

# Component-Based Representation in Automated Face Recognition

Kathryn Bonnen, Brendan Klare, *Student Member, IEEE* and Anil Jain, *Fellow, IEEE*

**Abstract**—This paper presents a framework for component-based face alignment and representation that demonstrates improvements in matching performance over the more common holistic approach to face alignment and representation. This work is motivated by recent evidence from the cognitive science community demonstrating the efficacy of component-based facial representations. The component-based framework presented in this paper consists of the following major steps: (i) landmark extraction using Active Shape Models (ASM), (ii) alignment and cropping of components using Procrustes Analysis, (iii) representation of components with Multi-Scale Local Binary Patterns (MLBP), (iv) per-component measurement of facial similarity, and (v) fusion of per-component similarities. We demonstrate on three public datasets and an operational dataset consisting of face images of 8,000 subjects, that the proposed component-based representation provides higher recognition accuracies over holistic-based representations. Additionally, we show that the proposed component-based representations: (i) are more robust to changes in facial pose, and (ii) improve recognition accuracy on occluded face images in forensic scenarios.

**Index Terms**—face recognition, component-based face representation, feature extraction, active shape model

## I. INTRODUCTION

Studies in cognitive science suggest that *both* global and local features are utilized for face perception and recognition [1], [2], [3], [4], [5], [6], [7]. However, despite these findings, research in automated face recognition has seen a disproportionate amount of attention spent developing systems that use holistic face descriptors to represent a face image. For example, holistic face representations such as densely sampled feature descriptors (e.g. Local Binary Patterns (LBP) [8], Scale Invariant Feature Transforms (SIFT) descriptors [9], biologically inspired features (BIF) [10]) and appearance-based (or pixel-based) representations [11], [12], [13] are commonly used in automated face recognition [14]. However, studies involving face recognition with component-based representations (i.e., representations that extract features per specific facial components<sup>1</sup>) are found at a much lower frequency in the literature, despite their demonstrated use in human face processing (see 1 for illustration of methods). Even methods which compute similarity measures at specific facial

landmarks, such as elastic bunch graph matching (EBGM) [15], do not operate in a per-component manner.

The discourse in cognitive science and psychology continues to argue about the precise roles of holistic, configural and component information in human face perception. However, evidence points to the presence of all three in human face perception [16], [17], [18], [2], [1], [19], [4], [3], [5], [6], [20], [7]. The earliest paradigms providing support for component processing created faces comprised of sets of components and then asked subjects to make decisions about whether faces were the “same” or “different”. Experimental results showed that the number of components by which the two images differed predicted subjects’ reaction times [1], [16], [2]. These results suggest a part-by-part processing of the face. Smith and Nielsen confirmed the presence of component processing using a slightly different matching paradigm comprised of schematic line drawings of faces presented with a delay of one or four seconds [17]. More recently, Schwaninger, et al. (2007) concluded that humans exhibit the capability to recognize faces based on information drawn from isolated features (i.e. components) [3]. In fact, Gold, et al. [21] present evidence supporting the idea that face processing is the result of the integration of individual component processing. They compare human component integration with an optimal Bayesian integrator that is based on component recognition performance. The comparison demonstrates that humans perform no better and perhaps slightly worse than the prediction of the Bayesian framework [21]. This means that for the purposes of human face recognition, holistic processing provides no improvement over an integration of the individual components of the face. Additional behavioral studies of humans have shown that certain facial components (i.e. the face outline, eyes and mouth) are more useful for perceiving faces than other components and, similarly, the upper face is more useful than the lower face [18], [22]. Moreover, humans have shown an increased capacity to determine an identity through a caricature sketch<sup>2</sup> than a true portrait of a face, which also suggests the importance of certain components of the face over others [23], [24]. Finally, investigations into the neural and physiological mechanisms of face processing suggest a system comprised of processes that include both holistic and face part (component) processing [7], [5].

As demonstrated by the literature in cognitive science, there is a high likelihood of the presence of component processing in human face perception. Yet, despite this strong evidence,

Kathryn Bonnen is with the Institute for Neuroscience, University of Texas, Austin, TX 78712. Brendan Klare and Anil K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA. Anil K. Jain is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea. E-mail: bonnenka, klarebre, jain@msu.edu

<sup>1</sup>Facial components are specific regions of the face, such as the nose, eyes, eyebrows, and mouth.

<sup>2</sup>Caricatures are drawings in which certain facial components and attributes are exaggerated beyond a realistic appearance.

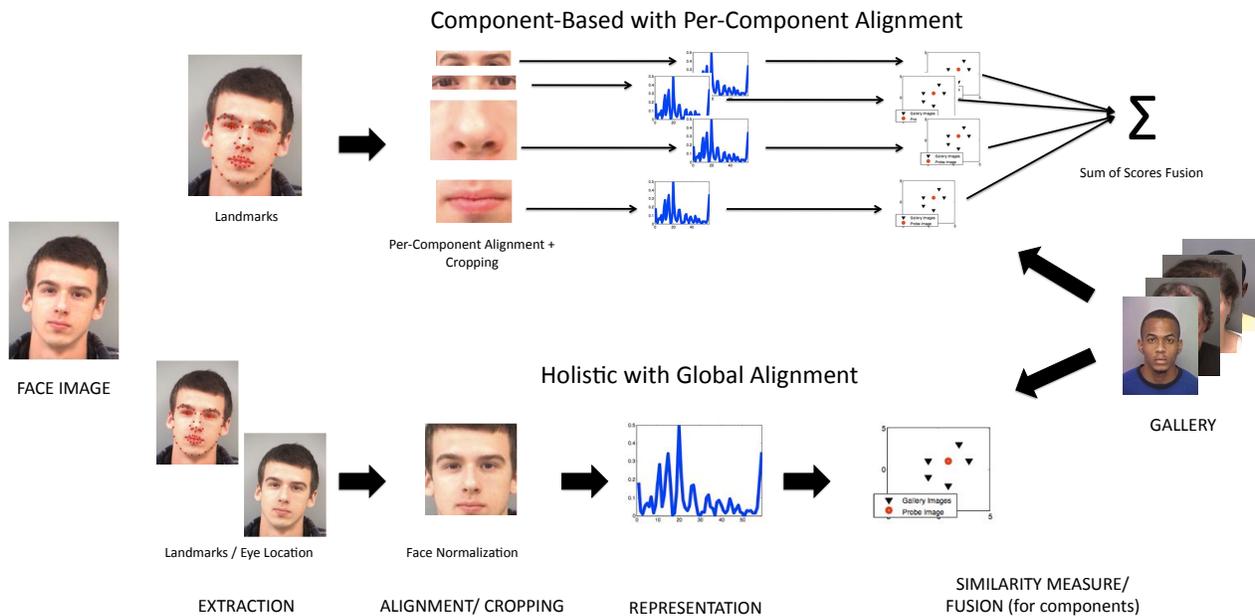


Fig. 1. Outline of the per-component alignment performed to yield the proposed component-based representations. This work demonstrates the value of representing faces in a per-component manner. When compared to a globally aligned holistic representation, and other representations found in the literature, the component-based representation offers strong accuracy improvements in a number of face recognition scenarios.

computer vision and biometrics literature in face recognition has focused heavily on holistic processing in which representations are derived from globally aligned faces (we call these holistic representations). Furthermore, the handful of studies in automated face recognition that have approached face recognition from the standpoint of component-based processing are generally limited in their approach [25], [26], [15], [27], [28], [29], [30], [31] (see Table I for a summary of a selection of these methods). For example, many of these methods use raw pixel-based representations, which have proven to be less robust feature representations [14]. Moreover, many of the previous methods roughly extract their facial components by cropping predefined regions of globally aligned face images, instead of using more precise methods that rely on facial landmark locations. Another similar line of research involves block-based representations, which break the face into blocks and create individual classifiers for each block and finally fuse the individual classifiers [32], [33]. While this paradigm is similar to a component-based representation, in this case the faces are again globally aligned and the blocks fail to take advantage of the concentration of information in components due to their arbitrary nature. Despite few contemporary component-based approaches to face recognition, a number of recent studies have addressed periocular recognition<sup>3</sup>, and have shown considerable success in identifying a person by the periocular region of the face alone [34], [35], [36]. These results bode well for modern approaches to component processing of the face.

Based on (i) the strong evidence of component processing in human face perception, and (ii) the lack of mature component-based methods in automated face recognition research, a more

<sup>3</sup>Periocular recognition involves the identification of persons based on the eyes and the local region of the face surrounding the eyes.

thorough investigation of the role of component-based processing in automated face recognition is warranted. In the same way that the field of cognitive science continues to investigate the precise roles of component and holistic processing in human face perception, automated face recognition algorithms also need to explore the role that component processing could have in leading to improved face recognition algorithms.

This work seeks to expand on the previous methods in component-based automated face recognition. Using facial components that are precisely extracted through automatically detected facial landmarks, we demonstrate that descriptors computed from these individually aligned components result in higher recognition accuracies than descriptors extracted using the more common approach of dense sampling from globally aligned faces. Both of these approaches to face recognition are detailed and compared in Figure 1. In addition to providing a component-based framework, we also demonstrate that: (i) component-based representations are relatively robust to changes in orientation, (ii) such representations are also useful in the presence of known occlusions, and (iii) the recognition accuracies have the potential to be further enhanced by learning algorithms.

This work is a study on feature representations, and not a study on learning algorithms. While we do demonstrate in Section III that the proposed component-based representations can be improved by performing statistical feature extractions, the component-based representations are not tied to any specific learning algorithm. Because the component-based representations are able to consistently improve recognition accuracy when compared to the commonly employed holistic and block-based feature representations, any learning-based method should benefit from the proposed component-based representations.

