UNIVERSITAT POLITÈCNICA DE VALÈNCIA

A system for modeling social traits in realistic faces with artificial intelligence

April, 2018

DEPARTAMENTO DE COMUNICACIONES

Author: Félix José Fuentes Hurtado

Supervisor: Prof. Valery Naranjo Ornedo
Prof. José Antonio Diego Más

A mis yayos y a mis padres.

Acknowledgments

Esta tesis es el resultado de un camino muy largo en el cual han participado muchas personas. Quiero agradecer, lo primero, el apoyo infinito y la confianza que me ha brindado siempre mi familia. Los que están y los que se fueron. Sin todos y cada uno de ellos nada de esto hubiera sido posible. También quiero agradecer a Paty los muchos consejos y ánimos a lo largo de esta aventura. Has sabido decirme las cosas que yo no era capaz de ver y eso me ha ayudado mucho. Además, de no ser por ti, no hubiera ido a hablar con Valery en un primer momento y nunca hubiera empezado esta tesis. Ahora que lo pienso, hubieramos vivido mucho más felices estos 3 años... no, es broma, te lo agradezco de verdad, con esto cumplo uno de los muchos sueños que tengo.

Como he dicho, esta tesis nació de ir a hablar un día con Valery a su despacho sin nada que perder. Sin ella, esta tesis no hubiera sido posible. Fue ella la que confió en mi desde un primer momento y me brindó la oportunidad de aprender muchísimo en un tema que me apasiona y con un equipo de gente maravilloso. Gracias Valery. Y gracias también a Adri, Sandra, Andrés, Fran, Fer, Reinier, Gabri, Rober, Jorge, Adri B., Conchi, Javi, Pablo y Jose por todos esos ratos de diversión que día a día te hacen la vida mucho más agradable, además de las muchas veces que me habéis ayudado.

También quiero agradecerle a Toni toda la ayuda que me ha prestado. Fue una sorpresa encontrarme con él en el camino y ahora echo la vista hacia atrás y entiendo que sin su ayuda, conocimiento y ánimos, esta tesis no hubiera salido adelante. Gracias Toni.

Por otra parte, agradecer también a mis madrileños el tiempo que pasé con ellos y la ayuda prestada. Ha sido un placer encontrarme con vosotros.

Por supuesto, no puedo olvidarme de todos mis amigos, los que durante este tramo de mi vida han seguido estando ahí para evitar que la tesis me volviera loco.

Por último, necesito darle las gracias a todas las personas que participaron en las dos validaciones que se llevaron a cabo, pues sin ellas, indudablemente, esta tesis no hubiera sido posible.

Me he equivocado, no era el último agradecimiento. Aunque probablemente él no se entere nunca porque vive en su maravilloso mundo perruno, necesito agradecerle a Tay lo que me ha ayudado. No hay nada como un rato con él para darse cuenta de que los agobios son innecesarios.

Muchas gracias a todos.

Abstract

Humans have specially developed their perceptual capacity to process faces and to extract information from facial features. Using our behavioral capacity to perceive faces, we make attributions such as personality, intelligence or trustworthiness based on facial appearance that often have a strong impact on social behavior in different domains. Therefore, faces play a central role in our relationships with other people and in our everyday decisions.

With the popularization of the Internet, people participate in many kinds of virtual interactions, from social experiences, such as games, dating or communities, to professional activities, such as e-commerce, e-learning, e-therapy or e-health. These virtual interactions manifest the need for faces that represent the actual people interacting in the digital world: thus the concept of avatar emerged. Avatars are used to represent users in different scenarios and scopes, from personal life to professional situations. In all these cases, the appearance of the avatar may have an effect not only on other person's opinion and perception but on self-perception, influencing the subject's own attitude and behavior. In fact, avatars are often employed to elicit impressions or emotions through non-verbal expressions, and are able to improve online interactions or even useful for education purposes or therapy. Then, being able to generate realistic looking avatars which elicit a certain set of desired social impressions poses a very interesting and novel tool, useful in a wide range of fields.

This thesis proposes a novel method for generating realistic looking faces with an associated social profile comprising 15 different impressions. For this purpose, several partial objectives were accomplished.

First, facial features were extracted from a database of real faces and grouped by appearance in an automatic and objective manner employing dimensionality reduction and clustering techniques. This yielded a taxonomy which allows to systematically and objectively codify faces according to the previously obtained clusters. Furthermore, the use of the proposed method is not restricted to facial features, and it should be possible to extend its use to automatically group any other kind of images by appearance.

Second, the existing relationships among the different facial features and the social impressions were found. This helps to know how much a certain facial feature influences the perception of a given social impression, allowing to focus on the most important feature or features when designing faces with a sought social perception.

Third, an image editing method was implemented to generate a completely new, realistic face from just a face definition using the aforementioned facial feature taxonomy.

Finally, a system to generate realistic faces with an associated social trait profile was developed, which fulfills the main objective of the present thesis.

The main novelty of this work resides in the ability to work with several trait dimensions at a time on realistic faces. Thus, in contrast with the previous works that use noisy images, or cartoon-like or synthetic faces, the system developed in this thesis allows to generate realistic looking faces choosing the desired levels of fifteen impressions, namely Afraid, Angry, Attractive, Babyface, Disgusted, Dominant, Feminine, Happy, Masculine, Prototypical, Sad, Surprised, Threatening, Trustworthy and Unusual.

The promising results obtained in this thesis will allow to further investigate how to model social perception in faces using a completely new approach.

Resumen

Los seres humanos han desarrollado especialmente su capacidad perceptiva para procesar caras y extraer información de las características faciales. Usando nuestra capacidad conductual para percibir rostros, hacemos atribuciones tales como personalidad, inteligencia o confiabilidad basadas en la apariencia facial que a menudo tienen un fuerte impacto en el comportamiento social en diferentes dominios. Por lo tanto, las caras desempeñan un papel fundamental en nuestras relaciones con otras personas y en nuestras decisiones cotidianas.

Con la popularización de Internet, las personas participan en muchos tipos de interacciones virtuales, desde experiencias sociales, como juegos, citas o comunidades, hasta actividades profesionales, como e-commerce, e-learning, e-therapy o e-health. Estas interacciones virtuales manifiestan la necesidad de caras que representen a las personas reales que interactúan en el mundo digital: así surgió el concepto de avatar. Los avatares se utilizan para representar a los usuarios en diferentes escenarios y ámbitos, desde la vida personal hasta situaciones profesionales. En todos estos casos, la aparición del avatar puede tener un efecto no solo en la opinión y percepción de otra persona, sino en la autopercepción, que influye en la actitud y el comportamiento del sujeto. De hecho, los avatares a menudo se emplean para obtener impresiones o emociones a través de expresiones no verbales, y pueden mejorar las interacciones en línea o incluso son útiles para fines educativos o terapéuticos. Por lo tanto, la posibilidad de generar avatares de aspecto realista que provoquen un determinado conjunto de impresiones sociales supone una herramienta muy interesante y novedosa, útil en un amplio abanico de campos.

Esta tesis propone un método novedoso para generar caras de aspecto realistas con un perfil social asociado que comprende 15 impresiones diferentes. Para este propósito, se completaron varios objetivos parciales.

En primer lugar, las características faciales se extrajeron de una base de datos de caras reales y se agruparon por aspecto de una manera automática y objetiva empleando técnicas de reducción de dimensionalidad y agrupamiento. Esto produjo una taxonomía que permite codificar de manera sistemática y objetiva las caras de acuerdo con los grupos obtenidos previamente. Además, el uso del método propuesto no se limita a las características faciales, y se podría extender su uso para agrupar automáticamente cualquier otro tipo de imágenes por apariencia.

En segundo lugar, se encontraron las relaciones existentes entre las diferentes características faciales y las impresiones sociales. Esto ayuda a saber en qué medida una determinada característica facial influye en la percepción de una determinada impresión social, lo que permite centrarse en la característica o características más importantes al diseñar rostros con una percepción social deseada.

En tercer lugar, se implementó un método de edición de imágenes para generar una cara totalmente nueva y realista a partir de una definición de rostro utilizando la taxonomía de rasgos faciales antes mencionada.

Finalmente, se desarrolló un sistema para generar caras realistas con un perfil de rasgo social asociado, lo cual cumple el objetivo principal de la presente tesis.

La principal novedad de este trabajo reside en la capacidad de trabajar con varias dimensiones de rasgos a la vez en caras realistas. Por lo tanto, en contraste con los trabajos anteriores que usan imágenes con ruido, o caras de dibujos animados o sintéticas, el sistema desarrollado en esta tesis permite generar caras de aspecto realista eligiendo los niveles deseados de quince impresiones: Miedo, Enfado, Atractivo, Cara de niño, Disgustado, Dominante, Femenino, Feliz, Masculino, Prototípico, Triste, Sorprendido, Amenazante, Confiable e Inusual.

Los prometedores resultados obtenidos en esta tesis permitirán investigar más a fondo cómo modelar la percepción social en las caras utilizando un enfoque completamente nuevo.

Resum

Els sers humans han desenvolupat especialment la seua capacitat perceptiva per a processar cares i extraure informació de les característiques facials. Usant la nostra capacitat conductual per a percebre rostres, fem atribucions com ara personalitat, intel·ligència o confiabilitat basades en l'aparença facial que sovint tenen un fort impacte en el comportament social en diferents dominis. Per tant, les cares exercixen un paper fonamental en les nostres relacions amb altres persones i en les nostres decisions quotidianes.

Amb la popularització d'Internet, les persones participen en molts tipus d'interaccions virtuals, des d'experiències socials, com a jocs, cites o comunitats, fins a activitats professionals, com e-commerce, e-learning, e-therapy o e-health. Estes interaccions virtuals manifesten la necessitat de cares que representen a les persones reals que interactuen en el món digital: així va sorgir el concepte d'avatar. Els avatars s'utilitzen per a representar als usuaris en diferents escenaris i àmbits, des de la vida personal fins a situacions professionals. En tots estos casos, l'aparició de l'avatar pot tindre un efecte no sols en l'opinió i percepció d'una altra persona, sinó en l'autopercepció, que influïx en l'actitud i el comportament del subjecte. De fet, els avatars sovint s'empren per a obtindre impressions o emocions a través d'expressions no verbals, i poden millorar les interaccions en línia o inclús són útils per a fins educatius o terapèutics. Per tant, la possibilitat de generar avatars d'aspecte realista que provoquen un determinat conjunt d'impressions socials planteja una ferramenta molt interessant i nova, útil en un ampla varietat de camps.

Esta tesi proposa un mètode nou per a generar cares d'aspecte realistes amb un perfil social associat que comprén 15 impressions diferents. Per a este propòsit, es van completar diversos objectius parcials.

En primer lloc, les característiques facials es van extraure d'una base de dades de cares reals i es van agrupar per aspecte d'una manera automàtica i objectiva emprant tècniques de reducció de dimensionalidad i agrupament. Açò va produir una taxonomia que permet codificar de manera sistemàtica i objectiva les cares d'acord amb els grups obtinguts prèviament. A més, l'ús del mètode proposat no es limita a les característiques facials, i es podria estendre el seu ús per a agrupar automàticament qualsevol altre tipus d'imatges per aparença.

En segon lloc, es van trobar les relacions existents entre les diferents característiques facials i les impressions socials. Açò ajuda a saber en quina mesura una determinada característica facial influïx en la percepció d'una determinada impressió social, la qual cosa permet centrar-se en la característica o característiques més importants al dissenyar rostres amb una percepció social desitjada.

En tercer lloc, es va implementar un mètode d'edició d'imatges per a generar una cara totalment nova i realista a partir d'una definició de rostre utilitzant la taxonomia de trets facials abans mencionada.

Finalment, es va desenrotllar un sistema per a generar cares realistes amb un perfil de tret social associat, la qual cosa complix l'objectiu principal de la present tesi.

La principal novetat d'este treball residix en la capacitat de treballar amb diverses dimensions de trets al mateix temps en cares realistes. Per tant, en contrast amb els treballs anteriors que usen imatges amb soroll, o cares de dibuixos animats o sintètiques, el sistema desenrotllat en esta tesi permet generar cares d'aspecte realista triant els nivells desitjats de quinze impressions: Por, Enuig, Atractiu, Cara de xiquet, Disgustat, Dominant, Femení, Feliç, Masculí, Prototípic, Trist, Sorprés, Amenaçador, Confiable i Inusual.

Els prometedors resultats obtinguts en esta tesi permetran investigar més a fons com modelar la percepció social en les cares utilitzant un enfocament completament nou.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

This chapter introduces the motivations behind this thesis, its objectives and its main contributions. In addition, it also presents the thesis outline.

Contents

1.1 Motivation

Humans have specially developed their perceptual capacity to process faces and to extract information from facial features (Bruce and A. Young, 1986; Damasio, 1985). Our brain has a specialized neural network for processing face information (Kanwisher et al., 1997) that allows to identify people, their gender, age and race, or even to judge their emotions or personality impressions. Using our behavioral capacity to perceive faces, we make attributions such as personality, intelligence or trustworthiness based on facial appearance (Bruce and A. W. Young, 2012). Furthermore, these attributions often affect face memory and have a strong impact on social behavior in different domains (Walker and Vetter, 2009). Therefore, faces play a central role in our relationships with other people and in our everyday decisions (Little et al., 2007; Todorov, 2011).

Nowadays, with the popularization of Internet, people participate in many kinds of virtual interactions, from social experiences (games, dating, communities...) to professional activities (e-commerce, e-therapy and e-health, meetings and interviews online, e-learning and education...). These virtual interactions manifest the need for faces that represent the actual people interacting. In this context the concept of “avatar” emerged. Avatars are defined as “general graphic representations that are personified by means of computer technology” (Holzwarth et al., 2006).

Types of avatars are varied and depend on the user, the context and the situation. The same person may be using different avatars in different scenarios. Some research has focused on the study of the social implications that the use of avatars has as a way of self-presentation and communication in Internet (Chung et al., 2007; Schultze, 2010). Avatars can also be used in professional situations as online collaborative design (Koutsabasis et al., 2012) or e-commerce. They can also be employed to treat disorders such as autism (Konstantinidis et al., 2009).

In all these cases, the appearance of the avatar may have an effect on another person’s opinion and perception. For example, anthropomorphic and less androgynous avatars are perceived as more credible and trustworthy (Nowak and Rauh, 2008). Large pupils and a slow flicker frequency make avatars look more sociable and attractive (Weibel et al., 2010)). Elderly avatars are perceived by adult users as more intelligent and reliable (Marin and S. Lee, 2013). Attractive and more elaborate avatars are more successful in their social interactions (Banakou et al., 2009), and even receive more favorable ratings in virtual job interviews (Behrend et al., 2012). Similar results can be found in Hasler et al. (2013).

A recent study on the motivations and strategies for designing avatars found that users design avatars considering the following objectives: the virtual exploration (living experiences that are only possible in the digital world), social navigation, contextual adaptation and identity representation (H. Lin and H. Wang, 2014). Furthermore, the attitude and behavior of people may be influenced by their avatar’s features. For example, Yee and Bailenson (2007) show how people propensity to establish online relationships and their self-confidence are affected by the attractiveness and height of their avatars.

On the other hand, avatars are also employed to elicit impressions or emotions through non-verbal expressions. For example, expressive avatars and emoticons can improve online interactions (Fabri and Moore, 2005) or even be useful for education purposes or therapy (Wiederhold and Riva, 2009). Moreover, the transmission of emotions is strongly related to credibility, which in turn affects the ability to persuade (Bărbat and Cretulscu, 2003). This ability is important for applications in areas ranging from e-commerce to e-therapy. For example, Holzwarth et al. (2006) showed that an expert avatar is more effective for selling products of high level of product involvement, while an attractive avatar is better at moderate levels. These attributions people rapidly make from faces (such as if a seller is trustworthy or attractive) are called trait impressions, and the human being creates them unconsciously in as little as 34 milliseconds (Todorov, Olivola, et al., 2015).

With the available technology, it is possible to generate realistic faces. However, the possibility to generate realistic faces with the ability to elicit a certain set of trait impressions is very limited. Dotsch and Todorov (2012) employed reverse correlation techniques to create faces that model social perception of faces by superimposing random noise on them. Vernon et al. (2014) modeled first impressions of faces from highly variable facial images and created computer-generated cartoon face-like images depicting how features affect impressions. Sutherland, Oldmeadow, et al. (2013) created a three-dimensional model to characterize social inferences from faces. Although all these works were successful at modeling social traits in faces, none of them achieved realistic looking faces nor could choose the intensity of several social traits at a time on one face. On the other hand, Walker and Vetter (2009) achieved realistic looking faces, but again only one social trait could be chosen at a time.

Therefore, the purpose of this thesis is the development of a system able to design realistic faces controlling the impressions they have to convey in a series of social traits. Since the avatars will be used by many people, the objective will be to elicit the desired impressions on most of them.

1.2 Objectives

The main objective of this thesis is to develop a system with the capability of creating avatar faces able to convey the most appropriate feelings to the viewer according to a given context. Since avatars will be used by many people, the goal is to transmit the desired sensation to most of them. The proposed system focuses on a combination of image processing and artificial intelligence algorithms, such as clustering methods and genetic algorithms, whose training is based on the human perception of a set of real faces. From the main objective, several secondary objectives arise:

- Development of an automatic method to group facial features by similarity in an objective way. This procedure will allow to create feature taxonomies which will be used to codify faces and create new ones.
- Development of an automatic method to codify or define any new incoming face.
- Using the aforementioned taxonomies obtained by the grouping method, determine how much each feature affects each impression formation.
- Implementation of a system able to create a new realistic face from a given codification or definition.
- Provide the created realistic faces with the ability to elicit a certain set of trait impressions.

1.3 Main contributions

This thesis provides novel methods for creating realistic faces able to convey certain impressions to the observer. One of its main contributions is to develop a method for automatic classification of facial features based on their appearance.

The methods presented in chapter 2 are defined by combining different morphological operators, dimensionallity reduction methods and clustering algorithms. Implementing an automatic classification method for facial features based on their appearance makes possible the creation of taxonomies of facial features. The importance of taxonomies lies in that by using them, any facial feature can be classified in a certain type. In other words, they allow for using a common terminology to define or codify face configurations providing a standardized way to describe them. The novelty in this work is that the procedure followed avoids the problems related to human limitations when classifying facial traits. On the one hand, facial features were classified using only their visual appearance, thus removing any possible human judgment. On the other hand, the method developed has the capability of classifying any new incoming feature in the already existent taxonomy. Furthermore, this method could easily be adapted to create taxonomies of any other facial feature or even other matters.

Another important contribution of this thesis is finding the relationships among facial features in creating social impressions. For example, it is now possible to certainly know how much a given type of eyes affect the social perception of *happiness*, or any of the other available impressions. This will allow to focus in the most important features when designing a new face which aim is to convey any of the fifteen impressions used in this work. Furthermore, the method is easily scalable and can be extended in the future to include new impressions or features.

Finally, the last contribution of this thesis is a system able to create realistic faces able to convey certain impressions to most observers. In particular, this system is able to obtain a face with a certain profile of 15 different impressions at the same time. The methods employed in chapter 3 to find the relationships among features and to generate faces able to convey impressions include the use of Genetic Algorithms and advanced image processing techniques such as Poisson image editing.

To sum up, the contributions of this thesis are:

- A method able to automatically and objectively classify facial features in groups based on their appearance.
- The existent relationships among facial features in the elicitation of fifteen different impressions.
- A method to codify new incoming faces in a standardized manner.
- A system able to create realistic faces that convey certain impressions to the gross of the population.

1.4 Outline

This thesis is divided into 4 chapters. This chapter has presented the motivations behind the research performed in the thesis, the objectives and its main contributions.

Chapter 2 introduces the necessity of an automatic method to classify facial features based on their appearance in order to create faces which elicit certain trait impressions and the importance of taxonomies is explained. Moreover, the most relevant works in this topic are reviewed and the methods followed to implement the system developed in this thesis to extract and cluster the facial features in order to create taxonomies are explained.

Chapter 3 presents the importance of automatic modeling of social impression of faces and its current state-of-the-art. Then, the methods followed to find an emotional function and to implement the face generator are explained.

General and final conclusions in addition to future prospects are presented in chapter 4. Figure 1.1 shows the implemented system flowchart.

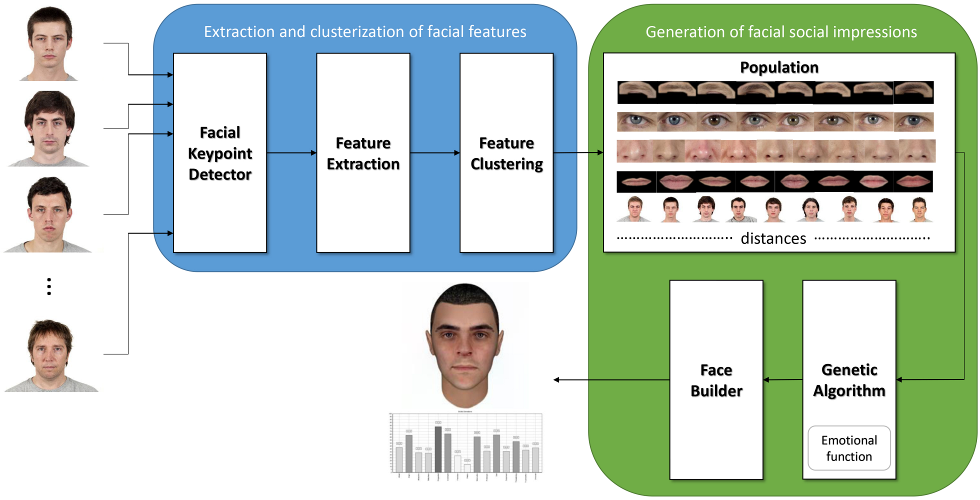


Figure 1.1: Implemented system flowchart.

Chapter 2

Extraction and clustering of facial features

In this chapter the procedure followed to extract and cluster facial features is explained. First, an introduction on why it is so important to create a method to automatically and objectively group features based on their appearance is presented to the reader. Then, the theoretical framework of the employed methods is explained. Next, the database used is depicted. Finally, the procedures implemented to extract the features and cluster them is thoroughly reviewed, resulting taxonomies are given and their validation is discussed.

Contents

2.1 Introduction

For centuries, artists and researchers have tried to develop procedures to measure and classify human faces. Anthropometric facial analysis is used in different fields like surgery (Arslan et al., 2008; Ferring and Pancherz, 2008; J. P. Porter and Olson, 2001), forensic science (Mane et al., 2010; Stefanie Ritz-Timme et al., 2011; Ritz-Timme et al., 2011), art (Hochscheid et al., 2015; Robins, 1984), face recognition, emotion recognition and facial trait judgments (Boberg et al., 2008; Buckingham et al., 2006; Rojas et al., 2011). In the last decades, new technologies have opened ways to automatically evaluate facial features and gestures, and computational methods for analysis of facial information are now applied to classify faces based on anthropometric or emotional criteria (Tian et al., 2005).

Classification or typology systems used to categorize different human body parts exist for many years. In 1940, William Sheldon developed somatotypes to describe the body build of an individual. Sheldon proposed a classification system in which all possible body types were characterized based on the degree to which they matched these somatotypes (Sheldon, 1954). Other taxonomies have been developed for the shape of the body (Alemany et al., 2010; Vinué et al., 2015), hands (Jee and Yun, 2016), feet (N.-S. Kim and Do, 2014) or head (Sarakon et al., 2014). Taxonomies, as classification system, allow using a common terminology to define body part configurations providing a standardized way to describe them, and are widely used in many fields such as ergonomics and bio-mechanics (Y.-L. Lin and K.-L. Lee, 1999; Preston and Singh, 1972), criminalistics (Stefanie Ritz-Timme et al., 2011), sports (Malousaris et al., 2008; Massidda et al., 2013), medicine (Koleva et al., 2002), design or apparel industry (Alemany et al., 2010). In general, this kind of typology systems is intended for qualitative categorization based on the global appearance of body parts, although, in some cases, a quantitative analysis of some selected features is developed to obtain the classification.

In the case of facial features, taxonomies are useful, for example, in ergonomics, forensic, anthropology or crime prevention. New human-machine interaction systems and online activities like e-commerce, e-learning, games, dating or social networks, are fields in which facial features classifications are needed. In these activities it is common to use human digital representations that symbolize the user's presence or that act as virtual interlocutor (Davis et al., 2009). The importance of communicative behaviors of avatars in new interaction systems (Carvalho et al., 2008; Fabri, Elzouki, et al., 2007; Fabri and Moore, 2005; Orvalho et al., 2009; Yee and Bailenson, 2007) has led to an increasing interest in creating realistic avatars able to convey appropriated sensations to users. In this context, it is common to synthesize faces and facial expressions combining facial features (Albin-Clark and Howard, 2009; Diego-Mas and Alcaide-Marzal, 2015; Sukhija et al., 2016; Trescak et al., 2012).

Several taxonomies of facial features can be found in the literature. For example, Vanezis's atlas (Vanezis et al., 1996) classifies 23 facial features, the Disaster Victim Identification Form (DVI) by Interpol categorizes 6, and the DVM database (Aßmann, 2007; Ohlrogge, 2009) 45 facial features. In Tamir (2011), different shapes of the human nose are classified into 14 groups based on the analysis of 1793 pictures of noses. A similar approach was used for classifying human chin (Tamir, 2013). In these works, a big set of photographs were analyzed and classified based on the similarity of the features.

This approach, while intuitively logical, has several problems not only in the development of taxonomies, but also in its later use. The classification of facial features is obtained from the opinion of a limited group of human observers. Classic behavioral work has shown that humans' brain integrates facial features into a gestalt whole when it processes face information (holistic face processing, (Richler et al., 2011)), decreasing our ability for processing individual features or parts of faces (Taubert et al., 2011). This part-whole effect makes difficult, for example, to recognize familiar faces from isolated features (Davidoff and Donnelly, 1990; Donnelly and Davidoff, 1999; Tanaka and Farah, 1993). Moreover, individual differences exist in face recognition ability (R. Wang et al., 2012), and some issues, like face race, influence the performance in processing features and the configuration of facial information (Hayward et al., 2008; Rhodes et al., 2009). This is reflected in low inter-observer and intra-observer agreement in the evaluation of facial features (Stefanie Ritz-Timme et al., 2011). Finally, apart from the difficulties of processing parts of faces, creating this kind of taxonomies implies classifying a very big set of elements (the number of possible different features) in an undefined number of groups, and this kind of tasks easily overcomes our capacities for information processing

(Miller, 1956; Scharff et al., 2011). To deal with these problems, we propose a new procedure to develop facial trait taxonomies based on its appearance using computational methods for automatically classifying features.

Recently, analysis of facial images has become a major research topic, and new computational methods for analysis of facial information have been developed. A comparison of these techniques shows two different approaches to deal with facial information (Rojas et al., 2011). The first one (structural approach) automatically encodes the geometry of faces using several significant points and relationships between them, doing a metric or morphological assessment of facial features. Examples of this kind of techniques are those based on SIFT feature descriptors (Meyers and Wolf, 2008), point distribution models (T. F. Cootes et al., 2001) or local binary patterns (Ahonen et al., 2006). On the other hand, the holistic approach uses appearance-based representations, considering all available information and encompassing the global nature of the faces. Holistic techniques include, for example, Fisherfaces (Belhumeur et al., 1997) or Eigenfaces (Turk and Pentland, 1991). Some work in facial features characterization has been done mixing structural and holistic techniques (Klare and Jain, 2010).

Classification methods of facial traits are needed in order to develop taxonomies. Research using computational methods is usually focused on the characterization of complete faces. However, less efforts have been done in facial trait classification based on its appearance. The objective of this work is to develop an appearance-based method to obtain a relatively low-dimensional vector of characteristics for facial traits. On this basis, large sets of five facial traits (eyebrows, eyes, noses, mouths and jawlines) of varying ethnicity (Asian, Black, Latino and White) were characterized. Using this characterization, the traits were clustered obtaining new taxonomies for each ethnic group. The procedure followed avoids the problems related to human limitations in classifying facial traits. On the one hand, the characterization and clustering of the traits were not based in human judgments. On the other hand, classifying new traits in one of the groups of the taxonomies can be done in an automatized way. Finally, the procedure was tested comparing human opinions with automatically generated groups of traits.

Next section shows the theoretical background of the methods implemented in this chapter. Then, the database employed in this thesis is depicted. Next, the procedures followed to extract the facial features from the face images are explained. Afterwards, the *eigenfaces* approach is employed to characterize the recently extracted, large sets of photographs of five facial features (eyebrows, eyes, noses, mouths, and jawlines). This holistic technique seems to be more consistent and reliable for categorizations than those that imply subjective judgments (Rojas et al., 2011). The clustering process used to group features is then described. Feature inter-distances are also extracted and clustered. The classifications obtained and the agreement between human judgments and these automatically generated taxonomies are then explained. Finally, the results are discussed and conclusions are exposed.

The rest of the chapter is organized as follows: Section 2.2 shows the theoretical background. Section 2.3 depicts the database utilized. Section 2.4 details the procedure followed to extract the facial features, and Section 2.5 shows the method followed to cluster the extracted facial features. Section 2.6 details the results and the facial feature taxonomies obtained with the methods explained above. Validation of the proposed procedure is provided in Section 2.7. Finally, Section 2.8 provides conclusions.

2.2 Theoretical background

2.2.1 Mathematical morphological operators

Mathematical morphology is a theory for the analysis of spatial structures. It is based on set theory, integral geometry and lattice algebra; and poses a powerful image analysis technique (Soille, 2013).

Let f be a gray-scale image which is defined as $f(\mathbf{x}) : E \to T$ where $\mathbf{x}$ is the pixel position. In the case of discrete valued images, $T = \{t_{min}, t_{min} + 1, \ldots, t_{max}\}$ is an ordered set of gray-levels. Typically, in digital 8-bit images $t_{min} = 0$ and $t_{max} = 255$. Furthermore, let $B(\mathbf{x})$ be a sub-set of Z^2 called structuring element (shape probe) centred at point $\mathbf{x}$, whose shape is usually chosen according to some a *priori* knowledge about the geometry and size of the relevant and irrelevant image structures.

Erosion and dilation are the two most basic mathematical morphology operators (Soille, 2013):

$$\begin{aligned} Dilation &: [\delta_B(f)](\mathbf{x}) = \max\nolimits_{b \in B(\mathbf{x})} f(\mathbf{x} + \mathbf{b}) \\ Erosion &: [\varepsilon_B(f)](\mathbf{x}) = \min\nolimits_{b \in B(\mathbf{x})} f(\mathbf{x} + \mathbf{b}). \end{aligned} \tag{2.1}$$

Their objective is to expand light or dark regions, respectively, according to the size and shape of the structuring element. These elementary operations can be combined to obtain a new set of operators or basic filters given by:

$$\begin{aligned} Opening &: \gamma_B(f) = \delta_B(\varepsilon_B(f)) \\ Closing &: \varphi_B(f) = \varepsilon_B(\delta_B(f)) \quad . \end{aligned} \tag{2.2}$$

Light or dark structures are respectively filtered out from the image by these operators regarding the structuring element chosen.

Another operator used in this work is the *geodesic dilation*. The *geodesic dilation* is the iterative unitary dilation of an image f (marker) which is contained within an image g (reference),

$$\delta_g^{(n)}(f) = \delta_g^{(1)}\delta_g^{(n-1)}(f), \text{ being } \delta_g^{(1)}(f) = \delta_B(f) \wedge g\,. \tag{2.3}$$

In order to define the *fill-holes operator*, we must first introduce the *geodesic reconstruction by dilation*, which performs the successive *geodesic dilation* of f regarding g up to idempotence,

$$\gamma^{rec}(g,f) = \delta_g^{(i)}(f), \text{ so that } \delta_g^{(i)}(f) = \delta_g^{(i+1)}(f)\,. \tag{2.4}$$

We can now define the *fill-holes operator*. Basically, this operator fills all holes in an image f that do not touch the image boundary f_∂ (used as marker):

$$\psi^{ch}(f) = [\gamma^{rec}(f^c, f_\partial)]^c\,, \tag{2.5}$$

where f^c is the complement image (i.e., the negative). For a gray-scale image, it is considered a hole any set of connected points surrounded by connected components of value strictly greater than the hole values.

The last algorithm used on this work is a *hit-or-miss* transform called *thickening* (Jain, 1989; Soille, 2013). In the case of binary *hit-or-miss* transformations, the structuring element is a set with two components, $\mathbf{B}(\mathbf{x})_{FG}$ and $\mathbf{B}(\mathbf{x})_{BG}$, placed so that both reference pixels are at the same position ($\mathbf{x}$) and are disjoint sets (i.e., $\mathbf{B}(\mathbf{x})_{FG} \cap \mathbf{B}(\mathbf{x})_{BG} = 0$). $\mathbf{B}(\mathbf{x})_{FG}$ defines the set of pixels that should match the foreground, while $\mathbf{B}(\mathbf{x})_{BG}$ does the same with the background. The hit-or-miss transform of a set f can be written in terms of an intersection of two morphological erosions:

$$f * \mathbf{B} = \varepsilon_{\mathbf{B}_{FG}}(f) \cap \varepsilon_{\mathbf{B}_{BG}}(f^c), \tag{2.6}$$

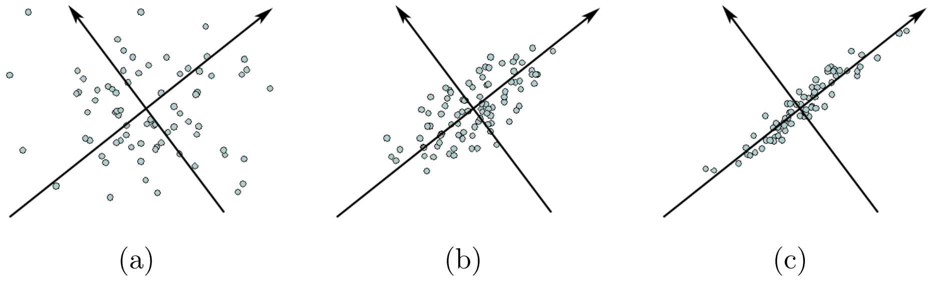
where f^c is the complement set of f, that is, the negative.

The *thickening* of a binary image f by a structuring element $\mathbf{B}$ is denoted by $f \odot \mathbf{B}$ and defined as the union of f and the *hit-or-miss* transform of f by $\mathbf{B}$:

$$f \odot \mathbf{B} = f \cup (f * \mathbf{B}). \tag{2.7}$$

2.2.2 Principal Component Analysis

The main idea of principal component analysis (PCA) is to transform a given feature space into a lower-dimensional subspace, while retaining as much as possible of the variation present in the original space (Toennies, 2012). If many of the original features are correlated, the distribution of samples in feature space actually occupies a lower-dimensional subspace. To achieve this reduction of dimensionality, the PCA produces an orthogonal transformation in the feature space such that all covariance values between features are zero and the coordinate axes are aligned or orthogonal to this subspace (Figure 2.1). Features corresponding to axes orthogonal to the subspace can be removed and thus it is possible to reduce dimensionality of the space.



(a) (b) (c)

Figure 2.1: PCA transforms the original axis system into one that decorrelates the data. To achieve it, new axes are oriented along the data distribution in the feature space, so features of this new coordinate system can be removed if their projection on the remaining axes produce only a small error. In example (a), no reduction is possible. In (b), reduction is possible but it removes information. In (c), reduction removes just noise.

In order to compute the PCA, the covariance matrix $\mathbf{C}$ of the original feature space needs to be estimated from sample covariances:

$$\begin{aligned}
\mathbf{C} &= \begin{pmatrix} c_{11} & c_{21} & \cdots & c_{N1} \\ c_{12} & c_{22} & \cdots & c_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1N} & c_{2N} & \cdots & c_{NN} \end{pmatrix} , \\
c_{ij} &\approx \frac{1}{K-1} \sum_{k=1}^{M} \left(f_{ik} - \bar{f}_i \right) \left(f_{jk} - \bar{f}_j \right) , \\
\bar{f}_i &\approx \frac{1}{K} \sum_{k=1}^{K} f_{ik} ,
\end{aligned} \tag{2.8}$$

where K is the number of samples available, and f_{ik} is the i^{th} feature of the k^{th} feature vector in the set of samples.

All off-diagonal elements of $\mathbf{C}$ would be zero if features were linearly uncorrelated. If features are uncorrelated and occupy only a lower-dimensional subspace that is aligned with features axes, some of the variances in the diagonal should be zero as well. Even if uncorrelated features are removed, any location of a sample in feature space can still be exactly represented.

However, covariance between features usually exists, so $\mathbf{C}$ usually contains nonzero off-diagonal elements. The PCA will create a new axis system in which new features $\mathbf{f}'$ are linear combinations of features of $\mathbf{f}$:

$$f'_j = \sum_{i=1}^{N} f_i e_{ij} \tag{2.9}$$

so that the covariance matrix of the sample distribution $\mathbf{f}'$ no longer has nonzero off-diagonal elements. The beauty of the method is that there exists a closed-form solution for computing weights e_{ij} (Figure 2.1).

Given an estimate of the original feature space covariance matrix $\mathbf{C}$, a new orthogonal system of feature axes $f'_1, \ldots, f'_N$ with covariance matrix $\mathbf{C}'$ is computed:

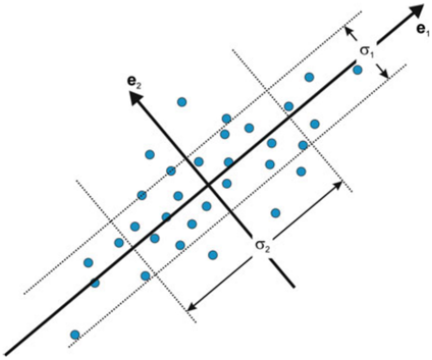


Figure 2.2: The PCA solves an eigenproblem for the covariance matrix **C**. A new coordinate system is created, in which the axes are the eigenvectors e_i and the eigenvalues λ_i are the variances (σ_i^2) along these axes. σ_i represents the standard deviations.

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_N^2 \end{pmatrix}, \tag{2.10}$$

where σ_i^2 represents the variance of the f_i feature. Eigenvalues λ_i and eigenvectors $\mathbf{e_i} = (e_{i1} \quad e_{i2} \ldots e_{iN})$ for **C** are computed as follows:

$$\begin{gathered} \mathbf{Ce_i} = \lambda_i \mathbf{e_i} \Rightarrow \mathbf{CE} = \mathbf{E\Lambda} \Leftrightarrow \mathbf{E^T CE} = \mathbf{\Lambda}\,, \\ \mathbf{E} = \begin{pmatrix} e_{11} & e_{21} & \cdots & e_{N1} \\ e_{12} & e_{22} & \cdots & e_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1N} & e_{2N} & \cdots & e_{NN} \end{pmatrix}, \end{gathered} \tag{2.11}$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of **C** and can be computed as:

$$
\begin{aligned}
\mathbf{\Lambda} = \mathbf{E}^T\mathbf{C}\mathbf{E} \approx \mathbf{E}^T \left[\frac{1}{K} \sum_{k=1}^{K} \left(\mathbf{f}_k - \bar{\mathbf{f}}\right) \times \left(\mathbf{f}_k - \bar{\mathbf{f}}\right)^T \right] \mathbf{E} \\
= \frac{1}{K} \sum_{k=1}^{K} \left[\mathbf{E}^T \left(\mathbf{f}_k - \bar{\mathbf{f}}\right)\right] \times \left[\left(\mathbf{f}_k - \bar{\mathbf{f}}\right)^T \mathbf{E}\right] \qquad (2.12) \\
= \frac{1}{K} \sum_{k=1}^{K} \mathbf{f}'_k \times \left(\mathbf{f}'_k\right)^T \ ,
\end{aligned}
$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C}$. They correspond to the feature variances of the covariance matrix $\mathbf{C}' = \mathbf{\Lambda}$ in a transformed system where a new feature f'_{ik} is computed by projecting the feature vector $\mathbf{f}_k$ on the i^{th} eigenvector:

$$f'_{ik} = (\mathbf{e}_i)^T(\mathbf{f}_k - \bar{\mathbf{f}})\,. \qquad (2.13)$$

Feature reduction is then carried out by inspecting the eigenvalues (*variances*) of $\mathbf{C}$. Features f'_i with their corresponding eigenvalues λ_i equal or close to zero can be removed, since that indicates high linear correlation.

To determine which features can be removed, they are sorted according to their variance. In order to choose the amount of feature reduction, a value $n < N$ is chosen such that the percentage $p_{var}(n) = \sigma^2_{accum}(n)/\sigma^2_{accum}(N)$ exceeds some threshold, where $\sigma^2_{accum}(n) = \sum_{i=1}^{n} \sigma_i^2$ accounts for the accumulated variance. For example, $p_{var}(n) > 0.95$ means that the first n features explain 95% of the variance in feature space.

Eigenfaces approach

In mathematical terms, Eigenfaces method aims to find the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating an image as a point (or vector) in a very high dimensional space. These eigenvectors can be thought of as a set of features that together characterize the variation between images, and are ordered accounting for a different amount of this variation. Each individual face can be represented exactly in terms of a linear combination of the eigenfaces (Figure 2.3) or using only the "best" eigenfaces (those that have the largest eigenvalues, and therefore account for the most variance within the set of images). The best M eigenfaces span an M-dimensional subspace of all possible images.

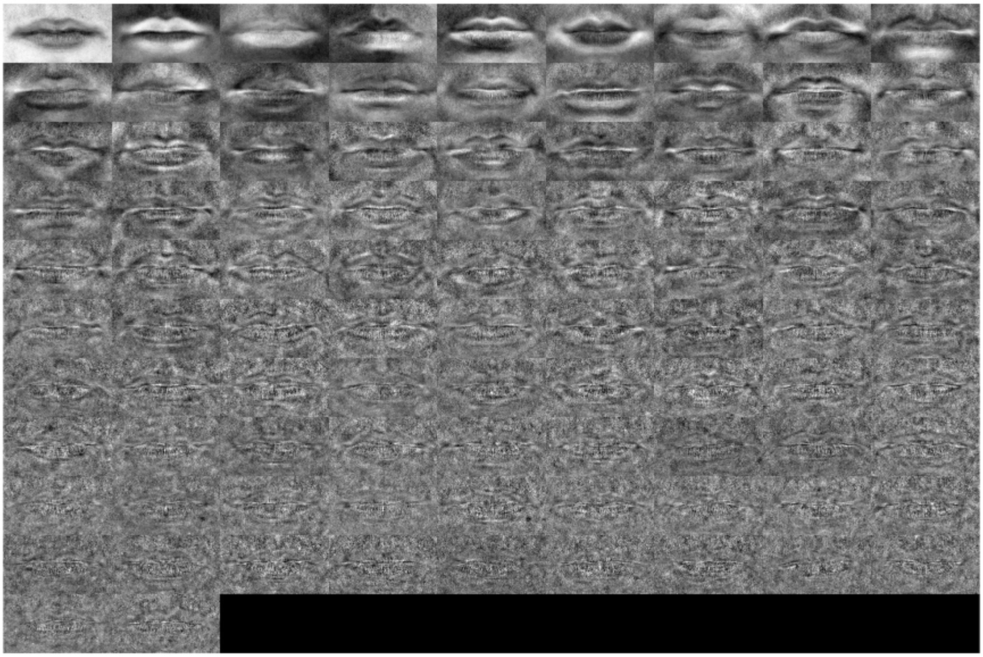


Figure 2.3: Set of basis obtained by applying the eigenfaces method on a set of mouths. This image shows the *eigenmouths* of a set of 92 mouths. Any original mouth can be recovered with more or less detail by computing a linear combination of a number of *eigenmouths*.

Let a face image f be a two-dimensional N by N array representing an image. The main idea of the principal component analysis is to find the vectors that best account for the distribution of face images within the entire image space. Each vector is of length N^2, describes an N by N image and is a linear combination of the original face images. Because these vectors are the eigenvectors of the covariance matrix corresponding to the original face images and they are face-like in appearance, they are refered to as "eigenfaces".

Let the training set of images be $f_1, f_2, f_3, \ldots, f_M$. The average face of the set is defined by $\Psi = \frac{1}{M}\sum_{n=1}^{M} f_n$. Each face differs from the average by the vector $\Phi_i = f_i - \Psi$. This set of very large vectors is then subject to principal component analysis, which seeks a set of M orthonormal vectors u_n which best describes the distribution of the data. The computation of the covariance of the matrix $A = [\Phi_1 \Phi_2 \ldots \Phi_M]$ is very expensive, but if $M > N^2$ (being M the number of images, and N^2 the total number of pixels in an image belonging to M), the solution can be obtained by first solving for the eigenvectors of an M by M matrix and then taking proper linear combinations of the face images Φ_i. Consider the eigenvectors v_i of $A^T A$ such that

$$A^T A \mathbf{v}_i = \mu_i \mathbf{v}_i \tag{2.14}$$

Pre-multiplying both sides by A

$$AA^T A \mathbf{v}_i = \mu_i A \mathbf{v}_i \tag{2.15}$$

from which $A\mathbf{v}_i$ are the eigenvectors of $C = AA^T$. Following this analysis, an M by M matrix $L = A^T A$ is constructed, where $L_{mn} = \mathbf{\Phi}_m^T \mathbf{\Phi}_n$, and the M eigenvectors ($\mathbf{v}_i$) of L can be obtained. These vectors determine linear combinations of the M training set face images to form the eigenfaces $\mathbf{u}_i$.

$$\mathbf{u}_i = \sum_{k=1}^{M} \mathbf{v}_{lk} \mathbf{\Phi}_k \qquad l = 1, ..., M \tag{2.16}$$

The associated eigenvalues allow to rank the eigenvectors according to their usefulness in characterizing the variation among the images (Turk and Pentland, 1991).

2.2.3 K-means clustering

K-means clustering is a variant of *partitional clustering.* It is an algorithm for putting N data points in an d-dimensional space into K clusters (MacKay, 2003).

Given a set of observations $x_i \in X$ for $i = 1, 2, \ldots, N$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the N observations into $K(\leq N)$ disjoint sets $\mathbf{S} = S_1, S_2, \ldots, S_K$ so as to minimize the within-cluster sum of squares (sum of distance functions of each point in the cluster to the K center)

$$\arg\min_{S} \sum_{i=1}^{K} \sum_{x \in S_i} \left\| x - \mu_i \right\|^2 \tag{2.17}$$

where μ_i is the mean of points in S_i. Any type of distance can be used with this algorithm.

It uses an iterative refinement technique. Given an initial set of k means $\mu_1^{(1)}, \ldots, \mu_K^{(1)}$, the algorithm proceeds by alternating between two steps:

- *Assignment step*: assign each observation to the cluster whose mean yields the least within-cluster sum of squares. Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean.

  $$S_i^{(t)} = \{x_p : \|x_p - \mu_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \; \forall j, 1 \leq j \leq K\} \tag{2.18}$$

  where each x_p is assigned to exactly one $S^{(t)}$, even if it could be assigned to two or more of them.

- *Update step*: Calculate the new means to be the centroids of the observations in the new clusters.

  $$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \tag{2.19}$$

  Since the arithmetic mean is a least-squares estimator, this also minimizes the within-cluster sum of squares (WCSS) objective.

The algorithm has converged when the assignments no longer change (Figure 2.4, bottom right plot). Since both steps optimize the WCSS objective, and there only exists a finite number of such partitionings, the algorithm must converge to a (local) optimum. However, there is no guarantee that the global optimum is found using this algorithm.

There are a number of cluster validity indexes that could be used to estimate K. In this work, some of the most popular and that work with general distance measures are used, which will be introduced below.

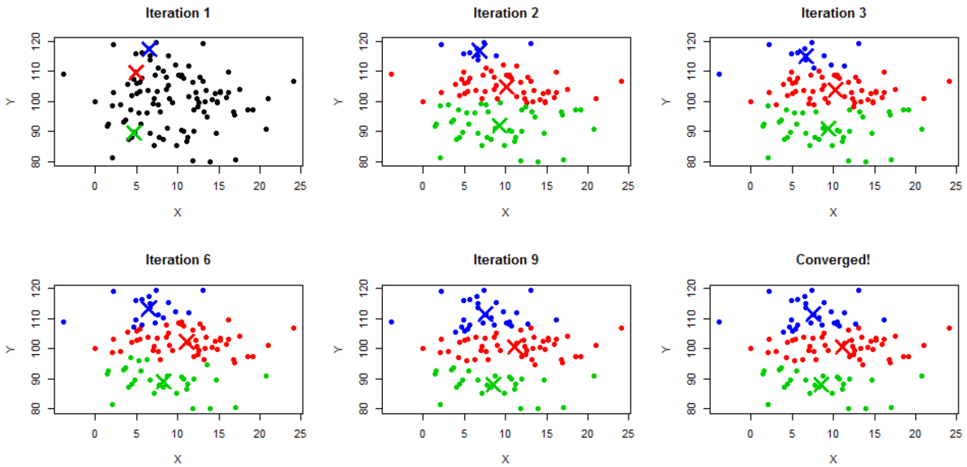


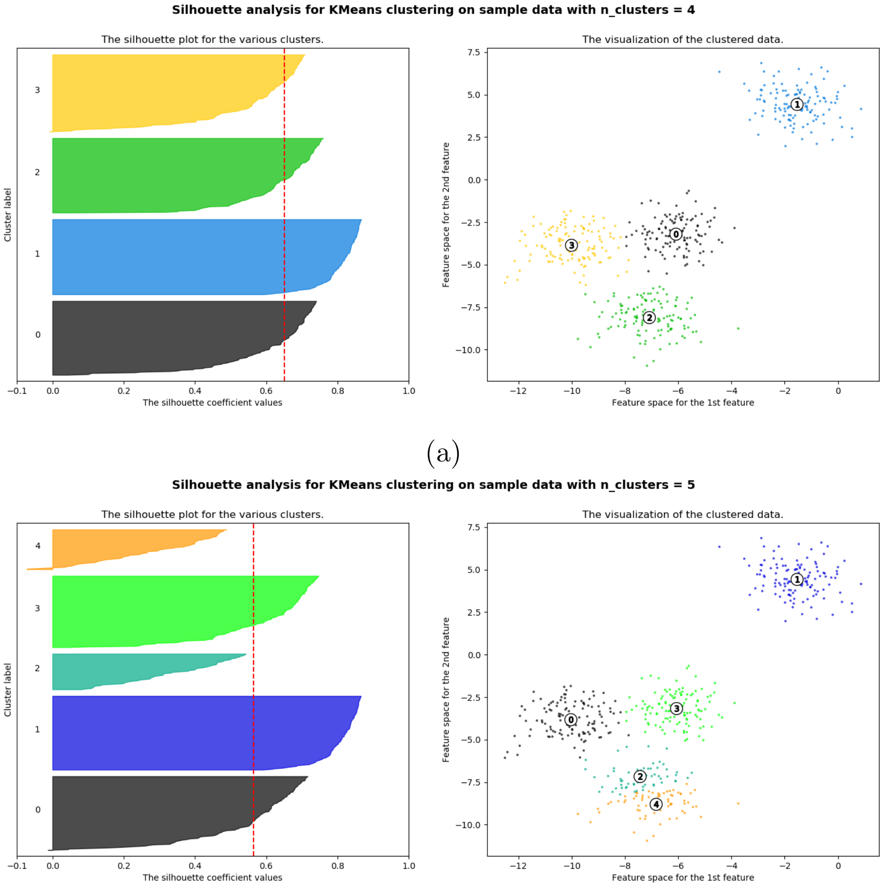
Figure 2.4: k-means progress. In iteration 1, the k centroids are used to assign to them their closest neighbours. Then, the mean of each of them is calculated, and the k centroids are updated (Iteration 2). This process is repeated until convergence is reached (last plot).

Silhouette index

The Silhouette is a method of interpretation and validation of consistency within clusters of data (Rousseeuw, 1987). Its value is a measure of how similar an object x_i is to its own cluster (cohesion) compared to other clusters (separation):

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}} \tag{2.20}$$

where $a(x_i)$ is the average dissimilarity of $x_i \in S_k$ to all other $x_j \in S_k$, $b(x_i)$ is the minimum dissimilarity over all clusters S_l, to which x_i is not assigned, of the average dissimilarities to $x_j \in S_l, l \neq k$. Therefore, the silhouette value $s(x_i)$ ranges from -1 to $+1$, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If $s(x_i)$ is around zero, the entity x_i could be assigned to another cluster without making cluster cohesion or separation any worse. A negative $s(x_i)$ suggests that assignment of x_i to this cluster is damaging cluster's cohesion and separation, whereas an $s(x_i)$ closer to 1 means the opposite. The validity of the whole clustering is then assessed by computing the Silhouette index, defined as $\frac{1}{N}\sum_{i\in X} s(x_i)$. This technique provides a succinct graphical representation of how well each object lies within its cluster (Rousseeuw, 1987). Figure 2.5 shows an example of a silhouette analysis. Making use of the silhouette, one can guess that the correct number of clusters is $k = 4$, as its silhouette is higher and instances are better distributed.



(a)

(b)

Figure 2.5: Example of a Silhouette analysis. (a) shows the silhouette of the clustering for $k = 4$, while (b) shows it for $k = 5$. As it is observed in this example, the silhouette is higher for $k = 4$, meaning that it is better to chose this clustering due to its clusters cohesion and compactness. Image extracted from `http://scikit-learn.org/`.

The silhouette can be calculated with any distance metric, such as the Euclidean distance or the Manhattan distance.

The literature indicates that there is no sole cluster validity index with a clear advantage over the others in every case (Bezdek and Pal, 1998). However, the Silhouette width index has performed well in many comparative experiments (Arbelaitz et al., 2013; Pollard and Van Der Laan, 2002).

Dunn's index

Dunn's index (Dunn, 1974) is defined as the ratio of the smallest distance between clusters, which estimates the separation of clusters, and the maximum cluster diameter, which estimates its cohesion. This index allows for general distance measures and was applied here with the Euclidean as well as general Minkowski distances.

Dunn's index is not without flaws. Possibly the most relevant in relation to this thesis is its sensitivity to the information in noisy features. However, this index does provide a rich and very general structure for defining cluster validity indexes for different types of clusters (Bezdek and Pal, 1998).

2.2.4 t-Distributed Stochastic Neighbor Embedding

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a machine learning algorithm developed by Geoffrey Hinton and Laurens van der Maaten (Maaten and G. Hinton, 2008) with the objective of reducing the dimensionallity of a given dataset for visualization purposes. To achieve it, t-SNE algorithm minimizes the divergence between two distributions: on the one hand, a distribution that measures pairwise similarities of the input objects and, on the another hand, a distribution that measures pairwise similarities of the corresponding low-dimensional points in the embedding. Let's assume a data set of high-dimensional input objects $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ and a function $d(\mathbf{x}_i, \mathbf{x}_j)$ that computes a distance between a pair of objects, for example, the Euclidean distance $d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2$. The objective of the algorithm is to learn an s-dimensional embedding in which each object is represented by a point, $\mathcal{E} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ with $\mathbf{y}_i \in \mathbb{R}^s$ (typical values for s are 2 or 3). To this end, t-SNE defines joint probabilities p_{ij} that measure the pairwise similarity between objects $\mathbf{x}_i$ and $\mathbf{x}_j$ by symmetrizing two conditional probabilities

$$p_{j|i} = \frac{\exp\left(-d(\mathbf{x}_i, \mathbf{x}_j)^2/2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-d(\mathbf{x}_i, \mathbf{x}_k)^2/2\sigma_i^2\right)}\,, \tag{2.21}$$

$$p_{i|i} = 0\,, \quad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}\,. \tag{2.22}$$

In Equation 2.21, the bandwidth of the Gaussian kernels (σ_i) is set in such a way that the perplexity of the conditional distribution P_i equals a predefined perplexity u. The perplexity can be interpreted as a smooth measure of the effective number of neighbors and is defined as

$$Perp(P_i) = 2^{H(P_i)}\,, \tag{2.23}$$

where $H(P_i)$ is the Shannon entropy of P_i measured in bits

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}\,. \tag{2.24}$$

Consequently, the optimal value of σ_i varies per object: in regions of the data space with a higher data density, σ_i tends to be smaller than in regions of the data space with lower density. The optimal value of σ_i for each input object can be found using a simple binary search (G. E. Hinton and Roweis, 2003) or using a robust root-finding method (Vladymyrov and Carreira-Perpinán, 2013).

In the s-dimensional embedding $\mathcal{E}$, the similarities between two points $\mathbf{y}_i$ and $\mathbf{y}_j$, that is, the low-dimensional models of $\mathbf{x}_i$ and $\mathbf{x}_j$, are measured using a normalized heavy-tailed kernel. In particular, the embedding similarity q_{ij} between the two points $\mathbf{y}_i$ and $\mathbf{y}_j$ is computed as a normalized t-Student kernel with a single degree of freedom (or Cauchy)

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j)\|^2)^{-1}}{\sum_{k \neq l}(1 + \|\mathbf{y}_k - \mathbf{y}_l)\|^2)^{-1}}\,, \quad q_{ii} = 0\,. \tag{2.25}$$

t-SNE employs the t-Student with one degree of freedom distribution for the map points in order to avoid imbalance in the distribution of the distances of a point neighbors. Using the Gaussian distribution for the map points would get such imbalance because even though the distribution of the distances is very different between a high-dimensional space and a low-dimensional space, the algorithm tries to reproduce the same distances in the two spaces. This imbalance would lead to an excess of attraction forces and a sometimes unappealing mapping. This is actually what happens in the original SNE algorithm, by Hinton and Roweis (G. E. Hinton and Roweis, 2003). However, the t-Student with one degree of freedom (or Cauchy) distribution has a much heavier tail than the Gaussian distribution, which compensates the original imbalance. Therefore, for a given similarity between two data points, the two corresponding map points will need to be much further apart in order for their similarity to match the data similarity. Figure 2.6 shows this effect.

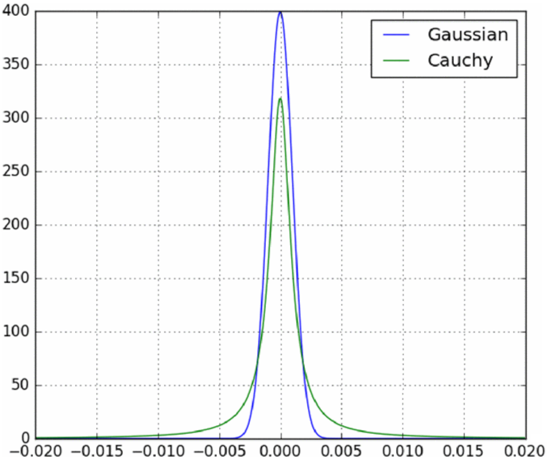


Figure 2.6: Gaussian distribution versus Cauchy distribution.

Concretely, the heavy tails of the normalized t-Student kernel allow dissimilar input objects $\mathbf{x}_i$ and $\mathbf{x}_j$ to be modeled by low-dimensional counterparts $\mathbf{y}_i$ and $\mathbf{y}_j$ that are too far apart. As aforementioned, this is desirable because it creates more space to accurately model the small pairwise distances (i.e., the local data structure) in the low-dimensional embedding.

The locations of the embedding points $\mathbf{y}_i$ are determined by minimizing the Kullback-Leibler divergence between the joint distributions P and Q:

$$C(\mathcal{E}) = KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \,. \tag{2.26}$$

Due to the asymmetry of the Kullback-Leiber divergence, the objective function focuses on modeling high values of p_{ij} (similar objects) by high values of

q_{ij} (nearby points in the embedding space). The objective function is non-convex in the embedding $\mathcal{E}$. It is typically minimized by descending along the gradient

$$\frac{\partial C}{\partial \mathbf{y}_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij}) q_{ij} Z (\mathbf{y}_i - \mathbf{y}_j) \,, \tag{2.27}$$

where the normalization therm $Z = \sum_{k \neq l} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}$.

It is straightforward to see that the evaluation of the joint distributions P and Q is $\mathcal{O}(N^2)$, because both distributions involve a normalization term that sum over all $N(N-1)$ pairs of unique objects. Since t-SNE scales quadratically in the number of objects N, its applicability is limited to data sets with only a few thousand input objects; beyond that, learning becomes too slow to be practical (and the memory requirements become too large) (Van Der Maaten, 2014). An example of how t-SNE is able to reduce the dimensionality of a complex dataset is shown in Figure 2.7, where a representation of the MNIST[1]data-set is shown.

2.2.5 Facial landmarks detection

The construction and alignment of generic deformable models able to capture the variability of a non-rigid object is one of the most popular and well-studied problems in the scope of computer vision. Particularly, the non-rigid object most studied is the face. The detection of facial landmarks, also known as facial feature points or facial fiducial points, plays an essential role in many face analysis tasks, such as face recognition (Scheenstra et al., 2005; Shi et al., 2006)), face morphing (Wolberg, 1998), face tracking (M. Kim et al., 2008), head pose estimation (Zhu and Ramanan, 2012), attribute inference (Kumar et al., 2009; Luo et al., 2013) or emotion recognition (Fasel and Luettin, 2003; Ko and Sim, 2010). Facial landmark detectors normally include a face detector prior to the landmark detection (the most commonly used is the Viola-Jones detector (Viola and Jones, 2004)) whose goal is to locate the face within the image and thus reduce the search space. After locating the face, a method to obtain the facial feature points is applied only in the space where the face should be (Figure 2.8). Facial landmarks have semantic meaning, that is, they are located around facial features such as eyebrows, eyes, nose, mouth and chin.

[1] The MNIST dataset is a large database of handwritten digits that is commonly used for training various image processing systems. It is comprised of 70000 gray-scale images of 28x28 pixels representing numbers from 0 to 9.

