Confidence Assessment On Eyelid and Eyebrow Expression Recognition

Javier Orozco
Computer Vision Center
Campus UAB, O-08193, Barcelona, Spain
jorozco@cvc.uab.es

Ognjen Rudovic
Computer Vision Center
Campus UAB, O-08193, Barcelona, Spain

F. Xavier Roca
Computer Vision Center
Campus UAB, O-08193, Barcelona, Spain

Jordi González
Institut de Robòtica i Informàtica Industrial
C. Llorens i Artigas 4-6, 08028, Barcelona, Spain

Abstract

In this paper, we address the recognition of subtle facial expressions by reasoning on the classification confidence. Psychological evidences have determined that eyelids and eyebrows are significant for the recognition of subtle facial expressions and the early perception of human emotions. This early perception results in a more complex problem, which requires a confidence assessment for any provided solution. Thus, traditional score-based classifiers (e.g. k-NN and NN) are not able to produce confident estimates. Instead, we first present five confidence estimators and a confidence classification assessment for Case-Based Reasoning (CBR). Second, we improve the expression retrieval from the database by learning the neighbourhood’s dimensions for the expected classification confidences. Third, we reuse the previous classified expressions and the confidence assessment to improve the classification achieved by k-NN. Fourth, we improve the database for generalization with new subjects by learning thresholds to minimize misclassification with low confidence, maximize correct classifications with high confidence and re-arrange misclassification with high confidence. The proposed system represents an effective contribution for both subtle expression recognition and CBR methodology. It achieves an average recognition of 97% ± 1% with a confidence of 96% ± 2% for expressiveness between 20% and 100%.

1. Introduction

Upper facial movements play an important role on the attempt of understanding human emotions and cognitive behaviours for Human Computer Interaction (HCI). Psychological studies establish the relevance of detecting and interpreting earlier when the facial emotion arises [7]. Eyelid and eyebrow expressions provide important non-verbal cues in human communication. In this context, irises, eyelids and eyebrows are known as gaze expression messengers to signal like, dislike, attentiveness, competence, dominance, credibility, intimacy and threat as well as regulating and initiating social interactions. Indeed, what Baron-Cohen et al [2] describe as the “language of the eyes” seems to be a significant channel for the communication of emotions and mental states. They showed that eyes could convey almost as much affective information as the whole face, see Figure 1. Likewise, Ekman and Friesen [8] applied these psychological foundations to computer science by describing specific facial muscle movements with a set of 44 action units (AUs), the Facial Action Coding System (FACS). Similarly, Paul Ekman [7] proposed six basic facial expressions categorizing the corresponding internal emotions.

The goal is to detect the facial actions and their intensity for a posterior classification. In this sense, several facial expression classifiers have been reported by using Neural Networks, Gabor Wavelets, Bayesian Networks, LDA, SVM, Neighbour Networks, etc [9]. All of them differ in robustness, accuracy, requirements for training and effectiveness. They report an average effectiveness of 86%, limited to the analysis of unseen subjects and intensity of expressions. Nonetheless, numeric score classifiers do not provide a confidence measure for the classification.

CBR has been used for facial expression recognition in a single subject [10], or user-assisted learning process of action units [13], without providing accurate solutions and robust processes for knowledge updating. Classification con-
confidence and emotion recognition are challenging tasks. In the literature, facial expression recognition is commonly addressed by classifying key features such as mouth, brows and wrinkles. However, subtle expressions are allocated near to the class boundaries representing a serious limitation for the aforementioned classifiers. Because of this, recognized expressions need a confidence assessment of the decided solution instead of the nearness to a belonging class. Toward this end, we use Case-Based Reasoning (CBR) that is more appropriate for complex and incomplete problem domains than eager methods, which replace the training data with abstractions obtained by generalization and which, in turn, require excessive amount of training data.

