Title
Computerized Face Recognition in Renaissance Portrait Art

Authors
Srinivasan, Ramya
Roy-Chowdhury, Amit, Prof
Rudolph, Conrad

Publication Date
2015-02-17

Peer reviewed

Computerized Face Recognition in Renaissance Portrait Art

Ramya Srinivasan, Conrad Rudolph and Amit Roy-Chowdhury

***Abstract*—In this work, 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. Works of portrait art bear the mark of visual interpretation of the artist. Moreover, the number of samples available to model these effects are often limited. 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. 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 uncertain 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.**

***Index Terms*—Face Recognition, Portraits, Style Modeling**

I. 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 [1]. 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" [1], 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 project *FACES* (*F*aces, *A*rt, and *C*omputerized *E*valuation *S*ystems) was conceptualized 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. *FACES* has been funded by the US National Endowment for the Humanities (Phase 1 and 2) under the Digital Humanities program. This paper will describe the challenges inherent in face recognition in art images, and summarize the results obtained in this project over the last two years. Some preliminary results have been presented in [12].

A. Roy-Chowdhury and R. Srinivasan are with the Department of Electrical and Computer Engineering, University of California, Riverside. CA,92521. E-mail: amitrc@ee.ucr.edu

C. Rudolph is with the Department of Art History, University of California, Riverside, CA 92521.

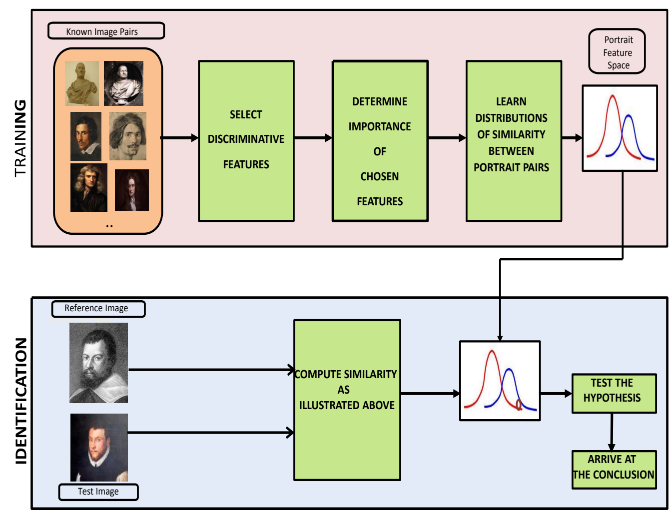

Fig. 1. Illustration of the learning (top) and identification framework (bottom)

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.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.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.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.

We begin by describing the *c*hallenges involved in face recognition of portraits. 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 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. Merely pulling images off the internet would lack scientific integrity.
(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 [2], 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.

II. Related Work

We review some image processing techniques employed for art analysis and also provide a survey of state-of-the-art in computerized face recognition.

Image Analysis in Art Works: Analysis of paintings using sophisticated computer vision tools has gained popularity in recent years [5]. Computer analysis has been used for identifying the artist [24] and for studying the effect of lighting in artworks [25], among others. A recent paper has explored application of computer-based facial image analysis [6] 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. Some preliminary results have been presented in our earlier paper [12] on a small set of data. In this work, extensive results are shown on a much richer dataset, and using a more sophisticated feature extraction algorithm. 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 [13]. This is further substantiated by the use of local features in matching forensic sketches (an art form) to human faces in works such as [7].

Computerized Face Recognition: A survey of still and video based face recognition research is provided in [3]. A vast majority of face recognition applications address surveillance and entertainment. These approaches can be classified into three categories, namely holistic methods, feature based structural matching methods or a combination of both depending on the representation in feature space. Recent research efforts have focused on cross spectral face recognition for comparing images taken in heterogeneous environments [8]. Such methods are not applicable for our study. First, since the images in the present scenario are obtained from museums across the world, we have no control on the kind of sensors used to capture them. Second, the quality of the image is not an issue here; the challenge is choice of appropriate features.

Some works [9], [27] model style factors such as a facial pose, expression, etc. and separate it from content, i.e., the identity of the person, and show promising results for face pose estimation, among others. In [11], the authors use attributes like chubby, attractive, etc. for face verification tasks. While models for separating style (e.g., an artist's rendition) from content (sitter's identity) can be useful for the present study, all of the existing methods require hundreds of images. Some works have looked at face recognition from sparse training data [10]. In [7], the authors leverage upon much larger mug shot gallery images or composite sketches for training. In [26], the authors evaluate the probability that two faces have the same underlying identity cause for recognition. However, these methods do not model style. In this work, we explore artist's style from the available sparse data.

III. 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.

A. Face as Seen by Artists

It is evident from [13] 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 [14].

A historical appraisal of facial anthropometry from antiquity upto Renaissance has been provided in [15] 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.

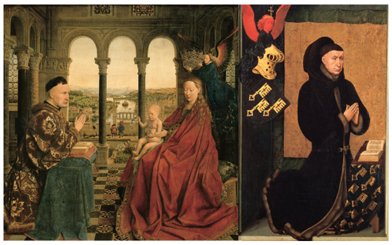

Fig. 2. Prominent facial landmarks such as pointed nose were retained in works of the same sitter Nicolas Rolin by different artists Jan Van Eyck and Rogier van der Weyden.

