UNIVERSITY OF CALIFORNIA
RIVERSIDE

Investigating the Role of Saliency for Face Recognition

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Ramya Malur Srinivasan

March 2015

Dissertation Committee:

Professor Amit K Roy-Chowdhury, Chairperson
Professor Ertem Tuncel
Professor Conrad Rudolph
Professor Tamar Shinar

The Dissertation of Ramya Malur Srinivasan is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am extremely grateful to my adviser Prof. Amit Roy-Chowdhury, who has been a great source of guidance and inspiration throughout these PhD years. He gave me some very interesting topics to explore and has always been there to help and suggest new ideas. Through his words of wisdom and kindness, Prof. Roy-Chowdhury has helped me tremendously during some very tough times. I have learned a lot from him not only from a research perspective, but also about many things in general. Indeed, it will be a life long effort for me to inculcate his good values.

I am also grateful to Prof. Conrad Rudolph for his suggestions and help in connection with the face recognition work on art images. It was a delight to know a bit about art history in the process. Prof. Rudolph has been very patient in clarifying my doubts and, in general, has been very accommodative. I would also like to thank Prof. Jeanette Kohl for her suggestions in this regard. I would also like to extend my sincere thanks to Prof. Ertem Tuncel who has helped me extensively and has been very kind to oblige my requests. Thanks are also due to Donato Sir and Dr. Abhishek of Samsung Research, Dallas where I did an internship.

I am very thankful to my lab colleagues for the various insightful discussions that we have had. In particular, I would like to thank Dr. Nandita and Abir for their invaluable help and time, they have been very kind to accommodate my requests. I would also like to thank Katya, Shu and Dr. Yingying for providing me with useful inputs at various points of time. It has been a very nice and pleasant experience knowing each of my lab colleagues.

I am deeply indebted to my late grandmother Smt. Ranganayaki and my late

grandfather Sri. Rangaswamy Iyengar for their heartfelt blessings, love, concern, care and affection towards me. I am immensely grateful to my mother-in-law Smt. Ambuja and father-in-law Sri. Santhanam for their encouragement, support and kindness. I am very grateful to my husband Dr. Arvind for his support and guidance. Words cannot describe the role of my mother Smt. Dr. Prabha and that of my father Sri. Srinivasan. They have sacrificed a lot for my sake, and with their unconditional love and affection, they have been there for me during the thick and thin of times. My mother has been my inspiration in pursuing a PhD and she has always boosted my confidence, corrected me and offered great moral strength to me. Finally, and most importantly, I wholeheartedly thank God for everything.

Acknowledgment of previously published or submitted materials: The text of this dissertation, in part or in full, is a reprint of the material as it appears in four previously published or submitted papers that I first authored. The co-author Dr. Amit K. Roy-Chowdhury, listed in all these publications, directed and supervised the research which forms the basis for this dissertation. The papers are, as follows.

1. Computerized Face Recognition in Renaissance Portrait Art, Accepted to the Signal Processing Magazine, Special Issue on Signal Processing for Art Investigations, 2015. My co-author Dr. Conrad Rudolph provided the dataset and contributed to the writing and analysis.

2. Recognizing the Royals—Leveraging Computerized Face Recognition for Identifying Subjects in Ancient Art Works, in ACM International Conference on Multimedia, 2013. My co-authors Dr. Conrad Rudolph and Dr. Jeanette Kohl provided the dataset and con-

tributed to the writing and analysis.

3. Modeling Style and Content from Very Limited Training Data—Application in Face Recognition, To be submitted to Pattern Recognition Letters.

4. Face Recognition based on Sigma Sets of Image Features, in IEEE International Conference on Acoustics, Speech and Signal Processing, 2014. My co-authors Dr. Abhishek Nagar, Donato Mitrani and Anshuman Tewari contributed to the writing and analysis.

To God.

ABSTRACT OF THE DISSERTATION

Investigating the Role of Saliency for Face Recognition

by

Ramya Malur Srinivasan

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, March 2015
Professor Amit K Roy-Chowdhury, Chairperson

There has been enormous interest in developing automatic face recognition techniques. Be it for government use such as law enforcement, surveillance and immigration, or for commercial use such as gaming industry, e-commerce, healthcare and banking, a large number of real world applications utilize face recognition. A variety of challenges are associated with a face recognition system. While modeling variations in facial expressions, age, pose and illumination is necessary in many applications, certain specific applications may also involve comparing face images taken over different media such as a facial sketch to a photo.

Selecting visually salient, i.e., highly informative and discriminative features, is critical to every face recognition task. Often, such features are selected based on expert knowledge and/or learned from training data. The choice of these features is largely governed by the application. While there has been extensive work on saliency-based feature selection strategies for object/activity recognition in general, the role of saliency in the context of face recognition is relatively unexplored. The primary focus of this work is to investigate the role of saliency for face recognition.

We discuss three face recognition applications illustrating the role of saliency in each of these problem domains. In the first, we propose a framework for identifying subjects (sitters) in ancient portraits belonging to the Renaissance era. Apart from the typical face recognition challenges, recognition in art works come with the additional challenges of having to deal with limited training data and the need to model variability in artistic renditions. In this direction, we propose a framework that is capable of learning salient characteristics of individual artists and subsequently perform identification based on statistical hypothesis testing.

We next discuss a related face recognition application of comparing an artistic sketch with a photograph. Here, we propose an unsupervised face recognition scheme based on computing saliency maps constructed from region covariance matrices of low level visual features. We also discuss the utility of such features for face recognition in unconstrained environments (often referred to as 'recognition in the wild') and subject to artificial distortions such as Gaussian blur and white noise. We conduct experiments on the Chinese University of Hong Kong Photo-Sketch database and the Quality Labeled Faces in the Wild (QLFW) to demonstrate the advantages of the proposed method.

Taking cue from the scenario of face recognition in art works wherein we have limited authenticated portraits, we next investigate the general problem of face recognition from very limited training data. This problem is relevant to many forensic science applications. We show that by learning salient features characteristic of a style such as a facial expression, pose, etc., one can obtain better recognition accuracies between face image pairs than with the case where such style information is not used. In particular, we leverage upon

statistical hypothesis testing frameworks which can learn and validate features specific to a style. We conduct experiments on the publicly available PubFig dataset wherein the annotated attributes such as smiling, frowning, etc. are used as style information. We show that as the number of training instances in a style class is reduced, the model performs better than state-of-the-art techniques.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Face recognition is a form of biometric identification [4] involving recognition of individuals based on the salient characteristics of their face images. A variety of challenges are associated with a typical face recognition task. Some typical challenges include modeling variations in facial expression, pose, illumination and age. Additionally, when dealing with videos, there may be variations due to non-rigid motion and background clutter. Certain specific face recognition tasks may also involve comparing images taken over different media such as comparing a sketch to a photograph, etc.

A typical face recognition system consists of two main stages, namely, a feature selection and extraction stage followed by a matching stage. The first stage primarily involves extracting low or high level informative features either manually or automatically from images or videos. The matching stage involves classification of an input image into one of the predetermined classes. In case of one-to-one matching, if the face image pair corresponds to the same person, then we consider them as a "match" class and if the

images depict two different people, we refer to them as a "non-match" class. In the case of one-to-many matching, the different classes indicate different identities registered in the system. The classifiers used in this stage are learned offline using training data, e.g, in a supervised framework [3] or may involve semi-supervised [2] or unsupervised learning [1].

The quality of features, which are extracted based on expert knowledge [9] and/or learned from available training data [10], is of critical importance. The focus of this work is on selecting *visually salient* features in order to achieve robust face recognition for different applications. In general, the term *visually salient* refers to those features/regions of an image that stand out relative to their neighbouring parts.

1.0.1 Motivation for the Use of Saliency in Face Recognition

Figure 1.1: Illustrations of the concept of saliency. From left, A represents photo and its corresponding sketch, B shows portraits depicting different sitters by artist Holbein, and C shows some images of smiling people.

The role of saliency in the context of face recognition can be best understood through illustrations. Consider, for instance, Fig.1A, which depicts a photo and its corresponding sketch. As can be noted from the figure, prominent facial features of the photo such as narrow eyes and broad nose are retained and emphasized in the sketch. In general, it is very likely that landmark features such as a facial scar, mole, and other facial assymetries are emphasized and retained in sketches.

Apart from the prominent features specific to the person depicted in the face, often, in sketches or artworks in general, there is an element of visual perception by individual artists. For example, some artists might like to employ dense brushstrokes in all their works, some might prefer to emphasize certain local features (e.g., eye corners, nose tips, etc.) more than other features across all the sitters depicted in their portraits, and some might choose to emphasize salient distances across the face such as mouth width, etc. in a characteristic manner. Due to such typical practices followed by individual artists, one can identify salient regions specific to an artist. Fig. 1B is an illustration of portraits depicting different sitters by Renaissance artist Holbein. A careful observer or an art connoisseur can note signature marks of the artist in these portraits.

Another interpretation for the concept of saliency in face recognition can be perceived from the nature of images under consideration. For example, consider Fig.1 C, which shows images of people smiling. In all these images, one's attention is immediately drawn to the mouth and the eye regions, which are characteristic of the smile. Thus, attributes that characterize various style factors such as a facial expression, pose, etc. can also help determine salient regions.

The underlying concept common to all these scenarios is to identify most informative regions of the face that can aid robust recognition. Depending upon the specific application under consideration, saliency can be perceived differently. Motivated by these illustrations, we investigate the role of saliency for various face recognition applications namely face recognition in art works, face recognition from very limited training data, face photo-sketch recognition and face recognition in unconstrained environments.

1.0.2 Challenges

Apart from the typical challenges associated with a face recognition system such as variations in pose, illumination, facial expression, age, background clutter, etc, there are additional constraints depending on the specific application. Below, we describe some of these challenges.

1. Sparsity of Training Data : In applications involving law enforcement, one can only offer a few or even a single image of the suspect due to non-availability of examples. In general, sparsity of training data is common to other forensic science applications such as studies of ancient art, architecture, biology or archaeology wherein experts are often required to answer questions related to authenticity (e.g., date estimation, identity verification, etc.). Most existing face recognition systems are designed to work under the condition that ample training instances are available across different poses and ages for each of the subjects under consideration. Such luxuries are not available in many forensic science applications. This necessitates design of robust training and validation algorithms that can work with very few training instances.

2. Modeling Artists' Styles: Whether it is the case of face recognition in portraits or a face photo-sketch recognition, the system has to take into account the mark of the visual interpretation of an artist. Different artists might depict the same person in different ways. Thus, styles of individual artists characterizing their aesthetic sensibilities (often biased by their socio-cultural backgrounds) have to be modeled. While there have been works related to artist identification based on statistics related to brushstrokes [11], there is little to no elaborate work on understanding how to model artist's style in face portraiture.

3. Choice of Features: Depending on the problem at hand, we need to choose features that possess high discriminative power in distinguishing the test image from the rest. Consider, for instance the problem of face recognition of portraits. In this case, ideally, the chosen features should not only be able to distinguish one sitter from another, but also be able to distinguish one artist from another. In certain other applications, computational cost is a constraint. So, design of highly efficient and possibly simple matching mechanisms becomes essential in these scenarios. This leads to interesting questions in machine learning on combinations of various algorithms that are pertinent to the specific application.

1.1 Related Work

We provide a review of related work on face recognition and saliency.

1.1.1 Face Recognition

A critical survey of still image and video based face recognition research is provided in [5]. Depending on the representation of the face descriptors, face recognition approaches can be broadly divided into two categories, namely, feature based methods [7] and holistic methods such as [6]. Some hybrid methods [8] involve both holistic and feature based procedures.

Feature-based approaches first process the input image to identify and extract (and measure) distinctive facial features such as the eyes, mouth, nose, etc. Then, they compute the geometric relationships among those facial points, thus reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition

techniques are then employed to match faces using these measurements. Early work carried out on automated face recognition was mostly based on these techniques.

A well known feature-based approach is the elastic bunch graph matching method proposed by Wiskott et al.[57], wherein each fiducial point on the face is a node of a fully connected graph, and is labeled with the Gabor filters responses applied to a window around the fiducial point. Each arch is labeled with the distance between the correspondent fiducial points. A representative set of such graphs is combined into a stack-like structure, called a face bunch graph. Using this architecture, a rank-1 recognition rate of 98% fwas obtained for a gallery of 250 individuals.

The main advantage offered by the featured-based techniques is that since the extraction of the feature points precedes the analysis done for matching the image to that of a known individual, such methods are relatively robust to position variations in the input image. The major disadvantage of these approaches is the difficulty involved in automatic fiducial point detection.

Holistic approaches attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. In the simplest version of the holistic approaches, the image is represented as a 2D array of intensity values and recognition is performed by direct correlation comparisons between the input face and all the other faces in the database. Though this approach has been shown to work [6] under limited circumstances (i.e., equal illumination, scale, pose, etc.), it is computationally very expensive and suffers from the usual shortcomings such as sensitivity to face orientation, variable lighting conditions, background clutter, and noise.

More recently, region covariance matrices (RCM) were proposed a set of holistic descriptors. Unlike many holistic descriptors, RCMs are robust to variations in pose/ illumination, and are low dimensional. RCMs have been applied to many applications such as texture classification and object detection [88, 95], object tracking [89, 97] and human detection [90]. These descriptors have shown best discriminative power for human detection tasks [98] and have also been recently studied in the context of face recognition [99].

The main advantage of holistic descriptors is that they do not require to know the precise location of certain fiducial points on the face. As a result, they are computationally simple. However, since most of these approaches start with the basic assumption that all pixels in the image are equally important, they require a certain degree of correlation between test and training images.

A vast majority of face recognition applications address surveillance and entertainment needs. Recent research efforts in the area are addressing new challenges such as heterogeous face recognition. For example, [13] focusses on cross spectral face recognition for comparing an infrared image to a photo . In [12], partial least squares (PLS) is used to linearly map images in different modalities to a common linear subspace where they are highly correlated. In [14], a generic heterogenous face recognition framework is proposed in which both probe and gallery images are represented in terms of non-linear kernal similarities to a collection of prototype face images.

Some works have looked at face recognition from sparse training data [15]. This is commonly referred to as the problem of undersampled face recognition. In [17], the authors

leverage upon much larger mug shot gallery images or composite sketches for training. In [71], the authors assume that the intra-class variations of one subject can be represented as a sparse linear combination of the variations of other subjects and use an extended sparse representation based classifier to perform recognition. In [16], the authors evaluate the probability that two faces have the same underlying identity cause and leverage this information for recognition.

Attribute based face verification methods have become popular over the years. Attributes are describable aspects of information about the face such as a facial expression, age, gender, pose or could be any other side information such as spectacles, beard, facial scar, etc. In [19], the authors proposed attribute classifiers that are basically binary classifiers trained to recognize the presence or absence of attributes. They also proposed "simile" classifiers that removed the manual labeling required for attribute classification and instead learned the similarity of faces, or regions of faces, to specific reference people. In [48], the authors proposed to use information about the relative strength of attributes for face verification. Some works such as [20], [18] model factors such as a facial pose, expression, etc. and separate it from the identity of the person, and show promising results for face pose estimation, among others. These are commonly referred to as the style-content separation methods, wherein attributes such a facial expression, pose are termed as style and the identity of the person is referred as the content.

1.1.2 Saliency

In recent years, many works have aimed to answer questions related to attentional mechanisms that determine the most salient regions in an image. The basis of many attention models dates back to Treisman and Gelade's [26] Feature Integration Theory, where they stated which visual features are important and how they are combined to direct human attention over pop-out and conjunction search tasks. Koch and Ullman [25] then proposed a feed-forward model to combine these features and introduced the concept of a saliency map which is a topographic map that represents conspicuousness of scene locations. They also introduced a winner-take-all neural network that selects the most salient location. Many other works followed this which specifically looked at digital images [22], [23], [24]. The first complete implementation and verification of the Koch and Ullman model was proposed by Itti et al. [21].

Based on the manner in which salient regions are determined, saliency models can be categorized into two broad groups, namely bottom-up models and top-down models. Regions of interest that are mainly driven by involuntary stimulus (e.g., a black line amidst many red lines) attract attention in a bottom-up manner [31]. These are typically determined by intrinsic low-level properties of the scenes. On the other hand, regions of interest that are propelled by a goal (e.g., searching for an object in a scene, etc.) attract attention in a top-down manner. These are usually determined by cognitive phenomena like rewards, expectations, etc. [32]. Some models integrate low-level bottom-up cues with some task-specific top-down knowledge such as face and object detectors [36] to improve their predictions.

There are a number of computational approaches to model saliency. Most of them consider bottom-up strategies in a supervised data-driven manner to compute the saliency maps [33], [34], [35]. Recently, unsupervised saliency learning approaches have been leveraged such as in [37] for person re-identification and in [38] for discovering new object categories. In general, modeling visual saliency has provided efficient strategies for various other computer vision applications such as image segmentation [27], object recognition [28], visual tracking [29], etc. Despite the considerable amount of work on saliency for different computer vision applications, there is very little work on saliency for face recognition [39], [40].

1.2 Contributions of the Thesis

The objective of this work is to design algorithms that can identify salient regions in order to achieve robust face recognition. The study has been carried out in three parts.

First, we explore the feasibility of face recognition technologies for analyzing works of portraiture, and in the process provide a quantitative source of evidence to art historians in answering many of their ambiguities concerning identity of the subject in some portraits and in understanding artists' styles [50], [49]. Based on an understanding of artistic conventions, we show how to learn and validate features that are robust in distinguishing subjects in portraits (sitters) and that are also capable of characterizing an individual artist's style. We show that this can be used to learn a feature space called Portrait Feature Space (PFS) that is representative of quantitative measures of similarities between portrait pairs known to represent same/different sitters. Through statistical hypothesis tests we analyze uncer-

tain portraits against known identities and explain the significance of the results from an art historian's perspective. Results are shown on our data consisting of over 270 portraits belonging largely to the Renaissance era.

Next, we propose an unsupervised face recognition framework that involves saliency maps computed from sigma sets of region covariance matrices [51], which are in turn estimated from simple image features. Encoding the features descriptors in this manner not only ensures robustness to variations in pose, illumination, etc., but also reduces computational complexity. Further, saliency maps provide a natural way of highlighting the most important regions and thereby improve recognition performance. We demonstrate the effectiveness of the proposed method for (a) face photo-sketch recognition and (b) unconstrained face recognition in the presence of artifacts.

Finally, given very limited training data, we propose a novel undersampled face recognition method by means of modeling style and content, where style can be attributes such as a facial expression, pose, etc. and content is usually the person's identity [52]. We show that by modeling style, face recognition accuracies can be significantly improved. In particular, a set of weighted features characteristic of a style class are learned, wherein the weights denote the importance of the chosen features in the class. The chosen style features are then validated by means of the robust Siegal-Tukey statistical test that is proven to work well for small sample sets. We analyze the PubFig and Yale dataset using the annotated attributes as style factors to illustrate the advantages of proposed method over state-of-the-art Bilinear models (BLM) and relative attributes method as sample size in the style class is reduced.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 discuses face recognition in portraits wherein, we propose a method that can automatically learn salient features characteristic of individual artists style and further propose an identification framework based on statistical hypothesis testing to identify ambiguous sitters in portraits. In Chapter 3, we present an unsupervised face recognition scheme that leverages saliency maps computed from region covariance matrices and demonstrate its advantages for face photo-sketch recognition and for face recognition in the wild. In Chapter 4, we discuss a framework that takes advantage of salient features specific to a style and demonstrate its advantages for face recognition from very limited training data. We provide a summary of the thesis and highlight directions for future work in Chapter 5.

Chapter 2

Face Recognition in Digital Arts

In this chapter, we will first discuss as to why the problem of face recognition is relevant to portriature and the challenges involved in it. Subsequently, we will provide a review of literature on various image processing techniques employed for art investigations with a particular focus on artists style modeling methods. We will then describe the overall framework involving the following stages namely (a) modeling the style of individual artists, (b) using the learned features for face recognition, (c) validation of recognition results through well-known cross validation tests, and (d) using the learned method for verifying the identity in test paradigms where the identity of the subject is unknown/controversial. We will finally discuss the significance of the results from an art historian's perspective.

2.1 Introduction

Renaissance portraits were depictions of some important people of those times. These encompass a wide range of art works such as sculptures, death masks, mosaics,

etc. Apart from being used for a variety of dynastic and commemorative purposes, they were used to depict individuals often to convey an aura of power, beauty or other abstract qualities [53]. A large number of these portraits, however, have lost the identities of their subjects through the fortunes of time.

Analysis of faces in portraits can offer significant insights into the personality, social standing, profession, etc. of the subject they represent. However, this is not a simple task since a portrait can be "subject to social and artistic conventions that construct the sitter as a type of their time" [53], thus resulting in large uncertainty regarding the identity of many of these portraits. Traditionally, identification of many of these portraits has been limited to personal opinion, which is often quite variable. The goal of this work is to evaluate the application of face recognition technology to portrait art and in turn aid art historians by providing a quantitative source of evidence to help answer questions regarding subject identity and artists' styles. This work is part of the project *FACES* (*F*aces, *A*rt, and *C*omputerized *E*valuation *S*ystems) which has been funded by the US National Endowment for the Humanities (Phase 1 and 2) under the Digital Humanities program. This project is a collaboration with the Department of Art History, UC, Riverside. This article presents FACES from the point of view of science. For the humanities point of view, please refer to [65].

There have been lingering ambiguities about the identity in some portraits–henceforth referred to as "test" images. The question has been whether they might represent a certain *known* identity, which we call as "reference images". As an instance, the test image in Fig. 2.1 is a portrait painted perhaps around 1590, and is believed by some to represent Galileo.

Through computerized face recognition technologies, we try to provide an alternate and quantitative source of evidence to art historians in answering such questions.

