Document Title: **Face Annotation at the Macro-scale and the Micro-scale: Tools, Techniques, and Applications in Forensic Identification**

Author(s): **Patrick J. Flynn, Anil K. Jain**

Document No.: **243277**

Date Received: **August 2013**

Award Number: **2009-DN-BX-K231**

Final Report for 2009-DN-BX-K231

Face Annotation at the Macro-scale and the Micro-scale: Tools, Techniques, and Applications in Forensic Identification

Principal Investigators:
Patrick J. Flynn, Univ. of Notre Dame
Anil K. Jain, Michigan State Univ.

Postdoctoral Scholars:
Gaurav Aggarwal, Univ. of Notre Dame
Unsang Park, Michigan State Univ.

University of Notre Dame

The Notre Dame team consisted of graduate RA Nisha Srinivas (joined the project in January 2010), postdoctoral scholar Gaurav Aggarwal (joined the project in June 2010), and PI Flynn. Research activity has focused in a number of areas, and multiple publications (attached) have resulted from this work. Software developed during the project was transferred to the National Institute of Justice Sensors, Surveillance, Biometrics Technology Center of Excellence (CoE) in Q3 of CY 2012 for evaluation.

Annotation and Detection of Facial Marks and use for Recognition:

We developed, evaluated, and documented a system [2][3] for facial verification that employs detected facial marks rather than holistic or larger-scale facial structures. The components of this research project and developed system are:

1. <u>Manual face mark annotation and comparison of human annotator performance</u>. Consistency between human annotators was assessed using a database of face images of identical twins, a specified catalog of facial mark types (e.g., mole, freckle, scar) and three human annotators. Significant variations were observed between human annotators with respect to the number of marks annotated and the labels applied to marks. This motivates the use of an automatic face mark detector, and additionally motivates the need for training of personnel in the area of mark localization and labeling. The consequences of human annotation inconsistency are documented below.

2. <u>Automatic face mark detection</u>. The multiscale facial mark detector developed by MSU (and described below) was evaluated as one of several techniques for facial mark detection. Ultimately, we employed a multiscale technique originally proposed by Loy and Zelinsky [4]. Other facial mark detectors are currently being examined in current work.

3. <u>Representation of facial mark sets, and exploitation for verification</u>. Representation of a set of "point" facial marks is relatively simple: the representation contains (minimally) a set of point locations in image coordinates and zero or more attributes (e.g., face mark type) on each point in the set. One may also precompute normalized coordinates for points in some appropriate face-centric coordinate system, and perhaps interpoint distances as well. In the future, means for expressing the extent of features that describe curves or regions containing one or more marks may be needed. Verification of a claimed facial identity requires the establishment of a correspondence between a subset of features in the probe (questioned) image and a subset of

features in the gallery image of the claimed identity's face. Figure 11 depicts such a correspondence, for a correct match. Our system uses a bipartite graph matching algorithm to form a set of correspondences with a matching score equal to the ratio of the number of compatible correspondences to the number of matchable marks. The miscorrespondences visible in Figure 1 are overcome by aggregating evidence for the correct matches using the RANSAC algorithm. Figure 2 depicts receiver operating characteristic (ROC) plots of matcher performance for several verification task designed to test distinguishability of twins (i.e., the set of non-match test pairs was restricted to contain matches between twin siblings). The different ROC curves depict different subsets of the different image data as well as different extraction strategies (manual versus automatic annotation). We consider the performance of this matcher to be good given the sparse nature of the features used. The sensitivity of the technique to annotation quality is a discouraging aspect of the results (this, however, can likely be partially mitigated by a training regime designed to yield repeatable results).

<u>Current work: Accommodation of non-frontal face pose (planned)</u>. The Notre Dame team is continuing to work on this problem after the end of the grant. The next technical challenge is to accommodate the nonuniform movement of facial marks observed when the face pose relative to the camera is changed. As a consequence of foreshortening, marks on portions of the face nearer the camera move a larger distance under a rotational pose change (pitch or yaw), and only knowledge of the 3D face shape can completely compensate for these effects. Thus, the question to be addressed is whether a 3D model should be estimated from training data, from training and testing data, or not at all. Answering this question will depend on the number of images with different facial pose available in the training and probe sets. We expect to investigate and possibly employ 3D modeling techniques or software packages to determine the feasibility of 3D face modeling at the necessary degree of accuracy given typical SLR camera imagery.

