the sample number for each feature doing the wrong classification with the same classifier. It is easy to understand that the higher error caused by each extraction method in the pair will reduce the entire function of the ensemble. High diversity and high accuracy are the target in determining the distinctive feature in AU fusion of the stacking ensemble system.

Further, Kuncheva examines the bound on the region for the dichotomous case where feasible kappa-error trade-offs are found. The paper derives bounds k_{\min} on kappa in terms of the error e , as in Equation (5). The pairwise closing of the bound has good performance benefit for fusion [25].

$$k_{\min} = 1 - \frac{1}{(1 - e)}, \quad 0 < e < 0.5 \quad (5)$$

5.2. Analyzing Feature Fusion Performance for AU Prediction Based on Kappa-Error Diagram Evaluation

Extracting a power feature not only reflects the essential characteristic of AUs, but also enhances the discriminative performance for AU prediction. Considering the differences of AUs' images, no feature displays stable description for all images. Actually, different features present different representational performance when describing the different objects or regions.

This paper designs a fusion experiment analysing the feature description ability for different AUs. Here, we chose four features (Gabor/geometry/LBP/WLD) for representing AU regions. Four-features extraction is done on the same public dataset at the same time. Three kinds of classifiers (MLP/SMO/IBK) were employed to train and learn AUs' features separately. Therefore, the corresponding kappa and error values are accounted for in each feature pair in predicting one AU region with one classifier. For four features, there are six pairs of feature combination. In Figure 4, three colours (red/green/blue) represent three kinds of classifiers, while six different symbols represent six feature pairs. Figure 4 shows the fusing results of six feature pairs for each AU prediction. The horizontal and vertical axes are the diversity parameter (kappa) and the average error parameter (error) respectively. The boundary curve of $K = 1 - 1/(1 - e)$ is shown in the diagram.

According to Figure 4, fusion performance of six feature pairs varies for different AUs' prediction. For the glabella region (AU10) and the middle region of face (AU4), feature combinations have high performance on AU recognition. The average error rate is generally lower than 0.7. Furthermore, diversity between four features exists in these areas. Additionally, the evaluation points are distributed close to the boundary, which indicates that the corresponding feature pair has good fusion characteristics. For eyes and mouth regions (AU2/AU13/AU15/AU16/AU24/AU27), the average error rate of feature pairs is about 0.7 and the kappa value in most situations is higher than others, which illustrates that the diversity of feature pairs in describing these regions is relatively weak compared to the above two regions. It also could be found from Figure 3 that the Gabor and LBP feature combination, denoted by "+", shows a better fusion performance than others. However, no other feature pairs display an overall advantage for all AUs' prediction. Four features show different representing ability in different AU regions, and fusion of any two features has a variety of representation performance. In conclusion, this paper adopts all four features to describe each AU and fuses them to enhance the overall describing ability for all selected expression AUs.

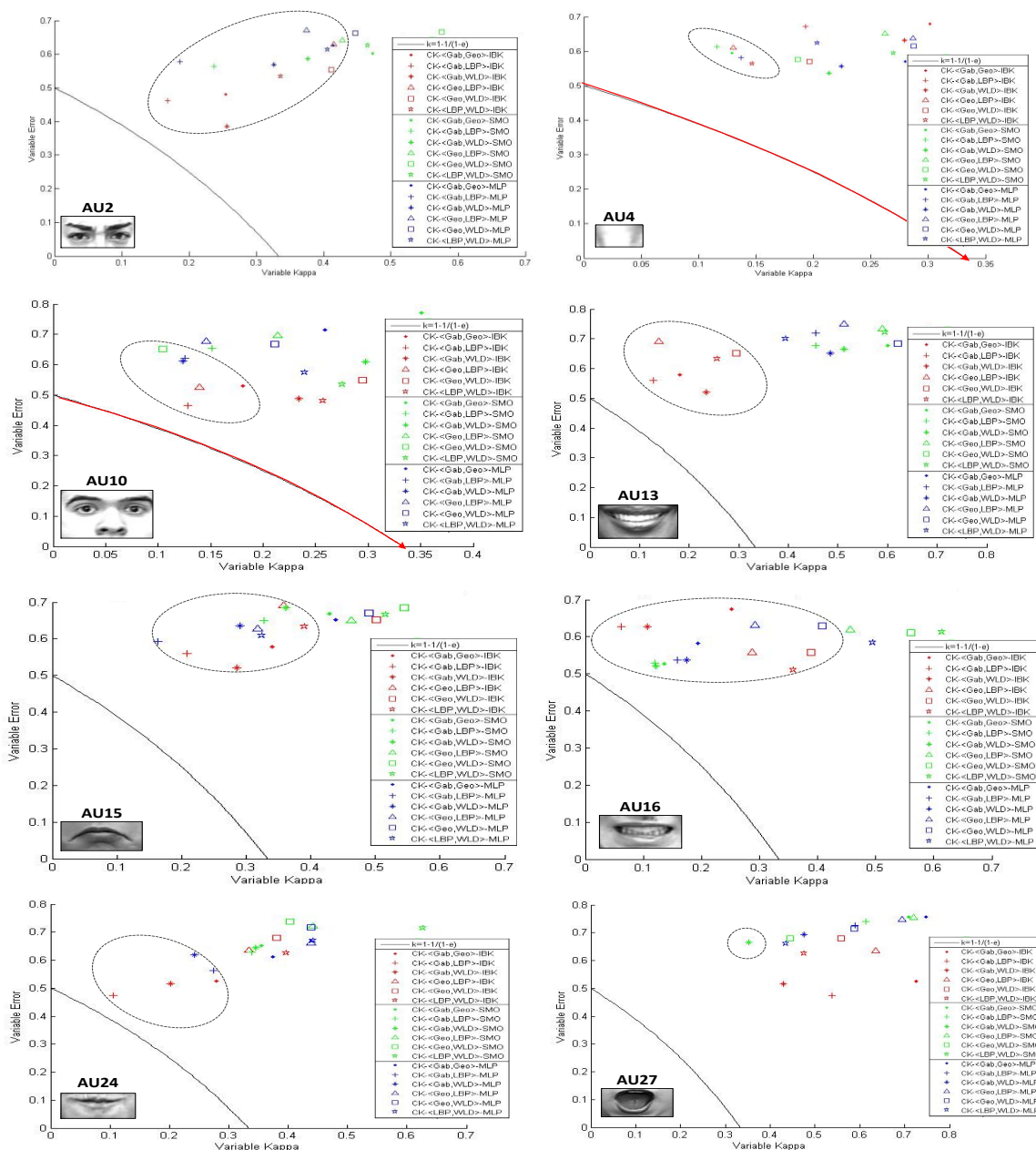


Figure 4. Kappa-error diagram of six pairs of feature descriptors. The points obtained by experimenting on the Cohn-Kanade (CK) dataset—the IBK, SMO, MLP classification results—are shown in red, green and blue. Feature pairs Gab-Geo/Gab-LBP/Gab-WLD/Geo-LBP/Geo-WLD/LBP-WLD are marked by *, +, “diamond”, “angle”, “square”, “star”.