The remainder of this paper is organized as follows. In Section II we detail our approach for aligning and extracting features from different facial components. In Section III we discuss an approach to learning per-component feature extractions. In Section IV the results of experiments using component-based representations of frontal view face images are discussed. Results from experiments on face images with orientation changes and occlusions are also detailed and discussed. Finally, in Section V, the results are summarized and some practical applications of the proposed representation are discussed.

## II. COMPONENT EXTRACTION AND REPRESENTATION

In order to align, crop and extract a feature vector from the facial components, the following three steps are performed:

- 1) Extraction of facial landmarks
- 2) Per-component alignment (and cropping) using Procrustes Analysis
- 3) Representation of components using Local Binary Patterns

This section will provide details regarding how each of these steps was performed.

### A. Landmark Detection

The first step in aligning the facial components is to extract a predefined set of 76 anthropometric landmarks<sup>4</sup>. For each of the facial components (i.e., the eyebrows, eyes, nose, and mouth), a subset of these anthropometric landmarks provides a general outline of the component (see Figure 2). In this section, we detail a method for automatic extraction of landmarks.

Given the variability in facial appearances, as well as the variability caused by pose and expression changes, the extraction of facial landmarks is often a difficult task. The use of Active Shape Models (ASM) [38] is a common approach for determining the location of facial landmarks. Due to the structural constraints afforded by the face, ASM model-based detection is able to handle minor variations in pose and expression. However, ASMs are sensitive to the initial placement of landmarks prior to the iterative updating of model parameters. If this initial placement is not closely aligned to the true landmark locations, then the ASM may converge on an inaccurate set of landmarks. To help mitigate this problem, a small subset of more stable landmarks (i.e., the center of the two eyes and the tip of the nose) can first be detected before applying the ASM.

Using PittPatt’s Face Recognition SDK [39], we first automatically detected the center of the two eyes, and the center of the nose. Because these three landmarks are also present in the ASM, we initialized the ASM landmarks by (i) solving the affine transformation from these three ASM points to the corresponding PittPatt detected points, and (ii) applying this transformation to the set of 76 ASM landmarks (representing the mean face in the model). The result of this step is an initial placement of facial landmarks that is well suited to correctly

converge on the proper locations. Indeed, we found this step (i) greatly improves the accuracy of landmark detection, and (ii) eliminates all failure to converge cases. Using this approach, the only cases where landmark extraction failed were when PittPatt was not able to detect the eyes and nose (this occurs at a frequency of roughly 1 in 1,000 faces for the neutral frontal view).

The ASM implementation used was provided through the Stasm open source software [40].

### B. Alignment and Cropping

Once the facial landmarks have been detected, each facial components can then be aligned. Component alignment was performed using Procrustes analysis [41], which finds the rigid transformation that minimizes the mean squared error between two ordered sets of coordinates. This step eliminates variations in translation, scale, and rotation, which allows for a more accurate similarity measure between facial components than previous techniques which used pre-defined cropping boundaries.

For a given component (e.g., the mouth), let  $P_i = (p_1^x, p_2^x, \dots, p_n^x, p_1^y, p_2^y, \dots, p_n^y)^T$  be the set of  $n$  landmark points for the  $i$ -th image. For a given image, let  $p_j^x, p_j^y$  be the  $x$  and  $y$ -coordinates of the  $j$ -th landmark point. Further, let  $p_j = (p_j^x, p_j^y)$ . Procrustes analysis (or alignment) is performed by:

- Removing the translational component for each image by subtracting the mean of the landmarks (i.e.,  $p_i \leftarrow p_i - \frac{1}{n} \sum_{i=1}^n p_i$ ).
- Normalizing the scale for each image by dividing the concatenated vector of points  $P_i$  by its  $L_2$  norm (i.e.,  $P_i \leftarrow P_i / \|P_i\|_2$ ).
- Removing rotational differences by using least squares minimization to solve for the rotation matrix based on the angle  $\theta_i$  that minimizes the difference between  $P_i$  and a reference set of points  $P$  ( $P$  can simply be the first image).

After performing Procrustes analysis on each component in each face image, we obtain rotation, translation and scaling parameters. Using these parameters, the facial components can be rigidly aligned.

Cropping is accomplished by creating a bounding box around the aligned landmarks. Consider the face on an  $xy$ -plane. The bounding box is created by first finding the horizontal cropping boundaries from the minimum and maximum  $x$  values. The vertical cropping boundaries are determined based on a ratio of the crop width (in order to maintain the images aspect ratio). A small pixel border around each set of landmarks is used to improve the subsequent descriptor extraction.

Upon completion of these steps, the components are scaled to the following sizes (and ratios):

- Eyebrows - 29x142
- Eyes - 29x156
- Nose - 114x161
- Mouth - 82x142

<sup>4</sup>Facial anthropometry is the science around the measurement of the human face.

TABLE I  
A SUMMARY OF THE PREVIOUS STUDIES ON COMPONENT-BASED AUTOMATED FACE RECOGNITION.

Description	Component Extraction Method	Components Used	Component Alignment	Component Representation	Face Database
Template Matching [25]	Rigid*	Eyes (including eyebrows), nose, mouth	none	Pixel representation, locally normalized pixel representation, gradient, and Laplacian	Private database (47 subjects, 188 images)
Elastic Bunch Graphing Matching (EBGM) [15]	EBGM (object-adapted grid)	N/A	EBGM (object-adapted grid)	Gabor wavelet coefficients	FERET (250 subjects, 1500 images); Bochum (108 subjects, 650 images)
Component-Based LDA Method with Component Bunches [27]	Rigid*	L. eye, R. eye, nose, R. mouth, L. mouth	none	Pixel representation	FERET (70 subjects per experiment, 140 images in experiment 1, 420 images in experiment 2)
Component-Based Cascade LDA [28]	Rigid*	4 overlapped regions (from the whole face)	none	Pixel representation	FERET (1,196 subjects, 2197 images)
Component-Based LDA Face Description [31]	Rigid*	Set 1: 14 components across whole face; Set 2: 5 components (eyebrows- forehead, L. eye- eyebrow, R. eye- eyebrow, L. nose- mouth, R. nose- mouth)	none	Pixel representation	Extended Version 1 MPEG-7 (635 subjects, 3175 images); Altkom (80 subjects, 1200 images); XM2VTS (MPEG-7) (295 subjects, 2950 images); FERET (875 subjects, 4000 images); Banca (MPEG-7) (52 subjects, 520 images)
Framework for High Resolution Face Recognition [29]	Rigid*	L. eye, R. eye, nose, mouth, forehead skin, cheek skin, irregularities	none	Pixel representation	Private database (200 subjects, 1600 images); XM2VTS (295 subjects, 1180 images)
Part-Based Face Recognition [30]	Rigid*	10 components (focused on regions with eyebrows, eyes, nose, mouth)	none	Local Binary Patterns	Near Infrared (1000 subjects, 10,000 images)
Component-Based Face Identification using 3D models [26]	Reference points on 3D head models	14 learned components	3D Morphable Models	Histogram equalized gray values	3D face database (100 subjects, 6,843 images) [37]; Private database (10 subjects, 30 images)

\* “Rigid” denotes the component extraction using a rigid alignment of the eyes (i.e., planar rotation and scaling), followed by component cropping using predefined cropping boundaries

These sizes were determined by tuning the size of the pixel border and horizontal cropping boundary per-component.

The same method for per-component Procrustes alignment can also be applied to the entire set of facial landmarks. Thus, the entire set of facial landmarks is used for a single, global alignment of the face. Referred to as holistic Procrustes alignment, our experiments will compare this more common alignment technique to the proposed component-based alignment. The mean components from both holistic (or global) Procrustes alignment and the proposed per-component alignment are shown in Figure 3. Qualitatively, we see the per-component alignment results in a more precise alignment of each facial component (see Figure 3 (c)). In turn, this approach will allow us to better understand the value of component-

based representations. For example, we will demonstrate in Section IV that despite the common belief that the eyes are the most accurate facial component in automated face recognition, the mouth component offers nearly the same recognition accuracy when properly aligned.

We originally applied the technique of aligning facial components to study identical twins [42] and facial carvings [43]. However, these earlier studies were focused on determining the distinctiveness of each component with respect to twins and carvings. They did not consider whether or not this technique resulted in improved recognition accuracy over holistic representations. The subsequent experiments in the paper will demonstrate that per-component face representations do indeed offer a significant improvement in recognition accuracy

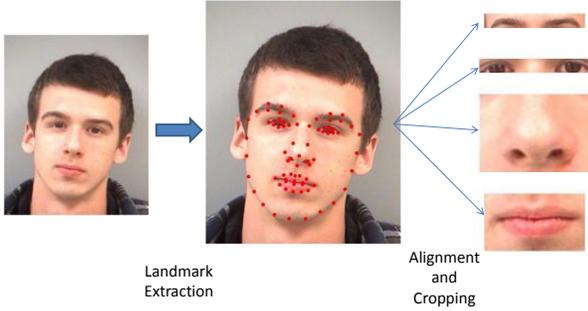


Fig. 2. The process for extracting facial components. Using automatically detected landmarks, each facial component is individually aligned using Procrustes analysis. In this paper we demonstrate the value of using this per-component alignment instead of the more common approach of globally aligning the entire face.

