Research Performance Final Report

Federal Agency and Organization Element to Which Report is Submitted	National Institute of Justice
Federal Grant or Other Identifying Number Assigned by Agency	2013-IJ-CX-K005
Project Title	A Simultaneous Low Resolution and Off-Pose Angle Face Matching Algorithm as an Investigative Lead Generative Tool for Law Enforcement
PD/PI Name, Title and Contact Information (e-mail address and phone number)	PI: Prof. Marios Savvides Director, CyLab Biometrics Center Associate Research Professor Electrical & Computer Engineering Carnegie Mellon University 5000 Forbes Ave, Pittsburgh, PA 15213 msavvid@ri.cmu.edu (412) 980-8939
Name of Submitting Official, Title, and Contact Information (e-mail address and phone number), if other than PD/PI	Rebecca Pawlikowsky, Research Administrator, rebeccap@andrew.cmu.edu (412) 268-6455
Submission Date	8/2/2017
DUNS and EIN Numbers	
Recipient Organization (Name and Address)	Carnegie Mellon University 5000 Forbes Avenue Pittsburgh, PA 15213-3815
Recipient Identifying Number or Account Number, if any	
Project/Grant Period (Start Date, End Date)	10/1/13-9/30/17
Reporting Period End Date	June 30, 2017
Report Term or Frequency (annual, semi-annual, quarterly, other)	Semi-Annual
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	

ACCOMPLISHMENTS: What was done? What was learned?

What are the major goals of the project?

The major goal of the project is to research and develop a forensic tool that can be disseminated to the law enforcement and criminal justice community, providing them with the capability to perform facial recognition using low quality, low resolution faces, such as those obtained from CCTV surveillance footage in many unsolved crimes.

What was accomplished under these goals?

A major accomplishment has been the development of a *unified face representation model* which is able to interpret various face degradation scenarios such as low resolution, pose, occlusions, etc. Specifically, we have re-interpreted the problem of face recognition & recovery under acquisition degradation as a missing-data recovery problem. This is an outcome of the face representation, which is able to analyze and incorporate the effects of resolution, pose, occlusions, etc. under a unified framework. Missing / corrupted data recovery techniques are then used to obtain a sanitized representation for recognition and synthesis. These techniques use models which typically need to be learned on similar images from a training dataset. We have developed, implemented and tested a few such techniques which are able to learn these models from large numbers of high-dimensional training data, enabling improved performance.
Our method can also be used as a pre-processing step to aid recognition by several commercial face recognition engines, thereby expanding the scope of these tools.

Using this representation of faces, we have been able to develop automated facial occlusion-recovery system. The system is able to reconstruct parts of the face which are not visible in an image due to an obstruction. Facial occlusions such as scarves, masks, sunglasses, eyeglasses, hair, etc., are often observed in low-resolution surveillance video, and these have a severe, irrecoverable impact on current facial recognition technology. The developed technique can alleviate these problems, thereby overcoming a pervasive shortcoming of super-resolution & other facial analysis techniques.

Additionally, we have developed a method for recovering both the representation vector and a set of confidences for off angle face images. Using these along with data completion techniques outlined in previous reports, frontal face images can be reconstructed and passed through face recognition engines.

We have also developed a periocular reconstruction & recognition technique. This technique aims to recover the full face image based off just the periocular region of the subject with high fidelity. The recovered full face can then be used for face recognition, and thereby overcoming the limit of matching subjects wearing masks or burkas. The improved models are able to show recovery on the ISIS “Jihadi John” suspect and are able to match the recovery to the true subject’s frontal face image with high accuracy.

In all of these approaches, the main problem was detecting these heavily degraded faces. To address this, we have developed new methods for detecting heavily occluded and off-angle faces have been developed and benchmarked on the WiderFace dataset.

There has also been new work done on alignment free 3D modeling for the purpose of pose correction or synthesis as requiring an end user to nudge landmark points was often too tedious to allow processing of large datasets. This new 3D modeling is much faster and extracts more accurate landmark points than previous methods.

More research has been done on the face matching component as well, using some of the latest techniques in the field using deep learning and new loss functions designed to promote discriminative feature extraction. The new face matcher has been trained on large datasets and is in the process of being benchmarked on datasets of interest such as IJB-A.

Besides algorithm development, we have selected and acquired multiple training and testing databases for experiments. The MBGC dataset has been manually processed for localization of landmark keypoints. The AR face database, used to demonstrate and evaluate the occlusion recovery technology, has been manually processed for localization of landmark keypoints and annotation of occluded facial regions. The CMU Multi-PIE and Labeled Faces in the Wild (LFW) allow us to show how the recognition performance functions at different degradations (pose, resolution, etc.) individually as well as in combination as in the case of LFW.

What opportunities for training and professional development has the project provided?

Nothing to Report

How have the results been disseminated to communities of interest?

The project has sparked collaborations between the CMU's CyLab Biometrics Center and two notable law enforcement agencies: (1) the NYPD's Real-Time Crime Center (RTCC), and (2) Pinellas County Sherriff's Office (PCSO).

The RTCC is a centralized technology center for the NYPD, which provides video, image and data analytics support to many branches of the NYPD and other law enforcement agencies. CMU researchers working on this project made a trip to the NYPD headquarters in New York City (funded through a collaboration with the CMU Software Engineering Institute), to introduce the NYPD personnel to the research and potential applications in real-world crime investigations. The research was well received NYPD detectives and analysts, and we are currently in the process of developing and transitioning usable software tools for the NYPD and other law enforcement agencies.

CMU researchers also met and have transitioned the occlusion recovery software to Pinellas County Sherriff's Office (PCSO), Florida, where it is currently being employed on ongoing criminal investigations. The PCSO is leading an extensive initiative in Florida's criminal justice system to incorporate facial recognition in investigative procedure, and works in conjunction with several law enforcement agencies across Florida with this goal.

What do you plan to do during the next reporting period to accomplish the goals?

No Change

PRODUCTS: What has the project produced?

Publications, conference papers, and presentations

Journal publications:

Ramzi Abiantun, Utsav Prabhu, and Marios Savvides, “Sparse Feature Extraction for Pose-Tolerant Face Recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, no.10, pp.2061-2073, Oct. 2014

Felix Juefei-Xu, Dipan K. Pal, and Marios Savvides, "Hallucinating the Full Face from the Periocular Region via Dimensionally Weighted K-SVD," *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW*), pp.1-8, June 2014

Felix Juefei-Xu, Dipan K. Pal, and Marios Savvides, "NIR-VIS Heterogenous Face Recognition via Cross-Spectral Joint Dictionary Learning and Reconstruction," *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW*)

Utsav Prabhu and Marios Savvides, “Face Recognition under Simultaneous Acquisition Degradations”, (Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence,* under review)

Other publications

Yutong Zheng*, Chenchen Zhu*, Khoa Luu, Chandrasekhar Bhagavatula, T. Hoang Ngan Le, and Marios Savvides. "*Towards a deep learning framework for unconstrained face detection.*" In Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, pp. 1-8. IEEE, 2016.
* means the authors contributed equally

Demonstrations

This work has been demonstrated at the Global Identity Summit (formerly known as the Biometrics Consortium Conference) in Tampa, FL. It was also demonstrated multiple times at Carnegie Mellon Unviersity CyLab Biometrics Center, Pittsburgh PA to multiple government agencies during visits throughout the year.

*Web*site(s) or other Internet site(s)

None.

Technologies or techniques

1. A Novel Representation for Faces

Faces are notoriously difficult to model; they are elastic, deformable shapes, are easily influenced by external factors, and are acquired using different methods. Typically, for most pattern recognition and data analysis tasks, a single instance of a face (which could be a 3D scan or 2D image) is represented as a single vector, with each element of the vector incorporating within it a certain characteristic of the face.

The representation of the face is a critical element in the design of any system; a poor representation can impede subsequent tasks significantly, and could prove to be the weakest link in the processing chain. From an information-theoretic perspective, many forms of acquisition degradation can be viewed as missing data, and can result in a local, measurable and accumulative impact on a prudently designed representation of the face. With an adequate understanding of the impact of each degradation, one can hence design a representation for which these properties are appropriate. In such a representation, recovery of the original face given an acquisition which contains a combination of different types of degradations, can be viewed as a *missing data completion* problem, which can be solved using a variety of techniques (along with sufficient training data).

We identify three fundamental properties that the ideal model space representation must satisfy in order for the proposed approach to be maximally effective: *Consistency*, *Completeness*, and *Predictable Degradation*. Given the varied types of face degradation, consistency is a non-trivial condition to always satisfy in the representation. The entire information contained in the face should represented in the model space, i.e. it should include a complete and highly detailed representation of the face. Most data completion techniques use available information to impute (i.e. recover) missing data. The inclusion of all possible information from the face enables the best possible result, independent of completion technique. Besides, a complete representation implies the possibility of re-synthesis of the face from the representation, and guarantees the retention of identity. The feature representation should also be predictable under degradation, i.e. should be such that a known degradation in acquisition affects a correspondingly calculable subset of the d dimensions of the model representation to a calculable degree, for all the expected types of degradation that are expected to be observed.

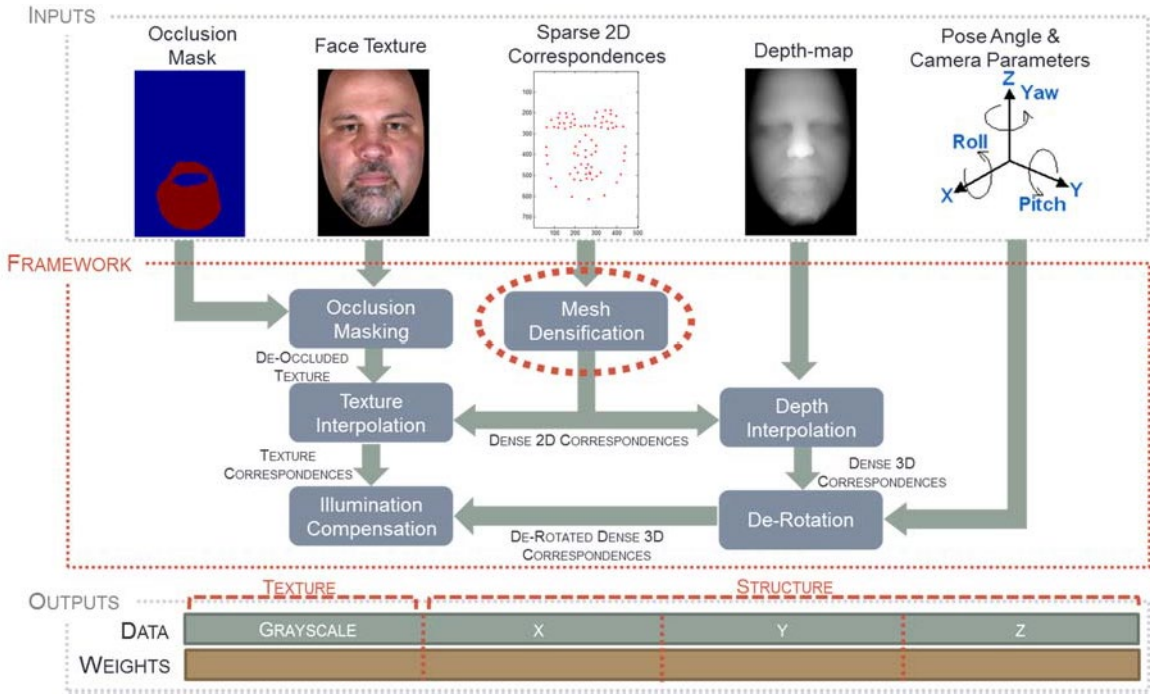


Figure 1: A flowchart of our entire representation process. We dis-associate shape from texture, and extract a data vector as well as a measurement confidence vector.

In order to satisfy these properties, our face representation technique is modeled by using a *Multi-Resolution Thin-Plate Spline deformation model*, which is learned from a sparse set of correspondences on the face. This allows us to model a dense facial structure with correspondences across a variety of occlusions, illuminations, resolutions, and other acquisition degradation conditions. This approach begins with the provided sparse fiducial point set of n points. We then use the 2D (x,y) coordinates of this sparse correspondence set to infer a thin-plate interpolating spline function. Let the sparse correspondences of a facial surface be represented by the points (x_i, y_i) and let be the set of the mean 2D coordinates computed over a database set of such faces. The thin-plate spline solution then allows us to find the smoothest possible real-valued function s which satisfies the interpolation problem

$$s(\bar{x}_i, \bar{y}_i) = (x_i, y_i), \quad \forall\ i = 1, 2, \ldots, 45$$

where s captures the deviation of the 2D feature points from the mean. The measure of smoothness used in this case is the integral

$$I(s_i) = \iint_{\Re^2} \left(\frac{\partial^2 s}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 s}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 s}{\partial y^2}\right)^2 dxdy$$

The variational problem of minimizing this smoothness measure subject to the interpolation constraints in specified above has been shown to have a unique solution given by the thin plate spline of the form

$$s(\bar{x}_i, \bar{y}_i) = a_0 + a_1 x + a_2 y + \sum_{i=1}^{45} \lambda_i r_i^2 \log r_i$$

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

where (a_0, a_1, a_2) represent the rigid deformation coefficients, λ_i represents the non-rigid spline deformation coefficients, and $r_i^2 = (x-x_i)^2 + (y-y_i)^2$. It turns out that, the integral smoothness measure is finite if and only if the non-rigid coefficients λ_i have the properties:

$$\sum_{i=1}^{n} \lambda_i = 0, \quad \sum_{i=1}^{n} \lambda_i x_i = 0, \quad \sum_{i=1}^{n} \lambda_i y_i = 0.$$

There thus exists an efficient and closed-form technique to determine these coefficients by matrix inversion. We select a set of uniformly distributed points on the mean face, and use the computed warping function to find the correspondences on the test face.

The thin plate spline interpolant allows us to overcome the negative aspects of other dense correspondence surfaces. Most importantly, the resulting vertex set can be selected to be uniformly distributed across the facial surface resulting in a more homogenous representation of the face, and achieving a similar overall representation density with fewer vertices. We select the granularity of the initial distribution of points based on the inter-pupillary distance of the observed face.

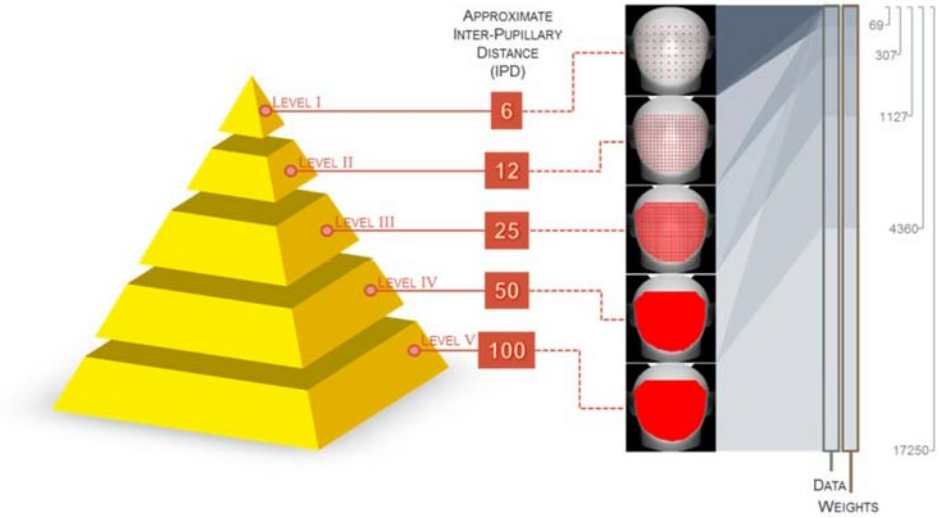


Figure 2: Our representation incorporates a multi-resolution framework to allow for analysis under a variety of resolutions

2. Novel Techniques for Data Recovery

A face acquired under any simultaneous combination of degradations is represented in our model space as a data vector, containing measurement information, and a confidence vector, containing corresponding measurement confidences. We develop multiple techniques to correct, de-noise, and complete the measurements by using machine learning algorithms, trained on very large amounts of similarly incomplete data. The construction and training of such machine learning models is an active field of research with a wide range of potential impact in many disciplines.