6. Multi-Layer Fusion with Weighted AU for Facial Expression Recognition

6.1. Multi-Layer AUs' Fusion with Stacking Framework

Stacking is a technique to fuse multiple classifiers applied to a specific classification problem [34]. It aims to improve the classification results over the individual classifier. It outperforms other fusion methods by simply voting or linear combination, which weights the function of individual classifiers through the sample training 29. This paper proposes a novel stacking-based ensemble structure combined with AUs' prediction in a facial expression recognition system. For each AU prediction, our proposed ensemble structure also takes full advantage of feature and classifier fusion to obtain better understanding of emotion action.

The stacking framework has two parts: the base level, which processes the input respectively with several base classifiers, and the meta-level, which relearns and integrates the results of base classifiers. The detailed procedure of stacking is illustrated as follows. Supposing there are n base classifiers marked as F_1, F_2, \dots, F_n , one meta-classifier marked as M , and m classes marked as C_1, C_2, \dots, C_m ; each sample S will be processed by the following procedure.

Stacking Procedure	
Step 1	$i = 1$.
Step 2	If $i \leq n$, go to Step 3. Otherwise, go to Step 4.
Step 3	Classifying with the base classifier F_i . For each sample S , the probability vector $P_i = \{P_{i1}, \dots, P_{ij}, \dots, P_{im}\}$ under the base classifier F_i is derived, where p_{ij} indicates the probability that sample S is assigned to class C_j ($j = 1, 2, \dots, m$). Then $i++$ go to Step 2.
Step 4	The classification results of all base classifiers are obtained, here marked as $P = \{P_1, P_2, \dots, P_n\}$. Then go to Step 5.
Step 5	The meta-classifier M processes the input data, viz. the matrix P from base classifiers, and outputs the ultimate recognition result.

Based on the stacking principle, we propose multiple channels for AUs' prediction at base level. As illustrated before, we pick eight affective related AUs, which include mouth, eyes, intercilium and middle of facial regions. At base level, we employ four alternate feature descriptors (Gabor, geometry, LBP, WLD) to represent AU blocks and use typical classifiers (SMO, MLP, IBK) to calculate the probability of whether the specific action happens or not. At the meta-level, we integrate the prediction probabilities of AUs as the input to the meta-classifier for further expression recognition. Figure 5 gives our proposed multi-layer facial expression recognition based on the stacking ensemble algorithm.

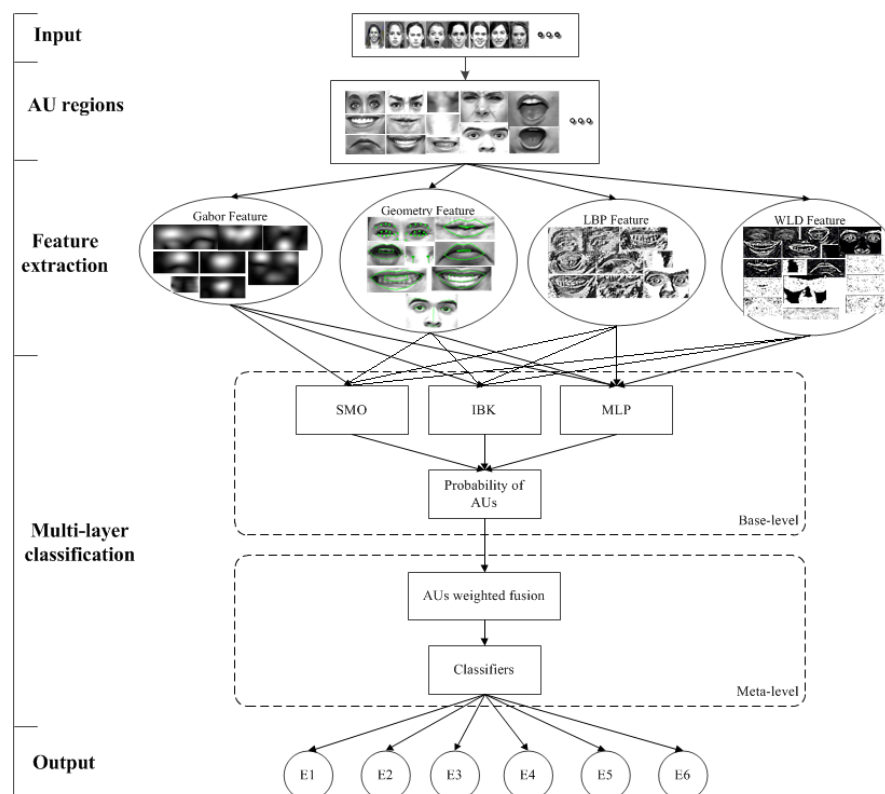


Figure 5. Multi-layer facial expression recognition based on the stacking ensemble strategy.

6.2. AUs' Weight Learning by Association Rules

To optimize the contribution and cooperation of eight AU channels for facial expression recognition, this paper provides a novel strategy, mining the strong relationship between AUs and emotions based on an a priori algorithm. Based on the learned rules, the weights of AUs are adjusted accordingly.

Association rule learning is a method for discovering relations between variables in large datasets. It is intended to identify strong rules in datasets using some measures of interestingness [35]. In order to select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support and confidence [36]. Support is the fraction of transactions that contain both X and Y items; confidence is the fraction of transactions containing items X that also contain items Y. Intuitively, support measures the significance of the rule, so we are interested in rules with relatively high support. Confidence measures the strength of the correlation, so rules with low confidence are not meaningful, even if their support is high. We employed the basic a priori algorithm for finding the sets of items with high support, then used confidence as a threshold to filter the strong relationship. Therefore, a rule in this context is there relationship among transaction items with enough support and confidence.

Motivated by this principle, we propose a strategy based on an a priori algorithm to help us find a strong relationship between AU and expressions. Counting the statistic rules from the data, we found the contribution of AU in expression prediction. The details of the process are described as follows.

First of all, this paper selected a geometric feature to extract the AU feature, which is identified in a more stable manner by different classifiers on expression recognition performance and outperforms any single feature descriptors among all the recognition results of AUs' prediction, and used classifiers from base level to distinguish the prediction expression label of eight AU regions. Note that the output in this learning process is the expression prediction label instead of the prediction probability of each AU. Then, the a priori algorithm was used to carry out AU strong association mining for each expression, which treated the eight AU prediction labels of one sample and its real expression label as an item set for finding a strong relationship between AUs' prediction expressions and terminal expression. Table 4 shows part of one emotion sample: for one image, there are eight AU prediction labels, which are in the set of angry, disgust, fear, happy, sad and surprise and the real expression label is in the last column (italic). Thus, one item-set consists of eight expression predicting labels and one expression real label, and all the item-sets form a whole dataset for one expression. We made them the input of an a priori algorithm.