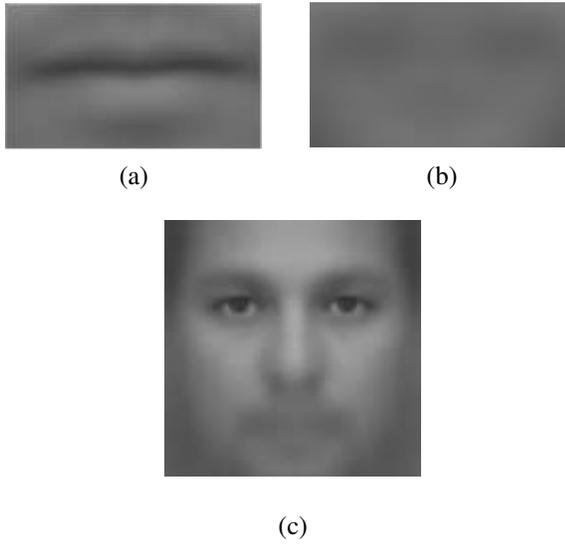


Fig. 3. Comparison of the mean mouth component image from (a) performing the proposed per-component alignment method, and (b) cropping using globally aligned faces. The poor alignment of certain components using globally aligned faces is observed from the mean globally aligned face (c). While most previous component-based methods extract facial components using globally aligned faces, this work aligns each facial component individually resulting in better representations, as shown in (a).

over the more commonly applied holistic and block-based representations.

### C. Representation

Once each facial component has been aligned, we extract a Multi-Scale Local Binary Patterns (MLBP) feature representation from each component. MLBP is the combination of Local Binary Patterns (LBP) [8] descriptors with different radii, and it is generally more effective for face recognition than LBP alone. MLBP and LBP descriptors have been successfully used in holistic-based approaches to automated face recognition by representing regions of the face with a descriptor that encodes the facial structure and shape [44].

Each facial component is divided into regions of  $d \times d$  pixels overlapping by  $m$  pixels ( $m < d$ ). Within each region, a histogram of LBP values is derived from comparisons at each pixel. The LBP value,  $V$ , calculated at each pixel is

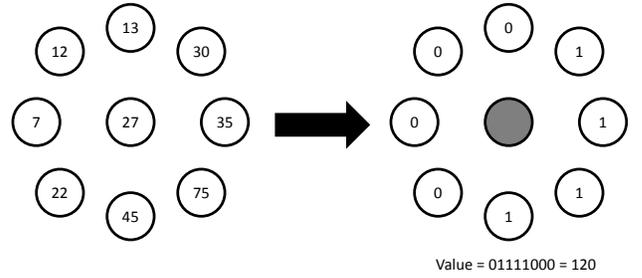


Fig. 4. Computing LBP value at each pixel

TABLE II  
PARAMETERS USED IN MULTI-SCALE LOCAL BINARY PATTERNS REPRESENTATION

	Eyebrows	Eyes	Nose	Mouth
Region Size	8x8	8x8	16x16	16x16
Amount Overlap	4	4	8	8
# Sample Points	8	8	8	4

computed using comparisons to  $P$  neighboring sample points at a radius of length  $R$ :  $V_{PR} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$  where  $g_p$  represents the gray value at each of the  $p$  surrounding pixels,  $g_c$  represents the gray value at the center, and  $s(x) = 1$  if  $x \geq 0$  and  $0$ , otherwise (see Figure 4). This creates a histogram of dimensionality  $2^P$ , though this can be further reduced by mapping LBP values without “uniform patterns” to the same value. A uniform pattern is an LBP binary string which produces 2 or fewer bitwise transitions [44]. The MLBP representation concatenates two or more LBP descriptors of an image patch at different radius lengths.

The choice of parameters (i.e. region size,  $d$ ; amount of overlap,  $m$ ; number of sampling points,  $P$ ; radius,  $R$ ; and the use of uniform binary patterns) were tuned per-component. All components use the radii combination,  $r = 1$  and  $r = 3$  and employ uniform binary patterns. Table II provides details of the remaining parameters used for each component.

### III. COMPONENT-BASED DISCRIMINANT ANALYSIS

Linear Discriminant Analysis (LDA) was first used in face recognition on pixel representations of faces by Belhumeur, et al. [45], and has become a common feature extraction/reduction technique in face recognition. The goal of LDA is to maximize the between subjects variance,  $S_b$ , and minimize the within subjects variance,  $S_w$ , by solving the generalized eigenvalue problem:  $S_b \cdot \Psi = \Lambda \cdot S_w \cdot \Psi$ . However, as facial representations have grown increasingly more complex, improved LDA methods have been developed to address the problem of high feature dimensionality in comparison to the size of the training set (e.g. random sampling LDA (RS-LDA) [46], direct LDA (D-LDA) [47] and regularized LDA (R-LDA) [48]). By partitioning the face into components, the high dimensionality problem is alleviated to some extent, however RS-LDA is still employed to further address these concerns.

The RS-LDA approach requires the following steps for training. First, the feature space is randomly sampled into  $k$

TABLE III  
PARAMETERS USED IN RANDOM SAMPLING LDA

	Eyebrows	Eyes	Nose	Mouth
% Features Sampled	0.5	0.5	0.5	0.5
# of Subjects	250	500	250	250
% Variance Retained	.98	.95	.95	.95
# Subspaces	25	7	25	7

subspaces, with each subspace sampling a fraction  $s$  ( $0 < s < 1$ ) of the available features. For each of the  $k$  random sample spaces, principal components analysis is performed in order to retain  $e$  percent of the variance. Finally,  $k$  LDA subspaces are learned from each of the PCA representations. Using these trained subspaces, images can then be sampled into each of the  $k$  random feature subspaces, projected into the corresponding PCA and LDA subspaces; each of the  $k$  subspace vectors can then be concatenated into a final feature vector. In addition to feature sampling, RS-LDA also performs bagging on the training subjects. That is, for each set of the  $k$  random samples,  $n_s$  subjects from the training set are randomly sampled from the  $n$  total subjects available ( $n_s < n$ ).

As discussed, RS-LDA has the following set of parameters: the percentage of the original set of features to be sampled during each stage ( $s$ ), the number of subjects to use during each stage ( $n_s$ ), the percentage of variance retained in each stage ( $e$ ), and the number of subspaces ( $k$ ). As before, these parameters are tuned per component. See Table III for the parameter values used in this work.

Our use of RS-LDA in conjunction with the component-based representation is to demonstrate the ability to improve this representation with statistical learning. The component-based representation is not tied to this particular discriminative subspace technique, and instead the component-based representation can conceivably be improved with any learning-based technique.

#### IV. EXPERIMENTAL RESULTS

The following experiments were designed in order to explore the effectiveness of component-based representation in face recognition. The primary baseline for these experiments is a holistic face representation. As discussed, the term “holistic representation” in this paper refers to densely sampled MLBP feature descriptors from globally aligned face images, which is one of the most common approaches in face recognition research.

In this work, face images are aligned in two different manners. The first alignment method is the common approach of alignment using the centers of the two eyes. The face is geometrically normalized using the eye locations to (i) perform planar rotation so the angle between the eyes is 0 degrees, (ii) scale the face so the inter-pupillary distance (IPD) between eyes is 48 pixels, and (iii) crop the face to 128x128 pixels. The second alignment method performs Procrustes alignment on the full set of facial landmarks using the same procedure that is performed on each facial component (see Section II).

Again the Procrustes-based method scales the face in order to achieve an IPD of 48 pixels.

Once each face is geometrically normalized using these two global alignment methods, MLBP feature descriptors are densely sampled from patches of size 12x12 across the face, with an overlap of 6 pixels. The parameters used for both the alignment and the MLBP descriptors have been optimized in a number of previous studies by our research group. Our prior research has relied on such holistic representations, and these parameters represent our best practices [49]. It is also important to note that we are not able to improve the accuracy of this holistic representation by increasing the size (or IPD) of the face image.

A block-based representation was also used as a baseline as it has appeared in recent research in face recognition [32], [33]. In this representation the face is aligned using Procrustes alignment on the full set of facial landmarks and then broken into 3x3 blocks. Each individual block can then be treated as a component to complete the rest of the representation and recognition processes.

##### A. Component Representations

This section contains experiments conducted to investigate the advantages of component-based representations, as opposed to the popular holistic representations. The experiments intend to isolate the recognition accuracy up to the feature representation stage, before any additional statistical learning (e.g. RS-LDA) is conducted. This deliberate decision to isolate the feature representation stage of face recognition is consistent with the primary goals and focus of this paper, to provide novel and improved feature representation techniques. These experiments were conducted on a large scale database of frontal face images from the Pinellas County Sheriff’s Office (PCSO), and unconstrained face images from the Labeled Faces in the Wild (LFW) [50] dataset.

1) *Component Representations - PCSO Database:* The PCSO dataset utilized for this experiment is comprised of pairs of frontal view images of nearly 16,000 subjects, each with one probe image and one gallery image (resulting in a total of 32,000 images). These images were randomly selected from a larger image dataset of 1.5 million operational face images from the PCSO. For our purposes, the dataset was partitioned into two non-overlapping subsets: (i) a training set of 8,000 subjects, and (ii) a test set of 8,000 subjects. The training set is not used in experiments that did not require training. For each set, roughly 10 subjects were removed because their probe or gallery image failed to enroll with the PittPatt SDK eye finder.

The component representations were created using the method outlined in Section II and their similarity was measured using the cosine similarity measure. Sum fusion was used to combine the component similarities after z-score normalization. Figure 5 displays the performance of the raw component representations. The True Accept Rate (TAR) at a fixed False Accept Rate (FAR) of 1% for each individual component is as follows: Eyebrows - 67.72%, Eyes - 66.63%, Nose - 65.16%, Mouth - 64.83%. The TAR of the fused result is 83.19%.