In this direction, we leverage upon a large number of portrait pairs that are *known* to represent a certain person as shown in top part of Fig. 2.1. The task then is to train the computer in identifying highly discriminative features that can not only distinguish one sitter from another, but also learn the importance of the chosen features depending on the amount of emphasis given to that feature by an artist. Using the learned features, quantitative measures of similarity between portrait pairs known to represent the same person can be computed to yield what we call "match scores". Analogously, similarity scores between portrait pairs not known to represent the same person yield "non-match scores". The resulting match (blue curve) and non-match scores (red curve) together constitute what we refer to as the Portrait Feature Space (PFS). Subsequently, using hypothesis tests, the similarity score between test and reference image, as shown by the brown ball in bottom part of Fig.2.1, is analyzed with respect to the learned PFS to arrive at appropriate conclusions of a possible match or non-match. If both match or non-match happen to be likely, then no decision can be made.

2.1.1 Challenges

Apart from the typical challenges associated with face recognition systems such as variations in pose, expression, illumination, etc., face recognition in portraits come with additional challenges. Some of these are described below.

1. Modeling Artists' Styles: Since portraits bear the mark of the visual interpretation of an

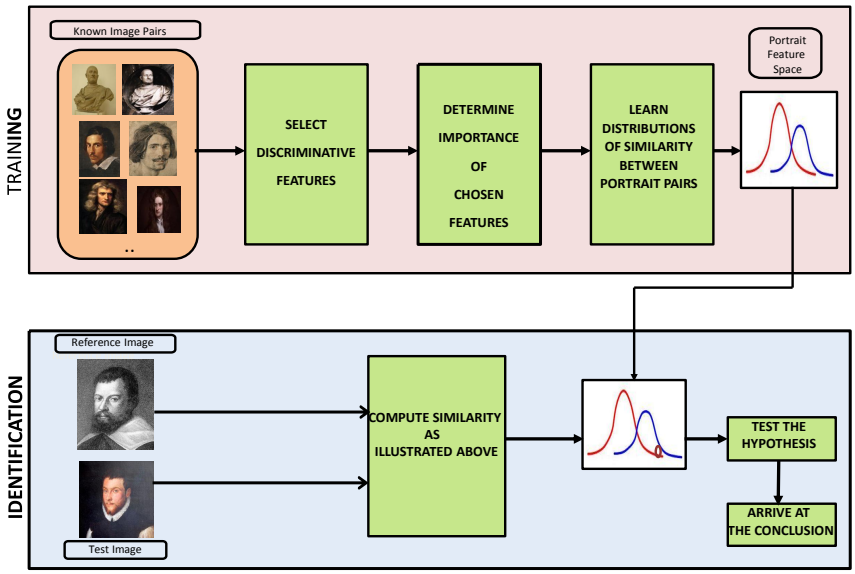

Figure 2.1: Illustration of the training (top) and identification framework (bottom)

artist, styles of individual artists characterizing their aesthetic sensibilities (often biased by their socio-cultural backgrounds) have to be modeled. Thus, portraits of the same sitter can vary from artist to artist. This results in considerable variability in the renditions, which has to be accounted for by the face recognition algorithms.

2. Lack of sufficient training data: Many existing feature selection methods rely on the availability of a significant amount of training data. This is rarely the case in our problem domain due to the following reasons:

(a) Lack of a significant body of images, the authenticity of which is well established.

(b) We need to logically choose a set of related images directed towards a particular demonstrative end and adhering to a particular period style.

3. Choice of Features: Given the aforementioned constraints, we need to choose features that best justify an artist's rendition and possess high discriminative power in distinguishing

the sitter from others. Although there has been some preliminary work on this [11], there is little to no elaborate work on understanding how to model style in face portraiture. This leads to interesting questions in machine learning on combinations of various algorithms that are pertinent here.

2.2 Related Work

We review some image processing techniques employed for art analysis. Analysis of paintings using sophisticated computer vision tools has gained popularity in recent years [44]. Computer analysis has been used for identifying the artist [46] and for studying the effect of lighting in artworks [47], among others. A recent paper has explored application of computer-based facial image analysis [45] using 3D shape information to identify one subject, namely Da Vinci in four artworks. The present work involves multiple sitters (both genders) by different artists portrayed across different media such as paintings, death masks, etc. Also, for the present analysis, shape information was found to be less discriminative when compared to other features such as anthropometric distances (AD) and local features (LF). This can be partly attributed to the evidence that artists often focused on LF and took some liberties with shape [43]. This is further substantiated by the use of local features in matching forensic sketches (an art form) to human faces in works such as [17].

2.3 Discriminative Feature Selection

A portrait is a visualization of an artist's aesthetic sensibilities blended with the sitter's personality. We therefore begin by understanding the relevant features for analysis based on a study of artistic trends during the period under study.

2.3.1 Face as Seen by Artists

It is evident from [43] that while drawing a human body, a lot of emphasis was laid upon maintaining the proportions of various parts. It is purported that the principles for the canons of human body may have been defined by Egyptian artists, who divided the entire body into different parts and provided baselines for their measurement. The importance of anthropometric ratios/distances was preserved even during the Renaissance era. According to Da Vinci, in a well proportioned face, the size of the mouth equals the distance between the parting of the lips and the edge of the chin, whereas the distance from chin to nostrils, from nostrils to eyebrows, and from eyebrows to hairline are all equal, and the height of the ear equals the length of the nose [54].

A historical appraisal of facial anthropometry from antiquity upto Renaissance has been provided in [55] to compare artists' concept of human profile. Flattened nose, tilted forehead and prominent upper lip were some of the features prevalent in Renaissance art works. In fact, prominent facial landmarks of a person were retained in works of the sitter by different artists as illustrated in Fig. 2.2.

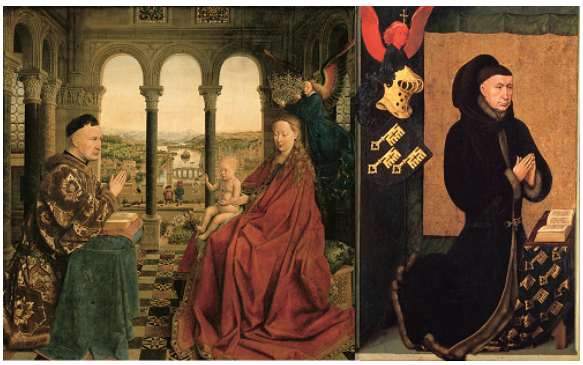

Figure 2.2: Prominent facial landmarks such as pointed nose were retained in works of the same sitter Nicolas Rolin by different artists Jan Van Eyck and Rogier van der Weyden.

2.3.2 Choice of Features

From the illustrations described above, it is clear that ancient Renaissance artists laid emphasis on two aspects in their renderings, namely, local features (LF) and anthropometric distances (AD), which we use for our analysis.

1. Local features: We use a set of 22 fiducial points to represent each face as listed in Table 2.1. The precise location of these points is determined by registering a generic mesh on the face. Gabor jets corresponding to 5 frequencies and 8 orientations are evaluated at each of these fiducial points. At a fiducial point n and for a particular scale and orientation j, the corresponding jet coefficient J_{n_j} is given by

$$J_{n_j} = a_{n_j} \exp(i\phi_{n_j}), \tag{2.1}$$

where a_{n_j} is the magnitude and ϕ_{n_j} is the phase.

2. Anthropometric distances: All images are normalized with respect to scale and orientation. A set of 11 salient distances is used to represent each face. These are listed in Table 2.2.

Number	Description of the feature
1	forehead tips (left)
2	forehead tip (right)
3	forehead center
4	chin bottom
5	nose top
6	nose bottom
7, 8	points on temple (left, right)
9, 10	chin ear corners (left and right)
11, 12	points on chin (left and right)
13, 14	cheekbones (left and right)
15, 16	mouth corners (left and right)
17, 18	iris (left and right)
19, 20	left eye corners (right and left eye)
21, 22	right eye corners (right and left eye)

Table 2.1: List of local features

2.3.3 Feature Extraction

Different artists are likely to depict and emphasize the aforementioned features in different ways. We wish to learn those features that are characteristic of an artist's style. We employ a method called the random subspace ensemble learning as it is capable of handling deficiencies of learning in small sample sizes [56]. Small sample sizes is very relevant to the present problem as we have very few works by an artist at our disposal (Sec 2.1.1). The random subspace method randomly samples a subset of the aforementioned features and performs training in this reduced feature space.

More specifically, we are given Z training portrait pairs and D features. Let L be the number of individual classifiers in the ensemble. We choose $d_i \leq D$ (without replacement) to be the number of features used in the i^{th} classifier. For each classifier, we determine the match and non-match scores (as appropriate) using the d_i features as follows.

Number	Description of the feature
1	distance between forehead tips
2	distance between forehead center and chin bottom
3	distance between nose top and bottom
4	distance between points on temples
5	distance between chin ear cornors
6	distance between points on chin
7	distance between iris
8	distance between cheekbones
9	distance between mouth corners
10	width of nose
11	distance between forehead center and nose bottom

Table 2.2: List of anthropometric distances

We compute

$$s_{LF}(I, I') = \frac{1}{d_i} \sum_{n=1}^{d_i} s_n(J, J^{'}), \tag{2.2}$$

where $s(J, J^{'})$ is an average local feature similarity measure between n corresponding Gabor jets computed across salient points in image pair (I, I'). In order to compute $s_n(J, J^{'})$, we use the normalized similarity measure mentioned in [57] given by

$$s_n(J, J^{'}) = \frac{\sum_j a_{n_j} a^{'}_{n_j}}{\sqrt{\sum_j a^2_{n_j} \sum_j a'^2_{n_j}}} \tag{2.3}$$

Similarly, we compute anthropometric distance similarity between image pairs (I, I') using the equation

$$s_{AD}(I, I') = e^{-\beta y}, \tag{2.4}$$

where y is the 2D Euclidean distance between the AD vectors $\vec{m}$, $\vec{n}$ that characterize images I, I' respectively (we use only those distances as selected by the random subspace classifier) and β is a co-efficient that is chosen suitably to obtain a discriminative range of values.

In order to identify features that give the highest separation between match and non-match scores, we then compute the Fisher Linear Discriminant function for each classifier. We choose the union of features from those classifiers that give the top k Fisher Linear Discriminant values as our style features.

2.3.4 Importance of the Chosen Features

Not all features identified by the above method are equally important in representing an artist's style. In order to understand the importance of the chosen features, we consider the non-parametric statistical permutation test [58]. Permutation test helps in assessing what features are same across all the instances belonging to an artist. Thus, features which are more invariant across the portraits by an artist can be perceived to be more characteristic of that artist and hence be assigned greater importance. Permutation tests have been used to determine invariant features in artworks [11].

Permutation test: The null hypothesis (H_0) is chosen to indicate that two portrait groups $G1, G2$ have the same average value in a particular feature; the alternate hypothesis (H_1) indicates that the average value of that feature is different in the two groups. Thus,

$$H_0 : \mu_{G1} = \mu_{G2}; H_1 : \mu_{G1} \neq \mu_{G2}, \tag{2.5}$$

where μ is the average value of a particular feature v under consideration in the two groups.

If the null hypothesis is true, then it should not matter when this feature v is randomly assigned among images in the group. For instance, let us assume that there is a certain way that the mouth corner is portrayed by Italian artist Bernini, whose works are included in our dataset. On an average, if this appearance is the same across all images by

Bernini, then the principle behind this test is that there will not be a significant difference if the mouth tips are randomly assigned across images in the group, i.e., assigning the feature of one sitter to the corresponding feature of another sitter.

Specifically, if there are N_s images by an artist Y, then we can divide these N_s images into 2 subgroups consisting of N_{s_1} and N_{s_2} images depicting different sitters. Let the feature values for the first group be $[v_1, v_2, ..., v_{N_{s_1}}]$ and in second group be $[v_{N_{s_1+1}}, v_{N_{s_1+2}}, ..., v_{N_{s_2}}]$. The permutation test is done by randomly shuffling $[v_1, ..., v_{N_s}]$ and assigning the first N_{s_1} values, $[v_{(1)}, v_{(2)}, ..., v_{(N_{s_1})}]$ to the first group and the rest N_{s_2} values $[v_{(N_{(s_1+1)}}, ..., v_{(N_{s_2})}]$ to the other group.

For the original two groups we compute,

$$\delta_0 = \left| \frac{1}{N_{s_1}} \sum_{i=1}^{N_{s_1}} v_i - \frac{1}{N_{s_2}} \sum_{i=1}^{N_{s_2}} v_{N_{s_1+i}} \right| \tag{2.6}$$

δ_0 denotes the variation in the feature v by artist Y as exhibited by various instances $I_1, ..., I_N$ in the two groups $G1$ and $G2$. Thus, $\delta_0 = |\mu_{G1} - \mu_{G2}|$. For any two permuted groups we compute

$$\delta_s = \left| \frac{1}{N_{s_1}} \sum_{i=1}^{N_{s_1}} v_{(i)} - \frac{1}{N_{s_2}} \sum_{i=1}^{N_{s_2}} v_{(N_{s_1+i})} \right| \tag{2.7}$$

δ_s denotes the variation in the feature v by artist Y after assigning v as depicted in I_i to an image not necessarily depicting the sitter in I_i.

We repeat this random shuffling of features among the images under consideration multiple times. The proportion of times $\delta_s > \delta_o$ is the p value. This value reflects the variation of the feature in the two groups. Smaller p denotes stronger evidence against the null hypothesis, meaning that the feature differed considerably in the two groups and thus

less characteristic of the artist's style. We compute p values for each feature as described above. The computed p values are used as scaling factors (weights) in estimating the similarity scores (s_p) in equations (2.2) and (2.4). It is to be noted that this method can be employed when we have ≥ 12 images by an artist [59]; in cases where enough images/artist is not available or when the artist is unknown, we use all the 22 LF and 11 AD features with equal weight (of 1 assigned to all the features) in obtaining the LF/AD similarity scores.

2.3.5 Feature Combination

The similarity scores obtained from LF and AD features may not be equally important in determining the similarity between portrait pairs. Further since the number of LF/AD features used are different, the scores need to be fused in a way such that the resulting distribution of match and non match scores are as peaked and disjoint as possible. We employ the following algorithm towards this.

1. We consider a convex combination of the scores from the two measures LF and AD, i.e.,

$$score = \lambda s_{LF} + (1 - \lambda) s_{AD} \tag{2.8}$$

λ being varied from 0 to 1 in steps of 0.1.

2. For every λ, we evaluate the mean and standard deviation of match and non-match scores using the RANSAC algorithm [62] to prune outliers.

3. At each λ, we evaluate $J = \frac{S_b}{S_w}$ where S_b is between class variance and S_w is within class variance. We choose that value of $\lambda = \lambda_{opt}$ that gives the maximum value of J. This is essentially computing the Fisher linear discriminant [63].

Using the procedure described above, we compute similarity scores between portrait pairs that are known to depict same sitters and different sitters to get match and non-match scores respectively. The resulting set of match and non-match scores, computed across various artists and sitters, are modeled as two Gaussians distributions (one for match scores and another for non-match scores). The mean and standard deviations of these distributions are estimated from training data. We refer to these match/non-match score distributions as the "Portrait Feature Space" (PFS).

2.3.6 Validation

We wish to ascertain if the learned features are good representations of the portraits considered. To verify this, we perform two-fold cross validation of the similarity scores.

Validation of Artist-Specific Similarity Scores: If the chosen features are robust representations of an artist Y, then the obtained match/non-match scores divided into two folds (groups), say A, B, should more or less be "similar" in that they come from the same artist. For this, we employ the Siegel-Tukey statistical test [61].

Siegel-Tukey Test: This is a non-parametric statistical method to test the null hypothesis (H_0) that two independent scores come from the same population (e.g., artist) against the alternative hypothesis (H_1) that the samples come from populations differing in variability or spread. Thus,

$$H_0 : \sigma_A^2 = \sigma_B^2, Me_A = Me_B; H_1 : \sigma_A^2 \geq \sigma_B^2 \tag{2.9}$$

where σ^2 and Me are the variance and medians for the groups A and B. The test is entirely

distribution-free. The absence of any normality assumption is an important feature of the test, because its parametric alternative, the F test for variance differences, is quite sensitive to departures from normality [60]. The p value obtained from this test, p_{st}, is given by

$$p_{st} = \Pr\left[X \leq U\right], \tag{2.10}$$

where U_A, U_B are the U statistics for groups A, B and $X \sim$ Wilcoxon (r, m) [59]. This is a measure of the confidence associated with the scores. Thus, if the learned features are good representations of an artist's style, they should be associated with a higher p_{st} value than the p_{st} value associated with scores obtained using all features.

Validation of PFS: In order to validate the PFS computed across various artists and sitters, we randomly divide the known instances into two groups to perform two-fold cross validation. In fold 1, we use group one to learn the PFS and use group 2 to validate and vice versa in fold 2. Ideally, the learned PFS from the two folds should have the same statistics.

2.4 Identification Framework

The goal of this work is to aid art historians by providing an alternate source of evidence in verifying uncertain portraits (test images) against a reference image by providing a quantitative measure of similarity. We use hypothesis testing for this purpose.

2.4.1 Hypothesis Testing

This is a method for testing a claim or hypothesis about a parameter in a population [64]. Below, we summarize it with respect to the learned PFS.

1. Null hypothesis claims that the match distribution accounts for the test's similarity score with reference better than non-match distribution. The alternate hypothesis is that non-match distribution models the score better.

2. We set level of significance α, i.e., the test's probability of incorrectly rejecting the null hypothesis, as 0.05, as per behavioral research standard.

3. We compute the test statistic using one independent non- directional z test [64], which determines the number of standard deviations the similarity score deviates from the mean similarity score of the learned match/non-match distributions.

4. We compute p values which are the probabilities of obtaining the test statistic that was observed, assuming that the null hypothesis is true. If $p < \alpha$, we reject null hypothesis.

2.4.2 Identity Verification

In order to examine the validity of the chosen approach, we consider similarity scores of the test image with artworks known to depict persons different from the one depicted in reference image. We call these images as distracters. In cases where enough works of the same artist is not available, we consider similar works of other artists. If a test image indeed represents the same sitter as in the reference image, not only should its score with the reference image be modeled by the match distribution, but also its scores with distracter faces should be modeled by the non-match distribution.

2.4.3 Analysis Scenarios

Following the procedure outlined earlier, we compute similarity scores of test cases with corresponding reference image and with distracters. Table 2.3 lists various hypothesis test scenarios that can arise [64] and the corresponding conclusions that one can infer. Match and non-match cases are straight forward to infer from Table 2.3. In cases where both match and non-match distributions are likely to account for the score in the same way as in green rows of Table 2.3, it can be said that the learned PFS cannot accurately describe the test data. If the match distribution is more likely to account for both test as well as distracters (magenta row in Table 2.3), it can be inferred that the chosen features do not possess sufficient discriminating power to prune outliers. Thus in these scenarios, it is not possible to reach any conclusion.

Reference		Distracters		Conclusion
Match	Non-match	Match	Non-match	
$p > \alpha$	$p < \alpha$	$p < \alpha$	$p > \alpha$	Match
$p < \alpha$	$p > \alpha$	$p < \alpha$	$p > \alpha$	No Match
$p > \alpha$	$p > \alpha$	NA	NA	No decision
$p < \alpha$	$p < \alpha$	NA	NA	No decision
$p > \alpha$	$p < \alpha$	$p > \alpha$	$p < \alpha$	No decision

Table 2.3: Various possibilities for p values of test with reference and distracters. NA stands for Not applicable. These refer to cases where the distracters are not applicable since the similarity score between the test and the reference image is likely to be both a match and a non-match score.

2.5 Dataset

2.5.1 Choice of Images

We have employed a set of images belonging to Western Europe between 15^{th} and early 18^{th} century. These images have been logically chosen by art historians in order to address different tasks such as (a) to test the relation of an unmediated image of the subject, e.g., a death mask to a work of portrait art like a painting, (b) to analyze a number of portraits of different sitters by the same artist to model artist's style, (c) to verify if the identity of the ambiguous subject in a given image is same as that of a known subject in a reference image. The images belong to different media such as drawings, prints, paintings, sculptures, death masks, etc. The dataset consists of works by over 35 artists such as Bernini, Algardi, Clouet, Kneller, etc.

2.5.2 Description

The dataset consists of about 271 images where the identity of the subject is known beyond doubt. There are about 20 test paradigms (with each having multiple image pairs to be compared) where the identity of the subject is in question and has to be compared against the reference image given in that paradigm. Table 2.4 provides a detailed description of the distribution of images in terms of the specific sitter and artist. Fig. 2.3 provides an illustration of the dataset.

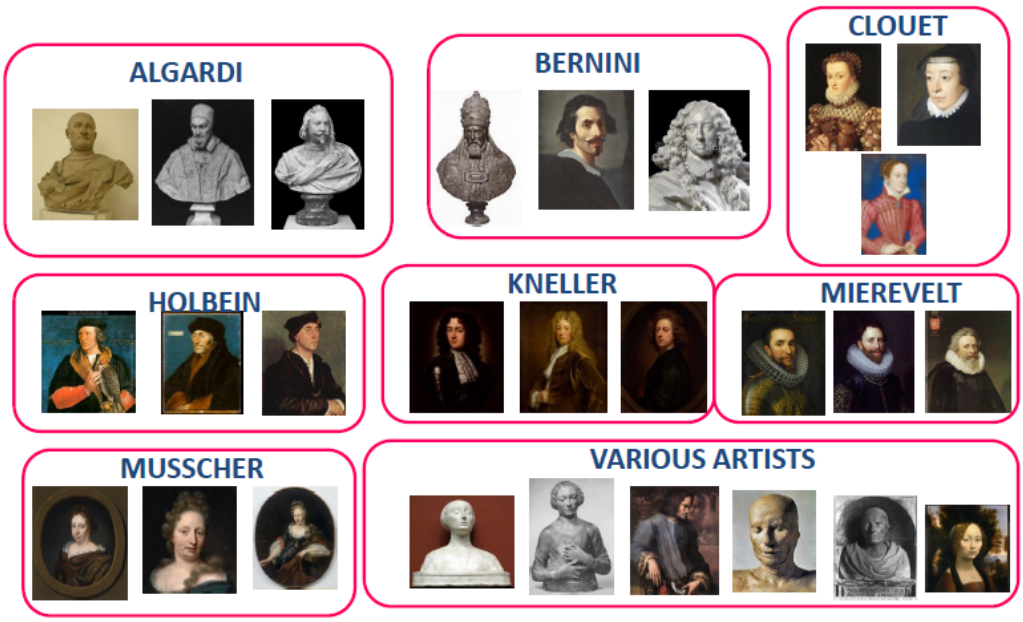

Figure 2.3: Illustration of the dataset across individual/multiple artists depicting different sitters.