In this paper, we address the recognition, confidence-classification and expressiveness evaluation of subtle expressions based on eyelid and eyebrow information according to the six facial emotions proposed by Paul Ekman. First, an Appearance-Based Tracker (ABT) is used to extract the eyelid and eyebrow movements from image sequences of facial emotions. Second, the confidence of a given solution is measured based on a set of confidence estimators, which sample the class boundaries in a different manner according to the target. Third, we prove that by using the obtained confidence value, it is possible to achieve higher classification, even with cases isolated from the class manifold while allowing extending the solving capability. Finally, we learn the optimum confidence-classification threshold to minimize the low classification with high confidence while maximizing the high classification and high confidence. As a result, our proposed framework establishes significant contributions for facial expression recognition, emotion understanding, confidence assessment and case base knowledge updating.

This paper starts explaining the facial movement extraction by applying ABTs in section 2. The CBR system is presented in section 3. Here, we describe the facial expression recognition problem. In section 4, the confidence assessment is presented, which is based on five confidence estimators sampling the boundaries in a different manner. In section 5, we demonstrate the strongest properties of the confidence value and estimators to support the classification decision. Section 6 presents some experimental results and discussion. Adopting some iterative strategies, the database can be improved dealing with the misclassification. The paper finishes with some conclusions and an outline of future work in section 7.

2. Appearance-Based Tracking

In our study, we propose to encode the movements of eyebrows and eyelids by applying tracking methods. In this way, the current framework will be extensible to emotion analysis that is based on time series of facial expressions.

Figure 2. (a) The 3D deformable mesh allows modelling eyebrows and eyelids by applying ABT. Therefore, we obtain (b) the mesh adaptation for each input face.

Appearance-Based Tracking (ABT) is a robust and accurate method for extracting facial movements from image sequences. Three mean components allow applying ABT; face modelling, observation and registration of appearances. Therefore, it is possible to obtain the temporal state of eyebrows and eyelids for head poses near to frontal position.

2.1. Face Modelling

To encode the movements of eyebrows and eyelids, we use a 3D deformable facial model, which provides a simple process for mapping images into an appearance model texture. Therefore, deformable shape and texture assemble an active appearance model. The 3D face model is given by the three spatial coordinates of each vertex of the mesh, which is encoded by the 3xN matrix \( F \), where \( N \) corresponds to the number of vertices:

\[
F = F_0 + A \gamma,
\]

where \( F_0 \) is the standard configuration rendering the zero-neutral expression, the matrix \( A \) contains the facial animation modes. The vector \( \gamma \in [-1, 1] \) encodes the intensity of three facial actions namely inner and outer eyebrows, and eyelids as it is shown in Figure 2.(a).

Consequently, the ABT algorithm can provide the head pose estimation and facial actions, which together match with the face pose at every frame of the image sequence, Figure 2.(b). These parameters are encoded by a tracking vector \( \vec{q} \), which contains the head pose (three Euler’s angles, scale and translation) and the animation parameters. The tracking vector is as follows:

\[
\vec{q} = [\theta_x, \theta_y, \theta_z, s, t_x, t_y, \gamma_0, \gamma_1, \gamma_2] = [\alpha, \vec{\gamma}].
\]
provides eyebrow and eyelid parameters \( \vec{\gamma} \), that deform the mesh \( F \), see Figure 3.

### 2.2.1 Appearance Observation

Let us consider an appearance sequence \( X = \{ \vec{x}_0, ..., \vec{x}_t \} \) corresponding to the input image sequence \( I = \{ I_0, ..., I_t \} \). The appearance sequence \( X \) is modelled as a multivariate normal distribution, as well as the single row and column vectors are single normal distributions.

\[
P(\vec{x}_t|\vec{q}_t) = \prod_{i=0}^{l} \frac{e^{-(x_i-\mu_i)^2/2\sigma_i^2}}{\sigma_i \sqrt{2\pi}}
\]  

We assume that the appearance model summarizes the past observations under an exponential envelope, that is, the past observations are exponentially forgotten. When the appearance is tracked for the current input image, we can update the appearance and use it to track in the next frame. It can be shown that the appearance parameters, i.e. \( \vec{\mu} \) and \( \vec{\sigma} \) can be updated using the following filtering technique that is time efficient for estimating the Gaussian parameters over time with respect to previous estimations.