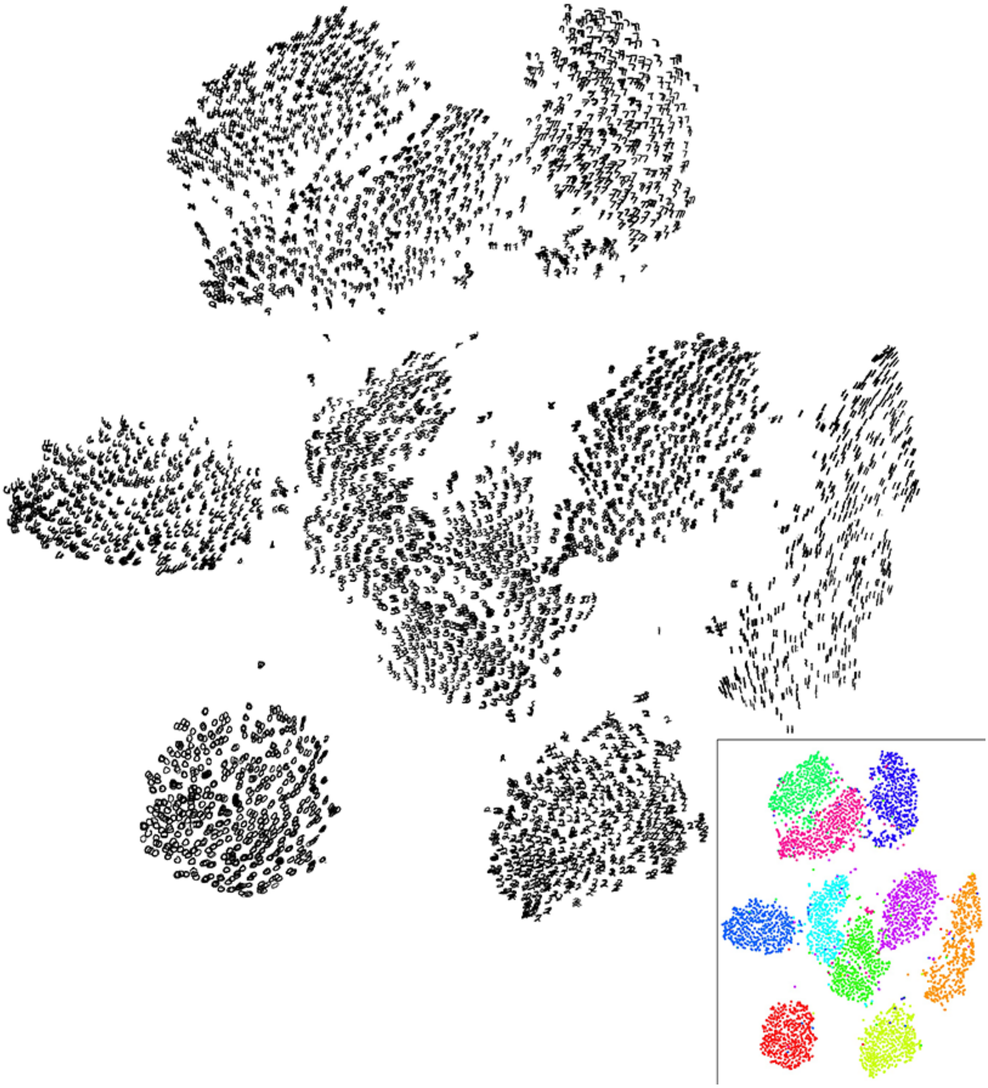


Figure 2.7: Graphical representation of instances of MNIST dataset using t-SNE. Image extracted from Maaten and G. Hinton, 2008.

According to T. Cootes, Taylor, et al. (1995), there are three types of facial feature points: application-dependent, such as the sharp corners of a boundary or the center of a mouth or eye; application-independent, such as the highest point on a face in some defined orientation (the highest point along

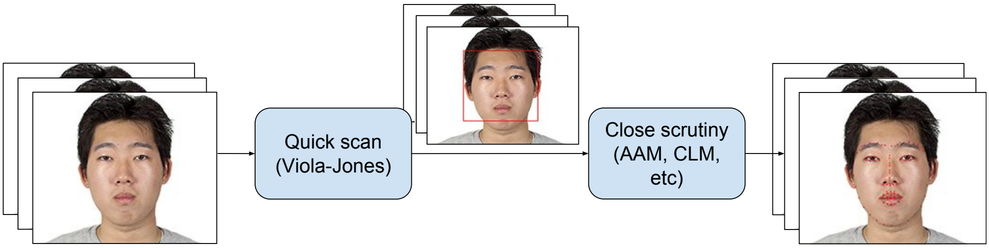


Figure 2.8: Workflow of facial landmark detection.

the nose's bridge); and hybrid points from the two previous types, such as points describing the chin. A model of facial landmarks can be composed of different number of points, for example, a 17-point model, 29-point model or 68 point-model. Following the division performed by N. Wang et al. (2017), the existing methodologies can be classified into four groups: constrained local model (CLM)-based methods, active appearance model (AAM)-based methods, regression-based methods and other methods.

CLM-based methods make use of local experts in order to calculate a response map accounting for the appearance variation around each facial feature point independently. These response maps are refined by a shape prior normally learned from training shapes and then used to predict the facial landmarks. AAM-based methods approach the appearance variation modelling from a holistic perspective and make use of training shapes as well. Regression-based methods, in contrast, directly estimate the shape from the appearance regression-based methods without any shape model or appearance model. "Other methods" can be further divided into four sub-categories: graphical model-based methods, joint face alignment methods, independent facial feature methods, and deep learning-based methods.

With independence of the number of facial landmarks detected and the approach employed, these points should locate several important commonly-used areas, such as the eyes, nose and mouth. These are the areas that carry most of the information for both generative and discriminative purposes (Asthana et al., 2014; N. Wang et al., 2017).

In this work, an AAM-based method was used to detect the internal features landmarks (eyebrows, eyes, nose and mouth) and a CLM-method to detect the face contour landmarks (Figure 2.9). Therefore, in the following subsections, these two methods will be explained in-depth.

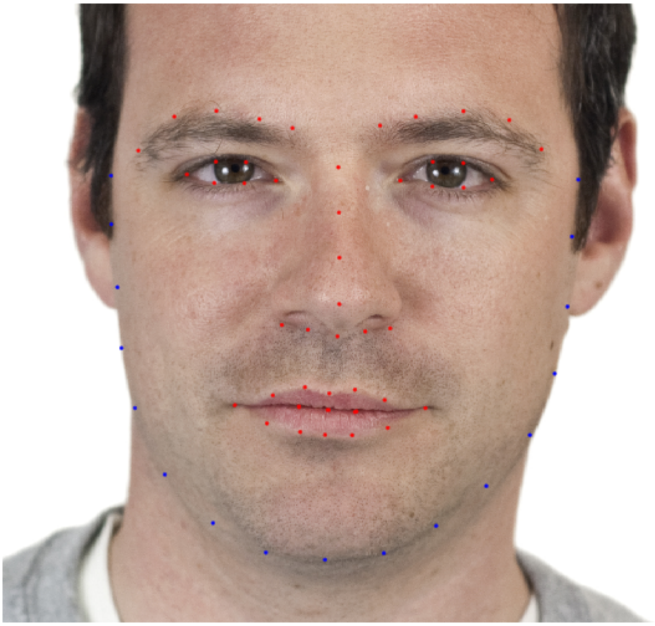


Figure 2.9: Facial feature point distribution indicating the internal features of the face (AAM-based method, red color) and the face contour (CLM-based method, blue color).

Active Appearance Models

The Active Appearance Model (AAM) is a generalization of the widely used Active Shape Model approach. The latter utilizes a statistical model of shape to match a set of model points to an object in an image while the former seeks to match both the model points and a representation of the texture of the object to an image (T. Cootes, Edwards, et al., 1999). This means that AAM models build a parametric model of the face shape and its appearance. This is normally done by using linear models, meaning that the shape and appearance is given as a linear combination of a set of template shapes and appearances, which are learned from examples of facial images manually annotated. These models, also called generative models, allow generating synthetic faces. In fact, in order to fit the generative model to the target face, the model searches for the most similar synthetic face to the input face.

An AAM model can be divided into three integral building blocks: a linear shape model, a linear texture (or appearance) model and a deformation model.

Linear shape model

A shape model is usually learned from training facial shapes and is taken as the prior refining the configuration of facial landmarks. This prior consists of a statistical distribution of facial landmarks (also known as the point distribution model (PDM) proposed by T. Cootes and Taylor (1992), Figure 2.10).

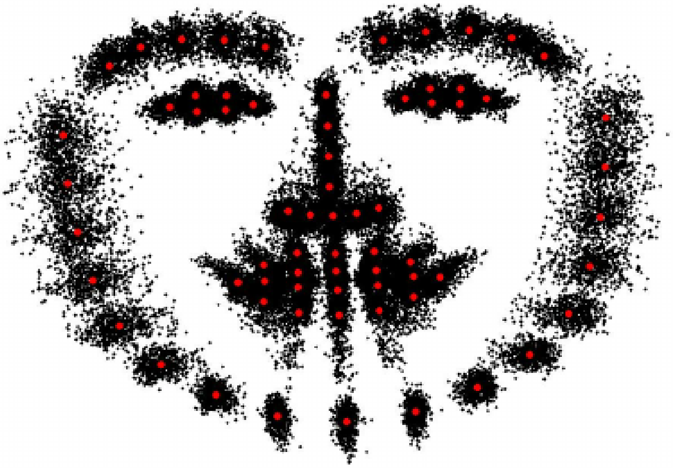


Figure 2.10: Illustration of statistical distribution of facial landmarks. There are 600 shapes (smaller dot points in black) normalized by Procustes analysis. The larger red dot points indicate the mean shape of the 600 shapes. Image extracted from N. Wang et al. (2017).

Let's see how a linear shape model is computed. A shape is represented by $\mathbf{s} = (x_1, y_1, \ldots, x_L, y_L)^\top \in \mathbb{R}^{2L}$, a vector composed of (x, y)-coordinates of L landmark points connected to a triangulated mesh. Let $\tau = \{(\mathbf{I}^1, \mathbf{s}^1), \ldots, (\mathbf{I}^m, \mathbf{s}^m)\}$ be a training set of facial images $\mathbf{I}^j$ and corresponding shapes $\mathbf{s}^j$. The shape model can be obtained by applying the principal component analysis (PCA) on all the aligned training shapes (Matthews and Baker, 2004; N. Wang et al., 2017). This alignment is performed by means of the Procustes Analysis (Goodall, 1991) and removes the similarity transformations from the original shapes $\mathbf{s}^j$. Then, a shape $\mathbf{s}$ generated by the model is represented as:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} \alpha_i \mathbf{s}_i \,, \tag{2.28}$$

where $\mathbf{s}_0$ is the mean of all these aligned training shapes, $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ are n template shapes which correspond to the n largest eigenvectors of the covariance matrix calculated from the similarity-free shapes (where n is usually chosen to preserve 90%-80% of the variance); and $\alpha = (\alpha_1, \ldots, \alpha_n)^\top \in \mathbb{R}^n$ is a vector of shape parameters.

Since the shapes utilized to construct the model had their similarity removed by the Procustes Analysis, global transformations like rotation, scaling or translation are not captured by it. This is resolved by a deformation model (Matthews and Baker, 2004).

Deformation model

Two frames can be differentiated, the reference frame and the image frame. The reference frame is composed by the transformation-free images, while the image frame consists of the aligned images. In order to obtain a shape $\mathbf{x}$ in the image frame, a deformation model is used. The deformation model is defined by a function $\mathbf{x} = \mathbf{W}(\mathbf{s}; \alpha)$ which warps pixel coordinates in the reference frame $\mathbf{s} = (x, y)$ to coordinates in the image frame $\mathbf{x} = (x, y)$. The vector α encapsulates the shape parameters as in the linear shape model. In other words, this function transforms the points of a shape (usually $\mathbf{s}_0$) to any possible shape defined by Equation 2.28:

$$\mathbf{x}_{point} = s\mathbf{R}\mathbf{s}_{point} + \mathbf{t}_{point} \,, \tag{2.29}$$

where s is a scale factor, R is a rotation matrix, $\mathbf{s}_{point}$ denotes a rearranged $2 \times N$ matrix where each column corresponds to one point in the shape $\mathbf{s}$ and $\mathbf{t}_{point}$ consists of N replications of the translation vector $\mathbf{t}$, one for each point of $\mathbf{s}_{\mathbf{point}}$. Similarly, $\mathbf{x}$ is the rearrangement of $\mathbf{x}_{point}$.

Linear texture model

The texture model is normally defined within the mesh obtained by triangulation of the mean shape $\mathbf{s}_0$. To build the texture model, every training face in τ must be warped to the mean-shape frame, so the resultant images are free of shape variation, thus accounting only for texture variations. These images are called shape-free textures, and are usually obtained by triangulation or thin plate spline method. Here, the deformation model $\mathbf{W}(\mathbf{s}; \alpha)$ is used to this end. Then, the shape-free textures are transformed into a grey-level vector $\mathbf{z}_i$, which is normalized by a scaling u and offset v to avoid the effect of global lighting:

$$\mathbf{a} = \frac{\mathbf{z}_i - v \cdot \mathbf{1}}{u}, \tag{2.30}$$

where u and v are the variance and the mean of the texture $\mathbf{z}_i$ respectively, and $\mathbf{1}$ is a vector of 1s with the same length as $\mathbf{z}_i$. Finally, the texture model is obtained by applying PCA on all the geometrically normalized textures:

$$\mathbf{a} = \mathbf{a}_0 + \sum_{i=1}^{m} \beta_i \mathbf{a}_i , \tag{2.31}$$

where $\mathbf{a}_0$ is the mean appearance, $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ are m template textures and $\beta = (\beta_1, \ldots, \beta_m)^\top \in \mathbb{R}^m$ is a vector of texture parameters.

In the above formulation, the shape and appearance models have their independent set of parameters (α and β, respectively). This is called *independent* AAM (Matthews and Baker, 2004). On the other hand, it is possible to merge their parameters into one only model, usually performing an additional PCA on top of the shape and texture templates. This variant is called *combined* AAM and is more general, often requiring more parameters to represent the same degree of accuracy as the independent AAM. In turn, fitting the model is more efficient and accurate, and coupling the parameters together is less restrictive to choice the fitting algorithm, as prevents the joint texture-shape templates to be orthogonal.

Fitting the models to an image

Once the shape and texture models have been obtained, the goal is to adjust them to an input image I by finding a set of parameters α_i $(i = 1, \ldots, n)$ and β_i $(i = 1, \ldots, m)$ which solve the following non-linear least squares problem:

$$\min_{\alpha,\beta} \left[\mathbf{a}_0 + \sum_{i=1}^{m} \beta_i \mathbf{a}_i - I(\mathbf{W}(\mathbf{s};\alpha)) \right]^2 . \tag{2.32}$$

Matthews and Baker (2004) shown that this problem can be decomposed into two independent sub-problems. First, finding the shape parameters α by solving Equation 2.33:

$$\min_{\alpha} \left[\mathbf{a}_0 - I(\mathbf{W}(\mathbf{s};\alpha)) \right]^2 . \tag{2.33}$$

Second, optimizing the texture parameters β, which have a closed solution. There exist a wide variety of methods for minimizing Equation 2.33, being the most commonly used the Lucas-Kanade algorihtm (Lucas, 1986; Lucas, Kanade, et al., 1981), a method based on the Gauss-Newton algorithm.

Constrained Local Models

Constrained Local Models (CLM) are composed of three main parts: a point distribution model (PDM), patch experts, and a fitting strategy. PDM is employed to model the localization of facial landmarks within the image using non-rigid shape and rigid global transformation parameters. Patch experts are used to model the appearance of local patches around landmarks of interest and are utilized to compute the response map which measures detection accuracy. Regarding the fitting approach, there are very varied strategies. One of the most popular is the Regularized Landmark Mean Shift (RLMS) (Saragih et al., 2011).

Point Distribution Model

The first stage, as with AAM, is to train the model on labelled examples. Once the model is trained, the objective is to estimate the rigid and non-rigid parameters $\mathbf{p}$, which fit the underlying image best:

$$\mathbf{p}^* = \arg\min_{P} \left[\mathcal{R}(\mathbf{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\mathbf{x}_i; \mathbf{I}) \right] , \tag{2.34}$$

where $\mathcal{R}$ is the regularization term that penalizes low likely or too complex shapes, and $\mathcal{D}$ accounts for the misalignment of the i_{th} landmark at the location $\mathbf{x}_i$ of the image $\mathbf{I}$. The location of the i_{th} feature $\mathbf{x}_i = [x_i, y_i, z_i]^T$ is controlled through the PDM by the parameter $\mathbf{p}$:

$$\mathbf{x}_i = s \cdot R_{2D} \cdot (\bar{\mathbf{x}}_i + \Phi_i \mathbf{q}) + \mathbf{t} \,, \tag{2.35}$$

where $\bar{\mathbf{x}}_i = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T$ is the mean value of the i^{th} feature, Φ_i is a $3 \times m$ principal component matrix, and $\mathbf{q}$ is a vector of parameters of length m controlling the non-rigid shape. The rigid shape parameters can be parameterized using the following 6 scalars: a scaling term s, a translation $\mathbf{t} = [t_x, t_y]^T$, and orientation $\mathbf{w} = [w_x, w_y, w_z]^T$. Rotation parameters $\mathbf{w}$ control the first two rows of the rotation matrix R_{2D}, which are in axis-angle form to facilitate their linearization. Then, the whole shape can be described by $\mathbf{p} = [s, \mathbf{t}, \mathbf{w}, \mathbf{q}]$.

Patch experts

Patch experts (also known as local detectors) compute response maps on the local region around the facial feature points, that is, they evaluate the probability of a facial feature point being aligned at a determined pixel location. The response of the i^{th} patch expert π_{x_i} at the image location $\mathbf{x}_i$ based on the surrounding support region is defined as:

$$\pi_{x_i} = \mathcal{C}_i(\mathbf{x}_i; \mathbf{I}) \,, \tag{2.36}$$

where $\mathcal{C}_i$ is the output of the i_{th} feature regressor. This regressor can model the misalignment from 0 (no alignment) to 1 (total alignment). There are numerous methods that have been used as patch experts: a distance metric such as the Mahalanobis distance, various Support Vector Regressor (SVR) models and logistic regressors, or even simple template matching techniques. The most commonly used patch expert is the linear Support Vector Regressor in combination with a logistic regressor (Baltrusaitis et al., 2013). The reason for using linear SVR is their efficient implementation on images due to the possibility of using convolutions. In Figure 2.11 some response maps computed with SVR can be observed.

Fitting the model

The fitting of CLM-based methods consists of two main steps: (1) predicting local displacements of shape model points and (2) constraining the configuration of all points to comply with the shape model. These two steps are iterated until a converge criterion is satisfied.

To fit the model to a new incoming face, CLM-based methods update an initial parameter estimate $\mathbf{p}_0$ (most usually from a face detector) to get closer to a solution $\mathbf{p}^* = \mathbf{p}_0 + \Delta\mathbf{p}$ (where $\mathbf{p}^*$ is the optimal solution). Therefore, the iterative fitting objective is as follows:

$$\arg\min_{\Delta\mathbf{p}} \left[\mathcal{R}(\mathbf{p}_0 + \Delta\mathbf{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\mathbf{x}_i; \mathbf{I}) \right] . \tag{2.37}$$

Equation 2.37 can be solved using several methods, but the most commonly used is the Regularized Landmark Mean Shift (Saragih et al., 2011).

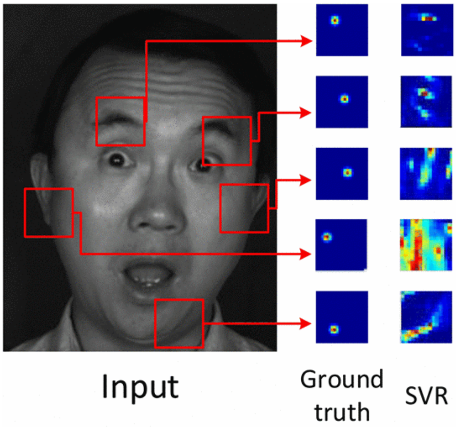


Figure 2.11: Sample response maps from patch experts of five features. The ground truth column is the ideal response, and the SVR column the one achieved by a SVR model used by CLM approaches. Red corresponds with high probability. Image extracted from Baltrusaitis et al. (2013).

2.3 Database

The database employed in this work is the Chicago Face Database (Ma et al., 2015), which is formed by 290 male and 307 female faces of ages ranging from 17 to 65 and varying ethnicity (Asian, Black, Latino and White). Each target in the database is represented with a neutral expression photo that has been normalized by an independent rater sample. In addition, for a small subset of targets some facial expressions are also available: happy (with open mouth and visible teeth), happy (with closed mouth), angry and fearful. As our purpose is to predict which impression the target conveys to the observer removing any possible trace of emotion, only neutral faces are used. By doing this, the target face will not be expressing any emotion, so the impression perceived by the observer will be due only to the structural configuration of the target face, instead of the emotion they could be expressing in case of a non-neutral face.

All photographs are normalized and have the same size, illumination conditions and position. In order to evaluate how the implemented evaluation model works, the rating information available with the database is employed. For each target, there are both physical and subjective attributes rated by a minimum of 20 independent judges. For this work, only subjective attributes were used, which were rated considering the person with respect to other people of the same race and gender using a 1-7 Likert scale (Likert, 1932). Table 2.1 shows a list of all the subjective attributes used.

Table 2.1: List of employed subjective attributes.

List of subjective attributes				
Afraid	Angry	Attractive	Babyface	Disgusted
Dominant	Feminine	Happy	Masculine	Prototypic
Sad	Surprised	Threatening	Trustworthy	Unusual

Most of the named subjective attributes are very basic and do not need an explanation, however, it is important to clarify what *Prototypic* and *Unusual* means. *Prototypic* measures racial prototypicality. In other words, it measures how much a face physical features resemble the features of people belonging to the same ethnicity. For example, for an Asian face, it would measure if its eyebrows, eyes, nose, cheeks, lips, and other physical features, are more Asian (i.e., typical of Asians) or less Asian (i.e., less typical of Asians). On the other hand, *Unusual* measures if the face being evaluated would stand out in a crowd (highly unusual) or not (little unusual). It is important to note that every face elicits every social trait to some extent. Therefore, the same face could be 60% *Happy* and 20% *Sad* (and would have a level for each of the social traits left), which would mean that to most observers, it resembles more a happy face than a sad one.

As perception and trait impressions are not ethnicity independent (Walker, Jiang, et al., 2011), the objective of this work was to create a dedicated predictor for each ethnicity individually. In addition, as this work is a proof of concept, only male gender was considered for the sake of simplicity. This allowed to avoid problems with long hair and make-up to some extent. Then, the procedure described in this work was applied to each male ethnicity, resulting in 290 male targets (52 Asian, 93 Black, 52 Latino and 93 White). The extension to female gender remains as future work, although it should require little to no modification in the method pipeline.

2.4 Automatic facial feature extraction

Faces are composed by different features, such as eyebrows, eyes, noses, mouths, etc. Furthermore, these features are located in a deterministic place within the face. If one is to define a face, the first step is to characterize these features, and the second one, to establish their locations. Then, the procedure followed in this work to completely characterize a face includes choosing which features are to be used and the distances to a reference point within the face. The facial features employed in this work are divided into three groups: internal, external and distance features. Internal features are composed by the eyebrow, the eye, the nose and the mouth. Only one eyebrow and eye need to be defined in order to completely characterize a face because, in this thesis, faces are considered symmetric. Therefore, the left side was randomly chosen to characterize the eyebrows and eyes, so the left hand side eyebrow and eye were chosen, and the right hand eyebrow and eye were obtained by horizontally flipping the left hand side ones. The external features refers to the jawline. On the other hand, distance features are composed by five distances which are used to specify the location of the before mentioned internal features, and are defined as follows:

- d_{eb}: distance from the lowermost jawline landmark to the centroid of the polygon formed by the eyebrow landmarks,
- d_e: distance from the lowermost jawline landmark to the centroid of the polygon formed by the eye landmarks,
- d_n: distance from the lowermost jawline landmark to the centroid of the polygon formed by the nose landmarks,
- d_m: distance from the lowermost jawline landmark to the centroid of the polygon formed by the mouth landmarks,
- d_{ee}: distance between the centroids of the polygons formed by both eyes landmarks.

Horizontal distances were not considered for any other feature than the eyebrow and eye because symmetry and centrality were assumed (faces in the real world are not, but it is assumed in this thesis for simplicity). Moreover, as eyebrows are always above the eyes, only one horizontal distance needs to be accounted for, the distance between eyes. This distance indicates how far of the vertical line of face symmetry eyebrows and eyes should be.

The procedure implemented to extract the facial features from the full-face images is completely automatic for all features. Hair is not considered due to the difficulty existent in automatic hair segmentation, which is an extensively studied problem which has not yet arrived to a good general solution (Aarabi, 2015; Guo and Aarabi, 2016; D. Wang et al., 2011). On the other hand, the jawline is extracted as a set of shapes and reference points, and thus has a different extraction procedure.

The procedure followed to extract the features starts with the facial landmarks detection. Once the landmarks are detected, those corresponding to internal features are used to create their respective masks. These are very coarse masks with the aim to define the region where each internal feature is located. Then, the extraction procedure begins.

Considering only *internal* features, that is, eyebrows, eyes, nose and mouth; the automatic extraction procedure is very similar for all of them. First, the face is aligned according to the landmarks of the processed feature. Then, the coarse mask is used to extract the target feature from the full-face image. At this point, the feature image is processed to remove all the non-desired regions. Finally, feature images are cropped so there is as few skin as possible while retaining the complete feature visible. Although the procedure is very similar for all of these features, there exist slight differences which will be pointed out in the next sections, where all the steps mentioned above are explained in-depth.

Although the pipeline for facial feature extraction was designed for the CFD database, it should be possible to employ the same procedure with any normalized face database (i.e. faces are the same size).

2.4.1 Facial landmarks detection

Facial landmarks are the base of the whole method. Given their importance, several facial landmark detection frameworks were tested. Finally, as mentioned in subsection 2.2.5, two algorithms were employed in this work:

- Chehra, a facial landmark detector developed by Asthana et al. (2014), who implemented an incremental discriminative facial deformable model trained by a cascade of regressors. Chehra is able to accurately locate the eyebrows, eyes, nose and mouth. However, it does not provide face contour landmarks.

- CLM-framework, which was developed by Thomas et al. (2016) and employs Constrained Local Models to locate the face landmarks, including the face contour.

Chehra was found to be more accurate and faster than CLM-framework, so this is the reason why two methods were used in order to extract the face landmarks: CLM for the jawline, and Chehra for the rest of facial traits. Figure 2.12 shows the facial landmark distribution for both frameworks.

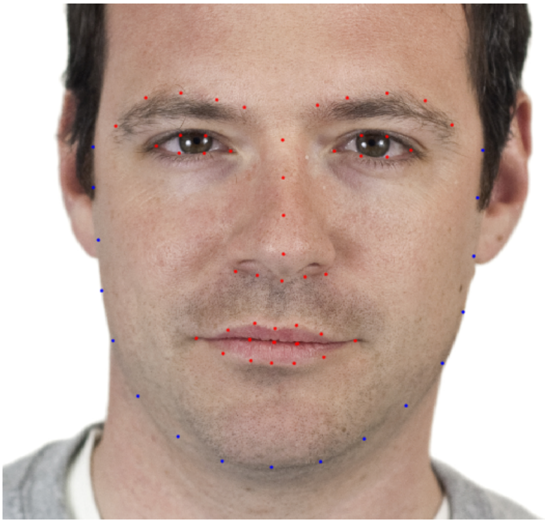


Figure 2.12: Facial landmark distribution for Chehra and CLM frameworks. In red, Chehra detected landmarks. In blue, CLM landmarks.

Then, in the first step of the developed method every face image is processed detecting its facial landmarks and storing them. These landmarks are used to identify and locate each feature so they can be properly extracted.

2.4.2 Feature masks creation

Masks are needed to extract each feature from the face removing as much skin and non-desired regions as possible. To do so, corresponding landmarks to each feature are identified and used to create a polygon by joining each of these feature landmarks (Figure 2.13).

These polygons are *thickened* with $n = \infty$ (i.e. unconnected objects are thickened by adding pixels to their exterior until when doing so would result in previously unconnected objects being 8-connected) obtaining the feature masks, which delimit the feature area. The result is shown in Figure 2.14.

Feature masks allow to extract each feature independently. Then, having introduced the facial landmarks and the feature masks, in the following sections the procedure followed to extract each feature is thoroughly explained.

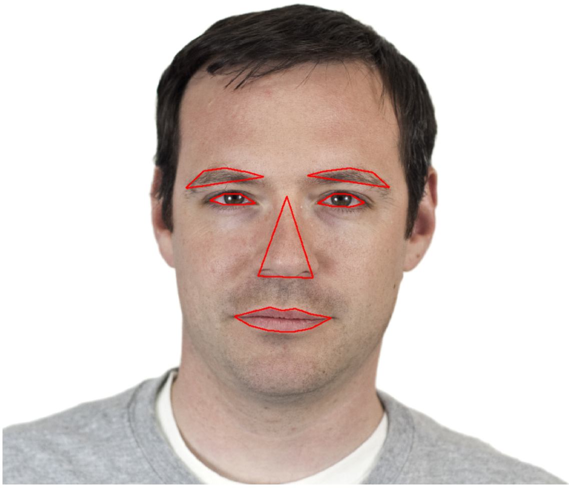


Figure 2.13: Polygonal feature masks created from the facial landmarks detected in Figure 2.8.

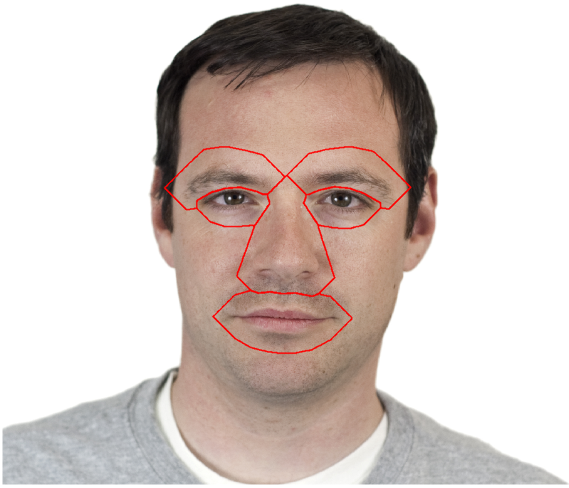


Figure 2.14: Feature masks are obtained by thickening polygonal feature masks with $n = \infty$.

2.4.3 Eyebrow extraction

The first facial feature extracted in this process is the eyebrow. In this case, as human faces have two eyebrows, CFD database is *doubled* by adding the horizontally flipped version of all the images to the existent ones. In such a way, original CFD images can be used to extract the left hand side eyebrows, and flipped CFD images provide the right hand side eyebrows flipped so they are comparable with the left hand side ones. Once the images are flipped, eyebrow landmarks are identified and all images are aligned using a point of the eyebrow as reference. Then, an improved mask to delimit the eyebrow is created from the thickened feature mask shown in Figure 2.14 and used to crop the full-face image to the eyebrow region only. Finally, in the last step, eyebrow images are cropped to the size of the biggest of the masks found. Figure 2.15 shows the flow of the eyebrow extraction procedure. In the following sections the different steps of this procedure are explained in-depth.

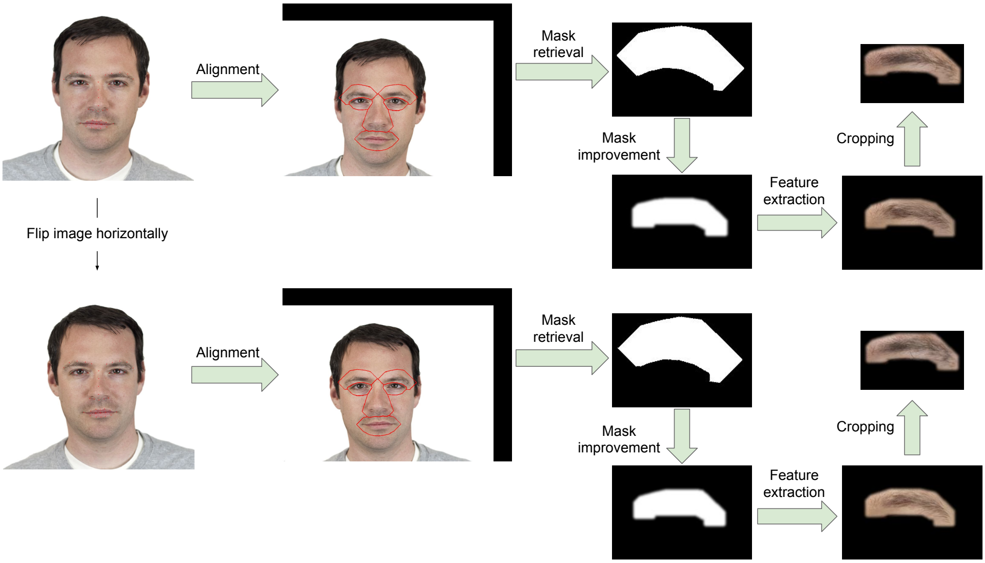


Figure 2.15: Eyebrow extraction flowchart.

Alignment

The alignment process of the eyebrows takes the top middle point of the detected eyebrow landmarks as reference and centers this point in the image. The top middle point is computed by assigning the x value of the middle point between the left-most and the right-most eyebrow landmarks to the x-coordinate and the y value of the up-most eyebrow landmark to the y-coordinate (Figure 2.16). Their alignment is completely necessary in order to be able to compare them. Without alignment, there would not be any feasible way to compare their sizes or shapes.

After this step, all top middle points for all eyebrows are at the same position along the image set, which means that all eyebrows are aligned.

Masking

In the previous section all images were aligned. Now, comparison is possible, but images contain a lot more information than just the eyebrow. Then, the objective of this step is to remove all this non-desired information. To do so, a precise mask is created for each eyebrow taking as starting point the feature mask obtained in subsection 2.4.2.

In the following lines each step is thoroughly explained following Figure 2.17, which illustrates the improved mask creation process. All the parameters utilized in this procedure were empirically chosen. The input to the process is the face image cropped to the size of the biggest existent thickened feature mask (a), which contains the eyebrow, but also the eye, skin and even hair. In order to remove these non-desired features, the eyebrow feature mask (b) is modified until a precise mask is obtained (k). The first step consists in eroding the feature mask with a vertical line of length $l = \lfloor (height(M_b)/r_f) \rfloor$, where M_b is the blob corresponding to the feature mask present in (b) and $r_f = 7$ is the reduction factor. This operation results in the reduced mask visible in (c). Next, vertical and horizontal *pikes* are removed from the mask. This is achieved by cropping the top, bottom, left and right of the reduced mask M_c, (c). Left and right sides are cropped by setting to black a number of pixels $dh = (rightmost\_point(M_c)) - leftmost\_point(M_c))/20$ on (c), resulting in (d). Top and bottom are handled in a different way. In order to get (e), (d) is *opened* with a vertical line of length $l = 6.5{\cdot}dh$ chosen empirically. At this point the mask has improved considerably by removing the hair and some skin, but the operations performed resulted in a misalignment with the eyebrow. Then,

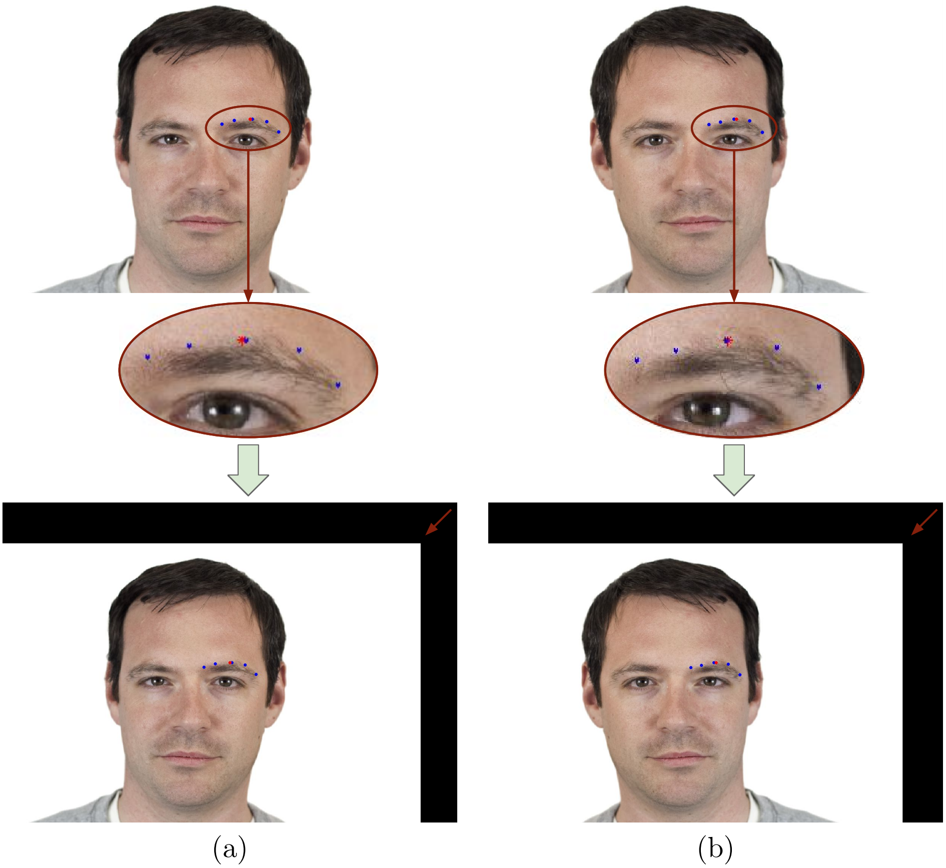


Figure 2.16: Eyebrow alignment process. (a) Refers to the original image, in which left hand eyebrow is centered. (b) Corresponds to the horizontally flipped image, in which the right hand eyebrow (now behaving as *left eyebrow*) is centered. Blue dots correspond to detected eyebrow landmarks, and the red dot to the computed reference (eyebrow landmark's top middle point).

the mask in (e) is shifted down a number of pixels $s_p = \lfloor (height(M_b)/s_r) \rfloor$, where $s_r = 10$, resulting in (f).

As was mentioned before, non-desired regions such as hair or skin have been already removed from mask (f), but there may still exist eye regions. To solve this, the blob in (f) is horizontally divided in 5 intervals, and a sub-image with the internal 3 intervals of the blow is taken (g) and eroded with a disk of radius $r = 2 \cdot dh$, resulting in (h). This sub-image is then pasted into (f), resulting

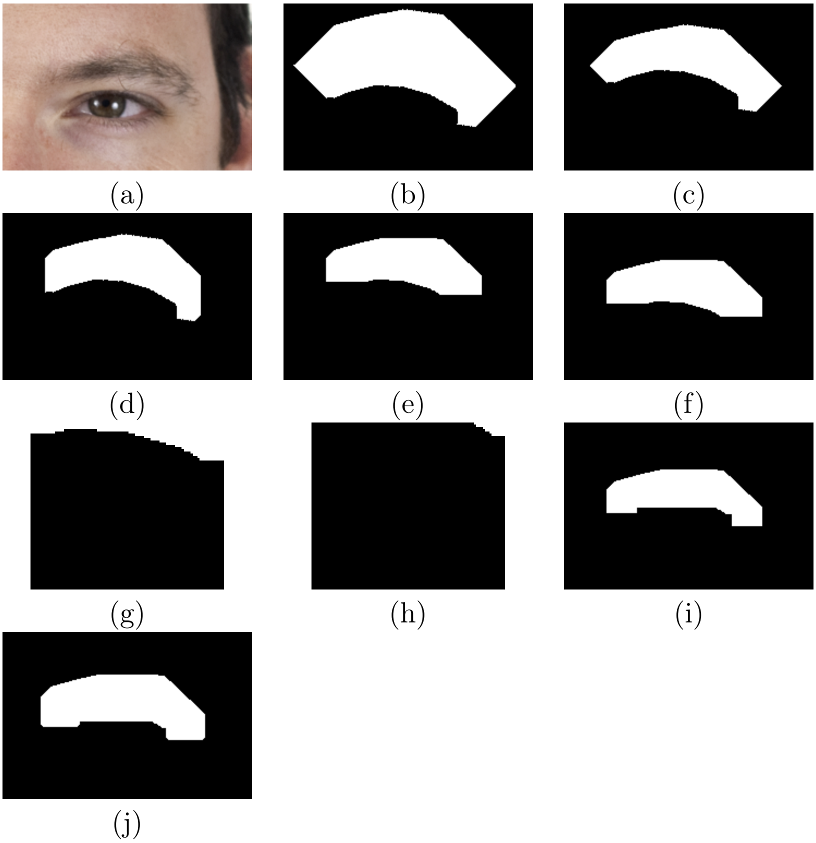


Figure 2.17: Eyebrow mask creation process.

in a mask which does not contain any of the non-desired eyes, (i). Finally, the last step is a dilation with a disk of radius $r = 5$ (j). At this point the mask needed to extract the eyebrow is already available.

This procedure is carried out on every eyebrow image, so at the end of the process, all eyebrows are saved independently and have a mask hiding all non-desired regions. However, these images are mostly composed by black pixels due to the improved mask, with a low amount of pixels used for eyebrow representation, which is a problem for the subsequent step, the clustering. Then, eyebrow images need to be cropped to remove as much black background as possible while keeping the same size for all of them.

Cropping

This step is not strictly necessary for feature extraction, however, it is for their posterior clustering. In this step eyebrows are cropped so the maximum amount of black pixels are removed while maintaining the same size for every image. This is achieved by computing the sizes of all the improved masks and cropping the eyebrow images to the biggest of the improved mask sizes (169 × 96 pixels). Figure 2.18 shows the result of this cropping.

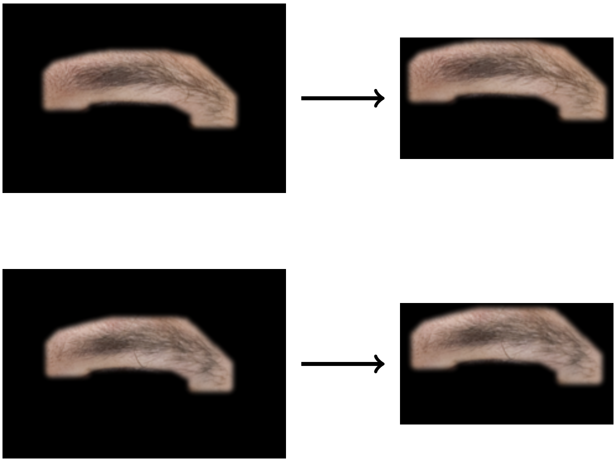


Figure 2.18: Eyebrow cropping for the clustering step.

At the end of this step, eyebrow images are ready to be clustered.

2.4.4 Eye extraction

As happened with the eyebrows, faces have two eyes as well, so the same procedure is followed to add the flipped version of all the CFD images to the existent ones so both eyes can be extracted as left hand side eyes. Once the images are flipped and included with the originals, eye landmarks and masks are identified. Eyes are then saved independently by means of the thickened feature masks shown in Figure 2.14. Then, all eyes are centered in the image using the centroid of the polygon formed by their landmarks as reference. To perform the centering, eye image canvas is expanded until the eye is placed in the center of the resulting image. This procedure is performed on every eye image. Finally, in the last step, eye images are cropped to an empirically found size which fits all eyes in a tight-fitting manner. Figure 2.19 shows the flow of the eye extraction procedure. In the following sections the different steps of this procedure are explained in-depth.

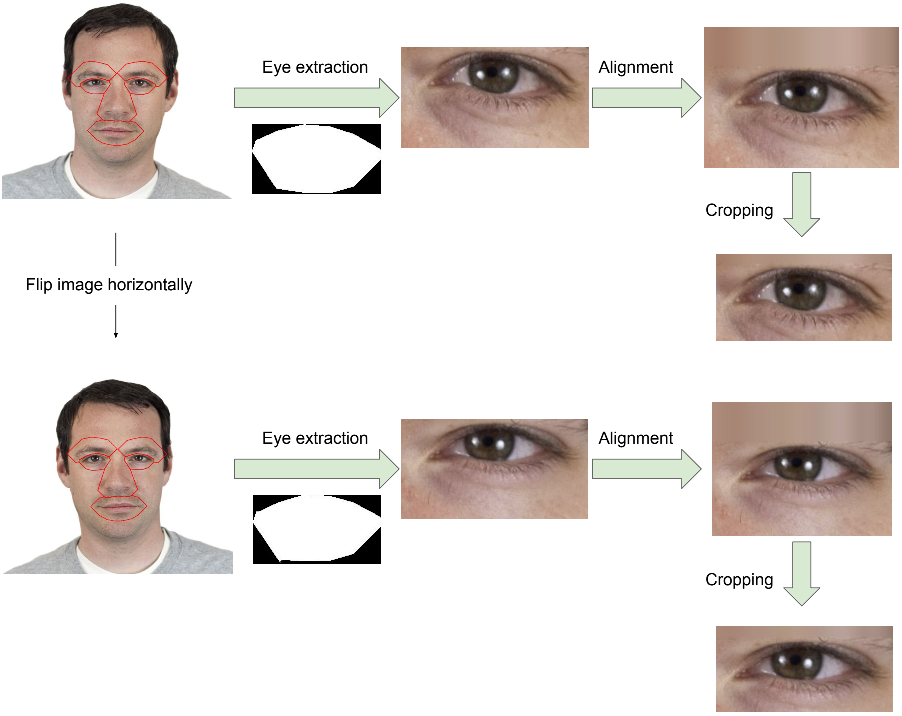


Figure 2.19: Eye extraction flowchart.

Extraction

The first step in the eye extraction pipeline is to extract and save each eye independently. To do so, the thickened feature masks are used, as shown in Figure 2.20 (a). For each eye, the bounding box of its corresponding thickened feature mask is computed (Figure 2.20 (b)). Then, the original image is cropped to this bounding box size (Figure 2.20 (c)). This is performed on every image so every eye is saved independently in a new image, but these images are not the same size and eyes are not aligned.

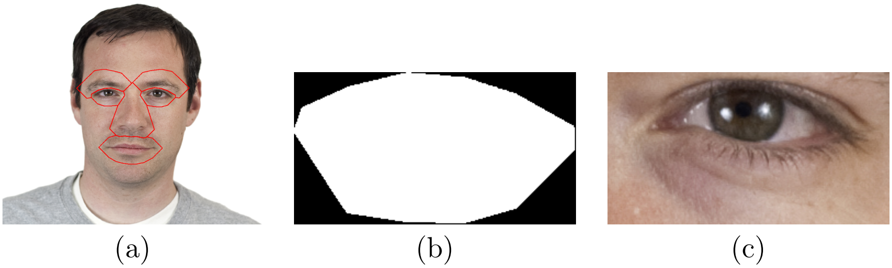


Figure 2.20: Eye extraction process. (a) Original image with the thickened feature masks visible. (b) Thickened feature mask employed to extract the eye. (c) Extracted eye.

Alignment

The alignment process of the eyes is different to the one used with eyebrows. In this case, a mask to hide the non-desired regions is not desirable, as it would hide expression lines and wrinkles. Therefore, no mask is employed with the eyes. Instead, eye images are cropped in a way that no eyebrow region is visible anymore. This is performed using each eye's thickened feature mask, and it usually yields eye images with the eye placed at the top half of the image. Then, to center the eye in the image, the distance from the eye's centroid to each extreme of the image is computed (Figure 2.21).

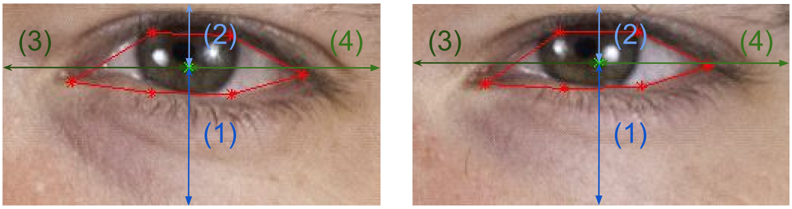


Figure 2.21: Alignment reference for left hand eye (left image) and flipped right hand eye (right image).

This yields four distances: $d_{h_{min}}$ (2), $d_{h_{max}}$ (1), $d_{v_{min}}$ (3) and $d_{v_{max}}$ (4), where $d_{h_{min/max}}$ and $d_{v_{min/max}}$ are the minimum and maximum horizontal and vertical distances from the centroid of the eye to the horizontal and vertical extremes of the image respectively. Next, eye image canvas is expanded to match $[d_{v_{max}}, d_{h_{max}}]$ pixels. The expansion is performed by repeating the last horizontal/vertical line of pixels respectively.

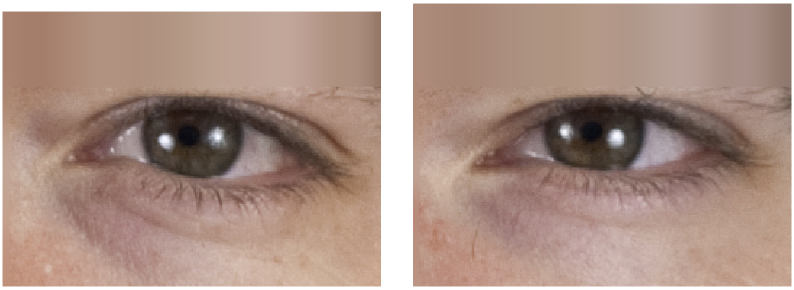


Figure 2.22: Eye alignment process. Left and right center aligned eyes.

This procedure allows for aligning the eye in the center of the image without using any kind of mask, and thus, without hiding any relevant information. However, although all eyes are now center aligned in their respective images, these images are still different sizes.

Cropping

In eye images obtained in last section there is still too much skin around the eye, which would lead the clustering algorithm to pay to much attention to skin, while disregarding eye features themselves. Furthermore, eye images are not the same size. Therefore, although this step is not strictly necessary for feature extraction, it is for their posterior clustering. Then, as with eyebrows, an extra step is carried out to remove as much skin as possible while retaining as much expression lines and wrinkles as possible and the same image sizes.

In this step eye images are cropped to an empirically established size which allows to remove a high amount of skin and at the same time keeps the whole eye and some expression lines within the final image. Figure 2.23 shows the result of this cropping, which results in 236 × 116 pixel images.

At the end of this step, eye images are ready to be clustered.

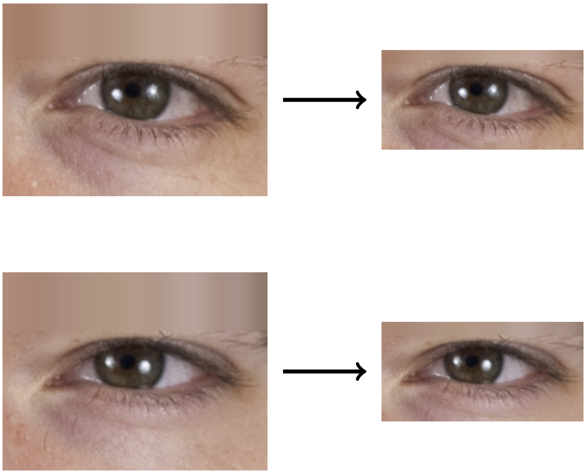


Figure 2.23: Eye cropping for the clustering step.

2.4.5 Nose extraction

Contrary to eyebrows and eyes, human face has only one nose, then it is not necessary to perform the *doubling* procedure done with previous features. With the nose, the first step is to identify its landmarks and mask. Then, images are aligned using the centroid of a triangle formed by some of the nose landmarks as reference. Next, the nose is extracted using the biggest thickened feature mask found among all nose masks. Finally, the resulting image is cropped so as few non-desired regions as possible remain visible. Figure 2.24 shows the flow of the nose extraction procedure. In the following sections the different steps of this procedure are explained in-depth.

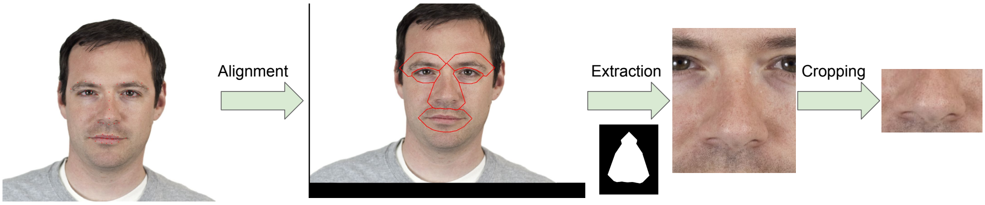


Figure 2.24: Nose extraction flowchart.

Alignment

The first step in the process carried out to extract the nose is the alignment. In this procedure, face images are processed so the nose ends up located in the center of the image. To perform the alignment, a reference point consistent across all images is necessary. In this case, the reference point was chosen as the centroid of a triangle formed by three of the nose landmarks (Figure 2.25). After performing this procedure, every nose is centered in its corresponding image.

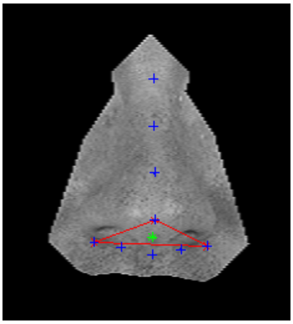


Figure 2.25: Method to compute the nose reference for alignment. In blue, the nose landmarks. In red, the triangle formed by the landmarks chosen to compute the reference point. In green, the centroid of the triangle, used for alignment.

Extraction

After performing the alignment, noses can be properly extracted from the full-face images by means of the biggest nose thickened feature mask found (Figure 2.26).

Thanks to the alignment, every resulting image will have the nose placed at the same point, making it easy to compare one to another. However, as can be observed in Figure 2.27, in this step some non-desired regions are still visible on the image. In next section, an explanation about how these regions are removed is presented.

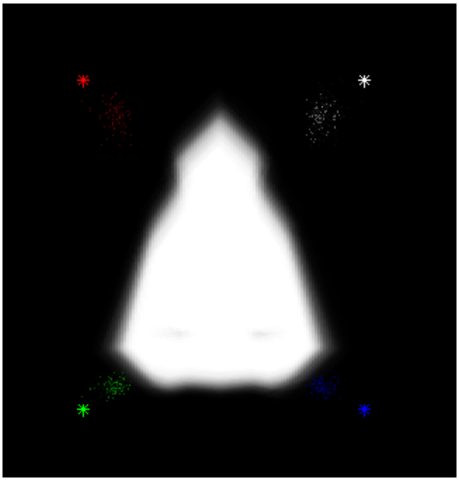


Figure 2.26: Biggest nose thickened feature mask found. Dots represent each existent nose thickened feature mask boundaries. Stars represent the boundaries of the biggest one.

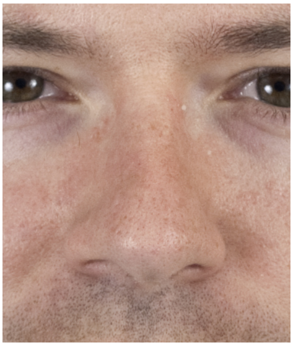


Figure 2.27: Nose after extraction.

Cropping

After alignment and masking of the nose images, the next step is the removal of the non-desired regions. To perform this task, all nose images are cropped to a sub-image frame empirically chosen. The result is an image of 187 × 118 pixels containing only the nose. An example is shown in Figure 2.28.

At the end of this stage, nose images are ready to be clustered.

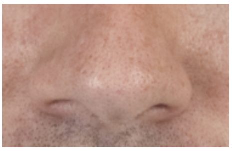


Figure 2.28: Extracted nose after masking.

2.4.6 Mouth extraction

Mouth extraction is performed in a similar manner to nose extraction. As with the previous features, the first step of the mouth extraction procedure consists in identifying its landmarks and mask. Then, images are aligned using the centroid of the mouth as reference, computed using its outer landmarks. Next, the mouth is extracted using the biggest thickened feature mask found among all mouth masks. Finally, the resulting image is masked, so beard and mustache (if any) are removed and only the mouth remains visible. Figure 2.29 shows the flow of the mouth extraction procedure. In the following sections the different steps of this procedure are explained in-depth.

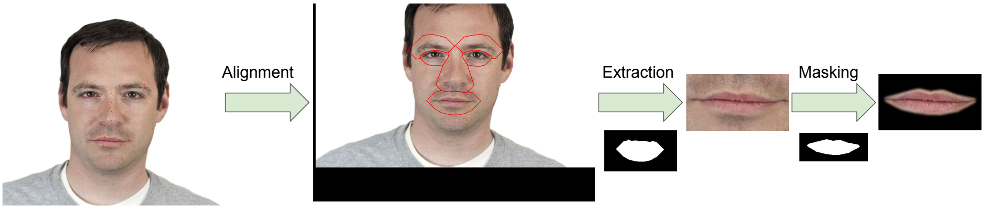


Figure 2.29: Mouth extraction flowchart.

Alignment

The first step in the process carried out to extract the mouth is the alignment. In this procedure, face images are processed so the mouth ends up located in the center of the image. To perform the alignment, a reference point consistent across all images is necessary. In this case, the reference point was chosen as the centroid of the outer mouth landmarks (Figure 2.30).

Thanks to the alignment, every resulting image will have the mouth placed at the same point, making it easy to compare one to another.

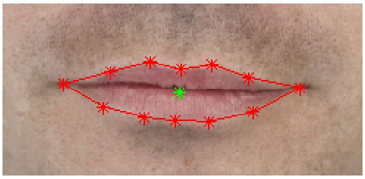


Figure 2.30: Method to compute the mouth reference for alignment. In red, the outer mouth landmarks. In green, its centroid, used for alignment.

Extraction

After performing the alignment, mouths can be properly extracted from the full-face images by means of the biggest mouth thickened feature mask found (Figure 2.31).

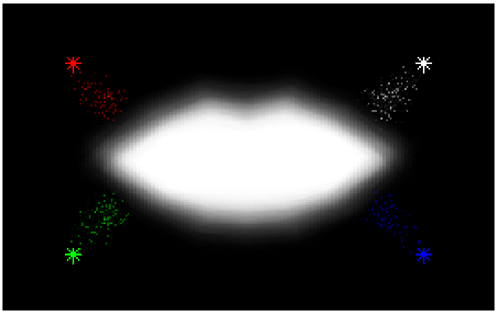


Figure 2.31: Biggest mouth thickened feature mask found. Dots represent each existent mouth thickened feature mask boundaries. Stars represent the boundaries of the biggest one.

In this case, the biggest mask found is 191 × 104 pixels. Therefore, all resulting images from this step will be this size. However, as can be observed in Figure 2.32, mouths might present beard or moustache, something undesired as it would distort the subsequent clustering. Then, in next section, a mask is created to hide all possible non-desired region.

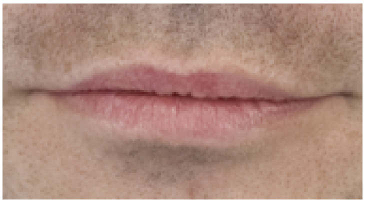


Figure 2.32: Extracted mouth.

Masking

After alignment and extraction of the mouth images the next step is the masking of the non-desired regions. These regions are those where beard or mustache might be present. To perform this task, the mouth polygonal feature mask explained in subsection 2.4.2 is employed (Figure 2.33).

The result is an image of 191 × 104 pixels with only the mouth visible. An example is shown in Figure 2.33.

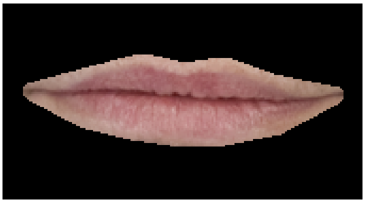


Figure 2.33: Mouth masking step.

At the end of this stage, mouth images are ready to be clustered.

2.4.7 Jawline extraction

Jawlines are characterized as a set of points delimiting the face contour and the centroids of the internal facial features. CFD images are already normalized and aligned, so there is no need for alignment in this case. Figure 2.34 shows the face contour landmarks distribution, composed by a total of 17 landmarks, and the computed centroids of the internal features, formed by the centroid coordinates of the eyes, the nose and the mouth. Internal feature centroids are also taken as part of the jawline characterization because not only the shape of the jawline is important, but how the features are distributed with regard to it. In total, 21 points are used in order to characterize the face jawline.

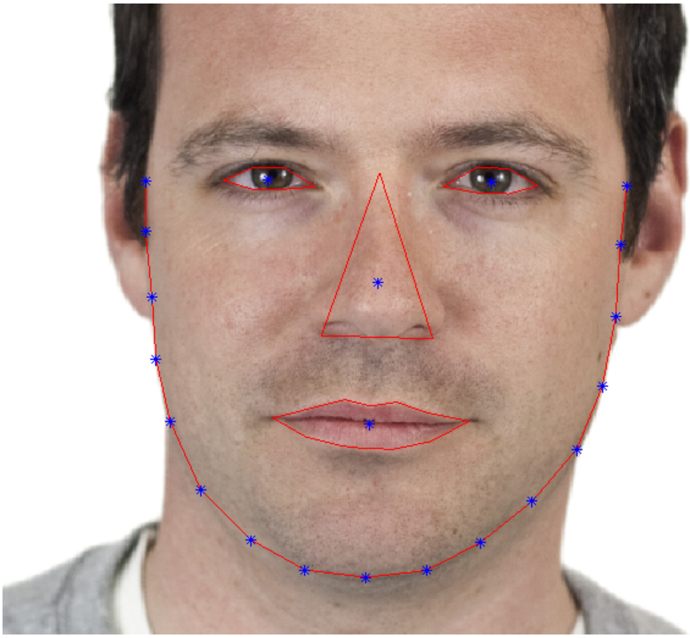


Figure 2.34: Jawline characterization.

2.4.8 Distances extraction

Human faces have different feature distributions, so it is important to model these differences if a realistic looking face is to be generated. Therefore, the following five distances were used to characterize the feature positions within the face, all taking the lowermost jawline landmark as reference:

- d_{eb}: distance from the mean point of the two eyebrow centroids to the lowermost jawline landmark (1),
- d_e: distance from the mean point of the two eye centroids to the lowermost jawline landmark (2),
- d_n: distance from the nose centroid to the lowermost jawline landmark (3),
- d_m: distance from the mouth centroid to the lowermost jawline landmark (4),
- d_{ee}: distance between both eye centroids (5).

