Understanding Facial Expressions

Feature Extraction from Facial Landmarks

Mirco Senes
Università degli Studi di Genova DIBRIS - 2024/2025

Abstract— **This project investigates facial expression recognition through geometric analysis of facial landmarks, focusing on four emotions: neutral, happiness, anger, and surprise. Using the KDEF dataset, a compact set of normalized features is extracted from 68 facial landmarks, emphasizing key facial regions such as the eyes, eyebrows, and mouth. The features are designed to capture expression-specific variations while minimizing identity-dependent traits. Performance is evaluated through pairwise distance comparisons, clustering, and supervised classification. Results show high accuracy for happiness and improved discrimination of anger and surprise after the introduction of a novel feature based on eyebrow–eyelid distance. While the approach performs well in controlled conditions, challenges remain under real-world variations, suggesting directions for future improvement.**

I. Introduction

Facial expression recognition is a key task in computer vision, with applications ranging from human-computer interaction to behavioral analysis. This project focuses on extracting a compact and robust feature vector from facial landmarks to distinguish between four specific expressions: **neutral**, **happiness**, **anger**, and **surprise**.

Using the KDEF dataset, we detect 68 facial landmarks through a shape prediction model and compute normalized geometric features that describe the spatial configuration of facial components. Particular attention is given to selecting features that reflect expression-related variations rather than individual-specific traits.

The analysis is structured in several stages: it begins with a comparison of individual subjects using feature vector distances, followed by experiments on a larger set of images from the KDEF dataset. To evaluate generalization, the method is also tested qualitatively on a small set of personal facial images captured in different conditions.

The remainder of this report is structured as follows. We begin by describing the methodology, including the landmark detection process and the extraction of normalized geometric features. This is followed by an overview of the KDEF dataset, highlighting its composition and relevance to facial expression analysis. We then present the experimental analysis, which includes subject-level comparisons, clustering results, and a supervised classification phase. The report concludes with a discussion of the findings.

II. Landmark Detection and Feature Extraction

This section describes the process of detecting facial landmarks and extracting geometric features used to represent facial expressions in a compact and interpretable form.

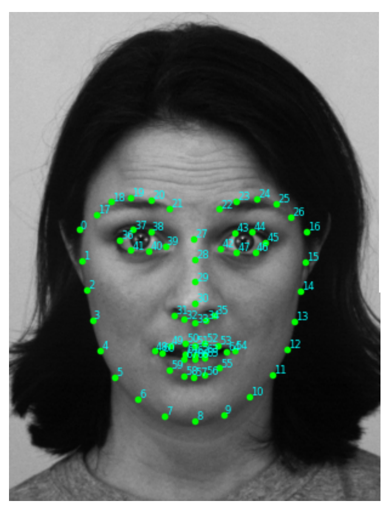
A. Landmark Detection

Facial landmark detection plays a central role in the analysis of facial expressions, as it enables the localization of key facial regions. In this project, landmark detection was performed using the *Dlib* library, which provides reliable tools for identifying facial reference points from images.

We focus on the standard 68-point landmark model, where each landmark corresponds to a specific location on the face, such as the corners of the eyes, the edges of the lips, or the contour of the jawline. These landmarks form a sparse structural representation of the face and serve as the basis for feature computation.

An example of this landmark configuration, with all points numbered according to their index, is shown in *Figure 1*.

The coordinates of these points are extracted from each image and used to compute a set of normalized geometric features, which will be described in the following subsection.



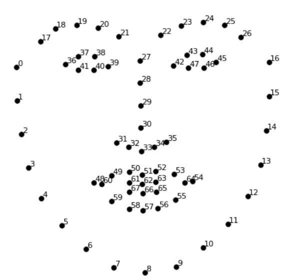


Figure 1. Visualization of the 68-point facial landmark configuration used in this project. Each point corresponds to a fixed reference location on the face.

B. Feature Extraction

Once the facial landmarks have been obtained, the process moves to the feature extraction phase. This step involves computing a series of normalized geometric measurements based on specific landmark indices, selected to capture meaningful facial variations associated with expressions.

The first operation is the calculation of the interocular distance, defined as the Euclidean distance between the outer corners of the eyes (landmarks 36 and 45). This value serves as a normalization factor, ensuring that all subsequent features can be compared across subjects regardless of face size or proportions.

The selected features were carefully chosen with the goal of minimizing the influence of identity-specific facial traits and maximizing sensitivity to expression-driven variations.