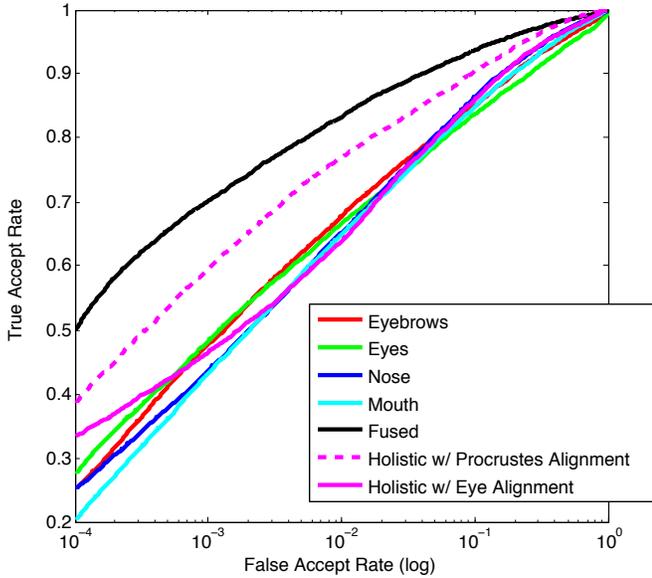


Fig. 5. Component-Based Performance (ROC)

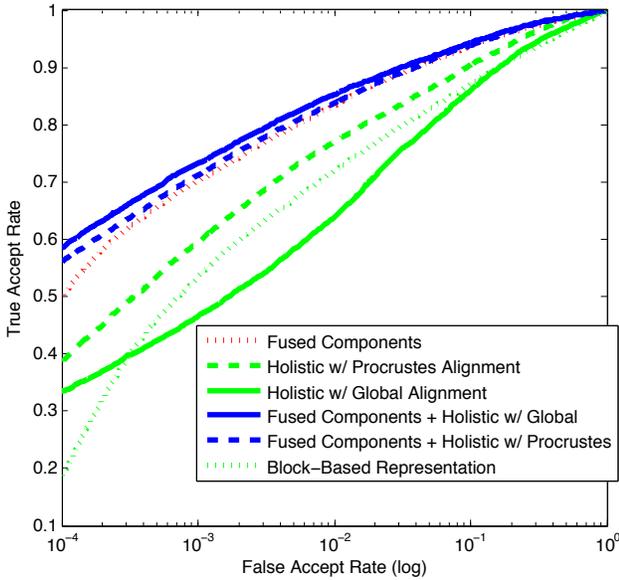


Fig. 6. Component-Based and Holistic Performance (ROC)

Figure 6 compares the accuracy of the proposed fused component representation to the more commonly employed holistic representation. If the face images are globally aligned using only the eye locations, then the holistic, densely sampled MLBP features result in a TAR (at FAR=1.0%) of 63.78%. If the face images are globally aligned using all the facial landmarks<sup>5</sup>, then the TAR of the holistic representation improves to 76.88%. The block-based representation results in a TAR of 71.90%. In all these baseline cases the recognition accuracies are significantly lower than the 83.19% accuracy achieved using the proposed per-component representation.

<sup>5</sup>The same points used for individual components are combined in global Procrustes alignment to create the set of points used to align the entire face.

2) *Component Representations - LFW Database*: The LFW dataset utilized for this experiment is comprised of 6,000 image pairs (12,000 images). The dataset has been partitioned into 10 folds as per the standard LFW protocol described in [50]. While this protocol makes provisions for training and testing sets, only the testing sets are used since there is no training required for this experiment. The feature representation methods are applied on each individual fold and their results are averaged to produce the results listed in Table IV.

The component representations were generated using the same procedure described in Section IV-A1. In addition to the holistic methods previously outlined (global eye alignment and Procrustes alignment), the LFW Database provides images aligned via commercial alignment and funneling. The holistic representation method was applied to this prealigned images to compare our method to other methods of alignment. Furthermore, the block-based representation method described in IV-A1 is also employed to provide an additional comparison. Roughly 15 subjects per fold of 600 subjects were removed because their probe or gallery image failed to enroll.

The results of this experiment are listed in Table IV (TAR at FAR = 10%). The performance on the LFW database is significantly lower than that of PCSO, which is due to the unconstrained nature of the LFW database. However, the performance of the fused components from the component-based method (31.39% TAR at FAR=10%) is observed to be higher than the performance of alternate holistic and block-based methods. This is consistent with the findings on the PCSO database.

The holistic representation, which is aligned using the two eye locations, has a higher accuracy on the LFW dataset than the Procrustes aligned holistic representation. This differs from the PCSO dataset, where the Procrustes aligned holistic method had a markedly higher accuracy than the eye-aligned holistic. These results suggests that the unconstrained nature of the LFW database introduces errors into the automated landmark extraction process. Thus, while the component-based representation still achieves the highest accuracy when compared to the baseline representations, improving automated landmark extraction should further improve the effectiveness of the component-based representation on unconstrained face images.

3) *Discussion*: The superior performance of the fused components over the baseline representations is likely due to several factors. One of these factors is that the local alignment of each facial component can more accurately measure the similarity of the most identifiable regions of a face. Similarly, the focus on these more identifiable areas of the face (i.e. the eyebrows, eyes, nose and mouth), highlights the facial regions with the highest degree of inter-person variability. The portions of the face that are left out, such as the cheek and forehead areas contain considerably less discriminative information. Another facet to consider is that the parameters of MLBP descriptors can be tuned to each individual component, allowing for a better description of each individual component and thus of the face overall.

While the component-based representation resulted in improved recognition accuracies over the globally aligned holis-

TABLE IV  
COMPONENT-BASED AND HOLISTIC RESULTS FOR LFW DATABASE  
(FALSE ACCEPT RATE AT 10%)

Region	TAR at FAR = 10%
Eyebrows	29.55%
Eyes	26.31%
Nose	24.37%
Mouth	23.59%
Fused	30.93%
Holistic	29.20 %
Holistic + Procrustes Alignment	28.94%
Holistic + Commercial Alignment	21.01%
Holistic + Funneling Alignment	21.22%
Block-Based	27.78%

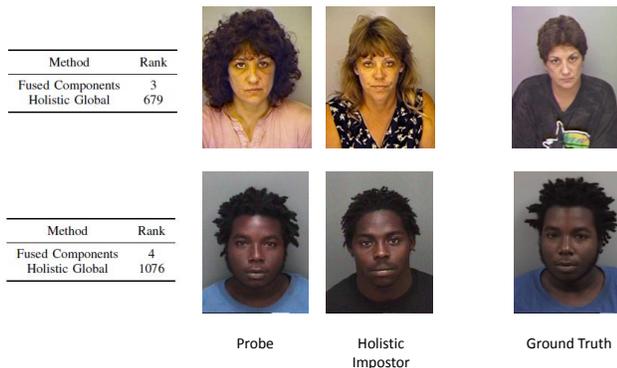


Fig. 7. Examples of face images from the PCSO database in which good performance was obtained by the component-based method proposed in this paper and poor performance was obtained by a globally aligned holistic method.

tic representation, as shown in Figure 6, the fusion of these two representations yielded the best accuracy (a TAR of 85.24% at FAR=1.0%). Thus, there is complementary information provided by component-based and holistic facial representations.

Figure 7 displays face images in which the component-based method performed well while the holistic method performed poorly. The holistic impostors provide a relatively good overall face match, but, especially in the case of the woman, the overall face shape and hair appears to have had a profound effect on the match. Similarly, Figure 8 displays face images in which the holistic method performed well while the component-based method performed poorly. Comparing individual components, there does appear to be a similarity, especially in the bottom example in which the impostor is a woman even though the subject is a man.

An interesting result in these experiments is the accuracy of the eyebrow and eye components. While they outperform the nose and mouth, they only do marginally better. However, in cognitive science and automated face recognition literature, the eyes and/or eyebrows have generally been regarded as the most useful component for recognition, while the nose and mouth are regarded as being less informative [18], [22], [29], [49].

A few considerations may explain our findings that the

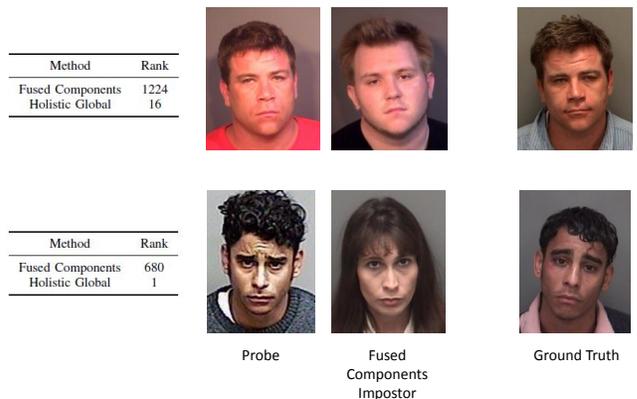


Fig. 8. Examples of face images from the PCSO database in which good performance was obtained by the globally aligned holistic method and poor performance was obtained by the component-based method proposed in this paper.

mouth and nose provide a similar level of discriminative information as the eyes and eyebrows. The first is a significant improvement in the alignment and cropping using ASM and Procrustes analysis, particularly for the nose and mouth. In previous works [25], [27], [28], [31], [29], [30], [26], the alignment was performed globally based on eye locations instead of locally per component (see Figure 3 for the difference between these two approaches). Due to the eye-based alignment, the previous works reported significantly lower performance for the nose and mouth components than what was achieved with per-component alignment shown here. With the local alignment method, components that are farther from the eyes can be more accurately aligned.

