Application of Augmented Reality to the recognition of historical figures portraits

Eugenio P. Concepcion

Master en aplicaciones Multimedia

Universitat Oberta de Catalunya

A thesis submitted for the degree of

Master in Multimedia

2014 June

1. Reviewer:

2. Reviewer:

Day of the defense:

Signature from head of PhD committee:

Abstract

The following work is centred on analysing the diverse face recognition techniques, in order to use them in an Augmented Reality application for recognizing historical figures in paintings. One of the main objectives of this project will be finding the most suitable recognition techniques for this purpose.

To my wife and my daughters

Acknowledgements

I would like to acknowledge all the people that have contribute for making this work possible, especially my family and my tutor, Enric Guaus.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Introduction

Although the development of face recognition techniques dates back to the 1970s, it has been a few years ago when their practical aspects seem to have been discovered. There are lots of applications and devices taking advantage of this technology, ranging from airport security cameras to smartphones applications for image processing. However, only a few of these techniques have been applied in the educational or edutainment context. On the other hand, Augmented Reality is a quite recent technology based on concepts developed a far as fifty years ago. Its main contribution has been to provide a natural way of combining real world images and video with virtual data. Thanks especially to smartphones and tablets, Augmented Reality has experienced a quick growth. The present project aims to combine these two concepts for providing the visitors of a museum a new experience, giving them a tool for discovering information about the historical figure in the painting they are watching.

1. INTRODUCTION

2

Proposal

2.1 Justification of topic interest

The objective of the present project is the determination of the most suitable facial recognition technique for identifying historical figures in paintings. This would be used in combination with Augmented Reality for developing an application for the visitors of a museum, who will be able to get detailed biographical information just by targeting the painting with a smartphone. The coming of Augmented Reality technology into the world of mobile application has brought a new catalogue of tools for education and cultural guidance. A mobile device can be used by visitors to discover scientific facts about every object exhibited. Moreover, these technologies not only provide an indubitably simple way of getting information, but also make unnecessary traditional means of data displaying like sign boards or graphics. Apart from the edutainment application of this research project, there is also a very interesting point for face recognition techniques. Most of actual techniques have been developed for and applied exclusively to real photographs. There is still a promising field of research to be developed related to painting analysis and recognition. The recognition of historical figures is just one of the possible applications that can be carried out. For example, a more mature image-based analysis can be performed for detecting forgeries, or for identifying the author.

2.2 State of the Art

In the last years, facial recognition has become increasingly important and the number of applications and devices that use it is growing up (1). Facial recognition offers several advantages over other biometric recognition techniques such as fingerprints or iris recognition. Besides being more natural and non-intrusive, its main advantage is that it can be applied remotely and discretely.

2.2.1 Historical context

The first documented system implementing facial recognition was developed by Takeo Kanade for his thesis (2). This first approach was followed by a period of inactivity, until the publication of the works of Sirovich and Kirby (3) (4). They proved that it is possible to analyze face images by means of Karhunen-Loeve transformation (also known as Principal Component Analysis, PCA) applied to a training collection of images. This technique performs a PCA analysis over a collection of face images to generate a set of features that can be used later to recognize faces. These feature sets were named eigenpictures (due to the existing relationship between these elements and the concept of eigenvectors o eigenvalues). The eigenpictures can be combined lineally to rebuild the original training set. Some time later, a pioneer work developed by Turk and Pentland about eigenfaces constituted a major step in the matter (5). Some other later developments were Fisherfaces method (6) (7) that applies Linear Discriminant Analysis (LDA) to enhance the accuracy of an initial processing stage based on PCA; or the application of local filters like Gabor Jets (8).

2.2.2 Recognition process

Habitually, every face recognition system follows a similar processing model (9). This model is composed of a sequence of stages:

- Face and features detection
- Face normalization, in which images are normalized geometrically and photometrically. This step is necessary for being able to operate under diverse illumination and pose conditions.
- Feature extraction.
- Feature comparison, in which detected features are compared with those features registered in the systems knowledge base.

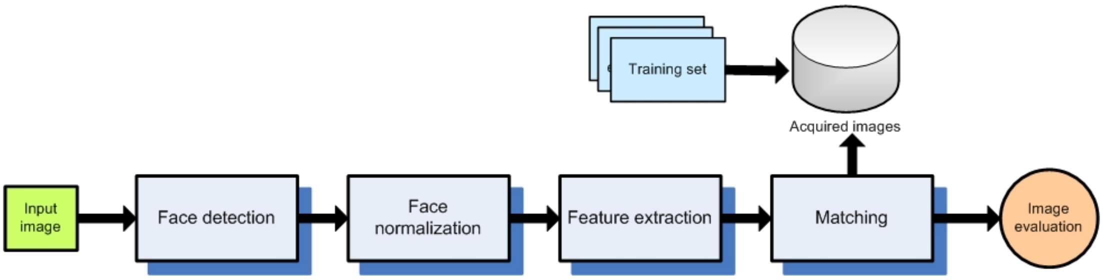


Figure 2.1: General overview of recognition process - Main stages in face recognition

2.2.3 Challenges and difficulties

In facial recognition, performance and effectiveness of the process depends strongly on external conditions such as illumination, pose, gestures, age, hair, facial wearing, movement, etc. Considering these factors, facial recognition conditions can be classified according to users' cooperation (1). So, there are cooperative scenarios in which users are willing to cooperate, looking for favourable conditions such as good illumination, gesture suppression, etc. Opposite to this, there are non-cooperative scenarios such as surveillance systems where users are not aware of being identified. Apart from these conditions, that are subject dependant, there are other difficulties related to the image capturing system itself (occlusion, exposure, lens aberration, etc.). All these issues affect every image processed by the system. Therefore, there is a great variability having influence in data extraction that hampers face identification.

2.3 Face recognition techniques

There are several face recognition techniques, but they can be classified in two main categories: image-based techniques and feature-based techniques. This classification is based on the way the images are analyzed, globally or focused on certain characteristics. The figure 2.2 shows a complete classification of existing techniques.

2.3.1 Feature-based techniques

Feature-based techniques use certain features that can be identified consistently in the different images to perform the facial recognition, instead of analysing intensity or the colours of the pixels from the area in which the face is, as image-based techniques do.