Table 4. Part of E1 (angry) item-set.

Num	Expression Prediction Label								Real Label of Expression
	AU2	AU4	AU10	AU13	AU15	AU16	AU24	AU27	
1	angry	disgust	angry	angry	angry	angry	angry	sad	<i>angry</i>
2	angry	happy	angry	angry	disgust	angry	angry	angry	<i>angry</i>
3	angry	angry	angry	angry	sad	angry	angry	angry	<i>angry</i>
...				

We picked out strong relationships between AUs and expressions for which min-support was greater than 0.85 and min-confidence was greater than 1. Parts of the results are shown in Table 5. One rule noted was that $P_{AUi} \rightarrow R_{Ej}$, where the AU prediction label was in the left set and the real label was in the right set.

Table 5. The result of association about AU and simple expression based a priori.

Expression	Association Rules of AU
E1 (39)	AU24 = angry \geq Expression = angry 39
E2 (43)	AU10 = disgust \geq Expression = disgust 41; AU27 = disgust \geq Expression = disgust 39
E3 (55)	AU16 = fear \geq Expression = fear 54
E4 (81)	AU13 = happy \geq Expression = happy 81; AU2 = happy \geq Expression = happy 75; AU4 = happy \geq Expression = happy 74
E5 (62)	AU15 = sad \geq Expression = sad 60; AU2 = sad \geq Expression = sad 56; AU16 = sad \geq Expression = sad 53
E6 (75)	AU27 = surprise \geq Expression = surprise 75; AU2 = surprise \geq Expression = surprise 70

In Table 5, we could find that each emotion is strongly related with several specific AUs. Therefore, we assign the different weights to corresponding AUs for emotion recognition. To make the quantitative analysis, the relationship is unified based on intentions. Noting the number of $P_{AUi \rightarrow R_{Ej}}$ as $\text{Num}(P_{AUi \rightarrow R_{Ej}})$ and total number of each emotion samples as $\text{Num}(D_{Ej})$, viz. $R_{Ej} = \text{Num}(P_{AUi \rightarrow R_{Ej}}) / \text{Num}(D_{Ej})$, i, j . Then, we mapped R_{Ej} into the three degrees of weight as 2, 1, 0.5, respectively; the mapping rules were:

$$0.95 \leq R_{Ej} \leq 1 \Rightarrow \text{weight} = 2 \quad (6)$$

$$0.90 \leq R_{Ej} \leq 0.95 \Rightarrow \text{weight} = 1 \quad (7)$$

$$0.85 \leq R_{Ej} \leq 0.90 \Rightarrow \text{weight} = 0.5 \quad (8)$$

Accordingly, the optimal AU weighted matrix with six expressions is produced. The size of matrix W is 6×8 , which is six expressions and eight AU channels, shown in Equation (9).

$$W = \begin{matrix} & \begin{matrix} AU2 & AU4 & AU10 & AU13 & AU15 & AU16 & AU24 & AU27 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 1 & 0.5 & 0 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 2 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} & \begin{matrix} angry \\ disgust \\ fear \\ happy \\ sad \\ surprise \end{matrix} \end{matrix} \quad (9)$$

$$F = W * \begin{bmatrix} AU2 \\ AU4 \\ AU10 \\ AU13 \\ AU15 \\ AU16 \\ AU24 \\ AU27 \end{bmatrix}_{8 \times n} \quad (10)$$

Then, we multiply the AU weight matrix and AU prediction probability matrix from the stacking multi-layer model as feature vectors F as in Equation (10), which is the input to the classifier for recognizing facial expressions.

Here, W is the AU weight matrix. The process of the weight AU matrix based on association rule learning is illustrated as in Figure 6. Note that in Figure 6, EPL is expression prediction label; SR is strong relationship; ERL is expression real label.

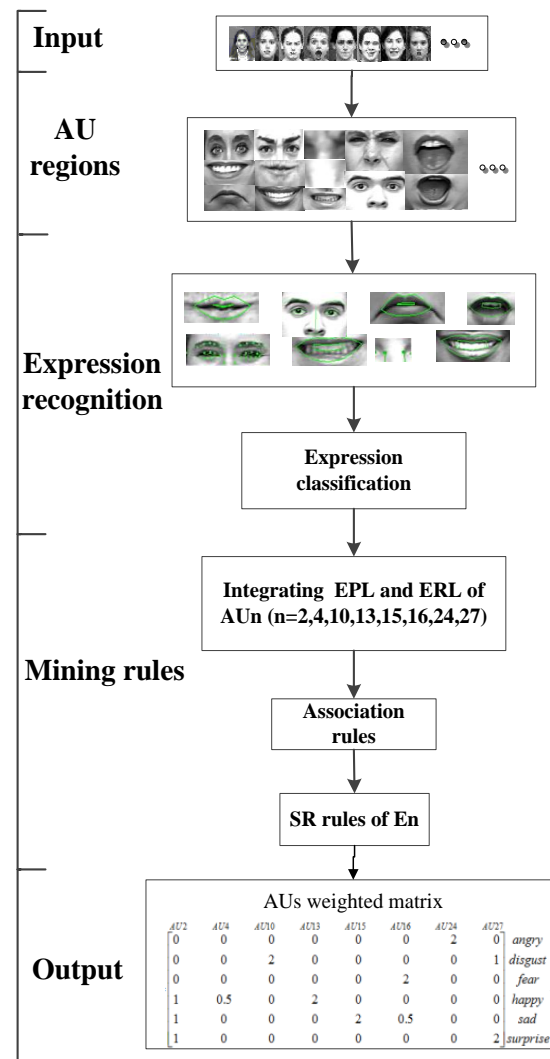


Figure 6. The framework of the a priori based strong relationship mining between AU and emotion.

6.3. AU Weighted Fusion Algorithm for Facial Expression Recognition

This algorithm combines the FACS technique and stacking framework. On the basis of AUs' prediction at each block with a multi-layer fusion model, weighted AUs' predictions are fused and processed for facial expression recognition. This two-layer model for expression recognition is shown in Figure 7.

As shown in Figure 7, the input at fusion level 2 is the prediction probability pool of eight AUs with the combination of three classifiers and four features respectively from level 1. The obtained combination of the AU prediction pool is multiplied with the learned AU weighted matrix as shown in Equation (6). Therefore, each expression only uses its strong related AU blocks and the input AU prediction pool will be reshaped as six rows of the weighted prediction pool. These six rows of the weighted prediction pool are generated inputs of six expression classifiers, viz. meta-classifier, respectively. The final expression is the one with the highest prediction possibility of these six binary expression classifiers.

