Suspect Identification Based on Descriptive Facial Attributes

Brendan F. Klare*   Scott Klum*   Joshua C. Klontz*   Emma Taborsky*   Tayfun Akgul†   Anil K. Jain‡

Abstract

We present a method for using human describable face attributes to perform face identification in criminal investigations. To enable this approach, a set of 46 facial attributes were carefully defined with the goal of capturing all describable and persistent facial features. Using crowd sourced labor, a large corpus of face images were manually annotated with the proposed attributes. In turn, we train an automated attribute extraction algorithm to encode target repositories with the attribute information. Attribute extraction is performed using localized face components to improve the extraction accuracy. Experiments are conducted to compare the use of attribute feature information, derived from crowd workers, to face sketch information, drawn by expert artists. In addition to removing the dependence on expert artists, the proposed method complements sketch-based face recognition by allowing investigators to immediately search face repositories without the time delay that is incurred due to sketch generation.

1. Introduction

Despite the continued ubiquity of surveillance cameras networks, a large number of crimes occur where only a witness description of a subject's appearance is available. The ability to accurately search a face database or videos from surveillance networks using verbal descriptions of a subject's facial appearance would have tremendous implications in the timely resolution of criminal and intelligence investigations. Pattern recognition technology *should* support such an identification paradigm where a human can describe the appearance of a subject's face to directly search a media repository. The goal of this research is to understand whether technology *can* currently support such a paradigm.

A major progression in the ability to search face image databases using verbal descriptions has been realized through a long line of research in matching hand drawn facial sketches to photographs [19, 15, 22, 21, 23, 9, 7, 3].

While automated sketch recognition technology offers a clear advantage over the legacy approach of disseminating a sketch through media outlets, issues with the sketch generation process can limit the use of sketch recognition to only high profile crimes. For example, while sketch recognition systems can leverage the expertise of a forensic sketch artist, they are equally limited by the requirement of having such an expert available to generate the sketch.

Another limitation with sketch recognition is the time delay between when a crime occurs, when a sketch artist can be deployed, when the artist finishes eliciting the witness for gathering enough information to draw the sketch, and when the sketch is finalized for dissemination. Such delays can prove costly in time sensitive investigations.

Finally, sketch-based face recognition is often hampered by noisy information provided by witnesses. A major reason for this is that a generated sketch provides no information regarding which regions of the face the witness feels most confident in describing. Because a witness may have varying degrees of confidence for different facial features, weighting (or removing) certain features to reflect witness's confidence should improve the retrieval process.

Despite its limitations, the use of hand drawn sketches has several distinct advantages: sketch artists often have specialized training to elicit witness memory descriptions, generated sketches can be disseminated to the public, and a sketch can be drawn with exact precision. Thus, the work in this paper is meant to supplement, not supplant, the use of sketch recognition technology.

The use of computer generated facial composites partially addresses the aforementioned issue by allowing non-experts (i.e., non-forensic sketch artists) to leverage witness descriptions of a person of interest. Computer generated facial composites typically provide a menu-based interface where each facial component (eyes, nose, mouth, etc.) may be selected to compose a rendered image of a suspect's face. Researchers have recently investigated algorithms that can match the computer generated composites to mug shot databases [5, 13]. However, despite the added benefit of having an image that can be disseminated to media outlets, searching face image databases using computer generated composites is a convoluted process that can be greatly simplified. That is, if the end goal is to search a face database, then the generation of a composite is both unnecessary and

---

*B. Klare, S. Klum, J. Klontz and E. Taborsky are with **Noblis, Falls Church, VA, U.S.A.**

†T. Akgul is with **Istanbul Technical University, Istanbul, Turkey**

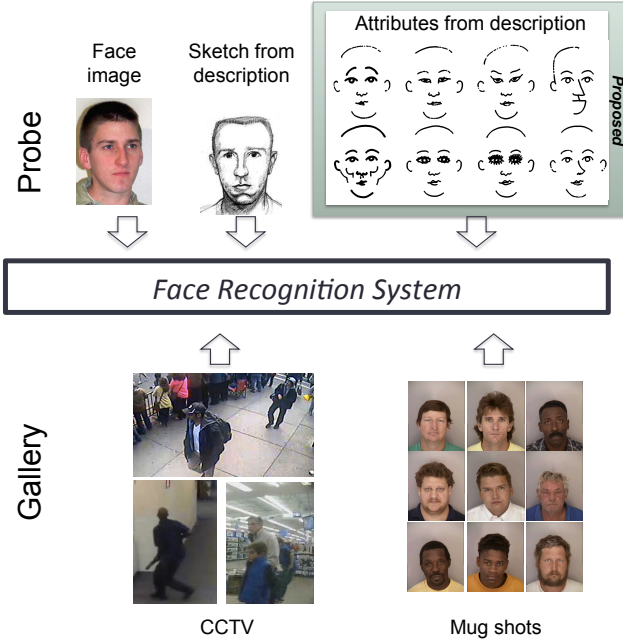‡A. Jain is with **Michigan State University, East Lansing, MI, U.S.A.**