2.6 Experiments

2.6.1 Style Modeling Results

We first extracted the 22 LF and 11 AD features for all the images. For those artists where we had enough images to model their style, we learned the features characteristic of their style. Top part of Fig. 2.4 depicts characteristic LF with dots denoting the relative importance of the feature as per the p value of permutation test. AD features representative of the style was similarly determined for these artists; these being AD features 4,8,3,7,2 for Algardi (Please see Table 2.2 for description of numbers), 1, 10, 7, 5,8 for Bernini, AD features 2, 1, 8, 9, 10, 5, 4 for Kneller, 5, 11, 2, 7 for Clouet, 4, 6, 11, 7, 3 for Mierevelt and 2, 8, 11, 3 for Holbein. Features are listed in decreasing order of importance for each artist. We verified the validity of these features using the p_{st} value computed from Siegel-Tukey

Artist	# Images	Artist	# Images
Algardi	14	Giotto	6
Bandini	1	Hansen	3
Bernini	33	Holbein	45
Botticelli	9	Kneller	19
Bronzino	5	Langel	1
Buggiano	2	Laurana	10
Cafa	2	Mantenga	3
Campin	4	Masaccio	4
Clouet	14	Raphael	5
da Fiesole	5	Signorelli	5
Da Vinci	7	Sittow	4
De Champaigne	7	Stringa	4
De Benintendi	3	Thronhill	3
Del Castagno	3	Torrigiano	1
Della Francesca	4	Van Mierevelt	24
Vasari	4	Van Musccher	18
Ghirlandaio	5	Verrocchio	6

Table 2.4: Illustration of image distribution : Number of images per artist.

test. As illustrated in bottom part of Fig. 2.4, for almost all cases, the confidence of the similarity scores increased upon using only the style features, thus validating the chosen LF. Similar results were obtained for AD features. It is to be noted that the Siegel-Tukey test validates both style-specific match and non-match scores; wherever there are not enough images to obtain match scores, only the available non-match scores are validated. The receiver operating characteristic (ROC) curve shown in Fig. 2.5 compares the performance for pair-wise sitter validation upon using (a) style features (b) all LF/AD features. The ROC demonstrates the improvement in pairwise validation upon using style features.

Significance of Style Modeling: These results could possibly aid art historians in attributing works to an artist that was not attributed to him/her before. Further, it could also help in identifying unrecognized portraits by these artists more confidently. It might

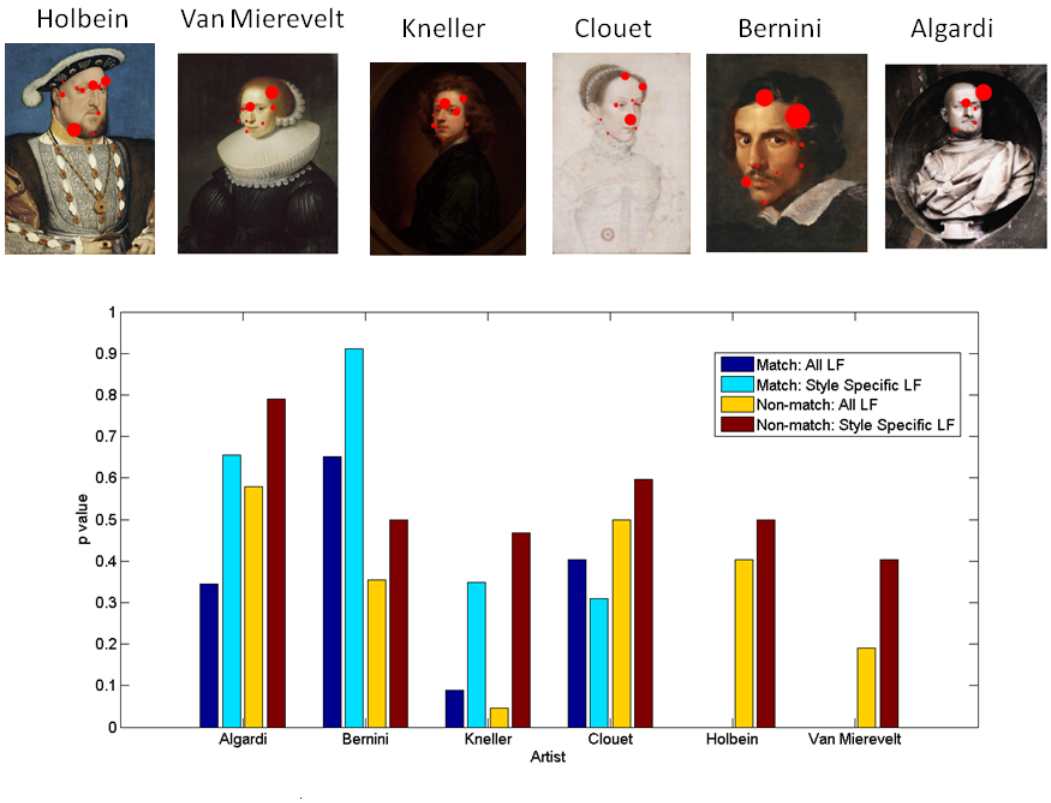

Figure 2.4: Top: Importance of chosen features with bigger dots indicating more important features; Bottom: Validation of style through Siegel-Tukey test

also be possible to understand the adherence to artistic canon and individual variations in art practices.

2.6.2 Validation with Known Sitters

From the set of known identities, we obtained match and non-match scores. It is to be noted that wherever an artist's style could be modeled, we used only those (weighted) features in obtaining the LF/AD similarity scores and otherwise used all the LF/AD features

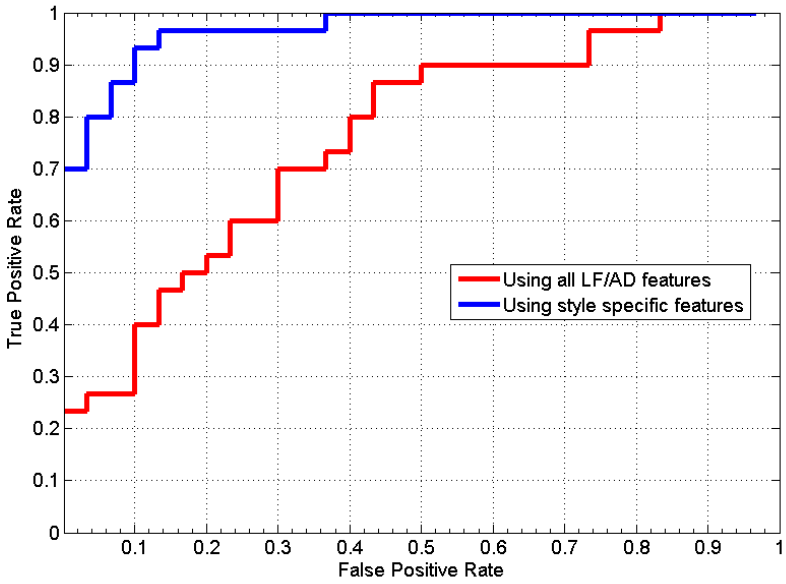

Figure 2.5: Pairwise sitter validation upon using style features.

followed by the feature combination strategy to fuse the similarity scores. The weight for LF feature was found to be 0.55 and that for AD features were 0.45. Experiments showed that there was improvement in the performance upon fusing scores from LF and AD as against using any one of them. The values of mean of PFS were 0.7246 (match) and 0.5926 (non-match) with standard deviations 0.043 and 0.052 respectively (See Fig. 2.6). As described earlier, the improvement in using style features as against all LF/AD features is evident from Fig. 2.5. Some notable tests that were correctly validated include comparison between a pair of busts by Bernini depicting Urban VIII, comparison of busts of Alexander VII by artists Bernini and Cafa, and comparison between a pair of self portraits by Bernini.

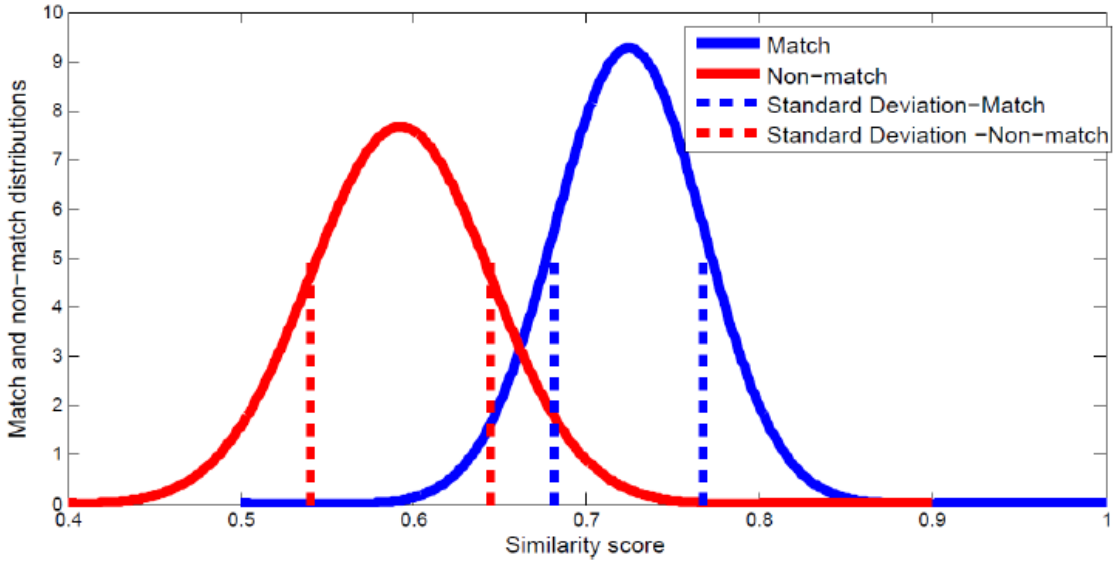

Figure 2.6: PFS showing the distribution of match and non-match scores along with their standard deviations.

2.6.3 Identity Verification

We want to provide quantitative measures of similarity to uncertain test paradigms provided to us by art historians. In this, we do not claim to provide the incontestable identity of the sitter in question, but to only provide a complementary viewpoint, which could serve the art history community.

Significance of Results from Art Perspective: In these identification tests, support was given to previous scholarly opinion on a number of important cases. Among these were the posthumous bust of Battista Sforza by Laurana in the Bargello and a death mask cast also by Laurana in the Louvre shown in col.1 of Fig.2.7. A match suggests that, as was thought, the mask was that of Battista. It also supports the idea that the cast was quite closely followed by Laurana as a model–rather than, say, Piero della Francesca's profile

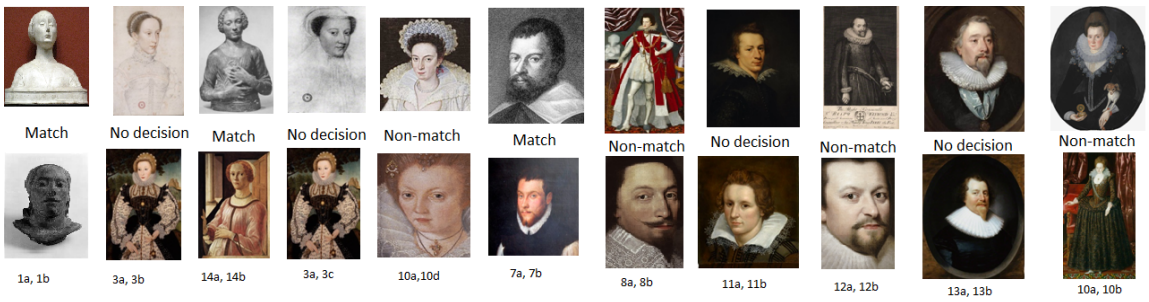

Figure 2.7: Illustrations of identification tests with conclusion in center. Bottom row shows images whose identity is uncertain; numbers refer to corresponding images in Appendix B.

portrait of Battista. A match was also indicated for Botticelli's Portrait of a Lady at the Window (c. 1475; widely thought to be a rendering of Smeralda Brandini) and Verrocchio's Lady with Flowers (c. 1475), the two portraits also sometimes being suggested by some to represent the same sitter, thus lending objective support to this position despite the two distinctly different personas conveyed in the images.

Tests strongly support the traditional supposition that Nicholas Hilliard's Young Man Among Roses, said to be "perhaps the most famous miniature ever painted," represents Robert Devereux, second earl of Essex. The results of test scores between a portrait of a woman at the National Portrait Gallery in London thought by some to represent Mary Queen of Scots and eight other portraits known to be of Mary were almost startling in their support for the identification of the unknown portrait as Mary. Results also lend new support to previous opinion that the portrait at the National Portrait Gallery thought by some to depict James Scott, Duke of Monmouth, first Duke of Buccleuch, does portray Monmouth lying in bed after his beheading for treason.

The portrait shown in bottom row of Col. 6 in Fig. 2.7 was sent to us by the Italian astronomer Paolo Molaro, of what he believes may be the earliest known likeness of Galileo Galilei, painted perhaps around 1590. When tested against a chronological spectrum of eight other known portraits of Galileo, the results gave decreasing similarity scores within the match range for the chronologically three closest likenesses (1601-1612). Thus, the test gives support to the identification of a previously unrecognized portrait as Galileo–possibly the earliest known portrait of Galileo. While age remains a challenge for FACES and requires more research, age differences of around ten years or so have not been too much of an obstacle.

A comparison between an unknown painting attributed to de Neve against a known portrait of George de Villiers, 1st Duke of Buckingham (col. 7, Fig. 2.7) and a comparison between an unknown portrait against a known portrait of Lady Arabella Stuart (col.5, Fig. 2.7) gave non-match scores. A list of identification paradigms with results is provided in the supplementary material. For a detailed description of these results from the art perspective, please refer to [65].

The results of *FACES* are only as dependable as the images tested. Areas that would benefit from further research include modeling wide age differences, strong angle views (including profile images) and even the use of different media (e.g., terracotta as opposed to marble, chalk in contrast to oil, etc.).

2.7 Conclusions

We presented a work that explores the feasibility of computer based face analysis for portraiture. After a careful understanding of artistic conventions, we arrived at relevant features for analysis. Subsequently, using machine learning tools, we learned a feature space describing the distribution of similarity scores for cases known to match/not match and also validated the same. We proposed a novel method to model artists' styles and to analyze uncertain portrait pairs. We believe that these results can serve as a source of complementary evidence to the art historians in addressing questions such as verifying authenticity, recognition of uncertain subjects, etc. As future work, we would like to explore modeling age variations in portraits and building family trees of artists/sitters.

Chapter 3

Robust Face Recognition Using Saliency Maps of Covariance Descriptors

In this chapter, we propose a robust unsupervised method for face recognition wherein saliency maps of second order statistics are employed as image descriptors. In particular, we leverage upon region covariance matrices (RCM) and their enhancement based on sigma sets for constructing saliency maps of face images. Sigma sets are of low dimension, robust to rotation and illumination changes and are efficient in distance evaluation. Further, they provide a natural way to combine multiple features and hence facilitate a simple mechanism for building otherwise tedious saliency maps. Using saliency maps thus constructed as the face descriptors brings in an additional advantage of emphasizing the most discriminative regions of a face and thereby improve recognition performance. We

demonstrate the effectiveness of the proposed method for two face recognition applications (a) face photo-sketch recognition, wherein we achieve performance comparable to state-of-the-art without having to do sketch synthesis, and (b) face recognition in unconstrained environments subject to additional artificial distortions.

3.1 Introduction

Face recognition is a form of biometric identification [4] involving recognition of individuals based on the salient characteristics of their face images. A large number of real world applications such as law enforcement, surveillance, gaming industry, healthcare, and banking utilize face recognition. As a result, there has been enormous interest in this area of research. A variety of challenges are associated with a typical face recognition task. It has been observed that the performance of several state-of-the-art face recognition methods degrades to a large extent in unconstrained and heterogenous environments.

One effective approach to alleviate these limitations is by designing highly discriminative, robust and yet computationally simple image features that can be efficiently matched. Typically, features are extracted based on expert knowledge [9] and/or learned from available training data [10]. Low level features such as color, gradient and filter responses are the simplest choices [84, 85] but they are not robust in the presence of illumination changes and non-rigid motion. Although statistical descriptors such as histograms of oriented gradients show robustness to rotation and translation, their dimensionality grows exponentially with feature dimensions.

Region covariance matrices, which are the second order statistics of a region, were first proposed in [89] as an image descriptor wherein its efficiency was demonstrated for texture classification. RCMs are shown to be robust to rotation/illumination changes and are also low dimensional compared to histograms. Furthermore, they provide a natural way of fusing multiple features. However, distance computation between RCMs is cumbersome as they do not lie on an Euclidean manifold. While there are Riemannian manifold based alternate distance computation strategies such as [83], they are usually time consuming. In order to overcome the aforementioned limitation of RCMs, a novel feature descriptor that possesses the effectiveness of second order statistics but which has an efficient distance computation scheme, was proposed in [75]. Sigma sets, as these descriptors are called, are obtained by performing a Cholesky decomposition of the RCMs and are thus equivalent to RCMs. They are low dimensional and robust. Since these descriptors are in the form of a set of points, distance evaluation is computationally simple unlike RCMs. In [51], the effectiveness of sigma setswas demonstrated for face recognition wherein we learnt scales specific to a region by means of block-wise scale selection. In this work, we explore if constructing saliency maps of sigma sets can achieve better performance and overcome the need to learn to region-specific features.

3.1.1 Motivation for the Use of Saliency Maps of Sigma Sets

Often in applications such as face recognition, it helps to emphasize the most discriminative regions in order to achieve superior recognition performance. Consider, for example, comparing a face photo to a sketch. It is very likely that the sketch artist would

Figure 3.1: Top row: Saliency maps help in emphasizing the most informative regions across the photo and its sketch and hence can be leveraged for face photo-sketch recognition. Bottom row: Saliency maps of a face image and its corrupted version. Despite the distortion, saliency map effectively captures the most discriminative regions in the face.

have emphasized the most prominent features of the face such as a pointed nose or long ears. Thus, prominent features can be understood to be those features which attract a viewer's attention almost instantly. Formally speaking, this refers to the concept of visual saliency. Visual saliency, or more generally known as visual attention, refers to a process which detects scene regions that stand out relative to their neighbouring parts. It mainly revolves around the principles of selection mechanisms and relevance to help determine the most characteristic regions of an image [30].

As an illustration, consider top row of Fig. 3.1, which shows a face photo, its corresponding sketch and their respective saliency maps computed from sigma sets of filter responses. The photo and sketch are part of the Chinese University of Hongkong (CUHK) dataset [41]. As can be noticed from the figure, the most discriminative regions, namely the

eyes and nose, are captured in the saliency maps of both the sketch and the photo. Bottom row of Fig. 3.1 illustrates the effectiveness of saliency maps in capturing most informative and discriminative regions even as the face images are subjected to artificial distortions. The images are part of the Quality Labeled Faces in the Wild (QLFW) dataset wherein face images are subjected to artificial distortions such as Gaussian blur and white noise [42]. These illustrations motivate us to employ saliency maps of sigma sets as the image descriptors.

Building saliency maps from sigma sets also provides an additional advantage of a simple mechanism for feature integration. Typically, in computing saliency maps, some basic visual features are extracted to form feature maps for each dimension, after which these individual maps are integrated to build a master saliency map. Different feature integration strategies (both linear and non-linear) have been proposed based on how individual feature dimensions contribute to the overall saliency. One of the main advantages of employing sigma sets constructed from RCM as feature descriptors is that they provide a natural way to integrate different feature maps by modeling their correlations.

3.1.2 Contributions

Thus main contributions of the paper can be summarized as follows. We propose a robust unsupervised face recognition scheme by employing saliency maps of sigma sets. The proposed descriptors can overcome variations in pose and illumination, are of low dimension and provide a natural way of integrating different features. The saliency maps emphasize the most discriminative regions and facilitate improved recognition performance.

The advantages of the proposed method is demonstrated for two different face recognition applications namely, (a) face photo- sketch recognition, without having to perform sketch synthesis and (b) face recognition in the presence of artificial distortions.

3.1.3 Related Work

Over the last few years, RCMs have become popular as a new set of holistic image descriptors. RCMs have been applied to many applications such as texture classification and object detection [95] andobject tracking [97].These descriptors have shown best discriminative power for human detection tasks [98] and have also been recently studied in the context of face recognition [99].

Based on the manner in which salient regions are determined, saliency models can be categorized into two broad groups, namely bottom-up models and top-down models. Regions of interest that are mainly driven by involuntary stimulus (e.g., a black line amidst many red lines) attract attention in a bottom-up manner [31]. These are typically determined by intrinsic low-level properties of the scenes. On the other hand, regions of interest that are propelled by a goal (e.g., searching for an object in a scene, etc.) attract attention in a top-down manner. These are usually determined by cognitive phenomena like rewards, expectations, etc. [32]. Some models integrate low-level bottom-up cues with some task-specific top-down knowledge such as face and object detectors [36] to improve their predictions.

There are a number of computational approaches to model saliency. Most of them consider bottom-up strategies in a supervised data-driven manner to compute the

saliency maps [33], [34], [35]. Recently, unsupervised saliency learning approaches have been leveraged such as in [37] for person re-identification and in [38] for discovering new object categories. In general, modeling visual saliency has provided efficient strategies for various other computer vision applications such as image segmentation [27], object recognition [28], visual tracking [29], etc. Despite the considerable amount of work on saliency for different computer vision applications, there is very little work on saliency for face recognition [39], [40].