We have explored the construction of principal subspaces by developing a novel streaming variant of *Generalized Hebbian Analysis* to enable a weighted least-squares recovery solution to obtain the recovered face. Let us assume that faces occupy a low-dimensional linear subspace of dimension k within the span of all d-dimensional objects, represented by the dxk basis matrix **B**. Let us further assume that

we have a test data vector $\mathbf{m}_t$, with corresponding data weight vector $\mathbf{w}_t$. Since we trust that the basis **B** completely and exactly defines the span of ``legal'' faces, we can compute a weighted projection of the test data onto this space, and then reconstruct the accurate face completely by minimizing the l-2 norm of the weighted reconstruction error $\mathbf{e} = \|\mathbf{w} \odot (\hat{\mathbf{m}}_t - \mathbf{m}_t)\|_2^2$ within the space spanned by the linear face basis **B**:

$$\hat{\mathbf{m}}_t = \mathbf{B}(\mathbf{B}^T\mathbf{W}_t^2\mathbf{B})^{-1}\mathbf{B}^T\mathbf{W}_t^2\mathbf{m}_t$$

For a given basis size k (i.e. number of columns in **B**), traditional PCA aims to find the basis which in which the projected variance of the data is maximized, or equivalently, in which the reconstruction error of any data sample is minimized. It hence aims to solve the low-rank matrix factorization problem (LR-MF):

$$(\hat{\mathbf{B}}, \hat{\mathbf{C}}) = arg\min_{\mathbf{B},\mathbf{C}} \|\mathbf{M} - \mathbf{B}\mathbf{C}\|_F^2$$

In our case, given the missing dimensions in the training data, we are required to solve the corresponding problem of weighted low-dimensional matrix factorization (WLR-MF), i.e.

$$(\hat{\mathbf{B}}, \hat{\mathbf{C}}) = arg\min_{\mathbf{B},\mathbf{C}} \|\mathbf{W} \odot (\mathbf{M} - \mathbf{B}\mathbf{C})\|_F^2 + \lambda_B\|\mathbf{B}\|_F^2 + \lambda_C\|\mathbf{C}\|_F^2$$

In order to construct a robust technique to find a minimizer for this objective which is relatively low on computational complexity and memory requirements, we first analyze a neuron-inspired stochastic streaming solution to the unweighted LR-MF problem based on the field of *Hebbian learning*. Consider a single-element neural network containing d inputs into a linear weighting engine with weights **b** leading to a single output y. We can consider a continuous stream of data inputs $\mathbf{m}_i$'s entering the system, and our task is to design an update rule for the weights **b** in order to maximize the variance of the expected output c (this is the equivalent objective of LR-MF).

We extend the GHA framework to solve the WLR-MF problem, by analyzing the relationship between the Hebbian update system and the ALS technique for computing the eigenbasis. We can consider the Hebbian update rule to also be an example of a similar alternating update rule, where the coefficient update remains the same, but the basis update is replaced by the new stochastic update rule which analyzes single data elements at a time rather than the entire data matrix. The kxk matrix inversion step disappears when there is no missing data. We hence arrive at the new Weighted GHA Rule, which is able to compute a solution to the WLR-MF problem, by altogether avoiding the nxn matrix inversion. The corresponding update rules for Weighted GHA are:

$$\mathbf{c}_n = (\mathbf{B}_n^T\mathbf{W}_{n+1}^2\mathbf{B}_n + \lambda_C\mathbf{I})^{-1}\mathbf{B}_n^T\mathbf{W}_{n+1}^2\mathbf{m}_{n+1}$$

$$\mathbf{B}_{n+1} = \mathbf{B}_n + \eta(\mathbf{w}_{n+1} \odot (\mathbf{m}_{n+1} - \mathbf{B}_n\mathbf{c}_n))\mathbf{c}_n^T$$

The resulting reduced dimension space, as specified by the basis matrix **B**, provides us with a compact, useful representation for any input face, allowing for both analysis and synthesis.

l2-based recovery techniques have been shown to suffer from inconsistency problems when dealing with incomplete data, particularly as the fraction of observable data decreases. The l1-minimization

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

technique returns a sparse solution in the coefficient space, i.e. **m** = **Bc**, where **c** contains very few non-zero items, hence representing the data using a small subset of basis vectors. This has proven to be much more resilient to large amounts of missing data (i.e. small data weights). To make optimal use of such techniques, we have also explored the construction of sparsity-inducing dictionaries from incomplete data by developing a modification of the K-SVD algorithm. The K-SVD dictionary learning algorithm aims to find a basis in which data elements have a sparse expansion. Typically, such a basis is overcomplete, so as to better model the space and enforce sparse basis expansion. However, this is not a necessary in the general sense; the basis only needs to be overcomplete with regards to the (linear) subspace that the data lies in. Traditional K-SVD aims to solve the following variant of the LR-MF problem outlined earlier: Given a data set {**m**}, find a basis representation **B** of size (i.e. number of columns) k, which best explains the data, while also enforcing a level of sparsity in the basis expansion:

$$(\hat{\mathbf{B}}, \hat{\mathbf{C}}) = arg\min_{B,C} \|\mathbf{M} - \mathbf{BC}\|_F^2 \qquad \text{s.t. } \|\mathbf{c}_{(\mathbf{i},\cdot)}\|_0 \leq \alpha \ \ \forall\, i$$

In traditional K-SVD, the algorithm alternates between a sparse coding step and a dictionary update step. In a typical alternation fashion, the sparse coding step solves for the coefficients given a current estimate of the basis, while the dictionary update step does the converse. In the case of missing/weighted data, we develop a similar algorithm for K-SVD, which we call WK-SVD, by updating both sparse coding as well as dictionary update steps. The algorithm is now intended to solve the weighted version of the problem, namely:

$$(\hat{\mathbf{B}}, \hat{\mathbf{C}}) = arg\min_{B,C} \|\mathbf{W} \odot (\mathbf{M} - \mathbf{BC})\|_F^2 \qquad \text{s.t. } \|\mathbf{c}_{(\mathbf{i}\cdot)}\|_0 \leq \alpha \ \ \forall\, i$$

Algorithm 1 The WK-SVD Algorithm

initialize: the dictionary matrix $\mathbf{B}$ to k elements of the data.

repeat

Sparse Coding Stage:

Use OMP (or other pursuit algorithm) to approximate the solution of

$$\hat{\mathbf{C}} = arg\min_{C} \|\mathbf{W} \odot (\mathbf{M} - \mathbf{BC})\|_F^2 \qquad \text{s.t. } \|\mathbf{c}_{(\mathbf{i},\cdot)}\|_0 \leq \alpha \ \ \forall\, i$$

Dictionary Update Stage:

for $i = 1, \ldots, k$ **do**

Identify the set $\mathbf{\Lambda_i}$ of data items which have a non-zero support in the i^{th} dictionary atom.

Compute the representation error for all data items in the support of this dictionary atom

$$\mathbf{E}_i = \mathbf{M}_{(\cdot,\mathbf{\Lambda_i})} - \sum_{i \neq j}^{k} \mathbf{b}_j \mathbf{c}_{(\mathbf{j},\mathbf{\Lambda_i})}$$

Book-keeping: Identify the set $\mathbf{\Gamma_i}$ such that $\sum_{l \in \Lambda_i} \mathbf{w}_{(\mathbf{\Gamma_i},\mathbf{l})} \neq 0$

repeat

PowerFactorization:

$$\hat{\mathbf{c}} = (\hat{\mathbf{b}}^T \mathbf{W}^{\mathbf{2}}_{(\mathbf{\Gamma_i},\mathbf{\Lambda_i})} \hat{\mathbf{b}} + \lambda_C \mathbf{I})^{-1} \hat{\mathbf{b}}^T \mathbf{W}^{\mathbf{2}}_{(\mathbf{\Gamma_i},\mathbf{\Lambda_i})} \mathbf{M}_{(\mathbf{\Gamma_i},\mathbf{\Lambda_i})}$$

$$\hat{\mathbf{b}} = (\hat{\mathbf{c}}^T \mathbf{W}^{\mathbf{2}}_{(\mathbf{\Gamma_i},\mathbf{\Lambda_i})} \hat{\mathbf{c}} + \lambda_B)^{-1} \hat{\mathbf{c}}^T \mathbf{W}^{\mathbf{2}}_{(\mathbf{\Gamma_i},\mathbf{\Lambda_i})} \mathbf{M}_{(\mathbf{\Gamma_i},\mathbf{\Lambda_i})}$$

$$\hat{\mathbf{b}} = \hat{\mathbf{b}} / \|\hat{\mathbf{b}}\|$$

until converged.

Update $\mathbf{b}_{(\mathbf{\Gamma_i},\mathbf{i})} = \hat{\mathbf{b}}$, $\mathbf{c}_{(\mathbf{\Gamma_i},\mathbf{i})} = \hat{\mathbf{c}}$

end for

until converged.

3. Off Angle Face Recognition

Off Angle Face Recognition

Perhaps the most commonly observed type of degradation that we analyze is that induced by projection of the 3D face onto a 2D image plane, such as when observed by a standard camera. The orientation of the camera with respect to the face determines the 3D pose angle of the observed face. Real-world images are seldom frontal mugshots, rather there is a significant 3D pose variation observed (i.e. rotation in pitch 0, yaw ¢ and roll 1/, as shown in Figure 3).

Figure 3: 3D pose angles and their induced effects. (L to R) Definition of the pitch 0, yaw ¢ and roll 1/ angles used, three different pose angles of the same subject, showing self-occlusions and appearance variations in different parts of the face.

The input image is first aligned using the facial landmarks to eradicate the effects of scaling, translation and roll 1/. Dense facial correspondence locations are then obtained using a Thin Plate Spline (TPS) warping technique, and the x, y, and grayscale intensity values are measured. The image formation model can be depicted using homogenous coordinates as:

$$\begin{bmatrix} \mathbf{x}'^T \\ \mathbf{y}'^T \end{bmatrix} = \mathbf{P}\mathbf{R}_{(\theta,\phi,0)} \begin{bmatrix} \mathbf{x}^T \\ \mathbf{y}^T \\ \mathbf{z}^T \end{bmatrix}$$

for all observed 2D facial correspondences (x^1, y^1) in the image, which are modeled by true 3D facial correspondences (x, y, z). P in this case represents a camera projection model, and $R_{(0,<f,O)}$ represents a 3D rotation by pitch 0 and yaw ¢ (and zero roll). The rotation matrix can be decomposed as a product of three corresponding independent rotation matrices:

$$\mathbf{R}_{(\theta,\phi,\psi)} = \mathbf{R}_{\theta}^{p}\mathbf{R}_{\phi}^{y}\mathbf{R}_{\psi}^{r}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & -sin(\theta) \\ 0 & sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} cos(\phi) & 0 & sin(\phi) \\ 0 & 1 & 0 \\ -sin(\phi) & 0 & cos(\phi) \end{bmatrix} \begin{bmatrix} cos(\psi) & -sin(\psi) & 0 \\ sin(\psi) & cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Each of these matrices captures the rotation of the viewpoint around a particular orthogonal axis.

We assume the degradation parameters are measured accurately either through manual input or a commercial pose estimator. We also negate roll by an image rotation based upon landmarks or the input pose to reduce R^r to an identity matrix. We make use of the generic depth model assumption as we have done in 3DGEM. This gives a value for the z coordinates in the representation. This can then be used to compute estimates of the derotated x and y coordinates as:

$$\begin{bmatrix} \mathbf{x}'^T \\ \mathbf{y}'^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & -sin(\theta) \\ 0 & sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} cos(\phi) & 0 & sin(\phi) \\ 0 & 1 & 0 \\ -sin(\phi) & 0 & cos(\phi) \end{bmatrix} \begin{bmatrix} \mathbf{x}^T \\ \mathbf{y}^T \\ \hat{\mathbf{z}}^T \end{bmatrix}$$

In order to use these measurements in our representation, we also need to have a set of observation confidences, w, to use our data completion techniques (WGHA and WKSVD). We leave the confidences for the z coordinates to be low since we do not have actual measurements but are operating under our generic depth assumption. The confidences for the x and y coordinates, w_x and w_y, can be determined from the yaw angle. The larger the yaw angle, the less confident we are in these coordinates. Therefore, we use a sigmoid to model this relationship with the confidence going to 0 as the yaw goes to 90 degrees as shown below.

$$\mathbf{w}_x = 1 - (1 + e^{(A_x - |\theta|)/S_x}), \quad \mathbf{w}_y = 1 - (1 + e^{(A_y - |\phi|)/S_y})$$

The values were empirically determined as $A_x = 45$, $A_y = 30$, $S_x = 10$, $S_y = 5$. In order to estimate the confidence in the texture value, we have to measure the vertex visibility at each point. As the face changes pose, certain points become self-occluded and therefore not visible, while others become more visible. We compute the vertex visibility using the generic depth assumption we made before. This allows us to compute the surface normal on all triangles in the mesh adjacent to a vertex. The average of these normal is used as the surface normal of the vertex as shown in Figure 4.

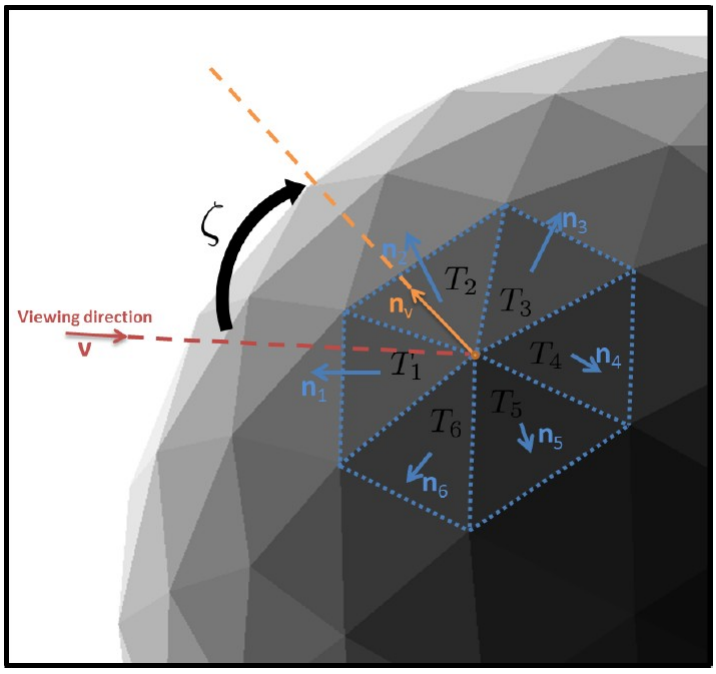


Figure 4: Computation of vertex visibility. Vertex normals are computed as a weighted average of triangle facet normals. The angle between viewing direction and the vertex normal provides a cue towards vertex visibility.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Using the pose of the face, we can know the viewing angle of the camera and compute the angle between the viewing direction and these normals. We again use a sigmoid relationship to determine the confidence in the texture using this angle.

Once the representation has been computed, WGHA or WKSVD can be used to generate the full representation, allowing us to generate a frontal face for recognition purposes.

Initial Experimental Results

We evaluated our technique by training a subspace model on a large conglomeration of data, which was collected and processed by the lab. This includes data from the MBGC dataset (>37k images), and a dataset of publicly available mugshot images downloaded from mugshots.com (~1k images). All the images were manually annotated with 79 landmark fiducial points. We test the pose correction technique on the CMU Multi-PIE dataset. Some examples of the reconstructed frontal faces can be seen in Figure 5.