The idea is to focus on features that, once normalized, are not strongly dependent on the person's facial structure, but instead vary consistently with emotional expressions. For example, absolute nose width or jaw size were avoided, while features like mouth curvature or eyelid spacing were prioritized.

The computed features include:

- **Left and right eye height**: Average vertical distances between the upper and lower eyelids (37–41, 38–40 for the left eye; 43–47, 44–46 for the right eye), which reflect eye openness.
- **Mouth width**: Horizontal distance between landmarks 48 and 54 (mouth corners).
- **Brow gap**: Horizontal distance between the inner ends of the eyebrows (21 and 22), which reflects eyebrow positioning and movement.
- **Smile indicators**: Distances from the mouth corners (48 and 54) to the respective nose points (31 and 35), helpful in detecting the upward tension of a smile.
- **Mouth angle**: The angle formed at landmark 51 (upper lip center) between vectors pointing to the mouth corners (48 and 54), indicating mouth curvature or openness.

These features were selected to strongly characterize the "happiness" expression, characterized by mouth widening and curvature. This is captured through measurements such as mouth width, mouth angle, and distances from the mouth to the nose.

In contrast, to discriminate between "anger" and "surprise", which are often confused due to similar mouth configurations, the focus shifts toward eye openness and brow gap. Surprise is typically marked by raised eyebrows and wide eyes, whereas anger is characterized by lowered eyebrows and slightly narrowed eyes. These differences are reflected in the eye height measurements and the distance between the inner eyebrows.

The final output is a compact and normalized feature vector.

During testing, it became evident that some expressions, particularly neutral, anger and surprise, were not always clearly distinguishable using the initial feature set. As a result, an additional feature was introduced: the vertical distance between the eyebrows and upper eyelids, which proved helpful in enhancing discrimination. Details and impact of this modification are discussed in Section IV.

III. KDEF Dataset

The Karolinska Directed Emotional Faces (KDEF) dataset is a widely used collection of facial images specifically designed for research on emotion recognition. It contains high-quality pictures of individuals displaying a range of facial expressions such as happiness, anger, surprise, and neutral.

This dataset was selected because it offers many subjects, each exhibiting multiple expressions. This variety allows for more robust conclusions and supports a more accurate and generalizable analysis of facial expression features across different identities.

However, the dataset also presents some limitations. All images were acquired under highly controlled conditions, with consistent subject-to-camera distance, uniform lighting, and high resolution. Although this setup enables precise and reliable landmark detection, it does not capture the variability found in real-world scenarios, such as changes in lighting, distance from the camera, or image quality. Further information on the dataset is available at [2].

IV. Experimental Analysis

To evaluate the effectiveness of the extracted facial features in capturing emotional expressions, we designed a four-phase experimental analysis.

The first phase examines the robustness of the features to inter-subject variability using a minimal and controlled subset of KDEF images.

The second phase extends the evaluation to a broader portion of the KDEF dataset through clustering and dimensionality reduction techniques, assessing the ability of the features to distinguish between different emotional categories across multiple subjects.

The third phase tests the method on a set of personal images acquired under less controlled conditions, to qualitatively assess its generalization capability beyond the training data domain.

Finally, a fourth phase introduces a supervised classification stage, designed to quantitatively assess the discriminative power of the proposed features.