3.2 Proposed Method

From the face image, simple low level features such as filter response is extracted. Region covariance matrices are built from these features by considering the correlations between different features within specific regions. A Cholesky decomposition of the RCMs yields sigma sets. Saliency maps are then constructed from sigma sets. We begin by explaining the Region Covariance Matrices (RCM) and the sigma sets followed by the saliency estimation procedure. Fig. 3.2 illustrates these steps.

3.2.1 Region Covariance Matrices

The region covariance matrix (RCM) of a local region in an image with respect to a set of low level features is essentially the local second order statistic of the given features. A major advantage of the covariance matrices is that they have much lower dimensionality when compared to the exact distributions of low level image features. RCMs are robust to small pose variations and provide a natural way for fusing multiple features which might

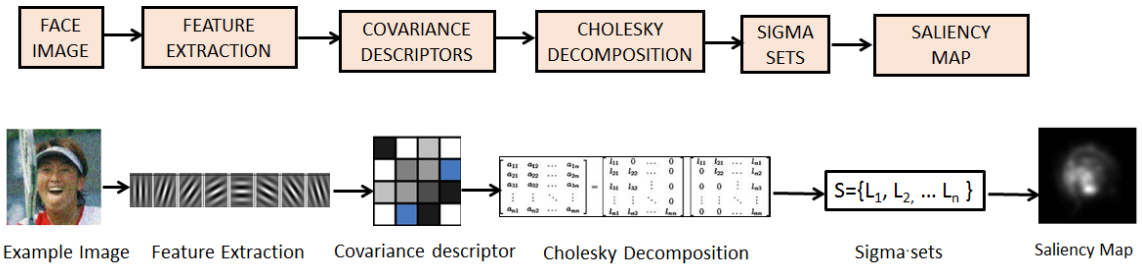

Figure 3.2: Steps involved in obtaining the proposed image descriptors: Top row depicts a block representation and bottom row provides a pictoral illustration.

be correlated.

Let I denote a face image. For a given rectangular region $R \in I$ with N pixels, let $\vec{f_i}$ be a y dimensional feature vector (consisting of Gabor responses) extracted from the i^{th} pixel in R , and $\vec{u}$ be the mean vector of the set of feature vectors $\vec{f_i}$ in R. The $y \times y$ covariance matrix $C(R)$ of R can be calculated as

$$C(R) = F_R F_R^T \tag{3.1}$$

where $F_R = [\hat{f}_1, ...\hat{f_N}]$ denotes the matrix of centered vectors $\hat{f}_i = \frac{1}{\sqrt{(N)}}(\vec{f_i} - \vec{u})$.

However RCMs do not lie in a Euclidean space, which makes distance computation between them hard. Different measures based on Riemannian geometry such as [83] were proposed to compute distance between RCMs. However, they are computationally inefficient as they involve computing matrix exponential and logarithm operations. The sigma set descriptors were proposed to overcome this drawback.

3.2.2 Sigma Sets

Sigma sets represent the covariance matrix as a small set of points S, that have same covariance values as the given matrix [75]. In other words, the set of points S is

equivalent to R in terms of second order statistics. Mathematically, it follows that for any matrix L that satisfies $C(R) = LL^T$, the set of columns of L has the same second order statistic as R. One way to obtain such a decomposition of the covariance matrix is using Cholesky decomposition. A Cholesky decomposition is used to represent any symmetric positive-definite matrix, such as a covariance matrix, as a product of a lower triangular matrix and its transpose. The fact that the component matrix of a Cholesky decomposition is lower triangular, it imposes an order on the set of points it represents and this is very helpful in devising a simple distance metric between two sigma sets.

The sigma set computation can be summarized algorithmically as follows:

Given: A face region R consisting of $N, y \times 1$ feature vectors.

Output: Sigma set $S = [L_1, ..., L_y]$ satisfying $C(S) = C(R)$.

Algorithm:

1. Calculate the $y \times y$ covariance matrix $C = C(R)$ of the face region R.

2. Perform Cholesky decomposition of C, $C = LL^T$, where L is a lower triangular matrix.

3. Multiply L by the scalar $\sqrt{y}$, i.e., $L = \sqrt{y} \times L$.

4. $S = [L_1, ..., L_y]$ where L_i is the i^{th} column of L.

Distance measure

The distance between sigma sets can be evaluated as summation of point to point distance and is equivalent to modified Hausdroff distance (MHD) [96], a widely used distance metric over closed and bounded sets. Given two sigma sets S_A and S_B, the modified Hausdroff distance is defined as

$$H(S_A, S_B) = \max\left\{h(S_A, S_B), h(S_B, S_A)\right\}, \tag{3.2}$$

where $h(S_A, S_B) = \frac{1}{2y}\sum_{a \in S_A} \min_{b \in S_B}(d_E(a, b))$. and $d_E(.)$ is some distance metric. Further, due to symmetric property of covariance matrices and their Cholesky decomposition,

$$h(S_A, S_B) = \frac{1}{y}\sum_{i=1}^{y} \min_{j=1}^{y}(d_E(L_i^A, L_j^B)), \tag{3.3}$$

where L_i^A and L_j^B,$i = 1, ...y$ denote the i^{th} and j^{th} points in S_A and S_B respectively.

Since the structure of the sigma set enforces the first i elements of i^{th} sigma point to be zero, the difference between two sigma sets can be obtained by the accumulation of differences between the corresponding non-zero elements in the columns of L. Thus, we can assume that

$$d_E(L_i^A, L_j^B) = \infty, i \neq j. \tag{3.4}$$

Thus, given two sigma sets S_A and S_B, the distance between them becomes

$$h(S_A, S_B) = \frac{1}{y}\sum_{i=1}^{y} d_E(L_i^A, L_i^B), \tag{3.5}$$

and the modified Hausdroff distance becomes

$$H(S_A, S_B) = \max\left\{\frac{1}{y}\sum_{i=1}^{y} d_E(L_i^A, L_i^B), \frac{1}{y}\sum_{i=1}^{y} d_E(L_i^B, L_i^A)\right\}, \tag{3.6}$$

Enriching Sigma Sets Feature Descriptors

Once sigma sets are computed, first order statistics such as the mean (μ) can be easily appended to the sigma sets, thus resulting in an enriched representation. Although sigma sets can effectively encode local structure information by using the second-order statistical relations among features, first-order statistics (mean) can be also valuable in capturing saliency of an image region with respect to its surroundings. This is particularly evident while dealing with contrast images. Since the means are different for low and high contrast regions, an analysis based on first-order statistics would make this region pop out from its surroundings[87]. The enriched feature descriptor is denoted as

$$\psi(R) = [S, \mu]^T, \tag{3.7}$$

where S is the sigma set of the region under consideration.

Computational Complexity of Sigma Sets

The amount of computations required for Cholesky decomposition (CD) of a $y \times y$ matrix is $O(y^3)$ which is relatively small given that most of the elements are zero.

3.2.3 Saliency Estimation

Similar to the definition in [87], we measure the saliency of a pixel by how much the pixel differs from its surroundings. In order to estimate this, we compared each rectangular image region (local neighborhood of a pixel) against its immediate context described by the nearby regions. The image regions are described by their RCMs, which provide a

natural way for non-linear integration of features using second-order statistics. We begin by explaning the covariance descriptors.

Given an input image I, we consider non-overlapping regions of square blocks, which are of size k × k pixels. The saliency of a block is estimated by comparing it with its immediate neighbouring blocks. If it locally displays distinct characteristics, it is regarded as salient. Specifically, saliency of a region R_i is determined as follows.

$$T(R_i) = \frac{1}{m}\sum_{j=1}^{m} d'(R_i, R_j) \tag{3.8}$$

wherein m denotes the number of immediate neighbors of a block and

$$d'(R_i, R_j) = \frac{||\psi(R_i) - \psi(R_j)||}{1 + ||x_i - x_j||} \tag{3.9}$$

Here, $\psi(R)$ denotes the enriched feature descriptor after incorporating mean with sigma sets, and x_i, x_i are the pixel centers for regions i, j respectively.

The saliency maps computed thus constitute the feature descriptor for each face. Constructing feature descriptors in the aforementioned manner provides a natural way of emphasizing prominent regions of a face. Thus, computing distance (e.g., 2D euclidean distance) between two such saliency based face descriptors is an efficient way of comparison. We employ 2D euclidean distance to compare two saliency based face descriptors.

3.3 Experiments

We demonstrate the utility of the proposed feature descriptor for two different face recognition applications. In the first, we study face photo-sketch recogniton wherein unlike existing methods, we do not need to perform an otherwise tedious sketch synthesis process.

Next, we discuss the benefits of saliency maps of sigma sets for face recognition in presence of distortions. We begin by providing a brief description of the datasets employed.

3.3.1 Datasets

For the purpose of demonstrating the advantages of using the saliency maps as feature descriptor in face photo-sketch recognition, we used the Chinese University of Hongkong (CUHK) Face photo-sketch dataset [41]. It includes 188 faces where for each face, there is a sketch drawn by an artist and a photo taken in frontal pose under normal lighting condition and neutral expression.

For the purpose of showing the effectiveness of saliency maps for face recognition under unconstrained environments, we employed the Quality Labeled Faces in the Wild (QLFW) dataset [42], which has over 13000 images. The QLFW database is derived from the LFW database. In the QLFW database, the LFW images are subjected to different types and levels of distortions, simulating distortions that can occur under real-world conditions, including impairments due to Gaussian blur and white noise. For each visual impairment type, the level of impairments were chosen such that the whole range of visual quality is represented from Poor (strong perceived impairment as compared to original source) to Excellent (no perceived impairment, original source).

3.3.2 Implementation Details

All the face images were cropped to size of 200×200 and subsequently divided into 8×8 blocks. We did not employ any pose correction on the images. We used Gabor

filters that essentially aggregate directional gradients and are able to effectively capture facial features. We used Gabor wavelets across 8 orientations and 2 scales to get a 16×16 dimensional RCM for each block in the image. From these, we constructed the sigma set descriptors by performing a Cholesky decomposition and subsequently appended the mean value of the region under consideration to obtain the enriched feature descriptor as described in eq. (3.7). We then computed the saliency maps by measuring the amount by which each region is different from its neighbors using eq. (3.8), the neighbors being the immediate adjacent blocks of the block under consideration. Thus, there can be a maximum of 8 neighbors and a minimum of 3 neighbors for any block.

3.3.3 Face Photo-Sketch Recognition Results

Many face photo-sketch recognition methods such as [41] first perform a sketch synthesis. Typically, from a training data which contains photo-sketch pairs, the joint photo-sketch model is learned at multiple scales. Subsequently, photo-sketch transofrmation is carried out after which recognition approaches can be applied in a straightforward way. Unlike such works, we do away with the sketch synthesis altogether. Since the saliency maps naturally highlight the most discriminative regions which are likely to be the same in both photo as well as the sketch, these maps can be directly employed to perform recognition. We compute 2D Euclidean distance between saliency descriptors of faces to obtain similarity scores. Subsequently, we use a nearest neighbor classifier for recognition.

Table 3.1 lists the rank-1 to rank-10 accuracy on the CUHK dataset and compares the performance with that of [41]. It can be noted that the proposed method provides

Table 3.1: Comparison of the proposed method with state-of-the-art in face photo-sketch recognition shown in terms of rank 1 to rank 10 percentage accuracies.

	1	2	3	4	5	6	7	8	9	10
[41]	96.3	97.7	98	98.3	98.7	98.7	99.3	99.3	99.7	99.7
Proposed	96	98	98.3	98.3	99.3	99.3	99.7	99.7	99.7	99.7

nearly similar performance to that of [41] *without having to do any sketch synthesis.* While the first match accuracy is better for [41], it is to be noted that the proposed method is only slightly less despite the fact that it does not require sketch synthesis. It is to be noted that rank-2 through rank-10 accuracies of the proposed method is either greater or on par with [41]. Given that the proposed method is unsupervised, the advantages are considerable not only in terms of recognition accuracy, but also in terms of computation.

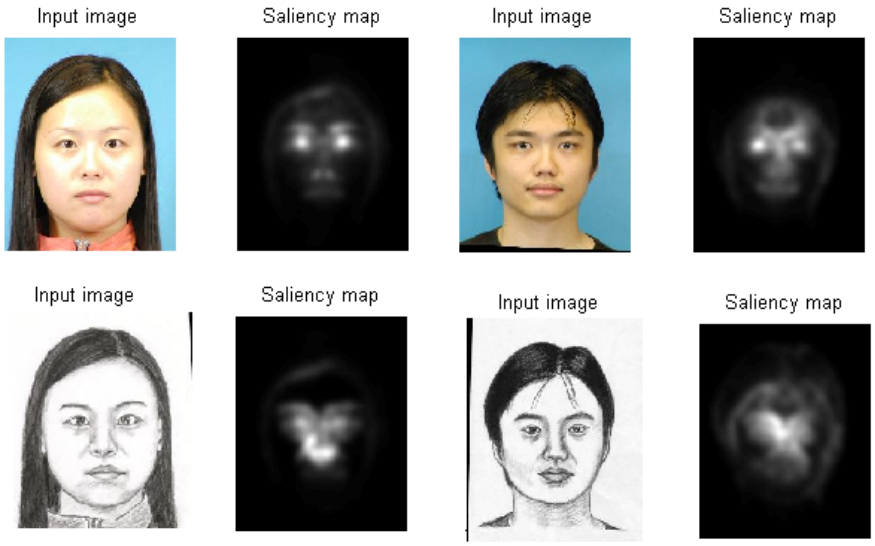

Figure 3.3: Illustration of the saliency map for photo and sketches across genders.

Some illustrations of the saliency maps of photos and sketches across genders can be seen in Fig. 3.3. While training to select salient regions particular to a gender can certainly improve performance, nevertheless the proposed unsupervised method is quite

effective in capturing some important regions across both the instances. Certain noticeable differences can be attributed to the fact that sketches are not identical to the photo.

3.3.4 Face Recognition in Presence of Distortions Results

We performed experiments on the QLFW dataset with Gaussian blur and white noise added on a scale of 1 to 4. Fig. 3.4 provides an illustration of images of the same person which have been corrupted with varying level of distortions. 1 denotes least distortion and 4 indicates highest amount of distortion. Gaussian blur and white noise are the two types of distortions shown. The corresponding saliency maps can be seen next to each image. As can be observed, despite the severe distortion induced in the images, the saliency maps are indeed very effective in consistently capturing the most informative regions across all levels of distortions.

Table 3.2 lists the pairwise recognition performance for varying levels of distortions. As the level of distortion increases, the recognition accuracy reduces for both types of distortions. The performance decreases drastically for white noise when compared to Gaussian blur. This can be attributed to the nature of images obtained after adding these distortions. There are no results reported yet on the QLFW for a direct comparison. Since we are working in the image restricted scenario with no outside training data for alignment, the closest setting for comparison happens to be the performance of sigma sets on the image restricted scenario of the LFW dataset. We noticed that there was almost 1.5% improvement upon using saliency maps of sigma sets than using just the sigma sets [51]. This clearly demonstrates the need to incorporate saliency maps of sigma sets.

Table 3.2: Pairwise recognition accuracies (in %) of the proposed method for varying levels of different distortions.

Distortion Type	Level 1	Level 2	Level 3	Level 4
Gaussian blur	84.12	71.38	60.22	52.84
White noise	83.62	70.12	59.23	41.71

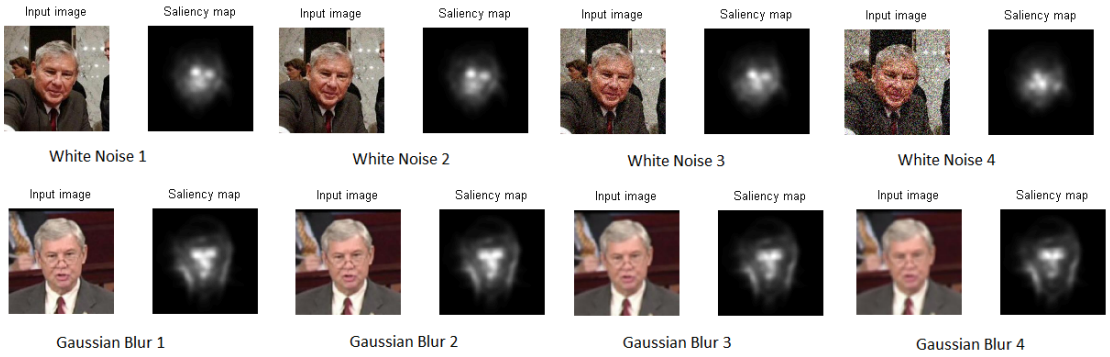

Figure 3.4: Saliency maps across various levels of different distortions. Even as the distortion level increases, the saliency maps of sigma sets are more or less similar suggesting that these descriptors are very robust for recognition in presence of distortions .

3.4 Conclusions

We analyzed the feasibility of a new set of face descriptors based on the saliency maps of sigma sets which are in turn constructed from simple filter responses. While sigma sets are low dimensional and robust to variations in pose and illumination, saliency maps effectively emphasize the most discriminative regions of the face, thus ensuring better recognition accuracies. Experiments showed promising performance for face photo sketch recognition without the need for sketch synthesis. We also demonstrated the effectiveness of these descriptors for face recognition in presence of distortions on the QLFW dataset wherein various types of distortions are added.

Chapter 4

Undersampled Face Recognition via Style-Content Modeling

In this chapter, we address the problem of undersampled face recognition by means of modeling style and content from very limited training data. Modeling style and content is a core vision problem that has been explored through bilinear models (BLM) where style can be attributes like facial expression, gait, etc. and content is usually the person's identity. Since most existing techniques require significant training data containing multiple instances of object/person's appearance, they are not robust when there are very limited training samples in a style class (e.g., < 10). We propose a novel algorithm for automatically inferring style and content from *very limited training samples* and demonstrate the advantage of knowing such information in pairwise face recognition. Further, unlike many state-of-the-art undersampled face recognition methods which leverage outside training data, the proposed method does not need outside training data and works with very limited training instances.

A set of weighted features characteristic of a style class are learned, wherein the weights denote the importance of the chosen features in the class. The chosen style features are then validated by means of the robust Siegel-Tukey statistical test that is proven to work well for small sample sets. We analyze the PubFig and Yale dataset using the annotated attributes as style factors to illustrate the advantages of proposed method over BLM and relative attributes method as sample size in the style class is reduced.

4.1 Introduction

Many applications involving law enforcement such as finger print analysis, handwriting analysis, face recognition, forgery detection, etc. can only offer a few or even a single image of the suspect due to non-availability of examples. In addition, sparsity of training data is common to many other forensic science applications such as studies of ancient art, architecture, biology or archaeology wherein experts are often required to answer questions related to authenticity such as for date estimation, identity verification, etc. In fact, practical illustrations of such scenarios can be found in [49, 11, 71] wherein the authors explain the difficulty in acquisition of images owing to their cost, availability and authenticity.

There have been works that have explored the problem of face recognition from limited training data, often referred to as undersampled face recognition. A number of these works leverage upon a larger set of images outside the gallery such as composite sketches [80], mugshot gallery images [82], or assume that the intra-class variations of one subject can be approximated by a sparse linear combination of those of other subjects and represent the variation between the training and testing images using additional images [71].

We propose a different approach to the aforementioned problem wherein we do not make use of any images from outside the gallery. The motivation for our approach comes from the fact that several vision problems are often characterized by two or more independent factors which interact to produce an observation. For example, in face recognition, a facial expression and the identity of the person interact or combine to yield the rendered face image. We may perceive one factor (facial expression) as the style and the other (identity of the person) as the content. In general, style can be various other attributes such as pose [66], gait [68], etc. The question we ask is *can we do style and content modeling from very limited training data and then use this information for face recognition?* Existing style-content analysis methods such as the bilinear/multilinear models [66, 67, 68, 69] are not applicable since these approaches rely upon obtaining multiple instances of object/person's appearance under various conditions (pose, expression, etc.), which are hard to procure in aforementioned scenarios.

Given limited training data ($\leq$ 10 per style class), our approach leverages upon statistical hypothesis testing to learn and validate style information. The choice of statistical hypothesis testing is motivated by the fact that certain non-parametric tests such the Siegel-Tukey test [61] do not assume any underlying distribution for the data while testing a claim. A claim can be any proposition about the images under consideration. For example, one such claim could be that two face images belong to a common style and represent the same person. Furthermore, the non-parametric permutation test [58] can help assess which features are more important to a style. As an instance, permutation test can help identity features such as the eye corners and mouth tips as being more important than features such

as forehead tips for images belonging to the style class "smile". The absence of normality is a particularly important feature of such tests given that we do not have enough training data to assume an underlying distribution. In fact, principles behind these methods have been used to address several problems in biology such as for investigating protective effect of treatment against infections in sparse samples [59].

In order to ensure robust hypothesis testing for learning the importance of different style features and for validating them, care has to be taken in initially selecting highly discriminative features characteristic of individual styles. In this direction, we make use of the random subspace ensemble learning techniques [56]. Specifically, the random subspace method is capable of handling deficiencies of learning in small sample sizes and has superior performance than a single classifier since an aggregation of the output of individual classifiers leads to reduction of variance in error. Further, diversity among weak classifiers contributes to robustness.

Contributions: In this paper, we address the *problem of undersampled face recognition via style-content modeling*. In particular, we propose a novel framework that leverages robust non-parametric statistical hypothesis testing strategies such as permutation tests [58] and Siegel-Tukey test [61] in conjunction with ensemble learning techniques such as random subspace [56] to learn and validate style-content from very limited training data. We demonstrate the advantage of knowing the style information in pairwise content recognition [i.e., in determining if 2 persons are same (match) or not (non-match)]. It is to be noted that the two face images in question could belong to a common style or depict different styles. We illustrate the proposed algorithm by considering the application of undersam-

pled face recognition in the wild and illustrate the results on two datasets [73], [72]. We compare the performance of the proposed algorithm against two baselines. The first is the standard bilinear models [66]. Since style could often be attributes such as an expression, we also compare our method against relative attributes [48], a very popular attribute based classification method.