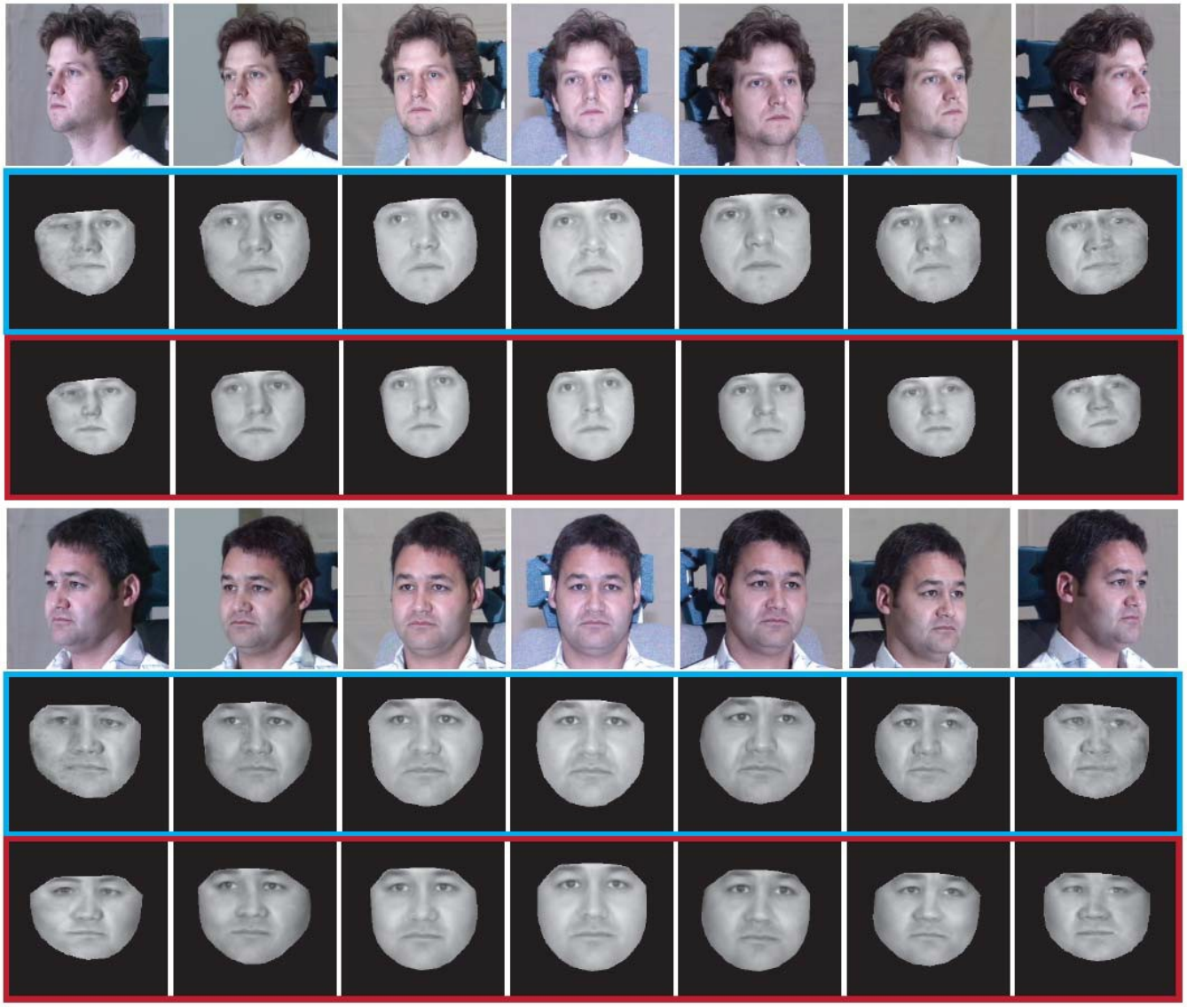
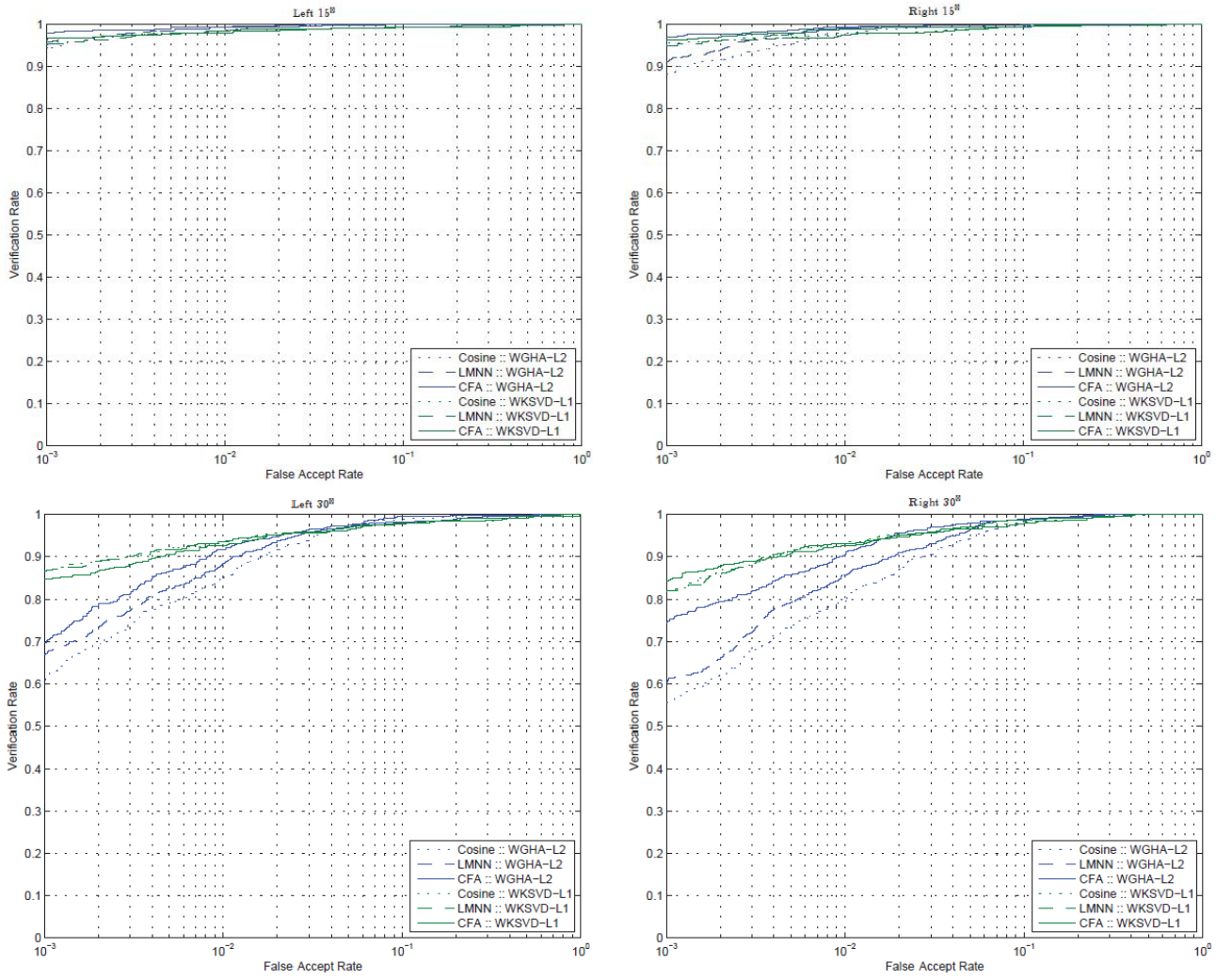


Figure 5: Sample images from the CMU Multi-PIE database demonstrating the facial pose correction capability

We experimented with using normalized cosine distance, Large Margin Nearest Neighbor, and Class Dependence Feature Analysis (CFA) classification schemes. As can be seen in the ROCs below, we achieve high recognition performance even at poses of 45 degrees.



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

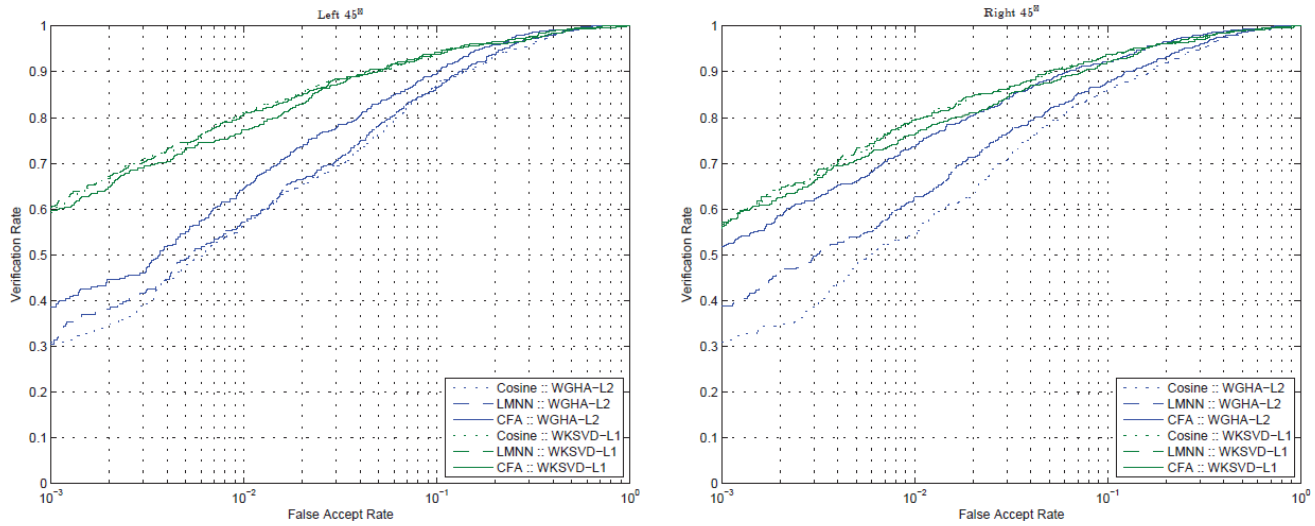


Figure 6: ROCs for yaw angles of 15 degrees (top), 30 degrees (middle), and 45 degrees (bottom) for both left and right angles.

We also compared to two commercial matching systems. The results show that adding this reconstruction as a preprocessing step allows us to greatly improve the matching performance as the pose gets more extreme.

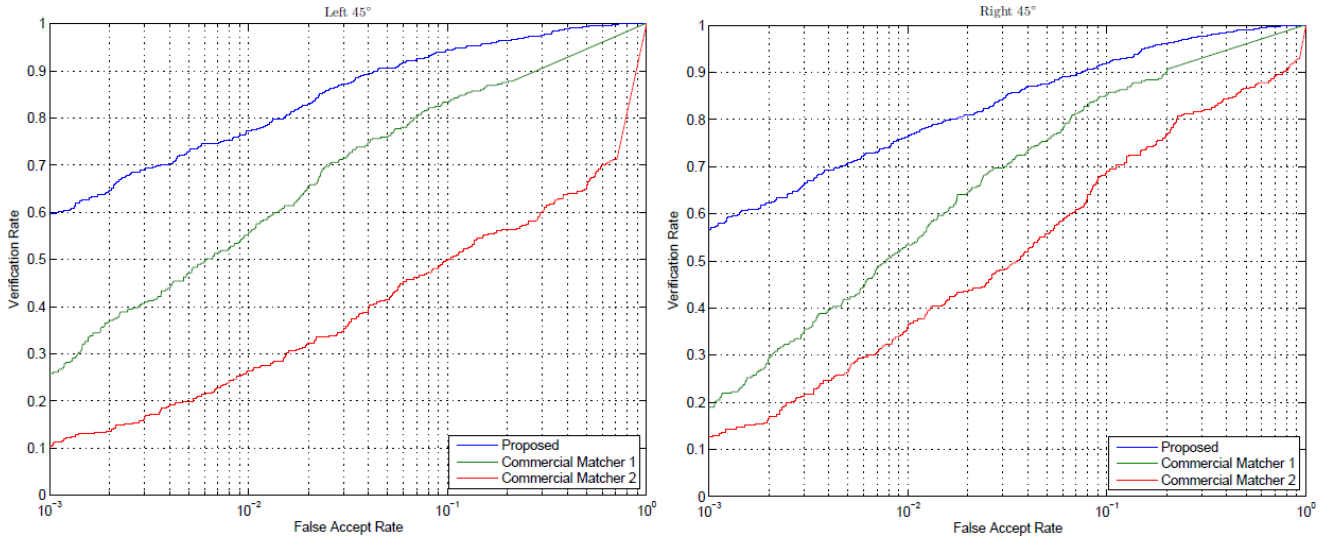


Figure 7: Comparison of commercial matchers (green and red) to our system (blue) at 45 degrees of yaw.

We have also run experiments seeing how a combination of pose and low resolution together impact face recognition. We accomplished this by running the Labeled Faces in the Wild (LFW) unsupervised recognition experiment. The results are shown below. As can be seen, we very well with only two methods outperforming us. Both of these methods, Pose Adaptive Filters and MRF-MLBP, extract very fine tuned features from the data.

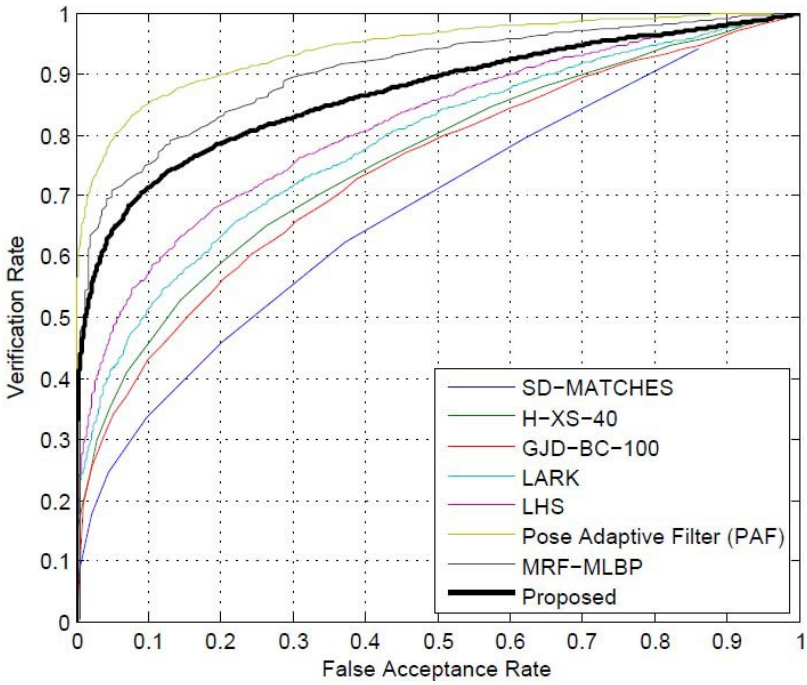


Figure 8: ROC for LFW using our method (black) and many state of the art methods.

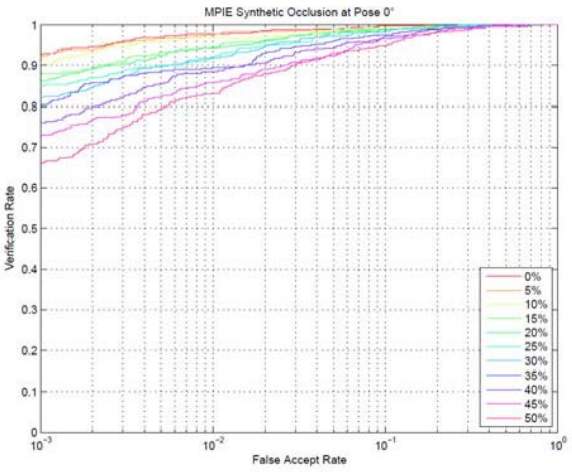
4. Off Angle, Occluded Face Recognition

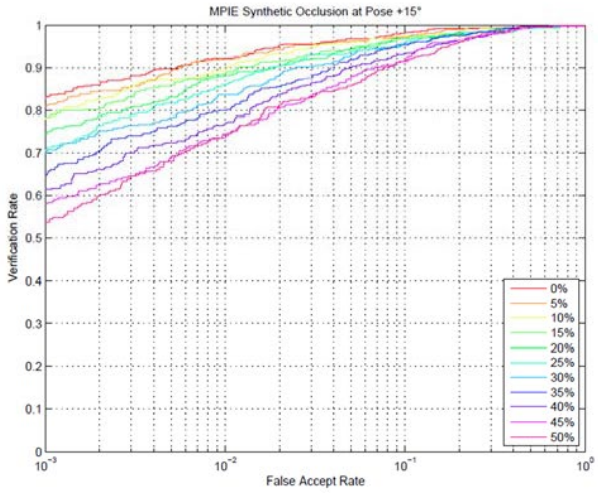
Off Angle, Occluded Face Recognition

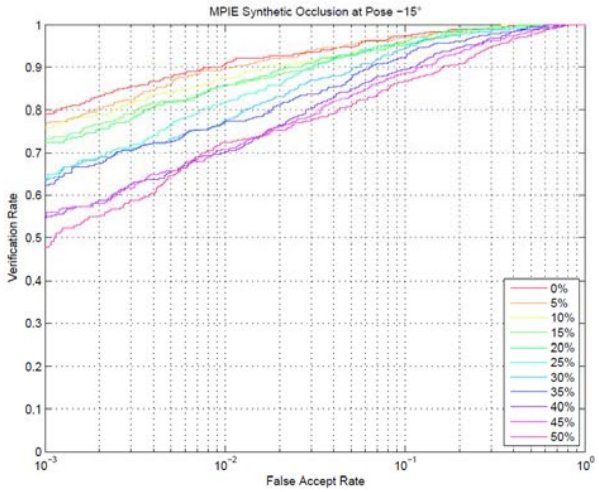
However, due to the nature of the LFW dataset, it is difficult to tell where the algorithm fails. In order to have a deeper understanding of the capabilities and limits of a system that uses our method, we have begun performing experiments on controlled combinations of these factors. So far, we have run experiments using the CMU MPIE dataset combining off-angle face recognition and synthetic occlusions. Since the MPIE dataset has fixed poses at which the face was captured but no occlusions, we need to synthetically add occlusions into the image. Since our representation does not use any region that is marked as occluded, we just need to set the mask for our weighting appropriately. However, to ensure there is no possibility of error, we artificially add another image in place and set the occlusion mask to the same location. Varying levels of occlusion are added from 5% to 50% of the face region covered as seen in Figure 9.

Figure 9: Varying levels of synthetic occlusion added to frontal image

In our experimental setup, the representations from the un-occluded frontal images are enrolled as a gallery set while the recovered representations from the occluded images at a specific pose are used as a probe set. With only the normalized cosine distance as the metric for matching two images, we see in Figure 10. Though we would prefer to be invariant to angle and occlusion, we can see that there is a graceful degradation in performance as more of the face becomes occluded and at different angles. This means that the algorithm will not dramatically fail just because a face under both conditions is given though there is a limit.







