# White Paper Report

Report ID: 106498

Application Number: HD-51625-12

Project Director: Conrad Rudolph (conrad.rudolph@ucr.edu)

Institution: University of California, Riverside

Reporting Period: 6/1/2012-5/31/2014

Report Due: 8/31/2014

Date Submitted: 8/22/2014

**FACES Phase I: White paper**

## FACES: Faces, Art, and Computerized Evaluation Systems, Phase I

### A feasibility proposal for the use of face recognition systems in the identification of unidentified works of portrait art

Conrad Rudolph, Professor, Department of the History of Art, University of California, Riverside
Amit Roy-Chowdhury, Professor, Department of Electrical Engineering, University of California, Riverside
Jeanette Kohl, Associate Professor, Department of the History of Art, University of California, Riverside
Ramya Srinivasan, Graduate Student Researcher, University of California, Riverside

Before the nineteenth century, most portraits were, almost by definition, depictions of people who were important in their own worlds. But, as a walk through almost any major art museum will show, a large number of these portraits from before the nineteenth century have lost the identities of their subjects through the fortunes of time. Traditionally, identification of many of these portraits has been limited to often quite variable personal opinion. FACES (Faces, Art, and Computerized Evaluation Systems) proposes to establish the initial parameters of the application of face recognition technology to works of portrait art--this highly subjective aspect of art history--while at the same time retaining the human eye as the final arbiter.

During this grant period, FACES began work establishing these parameters, asking such questions as: is face recognition technology, originally designed for actual (that is, photorealistic) human faces, applicable to works of portrait art, which are subject to a process of visual interpretation on the part of the artist? Which of the many different face recognition techniques should be used? Which functions of the many functions of a given technique would apply most effectively to our subjects? What culture and period would work best in this initial stage of testing? What types of portraits would best be used, sculpture (three-dimensional) or painting and drawing (two-dimensional)--or both? How will the identifying characteristics in a portrait of one sitter by an artist with a distinctive artistic style compare to a portrait of the same sitter but by a different artist who also has a distinctive style? If face recognition technology works with sculpture, will the identical process be able to be used for painting and drawing? If face recognition technology operates best with a straight-on view of the subject, how will the rate of successful tests be with three-quarter view portraits, the standard pose for portraits in early modern Western culture? For two-dimensional works, will the medium--oil painting, tempera, pencil, chalk, engraving, and so on--affect the test results differently? What about copies of portraits (for example, of famous sitters, like Isaac Newton) and copies of copies--how closely will they retain the identifying characteristics found in the original portrait? What about extreme or poor lighting in painting and drawing? What about aging as found in multiple portraits of the same sitter made over a long period of time? By the same artist? By different artists? What about the vary artistic ability of the individual artist?

In the course of initial investigation, it gradually became clear that of all the different methods of face recognition technology, two gave the most dependable results: the computation of anthropometric distances and of local features. These two methods were part of a larger, more complex process we call the FACES algorithm (detailed below).

While the FACES algorithm was constantly developed throughout the course of this first year and beyond, we began by testing the death mask of a known individual against an identified sculptural portrait of the same individual. That is, we tested an analogue--an unmediated image of the subject, not a work of art--against the image of 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.

We then left the relative security of the analogue and work-of-art pairing, and tested paradigms of exclusively three-dimensional works of art--that is, we then tested two works both of which were now subject to the subjectivity of artistic interpretation. (We use the term paradigm here to mean a logically chosen body of related images directed toward a particular demonstrative end.) More specifically, we tested a sculptural portrait of a known individual with another sculptural portrait of the same individual, both around the same stage of the individual's life and both depicted by the same artist--in other words, we proceeded with as much control over variables as possible.

Incrementally, we broadened our tests--too involved to fully detail here--introducing a similarly controlled but wide-ranging series of systematically chosen variations extending from more controlled paradigms to less controlled (that is, more challenging) ones. These included the same stage of an individual's life but by different artists, different stages of an individual's life by the same artist, and different stages of an individual's life by different artists--all in three-dimensional imagery.

Then we tested two-dimensional imagery, first simply comparing two two-dimensional images of the same subject by the same artist, and then mixing media by testing a number of sculpture vs painting (that is, three-dimensional vs two-dimensional) paradigms, employing a systematic series of distinctions similar to those already mentioned (different ages, different artists, and so on). Finally, we tested a few identified portraits of individuals against unidentified ones.