\[
\vec{\mu}_{t+1} = \omega \vec{\mu}_t + (1 - \omega) \vec{x}_t,
\]

\[
\vec{\sigma}_{t+1} = \omega \vec{\sigma}_t^2 + (1 - \omega) \vec{\chi}_t - \vec{\mu}_t)^2.
\]  

where \( \vec{\mu}_t \) and \( \vec{\sigma}_t \) are vectors of \( l \)-pixels according to the appearance’s size. Due to the binomial distribution approximation and the Central Limit Theorem for big sets of data, the Gaussian parameters gain significance after 50 frames. Learning factor \( \omega \) is \( 1/t \) until the \( 50^{th} \) frame, otherwise, it is a constant value. Here, we used a single Gaussian to model the appearance of each pixel. However, modelling the appearance with Gaussian mixtures can also be used on the expense of some additional computational load.

### 2.2.2 Appearance Registration

New pixel values are registered into the cumulative multivariate Gaussian \( X \), according to the state transition of the deformable model \( F \). Vector \( \vec{q}_{t+1} \), which embeds the shape deformation parameters, is estimated for the next frame by using an adaptive velocity model. This is a deterministic function defined by the last estimated vector and the increment vector \( \delta \vec{q}_t \) as follows:

\[
\vec{q}_{t+1} \approx \vec{q}_t + \delta \vec{q}_t
\]  

where \( \delta \vec{q}_t \) is the increment vector for the mesh deformation. The registration quality depends on estimating the optimal increment vector, which minimizes the error function between expected \( \vec{\mu} \), and estimated \( \vec{\chi} \) appearances.

This problem is usually solved by using an iterative first-order linear approximation based on the updated Gauss-Newton Iteration (GNI) algorithm [11, 12]. Therefore, starting from \( \vec{q} = \vec{q}_t \), we calculate the Jacobian matrix \( J \) for the appearance \( \vec{\chi}_t \) according to \( \vec{q}_t \) while minimizing the estimation error. Consequently, the increment vector is obtained and the corresponding appearance:

\[
\vec{q}_{t+1} = \vec{q}_t - \delta [J(\vec{q}_t)^T J(\vec{q}_t)]^{-1} J(\vec{q}_t) \vec{\chi}_t - \vec{\mu}_t
\]  

Thus, we apply a gradient descent method by partial differences, which is able to accommodate appearance changes while achieving precise estimations. Eyelid and eyebrow facial actions are encoded according to the Facial Animation Parameters (FAP) as continuous variables.

### 3. Case-Based Reasoning

#### 3.1. CBR Representation

Case-Based Reasoning imitates the way in which humans solve new situations by reusing previous solved problems [1]. In CBR approaches, the problems are called cases that follow a general structure of case descriptors and solution attributes. Hence, we understand that a case is a solved situation that is stored in the database, called the case-base, and a new testing data is called the problem.

The case attributes correspond to eyebrow and eyelid facial actions encoded by the vector \( \vec{\gamma} \), whose variables are...
3.2. Data Preparation

Given a sequence of facial action vectors $\mathbf{G} = \{\vec{\gamma}_0, ..., \vec{\gamma}_t\}$, from an image sequence $\mathbf{I}$, the $p$-norm is computed for each vector in the $L^p$-Space as follows:

$$||\vec{\gamma}||_p = \left( \sum_{i=1}^{p} |\gamma_i|^p \right)^{\frac{1}{p}} \text{ for } p = 0, ..., 6. \quad (7)$$

It can be seen in Figure 5 that all facial expressions tend to start near to the neutral expressiveness and drop once the peak is reached. Therefore, we modelled the $p-norm$ of all original neutral expressions as a singular normal distribution, $||\vec{\gamma}||_p^{neutral} \sim N(\mu_p, \sigma_p)$. Thus, it is possible to filter all facial expressions in the acceptance ratio of $(\mu_p \pm \sigma_p)$.