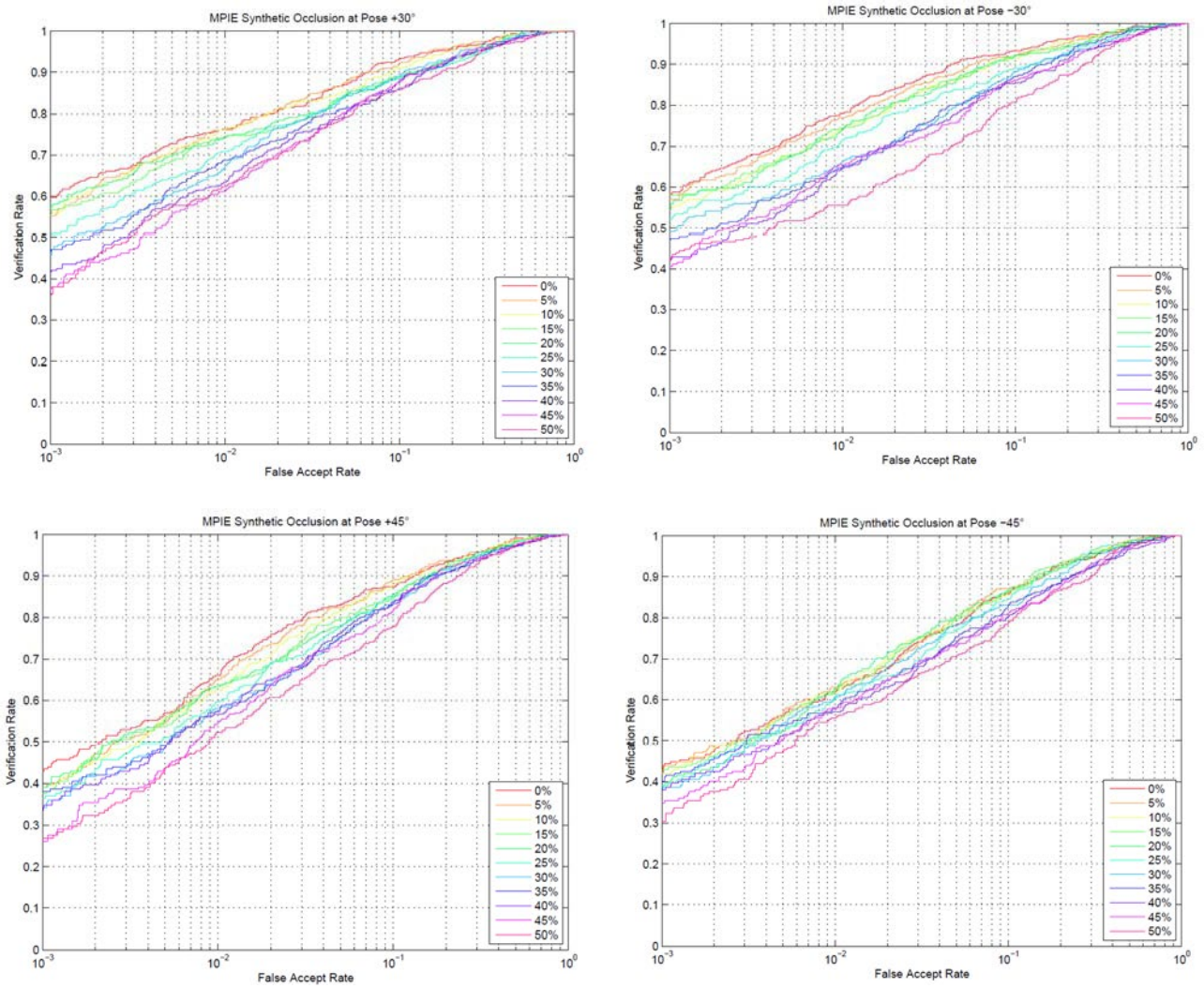


Figure 10: ROCs for verification at varying yaw angles (top to bottom, 0º, ±15º, ±30º, ±45º) and different levels of synthetic occlusion.

5. Low Resolution Face Recognition

By using the same representation, we can also super resolve face images by treating low resolution images as high resolution versions with many missing pixels. By setting the weights in the representation appropriately, we can generate super resolved versions by completing the feature vector. Some examples of super resolved versions from the MPIE dataset are shown in Figure 11. In order to verify that these super resolved images are indeed useful for recognition, we ran both the Pittpatt matcher and the Identix G6 matcher on the low resolution images and the super resolved images. The gallery set were the original high resolution images while the probe set was either the low resolution version or the super resolved version. As seen in Figure 12, Pittpatt shows a very dramatic improvement in using the super resolved faces for recognition while the Identix G6 matcher does not show the same improvement. This is possibly because Pittpatt must first be able to detect a face which is easier on the super resolved images. However, we can also see that when we use our own CFA matcher, we outperform both commercial systems on the low resolution scenario. However, it is difficult to compare all of these fairly as we cannot control the face detection step in Pittpatt whereas our method assumes some user input

as to the location of the face. This helps demonstrate, however, the importance of having some human interaction in this process at the moment as fully automated systems cannot currently handle such cases.

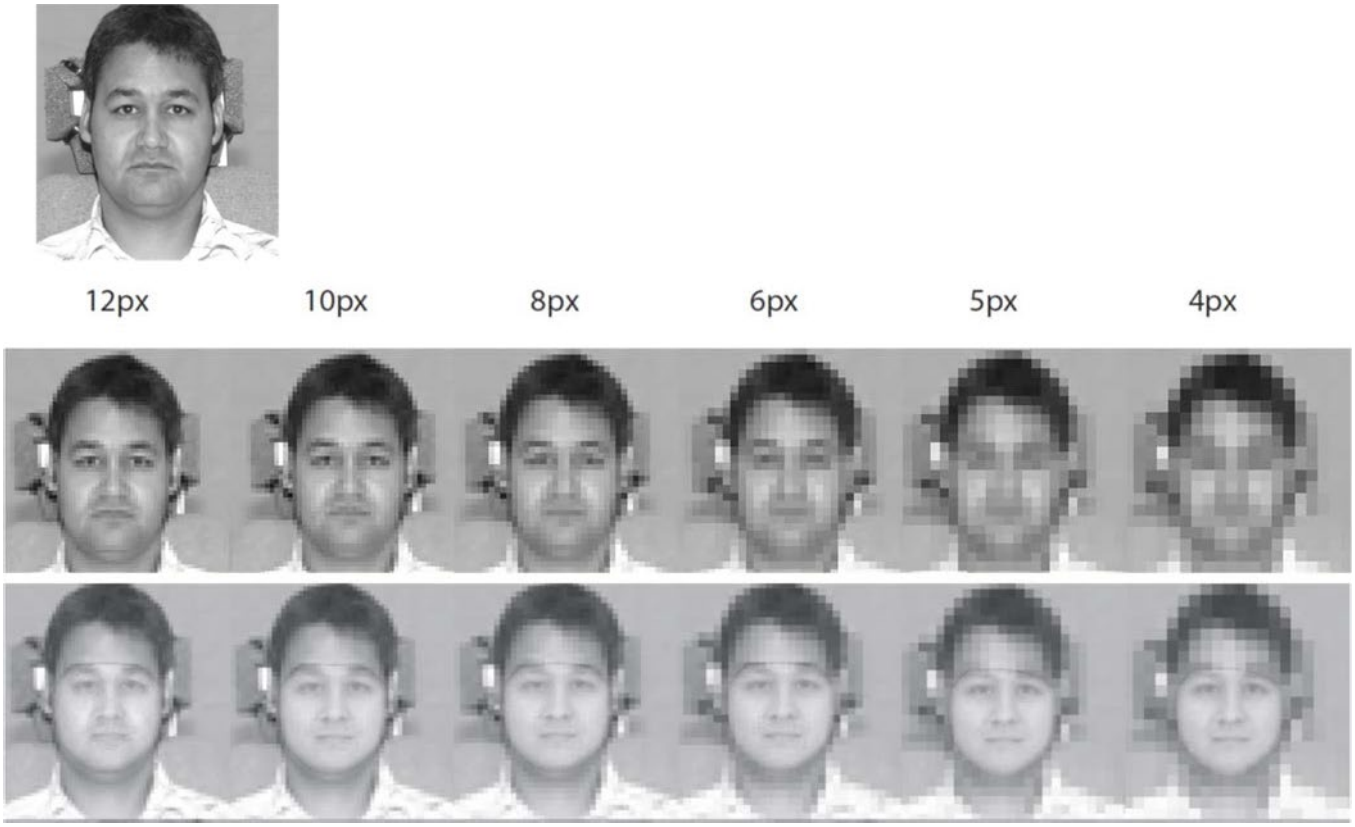


Figure 11: Example of super resolution applied to various levels of downsampling to approximate low resolution inputs. Original image (top), downsampled images (middle row), superesolved versions (bottom row)

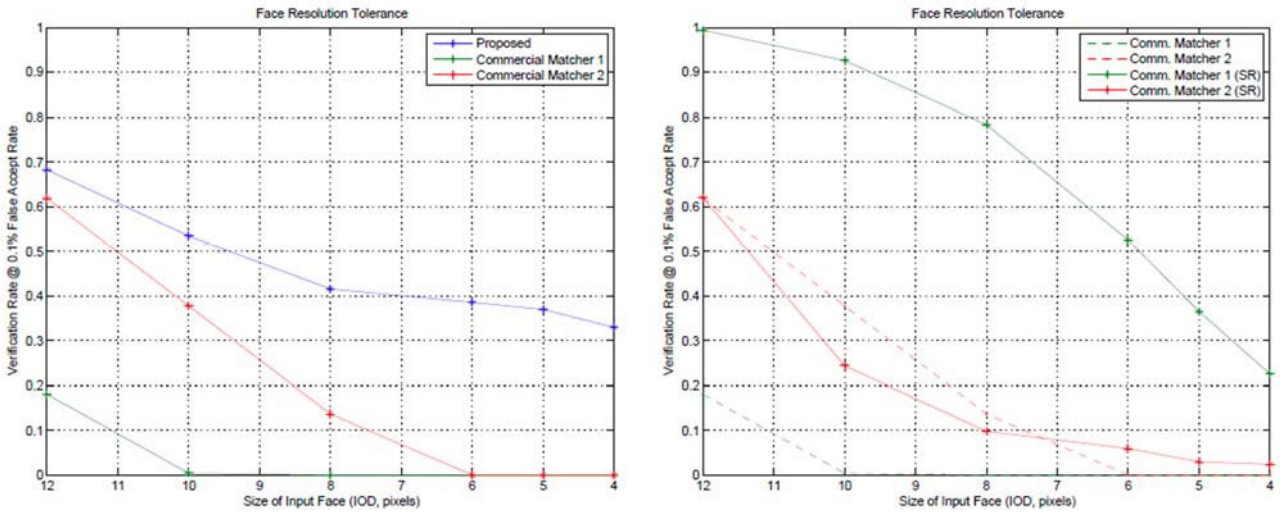


Figure 12: Proposed system vs. commercial systems (Left). Running from the original low resolution images, we outperform both Pittpatt green) and Identix G6 (red). Improvement to commercial systems using super resolution technique (right). Pittpatt (green) shows a large improvement when using the super resolved images (solid line) over the original images (dashed line). Identix G6 (red) does not show an improvement.

5. Using a Shape Free Representation for Occlusion-Removal from Face Images

A Shape-Free Representation for Accurate Texture Analysis

The representation of the face is a critical element in the design of any system; a poor representation can impede subsequent tasks significantly, and could prove to be the weakest link in the processing chain. Since we are interested in obtaining a feature vector from mostly texture information, it is beneficial to isolate the texture from shape when building the subspace. We define "shape" by the x and y coordinates of a specific set of predefined landmarks on a face. The shape information can be discounted by having all faces adopt the same shape before we model the subspace. In other words, for a true shape-free representation, all facial features (eyebrows, eyes, nose, mouth, etc.) should have the same dimensions and locations for all faces. Once this is accomplished, the only discriminative factor left between the shape-normalized faces is the texture. For this purpose, a global transformation, such as an affine transformation of the entire face, is not enough, since it multiplies every pixel of the input image by the same transformation matrix. Instead, local transformations, where every triangle bounded by any three control points can be transformed independently, are needed.

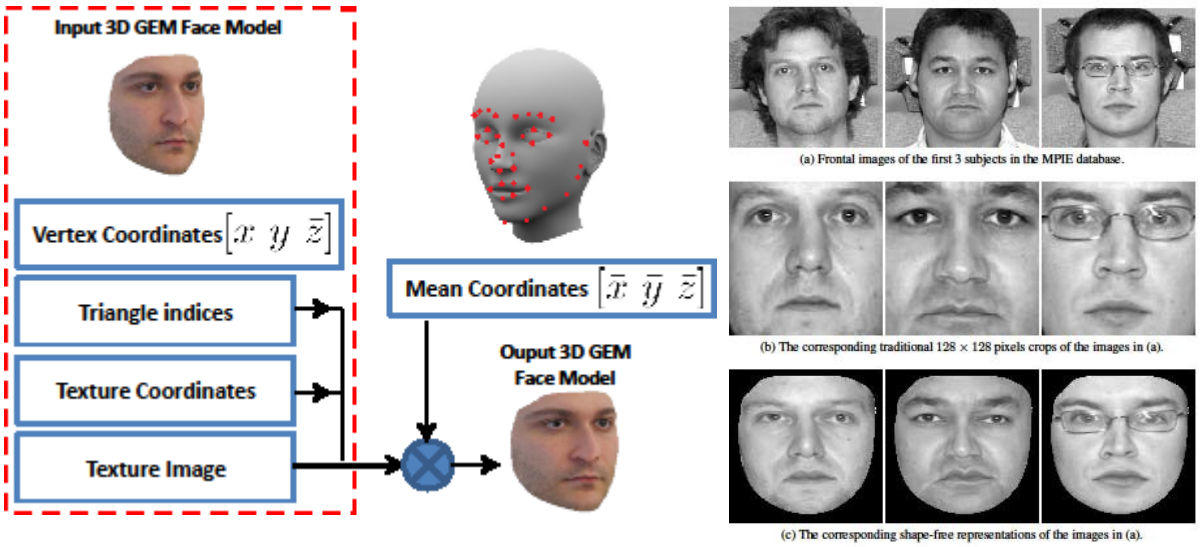


Figure 13: Generation of shape free images. The flowchart on the left depicts the procedure, while the images on the right depict example shape-free images (contrasted against traditional face crop)

Our approach to generate such a shape-free warp of the face devoid of discontinuities makes use of the 3DGEM face modeling framework. The 3DGEM framework generates a 3D of a face from a single 2D image. We build a 3D generic structure and render all face textures with this common structure to normalize the shape of faces as depicted in Figure 13. This will essentially render all the faces given a common shape denominator just like the previous method, but 3DGEM will ensure a much smoother rendering free of high frequency discontinuities and artifacts. Figure 13 also depicts three frontal faces from the MPIE database with a traditional crop using eye coordinates, and their equivalent shape-free

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

representation. With this shape-free representation, we can set out to build a subspace that accurately models all the texture variation.

ℓ_1-norm Sparse Feature Extraction in Subspace with Missing Dimensions

The aforementioned technique enables us to transition between a shape-free representation of the face, and the regular image representation. We construct a subspace-based technique for reconstruction occluded regions of the face, which can be represented as missing dimensions in the shape-free feature vector.

Let the test image be vectorized into a vector **x** of dimensionality d and represented by **x** = **Vc** +**m**, where **V** is the matrix of vectorized eigenfaces, c is the vector of coefficients, and **m** is the vectorised mean face. We compute the eigenbasis **V** by computing the principal subspace over a large training dataset, which has been processed to the shape-free representation. The problem is to now estimate the values of the missing pixels in a partially occluded face by utilizing the subspace. For notation simplicty, let **x**' be the vector of active pixels of **x**. **x'** is of size d' and d' < d pixels. Similarly, let **m**' be the mean of active mean pixels. For notational simplicity, we can also introduce $\mathbf{x}'_c$ = **x'** – **m'**, which represents the centered version of **x'** of size d'. **V'** is the matrix of active rows that are in **V**. We need to solve for the coefficient vector **c**.

We find a sparse solution for the coefficient vector **c** using ℓ_1-minimization. A sparse solution will allow us to represent individual faces in clusters of eigenvector bases, since the face data is more likely to be multimodal, while PCA assumes unimodality. Moreover, in ℓ_1-minimization literature, underdetermined problems are usually the norm rather than the exception, and the number of training bases N can largely exceed d', thereby enabling us to handle large amounts of occlusion on the face. The standard ℓ_1-minimization problem solves the following convex program:

$$\min J(\mathbf{c}) = \|\mathbf{c}\|_1 \text{ subject to } \mathbf{V}'\mathbf{c} = \mathbf{x}'_\mathbf{c}$$

which is known as Basis Pursuit (BP), and finds the vector with smallest ℓ_1 norm of vector c defined as

$$\|\mathbf{c}\|_1 = \sum_{i=1}^{d} |c_i|$$

The intrinsic details of specific ℓ_1 solvers is beyond the scope of this work, we have used an Augmented Lagrange Multiplier (ALM) approach to obtain the coefficients.

While these coefficients contain a compact yet discriminative representation of the individual faces, we are also able to use them to regenerate un-occluded face images. These images can then be processed by most existing facial recognition tools. We conducted our benchmarks by using the popular Pittpatt Face Recognition toolkit, which is widely used by several law enforcement agencies, so as to show real-world application of this technique.