Figure 2.35 shows these distances on a face within the CFD database.

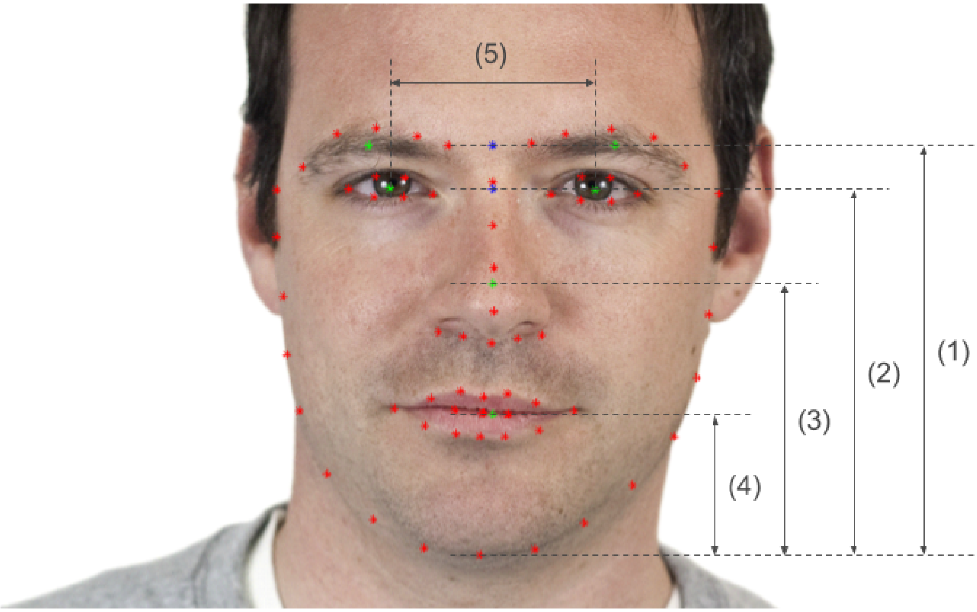


Figure 2.35: Distances characterization. In red, all detected facial landmarks. In green, feature computed centroids. Blue stars denote the eyebrows and eyes centroid, respectively.

2.5 Facial features clustering

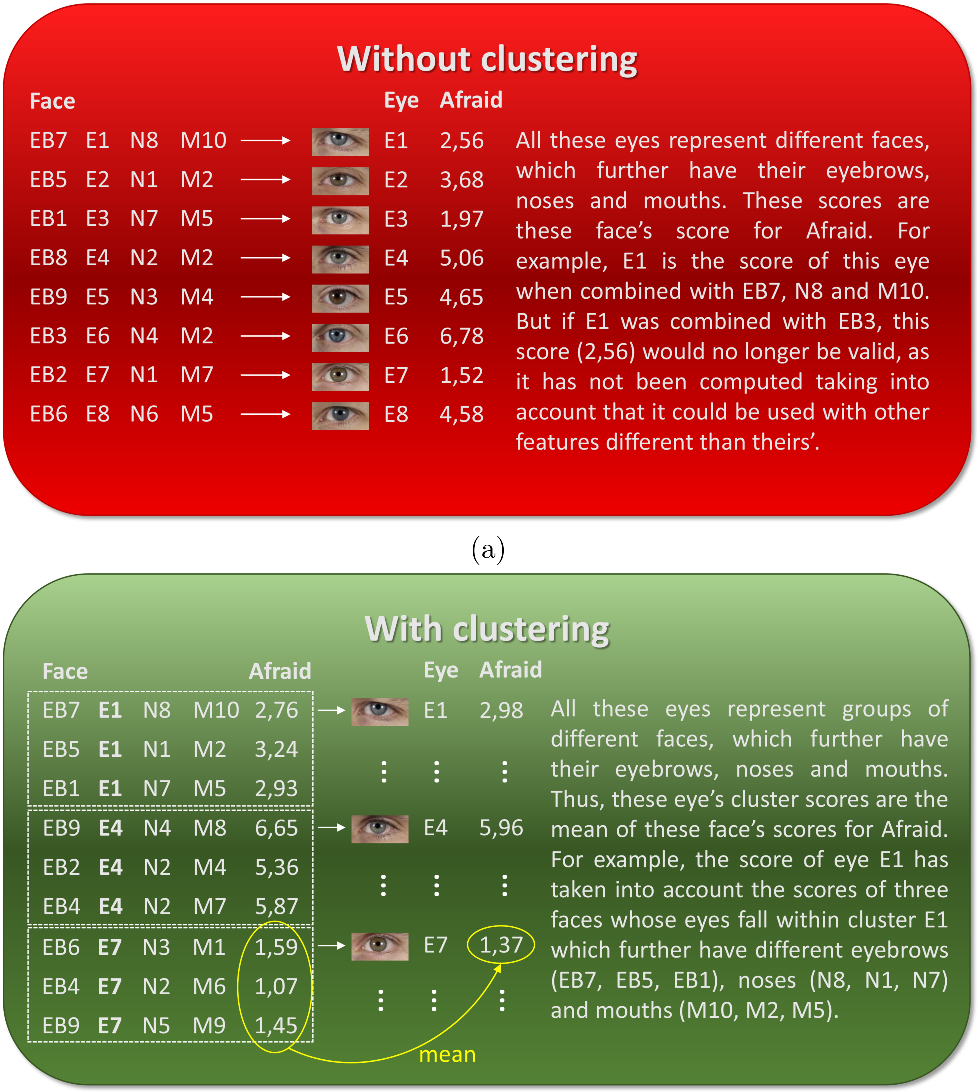
At this stage, sets of 580 eyebrows, 580 eyes, 290 noses, 290 mouths, 290 jawlines and 290 sets of distances are available[2]. The clustering or classification by appearance of the extracted features is necessary in order to create taxonomies, which are needed to evaluate faces. As the main objective of this thesis is the creation of new realistic faces by means of other face features, it is of sum importance to be able to combine features extracted from different faces and still have an appropriate score for these features. This way, it is possible to compute a score for a new face created with these features.

It might seem difficult to understand why, so in the following lines a clarifying explanation is provided. To make it easier to understand, only internal features are considered in the explanation (eyebrows, eyes, noses and mouths).

[2] Some samples of these images are available in Appendix B. To consult all the images employed and see the clustering obtained, please go to `https://github.com/flifuehu/facial-feature-clustering`.

Let's consider a list of faces with their features: eyebrows (EB), eyes (E), noses (N) and mouths (M). It may seem enough to find a function of similarity able to tell how similar are two features, and then take the score of the face where the most similar feature is found. However, by doing this, the assumption that features are independent among them is made, which is incorrect. When looking a face, people see a gestalt of features, meaning that features interfere one with another in order to form the impressions people experience when seeing faces. One way to gather this interferences or relationships among features is to group features by similarity. In this way, when a new face is processed, the eye is compared with all the groups available, and one is chosen. Then, inside this group, several eyes extracted from several different faces are present. Moreover, this group has a mean score computed having into account every present face. The key point is that this score has been computed by using the score of the faces to which the eyes belong instead of some sort of eye score (there is no available score for just the eyes). This links the score obtained by the cluster to the aforementioned type of eye, but also allows to consider that this score will be the same or very similar for any possible combination of the rest of facial features, because the eye cluster itself has been created accounting for some variance of these other features, as shows Figure 2.36 (b). However, if all the eyes were considered without any clustering, a similarity function would tell which is the most similar eye, but this eye would be inside a face with certain eyebrows, nose and mouth. Therefore, this score would be valid only in the case of a face with these features (Figure 2.36 (a)). Although the ideal situation would be to account for 2^{nd}, 3^{rd}, 4^{th} order (and so on) relationships between features, thus modeling relationships of every feature with each other, this requires a huge amount of rated faces unavailable at the moment, which makes it practically impossible to be implemented in this thesis due to time and resources limitations. Then, the implemented solution tries to overcome this problem to the extent possible, accounting for the feature variability present on the faces within a certain cluster. This explanation is valid for the rest of features employed in this thesis.

The procedure followed to cluster the distinct extracted features differs for some of them, according to the extraction and characterization method followed. On the one hand, internal features are extracted as images, so it is an image what characterizes them. On the other hand, jawlines and distances follow a different extraction and characterization method. Jawlines are characterized as a set of (x, y)-coordinates, and distances as a set of values representing distances.



Without clustering

Face				Eye	Afraid
EB7	E1	N8	M10	E1	2,56
EB5	E2	N1	M2	E2	3,68
EB1	E3	N7	M5	E3	1,97
EB8	E4	N2	M2	E4	5,06
EB9	E5	N3	M4	E5	4,65
EB3	E6	N4	M2	E6	6,78
EB2	E7	N1	M7	E7	1,52
EB6	E8	N6	M5	E8	4,58

All these eyes represent different faces, which further have their eyebrows, noses and mouths. These scores are these face's score for Afraid. For example, E1 is the score of this eye when combined with EB7, N8 and M10. But if E1 was combined with EB3, this score (2,56) would no longer be valid, as it has not been computed taking into account that it could be used with other features different than theirs'.

(a)

With clustering

Face				Afraid	Eye	Afraid
EB7	**E1**	N8	M10	2,76	E1	2,98
EB5	**E1**	N1	M2	3,24	⋮	⋮
EB1	**E1**	N7	M5	2,93		
EB9	**E4**	N4	M8	6,65	E4	5,96
EB2	**E4**	N2	M4	5,36	⋮	⋮
EB4	**E4**	N2	M7	5,87		
EB6	**E7**	N3	M1	1,59	E7	1,37
EB4	**E7**	N2	M6	1,07	⋮	⋮
EB9	**E7**	N5	M9	1,45		

All these eyes represent groups of different faces, which further have their eyebrows, noses and mouths. Thus, these eye's cluster scores are the mean of these face's scores for Afraid. For example, the score of eye E1 has taken into account the scores of three faces whose eyes fall within cluster E1 which further have different eyebrows (EB7, EB5, EB1), noses (N8, N1, N7) and mouths (M10, M2, M5).

(b)

Figure 2.36: Why clustering is necessary. (a) Shows why an approach without clustering is not appropriate. (b) Shows the process followed using clustering. EB, E, N, and M stand for eyebrow, eye, nose and mouth, respecitvely.

In order to cluster internal features, the pixels of the images themselves could be used. In this manner, eyebrows, eyes, noses and mouths would be clustered in spaces of 169 × 96, 236 × 116, 187 × 118 and 191 × 104 dimensions respectively. As can be noted, this approach yields very high dimensional spaces for clustering, which makes the method slow, very sensitive to noise and weak dealing with slight variations across images. Then, three possible approaches were tested in order to characterize these features: geometric-based characterization, appearance-based characterization and mixed characterization. The appearance-based approach was selected to characterize the features because the objective was to classify them based on their global appearance more than on their geometrical characteristics (structural approach). This method uses the *eigenfaces* approach to obtain a relatively low-dimensional vector of characteristics which characterizes the features (the term eigenfaces is maintained although the method is now used over facial features). As explained in subsubsection 2.2.2, the eigenfaces approach is a method to efficiently represent pictures of faces by a relatively low-dimensional vector. A principal component analysis can be used on an ensemble of face images to form a set of basis features (Sirovich and Kirby, 1987). These basis images, known as eigenpictures, can be linearly combined to reconstruct images in the original set.

Using this procedure over each set of features it was possible to characterize each feature by a set of M eigenvalues, thus reducing the quantity of information needed to describe the features while increasing speed, robustness and accuracy. This procedure allowed for automatic, robust, fast and objective classification of internal features from varying ethnic groups (Asian, Black, Latino and White) considering the global appearance of features while summarizing the central information to characterize them. It is important to note that this characterization was performed on a by-feature and by-ethnicity basis, that is, features were separated by type and by ethnicity. This was necessary due to the holistic character of the implemented method, completely based on appearance. For example, Asian and Black features are very different, so the basis computed by the eigenfaces approach would be very different for both of them. Thus, eigenfaces method was applied over each subset of facial features for each ethnicity independently.

On the other hand, jawlines were clustered using the coordinates extracted in the previous section. In the case of the distances, the clustering was performed making use of intervals.

In the following subsections the clustering process is explained in-detail for every feature.

2.5.1 Automatic internal features clustering

Internal features clustering includes the clustering of the eyebrows, the eyes, the noses and the mouths. All of them were characterized with only 45 eigenvalues. The same value was chosen for all of them in order to facilitate the subsequent clustering process, bearing in mind that the explained variances were about 85% or higher in all cases (Table 2.2).

Table 2.2: Percentages of variance explained by 45 eigenfaces for each dataset (feature and ethnicity).

	Ethnic group			
Feature	Asian	Black	Latino	White
Eyebrows	95.12	91.99	94.45	93.69
Eyes	91.15	84.98	88.61	86.22
Nose	98.55	93.49	98.19	91.36
Mouth	98.88	93.69	99.14	95.26

At this stage, the appearance of each feature can be characterized using 45 real values (eigenvalues). As an example of the information of the features that is captured using eigenfaces, Figure 2.37 shows a reduced set of original mouths (a), and the same set of mouths reconstructed using 45 eigenvalues before de-normalization (b).

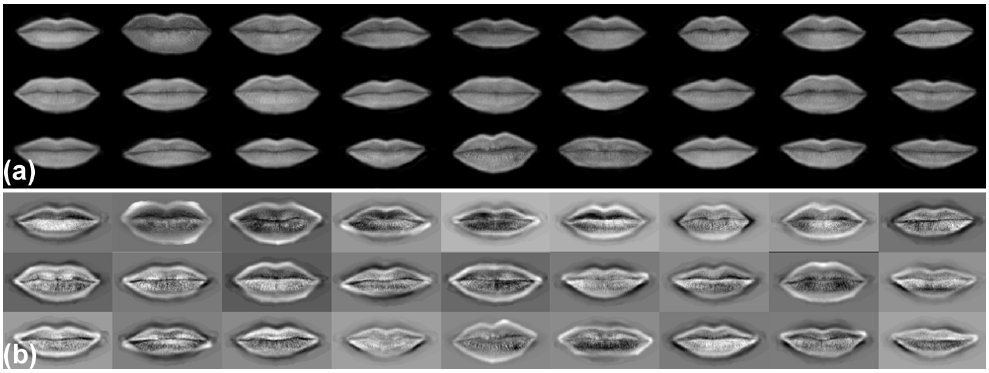


Figure 2.37: Original and reconstructed mouths before de-normalization using 45 eigenfaces. (a) Original mouths. (b) Reconstructed mouths before de-normalization.

Finally, and after reducing the dimensionality of the internal features data, the k-means algorithm (MacKay, 2003) is employed to cluster the features using their eigenvalues as characteristics. For each internal feature and ethnicity, the total number of instances is clustered in a 45 dimensional subspace conformed by the corresponding eigenvalues. A drawback of using this method is that the number of clusters k must be predefined, and this is unknown *a priori*. The approach followed to face this problem was to perform several k-means executions varying k, and to calculate the Dunn's Index (Dunn, 1974) for each set of clusters while monitoring the number of existent *mono-clusters*, which are defined as clusters with only one (for mouths and noses) or two instances (for eyebrows and eyes). The Dunn's Index measures the compactness and separation of the clusters obtained for each k. A higher Dunn's Index points to a small intra-cluster variance and a high inter-cluster distance, that is, the features included in each cluster are more similar among them, and more different from the features belonging to other clusters. Therefore, the number of clusters for each feature was selected as the k that maximized the Dunn's Index while keeping the number of mono-clusters equal or below 2.

2.5.2 Automatic jawlines clustering

The clustering of the jawlines is performed in the same manner as the internal features clustering, but choosing $M = 42$ eigenvalues instead. The election of this number of eigenvalues responds to the number of jawline points (17) and centroids (4) used to characterize each jawline. These points give a total of 21 (x, y)-coordinates, which unrolled to perform the clustering result in 42 parameters. In this way, all coordinates characterizing the jawline are used in their clustering, which results in an explained variance of 100%.

Similarly, the election of the k is still a problem, so the same procedure described for the internal features is followed for jawlines clustering.

2.5.3 Distances clustering

The case of the distances is completely different to the previous explained features. In this case, distances are clustered into 11 regular intervals created individually for each distance taking into account its minimum and maximum. By having 11 intervals it is possible to move the facial features to the most extreme top position in the CFD faces in 5 steps, leave them at their mean position, or move them 5 steps to the bottom. Taking as example the clustering for the distance from the eyebrows centroid to the lowermost point of the jawline (d_{eb}), intervals are created taking steps of

$$s = \frac{\max(d_{eb}) - min(d_{eb})}{12} \, . \tag{2.38}$$

Then, distances are classified inside the corresponding interval out of the 11 available.

2.6 Results

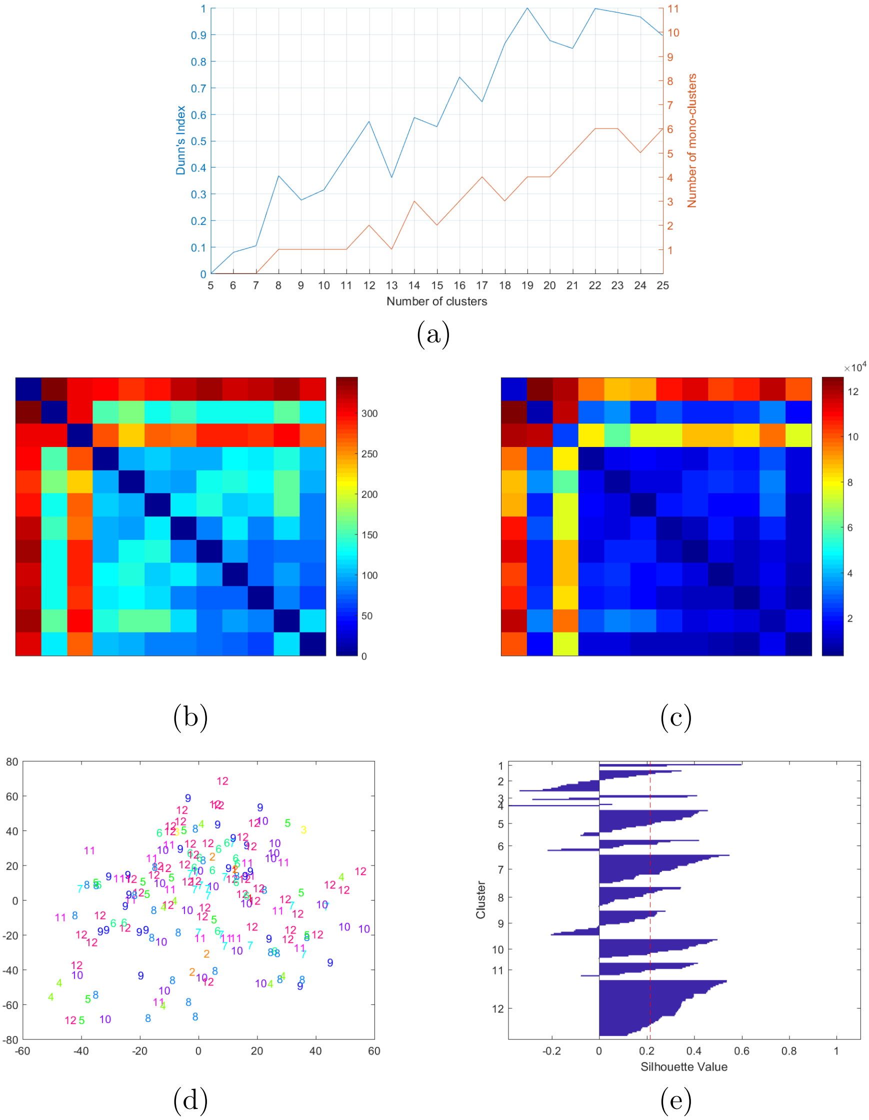
In this section the clusterings obtained are assessed and the obtained clusters are shown. For clarification purposes, metrics employed and computed in the clustering process are explained with an example in the following lines, and results are then presented for each ethnicity and feature in the subsequent subsections. Therefore, for each clustering, a figure like Figure 2.38 is presented. In this figure all the information relative to the presented clustering is shown.

Thus, Figure 2.38 (a) shows the Dunn's Index of each cluster compared to the number of existent mono-clusters. This information was used to find the best possible k. As was explained in subsection 2.5.1, the number of clusters was chosen as that of the clustering with the highest Dunn's Index with two or less mono-clusters. As can be drawn from this example, in this case $k = 12$.

In addition, several metrics were calculated in order to assess how good the clustering was. On the one hand, a distance matrix was computed to show how far one cluster is from another. Thus, the diagonal of this matrix will always be 0 (distance of cluster n to itself), while the element (i, j) shows the distance between clusters i and j (see Figure 2.38 (b)). Here, very low values indicate two nearby clusters and therefore it is not desired, since it would mean that those two clusters could possibly be taken as just one. In this example, the diagonal is 0 as advised, and the rest of the values are all above 50, which means that no cluster is very close to another.

On the other hand, a distance matrix was computed to check how distant are images of a given cluster from the rest of clusters. To do so, the mean of the distances of all instances of a determined cluster to each centroid of the other clusters are computed. This produces a symmetric matrix of distances, where position $(1, 1)$ represents the mean distance of instances belonging to cluster 1 to its centroid. Position $(1, 2)$ refers to the mean distance of instances belonging to cluster 1 to cluster 2 centroid, and so on. Then, this matrix should have a low valued diagonal, while the rest of values should be as high as possible. This would mean that the instances of a cluster are very close to their centroid and far from the rest of clusters (see Figure 2.38 (c)).

Figure 2.38 (d) represents the clustering representation in two dimensions. To be able to plot it in just two dimensions, the t-SNE algorithm was used.



(a)

(b)

(c)

(d)

(e)

Figure 2.38: Clustering metrics example. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids respectively. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

Figure 2.38 (e) shows the silhouette of the clustering, which allows to check graphically how well each instance lies within its cluster. The vertical red dashed line indicates the mean value of every cluster. As was explained in subsubsection 2.2.3, silhouette values range from -1 to +1, where values close to 1 indicate that the instance is far from the decision boundary of the cluster, values close to 0 show that the instance is on or close to the boundary, and values lower than 0 indicate that the instance might be incorrectly clustered.

As clusterings are performed for every ethnicity and feature independently, results are also presented by ethnicity and feature. In this manner, each subsection presents the results obtained for all the features of a given ethnicity, that is, the chosen k for each clustering, the Dunn's index, the Silhouette index and the resulting number of clusters after mono-cluster removal. Figures showing each of the aforementioned computed metrics are available. However, due to space limitations and for visualization purposes, these figures are given altogether in Appendix A. Furthermore, the obtained taxonomy for each ethnicity and feature is available in Appendix B. Validation of the clustering method here explained is provided in next section.

2.6.1 Asian

Automatic clustered features

The results of the clustering for the Asian ethnicity are available in Table 2.3.

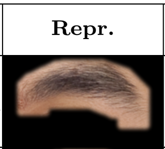
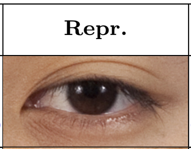
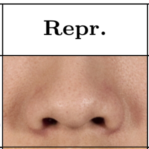
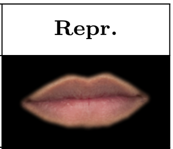
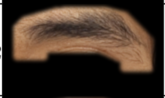
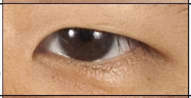
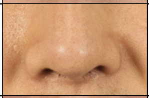
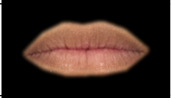
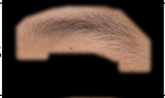
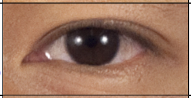
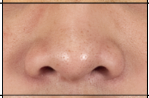
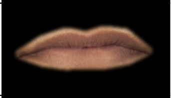
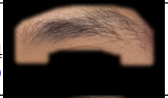
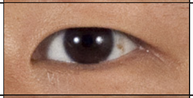
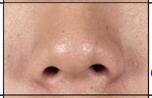
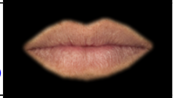
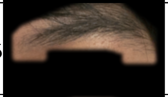
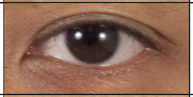
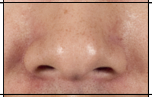
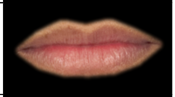
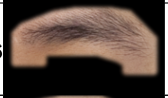
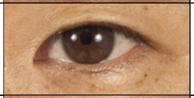
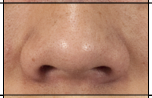
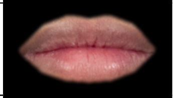
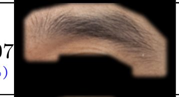
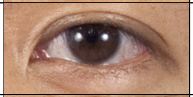
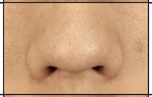
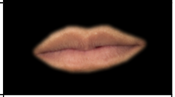
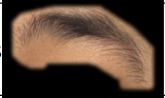
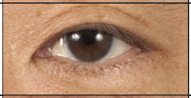
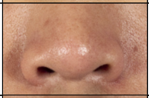
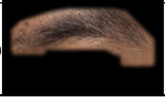
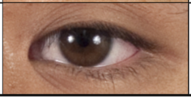
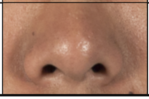
Table 2.3: Asian clustering metrics.

	k	k_{final}	**# of mc**	**DI**	**SI**
eyebrows	15	14	1	0.47	0.23
eyes	15	13	2	0.78	0.15
noses	16	14	2	0.30	0.20
mouths	9	7	2	0.19	0.20
jawlines	15	13	2	1.00	0.20

As can be observed, eyebrow's final number of clusters after monoclusters (mc) removal is $k = 14$, with a Dunn's Index of $DI = 0.47$ and a Silhouette Index of $SI = 0.23$ (Figure A.1). For eyes, the final number of clusters is $k = 13$, with $DI = 0.78$ and $SI = 0.15$ (Figure A.2). Regarding the noses, the final number of clusters is $k = 14$, $DI = 0.30$ and $SI = 0.20$ (Figure A.3). As for mouths, the final number of clusters is $k = 7$, with $DI = 0.19$ and $SI = 0.20$ (Figure A.4). Finally, jawline's final number of clusters is $k = 13$, with $DI = 1$

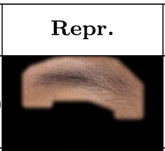
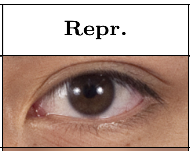
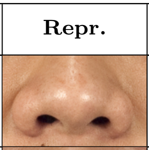
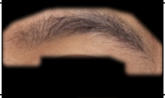
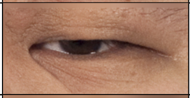
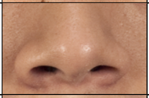
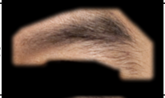
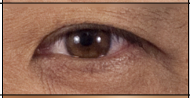
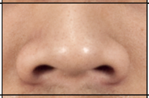
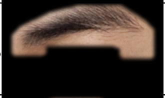
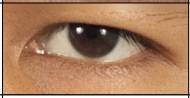
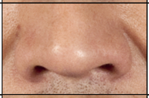
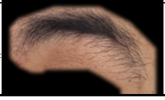
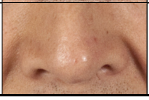
and $SI = 0.20$ (Figure A.5). Table 2.4 shows the cluster representatives of the internal features clusterings for Asian ethnicity. Along with each cluster representative's image is available the percentage of elements in that cluster over the total number of elements in the corresponding clustering. Representatives are sorted in descendent order according to this percentage. On the other hand, Table 2.5 does the same for the jawlines. Further, the obtained Asian taxonomies are completely available in Appendix B.1.

Table 2.4: Representatives of the Asian clustering for every automatically clustered internal feature. AEB identifies Asian eyebrow clusters, AE Asian eye clusters, AN Asian nose clusters and AM Asian mouth clusters.

Name (%)	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**
AEB01 (18.27%)		AE01 (11.54%)		AN01 (19.23%)		AM01 (28.85%)	
AEB02 (12.50%)		AE02 (11.54%)		AN02 (9.62%)		AM02 (23.08%)	
AEB03 (10.58%)		AE03 (10.58%)		AN03 (7.69%)		AM03 (13.46%)	
AEB04 (10.58%)		AE04 (8.65%)		AN04 (7.69%)		AM04 (11.54%)	
AEB05 (7.69%)		AE05 (8.65%)		AN05 (7.69%)		AM05 (9.62%)	
AEB06 (6.73%)		AE06 (7.69%)		AN06 (7.69%)		AM06 (7.69%)	
AEB07 (6.73%)		AE07 (7.69%)		AN07 (7.69%)		AM07 (5.77%)	
AEB08 (4.81%)		AE08 (6.73%)		AN08 (5.77%)			
AEB09 (4.81%)		AE09 (6.73%)		AN09 (5.77%)			

Continued on next page

Table 2.4 – *Continued from previous page*

Name (%)	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**
AEB10 (3.85%)		AE10 (6.73%)		AN10 (5.77%)			
AEB11 (3.85%)		AE11 (4.81%)		AN11 (3.85%)			
AEB12 (3.85%)		AE12 (4.81%)		AN12 (3.85%)			
AEB13 (2.88%)		AE13 (3.85%)		AN13 (3.85%)			
AEB14 (2.88%)				AN14 (3.85%)			

Distances

Distances were grouped in 11 regular intervals. Table 2.6 shows the minimum and maximum, the range and the interval size for each distance.

Table 2.5: Representatives of the Asian clustering for automatically clustered jawlines.

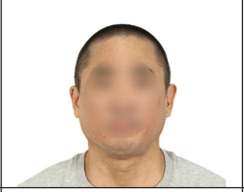
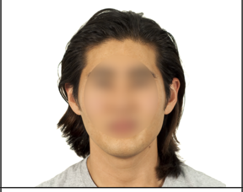
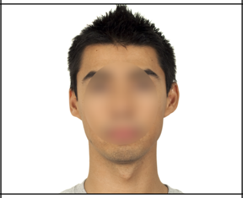
 AJ01 (15.38%)	 AJ02 (11.54%)	 AJ03 (11.54%)	 AJ04 (9.62%)
 AJ05 (7.69%)	 AJ06 (7.69%)	 AJ07 (7.69%)	 AJ08 (7.69%)
 AJ09 (5.77%)	 AJ10 (3.85%)	 AJ11 (3.85%)	 AJ12 (3.85%)
 AJ13 (3.85%)			

Table 2.6: Distances clustering for Asian ethnicity.

Distance	Min	Max	Range	Interval size
d_{eb}	671.36	843.60	172.25	15.66
d_e	587.67	716.97	129.30	11.76
d_n	428.53	550.46	121.93	11.09
d_m	213.01	319.40	106.39	9.67
d_{ee}	314.33	388.42	74.08	6.74

2.6.2 Black

Automatic clustered features

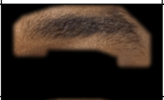
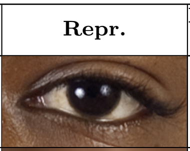
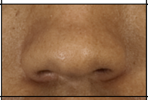
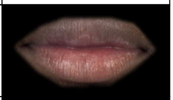
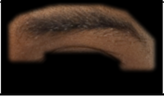
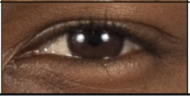
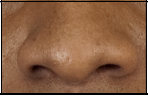
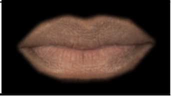
The results of the clustering for the Black ethnicity are available in Table 2.7.

Table 2.7: Black clustering metrics.

	k	k_{final}	**# of mc**	**DI**	**SI**
eyebrows	12	10	2	0.60	0.16
eyes	23	21	2	1.00	0.11
noses	21	20	1	0.46	0.17
mouths	9	8	1	0.53	0.18
jawlines	13	13	0	1.00	0.14

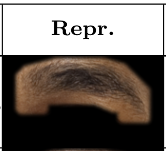
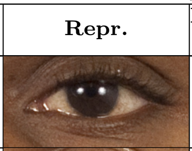
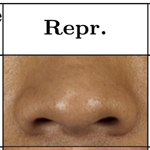
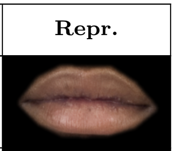
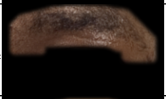
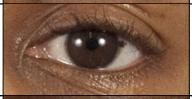
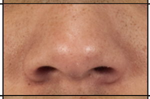
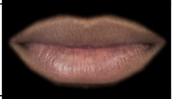
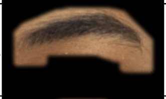
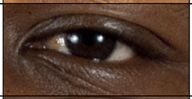
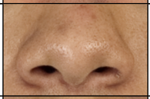
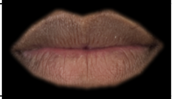
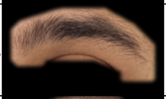
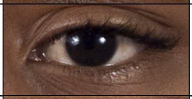
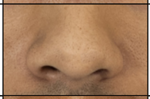
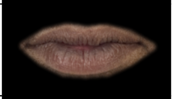
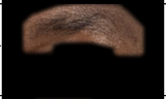
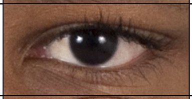
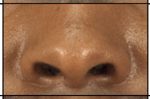
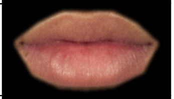
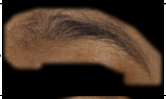
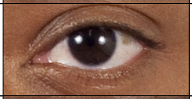
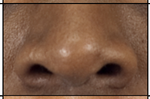
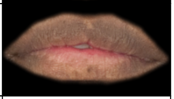
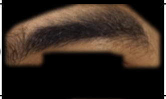
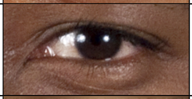
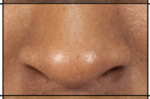
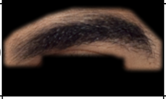
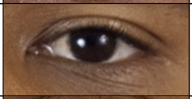
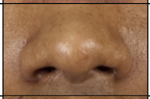
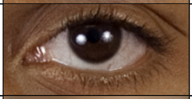
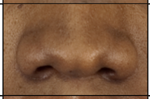
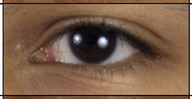
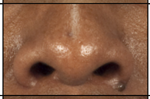
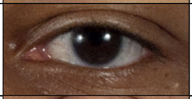
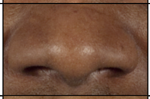
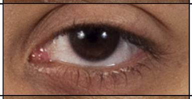
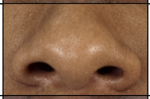
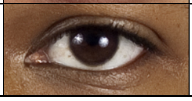
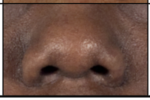
As can be observed, eyebrow's final number of clusters after monoclusters removal is $k = 10$, with a Dunn's Index of $DI = 0.60$ and a Silhouette Index of $SI = 0.16$ (Figure A.6). For eyes, the final number of clusters is $k = 21$, with $DI = 1$ and $SI = 0.11$ (Figure A.7). Regarding the noses, the final number of clusters is $k = 20$, $DI = 0.46$ and $SI = 0.17$ (Figure A.8). As for mouths, the final number of clusters is $k = 8$, with $DI = 0.53$ and $SI = 0.18$ (Figure A.9). Finally, jawline's final number of clusters is $k = 13$, with $DI = 1$ and $SI = 0.14$ (Figure A.10). Table 2.8 shows the cluster representatives of the internal features clusterings for Black ethnicity. Along with each cluster representative's image is available the percentage of elements in that cluster over the total number of elements in the corresponding clustering. Representatives are sorted in descendent order according to this percentage. On the other hand, Table 2.9 does the same for the jawlines. Further, the obtained Black taxonomies are completely available in Appendix B.2.

Table 2.8: Representatives of the Black clustering for every automatically clustered internal feature. BEB identifies Black eyebrow clusters, BE Black eye clusters, BN Black nose clusters and BM Black mouth clusters.

Name (%)	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**
BEB01 (19.35%)		BE01 (11.83%)		BN01 (9.68%)		BM01 (32.26%)	
BEB02 (18.28%)		BE02 (8.06%)		BN02 (9.68%)		BM02 (16.13%)	

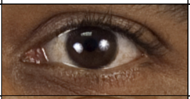
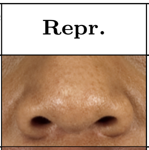
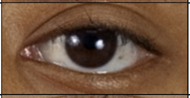
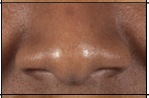
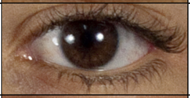
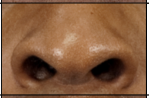
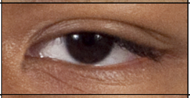
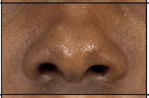
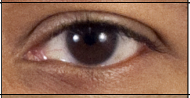
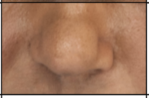
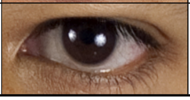
Continued on next page

Table 2.8 – *Continued from previous page*

Name (%)	Repr.	Name (%)	Repr.	Name (%)	Repr.	Name (%)	Repr.
BEB03 (15.05%)		BE03 (8.06%)		BN03 (8.60%)		BM03 (15.05%)	
BEB04 (14.52%)		BE04 (7.53%)		BN04 (7.53%)		BM04 (12.90%)	
BEB05 (13.98%)		BE05 (6.45%)		BN05 (7.53%)		BM05 (10.75%)	
BEB06 (6.45%)		BE06 (5.38%)		BN06 (6.45%)		BM06 (7.53%)	
BEB07 (4.84%)		BE07 (5.38%)		BN07 (5.38%)		BM07 (3.23%)	
BEB08 (3.23%)		BE08 (5.38%)		BN08 (5.38%)		BM08 (2.15%)	
BEB09 (2.15%)		BE09 (5.38%)		BN09 (5.38%)			
BEB10 (2.15%)		BE10 (4.84%)		BN10 (5.38%)			
		BE11 (4.30%)		BN11 (4.30%)			
		BE12 (3.76%)		BN12 (3.23%)			
		BE13 (3.23%)		BN13 (3.23%)			
		BE14 (3.23%)		BN14 (3.23%)			
		BE15 (3.23%)		BN15 (3.23%)			

Continued on next page

Table 2.8 – *Continued from previous page*

Name (%)	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**
		BE16 (3.23%)		BN16 (3.23%)			
		BE17 (2.69%)		BN17 (2.15%)			
		BE18 (2.69%)		BN18 (2.15%)			
		BE19 (2.15%)		BN19 (2.15%)			
		BE20 (1.61%)		BN20 (2.15%)			
		BE21 (1.61%)					

Distances

Distances were grouped in 11 regular intervals. Table 2.10 shows the minimum and maximum, the range and the interval size for each distance.

Table 2.9: Representatives of the Black clustering for automatically clustered jawlines.

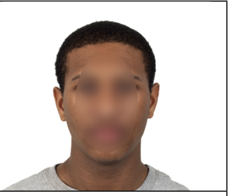
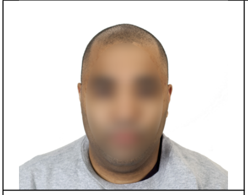
 BJ01 (15.05%)	 BJ02 (12.90%)	 BJ03 (11.83%)	 BJ04 (10.75%)
 BJ05 (10.75%)	 BJ06 (8.60%)	 BJ07 (6.45%)	 BJ08 (5.38%)
 BJ09 (5.38%)	 BJ10 (4.30%)	 BJ11 (4.30%)	 BJ12 (2.15%)
 BJ13 (2.15%)			

Table 2.10: Distances clustering for Black ethnicity.

Distance	Min	Max	Range	Interval size
d_{eb}	684.62	900.30	215.69	19.61
d_e	588.24	780.25	192.01	17.45
d_n	433.06	609.91	176.86	16.08
d_m	213.36	343.74	130.38	11.85
d_{ee}	323.48	415.51	92.03	8.37

2.6.3 Latino

Automatic clustered features

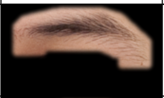
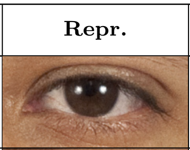
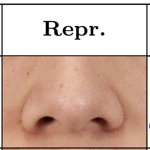
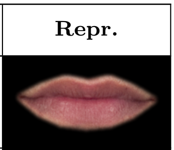
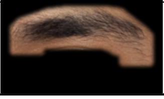
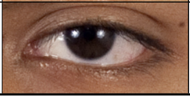
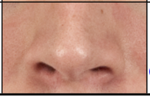
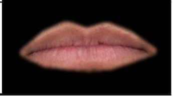
The results of the clustering for the Latino ethnicity are available in Table 2.11.

Table 2.11: Latino clustering metrics.

	k	k_{final}	**# of mc**	**DI**	**SI**
eyebrows	10	9	1	0.56	0.16
eyes	17	16	1	0.84	0.10
noses	16	14	2	0.36	0.18
mouths	5	5	0	1.00	0.22
jawlines	15	14	1	1.00	0.18

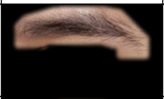
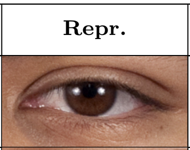
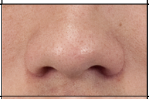
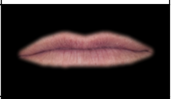
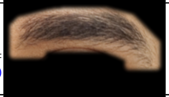
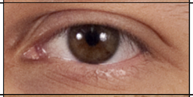
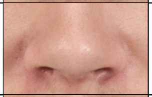
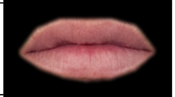
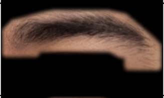
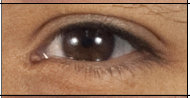
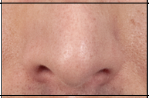
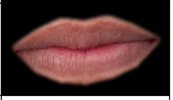
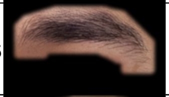
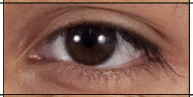
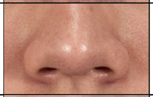
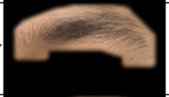
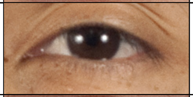
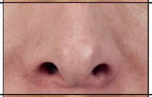
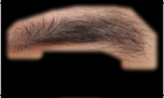
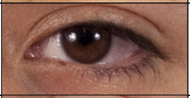
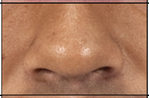
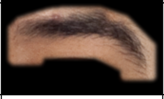
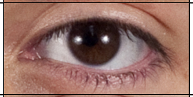
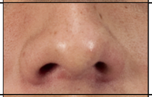
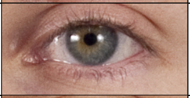
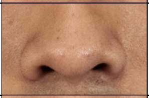
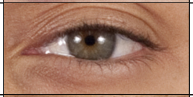
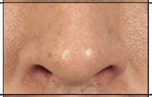
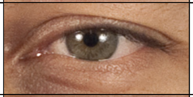
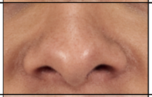
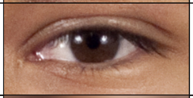
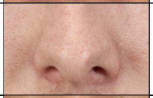
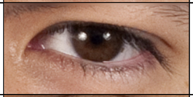
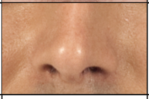
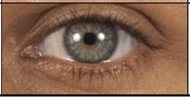
As can be observed, eyebrow's final number of clusters after monoclusters removal is $k = 9$, with a Dunn's Index of $DI = 0.56$ and a Silhouette Index of $SI = 0.16$ (Figure A.11). For eyes, the final number of clusters is $k = 16$, with $DI = 0.84$ and $SI = 0.10$ (Figure A.12). Regarding the noses, the final number of clusters is $k = 14$, $DI = 0.36$ and $SI = 0.18$ (Figure A.13). As for mouths, the final number of clusters is $k = 5$, with $DI = 1$ and $SI = 0.22$)Figure A.14). Finally, jawline's final number of clusters is $k = 14$, with $DI = 1$ and $SI = 0.18$ (Figure A.15). Table 2.12 shows the cluster representatives of the internal features clusterings for Latino ethnicity. Along with each cluster representative's image is available the percentage of elements in that cluster over the total number of elements in the corresponding clustering. Representatives are sorted in descendent order according to this percentage. On the other hand, Table 2.13 does the same for the jawlines. Further, the obtained Latino taxonomies are completely available in Appendix B.3.

Table 2.12: Representatives of the Latino clustering for every automatically clustered internal feature. LEB identifies Latino eyebrow clusters, LE Latino eye clusters, LN Latino nose clusters and LM Latino mouth clusters.

Name (%)	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**
LEB01 (24.04%)		LE01 (13.46%)		LN01 (13.46%)		LM01 (46.15%)	
LEB02 (15.38%)		LE02 (9.62%)		LN02 (11.54%)		LM02 (32.69%)	

Continued on next page

Table 2.12 – *Continued from previous page*

Name (%)	Repr.	Name (%)	Repr.	Name (%)	Repr.	Name (%)	Repr.
LEB03 (12.50%)		LE03 (7.69%)		LN03 (9.62%)		LM03 (9.62%)	
LEB04 (11.54%)		LE04 (7.69%)		LN04 (7.69%)		LM04 (7.69%)	
LEB05 (11.54%)		LE05 (7.69%)		LN05 (7.69%)		LM05 (3.85%)	
LEB06 (9.62%)		LE06 (6.73%)		LN06 (7.69%)			
LEB07 (7.69%)		LE07 (6.73%)		LN07 (7.69%)			
LEB08 (3.85%)		LE08 (6.73%)		LN08 (5.77%)			
LEB09 (3.85%)		LE09 (5.77%)		LN09 (5.77%)			
		LE10 (4.81%)		LN10 (5.77%)			
		LE11 (4.81%)		LN11 (5.77%)			
		LE12 (3.85%)		LN12 (3.85%)			
		LE13 (3.85%)		LN13 (3.85%)			
		LE14 (3.85%)		LN14 (3.85%)			
		LE15 (3.85%)					

Continued on next page

Table 2.12 – *Continued from previous page*

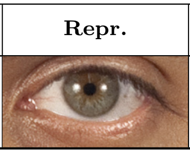
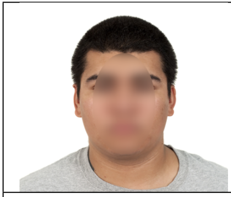
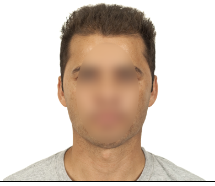
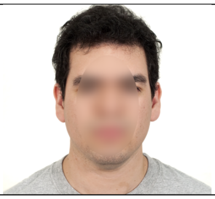
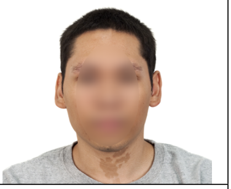
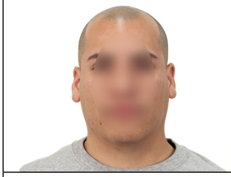
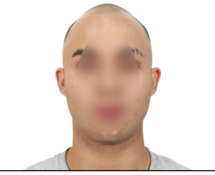
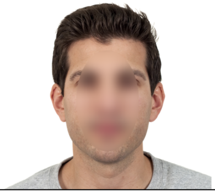
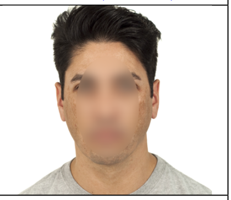
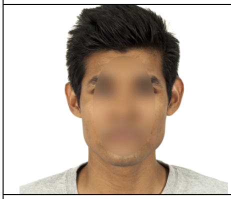
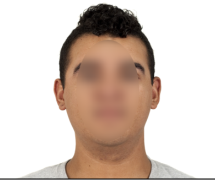
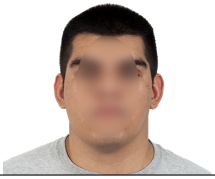
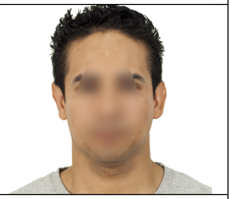
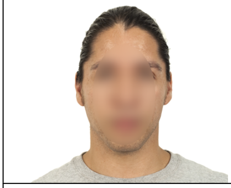
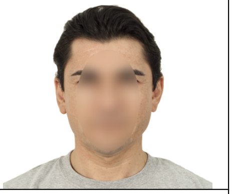
Name (%)	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**	**Name (%)**	**Repr.**
		LE16 (2.88%)					

Table 2.13: Representatives of the Latino clustering for automatically clustered jawlines.

 LJ01 (13.46%)	 LJ02 (11.54%)	 LJ03 (9.62%)	 LJ04 (9.62%)
 LJ05 (9.62%)	 LJ06 (7.69%)	 LJ07 (7.69%)	 LJ08 (7.69%)
 LJ09 (3.85%)	 LJ10 (3.85%)	 LJ11 (3.85%)	 LJ12 (3.85%)
 LJ13 (3.85%)	 LJ14 (3.85%)		

Distances

Distances were grouped in 11 regular intervals. Table 2.14 shows the minimum and maximum, the range and the interval size for each distance.

Table 2.14: Distances clustering for Latino ethnicity.

Distance	**Min**	**Max**	**Range**	**Interval size**
d_{eb}	698.87	845.41	146.54	13.32
d_e	600.55	738.47	137.92	12.54
d_n	437.19	581.85	144.65	13.15
d_m	213.70	339.41	125.71	11.43
d_{ee}	330.53	388.49	57.96	5.27

2.6.4 White

Automatic clustered features

The results of the clustering for the White ethnicity are available in Table 2.15.

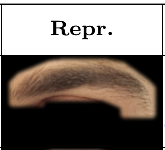
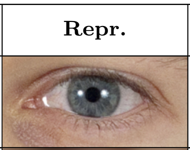
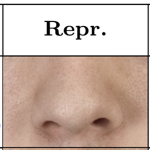
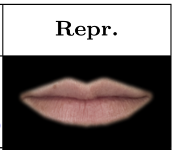
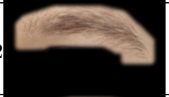
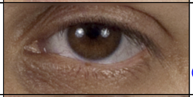
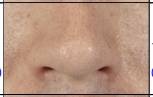
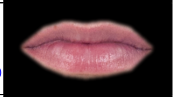
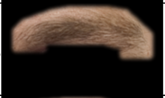
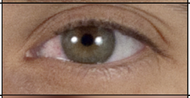
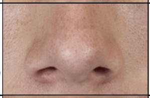
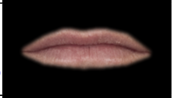
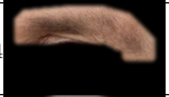
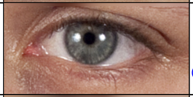
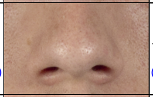
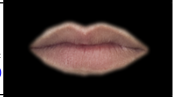
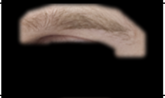
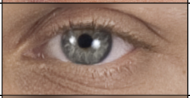
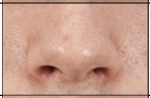
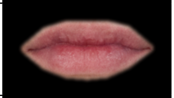
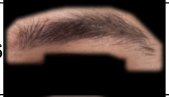
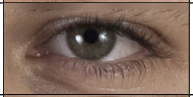
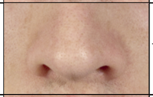
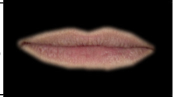
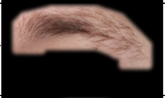
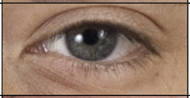
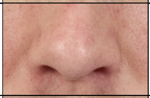
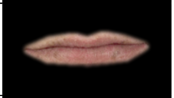
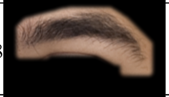
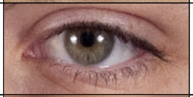
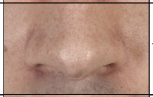
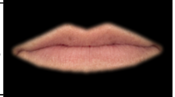
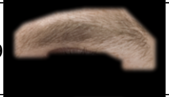
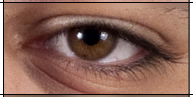
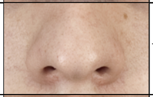
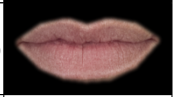
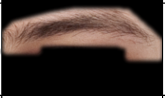
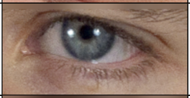
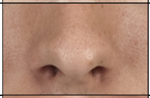
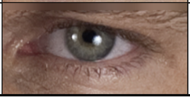
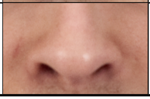
Table 2.15: White clustering metrics.

	k	k_{final}	**# of mc**	**DI**	**SI**
eyebrows	12	10	2	0.57	0.21
eyes	19	19	0	0.86	0.12
noses	12	12	0	0.62	0.17
mouths	11	9	2	0.55	0.21
jawlines	12	11	1	1.00	0.22

As can be observed, eyebrow's final number of clusters after monoclusters removal is $k = 10$, with a Dunn's Index of $DI = 0.57$ and a Silhouette Index of $SI = 0.21$ (Figure A.16). For eyes, the final number of clusters is $k = 19$, with $DI = 0.86$ and $SI = 0.12$ (Figure A.17). Regarding the noses, the final number of clusters is $k = 12$, $DI = 0.62$ and $SI = 0.17$ (Figure A.18). As for mouths, the final number of clusters is $k = 9$, with $DI = 0.55$ and $SI = 0.21$ (Figure A.19). Finally, jawline's final number of clusters is $k = 11$, with $DI = 1$ and $SI = 0.22$ (Figure A.20). Table 2.16 shows the cluster representatives of the internal features clusterings for White ethnicity. Along with each cluster representative's image is available the percentage of elements in that cluster over the total number of elements in the corresponding clustering. Representatives are sorted in descendent order according to this percentage.

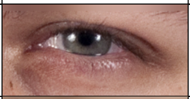
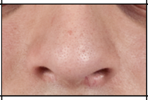
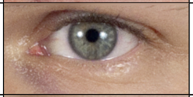
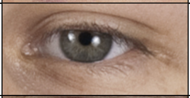
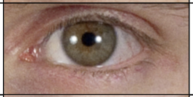
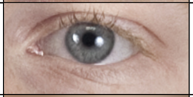
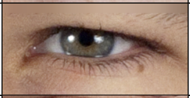
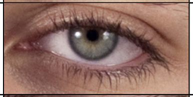
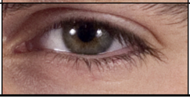
On the other hand, Table 2.17 does the same for the jawlines. Further, the obtained White taxonomies are completely available in Appendix B.4.

Table 2.16: Representatives of the White clustering for every automatically clustered internal feature. WEB identifies White eyebrow clusters, WE White eye clusters, WN White nose clusters and WM White mouth clusters.

Name (%)	Repr.	Name (%)	Repr.	Name (%)	Repr.	Name (%)	Repr.
WEB01 (24.19%)		WE01 (12.90%)		WN01 (12.90%)		WM01 (21.51%)	
WEB02 (12.37%)		WE02 (9.14%)		WN02 (11.83%)		WM02 (17.20%)	
WEB03 (11.29%)		WE03 (8.06%)		WN03 (10.75%)		WM03 (17.20%)	
WEB04 (10.75%)		WE04 (8.06%)		WN04 (10.75%)		WM04 (16.13%)	
WEB05 (9.14%)		WE05 (8.06%)		WN05 (9.68%)		WM05 (8.60%)	
WEB06 (8.60%)		WE06 (7.53%)		WN06 (9.68%)		WM06 (8.60%)	
WEB07 (8.60%)		WE07 (6.99%)		WN07 (9.68%)		WM07 (4.30%)	
WEB08 (8.06%)		WE08 (4.84%)		WN08 (6.45%)		WM08 (4.30%)	
WEB09 (4.84%)		WE09 (4.84%)		WN09 (6.45%)		WM09 (2.15%)	
WEB10 (2.15%)		WE10 (4.30%)		WN10 (4.30%)			
		WE11 (4.30%)		WN11 (4.30%)			

Continued on next page

Table 2.16 – *Continued from previous page*

Name **(%)**	**Repr.**	**Name** **(%)**	**Repr.**	**Name** **(%)**	**Repr.**	**Name** **(%)**	**Repr.**
		WE12 (3.76%)		WN12 (3.23%)			
		WE13 (3.76%)					
		WE14 (2.69%)					
		WE15 (2.69%)					
		WE16 (2.15%)					
		WE17 (2.15%)					
		WE18 (2.15%)					
		WE19 (1.61%)					

Distances

Distances were grouped in 11 regular intervals. Table 2.18 shows the minimum and maximum, the range and the interval size for each distance.

Table 2.17: Representatives of the White clustering for automatically clustered jawlines.

WJ01 (15.05%)	WJ02 (13.98%)	WJ03 (12.90%)	WJ04 (11.83%)
WJ05 (10.75%)	WJ06 (7.53%)	WJ07 (7.53%)	WJ08 (6.45%)
WJ09 (6.45%)	WJ10 (4.30%)	WJ11 (3.23%)	

Table 2.18: Distances clustering for White ethnicity.

Distance	Min	Max	Range	Interval size
d_{eb}	663.87	875.18	211.31	19.21
d_e	570.03	752.68	182.65	16.61
d_n	438.21	588.53	150.32	13.67
d_m	210.13	330.44	120.31	10.94
d_{ee}	321.24	405.13	83.88	7.63

2.7 Validation of the procedure

The described methodology proposes an automatic procedure to classify features based on their appearance. This procedure was used to group features of faces extracted from the Chicago Face Database. The intuitively logical approach to validate the procedure is to compare the obtained taxonomies with those generated by human evaluators. However, as it has been aforementioned, this last approach has important drawbacks. Classifying a big set of features in an undefined number of groups is a hard task considering human capabilities for information processing (Miller, 1956; Scharff et al., 2011). On the other hand, important problems of using this approach are the part-whole effect (Taubert et al., 2011) that decreases human ability for processing individual features, and the influence of face race on the performance in processing facial information (Hayward et al., 2008; Rhodes et al., 2009). Previous works have reported low inter-observer and intra-observer agreement in the evaluation of facial features (Stefanie Ritz-Timme et al., 2011), therefore, a different approach must be used to validate the proposed procedure.

Instead of comparing the obtained taxonomies with those generated by humans, we measured the agreement of human evaluators with the proposed taxonomies. The main objectives were to reduce the number of features presented simultaneously to the human evaluators to make a decision and to simplify the decision that must be made. To do this, a survey composed of several stages was developed (Figure 2.39).

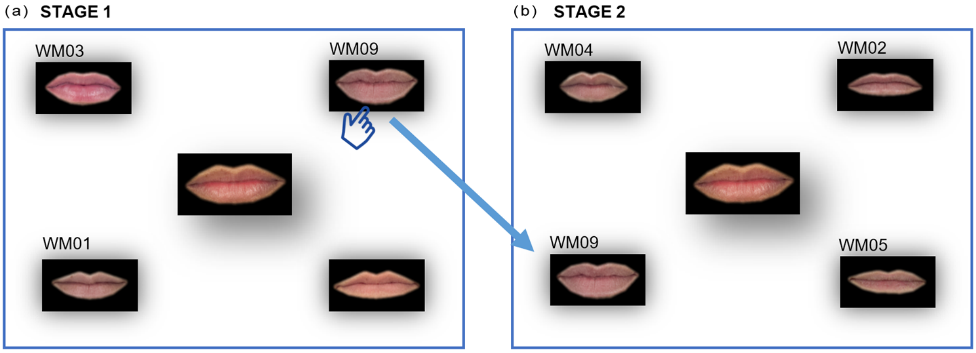


Figure 2.39: Stages 1 and 2 of the survey procedure.

Initially, the image of one feature was selected from the entire dataset in a random way (target feature). Four different representative features were randomly selected (representative features are the features designated as representatives of their groups in the obtained taxonomy). In the first stage of the survey, the five features were presented to the evaluator in a web form (Figure 2.39 (a)). The target feature was in the center of the form, and the four representative features were at the corners. The evaluator was asked to select the representative feature most similar to the target feature by clicking it using the mouse. Then, the selected representative feature passed to the second stage in which a new form was composed like in Figure 2.39 (b). The target feature was in the center again, and the selected representative feature was at a corner of the form. Three new different representative features were randomly selected and situated in the remaining corners. This process was repeated until every representative feature was shown at least once. The cluster of the representative feature selected in the last stage was considered the result of the survey (i.e. the cluster to which the target feature belongs according to the opinion of the respondent). Using this procedure, the decision process was simplified because the number of simultaneous alternatives was reduced to four. As a drawback, the probability of one representative feature to be finally selected depends slightly on the stage in which it is shown.

21 White males and 11 White females aged between 25 and 46 years old participated in three surveys: eyes, noses and mouths. In each survey, 200 target features were selected at random from each White features dataset excluding the representatives. The target features were presented in the survey web form, and the cluster of the representative feature finally selected by the evaluators was registered. This validation procedure was performed only on White eyes, noses and mouths due to time limitations, as it results practically impossible to validate every feature for every ethnicity. Moreover, we intended to validate the method, not the taxonomies themselves, and this approach allowed us to achieve this goal.

Therefore, Tables 2.19, 2.20 and 2.21 show the results of the survey for White eyes, noses and mouths respectively. The first column of the table presents the finally selected cluster. In this column, *Expected* means the cluster in which the target feature was grouped by the automatic procedure. 82 target mouths, 62 target eyes and 93 target noses were classified in the expected cluster. The distance between clusters can be measured through the eigenvalues of their representative features; therefore, it is possible to determine the distance from the expected cluster to each of the other clusters. The closer two clusters are, the more similar are the features they contain. In the aforementioned tables, 1^{st} closest is the cluster closest to the expected cluster, 2^{nd} closest is the second cluster closest to the expected cluster and so on. The number, the percentage and the cumulative percentage of features classified in each cluster are shown.

The percentages of features classified in the expected cluster or in the three clusters closest to it were 73.0% for eyes, 81.0% for noses and 75.5% for mouths.

Table 2.19: Results of the validation survey for White eyes.