The expressiveness is calculated for the remaining frames in each expression cluster. Including both forced and subtle facial expressions from both databases, the expressiveness is defined as the percentage of the $p$-norm over its maximum variation inside of the cluster. Therefore, for a given facial expression $\vec{\gamma}$ belonging to the expression class $c$, the expressiveness $\epsilon$ is computed as follows:

$$\epsilon_c = \frac{||\vec{\gamma}||_p}{(||\vec{\gamma}||_{p_{\text{max}}} - ||\vec{\gamma}||_{p_{\text{min}}})} \times 100 \quad (8)$$

Consequently, facial expressions whose expressiveness is lower than 50% are considered subtle whereas an expressiveness greater than 80% indicates the peak of the expression and therefore a forced expression. Thus, we adapt the sequence-domain labels to picture-domain classification.

4. Confidence Assessment

CBR generalizes decision rules for a target problem according to local approximations. However, the efficiency of the solving process depends on the discrimination capabilities, mainly, nearby to the boundaries of clusters. Cheetham and Price [4] have pointed the importance of providing a confidence indication with any proposed solution. Some CBR approaches handle the solution quality by using as similarity criteria, the amount of cases that indicate confidence. Other contributions on confidence assessment highlight whether a feature has a negative or positive correlation with respect to the classification. Consequently, the development of CBR systems has increased the necessity of supporting the analysis of the case-base structure while providing solutions with a required accuracy and stability [4]. According to these authors, we are also concerned on proving the classification stability of a set of estimators. Therefore, a good confidence estimator remains stable while predicting on several tries of different classification models. The smaller bias from the average prediction indicates the most confident estimator.

4.1. Confidence Estimators

In order to evaluate confidence, we propose five similarity measures as confidence estimators. All the similarity measures are based on $k$-NN with the aim of identifying those cases that are close (i.e. with high similarity) to cases of the same class as the target case and are far (i.e. low similarity) from cases of a different class. The closer a target case is to cases from a different class, the higher the chance that the target case is laying near or at the decision surface. Whereas the closer a target is to other cases of the same class, the higher the chance that it is further from the decision surface.

We assess the confidence of the solution by computing five predictors based on $k$-NN classification and the Euclidean distance, at the Case Revise step [6]. Given a retrieved neighbourhood, we compute the confidence...
estimators for each solution proposed by the retrieved neighbours according to Relevant Neighbours (RN) and Irrelevant Neighbours (IN), as follows:

1. **Average RN-Nearness** indicates the average nearness similarity of the \( r \)-RN (\( r \) RN within \( k \)-NN) to the target that is assumed belonging to class \( c \).

\[
S_1(c) = 1 - \frac{\sum_{i=1}^{r} \| \vec{\gamma}_t - \vec{\gamma}_i \|}{r},
\]

where \( \vec{\gamma}_t \) is the target and \( \vec{\gamma}_i \), each one of the RN.

2. **Similarity Nearness Ratio** compares the Average RN-Nearness and Average IN-Nearness when assuming the target in the class \( c \).

\[
S_2(c) = \frac{1 - \frac{\sum_{i=1}^{r} \| \vec{\gamma}_t - \vec{\gamma}_i \|}{r}}{(2 - \frac{\sum_{i=1}^{n} \| \vec{\gamma}_t - \vec{\gamma}_i \|}{n} - \frac{\sum_{i=1}^{r} \| \vec{\gamma}_t - \vec{\gamma}_i \|}{n})}
\]

where \( n \) is the number of IN according to a supposed solution \( c \). \( S_2 \) provides both the nearness to boundaries and the boundary’s side location of the target.

3. **Similarity RN-Density** compares the average similarity inside of the RNs, inside of the INs and their average similarity to the target. Uniformly distributed RNs with the target and scattered INs, indicate high similarity.

\[
S_3(c) = \frac{1 - \frac{\sum_{i=1}^{r} \| \vec{\gamma}_t - \vec{\gamma}_i \|}{r^2} - \frac{\sum_{i=1}^{n} \| \vec{\gamma}_t - \vec{\gamma}_i \|}{n^2}}{r^2}
\]

4. **Average Similarity Norm** compares the average expressiveness of RNs to target’s expressiveness.

\[
S_4(c) = 1 - \frac{|r||\vec{\gamma}_t||p - \sum_{i=1}^{r} ||\vec{\gamma}_i||p|}{r}
\]

Grouping relationships are not always the stronger criteria of belonging to the cluster.