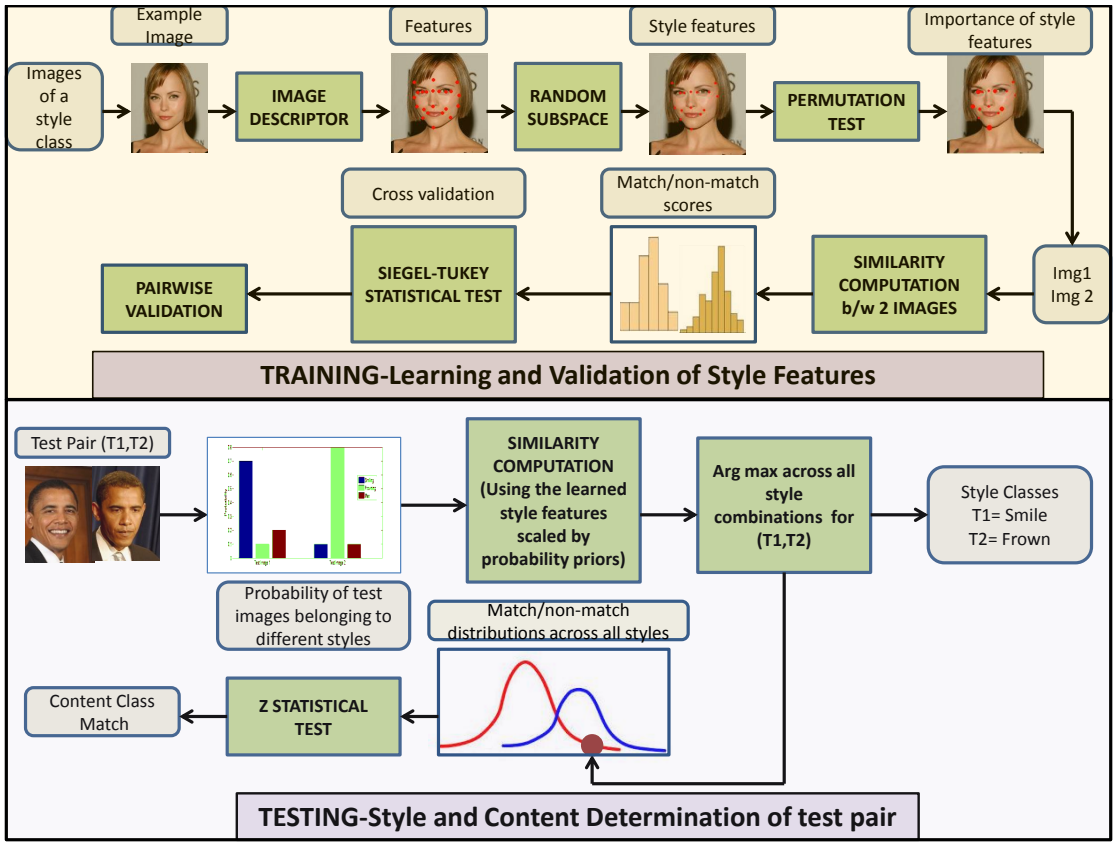

Figure 4.1: Overview of the proposed method for infering style and content from sparse image samples.

4.1.1 Overview

From a very limited set of images known to belong to a style, region covariance matrices (computed from the second order statistics of a set of low level descriptors) [89] represented in the form of sigma sets [75] are extracted and averaged across all images belonging to that style to yield characteristic descriptor of the style. These low level descriptors could be filter responses, intensity, color, etc.

Next, we consider a set of salient points on the image and extract the low level descriptors around these points. The descriptors around a salient point constitute a feature. We automatically learn a subset of weighted features that are characteristic of a style wherein the weights denote the relative importance of the chosen features in the style. Towards this, we first employ the random subspace method to arrive at a subset of features that are characteristic of a style. Since not all the chosen features will be equally important for a style, we then employ the statistical permutation tests to learn the relative importance (weights) of the chosen features. The chosen features are then used to obtain a similarity score between pairs of instances in each style to yield match (if the image pairs denote the same person) or non-match scores (if the image pairs denote different people) as appropriate. The chosen features are validated by showing that there is significant improvement in the confidence of the style-specific similarity scores over style independent similarity scores (obtained using all features) by means of the Siegel-Tukey test [61]. Siegel-Tukey test does not assume any underlying distribution for the similarity scores and hence applicable for this scenario wherein we have very few scores.

Given test image pairs (whose style and content is unknown), probabilities of the images to belong to each of the styles is estimated by considering their sigma set descriptors with respect to the learned sigma sets in a mixture model framework. Subsequently, these probabilities are used as scaling factors in computing the similarity scores between the image pairs. The styles of the images is then determined by the maximum similarity score across all combinations of styles. The content (match/non-match) is then determined by computing the p value of the test pair with respect to the entire set of match and non-match scores. Fig 4.1 illustrates the approach.

4.2 Related Work

We review related work on style-content modeling, attribute based methods and methods which deal with learning when the training data is very limited.

4.2.1 Style-Content Modeling Methods:

A model for separating content from style was introduced in [66], wherein promising results were shown for face pose estimation, among others. In [68], separating content from styles on non-linear manifolds was addressed. In [69], face transfer based on multilinear models was considered. However, all of these require significant training data. The authors in [81] propose to learn complex manifolds of object appearance from a small set of training data and show that their method overcomes the limitations of bilinear/multilinear models. However, the method requires the 3D shape of the object. In contrast, our approach can perform well with few training instances and does not need 3D shape information.

4.2.2 Learning in Sparse Samples:

In [76], the authors exploited qualitative influences as constraints on probability distributions in Bayesian learning to learn from few samples. In [77], a novel graphical model based framework was proposed for consumer produced sparse multimedia. To deal with sparse data, training images were shared among semantically similar classes in [78],[79]. Often, in forensics, from a small sample signature set (e.g. a few latent fingerprints), one has to determine the identity of the person with whom the signatures are associated. To a large extent, such problems leverage upon a much larger set of composite sketches [80] and mugshot gallery images [82] to perform recognition. However, none of these methods address style-content modeling, which we demonstrate to be very useful for recognition from limited trainind data.

4.3 Learning Style Features

From very limited training data, we learn features that are typical of the style and use that information for face recognition. In this direction, we first learn characteristic image descriptors of a style from training data.

4.3.1 Image Descriptors

We learn two types of descriptors. The first type is computed across the entire image and the second is computed around landmark points of the image. While the former provides a holistic description of the image and is used to estimate the probability of a test image to belong to a style (weak classifier), the latter provides a detailed description of

discriminative regions characteristic of a style (strong classifier). We describe these steps in detail in the following paragraph.

We employ region covariance matrices (RCM) [89] represented in the form of sigma sets [75] that essentially capture the local 2^{nd} order statistic with respect to a set of low level characteristics as our image descriptors. The low level characteristics could be filter responses, color, etc. RCMs have much lower dimensionality when compared to the exact distributions of low level image characteristics, can be computed efficiently using integral images and are robust to small pose variations. These descriptors have shown best discriminative power for human detection tasks, texture recognition, etc. [86].

We divide the image I into 8×8 blocks. For a given region $R \in I$ with N pixels, let $\vec{f_i}$ be a y dimensional vector (consisting of low level characteristics) extracted from the i^{th} pixel in R , and $\vec{u}$ be their mean vector. The $y \times y$ covariance matrix $C(R)$ of R can be calculated as

$$C(R) = F_R F_R^T \tag{4.1}$$

where $F_R = [\hat{f}_1, ...\hat{f_N}]$ denotes the matrix of centered vectors $\hat{f}_i = \frac{1}{\sqrt{(N)}}(\vec{f_i} - \vec{u})$.

Since RCMs do not lie on a Euclidean space, distance between RCMs is computed using Riemmannian geometry that is usually time consuming. In order to overome this, a small set of vectors that can be uniquely determined through Choloesky decomposition on the RCM and that possess the effectiveness of RCMs called the sigma sets (SS) is computed [75]. Thus,

$$SS = [L_1, ..., L_y] \tag{4.2}$$

where L_i is the i^{th} column of L in the Cholesky decomposition of $C(R) = LL^T$. Distance

between two sigma sets can be efficiently computed using an approximation of Hausdorff distance metric [75]. The SS descriptors averaged across all images belonging to a style class j represents the characteristic descriptor of that style SS_j. We will use these descriptors in estimating the probability of test image to belong to a style.

We also compute the low level descriptors (filter responses, color, etc.) around salient points in the image. Depending upon the image, salient points can vary. For instance, it could be corner of eyes, nose, etc. for a face image. In this paper, the term "feature" is used to refer to these extracted low level image descriptors around a salient point. Thus, we have D features characterizing each image if there are D salient points considered for each image in the style class. These features are the inputs to the random subspace ensemble classifier described next.

4.3.2 Random Subspace Ensemble Method for Learning Style Features

From very limited training data, we learn a subset of the D features characterizing the style of the category under consideration. The random subspace method randomly samples a subset of these features and performs training in this reduced feature space. Multiple sets (or bags) of randomly sampled features are generated, and for each bag the parameters are learned. An aggregation of the output of individual classifiers leads to reduction of variance in the error; further diversity among the weak classifiers contributes to the robustness [56].

More specifically, we are given Z training image pairs and D features. Let L be the number of individual classifiers in the ensemble. We choose $d_i \leq D$ (without replacement)

to be the number of features to be used in i^{th} classifier. For each classifier, we determine the match and non-match scores [similarity scores between image pairs of a style class depicting the same person (match score) and different people (non-match score)] using the d_i features as follows . We compute

$$s(I, I') = \frac{1}{d_i} \sum_{n=1}^{d_i} S_n(I, I^{'}), \tag{4.3}$$

where $s(I, I^{'})$ is an average similarity measure between n corresponding salient points in image pair (I, I'). We can use any normalized similarity measure (such as that mentioned in [57] for face verification). In order to identify features that give the highest separation between match and non-match scores, we then compute the Fisher Linear Discriminant function (ratio of variance between match/non-match score set to that within each of these sets) for each classifier. We choose the union of features from those classifiers that give the top k Fisher Linear Discriminant values as our style features; k chosen experimentally. The Fisher measure does not require the data to be normally distributed and thus applicable for sparse data.

4.3.3 Importance of the Chosen Style Features

Not all features identified by the above method are equally important in representing a style. In order to understand the importance of the chosen features, we consider the non-parametric permutation statistical test [58]. Permutation tests helps in assessing what features are same (in other words invariant) across all the instances belonging to a class. Thus, features which are more invariant across the instances of the class can be perceived to be more characteristic of the class and thus be assigned greater importance.

Permutation test: The null hypothesis (H_0) is chosen to indicate that two image groups $G1, G2$ have the same average value in a particular feature; the alternate hypothesis (H_1) indicating that the average value of that feature is different in the two groups. Thus,

$$H_0 : \mu_{G1} = \mu_{G2}; H_1 : \mu_{G1} \neq \mu_{G2}, \tag{4.4}$$

where μ is the average value of a particular feature v under consideration in the two groups.

If the null hypothesis is true, then it should not matter when this feature v is randomly assigned among images in the group. For instance, there is a certain way that the mouth corner looks when a person smiles. On an average, if this appearance is same across all images and groups, then the principle behind the test is that there will not be a significant difference if the mouth tips are randomly assigned across images in the group (i.e. assigning the feature of one person to the corresponding feature of another person).

Specifically, if there are N_s images of a style class S, then we can divide these N_s images into 2 subgroups consisting of N_{s_1} and N_{s_2} images. Let the feature values for the first group be $[v_1, v_2, ..., v_{N_{s_1}}]$ and in second group be $[v_{N_{s_1+1}}, v_{N_{s_1+2}}, ..., v_{N_{s_1+s_2}}]$. The two sided permutation test is done by randomly shuffling $[v_1, ..., v_{N_s}]$ and assigning the first N_{s_1} values, say, $[v_{(1)}, v_{(2)}, ..., v_{(N_{s_1})}]$ to the first group and the remaining N_{s_2} values $[v_{N_{(s_1+1)}}, ..., v_{(N_{s_1+s_2})}]$ to the second group.

For the original two groups we compute,

$$\delta_0 = \left| \frac{1}{N_{s_1}} \sum_{i=1}^{N_{s_1}} v_i - \frac{1}{N_{s_2}} \sum_{i=1}^{N_{s_2}} v_{N_{s_1}+i} \right| \tag{4.5}$$

δ_0 denotes the variation in the feature v of style class S as exhibited by various instances $I_1, ...I_N$ in the two groups $G1$ and $G2$. Thus, $\delta_0 = |\mu_{G1} - \mu_{G2}|$. For any two permuted

groups we compute,

$$\delta_s = \left| \frac{1}{N_{s_1}} \sum_{i=1}^{N_{s_1}} v_{(i)} - \frac{1}{N_{s_2}} \sum_{i=1}^{N_{s_2}} v_{(N_{s_1}+i)} \right| \tag{4.6}$$

δ_s denotes the variation in the feature v of style class S after assigning the feature as depicted by $I_i, i = 1, 2, ...l$ to an image not necessarily of I_i.

We repeat this random shuffling of features among the images under consideration multiple times and count the number of times $\delta_s > \delta_o$. The proportion of times $\delta_s > \delta_o$ is the p value which we denote as p. This value reflects the variation of the feature in the two groups. Smaller p denotes stronger evidence against the null hypothesis, meaning that the feature differed considerably in the two groups. If a certain feature showed no difference in the 2 groups, then it does not matter to which image this feature is associated since the average value does not change; thus it can be considered as a random assignment into any image in the pool. We compute p values for each feature as described above.

The resulting p values indicate the importance of the chosen features. Greater the p values, the smaller is the variance of that feature in the class (hence it is more characteristic of the class) and vice versa. We use these p values as weights in computing the similarity scores between image pairs. Doing so ensures greater robustness in separating match and non-match scores by assigning a higher weight to a more important feature. Thus, (4.3) now becomes

$$s_p(I, I') = \frac{1}{M} \sum_{v=1}^{M} p_v S_v(I, I') \tag{4.7}$$

where M is the number of features as determined by the random subspace method, p_v is the p value for feature v, $S_v(I, I')$ is the similarity between the images I and I' with respect to the feature v and s_p is the p-normalized similarity score.

4.3.4 Validation of the Style Features

Our goal here is to show that there is a higher confidence associated with style-specific scores than with the scores computed using all features. Towards this, we employ a robust non-parametric statistical test called the Siegel-Tukey test that basically checks the null hypothesis (H_0) that two independent score sets come from the same population (style) against the alternative hypothesis (H_1) that they come from populations differing in variability or spread. If the style features are indeed good representators of the class, then there should be a higher level of confidence associated with the null hypothesis when compared with style independent features. The hypotheses can be formally written as follows,

$$H_0 : \sigma_A^2 = \sigma_B^2, Me_A = Me_B; H_1 : \sigma_A^2 \geq \sigma_B^2 \tag{4.8}$$

where σ^2 and Me are the variance and medians for the groups A and B, which we define below. The test is entirely distribution-free. The absence of any normality assumption is an important feature of the test, because its parametric alternative, the F test for variance differences, is quite sensitive to departures from normality [60].

Application to our Problem: Let us say we have T match scores (can be non-match scores also) of a style class. We divide these scores into two groups A and B with r style scores for the first group and m style scores for the second. As per the procedure of this test, we first arrange all these T style scores in ascending order. Next, we assign ranks 1 to T in an alternate extreme manner i.e., assign rank 1 to the lowest number of the sequence, ranks 2 and 3 to the two highest members in the sequence, ranks 4 and 5 to the next two lowest, etc. (If the total number of observations is odd, the middlemost score is dropped.) Assigning

the ranks in this way puts the lower ranks at the extremes of the ordered sequence and the higher ranks in the middle of the sequence. If the null hypothesis were true, the scores from the two groups would tend to be well mixed, so that the mean rank assigned to one of the groups would tend to equal the mean rank assigned to the other group and if the alternate hypothesis is true, the mean rank assigned to the scores from the more variable group will be smaller than the mean rank assigned to the scores from the less variable group. An illustration of this ranking procedure is provided in supplementary material. Once we have obtained the ranks, the next step in this test is to sum the ranks within each group. Let these be denoted as W_A, W_B. If the null hypothesis is true, it is expected that the sum of the ranks (taking into account the size of the two groups) will be roughly the same. If one of the two groups is more dispersed (thus indicating that the obtained scores of the two of groups are less characteristic of a common style), its sum will be lower; while the other group will receive more of the high scores assigned to the center.

From the rank sums, the U statistics associated with Wilcoxon rank sum test [59] are calculated by subtracting off the minimum possible score from the rank sums of each group. Thus,

$$U_A = W_A - \frac{r(r+1)}{2}; U_B = W_B - \frac{m(m+1)}{2}; \tag{4.9}$$

where $U = \min(U_A, U_B)$. In general, the minimum possible rank sum for a group with r elements occurs when its elements get the ranks $1, 2, ...r$. Thus $\frac{r(r+1)}{2}$ is the minimum possible rank sum score for that group.

The p_{st} value given by

$$p_{st} = \Pr\left[X \leq U\right], \tag{4.10}$$

where U_A, U_B are the U statistics for groups A, B and $X \sim Wilcoxon(r, m)$, is a measure of the confidence associated with the scores. Thus, if the style features are good representators of the class, they should be associated with a higher p_{st} value than the p_{st} value associated with scores obtained using all features.

4.4 Determining Style-Content of Test Images

After learning style information from very limited training data, we would like to verify if the two test images represent the same person (match) or not by inferring their individual styles.

4.4.1 Estimating Style Probabilities

We first estimate the probability of test images to belong to a style. Towards this, we use average values of SS descriptors for different styles (Sec 4.3.1) in a mixture model framework. Let SS_T denote the sigma set descriptor of a test image and SS_j denote the average value of SS descriptor for style j. Let b denote the $w \times x$ dimensional matrix whose rows are the SS descriptors for a style (where there are w style classes and x is the dimension of SS descriptor) and let $\vec{P} = [p_1, p_2, \cdots, p_w]$ where p_i is the probability with which a test image can belong to style c_i. We then seek to express $\vec{SS}_T$ as

$$\vec{SS}_T = p_1[\vec{SS}_1] + p_2[\vec{SS}_2] + \cdots p_w[\vec{SS}_w] \tag{4.11}$$

The solutions of the above equation is equivalent to computing the following.

$$\min_{\vec{P}} \left|\left|\vec{SS}_T - \vec{P}\vec{b}\right|\right|_2^2 s.t \left|\left|\vec{P}\right|\right|_1 = 1, p_i \geq 0 \tag{4.12}$$

The above is a constrained optimization problem which can be solved by applying Karush-Kuhn-Tucker (KKT) conditions i.e.

$$D\left|\left|\vec{S}\vec{S}_T - \vec{P}\vec{b}\right|\right|_2^2 + \lambda D(\vec{P}.\vec{1} - 1) - \vec{\mu}.\vec{P}^T = 0 \tag{4.13}$$

where D stands for derivative, $\vec{1}$ is a unit column vector with k rows, λ is the Lagrange multiplier and $\vec{\mu} = [\mu_1, \mu_2 \cdots \mu_w]$ is the KKT multiplier such that $\mu_i \geq 0$ for all i and $\vec{\mu}^T\vec{P} = 0$. Thus we get the probabilities with which each image in the test pair can belong to different style classes. It is to be noted that these are merely weak estimates and are used as a scaling factor in determining the final category, the details of which is described in the following sections.

4.4.2 Computing Similarity Scores for Test Pairs

We estimate similarity scores for all $w \times w$ possible combinations of style classes between the test image pair, wherein only the corresponding learned weighted style features are used in computing the score. This score is further scaled by the prior probabilities of the image pair to belong to the style combination to yield the similarity score for each combination.

Specifically, let (T_1, T_2) denote the test image pair. For each of the $w \times w$ possible style combinations which (T_1, T_2) can take, we compute the similarity score for the particular style combination (i, j) as

$$s_{ij}(T_1, T_2) = \exp^{-\beta[p_i(T_1)p_j(T_2)]} * s_{p_{ij}}(T_1, T_2) \tag{4.14}$$

where β is a scaling factor chosen experimentally, $p_i(T_1), p_j(T_2)$ are the probabilities with

which image (T_1, T_2) can belong to style classes (i, j) respectively, and $s_{p_{ij}}(T_1, T_2)$ is computed using (7) between style classes i and j. It is to be noted that, in general, for 2 style classes (i, j) where $i \neq j$, the style features and their importance are different. Therefore for computing $s_{p_{ij}}(T_1, T_2)$ in (4.14) using (4.7), we use the union of the style features in (i, j) with their corresponding importance determined by the average importance across i and j. Doing this ensures all characteristic features from both styles i and j (appropriately weighted by their importance) is considered in determining the similarity.

4.4.3 Style-Content Classification

The style classes (SC_1, SC_2) of the image pair (T_1, T_2) is determined to be the combination that yields the maximum value for (14). Note that images T_1, T_2 could belong to the same style or different styles. Thus,

$$SC_1, SC_2 = arg \max_{i,j} s_{ij}(T_1, T_2) \tag{4.15}$$

In order to determine the content class (match/non-match) between the pair, we leverage upon the match and non-match similarity scores computed across all image pairs in the training data. We then fit Gaussian distributions to each of these scores to get a characteristic match distribution and a non-match distribution. Please note that since these scores are pooled across all styles in the data, their number is sufficiently large to be modeled by a Gaussian unlike the case of style-specific scores which are sparse and hence require non-parametric methods for modeling and validation.