Eyes			
Selected cluster	**Nº**	**%**	**Cum %**
Expected	62	31.0%	31.0%
1st closest	38	19.0%	50.0%
2nd closest	27	13.5%	63.5%
3rd closest	19	9.5%	73.0%
4th closest	12	6.0%	79.0%
5th closest	8	4.0%	83.0%
6th closest	5	2.5%	85.5%
7th closest	9	4.5%	90.0%
8th closest	4	2.0%	92.0%
9th closest	1	0.5%	92.5%
10th closest	5	2.5%	95.0%
11th closest	3	1.5%	96.5%
12th closest	1	0.5%	97.0%
13th closest	0	0.0%	97.0%
14th closest	2	1.0%	98.0%
15th closest	0	0.0%	98.0%
16th closest	1	0.5%	98.5%
17th closest	0	0.0%	98.5%
18th closest	1	0.5%	99.0%
19th closest	0	0.0%	99.0%
20th closest	0	0.0%	99.0%
21th closest	1	0.5%	99.5%
22th closest	1	0.5%	100.0%
23th closest	0	0.0%	100.0%
24th closest	0	0.0%	100.0%

Table 2.20: Results of the validation survey for White noses.

Noses			
Selected cluster	**Nº**	**%**	**Cum %**
Expected	93	46.5%	46.5%
1st closest	36	18.0%	64.5%
2nd closest	21	10.5%	75.0%
3rd closest	12	6.0%	81.0%
4th closest	8	4.0%	85.0%
5th closest	14	7.0%	92.0%
6th closest	3	1.5%	93.5%
7th closest	5	2.5%	96.0%
8th closest	6	3.0%	99.0%
9th closest	1	0.5%	99.5%
10th closest	0	0.0%	99.5%
11th closest	1	0.5%	100.0%

Table 2.21: Results of the validation survey for White mouths.

Mouths			
Selected cluster	**Nº**	**%**	**Cum %**
Expected	82	41.0%	41.0%
1st closest	24	12.0%	53.0%
2nd closest	23	11.5%	64.5%
3rd closest	22	11.0%	75.5%
4th closest	17	8.5%	84.0%
5th closest	14	7.0%	91.0%
6th closest	6	3.0%	94.0%
7th closest	6	3.0%	97.0%
8th closest	6	3.0%	100.0%

2.8 Conclusions

Classification systems to categorize human body parts, or taxonomies obtained from them, provide a standardized way to describe or configure the human body, and a lot of work has been done to categorize many different body parts. Describing facial features using a common terminology is essential in disciplines like ergonomics, forensics, surgery or criminology. Moreover, the growth of new technologies that use virtual interlocutors or avatars has led to an increasing interest in synthesizing faces and facial expressions that symbolize the user's presence in new human-machine interaction systems and online activities.

However, there are very few classification systems or taxonomies for features, probably due to the complexity of this task, and to the human limited capacity for processing individual features compared to processing whole faces. Classifying the appearance of facial features requires a holistic approach considering all visible information. Then, encoding the geometry and doing a metric or morphological assessment is not enough to obtain facial features taxonomies based on the appearance. Therefore, we employed appearance-based representations of the features (*eigenfaces*) in order to classify them. The developed procedure groups the features considering all available information and encompassing their global nature.

This procedure was used to classify the features of 290 images of males with neutral expression from the Chicago Face Database, obtaining taxonomies of eyebrows, eyes, noses, mouths and jawlines for several ethnic groups. To validate the procedure, the agreement of human evaluators with the proposed taxonomies was measured. Out of 200 cases for each feature, 41.0% of mouths, 31.0% of eyes and 46.5% of noses, were classified by humans in the same cluster as the automatic procedure. More than 73.0% of the features were classified in the expected cluster or in the three clusters closest to it (75.5% of mouths, 73.0% of eyes and 81.0% of noses).

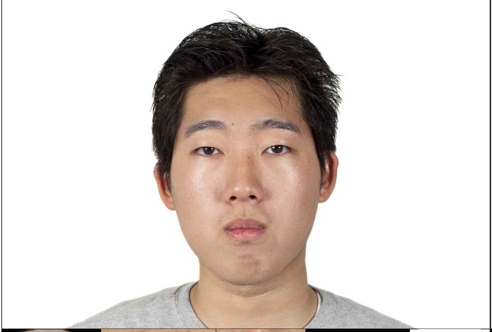
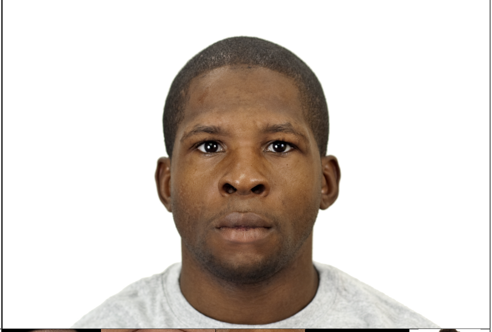
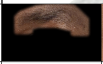
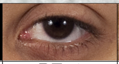
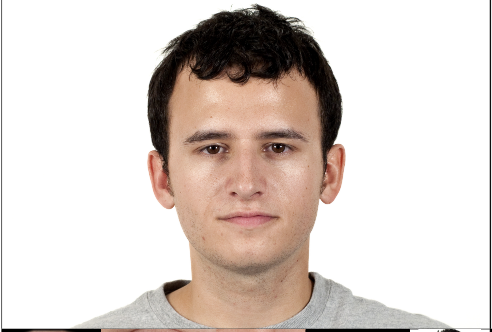
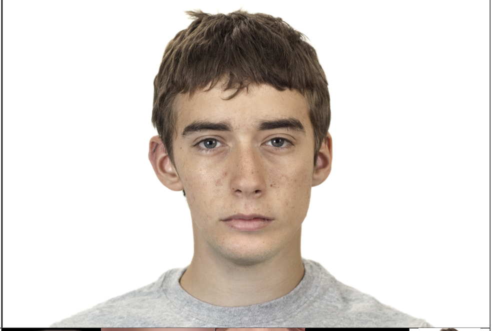
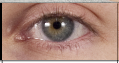
To the best of our knowledge, there are not similar studies to compare these results. In (Stefanie Ritz-Timme et al., 2011), the applicability and feasibility of the DMV atlas (Aßmann, 2007) was tested measuring the inter-observer and intra-observer errors when classifying several morphological features of male faces (e.g. head shape, nose bridge length, chin shape...). As an example, in this test the shape of the chin (*jawline*) was classified into three classes. Despite the low number of classes, the inter-observer error was approximately 39%, while the intra-observer error was 30% for no-experienced observers. These results reflect the subjectivity when judging facial features and the wide variability; every observer showed a specific recognition pattern for the individual facial features. Moreover, this study also concluded that the morphologic assessment of faces is affected by cultural variables. Although more tests must be done, on the light of these results it can be concluded that the proposed automatic procedure is a good approach to classify facial features. Furthermore, the use of the proposed method is not restricted to facial features, and it should be possible to extend its use to automatically group any other kind of images by appearance.

Nevertheless, this study has some limitations. The experiment carried out employed 290 images of males with neutral expression from the Chicago Face Database. Therefore, the taxonomies obtained are only representative of the features of the faces belonging to this database. The representativeness of these taxonomies with respect to other populations must be carefully analyzed before using them. The objective of this work was not to achieve the taxonomies but to develop the automatic procedure to classify facial features based on their appearance. A more comprehensive face database can be used to obtain more representative taxonomies. On the other hand, future work must be done to extend this procedure to other facial features such as hair, and to obtain features taxonomies from faces of females. Moreover, the asymmetry of the face could be taken into account by introducing more horizontal distances to characterize the face.

To sum up, although judging the similarity of facial features is a subjective process with wide inter-observer and intra-observer variability, the results of the validation survey developed in this work show that the proposed procedure can be considered appropriate for the automatic classification of facial features based on their appearance. This procedure deals with the difficulties associated to classify features using judgments from human observers, and facilitates the development of facial features taxonomies. Table 2.22 shows an example of facial feature clustering for each ethnicity using the method implemented in this thesis.

The facial feature clustering method explained in this chapter will be used in this thesis to generate new faces which elicit certain impressions. The methods followed to achieve it are explained in next chapter.

Table 2.22: Example of the clustering of the facial features for each ethnicity.

Asian					Black				
									
									
AEB7	AE6	AN10	AM6	AJ13	BEB7	BE14	BN2	BM4	BJ10
ADEB10	ADE10	ADN9	ADM8	ADEE8	BDEB4	BDE5	BDN6	BDM4	BDEE5
Latino					White				
									
									
LEB6	LE10	LN3	LM4	LJ12	WEB10	WE12	WN12	WM8	WJ1
LDEB1	LDE2	LDN3	LDM5	LDEE3	WDEB4	WDE4	WDN2	WDM2	WDEE5

Chapter 3

Generation of facial social impressions

This chapter explains the procedure followed to develop and implement the model of social trait impressions. First, an introduction on the importance of social trait impressions and previous works is given to the reader. Then, the theoretical framework is explained. Next, the procedure followed is thoroughly reviewed. Finally, results and their validation are shown, and conclusions are drawn in the light of these results.

Contents

3.1 Introduction

People are constantly making attributions from faces, such as whether a person is trustworthy or threatening. It is with good reason that the face is the primary source of visual information for identifying people and reading their emotional and mental states (Todorov, Dotsch, et al., 2011). In fact, these attributions are formed very fast in our brains: an exposure of as little as 34 milliseconds is enough to form an impression, and they do not change with exposures longer than 200 milliseconds. Furthermore, attributions made from faces are very important, as they are very likely to influence our behavior towards people, such as whom we help, whom we hire, or whom we ask for a date (Rule and Ambady, 2010; Zebrowitz and Montepare, 2008).

As Kahneman (2003) suggested, impressions from faces are natural assessments that have more to do with perception than with thinking. The face has been considered as a window to a person's true nature ever since ancient cultures. However, it was in the nineteenth century when this assumption reached his heyday. Lavater (1800), a Swiss pastor, spread the ideas of physiognomy, the "art" of reading personality in faces. Lombroso (2006), the founding father of criminal anthropology, wrote about how it could be possible to identify criminals by external, physical characteristics. Galton (1883) developed the first morphing techniques in his work to identify specific human types ranging from the ideal English man to the criminal. Their ideas were discarded in the twentieth century, however, they were onto something.

A large body of research shows that facial appearances influence significant social outcomes in very diverse domains, such as politics, law, business, and the military. Many of the performed studies find that particular facial characteristics can help to experience desirable outcomes (e.g., winning an election) or avoid undesirable outcomes (e.g., being convicted of a crime) (Olivola et al., 2014). So, what differences in facial structure and appearance lead to these social inferences? For example, what information is used by people to decide if a face looks trustworthy or untrustworthy?

A big problem when trying to model the possible relationships among facial characteristics and social traits is that the space of possible variables driving the perceptions of these traits is infinitely large, so it results almost impossible to solve this problem with conventional approaches. Then, the standard approach is to systematically manipulate a facial feature and ask people to assess the modified face in the range of social traits of interest. However, this approach does not necessarily show that this facial feature is the most important feature for the assessed social trait. First, changes of other social features could produce similar effects. Second, the same feature could be perceived differently in company of other features. Furthermore, it is not even clear how to define facial features. Thus, following the idea that faces are perceived holistically as integrated gestalts rather than as a collection of independent features, two set of techniques can be differentiated: psychological reverse correlation methods (PRCM) and reverse correlation methods in the context of face space models (FSRCM) (Todorov, Dotsch, et al., 2011). PRCM are based on judgments of images that are visually degraded or altered with randomly generated noise. FSRCM, on the other hand, are based on judgments of randomly generated faces from a multidimensional face space model.

3.1.1 Psychological Reverse Correlation Methods

In traditional paradigms, responses depend on meaningful manipulation of stimulus attributes, and thus, relationships are quantified by correlating fixed stimulus attributes with responses. In “reverse” correlation paradigms, on the other hand, variations in stimulus attributes are random, and the correlation between stimuli and responses is used to model those variations in stimulus attributes that caused the acquired response pattern.

There are two popular PRCM techniques, both of them consisting in superimposing noise on visual images. In the first approach, the base image is unambiguous (e.g., a prototypical sad face), while in the second technique, the image is ambiguous (e.g., a morph of two facial expressions). These two approaches also differ in their objectives: while the former, also known as informational diagnosticity PRCM, aims at revealing the diagnostic information used by the perceiver for the specific judgment, for example, sadness; the objective of the latter, also called internal representation PRCM, is to infer the perceiver’s internal representation of the perceptual category (e.g., expression of sadness).

The informational diagnosticity PRCM technique relies on the selection of a predefined signal (e.g., comparing happy an sad faces). In contrast, the internal representation PRCM technique is used when signal attributes are unknown or when researchers want to examine participants’ subjective internal representation of a category, without making any assumption about what typical category members look like.

Some works in internal representation PRCM include those of Mangini and Biederman (2004), who demonstrated that the method works well for gender, identity classification and emotional expression; Dotsch, Wigboldus, et al.

(2008) who employed this method to reveal potential biases in the representation of a stigmatized ethnic out-group; or Dotsch and Todorov (2012), who applied the method to judgments of trustworthiness, dominance and threat.

3.1.2 Reverse Correlation Methods in the Context of Face Space models

While the previous approach made use of noise to achieve its objective, FSRCM is focused on varying properties of the faces directly. It can be divided into two tasks: creating a statistical model of face representation and using this model to derive the changes in facial features that lead to corresponding changes in social judgments. Similarly to PRCM, FSRCM does not explicitly manipulate facial features.

This approach makes use of a face space, where faces are represented as points in a multidimensional space and each dimension is a property of the face. These multidimensional models provide a powerful representational framework that can account for variations in face identity and facial expressions Calder and A. W. Young (2005). Moreover, these models can generate an unlimited number of faces, where each face is a point defined by its coordinates on all face dimensions.

Oosterhof and Todorov (2008) followed this approach to generate models of perceived face trustworthiness, dominance and threat. In a posterior work, they also built models of several other social dimensions, such as attractiveness (Said and Todorov, 2011; Todorov, Dotsch, et al., 2011). On the other hand, Walker and Vetter (2009) built models of six different social dimensions: aggressiveness, extroversion, likeability, risk-seeking, social skills, and trustworthiness. In addition, they applied these models to real faces showing how subtle manipulations in their photographs can lead to the expected social attributions.

3.1.3 Limitations

These previously described methods are powerful tools for identifying diagnostic visual information for social traits perception in faces and for identifying internal representations of particular social traits. However, these methods have some inherent limitations.

First, Mangini and Biederman (2004) showed that the outcomes of reverse correlation paradigms do not really correlate to actual mental representations, but to a quantification of the strategy used when performing a task (which to a large extent correlates with mental representations). This is particularly worrying when using a PRCM approach, where participants are asked to judge a very large number of artificially degraded stimuli, as the motivation of the participant to perform the task will most likely decay with time. FSRCM approaches improve this aspect, as they use non-degraded images of faces that approximate natural social perception as stimuli, which makes the evaluation task more natural and easier for the participant.

Second, not every topic allows for reverse correlation application. Due to the number of necessary trials (relative to the number of stimulus attribute variations) reverse correlation data are commonly analyzed in a linear fashion and interactions between features are mostly disregarded.

Moreover, both approaches need a large number of trials to model interactions, which again has the risk of loosing the participant's motivation and therefore the quality of the procedure would decrease.

Finally, reverse correlation methods have been limited to two categories (e.g. happy versus sad) or one dimension (e.g., trustworthy, dominant, etc.) per task. Outcomes may change considerably when a participant simultaneously considers multiple categories or dimensions.

3.1.4 Our method

Systems able to model face perception are very important, as denoted previously. While there already exist some methods trying to model the perception of social traits from faces, these methods are far from being complete. Thus, this thesis proposes a new method able to work with realistic faces in fifteen categories or trait dimensions at a time.

The rest of the chapter is organized as follows: section 3.2 shows the theoretical background of the implemented methods. Section 3.3 details the procedure followed to assess the groups of features obtained in the previous chapter, the development of a face evaluation function and its optimization, and the face generation algorithm. Section 3.4 details the results and the faces obtained following the methods explained above. Validation of the proposed procedure is provided in section 3.5. Finally, section 3.6 provides conclusions.

3.2 Theoretical background

3.2.1 Genetic Algorithms

Genetic Algorithms (GAs) are search algorithms based on the mechanics of natural selection and natural genetics (Goldberg, 1989). They belong to the field of Evolutionary Computation, which comprises Genetic Algorithms, Evolutionary Strategies and Evolutive Programming. Developed by Holland (1975) and his colleagues at the University of Michigan, GAs are adaptive methods, usually employed in search and optimization problems, based on sexual reproduction and the survival of the fittest. More formally, according to Goldberg (1989), genetic algorithms "combine survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search". They are often understood as function optimizers, even though they can be applied to a high variety of different problems (Whitley, 1994).

According to the current literature search methods can be divided into three branches: calculus-based techniques, enumerative techniques and guided random search techniques (Goldberg, 1989). The main line of research on search techniques has been robustness, that is, the balance between efficacy and efficiency needed for survival in many different environments. GAs have been theoretically and empirically proven to provide robust search in complex spaces while being relatively simple and not limited by restrictive assumptions about the search space (continuity, existence of derivatives, unimodality, etc). Furthermore, they have shown several advantages when compared with the other conventional search methods such as calculus-based or enumerative methods. For instance, calculus-based methods are local in scope, which means that they have no possibility at escaping to local minima, and depend upon the existence of derivatives, which is a severe shortcoming and makes them suitable to a very limited problem domains. On the other hand, enumerative methods rely on computing the objective function at every point in the search space, one at a time, which makes them very inefficient. In contrast, genetic algorithms (which are encompassed into guided random search techniques, Figure 3.1) use random choice as a tool to guide a highly exploitative search through a coding of a search space, overcoming the shortcomings before mentioned. These are the main differences between genetic algorithms and conventional optimization and search procedures:

- GAs work with a coding of the parameter set, not the parameters themselves,

- GAs search from a population of points, not a single point,
- GAs use payoff (objective function) information, not derivatives or any other auxiliary knowledge,
- GAs use probabilistic transition rules, not deterministic rules.

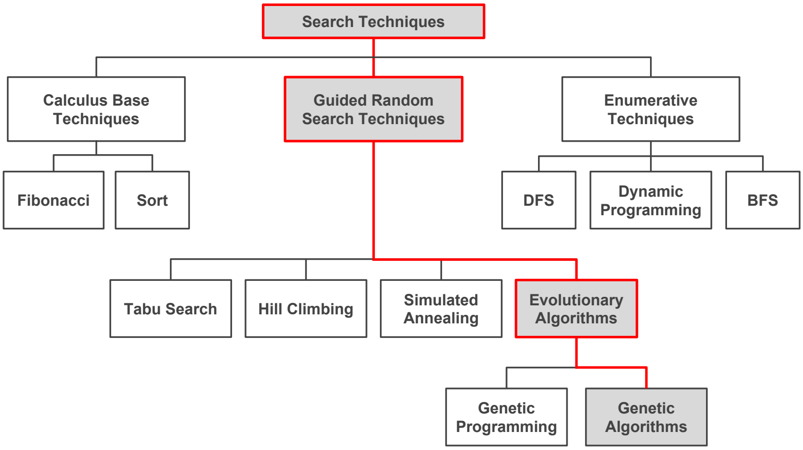


Figure 3.1: Search methods classification.

Components, structure and terminology

Since genetic algorithms are designed to mimic a biological process, many of the terms employed are borrowed from biology (in a much simplified way). The basic components of almost any genetic algorithm are:

- a *fitness function* for optimization,
- a population of *chromosomes*,
- the genetic operators: a *selection*, *crossover* and *mutation* mechanisms to produce the population of the next generation.

The *fitness function* tests and quantifies how “fit” each potential solution is, therefore it is the function that the algorithm is trying to optimize. The *population* is the subset of all the possible *encoded* solutions to the given problem. The *chromosome* refers to the values that represent a candidate solution or *individual*. Each chromosome is composed of several positions or *genes*, which can take a defined range of values or *alleles* (Figure 3.2).

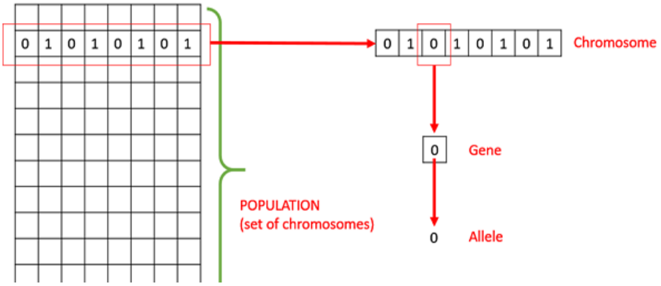


Figure 3.2: Basic terminology of genetic algorithms: population, chromosomes, genes and alleles.

If a problem has N dimensions, then each chromosome is typically encoded as an N-element array $[p_1, p_2, \ldots, p_N]$ where each p_i is a particular value or *allele* of the i^{th} parameter. Here, two solution spaces are distinguished: the computation space or *genotype* and the real world space or *phenotype*. In the former one, the population is encoded in a way it can be easily understood and manipulated using a computer, while in the latter one, solutions are represented in the same way they are represented in the real world. The way to transform elements in one space into the another is by employing the encoding and decoding functions (Figure 3.3). It is up to the creator to decide how to implement the encoding and the decoding functions. One possible implementation of the encoding function is to convert each parameter value into a bit string (sequence of 0's and 1's) and then concatenate them end-to-end like genes in DNA strand to create the chromosomes, but it is also possible to include permutations, real numbers and many other objects inside the chromosomes.

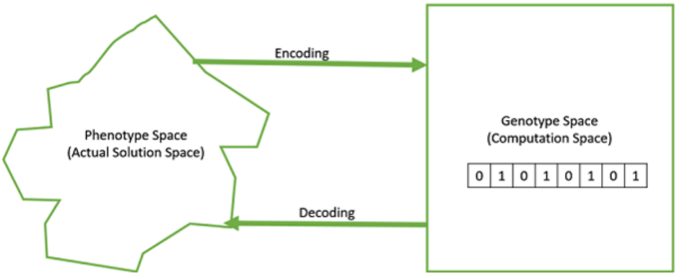


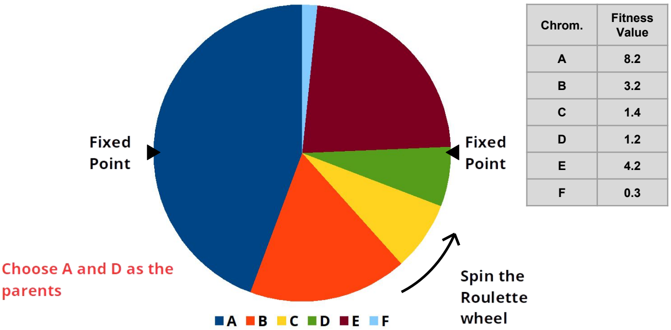
Figure 3.3: Basic terminology of genetic algorithms: genotype, phenotype end encoding and decoding functions.

The initialization of a genetic algorithm consists of a randomly created set of chromosomes which serves as the initial population (first generation). Then, this initial population is evaluated with the fitness function to test how "good" each solution is to the problem at hand. Now, the *selection operator* chooses some of the chromosomes for reproduction based on a user defined probability distribution. The fitter a chromosome is, the more likely it is to be selected. For instance, let's define f as a non-negative fitness function. Then, the probability that chromosome C_{36} is chosen to reproduce from a population N_{pop} might be

$$P(C_{36}) = \left| \frac{f(C_{36})}{\sum_{i=1}^{N_{pop}} f(C_i)} \right| . \tag{3.1}$$

Selection operation is performed with replacement, so it is possible that the same chromosome is chosen more than once. There exist a wide variety of methods to perform chromosome selection, such as fitness proportionate selection, tournament selection, rank selection, random selection or elitism. The most basic one is the fitness proportionate selection (FPS), in which every individual can become a parent with a probability proportional to its fitness. There are several approaches within the FPS methods, such as the Roulette Wheel Selection (RWS) and the Stochastic Universal Sampling (SUS). The RWS is the most basic one, in which as many chromosomes are chosen according to where a randomly generated number is located within the fitness distribution. However, this method punishes in excess chromosomes with low fitness value. In contrast, SUS is a development of fitness proportionate selection (FPS) which exhibits no bias and minimal spread. Where RWS chooses several solutions from the population by repeated random sampling, SUS uses a single random value to sample all of the solutions by choosing them at evenly spaced intervals (Figure 3.4). An important limitation of FPS methods is that they cannot deal with negative fitness values.

The *crossover operator* is analogous to reproduction and biological crossover. In it, two parents are selected and one or more off-spring are produced using the genetic material of the parents. It is usually applied with a high probability $P_{crossover}$. Some of the most used crossover methods are the one-point crossover, multi-point crossover, uniform crossover, whole arithmetic recombination or Davis' order crossover (OX1). The most basic crossover operator is the one-point crossover, in which a random crossover point is selected and the tails of its two parents are swapped to get new off-springs (Figure 3.5).



Chrom.	Fitness Value
A	8.2
B	3.2
C	1.4
D	1.2
E	4.2
F	0.3

Figure 3.4: Stochastic Universal Sampling (SUS) implementation of fitness proportionate selection.

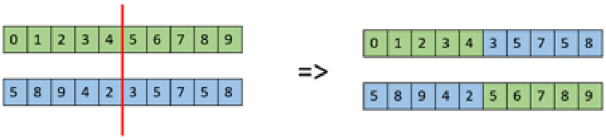


Figure 3.5: One-point crossover.

The *mutation operator* can be defined as a small random tweak in the chromosome to get a new solution. It is used to maintain and introduce diversity in the genetic population and is usually applied with a low probability ($P_{mutation}$). In fact, if the probability is very high, the GA behaves like a random search. In turn, if the probability is very low, the algorithm can get stuck in a local optimum instead of finding the global optimum (R. L. Haupt and S. E. Haupt, 2004). Some commonly used mutation operators are the bat flip mutation, random resetting, uniform mutation, swap mutation, scramble mutation and inversion mutation. The most simple mutation operator is the bat flip mutation, used for binary encoded GAs, that consists in selecting one or more bits and flip them (Figure 3.6).

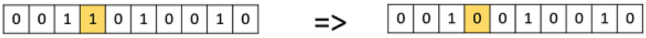


Figure 3.6: Bat flip mutation.

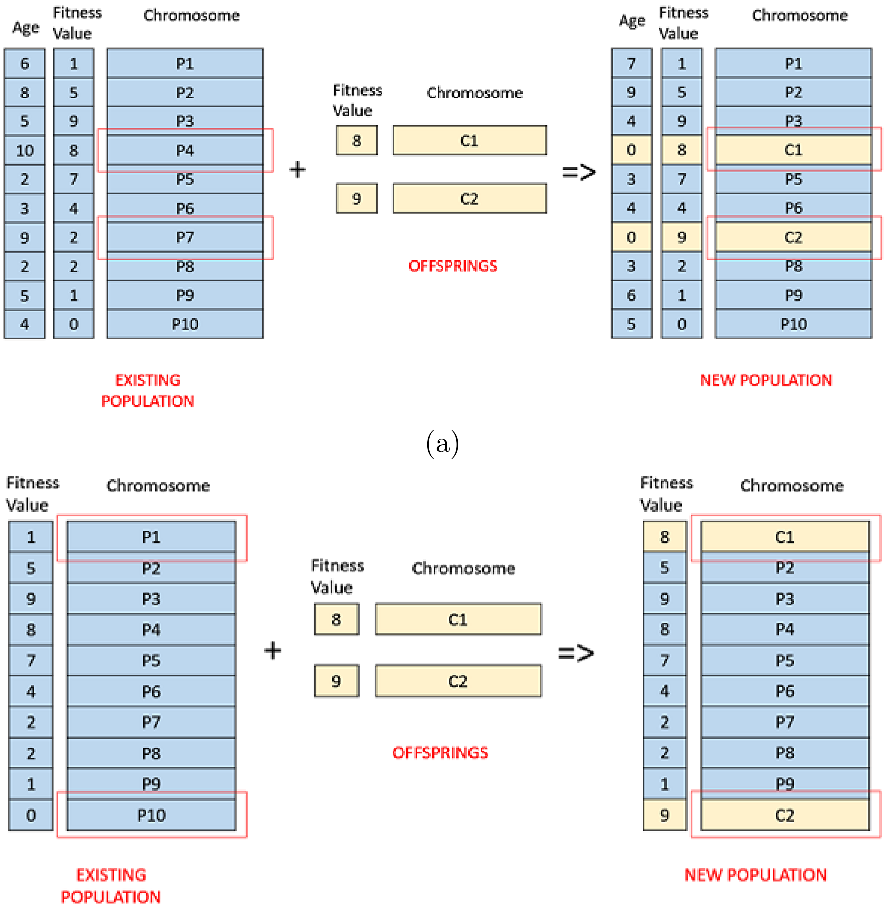
However, in this work, due to the nature of the representation employed (floating-point representation), the uniform mutation method was used (an extension of the bit flip for the float representation). In this method, a probability of mutation is randomly computed for each allele, and those which are higher than the mutation probability ($P_{mutation}$), are muted. Alleles which are to be muted take their new values from a uniform distribution that goes from the lower to the upper established boundaries.

Typically, selection, crossover and mutation processes continue until the number of offspring is the same as the initial population, so that the first generation can be completely replaced by the new offspring. This replacement is performed according to the Survivor Selection Policy, which determines which individuals are to be kicked out and which are to be kept in the next generation based on the fitness value. Some GAs employ *elitism*, ensuring that the current *fittest* member of the population is always propagated to the next generation. The easiest policy is to kick random members out of the population, but such an approach frequently has convergence issues, therefore, two strategies are widely used: age-based selection and fitness-based selection. In age-based selection each individual is allowed in the population for a finite number of generations after which it is kicked out. An example is shown in Figure 3.7 (a), where chromosome P4 and P7 are the ones to be replaced due to their *age*. On the other hand, in fitness-based selection, the children tend to replace the least fit individuals in the population. The selection of these least fit individuals can be done using a variation of any of the selection methods mentioned before (tournament selection, fitness proportionate selection, rank selection, etc.). Figure 3.7 (b) shows an example in which P1 and P10 are the least fit individuals and are the ones to be replaced. In addition, the best chromosome (the one with the highest fitness value) is stored along with its fitness value (S_{best}).

The whole process of iterations is called a run, and it is usually repeated until the algorithm converges, that is, until the fitness of the "*best-so-far*" chromosome stabilizes and does not change for a number of generations. However, there are other stop criteria, such as an absolute number of generations or a certain pre-defined fitness value to be reached. An advantage of GAs is that there is a "*best-so-far*" solution available since iteration 1.

The "performance" of a genetic algorithm depends highly on the encoding used and the fitness function, as well as the crossover and mutation probabilities. Then, it is of sum importance to find a proper encoding and a good fitness function. The crossover and mutation probabilities are usually chosen empirically after performing a few trial runs.

The pseudo-code of a genetic algorithm can be observed in algorithm 3.1.



Age	Fitness Value	Chromosome
6	1	P1
8	5	P2
5	9	P3
10	8	P4
2	7	P5
3	4	P6
9	2	P7
2	2	P8
5	1	P9
4	0	P10

Fitness Value	Chromosome
8	C1
9	C2

Age	Fitness Value	Chromosome
7	1	P1
9	5	P2
4	9	P3
0	8	C1
3	7	P5
4	4	P6
0	9	C2
3	2	P8
6	1	P9
5	0	P10

(a)

Fitness Value	Chromosome
1	P1
5	P2
9	P3
8	P4
7	P5
4	P6
2	P7
2	P8
1	P9
0	P10

Fitness Value	Chromosome
8	C1
9	C2

Fitness Value	Chromosome
8	C1
5	P2
9	P3
8	P4
7	P5
4	P6
2	P7
2	P8
1	P9
9	C2

(b)

Figure 3.7: Survivor Selection Policies. In (a), age-based selection policy. In (b), fitness-based selection policy.

3.2.2 Image seamless cloning

Seamless cloning consists in copying an image region from a foreground image onto a background image without visual seams. These visual seams are usually produced by differences in the color and texture of the two images. Naive cloning, that is, cropping the desired region from the foreground image and pasting it onto the background image produces an artificial result due to these

Algorithm 3.1: Genetic algorithm pseudo-code.

```
Input: $Population_{size}, Problem_{size}, P_{Crossover}, P_{mutation}$
Output: $S_{best}$
Population → InitializePopulation ($Population_{size}$, $Problem_{size}$) ;
EvaluatePopulation (Population) ;
$S_{best}$ → GetBestSolution (Population) ;
while ¬StopCondition do
    Parents ← SelectParents (Population, $Population_{size}$) ;
    Children ← ∅ ;
    foreach $Parent_1, Parent_2$ ∈ Parents do
        $Child_1, Child_2$ ← Crossover ($Parent_1, Parent_2, P_{crossover}$) ;
        Children ← Mutate ($Child_1, P_{mutation}$) ;
        Children ← Mutate ($Child_2, P_{mutation}$) ;
    end
    EvaluatePopulation (Children) ;
    $S_{best}$ ← GetBestSolution (Children) ;
    Population ← Replace (Population, Children) ;
end
return $S_{best}$;
```

seams. On the contrary, seamless cloning is able to remove these seams. Figure Figure 3.8 shows the difference between these two cloning methods.

Poisson method

One of the most famous seamless cloning method is that solving the Poisson equation, which works by interpolating an image aided with a guidance vector field (Pérez et al., 2003). As a color image has three channels, the interpolation problem needs to be solved for each color component separately. Working with each component independently allows for considering only scalar image functions. Let S, a closed subset of $\mathbb{R}^2$, be the image definition domain, and Ω a closed subset of S with boundary $\partial\Omega$. Let f^* be a known scalar function defined over S minus the interior of Ω and f an unknown scalar function defined over the interior of Ω. Finally, let $\mathbf{v}$ be a vector field defined over Ω (Figure 3.9).

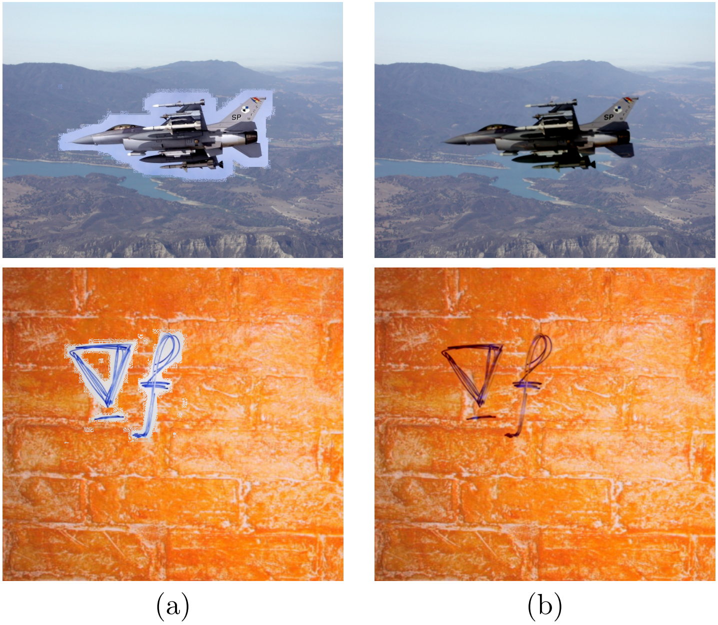


Figure 3.8: Differences between naive cloning (a) and seamless cloning (b). Image extracted from Pérez et al. (2003).

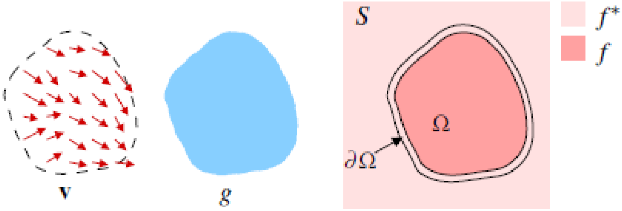


Figure 3.9: Guided interpolation notations. Unknown function f interpolates in domain Ω the destination function f^* under guidance of vector field $\mathbf{v}$, which might be or not the gradient field of a source function g. Image extracted from Pérez et al. (2003).

The simplest function f which interpolates f^* over Ω is the membrane interpolant defined as the solution of the minimization problem:

$$\min_f \iint_\Omega |\nabla f|^2 \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}\,, \tag{3.2}$$

where $\nabla = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right]$ is the gradient operator. The minimizer must satisfy the Euler-Lagrange equation

$$\Delta f = 0 \text{ over } \Omega \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}\,, \tag{3.3}$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. Equation 3.3 is a Laplace equation with Dirichlet boundary conditions. This simple method produces an unsatisfactory, blurred interpolant when used for image editing. To overcome this limitation, further constraints are introduced to the problem in the form of a guidance vector. A *guidance vector* is a vector field $\mathbf{v}$ used in an extended version of the minimization problem (3.2) above:

$$\min_f \iint_\Omega |\nabla f - \mathbf{v}|^2 \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}\,, \tag{3.4}$$

whose solution is the unique solution of the following Poisson equation with Dirichlet boundary conditions:

$$\Delta f = \operatorname{div}(\mathbf{v}) \text{ over } \Omega \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}\,, \tag{3.5}$$

where $\operatorname{div}(\mathbf{v}) = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$ is the divergence of $\mathbf{v} = (u, v)$. Equation 3.5 allows to fill in the region of interest following the color and intensity variations corresponding to the scene to be inserted while keeps the solution coherent with the intensities of the background image. When dealing with color images, three Poisson equations (3.5) are solved independently in the three color channels of the utilized color space.

The basic choice for the guidance field $\mathbf{v}$ to compute it from the source image. However, there are situations where is desirable to combine both the source and destination images. In this situations a mixing gradient is used. This will be explained in-detail in the following section.

When the guidance field $\mathbf{v}$ is conservative, that is, it is the gradient of some function g, it is helpful to understand the Poisson interpolation by defining a correction function $\tilde{f}$ on Ω such that $f = g + \tilde{f}$. Then, the Poisson equation (3.5) becomes the following Laplace equation with boundary conditions:

$$\Delta \tilde{f} = 0 \text{ over } \Omega \text{ with } \tilde{f}|_{\partial\Omega} = \tilde{f}^*|_{\partial\Omega}\,. \tag{3.6}$$

Therefore, inside Ω, the additive correction $\tilde{f}$ is a membrane interpolant of the mismatch $(f^* - g)$ between the source and the destination along the boundary $\partial\Omega$.

Discrete Poisson solver

The variational problem (3.4) and the associated Poisson equation with Dirichlet boundary conditions (3.5) for discrete images can be discretized and solved by using the underlying discrete pixel grid. Let's then define S and Ω as finite point sets taken from an infinite discrete grid. S can include all the image pixels or only a subset of them. For each pixel $p \in S$, let N_p be the set of its 4-connected neighbors which are in S, and let $\langle p, q \rangle$ denote a pixel pair such that $q \in N_p$. Then, the boundary of Ω is $\partial\Omega = \{p \in S \setminus \Omega : N_p \cap \Omega \neq \emptyset\}$. The task is then to compute the set of intensities $f|_\Omega = \{f_p, \, p \in \Omega\}$, being f_p the intensity value of f at p.

When dealing with Dirichlet boundary conditions defined on an arbitrary shape boundary, it is better to discretize the variational problem (3.4) directly rather than the Poisson equation (3.5). The finite difference discretization of (3.4) yields the following discrete, quadratic optimization problem:

$$\min_{f|\Omega} \sum_{\langle p,q \rangle \cap \Omega \neq \emptyset} \left(f_p - f_q - v_{pq}\right)^2, \text{ with } f_p = f_p^* \text{ for all } p \in \partial\Omega\,, \tag{3.7}$$

where v_{pq} is the projection of $\mathbf{v}\left(\frac{p+q}{2}\right)$ on the oriented edge $[p, q]$, that is, $v_{pq} = \left(\frac{p+q}{2}\right) \cdot \overrightarrow{pq}$. The solution of (3.7) satisfies the following simultaneous linear equations:

$$\text{for all } p \in \Omega, |N_p| f_p - \sum_{q \in N_p \cap \Omega} f_q = \sum_{q \in N_p \cap \partial\Omega} f_q^* + \sum_{q \in N_p} v_{pq}\,. \tag{3.8}$$

When Ω contains pixels on the border of S (edges of the pixel grid), they have a truncated neighborhood ($|N_p| < 4$). On the contrary, pixels in the interior of Ω ($N_p \in \Omega$) have no boundary terms in the right hand side of (3.8):

$$|N_p| f_p - \sum_{q \in N_p} f_q = \sum_{q \in N_p} v_{pq}\,. \tag{3.9}$$

Equation 3.8 is a classical, sparse, symmetric, positive-definite system. Due to the arbitrarity of the shape boundary $\partial\Omega$ iterative solvers such as Gauss-Seidel iteration with successive over-relaxation or V-cycle must be used to solve it.

Guidance fields

As was briefly mentioned in the previous section, there exist two options as to choose the guidance field. The first and most basic one consists in choosing the guidance field $\mathbf{v}$ as a gradient taken directly from the source image (*imported gradient*), while the second and more powerful approach combines the gradients of both the source and destination image (*mixed gradient*), which is useful to add objects with holes or partially transparent, for example.

Imported gradients

As was previously mentioned, the basic choice for the guidance field $\mathbf{v}$ is a gradient field computed directly from a source image. Let g be the source image. The guidance field is defined as:

$$\mathbf{v} = \nabla g \,. \tag{3.10}$$

Equation 3.5 then reads

$$\Delta f = \Delta g \text{ over } \Omega \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega} \,. \tag{3.11}$$

Regarding the numerical implementation, the continuous specification (3.10) translates into

$$\text{for all } \langle p, q\rangle \,, v_{pq} = g_p - g_q \,, \tag{3.12}$$

which must be plugged into Equation 3.8.

Mixed gradients

Imported gradients described in the previous section have a limitation: no trace of the image f^* is kept inside Ω, that is, there is no possibility of combining properties of f^* with those of g. This could be useful to add objects with holes or partially transparent on top of a textured or cluttered background (Figure 3.10).

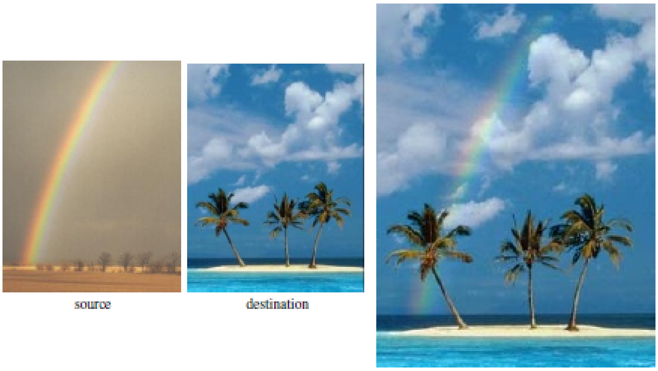


Figure 3.10: Inserting transparent objects. Seamless cloning with mixed gradients as guidance field facilitates the transfer of partially transparent objects. Image extracted from Pérez et al. (2003).

Mixed gradients overcome this limitation by effectively combining both the source and destination image gradients. One possibility would be to define the guidance field $\mathbf{v}$ as a linear combination of the source and destination gradient fields. However, this approach has the effect of washing out the textures, which is not desired (Figure 3.11 (c)).

Another better possibility is to retain the stronger of the variations in f^* or g at each point $\mathbf{x} \in \Omega$ using the following guidance field:

$$\mathbf{v}(\mathbf{x}) = \begin{cases} \nabla f^*(\mathbf{x}) & \text{if } |\nabla f^*(\mathbf{x})| > |\nabla g(\mathbf{x})|, \\ \nabla g(\mathbf{x}) & \text{otherwise.} \end{cases} \tag{3.13}$$

This is possible because the Poisson methodology allows for the use of non-conservative guidance fields. As for the numerical implementation of this guidance field:

$$\text{for all } \langle p, q \rangle, v_{pq} = \begin{cases} f_p^* - f_q * & \text{if } |f_p^* - f_q^*| > |g_p - g_q|, \\ g_p - g_q & \text{otherwise.} \end{cases} \tag{3.14}$$

Mixed gradient seamless cloning is also very useful when adding one object from a source image very close to another object in the destination image (Figure 3.12).

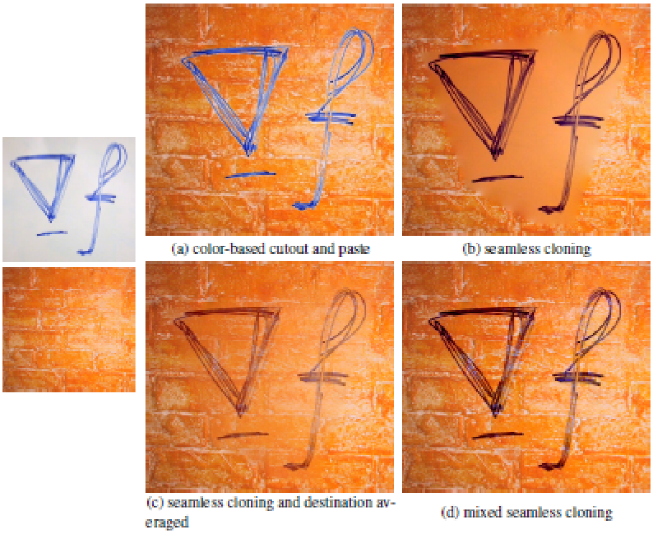

(a) color-based cutout and paste
(b) seamless cloning
(c) seamless cloning and destination averaged
(d) mixed seamless cloning

Figure 3.11: Inserting objects with holes. (a) The classic method, color-based selection and alpha masking. (b) Seamless cloning with imported gradient. (c) Seamless cloning with mixed gradient as a linear combination of source and destination gradients. (d) Seamless cloning with mixed gradient based on gradient selection for each point $\mathbf{x} \in \Omega$. Image extracted from Pérez et al. (2003).

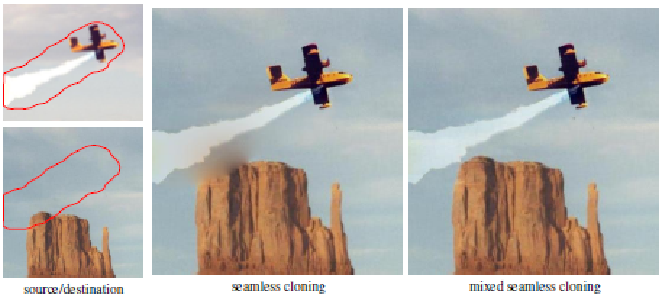


Figure 3.12: Inserting one object close to another. Seamless cloning with mixed gradients as guidance field inhibits the bleeding produced with imported gradient when an object in the destination image touches the selected region Ω. Image extracted from Pérez et al. (2003).

3.3 Generation of facial social impressions

Faces are the most exposed part of the body when interacting with other people. Therefore, the human being is able to process a face and extract information regarding the social traits that person has very quickly (Bar et al., 2006). As has been already mentioned in several occasions, the main objective of this thesis is to develop a system able to create faces with a certain subset of these social traits in order to have influence on the opinion the observer forms when seeing a face. The method developed to achieve this objective is split into three stages:

- Assessment of the created facial feature clusters in the available social traits.
- Optimization of the function employed to compute the score of a face given a certain set of facial features.
- Implementation of the face generation system.

The creation of a database of facial features grouped by similarity was explained in chapter 2. However, these feature clusterings need to have an associated score for each available social trait. This is performed by employing the data available in CFD, consisting of assessments for 15 *social traits*[1]. Then, with the feature clusters obtained by the process described in the previous chapter and the scores available for each face and social trait within the database, it is possible to give each cluster a score for each of these social traits. This process was already mentioned when explaining the necessity of the clustering, in section 2.5. However, it is explained in depth in subsection 3.3.1.

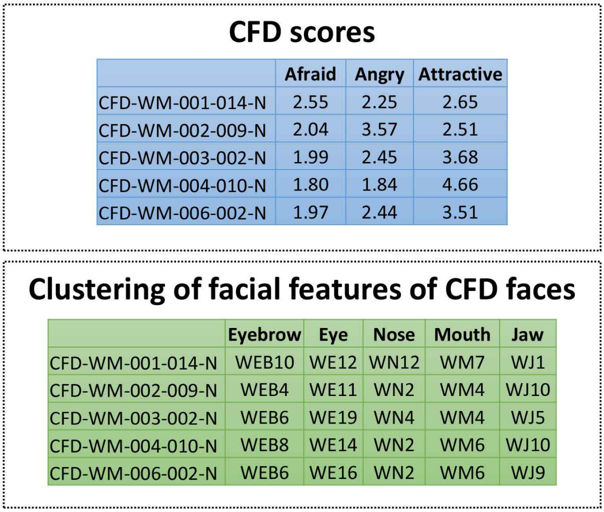
Another stage of the system is the optimization of the function used to evaluate a face. This function is needed in order to compute the impression scores of a given set of facial features. Basically, this function was implemented as a weighted sum of each facial feature score. In order to find the best possible combination of weights for this function, a GA was employed. This process is depicted in subsection 3.3.2.

Finally, the last stage involves the creation of a face with the most suitable combination of features for the desired social trait profile. Section 3.3.3 explains this process in detail.

[1] CFD database consists of photographs of faces in several emotional states (neutral, happy, afraid, etc). For this thesis purpose, only faces which were not expressing any emotion were selected (i.e. neutral faces). Therefore, it is possible to assume that scores given to these neutral faces correspond to their social traits, as they are not expressing any emotion.

3.3.1 Social trait assessment of the facial feature clusters

In order to build a new face from the existing facial features clustered in the previous chapter, a score for each trait impression needs to be associated to these clusters. These scores are called *clustering mean scores*. The process followed to compute them is explained thoroughly in the following lines. For the sake of simplicity, let's consider only 3 trait impressions and 5 facial features, as Figure 3.13 shows. These are the data available to give each cluster a score: on one side, the CFD scores for each face and impression; on the other side, the clustering obtained in chapter 2.



CFD scores

	Afraid	Angry	Attractive
CFD-WM-001-014-N	2.55	2.25	2.65
CFD-WM-002-009-N	2.04	3.57	2.51
CFD-WM-003-002-N	1.99	2.45	3.68
CFD-WM-004-010-N	1.80	1.84	4.66
CFD-WM-006-002-N	1.97	2.44	3.51

Clustering of facial features of CFD faces

	Eyebrow	Eye	Nose	Mouth	Jaw
CFD-WM-001-014-N	WEB10	WE12	WN12	WM7	WJ1
CFD-WM-002-009-N	WEB4	WE11	WN2	WM4	WJ10
CFD-WM-003-002-N	WEB6	WE19	WN4	WM4	WJ5
CFD-WM-004-010-N	WEB8	WE14	WN2	WM6	WJ10
CFD-WM-006-002-N	WEB6	WE16	WN2	WM6	WJ9

Figure 3.13: Data available to create the *cluster mean scores*. The cluster nomenclature is as follows: W stands for White ethnicity, EB for eyebrow, E for eye, N for nose, M for mouth and J for jawline.

In order to assign scores to each cluster, the mean of all the faces to which the instances of the clustering belong is computed. This is done for every impression. Then, following this example, the *cluster mean score* of Afraid for cluster 6 of White eyebrows would be computed as:

$$\mathrm{WEB6}_{Afraid} = \frac{1.99 + 1.97}{2} = 1.98\,. \tag{3.15}$$

This is easy to understand looking at Figure 3.13, where it is shown that image CFD-WM-003 and CFD-WM-006 belong to the cluster WEB6. Therefore, looking the values for these two faces corresponding to Angry, it is possible to build Equation 3.15.

Following the same procedure, the *cluster mean score* of Angry for WN2 would be:

$$\mathrm{WN2}_{Angry} = \frac{3.57 + 1.84 + 2.44}{3} = 2.62\,. \tag{3.16}$$

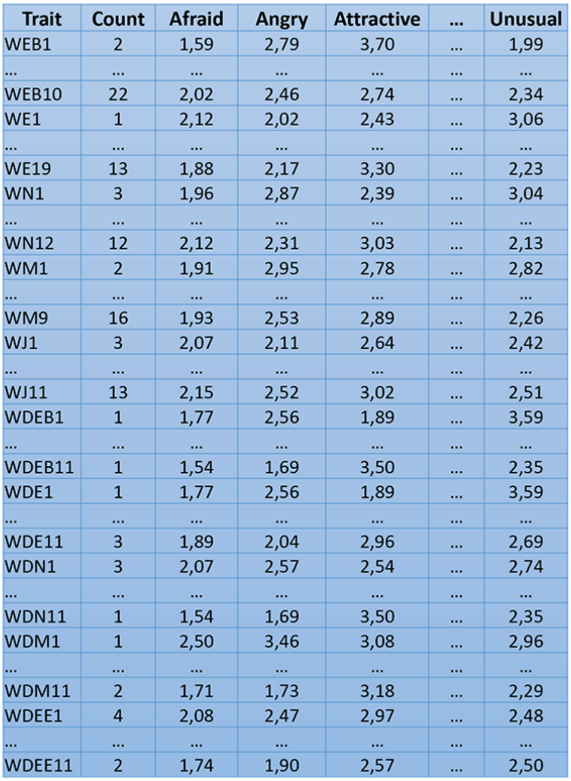
Similarly, the *cluster mean score* of Attractive for cluster 4 of White mouths would be defined as:

$$\mathrm{WM4}_{Attractive} = \frac{2.51 + 3.68}{2} = 3.10\,. \tag{3.17}$$

These examples illustrate the method followed to compute the *cluster mean score* of each facial feature, impression and ethnicity. Figure 3.14 shows an example for the computation of the cluster mean scores for White ethnicity. Rows represent each cluster and columns the score for each impression computed as indicated above.

3.3.2 Optimization of the face evaluation function

By having the *cluster mean score* of each social trait and cluster, it is possible to compute the social trait profile of a face. To obtain the score of a face for a given social trait, the *cluster mean scores* of this trait corresponding to the selected facial features are pre-multiplied by the set of weights corresponding to the selected trait (one per facial feature) and added up. These weights are optimized by means of a Genetic Algorithm as will be shown later in this section. Performing this operation for the fifteen traits available gives the predicted face social trait profile. Equation 3.18 shows the calculation of the *global face score (GFS)* for a social trait, where t denotes the social trait for which the score is being calculated.



Trait	Count	Afraid	Angry	Attractive	...	Unusual
WEB1	2	1,59	2,79	3,70	...	1,99
...	...	...	...	...	...	...
WEB10	22	2,02	2,46	2,74	...	2,34
WE1	1	2,12	2,02	2,43	...	3,06
...	...	...	...	...	...	...
WE19	13	1,88	2,17	3,30	...	2,23
WN1	3	1,96	2,87	2,39	...	3,04
...	...	...	...	...	...	...
WN12	12	2,12	2,31	3,03	...	2,13
WM1	2	1,91	2,95	2,78	...	2,82
...	...	...	...	...	...	...
WM9	16	1,93	2,53	2,89	...	2,26
WJ1	3	2,07	2,11	2,64	...	2,42
...	...	...	...	...	...	...
WJ11	13	2,15	2,52	3,02	...	2,51
WDEB1	1	1,77	2,56	1,89	...	3,59
...	...	...	...	...	...	...
WDEB11	1	1,54	1,69	3,50	...	2,35
WDE1	1	1,77	2,56	1,89	...	3,59
...	...	...	...	...	...	...
WDE11	3	1,89	2,04	2,96	...	2,69
WDN1	3	2,07	2,57	2,54	...	2,74
...	...	...	...	...	...	...
WDN11	1	1,54	1,69	3,50	...	2,35
WDM1	1	2,50	3,46	3,08	...	2,96
...	...	...	...	...	...	...
WDM11	2	1,71	1,73	3,18	...	2,29
WDEE1	4	2,08	2,47	2,97	...	2,48
...	...	...	...	...	...	...
WDEE11	2	1,74	1,90	2,57	...	2,50

Figure 3.14: *Cluster mean scores* of each impression and feature for White ethnicity.

$$GFS^t = \begin{pmatrix} score^t_{EB} \\ score^t_{E} \\ score^t_{N} \\ score^t_{M} \\ score^t_{J} \\ score^t_{DEB} \\ score^t_{DE} \\ score^t_{DN} \\ score^t_{DM} \\ score^t_{DEE} \end{pmatrix} * \begin{pmatrix} weight^t_{EB} \\ weight^t_{E} \\ weight^t_{N} \\ weight^t_{M} \\ weight^t_{J} \\ weight^t_{DEB} \\ weight^t_{DE} \\ weight^t_{DN} \\ weight^t_{DM} \\ weight^t_{DEE} \end{pmatrix}^{\top} \tag{3.18}$$

However, as the values employed in order to compute this calculation are the means of the facial features belonging to each cluster, the variance is diminished in excess. Therefore, this prediction function cannot reproduce extreme values present in the CFD database, as the means have flattened them. In order to solve this situation, the predicted values are transformed so they have the same mean and standard deviation as the original CFD scores. As each social trait has a different mean and standard deviation, this process is performed separately for each of them. Equation 3.19 illustrates this operation.

$$GFS_{expanded} = \frac{GFS - \mu_{GFS}}{\sigma_{GFS}} \cdot \sigma_{CFD} + \mu_{CFD} \tag{3.19}$$

GFS stands for the global face score for a given impression. μ_{GFS} and σ_{GFS} are defined as the mean and the standard deviation of the global face score obtained with the evaluation function. Finally, μ_{GFS} and σ_{GFS} are the mean and the standard deviation of the CFD scores. By performing this operation, original global scores are closer to the scale employed in CFD scores, so they can be compared properly. The metric used to compare the predicted GFS and the CFD scores was the mean squared error.

The capability of the developed model lies on achieving a good fitting of the evaluating function to the data available on the CFD. Using a weight per feature and impression allows to grant a different level of importance to each facial feature on the formation of each impression. Therefore, each model is more specific and captures in a better way the relationships among features to obtain the global face score of a given impression. Then, it is very important to find a good combination of weights that allows to obtain a global face score in harmony with what most people would think of that face.

The intuitively way to compute these weights would be to use an optimization method or a grid search. However, due to the high number of variables and the size of the search space, a Genetic Algorithm was chosen instead in order to find a solution in the most efficient manner.

Weight optimization

The process of weight optimization allowed to find relationships among facial features and social traits. The GA employed to perform this optimization was configured to perform single-point crossover and uniform mutation with a probability of $P_{crossover} = 0.6$ and $P_{mutation} = 0.4$ respectively on a population of 50 individuals. The permitted range for the weights was set to the interval $[0, 1]$. The number of iterations was established at 200 000, however, it was never reached due to the early stopping condition implemented. This condition allowed for a maximum of 100 consecutive iterations without a change higher than 0.0001 in the fitness function solution. If iteration number 101 was reached without a change higher than the before mentioned tolerance, the best individual of the last iteration was chosen as the solution. The selection method employed was Stochastic Universal Sampling, and the Survivor Selection Policy was fitness-based with *elitism*, that is, the best individual was always selected for the next iteration. Finally, the fitness function was defined as the mean squared error between the predictions made with a given combination of weights and the actual face scores of the used faces. For example, if there were 50 individuals (*solutions*) in iteration 1 with 50 different combination of weights, the fitness function would compute the mean squared error between the predictions obtained with each combination of weights and the actual scores of the White faces. This process would result in 50 fitness values, where lower is better. Table 3.1 shows an example of this process.

Table 3.1: Example of the fitness computation for an iteration of the Genetic Algorithm for weights optimization. The best solution present in this iteration is marked in blue. FV refers to Fitness Value.

Eyebrow	Eye	Nose	Mouth	Jawline	DEB	DE	DN	DM	DEE	FV
0.0776	0.2023	0.0839	0.1171	0.1255	0.0358	0.0365	0.1520	0.0505	0.1188	0.2371
0.1189	0.0409	0.0813	0.1115	0.0672	0.1137	0.0862	0.1071	0.1307	0.1426	0.5436
0.1617	0.2092	0.0210	0.2580	0.0033	0.0860	0.0727	0.1245	0.0421	0.0216	0.1874
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.1261	0.1729	0.0124	0.0298	0.0773	0.1212	0.1500	0.0192	0.1894	0.1015	0.3694

Following the example showed in Table 3.1, the best combination of weights would be the one marked in blue, with $FV = 0.1874$. With this configuration, the optimization was performed individually for each ethnicity and impression, resulting in a total of 60 weight configurations, one for each trait impression and ethnicity (15 trait impressions × 4 ethnicities = 60 weight configurations).

3.3.3 Implementation of the face generation system

In order to create a face which expresses a certain profile of impressions, it would be necessary to compare the desired social trait profile with that of the faces created with every possible combination of features. Again, this is practically impossible due to the amount of time needed. Therefore, a GA was employed to find the best combination of features needed to create a face expressing the desired profile of social traits. Figure 3.15 shows the flowchart of the implemented face generation system.

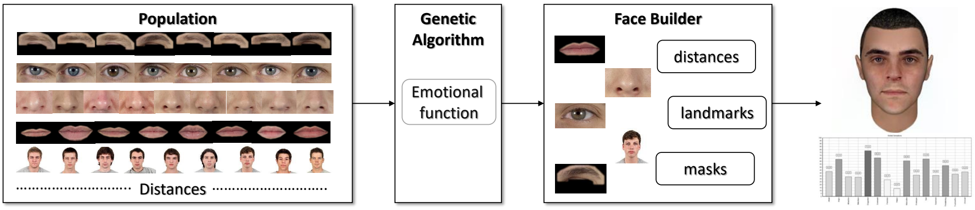


Figure 3.15: Face generator flowchart.

The process of creation of a new face involves two steps: face definition and face generation. Face definition refers to the process of finding the best combination of facial features for a certain social trait profile. Facial surgery is the process of adding the features into a *base face* to obtain the desired new face.

The following sections explain the process of creation of the *base faces*, the face definition procedure and the face generation process.

Base face generation

Base faces is the way to introduce the jawline in the model. In order to avoid automatically morphing the base face according to the jawline chosen by the GA each time a new face had to be generated, it was done manually beforehand.