Initial Experimental Results

We evaluated our technique by training a subspace model on a large conglomeration of data, which was collected and processed by the lab. This includes data from the MBGC dataset (>37k images), and a dataset of publicly available mugshot images downloaded from mugshots.com (~1k images). All the images were manually annotated with 79 landmark fiducial points. We test the occlusion recovery

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

technique on the popular AR face database, which contains images of 134 individuals (75 Men, 59 Women) under various conditions, including wearing sunglasses and scarves. We provided manual occlusion annotations and fiducial points for the images.

We have conducted several experiments to demonstrate the use of the technique at extending the range of face recognition tools. The first experiment is to de-occluded natural commonly-observed occlusions on the face, namely sunglasses and scarves. Some visual results from this removal are shown below:

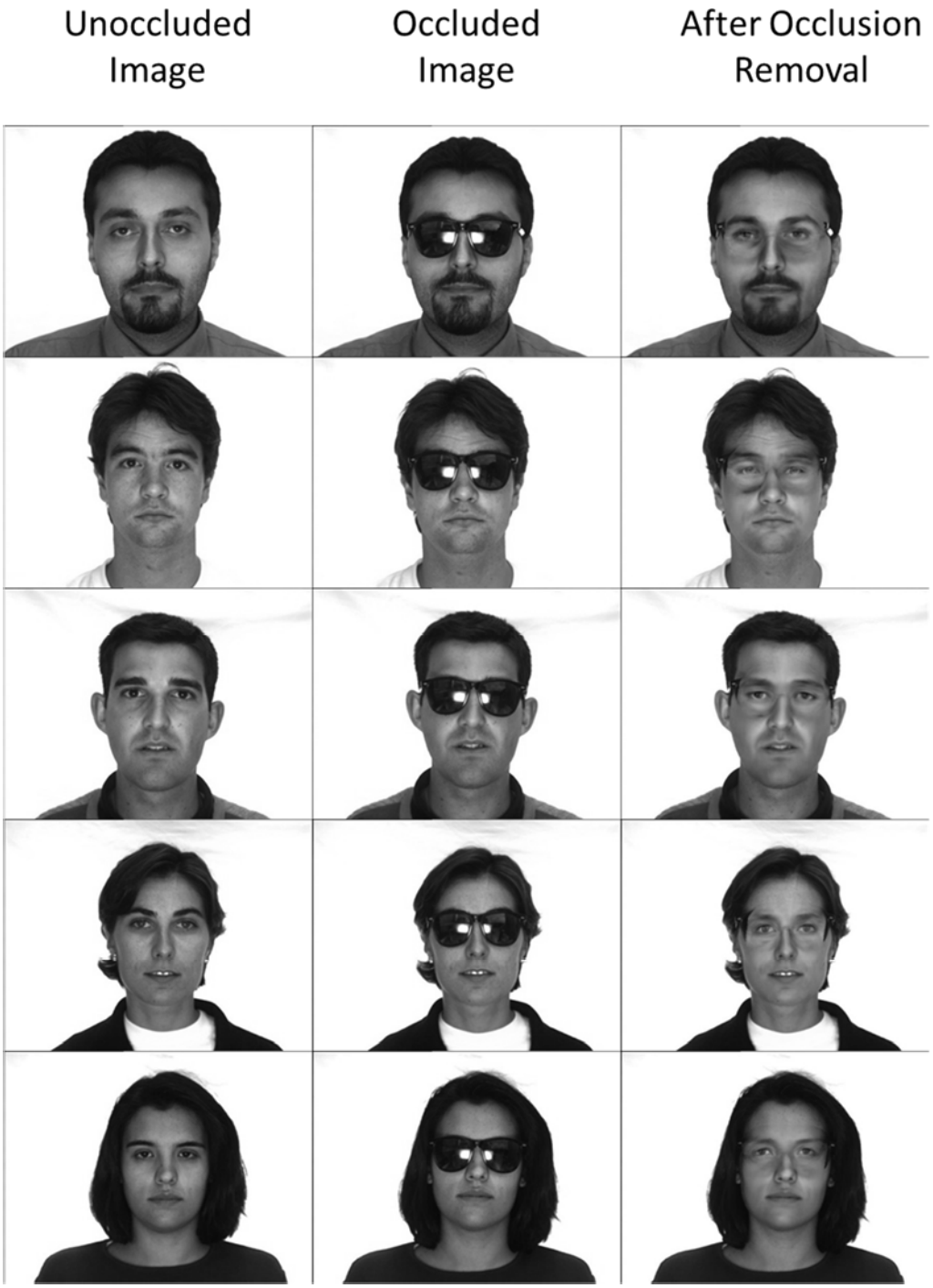
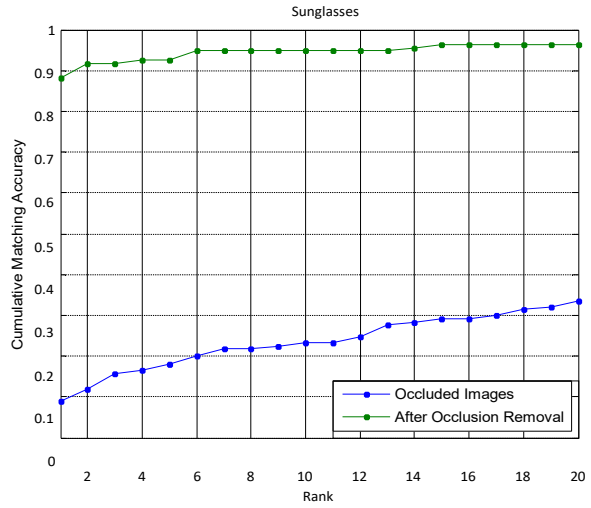


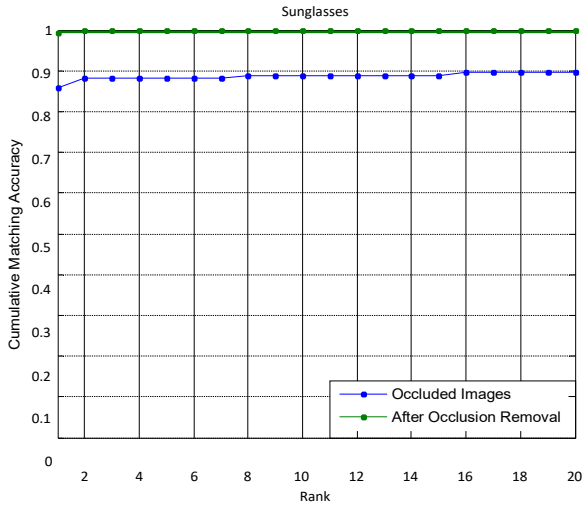
Figure 14: Results from the “sunglasses” experiment

Figure 15: Results from the “scarf” experiment

We also conducted a recognition experiment, using the widely popular commercial Pittpatt face recognition engine. The gallery in this experiment contains one unoccluded image of each individual. The probe images are occluded and de-occluded images under the two occlusions. The resulting Cumulative

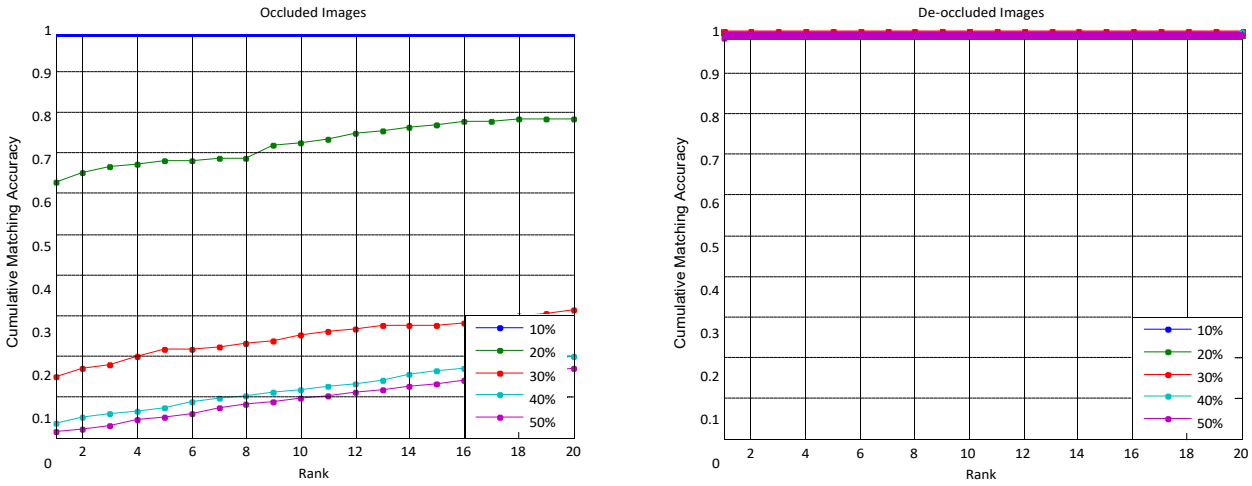
Match Characteristic curves clearly demonstrate the efficacy of the system at boosting recognition performance under severe occlusions:





In order to better understand the performance of the system, and tolerance to increasing levels of occlusion, we also evaluated the system using increasing amounts of synthesized occlusions. For this experiment, we used one un-occluded image of each identity as the gallery, and synthesized many artificially occluded images. The degree of occlusion is determined as the area fraction of the face in the image that is overlaid, from 10% to 50%. The locations of the occluded regions were randomized (with the constraint that the entire occlusion lies inside the face). An example of the synthesized occlusions are shown below:

As is evident from these images, a 50% occlusion by area encompasses a large amount of the discriminatory features of the face, covering most of the eyes, nose, and mouth. These occlusions are recovered by using the proposed technique, and the result is evaluated with the help of the Pittpatt face recognition system. The gallery used is the unoccluded faces as before. The resulting CMC curves for the occluded and recovered images are shown below:



This clearly indicates that the system is robust to a significantly large degree of occlusion, and is able to greatly improve recognition accuracy for these images.

6. Full-Face Hallucination from the Periocular Region

On August 19 2014, the ISIS militant group released a video of the execution of an American journalist James Foley. In the video the terrorist is seen wearing a mask revealing only a small portion of the periocular region. It is a very challenging problem to identify the suspect with simply the small visible portion of the face given the illumination and low resolution conditions. Commercial matchers, as used by law enforcement agencies, are unable to process such a partial face. However, we are able to generate a full face reconstruction based off only the left eye of the suspect. Further, we simulate a facial database search using a commercial matcher. We find that we are able to match the suspect from the ISIS video (using visual features only) to his true identity (confirmed by other sources) as rank-3 in a database of over 1,000,000 subjects.

Our goal is to reconstruct or hallucinate the rest of the face given a part of the face. Out approach is based on the method of dictionary learning in signal processing. Keeping in mind the issues related to dictionary learning, we arrive at the problem of jointly optimizing the learning procedure for two goals. The first is to learn a dictionary of whole faces so as to include prior knowledge about the spatial relationships between the facial features and the features in the partially visible region. The second is to obtain a dictionary in which the reconstruction error for the partially visible region is penalized more than the entire face and both are jointly minimized for the same sparse coefficients. The second condition ensures that the reconstruction is faithful to the biometric information that is actually present.

We propose a simple approach which promotes the approximation coefficients to be jointly shared for the periocular region and the entire face. Our first objective is to learn a dictionary by solving

$$\underset{\mathbf{D},\mathbf{X}}{\text{minimize}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \;\; \text{such that} \;\; \forall i, \|\mathbf{x}_i\|_0 < K$$

However, we would also like to have a low reconstruction error using the same sparse coefficients restricted to the periocular region set Λ. Thus we also desire to solve

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

$$\underset{\mathbf{D}_\Lambda,\mathbf{X}}{\text{minimize}} \|\mathbf{Y}_\Lambda - \mathbf{D}_\Lambda \mathbf{X}\|_F^2 \text{ such that } \forall i, \|\mathbf{x}_i\|_0 < K$$

Combining the two objectives to solve them jointly allows us to force a common K-sparse representation and also provides a trade-off between errors with an efficient algorithmic solution. Our primary problem is therefore

$$\underset{\mathbf{D},\mathbf{X},\mathbf{D}_\Lambda}{\arg\min} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \beta\|\mathbf{Y}_\Lambda - \mathbf{D}_\Lambda \mathbf{X}\|_F^2$$
$$\text{such that } \forall i, \|\mathbf{x}_i\|_0 < K$$

Here β provides a trade-off between the reconstruction error of the periocular dimensions versus the entire face. Obtaining a consistent sparse encoding between the two sets of dimensions allows for a more meaningful reconstruction. This is apparent if one considers the reconstruction procedure. Given a novel periocular image, we would first obtain the sparse representation x in D_Λ. We then obtain the reconstruction using Dx. Using the original K-SVD training method, there was no reason to expect a low reconstruction error in obtaining the entire face. Thus, relationships between periocular and other facial features are not explicitly learned. However, by forcing consistent sparse representations x during training, we optimize for a low reconstruction error for both regions jointly and simultaneously.
Solving the formulation is achieved by a simple rearrangement before using the standard K-SVD as previously observed:

$$\underset{\mathbf{D},\mathbf{D}_\Lambda,\mathbf{X}}{\arg\min} \left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\beta}\mathbf{Y}_\Lambda \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\beta}\mathbf{D}_\Lambda \end{pmatrix} \mathbf{X} \right\|_F^2$$
$$\text{such that } \forall i, \|\mathbf{x}_i\|_0 \leq K$$

This translates to the standard K-SVD problem. In effect the formulation is equivalent to re-weighting dimensions belonging to Λ by $1 + \mathbf{j}\sqrt{3}$. Note that one can easily generalize this framework to include multiple subsets of other dimensions with different weights. Further, this method along with PCA based and K-SVD based methods, is open set thereby enabling reconstruction of any face that is not present in the training set. For convenience, we call this method Dimensionally Weighted K-SVD or DW-KSVD.

We also developed an alternate algorithm called Hierarchical OMP, which heavily utilizes the sparse signal processing algorithm called OMP (Orthogonal Matching Pursuit). The key advantage of Hierarchical OMP over DW-KSVD, is the fact that it does not require retraining of a dictionary given a periocular mask. We also re-run the experiment of identifying the ISIS suspect out of a million mugshot images with the true identified suspect. We show that our reconstructions using DW-KSVD can match the suspect at ***Rank-1*** and ***Rank-2*** out of three independent reconstruction attempts.

The previous approach called DW-KSVD for full-face hallucination was based on the method of joint dictionary learning from sparse signal processing. DW-KSVD was based on the sparse dictionary learning algorithm called K-SVD. K-SVD however, optimized for a single objective function. For robust reconstruction one requires high reconstruction fidelity between the original and reconstructed full-face (goal 1) and the original and reconstructed periocular region (goal 2). DW-KSVD allows one to move the solution towards both goals simultaneously. This helps the final solution to be a more faithful reconstruction.

One limitation of the joint dictionary learning techniques (including DW-KSVD) that, given a new periocular mask or region of interest, complete retraining is required in order to find the optimal mapping from the periocular region to the full face. Although this is guarantees a better performance, during test time, a significant amount of computation is required to learn that map. In order to investigate alternate algorithms which do not require retraining during test time, we propose Hierarchical OMP as a potential alternate algorithm.

Let the training set be $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_n}\}$ and the support of the periocular region be Λ. Thus, $\mathbf{x}_{1|\Lambda}$ is the periocular region. We divide the training set into L distinct non-overlapping segments such that each segment is underdetermined by a manageable amount where OMP's theoretical guarantees still hold. We have

$$\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2 \cup ...\mathbf{X}_L$$

Let $\mathbf{y}_{|\Lambda}$ be a novel testing image that we would like to reconstruct. We present the following algorithm capable of reconstruction from arbitrary unseen Λ.

Hierarchical OMP $(\mathbf{X}, \mathbf{y}_\Lambda, K, \Lambda)$:
$\mathbf{y} \leftarrow \mathbf{0}$ (Initial reconstruction)
$\mathbf{X}_\Omega \leftarrow \mathbf{0}$
while all L segments not processed **do**
  $\mathbf{a}_l \leftarrow \text{OMP}(\mathbf{X}_{l|\Lambda}, \mathbf{y}_\Lambda, K)$ (K Sparse approximation for l_{th} segment)
  $\mu_l \leftarrow \text{supp}(\mathbf{a}_l)$
  $\mathbf{X}_{\Omega|\Lambda} \leftarrow \mathbf{X}_{\Omega|\Lambda} \cup \mathbf{X}^{\mu}_{l|\Lambda}$ ($\mathbf{X}^{\mu}$ implies $\mathbf{X}$ restricted to examples indexed in μ)
end while
$\mathbf{a} \leftarrow \text{OMP}(\mathbf{X}_{\Omega|\Lambda}, \mathbf{y}_\Lambda, K)$
$\mathbf{y} \leftarrow \mathbf{X}_\Omega \mathbf{a}$
return $\mathbf{y}$

Hierarchical OMP can handle arbitrary periocular regions of interest at test time without requiring retraining. However, it does not learn a dictionary and simply utilizes the training set to find which samples can be best used to explain the novel periocular region. One in the first stage it determines the best samples, in the second stage it uses those samples to find an even better approximation. The full face of the samples are then used to reconstruct the full face.