Given the similarity score as determined by (4.15), we compute the test statistic with respect to the match and non-match distribution using the non-directional Z hypothesis

test [64]. We set the null hypothesis as that the match distribution accounts for the test pair similarity score better than non-match distribution and the alternate hypothesis as that the non-match distribution models the score better. We set level of significance α (probability of incorrectly rejecting the null hypothesis) as 0.05, as per behavioral research standard. We then compute the test statistic p_Z using one independent non- directional z test which determines the number of standard deviations the similarity score deviates from the mean similarity score of the learned distributions.

If $p_Z > \alpha$ with respect to match distribution and $p_Z < \alpha$ with respect to non-match distribution, we conclude the similarity score corresponds to a match score i.e., images T_1, T_2 represent the same person. If $p_Z < \alpha$ with respect to match distribution and $p_Z > \alpha$ with respect to non-match distribution, we conclude the similarity score corresponds to a non-match score i.e., images T_1, T_2 represent different people. However, in cases where $p_Z < \alpha$ (or $p_Z > \alpha$) with respect to both match/non-match distribution, no conclusion can be made regarding the content class as both match/non-match is equally likely.

4.5 Experiments

4.5.1 Datasets

We evaluated our approach on two datasets. The first dataset we considered is a subset of the PubFig database [73] consisting of instances pertaining to style attributes (a) smiling, (b) frowning, (c) female, (d) male, (e) youth. Most of these classes have many images, but our goal is to show that we can infer the style and content from a *small number*

of images. We chose 20 image pairs/class for training and tested on 600 images across the entire dataset. The second dataset is the Yale dataset [72] consisting of 165 images of 11 subjects across different 4 configuration such as (a) happy, (b) sleepy, (c) surprised (d) wink and (e) glasses wherein we used images from 5 subjects for training and rest for testing.

4.5.2 Image Descriptors

To obtain the sigma set descriptors, we considered the covariance of Gabor wavelet responses across 8 orientations and 5 scales. We also considered these filter around 22 salient fiducial points on the face to obtain the "features" as described earlier. An illustration of these locations is provided in Fig. 4.2. These points are often used for face verification [57]. These image descriptors formed the inputs to the random subspace ensemble classifier. We considered a total of $L = 6$ classifiers varying the number of features (d_i) in steps of 2 from 10 to 20 in each of the classifiers with these features being randomly chosen as per the random subspace algorithm.

4.5.3 Experimental Scenarios

We want to demonstrate the advantages of modeling style for undersampled face recognition. Hence we consider two settings—(1) when style is unknown for test images, and (2) when style is known.

Baseline 1 : We compared our approach with the standard bilinear model [66] as the first baseline. From the training images with N_b style classes and N_c content classes, we learn the style matrix and content vector using the procedure described in [66]. For test images

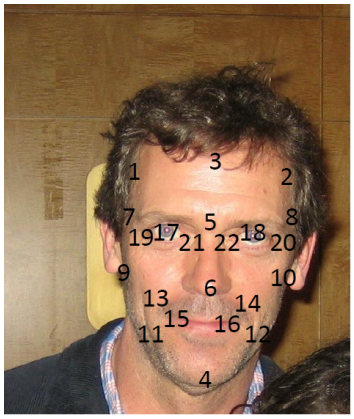

Figure 4.2: Location of salient points on an example image.

with unknown style and content, we use the EM algorithm to classify. When style is known apriori, we simply compute the similarity between the image pairs using the learned style parameters to classify. The match/non-match scores obtained from these experiments is used to plot the ROC.

Baseline 2: We compared our approach against an attribute based identification method namely relative attributes [74] on the PubFig data. Since the method requires the user to provide orderings regarding the relative strengths of attributes in the training data, the method is not applicable when style information is not known aprioir. We provided the attribute relationships (in terms of ordering and similarity as in [74]) for the sparse training images. We used the code provided by the authors in order to obtain the rank scores. From these rank scores, we computed the similarity between image pairs to get match and non-match scores. We used these scores to generate the ROC curves.

4.5.4 Style Modeling Results

Using the procedure described earlier, we computed match scores and non-match scores for each style class under consideration. We considered the union of features from the top 2 Fisher discriminant scores as style features. Table 4.1 and 4.2 list the features characteristic of each style class for the PubFig and Yale respectively; with most important features (as determined from the permutation test) listed first.

Table 4.1: List of style features written in the decreasing order of their importance for each style class for PubFig dataset. Numbers denote features secribed in text and styles (a-e) for PubFig correspond to the classes described in text.

Style	**Style features**
(a)	16,15,20,22,14,21,11, 19
(b)	2,3,5,22,20,14,6,16, 10,8,4
(c)	7,1,12,10,20,13, 6,8
(d)	5,20,2,13,4,1,9,17, 10,16
(e)	14,12,4,6,7,10,9,19

Table 4.2: List of style features written in the decreasing order of their importance for each style class for Yale dataset. Numbers denote features secribed in text and styles (a-e) for Yale correspond to the classes described in text.

Style	**Style features**
(a)	16,15,21,19,11
(b)	22,18,17,13
(c)	5,22,3,19,20,13, 6,8
(d)	19,20,17, 22
(e)	20,19,5,18,17

As can be noted from Table 4.1, for the smiling class, the most important features are the mouth tips, followed by eye corners and cheekbones. Similarly, for frowning people, the most important features are forehead tip, followed by center point on the forehead and nose top. These results are reflective of the patterns generated by these expressions on the face. Similar explanation applies to the Yale dataset (Table 4.2). While mouth tips, eye

corners are indicative of "happy" style, predominantly the features around the eyes (corners and iris) and also the cheek bones are represeantative of "sleepy" style. For the "surprise" class, nose top, eye corners, mid point on the forehead are some important features.

4.5.5 Validation of the Chosen Style Features

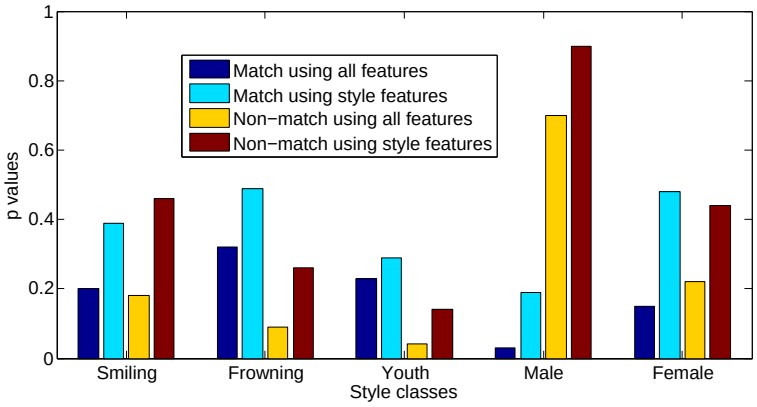

Figure 4.3: p values computed using Siegel-Tukey test for validating the match/non-match scores of each style on PubFigdataset

We validate the chosen style features by showing advantage of style features by demonstrating improvement in the confidence of the style scores. Using the procedure described earlier, we calculated the p values associated with the style-specific scores and style independent scores. Fig. 4.3 and 4.4 shows the p values (computed from Siegel-Tukey test) across different style classes for the PubFig and Yale datasets respectively (for match and non-match scores separately) using all features and using only the style features. As can be noted from Fig.4.3 and Fig 4.4, the p values are higher for style-specific scores for most cases, thus demonstrating the higher confidence associated in using style features.

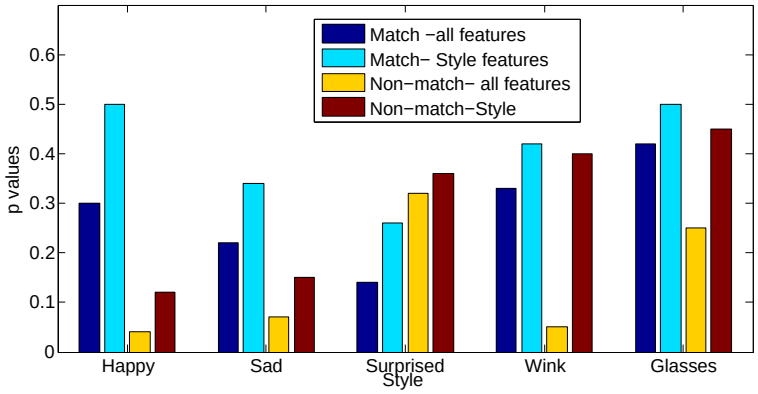

Figure 4.4: p values computed using Siegel-Tukey test for validating the match/non-match scores of each style on Yale dataset

An illustration of correct and incorrect style identification is provided in Fig. 4.5. We also wanted to quantitatively analyze the performance of bilinear models in recognizing the style as the number of training instances were reduced. Using the procedure outlined earlier, we computed the style-specific similarity scores (Sec II B-C) for the proposed method and also computed the corresponding scores using the BLM method for different values of training instances/style (5-9). Fig. 4.6 provides the p values obtained from the Siegel-Tukey test for various number of training instances for the "smiling" style class. From Fig. 4.6, it can be noted that the p value associated with the proposed method is greater than BLM across all values of sample sizes ≤ 8. Similar results were observed for other styles as well. This demonstrates the advantages of the proposed method for style estimation with less data.

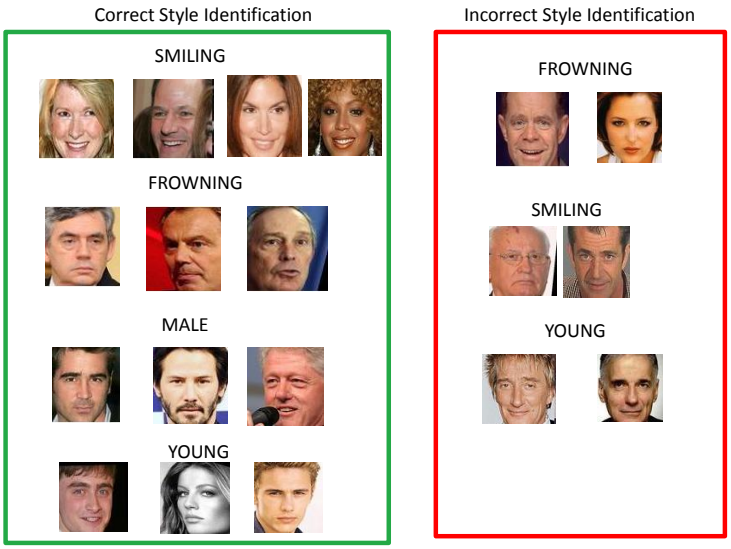

Figure 4.5: Illustration of style identification performance–Example image pairs that were correctly/incorrectly classified by each of the three methods mentioned.

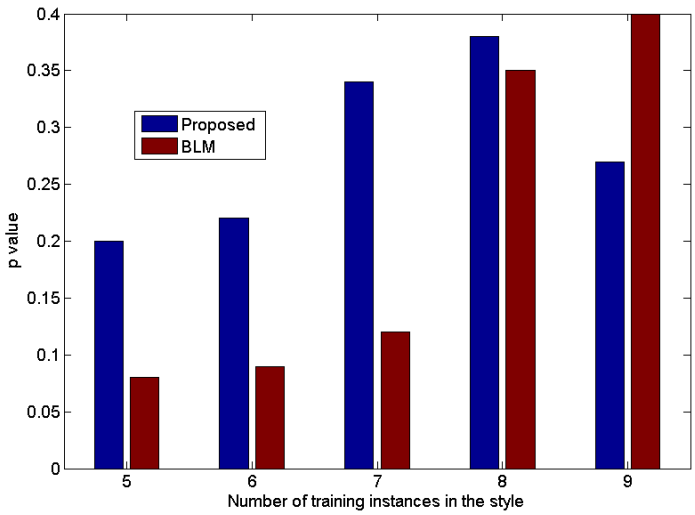

Figure 4.6: Comparison of style estimation performance between the proposed method and BLM. Result shown with respect to style "smiling" .

4.5.6 Recognition Performance:

We analyzed the performance of the proposed model in terms of pairwise recognition accuracy a) by using only the selected style features and b) using all features (i.e. without style modeling). We also compared the performance with bilinear models and relative attributes methods (when style is known apriori). Fig. 4.7 and 4.8 show the resulting ROC curves obtained for the PubFig database and Yale dataset respectively. We roughly used 8 image pairs/style for training. As can be noted, for all datasets, the performance of the selected style features is superior compared to the other methods, thus validating the chosen style features. It is to be noted that when style information is known, we also compare with the relative attributes method. An illustration of the correctly classified images is provided in Fig. 4.9.

4.5.7 Performance with Variation in Sample Size

In order to understand the benefits of the proposed method over bilinear models in terms of sample size, we varied the sample size in match/non-match scores of each style class from 4 to 10 in steps of 1 (for PubFig and Yale database). The average recognition

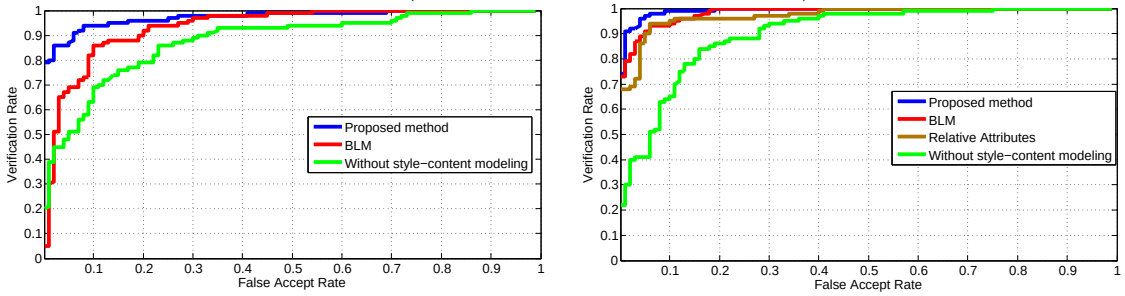

Figure 4.7: Improvement in pairwise content recognition performance using the proposed method on PubFig database when style is unknown (left) and known apriori (right)

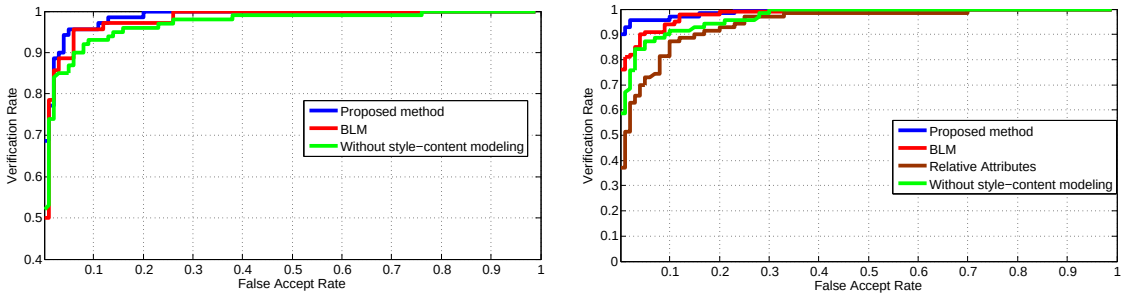

Figure 4.8: Improvement in pairwise content recognition performance using the proposed method on Yale database when style is unknown (left) and known apriori (right)

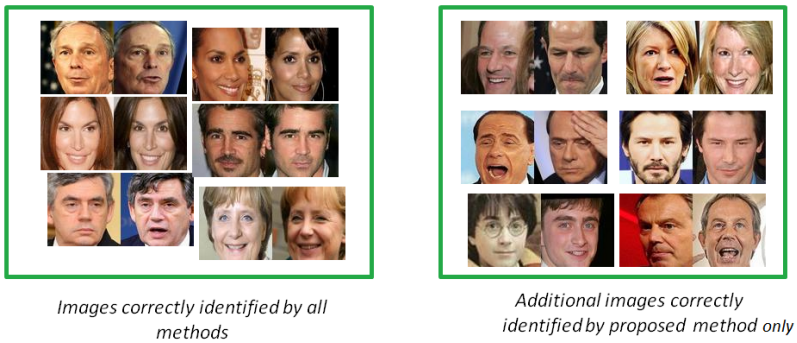

Figure 4.9: Illustration of recognition performance–Greater number of image pairs were correctly classified by proposed method.

rates across all styles (for PubFig dataset) at a sample size of 4/style was approximately 72% (proposed method), 67% (bilinear model), 65% (relative attributes) and 71% (proposed method), 74% (bilinear model), 67% (relative attributes) at a sample size of 10/style. Also, at each sample size, we calculated the corresponding p value from Siegel-Tukey test to understand the confidence of the similarity scores obtained. Fig. 4.10-4.11 show the variation in average p values computed across all style classes under the scenarios of style being known beforehand and not known initially for both match and non-match scores. Initially, with small sample sizes, the confidence of the proposed model is better than bilinear models. As the number of instances per style class increases, the bilinear model starts to performs better than the proposed model. It can also be inferred from Fig. 4.10-4.11, that the p values are slightly higher when the style information is known beforehand indicating a greater confidence than for the case when style information is not known beforehand. However, it is to be noted that even when style information is not known apriori, the confidence of the similarity scores is not significantly lesser than the case when style was known beforehand.

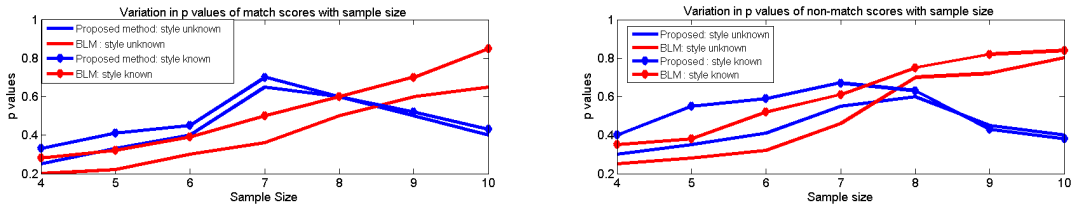

Figure 4.10: Variation in average p value (computed across all styles in PubFig database both when style information is known and unknown apriori) with sample size for match scores (left) and non-match scores (right)

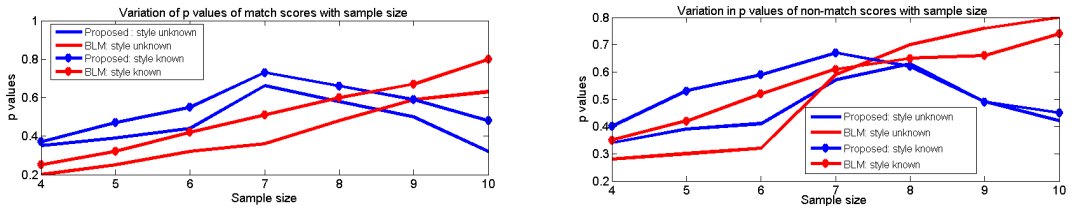

Figure 4.11: Variation in average p value (computed across all styles in Yale database both when style information is known and unknown apriori) with sample size for match scores (left) and non-match scores (right)

4.6 Conclusion

We proposed a novel technique to automatically recognize style and content from limited training samples and in turn demonstrated its usefulness in the context of undersampled face recognition. Through random subspace ensemble learning and statistical permutation test, we arrived at a set of style-specific similarity scores. We validated these scores by showing improvement in pairwise content recognition over standard bilinear models and relative attributes. We also showed improvement in the confidence of the style-specific similarity scores over style independent similarity scores and analyzed the performance of the proposed model with variation in training data sample sizes to show that for sample sizes in the range 4-8 images/style, the proposed method outperforms bilinear models. We also compared the performance of the method when style information is known apriori against when style is not known apriori. This method can potentially benefit other vision problems that involves analysis of style and content.

Chapter 5

Conclusions and Future Work

5.1 Thesis Summary

In this thesis, we proposed novel face recognition algorithms centered around the core idea of saliency. In particular, we investigated the role of saliency for different face recognition applications such as recognition in portraits, photo-sketch recognition, recognition in presence of distortions, and recognition from very limited training data. The contributions can be broken down in the following manner.

In Chapter 2, we proposed a novel framework for understanding salient characteristics of artists' renditions, which in turn is essential for robust face recognition in portraits. Using the ensemble learning and statistical hypothesis tests, we carried out experiments on over 270 protraits belonging to the Renaissance era in an attempt to provide an independent and quantitative source of evidence to art historians in answering questions related to the identity of some ambiguous portraits.

In Chapter 3, we proposed an unsupervised face recognition framework leveraging a novel feature descriptor based on saliency maps of region covariance matrices. The efficiency of the method was demonstrated for two applications, namely, photo-sketch recognition and for face recognition in unconstrained environments in presence of distortions.

Finally, in Chapter 4, we demonstrated the use of knowing information regarding certain salient attributes of an image (referred to as the style) in order to perform face recognition from very limited training data. We showed that the proposed method performs better than state-of-the-art methods as the number of training instances belonging to a style reduces.

5.2 Future Work

There are many ways in which topics dealt herewith can be pursued further. One possible direction is to explore if the proposed style-content modeling framework can be leveraged to address face recognition across ages given very limited training data. In this case, age could be the style factor and content could be the identity of subjects. Another possible avenue for future work could be to investigate the problem of unconstrained face recognition given very limited training data. In particular, one could explore if learning saliency models can be beneficial for the aforementioned applications, i.e., to check if saliency can aid unconstrained face recognition given very limited training data, and for face recognition across ages from limited training data.

Bibliography

[1] B. Raytchev, and H. Murase. Unsupervised Face Recognition by Associative Chaining. *Pattern Recognition*, 36(1):245–257, 2003.

[2] Z. Cui, W. Li, D. Zhu, S Shan, and X. Chen. Fusing Robust Face Region Descriptors via Multiple Metric Learning for Face Recognition in the Wild. *Computer Vision and Pattern Recognition*, 2013.

[3] D. Rim, K. Hassan, and C. Pal. Semi Supervised Learning in Wild Faces and Videos. *British Machine Vision Conference*, 14(1):3.1–3.12, 2011.