Then, to create the base faces, a face was generated with FaceGen (Inversions, 2008). This face was chosen as the mean male face of each ethnicity. Figure 3.16 shows the mean male face for White ethnicity. Due to time limitations, and to follow the line of the thesis, base faces were generated only for White ethnicity.

In order to know where the facial features were located within the base face, the same procedure as for CFD images was applied. This procedure consisted in detecting the facial landmarks and creating the corresponding *thickened masks*. After this procedure, a mask for each facial feature was available, as shows Figure 3.17. These masks were necessary to indicate the image editing algorithm which regions were to be used in each step.

The base face was then morphed to match each jawline representative using Adobe Photoshop.

Table 3.2 shows the 11 base faces created corresponding to the 11 jawline representatives of the jawline clustering for Whites.

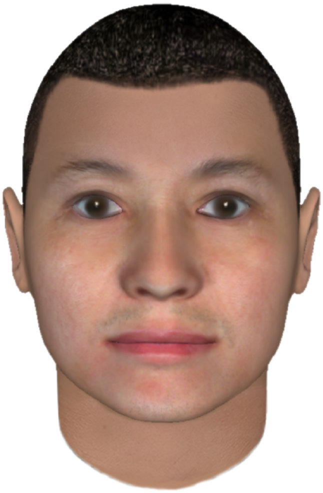


Figure 3.16: Base face for White ethnicity.

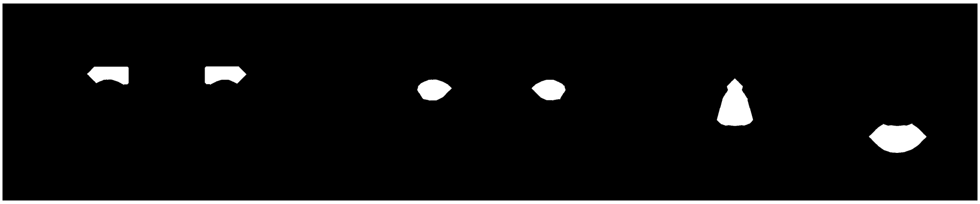
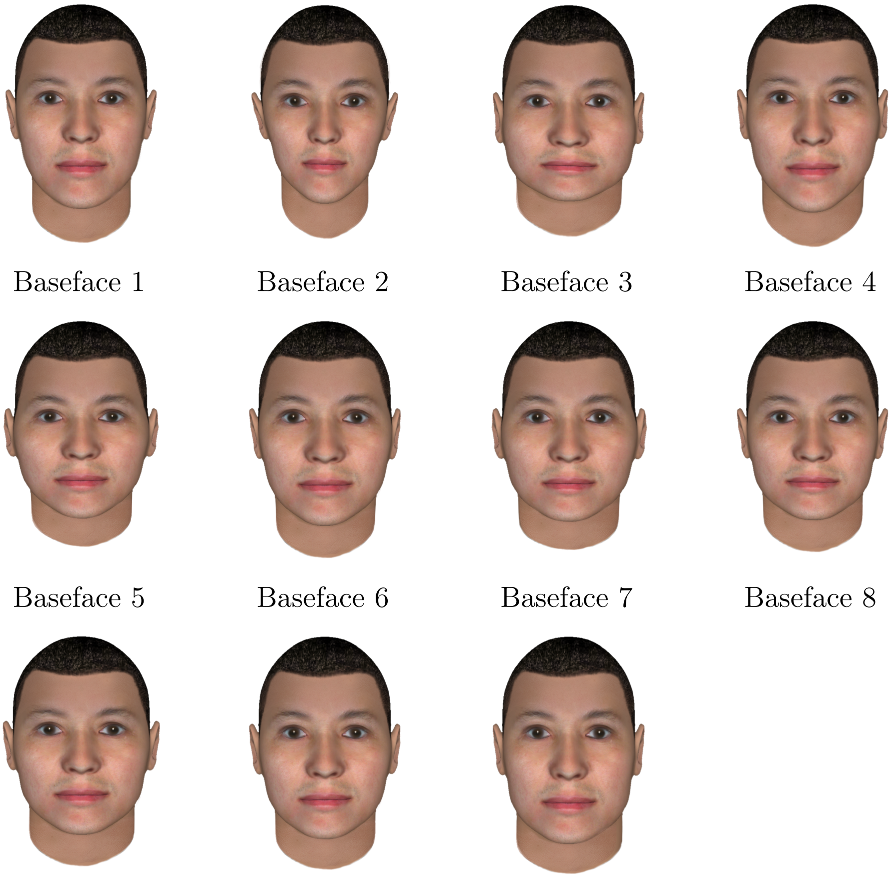


Figure 3.17: Base face masks for face generation.

Face definition

In order to create a face which expresses a certain profile of social traits, the first step was to find the combination of facial features which better expressed the desired social traits. This was performed by means of a GA which searched the best combination of facial features representing the sought social traits.

Although the simplest way to find this combination of facial features would be to compute every possible facial feature combination and compare the generated profile with the sought one, this is a difficult and time-consuming task with an enormous search space, and therefore unpractical. For example, to create a White face, it is necessary to choose between 10 eyebrows, 19 eyes, 12



Baseface 1 Baseface 2 Baseface 3 Baseface 4

Baseface 5 Baseface 6 Baseface 7 Baseface 8

Baseface 9 Basefacel 10 Baseface 11

Table 3.2: Base faces for White ethnicity.

noses, 9 mouths, 11 jawlines, and 161 051 combinations of distances[2], which gives a total of 36 352 431 720 possible combinations, that is, more than 35 billions. This means that with a computer able to check 1 000 combinations per second, it would take 421 days to check every possible combination, or in other words, more than a year.

Therefore, the face definition function requires the use of a *meta-heuristic* search algorithm to reduce the time needed to find a solution. For this reason,

[2] 11 intervals of 5 different distances give $11^5 = 161\,051$ possible combinations.

a Genetic Algorithm was employed in order to find the combination of facial features that better fitted the sought social trait profile. The GA employed to perform this optimization was configured to perform single-point crossover and uniform mutation with a probability of $P_{crossover} = 0.8$ and $P_{mutation} = 0.65$ respectively on a population of 100 individuals. The codification of the chromosome was performed by assigning a gene to each facial feature. Then, the alleles of each chromosome were created independently for each gene as the number of clusters available for each facial feature. Table 3.3 shows the described codification.

Table 3.3: Codification used in the GA for face definition. DEB, DE, DN, DM and DEE stand for the distances from the lowermost point of the jawline to the eyebrow, to the eye, to the nose and to the mouth and the distance between eyes, respectively.

Eyebrow	**Eye**	**Nose**	**Mouth**	**Jawline**	**DEB**	**DE**	**DN**	**DM**	**DEE**

The uniform mutation was implemented in two steps. First, a coin was flipped to decide if the chromosome had to be muted or not. If so, one gene was randomly selected and its value randomly chosen from the alleles. The number of iterations was established at 1 000. The selection method employed was Stochastic Universal Sampling, and the Survivor Selection Policy was fitness-based with *elitism*. Finally, the fitness function was defined as the sum of the absolute values of the differences between each impression prediction and its desired value. Table 3.4 shows an example of this process.

Table 3.4: Example of the fitness computation for an iteration of the Genetic Algorithm for face definition. Only 10 individuals are shown for clarity reasons. The list of impressions is: Afraid, Angry, Attractive, Babyface, Disgusted, Dominant, Feminine, Happy, Masculine, Prototipical, Sad, Surprised, Threatening, Trustworthy and Unusual. FV refers to the fitness function value, and the star denotes the desired social trait profile. The best solution present in this iteration is marked in blue.

#	**Afr.**	**Ang.**	**Att.**	**Bab.**	**Dis.**	**Dom.**	**Fem.**	**Hap.**	**Mas.**	**Pro.**	**Sad**	**Sur.**	**Thr.**	**Tru.**	**Unu.**	**FV**
1	2.55	2.25	2.65	3.59	2.18	1.86	2.10	1.97	3.58	3.86	3.49	1.70	1.83	3.22	2.32	26.47
2	2.04	3.57	2.51	1.79	3.03	3.86	1.70	2.01	4.88	1.87	2.78	1.79	3.53	2.98	3.31	27.27
3	1.99	2.45	3.68	2.88	2.18	3.36	1.92	2.54	4.89	4.19	2.67	1.89	2.57	3.58	2.64	20.55
4	1.80	1.84	4.66	2.47	1.74	3.15	1.80	2.63	4.85	4.29	2.43	1.47	1.76	3.56	1.93	16.74
5	1.97	2.44	3.51	3.05	2.02	3.48	2.58	2.75	4.48	3.86	2.44	1.76	2.48	3.43	2.93	21.64
6	2.12	2.78	4.08	2.91	2.38	3.35	2.03	2.26	4.73	3.49	2.39	1.86	2.46	3.51	2.23	21.02
7	2.53	3.61	2.28	2.38	2.98	2.80	1.89	1.49	3.79	3.68	3.86	1.78	3.50	2.46	2.75	30.22
8	2.05	2.75	3.17	2.51	2.56	2.38	2.11	1.92	4.25	3.57	3.28	1.69	2.44	3.06	2.35	25.01
9	2.48	2.84	2.84	2.50	2.40	2.48	1.67	1.92	4.55	3.96	3.16	1.79	2.51	3.36	1.89	23.55
10	2.49	2.72	2.76	2.03	2.51	2.64	1.80	2.12	4.33	3.58	3.48	1.87	2.55	3.21	1.97	24.04
*	1	1	5	2	1	6	1	7	5	4	1	2	1	5	1	

This procedure ends when the GA finds the best combination of facial features which convey a sought profile of social traits. Let's consider that iteration shown in Table 3.4 is the last one. Then, the solution obtained for the desired social trait profile is the one with $FV = 16.74$. Then, chromosome number 4 is used to build the face. Table 3.5 shows chromosome number 4. These values are the input to the face generation stage.

Table 3.5: Example of a face definition. DEB, DE, DN, DM and DEE stand for the distances from the lowermost point of the jawline to the eyebrow, to the eye, to the nose and to the mouth and the distance between eyes, respectively.

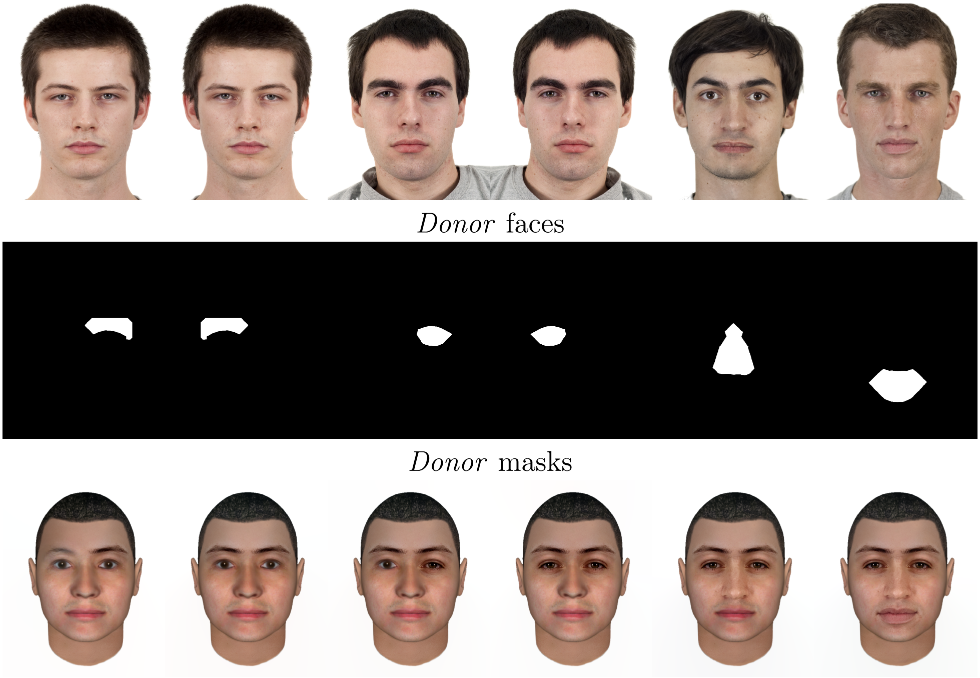
Eyebrow	**Eye**	**Nose**	**Mouth**	**Jawline**	**DEB**	**DE**	**DN**	**DM**	**DEE**
1	3	4	1	8	9	1	5	1	4

Facial surgery

In order to achieve a realistic face it is important to use a seamless fusion method, which further adapts the illumination and tone of the different patches being sewed. The algorithm used in this work to achieve this task is the one described by Pérez et al. (2003), that is, the Poisson Image Editing method. As explained in subsection 3.2.2, this algorithm makes use of the Poisson Equation and information of the gradient of the images in order to achieve a seamless fusion.

The first step is to look for the base face corresponding to the jawline indicated by the definition given by the GA. With the base face selected, the next stage requires the *donor* features, that is, the features to be pasted onto the *patient* (base) face in order to configure the new face.

On the other hand, the distances given by the GA are also taken into account and internal features are pasted into the *patient* face in the positions indicated by these distances. Following with the example face definitions showed above, Figure 3.18 shows the process of the *facial surgery*, in which facial features are pasted one at a time starting with the left eyebrow, and finalizing with the mouth. The first row of the figure shows the *donor*'s face, which appears twice in the case of eyebrows and eyes due to the symmetry assumed. The second row shows the masks indicating the region of the *donor* face to be extracted and pasted into the *patient* face. Finally, the third row shows the *patient* face after each feature insertion.



Donor faces

Donor masks

Patient face after surgeries.

Figure 3.18: Facial surgery progress.

3.4 Results

3.4.1 The social trait model

The implemented evaluation function, by means of the optimized weights, is able to predict a set of social trait scores for any combination of facial features. In this section, the fitting achieved for the existent CFD faces is shown as a measure of what social traits does the model predict better. Table 3.6 shows these results for White ethnicity without and with dispersion. As was explained in subsection 3.3.2, the global face scores (GFS) obtained with the evaluating function need to be transformed (i.e. add dispersion removed when doing the cluster mean scores) in order to better reproduce the variation present in CFD scores. Correlations are the same in both cases due to the linearity of the operations performed to add the lost dispersion to the GFS. Three metrics are shown, the correlation (with its corresponding p-value), the coefficient of determination (r^2), and the mean squared error (MSE).

Table 3.6: Results of the implemented social traits model for each social impression.

	Correlation	**p-value**	**r^2**	**MSE**	
				Without dispersion	**With dispersion**
Afraid	0.7018	3.28e-15	0.4925	0.1088	0.1013
Angry	0.7274	1.01e-16	0.5292	0.2745	0.1872
Attractive	0.7661	2.35e-19	0.5869	0.2540	0.1923
Babyface	0.8124	2.86e-23	0.6600	0.3486	0.1661
Disgusted	0.7008	3.71e-15	0.4912	0.1303	0.0841
Dominant	0.7429	1.01e-17	0.5519	0.3461	0.2480
Feminine	0.8086	6.56e-23	0.6538	0.1014	0.0528
Happy	0.7551	1.47e-18	0.5702	0.1966	0.1222
Masculine	0.7927	1.76e-21	0.6283	0.2023	0.1051
Prototypical	0.8208	4.26e-24	0.6737	0.4139	0.7067
Sad	0.7273	1.02e-16	0.5290	0.2486	0.2183
Surprised	0.7755	4.43e-20	0.6015	0.0403	0.0220
Threatening	0.7559	1.30e-18	0.5713	0.2684	0.1730
Trustworthy	0.7492	3.83e-18	0.5612	0.0901	0.0633
Unusual	0.7536	1.87e-18	0.5680	0.2260	0.1896
Mean	0.7593	4.82e-16	0.5779	0.2167	0.1755

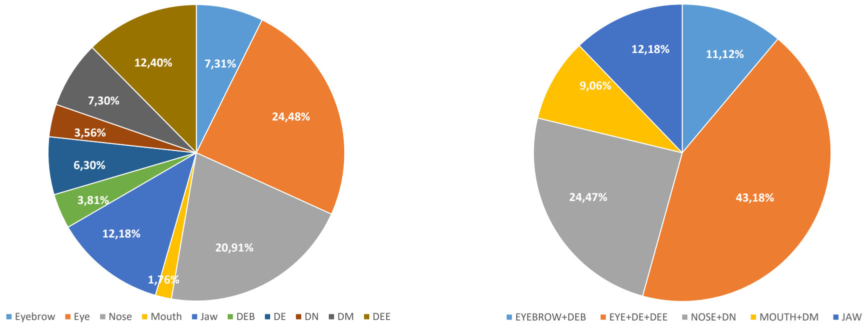
3.4.2 How facial features affect social impressions

In the following lines, the results for each ethnicity are presented. In order to understand the results, the term *facial feature group* needs to be defined. A *facial feature group* is a group of the internal facial feature with its corresponding distance(s). For example, *eyebrow group* includes the eyebrow and the distance from the jawline to the eyebrow (d_{eb}). Similarly, *eye group* accumulates the importance of the eye, the distance from the jawline to the eyes (d_e) and the distance between eyes (d_{ee}). The rest of the groups are formed in the same manner.

Two kind of graphs are shown in the results: a pie chart and a stacked bars chart. The former shows the mean of each facial feature weight and the mean of each *facial feature group* on each impression. The latter shows the same data broken down into each impression.

Asian

In general, eyes and nose are the most important facial features in Asian ethnicity regarding impression formation. In fact, just eyes and nose account for 45.39% of the importance (Figure 3.19 (a)), and these numbers increase to 67.65% if their groups are considered (Figure 3.19 (b)). It is important to note the importance of the distance between eyes, with 12.40%. Jaw (12.18%),



(a) (b)

Figure 3.19: Overall importance of each facial feature (a) and facial feature group (b) in impression formation for Asian ethnicity.

eyebrows (11.12%) and mouth (9.06%) complete the distribution of weights of facial feature groups for facial impression formation for Asian ethnicity.

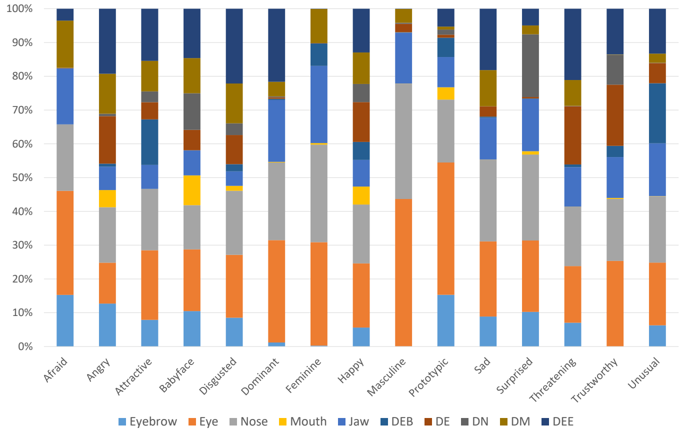


Figure 3.20: Importance of each facial feature in the formation of each impression for Asian ethnicity.

This is the mean trend, however, each impression is different. For example, considering eyes alone, they are very important for Afraid, Masculine and Prototypical, whilst they are less for Angry, Threatening or Unusual (Figure 3.20). However, if the eye group is considered, it is very important for Angry and Threatening (Figure 3.21). This is due to the importance of the eye distances

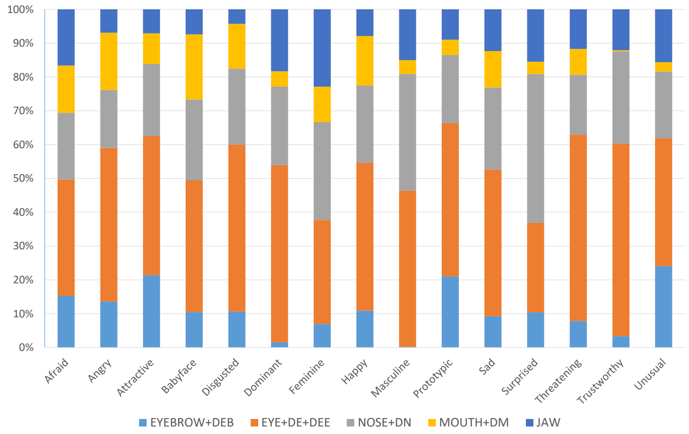


Figure 3.21: Importance of each facial feature group in the formation of each impression for Asian ethnicity.

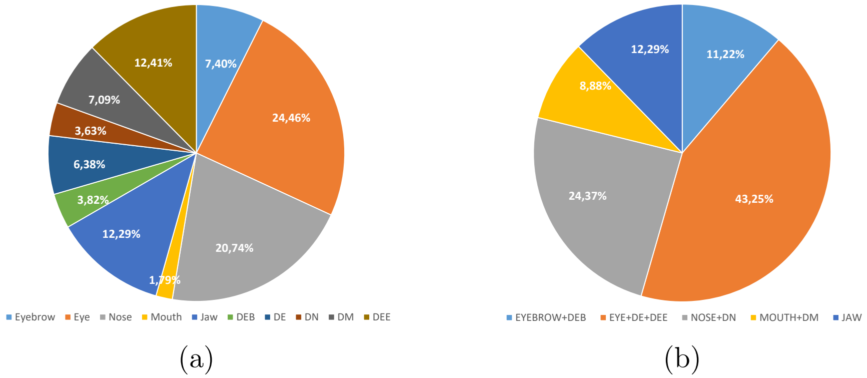
(d_e and d_{ee}) for these impressions. Therefore, it is evident that not only the facial features themselves are important, but also their positions are.

On the other hand, noses keep a similar importance among all impressions, whilst mouths show very few importance in general. Only in Angry, Babyface and Happy mouths have a notable contribution. Nevertheless, if distances are included, nose importance increases significantly for Surprised and mouths achieve a considerable amount of importance in most impressions. Jaw and hair do not have any associated distance, thus, they do not vary between features alone and feature groups. Jaw reaches its highest importance for Afraid, Dominant, Feminine and Surprised.

Finally, it is also important to note how eyebrows alone are not very important in facial trait impression formation, but the eyebrows distance to the jawline (d_{eb}) is.

Black

As for Asian, for Black ethnicity eyes and nose are the most important facial features as well regarding impression formation. In fact, just eyes and nose account for 45.20% of the importance (Figure 3.22 (a)), and these numbers increase to 67.62% if their groups are considered (Figure 3.22 (b)). It is important to note the importance of the distance between eyes, with 12.41%. Jaw (12.29%), eyebrows (11.22%), and mouth (8.88%) complete the distribu-



(a) (b)

Figure 3.22: Overall importance of each facial feature (a) and facial feature group (b) in impression formation for Black ethnicity.

tion of weights of facial feature groups for facial impression formation for Black ethnicity.

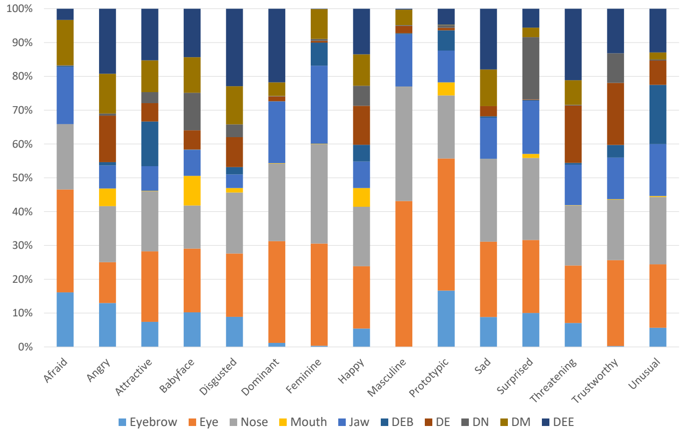


Figure 3.23: Importance of each facial feature in the formation of each impression for Black ethnicity.

Again, this is the mean trend, but there exist differences among impressions. For example, for this ethnicity, eyes alone are very important for Afraid, Dominant, Masculine and Prototypical; and less for Angry and Unusual (Figure 3.23). However, if distances are included, eyes account for more than 50%

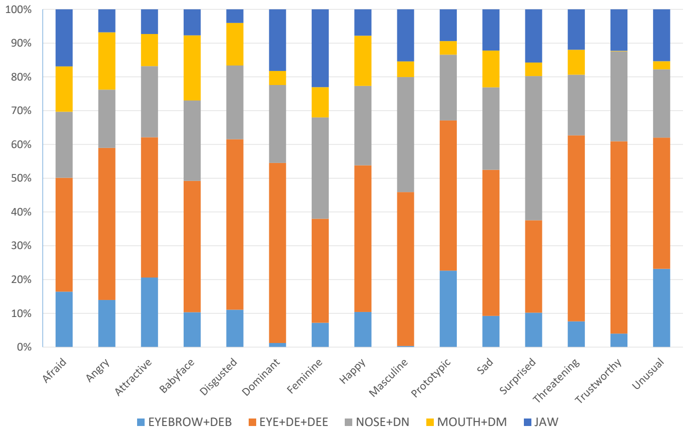


Figure 3.24: Importance of each facial feature group in the formation of each impression for Black ethnicity.

of the importance for Disgusted, Dominant, Threatening and Unusual (Figure 3.24).

In this case, for Black ethnicity, and considering distances, eyebrows account for one fifth of the total importance for Attractive and Prototypical, and less than 5% for Dominant and Trustworthy; noses play an important role in Surprised, Trustworthy and Masculine, with more than 25% of importance; whilst jaw and mouth do not achieve more than 20% of importance for any impression.

Finally, as for Asian ethnicity, it is also important to note how eyebrows distance to the jawline (d_{eb}) has an important weight in most impression formation.

Latino

Similarly to Asian and Black, eyes and nose are the most important facial features as well regarding impression formation for Latino ethnicity. In fact, just eyes and nose account for 45.31% of the importance (Figure 3.25 (a)), and these numbers increase to 67.46% if their groups are considered (Figure 3.25 (b)). It is important to note the importance of the distance between eyes, with 12.36%. Jaw (12.24%), eyebrows (11.46%), and mouth (8.83%) complete the distribution of weights of facial feature groups for facial impression formation for Latino ethnicity.

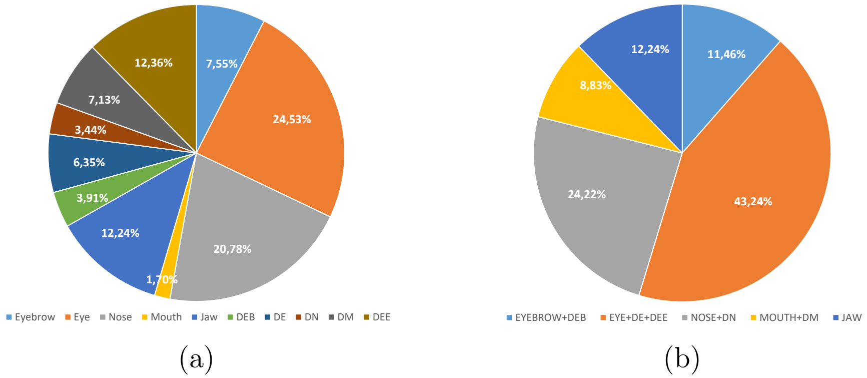


Figure 3.25: Overall importance of each facial feature (a) and facial feature group (b) in impression formation for Latino ethnicity.

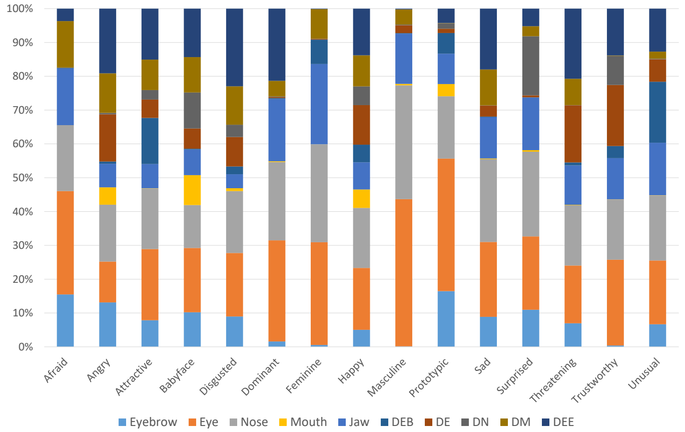


Figure 3.26: Importance of each facial feature in the formation of each impression for Latino ethnicity.

For Latino ethnicity, eyes achieve their maximum importance in the definition of Afraid, Dominant, Masculine and Prototypical; and the minimum for Angry and Unusual (Figure 3.26). However, when distances are also considered, the eye group achieves its maximum importance for Disgusted, Dominant, Threatening and Trustworthy (Figure 3.27). This implies that eye distances are very important for Disgusted, Threatening and Trustworthy.

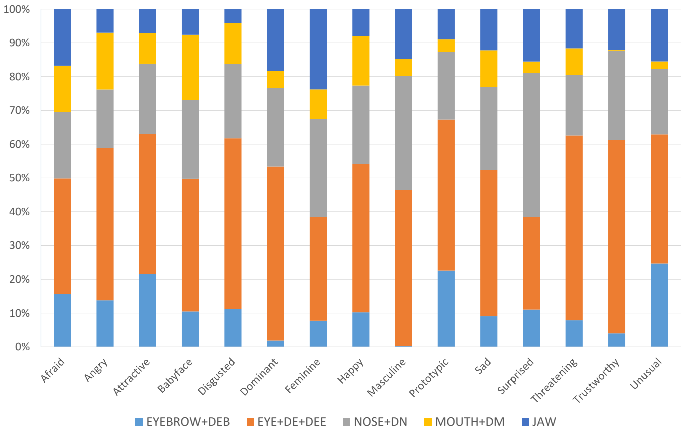


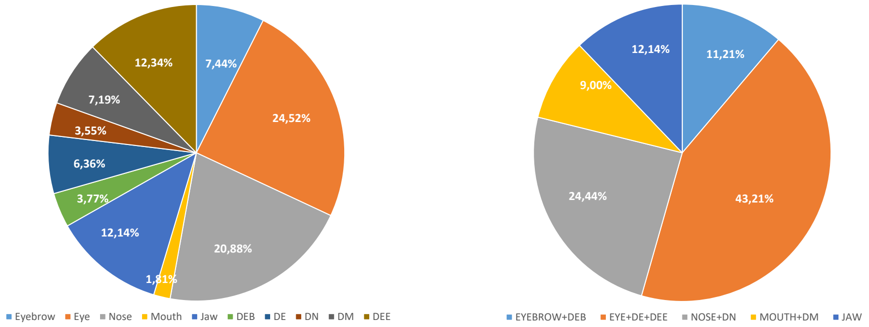
Figure 3.27: Importance of each facial feature group in the formation of each impression for Latino ethnicity.

Noses, on the other hand, have a more regular influence in the definition of the different impressions, without very large changes among them. Furthermore, the nose distance is very important for Surprised. Mouths alone do not obtain more than 10% of importance in any of the studied impressions, but they do if their distance is also considered for Afraid, Angry, Babyface, Disgusted and Happy. This is due to the importance of the mouth distance (d_m) for these impressions. Jaw reaches its highest importance for Afraid, Dominant, Feminine and Surprised.

Finally, eyebrow group obtains a high importance for the definition of Attractive, Prototypical and Unusual; whilst eyebrows alone do so for Afraid and Prototypical.

White

Finally, White ethnicity behaves similarly to all the previous ones. Eyes and nose are the most important facial features as well regarding impression formation, with just eyes and nose accounting for 45.40% of the importance (Figure 3.28 (a)), and 67.65% if their groups are considered (Figure 3.28 (b)). It is important to note the importance of the distance between eyes, with 12.34%. Jaw (12.14%), eyebrows (11.21%), and mouth (9.00%) complete the distribution of weights of facial feature groups for facial impression formation for White ethnicity.



(a) (b)

Figure 3.28: Overall importance of each facial feature (a) and facial feature group (b) in impression formation for White ethnicity.

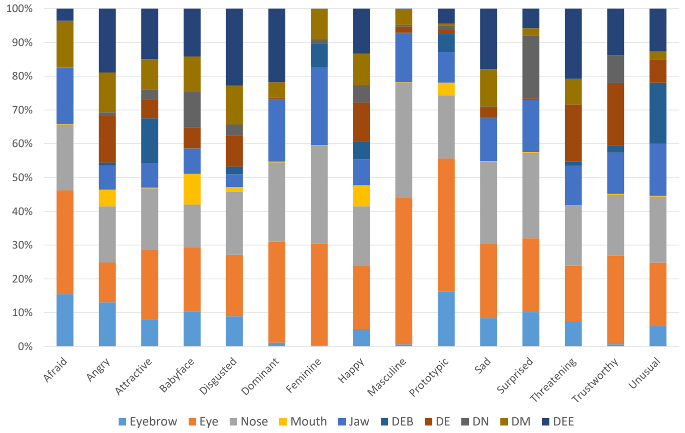


Figure 3.29: Importance of each facial feature in the formation of each impression for White ethnicity.

As for the other ethnicities, White obtains differences among impressions as well. For example, eyes alone are very important for the definition of Afraid, Dominant, Masculine and Prototypical; but very few for Angry (Figure 3.29). Nevertheless, when distances are considered as well, the maximum importance is obtained for Disgusted, Dominant, Threatening and Trustworthy (Figure 3.30). This implies a high importance of the eye distances for Disgusted, Threatening and Trustworthy.

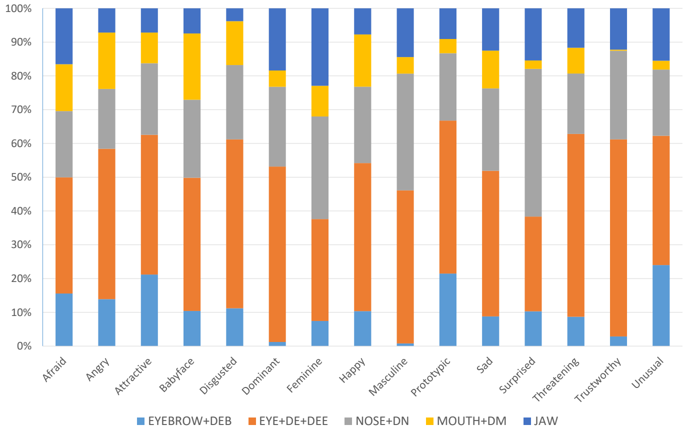


Figure 3.30: Importance of each facial feature group in the formation of each impression for White ethnicity.

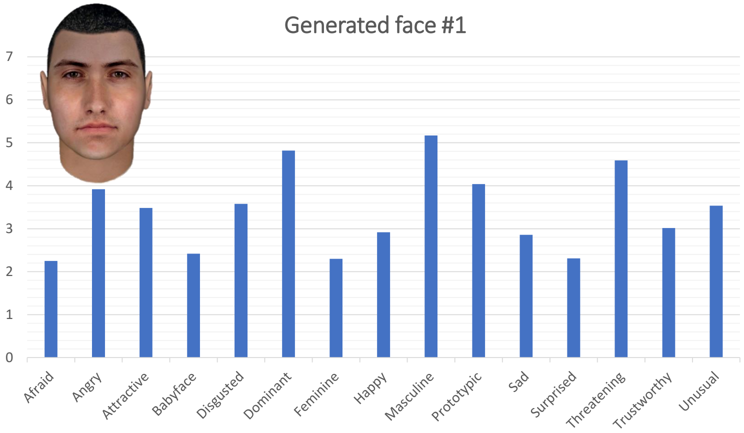
On the other hand, nose importance is higher for Masculine, Sad and Surprised, and lower for Babyface and Unusual. Mouth is very few important for every impression, obtaining its maximum for Babyface, with less than 10%. Jaw reaches its highest importance for Afraid, Dominant, Feminine and Surprised.

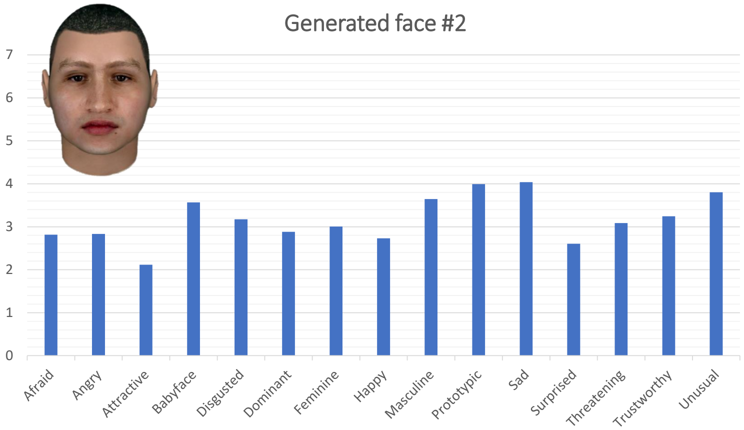
Finally, it is also important to note how distances are very important, obtaining near 50% of importance for several impressions.

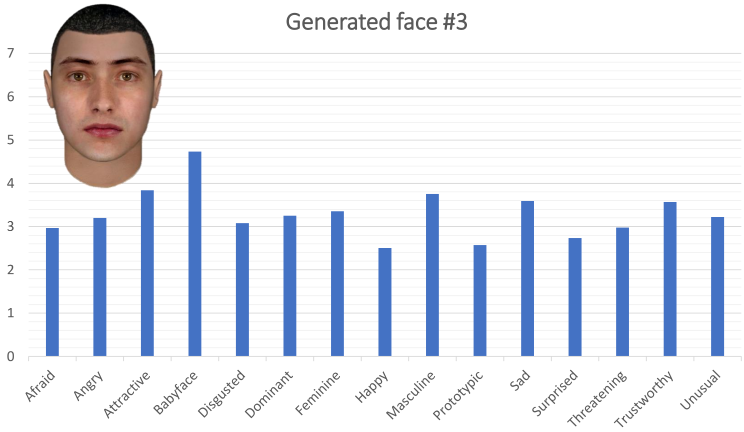
3.4.3 Generation of faces that convey certain social impressions

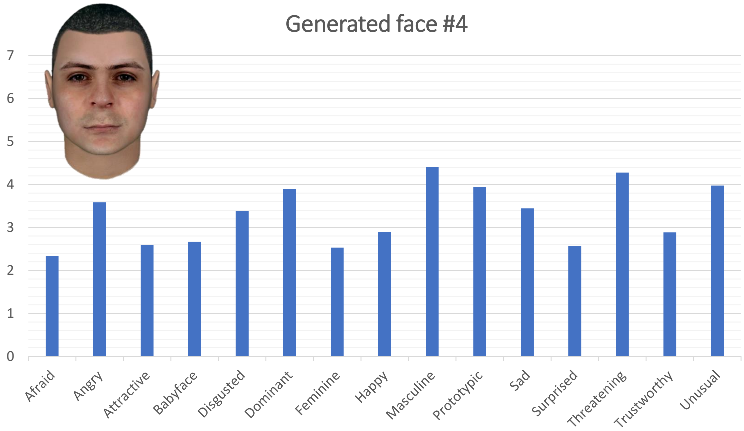
Ten different faces were created and are shown along their respective social trait profiles in Table 3.7. These results were obtained only for White ethnicity due to time limitations, as the procedure required for the creation of the base faces was time consuming and for validation reasons that will be explained later in this chapter.

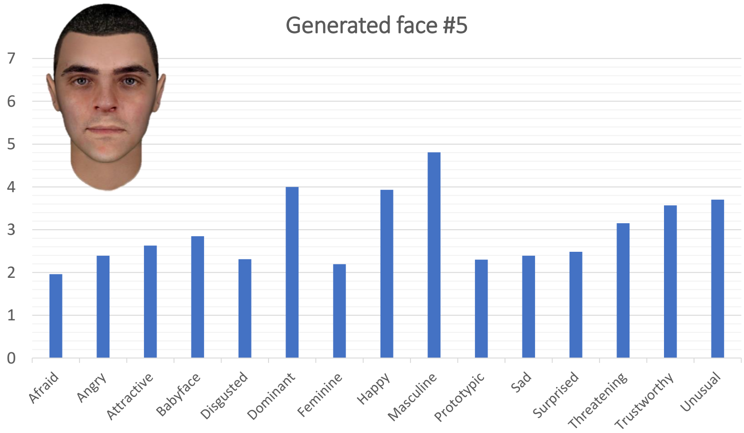
Table 3.7: Generated faces for white ethnicity.

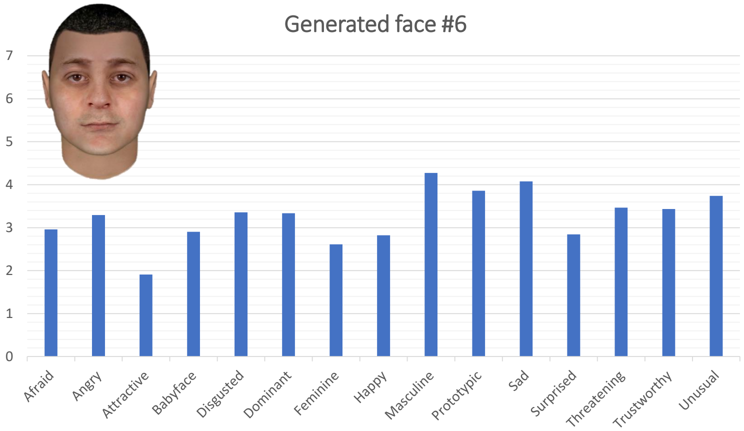


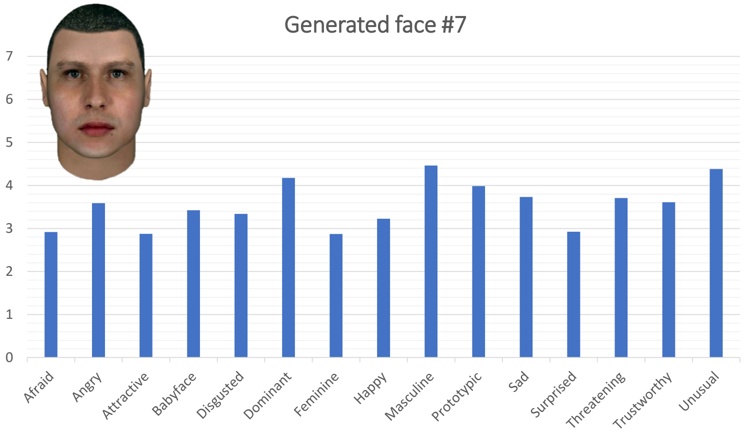


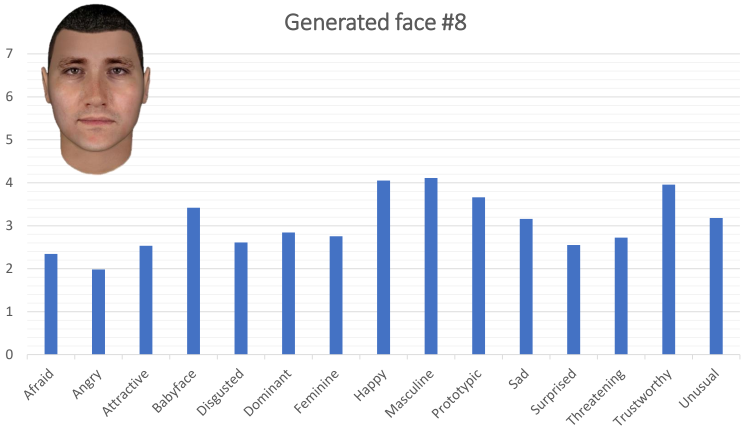

Generated face #3
0
1
2
3
4
5
6
7
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

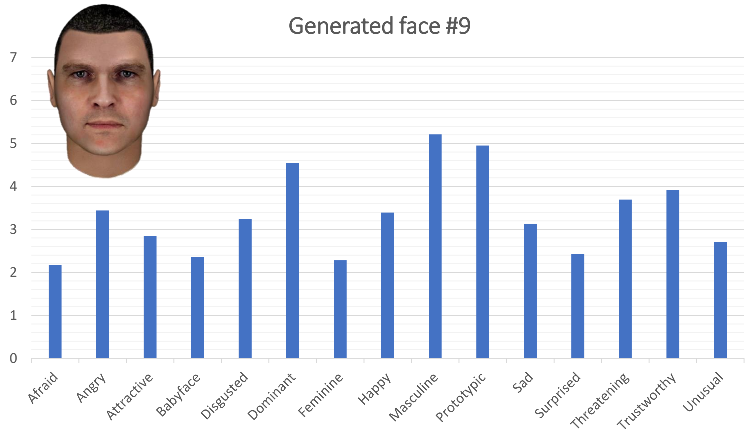

Generated face #4
0
1
2
3
4
5
6
7
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

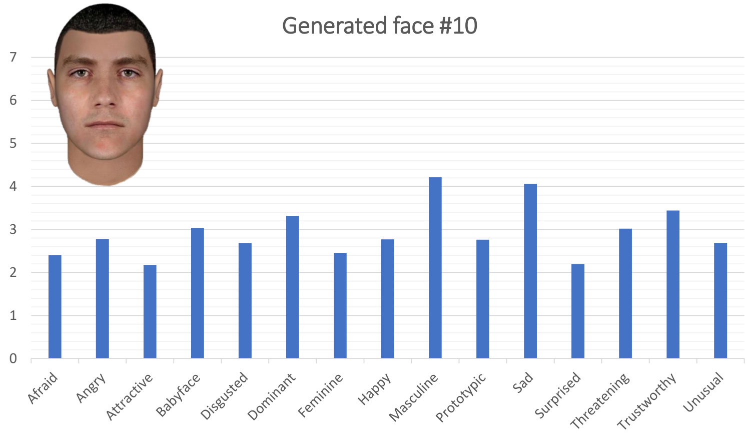

Generated face #5
0
1
2
3
4
5
6
7
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


Generated face #6
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


Generated face #7
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


Generated face #8
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

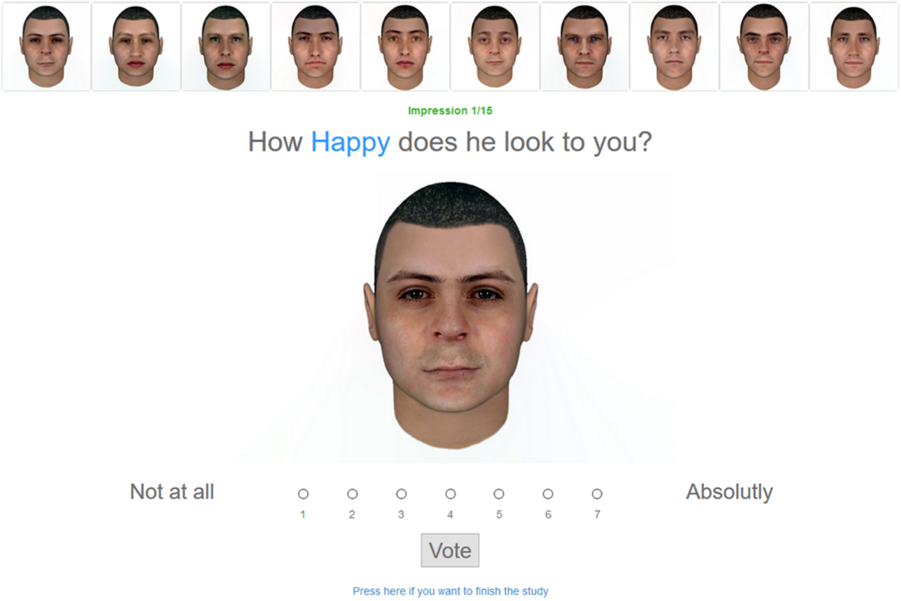




3.5 Validation of the face generator

The described methodology proposes an automatic method to generate faces able to convey certain social traits to the observer. This task has the inherent difficulty of the existent subjectivity and the inter-observer and intra-observer variability produced when assessing social traits (Sutherland, A. W. Young, et al., 2017; Todorov and J. M. Porter, 2014).

Some works employ emotional classifiers in order to evaluate how good their predictions are (Said, Sebe, et al., 2009). By doing this, they remove this inter-observer and intra-observer variability. However, this kind of validation lacks



Impression 1/15

How Happy does he look to you?

Press here if you want to finish the study

Figure 3.31: Platform used to validate the face generation system.

strength, as the goal is to match what people feel when they see the generated faces, and this approach looks for emotions in the face instead of impressions. Moreover, to the best of our knowledge, there is no classifier trained to work with all the impressions employed in this work. Another possible method for validating this work would be to show people two faces and ask them to decide which one is more extreme with respect to each trait (Walker and Vetter, 2009). However, this kind of validation would allow us to assert that our system is able to tell if a face is better or worst at eliciting a certain impression, and thus, to sort the generated faces, but it would not permit us to obtain a profile of scores for a set of social traits for a given face. Therefore, we decided to ask people to assess our created faces using the same scale as the CFD database (1-7 Likert) in order to compare the profile obtained with our system and the profile extracted from people assessments. To do so, a web page was created in which the 10 generated faces where shown once for each impression, and the user had to assess them. To avoid the effect of learning the faces, impression and face order was randomized for each user. Figure 3.31 shows the interface of the implemented tool.

35 people participated in the validation, 16 men and 19 women. The ages of the participants were comprised between 71 and 18 years old, with a mean of 37.

The validation is presented in two ways. First, the ability of the system to elicit the social traits available is shown. Then, the accuracy of the system at creating faces with certain profiles of social traits is described.

3.5.1 Social trait validation

Table 3.8 shows correlation (with its corresponding p-value), the coefficient of determination (r^2), and the MSE of our system working on the CFD faces *versus* the new faces generated by our system. This means that we compared the scores given by our system with those given by people during the validation for the new created faces. It must be noted that while the "CFD faces" related metrics are computed with 93 faces, the "Generated faces" metrics are computed with only 10 faces (the generated ones). In addition, last row of the table shows the mean taking only correlations statistically significant with $p < 0.05$.

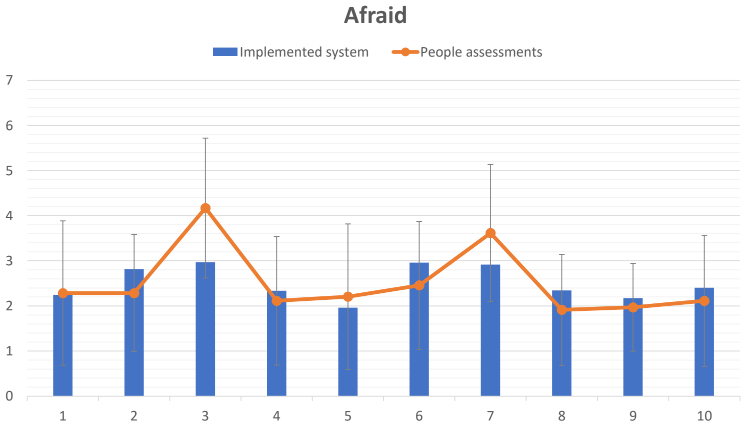
Table 3.8: Comparison of the metrics obtained by the developed model for existent CFD faces and for the generated with the implemented system. In blue, correlations statistically significant ($p < 0.05$).

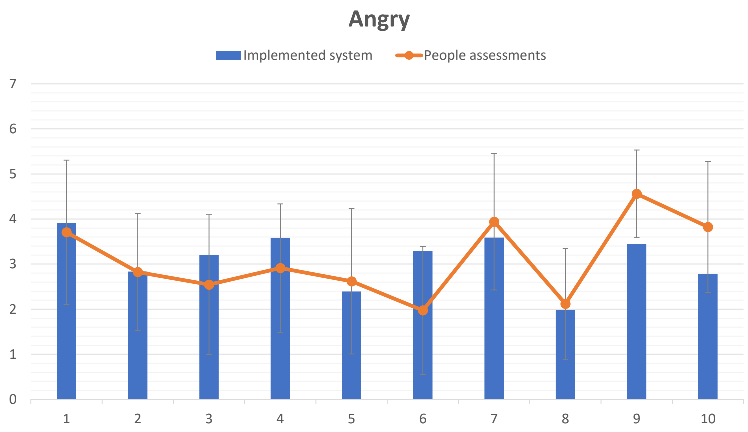
	CFD faces				**Generated faces**			
	Corr.	**p-value**	**r^2**	**MSE**	**Corr.**	**p-value**	**r^2**	**MSE**
Afraid	0.7018	3.29e-15	0.4925	0.1013	0.6922	0.0266	0.4791	0.2894
Angry	0.7274	1.01e-16	0.5292	0.1872	0.5005	0.1407	0.2505	0.5222
Attractive	0.7661	2.35e-19	0.5869	0.1923	0.5131	0.1294	0.2632	0.4603
Babyface	0.8124	2.87e-23	0.6600	0.1661	0.7226	0.0182	0.5222	0.4023
Disgusted	0.7008	3.72e-15	0.4912	0.0841	0.2980	0.4030	0.0888	0.3737
Dominant	0.7429	1.02e-17	0.5519	0.2480	0.7391	0.0146	0.5462	0.5489
Feminine	0.8086	6.57e-23	0.6538	0.0528	0.7556	0.0115	0.5710	0.1476
Happy	0.7551	1.47e-18	0.5702	0.1222	0.3586	0.3088	0.1286	0.6217
Masculine	0.7927	1.76e-21	0.6283	0.1051	0.8320	0.0028	0.6923	0.1440
Prototypical	0.8208	4.27e-24	0.6737	0.7067	0.1689	0.6410	0.0285	0.7503
Sad	0.7273	1.03e-16	0.5290	0.2183	0.6878	0.0279	0.4731	0.1986
Surprised	0.7755	4.44e-20	0.6015	0.0220	0.5767	0.0810	0.3325	0.6020
Threatening	0.7559	1.30e-18	0.5713	0.1730	0.5983	0.0676	0.3580	0.5347
Trustworthy	0.7492	3.83e-18	0.5612	0.0633	0.3142	0.3766	0.0987	0.6062
Unusual	0.7536	1.87e-18	0.5680	0.1896	0.7579	0.0111	0.5743	0.1377
Mean	0.7593	4.82e-16	0.5779	0.1755	0.5677	0.1507	0.3605	0.4226
Mean ($p < .05$)	0.7593	4.82e-16	0.5779	0.1755	0.7410	0.0161	0.5512	0.2669

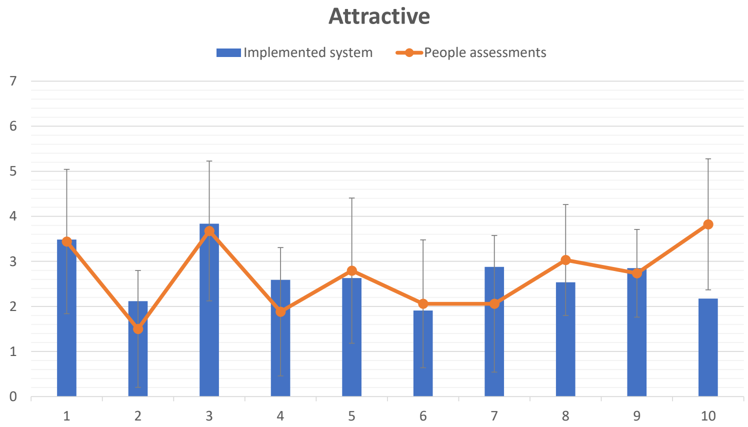
As is shown in Table 3.8, our system obtains high statistically significant ($p < 0.05$) correlations with low error ($MSE < 0.3$, except for Babyface and Dominant) for 7 of the social traits studied, namely Afraid, Babyface, Dominant, Feminine, Masculine, Sad and Unusual. Furthermore, according to the coefficient of determination, our model is able to explain at least 47.31% of the mentioned impressions, reaching a 69.23% for Masculine. The last row of the table shows the mean of these impression metrics. Moreover, we also obtained good correlation for Surprised and Threatening, with $p < 0.1$, although the error in this case is higher ($MSE > 0.5$) and r^2 is lower ($r < 0.4$). On the other hand, impressions as Angry and Attractive obtain a correlation near 50% with $p < 0.2$ and acceptable errors. Finally, our model does not work well with Disgusted, Happy, Prototypical and Trustworthy.

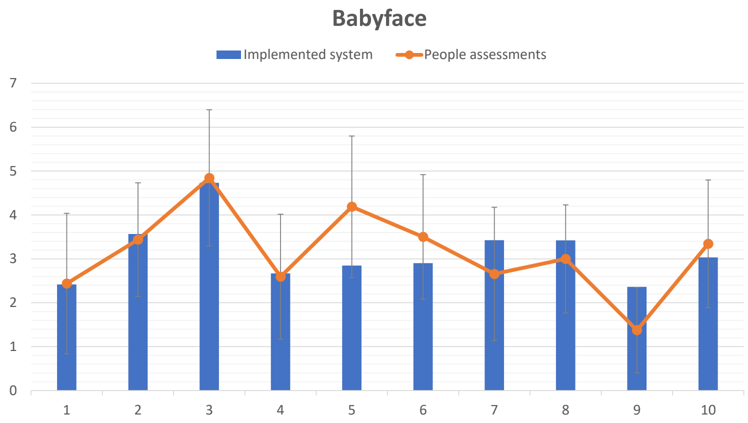
Table 3.9 shows how well the developed model is able to shape each impression.

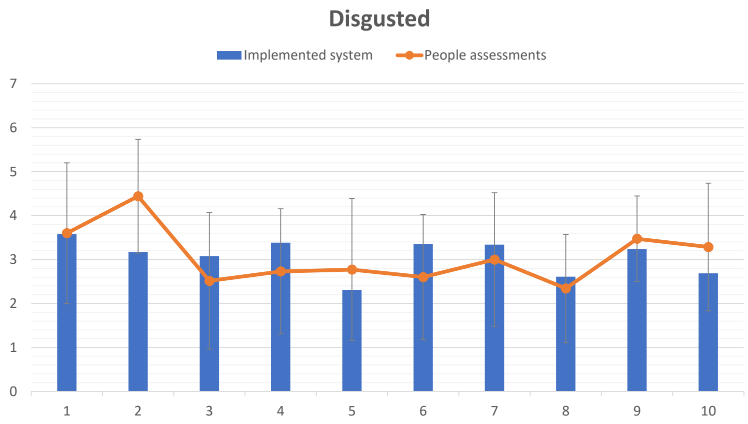
Table 3.9: Validation results of the implemented social traits generation model. In blue, score of our model for each of the faces generated. In orange, the mean of the people assessments with one standard deviation ($\pm\sigma$) in gray.