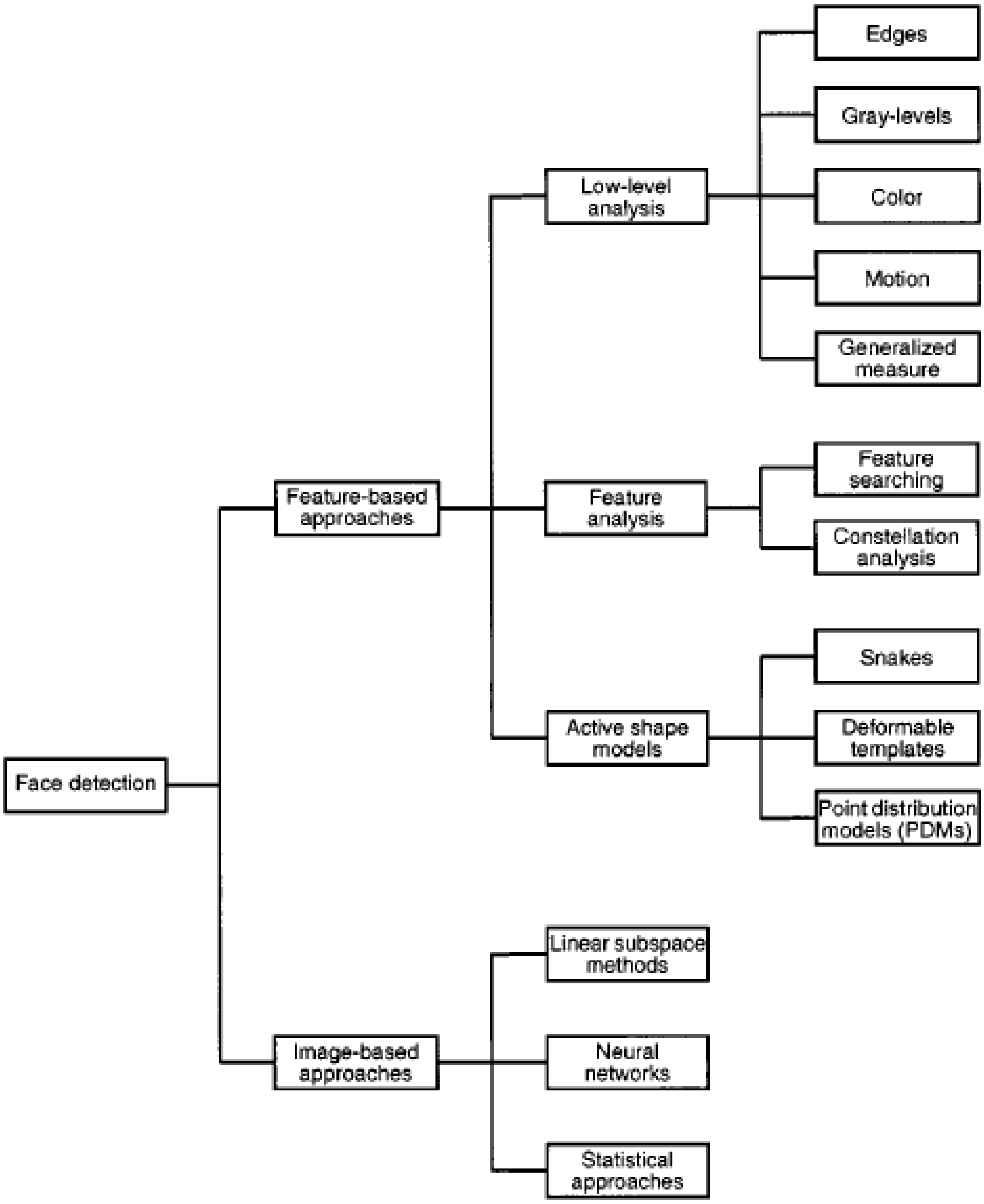


Figure 2.2: Classification of existing techniques - Taken from (10)

Some of these features include the centres of the eyes, the eyebrow, the line of the lips or the chin. In the same way as pixels where analysed by image-based techniques, features can be modelled statistically. These data can be analysed by means of covariant analysis. Feature-based techniques can be grouped in the categories: low-level analysis, feature analysis and active shape analysis. Low-level analysis is centred on resolving the common face detection problem: face detection in an image with mixed background elements. In feature analysis technique, visual features are organized in a more global concept, extracting facial features and using geometric information from faces. By means of features analysis ambiguities are reduced and the position of the face and its features are easier to determine. Finally, active shape models (ASM) encompass several techniques. These techniques have been defined to extract complex elements and non-rigid features like pupils or lips tracking (10).

2.3.1.1 Feature analysis

Feature analysis is a high level technique developed for overcoming the difficulties associated with a low level analysis of an image. For example, in a low level approach the system would try to locate similar regions in a face by means of a colour model of the skin. However, certain objects in the background could also be recognized as faces if their colour fit with skin model. This false positive can be avoided applying a higher level analysis (10). Many face recognition techniques take advantage of face geometrical data for building and later verifying those features subject of ambiguity. There are two main approaches to face geometry (10). The first strategy performs a sequential search of features, being based on the relative position of individual features. The certainty of having correctly identified a particular feature is supported by the relative position of nearest features. The second approach consists of grouping the features in flexible constellations, using several facial models. The constellation model is a probabilistic generative model employed in artificial vision for recognizing categories of objects. This model tries to represent object classes by means of a set of N parts defined under mutual geometrical restrictions. Features search-based techniques usually start by determining the most prominent facial features. Once these features have been detected, and after having taken advantage of known anthropometric relations in the face, it is possible to define hypothesis about the rest of features. One example of this strategy is the algorithm proposed by De Silva (10). It begins by establishing a hypothesis about the start position of the head, to later go over the face from up to down, up to locate the eye line. The distance between eye line and the top of the head is taken as a reference measure. This dimension can be used later as flexible template for calculating the remaining features. The process of generating the template also takes into account biometric data from a knowledge base. This template is flexible, and will be adjusted according to the location of the whole set of features.

2.3.1.2 Active Shape Models

Active Shape Models (ASM) technique was developed by Tim Cootes and Chris Taylor in 1995 (11). It is based on using a Point Distribution Model (PDM) for representing the average geometry of a shape, and applies statistical models of geometric variation inferred from a training sample. A PDM consists of a set of distinctive feature locations. This PDM captures variations in the shape of faces, such as the general size, and main features such as eyes and lips. The larger the variety inside the training sample is, the larger the set of feature points to be considered is. Intuitively, an ASM is a statistical model of the shape of an object, which is iteratively modified to fit to an example of the object in a new image. These shapes are constrained by the PDM to vary only in a range determined by a training set of labelled examples. The shape of an object is represented by a set of points (controlled by the shape model). The ASM algorithm aims to match the model to a new image. The fit process in an ASM consists of the PDM initialized in its average shape and scaled and rotated to a standard position for detection. After that, an iterative process takes place until convergence:

- Search in every point surrounding area for the best location of this point in the local appearance model.
- Restrict new points to a plausible shape.