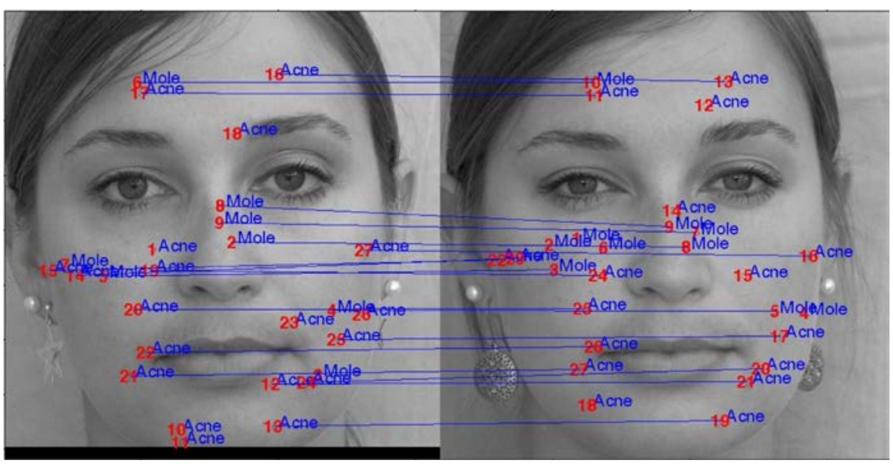

Figure 1. Depiction of face mark correspondences between two images of the same person. There are also a few miscorrespondences present; the RANSAC method allows us to discard such correspondences by aggregating evidence from multiple sets of correct matches.

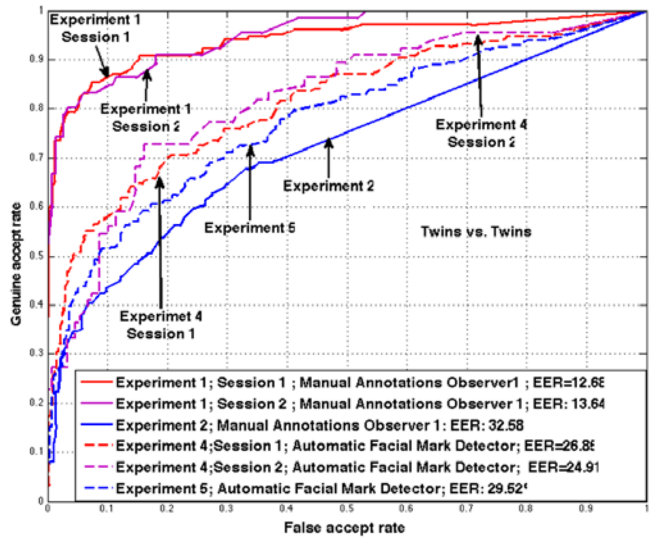

Figure 2. Performance of the face mark matcher on a twins-versus-twins verification experiment (From Reference 1).

Application to discrimination of identical twins:

The face mark extraction and matching approaches were applied to a data set of identical twin face images captured in 2009 and 2010 at the Twins Days festival in Twinsburg, Ohio. Identical twins represent a strong challenge to face recognition using traditional methods and it was hypothesized that facial marks may help to distinguish twin siblings.

Transition activity:

By agreement with the Program Manager, the face mark detection and matching software will be provided to the CoE's evaluation team for testing, along with data sets used. We anticipate some interactions after the end of the project regarding the results of testing and possible next steps.