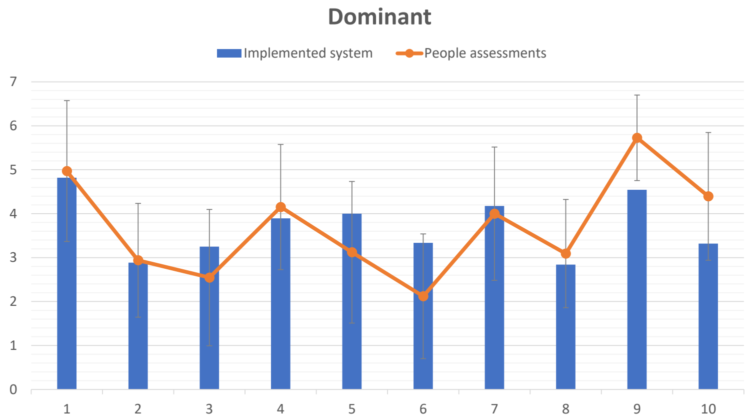


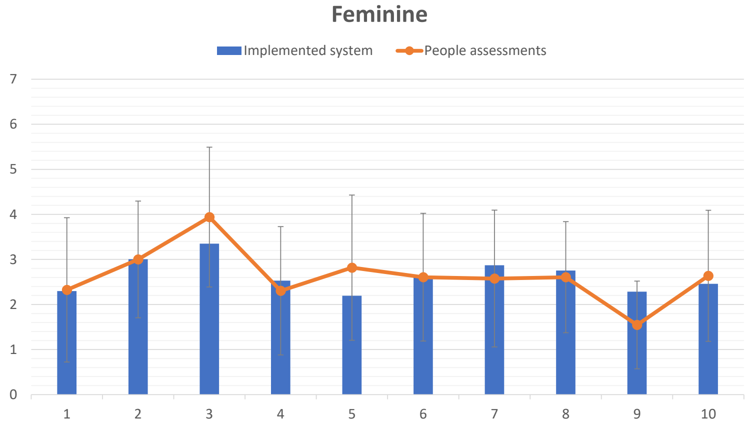

Angry
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

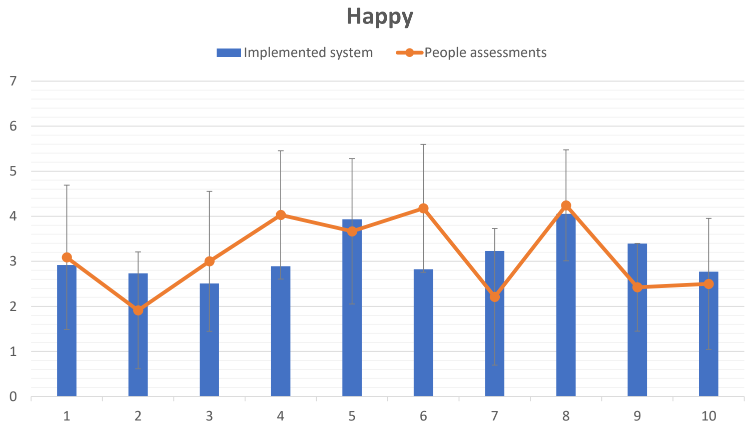

Attractive
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

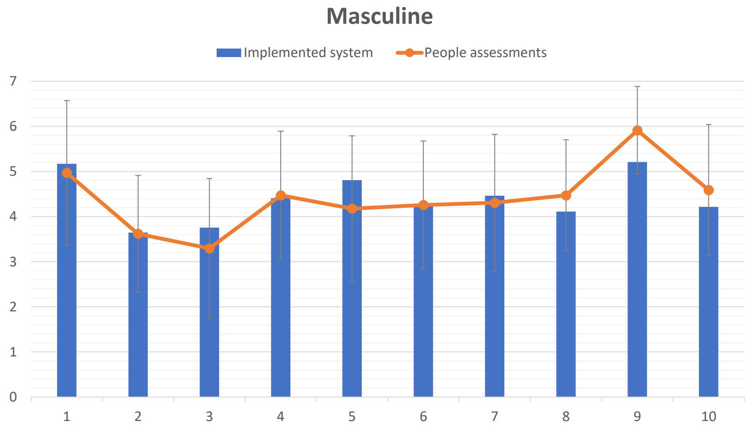

Babyface
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

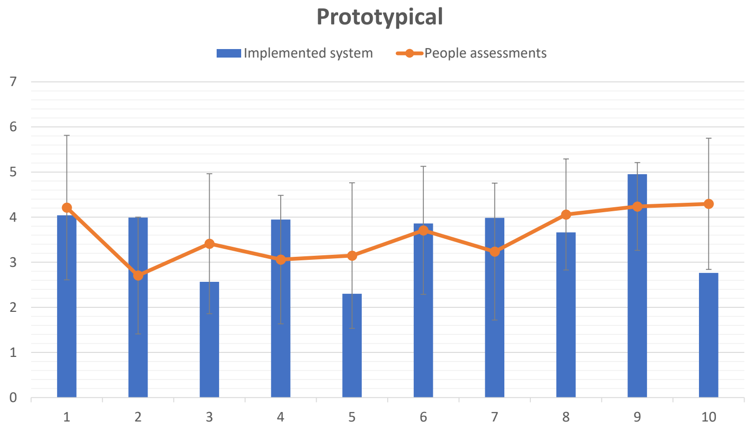

Disgusted
Implemented system
People assessments
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7
8
9
10

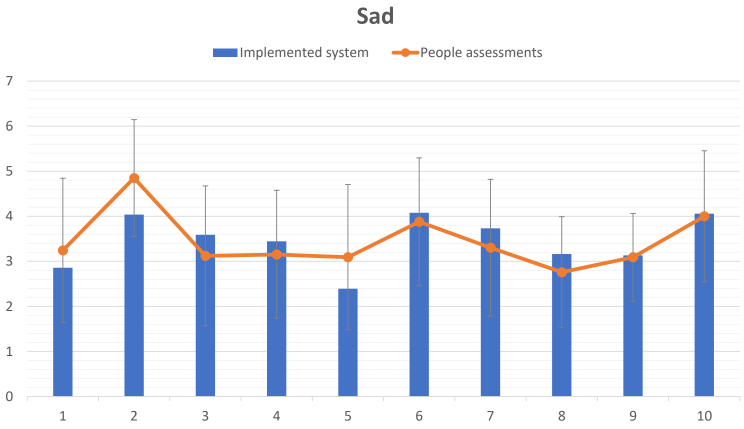

Dominant
Implemented system
People assessments
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7
8
9
10

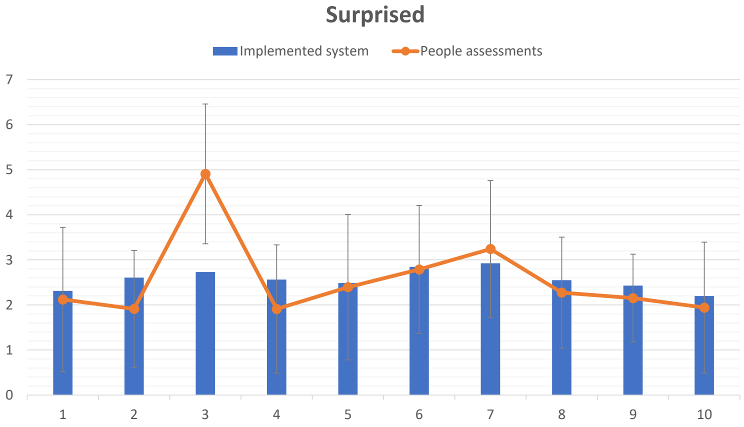

Feminine
Implemented system
People assessments
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7
8
9
10

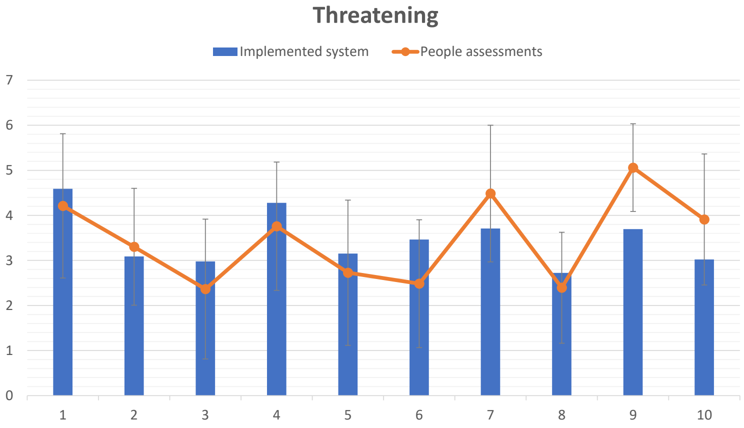

Happy
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

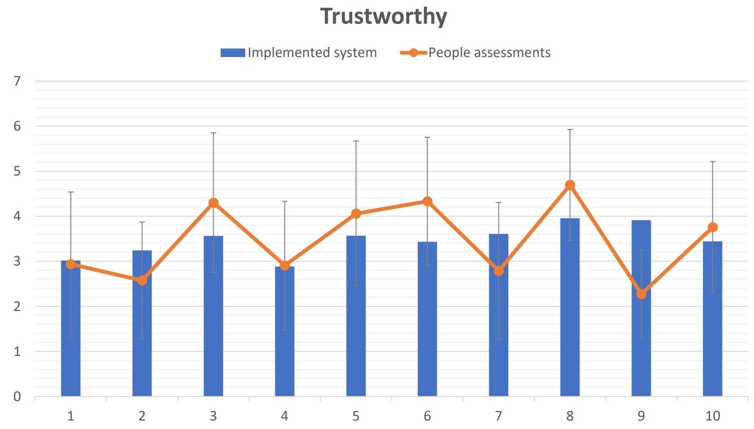

Masculine
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

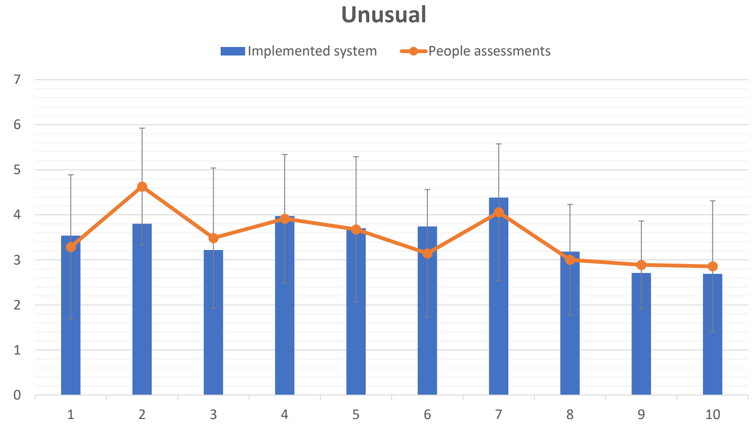

Prototypical
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10


Sad
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10


Surprised
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10


Threatening
Implemented system
People assessments
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10





3.5.2 Face validation

In this section, the accuracy at creating faces with a desired social trait profile is shown. Table 3.10 shows the correlation, the MSE, and the coefficient of determination (r^2) of the profiles obtained by our system *versus* the people's opinion for each generated face. In addition, last row of the table shows the mean taking only correlations statistically significant with $p < 0.05$.

As is shown in Table 3.10, our model is able to create faces with a desired profile of trait impressions with a correlation $r > 0.65$ with $p < 0.05$ for 8 out of the 10 generated faces. Furthermore, for two of the generated faces, the correlation obtained is higher than 0.9, explaining more than 85% of the

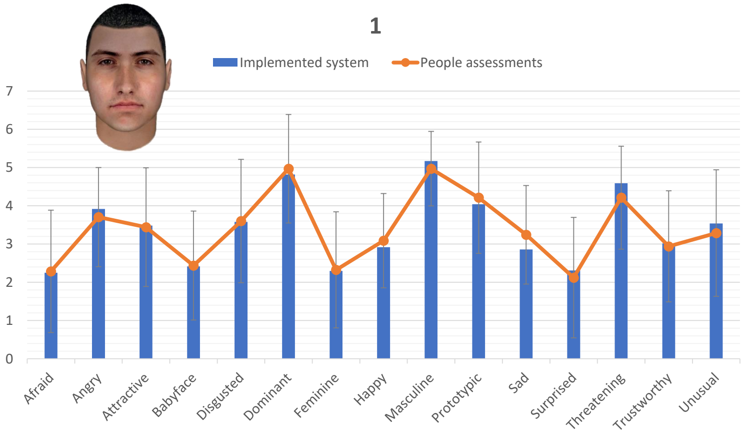
Table 3.10: Metrics of the validation of the implemented system for faces.

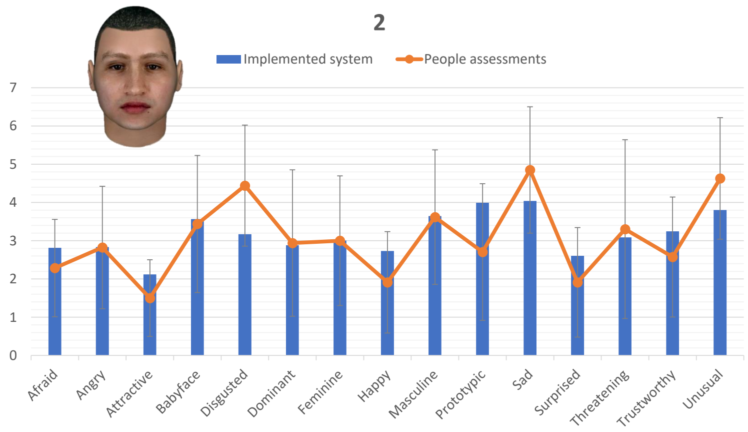
	Correlation	p-value	r^2	MSE
1	0.9778	1.03e-06	0.9561	0.0378
2	0.7535	0.0118	0.5678	0.4620
3	0.3373	0.3405	0.1137	0.6799
4	0.7975	0.0057	0.6360	0.2950
5	0.7104	0.0213	0.5047	0.3609
6	0.4971	0.1437	0.2471	0.5637
7	0.6670	0.0351	0.4449	0.3598
8	0.9229	0.0001	0.8517	0.1356
9	0.8206	0.0036	0.6735	0.7261
10	0.6597	0.0379	0.4352	0.6054
Mean	0.7144	0.0600	0.5431	0.4226
Mean $(p < .05)$	0.7887	0.0145	0.6337	0.3728

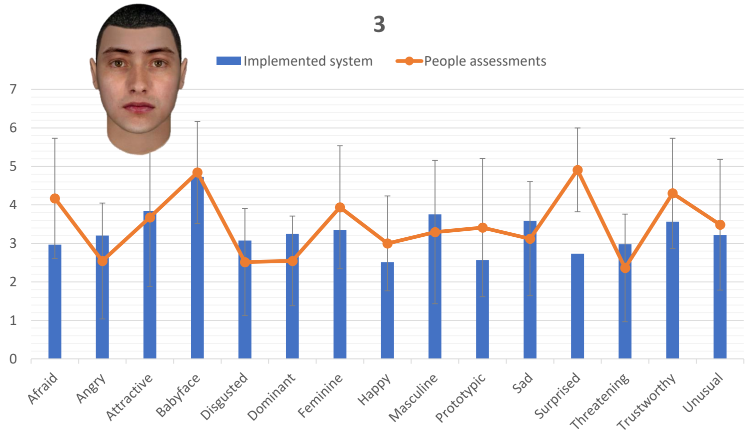
variation present in the social trait profile, with a MSE lower than 0.15. If only correlations statistically significant with $p < 0.05$ are taken, the mean correlation achieved is 0.7887, with a MSE of 0.3728 and a p-value of 0.0145.

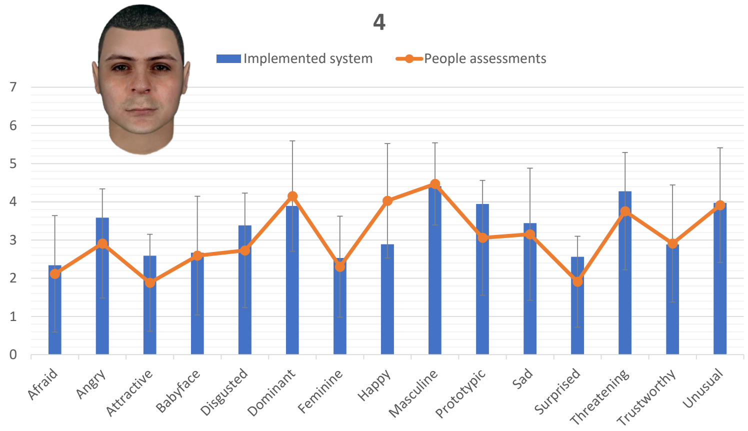
Table 3.11 shows the generated faces along with their social trait profiles. The best results are obtained by faces 1 and 8, with 4 and 9 being very good as well; whilst faces 3 and 6 obtained the worst results.

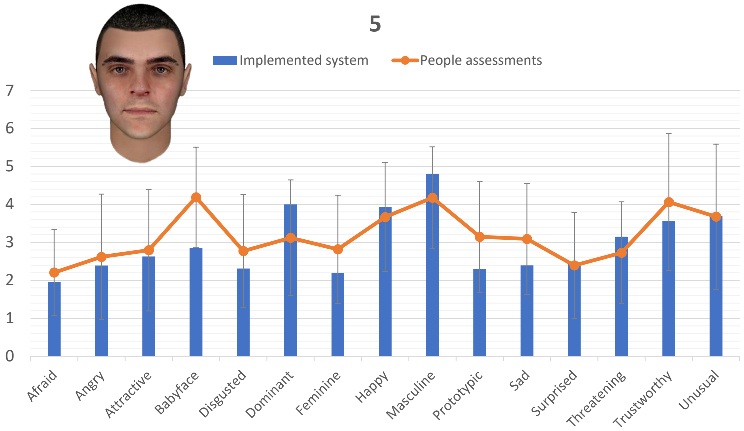
Table 3.11: Validation results of the generated faces. In blue, the social trait profile obtained by our model. In orange, the mean of the people assessments. Finally, it is shown one standard deviation of people assessments ($\pm\sigma$).

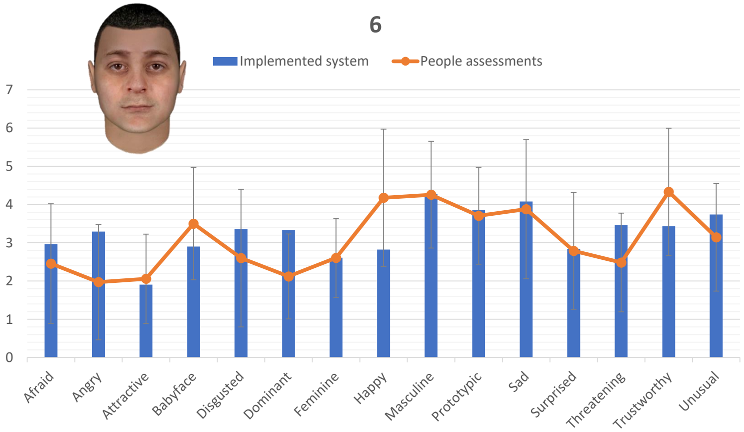


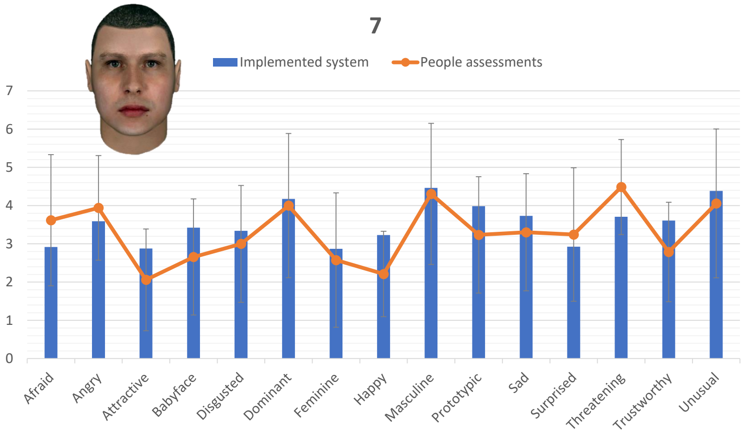

2
Implemented system
People assessments
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

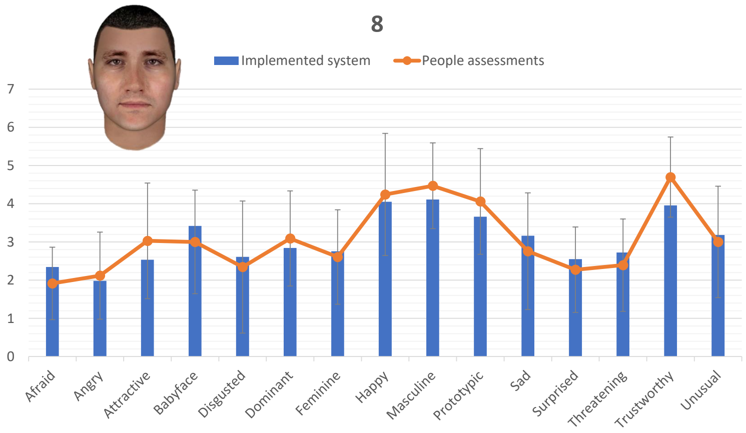

3
Implemented system
People assessments
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

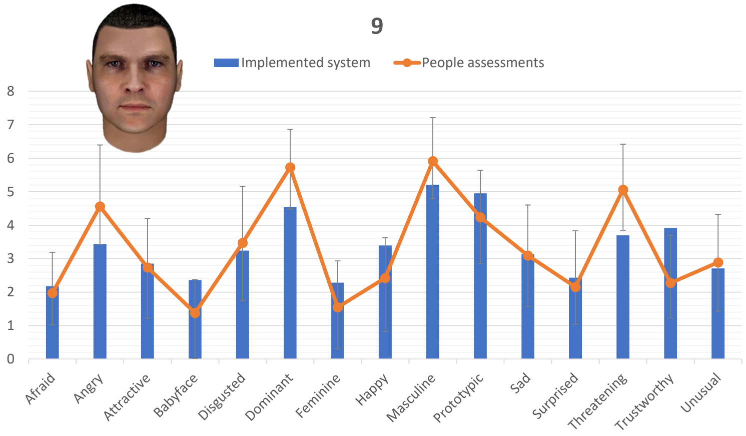

4
Implemented system
People assessments
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

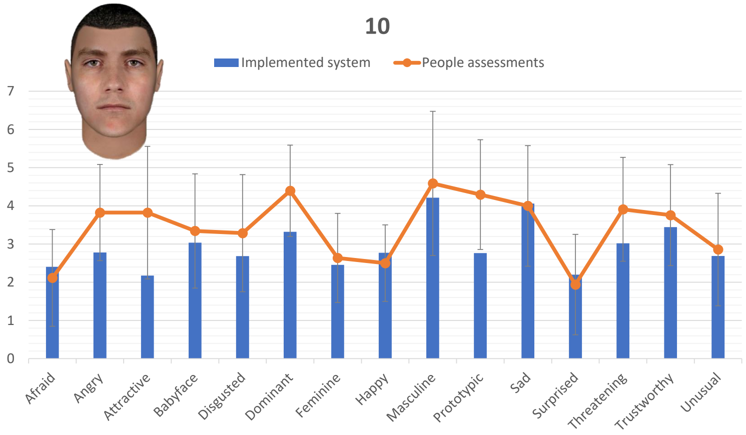

5
Implemented system
People assessments
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


6
Implemented system
People assessments
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


7
Implemented system
People assessments
7
6
5
4
3
2
1
0
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


8
Implemented system
People assessments
0
1
2
3
4
5
6
7
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


9
Implemented system
People assessments
0
1
2
3
4
5
6
7
8
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual


10
Implemented system
People assessments
0
1
2
3
4
5
6
7
Afraid
Angry
Attractive
Babyface
Disgusted
Dominant
Feminine
Happy
Masculine
Prototypic
Sad
Surprised
Threatening
Trustworthy
Unusual

3.6 Conclusions

In this chapter, a new approach to model social traits in faces has been presented. This method employs the previously clustered facial features (see chapter 2) to form new faces with a desired social trait profile. The novelty in this work lies in the use of real facial feature traits to form new realistic faces able to elicit up to 15 different impressions.

Even though there already exist different works on this topic, it is difficult to find one which allows to create realistic faces, and even more if it is desired to convey more than one social trait at a time. In fact, the closest work to the described objective is the one of Walker and Vetter (2009), which is capable of crating realistic faces expressing one social trait at a time. Therefore, to the best of our knowledge, this thesis is the first work which allows to create a face with a specific social trait profile.

Moreover, results obtained give important insights on how important facial features are in the formation of each impression. This is a very interesting result, as it reveals where do we need to focus if we seek to convey a certain set of impressions.

On the other hand, very few works have achieved to generate realistic faces, even if only one impression was to be elicited. In contrast, our model obtains good results for 7 of the 15 social traits employed, namely Afraid, Babyface, Dominant, Feminine, Masculine, Sad and Unusual; with correlations higher than 0.65 and MSEs lower than 0.7.

It is also important to note that our model achieves very similar results with the existent CFD faces and with the generated ones. This means that the model is able to capture and reproduce these impressions. However, cases as Disgusted, Happy and Trustworthy suffer a great diminution, showing how the model is not able to properly mimic these impressions. In the case of Prototypical, its correlation falls from 0.8208 to 0.1689. After carefully reviewing the results and talking with people who did the validation, the most plausible reason for this is that there was a misunderstanding regarding what “Prototypical” meant.

Moreover, the results presented until now talk about how well our model is able to mimic certain impressions. However, the main objective of this thesis is to generate faces that elicit a certain set of impressions on most observers. Therefore, how the model performs on each impression separately is not as important as how it performs on generating a face with a desired set of impressions. In this case, our model is able to generate faces with a desired set of impressions with correlations higher than 0.65 ($p < 0.05$) for 8 of the 10 generated faces and MSEs lower than 0.61 for 7 of the 10 generated faces.

Nevertheless, this model has some important limitations, as the absence of hair modeling, the assumption of face symmetry or the absence of a female model. In addition, interactions between features are disregarded, which could vastly improve the performance of the model. However, we did not dispose of enough photographs as to consider these inter-relationships. Thus, the low quantity of faces available has been a limiting factor as well. In future works, more images should be gathered and labeled, and the face model should include hair, asymmetry and features inter-relationships. Furthermore, a female model should also be created.

To sum up, considering the great subjectivity and variability present in social trait impressions, we consider that the developed model is a step forward in the understanding of how these impressions are produced and how to reproduce them.

Chapter 4

Conclusions

As was indicated in chapter 1, the main objective of this thesis was to develop a system with the capability of creating avatar faces able to convey a desired set of impressions to most observers.

In order to achieve this main objective, the first step was to develop an automatic method to group facial features by similarity. Facial features are not independent one from another, that is, an eye may give a completely different impression depending on which mouth is chosen, for example. Therefore, the best way to model these relationships among features would be to find 2$^{\text{nd}}$, 3$^{\text{rd}}$, . . . , (n$_{\text{features}}$-1)$^{\text{th}}$ order relationships. The problem is the amount of data needed in order to accomplish this task. We only had 290 images to work with, what was not enough even to find significant 2$^{\text{nd}}$ order relationships. To overcome this limitation to the extent possible, we decided to perform a clustering with the facial features and thus account for the variance present within the cluster. Thus, taking as example the eyes, a score computed for a cluster including 10 eyes extracted from 10 different faces would also account for the influence of the different eyebrows, noses, mouths, etc., present in these faces. While this is not the best approach, it turned out to be enough to obtain reasonable results. Gathering more images in order to be able to model 2$^{\text{nd}}$ and higher order relationships remains as future work.

The clustering method developed in chapter 2 of this thesis follows a holistic approach which allows to objectively and automatically group the available facial features by appearance. This automatic clustering method further allows to codify faces in a standardized manner, which helps to define faces in a consistent way. Furthermore, we obtained an objective, appearance-based facial feature taxonomy which can be used in any domain of interest.

Chapter 3 shows how the extracted facial features grouped by similarity are employed to build a model which allows to assess a given face and give it a social trait profile. To perform this step, a set of weights was optimized which, in turn, allows to explain how much each facial feature affects the formation of each impression. This optimization is needed and cannot be avoided, as literature shows that not all facial features have the same importance when inferring social attributes (Santos and A. W. Young, 2011).

Using this model, we can generate new definitions of faces with a certain social trait profile. On the one hand, when dealing with social traits, this approach allowed us to obtain a model which achieves correlations higher than 0.68 and MSEs lower than 0.55 for 7 of the 15 social traits employed (Afraid, Babyface, Dominant, Feminine, Masculine, Sad and Unusual). On the other hand, our model is able to generate faces expressing a desired set of impressions with correlations higher than 0.65 for 8 of the 10 generated faces and MSEs lower than 0.61 for 7 of the 10 generated faces. We consider this a significant breakthrough and an important contribution to the state-of-the-art, as we have been unable to find any other previous work able to model so many trait dimensions on realistic face images.

Furthermore, we implemented the image fusion method developed by Pérez et al. (2003), which combined with our model, allowed us to combine different facial features and form a realistic face seamlessly. We are then capable of creating a realistic face from just a face definition as input which, in addition, conveys a sought social trait profile to most observers.

Nevertheless, the work has some limitations. The first and most important one is that we focused on male faces, leaving female gender unattended. We made this decision as it was easier to extract the facial features from male subjects rather than from female ones due to their hair, and because of time limitations.

It is also important to note that no 2$^{\text{nd}}$ and higher order relationships among facial features have been accounted for in this model. This is a very important limitation to bear in mind and it will be mitigated in following versions.

However, as gathering a sufficient quantity of face images and rating them is a very time-consuming task, it is not possible to include it in this thesis.

Another important limitation is the absence of hair and asymmetry in the model. Human faces are not completely symmetric, and small displacements of facial features could play an important role in the final perception of the face. Then, a model which could account for these displacements and asymmetries would most likely perform better. Finally, a better and more robust model could have been obtained if more evaluated face images had been available.

Therefore, in future works, a model should be developed for female gender. More images should be gathered and evaluated by humans so higher order relationships among features could also be modeled and included within the model. Finally, hair, asymmetry and feature displacements would also need to be included in the model in order to better characterize the human face.

Merits

Journal papers

Diego-Mas, J. A., **Fuentes-Hurtado, F.**, Naranjo, V., & Raya, M. A. Influences of each facial feature on how we perceive and interpret faces. Under review in *Scientific Reports.*

Fuentes-Hurtado, F., Diego-Mas, J. A., Naranjo, V., & Raya, M. A. Automatic classification of human facial features based on their appearance. Under review in *PLOS ONE.*

Martín, A., Lara-Cabrera, R., **Fuentes-Hurtado, F.**, Naranjo, V., & Camacho, D. EvoDeep: a new Evolutionary approach for automatic Deep Neural Networks parameterization. In press in *Journal of Parallel and Distributed Computing.*

Fuentes-Hurtado, F., Diego-Mas, J. A., Naranjo, V., & Raya, M. A. A hybrid method for accurate iris segmentation on at-a-distance visible wavelength images. Under review in *EURASIP Journal on Image and Video Processing.*

Granero, A. C., **Fuentes-Hurtado, F.**, Ornedo, V. N., Provinciale, J. G., Ausín, J. M., & Raya, M. A. (2016). A comparison of physiological signal analysis techniques and classifiers for automatic emotional evaluation of audiovisual contents. *Frontiers in Computational Neuroscience*, 10.

International conferences

Martín, A., **Fuentes-Hurtado, F.**, Naranjo, V., & Camacho, D. (2017, June). Evolving deep neural networks architectures for Android malware classification. In *Congress on Evolutionary Computation (CEC), 2017 IEEE* (pp. 1659-1666).

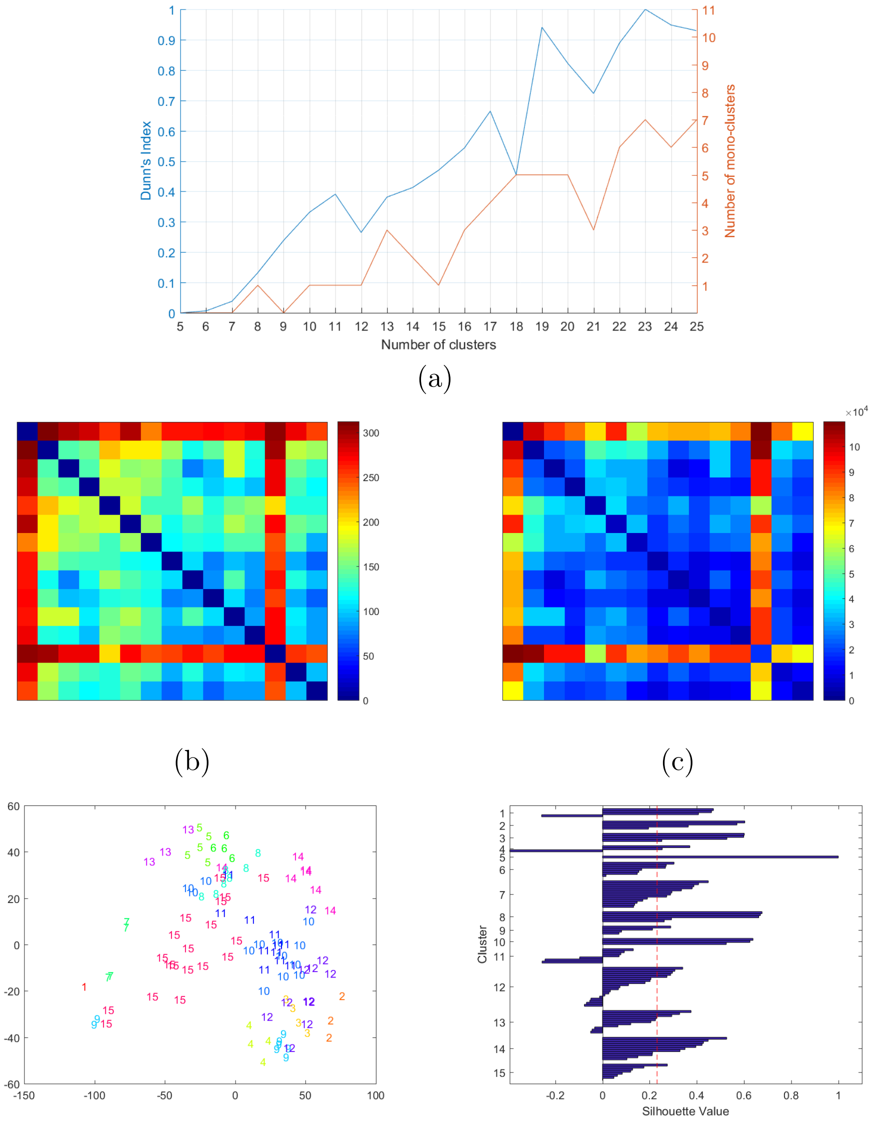
Fuentes-Hurtado, F., Diego-Mas, J. A., Naranjo, V., & Raya, M. A. How important are the eyes in trustworthiness perception? A system to tell the importance of facial features in social trait perception. Under review in *EUSIPCO 2018.*

Fuentes-Hurtado, F., Diego-Mas, J. A., Naranjo, V., & Raya, M. A. A Holistic automatic method for grouping facial features based on their appearance. Under review in *EUSIPCO 2018.*

Appendices

Appendix A

Clustering results



(a)

(b)

(c)

(d)

(e)

Figure A.1: Asian eyebrows clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

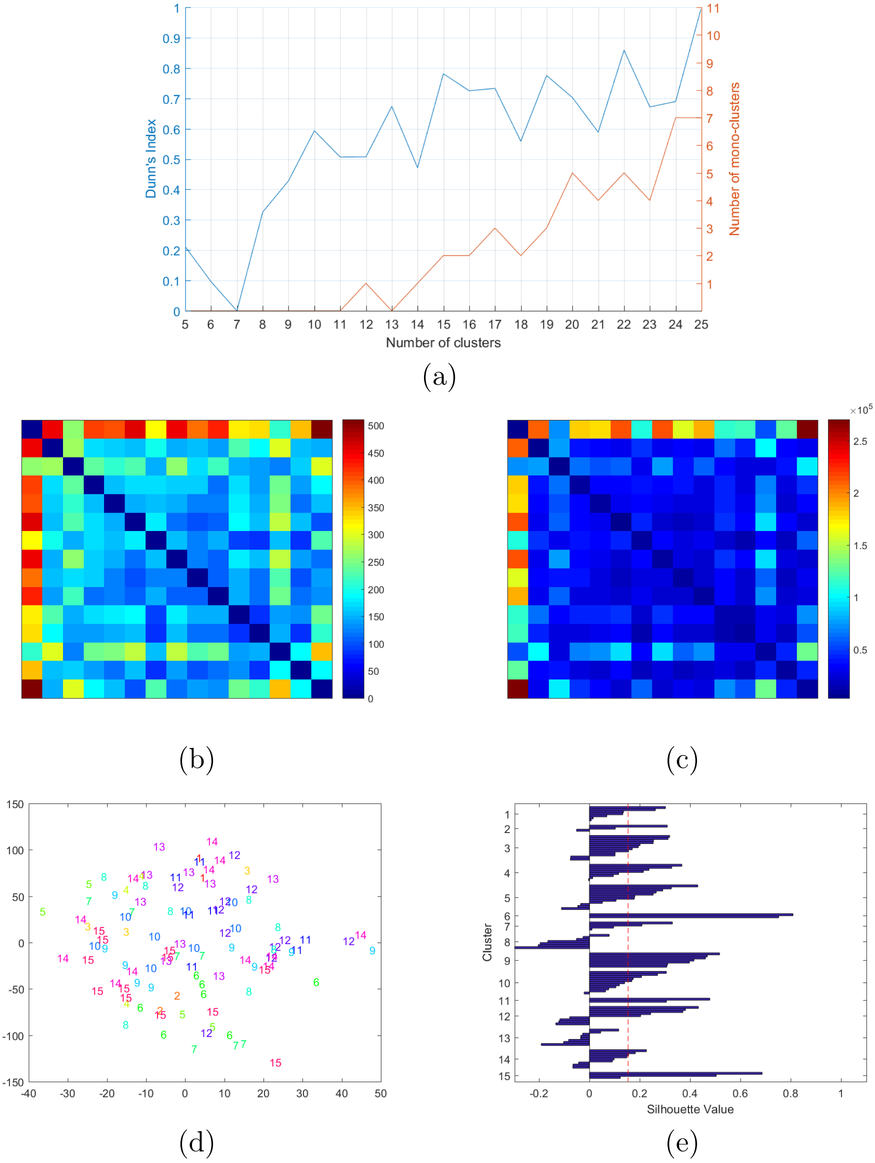
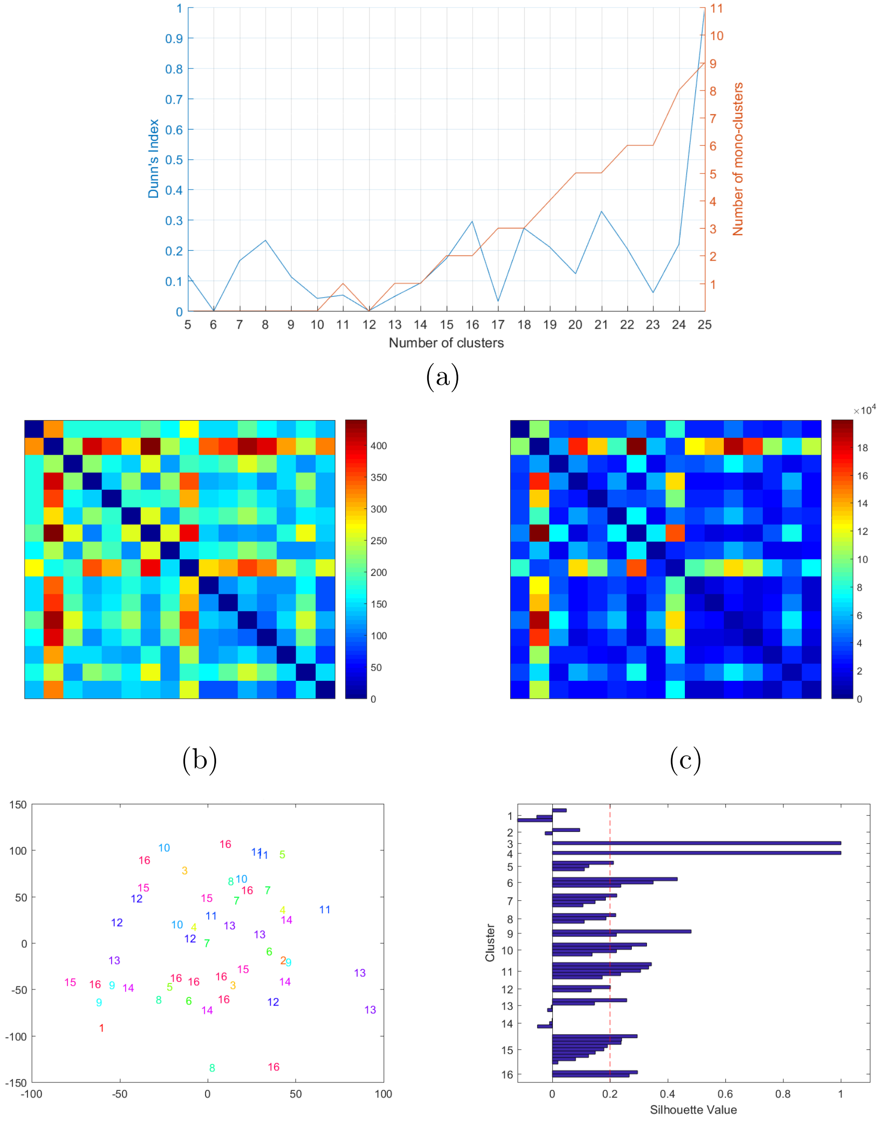


Figure A.2: Asian eyes clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

(b)

(c)

(d)

(e)

Figure A.3: Asian noses clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

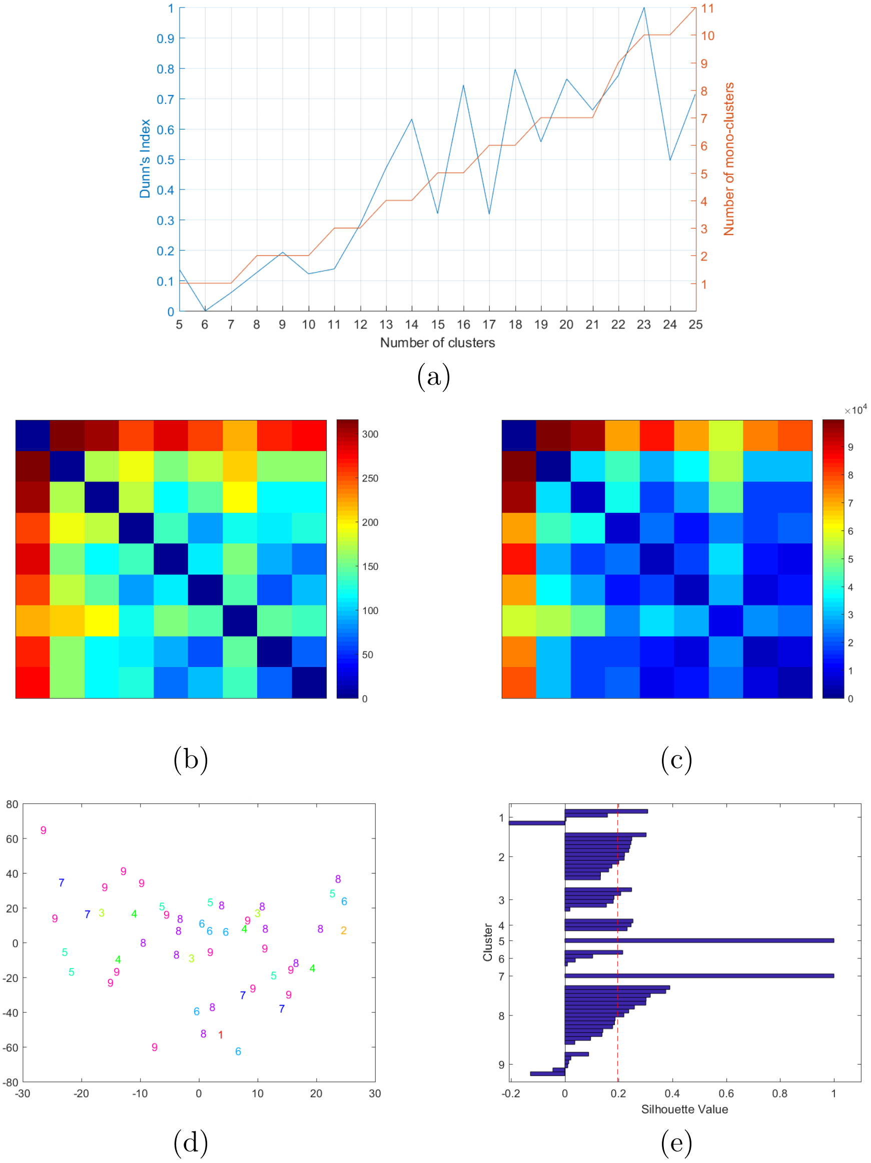
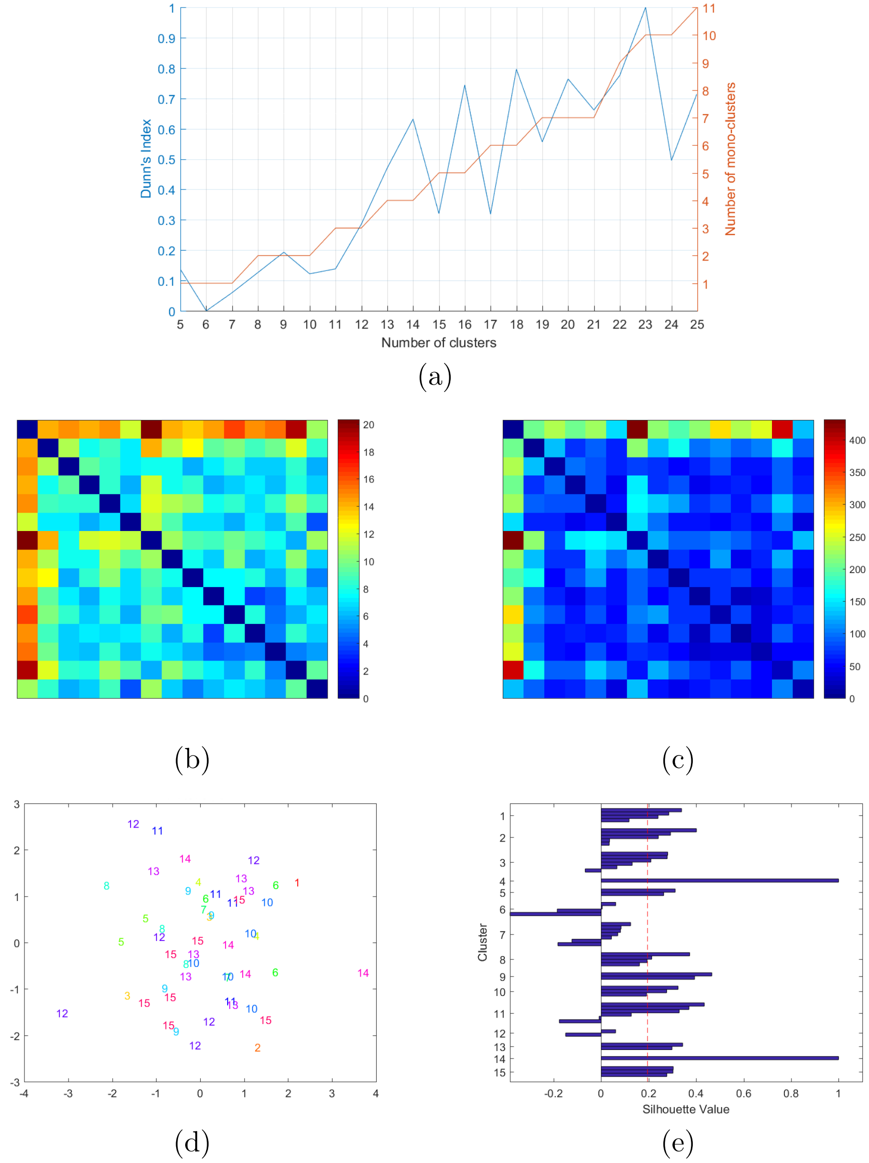


Figure A.4: Asian mouths clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

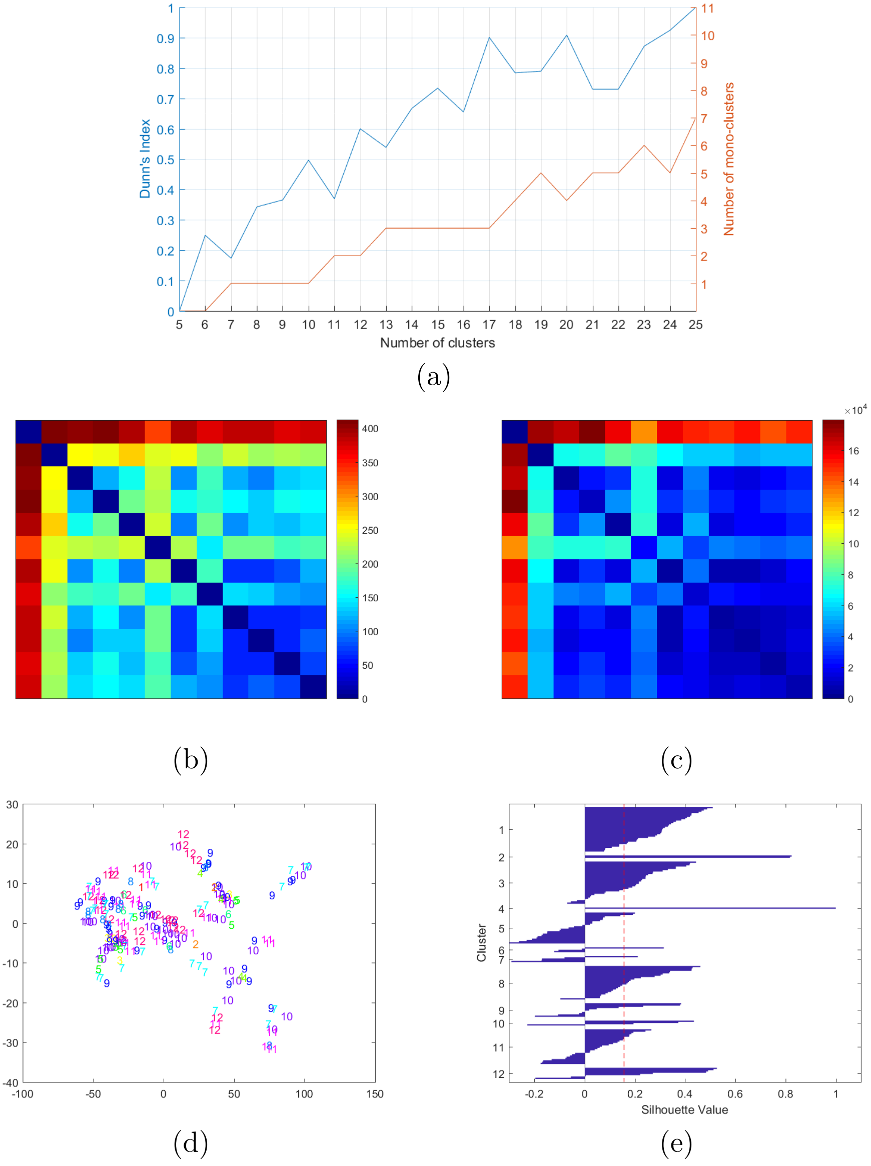
(b)

(c)

(d)

(e)

Figure A.5: Asian jawlines clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

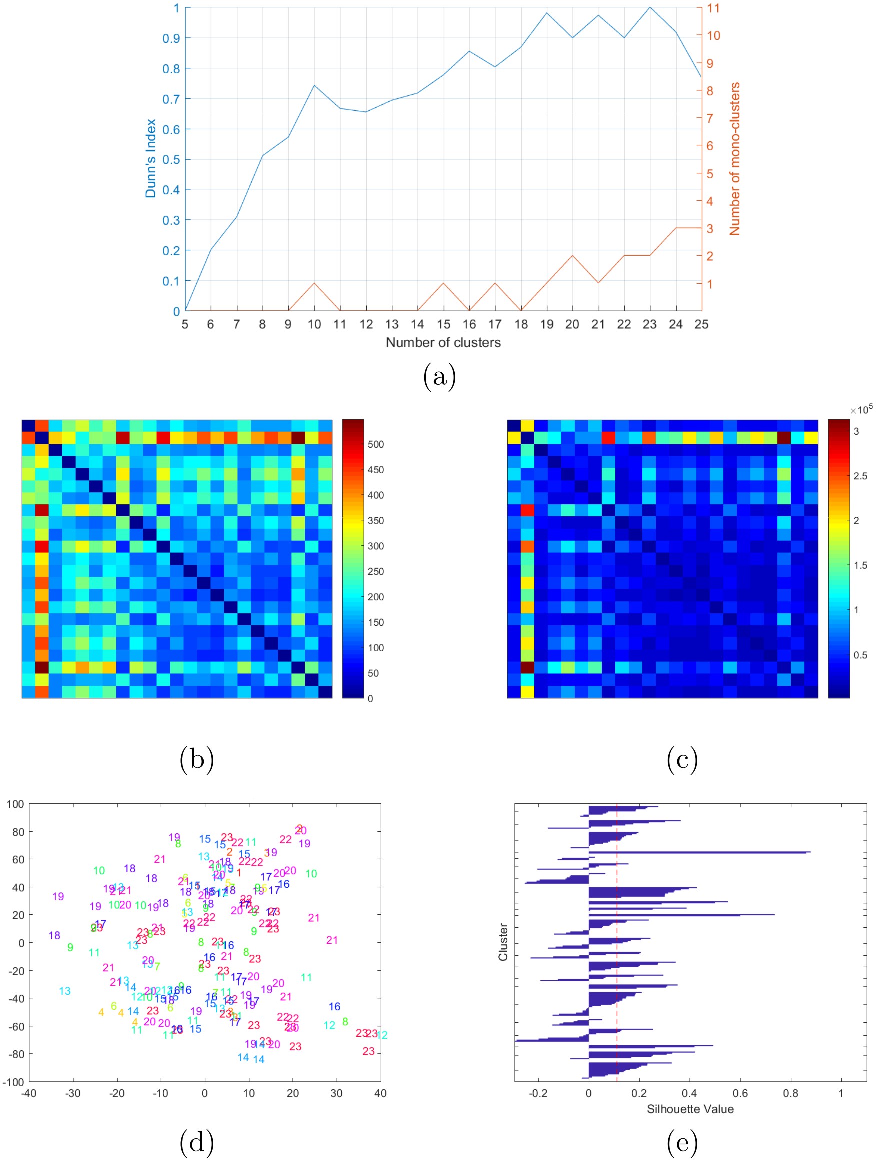
(b)

(c)

(d)

(e)

Figure A.6: Black eyebrows clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

(b)

(c)

(d)

(e)

Figure A.7: Black eyes clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

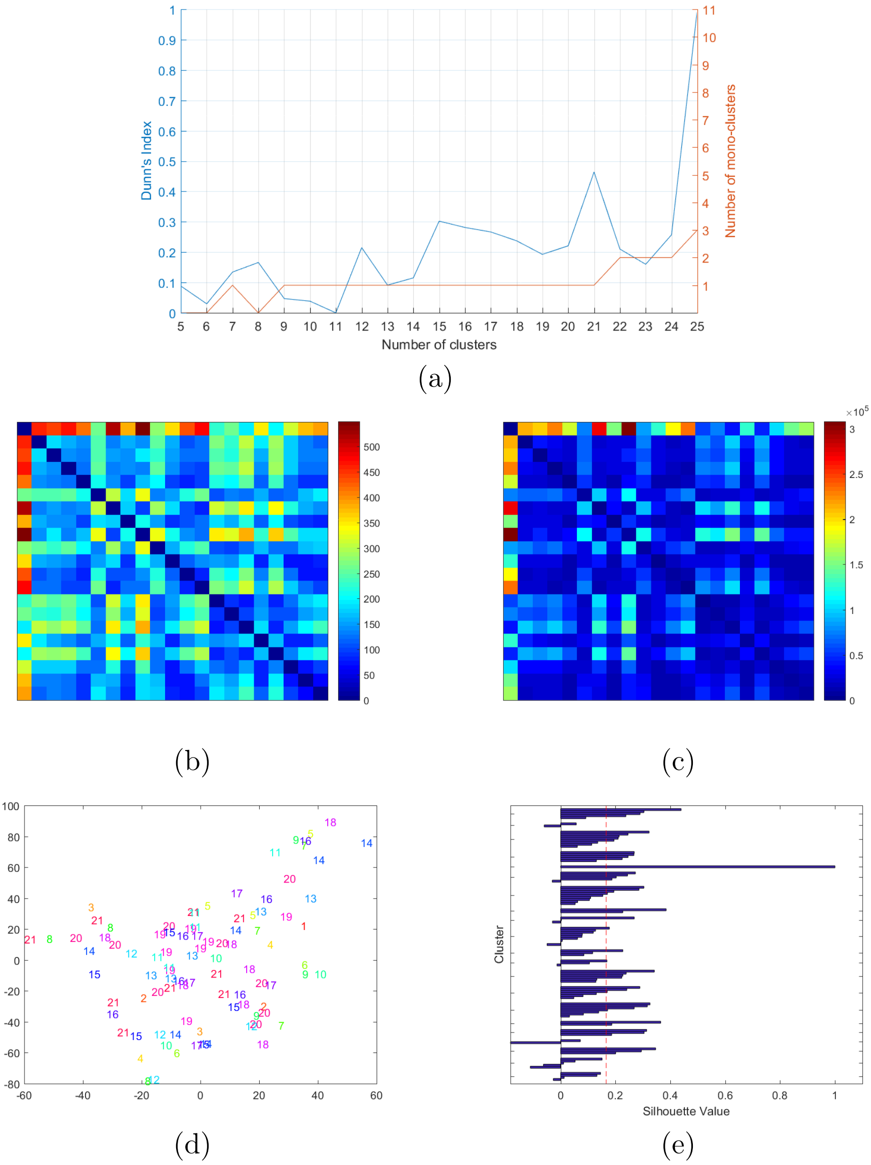
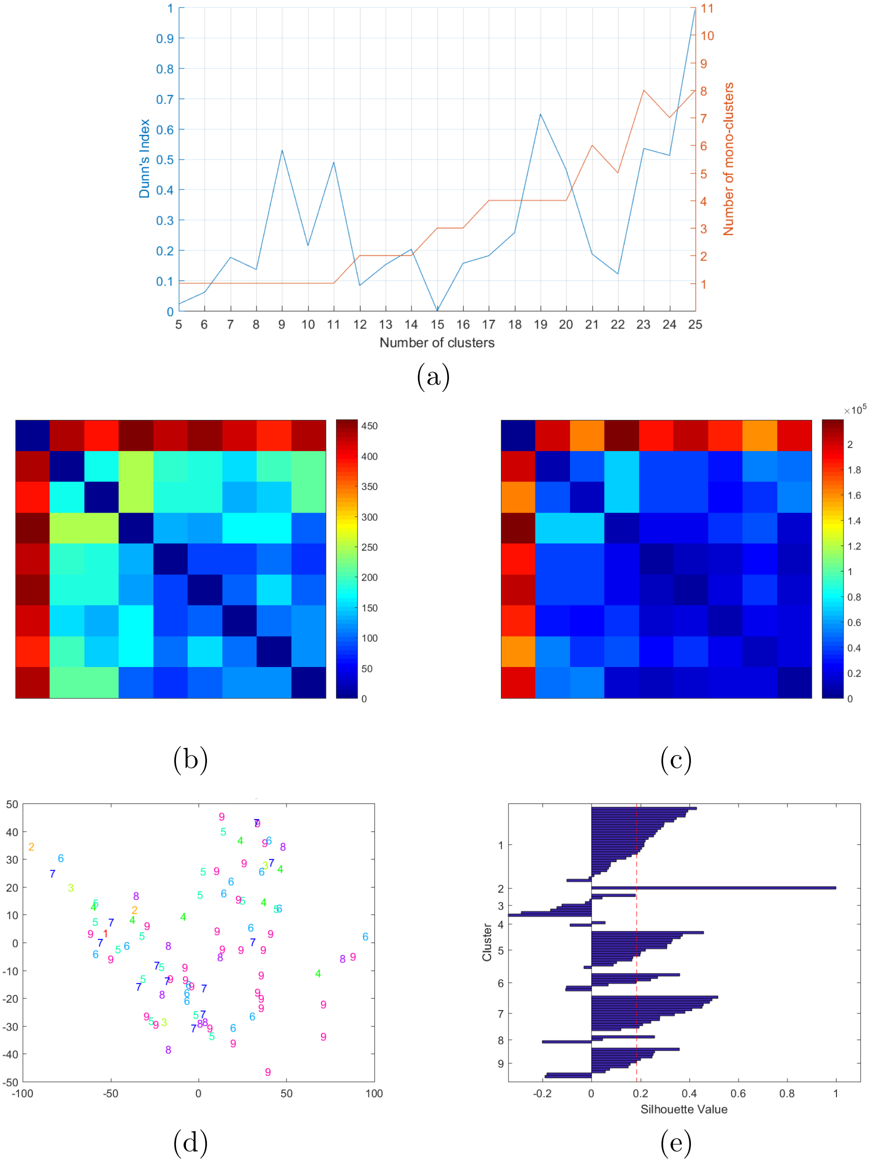


Figure A.8: Black noses clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

(b)

(c)

(d)

(e)

Figure A.9: Black mouths clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

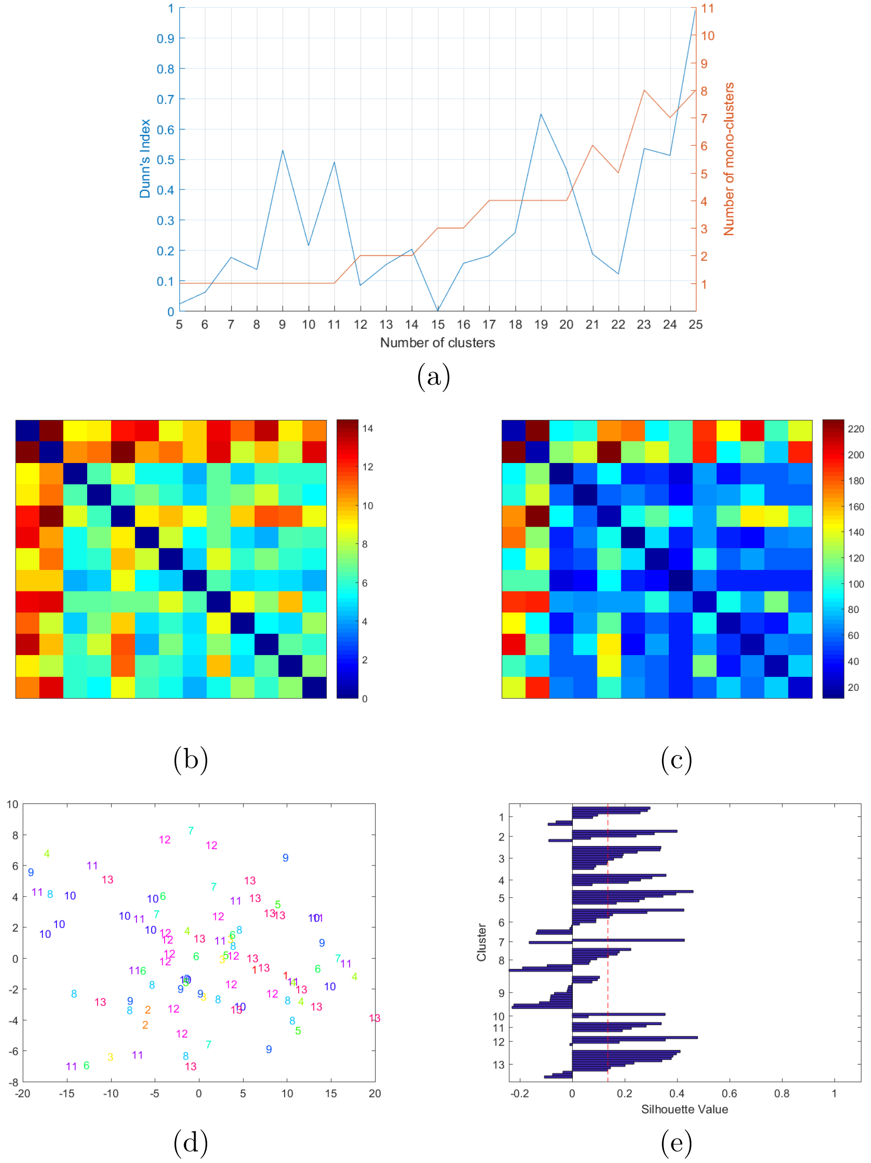
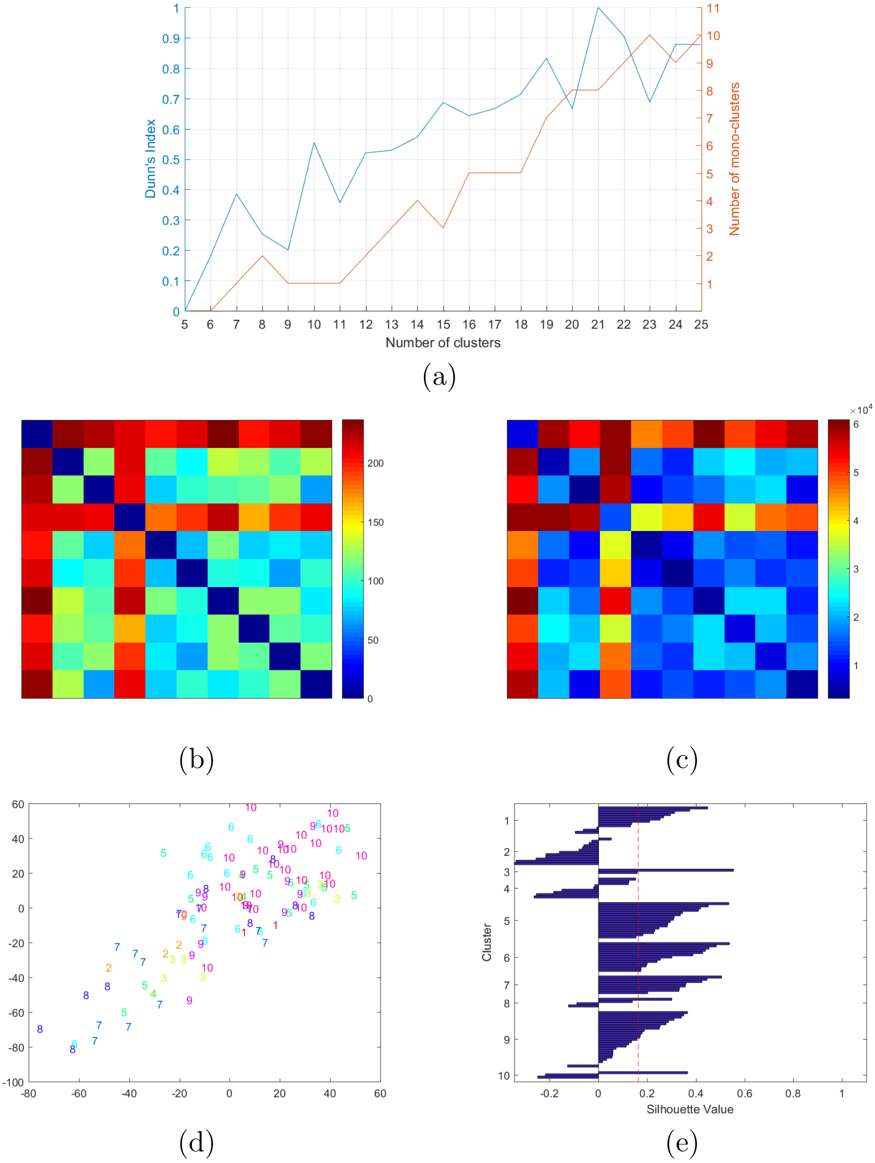