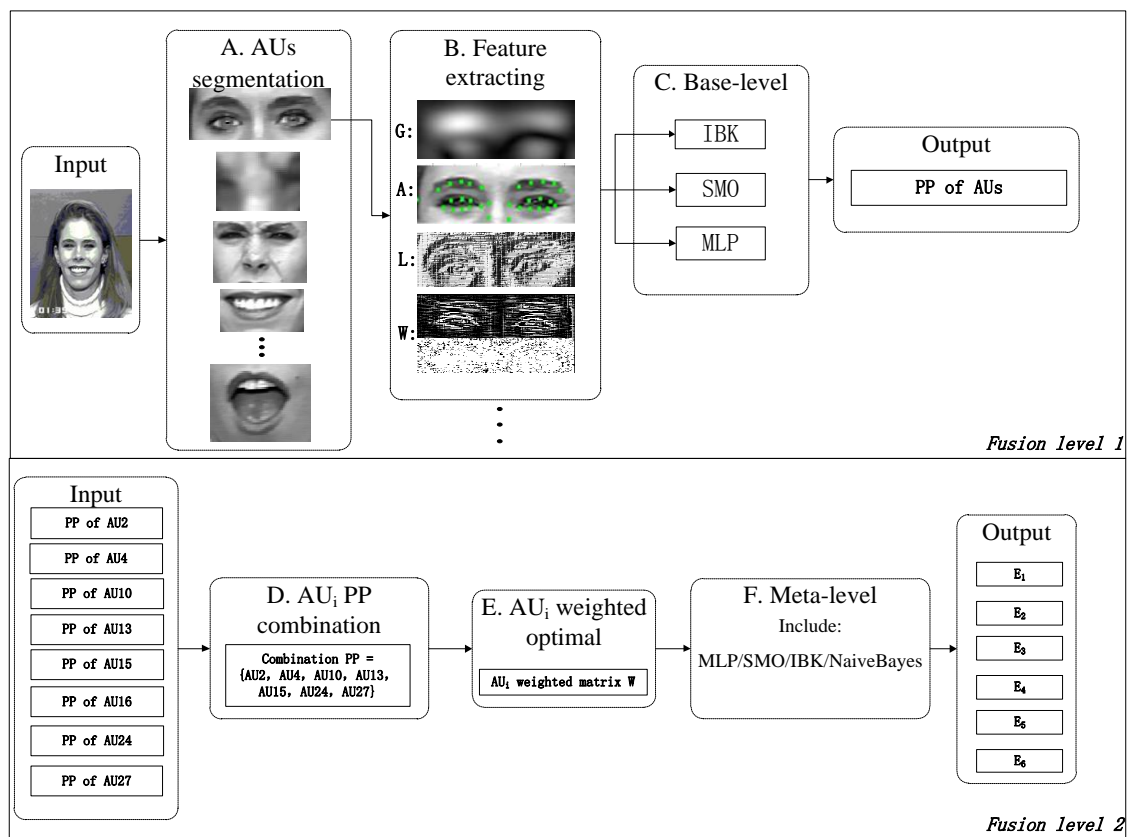


Figure 7. Multi-level fusion model of expression recognition.

Further details of the overall procedure for the approach in Figure 7 are as follows.

Abbreviations and notation

AU—action unit—active for each instance/image

AU_i —action unit block—e.g., use (2, 4, 10, 13, 15, 16, 24, 27)

G, A, L, W—features—e.g., use (Gabor, AAM, LBP, WLD)

PP—prediction probability

E_i —facial expressions—e.g., angry, disgust, fear, happy, sad, surprise

Fusion level 1

A. AUs' segmentation: divide facial image into eight AU blocks which related with six basic expressions.

AU_i consists of $AU_2, AU_4, AU_{10}, AU_{13}, AU_{15}, AU_{16}, AU_{24}, AU_{27}$;

B. Feature extraction: select G, A, L, W feature descriptors; extract each AU region. For one kind of AU, $AU_i = \{F_G, F_A, F_L, F_W\}, i \in \{1, 2, \dots, 8\}$;

C. Base-classifiers recognition: treat SMO, MLP, IBK as a base classifier to recognize AUs; calculate PP of each AU_i . For one kind of AU, the $P_j = \{P_G, P_A, P_L, P_W\}$, and j is SMO, MLP, IBK;

Fusion level 2

D. AU_i PP combination: integrate PP from three classifiers matched to four features, for one AU region, format $12(3 \times 4)$ dimensions vector. $AU_i = \{p_1, p_2, \dots, p_3\}$, and $i \in \{1, 2, \dots, 8\}, m \in \{1, 2, 3, \dots, 12\}$;

E. AU_i weighted optimal: employ an a priori algorithm, mining the strong relationship between AU_i and expressions; produce the AU tuning matrix $W_{6 \times 8}$; Note: Each row is a weighted AU feature input for one basic expression classification.

F. Meta-classifiers recognition: use one type of classifier to recognize facial expression; including MLP, SMO, IBK and NB. Note: there are six binary classifiers for determining the possibility of each expression. The final expression is the one with the highest prediction possibility.

7. Experimental Results

7.1. Image Datasets and Experimental Setup

To demonstrate the effectiveness of the proposed multi-layer fusion with the AU weight optimal strategy, we have performed experiments on two public facial expression datasets: CK dataset (CK) and JAFFE dataset (JAFFE). The number of images for each emotion is shown in Table 6.

Table 6. Sample size of each emotion on CK and JAFFE datasets.

Emotions	CK	JAFFE
1: angry	39	30
2: disgust	43	29
3: fear	55	32
4: happy	81	32
5: sad	62	30
6: surprise	75	30
Total images	355	183

At the base-level, three classifiers are selected: SVM, MLP and IBK. SVM used in the paper is a simple support vector machine SMO with linear kernel. Parameters of used MLP are set with Weka's default hidden layer (MLP with hidden layer default "a" signifying the arithmetic mean of the input and output layer sizes). Instance-based learning IBK is configured with Weka's default of 1 nearest neighbour. The classifier at the meta-level employs Naive Bayes (NB), based on so-called Bayesian theory, which can often outperform more sophisticated classification methods [37]. All the training and testing processes are done in a cross-validation way with 10 folds in weka.

7.2. Comparison on Multi Features or Classifiers Fusion Performance Analysis

To verify the rationality and validity of our proposed multi-layer fusion strategy, we designed two comparison experiments. One compares one feature with multiple feature fusion, and the other compares one classifier and multiple classifiers. This paper makes comparison on the recognition performance (accuracy rate) between our proposed full fusion methods with multiple features and multiple classifiers with two partial fusions.

