Face and Gesture Analysis: Emotions

Alejandro Fernández Alburquerque, 242349
Andreu Garcies Ramon, 240618

February 2024

NOTE BEFORE READING: we provide the code for this practice in the GitHub repository of the project: `https://github.com/Alejandro-FA/UPF-Face-and-Gesture-Analysis`

1 Introduction

In this project, we explore the circumplex model of affect as one of the predominant theories of emotional analysis through facial expressions. The goal is two-fold: to understand the mathematical details of multidimensional scaling and to grasp the intrinsic challenges of emotion classification. We will also perform statistical analysis to determine the validity of our results and compare them with those obtained in the literature.

2 Literature review

Emotion classification is a significant area of study in computer vision, and it has gained even more relevance with the expansion of Machine Learning and Deep Learning techniques [Pantić, 2009]. After more than 50 years of studies, two main paradigms have emerged: the idea that there exists a discretized set of universal emotions (also known as "basic emotions") that originate from separate neurophysiological systems, and the circumplex model of affect, a dimensional understanding of the matter that states that there are only two neurophysiological systems from which all emotions appear [Posner et al., 2005].

Several researchers hypothesize that facial expressions are an evolution result [Du and Martínez, 2015]. For example, facial expressions of disgust could provide an evolutionary advantage to protect ourselves from germ contamination, while other expressions could be related to social interaction. On the other hand, the circumplex model of affect stems from the fact that humans tend to recognize emotions as ambiguous and overlapping experiences rather than discretized ones. Another critical consideration is that humans are not very good at labeling emotions without more context than a facial image. Fortunately, we are much better at comparing the similarity between two images, which we can use to build a (subjective) similarity matrix. Thanks to multidimensional scaling, we can convert these similarity matrices into a continuous 2D space of emotions, in which each of the subjective emotions we experience is specified as a linear combination of two independent systems (often called valence and arousal).

As mentioned above, the consensus is to identify two dimensions when applying multidimensional scaling to emotion classification with facial expressions. Whether more dimensions should be chosen is still debatable, though, because multidimensional scaling does not allow us to find the original dimensionality of the data (we only have their distances) [Izenman, 2008]. In this project, we will perform some fundamental statistical analysis to find the reason behind this choice.

3 Methodology and mathematical concepts of multidimensional scaling

As mentioned in the previous section, identifying discrete emotions is complex and unreliable. For this reason, researchers have proposed a dimensional paradigm to explain the nature of emotions, and they have used multidimensional scaling to obtain the intrinsic dimension of the "emotions space".

In this section of the report, we will provide the explanations required to understand multidimensional scaling, which we use to extract the meaningful basis of the intrinsic dimensions of emotions.

Multidimensional Scaling (MDS) is a statistical technique for analyzing the pairwise dissimilarities between a set of observations [Izenman, 2008]. Given a two-way dissimilarity matrix of emotions, MDS aims to

obtain a higher dimensional space where we can represent emotions as points in the space. A characteristic of MDS is that the number of dimensions of the original data is unknown. In our case, we do not know how many dimensions we need to describe any emotion accurately.

There exist different methods for MDS. The one that we have applied and the one that we will explain is known as classical scaling, which is equivalent to principal component analysis (PCA) if the distance metric used is the Euclidean distance.

In order to be able to apply MDS, we need a distance matrix that holds the distance between every pair of samples of our data. These distances do not need to be Euclidean distances, but for the moment, we will assume so. If we define d_{ij} as the Euclidean distance between sample i and sample j, we can build a matrix of distances with the following equation:

$$dij^2 = \|\mathbf{X}_i\|^2 + \|\mathbf{X}_j\|^2 - 2\mathbf{X}_i^T\mathbf{X}_j \tag{3.0.1}$$

We will now define a matrix B whose elements are $b_{ij} = \mathbf{X}_i^T\mathbf{X}_j$. Therefore, equation (3.0.1) can be written in terms the entries of matrix B as

$$d_{ij}^2 = b_{ii} + b_{jj} - 2 \cdot b_{ij} \tag{3.0.2}$$

The problem is that we do not have access to $\mathbf{X}_i$ and $\mathbf{X}_j$. However, if we define a matrix A, with $a_{ij} = -\frac{1}{2}d_{ij}^2$, and a double centering matrix $H = I_n - \frac{1}{n}\mathbf{1}_{n\times n}$, where I_n is the identity matrix and n is the number of images that we are comparing, then it can be shown that matrix B can be computed as

$$B = HAH \tag{3.0.3}$$

At this point, to obtain the representation of the different emotions represented in our dataset of images in the embedding space, the problem has been reduced to compute the eigendecomposition of matrix B.