Figure A.10: Black jawlines clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

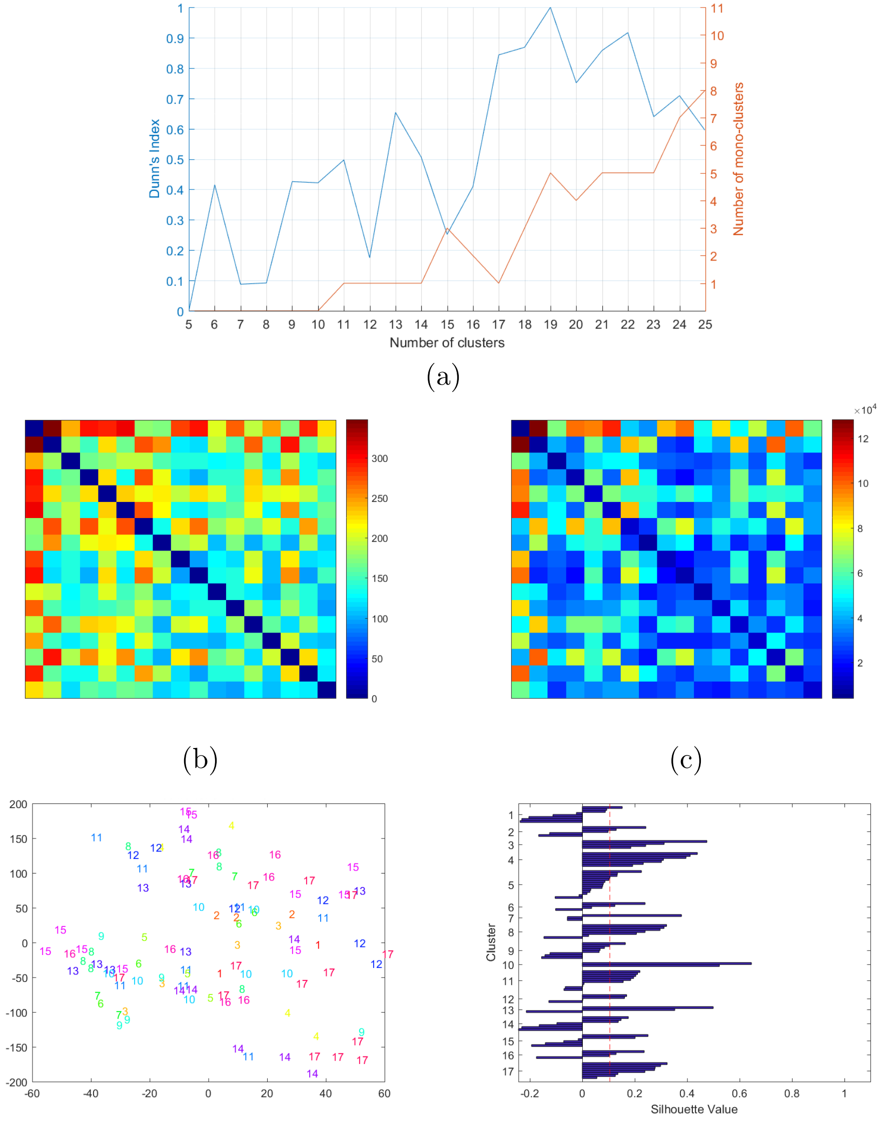
(b)

(c)

(d)

(e)

Figure A.11: Latino eyebrows clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

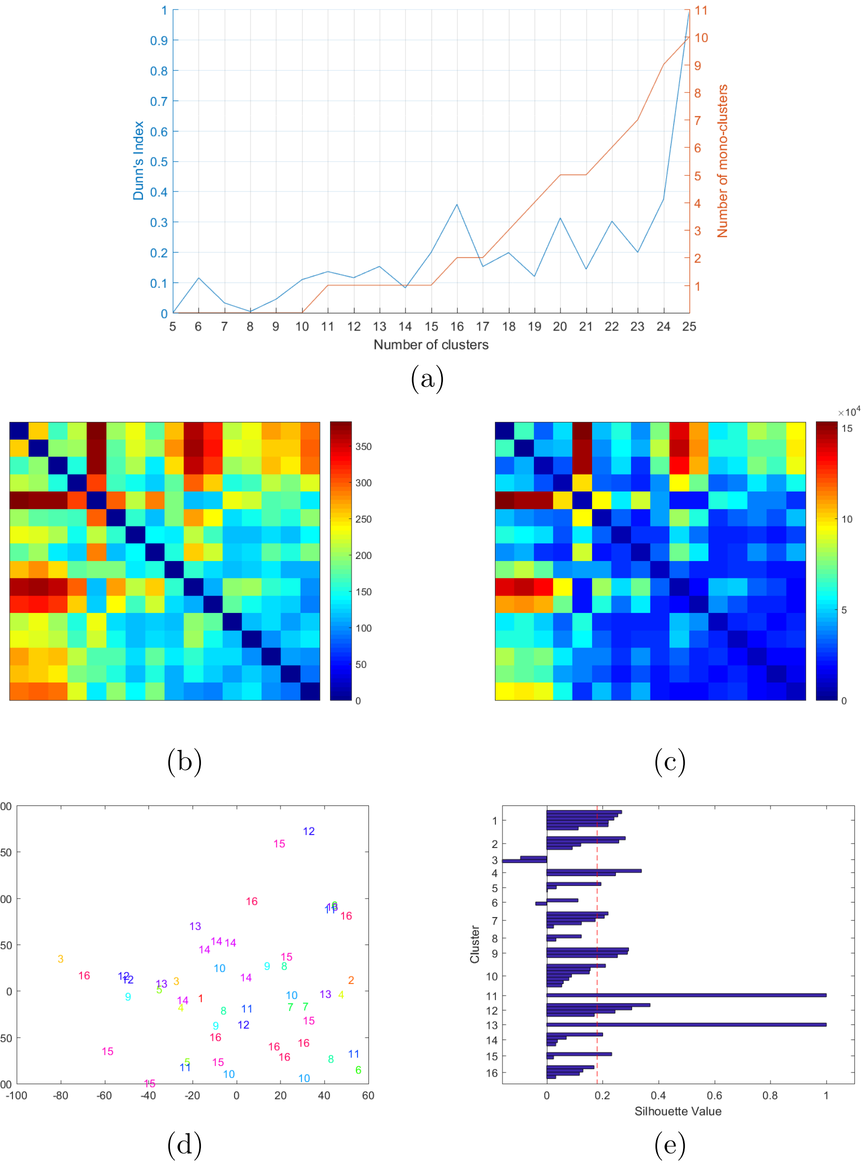
(b)

(c)

(d)

(e)

Figure A.12: Latino eyes clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

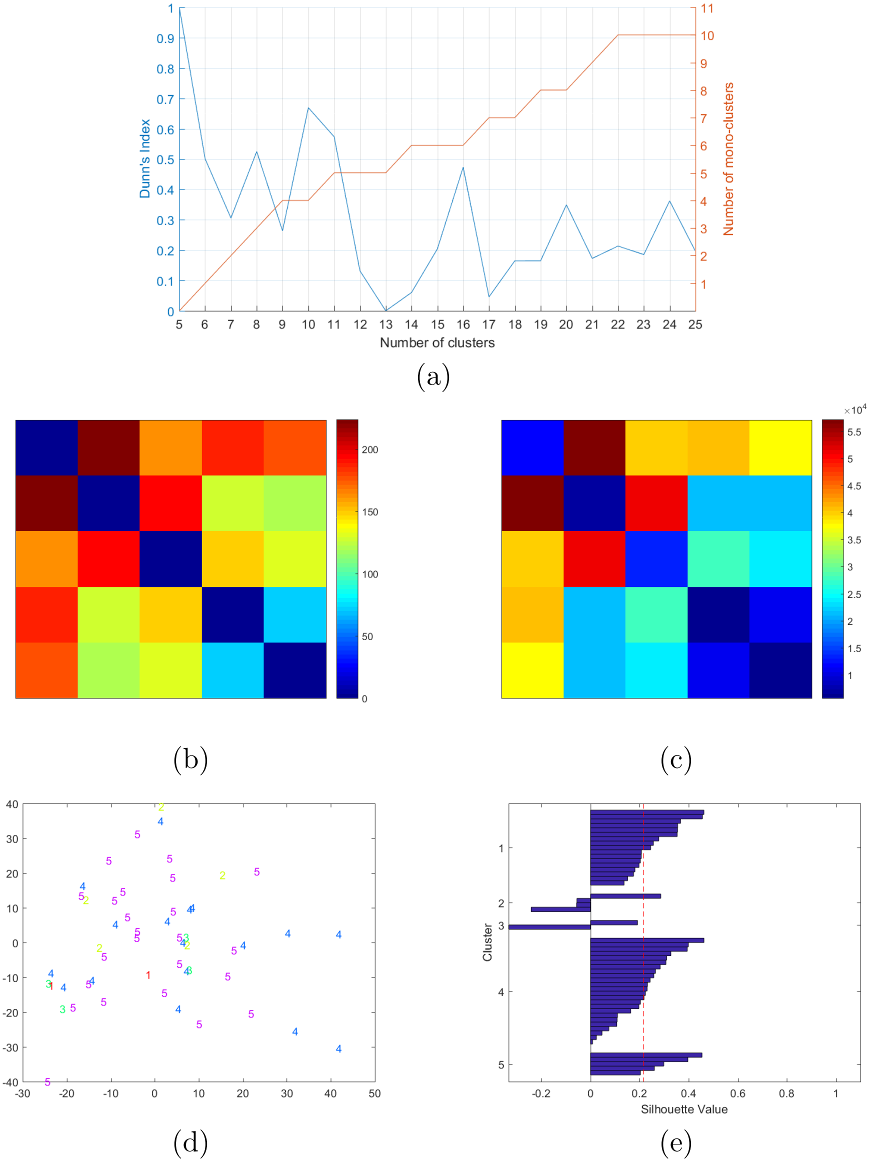
(b)

(c)

(d)

(e)

Figure A.13: Latino noses clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

(b)

(c)

(d)

(e)

Figure A.14: Latino mouths clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

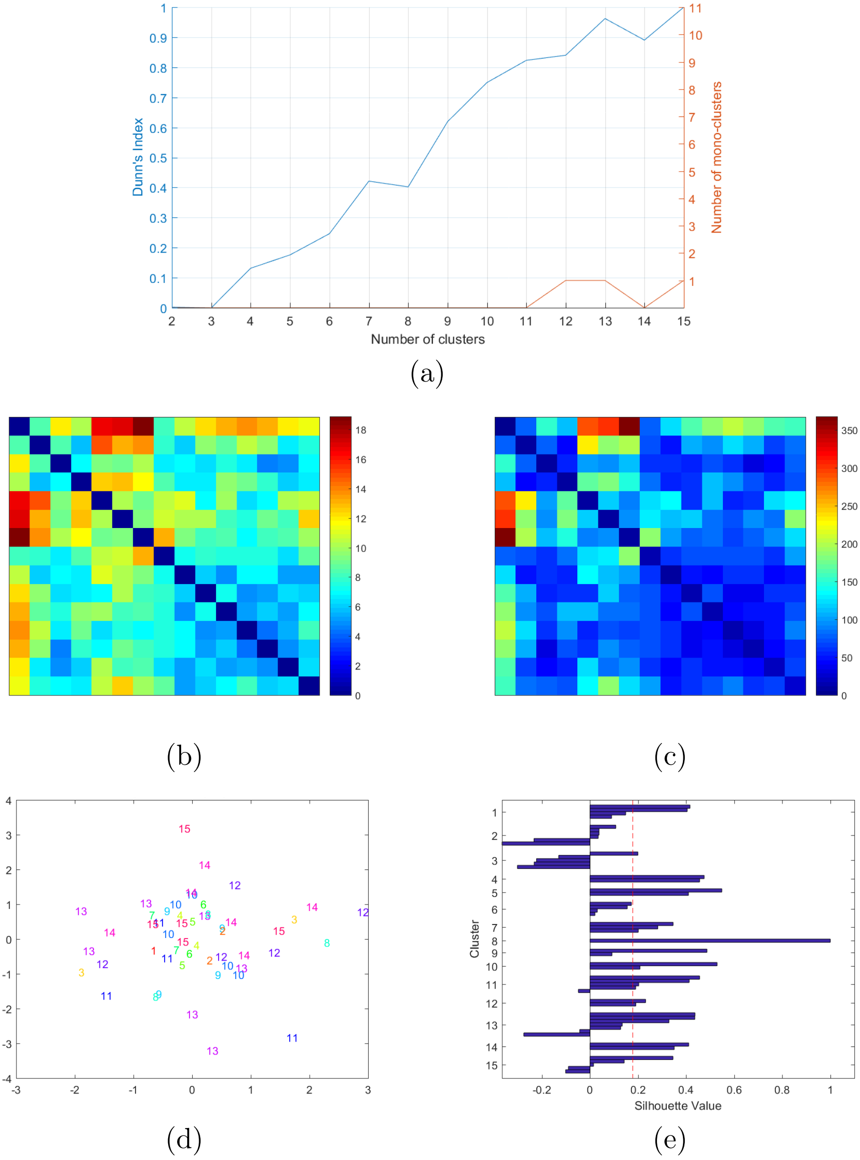


Figure A.15: Latino jawlines clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

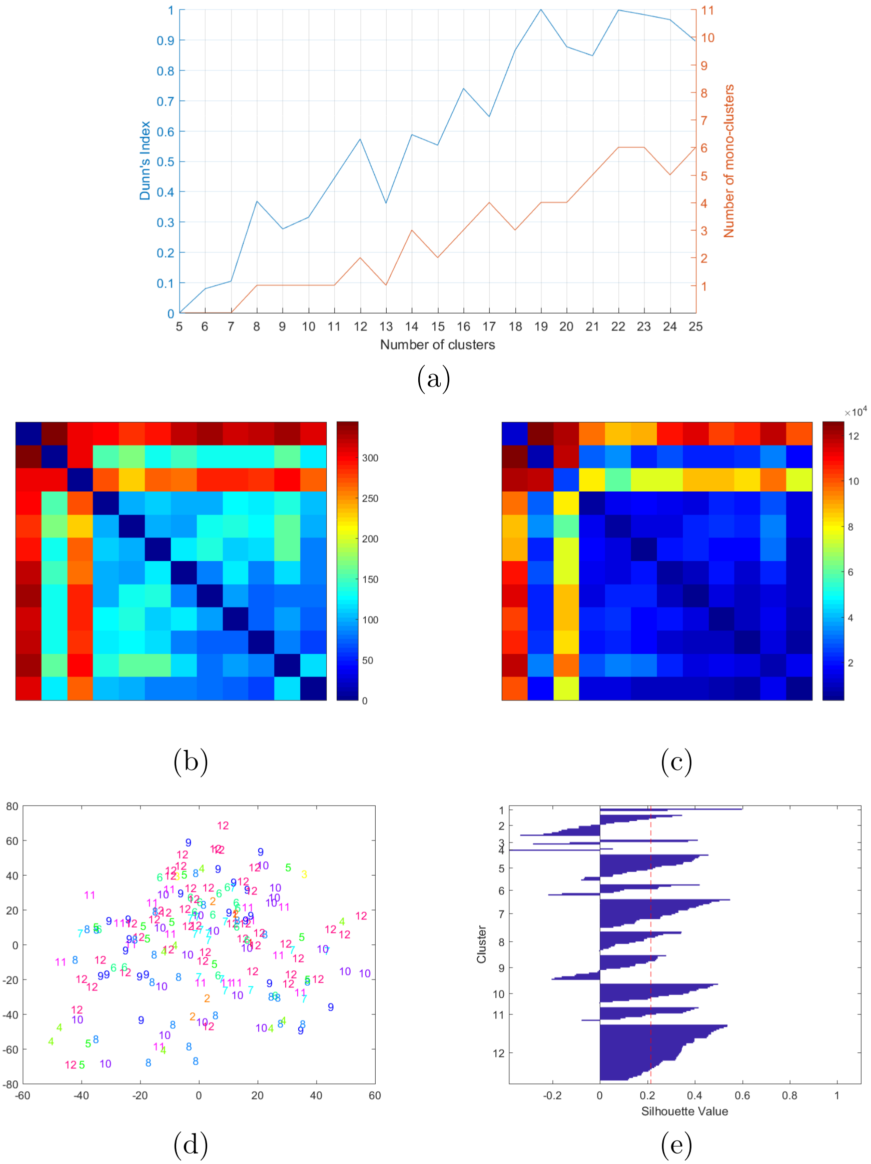
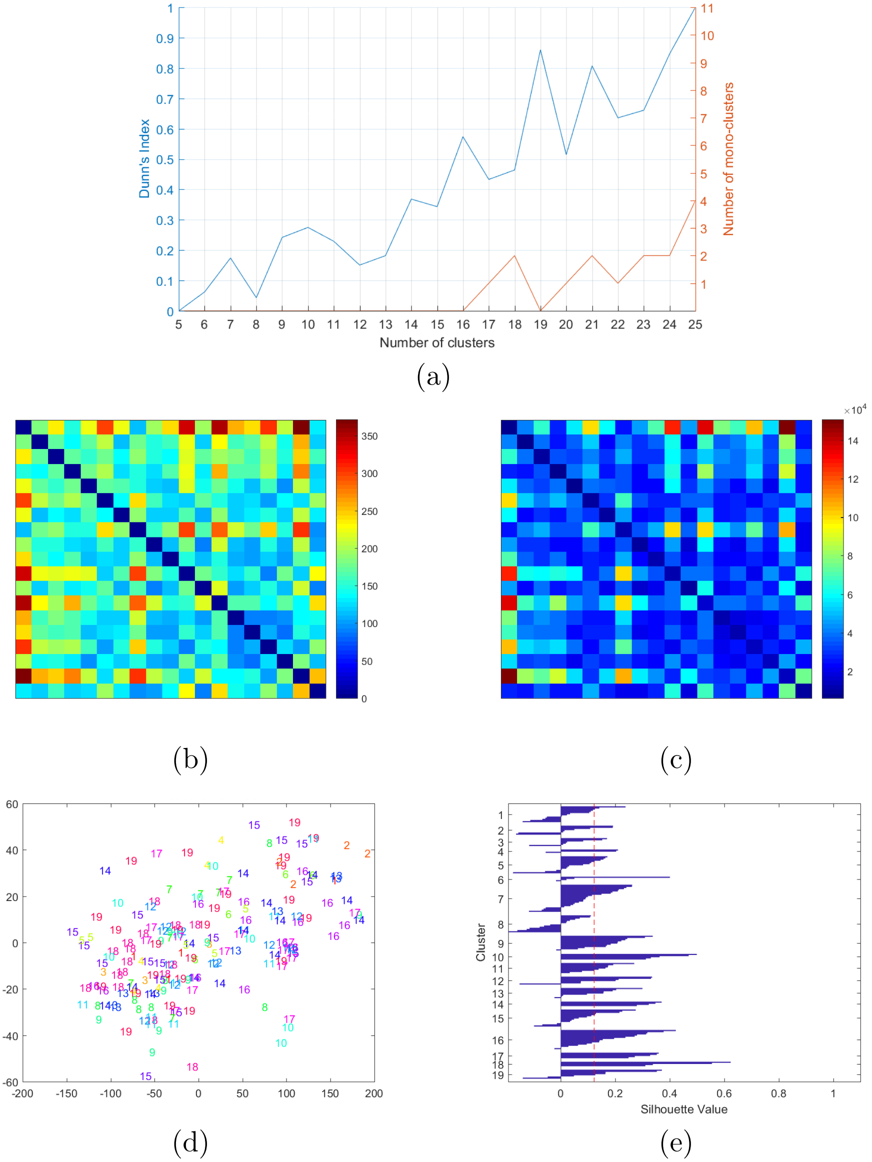


Figure A.16: White eyebrows clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

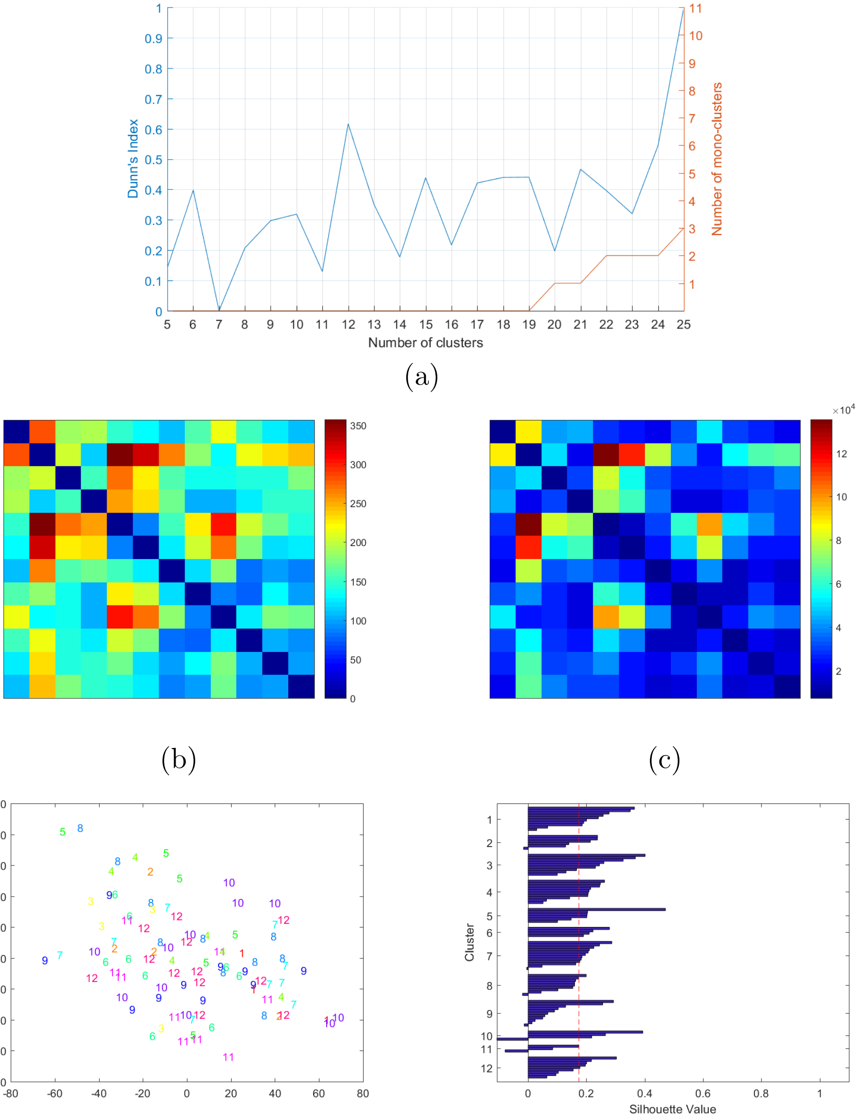
(b)

(c)

(d)

(e)

Figure A.17: White eyes clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

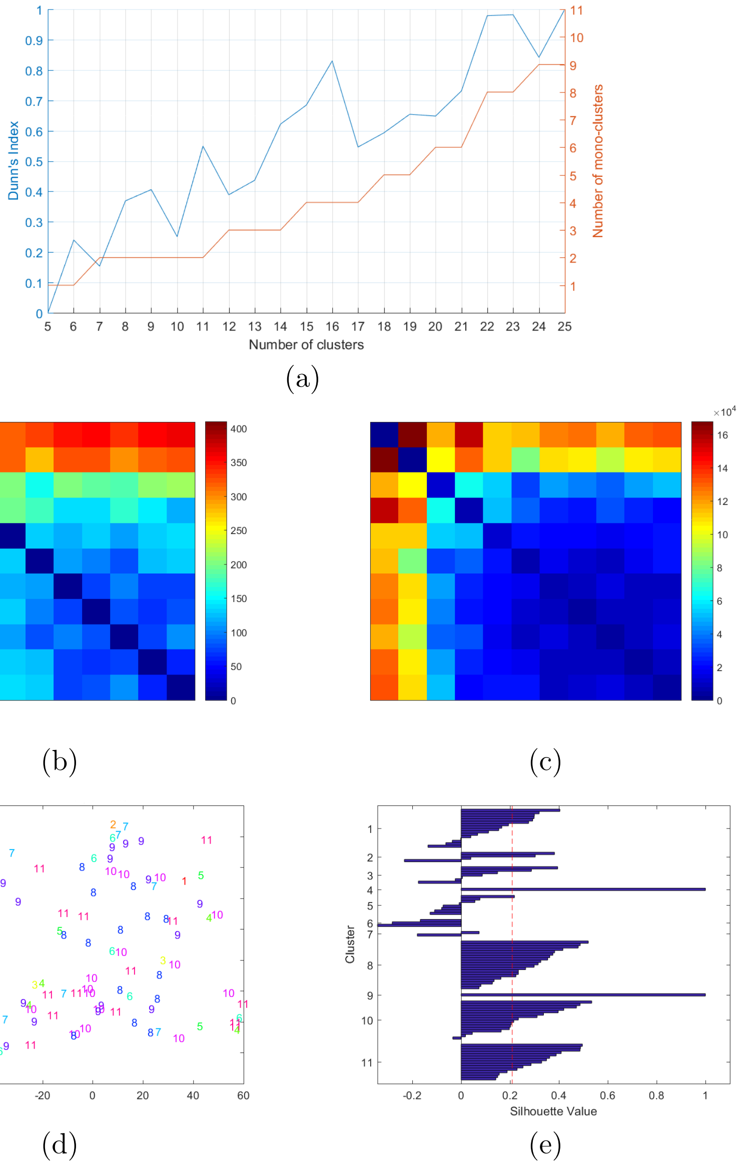
(b)

(c)

(d)

(e)

Figure A.18: White noses clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

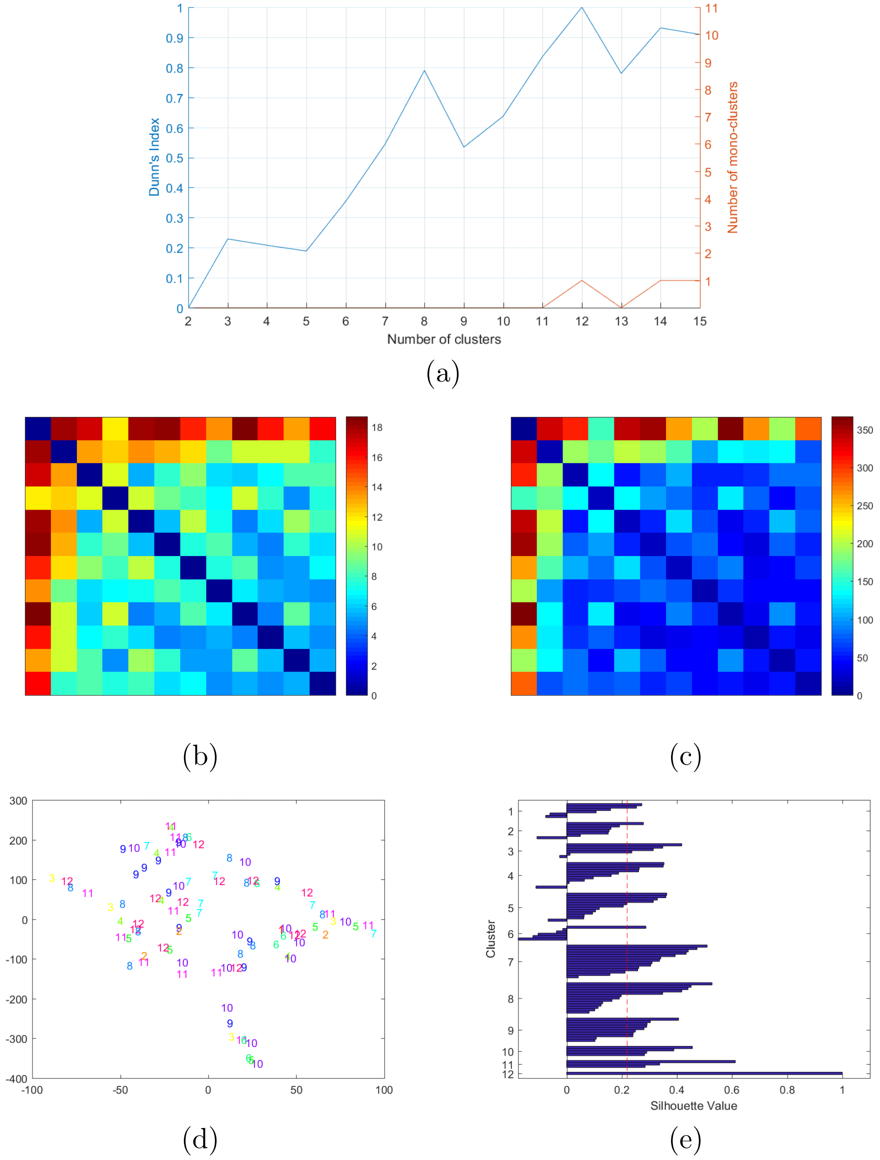
(b)

(c)

(d)

(e)

Figure A.19: White mouths clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.



(a)

(b)

(c)

(d)

(e)

Figure A.20: White jawlines clustering metrics. (a) shows the data employed in order to chose k. (b) and (c) represent the inter-cluster distances and the distances of every instance within a cluster to its centroid and all other cluster centroids. (d) shows the representation of the clustering in a two-dimensional space. (e) presents the silhouette of the clustering.

Appendix B

Taxonomies

B.1 Asian taxonomy

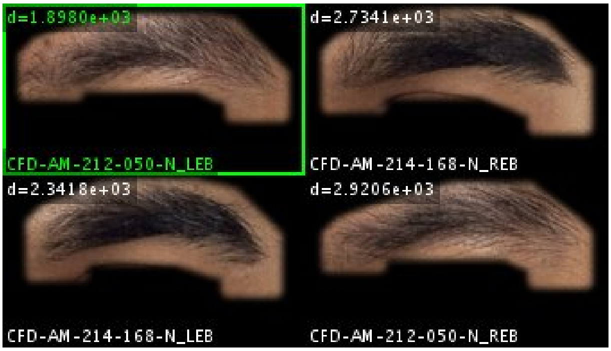


Figure B.1: Asian eyebrows clustering example 1.

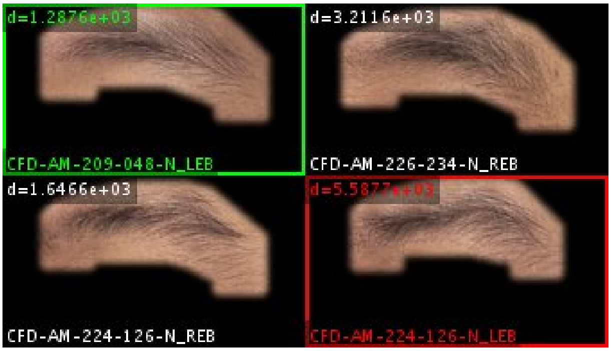


Figure B.2: Asian eyebrows clustering example 2.

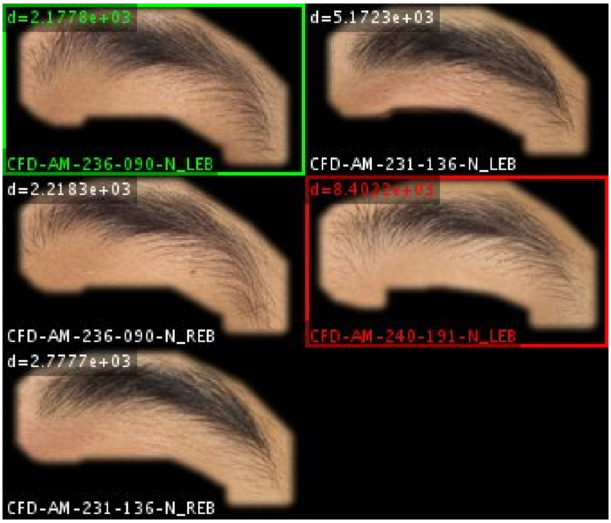


Figure B.3: Asian eyebrows clustering example 3.

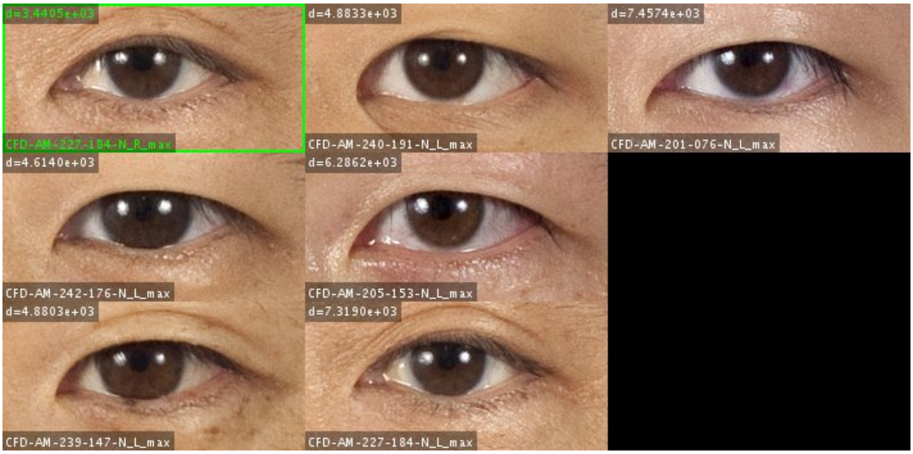


Figure B.4: Asian eyes clustering example 1.

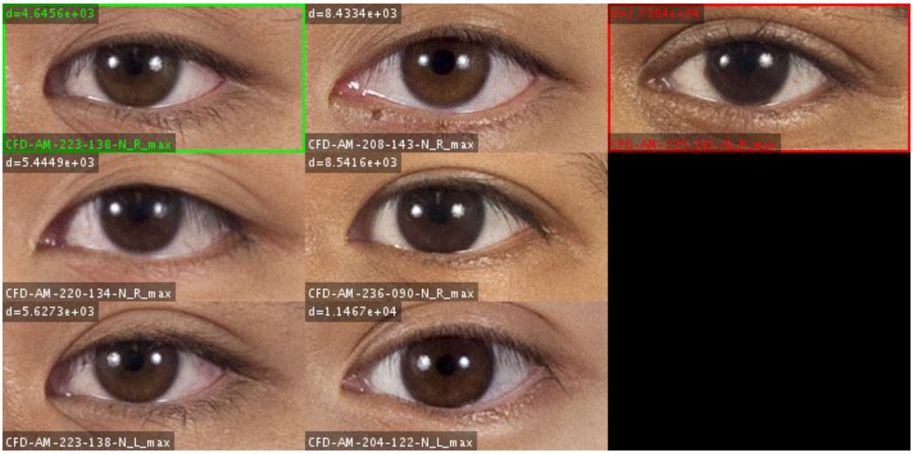


Figure B.5: Asian eyes clustering example 2.

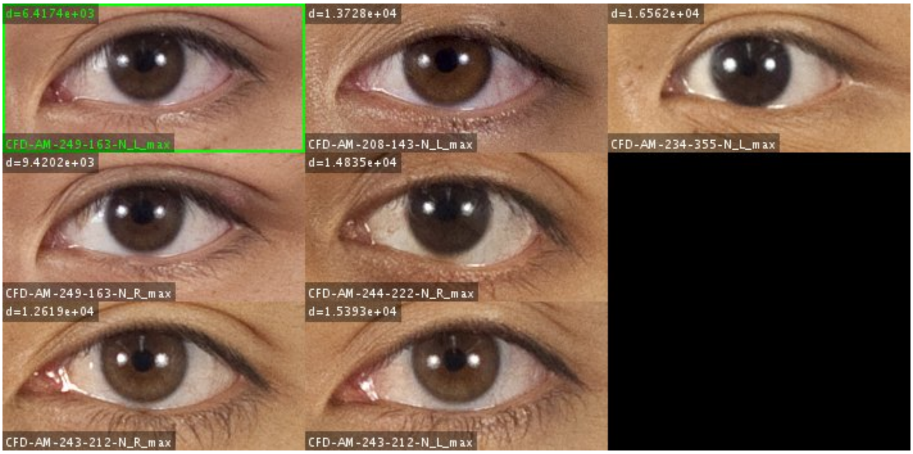


Figure B.6: Asian eyes clustering example 3.

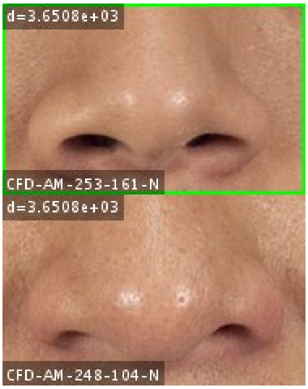


Figure B.7: Asian noses clustering example 1.

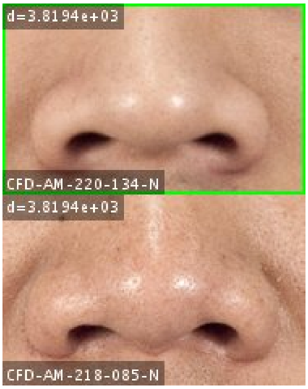


Figure B.8: Asian noses clustering example 2.

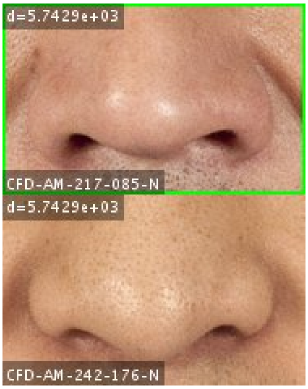


Figure B.9: Asian noses clustering example 3.

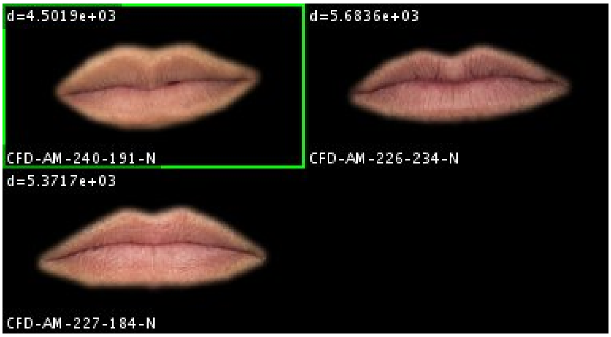


Figure B.10: Asian mouths clustering example 1.

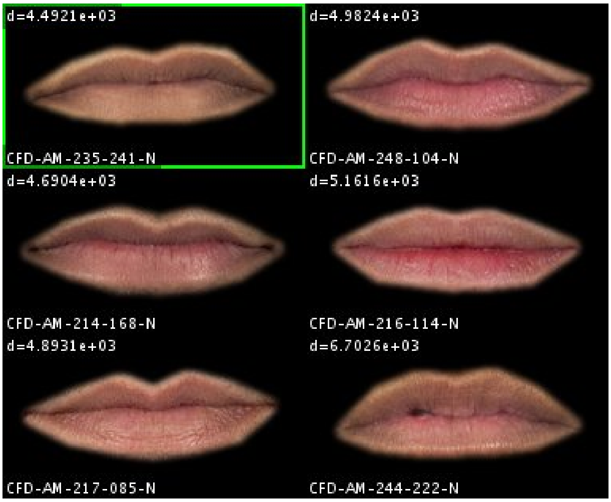


Figure B.11: Asian mouths clustering example 2.

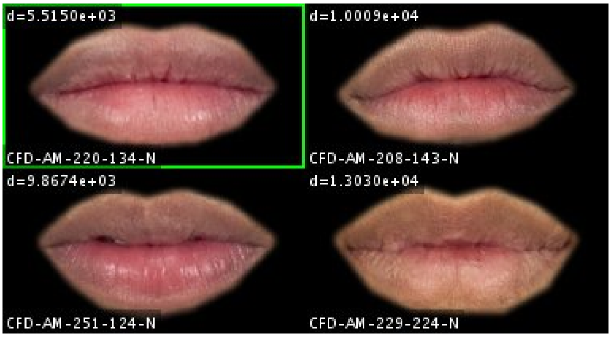


Figure B.12: Asian mouths clustering example 3.

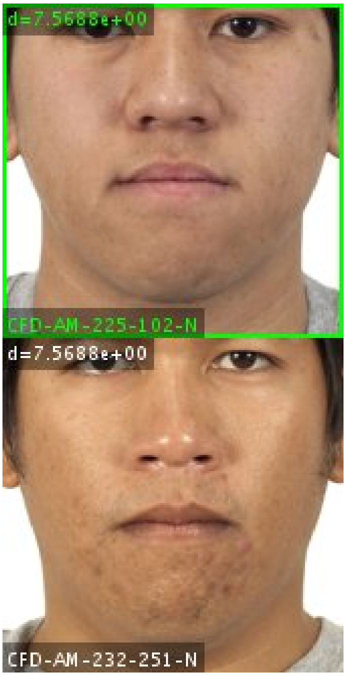


Figure B.13: Asian jawlines clustering example 1.

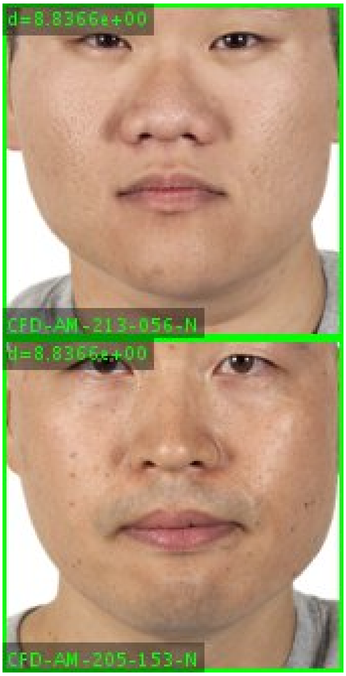


Figure B.14: Asian jawlines clustering example 2.

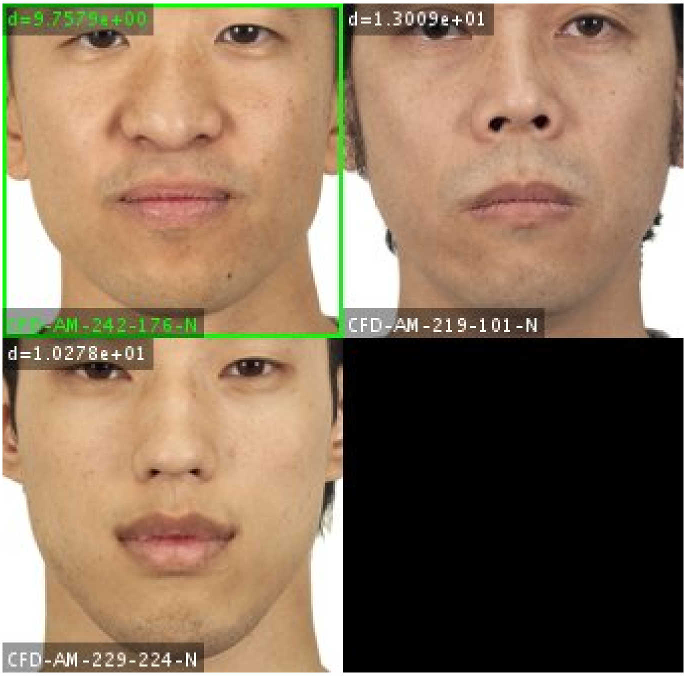


Figure B.15: Asian jawlines clustering example 3.

B.2 Black taxonomy

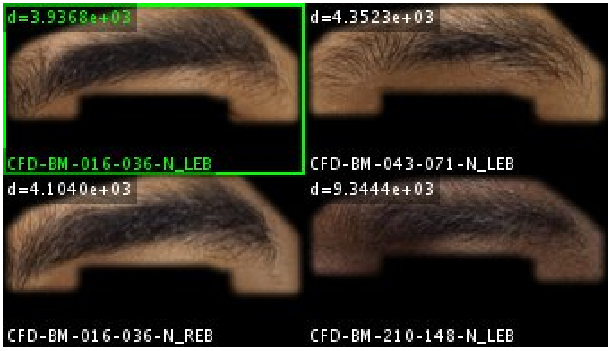


Figure B.16: Black eyebrows clustering example 1.

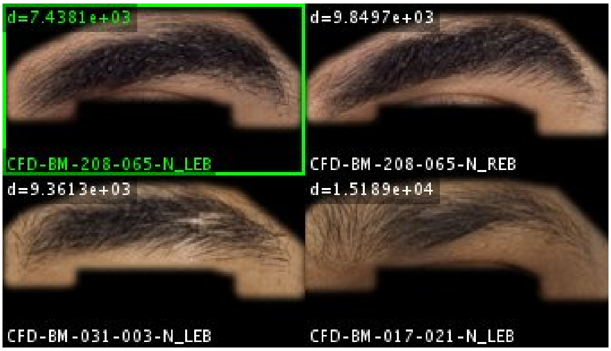


Figure B.17: Black eyebrows clustering example 2.

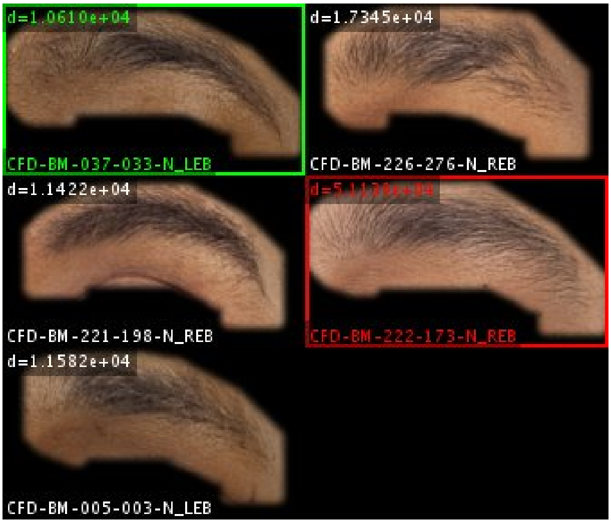


Figure B.18: Black eyebrows clustering example 3.

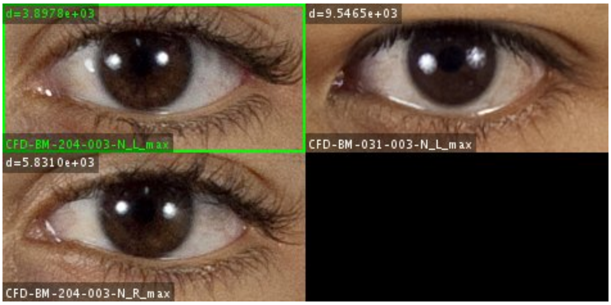


Figure B.19: Black eyes clustering example 1.

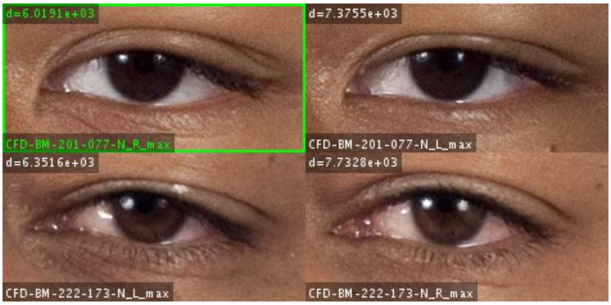


Figure B.20: Black eyes clustering example 2.

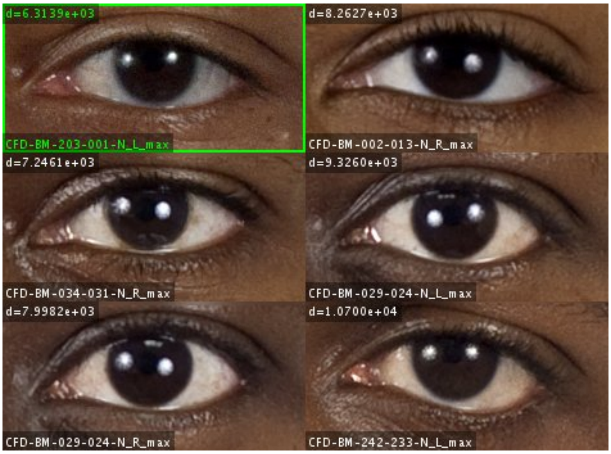


Figure B.21: Black eyes clustering example 3.

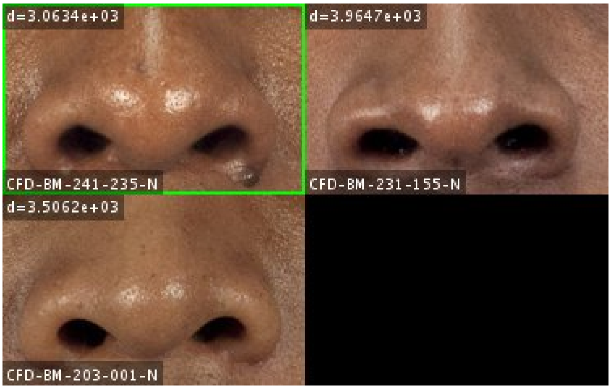


Figure B.22: Black noses clustering example 1.

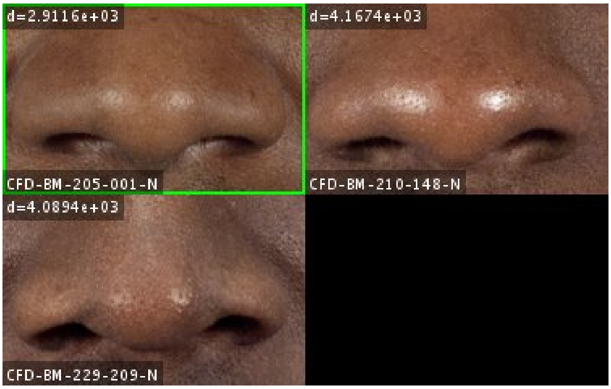


Figure B.23: Black noses clustering example 2.

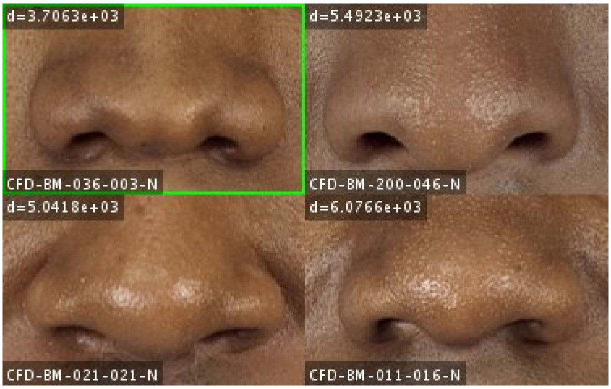


Figure B.24: Black noses clustering example 3.

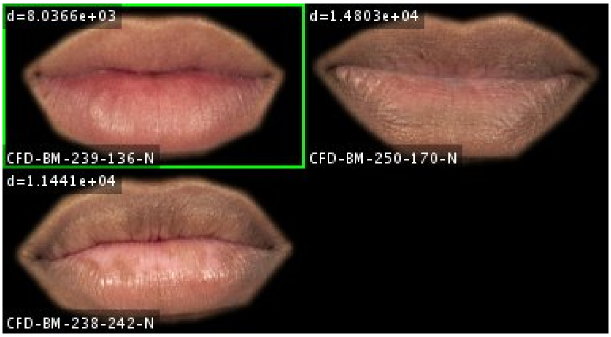


Figure B.25: Black mouths clustering example 1.

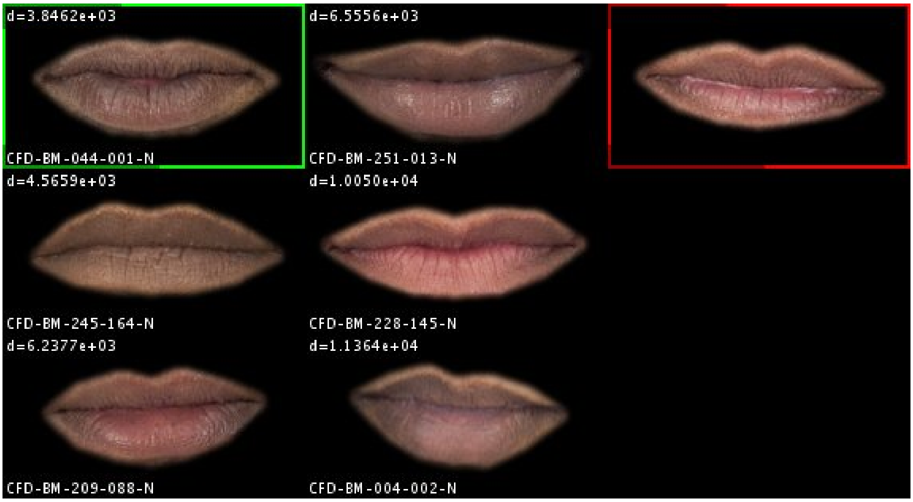


Figure B.26: Black mouths clustering example 2.

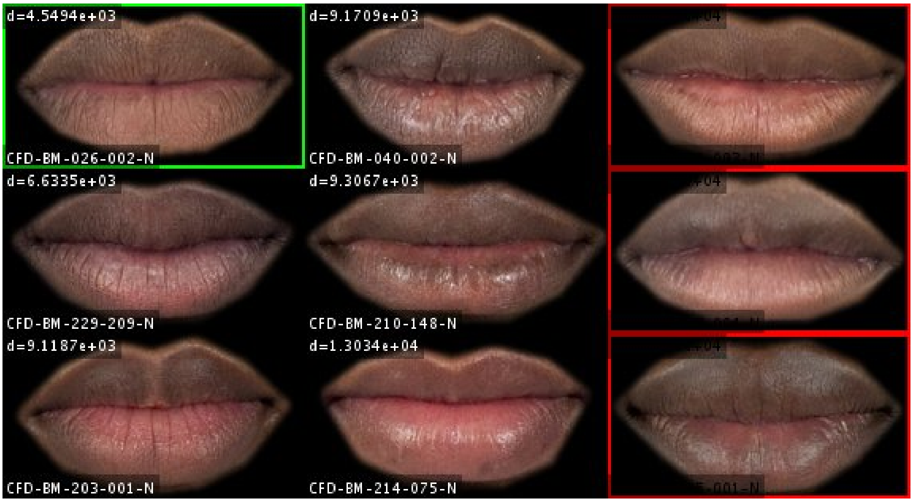


Figure B.27: Black mouths clustering example 3.

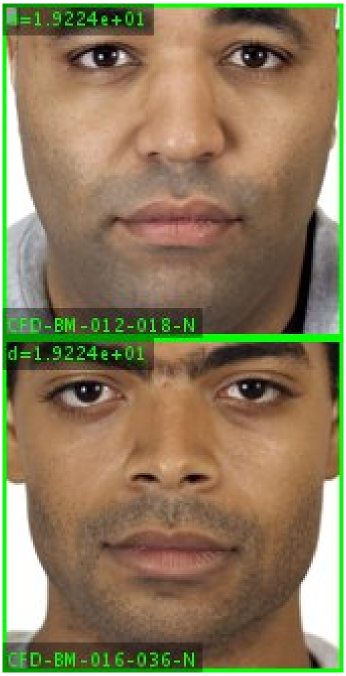


Figure B.28: Black jawlines clustering example 1.

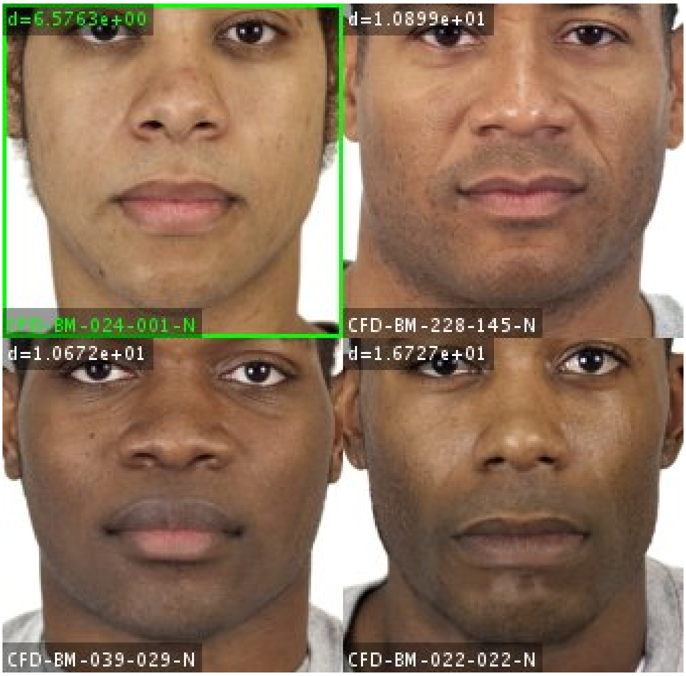


Figure B.29: Black jawlines clustering example 2.

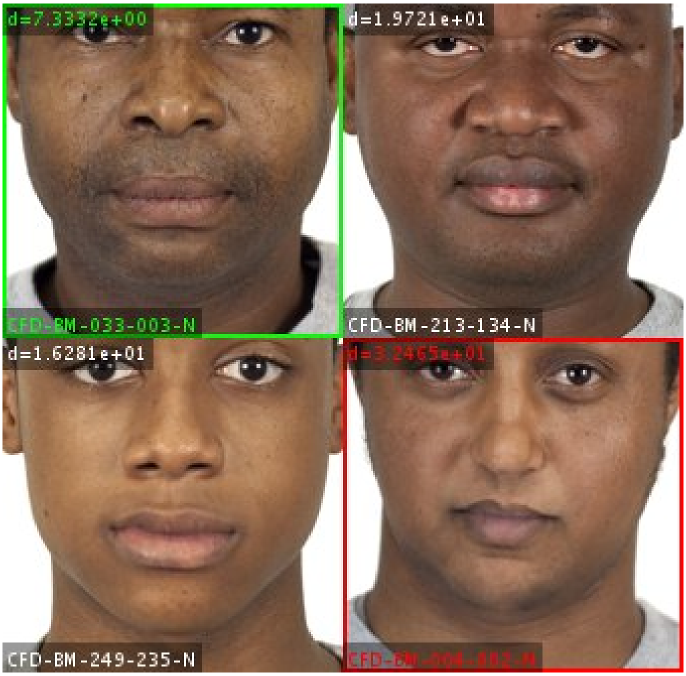


Figure B.30: Black jawlines clustering example 3.

B.3 Latino taxonomy

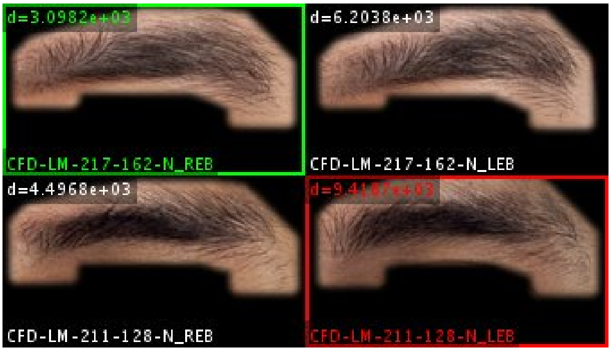


Figure B.31: Latino eyebrows clustering example 1.