Here, we use one of four features (Gabor/geometry/LBP/WLD) as the feature for AUs and employ combined classifiers with MLP, SMO, IBK and NB to do facial expression recognition. From the experimental results (Table 7), compared with using single feature, the multiple feature fusion based method exhibits the best expression recognition results even in different classifiers and different datasets.

Similarly, this paper used one of three classifiers (IBK/MLP/SMO) as a single classifier. Comparison is made between this classifier and a multi-classifier combination (SMO + IBK + MLP) under the same AUs feature for facial expression recognition. From the experiment (Table 8), we found that the recognition rate of all single classifiers alone does not exceed the multiple classifiers fusion method.

As is shown in the Tables 7 and 8, all the cases with highest prediction accuracy are specified in bold. The results show that the multiple features or the multiple classifiers fusion method achieves the highest accuracy rate in most cases and comparative results with the highest ones in a few other cases. It indicates that the fusion methods generally outperform any single feature or classifier-based methods. The results also verify that there is a great complementarity between different features/different classifiers. Therefore, benefiting from making full use of the diversity among different features and classifiers, fusion-based methods overcome the shortcomings of a unilateral aspect. Conclusively, the fusion strategy for AUs' prediction based on a stacking framework is feasible.

Table 7. Comparing recognition results with simple feature and multi-feature based on AU.

Accuracy (%)	Feature Description	MLP	SMO	IBK	NB
JAFPE	Gabor feature	63.19	59.89	62.09	53.30
	Geometry feature	83.52	79.67	84.07	78.02
	LBP feature	46.15	45.60	47.25	47.80
	WLD feature	78.02	71.98	75.82	70.33
	Gab + Geo + LBP + WLD	89.07	90.16	81.42	83.61
CK	Gabor feature	64.97	55.65	63.84	62.99
	Geometry feature	88.98	84.18	86.44	87.29
	LBP feature	62.43	53.67	59.32	62.71
	WLD feature	72.32	65.82	72.32	73.73
	Gab + Geo + LBP + WLD	94.65	91.83	85.92	93.52

Table 8. Comparison of recognition results with a simple classifier and multi-classifiers based on AU.

Accuracy (%)	Base Classifiers	Gabor Feature	Geometry Feature	LBP	WLD
JAFPE	IBK	53.30	64.29	38.87	60.85
	MLP	50.27	74.45	48.764	64.42
	SMO	50.27	60.16	42.99	60.99
	SMO + IBK + MLP	59.62	81.32	46.70	74.04
CK	IBK	53.74	74.79	50.49	58.55
	MLP	57.98	87.64	54.94	64.55
	SMO	58.90	80.00	58.05	64.97
	SMO + IBK + MLP	61.86	89.22	59.53	71.04

7.3. Performance Analysis with a Variety of Single Classifier Performance

As mentioned before, the selected features and classifiers are basic. As we aim to prove the stacked based fusion scheme, multiple features and multiple classifiers are beneficial to integrate the advantage of performance and overcome the limitation of a single feature and classifier. Therefore, the overall performance would be improved if each single source, including features and classifiers, is more powerful. To demonstrate this, we discuss the performance when changing the performance of the meta-classifier for facial expression recognition. To do this evaluation, we employ the different types of SVMs, from linear SVM to kernel SVM, with polynomial and radial basis kernel function respectively as the meta-classifier. The parameters are configure as default parameters of Weka's SVM. The first layer is still fusion of three classifiers (IBK/MLP/SMO) and four features (Gab + Geo + LBP + WLD) for AU recognition. The experiment result with the 10-fold approach is shown in Table 9.

Table 9. Comparison of recognition results with a simple classifier and multi-classifiers based on AU.

CK	Linear Support Vector Machine	Polynomial Kernel Support Vector Machine	Radial Basis Function Kernel Support Vector Machine
Accuracy (%)	91.83%	92.39%	93.24%

The experiment result demonstrates that overall performance improves when we enhance the single classifier performance. This conclusion could be expended to the first layer doing AU recognition. If processing with a more powerful classifier or feature, the overall performance will increase. However, we prefer to employ the basic classifiers and features to prove our fusing strategy and to prove that even an ensemble with a basic diversity source will, by learning, still achieve the better classification result.

In this part, we discuss the classification performance with the confusion matrix to clarify the possible reason causing the false prediction. The confusion matrix of facial expression recognition with the above three SVM as meta-classifiers is shown in Figure 8.

a = angry	a	b	c	d	e	f	<--	a	b	c	d	e	f	<--	a	b	c	d	e	f	<--
b = disgust	31	3	0	0	5	0	a	35	3	0	0	1	0	a	34	2	0	0	3	0	a
c = fear	0	40	0	1	2	0	b	2	39	1	1	0	0	b	2	38	1	1	1	0	b
d = happy	0	1	48	6	0	0	c	0	0	44	8	2	1	c	0	0	43	8	2	2	c
e = sad	0	0	3	77	1	0	d	0	0	0	81	0	0	d	0	0	0	81	0	0	d
f = surprise	6	1	0	0	55	0	e	4	0	1	0	57	0	e	3	0	2	0	57	0	e
	0	0	0	0	0	75	f	0	0	0	0	0	75	f	0	0	0	0	0	75	f
	(a)							(b)							(c)						

Figure 8. Confusion Matrix. (a) Linear SVM; (b) Polynomial kernel SVM; (c) Radial basis function Kernel SVM.

The above fusion matrix shows that most mixing expressions happen mostly at the same pairs, for example, the sad–angry pair, the disgust–angry pair, and the fear–happy pair. It is interesting to find that most situations with the false recognition reduction by the Kernel SVM compare to the linear SVM. We also find that confusing expressions, such as disgust and angry, are to some extent easily confused emotions. However, the happy emotion is still confused with fear. We analyse the images. Some are shown in Figure 9. We can see that one happy expression image and one fear expression image for the left subject have bigger differences than for the happy and fear expression from the right subject. We deduct that this false recognition is partly because the images are similar. Therefore, a more powerful discriminate feature or classifier could help to improve the overall performance.



Figure 9. Happy and fear expression examples.

7.4. Experiment on AU Weighting Multi-Layer Fusion Expression Recognition Performance Analysis

To test the performance of our fusion expression recognition approach with or without the learned weights of AU, we did comprehensive experiments to classify six expressions (angry, disgust, fear, happy, sad, surprise), which will ensure the comprehensiveness of final recognition results. Several different metrics are used for evaluating algorithm performance, which are the recognition rate, false positive rate and receiver operating characteristic curve (ROC) area for each expression. Figures 10 and 11 show the results of the multi-layer model with or without AU optimal weight respectively. From top to bottom, performance comparison on the CK and JAFFE datasets is shown in terms of (a) recognition rate; (b) false positive rate (FTP) and (c) ROC area(ROCA) for six basic expressions.