A second contributing factor to the relatively even performance of the eye and eyebrow regions and the nose and mouth regions has to do with the cropping boundaries of the components. Consider the eyes and eyebrows components extracted by our method in Figure 2. In comparison to previous component-based methods (e.g. [31], and Figure 9), our components are smaller (i.e. the “eyes” component does not include the eyebrows and the “eyebrows” component is tightly cropped such that it does not include the forehead or the eyes). The performance gap between the tightly cropped eyes and the eyes containing eyebrow regions is shown in Figure 10, demonstrating how the superiority of the “eye” region can be restored by cropping in a similar fashion as previous studies (see Figure 11 for example of larger cropping region). However, the goal of this paper is to explore the use of components. As components, the eyes and eyebrows do nearly as well as the larger eye region but contribute a greater weight to the overall component framework as two components than they would as one.

## B. RS-LDA

The following experiments were conducted to demonstrate that the component-based approach to representing faces can be further enhanced through learning algorithms. Both experiments measure the component-based method using RSLDA discriminative subspaces to boost performance. The first exper-

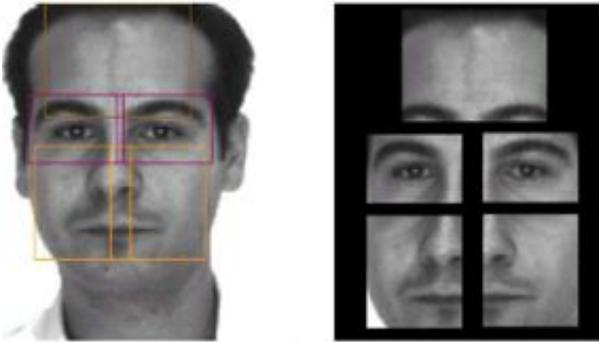


Fig. 9. An example of the global eye cropping from [31].

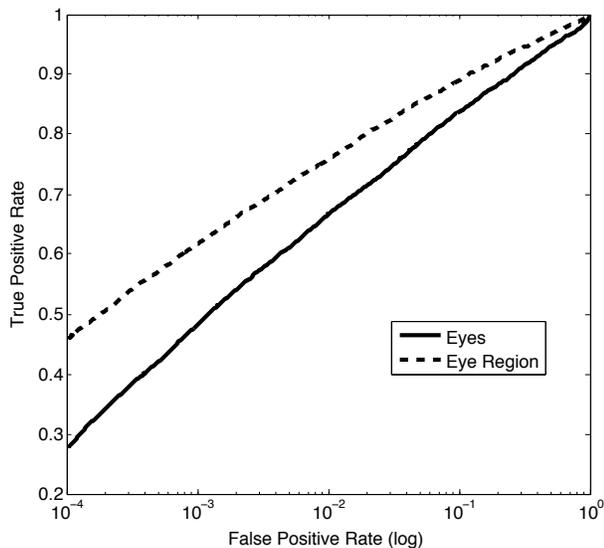


Fig. 10. Comparison of the recognition accuracy for differently sized eye components. The “Eye Region” refers to the eye and eyebrow area shown in Figure 11 while the “Eyes” refers to a component consisting of solely the tightly cropped eyes shown in Figure 2.

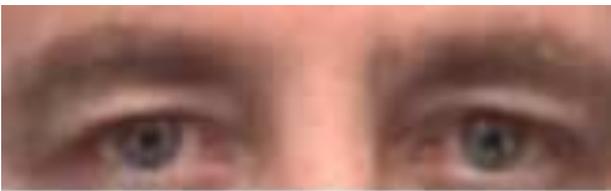


Fig. 11. Large cropping boundaries around the eyes is one factor for previous studies concluding that the eye component is more informative than the nose and mouth. Unlike the eye cropping shown here, if the eyebrows are removed from the cropping of the eyes then the recognition accuracy is generally similar to nose and mouth components which have been locally aligned.

TABLE V  
COMPONENT-BASED PERFORMANCE AFTER RS-LDA TRAINING (FALSE ACCEPT RATE AT 1.0%).

Region	FAR = 1%
Eyebrows	79.36%
Eyes	83.20%
Nose	84.31%
Mouth	74.17%
Fused	94.84%
Holistic	90.43%

iment uses the PCSO database and the second leverages the FERET database in order to make comparisons to previous results for other local feature-based methods.

1) *PCSO*: The PCSO dataset described in Section IV-A1 was utilized for this experiment. RS-LDA subspaces were learned using the training set (set 1), and then applied on the testing set (set 2) as outlined in Section III. The similarity between the extracted component features was measured using the cosine similarity. The components were combined using sum fusion of the component similarities after z-score normalization. Figure 12 displays the performance of each individual component, the fused component performance, and the holistic performance. The TAR at FAR = 1.0% was: Eyebrows - 79.36%, Eyes - 83.20%, Nose - 84.31%, Mouth - 74.17%, Fused - 94.84% (see Table V for a summary). This is compared to the holistic performance of 90.43% using the RS-LDA algorithm on the holistic representation in the same manner detailed in [49]. The accuracy of 94.84% is significantly higher than 83.19% when the component-based representation did not go through RS-LDA training. As discussed previously, while RS-LDA was utilized for this experiment, the learning method was chosen simply to demonstrate the viability of applying learning algorithms to the feature representation proposed by this work. This work leaves open the problem of studying the most appropriate learning algorithms for component representations. The authors believe that effective learning methods for component-based descriptors (such as R-LDA [48], RS-LDA [46], and ERE [51]) should be explored in future work.

2) *FERET*: The FERET dataset [52] is utilized as described in Section 5.1.1 of [53]. The dataset is comprised of 1,194 persons (2,388 images, 2 images Fa/Fb per person). There are two experiments based on this dataset: the first (FERET 1a) divides the dataset into a training set of 250 subjects and a testing set of 944 subjects; the second (FERET 1b) divides the dataset into a training set of 497 subjects and a testing set of 697 subjects. The component-based method is applied as described in section IV-B1. Table VI reports the rank 1 accuracy of each individual component, the fused component performance and the performance of other methods reported in [53] (rank 1 accuracy is used here as opposed to ROC curves in order to compare directly with the methods tested in [53]). For both the smaller and larger training sets the component-based method outperforms other methods, including other local feature-based methods for both the smaller and larger training sets.

TABLE VI  
COMPONENT-BASED PERFORMANCE (RANK-1 ACCURACY) AFTER RS-LDA TRAINING ON FERET DATABASE.

Experiment	Eyebrows	Eyes	Nose	Mouth	Fused	PCA*	LDA*	ERE*	SIFT*	PFD-SIFTM*	PFDM*
FERET 1a	99.68%	85.38%	99.89%	77.86%	100%	83.16%	89.72%	94.81%	93.33%	97.67%	97.88%
FERET 1b	100%	85.94%	99.85%	79.20%	100%	85.80%	96.41%	97.13%	94.41%	98.42%	98.71%

\*Results reported in [53]

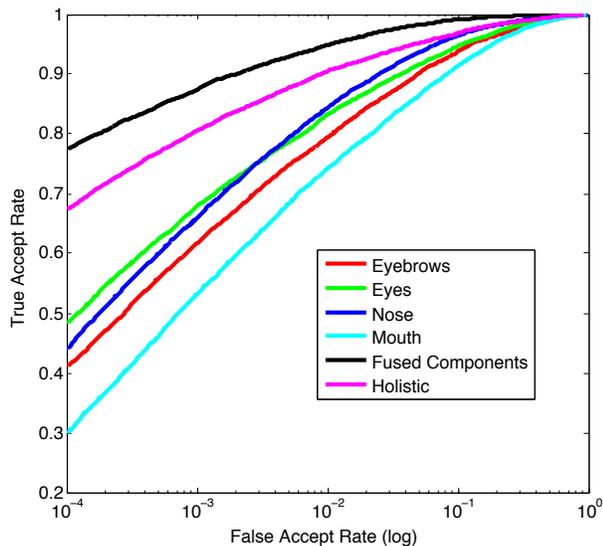


Fig. 12. Component-Based Performance (ROC) after RS-LDA training on PCSO database.

### C. Robustness to Changes in Facial Pose

State of the art automated face recognition performs well in controlled situations, that is, when factors like illumination, pose, expression and occlusion are eliminated. However, when such variates cannot be constrained, automated face recognition algorithms generally exhibit a decrease in recognition accuracy. We demonstrate in this section and in Section IV-D that the component-based approach has the potential to be very useful in the presence of occlusions and changes in facial pose. In addition to the holistic representations used for previous experiments, the PittPatt Multi-Pose [39] and FaceVACS [54] commercial off the shelf (COTS) face recognition systems were also used for comparison. We also find that the component-based method performs largely the same as the holistic methods in the presence of changes in facial expression though the fusion of holistic and component-based representations can result in a boost in accuracy.

The following experiment explores the robustness of the component-based approach for varying pitch values<sup>6</sup>. The FERET database (with  $n = 200$  subjects) was used for this experiment [52]. In addition to the frontal view images for 200 subjects, the FERET database provides 8 additional face images at different pitch values:  $\pm 15^\circ$ ,  $\pm 25^\circ$ ,  $\pm 40^\circ$ ,  $\pm 60^\circ$ . Facial landmarks are extracted automatically and used for

<sup>6</sup>Pitch rotation refers to rotations performed about the y-axis or specifically in the case of faces, rotations performed by twisting the head to look left or right.

alignment as described in Section II. The component-based approach is compared to both the holistic approach with global eye alignment and the holistic approach with global Procrustes alignment. As before, the cosine similarity metric and sum score fusion of z-score normalized similarities was used in all cases. The recognition accuracies on this multi-pose dataset are displayed in Table VII. For each pose, only subjects that enrolled with both PittPatt and FaceVACS were used in all cases. Thus, the number of subjects  $n$  varies in the extreme poses.