5. **Similarity Norm Ratio** compares the relative expressiveness of RNs and INs with respect to the target.

\[
S_5(c) = \frac{S_4(c)}{|S_4(c) + 1 - \frac{\sum_{i=1}^{n} ||\vec{\gamma}_t||p - \sum_{i=1}^{r} ||\vec{\gamma}_i||p|}{n^2}}
\]

All these similarity measures are standardized such that 0.0 indicates less confident solution and conversely 1.0 indicates the best confidence. Once the retrieved solutions are available, these five estimators are calculated as many times as different solutions are proposed by the \( k \) neighbourhood.

4.2. Classification Confidence

While it is evident that assessing confidence in classification is useful for CBR systems, it is also clear that it is not straightforward to assess confidence for multi-class domains. Cheetham and Price [4] describe different measures of confidence that can be applicable for \( k \)-NN classifiers. However, their confidence estimators contain similar, redundant information about nearness, which does not allow dichotomizing near classes and decision surfaces for complex domains.

The decision surface presents difficulties due to the nearness of the classes as well as the complexity of the case base. After computing the confidence estimators, the confidence agrees with the best-scored measures that propose the majority class. Therefore, the confidence \( \vartheta \), is the percentage of the highest estimators voting the majority solution. For example, given a target problem \( T \) and \( k = 5 \)-NN, let consider that the \( k \) nearest neighbours correspond to classes \([a, b, c, a] \): this is one of the typical tie situations where the majority of \( k \)-NN cannot solve. Now, consider that the computed confidence assessment results as follows:

<table>
<thead>
<tr>
<th>Table 1: Confidence assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_1 )</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>c</td>
</tr>
</tbody>
</table>

Consequently, the confidence \( \vartheta = 4/5 \) corresponds to the highest estimators voting the majority solution \( c \).

Two validation methods are performed to prove that these measures are statistically good confidence estimators: Leave-One-Out (LOO) and Actor-Fold-Cross-Validation processes for \( k \)-values between 2 and 20. LOO tests the system’s solving capability intra-class (person-dependent), whereas AFCV tests inter-class solution capability (person-independent).

The classification stability of the estimators and the confidence, Figure 6.(a), demonstrate that they are statistically good confidence estimators of the classification function. Classifying unseen data as the AFCV test reveals, we achieve better classification than the majority \( k \)-NN. The average classification bias is 2\%, which is equivalent to a stability of 98\%. Furthermore, it is possible to obtain the performance of the classification-confidence \( \vartheta \), see Figure 6.(b). The Confidence bias is 3\% while achieving an average classification-confidence of 94\%. Therefore, by comparing models with \( k \)-values between 2 and 20, it is possible to achieve an average classification of 73\% ± 2\%, with an average classification-confidence 94\% ± 3\%.

The previously obtained results are significant contributions for CBR, since they provide confidence estimations in the solving process. The classification-confidence \( \vartheta \) has the strengths of any statistical confidence estimator beyond a
similarity measure indicating confidence. On one hand, the case-retrieve step is improved by using a \( k \) value for an expected confidence in the solution of the case-reuse. On the other hand, it is possible to obtain solutions as Expressiveness-Confidence-Expression, \( <\epsilon, \vartheta, \lambda> \).

5. Confidence Classification

Due to the stability of classification-confidence solving target problems, we use the \( \vartheta \) value and the corresponding solution in the CBR-cycle. Next, we describe a CBR-cycle using confidence for classification.

**Case-Retrieve:** Given a target problem \( T \) described by the facial action \( \vec{\gamma}_t \) and an expected confidence \( \vartheta \), we apply \( k \)-NN according to Figure 6(b), to retrieve \( k \)-Neighbours.

**Case-Reuse:** The confidence \( \vartheta \) is computed for the target \( T \) according to all solutions from the retrieved neighbourhood. Adopting the estimators’ agreement (e.g. as in the Table 1), the confidence-classification is obtained for the target, \( T = [\vec{\gamma}, \epsilon, \vartheta, \lambda] \).