B. Choice of Features

From the illustrations described above, it is clear that ancient Renaissance artists laid emphasis on two aspects in their renderings, which we use for our analysis.

1. Local features: We use a set of 22 fiducial points to represent each face, these being (1, 2) forehead tips (left, 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) and (21, 22) right eye corners (right and left eye). 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

$$J_{n_j} = a_{n_j} \exp(i\phi_{n_j}), \tag{1}$$

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, namely, (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 corners, (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.

C. 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 [16]. Small sample sizes is very relevant to the present problem as we have very few works by an artist at our disposal (Sec 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. We compute

$$s_{LF}(I, I') = \frac{1}{d_i} \sum_{n=1}^{d_i} s_n(J, J^{'}), \tag{2}$$

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 [4] given by

$$s_n(J, J^{'}) = \frac{\sum_j a_{n_j} a^{'}_{n_j}}{\sqrt{\sum_j a^2_{n_j} \sum_j a'^2_{n_j}}} \tag{3}$$

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

$$s_{AD}(I, I') = e^{-\beta y}, \tag{4}$$

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 dynamic range of values. In our experiments, we set β to be 5.

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.

D. 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 [17]. 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 [2].

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,

$$H_0 : \mu_{G1} = \mu_{G2}; H_1 : \mu_{G1} \neq \mu_{G2}, \tag{5}$$

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,

$$\delta_0 = \left| \frac{1}{N_{s_1}} \sum_{i=1}^{N_{s_1}} v_i - \frac{1}{N_{s_2}} \sum_{i=1}^{N_{s_2}} v_{N_{s_1+i}} \right| \tag{6}$$

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

$$\delta_s = \left| \frac{1}{N_{s_1}} \sum_{i=1}^{N_{s_1}} v_{(i)} - \frac{1}{N_{s_2}} \sum_{i=1}^{N_{s_2}} v_{(N_{s_1+i})} \right| \tag{7}$$

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) and (4). It is to be noted that this method can be employed when we have ≥ 12 images by an artist [21]; 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.

E. 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.,

$$score = \lambda s_{LF} + (1-\lambda) s_{AD} \tag{8}$$

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 [18] 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 [20].

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).

F. Validation of the Learned Features

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.

1) 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 [23].

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,

$$H_0 : \sigma_A^2 = \sigma_B^2, Me_A = Me_B; H_1 : \sigma_A^2 \geq \sigma_B^2 \tag{9}$$

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 [22]. The p value obtained from this test, p_{st}, is given by

$$p_{st} = \Pr\left[X \leq U\right], \tag{10}$$

where U_A, U_B are the U statistics for groups A, B and $X \sim$ Wilcoxon (r, m) [21]. 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.

2) Validation of PFS: In order to validate the PFS computed across various artists/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.

IV. 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.

A. Hypothesis Testing

This is a method for testing a claim or hypothesis about a parameter in a population [19]. 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 [19], 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.

B. 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.

C. Analysis Scenarios

Following the procedure outlined earlier, we compute similarity scores of test cases with corresponding reference image and with distracters. Table I lists various hypothesis test scenarios that can arise [19] and the corresponding conclusions that one can infer. Match and non-match cases are straight forward to infer from Table I. 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 I, 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 I), 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 I
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.

V. Dataset

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. A part of the dataset can be accessed here https://www.dropbox.com/s/ai6vux9m06pwy72/Sample%20Data.zip?dl=0
Description: The dataset consists of about 271 images where the identity of the subject is known beyond doubt. There are 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 II provides a detailed description of the distribution of images in terms of the specific sitter and artist. Fig. 3 provides an illustration of the dataset.

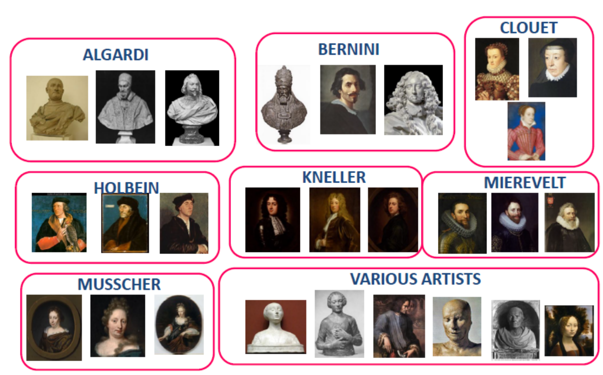

Fig. 3. Illustration of the dataset across individual/multiple artists depicting different sitters.

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 II
ILLUSTRATION OF IMAGE DISTRIBUTION : NUMBER OF IMAGES PER ARTIST.