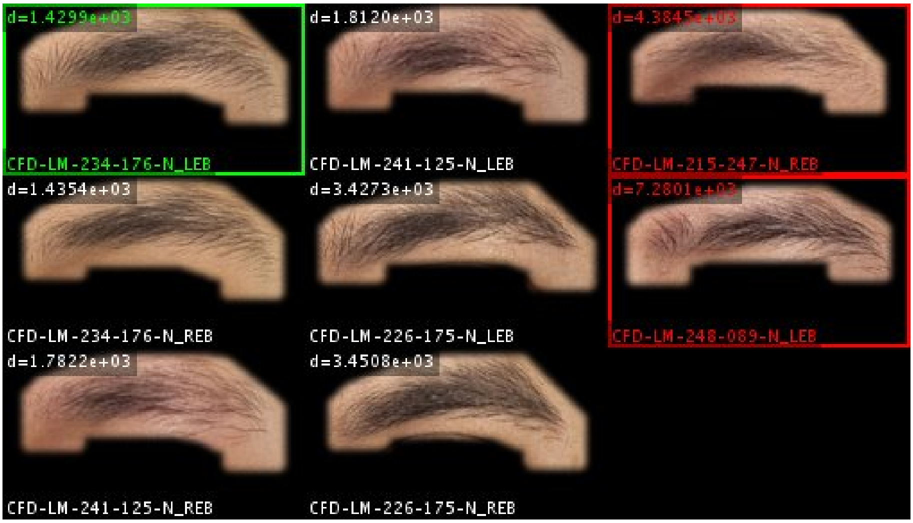


Figure B.32: Latino eyebrows clustering example 2.

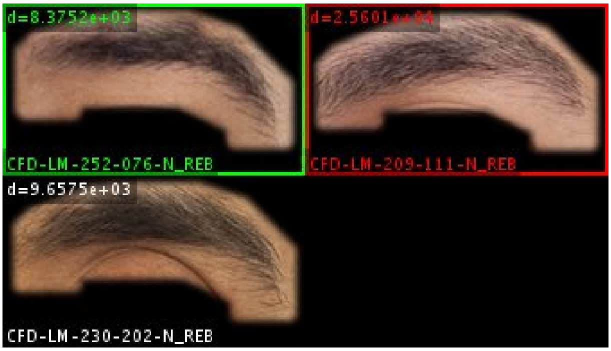


Figure B.33: Latino eyebrows clustering example 3.

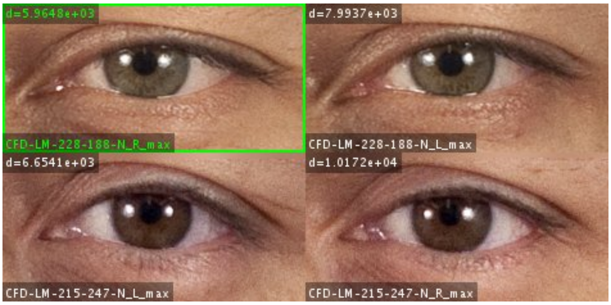


Figure B.34: Latino eyes clustering example 1.

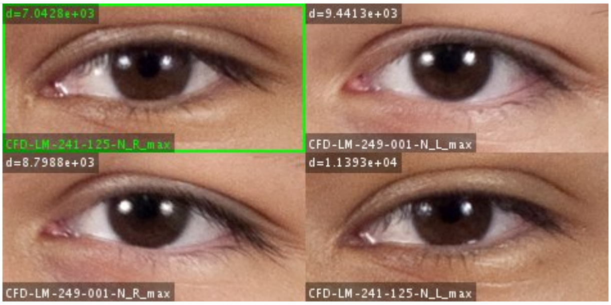


Figure B.35: Latino eyes clustering example 2.

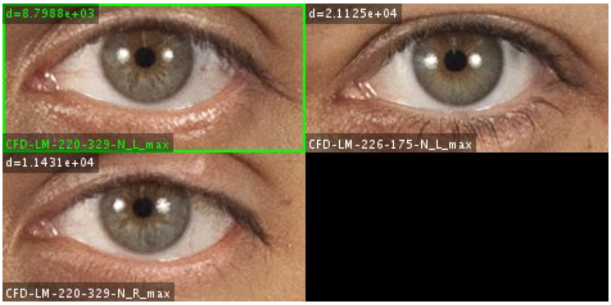


Figure B.36: Latino eyes clustering example 3.

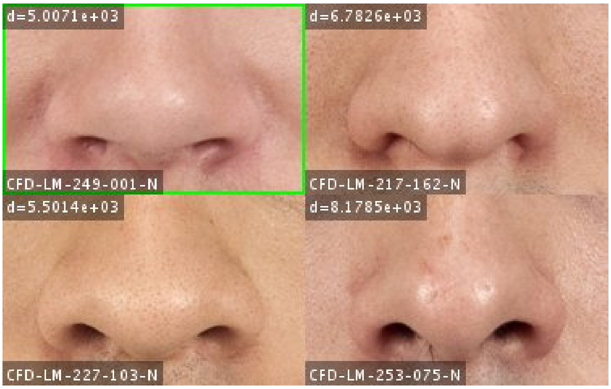


Figure B.37: Latino noses clustering example 1.

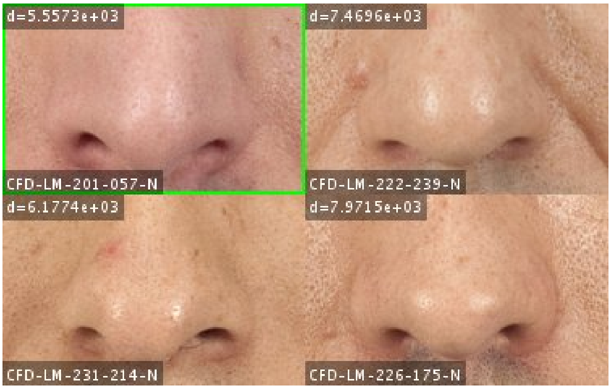


Figure B.38: Latino noses clustering example 2.

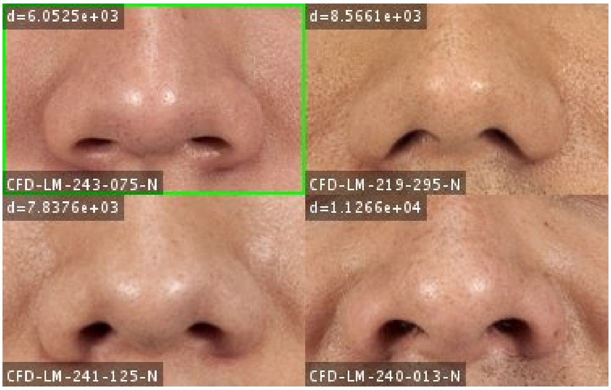


Figure B.39: Latino noses clustering example 3.

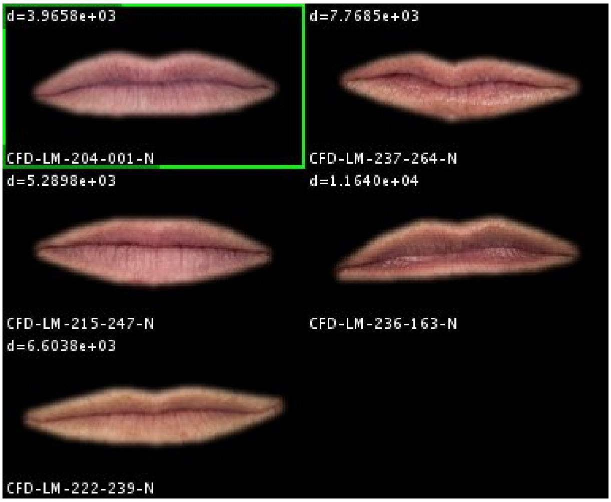


Figure B.40: Latino mouths clustering example 1.

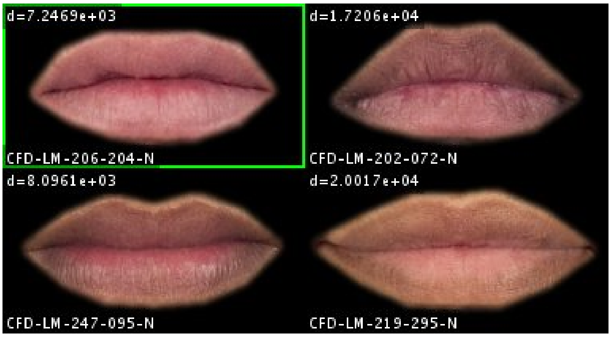


Figure B.41: Latino mouths clustering example 2.

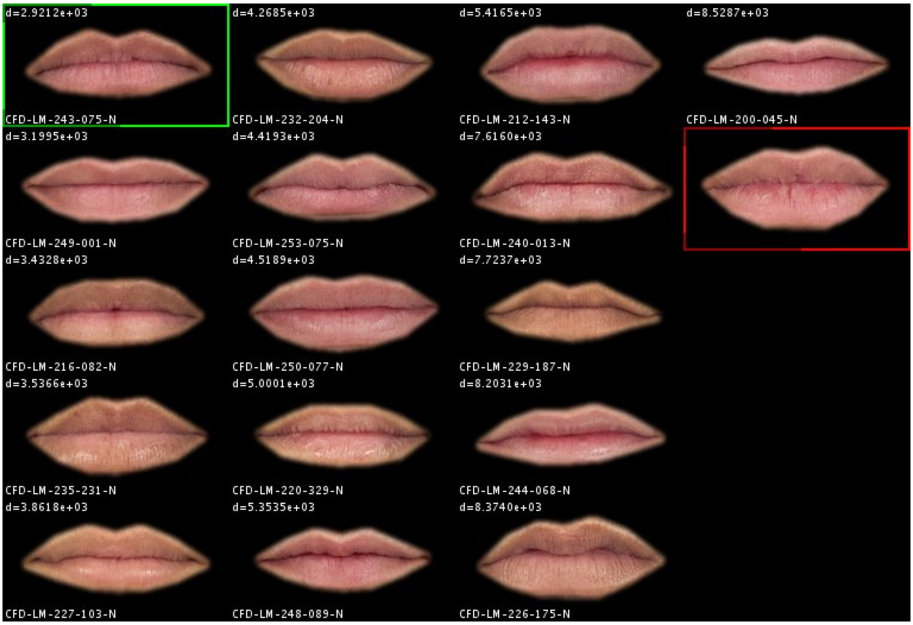


Figure B.42: Latino mouths clustering example 3.

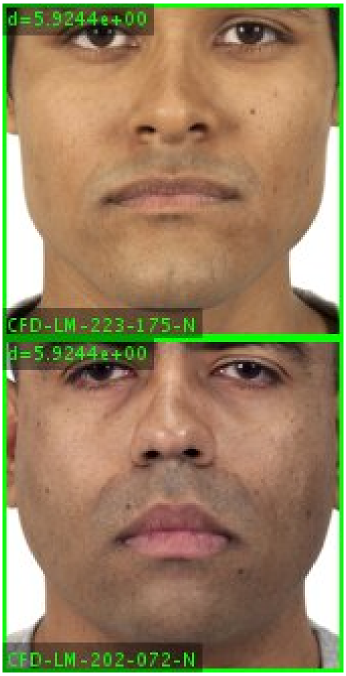


Figure B.43: Latino jawlines clustering example 1.

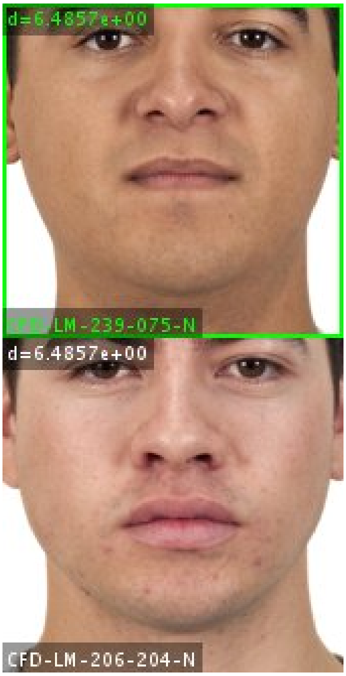


Figure B.44: Latino jawlines clustering example 2.

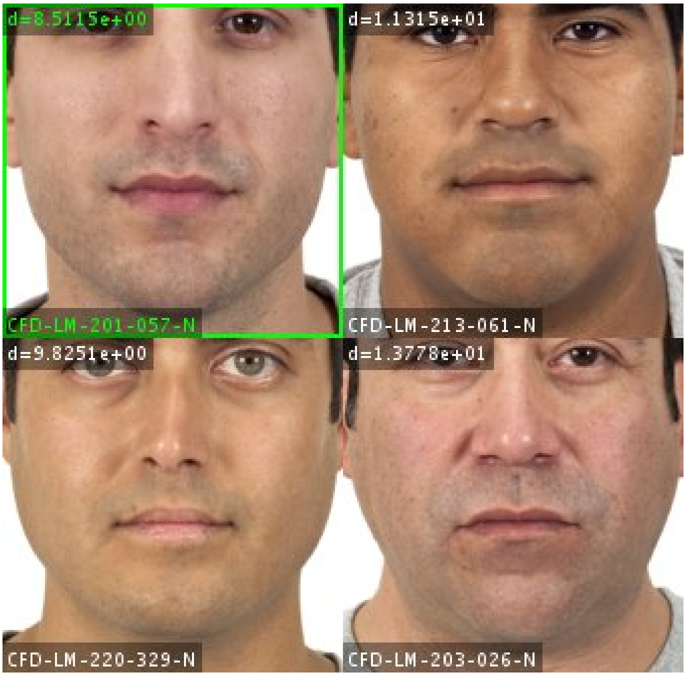


Figure B.45: Latino jawlines clustering example 3.

B.4 White taxonomy

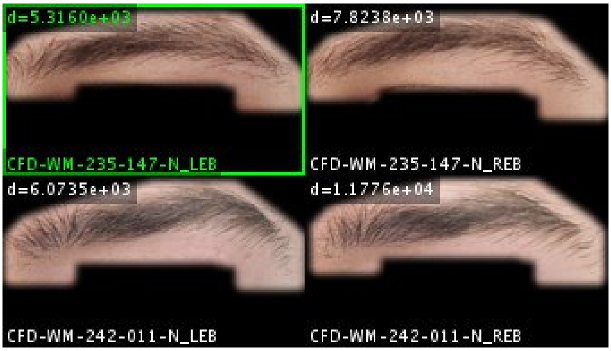


Figure B.46: White eyebrows clustering example 1.

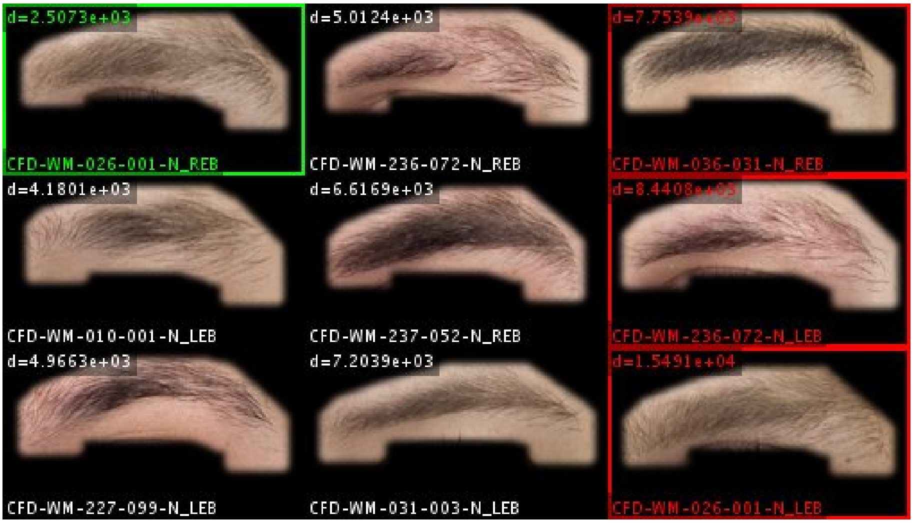


Figure B.47: White eyebrows clustering example 2.

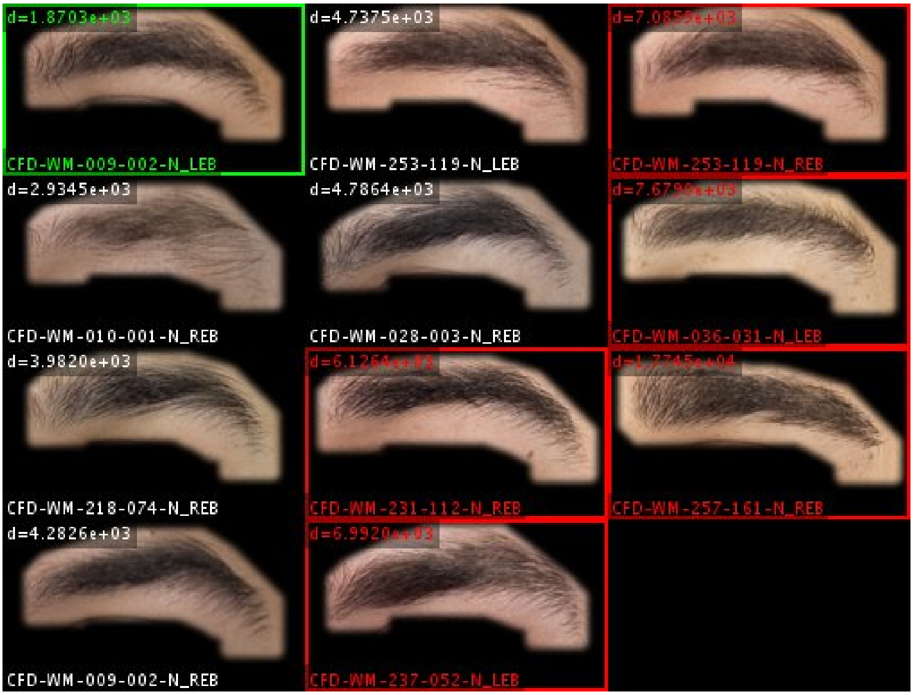


Figure B.48: White eyebrows clustering example 3.

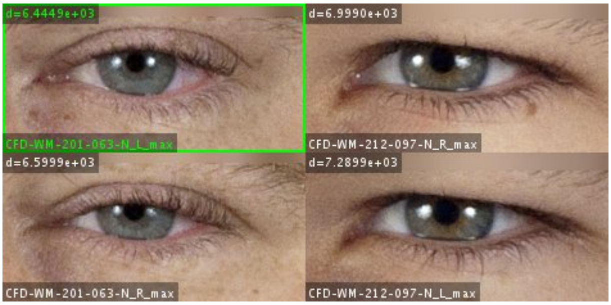


Figure B.49: White eyes clustering example 1.

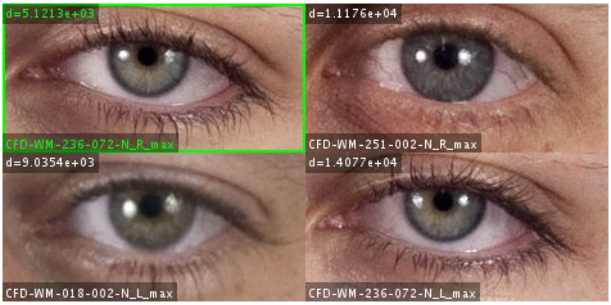


Figure B.50: White eyes clustering example 2.

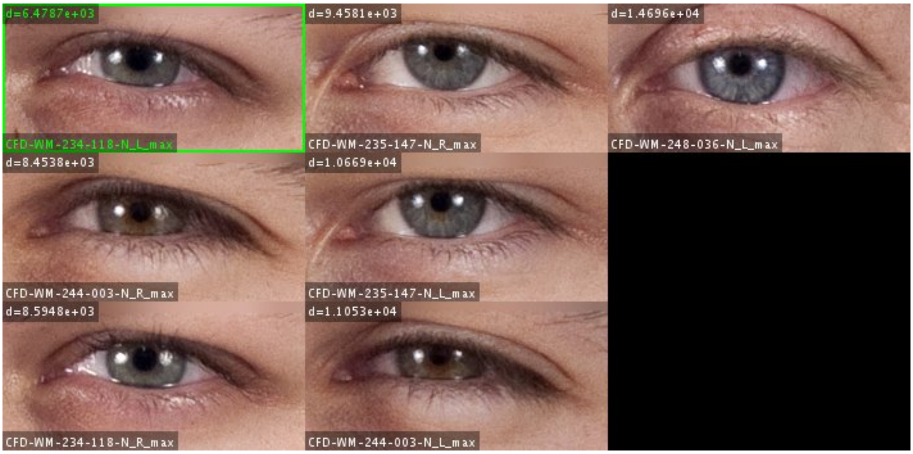


Figure B.51: White eyes clustering example 3.

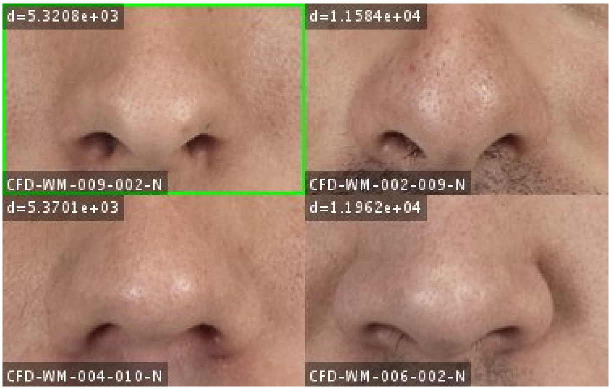


Figure B.52: White noses clustering example 1.

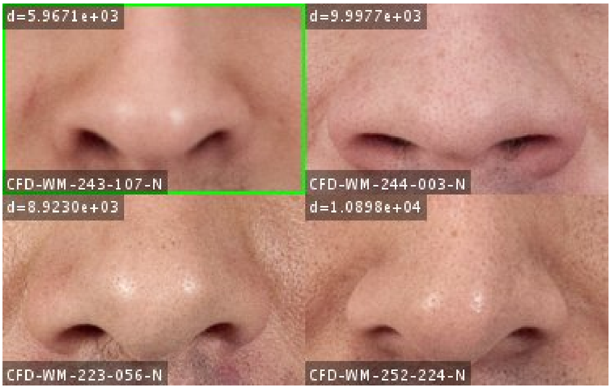


Figure B.53: White noses clustering example 2.

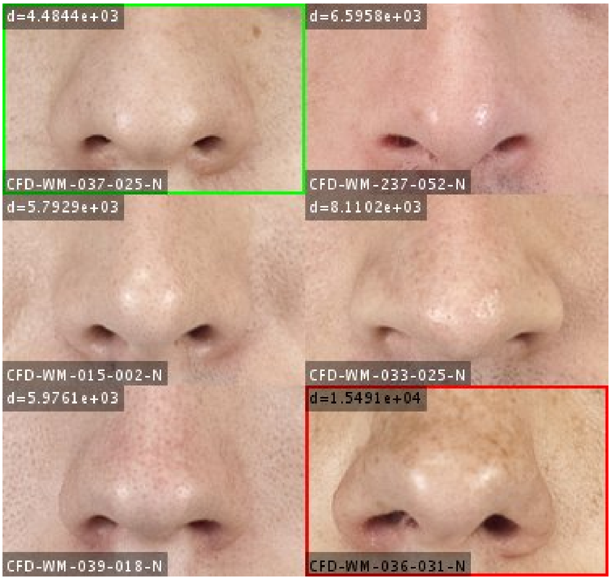


Figure B.54: White noses clustering example 3.

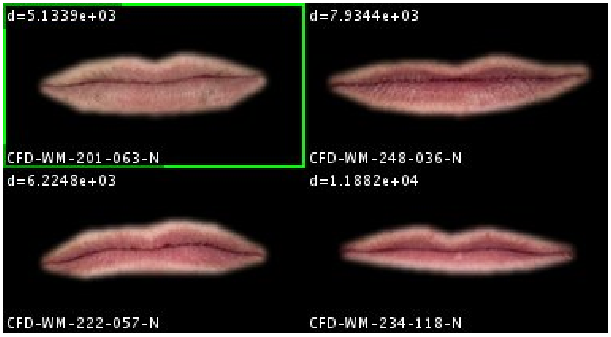


Figure B.55: White mouths clustering example 1.

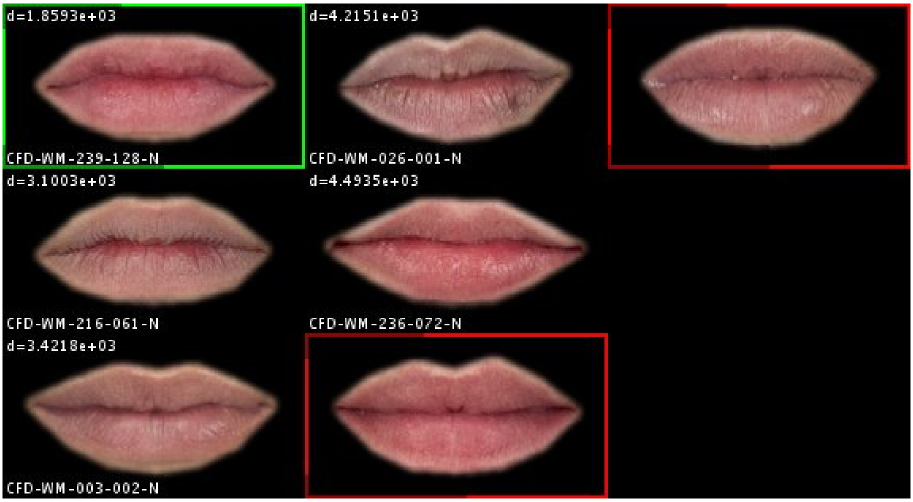


Figure B.56: White mouths clustering example 2.

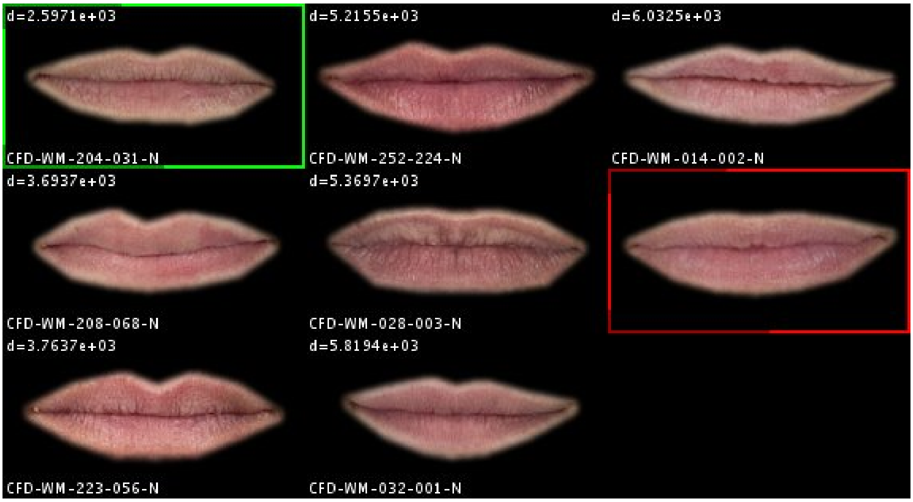


Figure B.57: White mouths clustering example 3.

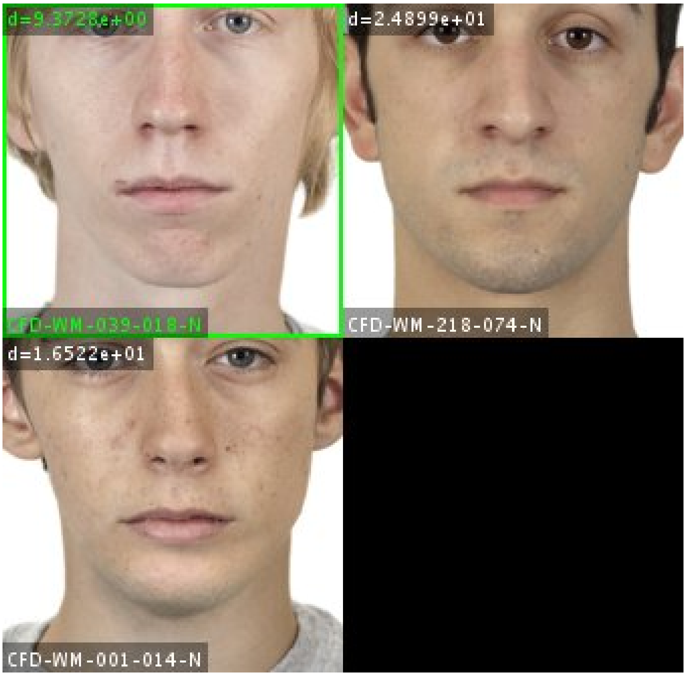


Figure B.58: White jawlines clustering example 1.

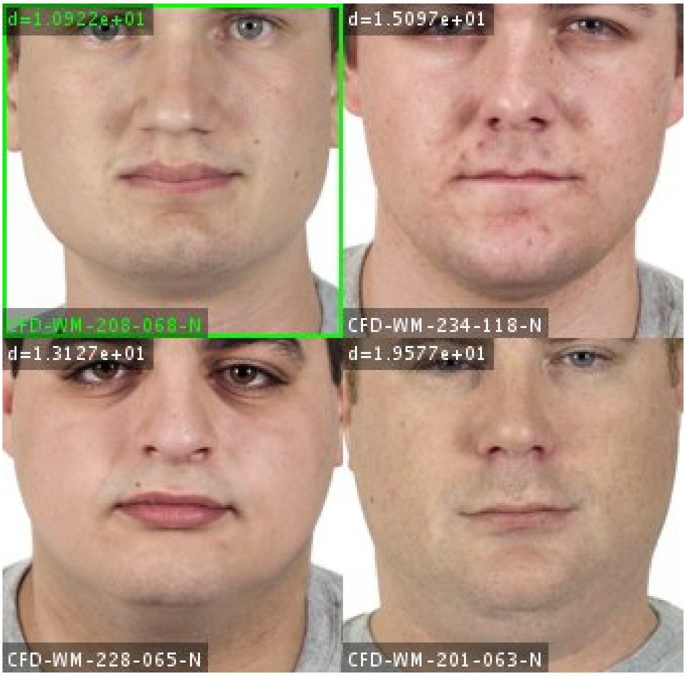


Figure B.59: White jawlines clustering example 2.

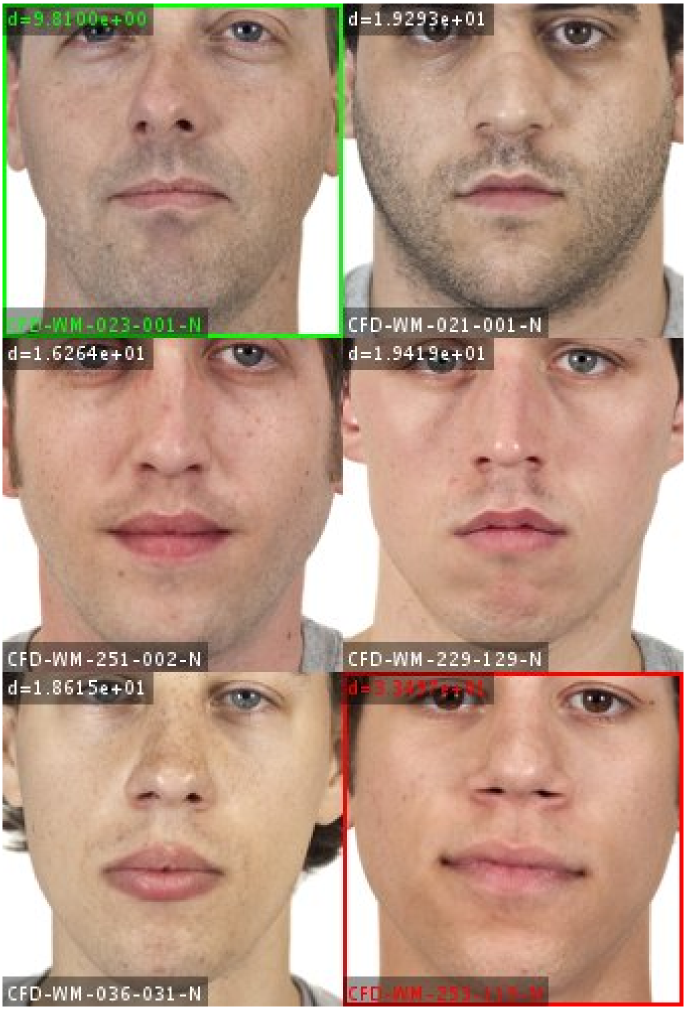


Figure B.60: White jawlines clustering example 3.

Bibliography

Parham Aarabi (2015). "Automatic segmentation of hair in images". In: *Multimedia (ISM), 2015 IEEE International Symposium on.* IEEE, pp. 69–72 (cit. on p. 36).

Timo Ahonen, Abdenour Hadid, and Matti Pietikainen (2006). "Face description with local binary patterns: Application to face recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 28.12, pp. 2037–2041 (cit. on p. 10).

Adrian Albin-Clark and Toby Howard (2009). "Automatically Generating Virtual Humans using Evolutionary Algorithms." In: *TPCG*, pp. 61–64 (cit. on p. 9).

Sandra Alemany, Juan Carlos González, Beatriz Nácher, Carol Soriano, Carlos Arnáiz, and H Heras (2010). "Anthropometric survey of the Spanish female population aimed at the apparel industry". In: *Proceedings of the 2010 International Conference on 3D Body Scanning Technologies. Lugano, Switzerland* (cit. on p. 8).

Olatz Arbelaitz, Ibai Gurrutxaga, Javier Muguerza, Jesús M Pérez, and Inigo Perona (2013). "An extensive comparative study of cluster validity indices". In: *Pattern Recognition* 46.1, pp. 243–256 (cit. on p. 22).

Seher Gündüz Arslan, Celal Genç, Bahadır Odabaş, and Jalen Devecioğlu Kama (2008). “Comparison of facial proportions and anthropometric norms among Turkish young adults with different face types”. In: *Aesthetic plastic surgery* 32.2, pp. 234–242 (cit. on p. 8).

Sabine Aßmann (2007). *Anthropological atlas of male facial features.* Verlag für Polizeiwissenschaft (cit. on pp. 9, 82).

Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic (2014). “Incremental face alignment in the wild”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1859–1866 (cit. on pp. 27, 37).

Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency (2013). “Constrained local neural fields for robust facial landmark detection in the wild”. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 354–361 (cit. on pp. 33, 34).

Domna Banakou, Konstantinos Chorianopoulos, and Kostas Anagnostou (2009). “Avatars’ appearance and social behavior in online virtual worlds”. In: *Informatics, 2009. PCI’09. 13th Panhellenic Conference on.* IEEE, pp. 207–211 (cit. on p. 2).

Moshe Bar, Maital Neta, and Heather Linz (2006). “Very first impressions.” In: *Emotion* 6.2, p. 269 (cit. on p. 104).

Boldur E Bărbat and Radu Cretulscu (2003). “Affordable affective avatars. Persuasion, emotions and language (s)”. In: *Proc. of the 1st Balkan Conference in Informatics (BCI’2003)* (cit. on p. 3).

Tara Behrend, Steven Toaddy, Lori Foster Thompson, and David J Sharek (2012). “The effects of avatar appearance on interviewer ratings in virtual employment interviews”. In: *Computers in Human Behavior* 28.6, pp. 2128–2133 (cit. on p. 2).

Peter N. Belhumeur, João P Hespanha, and David J. Kriegman (1997). “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”. In: *IEEE Transactions on pattern analysis and machine intelligence* 19.7, pp. 711–720 (cit. on p. 10).

James C Bezdek and Nikhil R Pal (1998). “Some new indexes of cluster validity”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 28.3, pp. 301–315 (cit. on p. 22).

Marion Boberg, Petri Piippo, and Elina Ollila (2008). “Designing avatars”. In: *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*. ACM, pp. 232–239 (cit. on p. 8).

Vicki Bruce and Andrew W Young (2012). *Face perception*. Psychology Press (cit. on p. 1).

Vicki Bruce and Andy Young (1986). “Understanding face recognition”. In: *British journal of psychology* 77.3, pp. 305–327 (cit. on p. 1).

Gavin Buckingham, Lisa M DeBruine, Anthony C Little, Lisa LM Welling, Claire A Conway, Bernard P Tiddeman, and Benedict C Jones (2006). “Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness”. In: *Evolution and Human Behavior* 27.5, pp. 381–389 (cit. on p. 8).

Andrew J Calder and Andrew W Young (2005). “Understanding the recognition of facial identity and facial expression”. In: *Nature Reviews Neuroscience* 6.8, pp. 641–651 (cit. on p. 88).

Paulo VR Carvalho, Isaac L dos Santos, Jose Orlando Gomes, Marcos RS Borges, and Stephanie Guerlain (2008). “Human factors approach for evaluation and redesign of human–system interfaces of a nuclear power plant simulator”. In: *Displays* 29.3, pp. 273–284 (cit. on p. 9).

Donghun Chung, Brahm Daniel deBuys, and Chang S Nam (2007). “Influence of avatar creation on attitude, empathy, presence, and para-social interaction”. In: *International Conference on Human-Computer Interaction*. Springer, pp. 711–720 (cit. on p. 2).

Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor (2001). “Active appearance models”. In: *IEEE Transactions on pattern analysis and machine intelligence* 23.6, pp. 681–685 (cit. on p. 10).

Timothy Cootes, Gareth J Edwards, Christopher J Taylor, et al. (1999). "Comparing active shape models with active appearance models." In: *Bmvc.* Vol. 99. 1, pp. 173–182 (cit. on p. 28).

Timothy Cootes and Christopher J Taylor (1992). "Active Shape Models-'smart snakes'." In: *BMVC.* Vol. 92, pp. 266–275 (cit. on p. 29).

Timothy Cootes, Christopher J Taylor, David H Cooper, and Jim Graham (1995). "Active shape models-their training and application". In: *Computer vision and image understanding* 61.1, pp. 38–59 (cit. on p. 26).

Antonio R Damasio (1985). "Prosopagnosia". In: *Trends in Neurosciences* 8, pp. 132–135 (cit. on p. 1).

Jules Davidoff and Nick Donnelly (1990). "Object superiority: A comparison of complete and part probes". In: *Acta psychologica* 73.3, pp. 225–243 (cit. on p. 9).

Alanah Davis, John Murphy, Dawn Owens, Deepak Khazanchi, and Ilze Zigurs (2009). "Avatars, people, and virtual worlds: Foundations for research in metaverses". In: *Journal of the Association for Information Systems* 10.2, p. 90 (cit. on p. 9).

Jose Antonio Diego-Mas and Jorge Alcaide-Marzal (2015). "A computer based system to design expressive avatars". In: *Computers in Human Behavior* 44, pp. 1–11 (cit. on p. 9).

Nick Donnelly and Jules Davidoff (1999). "The mental representations of faces and houses: Issues concerning parts and wholes". In: *Visual Cognition* 6.3-4, pp. 319–343 (cit. on p. 9).

Ron Dotsch and Alexander Todorov (2012). "Reverse correlating social face perception". In: *Social Psychological and Personality Science* 3.5, pp. 562–571 (cit. on pp. 3, 88).

Ron Dotsch, Daniël HJ Wigboldus, Oliver Langner, and Ad van Knippenberg (2008). "Ethnic out-group faces are biased in the prejudiced mind". In: *Psychological Science* 19.10, pp. 978–980 (cit. on p. 87).

Joseph C Dunn (1974). “Well-separated clusters and optimal fuzzy partitions”. In: *Journal of cybernetics* 4.1, pp. 95–104 (cit. on pp. 22, 60).

Marc Fabri, Salima Y Awad Elzouki, and David Moore (2007). “Emotionally expressive avatars for chatting, learning and therapeutic intervention”. In: *International Conference on Human-Computer Interaction.* Springer, pp. 275–285 (cit. on p. 9).

Marc Fabri and David Moore (2005). “The use of emotionally expressive avatars in collaborative virtual environments”. In: *Virtual Social Agents* 88 (cit. on pp. 2, 9).

Beat Fasel and Juergen Luettin (2003). “Automatic facial expression analysis: a survey”. In: *Pattern recognition* 36.1, pp. 259–275 (cit. on p. 25).

Verena Ferring and Hans Pancherz (2008). “Divine proportions in the growing face”. In: *American Journal of Orthodontics and Dentofacial Orthopedics* 134.4, pp. 472–479 (cit. on p. 8).

Francis Galton (1883). *Inquiries into the human faculty & its development.* JM Dent and Company (cit. on p. 86).

David E Goldberg (1989). “Genetic algorithms in search, optimization, and machine learning, 1989”. In: *Reading: Addison-Wesley* (cit. on p. 90).

Colin Goodall (1991). “Procrustes methods in the statistical analysis of shape”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339 (cit. on p. 29).

Wenzhangzhi Guo and Parham Aarabi (2016). “Hair segmentation using heuristically-trained neural networks”. In: *IEEE transactions on neural networks and learning systems* (cit. on p. 36).

BéAtrice S Hasler, Peleg Tuchman, and Doron Friedman (2013). “Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds”. In: *Computers in Human Behavior* 29.4, pp. 1608–1616 (cit. on p. 2).

Randy L Haupt and Sue Ellen Haupt (2004). *Practical genetic algorithms.* John Wiley & Sons (cit. on p. 94).

William G Hayward, Gillian Rhodes, and Adrian Schwaninger (2008). "An own-race advantage for components as well as configurations in face recognition". In: *Cognition* 106.2, pp. 1017–1027 (cit. on pp. 9, 78).

Geoffrey E Hinton and Sam T Roweis (2003). "Stochastic neighbor embedding". In: *Advances in neural information processing systems*, pp. 857–864 (cit. on pp. 23, 24).

Helle Hochscheid, Ronald Hamel, W Wootton, B Russell, and E Libonati (2015). "Shaping space: facial asymmetries in fifth-century Greek sculpture". In: *The Art of Making in Antiquity* (cit. on p. 8).

John H Holland (1975). "Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence." In: (cit. on p. 90).

Martin Holzwarth, Chris Janiszewski, and Marcus M Neumann (2006). "The influence of avatars on online consumer shopping behavior". In: *Journal of marketing* 70.4, pp. 19–36 (cit. on pp. 2, 3).

Singular Inversions (2008). "FaceGen modeller (Version 3.3)[computer software]". In: *Toronto, ON: Singular Inversions* (cit. on p. 110).

Anil K Jain (1989). *Fundamentals of digital image processing.* Prentice-Hall, Inc. (cit. on p. 12).

Soo-chan Jee and Myung Hwan Yun (2016). "An anthropometric survey of Korean hand and hand shape types". In: *International Journal of Industrial Ergonomics* 53, pp. 10–18 (cit. on p. 8).

Daniel Kahneman (2003). "A perspective on judgment and choice: mapping bounded rationality." In: *American psychologist* 58.9, p. 697 (cit. on p. 86).

Nancy Kanwisher, Josh McDermott, and Marvin M Chun (1997). "The fusiform face area: a module in human extrastriate cortex specialized for face perception". In: *Journal of neuroscience* 17.11, pp. 4302–4311 (cit. on p. 1).

Minyoung Kim, Sanjiv Kumar, Vladimir Pavlovic, and Henry Rowley (2008). "Face tracking and recognition with visual constraints in real-world videos". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.* IEEE, pp. 1–8 (cit. on p. 25).

Nam-Soon Kim and Wol-Hee Do (2014). "Classification of Elderly Womens Foot Type". In: *Journal of the Korean Society of Clothing and Textiles* 38.3, pp. 305–320 (cit. on p. 8).

Brendan Klare and Anil K Jain (2010). "On a taxonomy of facial features". In: *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on.* IEEE, pp. 1–8 (cit. on p. 10).

Kwang-Eun Ko and Kwee-Bo Sim (2010). "Development of a Facial Emotion Recognition Method based on combining AAM with DBN". In: *Cyberworlds (CW), 2010 International Conference on.* IEEE, pp. 87–91 (cit. on p. 25).

M Koleva, A Nacheva, and M Boev (2002). "Somatotype and disease prevalence in adults". In: *Reviews on environmental health* 17.1, pp. 65–84 (cit. on p. 8).

Evdokimos I Konstantinidis, Magda Hitoglou-Antoniadou, Andrej Luneski, Panagiotis D Bamidis, and Maria M Nikolaidou (2009). "Using affective avatars and rich multimedia content for education of children with autism". In: *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments.* ACM, p. 58 (cit. on p. 2).

Panayiotis Koutsabasis, Spyros Vosinakis, Katerina Malisova, and Nikos Paparounas (2012). "On the value of virtual worlds for collaborative design". In: *Design Studies* 33.4, pp. 357–390 (cit. on p. 2).

Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar (2009). "Attribute and simile classifiers for face verification". In: *Computer Vision, 2009 IEEE 12th International Conference on.* IEEE, pp. 365–372 (cit. on p. 25).

Johann Caspar Lavater (1800). *Essays on physiognomy: for the promotion of the knowledge and the love of mankind; written in the German language* *by JC Lavater, abridged from Mr. Holcrofts translation.* printed for GGJ & J. Robinson (cit. on p. 86).

Rensis Likert (1932). “A technique for the measurement of attitudes.” In: *Archives of psychology* (cit. on p. 35).

Hsin Lin and Hua Wang (2014). “Avatar creation in virtual worlds: Behaviors and motivations”. In: *Computers in Human Behavior* 34, pp. 213–218 (cit. on p. 2).

Ya-Lih Lin and Kun-Lung Lee (1999). “Investigation of anthropometry basis grouping technique for subject classification”. In: *Ergonomics* 42.10, pp. 1311–1316 (cit. on p. 8).

Anthony C Little, Robert P Burriss, Benedict C Jones, and S Craig Roberts (2007). “Facial appearance affects voting decisions”. In: *Evolution and Human Behavior* 28.1, pp. 18–27 (cit. on p. 1).

Cesare Lombroso (2006). “Criminal Man, translated and with a new introduction by Mary Gibson and Nicole Hahn Rafter”. In: *Durham, NC: Duke University Press.(Original work published 1876 and 1897)* (cit. on p. 86).

Bruce D Lucas (1986). “Generalized image matching by the method of differences.” In: (cit. on p. 32).

Bruce D Lucas, Takeo Kanade, et al. (1981). “An iterative image registration technique with an application to stereo vision”. In: (cit. on p. 32).

Ping Luo, Xiaogang Wang, and Xiaoou Tang (2013). “A deep sum-product architecture for robust facial attributes analysis”. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2864–2871 (cit. on p. 25).

Debbie S Ma, Joshua Correll, and Bernd Wittenbrink (2015). “The Chicago face database: A free stimulus set of faces and norming data”. In: *Behavior research methods* 47.4, pp. 1122–1135 (cit. on p. 34).

Laurens van der Maaten and Geoffrey Hinton (2008). “Visualizing data using t-SNE”. In: *Journal of Machine Learning Research* 9.Nov, pp. 2579–2605 (cit. on pp. 22, 26).

David JC MacKay (2003). *Information theory, inference and learning algorithms.* Cambridge university press (cit. on pp. 18, 59).

Grigoris G Malousaris, Nikolaos K Bergeles, Karolina G Barzouka, Ioannis A Bayios, George P Nassis, and Maria D Koskolou (2008). “Somatotype, size and body composition of competitive female volleyball players”. In: *Journal of science and medicine in sport* 11.3, pp. 337–344 (cit. on p. 8).

Deepa R Mane, Alka D Kale, Manjula B Bhai, and Seema Hallikerimath (2010). “Anthropometric and anthroposcopic analysis of different shapes of faces in group of Indian population: A pilot study”. In: *Journal of forensic and legal medicine* 17.8, pp. 421–425 (cit. on p. 8).

Michael C Mangini and Irving Biederman (2004). “Making the ineffable explicit: Estimating the information employed for face classifications”. In: *Cognitive Science* 28.2, pp. 209–226 (cit. on pp. 87, 88).

Angie Lorena Marin and Sukhan Lee (2013). “Interaction Design for Robotic Avatars Does Avatar’s Aging Cue Affect the User’s Impressions of a Robot?” In: *International Conference on Universal Access in Human-Computer Interaction.* Springer, pp. 373–382 (cit. on p. 2).

Myosotis Massidda, Stefania Toselli, Patricia Brasili, and Carla M Calo (2013). “Somatotype of elite Italian gymnasts”. In: *Collegium antropologicum* 37.3, pp. 853–857 (cit. on p. 8).

Iain Matthews and Simon Baker (2004). “Active appearance models revisited”. In: *International journal of computer vision* 60.2, pp. 135–164 (cit. on pp. 29–31).

Ethan Meyers and Lior Wolf (2008). “Using biologically inspired features for face processing”. In: *International Journal of Computer Vision* 76.1, pp. 93–104 (cit. on p. 10).

George A Miller (1956). “The magical number seven, plus or minus two: some limits on our capacity for processing information.” In: *Psychological review* 63.2, p. 81 (cit. on pp. 10, 78).

Kristine L Nowak and Christian Rauh (2008). “Choose your “buddy icon” carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions”. In: *Computers in Human Behavior* 24.4, pp. 1473–1493 (cit. on p. 2).

S. Ohlrogge (2009). *Anthropological Atlas of Female Facial Features*. Schriftenreihe angewandte forensische Anthropologie. Verlag für Polizeiwissenschaft Prof. Dr. Clemens Lorei. ISBN: 9783866760721. URL: `https://books.google.es/books?id=3ltqPgAACAAJ` (cit. on p. 9).

Christopher Y Olivola, Friederike Funk, and Alexander Todorov (2014). “Social attributions from faces bias human choices”. In: *Trends in Cognitive Sciences* 18.11, pp. 566–570 (cit. on p. 86).

Nikolaas N Oosterhof and Alexander Todorov (2008). “The functional basis of face evaluation”. In: *Proceedings of the National Academy of Sciences* 105.32, pp. 11087–11092 (cit. on p. 88).

V Orvalho, J Miranda, and AA Sousa (2009). “Facial Synthesys of 3D Avatars for Therapeutic Applications.” In: *Studies in health technology and informatics* 144, p. 96 (cit. on p. 9).

Patrick Pérez, Michel Gangnet, and Andrew Blake (2003). “Poisson image editing”. In: *ACM Transactions on graphics (TOG)*. Vol. 22. 3. ACM, pp. 313–318 (cit. on pp. 97, 98, 102, 103, 114, 144).

Katherine S Pollard and Mark J Van Der Laan (2002). “A method to identify significant clusters in gene expression data”. In: (cit. on p. 22).

Jennifer Parker Porter and Krista L Olson (2001). “Anthropometric facial analysis of the African American woman”. In: *Archives of facial plastic surgery* 3.3, pp. 191–197 (cit. on p. 8).

TA Preston and Mohan Singh (1972). “Redintegrated somatotyping”. In: *Ergonomics* 15.6, pp. 693–700 (cit. on p. 8).

Gillian Rhodes, Louise Ewing, William G Hayward, Daphne Maurer, Catherine J Mondloch, and James W Tanaka (2009). “Contact and other-race effects in configural and component processing of faces”. In: *British Journal of Psychology* 100.4, pp. 717–728 (cit. on pp. 9, 78).

Jennifer J Richler, Olivia S Cheung, and Isabel Gauthier (2011). “Holistic processing predicts face recognition”. In: *Psychological science* 22.4, pp. 464–471 (cit. on p. 9).

Stefanie Ritz-Timme et al. (2011). “A new atlas for the evaluation of facial features: advantages, limits, and applicability”. In: *International journal of legal medicine* 125.2, pp. 301–306 (cit. on pp. 8, 9, 78, 82).

S Ritz-Timme et al. (2011). “Metric and morphological assessment of facial features: a study on three European populations”. In: *Forensic science international* 207.1, 239–e1 (cit. on p. 8).

Gay Robins (1984). “Analysis of Facial Proportions in Egyptian Art”. In: *Gottinger Miszellen Gottingen* 79, pp. 31–41 (cit. on p. 8).

Mario Rojas, David Masip, Alexander Todorov, and Jordi Vitria (2011). “Automatic prediction of facial trait judgments: Appearance vs. structural models”. In: *PloS one* 6.8, e23323 (cit. on pp. 8, 10, 11).

Peter J. Rousseeuw (1987). “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics* 20.Supplement C, pp. 53–65. ISSN: 0377-0427. DOI: `https://doi.org/10.1016/0377-0427(87)90125-7`. URL: `http://www.sciencedirect.com/science/article/pii/0377042787901257` (cit. on pp. 20, 21).

Nicholas Rule and Nalini Ambady (2010). “First impressions of the face: Predicting success”. In: *Social and Personality Psychology Compass* 4.8, pp. 506–516 (cit. on p. 86).

Christopher P Said, Nicu Sebe, and Alexander Todorov (2009). “Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces.” In: *Emotion* 9.2, p. 260 (cit. on p. 128).

Christopher P Said and Alexander Todorov (2011). “A statistical model of facial attractiveness”. In: *Psychological Science* 22.9, pp. 1183–1190 (cit. on p. 88).

Isabel M Santos and Andrew W Young (2011). “Inferring social attributes from different face regions: Evidence for holistic processing”. In: *The Quarterly Journal of Experimental Psychology* 64.4, pp. 751–766 (cit. on p. 144).

Jason M Saragih, Simon Lucey, and Jeffrey F Cohn (2011). “Deformable model fitting by regularized landmark mean-shift”. In: *International Journal of Computer Vision* 91.2, pp. 200–215 (cit. on pp. 32, 33).

Pornthep Sarakon, Theekapun Charoenpong, and Supiya Charoensiriwath (2014). “Face shape classification from 3D human data by using SVM”. In: *Biomedical Engineering International Conference (BMEiCON), 2014 7th*. IEEE, pp. 1–5 (cit. on p. 8).

Alec Scharff, John Palmer, and Cathleen M Moore (2011). “Evidence of fixed capacity in visual object categorization”. In: *Psychonomic bulletin & review* 18.4, pp. 713–721 (cit. on pp. 10, 78).

Alize Scheenstra, Arnout Ruifrok, and Remco Veltkamp (2005). “A survey of 3D face recognition methods”. In: *Audio-and Video-Based Biometric Person Authentication*. Springer, pp. 325–345 (cit. on p. 25).

Ulrike Schultze (2010). “Embodiment and presence in virtual worlds: a review”. In: *Journal of Information Technology* 25.4, pp. 434–449 (cit. on p. 2).

William Herbert Sheldon (1954). *Atlas of men: A guide for somatotyping the adult male at all ages*. Harper (cit. on p. 8).

Jiazheng Shi, Ashok Samal, and David Marx (2006). “How effective are landmarks and their geometry for face recognition?” In: *Computer vision and image understanding* 102.2, pp. 117–133 (cit. on p. 25).

Lawrence Sirovich and Michael Kirby (1987). “Low-dimensional procedure for the characterization of human faces”. In: *Josa a* 4.3, pp. 519–524 (cit. on p. 58).

Pierre Soille (2013). *Morphological image analysis: principles and applications.* Springer Science & Business Media (cit. on pp. 11, 12).

Pratibha Sukhija, Sunny Behal, and Pritpal Singh (2016). “Face recognition system using genetic algorithm”. In: *Procedia Computer Science* 85, pp. 410–417 (cit. on p. 9).

Clare AM Sutherland, Julian A Oldmeadow, Isabel M Santos, John Towler, D Michael Burt, and Andrew W Young (2013). “Social inferences from faces: Ambient images generate a three-dimensional model”. In: *Cognition* 127.1, pp. 105–118 (cit. on p. 3).

Clare AM Sutherland, Andrew W Young, and Gillian Rhodes (2017). “Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties”. In: *British Journal of Psychology* 108.2, pp. 397–415 (cit. on p. 128).

Abraham Tamir (2011). “Numerical survey of the different shapes of the human nose”. In: *Journal of Craniofacial Surgery* 22.3, pp. 1104–1107 (cit. on p. 9).

— (2013). “Numerical Survey of the Different Shapes of Human Chin”. In: *Journal of Craniofacial Surgery* 24.5, pp. 1657–1659 (cit. on p. 9).

James W Tanaka and Martha J Farah (1993). “Parts and wholes in face recognition”. In: *The Quarterly journal of experimental psychology* 46.2, pp. 225–245 (cit. on p. 9).

Jessica Taubert, Deborah Apthorp, David Aagten-Murphy, and David Alais (2011). “The role of holistic processing in face perception: Evidence from the face inversion effect”. In: *Vision research* 51.11, pp. 1273–1278 (cit. on pp. 9, 78).

Peter BM Thomas, Tadas Baltrušaitis, Peter Robinson, and Anthony J Vivian (2016). “The Cambridge Face Tracker: Accurate, Low Cost Measurement of Head Posture Using Computer Vision and Face Recognition Software”. In: *Translational vision science & technology* 5.5 (cit. on p. 37).

Ying-Li Tian, Takeo Kanade, and Jeffrey F Cohn (2005). "Handbook of face recognition". In: *Ch Facial Expression Analysis; Springer: Berlin/Heidelberg, Germany*, pp. 487–519 (cit. on p. 8).

Alexander Todorov (2011). "Evaluating faces on social dimensions". In: *Social neuroscience: Toward understanding the underpinnings of the social mind*, pp. 54–76 (cit. on p. 1).

Alexander Todorov, Ron Dotsch, Daniel HJ Wigboldus, and Chris P Said (2011). "Data-driven methods for modeling social perception". In: *Social and Personality Psychology Compass* 5.10, pp. 775–791 (cit. on pp. 86–88).

Alexander Todorov, Christopher Y Olivola, Ron Dotsch, and Peter Mende-Siedlecki (2015). "Social attributions from faces: Determinants, consequences, accuracy, and functional significance". In: *Annual Review of Psychology* 66 (cit. on p. 3).

Alexander Todorov and Jenny M Porter (2014). "Misleading first impressions: Different for different facial images of the same person". In: *Psychological science* 25.7, pp. 1404–1417 (cit. on p. 128).

Klaus D Toennies (2012). "Guide to medical image analysis". In: *Methods and Algorithms* (cit. on p. 13).

Tomas Trescak, Anton Bogdanovych, Simeon Simoff, and Inmaculada Rodriguez (2012). "Generating diverse ethnic groups with genetic algorithms". In: *Proceedings of the 18th ACM symposium on Virtual reality software and technology.* ACM, pp. 1–8 (cit. on p. 9).

Matthew Turk and Alex Pentland (1991). "Eigenfaces for recognition". In: *Journal of cognitive neuroscience* 3.1, pp. 71–86 (cit. on pp. 10, 18).

Laurens Van Der Maaten (2014). "Accelerating t-SNE using tree-based algorithms." In: *Journal of machine learning research* 15.1, pp. 3221–3245 (cit. on p. 25).

Pater Vanezis, D Lu, J Cockburn, A Gonzalez, G McCombe, O Trujillo, and M Vanezis (1996). "Morphological classification of facial features in adult

Caucasian males based on an assessment of photographs of 50 subjects". In: *Journal of Forensic Science* 41.5, pp. 786–791 (cit. on p. 9).

Richard JW Vernon, Clare AM Sutherland, Andrew W Young, and Tom Hartley (2014). "Modeling first impressions from highly variable facial images". In: *Proceedings of the National Academy of Sciences* 111.32, E3353–E3361 (cit. on p. 3).

Guillermo Vinué, Irene Epifanio, and Sandra Alemany (2015). "Archetypoids: A new approach to define representative archetypal data". In: *Computational Statistics & Data Analysis* 87, pp. 102–115 (cit. on p. 8).

Paul Viola and Michael J Jones (2004). "Robust real-time face detection". In: *International journal of computer vision* 57.2, pp. 137–154 (cit. on p. 25).

Max Vladymyrov and Miguel A Carreira-Perpinán (2013). "Entropic Affinities: Properties and Efficient Numerical Computation." In: *ICML (3)*, pp. 477–485 (cit. on p. 23).

Mirella Walker, Fang Jiang, Thomas Vetter, and Sabine Sczesny (2011). "Universals and cultural differences in forming personality trait judgments from faces". In: *Social Psychological and Personality Science* 2.6, pp. 609–617 (cit. on p. 35).

Mirella Walker and Thomas Vetter (2009). "Portraits made to measure: Manipulating social judgments about individuals with a statistical face model". In: *Journal of Vision* 9.11, pp. 12–12 (cit. on pp. 1, 3, 88, 129, 141).

Dan Wang, Xiujuan Chai, Hongming Zhang, Hong Chang, Wei Zeng, and Shiguang Shan (2011). "A novel coarse-to-fine hair segmentation method". In: *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.* IEEE, pp. 233–238 (cit. on p. 36).

Nannan Wang, Xinbo Gao, Dacheng Tao, Heng Yang, and Xuelong Li (2017). "Facial feature point detection: A comprehensive survey". In: *Neurocomputing* (cit. on pp. 27, 29).

Ruosi Wang, Jingguang Li, Huizhen Fang, Moqian Tian, and Jia Liu (2012). "Individual differences in holistic processing predict face recognition ability". In: *Psychological science* 23.2, pp. 169–177 (cit. on p. 9).

David Weibel, Daniel Stricker, Bartholomäus Wissmath, and Fred W Mast (2010). "How socially relevant visual characteristics of avatars influence impression formation". In: *Journal of Media Psychology* (cit. on p. 2).

Darrell Whitley (1994). "A genetic algorithm tutorial". In: *Statistics and computing* 4.2, pp. 65–85 (cit. on p. 90).

BK Wiederhold and G Riva (2009). "Facial synthesys of 3D avatars for therapeutic applications". In: *Annual Review of Cybertherapy and Telemedicine 2009: Advanced Technologies in the Behavioral, Social and Neurosciences* 144, p. 96 (cit. on p. 2).

George Wolberg (1998). "Image morphing: a survey". In: *The visual computer* 14.8, pp. 360–372 (cit. on p. 25).

Nick Yee and Jeremy Bailenson (2007). "The Proteus effect: The effect of transformed self-representation on behavior". In: *Human communication research* 33.3, pp. 271–290 (cit. on pp. 2, 9).

Leslie A Zebrowitz and Joann M Montepare (2008). "Social psychological face perception: Why appearance matters". In: *Social and personality psychology compass* 2.3, pp. 1497–1517 (cit. on p. 86).

Xiangxin Zhu and Deva Ramanan (2012). "Face detection, pose estimation, and landmark localization in the wild". In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE, pp. 2879–2886 (cit. on p. 25).