Experimental results in Figures 9 and 10 show that the proposed multi-layer face recognition method with the AU weighted strategy has a significantly increased recognition rate. For six expressions, the mean accuracy achieved is 96.62% for CK and 94.63% for JAFFE, which exceed, by 2.02% and 2.23%, that without AU weight tuning. The average FTP of the AU optimal weight is 0.103 in CK and 0.147 in JAFFE, much lower than un-tuning AU weighted. The average of ROCA is 0.94 for CK and 0.91 for JAFFE, which shows the capability of classifier recognition: the higher the value, the stronger the performance. We can obviously conclude that our proposed AU optimal weighted fusion is effective and accurate for facial expression recognition.

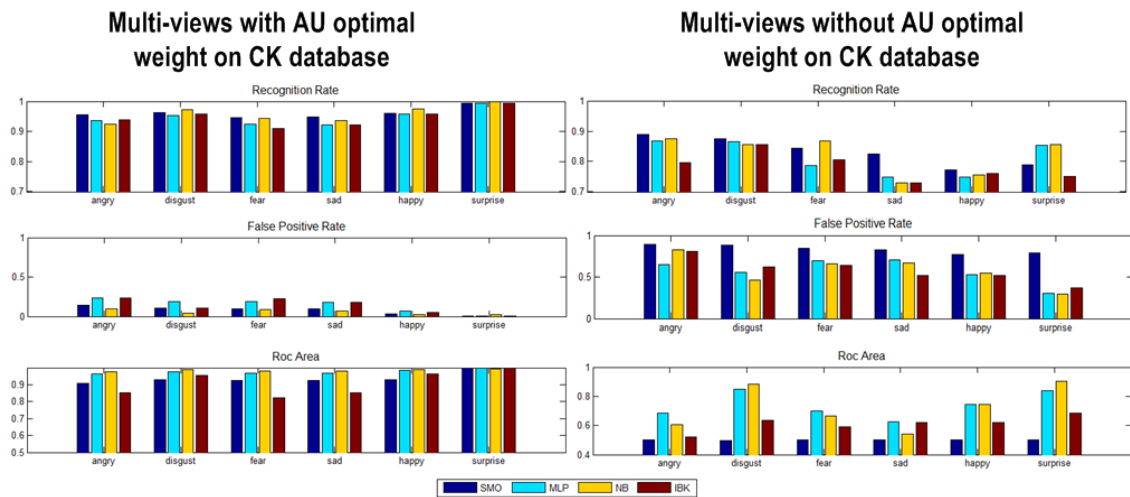


Figure 10. The results of the multi-layer AU fusion strategy on the CK dataset.

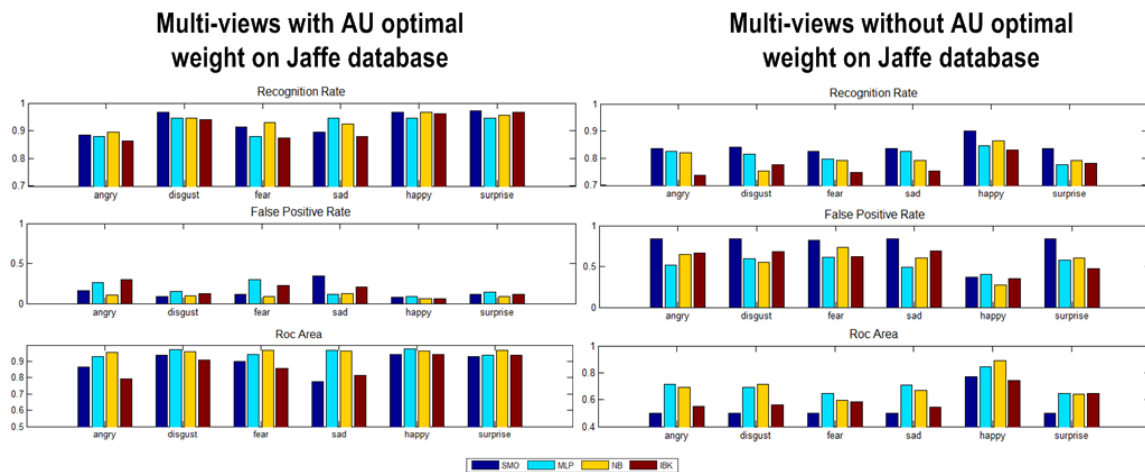


Figure 11. The results of the multi-layer AU fusion strategy on the JAFFE dataset.

The computational complexity of this approach is much less than for any kind of deep learning method; input image is 355 for CK and 187 for JAFFE. The classifiers on average took about 0.35 s (CK) and 0.21 s (JAFFE) to complete classification in a 10-fold experiment. The recognition results using the highest performance classifiers respectively on CK and JAFFE are shown in Table 10.

Table 10. The recognition results of the best performed classifier on CK and JAFFE.

Database	Mean Accuracy (%)	Angry	Disgust	Fear	Happy	Sad	Surprise
CK	96.62%	95.49% (SMO)	97.18% (NB)	94.65% (SMO)	97.46% (NB)	94.93% (SMO)	100.00% (NB)
JAFFE	94.63%	89.61% (NB)	96.72% (SMO)	92.90% (NB)	96.72% (SMO/NB)	94.54% (MLP)	97.26% (SMO)

7.5. Comparison with Several State-Of-The-Art Expression Recognition Methods

We compared our multi-level fusion based expression recognition method with some work in the current literature using either the CK or JAFFE dataset. Comparisons of recognition accuracies are shown in Tables 11 and 12. The accuracy rate achieved in our proposed approach is highlighted in bold.

Table 11. Comparison with existing methods on CK database.

Reference	Accuracy	Method
Cohen et al. (2003) [38]	73.2%	Geometric feature + tree-augmented-NB
Bartlett et al. (2005) [2]	89.1%	Gabor filter + Adaboost + SVM
Shan et al. (2009) [39]	88.9%	Boosted-LBP + SVM
Thiago et al. (2013) [26]	88.9%	Ensemble Gabor and LBP
Xu et al. (2009) [40]	89.1%	ASM + geometry feature + LSVM classifier
Jia et al. (2013) [29]	90.7%	Dynamic geometry feature + stacking
Tian Y et al. (2002) [3]	92.7%	Gabor wavelets + geometry with FACS
Chen-Chiung Hsieh et al. (2016) [31]	94.7%	dynamic face regions + A multi-class SVM
Deepak Ghimire et al. (2013) [23]	95.17%	Feature selective multi-class AdaBoost
Our proposed approach	96.62%	Multi-layer of optimally weighted AU
Ping et al. (2014) [41]	96.7%	Boosted deep belief network
Deepak Ghimire et al. (2013) [23]	97.35%	SVM on boosted features

Table 12. Comparison with existing methods on JAFFE database.