Publications developed with support from this project

[1] Nisha Srinivas, Gaurav Aggarwal, Patrick Flynn, and Richard Vorder Bruegge, "Analysis of Facial Marks to Distinguish Between Identical Twins," *IEEE Trans. Information Forensics and Security* 7(5):1536-1550, October 2012. DOI: 10.1109/TIFS.2012.2206027.
[2] N. Srinivas, G. Aggarwal, P. Flynn and R. Vorder Bruegge, "Facial Marks as Biometric Signatures to Distinguish between Identical Twins," *Proc. CVPR Workshop on Biometrics*, June 2011, pp. 106-113. DOI: 10.1109/CVPRW.2011.5981818
[3] N. Srinivas, M. Pruitt, G. Aggarwal, P. Flynn and R. Vorder Bruegge, "Preliminary Assessment of Discrimination of Twins in Photographs Based on Facial Blemishes," abstracted in *Proc. 63rd AAFS Annual Meeting (American Academy of Forensic Sciences)*, p. 140, Chicago, January 2011.

Additional References

[4] G. Loy and A. Zelinsky, "A fast radial symmetry transform for detecting points of interest," *Proc. 7th European Conf. on Computer Vision-Part I, ECCV '02*, (London, UK), pp. 358–368, Springer-Verlag, 2002.

Michigan State University

Our contributions on this project can be summarized into three major topics: (i) development of a robust facial mark detection method using scale space analysis, (ii) showing the utility of facial marks in face image retrieval tasks using a large scale face image database and (iii) symbolic query based face image retrieval study using facial marks when query face image is not available.

1. Robust Facial Mark Detection Using Scale Space Property

Local keypoint detection method using scale space analysis has been extensively used in object matching and retrieval. The Scale Invariant Feature Transformation (SIFT) is a well known keypoint detection and image matching method using the scale space analysis. We employed the scale space analysis for the facial mark detection process. The scale space extrema (keypoint) detection starts by constructing a normalized multi scale representation of the face image by convolving the input image with a Laplacian of Gaussian (LoG) filter with a sequence of standard deviation values. Next, we detect local extrema over both spatial and scale direction in every 3×3×3 image pixel block. The detected candidate mark locations have the following properties.

- Location of local extrema contains the candidate facial mark.
- The detected scale (standard deviation) indicates the size of the corresponding facial mark.
- The absolute value of LoG filter response reflects the strength of the response. This strength is used as the confidence value to select stable marks.
- The sign of LoG filter response provides an evidence for the type of detected facial mark. Positive (negative) sign represents a dark (bright) facial mark with brighter (darker) surrounding skin.

A test with a face image database with 200 subjects shows that the proposed mark detection algorithm improves the accuracy (58.7% recall) over our earlier mark detection method (42.9% recall) [1]. We also conducted face matching experiments using a large face image database obtained from Pinellas County Sheriff's Office (PCSO) and the twin database collected by the Univ. of Notre Dame containing 89 different identical twin pairs. The experimental results are summarized below [2].

- The rank-1 matching accuracy improved from 56.3% (COTS FR engine) to 57.2% (COTS FR engine + face mark matcher) in the matching experiment using a subset of PCSO database (1,000 probe and 100,000 gallery images, one image per subject).

- The rank-1 matching accuracy improved from 90.8% (COTS FR engine) to 92.1% (COTS FR engine + face mark matcher) in the matching experiment using the Univ. of Notre Dame identical twin database (404 images of 178 subjects, 89 twin pairs, each image is used as probe and the rest as gallery).

2. Image Query Based Retrieval

We have studied the utility of facial marks as a filter to quickly narrow down candidate matching face images on a large scale database. The basic idea is to generate a short fixed length histograms containing facial mark information (i.e., type and location) and using them as index to quickly retrieve candidate matching face images. Some of the facial marks are not stable due to their temporary nature (e.g., pimples or zits) or poor image quality (e.g., low resolution or image blur). Therefore, the mark based index cannot be used to filter the database with the same high confidence as the demographic information such as gender or ethnicity. We propose a conditional filtering process using a leading COTS face recognition engine (FaceVACS) as a measure of the quality of the retrieved results. The conditional filtering process includes the following steps.

(i) Mark index based filtering is first applied to retrieve a short list of gallery images.
(ii) FaceVACS is used to sort the subset of gallery images according to their similarity to the probe image in a descending order.
(iii) The maximum match score from FaceVACS in the filtered list is subjected to the following test; if it is lower than a threshold value, t_{FV}, then FaceVACS is used to match all the images in the gallery (assuming the correct mate(s) is accidentally filtered out); otherwise, the filtered list serves to generate the retrieval result.