Convergence is reached when:

- The number of complete iterations has reached a certain threshold
- The percentage of moved points is lower than a fraction of search distance in the last iteration

2.3.1.3 Active Appearance Model

The Active Appearance Model (AAM) is a generalization of the Active Shape Model (ASM). An AAM can be defined as an integrated statistical model which combines a model of shape variation with a model of the appearance variations in a shape-normalized frame. In an AAM there has been built a statistical model of the shape and a grey-level representation of the target in a way that can be generalized to almost any sample. The process of matching an image involves finding the model parameters which minimize the difference between the image and the synthesized model example projected into the image. This model uses all the information of the surrounding area instead of just analysing nearest points or the borders. In the same way as ASM, AAM training process requires the corresponding points of a PDM to be marked in a face set. However, there is an important difference between AMM and ASM. Instead of updating PDM using local points' searches that will be later restricted by the PDM, the effects in the appearance due to the changes in the parameters of the model will be learnt. Learning process associates changes in the parameters by projecting errors in the ASM. The fit process implies an initialization as described. The model is re-projected in the image and the resulting calculated difference. This error is used later to update model parameters, and these parameters are restricted to realistic values. The process iterates until error is below a certain threshold.

2.3.2 Image-based techniques

This category contains techniques which exclusively employ the pixel intensity or the colour within an area where the face has been detected for determining if the face belongs to a learnt data set. Generally speaking, these techniques usually assume the area where the face is located has already been detected and the input image is normalized (lighting and pose), and it is reduced to a fixed size area (NxM pixels). These techniques make use of a strong statistical support usually related to the concept of supervised classification. An individual is classified within a group according to the information of a set of variables previously observed in a set of individuals who are known to be correctly classified in a certain group, in this case originating from a training sample. As mentioned before, techniques based on face features show some difficulties, such as unpredictability of the face appearance and the environmental conditions. Despite these techniques advances have improved the behaviour facing these problems, they are still limited to frontal images of the face and shoulders at best. Techniques able to work in adverse conditions are still needed, such us the detection in environments with many people or images with a distinction between background-vague shape (10) (12). These and other needs have inspired the development of the set of techniques analysed in this point. In this view, face detection is part of a general problem of pattern identification. Stating the problem as a question of learning and pattern recognition based on samples, the use of a specific knowledge of the face is eliminated. Thus, the potential error during modelling due to an incomplete or imprecise knowledge is avoided.

2.3.2.1 The space of faces and its dimensionality

The computerized analysis of face images must interpret a visual signal, understood in terms of light reflected on the surface of a face and registered by a digital sensor as an array of pixels (1). Those pixels contain coded information about colours, or just light intensity (grey scale image). After normalization and size adjustment to MxN dimensions, the array of pixels can be considered a point (vector) in an MxN-dimensional space of images, simply because it is written in a certain fixed order. The usual problem in the analysis of multidimensional data is dimensionality, that is, the number of coordinates necessary to specify a point of data. To specify an arbitrary image in the space of images, it is necessary to state every pixel forming it. Thus, the nominal dimension of the space is determined by the representation of a pixel and its value is MxN, a really large value, even for small size images. Recognition methods using this kind of representation must work with this complexity, commonly known as the dimensionality curse (1). Managing multi-dimensional samples, especially in the context of image similarity and pattern detection, is mostly burdensome, computationally speaking. In the case of parametric methods, the number of parameters to be estimated typically grows exponentially with the dimensionality. Usually, this value is far higher than the number of images available to the training, making the estimation task harder. Similarly, for non-parametric methods, the sample complexity, this is, the number of samples needed to represent the underlying distribution, is unaffordable high. Nevertheless, there is an advantage: most of the face surface is smooth and shows a regular texture. Consequently, a sampling of individual pixels is unnecessarily dense, and a pixel value can be correlated to those of surrounding pixels. Moreover, the appearance of the faces can be easily fitted into a model: a front view of a face is basically symmetric, showing eyes in both sides, a nose in the centre and other similar characteristics. Thus, a not so high proportion of the image points give relevant information about the face. These natural restrictions force face images to be confined in a subspace named face space.

The general procedure of this technique consists of an initial training in which samples provided to the system are classified as prototypes of faces or non-faces. The comparison between the classification generated from these learning and the information extracted from the input image lead the system to make a decision about a face.

2.3.2.2 Eigenfaces

The problem with the image representation in image-based approaches is its high dimensionality. Two-dimensional pxq grayscale images span a pxq-dimensional vector space, so an image with 100x100 pixels lies in a 10,000-dimensional image space already. The question is: Are all dimensions equally relevant? The answer depends on if there is any variance in data. In this approach, the challenge is to look for those components that account for most of the information. The Eigenfaces algorithm focuses mainly on decodifying face images, identifying a set of local and global features. These features are not necessarily related to facial features like eyes, lips or nose (1). This technique would be applied in a context in which all available images are front views of faces, and the conditions of lightning and pose are clearly favourable. Despite of this limitation, there are variations of this technique that tolerate certain degree of variability (1). The original Eigenfaces algorithm was developed by Kirby and Sirovich (3) (4). They were focused on getting a more efficient representation of faces applying Principal Component Analysis (PCA). The Principal Component Analysis was derived from Karhunen-Loeve's transformation by Harold Hotelling in 1933 as of previous works of Karl Pearson (13), to turn a set of possibly correlated variables into a smaller set of uncorrelated variables. The basic idea can be explained as follows: given a high-dimensional data set (like a training images set), it can be described by a subset of correlated variables. This means simply that just considering few meaningful dimensions is enough for representing most of the information. The PCA method finds the directions with the greatest variance in the data, called principal components. Thanks to this information, it is feasible to decide which elements must be preserved because of they are meaningful (5). The result of applying these techniques is a vector of processed images. These images, or eigenfaces, can be linearly combined for reconstructing the original training set images. From a mathematical point of view, this objective is discovering the major components of the face distribution studied. In other words, it aims to calculate the eigenvectors of the covariance matrix corresponding the set of face images. This approach deals with every image as a point (or vector) in an N-dimensional space. The eigenvectors are sorted according to the amount of variation between images of faces. These eigenvectors form a set of features that, taken together, characterize the variation between face images. Let be a sample containing M images, PCA would generate a set of N eigenfaces, where N<M. The error during reconstruction can be reduced by increasing the number of eigenfaces, although this number will ever be lower than M. In every set of eigenfaces, each image of the original sample can be expressed as a linear combination of eigenfaces. For example, let f be the resulting image, then it can be calculated from the set E of eigenfaces: $\mathrm{f} = \mathrm{w}_1\mathrm{e}_1 + \mathrm{w}_2\mathrm{e}_2 + ... + \mathrm{w}_N\mathrm{e}_N$.