Reference	Accuracy	Method
Bashyal and Venayagamoorthy (2008) [42]	90.2%	Gabor filters + Learning Vector Quantization (LVQ)
Koutlasand Fotiadis (2008) [43]	92.3%	Gabor filters + artificial neural networks
Yu et al. (2013) [44]	85.7%	WLD + pool + SVM
Kyperountas M et al. (2010) [45]	85.9%	Gabor filters + Salient feature and reliable classifier selection (SFRCS)
Fu (2008) [46]	87.5%	Convolution Neural Network (CNN)
Jia et al. (2013) [29]	92.3%	Gabor filters + stacking
Ping et al. (2014) [41]	93.0%	Boosted deep belief network
Our proposed approach	94.63%	Multi-layer of optimal weighted AU

The results in Tables 11 and 12 show that our proposed fusion method outperforms the others in both CK and JAFFE datasets. Comparing with the deep belief network of Ping et al. [41], our proposed method achieves a comparative recognition rate only 0.1% less than the current prevalent deep learning based method. Ping et al. explored a deep learning method to find the discriminative feature and strong classifiers to recognize expressions. It achieves significantly high recognition results. However, it has a potential risk of over fitting based on the learning scheme. While our proposed method is based on fusion of the various typical features, it overcomes the shortage of dependency on datasets. Actually, the typical features have an advantage in representing the face, as they are derived from the facial analysis problem, compared with the data-driven based deep learning scheme. By making full use of the complementary contribution of different typical features, the generalization ability in representing expression action blocks is improved.

As listed in Table 11, our proposed approach also has a lower accuracy rate compared with the others, such as that in Deepak Ghimire et al., especially with SVM + boosting feature [23]. This is, to some extent, due to our feature and classifier selection. In this comparison, Deepak Ghimire et al. propose a well-designed dynamic feature, whilst we do not make full use of temporal features of facial expression, which is a significant attribution for expression recognition. In fact, our fusion solution has the great potential to improve the overall performance by reselecting the powerful well-designed features or classifiers. From this viewpoint, what we proposed has the advantage of being a flexible scheme to integrate the different sources, which has no limitation of similar mechanism. In general, our proposed method performs well in different datasets and displays generally good recognition performance.

8. Conclusions and Future Work

Facial expression recognition plays a significant role in intelligent human–computer interaction. In this paper, we present some performance analyses on the fusion ensemble based facial expression recognition method with multiple typical features and classifiers. In addition, we propose learning the contribution of AUs in each facial expression prediction by associating rule learning and optimizing the AU weights accordingly. In contrast to previous work, we aim to explore the fusion performance of AU features based on a kappa-error evaluation technique, and design an AU weight optimization idea based on an a priori algorithm. With the stacking framework, a multi-level calculating model is estimated. At the base-level, we engage complementary features and multiple classifiers to predict the probability of AUs' occurrence; at the meta-level, the model completed the facial expression classification with AUs' optimally weighted fusion according to the base-level output. Extensive experiments show that multi-layers of the AU descriptor obtain the discriminative ability for six expressions well, and the recognition performance improves by adjusting AU weights. Moreover, the fusion strategy with several feature descriptors has approximately a 12% higher recognition rate than any single feature descriptor, and it also verifies that each descriptor plays complementary roles in AU representation. Each AU has different contribution roles for different expressions. As a result, our fusion strategy, based on the typical features and classifiers, provides a more stable and better solution of facial expression recognition. The experiment results also prove that the fusion strategy not only outperforms the non-fusion strategy, but also achieves higher recognition results than some state-of-the-art facial expression methods, including one deep learning solution.

In the future, more diverse data, for example, with no lab capturing restriction, will be used for performance evaluation of the proposed fusion strategy. Furthermore, to improve the system's performance, we will analyse the movement of the muscles and global features, minimize the computation to satisfy the demand for automatic and real-time applications, and overcome the illumination and gesture changes in facial expression recognition.

Acknowledgments: This research is partially sponsored by Natural Science Foundation of China (Nos. 61370113, 91546111, 91646201, 61672070 and 61672071), Beijing Municipal Natural Science Foundation (4152005), Key Projects of Beijing Municipal Education Commission (No. KZ201610005009), the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions of 2014 (No. 067145301400), and the international cooperation seed grant from BJUT of 2016 (No. 007000514116520).

Author Contributions: Xibin Jia and Shuangqiao Liu conceived and designed the experiments; Shuangqiao Liu performed the experiments; David Powers conceived and designed the experiments specially for the fusion model and instructed Shuangqiao Liu analyzed the data; Xibin Jia and Shuangqiao Liu wrote the paper. David Powers especially helped to write and modify the early version of the paper. Barry Cardiff helped with modification of later version.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mehrabian, A. Communication without words. *Psychol. Today* **1968**, *2*, 53–56.
2. Bartlett, M.S.; Littlewort, G.; Frank, M.; Lainscsek, C.; Fasel, I.; Movellan, J. Recognizing facial expression: Machine learning and application to spontaneous behavior. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; pp. 568–573.
3. Tian, Y.; Kanade, T.; Cohn, J.F. Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 20–21 May 2002; pp. 229–234.
4. Lu, H.; Wang, Z.; Liu, X. Facial expression recognition using NKFDA method with Gabor features. In Proceedings of the Sixth World Congress on Intelligent Control and Automation (WCICA), Dalian, China, 21–23 June 2006; pp. 9902–9906.