Figure 1. Existing approaches to automted face identification of a suspect are generally limited to querying target repositories with either face images, or hand drawn sketches from verbal descriptions. We propose a system for performing suspect identification based on described facial attributes.

may even be a source of noise since such a software system is not designed with the intention of performing automated face recognition. Furthermore, the issue of low confidence regions still manifests with computer generated composites; the output composite has no indication of the witness's confidence in a given facial region.

In this paper we perform an initial investigation into witness-based identification using facial attribute recognition. The objective is to use witness provided attribute descriptions to search large scale face image repositories (e.g., mug shots databases or videos from surveillance networks) of automatically extracted attribute information. This approach is motivated by the aforementioned deficiencies in sketch recognition algorithms.

The goal of this paper is to address many of the fundamental challenges in using verbal descriptions of facial appearnce to search media repositories for persons of interest [2]. Our research focuses on the scenario of manually labelled query images to measure the inherent feasibility of the proposed search paradigm.

A critical factor when using face attributes to search a database is the development of a set of attributes (features) which compactly, yet concisely, represent the face. As previous research has suggested [20], caricature recognition is enabled by targeting the postulated sparse encoding of predominant facial attributes in the human brain. The facial attributes developed in this project were motivated by such

findings, and expand on other research related to matching caricatures to photographs [10]. Section 4 discusses the development of these attributes, the use of crowd sourced annotation to label a corpus of data, and provides an analysis of the consistency and discriminability of the choosen attributes.

Witness-based identification from facial attributes uses manually provided attributes as query information. However, practical applications require that the target gallery be automatically encoded with the facial attributes. In Section 5 we provide an algorithm to perform such automated extraction. The proposed algorithm operates by performing face component localization and alignment, followed by texture descriptor encoding and support vector regression.

The results shown in Section 6 demonstrate that the proposed method has efficacy in searching face image databases. As such, we provide a sound basis for performing witness-based identification with the following advantages over sketch recognition: (i) facial attribute descriptions can be provided by non-experts using menu driven software, (ii) face image databases can be immediately searched using attributes, and (iii) witness attribute search replaces the indirect path of computer/artist generated sketches followed by mug shot retrieval. Other contributions of the research described in this paper include the development of a set of 46 carefully crafted attribute features, an algorithm for automatically extracting attributes from face images, and the ability to improve sketch recognition accuracy by fusing sketch recognition with facial attribute recognition.

2. Related Research

Two lines of face recognition research have motivated this study: sketch recognition and attribute-based face recognition.

The notion of automatically matching a hand drawn sketch of a face to photographs was popularized by Tang *et al*. through a series of early papers that sought to synthesize a photograph from a sketch [19, 15, 21]. These approaches were evaluated on databases consisting of a photograph and a "viewed sketch" (or "viewed composites"). A viewed sketch refers to a hand drawn sketch where the sketch was drawn while looking at a photograph of the subject. This approach is hypothetical: if a high quality photograph of a suspect were available his sketch would not be needed. However, it was important for the problem to be presented in this manner to isolate the heterogeneity between sketches and photographs and perform an initial investigation of the problem. We are similarly motivated to isolate and understand the challenge of matching human (witness) provided facial descriptions to automatically derived features.

Klare et al. extended the work of Tang and Wang by examining the case of matching sketches drawn based on wit-
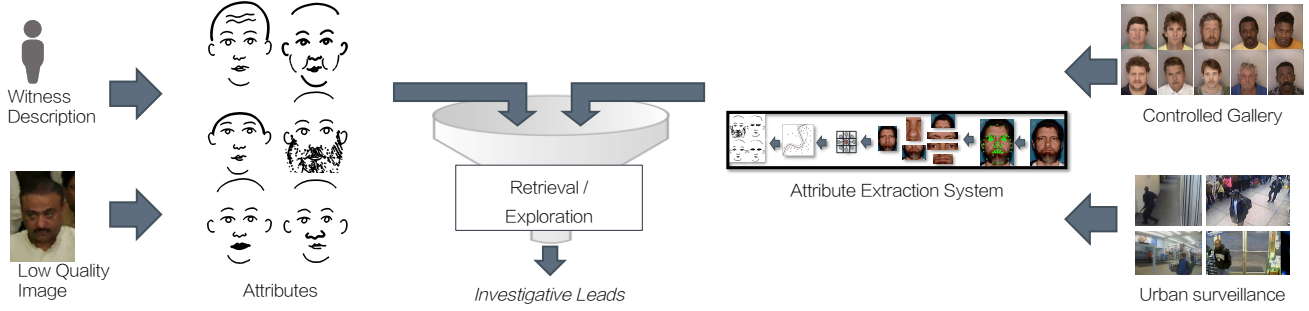
2