2.3.2.3 Fisherfaces

The Principal Component Analysis, which is the core of the Eigenfaces method, finds a linear combination of features that maximizes the total variance in data. While this is clearly a powerful way to represent data, it doesnt consider any classes and so a lot of discriminative information may be lost when throwing components away. Imagine a situation where the variance in your data is generated by an external source, let it be the light. The components identified by a PCA do not necessarily contain any discriminative information at all, so the projected samples are smeared together and a classification becomes impossible. The Eigenfaces method described before took a holistic approach to face recognition: A facial image is a point from a high-dimensional image space and a lower-dimensional representation is found, where classification becomes easy. The lower-dimensional subspace is found with Principal Component Analysis, which identifies the axes with maximum variance. While this kind of transformation is optimal from a reconstruction standpoint, it doesnt take any class labels into account. In a context where the variance is generated from external sources, for example light, the axes with maximum variance do not necessarily contain any discriminative information at all; hence a classification becomes impossible (6). A class-specific projection with a Linear Discriminant Analysis was applied to face recognition in Fisherfaces. The basic idea is to minimize the variance within a class, while maximizing the variance between the classes at the same time. The Linear Discriminant Analysis (LDA) performs a class-specific dimensionality reduction and was developed by the statistician R. A. Fisher. He initially applied it for solving taxonomic problems when there are multiple measurements. Applied to face recognition, this technique aids finding the combination of features that separates best between classes, because it maximizes the ratio of between-classes to within-classes scatter, instead of maximizing the overall scatter. The idea is simple: same classes should cluster tightly together, while different classes are as far away as possible from each other in the lower-dimensional representation.

2.3.2.4 Neural networks

Neural networks have become a commonly used technique in pattern recognition (10). In the specific case of face recognition these techniques are suitable to being applied in all processing stages (14) (15). The first approaches to facial recognition using neural networks date back to the 1990s, and were based on a multilayer perceptron (16) (17). These models gathered promising results, even starting with reduced datasets as entries. Rowley was the first one to identify the limitations of this approach and develop a solution (18). He worked with a set of images that were pre-processed for lightning adjust and histogram equalization. The problem with this method was overlapping detections, and the proposed solution applies two different heuristics (10):

- Thresholding: the number of detections in the immediate surrounding region the current location is counted, and if it is above a certain threshold, a face is present at this location.
- Overlap elimination: when a region has been classified as a face according to prior criterion, then overlapping detections are likely to be false positives and thus are rejected.

In order to improve the performance of the process, different neural networks were trained and their outputs were combined applying a discriminative strategy (AND, OR, polling, or even an independent neural network). Some authors (19) have proposed new strategies based on constrained generative models (CGM). A CGM is a type of multilayer auto-associative perceptron fully connected that uses three layers of weights. A multilayer perceptron is a neural network structured in layers so that it can solve non-linearly separable problems. This perceptron can be totally or locally connected. In the first case, every output of layer i is the input of every neuron in layer i+1; in the other case, every output of layer i is just the input of a subset of neurons in layer i+1. The objective of this approach was applying a non-linear PCA by forcing the projection of samples containing no faces similar to those that contain faces, and make a decision using error analysis.

2.3.2.5 Statistical approaches

Apart from the explained methods, there should be considered other approaches to face detection. Statistical approaches consider systems based on different concepts such as information theory, support vector machine and Bayesian decision rules (1). Statistical methods are in the same core of every face recognition system, they can be considered as the most common approach in the learning stage of a face recognition system. However, several authors have proposed face recognition solutions that also apply a statistical strategy for detecting or recognizing faces. Based on their prior work of maximum likelihood face detection (20), Colmenarez and Huang proposed a system based on Kullback divergence. This divergence is defined as the non-negative measure of the difference between two probability density functions. The proposed system creates a join-histogram for each pair of pixels in the training set and then this histogram is used to create probability functions for the classes of faces and non-faces.

2.3.3 Advantages and disadvantages

Table 1 summarizes the advantages and disadvantages of the studied techniques.

As stated, both groups have advantages and disadvantages. In the particular case of image-based techniques, most of the algorithms have been enhanced to compensate their main disadvantage, that is, variations processing. Globally, all these modifications have brought a better generation of algorithms, bringing the best performance results in FERET evaluations (1). Moreover, Probabilistic Eigenfaces, Fisherfaces and EBGM have been classified as the three best face recognition techniques. Despite the fact that EGBM is a feature-based algorithm, its success is mainly due to the application at feature-level of image-based techniques in a neural network.

Family	Advantage	Disadvantage
Image-based techniques	The main advantage is that there is no information loosing due to focusing only in certain areas or potentially interesting features (21).	The main disadvantage of these techniques is precisely their main feature: all of them presuppose that all the pixels of the image are equally important (22). Consequently, these techniques are computationally expensive and require a high degree of correlation between the images to be tested and the training set. They are also more sensitive with pose, scale and lightning variations than feature-based techniques (Beumier y Acheroy, 2000).
Feature-based techniques	The main advantage of these techniques is their robustness to pose variations (10). Feature-based schemes tend to be invariant to size or illumination changes. Another advantage of these models is the high compactness of face representation and high matching speed.	The main disadvantage is related to the fact that for performing automatic features detection it is necessary to make first an arbitrary decision about which features are important for performing the recognition. In general, these techniques do not offer such discrimination ability neither a further processing able to compensate this intrinsic weakness.

Table 2.1: Comparison of face recognition techniques - Overview of advantages and disadvantages.

2.3.4 Evaluation technologies

Face recognition from still frontal images has made great strides over the last twenty years. Over this period, error rates have decreased by three orders of magnitude when recognizing frontal face in still images taken with consistent controlled illumination in an environment similar to a studio (1). Under these conditions, error rates below 1% according to the Face Recognition Vendor Test (FRVT) 2006 and the Multiple Biometric Evaluation (MBE) 2010 (1). These tests are based on the two main test protocols for face recognition: FERET y FRVT 2002. Both protocols have been applied to perform FRVT 2006 y MBE 2010. The FRVT 2002 protocol was designed to evaluate biometrical analysis algorithms, not only face recognition, and computing a wide range of performance statistics. This includes the standard performance tasks of open-set and closed-set identification, and verification. It also allows any other techniques such as similarity score normalization, performance statistics variability measuring, and covariate analysis. In face recognition, and also biometrics, performance is usually worked out using three standard measurements: verification, open-set and closed-set identification. Each one has its own set of performance measures. All three tasks are closely related, and open-set identification is considered the general case (1). Measuring algorithm performance requires three sets of images. The first is a gallery G, which contains images of the people known to the system. The other two are probe sets. A probe is a biometric sample that is presented to the system for recognition, where recognition can be verification or identification. The first probe set is PG that contains biometric samples of people in a gallery (these samples are different from the samples in the gallery). The other probe set is PN, which contains biometric samples of people that are not in a gallery. Closed-set identification is a common performance measure where the question to be answered is: whose face is this? This question is meaningful for closed-set identification, since the biometric sample in a probe is always someone in the gallery. Thus, the general case of closed-set identification is open-set identification. In open-set identification, the face in the probe does not have to belong to the original training set. In this case, the basic question to be answered is: do we know this face? So, the system has to decide if the probe contains an image of a person in the gallery, and if so, report the identity of that person.

2.4 Augmented Reality

The Augmented Reality (AR) can be defined as an enhanced vision of the world by adding computer generated virtual elements (23). It can also be defined more formally as a real-time, direct or indirect, view of a physical real-world environment that has been enhanced or augmented by adding virtual computer-generated information to it (24). The objective of AR is to create an interactive vision, which combines real objects with virtual objects. The purpose of this interaction is to increase the information perceived by a user by providing him virtual data combined with reality. This reality is not forced to be originated just in the users environment, it can be generated indirectly (like a video received by streaming) and then virtual information is composed.