A core component of Hierarchical OMP is OMP. In order to make sure that the sparse representation is accurate in terms of reconstruction error, OMP is operated well-within the phase transition boundaries. However, those boundaries have been established for random Gaussian measurement matrices and not structured matrices such as dictionaries. Since neither theoretical nor empirical results exist for reconstruction from structured measurement matrices, we leave a large enough margin for operating point of OMP on the phase transition graph from the phase transition boundary.

Experimental Results

We ran an experiment with the image of "Jihadi John" from the video ISIS released. The entire experiment has two major parts: hallucinating the full face from the heavily occluded facial image, and facial identification. We detail each of the two components below.

Full Face Hallucination from Heavily Occluded Face

Figure 16 depicts the process flow to reconstruct the full face from the periouclar region.

Figure 16 Process flow for the reconstruction/hallucination of the full face of the suspect from just the sample video frame. The frame is first cropped and then de-rotated to obtain a near frontal orientation. Illumination normalization then helps in removing some of the effects of shadow. We identify a small usable periocular region which seems to contain the most amount of biometric identity. We then hallucinate the full face using DW-KSVD.

The DW-KSVD method is used to reconstruct the full-face from the small periocular region that is depicted in the figure as the "usable preiocular region". A total of three independent attempts were made to choose a trust region leading to three reconstructions; however, we find that the three reconstructions are similar to each other. This shows reproducibility of our method. The three sparse dictionaries were trained on a separate database of 200,000 faces all of whom were cropped with the same mask as the three independent periocular masks.

Face Identification

Once we have the three reconstructions from the previous experiment, we re-run the face identification experiment, this time with the true suspect (confirmed from other sources) Mohammad Emawazi. To create the gallery set we add in the suspect's frontal image available on the Internet, into the database of 1,000,000 mugshots.

Following the previous report, we re-run a commercial face matcher (Pittpatt) with the three independent reconstructions as the query images and the (1 million + 1) mugshot images as the gallery database. The top 10 matches are shown in the Figure below. As can be seen from the figure, we are able to identify the suspect in ***Rank-1*** from a database of 1 million images. Moreover, the other two attempts matched at ***Rank-2*** and ***Rank-1*** respectively illustrating the robustness of the method.

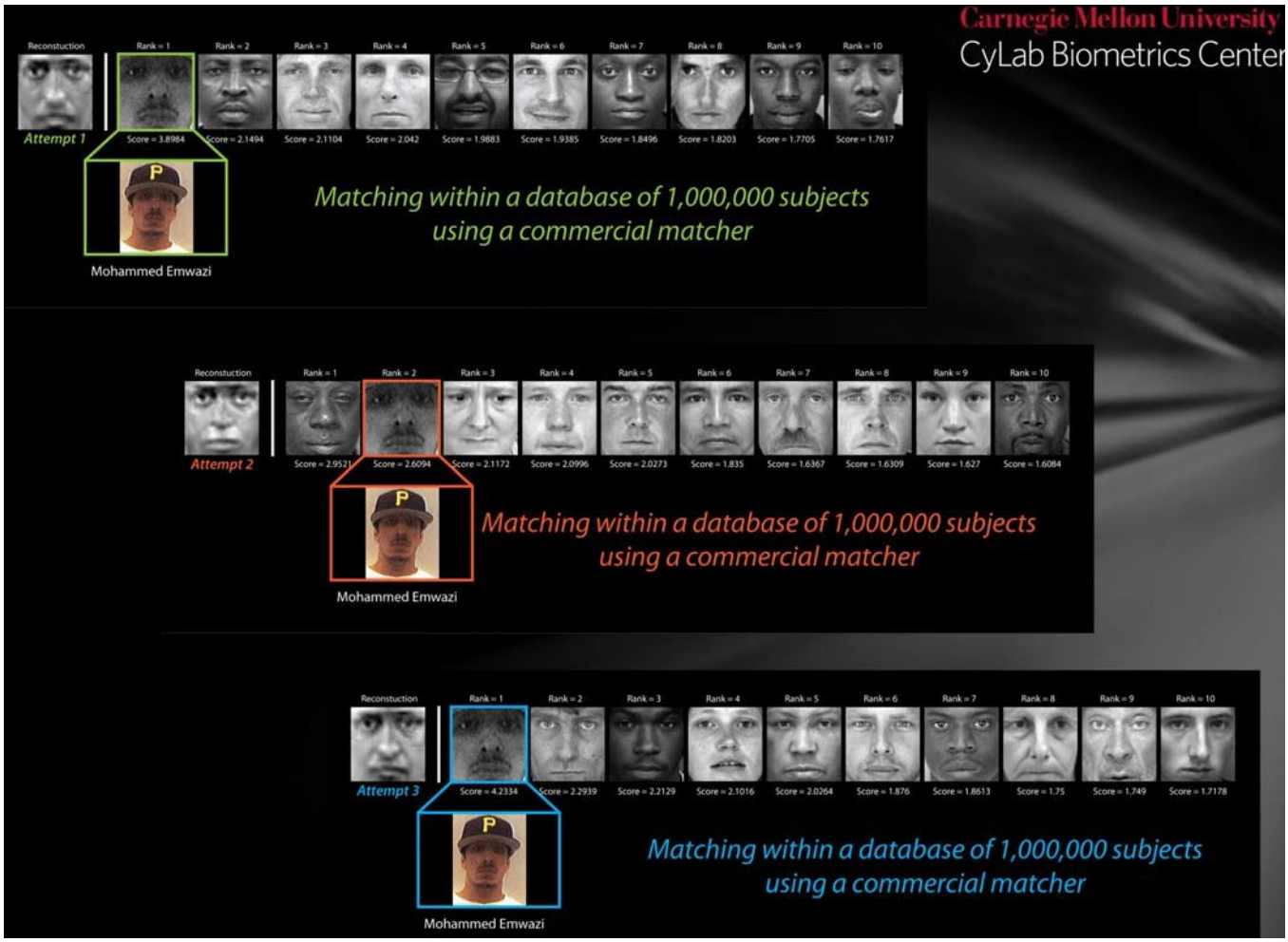


Figure 17: Rank lists of the top matches using a commercial matcher returned when the query images are the hallucinated images. The database or gallery contains more than 1,000,000 subjects. We identify the suspect as the top-match in two of the three attempts

Experimental Results (New Jihadi John)

There was another video with a new "Jihadi John" suspect. We have repeated the process form before with our recovery methods to generate a full face from the periocular region. We reconstructed the face using a dictionary with facial hair present and one without facial hair to show both possibilities. Similarly, if we have more information about the suspect, we can tailor the dictionary to any extra information to generate more possible faces. As we don't know who the suspect is, we cannot run recognition experiments though we plan on doing so as soon as his identity becomes known to verify our hallucinated face. As can be seen below, the recovered face is still very faithful to the visible region and is a potential face for the suspect under the mask. Unfortunately, we cannot verify the accuracy of this reconstruction at the current time.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

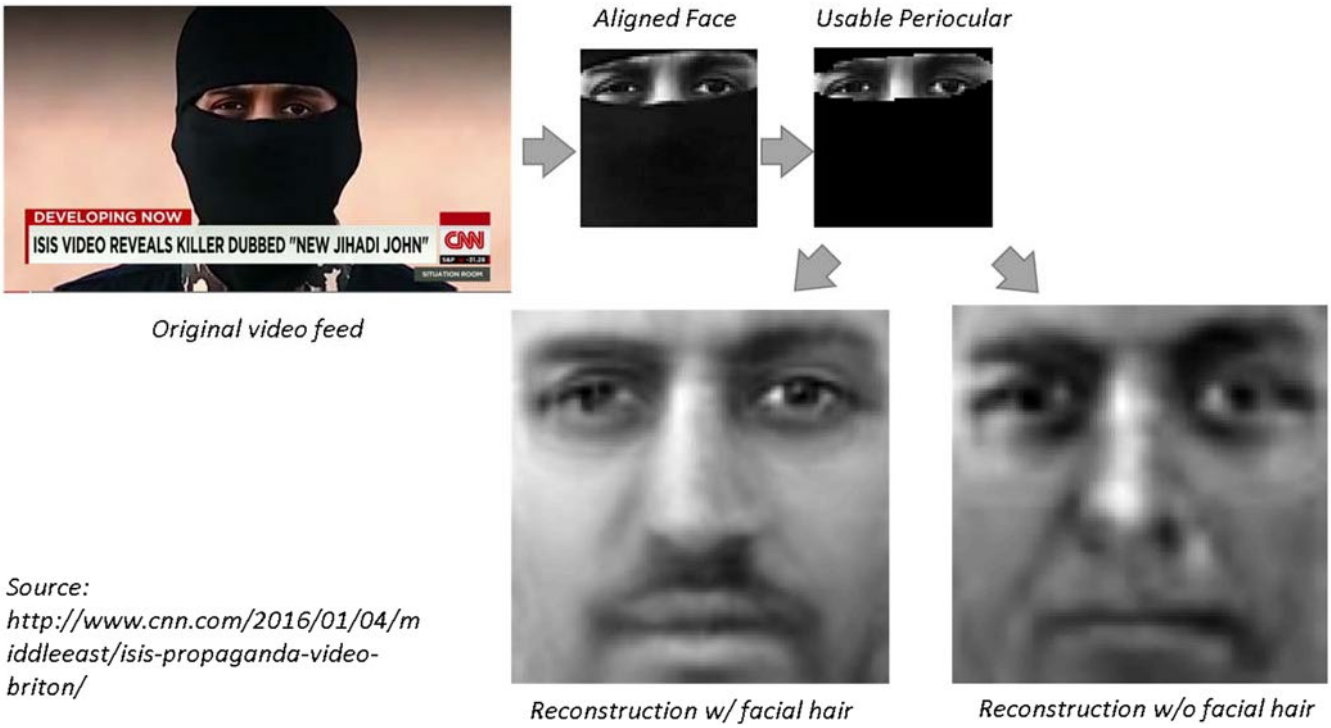


Figure 18: Reconstruction workflow of new "Jihadi John" showing original image, aligned face image, masked out periocular region, and reconstruction with and without facial hair. No illumination normalization was necessary on this image.

7. Full Craniofacial 3D modelling

Our current representation works well for the interior portions of the face which is what is mostly used by commercial applications as well as our own internal matchers. However, there is a possibility that more information is needed by another system or by an end user who wishes to visually inspect the 3D models. To that end, we have started work on developing a full craniofacial 3D modelling technique that will allow us to use as much information as possible from the original images.

Generating Shape

Our approach is based on using a generic full craniofacial 3D model and computing a camera projection matrix from the 3D space to the image. In order to compute the projection matrix, **M**, we need to have a set of correspondences between the 3D model and the 2D image plane. We accomplish this by hand selecting 79 vertices on the 3D model corresponding to the 79 landmark points we can automatically detect on a 2D image of a face. By using these as correspondences between the 3D model and the image plane, we can formulate the relationship between the homogenous versions of these points as

$$p_c \cong Mp_w$$

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

where p_c is the homogenous point in the camera's image plane, p_w is the homogenous point in the world coordinate system (i.e. the 3D model coordinate system) and M is the 3x4 camera projection matrix. Since these are all in homogenous coordinates, the equality can only be defined up to a scale change. Since both sides represent vectors in the same direction however, we can show the cross product of the two must be equal to 0 exactly. This allows us to rewrite the set of equations as

$$Pm = O$$

Where P is a 2Nx12 matrix containing the coefficients from rearranging the cross product of the vectors and m is a vector containing all elements of M. From this we can compute M as the row null space of P. However, since we have more than 12 rows, there will most likely not be a null space. This occurs because we have more than the minimum number of points needed but this does allow us to be tolerant to some noise in the measurements of the landmarks. By computing the SVD of P, we can find the singular vector corresponding to the smallest singular value. This is the solution that gets us closest to satisfying all constraints. This gives us the camera projection matrix.

Once we have computed M, we need to readjust the 3D model to properly fit the points on the subject in the image. Otherwise, every subject would have the same 3D model in the end. To accomplish this, we can use M to compute a ray in 3D space from the camera origin through each landmark point in the image plane. The camera center is embedded into the camera projection matrix as $C = -P_{3x3}^{-1} p_4$ where $P = [P_{3x3}|p_4]$ or P_{3x3} is the first 3 columns of P and p_4 is the last column of P. The optical ray through the camera center and any image point, p_c is

$$R = \begin{pmatrix} C \\ 1 \end{pmatrix} + A \begin{pmatrix} P_{3x3}^{-1} p_c \\ 0 \end{pmatrix}$$

By adjusting A we can compute all points on this ray. For each landmark on the 3D model, we can compute the closest point on the corresponding ray and use that as the new point that the 3D model should have. This way, all landmarks will image onto their correspondences exactly. Once we have these new 3D points, we use a thin plate spline interpolation to generate the new full craniofacial 3D model that is specific to this image.

Generating Texture

In order to texture the new 3D model, we use the camera projection matrix to project all 3D points onto the image plane. We can then determine the texture value by interpolating it based on the pixel values in the image. However, this will necessarily texture models with the same values when they are occluded by other points on the model since we lose all information about depth as can be seen in Figure 19.

Figure 19: 3D Reconstruction showing self occlusion problem

To account for this, for each vertex in the model, we look at the points that projected to a close region around it and find the triangles in the 3D model comprised of those points, excluding any containing the original vertex itself. We then compute a ray-triangle intersection test for each of these triangles and the ray going from the camera center to the vertex. If any of them intersect, we know that the point would be occluded in the projection and so we can fill it with black instead. When we do this, the new 3D model makes much more sense. The original image along with landmarks and the resulting 3D model from various views are shown in Figure 20.

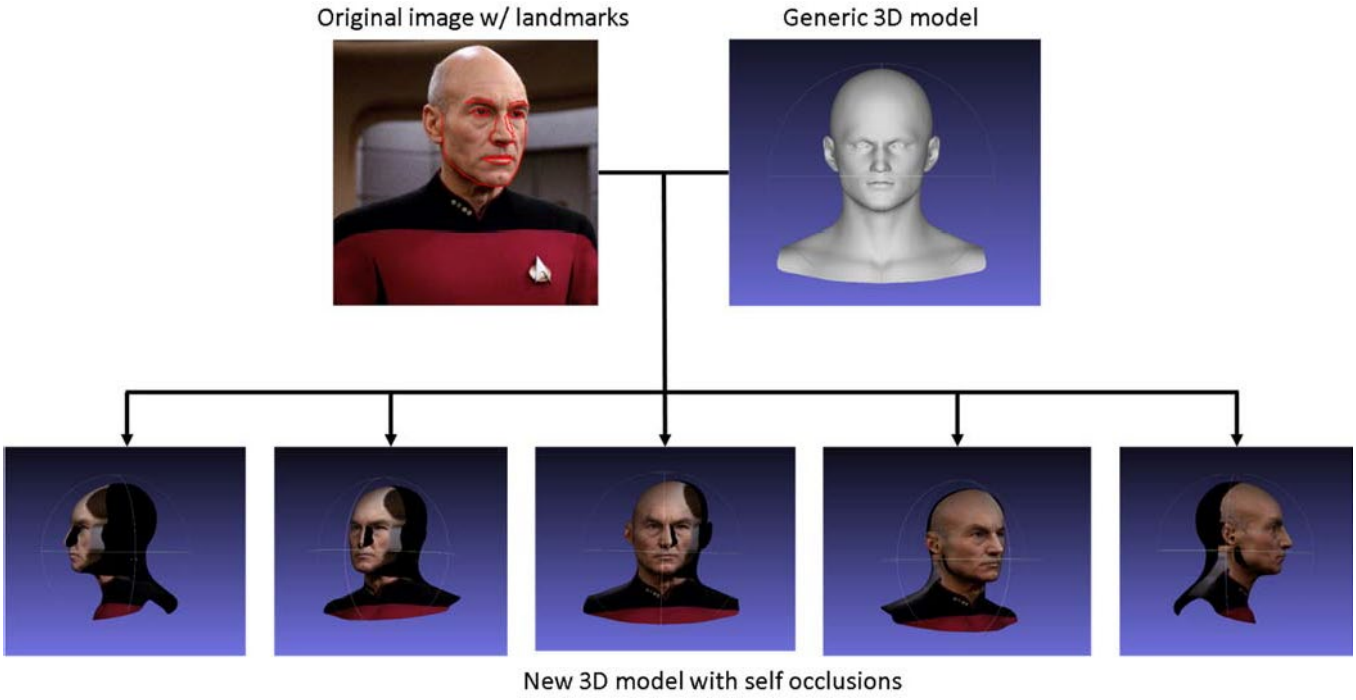


Figure 20: 3D model at various views (bottom) generated from original image, landmarks, and generic 3D model (top)

As can be seen, the missing regions need to be filled in. This can be accomplished with techniques we have already laid out previously. We plan on investigating the best way to fill in the missing portions in the future.