Figure 2. The use of human derived face attributes to query mug shot galleries and nearby surveillance cameras is proposed to aid in the timely generation of investigative leads in criminal investigations. This paper focuses on the application of using witness provided descriptions as query, and a controlled gallery as the target. However, the other use case illustrated here (low quality image as query and surveillance imagery as target) is also supported by the research presented in this paper.

ness descriptions (i.e., forensic sketches) [9]. The increased difficulty of matching operational sketches necessitated an alternative algorithmic solution, which was provided in the form of encoding sketches and photos with feature descriptors, and extracting features that were consistent between the sketch and photo modalities [8]. This approach was used in several subsequent works that improved the state of the art in sketch recognition [23, 3]. More recently, the problem of matching computer generated composites to photographs has been explored [5, 13].

A major contribution to the face recognition community has been the research by Kumar et al. on matching faces using attributes [14]. Motivated by the demonstrated properties of attribute-based representations in other pattern recognition problems, and the desire to semantically search for face images, the solution achieved notable recognition accuracies and the desirable properties of a compact representation and human interpretable features.

Our work seeks to build on this previous work on attribute-based face recognition [14], and adapt it to the problem at hand: witness-based face identification. As such, our work can be differentiated in two ways. First, the proposed attribute-based facial features have been crafted with persistence and uniqueness properties in mind [6]. For example, the previous approach [14] to attribute-based face recognition used features that have no identity information, such as "black and white photo", "flash", "posed photo", or "teeth visible". Other features used are rather subjective such as "attractive woman". By contrast, we have developed features using a human face representation expert to systemically describe the facial characteristics that most concisely convey the identity of a person of interest. The second differentiation is that our approach is motivated by the need to perform face recognition queries using descriptions from witnesses of a crime. As such, we are primarily conducting experiments to understand the trade offs between attribute-based face recognition and sketch recognition.

3. Datasets

The majority of the experiments in this paper were conducted on the CUHK Face Sketch FERET Database (CUFSF) [21, 23]. This database consists of 1,194 photographs from the FERET [17] database, and 1,194 hand drawn sketches of the subjects in the photographs which were generated at the Chinese University of Hong Kong. Each sketch in the CUFSF database is drawn by an artist while viewing the corresponding photograph. Thus, the appearance and structure of the sketches is highly accurate. Of these 1,194 subjects, 175 had an additional photograph in the FERET database with time lapse between the original image. These additional 175 images were used in certain experiments.

4. Facial Attributes

The success of attribute-based face recognition hinges on the development of a set of facial attributes that both capture all the variations in facial appearance, and are terse enough for a witness to provide soon after a crime occurs. To yield such attributes, our research was performed in collaboration with an artist who specializes in the caricaturing of faces using a minimal number of features. The result was a set of 46 facial attributes that capture component level information (the appearance of the eyes, nose, mouth, etc.), the relationship between components (e.g., distance between the nose and mouth), and holistic information such as gender or wrinkles. The developed features are qualitative, and categorical. Of the 46 features, 19 are binary with two feature categories (e.g., unibrow[1] vs. no unibrow). An additional 19 features have three categories (e.g., nose size small, normal and large). Of the remaining features, six features have four categories, one feature has five categories, and one feature has seven categories. The features with a large number of categories are generally holistic features that described