In the worst case scenario, the matching time of this scheme will be slightly increased because of the cost of additional mark index based matching in addition to the FaceVACS. However, given the speed of face matchers and mark-based matching, this increase in computation is insignificant.

We have measured the average matching time using a small database with 1,000 probe and 1,000 gallery face images both for FaceVACS and mark index based matcher (template construction time is not included). The entire matching time on a subset of PCSO database (1K probe and 100K gallery, DB1) is extrapolated as 6 hours and 12.5 minutes with FaceVACS and mark index, respectively. Therefore, it is expected that applying mark index based filtering first, followed by FaceVACS on the filtered set of candidate images will lead to a more efficient search process. In the filtering experiment with DB1, the overall matching time is decreased by 66 mins. (18.4% speed up) with 7.1% loss of matching accuracy (t_{FV}=0.6), or 24.8 mins. (7% speed up) with 0.5% loss of rank-1 matching accuracy (t_{FV}=0.9) (match scores of FaceVACS are in the range [0,1]. The instability of some of the marks requires a large penetration rate (about 50%) to achieve high retrieval accuracy.

3. Symbolic Query Based Retrieval

There are situations when the gallery database needs to be searched in the absence of a query image. These situations occur when there is no surveillance or any other type of camera available to capture a suspect's facial image during a criminal activity. However, a witness to the crime is able to provide a verbal description of the suspect in the form of a statement like “white Caucasian male, about 30 years old, tall with a dark mole on the right cheek.” Our goal is to build a face retrieval system that can utilize the demographic information as well as the facial mark information to drastically filter the gallery. It is hoped that the witness will now be able to examine this short list of gallery face images and positively identify the suspect. Our mark based indexing scheme can be used for this non-image (or symbolic query) based retrieval. The mark index is constructed based on any mark observed in the suspect's face in terms of the morphology, color, and location. This mark index is then compared with the indices of the gallery images to retrieve the candidate images.

We used 100 face images as the probe images in the PCSO database to evaluate the symbolic query based retrieval method. We manually identified distinctive marks on these 100 probe face images, and derived symbolic queries based on the labeled marks. For example, a point shaped dark mark can be interpreted as a mole, and a white spot can be interpreted as a blemish. Some of the example queries are listed below.

- A mole on the right side of forehead
- A mole on the left side of the chin
- A scar on the right cheek

Since we know which symbolic query is generated from which subject in the probe set, and since we also have the mate images in the gallery database, we can evaluate these symbolic queries in terms of the retrieval accuracy. About 48% of the symbolic queries successfully retrieved the correct mates in the filtered results at 42% of penetration rate. This is quite a good performance considering that a conventional image based face matcher cannot operate with these symbolic queries. Fig. 1 shows example retrieval results from symbolic queries. The retrieved face images of the correct mates from whom the symbolic queries are constructed are enclosed with red boxes. Additional demographic constraints can be incorporated into the retrieval process (e.g., skin color) but it should also be possible to execute the system without these constraints in order to make it robust to deliberate attempts to obscure those constraints (e.g., facial cosmetics, gender attributes).