$$B = \underbrace{\mathbf{X}^T \cdot \mathbf{X}}_{\text{unknown}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \left(\mathbf{V}\mathbf{\Lambda}^{1/2}\right) \cdot \left(\mathbf{\Lambda}^{1/2}\mathbf{V}^T\right) = \mathbf{Y}\mathbf{Y}^T \tag{3.0.4}$$

with $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_t\}$ the first $t < n$ positive eigenvalues of B and $\mathbf{V}$, its eigenvectors. Note that from equation (3.0.4), we are able to obtain $\mathbf{X}$ which was unknown from $\mathbf{Y}^T$ ($\mathbf{X} = \mathbf{Y}^T$), which we have been able to compute as $\mathbf{Y}^T = \mathbf{\Lambda}^{1/2}\mathbf{V}^T$.

3.1 Multidimensional Scaling for non-Euclidean distances

As mentioned previously, MDS can also be applied to other distance metrics besides the Euclidean distance, and the same procedure can be followed. However, matrix B might no longer be positive definite, which will cause negative eigenvalues to appear. Consequently, the corresponding eigenvectors will have complex coordinates, and complex coordinates do not have any meaningful interpretation in the realm of MDS. One common approach is to discard negative eigenvalues. Recall that a matrix M is positive definite whenever $\langle x, Mx\rangle > 0, \forall x \in \mathbb{R}^n$.

4 Results

4.1 Creating a similarity matrix

Our dataset consists of 24 images distributed in 8 different categories: *anger, boredom, disgust, friendliness, happiness, laughter, sadness* and *surprise*. Figure 1 shows some of the images used. In order to build the similarity matrix, we have compared each possible combination of two images and granted them a similarity score between 0 and 9 based on the criterion from Table 1. Figure 2 shows the resulting similarity matrix. We will give more details on its interpretation in section 4.1.1.

To apply MDS, explained in section 3, we need to convert the similarity matrix (which we will name as C) into a dissimilarity matrix, D using equation (4.1.1)

$$d_{ij} = \sqrt{c_{ii} - 2 \cdot c_{ij} + c_{jj}} \tag{4.1.1}$$

where c_{ij} is the similarity score between pictures i and j, and d_{ij} are the resulting distances between pictures i and j.

Figure 1: Sample images for each emotion. From left to right, the emotions are anger, boredom, disgust, friendliness, happiness, laughter, sadness and surprise.

Score	Criterion
0-1	emotions not similar at all
2-3	not very similar emotions
4-5	more or less similar emotions
6-7	quite similar emotions
8-9	same emotion

Table 1: Emotion Similarity Scores and Criteria

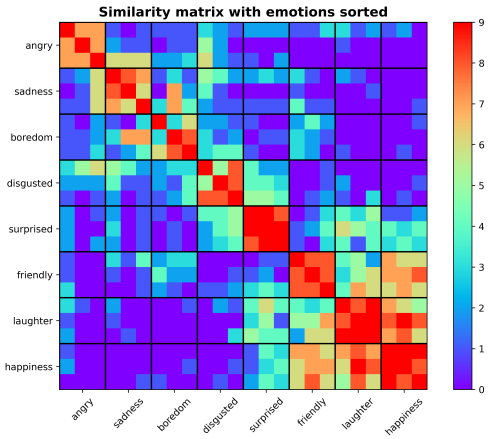

Figure 2: Similarity matrix

4.1.1 Verifying the reliability of the similarity matrix

Apart from a similarity matrix, we also have a consistency matrix, which we will use to determine the reliability of our results. The consistency matrix is a partially filled similarity matrix. While labeling the similarity between each pair of images, we were presented with some repeated pairs. The result of labeling the repeated pairs is stored in the consistency matrix, which we can then compare with the similarity matrix to ensure that the same pair of facial expressions is evaluated similarly at different points in time.

We have used the mean absolute error to determine the reliability of our consistency matrix, which is defined as follows:

$$E = \frac{1}{n^2} \sum_{i}^{n} \sum_{j}^{n} |X_{ij} - Y_{ij}| \tag{4.1.2}$$

Since the consistency matrix has no value for all its elements, we only compute the mean absolute error for the filled entries. The error obtained can be seen in table 2.