---

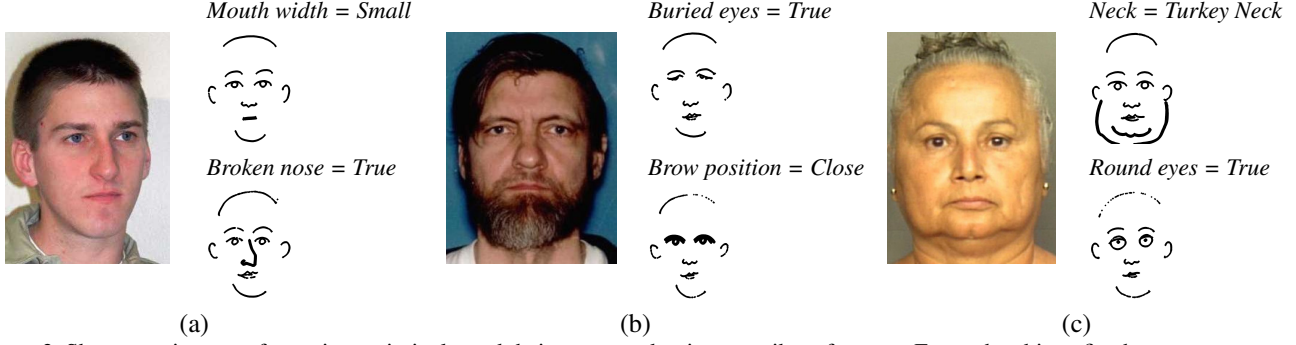[1]A unibrow is when the two eyes brows are connected.

Figure 3. Shown are images of notorious criminals, and their more predominant attribute features. For each subject, five human annotators labelled these images with all 46 attributes defined in this study. (a) Timothy McVeigh was consistently labelled as having a small mouth and a broken nose; (b) Ted Kaczynski was labelled as having buried eyes and eyebrows close to his eyes; and (c) Griselda Blanco was labelled as having round eyes and a turkey neck. The goal of this study is to understand whether attributes derived from amateurs can improve on limitations of hand drawn facial sketches.

the appearance of hair or other specific attributes.

With a total of 46 attributes defined, the next step was to have the datasets described in Section 3 manually annotated. To obtain attribute type labels from human annotators, we created Human Intelligence Tasks (HITs) using the Amazon Mechanical Turk (AMT) service. A HIT Type was created for every attribute, and a photograph of the subject was displayed along with simple line drawings representing the different categories for each attribute. Annotators were then asked to choose the drawing that best matched the subject in the image. In most cases, an example photograph exhibiting the attribute type was provided to the annotators and displayed along with the line drawings. To eliminate any possible language barrier, tasks included a minimal amount of verbage describing the attributes.

Every HIT had three assignments, meaning each attribute type was labelled by three different Turk workers for every image. This variety of responses allows for an attribute confidence level to be inferred from the amount of consensus.

In order to better understand the nature of the attributes and data collected from Turk workers, we measured the *entropy* and the *consistency* of the responses. The entropy of an attribute estimates how much information each attribute might provide about a face. The greater the entropy of an attribute, the more potential it has to discriminate between faces. For a given attribute feature f, the entropy of that feature $H(f)$ is defined as $H(f) = -\sum_{i=1}^{n_f} p(f_i) \log_2 p(f_i)$, where n_f is the number of possible categories for a given features, and $p(f_i)$ is the probability that the i-th category of feature f will occur. $p(f_i)$ is measured empirically based on the statistics of manual annotations.

The overall consistency is a simple way to assess how difficult or subjective an attribute is for human labelers. If all labelers chose the same attribute type for an image, that attribute is considered 'consistent' for that image. If label-

ers chose two attribute types it is labeled as 'partial', and all three labelers disagreeing means 'inconsistent' attribute.

Figure 4 lists all the attributes used in this study, as well as the entropy and consistency of those features.

5. Automated Attribute Extraction

While the attributes for the query (person of interest) can be manually elicited from the source (witness memory or low quality image), the target gallery is typically very large, and must be automatically processed. As such, the algorithmic extraction of attribute labels from face images is required to support the proposed system. In this section we describe the process for attribute extraction. A summary of

Attribute	Entropy	Consistency			
Lip Thickness	1.90	0.20	Face Marks	1.20	0.48
Eye Slant	1.69	0.28	Sharp Eyes	1.00	0.31
Thick Eyebrows	1.55	0.34	Face Length	1.00	0.57
Ear Size	1.54	0.26	Small Eyes	1.00	0.61
Face Shape	1.53	0.34	Gender	0.98	0.95
Eye Separation	1.53	0.21	Smiling	0.95	0.64
Cheek Density	1.53	0.28	Baggy Eyes	0.94	0.42
Nose Size	1.53	0.31	Buried Eyes	0.93	0.47
Eyebrow Position	1.52	0.33	Bent Eyes	0.92	0.40
Ear Pitch	1.52	0.38	Eyelash Visibility	0.90	0.41
Chin Size	1.51	0.22	Almond Eyes	0.90	0.41
Nose-Eye Distance	1.50	0.22	Beard	0.89	0.81
Nose-Mouth Distance	1.47	0.28	Mouth Asymmetry	0.86	0.52
Mouth Width	1.45	0.30	Line Eyes	0.86	0.53
Forehead Size	1.45	0.32	Round Eyes	0.85	0.58
Nose Width	1.44	0.38	Sleepy Eyes	0.83	0.56
Eye Color	1.43	0.36	Widows Peak	0.79	0.51
Eyebrow Orientation	1.41	0.33	Hairstyle	0.77	0.64
Mouth Bite	1.32	0.34	Hair Density	0.70	0.82
Nose Orientation	1.31	0.37	Broken Nose	0.69	0.60
Neck Thickness	1.30	0.37	Forehead Wrinkles	0.62	0.84
Hair Color	1.29	0.47	Unibrow	0.47	0.82
Mustache	1.22	0.75	Glasses	0.36	0.97