Development of the FACES algorithm was painstaking and gradual--and too much to fully accomplish in one year of work. What we did do was establish proof of concept. Practically speaking, this means that we identified the issues, established the basic methodology (even if not fully worked out yet), and applied this methodology to a particular set of paradigms.

At the same time, we became aware of areas that needed more work in the future. First, it became apparent that more work needed to be done on establishing an optimum feature set (the most effective body of identifying facial features, given the unique demands of portrait art). Second, we recognized that we needed to develop a gallery of images somewhere in the low hundreds with which to establish non-match averages (that is, a standard with which to compare a given image under investigation). A distinct research issue, such a gallery would also help identify the elements of our optimum feature set. Third, we realized that we had to model styles and normalize for effects. This means that we needed to objectively characterize the individual style of an artist and the period style, and then find a way to systematically take these into account in relation to the variations in human faces that already occur naturally. Fourth, we knew that we needed to develop an algorithm robust enough to deal with the vexing problem of angle views of individuals. And, fifth, we saw that aging had shown itself to be a consistent challenge. From the beginning, we expected that the amount of work required to establish the parameters of face recognition technology to works of portrait art would require more than one year of work.

**The FACES algorithm**

Put as succinctly as possible, the FACES algorithm works as follows.

Overview of the Algorithm

Figure 1 illustrates the procedure adopted in our work. Artists' renditions are examined to arrive at relevant features for analysis - these being local features (LF) and anthropometric distances (AD). For the pairs of images known to represent the same person, we compute measures of LF and AD similarity to get what we refer to as the "match scores". Similarly, set of non-match scores are obtained from instances that are known to not represent the same person.

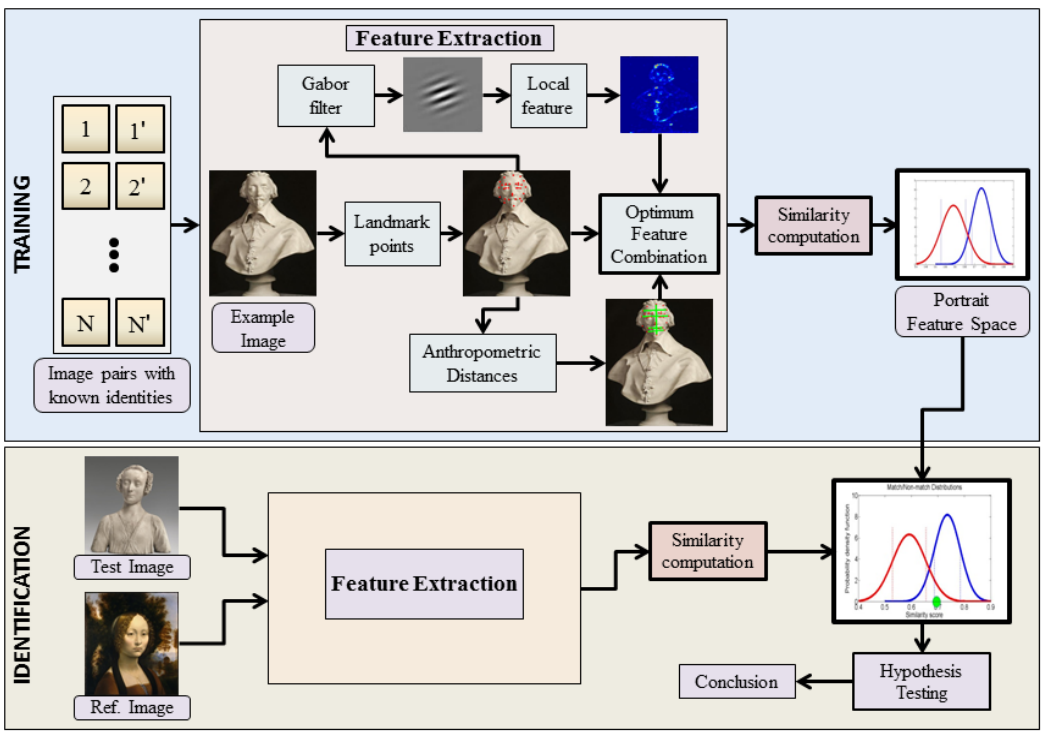



Figure 1: Overview of the Algorithm

Using Fischer linear discriminant analysis (a method used in pattern recognition and statistics to categorize or separate two or more groups/classes of data), similarity scores from LF and AD are fused in a way so as to maximize the variance between classes (match/non-match) while maintaining a small variance within each class. Thereafter, using RANSAC algorithm (an iterative method used to estimate the parameters of a mathematical model that contains outliers, i.e., instances that do not fit to the model), we fit probability density functions (PDF-distributions that describe the probability of similarity scores) to the match and non-match scores and prune outliers to obtain distributions of match and non-match scores. We call these learned distributions as Portrait Feature Space (PFS), which is then validated on known instances. For identification purposes, the position of the similarity score between test and reference image in the PFS is used to arrive at conclusions using statistical hypothesis tests (a statistical method of inference using the observed data).