**Case-Revise:** The classification results are revised upon the relative difference in values of confidence and quality of target’s RNs. We denote by \( Q \), the quality, i.e. acquired average score from the agreed confidence estimators.

**Case Retain:** A solved target is retained when the case-base improves according to confidence, the quality on anyone of the confidence estimators. In fact, estimators \( S_1 \) and \( S_2 \) indicate how large and well delimited the clusters are, respectively. \( S_3 \) reveals the density of the class or seeds of the class. \( S_4 \) and \( S_5 \) provide expressiveness information indicating how flat is the class or its boundaries with respect to near classes.

6. Experimental Results

In our experiments, we have used two public databases, FGnet and MMI [14, 15]. They were gathered to apply a standardization of expression intensity, which extends the generalization capabilities. Both databases complete 25,000 images split in 200 sequences and 30 subjects performing the seven basic facial expressions.

We first extract the facial actions for a given image sequence, which constitute the data for the facial expression classification system, see again Fig. 3. Next, we construct the case structure for all facial expressions \( G = \{\vec{\gamma}_0, ..., \vec{\gamma}_t\} \), see again Fig. 4(b).

This database has been split in four samples of one thousand images, containing frames uniformly chosen from each sequence and expressiveness between 20% and 100%. Two of the four samples contain the same number of images per emotion, and the other two samples contain the same number of sequences per emotion (more frames from longer emotions). We performed three experiments:

1. **Confidence Assessment** for each of the four datasets, for \( k \)-values between 2 and 20, and performing both LOO and AFCV. 152 tries to learn the \( k \)-values for the optimum expected classification confidences.

2. **Confidence Classification Thresholds** are learnt by testing each dataset with \( k \)-values between 2 and 20, both LOO and AFCV, independently classification and confidence rates. In total, 152 tries by 9 classification thresholds and 9 confidence thresholds.

3. **Confidence Classification Updating** by iteratively using the previous learnt thresholds up to minimizing misclassification with low and high confidence.
The AFCV confidence threshold, which minimizes misclassification with high confidence (FP) while maximizing correct classifications with high confidence (TP).

(True Negatives (TN) and False Positives (FPs), respectively) and re-arranging misclassification with high confidence (False Negatives (FN)) and maximizing classification with high confidence (True Positives (TPs)).

Starting from zero – Confidence and zero – Classification, we iteratively identify TPs, FPs, TNs and FNs in order to calculate TPR vs. FPR of the CBR system. Consequently, we obtain optimum confidence-classification thresholds to retain in the case base those cases that maximize the TPR and reject those cases that do not minimize FPR, see Fig. 7.

Firstly, we deal with the TNs by extracting them from the case-base while iteratively computing LOO and AFCV. As soon as the case-base is free of TNs, we proceed to deal with FPs. In this case, we reconsider other proposed solutions in the case-retrieve step. It is possible that actors mix expressions in the performance, which makes the expression clusters to be closer and spread near to the decision surface. Finally, if a data is continuously detected as FP, we proceeded to extract it also from the case-base.

As a result, all cases in the case-base are updated with the average classification-confidence \( \vartheta \), after applying AFCV for \( k \)-values between 2 and 20, obtaining the final structure \( C = \langle \gamma, \epsilon, \vartheta, \lambda \rangle \).

Experimental results before and after training are summarized in the Table 2. Here, we use the learnt confidence-classification threshold to apply AFCV for \( k \)-values between 2 and 20. However, the optimal classification-confidence threshold allows obtaining better results.

Finding the best confidence-classification threshold, both the \( k \)-NN majority and the confidence-classification increase effectiveness while reducing error detections. For example, in Table 2, we can see that the obtained results are comparable to eager classifiers. In [5], Cohen et al. present a comparison of eager classifiers such as Neural Networks, Naïve Bayesian Networks and TAN Bayesian Networks. The best average classification rate that they reported was 86% by using TAN-BN.