Figure 4. Listed are names of all 46 facial attributes used in this study, along with their entropy and consistency based on manually assigned values. Consistency values shown are the average for a given attribute across all 1,194 images.

| Input Image | Landmark Detection | Component Cropping | LBP Descriptor Representation | SVM Regression | Attributes Likelihoods |

Figure 5. The process for automatically extracting attributes from face images is illustrated above. Each attribute regressor function operates on the component cropping corresponding to the region of the face related to that attribute. For example, mouth attributes use descriptor representations extracted from the cropped mouth component.

this approach is provided in Figure 5. The software implementation of the described algorithm has been performed within the OpenBR framework [11].

The first step in automatic extraction of facial attributes is to localize the face. Localization is performed by detecting the eye locations using a pre-trained black box eye detector. An affine normalization transformation is applied to the image such that, after cropping the image to 192x240 pixels, the normalized eye locations are 34.5% in from the sides and 47.5% in from the top. Next, facial landmarks are localized in the normalized image using an active shape model (ASM) via the open source library STASM [16]. To improve the accuracy of detected landmarks, the ASM is seeded with the normalized eye locations detected in the previous step.

The detected landmarks provide the information needed to extract the attributes from the corresponding region of the face. That is, if an attribute is only related to the eyes, it is important to not perform classification using other regions of the face. As such, based on the landmark information provided, a bounding box containing relevant landmarks is used to segment individual facial components (nose, mouth, eyes, hair, brow, and jaw). For attributes that describe a large region of the face (e.g. cheek density, gender), the normalized face image is simply cropped to a tighter region around the face. Each of the attributes used in this study is then assigned to one of the cropped regions. For example, all of the eye attributes are assigned to the eye cropping, holistic attributes are assigned to the face cropping, etc.

After cropping the components, scale space normalization is performed: the mouth, eyes, hair, and brow components are resized to 24x36 pixels; the jaw and face to 36x36 pixels; and the nose to 36x24 pixels. Empirical results show that discarding the aspect ratios had no noticeable impact on recognition performance.

With each component cropped and aligned, the appearance is normalized using the Tan & Triggs [18] preprocessing pipeline to reduce illumination variations. The texture of the components is then encoded using uniform local binary pattern (LBP) [1] features with radius $r = 1$. Descriptor histograms are then computed for 8x8 patches (with a 4

pixel overlap), which are concatenated to form a single feature vector per component x_c (where c specifies the component). The dimensionality of each x_c vector is reduced by projecting into a learned PCA subspace V_c, which preserves 95.0% of the training data variation.

For each facial attribute, the set of n (in our case n is half the available dataset) training vectors $X_c = [x_c^1, \dots, x_c^n]$ is used to train an epsilon support vector regression function using the LibSVM library [4]. Attribute features from the same component will have the same vector as training input, however the target values y_a^i could be different. For attribute a, y_a^i is the percentage of votes that manual annotators provided for that given feature. Thus, with three annotations per feature, the regression function can still leverage inconsistent labels as $y_a^i \in \{0, 1/3, 2/3, 1\}$. The SVR parameters of the RBF kernel (C, γ) are computed automatically during training via cross-validation.

Given the possible designations for the 46 attributes, a total of 86 SVRs are trained based on the use of one-vs-all SVRs for the attributes with more than two possible labels. The output of the 86 regressors are concatenated into a final attribute feature vector v_i for image i. To generate the similarity between two faces i and j, the output of the regressors are concatenated into a final attribute feature vector v_i. Several distance metrics were explored, including L_2, L_1, dot product, and weighting schemes using the entropy and consistency metrics discussed in Section 4. The L_1 distance metric yielded the best results. Thus, the distance between faces i and j is computed as $d(i, j) = \|v_i - v_j\|_1$.