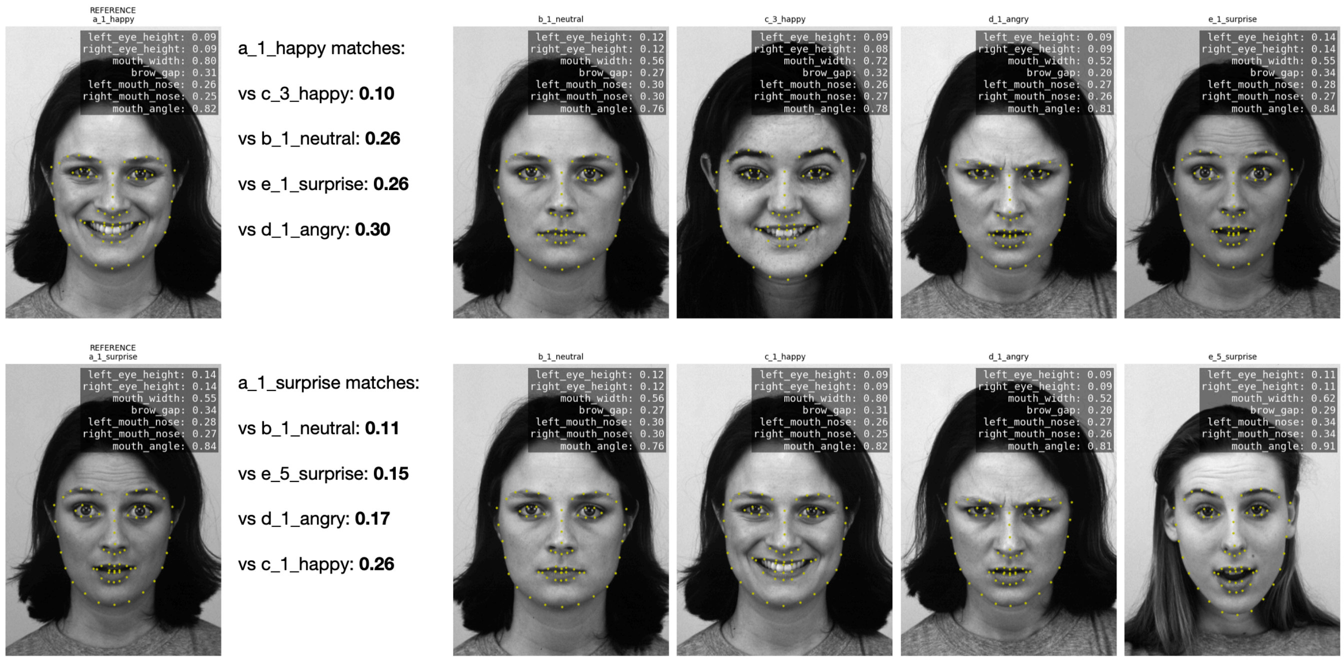


Figure 2. Each reference expression is compared both with the same expression performed by a different person and with different expressions performed by the same person. Two representative results are shown: one for happiness and one for surprise, along with their corresponding distance values.

A. *Controlled Comparison on a Minimal Set of Images*

As previously mentioned, selected features are designed to be expression-sensitive while minimizing the influence of person-specific facial traits.

To validate this, we conduct a controlled test using a small subset of images:

- One individual is compared with themselves showing different expressions.
- That same individual is also compared to a different person showing the same expression.

Ideally, the feature vectors of two different individuals showing the same expression should be closer than those of the same person showing different expressions.

The results show that in three out of four test cases, the correct match corresponds to the minimum Euclidean distance between feature vectors. *Figure 2* presents two representative examples: the top row illustrates the most effective case, while the bottom row illustrates the least effective case.

In the first row, the reference image shows a happy expression. Its closest match is correctly identified as another happy face (distance 0.10), while the distances to neutral, surprise, and angry expressions are significantly higher (0.26, 0.26, and 0.30 respectively). This suggests that happiness may be the most distinguishable emotion, with low intra-class distances and clearly higher inter-class distances.

In contrast, the second row presents a surprise expression as reference. Although the surprise sample is correctly matched at a relatively low distance (0.15), the neutral face shows an even smaller distance (0.11), indicating confusion between the two. This suggests that the current feature set may not yet capture the distinctive characteristics of all expressions effectively.

However, this might be an isolated case. For this reason, we broaden the evaluation across a larger portion of the dataset to confirm the consistency of the observed patterns before introducing any modifications to the selected features.

B. *Clustering on More Data*

Following the current implementation, a broader analysis was conducted using a set of 30 subjects, each exhibiting all four target expressions.

Feature vectors were extracted from all samples, and clustering was performed using the k-means algorithm with k=4, reflecting the four target emotion categories: happy, neutral, angry, and surprise.

For visualization purposes, Principal Component Analysis (PCA) was applied to project the high-dimensional feature space into two dimensions, allowing clearer interpretation of the clustering results.

As shown in *Figure 3*, the clustering results reveal a clear pattern: happy expressions (H) form a well-defined and consistent group, further supporting earlier observations that happiness is effectively captured and readily distinguishable using the current feature set.

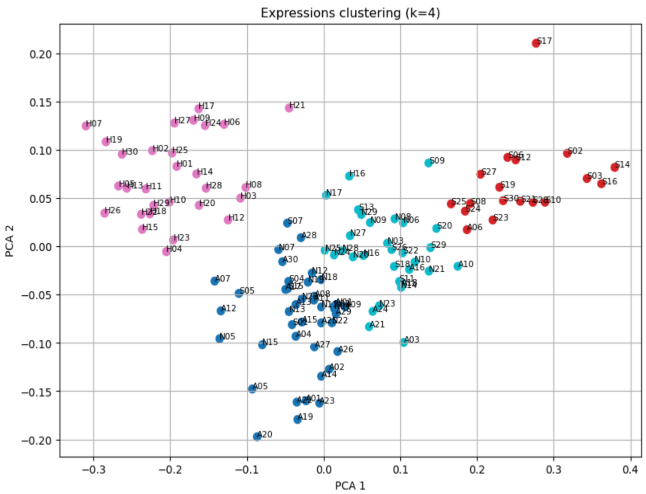


Figure 3. PCA plots after k-means clustering using the first set of features. Each point is labeled with the first letter of the actual emotion and the subject ID.