[4] A. K. Jain, A. Ross, and K. Nandakumar. Introduction to Biometrics: A textbook. *Springer Publishers ISBN 978-0-387-77325-4*, 2011.

[5] W. Zhao, and R. Chellappa. Face Processing: Advanced Modeling and Methods. *Academic Press*,2006

[6] M. Turk, and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–86, 1991.

[7] Wiskott et al. Face Recognition by Elastic Graph Bunch Matching. *IEEE. Trans. Patt. Anal. Mach. Intell.*,7, 775–779, 1997.

[8] A. Pentland, B. Moghaddam, and T. Starner. View Based and Modular Eigenspaces for Face Recognition. *IEEE. Conf. Comp. Vision and Patt. Reco.*,1994.

[9] C. Liu, and H. Wechsler. Gabor Feature Based Classification Using the Enhanced Fischer Linear Discriminant Model for Face Recognition. *IEEE. Trans. on Image Proc.*, 11, 467–476, 2002.

[10] Z. Cao, Q. Yin , X. Tang, and J. Sun. Face Recognition with Learning Based Descriptors. = *IEEE. Conf. Comp. Vision and Patt. Reco.*,2707–2714, 2010.

[11] J. Li and L. Yao and J. Wang, *Rhythmic Brushstrokes Distinguish Van Gough from his contemporaries: Findings via Automated Brushstrokes Extraction*, IEEE Trans. Patt. Anal. Mach. Intell., 2012.

[12] A. Sharma and D. W. Jacobs, *Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch*, Proc of Comp. Vision and Patt. Reco., 2011.

[13] N. D. Kalka and T. Bourla and B. Cukic and L. Hornak, *Cross-spectral face recognition in heterogeneous environments: A case study on matching visible to short-wave infrared imagery*, Proc. International Joint Conference on Biometrics, pp 1-8, 2011.

[14] B. Klare and A. K. Jain,*Heterogeneous face recognition using kernel prototype similarities*, IEEE Trans. Patt. Anal. Mach. Intell., vol.2, pp. 1-6, 2011.

[15] W. Deng and J. Hu and J. Guo,*Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary*, IEEE. Trans. on Patt. Analysis and Machine Intelligence, 2012.

[16] L. Peng, et al. *Probabilistic models for inference about identity*, IEEE Trans. on Patt. Anal. and Mach. Intelli. ,vol 34, no.1, pp 144-157, 2012.

[17] B. Klare and Z. Li and A. K. Jain, *Matching Forensic Sketches to Mug Shot Photos,* IEEE Trans. Patt. Anal. Mach. Intell., vol.33, no.3, pp. 639-646, 2011.

[18] M. Vasilesc and O. Alex and D. Terzopoulos, *Multilinear analysis of image ensembles: Tensorfaces*, European Conf. on Comp. Vision, pp 447–460, 2002.

[19] N. Kumar and A. Berg and P. Belhumeur and S. Nayar,*Attribute and simile classifiers for face verification*, IEEE. Intl. Conf. on Comp. Vision, 2009.

[20] J. B. Tenenbaum and W. T. Freeman,*Separating style and content with bilinear models*, Neural Computation,vol 12,no 6, pp 1247–1283, 2000.

[21] L. Itti, C. Koch, and E. Niebur, *A Model of Saliency-Based Visual Attention for Rapid Scene Analysis*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.

[22] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, *Modeling Visual Attention via Selective Tuning*, Artificial Intelligence, vol. 78, nos. 1-2, pp. 507-545, 1995.

[23] R. Milanese, *Detecting Salient Regions in an Image: From Biological Evidence to Computer Implementation*, PhD thesis, Univ. Geneva, 1993.

[24] S. Baluja and D. Pomerleau, *Using a Saliency Map for Active Spatial Selective Attention: Implementation and Initial Results*, Proc. Advances in Neural Information Processing Systems, pp. 451- 458, 1994.

[25] C. Koch and S. Ullman, *Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry*, Human Neurobiology, vol. 4, no. 4, pp. 219-227, 1985.

[26] A. M. Treisman and G. Gelade, *A Feature Integration Theory of Attention*, Cognitive Psychology, vol. 12, pp. 97-136, 1980.

[27] J. Han, K. Ngan, M. Li, H. Zhang, *Unsupervised extraction of visual attention objects in color images*, Trans. Circuits and Systems for Video Technology, 16(1), 2006.

[28] U. Rutishauser, D. Walther, C. Koch, P. Perona, *Is bottom-up attention useful for object recognition?*, CVPR, 2004.

[29] V. Mahadevan, N. Vasconcelos,*Saliency-based discriminant tracking*, CVPR, 2009

[30] A. Borji, L. Itti, *State-of-the-art in visual attention modeling*, IEEE. Trans. on Patt. and Mach. Intelli., 2012.

[31] H. C. Nothdurft, *Salience of Feature Contrast*, Neurobiology of Attention, Academic Press, 2005.

[32] L. Itti, C. Koch, *Computational Modeling of Visual Attention*, Natural Rev. Neuroscience, vol. 2, no. 3, pp. 194-203, 2001.

[33] T. Judd and K. Ehinger, F. Durand, A. Torralba, *Learning to predict where humans look*, IEEE International Conference on Computer Vision,pp. 21062113, 2009.

[34] T. Liu, X. Sun, T. Nan-Ning Zheng, H. Shum, *Learning to detect a salient object*, IEEE Conference on Computer Vision and Pattern Recognition, pp1-8, 2007.

[35] Q. Zhao, C. Koch , *Learning visual saliency by combining feature maps in a nonlinear manner using AdaBoost*, Journal of Vision, 12 (6): 22, 115, 2012.

[36] M. Cerf, J. Hareland, W. Einhaeuser, C. Koch, *Predicting human gaze using low-level saliency combined with face detection*, Advances in Neural Information Processing Systems, pp. 241248. Cambridge, MIT Press, 2007.

[37] R. Zhao,W. Ouyang, X. Wang, *Unsupervised Salience Learning for Person Re-identification*, IEEE Conference on Computer Vision and Pattern Recognition, 2013.

[38] J. Zhu, J. Wu, Y. Wei, E. Chang, Z. Tu, *Unsupervised Object Class Discovery via Saliency-Guided Multiple Class Learning*, IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[39] R. Cappelli, A. Franco, D. Maio*Gabor Saliency Map for Face Recognition*, International Conference on Image Analysis and Processing, 2007.

[40] A. Salah, B. Bicego, L. Akarun, E. Grosso, M. Tistarelli, *Hidden Markov Model-based face recognition using selective attention*, Human Vision and Electronic Imaging, 2007.

[41] X. Wang, X. Tang, *Face Photo-Sketch Synthesis and Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 31, 2009.

[42] L. J .Karam, T. Zhu, *Quality Labeled Faces in the Wiild (QLFW): A Database for Studying Face Recognition in Real-World Enviroments*, Arizona State University, IVU Lab Technical Report 03-2014-1, March 2014.

[43] A. Perrig, *Drawing and the Artist's Basic Training from the 13th to the 16th Century*, The Art of the Italian Renaissance. Architecture, Sculpture, Painting, Drawing,Cologne/Germany, pp 416-441, 2007.

[44] D. Stork, *Computer vision and computer graphic analysis of paintings and drawings: An introduction to the literature,* Proc. Comp Anal. Images. Patt., 2009.

[45] C. Tyler and W. Smith and D. Stork, *In search of Leonardo : Computer based facial image analysis of Renaissance artworks for identifying Leonardo as subject,* Human vision and electronic imaging, 2012.

[46] C. Johnson, et. al, *Image Processing for Artist Identification - Computerized Analysis of Vincent van Gogh's Painting Brushstrokes*,' IEEE Signal Processing Magazine, Special Issue on Visual Cultural Heritage, vol. 25, no. 4, pp. 37-48, 2008.

[47] D. Stork and M. Johnson, *Recent developments in computer-aided analysis of lighting in realist fine art*, Image processing for art investigations, Museum of Modern Art, 2010.

[48] D. Parikh andK. Graumman, *Relative Attributes*, International Conference on Computer Vision, 2011.

[49] R. Srinivasan and A. Roy-Chowdhury and C. Rudolph and J. Kohl, *Recognizing the Royals-Leveraging Computerized Face Recognition for Identifying Subjects in Ancient Artworks*, 21st ACM Intl. Conf. on Multimedia, 2013.

[50] R. Srinivasan and A. Roy-Chowdhury and C. Rudolph, *Computerized Face Recognition in Renaissance Portrait Art*, Signal Processing Magazine, Special Issue on Signal Processing for Art Investigations, 2015.

[51] R. Srinivasan and A. Nagar and A. Tewari and D. Mitrani and A. Roy-Chowdhury, *Face Recognition based on Sigma Sets of Image Features*, IEEE International Conference on Acoustics, Speech and Signal Processing, 2014.

[52] R. Srinivasan and A. Roy-Chowdhury, *Modeling Style and Content from Very Limited Training Data—Application in Face Recognition*, forthcoming.

[53] S. West, *Portraiture*, Oxford University Press, 2004.

[54] F. Vegter and J. Hage, *Clinical Anthropometry and Canons of the Face in Historical Perspective*, History of Facial Anthropometry, vol.106, no.5, 2005.

[55] Farkas et.al., *Inclinations of Facial Profile: Art versus Reality*, Plastic and Reconstructive Surgery, vol. 75, no 4, pp 509-518, 1984.

[56] T. Ho, *The random subspace method for constructing decision forests*, IEEE. Trans. on Patt. Anal. and Mach. Intell., vol 20, n0 8, pp 832-844,1998.

[57] Wiskott et. al., *Face Recognition by elastic graph bunch graph matching,* IEEE Trans. Patt. Anal. Mach. Intell., vol.7, pp. 775-779,1997.

[58] P. I. Good, *Permutation, Parametric, and Bootstrap Tests of Hypothesis*, 3rd Ed., Springer, 2009.

[59] H. Mann and D. Whitney, *On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other*, Annals of Mathematical Statistics, vol 18, no 1, pp 50-60,1947.

[60] A. Gayen, *The distribution of the variance ratio in random samples of any size drawn from non-normal universe*, Biometrika,vol 37, pp 236-255, 1950.

[61] S. Siegel and J. Tukey,*A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples*, Journal of the American Statistical Association, 1960.

[62] M. Fischler and R. Bolles, *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*, Proc of Comm. of the ACM, vol. 24, no.6, pp. 381-395, 1981.

[63] R. Duda and P. Hart and D. Stork, *Pattern Classification*, Wiley-Interscience, 2001.

[64] M. Asadoorian, *Essentials of Inferential Statistics*, University Press of America, 2005.

[65] C. Rudolph, A. Roy-Chowdhury, R. Srinivasan, and J. Kohl, *FACES: Faces, Art, and Computerized Evaluation Systems–A Feasibility Study of the Application of Face Recognition Technology to Works of Portrait Art*, (forthcoming).

[66] J. B. Tenenbaum and W. T. Freeman, emph Separating style and content with bilinear models, Neural Computation, vol 12, no 6, pp1247–1283, 2000.

[67] M. Vasilescu and D. Terzopoulos, *Multilinear independent component analysis*, IEEE. Intl. Conf. on Comp. Vision and Patt. Reco., 2005.

[68] A. Elgammal and C. Lee, *Separating style and content on an nonlinear manifold*, IEEE. Intl. Conf. on Comp. Vision and Patt. Reco., pp 478–485, 2004.

[69] D. Vlasic and M. Brand and H. Pfister and J. Popovic, *Face transfer with multilinear models*, ACM Trans. on Graphics, pp 426–433, 2005.

[70] R. Srinivasan and A. Roy-Chowdhury and C. Rudolph and J. Kohl, *Quantitative Modeling of Artists Styles in Renaissance face portraiture*, 2nd Intl. Workshop on Historical Document Imaging and Processing, pp 94–101, 2013.

[71] W. Deng and J. Hu and J. Guo, *Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary*, IEEE Trans. on Patt. Anal. and Mach. Intell., vol 34, no 9, 2012.

[72] P. Belhumeur and J. Hespanha and D. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.

[73] N. Kumar and A. Berg and P. Belhumeur and S. Nayar, *Attribute and simile classifiers for face verification*, IEEE. Intl. Conf. on Comp. Vision, 2009.

[74] D. Parikh and K. Grauman, *Relative Attributes*, IEEE. Intl. Conf. on Comp. Vision, 2011.

[75] X. Hong and T. Yuan and H. Chang and S. Shan and X. Chen and W. Gao, *Sigma Set: A Small Second Order Statistical Region Descriptor*, IEEE.Conf. on Computer Vision and Pattern Recognition, 2009.

[76] E. Altendorf, *Learning from sparse data by exploiting monotonicity constraints*, Conf. on Uncertainty in Artificial Intelligence, 2005.

[77] J. Choi, et.al, *Multimodal Location Estimation of Consumer Media: Dealing with Sparse Training Data*, International Conference on Multimedia and Expo, 2012.

[78] R. Fergus and H. Bernal and Y. Weiss and A. Torralba, *Semantic label sharing for learning with many categories*, IEEE. European. Conf. on Comp. Vision, 2010.

[79] G. Wang and D. Forsyth and D. Hoiem, *Comparitive object similarity for improved recognition with few or no examples*, IEEE. Intl. Conf. on Comp. Vision and Patt. Reco., 2010.

[80] S. Klum and H. Han and A. K. Jain and B. Klare, *Sketch Based Face Recognition: Forensic vs. Composite Sketches*, International Conference on Biometrics, 2013

[81] Y. Xu and A. Roy-Chowdhury, *Learning a geometry integrated image appearance manifold from a small training set*, IEEE. Conf. on Comp. Vision and Patt. Reco., 2008

[82] B. Klare and Z. Li and A. K. Jain, *Matching forensic sketches to mug shot photos*, IEEE. Trans. on Patt. Analysis and Machine Intelligence, vol 33, no 3, pp 636–646, 2011.

[83] W. Forstner, B Moonen *A metric for covariance matrices*, Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University ,1999.

[84] A. Rosenfeld, and G. Vanderburg. Coarse-fine Template Matching. *IEEE. Trans. Syst. Man. Cyb.*, 7:104–107, 1977.

[85] R. Brunelli, and T. Poggio. Face Recognition: Features versus Templates. *IEEE. Trans. Patt. Anal. Machine. Intell.*, 15:1042–1052, 1993.

[86] B. Wu and R. Nevatia, *Optimizing Discrimination Efficiency Tradeoff in Integrating Heterogenous Local Features for Object Detection*, IEEE. Conf. on Comp. Vision and Patt. Reco., 2008.

[87] E. Erdem, A. Erdem, *Visual saliency estimation by nonlinearly integrating features using region covariances*, Journal of Vision, 2013.

[88] O. Tuzel, F. Porikli, P. Meer, *Region covariance: A fast descriptor for detection and classification*, European Conference of Computer Vision, 2006.

[89] O. Tuzel, F. Porikli, P. Meer, *Covariance Tracking Using Model Update Based on Lie Algebra*, IEEE. Conf. on Computer Vision and Pattern Recognition, 2006.

[90] O. Tuzel, F. Porikli, P. Meer, *Pedestrian Detection via Classification on Riemmanian Manifolds*, IEEE. Trans. Patt. Anal. Machine. Intell., 30:1713–1727, 2008.

[91] X. Hong, T. Yuan, H. Chang, S. Shan, X. Chen, W. Gao, *Sigma Set: A Small Second Order Statistical Region Descriptor*, IEEE.Conf. on Computer Vision and Pattern Recognition, 2009.

[92] R. Srinivasan, A. Nagar, A. Tewari, D. Mitrani, A. Roy-Chowdhury *Face Recognition based on Sigma Sets of Image Features*, IEEE.Intl. Conf. on Acoustics, Speech and Signal Processing, 2014.

[93] D. Yi, Z. Lei, and S . Li. *Towards Pose Robust Face Recognition*, IEEE Conf. on Computer Vision and Pattern Recognition, 2013.

[94] S. Arashloo and J. Kittler, *Efficient processing of MRFs for unconstrained-pose face recognition*, Biometrics: Theory, Applications and Systems, 2013 .

[95] F. Porikli and T. Kocak. *Robust License Plate Detection Using Covariance Matrices in a Neural Network Framework*, IEEE.Adv. Video Signal Based Surveillance Conf., 2006.

[96] M. Dubuisson and A. Jain. *A Modified Hausdroff Distance for Object Matching*, International Conference on Pattern Recognition, 1994.

[97] F. Porikli and T. Kocak. *Fast Construction of Covariance Matrices for Arbitrary Size Image Windows*, IEEE.Intl. Image Proc. Conf., 1581–1584, 2006.

[98] B. Wu, and R. Nevatia. *Optimizing Discrimination Efficiency Tradeoff in Integrating Heterogenous Local Features for Object Detection*, IEEE. Conf. on Computer Vision and Pattern Recognition, 2008.

[99] Y. Pang, T. Yuan and X. Li. *Gabor-Based Region Covariance Matrices for Face Recognition*, IEEE. Transactions on Circuits and Systems for Video Technology, 2008.

Appendix A

Source Description for the Portraits Illustrated in Chapter 1

Fig. 2.2 : S. West, *Portraiture*, Oxford University Press, 2004.

Fig. 2.3 : Algardi Images (from left to right): Montagu, Alessandro Algardi (New Haven 1985) v.2, fig. 150, Jennifer Montagu, Alessandro Algardi (New Haven 1985) fig. 176, Jennifer Montagu, Alessandro Algardi (New Haven 1985) fig. 160.

Bernini Images (from left to right): ed Bacchi, et al, Bernini and the Birth of Baroque Portrait Sculpture (Los Angeles 2008) p.77, 289, Grazia Bernardini, Gian Lorenzo Bernini (1999) fig. 6, ed Bacchi, et al, Bernini and the Birth of Baroque Portrait Sculpture (Los Angeles 2008) p. 38, 256, 292.

Clouet Images (from left to right) : `http://en.wikipedia.org/wiki/File:Elizabeth_d_Autriche_by_Francois_Clouet_1510_1572.jpg`, `http://en.wikipedia.org/wiki/File:Catherine_de_Medicis.jpg`, `http://www.royalcollection.org.uk/collection/401229/`

mary-queen-of-scots-1542-87.

Holbein Images (from left to right): Holbein-Cheseman-Hague, Holbein-Erasmus-Met, Holbein-Southwest II, Uffizi.

Kneller Images (from left to right): James Scott, Duke of Monmouth and Buccleuch; 1678 (British 17th-18th centuries)- by studio of Godfrey Kneller (National Portrait Gallery, London; NPG 5225); Duke of Buccleuch (priv. col.) , http://en.wikipedia.org/wiki/File:Catherine_de_Medicis.jpg , (British 17th-18th centuries); (London, National Portrait Gallery; NPG 3794).

Mierevelt Images (L-R): portrait of Ambrogio Spinola; 1609 (Northern Baroque); Amsterdam, Rijksmuseum (SK-A-554) portrait of Dudley Carleton, Viscount Dorchester; 1620 (Northern Baroque); London, National Portrait Gallery, NPG 3684, portrait of Jacob Van Dalen; 1640 (Northern Baroque); New York, Metropolitan Museum of Art, 25.110.13.

Musscher Images (L: R): Amsterdam, Rijksmusseum, SK-A-4233, Anna Verbie (unk-unk), wife of Sybrandt Oosterling; 1694 (Northern Baroque); (private collection, Netherlands), wife of Adriaen van Loon (1631-1722); 1681 (Northern Baroque); (Museum Van Loon, Amsterdam; Inv. 95).

Various Artists(L-R): Museo nazionale del Bargello, Florence, Lady with the Primroses; c. 1475 (Italian Early Renaissance); attributed to Verrocchio (Bargello), painting; 1533-1534 (Italian Late Renaissance); by Vasari (Uffizi), death mask casting; 1446; by Buggiano (Andrea Cavalcanti) (Duomo, Florence) late 16th century (Italian Late Renaissance); by Giovanni Bandini (Duomo, Florence), painting; c. 1474 (Italian Early Renaissance); by Leonardo da Vinci (National Gallery, Washington).

Fig. 2.4 (top from L–R) Holbein-Henry VIII, Madrid, portrait of a young woman; 1630 (Northern Baroque); Vienna, Kunsthistorisches Museum, Kneller self-portrait; 1685 (British 17th-18th centuries); (London, National Portrait Gallery; NPG 3794), by Franois Clouet or Jacques Decourt (BnF, Estampes, Paris; Rs. Na 22), ed Bacchi, et al, Bernini and the Birth of Baroque Portrait Sculpture (Los Angeles 2008) p.86,Montagu, Alessandro Algardi (New Haven 1985) v.2, fig. 150.

Fig. 7 From left to right- Top row: Museo nazionale del Bargello, (London, National Portrait Gallery; NPG 3794), by Franois Clouet or Jacques Decourt (BnF, Estampes, Paris; Rs. Na 22),Lady with the Primroses; c. 1475 (Italian Early Renaissance); attributed to Verrocchio (Bargello), painting, `http://expositions.bnf.fr/renais/grand/071.html`, Marcus Gheeraerts the Younger, c. 1605-1610 (Norton Simon Museum, Pasadena; F.1965.1.027.P), published 1793 after a lost painting of 1601 by Santi di Tito (from Clemente de Nelli, Vita e commercio letterario di Galileo Galilei matematico e filosofo); c. 1616; attributed to William Larkin (National Portrait Gallery; NPG 3840, attributed to Abraham van Blijenberch (Scottish National Portrait Gallery; PG 1096), engraving by George Vertue (1723) after a painting by Michiel Van Mierevelt (1613) (National Portrait Gallery, London; NPG D17977) , 1st Earl of Portland; after Anthony Van Dyck; late 1620s (Government Art Collection; GAC 1507) ,by Robert Peake the Elder; 1605 (National Galleries of Scotland; PG 9)

Fig. 7 bottom row from left to right: death mask casting; c. 1472; by Francesco Laurana (Louvre; RF 1171), National Portrait Gallery, London; NPG 96, c. 1475 (Italian Early Renaissance); attributed to Botticelli (Victoria and Albert), National Portrait Gallery, Lon-

don; NPG 96, c. 1475 (Italian Early Renaissance), c. 1595-1600 (National Portrait Gallery, London; NPG 1723), c. 1590; artist unknown (private collection), painting 1627; attributed to Cornelius de Neve (National Portrait Gallery, London; NPG 1346), artist unknown; c. 1610 (National Portrait Gallery, London; NPG 1195), c. 1617-1620; artist unknown (National Portrait Gallery, London; NPG 40) , 1627 (National Portrait Gallery, London; NPG 1344), c. 1595-1600 (National Portrait Gallery, London; NPG 1723).