6. Experiments

6.1. Facial Attributes vs. Hand Drawn Sketches

The first set of experiments are designed to compare the use of facial attributes to hand drawn sketches. Using the 1,196 FERET photographs, we randomly partitioned the dataset into training and testing sets using two-fold cross validation. Thus, one half of the images/subjects were available for training, the other half of the images/subjects were available for testing, and this process was repeated two times (with the roles of training and testing sets inter-
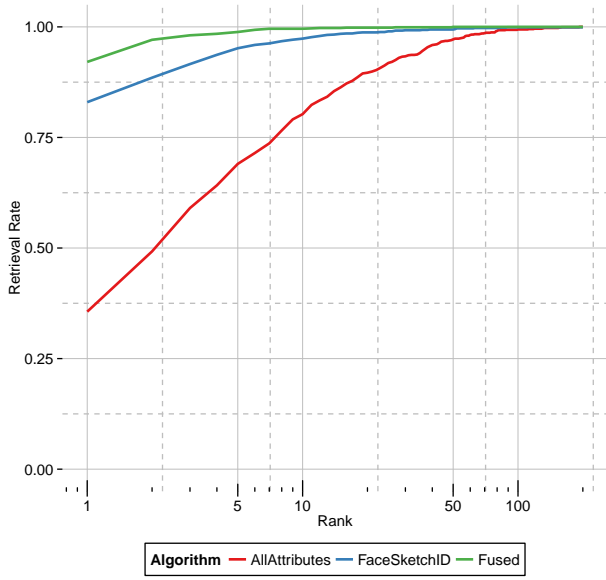
Figure 6. CMC accuracies comparing the use of facial attributes to hand drawn sketches. As expected, the use of attributes provided by amateur annotators does not match the accuracy achieved through expert artists and a sketch recognition system. However, the attributes achieve good accuracy given they contain limited information. Further, the attributes improve the accuracy of the sketch system when fused.

changed); the average results are presented. The training portion of the dataset was used to train the automated attribute extraction system. No subjects used in training were used for testing.

Recognition results were generated on the test data by using the human provided attributes as query/probe, and the features automatically extracted from the photographs for the target/gallery. The baseline used was the MSU FaceSketchID sketch recognition algorithm [12], which achieves state of the art accuracy in sketch recognition. The training data was used to train the discriminative subspaces in the baseline sketch recognition algorithm. The hand drawn sketches from the testing partition were used as probes, and the corresponding photographs were used as the gallery (consistent with the attribute system).

In both systems (the attribute-based and the sketch recognition systems), the gallery images are the same: photographs of subjects from the testing partition. The difference is the query information used to probe these gallery images. In the case of sketch recognition, hand drawn sketches were created by *expert* artists looking at the gallery photograph (again, these are called viewed sketches). In the case of the attribute system, attributes features were selected by *amateur* AMT laborers. While this scenario is hypothetical, our main motivation in this experiment is to answer
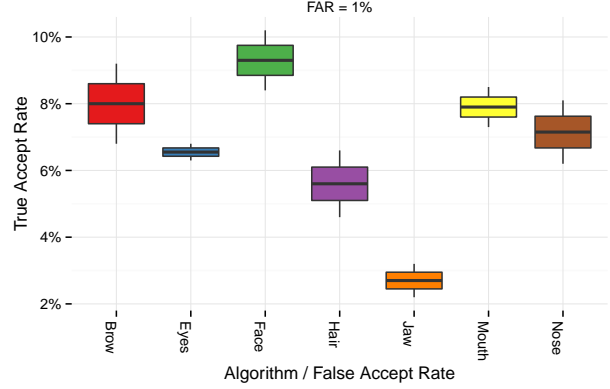

Figure 7. The accuracy of attribute feature categories extracted from different facial components are shown. While the eyes are highly informative in standard face recognition, there is a lot of difficulty in categorically describing eyes, which likely explains its relatively low accuracy. For each component, we list the true accept rate at a fixed face accept rate of 1.0%.

the following question: *can information derived from non-experts (annotators) achieve comparable results to information derived by experts (sketch artists)?*

The cumulative match characteristic (CMC) plot of the recognition accuracies of the attribute algorithm can be found in Figure 6. As expected, the accuracy of the attribute-based representation does not match that of the sketch-based recognition algorithm. This is because of the high precision of the sketches, and that nearly five years was spent developing the sketch recognition algorithm. However, it is notable that the proposed approach takes amateur annotators, and is able to achieve identification accuracy similar to previously reported results on forensic sketches [7]. Further, the fusion with attributes increases the sketch recognition accuracy from 84.5% to 92.0% at Rank-1. Thus, while the intent of this research is to offer an alternate paradigm to leverage witness descriptions, the proposed method can also improve the accuracy of a well tuned system. That is, even when a expert forensic artist is used to create a facial sketch, it may be valuable to use the verbal description.