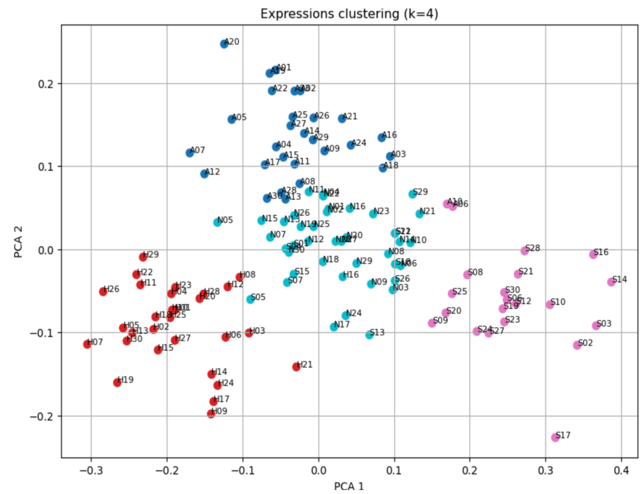


Figure 4. PCA plots after k-means clustering after improving the feature set. Each point is labeled with the first letter of the actual emotion and the subject ID.

Surprise expressions (S) also tend to group together, although with slightly less separation.

In contrast, neutral (N) and angry (A) expressions frequently overlap, often appearing in mixed groups along with some surprise samples.

This suggests that these emotions are harder to distinguish based on the current representation—an outcome that aligns with the known visual similarity between neutral and angry faces, as some individuals naturally appear angry even when neutral.

These findings motivate a refinement of the feature set to better separate these more ambiguous expressions.

C. Feature Refinement: Eyebrow-Eyelid Distance

One of the most intuitive visual cues distinguishing an angry face from a neutral one is the frowning of the eyebrows. However, computing features such as the slope or curvature of the eyebrows can be unreliable, as these characteristics vary significantly between individuals due to differences in eyebrow shape and natural positioning.

To address this, I introduced a more robust and consistent feature: the vertical distance between the upper eyelid and the eyebrow. This measure is less sensitive to personal differences in eyebrow shape and better reflects the eyebrow movement typically associated with anger— such as lowered brows positioned closer to the eyes.

The new feature uses the following landmarks:

- **Left eye and eyebrow**: top of the eyebrow (landmark 20) and the average position of the upper eyelid (landmarks 37 and 38)
- **Right eye and eyebrow**: top of the eyebrow (landmark 23) and the average position of the upper eyelid (landmarks 43 and 44)

Interestingly, this feature is also helpful in distinguishing surprised expressions, which tend to increase this distance as the eyebrows rise above widened eyes.
Therefore, it contributes to improving the separation of both angry and surprise expressions within the feature space.

This improvement is confirmed by *Figure 4*, where the emotion clusters appear more internally consistent than in the previous configuration, with each group now exhibiting more accurate and coherent emotion labels.

Happy expressions (H) form a clean and well-defined group, consistently composed of happy samples, as already observed in previous clustering attempts. Surprise expressions (S) are also grouped more reliably, with fewer mismatches, though some overlap with a couple of angry samples remains.

Angry expressions (A) are now better recognized, reducing confusion with neutral faces. Neutral expressions (N), although still somewhat mixed—particularly with surprise and occasional angry or happy samples—are more cohesively grouped compared to earlier results.

D. Generalization on Personal Data

In a final stage, the method was also tested on a small set of personal photos to further validate its effectiveness.

Additional experiments were conducted by partially occluding the face; however, this prevented correct face detection, making it impossible to extract facial landmarks.

Tests were also carried out with the subject positioned farther from the camera. In these cases, the change in perspective likely introduced geometric distortions in facial shape, affecting the relative distances between landmarks and leading to inconsistent results compared to close-up images.

The most reliable outcomes are shown in *Figure 5*, demonstrating that under conditions like those of the KDEF dataset, the method performs satisfactorily—even in the presence of slight variations in head orientation or lighting direction.