Appendix B

Discussion of Identification Paradigms

Note: Test results are indicated as match/non-match/no decision as per the analysis procedure described in this paper. The images in each test are marked alphabetically and the result between possible image pairs is given. For example, for paradigm 1, the test result "match" indicates that images a and b gave a match score. "(?)" indicates that the identity of the sitter is hypothesized but uncertain.

-1: Battista Sforza paradigm

-a: Battista Sforza; bust; c. 1474; by Francesco Laurana (Museo nazionale del Bargello, Florence)

-b: Battista Sforza (?); death mask casting; c. 1472; by Francesco Laurana (Louvre; RF 1171)

Table B.1: Test result for the image pair under consideration

Image pair	Result
1a, 1b	Match

This paradigm tested an analogue (an unmediated image of the subject, not a work of art) against a three-dimensional work of art that, in this case, physically approaches the subject in form and size but that nevertheless partakes of the subjectivity of artistic interpretation. The match score indicates the probability of a match, despite the obvious challenges in testing an image rendering the death throes of an individual against a work of portrait art.

-2: Eva Visscher paradigm

-a: Eva Visscher; c. 1685; by Michiel Van Musscher (Amsterdam, Rijksmusseum, SK-A-4233)

-b: Family of the Artist; 1694-1701; by Michiel Van Musscher; the figure of the adult female is unknown, with some scholars believing that it represents the artist's first wife, Eva Visscher, and others that it portrays his second, Elsje Klanes (Antwerp, Royal Museum of Fine Arts; Inv. 739)

-c: Interior with Eva Visscher (1651-1684), the artists first wife, and their two children, and with his self-portrait on the wall; c. 1683; by Michiel Van Musscher (Private collection, The Netherlands).

-d: Eva Visscher (1651-1684), first wife of Michiel van Musscher; c. 1690; by Michiel Van Musscher (Collection of the Wawel Royal Castle, Cracow; Inv. 1126).

Table B.2: Test result for the image pair under consideration

Image pair	Result
2a, 2b	Match
2a, 2c	Non-match
2a, 2d	Match

This paradigm tested four portraits of the artist's wife, three of which are known to be of his first wife (Eva Visscher) and the fourth is thought by some to represent his second wife. Some of the figures involved made around ten to fifteen years apart. The test resulted in two matches and a non-match, the non-match possibly being accounted for by strong profile of the portrait.

-3: Mary Queen of Scots paradigm

-a: Unknown: Mary Queen of Scots (?); formerly identified as Mary of Lorain, Queen of James V of Scotland; painting; c. 1570; (National Portrait Gallery, London; NPG 96)

-b: Mary Queen of Scots; drawing; c. 1558; by Franois Clouet or Jacques Decourt (BnF, Estampes, Paris; Rs. Na 22)

-c: Mary Queen of Scots; c. 1560; Franois Clouet (BnF, Estampes, Paris; Rs. Na 22)

-d: Mary Queen of Scots; painting; c. 1560-1592; artist unknown (National Portrait Gallery, London; NPG 1766)

-e: Mary Queen of Scots; cast of head from tomb in Westminster; 1606-1616; Cornelius and William Cure (National Portrait Gallery, London; NPG 307a and 307B)

-f: Mary Queen of Scots; c. 1558; by Franois Clouet; miniature (Royal Collection; RCIN 401229)

-g: Mary Queen of Scots; c. 1560-1561; by Franois Clouet (Royal Collection; RCIN 403429)

-h: Mary Queen of Scots; c. 1555; by Franois Clouet and workshop (Zakad Narodowy im. Ossoliskich, Wrocaw)

-i: Mary Stuart (and Francois II of France); c.1558; workshop of Franois Clouet (from Catherine de' Medici's Book of Hours, Bibliotheque nationale de France MS NAL 82)

-j: Mary Queen of Scots; miniature; 1579; by Nicholas Hilliard (Royal Collection; RCIN 420641)

This paradigm tested a portrait thought at one time to be of Mary Queen of Scots against eight other portraits known to be of Mary returning four scores in the match range and four in the "no decision" range, some of the latter being quite close to the match range, despite the differences in style in the representation of the face. The score in the non-match range was with her tomb sculpture in Westminster Abbey, made 18-28 years after her death. The results indicate the strong probability that the test portrait is of Mary.

-4: James Scott paradigm

-a: Unknown: James Scott, Duke of Monmouth and Buccleuch (?); Edward Sackville, son

Table B.3: Test result for the image pair under consideration

Image pair	Result
3a, 3b	No decision
3a, 3c	No decision
3a, 3d	No decision
3a, 3e	Non-match
3a, 3f	Match
3a, 3g	Match
3a, 3h	Match
3a, 3i	No decision
3a, 3j	Match

of the 4th Earl of Dorset (?); Launcelot Northbrook (?); 1640s; artist unknown (National Portrait Gallery, London; NPG 1566)

-b: James Scott, Duke of Monmouth and Buccleuch; 1678; by studio of Godfrey Kneller (National Portrait Gallery, London; NPG 5225)

-c: James Scott, Duke of Monmouth and Buccleuch; c. 1683; possibly after William Wissing (National Portrait Gallery, London; NPG 151)

-d: James Scott, Duke of Monmouth and Buccleuch; c. 1660-1690; by Abraham Blooteling, after Sir Peter Lely (National Portrait Gallery, London; NPG D19810)

Some scholars believe that the test image (a) represents James Scott, nephew of King James II, after having been beheaded for treason. Match scores with two portraits made during his life lend support to this view. The "no decision" comes from a poor quality engraving of an earlier, lost portrait.

-5: Anne Boleyn paradigm

-a: The "Moost Happi" medal; 1534; lead (London, British Museum)

-d: Anne Boleyn; late 16th-century copy of a portrait whose source ultimately goes back to around 1533-1536; painting; this is considered the oldest and best iteration of this group (London, National Portrait Gallery; NPG 668)

Table B.4: Test result for the image pair under consideration

Image pair	Result
4a, 4b	Match
4a, 4c	Match
4a, 4d	No decision

-e: Anne Boleyn, Hever Castle portrait; late 16th-century copy of a portrait whose source ultimately goes back to around 1533-1536; painting (Hever Castle, Broadland Properties Limited)

-f: Anne Boleyn; 1590-1610 copy of a portrait whose source ultimately goes back to around 1533-1536; painting (London, National Portrait Gallery; NPG 4980)

-g: Anne Boleyn (?), Nidd Hall portrait; painting; late 16th century; although the figure wears jewelry believed to have belonged to Anne Boleyn, many think that the image is of her successor, Jane Seymour, Henry VIII's third wife (private collection/Bradford Art Galleries and Museums)

-31h: Jane Seymour; chalk; 1536-1537; chalk, ink; by Hans Holbein the Younger (The Royal Collection)

-31i: Jane Seymour; 1536-1537; by Hans Holbein the Younger (Kunsthistorisches Museum, Gemldegalerie, Vienna)

While it seems certain that a number of portraits of Anne were made during her life, only one has survived that can said to be of her with a historical basis, a small, worn lead medal. When tested against four portraits probably most commonly said to be of

Table B.5: Test result for the image pair under consideration

Image pair	Result
5a, 5d	No decision
5a, 5e	Non-match
5a, 5f	Non-match
5a, 5g	Match
5a, 31h	Match
5a, 31i	Match

Anne, the results were a "no decision" (a copy of an older portrait), two non-matches, and match. However, we also tested the medal against two portraits of Jane Seymour, third wife of Henry VIII (a painting and preparatory drawing for the painting) and received match scores. This suggested to us that the test image, the medal, did not provide sufficient portrait data for a reliable test, and that match returns must be understood in the larger context.

-6: Shakespeare paradigm

-b: William Shakespeare; sculpted bust; believed to be before c. 1620; probably by Gerard Johnson (Gheerart Janssen) the Younger (Holy Trinity Church, Stratford; cast NPG 1735)

-c: William Shakespeare, Droeshout engraving, second state; from the First Folio, first published 1623, but probably copied from an existing portrait; by Martin Droeshout; implied to be an accurate rendering by Ben Jonson in the prefatory matter to the First Folio (Folger Shakespeare Library)

-d: William Shakespeare, Chandos portrait; c. 1600-1610; painting and infrared reflectogram; attributed to John Taylor; believed by some to be the only image done from life (London, National Portrait Gallery; NPG 1)

-g: William Shakespeare (?), Janssen portrait; c. 1610; painting, x-ray, and infrared reflectogram; this is said by some to have engendered its own group of copies; others feel it is an early copy of the Cobbe portrait, which they believe acted to engender the group; others still are convinced that it is a portrait of Thomas Overbury, a contemporary courtier (Washington, Folger Shakespeare Library; FPS 17)

-h: William Shakespeare (?), Cobbe portrait; c. 1610; painting; said by some to be the

original source of the Janssen portrait group, by others to be an early copy in that group, and by others to represent Thomas Overbury (Hatchlands Park, Surrey, UK)

-l: William Shakespeare (?), Hampton Court Palace portrait; x-ray of painting; thought by some to represent William Herbert, 3rd Earl of Pembroke; Hampton Court Palace.

Two poor quality posthumous portraits of Shakespeare exist that are thought by many to have been made from a now lost portrait made from life. A relatively large number of other portraits are sometimes said to also be of Shakespeare. This paradigm received a very particular pattern of results suggesting that, as this technology advances future research on this and perhaps a broader body of portraits may very well contribute to untangling this very tangled body of portraits to a useful degree.

-7: Galileo paradigm

-a: Galileo Galilei (?); painting; c. 1590; artist unknown (private collection)

-b: Galileo Galilei; engraving made in 1750 by Giuseppe Calendi and Rafaello Morghen (published 1793) after a lost painting of 1601 by Santi di Tito (from Clemente de Nelli,

Table B.6: Test result for the image pair under consideration

Image pair	Result
6b, 6g	No decision
6b, 6h	Non-match
6b, 6l	Non-match
6c, 6g	No decision
6c, 6h	No decision
6c, 6l	Non-match
6d, 6g	No decision
6d, 6h	No decision
6d, 6l	No decision

Vita e commercio letterario di Galileo Galilei matematico e filosofo)

-c: Galileo Galilei; painting; c. 1604; by Domenico Tintoretto (National Maritime Museum, Greenwich)

-d: Galileo Galilei; painting; c. 1612; attributed to Filippo di Nicola Furini (Vienna Kunsthistorishe; INV GG 7976)

-e: Galileo Galilei; drawing; 1624; by Ottavio Leoni (Biblioteca Marucelliana, Florence)

-f: Galileo Galilei; painting; 1624; by Domenico Passignano (private collection, Helsinki)

-g: Galileo Galilei; painting; 1636; by Justus Sustermans (Suttermans) (Uffizi Florence)

-h: Galileo Galilei; painting; 1639; Mellin (private collection)

-i: Galileo Galilei; painting; 1640; by Justus Sustermans (Suttermans) (Palazzo Pitti, Florence)

A painting believed by some to be the earliest known portrait of Galileo, when tested against a broad spectrums of portraits known to be of Galileo, returned results that neatly broke down into consistently decreasing similarity scores: first, within the match range for the chronologically three closest likenesses (1601-c. 1612); then, within the "no decision" range for the next two (1624); and, finally, within the non-match range for the

Table B.7: Test result for the image pair under consideration

7a, 7b	Match
7a, 7c	Match
7a, 7d	Match
7a, 7e	No decision
7a, 7f	No decision
7a, 7g	Non-match
7a, 7h	Non-match
7a, 7i	Non-match

final three (1635-1640). Not only does this tend to support the identification of the test image, but it provides a convincing spread of test results regarding age as a factor in portrait identification.

-8: George Villiers paradigm

-a: Unknown; painting; 1627; attributed to Cornelius de Neve (National Portrait Gallery, London; NPG 1346)

-b: George Villiers, 1st Duke of Buckingham; painting; c. 1616; attributed to William Larkin (National Portrait Gallery; NPG 3840)

-c: George Villiers, 1st Duke of Duke of Buckingham, and his family; painting; 1628; after Gerrit van Honthorst (National Portrait Gallery, London; NPG 711)

-d: George Villiers, 1st Duke of Buckingham; painting; 1625; by Peter Paul Rubens (Palazzo Pitti, Florence)

-e: George Villiers, 1st Duke of Buckingham; painting; 1625; by Michiel Van Mierevelt (Cambridge University Library; accession no. 30)

-f: George Villiers, 1st Duke of Buckingham; chalk; 1625; by Peter Paul Rubens (Graphische Sammlung Albertina, Vienna)

A portrait thought by some to represent George Villiers, 1st Duke of Buckingham (1592-1628), when tested against a body of other portraits known to be of Villiers, returned

Table B.8: Test result for the image pair under consideration

8a, 8b	Non-match
8a, 8c	Non-match
8a, 8d	No decision
8a, 8e	No decision
8a, 8f	Non-match

a mixed body of non-match and "no decision" results, something that does not lend support to the identification of this painting as Villiers.

-9: Michelangelo paradigm

-a: Michelangelo; chalk; 1548-1553; by Daniele da Volterra (Haarlem, Netherlands, Teylers Museum)

-b: Michelangelo; bronze bust; 1564-1566; by Daniele da Volterra (Florence, Museo Nazionale del Bargello)

-c: Michelangelo; bronze bust; 1564; by Daniele da Volterra (Galleria dell'Accademia, Florence)

-d: Michelangelo (?), the figure of Nicodemus in The Entombment of Christ ; painting; 1602-1604; by Caravaggio (Vatican Museums)

It has recently been proposed that the famous seventeenth-century Italian painter, Caravaggio, gave one of the figures in his painting of The Entombment of Christ the features of Michelangelo. In this particular case, our results may have been affected by angle-views of some of the images. Further testing with better angles might give more dependable results for this possible crypto-portrait.

-10: Arabella Stuart paradigm

-a: Unknown, possibly Lady Arabella Stuart; artist unknown; c. 1595-1600 (National Portrait Gallery, London; NPG 1723)

Table B.9: Test result for the image pair under consideration

9a, 9d	No decision
9b, 9d	Non-match
9c, 9d	No decision

-b: Lady Arabella Stuart; by Robert Peake the Elder; 1605 (National Galleries of Scotland; PG 9)

-c: Lady Arabella Stuart; artist unknown; c. 1605 (Government Art Collection; GAC 399)

-d: Lady Arabella Stuart; Marcus Gheeraerts the Younger, c. 1605-1610 (Norton Simon Museum, Pasadena; F.1965.1.027.P)

The National Portrait Gallery in London possesses a portrait thought by some to represent Lady Arabella Stuart (1575-1615). When tested against three other portraits known to represent Stuart, the results were mixed. We also tested the known portraits against each other, again with mixed results. This reinforced the obvious point that the ability of the individual artist must always be taken into consideration in this type of testing.

-11: William Drummond paradigm

-a: Unknown; painting; artist unknown; c. 1610 (National Portrait Gallery, London; NPG 1195)

-b: William Drummond; painting; 1612; attributed to Abraham van Blijenberch (Scottish National Portrait Gallery; PG 1096)

In some ways, this might be thought of as a "classic" paradigm: the comparison of a single unknown portrait against a single known one. The result, "no decision," illustrates the challenge of such a paradigm. When testing for identification, it is always preferable

Table B.10: Test result for the image pair under consideration

10a, 10b	Non-match
10a, 10c	Match
10a, 10d	Non-match
10b, 10c	No decision
10c, 10d	Non-match

Table B.11: Test result for the image pair under consideration

11a, 11b	No decision

to have a larger number of reference images. Whereas in larger paradigms, a "no decision" can sometimes be interpreted, no matter how partially, here it remains a simple statement of uncertainty.

-12: Ralph Winwood paradigm

-a: Unknown; painting; c. 1617-1620; artist unknown (National Portrait Gallery, London; NPG 40)

-b: Sir Ralph Winwood; engraving by George Vertue (1723) after a painting by Michiel Van Mierevelt (1613) (National Portrait Gallery, London; NPG D17977)

-c: Sir Ralph Winwood; painting; 1613; attributed to Abraham Blyenberch (Collection of the Duke of Buccleuch and Queensberry)

This paradigm illustrates the potential complexities of testing: while there is a match between the test image and one of the reference images, there is a non-match between the test image and another of the reference images, in this case an engraving of an earlier lost portrait. At the same time, there is a "no decision" in the result of testing of the engraving and the painted reference image. This suggests the possibility that the test image represents Winwood, while the engraving carries a different set of facial signifiers (either because the

Table B.12: Test result for the image pair under consideration

12a, 12b	Non-match
12a, 12c	Match
12b, 12c	No decision

lost image it copies did or in its own right).

-13: Richard Weston paradigm

-a: Unknown; painting; by Cornelius Johnson; 1627 (National Portrait Gallery, London; NPG 1344)

-b: Richard Weston, 1st Earl of Portland; after Anthony Van Dyck; late 1620s (Government Art Collection; GAC 1507)

Like the William Drummond paradigm, the Richard Weston paradigm is an example of the desire for a larger body of reference images with which to compare the test image.

-14: Lady at the Window/Lady with the Primroses paradigm

-a: Portrait of a Lady at the Window, Smeralda Brandini (?); c. 1475; attributed to Botticelli (Victoria and Albert)

-b: Lady with the Primroses; c. 1475; attributed to Verrocchio (Bargello)

Having noted the desire for a large body of reference images with which to compare the test image–and insisting that it is a principle of FACES that this technology does not prove the identity of its subjects–the match return from this limited paradigm is high and convincing, despite the two distinctly different personas conveyed in the images.

-15: Ginevra de'Benci paradigm

-a: Bust of a Young Woman; bust; 1465-1466; attributed to Verrocchio (Frick, New York)

Table B.13: Test result for the image pair under consideration

13a, 13b	No decision

Table B.14: Test result for the image pair under consideration

14a, 14b	Match

-b: Ginevra de'Benci; painting; c. 1474; by Leonardo da Vinci (National Gallery, Washington)

While the negative results of this test are no less subject to the "no decision" or match results of other limited tests, the low score tends to refute the position that the two different images represent the same person.

-16: Niccol Strozzi/Robert de Masmines (?) paradigm

-a: Niccol Strozzi; bust; 1454; by Mino da Fiesole (Berlin, Bodemuseum)

-b: Robert de Masmines (?); painting; before 1444; by Robert Campin (Master of Flmalle) (Thyssen-Bornemisza Collection, Lugano-Castagnola, Switzerland) The "no decision" here is complicated by an approximately ten year age difference between the two images (and so presumably between the ages of the sitter at the time of sitting). The effect of age differences varies widely among various individuals. While ten years difference in age usually doesn't play a strong part in test results, the effectiveness of the testing depends upon other factors as well, such as the age of the person, weight gain, and so on. In a paradigm of Isaac Newton, probable age differences of up to twenty-nine years did not strongly affect test results.

-17: Battista Sforza painting paradigm

Table B.15: Test result for the image pair under consideration

15a, 15b	Non-match

Table B.16: Test result for the image pair under consideration

16a, 16b	No decision

-a: Battista Sforza; painting; c. 1472; by Piero della Francesca (Uffizi, Florence)

-b: Battista Sforza (?); death mask casting; c. 1472; by Francesco Laurana (Louvre, Paris)

In an earlier Battista Sforza paradigm, we established a probable match between a known bust of Sforza and a death mask casting long thought to be of her. Here, the "no decision" between the casting and a painting known to be of her is in all likelihood the result of the relatively limited information available from the profile view of the painting, an indictor that the profile view needs further research.

-18: Andrea Mantegna paradigm

-a: Mantegna, self-portrait; bronze bust; c. 1490; (on his tomb, SantAndrea, Mantua)

-b: Warrior, Saint James before Herod, self-portrait of Mantegna (?); 1452; (Eremitani, Padua)

-c: Face in vegetal design, Mantegna (?); c. 1474; by Mantegna (Camera degli Sposi, Palazzo Ducale, Mantua)

Some scholars have believed that images b and c represent crypto-portraits of the artist, Andrea Mantegna. When tested against a self-portrait by Mantegna, the first came up "no decision" while the second was a non-match.

Table B.17: Test result for the image pair under consideration

17a, 17b	No decision

Table B.18: Test result for the image pair under consideration

18a, 18b	No decision
18a, 18c	Non-match

-19: Vincenzo Giustiniani/Antonio Cepparelli paradigm

-a: Unknown: thought by some to be Vincenzo Giustiniani; bust; c. 1670; by Bernini (private collection)

-b: Vincenzo Giustiniani; chalk; 1631; by Claude Mellan (Vienna, Albertina Sammlung)

-c: Antonio Cepparelli; bust; 1622-1623; by Bernini (San Giovanni de Fiorentini, Rome)

Table B.19: Test result for the image pair under consideration

19a, 19b	No decision
19a, 19c	No decision
19b, 19c	No decision

It has been suggested that a bust by Bernini in a private collection represents Vincenzo Giustiniani. When tested against a known image of Giustiniani, however, it received a "no decision," as it did when tested against a portrait of Antonio Cepparelli (both the Giustiniani and Cepparelli portraits also receiving a "no decision" when tested against teach other). This paradigm is a cautionary tale about possible similar physiognomies, the alteration of facial components (the eyes in the Giustiniani portrait), and sometimes possibly even the testing of images in different mediums (e.g., chalk against sculpture), the latter being a subject that requires further research.