2.4.1 Augmented Reality Displays

There are three major types of displays used in Augmented Reality (24): head-mounted displays, handheld displays and spatial displays.

2.4.1.1 Head-mounted displays

A head-mounted display (HMD) is a display device that can be worn on the head, independently or as part of a helmet. It combines images from the real world with virtual elements, creating an enhanced environment over the user's view of the world. HMD can be monocular or binocular, providing a more or less integrated vision of reality. One of the most popular appliances of this kind is the Google Glass project.

2.4.1.2 Handheld displays

Handheld displays are devised as small computing devices with a display that the user can hold in their hands. They overlay virtual graphics onto the real environment and employ sensors, such as digital compasses and GPS units for offering a natural behaviour to the users. There are currently three distinct classes of commercially available handheld displays that are being used for augmented reality system: smartphones, PDAs and tablets. Smartphones are both portable and widespread, and in the last years they have been provided with powerful CPU, camera, accelerometer, GPS, and solid state compass, making them a quite convenient platform for AR. However, their small display size is less than ideal for 3D user interfaces. PDAs present much of the same

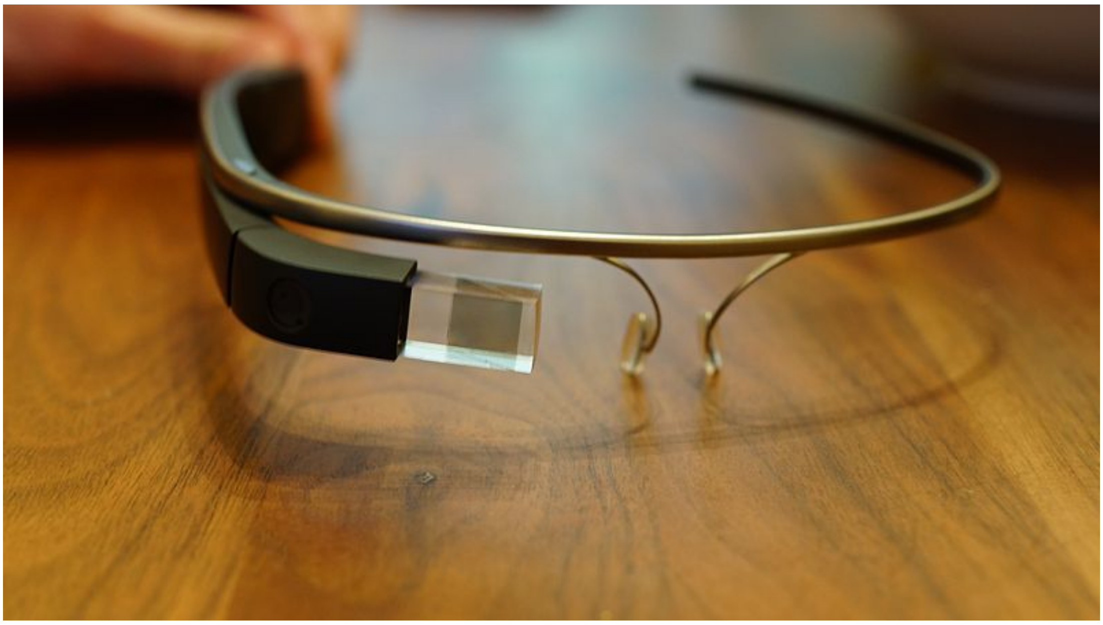


Figure 2.3: Google Glass - Courtesy of Wikipedia

advantages and disadvantages of the smartphones, but they are becoming progressively less widespread than them. Finally, tablets are a lot more powerful than the rest of devices, but they are a bit more expensive and harder to handle, even using both hands.

2.4.1.3 Spatial Augmented Reality

Spatial Augmented Reality is the most complex of Augmented Reality appliances. Commonly, it makes use of video-projectors, optical elements, holograms, and any other tracking technologies to project graphical information directly onto physical objects instead of requiring the user to have a display. Spatial displays separate most of the technology from the user and integrate it into the environment. This characteristic makes SAR very suitable for developing performances for groups of users, thus allowing an interaction between them.

2.4.2 Augmented Reality Activation

Prior to Augmented Reality can be applied over an object, it must be activated. There are several ways of activating an AR application (25). The most common are those based on markers. In order the AR to be activated, the AR device must detect the marker using its sensors. Typically it is a visual mark, but it can also be any other element detected by device sensors, like a code, an electromagnetic signal, or a GPS position. Once the marker has been read and identified, virtual information will be shown. Another more complex way of AR activation is by means of automatic recognition. This way of tracking employs artificial vision techniques for recognizing the target objects. It is called markerless tracking due to it does not use any markers for activating the AR. Nevertheless, marker-based AR is still the most common technique for AR activation (24). Typically, AR device employs its sensors for detecting the marker and activate the AR over the target object. This kind of detection can be based on visual codes (for example, QRCodes), or in the physical location of the user (inside a museum, or in a city, etc.). The common point for all marker-based systems is that AR is not activated by object recognition, but for any other means. The main strong aspects of the marker-based activation are the following:

- Simple detection algorithm
- Robust against lighting changes

Markers also have several weaknesses. They are as follows:

- Do not work if partially overlapped
- Marker image usually has to be black and white
- Marker has to be contained in a square form in most cases (because this make it easier to be detected)
- Non-esthetic visual look of the marker
- Has nothing in common with real-world objects

In the case of location-based markers, the main shortcoming is due to the fact that the accuracy of positional sensors (like GPS) can be quite variable. This issue can be compensated indoors using emitters to help the device to detect its location accurately. Image-based AR employs pictures of the objects itself for activation. This is the most common strategy when applying AR over images, painting, sculptures or advertisement. This seems to be the most appropriate approach for this research project. There are two ways of applying this technique; either a visual marker has been previously stored in the application, or visual recognition techniques are applied for detecting the targets. Although this last approach is the most complex, it is also the most flexible, allowing the system to recognize the same historical figure in different portraits.

2.4.3 Augmented Reality Browsers

The term Augmented Reality Browser was coined by the vendors for naming an interactive AR environment that allows users to navigate through digital contents. AR Browsers provide real-time information from the Web overprinting images and video caught by the smartphone camera (26). That is, AR Browsers allow an enhanced perception of the world adding interaction and information gathering in an integrated and intuitive way. AR Browsers are optimized for showing Web contents in small screens in an efficient way. They provide users a very convenient real-time image processing, giving them a realistic perception. Information and contents can be accessed by several means (26) :

- Geolocation data (latitude, longitude, accelerometer, compass, or any other available sensor)
- Image recognition (people or objects)
- Physical interaction with the target (i.e. artificial vision tracking)

AR Browser capabilities have quickly evolved. The first models had limited tracking and detection functions, and the new ones offer advanced integrated abilities. Location model has been extended, and for inside navigation it is possible to use LLA Markers (Junaio), which have a codified georeference location, instead of GPS signal. This technique enhances accuracy.