<u>Details of the Algorithm</u>

1. *Features Extraction*- We first describe the process of extracting the above mentioned features.

*1.1 Local Feature Extraction*

A set of 22 fiducial points is used to represent each face. These include forehead tips (left, right), forehead center, chin bottom, eye corners (right, left of each eye), iris (left, right), cheekbones (left, right), nose top, nose bottom, mouth corners (left, right), chin ear corners (left, right), points on temple (left, right), and points on chin (left, right). The precise location of these points is determined by registering a generic mesh on the face and finding the corresponding points between them. Gabor jets are evaluated at each of these fiducial points. A jet describes a small patch of grey values in an image around the fiducial points described above. It is based on convolution (a mathematical operation that determines the degree of overlap between two functions, i.e., image patch and the Gabor filter in this case to estimate the edges present) the image with Gabor wavelets corresponding to 5 frequencies and 8 orientations. LF similarity score between two portraits is evaluated as the average of jet similarities over all fiducial points considered in the image (i.e. 22).

*1.2 Anthropometric Distances Extraction*

All images are normalized with respect to scale and orientation. A set of 11 salient anthropometric distances, represented as a vector, characterizes each face. These distances include distance between iris, between forehead center and chin bottom, between forehead tips, between nose top and bottom, between chin ear corners, between mouth corners, between cheekbones, between points on chin, between forehead center and nose bottom, between points on temples and width of nose. The similarity between two AD vectors is evaluated by converting the distance into a similarity measure using appropriate conversion schemes.

2. *Portrait Feature Space (PFS) Learning Framework*

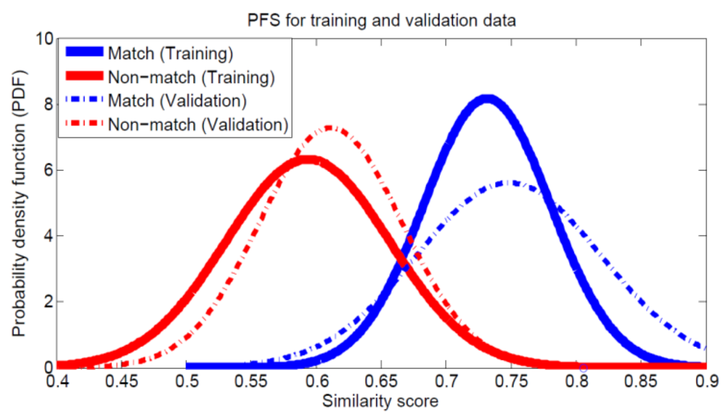
A set of portrait pairs authenticated to be of the same subject are used as training examples to learn PFS and the remaining is used to validate it. We fuse scores obtained from LF and AD features of these images in a way such that the resulting distribution of match and non match scores are as peaked and disjoint as possible so as to enable efficient decision making in identification cases. Towards this, we employ the following methodology.

1. We consider a convex combination of the scores from the two measures LF and AD as $\lambda * s_{LF} + (1-\lambda) * s_{AD}$, $\lambda$ being varied from 0 to 1 in steps of 0.1.
2. For every $\lambda$, we evaluate the mean (average) and standard deviation of match and non-match scores using the RANSAC algorithm [4] to prune outliers.

3. At each $\lambda$, we evaluate the Fisher linear discriminant function, $J= S_b/S_w$, 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.
4. The distributions of match and non-match scores, with obtained $\lambda_{opt}$ in Step 3, for the combined (LF and AD) feature set are modelled as Gaussians distributions (a type of PDF) with means and standard deviations estimated from Step 2.

*3. Validation of learned PFS*

The goal of validating the learned PFS is to verify that the match/non-match scores are indeed representative of the similarity between portrait pairs. Towards this, we perform two-fold cross validation on the set of images, i.e., we divide the set of instances into two groups, 1 and 2. First, we learn the PFS from images of group 1 and validate on 2 (verify if the similarity scores of images under group 2 agree with the PFS as learned from images in group 1). Next, we learn PFS from group 2 and validate on group 1. Mean and standard deviations of match and non-match scores from two folds are averaged to obtain the resulting curves shown in Fig 2. It is to be noted that these curves are dependent on the data under consideration.