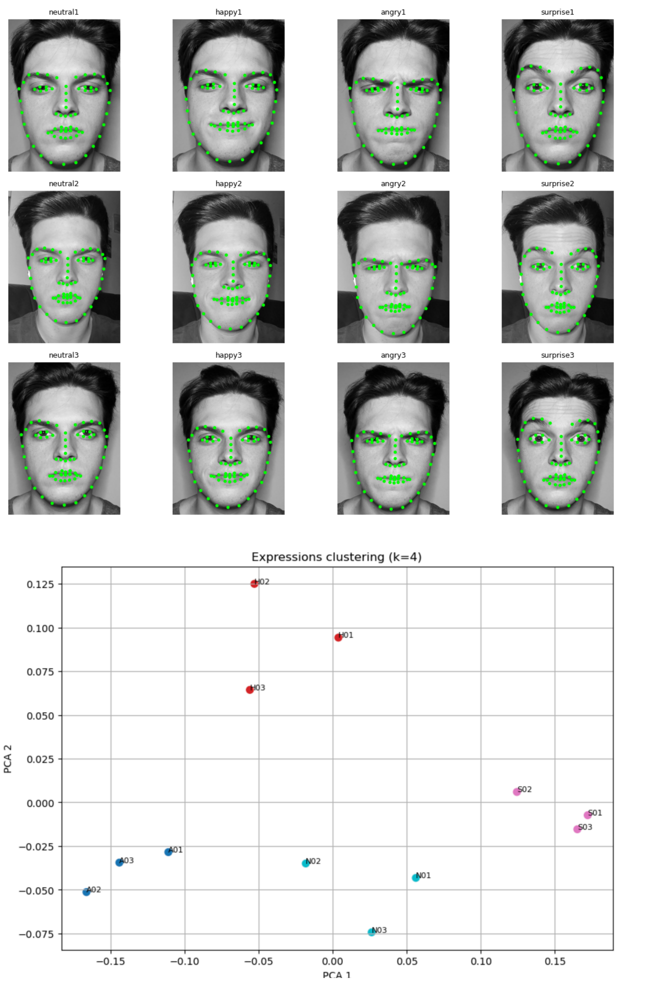


Figure 5. Sample images from the personal dataset and the corresponding clustering results.

E. Supervised Classification

To further assess the quality of the extracted features, a supervised classification stage was introduced using a kernel based SVM classifier.

The model was trained on the same KDEF set of images previously used for the clustering analysis (30 subjects across 4 expressions) and achieved an overall accuracy of 93%.

The classifier was also tested on the same personal image set used earlier. Despite the different acquisition conditions, it maintained an accuracy of 83%, confirming the method's potential to generalize beyond standardized datasets.

V. Conclusions

This project aimed to develop a feature-based method for recognizing facial expressions using geometric properties extracted from facial landmarks. By focusing on a compact and interpretable set of normalized features, we wanted to emphasize expression-driven variation while minimizing identity-specific traits.

We selected specific geometric features related to the eyes, mouth, and eyebrows, as these facial regions show the most significant movement during expressions.

Through successive testing, the method demonstrated a strong ability to identify happiness, which consistently formed well-separated clusters and showed low intra-class distances.

Other emotions—particularly anger, surprise, and neutrality—initially proved more difficult to distinguish. To improve this, we introduced an additional feature: the vertical distance between the upper eyelid and eyebrow. This addition helped reduce inter-cluster mixing, resulting in more coherent and better-structured emotion groupings, as observed in the updated PCA visualizations.

The results indicate that certain expressions are more readily distinguishable than others, with some consistently forming well-defined and coherent feature clusters. Feature refinement contributed to a general improvement in expression separability, although partial overlap between some categories persists. Testing on personal images confirmed the method's potential generalizability, but performance dropped significantly under occlusion or extreme changes in perspective, highlighting limitations in uncontrolled settings.

The method has shown promising results, but the findings remain preliminary. Future work should focus on enhancing robustness in real-world conditions, improving the distinction between similar expressions, and investigating temporal features—capturing expression changes over time—as well as learned features to achieve higher accuracy.

All code developed for this project is available in the provided Python notebook [1].

References

[1] Mirco Senes. *Facial Expression Recognition from Landmarks – Project Code and Notebook*. GitHub Repository: https://github.com/mircosenes/understanding_facial_expressions

[2] Karolinska Directed Emotional Faces (KDEF) Dataset. Available at: https://www.kdef.se/