2.4.4 Augmented Reality applications in the educative context

The development of AR applications in the educative context is not a new thing. They have been used for years in several contexts (24). For example, there is an application that allows users to visit historical ruins and watch a virtual reconstruction of the buildings (27). The application of AR in this context provides a number of advantages (24): an effective communication with users by means of multimedia, a way of presenting information that is both natural and intuitive, and a low maintenance and acquisition costs. To illustrate this, a user will find easier to walk through an exhibition and target an object with his mobile to get additional data about it, than to look for the object description in a catalogue. In addition to this straightforwardness, this textual information could be enhanced with presentations, videos, and any other multimedia material if convenient. In Spanish context, edutainment is a field in which AR has been successfully applied. There is a wide catalogue of scientific disciplines. There are applications for simulate the anatomy of animals, and also the detailed analysis of paintings.

2.5 Hypothesis, research questions and objectives

2.5.1 Hypothesis

The project's working hypothesis states as follows: image-based recognition techniques are more suitable than feature-based techniques for characters recognition in a picture. The main advantage of these techniques is that there is no information loosing due to focusing exclusively on specific regions or determined features (21).

2.5.2 Research question

The research question of the present project can be stated like this: Which is the most suitable facial recognition technique for identifying historical figures in paintings? To answer this question, the research project must cover before a number of tasks:

- Analyzing artificial vision techniques for determining the best image recognition method.
- Analyzing the existing AR development frameworks for proposing the most suitable technology for implementing the application.
- Designing AR application in accordance with the conclusions of the previous tasks.

2.5.3 Objectives

The main objective of the research project is the determination of the most suitable facial recognition technique for historical characters identification in paintings. This objective can lead to others. For example, analysing a painting is not equivalent to photography analysis. A painting is an abstraction from reality that has been represented manually in a flat image. Difficulties associated to such process are common to other contexts. For example, facial recognition applied to images that have been manipulated or taken under unfavourable conditions. Another interesting feature of characters recognition is the discovery of unknown characters in paintings. In many pictures from the Renaissance or the Late Middle-Ages there are historical personalities depicted as benefactors. This person usually defrays the costs of painting elaboration, and was depicted as a worshipper. Sometimes this person is publicly known, such as Lorenzo de Mdici. But there are also unknown benefactors that are depicted in several paintings. In some other paintings there are historical personalities depicted as extras in a scene, such as Rembrandt did. If the proposed system is accurate enough, it could also be used for detecting people across different paintings and help to recognize them. In this way, there is a project developed under FACE initiative (28).

2.6 Research methodology

From a methodological point of view, the first part of the projected work fits design and creation model while the second one is a survey. Design and creation strategy allows several approaches, and the outcomes of research are not mandatory to be an application. The benefits of this stage are not exclusively the prototype but the empirical data that its use provides. Thus, the prototype can be considered a tool for supporting a research process that provides a convenient sandbox for proposed techniques. Beside this, survey will be performed to evaluate users' experience. Gathered data will allow analysing their perception of the tool.

2.6.1 Phase I: Draft

The analysis of the problem will be developed under an opened approach. The reason is that the projects main objective is to determine which option will be the most suitable to apply to a picture for facial recognition. A priori, considering the conclusions of the articles studied in the State of the Art, image-based seem to be the most suitable techniques. Nevertheless, narrowing the project perspective to these techniques and ignoring the rest could lead to biased conclusions.

2.6.2 Phase II: Theoretical framework

The study of the theoretical framework must serve as a stage prior to the development of a prototype. It is not only about studying the State of the Art in order to write the thesis. We must go in depth about those more promising techniques a priori. These knowledge will allow to select the technologies to be applied in the prototype. A priori, at least two image-based and one feature-based technique should be considered. This initial set could be widened for the inclusion of more techniques of each type, especially statistical techniques. The State of the Art study, although having been performed in the initial stages of the development of the research planning, does not have to be considered as a close element once the main body is written. It must be taken into account that the research work will last for two or three years, and during this period new papers can appear or new interesting techniques can be developed, and this new information will be gathered in the final statement.

2.6.3 Phase III: Prototype implementation

As it has been pointed out previously, the project considers the development of a prototype to test different face recognition techniques. This phase is planned to host several tasks related to prototype implementation, such as its scope, the objectives to achieve, the most suitable technologies for building it, and finally the design and programming. Generally speaking, the prototype will target two objectives. On one side, it should be used for measuring the performance of the applied algorithms in a formal way, providing response times for every single algorithm, resources consumption, etc. All these data will be persisted in a database in order to be analysed later. On the other side, it should allow to study the perception of the users, measuring the quality of user experience. Due to the working hypothesis is that image-based techniques are more suitable for face recognition in a painting than feature-based techniques, the prototype should ideally implement a good number of different techniques of this type. Nevertheless, for being able to establish an unbiased comparative, the prototype should also considerate at least the most significant feature-based algorithms. The agilest technological choice for implementing the prototype seems to be OpenCV (29). OpenCV is an open-source library centred on artificial vision, developed originally by Intel in 1999. It is a very popular tool that can be applied in multiple artificial vision related contexts. OpenCV offers many built-in functions for face detection and face recognition, including implementations for Eigenfaces, Fisherfaces and LBP. It was originally developed for C++, but nowadays it is partially or totally available for other programming languages such as Java or Python. Despite OpenCV offers an implementation for certain algorithms, there will be a number of unimplemented techniques that should be programmed. It is important to underline that nearly all face recognition algorithms are conceptually complex, and that all developed code should be formally validated. An erroneous implementation could add biased data in the later analysis. So, it should be considered a previous validation stage prior to add a new algorithm to the prototype.

2.6.4 Phase IV: Data collection and analysis

Data collection process will be focused on developing a survey. This stage will require the prototype developed in the previous stage. The prototype needs to be fully functional in order to apply recognition techniques. Ideally, it should be able to offer a wide catalogue of techniques for testing.

2.6.4.1 Survey development

A group of volunteers will be gathered to carry out the survey. This group will be divided up to cover different techniques separately. Let N be the number of techniques implemented in the prototype, then there will be N+1 segments, one per each technique plus a control group that will test all the techniques. The subjects will not know about which techniques they are using. All of them will follow the same itinerary at the museum, visiting the pictures used for training the system. Once that a user has finished his itinerary, he will be asked to score certain characteristics such as system accuracy, response time, and the overall user experience. Apart from this information obtained from users, formal measures will be taken from the own application to determine precise response time.