VI. EXPERIMENTS

A. 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. 4 depicts characteristic LF with dots denoting the relative importance of the feature as per the p value of permutation test. AD features representative

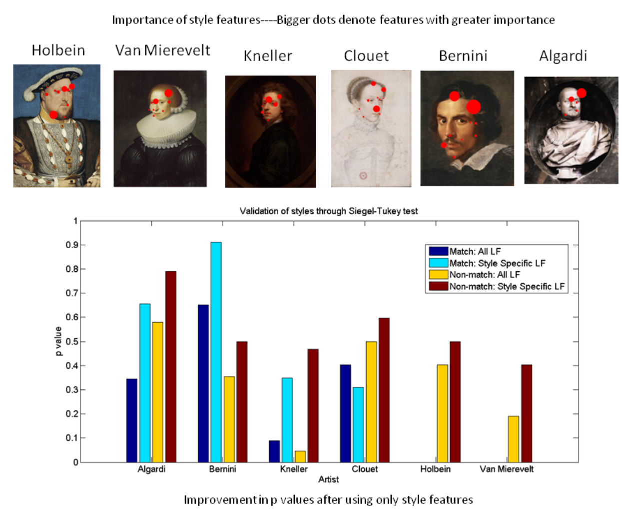

Fig. 4. Top: Importance of chosen features with bigger dots indicating more important features; Bottom: Validation of style through Siegel-Tukey test

of the style was similarly determined for these artists; these being AD features 4,8,3,7,2 for Algardi (Please see Sec III B 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 test. As illustrated in bottom part of Fig. 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. 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.

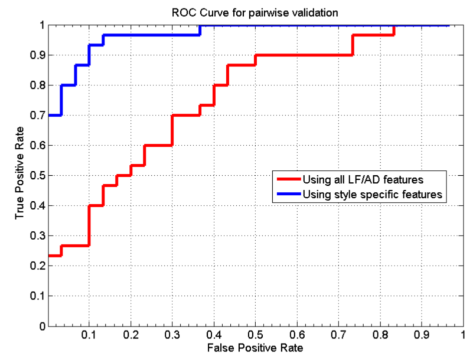

Fig. 5. Pairwise sitter validation upon using style features.

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 also be possible to understand the adherence to artistic canon and individual variations in art practices. Lastly, these results could help in understanding the differences in works from different geographical regions.

B. 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 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. 6). As described earlier, the improvement in using style features as against all LF/AD features is evident from Fig. 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.

C. Identity Verification

We want to provide quantitative measures of similarity to uncertain test paradigms provided to us by art historians. In

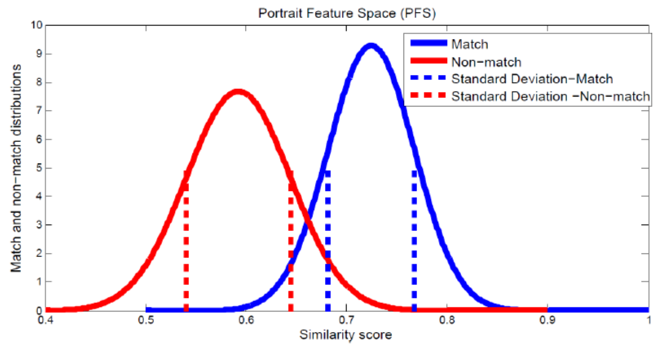

Fig. 6. PFS showing the distribution of match and non-match scores along with their standard deviations.

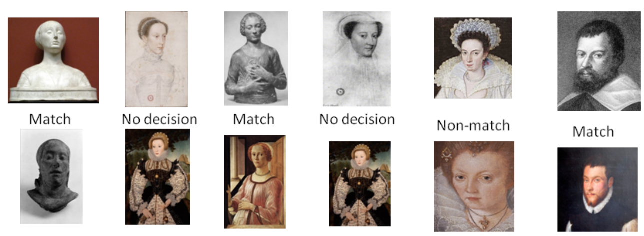

Fig. 7. Illustrations of identification tests with conclusion in center.

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.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 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

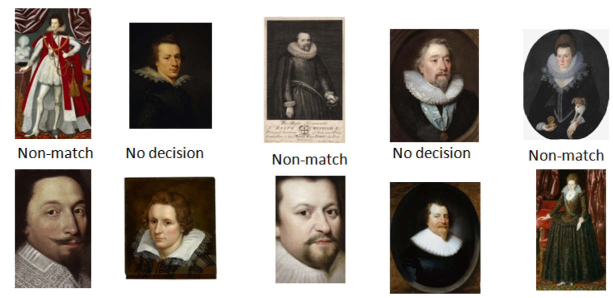

Fig. 8. Illustrations of identification tests with conclusion in center.

lying in bed after his beheading for treason.

The portrait shown in bottom row of Col. 6 in Fig. 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-c. 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. 1, Fig. 8) and a comparison between an unknown portrait against a known portrait of Lady Arabella Stuart (col.5, Fig. 7) gave non-match scores. A complete list of identification tests with results can be found in the link https://www.dropbox.com/s/u2dxu0uecd7q310/identification_faces_list.doc?dl=0.

The results of *FACES* are only as dependable as the images tested. For instance, some radical angle view images including profiles are hard to identify. Also, comparing portraits of sitters separated by an age difference of more than 30 years remains a challenge to *FACES*.

VII. 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.

References

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

[2] 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.

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

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

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

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

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

[16] 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.

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

[18] 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.

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

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

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

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