In obtaining the above, we were provided 34 pairs of images where the identities of the subjects were known. A part of these images was used to learn the PFS and the rest was used to validate it. We were also provided 11 pairs of reference and test images wherein we had to find whether the subject depicted in test image (whose identity is unknown) is same as that depicted in reference image (whose identity is known). The art works consisted of death masks, paintings and sculptures of several aristocrats.

*4. Identification Framework*

Given the learned PFS, the question now is to verify an unknown test image against a reference image. Towards this, we employ hypothesis testing.

*5.1 Hypothesis Testing*

This is a method for testing a claim or hypothesis (in this case that of a match/non-match between portrait pairs) [5]. Below, we summarize it with respect to the learned PFS in arriving at the conclusion for a match.

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 (test's probability of incorrectly rejecting the null hypothesis) as 0.05, as per common practice in such problems.
3. We compute the test statistic using one independent non- directional z test [5], which determines the number of standard deviations the similarity score deviates from the mean similarity score of the learned distributions
4. We compute *p* values which are the probabilities of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. If $p<\alpha$ we reject the null hypothesis.

*5.2 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 different persons other than the one depicted in reference image. We call these images as distracters. Depending on availability, we choose similar works by the same artist (artist of reference image) as distractors. If a test image indeed represents the same subject as in the reference image, not only should its score with the reference image be modeled through match distribution, but also its scores with distracter faces should be modeled by non-match distribution.

*5.3 Analysis Scenarios*

We computed similarity scores of test cases with corresponding reference image and with 10 distracters. Table 1 lists various hypothesis test scenarios that can arise [5] and the corresponding conclusions that one can infer. Match and non-match cases are straight forward to infer from Table 1. In cases where both match and non-match distributions are likely to account for the test data in the same way, it can be said that the learned PFS cannot accurately describe the test data (black rows in Table 1). If either match or non-match distribution is more likely to account for both test as well as distracters (magenta rows in Table 1), 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 | | Distractors | | Conclusion |
|---|---|---|---|---|
| Match | Non-match | Match | Non-match | |
| $p > \alpha$ | $p \leq \alpha$ | $p \leq \alpha$ | $p > \alpha$ | Match |
| $p \leq \alpha$ | $p > \alpha$ | $p < \alpha$ | $p \geq \alpha$ | Non Match |
| $p > \alpha$ | $p > \alpha$ | $NA$ | $NA$ | No decision |
| $p < \alpha$ | $p < \alpha$ | $NA$ | $NA$ | No decision |
| $p > \alpha$ | $p \leq \alpha$ | $p > \alpha$ | $p \leq \alpha$ | No decision |
| $p \leq \alpha$ | $p > \alpha$ | $p \leq \alpha$ | $p > \alpha$ | No decision |

References:
[1] J.Li, L.Yao, and J.Wang. *Rhythmic brushstrokes distinguish Van Gogh from his contemporaries: Findings via automated brushstrokes extraction* in IEEE Trans. Patt. Anal. Mach. Intell., 34(6):159--176, 2012.

[2]  L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg. *Face recognition by elastic graph bunch graph matching* in  IEEE Trans. Patt.Anal. Mach. Intell.}, 7:775--779, 1997.
[3] L.Farkas. *International anthropometric study of facial morphology in various ethnic groups/races* in  Journal of Craniofacial Surgery, 16(4):615--646, 2005.
[4] M. Fischler and R. Bolles. *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. In Comm. of the ACM, 24(6):381--395, 1981.
[5] J. Shao. *Mathematical Statistics*. in Springer New York,ISBN-13: 9781441929785, 2nd Edition}, 2012.

We had two provisional articles published on FACES during this period [Ramya Srinivasan, Amit Roy-Chowdhury, Conrad Rudolph, and Jeanette Kohl, "Recognizing the Royals--Leveraging Computerized Face Recognition to Identify Subjects in Ancient Artworks," *ACM International Conference on Multimedia* (2013) 581-584; and Ramya Srinivasan, Amit Roy-Chowdhury, Conrad Rudolph, and Jeanette Kohl, "Quantitative Modeling of Artists Styles in Renaissance Face Portraiture," *Second International Workshop on Historical Document Imaging and Processing* (2013) 94-101.]  These are provisional papers; the final primary publications (one oriented toward computer science, one oriented toward the humanities) for this project have not yet reached the publication stage.

We also plan to disseminate our findings through a dedicated website.