6.2. Experiment 2: Comparison of Components

Next we compare the recognition accuracy of each facial component. We use manual annotations as the probe information, automated extractions as the gallery information, and two-fold cross validation in the same manner as the first experiment in Section 6.1. Figure 7 lists the recognition accuracies (true accept rates at a fixed false accept rates of 1.0%) for each of the different facial components used in this study. As discussed in Section 5, each of the 46 attributes was assigned to one of the seven cropped facial regions (eyes, brow, nose, mouth, jaw, hair, and the entire
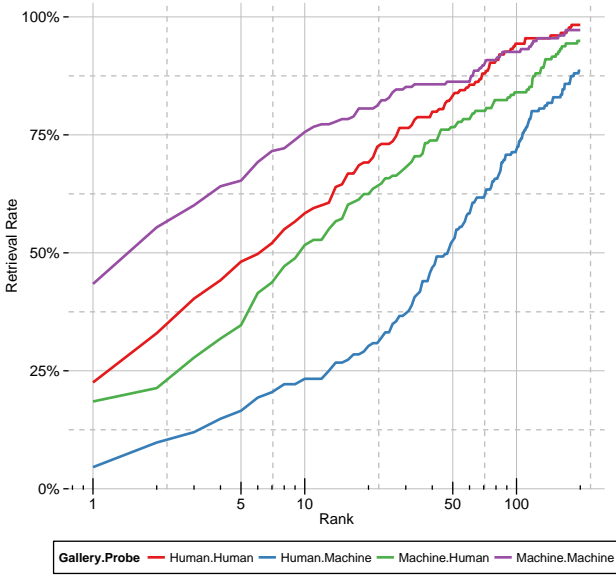
Figure 8. CMC accuracies when using human provided facial attributes and machine extracted attributes.

face). The results listed here are the fusion of all attributes for a given component.

Perhaps the most surprising result presented here is the relatively poor accuracy of the eye component compared to other interior facial features (nose, mouth, and brow). We believe this is due to the difficulty in characterizing eye variations into categorical features given the complex shape of the eyes. In fact, this finding also agrees with recent research in computer generated composite recognition, where the relative performance of the eye component was poor [5]. Because computer generated composites are equally limited by the discrete set of options, the difficulty in capturing the complex shape of the eyes would similarly seem to be a limiting factor.

6.3. Experiment 3: Human vs. Machine

In our next experiment we measure the accuracy of manually labelled attributes vs. automatically extracted attributes. To accomplish this experiment, we had AMT annotators label the 175 mated photographs in FERET with attribute information. Thus, for 175 subjects among the 1,196 total subjects in the CUFSF database, we have manually labelled attribute information from two separate images.

Two-fold cross validation was performed using the exact splits used in the previous experiments. Of the 175 subjects, only those subjects who were in the test partition were used to generate recognition accuracies. The first image from the other subjects was used in the training process. This resulted in 84 testing subjects in the first split with a gallery of 604 subjects, and 91 testing subjects in the second split

with a gallery of 508 subjects.

We compared three methods for matching attribute representations: (i) human labelled versus human labelled, (ii) human labelled versus automatically extracted, and (iii) automatically extracted versus automatically extracted. In the case of the automatically (or machine) extracted attributes, the extraction algorithm was trained on the training partition, which contained no subjects in the testing partition.

Figure 8 lists the CMC results comparing human vs. automatic extraction. It is quite notable that machine derived attributes achieve nearly twice the Rank-1 accuracy as human attributes when comparing the human vs. human to machine vs. machine. This speaks perhaps to the difficulty that humans face in assigning discrete values to the attributes, whereas the algorithm can assigned a numeric value based on the regression output.

When comparing human vs. machine performance, an interesting observation can be made: when using human/manual attributes in the gallery and machine/automated attributes as probes, the accuracy is significantly worse than using human attributes as probes and machine attributes as gallery (as shown in Figure 8). While the only practical case (fortunately) is the use of machine derived attributes in the gallery, the discrepancy between these results is quite informative. We believe this to be a manifestation of the discrete values provided by humans, versus the numeric values provided by machines. When the attribute values are numeric and continuous (as in the case of machine derived attributes), subjects are more naturally separated in the feature space.