For small changes in pose (close to pitch value  $0^\circ$ ) the performances of the commercial systems, the holistic approaches, and the component-based approaches are essentially the same. However, for large changes in pose, each individual component alone outperforms the holistic approaches, with the exception of the nose (see Table VII and Figure 13). Given the drastic difference in the appearance of the nose in the side versus frontal face view, the low performance of the nose is to be expected. We also observe relatively asymmetrical performance (e.g., difference in accuracies between poses of  $+60$  and  $-60$ ) but this is likely due to slight imprecision in the actual pitch values of the subject's face and the recorded values (i.e., there is noise in the exact pose of the face).

As for the commercial systems, PittPatt's multi-pose face recognition system has traditionally been the strongest for pose variation and that is also evident in our results. PittPatt provides the best individual performance on face recognition for extreme poses, followed by the proposed technique of fused components. When fusing PittPatt and the proposed component-based representation (using sum of scores fusion), the recognition accuracy improved to over 78% for  $\pm 60^\circ$  pitch values. This performance is significantly higher than the performance of PittPatt alone, indicating additional and complementary discriminating information is captured in the proposed component-based representation. Overall, the use of a component-based face representation shows tremendous potential to help decrease the within-subject variation caused by changes in pose.

There is some difficulty in automatically detecting landmarks for the extreme poses variations. The previous results are presented for those images which can be commonly enrolled between our holistic and component-based approaches, and the commercial systems, FaceVACS and PittPatt (see Table VII). We will address these issues from a forensic recognition perspective in Section IV-F.

### D. Robustness to Occlusion

A distinct advantage of the proposed component-based representation is in forensic recognition scenarios, where analysts

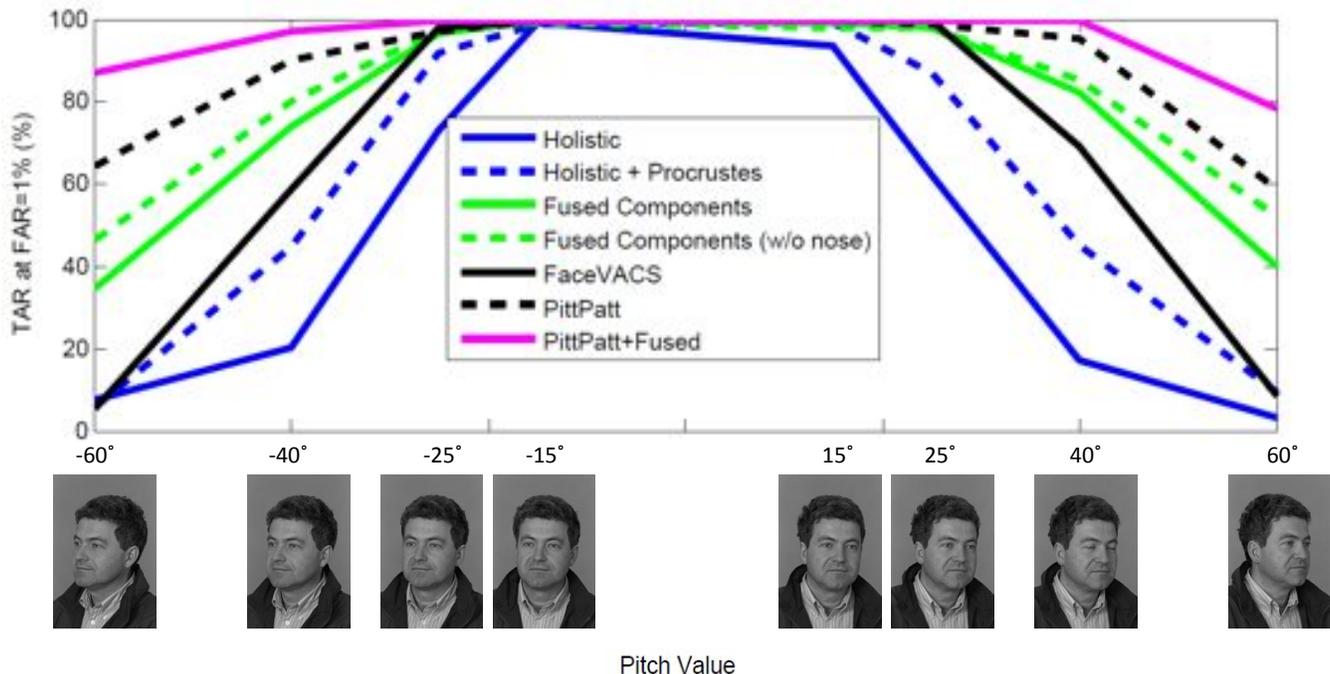


Fig. 13. Recognition accuracies (TAR at FAR=1%) on multi-pose images from the FERET dataset. The more traditional approach of performing holistic face alignment results in severely degraded recognition accuracy as the pose of the face changes. By contrast, the proposed component-based approach remains relatively robust across pose variations. Without leveraging any training data, the proposed component-based method even surpasses one of the most accurate commercial face recognition systems, FaceVACS. When fusing the component-based representation with PittPatt’s commercial face recognition system, state-of-the-art accuracies are achieved.

TABLE VII  
RECOGNITION ACCURACIES (TAR AT FAR=1%) ACROSS POSE VARIATIONS.

Pitch Value	$n$ sub-jects	Holistic Global Alignment	Holistic Procrustes Alignment	Eyebrows	Eyes	Nose	Mouth	Fused Components	Fused Components w/o Nose	FaceVACS	PittPatt	PittPatt+Fused Components
+60°	141	3.31%	9.45%	22.69%	12.43%	6.37%	36.30%	40.00%	52.16%	8.60%	58.53%	78.39%
+40°	199	17.32%	45.06%	51.81%	45.07%	21.52%	79.41%	82.20%	85.37%	69.10%	95.59%	99.62%
+25°	200	62.33%	87.15%	81.68%	85.35%	59.63%	90.48%	98.09%	97.83%	99.68%	98.91%	100.00%
+15°	200	93.88%	99.24%	93.88%	93.71%	98.01%	93.77%	99.50%	98.11%	100.00%	99.81%	100.00%
-15°	200	99.53%	98.86%	98.18%	94.50%	98.08%	92.95%	100.00%	99.23%	100.00%	100.00%	100.00%
-20°	200	72.94%	91.98%	83.97%	84.05%	66.83%	89.48%	96.86%	97.80%	98.22%	97.25%	99.87%
-40°	198	20.41%	44.79%	47.03%	52.95%	20.22%	67.45%	74.04%	79.99%	58.62%	90.04%	97.01%
-60°	174	7.55%	5.86%	16.01%	22.03%	5.75%	40.33%	34.70%	46.72%	5.41%	64.39%	87.21%

submit face queries and examine retrieval results [55]. In this section we will consider the case of face occlusions, which many criminals use to evade identification. When representing faces using facial components, an analyst could simply choose to omit occluded facial regions.

The robustness of the component-based approach to occlusion was explored using a subset of the AR database (with  $n = 136$  subjects) [56]. The AR database contains frontal views of faces under varying expression, illumination and occlusion. Frontal views under occlusion (i.e. faces wearing sunglasses and faces wearing scarves) were utilized in this experiment. The frontal view faces were represented using a holistic approach with global eye alignment, a holistic approach with global Procrustes alignment and the proposed component-based approach. Again, we used the cosine similarity metric

to compare facial representations. The individual components in the component-based approach were combined using sum of scores fusion but did not include occluded components (i.e. for faces wearing sunglasses, fusion is only done for nose and mouth and for faces wearing scarves, fusion is only done for eyebrows, eyes and nose). As discussed above, this replicates a common scenario in forensic face recognition, where an analyst would be able to provide the missing information (in this case the location of the occlusion) to improve face retrieval accuracy.

Table VIII provides a summary of the recognition results for the proposed component-based and holistic approaches excluding images that fail to enroll. The column labelled  $n$  indicates the number of images which were successfully enrolled. Because matching for the fused component-based

TABLE IX  
 OCCLUSION: COMPARISON OF COMPONENT-BASED APPROACH TO TWO  
 COMMERCIAL SYSTEMS (TAR AT FAR=1%)

Occlusion Type	$n$	Fused Components	FaceVACS	PittPatt
Wearing Sunglasses	63	72.77%	20.66%	15.53%
Wearing Scarf	111	97.43%	99.11%	89.86%

approach is based on components that are not occluded, we observe a marked increase in performance over the holistic approach, especially for the faces wearing sunglasses. Notably, in the case of the sunglasses, the performance of the mouth alone outperforms the fused performance of the nose and mouth. This result is likely due to the fact that the presence of the sunglasses near the nose actually decreases the quality of the landmark extraction. However, in forensic applications, automatic extraction of these landmarks is not essential - if a person's face is occluded in a way that interferes with landmark extraction for visible components, law enforcement personnel could mark these landmarks manually and likely produce an improved annotation (see Section IV-F for further discussion). Manual annotation is also a mitigating procedure for those face images for which automatic extraction of facial components fails. Table IX displays results for our component-based approach in comparison to commercial systems for face recognition on commonly enrolled images. Because the images must be successfully enrolled in both the commercial system and our component-based system, the number of available subjects to make this comparison is somewhat reduced. The proposed component-based approach achieves accuracies similar to PittPatt and FaceVACS on faces occluded by a scarf. However, the fused component approach for faces with sunglasses improves accuracy nearly fourfold over the commercial systems. This result continues to demonstrate the potential of the component-based system in the presence of occlusion although further validation may be necessary given the relatively small number of commonly enrolled subjects.