Our system achieves an average of 97% ± 1% with a confidence of 96% ± 2%. Detailed experimental results are shown in Table 3, while testing with unseen subjects and expressiveness between 20% and 100%.

Although our system is accurate in deciding near to the decision surface, there are classes containing more subtle expressions that are harder to distinguish. This is the case of the Anger class, which presents a 12% of confusion with the Sadness class [3]. However, Sad is rarely confused with Anger, which allows concluding that it is possible to overcome the problem by retaining new Anger cases.

### 7. Conclusions

This paper has shown the strengths of combining ABT and CBR in a stand-alone system with capabilities to attach confidence estimations to its own solutions. The confidence estimators provide significant information for the facial expression recognition problem toward the real approach of human emotion understanding in video sequences.

Six basic expressions plus the neutral one compose the case base of the CBR system. This is a multi-dimensional domain since classes are spread in the space forming seeds inside or surrounded by other clusters. However, it has been proved that the five confidence estimators efficiently provide information of the decision surface. Thus, it is possible to distinguish subtle facial expressions allocated at the boundaries of expression clusters.

We demonstrated that our confidence measures are good confidence estimators and stable classifiers. They increase the classification efficiency when including subtle expressions. Moreover, they were extended for maintenance policies, contributing to the four steps of the CBR cycle; retrieving the optimal \( k \)-neighbourhood for expected classification error and confidence. Reusing previous solved cases to provide confident solutions. It is possible to revise the
solutions based on updated knowledge of the decision surface and the case base. By learning the optimum confidence thresholds, it is possible to retain or reject new solved cases. Experimental results have proven that the classification confidence improves the facial expression recognition rather providing additional information for the solution. After training, the system confidently recognizes facial expressions of unseen actors, can delete cases with low confidence and retain those with high confidence. As a result, the system achieves an average classification rate of 93% ± 1% with an average confidence of 96% ± 3%. The solutions are composed by expressiveness, confidence and expression class (ε, θ, Λ).

Ekman stated that subtle expressions are highly similar to the neutral expression, since they early reveal the raising of the expression [7]. This is detectable when including eye information as also mentioned by Baron and Cohen [2]. We have proven that subtle expressions, which are allocated near to the class boundaries, are confidently recognized. The low confidence is an evident weakness when the boundary of the class is not well delimited. It is possible to overcome this situation by analysing separately the confidence estimators. They may indicate to retain cases for specific areas of the case base.

We presented four contributions that represent important advances for facial expression recognition in dynamical environments, which include contextual information, different people’s cultures and expressivity. CBR and confidence assessment allow improving the recognition capabilities by making decision rules to increase the classification confidence. The low cost of knowledge acquisition and the strengths of dichotomizing the decision boundaries in multi-class and multi-modal domains, make CBR suitable for adaptive learning systems.

Acknowledgements This work is supported by EC grants IST-027110 for the Hermes project and IST-045547 for the Vidi-video project, by the Spanish MEC under projects TIN2006-14606 and CONSOLIDER INGENIO 2010 (CSD2007-00018). Jordi Gonzalez also acknowledges the support of a Juan de la Cierva Postdoctoral fellowship from the Spanish MEC.

References


<table>
<thead>
<tr>
<th>Emotion</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fears</th>
<th>Happy</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>82.09</td>
<td>3.33</td>
<td>0.00</td>
<td>1.22</td>
<td>0.48</td>
<td>2.22</td>
<td>0.00</td>
<td>94.93</td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>93.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>96.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>91.54</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>96.34</td>
<td>0.48</td>
<td>0.00</td>
<td>0.00</td>
<td>94.15</td>
</tr>
<tr>
<td></td>
<td>4.48</td>
<td>3.33</td>
<td>0.00</td>
<td>2.44</td>
<td>99.04</td>
<td>2.22</td>
<td>0.00</td>
<td>99.13</td>
</tr>
<tr>
<td></td>
<td>11.94</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>95.56</td>
<td>2.13</td>
<td>93.78</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>97.87</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>95.32</td>
</tr>
</tbody>
</table>

Table 3. Personal-Independent CBR Facial Expression Recognition by Assessing Confidence.