Landmark Free 3D Modeling

All of our previous work relied on an accurate set of 2D landmarks being provided along with the image. Without the accurate landmarks, the 3D modeling and the data completion gave poor results. Additionally, as the pose of the face became more extreme, there was a higher probability that automatic landmarking would fail as well as many of the landmarks becoming occluded by the face itself. In order to alleviate this problem and automate the 3D modeling process, we have been performing research into using deep networks to estimate the same TPS and camera projection parameters needed. By using a deep network as shown in Figure 21, we can warp a generic model and render a frontal image without the need for explicit landmarks to be given.

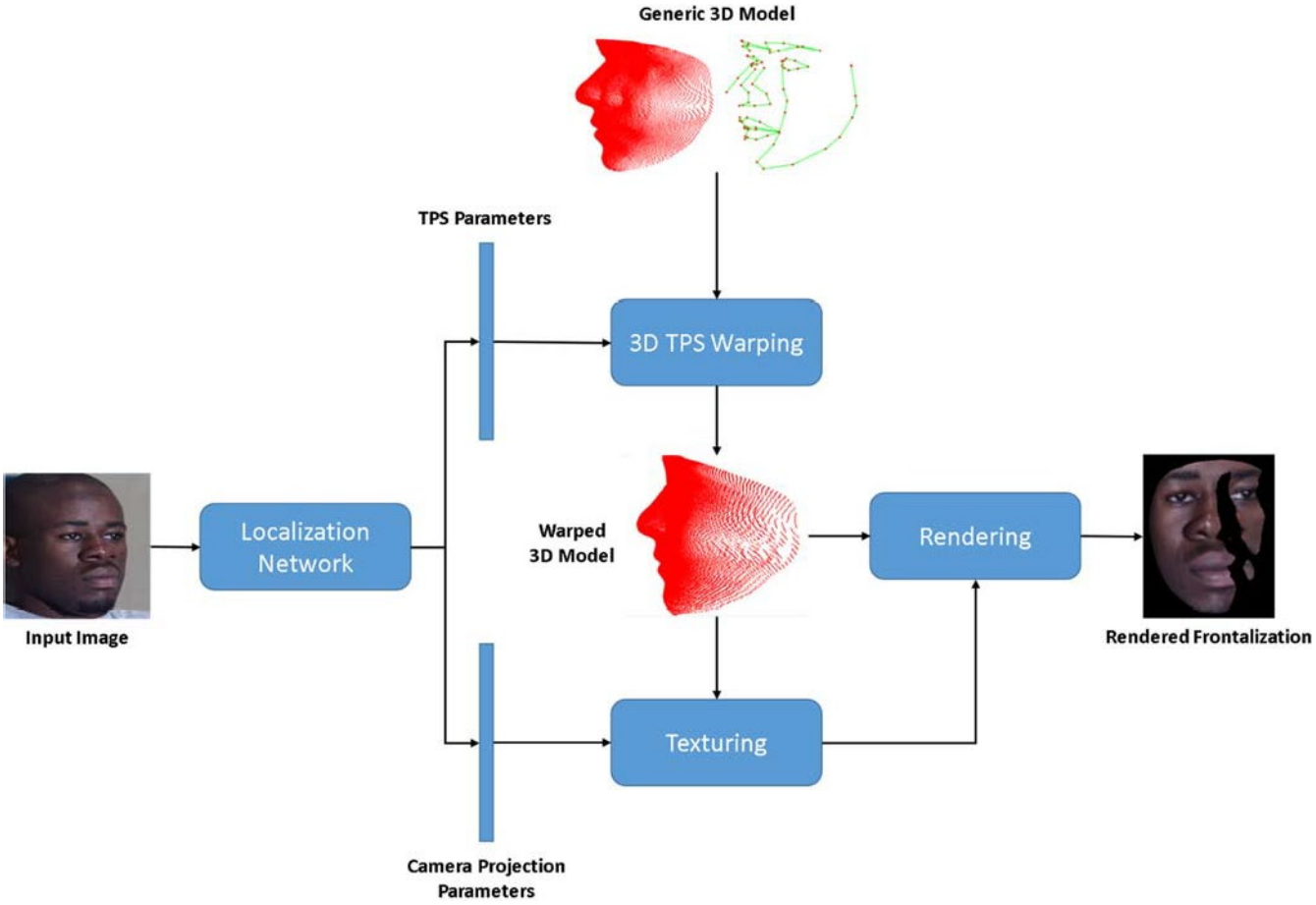


Figure 21: Network architecture for generating 3D models

In order to train deep networks, one needs many samples of data to train with. However, there is not such an abundance of 2D images with corresponding 3D models available like there is in many other face applications. In order to be able to train the network, we have to use synthetic data to train the model. Due to our previous methods and the large amount of work we have done in the past on 3D modeling, we are able to generate many synthetic views of faces from a frontal image with landmarks as shown in Figure 22. Since we render the images ourselves, we have perfect knowledge of where every 3D vertex is and where it projects onto the 2D synthesized image. This allows us to train our network to at least give the same performance as our previous methods in the 3D modelling stage.

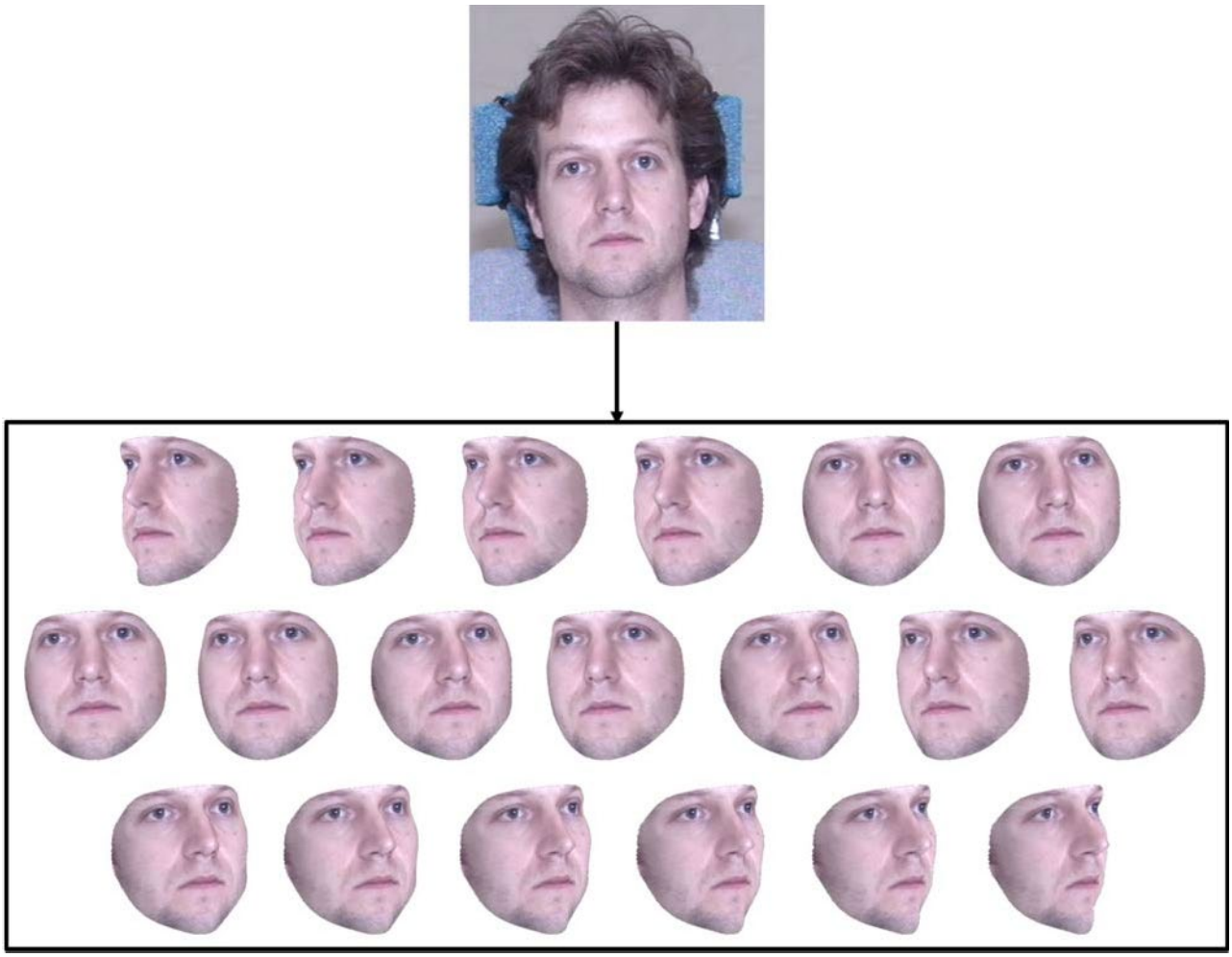


Figure 22: Example of synthesized pose images for training deep network. The original image (top) is landmarked by hand to generate the most accurate 3d model which is used to synthesize the images in the box

Once the network is trained, any image can have the 3Dmodel fit to it without the need for input landmarks. This new model can be used along with the camera position estimated to determine a pose and find the non-visible regions. When these models are rendered from a frontal viewpoint, it becomes clear that the newly rendered faces do indeed look as if they are facing the camera as shown in Figure 23.

The models inherently contain the landmarks we usually use as well and can be used as both a way to evaluate the fitting of the model as well as a fast way to landmark such images for any future task that might require such landmarks as shown in Figure 24. These landmarks are actually more consistent than a 2D approach due to the 3D position not shifting when a face is at a different pose. In a traditional landmarking method, the boundary points move to the visible boundary which mans the landmarks move in a 3D sense as well as a 2D sense. This approach does not suffer the same problem.

Figure 23: Original images (top) with projected 2D coordinates of 3D model estimated by the network (middle). The frontalized rendering with non-visible regions blacked out are shown (bottom).

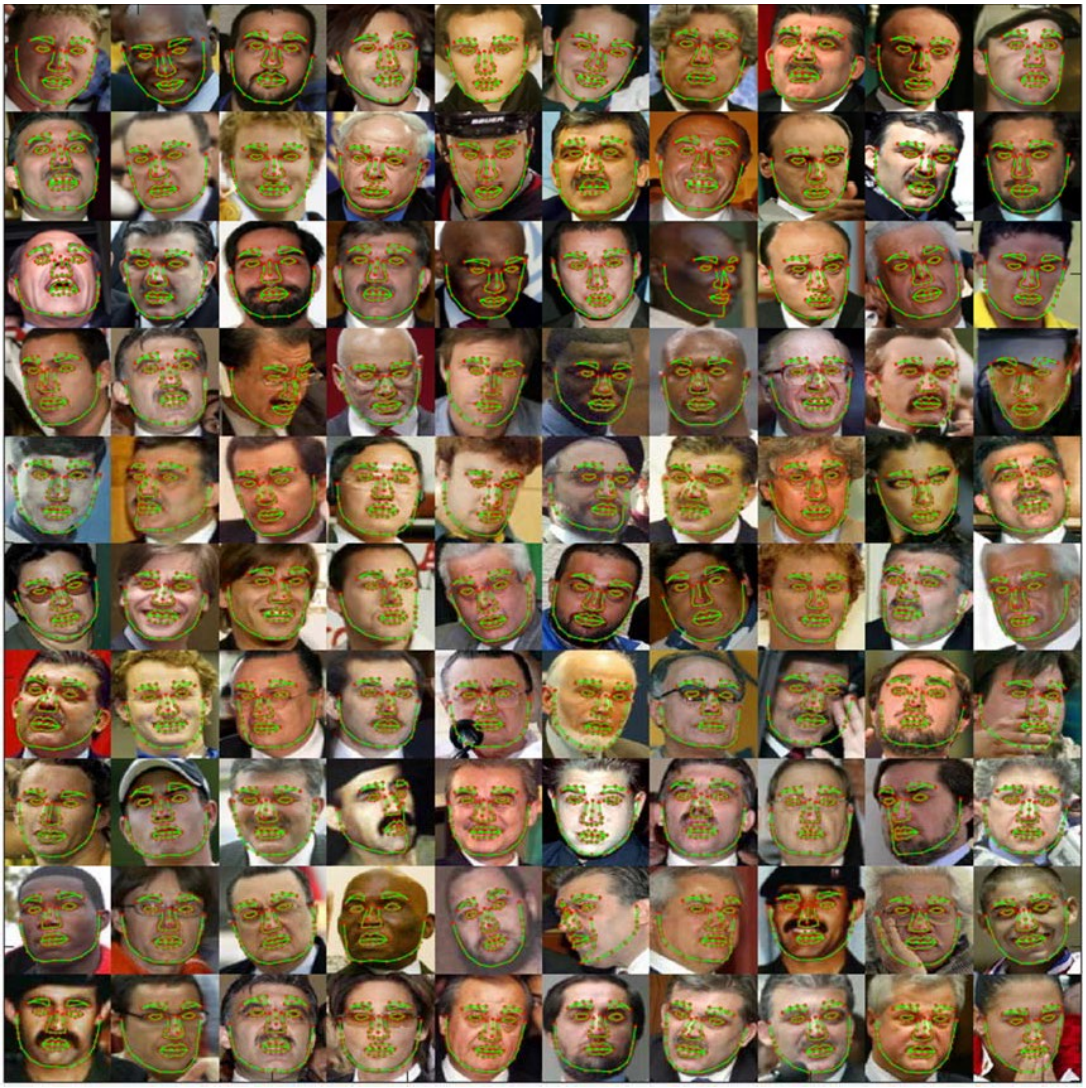


Figure 24: Landmarks found by the 3D modeling on 100 images from the LFW dataset.

Extensions to the 3D Landmarking

We have extended that work to extract more accurate landmarks and 3D models. Instead of using data synthesized form the 3DGEM approach, we now use the data in the 300W-LP dataset. This dataset contains all of the faces from the 300W dataset but with 3D models associated with each face. The 3D models were generated by hand fitting the Basel Face Model (BFM) to the data and then synthetically rotating the 3D heads while maintaining the background in the image. This leads to more realistic looking synthetic images as can be seen in Figure 25.

Figure 25: Examples of synthetically generated pose images from 300W-LP dataset

By training on these kinds of images, the model is able to ignore background information much more reliably and results in more accurate 3D models and landmarks. We also changed the architecture of the network itself to add in a regression step for the landmarks themselves. Since the landmarks were found by projecting the 3D model onto the image, often, there would be errors in the landmarks due to no knowledge of the local features around the landmarks. In order to address this, we sample a refined feature map at the initial landmark locations and pass these features through a fully connected layer that outputs the change in landmark position. The new architecture can be seen in Figure 26.

This new section is trained after the old model is trained to avoid any "moving target" problems of the minimum in the loss function. This new model has achieved state-of-the-art performance in 2D landmarking and can run on a GPU at faster than real time speeds with our most accurate model landmarking about 50 faces per second. The results on the AFLW and AFLW2000-3D datasets can be seen in Table 1.

Table 1: Landmarking results on the AFLW and AFLW2000-3D datasets. The AFLW2000-3D dataset is designed to ensure the landmarks are fitting a 3D model and not moving to the visible boundary. This is important for 3D modeling tasks.