In order to determine the quality of the error obtained, we performed a bootstrap analysis. The null hypothesis is that the values of the consistency matrix are not correlated with the values of the similarity matrix, and the alternative hypothesis is that they are positively correlated. In order to see if we can reject the null hypothesis, we have compared the error between our consistency matrix and the similarity matrix with the error of 1000 randomly generated consistency matrices. More precisely, we have only resampled the non-empty entries of our consistency matrix (we still ignore other elements). With this analysis, we obtained a p-value of less than 0.001, so we can confidently reject the null hypothesis and say that our consistency matrix is positively correlated with the similarity matrix.

Number of re-evaluated images	Mean absolute error	p-value
26	0.615	< 0.001

Table 2: Reliability of the similarity matrix based on the results of the consistency matrix

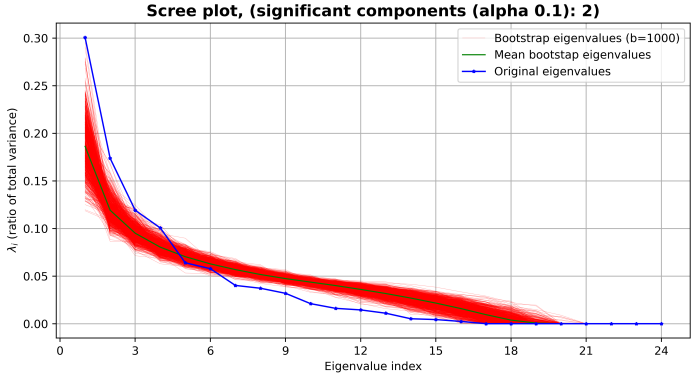

Figure 3: Scree plot. We use $\alpha = 0.1$ because we have a low number of samples. With $\alpha = 0.05$, we only obtain one significant component. We have set negative eigenvalues to 0 for convenience.

4.2 Extracting a meaningful basis

Having applied MDS to our distance matrix, we obtain a basis of $n = 17$ (17 out of the 24 eigenvalues are positive) components. However, not all components are needed to properly represent the emotions in the embedding space. For this reason, we have used bootstrap to determine how many of the 17 components are meaningful.

With this goal, let us define $p = 17$ tests. For each of these p tests, we define a null hypothesis indicating that *the variance explained by component p of the original data is the same as that of the relative variance explained by component p of the MDS space*. Here, we assume that the MDS basis is sorted in descending order according to the eigenvalues and that negative eigenvalues (and their corresponding eigenvectors) have been set to zero. Mathematically, the null hypothesis is defined as $H_{0,p} : \lambda_p = \hat{\lambda}_p$, where $\hat{\lambda}_p$ corresponds to the eigenvalue of uncorrelated data. The alternative hypothesis is $H_{A,p} : \lambda_p > \hat{\lambda}_p$.

It is important to note that bootstrap resamples must be generated according to H_0. The distance matrix is symmetric; therefore, the permuted distance matrix has to be symmetric as well. To achieve this, we need to randomly permute the upper triangular part of the original distance matrix and transpose it into the lower triangular part to ensure the symmetry of the bootstrap resample.

We have used $B = 1000$ permutation tests to test all the hypotheses. For each of these tests, we compute a p-value. This p-value represents the probability of observing a result as extreme as (or more) than the ones observed, assuming that H_0 is true. At significance level α, H_0 cannot be rejected if p-value $< \alpha$ is true. Finally, once all tests have been performed, we consider that there are k significant components if we have been able to reject the null hypothesis with $\alpha = 0.1$ of k tests. The results of these tests for images and landmarks can be seen in Figure 3.

For the 24 images we have compared, our results indicate that the meaningful basis with $\alpha = 0.1$ contains the first two components. This means that, on average, 90 % of the time, the variance explained by each of the first two principal components (47.5%) of our basis will be higher than if we were using uncorrelated data. Despite being two the number of significant components, the amount of variance retained by them is not very high.

The performance of MDS can be assessed visually with a distance-distance plot, in which we compare the original distances against the distances between the points in the reconstructed space. Ideally, we expect the scattered points to lay on a straight line with slope 1. However, points are far more dispersed due to the amount of variance lost by projecting to the MDS space. Figure 4 shows two distance-distance plots.

4.3 Analysis of results

Once we have the principal components obtained with MDS, we can plot our samples using the linear space determined by the first two principal components. Since one of the project's goals is to compare our results with the circumplex model of affection, we have mapped our linear space into polar coordinates. We have also flipped one of the directions to follow the typical representation of the circumplex model. The results can be seen in Figure 5.