2.6.4.2 The R programming language

R is a statistical programming language with an integrated collection of programs for data manipulation, calculus and graphics. Mainly, R is a tool for developing new interactive methods of data analysis (30). Amongst its main features, R offers:

- A well-grounded programming language, simple and effective (containing conditionals, loops, recursion, etc.)
- Data manipulation and storage
- Operators for indexed variables calculus
- An integrated collection of data analysis tools
- Graphical functions for data analysis
- A consistent environment, not a simple aggregation of separate tools

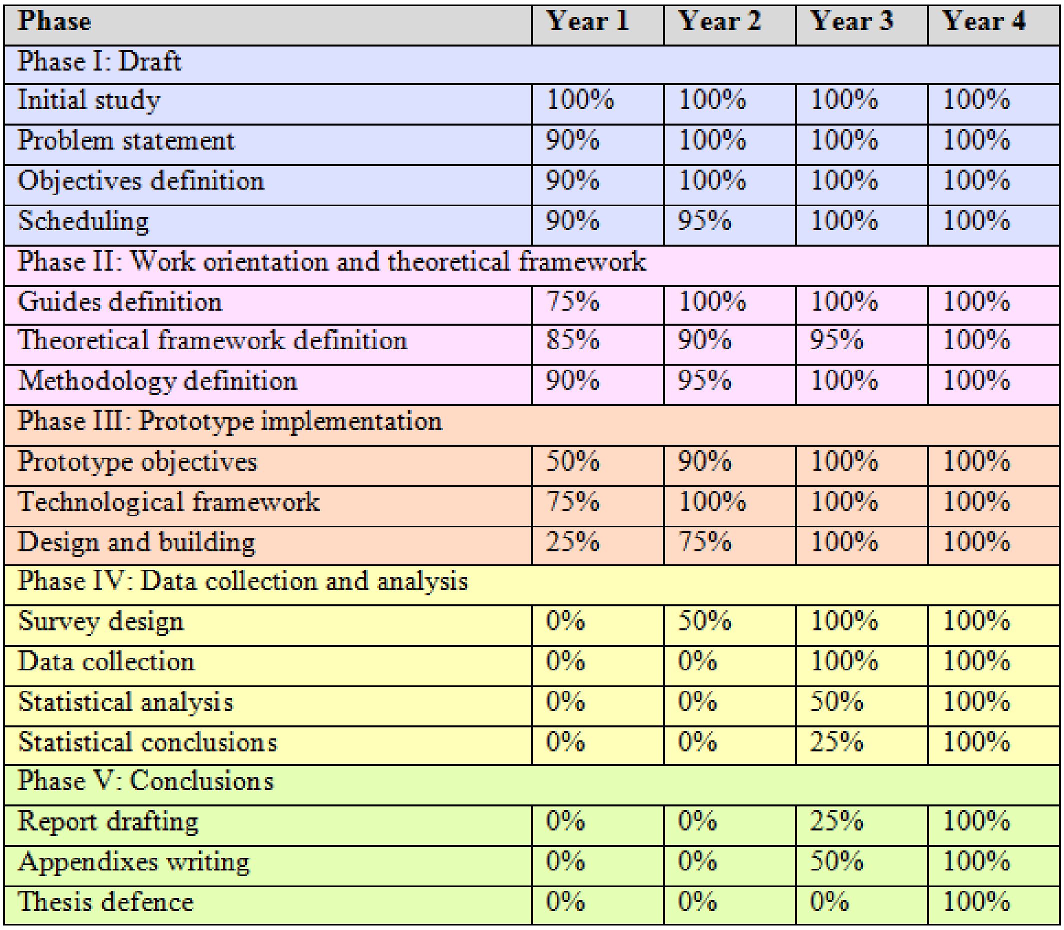
A study will be performed using all gathered data. After a statistical analysis, conclusions will be studied to determine if there are significant variations among the used techniques.

2.6.5 Phase V: Conclusions

During this last phase, the results obtained in prior stages will be reviewed, and the final thesis will be written. If experiments results are in accordance with initial hypothesis, this should be correctly justified by means of quantitative data. For example, the average performance measured, or the hit rate, could be considered proper indicators. These data will complement with user experience data, gathered during the survey. All the quantitative data should be properly structured to support thesis defence. That is the reason why all statistical data should have been compiled carefully. It is very important that statistical results were significant and had enough statistical power.

2.7 Scheduling

A 4-years calendar has been defined and each column reflects the progress of the tasks at the end of every year. Project scheduling is summarized in Figure 2.4. Generally speaking, the working plan defines an initial period for definition, analysis and prototype design, whose completion will take two years. After that, a completely functional prototype should be available and all necessary data gathered when the third year has finished. The last year will be used for data analysis and conclusions. A number of appendixes have been considered, in order to gather any other secondary outcomes or explanatory notes of the research process.



Phase	Year 1	Year 2	Year 3	Year 4
Phase I: Draft				
Initial study	100%	100%	100%	100%
Problem statement	90%	100%	100%	100%
Objectives definition	90%	100%	100%	100%
Scheduling	90%	95%	100%	100%
Phase II: Work orientation and theoretical framework				
Guides definition	75%	100%	100%	100%
Theoretical framework definition	85%	90%	95%	100%
Methodology definition	90%	95%	100%	100%
Phase III: Prototype implementation				
Prototype objectives	50%	90%	100%	100%
Technological framework	75%	100%	100%	100%
Design and building	25%	75%	100%	100%
Phase IV: Data collection and analysis				
Survey design	0%	50%	100%	100%
Data collection	0%	0%	100%	100%
Statistical analysis	0%	0%	50%	100%
Statistical conclusions	0%	0%	25%	100%
Phase V: Conclusions				
Report drafting	0%	0%	25%	100%
Appendixes writing	0%	0%	50%	100%
Thesis defence	0%	0%	0%	100%

Figure 2.4: Project scheduling - Diagram shows task degree of completion

2.7.1 Phase I: Draft

As shown in project scheduling, most of the tasks scheduled for this stage should be finished by the end of the first year of project development. Nevertheless, there are two tasks that could be extended until the second or even the third year. Tasks related to objectives and problem definition could be affected by prototype development. That is the reason why their progress has been estimated in a 90% for the first year. The fact that the main activity during this phase is the definition of the theoretical framework should be taken into account when developing the prototype, because both tasks could benefit from the other. Equally, the proposed scheduling could be modified during project development. Although main milestones are clear, certain tasks are subject to being changed. All the same, it has been established that once the third year has finished, planning has to be completed at 100%.

2.7.2 Phase II: Theoretical framework

Mainly, the development of the theoretical framework will be carried out during the first year. Although it is important to remark that it will not be terminated until prototype related work has been finished. As stated before and owing to research project will take several years, it is perfectly feasible that research advances in this area develop new techniques or any other matter of interest that should be considered for the concluding version.

2.7.3 Phase III: Prototype implementation

It has been planned this phase to be developed in the second year. Nevertheless, there are several relevant issues that will be started in the first year. At the end of this year, all the important decisions concerning technologies and objectives should be made.