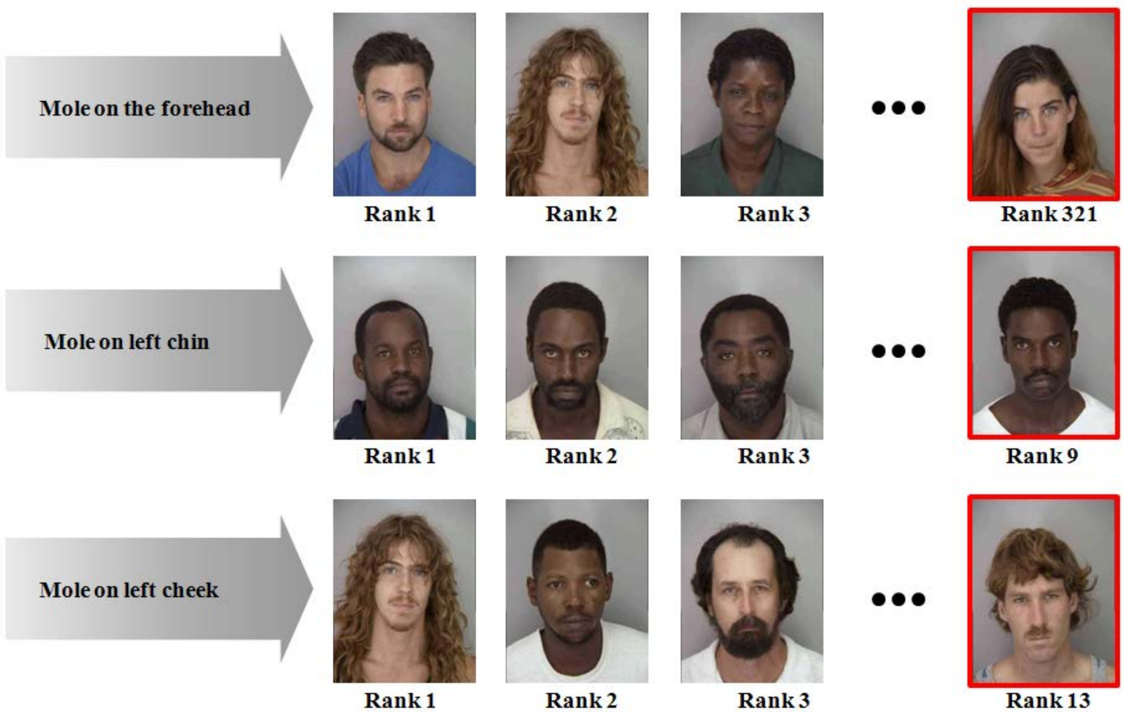

Figure 1. Example retrieval results using symbolic queries.

Besides the aforementioned three major contributions, we also studied the utility of facial marks with facial pose variations. A subset of face image database with 90 degree profile images was collected from the FERET database (132 images from 132 subjects). This database along with the earlier face image database collected from FERET (426 images from 213 subjects) [1] was used to study the facial mark detection and matching with pose variations. All facial marks are manually labeled using the six mark categories (based on morphology and color) and the previously proposed automatic facial mark detection method is applied. The automatic facial mark detection performed with 41.8% of recall and 26.8% of precision with profile face images. We have tried two matching/retrieval schemes: (i) using location only and (ii) using both location and mark class information. It was observed that the matching scheme using location information alone performed better than using both location and mark class when there is a large pose variation. This is because the automatic mark classification is not very stable in the presence of large pose variations.

To summarize, we have developed a robust facial mark detection method that provides both location and scale information of facial marks. We have shown the utility of facial marks in face image retrieval tasks using a large scale face image database. Our proposed facial mark based matching scheme can be combined with COTS face matcher to improve the overall matching accuracy. The improvements of face matching accuracy are shown both on general face image database and identical twin face image database. Next, we proposed a conditional filtering scheme to reduce the retrieval time on a large scale face image database. Our study on symbolic query based face image retrieval shows a great utility of facial marks when query image is not available.

Even though we have shown a number of cases where facial marks can be used for face image retrieval tasks, we have observed that the instability of facial marks is still a major bottleneck in using of facial marks in practice. The instability of facial marks are caused partly by the nature of facial marks, but mostly caused by the low quality face mug shot images. Therefore, the development of a more robust facial mark detection and representation scheme and establishing a standard practice in face mug shot image capture process to ensure sufficient quality of images for accurate mark detection are some of the imminent requirements to extend the use of facial marks.

[1] U. Park and A. K. Jain, Face Matching and Retrieval Using Soft Biometrics, *IEEE Transactions on Information Forensics and Security (TIFS)*, Vol. 5, No. 3, pp. 406-415, 2010.

[2] U. Park, S. Liao, B. Klare, J. Voss, A. K. Jain, Face Finder: Filtering a Large Database using Scars, Marks and Tattoos (SMT), MSU Technical Report, MSU-CSE-11-15, 2011.