5. Zhang, L.; Tjondronegoro, D. Selecting, optimizing and fusing ‘salient’ Gabor features for facial expression recognition. In Proceedings of the 16th International Conference, ICONIP 2009, Bangkok, Thailand, 1–5 December 2009.
6. Lyons, M.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding facial expressions with Gabor wavelets. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998; pp. 200–205.
7. Hu, Y.; Zeng, Z.; Yin, L.; Wei, X.; Zhou, X.; Huang, T.S. Multiview facial expression recognition. In Proceedings of the 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–6.
8. Dahmane, M.; Meunier, J. Emotion recognition using dynamic grid-based HoG features. In Proceedings of the 2011 IEEE International Conference Automatic Face & Gesture Recognition and Workshops (FG 2011), Santa Barbara, CA, USA, 21–25 March 2011.
9. Zhao, G.; Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [[CrossRef](#)] [[PubMed](#)]
10. Valstar, M.F.; Mehu, M.; Jiang, B.; Pantic, M.; Scherer, K. Metaanalysis of the first facial expression recognition challenge. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2012**, *42*, 966–979. [[CrossRef](#)] [[PubMed](#)]
11. Jain, S.; Hu, C.; Aggarwal, J.K. Facial expression recognition with temporal modeling of shapes. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 1642–1649.
12. Chen, J.; Shan, S.; He, C.; Zhao, G.; Pietikainen, M.; Chen, X.; Gao, W. WLD: A robust local image descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1705–1720. [[CrossRef](#)] [[PubMed](#)]
13. Gong, D.; Li, S.; Xiang, Y. Face recognition using the Weber Local Descriptor. *Neurocomputing* **2011**, *122*, 589–592.
14. Whitehill, J.; Bartlett, M.S.; Littlewort, G.; Fasel, I.; Movellan, J.R. Towards practical smile detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2106–2111. [[CrossRef](#)] [[PubMed](#)]
15. Yang, P.; Liu, Q.; Metaxas, D.N. Boosting coded dynamic features for facial action units and facial expression recognition. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–6.
16. Hui, J.; Wen, G. Analysis and application of the facial expression motions based on Eigen-flow. *J. Softw.* **2003**, *14*, 2098–2105.
17. Cohn, J.F.; Zlochower, A. J.; Lien, J.; Kanade, T. Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology* **1999**, *36*, 35–43. [[CrossRef](#)] [[PubMed](#)]
18. Senechal, T.; Rapp, V.; Salam, H.; Segulier, R.; Bailly, K.; Prevost, L. Combining AAM coefficients with LGBP histograms in the multi-kernel SVM framework to detect facial action units. In Proceedings of the Automatic Face & Gesture Recognition and Workshops (FG 2011), Barbara, CA, USA, 21–25 March 2011; pp. 860–865.
19. Eckert, M.; Gil, A.; Zapatero, D.; Meneses, J.; Martínez Ortega, J.F. Fast facial expression recognition for emotion awareness disposal. In Proceedings of the 6th International conference on Consumer Electronics, Berlin, Germany, 5–7 September 2016.
20. Vezzetti, E.; Marcolin, F.; Fracastoro, G. 3D face recognition: An automatic strategy based on geometrical descriptors and landmarks. *Robot. Auton. Syst.* **2014**, *62*, 1768–1776. [[CrossRef](#)]
21. Vezzetti, E.; Marcolin, F. 3D Landmarking in multiexpression face analysis: A Preliminary study on eyebrows and mouth. *Aesthet. Plast. Surg.* **2014**, *38*, 796–811. [[CrossRef](#)] [[PubMed](#)]
22. Essa, I.A.; Pentland, A.P. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 757–763. [[CrossRef](#)]
23. Ghimire, D.; Lee, J. Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines. *Sensors* **2013**, *13*, 7714–7734. [[CrossRef](#)] [[PubMed](#)]
24. Li, J.; Lam, E.Y. Facial expression recognition using deep neural networks. In Proceedings of the 2015 IEEE International Conference on Imaging Systems and Techniques (IST), Macau, China, 16–18 September 2015.
25. Kuncheva, L.I. A bound on kappa-error diagrams for analysis of classifier ensembles. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 494–501. [[CrossRef](#)]
26. Zavaschi, T.H.H.; Britto, A.S., Jr.; Oliveira, L.E.S.; Koerich, A.L. Fusion of feature sets and classifiers for facial expression recognition. *Expert Syst. Appl.* **2013**, *40*, 646–655. [[CrossRef](#)]

27. Koelstra, S.; Pantic, M.; Patras, I. A dynamic texture-based approach to recognition of facial actions and their temporal models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1940–1954. [[CrossRef](#)] [[PubMed](#)]
28. Valstar, M.; Pantic, M. Fully automatic facial action unit detection and temporal analysis. In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; p. 149.
29. Jia, X.; Zhang, Y.; Powers, D.; Ali, H.B. Multi-classifier fusion based facial expression recognition approach. *KSII Trans. Internet Inf. Syst.* **2014**, *8*, 196–212.
30. Cunningham, D.W.; Kleiner, M.; Wallraven, C.; Lthoff, H.H. Manipulating video sequences to determine the components of conversational facial expressions. *ACM Trans. Appl. Percept.* **2005**, *2*, 251–269. [[CrossRef](#)]
31. Hsieh, C.C.; Hsieh, M.H.; Jiang, M.K.; Cheng, Y.M.; Liang, E.H. Effective semantic features for facial expressions recognition using SVM. *Multimed. Tools Appl.* **2016**, *75*, 6663–6682. [[CrossRef](#)]
32. Ekman, P.; Friesen, W.V. *Facial Action Coding System*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
33. Margineantu, D.D.; Dietterich, T.G. Pruning adaptive boosting. In Proceedings of the 14th International Conference on Machine Learning, Nashville, TN, USA, 8–12 July 1997; Volume 97, pp. 211–218.
34. Kittler, J.; Hatef, M.; Duin, R.P.W.; Matas, J. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 226–239. [[CrossRef](#)]
35. Piatetsky-Shapiro, G. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*; AAAI Press: Menlo Park, CA, USA, 1991.
36. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules in large databases. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile, 12–15 September 1994; pp. 487–499.
37. John, G.H.; Langley, P. Estimating continuous distributions in bayesian classifiers. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montréal, QC, Canada, 18–20 August 1995.
38. Cohen, I.; Sebe, N.; Garg, A.; Chen, L.; Huang, T.S. Facial expression recognition from video sequences: Temporal and static modeling. *Comput. Vis. Image Underst.* **2003**, *91*, 160–187. [[CrossRef](#)]
39. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comput.* **2009**, *27*, 803–816. [[CrossRef](#)]
40. Xu, W.; Sun, Z.X. Facial expression recognition from image sequences with LSVM. *J. Comput. Aided Des. Comput. Gr.* **2009**, *21*, 542–548.
41. Liu, P.; Han, S.; Meng, Z.; Tong, Y. Facial Expression Recognition via a Boosted Deep Belief Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1805–1812.
42. Bashyal, S.; Venayagamoorthy, G.K. Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Eng. Appl. Artif. Intell.* **2008**, *21*, 1056–1064. [[CrossRef](#)]
43. Koutlas, A.; Fotiadis, D.I. An automatic region based methodology for facial expression recognition. In Proceedings of the IEEE Conference on Systems, Man and Cybernetics, Singapore, 12–15 October 2008; pp. 662–666.
44. Yu, K.; Wang, Z.; Zhuo, L.; Wang, J.; Chi, Z.; Feng, D. Learning realistic facial expressions from web images. *Pattern Recognit.* **2013**, *46*, 2144–2155. [[CrossRef](#)]
45. Kyperountas, M.; Tefas, A.; Pitas, I. Salient feature and reliable classifier selection for facial expression classification. *Pattern Recognit.* **2010**, *43*, 972–986. [[CrossRef](#)]
46. Fu, X.F. *Research on Binary Pattern-Based Face Recognition and Expression Recognition*; Zhejiang University: Hangzhou, China, 2008.