#### E. Facial Expression

This experiment investigates the robustness of the component-based approach to facial expression using a subset of the AR database ( $n = 136$  subjects) [56]. As previously described, the AR database contains frontal views under varying expression, illumination and occlusion. The neutral images were used as the gallery and the images with varied expressions (smile, anger, and scream) were used as the probe set. As before, each image was represented using a holistic approach with global eye alignment, a holistic approach with global Procrustes alignment and the proposed component-based approach. In each of the experiments 2-3 images were removed due to ASM extraction failure. The cosine similarity metric was used to measure the similarity between representations. All four components were fused together using sum of scores fusion. Table X provides a summary of the results. The fused component method performs comparably to the holistic methods for the given expressions (smile, anger,

and scream). Fusing together the component-based similarities with the holistic global alignment results in a very slight boost to performance for smile and anger, which are already near 100% TAR at FAR = 1%. However, for the scream expression this fusion results in a 10% boost in the TAR. Though the component-based method is comparable or slightly worse than the holistic methods for facial expressions, these results indicate that the component-based method might be leveraged in addition to the holistic method in order to boost the overall accuracy of a face recognition method.

#### F. Manual Landmark Annotation to Mitigate Poor/Failed Automated Landmark Detection

One of the weaknesses of the proposed algorithm is that it hinges on the ability to detect and align individual components of the face. The results displayed in Sections IV-C and IV-D demonstrate that in the presence of occlusion and pose variations, ASM sometimes fails to converge on a set of facial landmarks, or results in highly inaccurate landmarks. This was true especially for sunglasses occluding the eyes and poses with pitch values of  $\pm 40^\circ$  and  $\pm 60^\circ$ . In the case of inaccurate facial landmarks the matches are very poor, and in the case of landmark detection failure, recognition cannot be performed. In order for the component-based approaches to be robust under various types of noise, a process for addressing these failures and inaccuracies is crucial. In typical forensic applications of face recognition, automatic extraction of landmarks is preferred but not mandatory; an analyst could be easily trained to manually annotate the landmarks for the images in which automatic detection fails [55]. This section will demonstrate that the use of manual landmarks is a viable procedure when automatic detection fails.

For occlusion it was determined that the automatic landmark detection was consistently poor especially near the occlusions. Thus, for this occlusion experiment, we utilized the same three images (neutral frontal, with sunglasses, and with scarf) from the 136 subjects in the AR database as in Section IV-D. Some images are missing for the 136 subjects, resulting in 135 gallery neutral frontal images, 135 images with sunglasses and 136 images with scarves. The automatically detected landmarks were used for all of the neutral frontal views as ASM is extremely effective for these cases. All images with occlusions caused by sunglasses or a scarf were manually annotated to produce sets of landmarks. After the automatically detected landmarks and the manually generated landmarks are produced, the procedure continues beginning at the alignment step as previously outlined in Section II.

Using these manual annotations there is a small performance decline for images occluded by a scarf, 91.85% compared to 95.59% (TAR at FAR=1%). However, there is a marked improvement for the face images occluded by sunglasses, 93.84% compared to 72.19% (see Tables VIII and XI). This boost in performance demonstrates the viability of manual annotation for probe images in a component-based system. Furthermore, the performance on the sunglasses occluded images in comparison to commercial face recognition systems (see Table IX) is the overwhelming evidence in support of value of properly aligned components in face recognition.

TABLE VIII  
ROBUSTNESS TO OCCLUSION (TAR AT FAR=1%) IN THE FERET DATABASE

Occlusion Type	$n$ subjects	Holistic Global Alignment	Holistic Procrustes Alignment	Eyebrows	Eyes	Nose	Mouth	Fused Components
Wearing Sunglasses	102	36.39%	44.75%	1.04%*	5.68%*	40.37%	85.30%	72.34%
Wearing Scarf	117	92.19%	70.74%	89.57%	91.50%	90.95%	0.85%*	96.67%

\* These components were not included in fusion.

TABLE X  
FACIAL EXPRESSION (TAR AT FAR=1%) IN THE AR DATABASE

Expression	$n$ subjects	Fused Components	Holistic Global Alignment	Holistic Procrustes Alignment	Holistic Global Alignment + Fused
Smile	134	98.60%	99.18%	98.03%	99.20%
Anger	133	99.27%	99.33%	100%	100%
Scream	133	74.36%	75.22%	71.06%	86.10%

TABLE XI  
OCCLUSION: MANUAL ANNOTATION (TAR AT FAR=1%) IN FERET DATABASE

Occlusion Type	Holistic Global Alignment	Holistic Procrustes Alignment	Eyebrows	Eyes	Nose	Mouth	Fused Components
Wearing Sunglasses	56.29%	53.22%	3.92%*	2.45%*	66.38%	93.98%	95.00%
Wearing Scarf	94.35%	90.19%	91.52%	94.59%	84.93%	1.71%*	97.30%

\* These components were not included in fusion.

Facial component extraction for extreme pose variations is very successful for the FERET database except for the extreme poses at pitch values of  $\pm 40^\circ$  and  $\pm 60^\circ$ . In general, when the automatic extraction procedure succeeds on a face image with an extreme pose the localization of the inner components is acceptable, but the outer face landmarks (denoting the face outline) are not. For our purposes it is sufficient for only the inner landmarks to be accurate because these are the landmarks which both the holistic and component approaches utilize (i.e., alignment is based solely on these inner landmarks). However, extraction does fail in the cases where PittPatt fails to detect the eyes and nose. In these failure to enroll cases we manually marked the landmarks. Once the automatic landmarks (for faces that are properly enrolled) and manual landmarks (for failure to enroll faces) have been extracted, the verification procedure described in Section IV-C was then repeated for all face images at pitch values of  $\pm 40^\circ$  and  $\pm 60^\circ$  (see Table XII).

We observe a comparable performance in this experiment using a combination of manually annotated and automatically extracted landmarks. That is, the faces that we needed to manually annotate matched with the same accuracy as those faces where automatic landmark extraction was successful. This result supports the viability of manual annotation for face images under pose variation, and, thus, for forensic face recognition applications. There are some differences between the matching performance for the face images at  $\pm 60^\circ$  pitch, though we would expect them to be largely symmetrical. This could be due in large part to the fact that the automatic extraction failed for face images at  $+60^\circ$  pitch significantly more (approximately double) than at  $-60^\circ$  pitch. Thus, manual

annotation is performed on significantly more of these images, thereby causing differences in performance. However, we observe that this increased manual annotation for the  $+60^\circ$  face images does not decrease performance but appears to have improved the performance. This asymmetrical enrollment failure may be due to differences within the image set or biases in the automatic extraction method that favor one orientation over the other.

## V. CONCLUSIONS

Motivated by studies in cognitive science literature, the main objective of this work is to demonstrate the potential of component-based representations in automated face recognition. In the previous studies on this topic, the difficulties in extracting individual facial components prevented the effective use of component-based approaches in automatic face recognition. However, using more precise approaches for face alignment, it is possible to effectively align and extract such components. Indeed, face recognition accuracies using our proposed component-based representations greatly exceed the accuracy using similar holistic (or globally aligned) representations. Further, we demonstrated the effectiveness of the proposed component-based approach when addressing facial occlusions and variations in orientation.

There are a number of practical advantages of component-based approaches to face recognition. The use of face recognition systems by law enforcement agencies is not a fully automated endeavor for all scenarios [55]. The ability to exclude certain facial components and more generally to analyze the per-component similarity in face comparisons could be

TABLE XII  
POSE VARIATION: MANUAL ANNOTATION (TAR AT FAR=1%) IN AR DATABASE

Pitch Value	Holistic Global Alignment	Holistic Procrustes Alignment	Eyebrows	Eyes	Nose	Mouth	Fused Components	Fused Components w/o Nose
+60°	3.31%	9.16%	18.74%	11.99%	7.16%	35.21%	47.03%	55.55%
+40°	17.26%	41.38%	51.24%	49.66%	21.17%	78.11%	84.33%	85.94%
-40°	20.16%	40.58%	46.25%	52.91%	19.81%	66.71%	75.26%	78.44%
-60°	7.51%	9.63%	16.74%	26.91%	6.80%	40.46%	31.75%	45.71%

extremely useful to law enforcement personnel and biometric analysts. As previously demonstrated, it might allow the user of the system to exclude the eyes when a person is wearing sunglasses, or the mouth when they are wearing a scarf (for example). Furthermore, if the automated landmark extraction fails or is not ideal for the desired components, a user could mark components manually resulting in an improved set of landmarks.

Having demonstrated the improvement in recognition accuracy yielded from component-based representations, a viable future research topic is a dedicated study on how to best tailor learning-based methods to component-based representations. While existing algorithms can readily be applied to these representations (as we demonstrated using RS-LDA), nuances of this representation may afford further performance increases by specially tailored learning algorithms.

In conclusion, we have demonstrated the potential of component-based alignment and representations towards improving automated face recognition algorithms. Given similar evidence in the cognitive science domain, it is important that component-based face representations be considered as a viable strategy in designing automated face recognition systems.

#### ACKNOWLEDGEMENTS

The authors would like to thank The Pinellas County Sheriff's Office for providing a large database of face images. The work of A.K. Jain was supported in part by the World Class University (WCU) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0). All correspondence should be directed to A.K. Jain.