We are pleasantly surprised with the final result of the project. The two directions captured by multi-dimensional scaling largely correspond to the ones proposed by the circumplex model. Each emotion of our

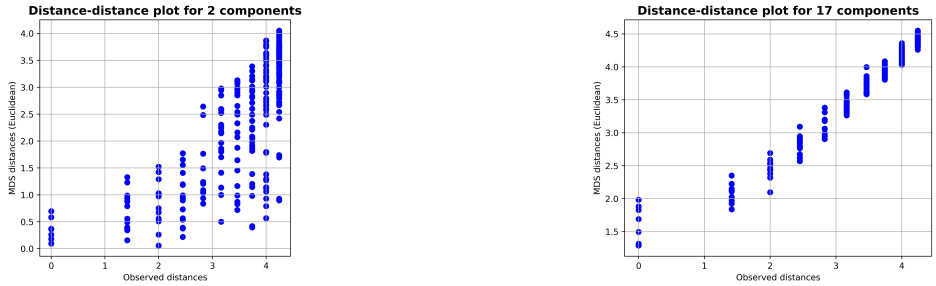

Figure 4: Distance-distance plots using 2 (left) and 17 (right) components

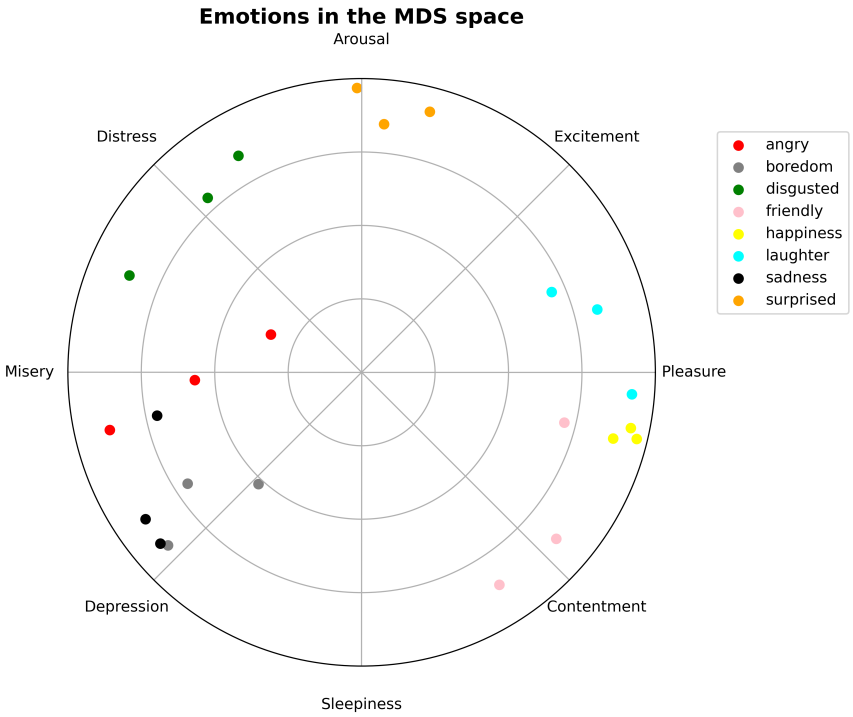

Figure 5: Obtained embedding space. We have based the labels for each angular direction on [Posner et al., 2005].

dataset is approximately mapped to its appropriate position in the space.

Some emotions are better classified than others. For example, boredom expressions have been mapped with a negative valence, while our intuition suggests a more neutral one would be more fitting. Similarly, expressions of anger have been mapped to a neutral activation level, although we expected a positive activation level. Emotions have been better mapped on the positive side of the valence axis. Our representation shows that laughter is the emotion with more activation, followed by happiness and friendly emotions, which is the expected outcome.

It is important to note that 24 images is a small dataset, and it is to be expected to have some inconsistencies and limitations. A more significant number of samples would be necessary to evaluate other characteristics of our results, like the dispersion of the different samples within the same emotion.

References

[Du and Martínez, 2015] Du, S. and Martínez, A. M. (2015). Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in Clinical Neuroscience*, 17(4):443–455.

[Izenman, 2008] Izenman, A. J. (2008). *Modern multivariate statistical techniques*, volume 1. Springer.

[Pantić, 2009] Pantić, M. (2009). Machine analysis of facial behaviour: naturalistic and dynamic behaviour. *Philosophical transactions of the Royal Society of London*, 364(1535):3505–3513.

[Posner et al., 2005] Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03).