2.7.4 Phase IV: Data collection and analysis

This stage has been planned to start on the third year, excluding survey design, which could start in the second year. Thus, this last task could be developed in parallel with prototype implementation, and both tasks could complement each other. In any case, survey design completion will be in the third year. All other tasks associated with this phase would be completed during the third year, with the exception of the development of the study's findings. They will begin to be drafted at that time, but termination would also be associated with the last phase, and therefore the last year.

2.7.5 Phase V: Conclusions

This last phase is mainly centred on developing the last version. Some of the conclusions may have been written in a first draft, summarizing the results of the survey based on the prototype, but they will be finished in the last year.

3

Director thesis

3.1 Director proposal

Enric Guaus (Barcelona, 1974) is a researcher in sound and music computing at the Music Technology Group, Universitat Pompeu Fabra (UPF), and professor at the Sonology Department, at the Escola Superior de Música de Catalunya (ESMUC). He obtained a PhD in Computer Science and Digital Communications (UPF), in 2009, with a dissertation on automatic music genre classification. His research interests cover music information retrieval and human interfaces for musical instruments. He is assistant professor in acoustic engineering at the Universitat Pompeu Fabra (UPF) and lecturer in maths, electronics and computer science at the Escola Superior de Música de Catalunya (ESMUC). He is member of the Observatori de de prevenció auditiva per als msics (OPAM) i de la Barcelona Laptop Orchestra (BLO).

3.2 Relation to UOC

Currently, Enric Guaus is also a consultant professor at Universitat Oberta de Catalunya (UOC) and collaborator at different master programs. He is actively collaborating with David García Solórzano, professor of Master Thesis course.

References

[1] Anil K. Jain and Stan Z. Li. *Handbook of face recognition.* 2011. 3, 5, 10, 11, 14, 16

[2] Takeo Kanade. **Picture processing system by computer complex and recognition of human faces**. 1974. 4

[3] Lawrence Sirovich and Michael Kirby. **Low-dimensional procedure for the characterization of human faces**. *JOSA A*, **4**(3):519–524, 1987. 4, 11

[4] Michael Kirby and Lawrence Sirovich. **Application of the Karhunen-Loeve procedure for the characterization of human faces**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **12**(1):103–108, 1990. 4, 11

[5] M.A. Turk and A.P. Pentland. **Eigenfaces for recognition**. *J. Cogn. Neurosci. 3(1)*, pages 71–86, 1991. 4, 12

[6] Peter N. Belhumeur, João P Hespanha, and David Kriegman. **Eigenfaces vs. fisherfaces: Recognition using class specific linear projection**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **19**(7):711–720, 1997. 4, 12

[7] Kamran Etemad and Rama Chellappa. **Face recognition using discriminant eigenvectors**. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, **4**, pages 2148–2151. IEEE, 1996. 4

[8] Hiroshi Yoshimura, Minoru Etoh, Kenji Kondo, and Naokazu Yokoya. **Gray-scale character recognition by Gabor jets projection**. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, **2**, pages 335–338. IEEE, 2000. 4

[9] Tim Rawlinson, Abhir Bhalerao, and Li Wang. **Principles and Methods for Face Recognition and Face Modelling**. pages 1–29, 2009. 4

[10] Erik Hjelmå s and Boon Kee Low. **Face Detection: A Survey**. *Computer Vision and Image Understanding*, **83**(3):236–274, September 2001. 6, 7, 10, 13, 15

[11] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. **Active shape models-their training and application**. *Computer vision and image understanding*, **61**(1):38–59, 1995. 8

[12] Rabia Jafri and Hamid R. Arabnia. **A Survey of Face Recognition Techniques**. *Journal of Information Processing Systems*, **5**(2):41–68, June 2009. 10

[13] Mikkel B Stegmann and David Delgado Gomez. **A brief introduction to statistical shape analysis**. *Informatics and Mathematical Modelling, Technical University of Denmark, DTU*, **15**, 2002. 11

[14] Rama Chellappa, Charles L Wilson, and Saad Sirohey. **Human and machine recognition of faces: A survey**. *Proceedings of the IEEE*, **83**(5):705–741, 1995. 13

[15] Shang-Hung Lin, Sun-Yuan Kung, and Long-Ji Lin. **Face recognition/detection by probabilistic decision-based neural network**. *Neural Networks, IEEE Transactions on*, **8**(1):114–132, 1997. 13

[16] Gilles Burel and Dominique Carel. **Detection and localization of faces on digital images**. *Pattern Recognition Letters*, **15**(10):963–967, 1994. 13

[17] Paul Juell and Ron Marsh. **A hierarchical neural network for human face detection**. *Pattern Recognition*, **29**(5):781–787, 1996. 13

[18] Henry A Rowley, Shumeet Baluja, and Takeo Kanade. **Neural network-based face detection**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **20**(1):23–38, 1998. 13

[19] Feraud Raphaël, Bernier Olivier, and Collobert Daniel. **A constrained generative model applied to face detection**. *Neural Processing Letters*, **5**(2):11–19, 1997. 13

[20] Antonio J Colmenarez and Thomas S Huang. **Maximum likelihood face detection**. In *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, pages 307–311. IEEE, 1996. 14

[21] Tony S Jebara. *3D pose estimation and normalization for face recognition.* PhD thesis, McGill University, 1995. 15, 21

[22] Raphael Cendrillon and Brian Lovell. **Real-time face recognition using eigenfaces**. In *Visual Communications and Image Processing 2000*, pages 269–276. International Society for Optics and Photonics, 2000. 15

[23] Ben Butchart. **Augmented reality for smartphones**. 2011. 17

[24] Julie Carmigniani and Borko Furht. *Handbook of Augmented Reality.* Springer New York, New York, NY, 2011. 17, 19, 20

[25] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. **Real-time detection and tracking for augmented reality on mobile phones.** *IEEE transactions on visualization and computer graphics*, **16**(3):355–68, 2010. 18

[26] Lester Madden. *Professional augmented reality browsers for smartphones: programming for junaio, layar and wikitude.* John Wiley & Sons, 2011. 20

[27] Erich Bruns, Benjamin Brombach, Thomas Zeidler, and Oliver Bimber. **Enabling mobile phones to support large-scale museum guidance**. *IEEE multimedia*, **14**(2):16–25, 2007. 20

[28] B Miller. **Research on Application of Face-recognition Software To Portrait Art Shows Promise**. http://ucrtoday.ucr.edu/15392, 2008. [Online; accessed 29-June-2014]. 22

[29] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library.* " O'Reilly Media, Inc.", 2008. 24

[30] Bill Venables, Dave Smith, Robert Gentleman, and Ross Ihaka. *Notes on R: a programming environment for data analysis and graphics*. University of Auckland, 1998. 25

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other German or foreign examination board.

The thesis work was conducted from 2014 to 2014 under the supervision of Enric Guaus at Universitat Oberta de Catalunya.

MADRID, June, 29th 2014