#### REFERENCES

- [1] J. Bradshaw and G. Wallace, "Models for the processing and identification of faces," *Perception and Psychophysics*, vol. 9, pp. 443–448, 1971.
- [2] J. Sergent, "An investigation into component and configural processes underlying face perception," *British Journal of Psychology*, vol. 75, pp. 221–242, 1984.
- [3] A. Schwaninger, S. Schumacher, C. Wallraven, and H. Buelthoff, "Using 3d computer graphics for perception: The role of local and global information in face processing," *Applied Perception in Graphics and Visualization*, 2007.
- [4] A. Schwaninger, C.-C. Carbon, and H. Leder, *Development of Face Processing*. Goettingen: Hogrefe, 2003.
- [5] A. Schwaninger, C. Wallraven, D. Cunningham, and S. Chiller-Glaus, "Processing of identity and emotion in faces: A psychophysical, physiological and computational perspective," *Progress in Brain Research*, vol. 156, pp. 321–343, 2006.
- [6] D. Maurer, R. Le Grand, and C. Mondloch, "The many faces of configural processing," *TRENDS in Cognitive Sciences*, vol. 6, no. 6, June 2002.
- [7] N. Sagiv and S. Bentin, "Structural encoding of human and schematic faces: Holistic and part-based processes," *Journal of Cognitive Neuroscience*, vol. 13, no. 7, pp. 937–951, 2001.
- [8] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [9] B. Klare, Z. Li, and A. Jain, "Matching forensic sketches to mugshot photos," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639–646, 2011.
- [10] E. Meyers and L. Wolf, "Using biologically inspired features for face processing," *Int. Journal of Computer Vision*, vol. 76, no. 1, pp. 93–104, 2008.
- [11] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [12] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, 1997.
- [13] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [14] B. Klare and A. Jain, "On a taxonomy of facial features," in *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2010.
- [15] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," in *Proc. IEEE Int. Conference on Image Processing*, vol. 1, p. 129, 1997.
- [16] M. Matthews, "Discrimination of Identikit construction of faces: Evidence for a dual processing strategy," *Perception and Psychophysics*, vol. 23, pp. 153–161, 1978.
- [17] E. Smith and G. Nielsen, "Representations and retrieval processes in short-term memory and recall of faces," *Journal of Experimental Psychology*, vol. 85, pp. 397–405, 1970.
- [18] J. Shepherd, G. Davies, and H. Ellis, "Studies of cue saliency," *Perceiving and Remembering Faces*, 1981.
- [19] M. Farah, K. Wilson, M. Drain, and J. Tanaka, "What is "special" about face perception," *Psychological Review*, vol. 105, no. 3, pp. 482–498, July 1998.
- [20] J. Richler, O. Cheung, and I. Gauthier, "Holistic processing predicts face recognition," *Psychological Science*, vol. 22, no. 4, pp. 464–471, 2011.
- [21] J. Gold, P. Mundy, and B. Tjan, "The perception of a face is no more than the sum of its parts," *Psychological Sciences*, March 2012.
- [22] J. Sadr, I. Jarudi, and P. Sinha, "The role of eyebrows in face recognition," *Perception*, vol. 32, pp. 285–293, 2006.
- [23] R. Mauro and M. Kubovy, "Caricature and face recognition," *Memory and Cognition*, vol. 20, no. 4, pp. 433–440, 1992.
- [24] G. Rhodes, S. Brennan, and S. Carey, "Identification and ratings of caricature: Implications for mental representations of faces," *Cognitive Psychology*, vol. 19, no. 4, pp. 473–497, Oct. 1983.
- [25] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1053, Oct. 1993.
- [26] B. Heisele, T. Serre, and T. Poggio, "A component-based framework for face detection and identification," *International Journal of Computer Vision*, vol. 74, no. 2, 2007.
- [27] J. Huang, P. Yuen, W.-S. Chen, and J. Lai, "Component-based lda method for face recognition with one training sample," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 120–126, Oct. 2003.

- [28] W. Zhang, S. Shan, W. Gao, Y. Chang, and B. Cao, "Component-based cascade linear discriminant analysis for face recognition," *Advances in Biometric Person Authentication Lecture Notes in Computer Science*, vol. 3338, pp. 19–79, 2005.
- [29] D. Lin and X. Tang, "Recognize high resolution faces: From macrocosm to microcosm," *Computer Vision and Pattern Recognition*, pp. 1355–1362, 2006.
- [30] K. Pan, S. Liao, S. Zhang, S. Z. Li, and P. Zhang, "Part-based face recognition using near infrared images," *IEEE conference on Computer Vision and Pattern Recognition*, 2007.
- [31] T.-K. Kim, H. Kim, W. Hwang, and J. Kittler, "Component-based lda face description for image retrieval and mpeg-7 standardisation," *Image and Vision Computing*, vol. 23, pp. 631–642, 2005.
- [32] Y. Su, S. Shiguang, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Transactions on Image Processing*, vol. 18, August 2009.
- [33] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of gabor magnitude and phase for face recognition," *IEEE Transactions on Image Processing*, vol. 19, May 2010.
- [34] U. Park, R. Jillela, A. Ross, and A. Jain, "Periocular biometrics in the visible spectrum," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 1, pp. 96–106, 2011.
- [35] P. Miller, J. Lyle, S. Pundik, and D. Woodard, "Performance evaluation of local appearance based periocular recognition," *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2010.
- [36] S. Bharadwaj, H. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2010.
- [37] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, 1999.
- [38] T. Cootes, C. Taylor, and D. Cooper, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, 1995.
- [39] PittPatt Face Recognition SDK, Pittsburgh Pattern Recognition, <http://www.pittpatt.com>.
- [40] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," *ECCV*, 2008, <http://www.milbo.users.sonic.net/stasm>.
- [41] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
- [42] B. Klare, A. Paulino, and A. Jain, "Analysis of facial features in identical twins," *IJCB*, 2011.
- [43] B. Klare, P. Mallapragada, A. K. Jain, and K. Davis, "Clustering face carving: Exploring the devatas of angkor wat," *Proc. International Conference on Pattern Recognition*, 2010.
- [44] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [45] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear project," *IEE Trans. Pattern Analysis & Machine Intelligence*, vol. 19, no. 7, pp. 711–721, 1997.
- [46] X. Wang and X. Tang, "Random sampling for subspace face recognition," *Int. Journal of Computer Vision*, vol. 70, no. 1, pp. 91–104, 2010.
- [47] H. Jie, H. Yu, and J. Yang, "A direct lda algorithm for high-dimensional data - with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.
- [48] J. Lu, K. Plataniotis, and A. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letters*, vol. 26, no. 2, pp. 181–191, 2005.
- [49] B. Klare and A. Jain, "Face recognition across time lapse: On learning feature subspaces," *International Joint Conference on Biometrics*, 2011.
- [50] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *University of Massachusetts, Amherst, Technical Report 07-49*, October 2007.
- [51] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *Pattern Analysis and Machine Intelligence, IEE Transactions on*, vol. 30, no. 3, pp. 383–394, March 2008.
- [52] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The feret evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, October 2000.
- [53] C. Geng and X. Jiang, "Face recognition based on the multi-scale local image structures," *Pattern Recognition*, vol. 44, pp. 2565–2575, 2011.
- [54] FaceVACS Software Developer Kit, Cognitec Systems GmbH, <http://www.cognitec-systems.de>.
- [55] A. K. Jain, B. Klare, and U. Park, "Face matching and retrieval: Applications in forensics," *IEEE Multimedia*, vol. 19, no. 1, pp. 20–28, 2012.
- [56] A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, no. 24, 1998.



research interests include pattern recognition, image processing, and computer vision.



Fingerprint Recognition (2009), Handbook of Biometrics (2007), Handbook of Multibiometrics (2006), BIOMETRICS: Personal Identification in Networked Society (1999), and Algorithms for Clustering Data (1988). He served as a member of the Defense Science Board and The National Academies committees on Whither Biometrics and Improvised Explosive Devices.

Dr. Jain received the 1996 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award and the Pattern Recognition Society best paper awards in 1987, 1991, and 2005. He is a fellow of the AAAS, ACM, IAPR, and SPIE. He has received Fulbright, Guggenheim, Alexander von Humboldt, IEEE Computer Society Technical Achievement, IEEE Wallace McDowell, ICDM Research Contributions, and IAPR King-Sun Fu awards.



**Brendan F. Klare** is a lead scientist at Noblis. He received the B.S. and M.S. degrees in Computer Science from the University of South Florida in 2007 and 2008, and the Ph.D. degree in Computer Science from Michigan State University in 2012. From 2001 to 2005 he served as an airborne ranger infantryman in the 75th Ranger Regiment. Brendan has authored several papers on the topic of face recognition, and he received the Honeywell Best Student Paper Award at the 2010 IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS). His other

**Anil K. Jain** is a University distinguished professor in the Department of Computer Science and Engineering at Michigan State University, East Lansing. His research interests include pattern recognition and biometric authentication. He served as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (1991-1994). The holder of six patents in the area of fingerprints, he is the author of a number of books, including Introduction to Biometrics (2011), Handbook of Face Recognition (2011), Handbook of Fingerprint Recognition (2009), Handbook of Biometrics (2007), Handbook of Multibiometrics (2006), BIOMETRICS: Personal Identification in Networked Society (1999), and Algorithms for Clustering Data (1988). He served as a member of the Defense Science Board and The National Academies committees on Whither Biometrics and Improvised Explosive Devices.

**Kathryn Bonnen** is a graduate student in the Institute for Neuroscience at the University of Texas at Austin. In 2011, she received B.S. degrees in Computer Science and Psychology from Michigan State University. She is a National Science Foundation graduate fellow and a Harrington fellow. Her research interests include human learning and vision, pattern recognition, and image processing.