7. Conclusions

We have proposed a method for suspect identification using facial attribute descriptions. This was achieved by developing a set of 46 categorical facial attributes. In operational scenarios, the attribute values from a probe image would be provided by human witnesses, or derived from low quality imagery. In turn, these attribute descriptions could be used to search mug shot galleries or surveillance videos for investigative leads. In order to search such large repositories with manually derived query information, faces in the repositories must have their attributes automatically extracted. As such, we have developed an algorithm to automatically extract attribute information from face images.

Experiments were conducted to compare the proposed attribute-based recognition paradigm to hand drawn sketch recognition, as each method seeks to perform identification using witness descriptions. The proposed method does not achieve the same accuracy as sketch recognition, nor was it expected to meet this upper bound. Instead, we demonstrate that the proposed method can achieve accuracies of roughly the same order as sketch recognition, and can also improve the accuracy of sketch recognition through fusion. Addi-

tional experiments demonstrate the strengths of automatically extracted features versus manually labelled features.

The results of our initial investigation into attribute-based suspect identification are very compelling, and have prompted us to broaden the investigation. We will explore the impact of face descriptions provided from human memory, to better understand the operational use cases of witness identification. The use of attributes from low quality imagery will also be explored, as commercial face recognition algorithms that fail on such low quality imagery. By contrast, attribute-based suspect identification is premised on the use of high recall matching, and thus should perform well in this scenario. Further, we will investigate the use of an incomplete set of attributes to replicate scenarios in which a witness is not able to provide an exhaustive description. Finally, we will explore the use of confidence-based matching, where users can provide a confidence in their attribute assignments.

Acknowledgements

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, December 2006.

[2] L. Best-Rowden, H. Han, C. Otto, B. Klare, and A. K. Jain. Unconstrained face recognition: Identifying a person of interest from a media collection. In *MSU Technical Report, MSU-CSE-14-1*, 2014.

[3] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Memetically optimized MCWLD for matching sketches with digital face images. *IEEE Trans. Information Forensics Security*, 7(5):1522–1535, October 2012.

[4] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[5] H. Han, B. Klare, K. Bonnen, and A. Jain. Matching composite sketches to face photos: A component based approach. *IEEE Trans. Information Forensics and Security*, 8(1):191–204, January 2013.

[6] A. K. Jain, P. Flynn, and A. A. Ross. *Handbook of Biometrics*. Springer, 2007.

[7] B. Klare and A. Jain. Heterogeneous face recognition using kernel prototype similarities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(6):1410–1422, June 2013.

[8] B. Klare and A. K. Jain. Sketch-to-photo matching: a feature-based approach. In *SPIE Defense, Security, and Sensing*, pages 766702–766702, 2010.

[9] B. Klare, Z. Li, and A. Jain. Matching forensic sketches to mug shot photos. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(3):639–646, March 2011.

[10] B. F. Klare, S. S. Bucak, A. K. Jain, and T. Akgul. Towards automated caricature recognition. In *International Conference on Biometrics*, pages 139–146. IEEE, 2012.

[11] J. Klontz, B. Klare, S. Klum, A. Jain, and M. Burge. Open source biometric recognition. In *Proc. IEEE Biometrics: Theory, Applications, and Systems*, 2013.

[12] S. Klum, H. Han, B. Klare, and A. K. Jain. The FaceSketchID system: Matching facial composites to mugshots. In *MSU Technical Report, MSU-CSE-14-6*, 2014.

[13] S. Klum, B. Klare, H. Han, and A. Jain. Sketch based face recognition: Forensic vs. composite sketches. In *International Conference on Biometrics*, 2013.

[14] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, 2011.

[15] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *Proc. IEEE CVPR*, pages 1005–1010, June 2005.

[16] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *Proc. ECCV*, 2008.

[17] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The feret evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[18] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Processing*, 19(6):1635 – 1650, June 2010.

[19] X. Tang and X. Wang. Face sketch recognition. *IEEE Trans. Circuits System for Video Technology*, 14(1):50 – 57, January 2004.

[20] T. Valentine and V. Bruce. The effects of distinctiveness in recognising and classifying faces. *Perception*, 15(5):525–535, 1986.

[21] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(11):1955–1967, November 2009.

[22] P. Yuen and C. Man. Human face image searching system using sketches. *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, 37(4):493–504, July 2007.

[23] W. Zhang, X. Wang, and X. Tang. Coupled information-theoretic encoding for face photo-sketch recognition. In *Proc. IEEE CVPR*, 2011.