	AFLW Dataset (21 pts)					AFLW 2000-3D Dataset (68 pts)				
Method	[0, 30]	(30, 60]	(60, 90]	mean	std	[0, 30]	(30, 60]	(60, 90]	mean	std
CDM	8.15	13.02	16.17	12.44	4.04	-	-	-	-	-
RCPR	5.43	6.58	11.53	7.85	3.24	4.26	5.96	13.18	7.80	4.74
ESR	5.66	7.12	11.94	8.24	3.29	4.60	6.70	12.67	7.99	4.19
SDM	4.75	5.55	9.34	6.55	2.45	3.67	4.94	9.76	6.12	3.21
3DDFA	5.00	5.06	6.74	5.60	0.99	3.78	4.54	7.93	5.42	2.21
3DDFA+SDM	4.75	4.83	**6.38**	5.32	**0.92**	**3.43**	**4.24**	**7.17**	**4.94**	1.97
Ours (AlexNet)	**4.11**	**4.69**	6.61	**5.14**	1.31	3.71	5.33	7.19	5.41	**1.74**
Ours (VGG-16)	**3.55**	**3.92**	**5.21**	**4.23**	**0.87**	**3.15**	**4.33**	**5.98**	**4.49**	**1.42**

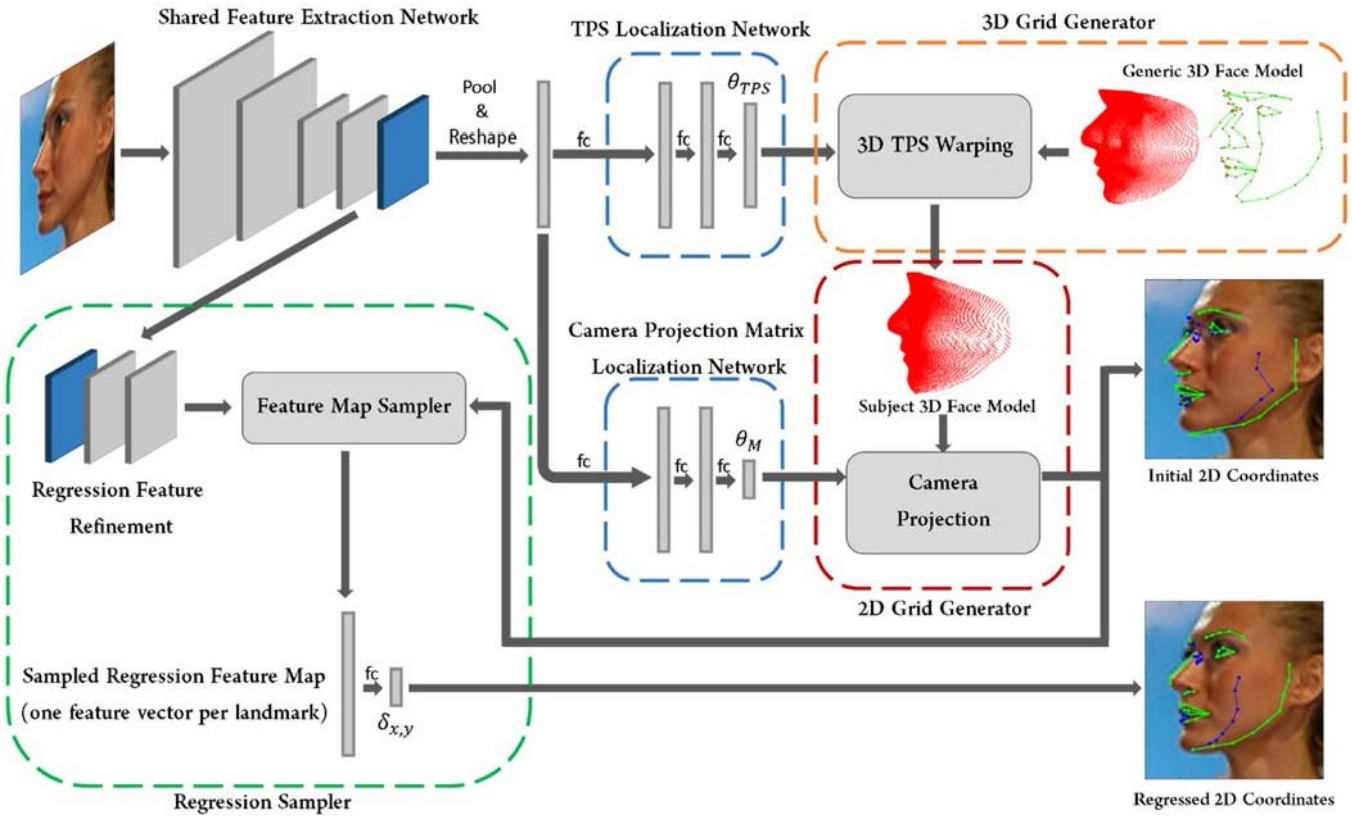


Figure 26: New 3D modeling and landmarking architecture. Note that there is now a regression sampler that outputs refined landmarks.

This increase in accuracy in the landmarks means that the 3D models generated from the landmarks are also able to handle more variations, such as expression. The final model is generated in the same way as detailed in previous reports with the landmarks being used to back project the camera rays into 3D space and find the closest points. Since an initial 3D model is found in the network itself, not much warping needs to be done which results in better looking 3D models as can be seen in Figure 27.

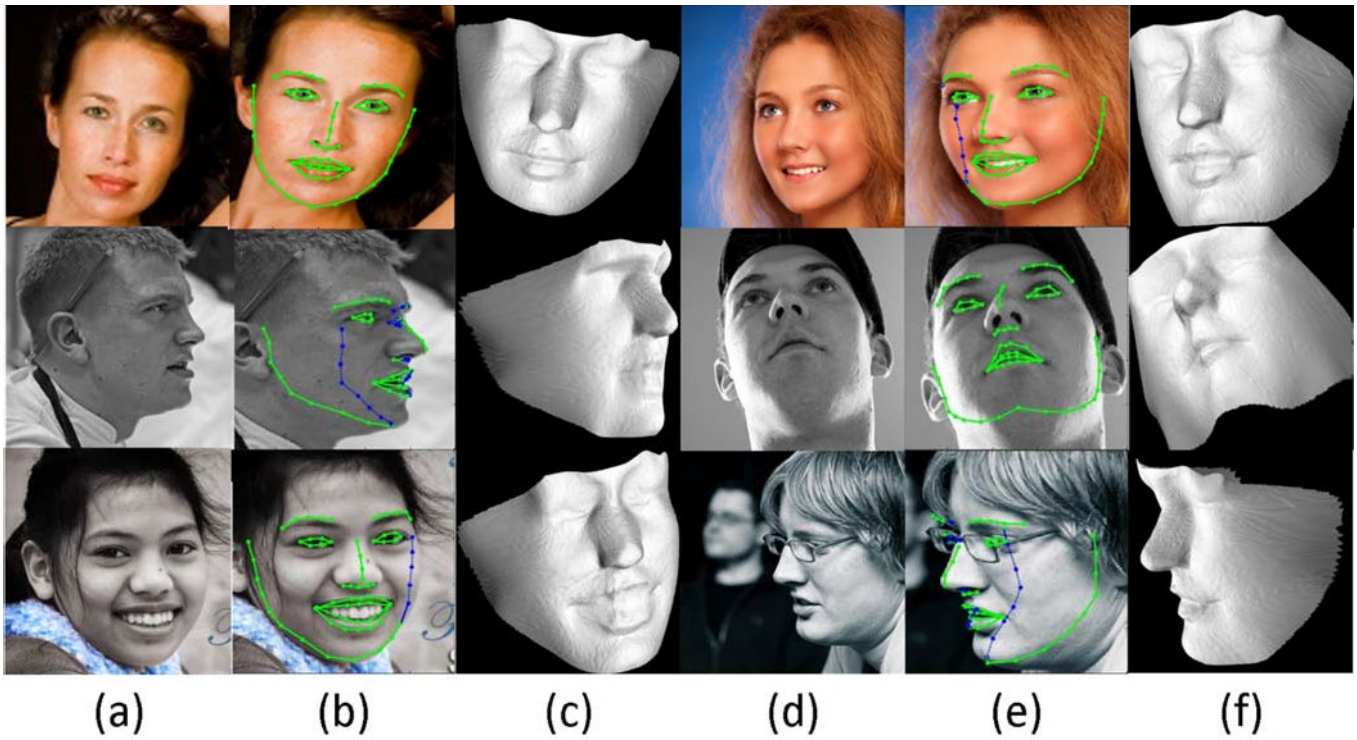


Figure 27: (a & d) Original images. (b & e) Landmarks found using new method. Landmarks in blue are determined to be self occluded due to pose. (c & f) Extracted 3D model.

8. Robust Face Detection

Robust face detection is one of the most important pre-processing steps to support facial expression analysis, facial landmarking, face recognition, pose estimation, building of 3D facial models, etc. Although this topic has been intensely studied for decades, it is still challenging due to numerous variants of face images in real-world scenarios.

We have developed a novel end-to-end network named Contextual Multi-Scale Region-based Convolutional Neural Network (CMS-RCNN) to robustly detect human facial regions from images collected under various challenging conditions, e.g. large occlusions, extremely low resolutions, facial expressions, strong illumination variations, etc. Two main contributions of our work include 1) multi-scale features for tiny faces, and 2) explicit reference of body context for challenging faces.
The proposed approach was benchmarked on two challenging face detection databases, i.e. the Wider Face database and the Face Detection Dataset and Benchmark (FDDB), and compared against recent other face detection methods, e.g. Two-stage CNN, Multi-scale Cascade CNN, Faceness, Aggregate Chanel Features, HeadHunter, Multi-view Face Detection, Cascade CNN, etc. The experimental results show that our proposed approach consistently achieves highly competitive results with the state-of-the-art performance against other recent face detection methods.

With the use of this robust face detector, we can normalize faces more robustly for input to the 3D modeling and for face matching. Additionally, without a face detector capable of detecting off-angle and

occluded faces accurately, much of the rest of the work presented in these reports would require more human intervention as undetected faces cannot be automatically processed. To that end, the CMS-RCNN face detector will lead to better performance in many of the other tasks.

Method

Our designed CNN architecture allows the network to simultaneously look at multi-scale features, as well as to explicitly look outside facial regions as the potential body regions. In other words, this process tries to mimic the way of face detection by human in a sense that when humans are not sure about a face, seeing the body will increase our confidence. Additionally this architecture also helps to synchronize both the global semantic features in high level layers and the localization features in low level layers for facial representation. Therefore, it is able to robustly deal with the challenges in the problem of unconstrained face detection.

Our CMS-RCNN method introduces the Multi-Scale Region Proposal Network (MS-RPN) to generate a set of region candidates and the Contextual Multi-Scale Convolution Neural Network (CMS-CNN) to do inference on the region candidates of facial regions. A confidence score and bounding box regression are computed for every candidate. In the end, the face detection system is able to decide the quality of the detection results by thresholding these generated confidence scores in given face images. The architecture of our proposed CMS-RCNN network for unconstrained face detection is illustrated in Figure 28.

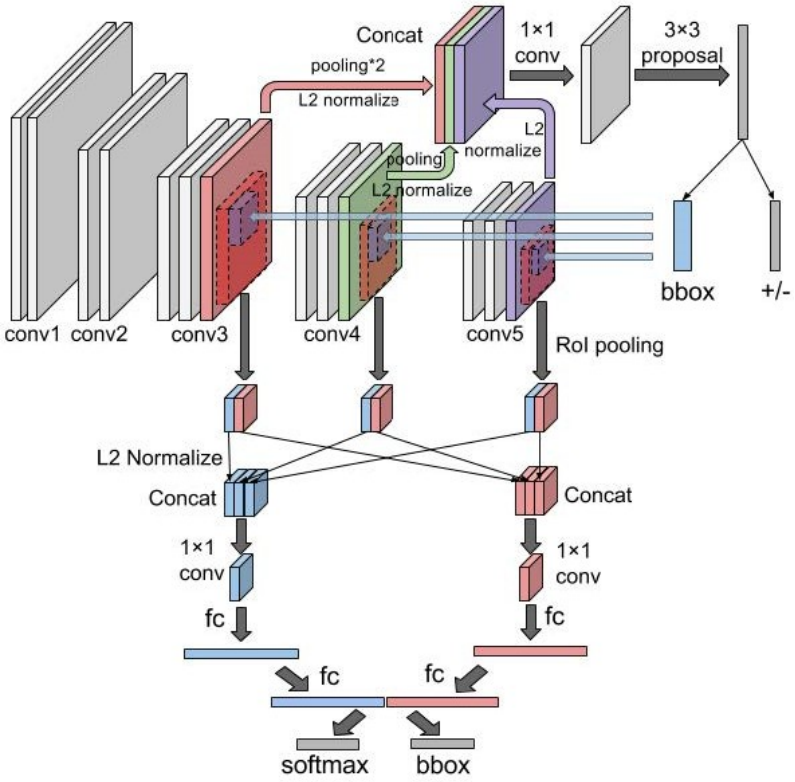


Figure 28: Our proposed Contextual Multi-Scale Region-based CNN model

In this architecture, the feature maps are incorporated from lower level convolution layers with the last convolution layer for both MS-RPN and CMS-CNN. Features from lower convolution layer help get more information for the tiny faces, because stride in lower convolution layer will not be too small. Another benefit is that both low-level feature with localization capability and high-level feature with semantic information are fused together, since face detection needs to localize the face as well as to identify the face. In the MS-RPN, the whole lower level feature maps are down-sampled to the size of high level feature map and then concatenated with it to form a unified feature map. Then we reduce the dimension of the unified feature map and use it to generate region candidates. In the CMS-CNN, the region proposal is projected into feature maps from multiple convolution layers. And RoI-pooling is performed in each layer, resulting in a fixed-size feature tensor. All feature tensors are normalized, concatenated and dimension-reduced to a single feature blob, which is forwarded to two fully connected layers to compute a representation of the region candidate.

In our proposed network, the contextual body reasoning is implemented by explicitly grouping body information from convolution feature maps shown as the red blocks in Figure 28. Specifically, additional RoI-pooling operations are performed for each region proposal in convolution feature maps to represent the body context features. Then same as the face feature tensors, these body feature tensors are normalized, concatenated and dimension-reduced to a single feature blob. After two fully connected layers the final body representation is concatenated with the face representation. They together contribute to the computation of confidence score and bounding box regression. With projected region proposal as the face region, the additional RoI-pooling region represents the body region and satisfies a pre-defined spatial relation with the face region. In order to model this spatial relation, we make a simple hypothesis that if there is a face, there must exist a body, and the spatial relation between each face and body is fixed. This assumption may not be true all the time but should cover most of the scenarios since most people we see in the real world are either standing or sitting. Therefore, the spatial relation is roughly fixed between the face and the vertical body.

Results

WIDER FACE Dataset

WIDER FACE is a public face detection benchmark dataset. It contains 393,703 labeled human faces from 32,203 images collected based on 61 event classes from internet. The database has many human faces with a high degree of pose variation, large occlusions, low-resolutions and strong lighting conditions. The images in this database are organized and split into three subsets, i.e. training, validation and testing. Each contains 40%, 10% and 50% respectively of the original databases.

The images and the ground-truth labels of the training and the validation sets are available online for experiments. However, in the testing set, only the testing images (not the ground-truth labels) are available online. All detection results are sent to the database server for evaluating and receiving the Precision-Recall curves.

In our experiments, the proposed CMS-RCNN is trained on the training set of the WIDER FACE dataset containing 159,424 annotated faces collected in 12,880 images. The trained model on this database are used in testing of all databases.

The proposed CMS-RCNN model is compared against recent strong face detection methods, i.e. Two-stage CNN, Multiscale Cascade CNN, Faceness, and Aggregate Channel Features (ACF). All these methods are trained on the same training set and tested on the same testing set. The Precision-Recall curves and AP values are shown in Figure 29. Figure 3 shows some examples of face detection results using the proposed CMS-RCNN on this database.

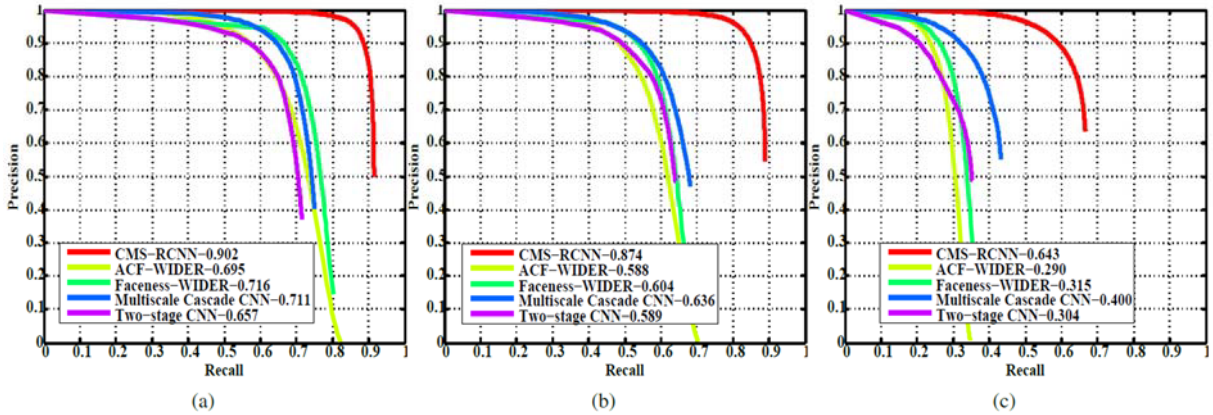


Figure 29: Precision-Recall curves obtained by our proposed CMS-RCNN (red) and the other baselines, i.e. Two-stage CNN, Multi-scale Cascade CNN, Faceness, and Aggregate Channel Features (ACF).

Figure 30: Some examples of face detection results using our proposed CMS-RCNN method on WIDER FACE database.

FDDB Face Database

To show that our method generalizes well to other database, the proposed CMS-RCNN is also benchmarked on the FDDB database. It is a standard database for testing and evaluation of face detection algorithms. It contains annotations for 5,171 faces in a set of 2,845 images taken from the Faces in the Wild dataset. Most of the images in the FDDB database contain less than 3 faces that are clear or slightly occluded. The faces generally have large sizes and high resolutions compared to WIDER

FACE. We use the same model trained on WIDER FACE training set to perform the evaluation on the FDDB database.

The evaluation is proceeded following the FDDB evaluation protocol and compared against the published methods provided in the protocol. The proposed CMS-RCNN approach outperforms most of the published face detection methods and achieves a very high recall rate comparing against all other methods shown in Figure 31. Figure 32 shows some examples of the face detection results using the proposed CMS-RCNN on the FDDB dataset.

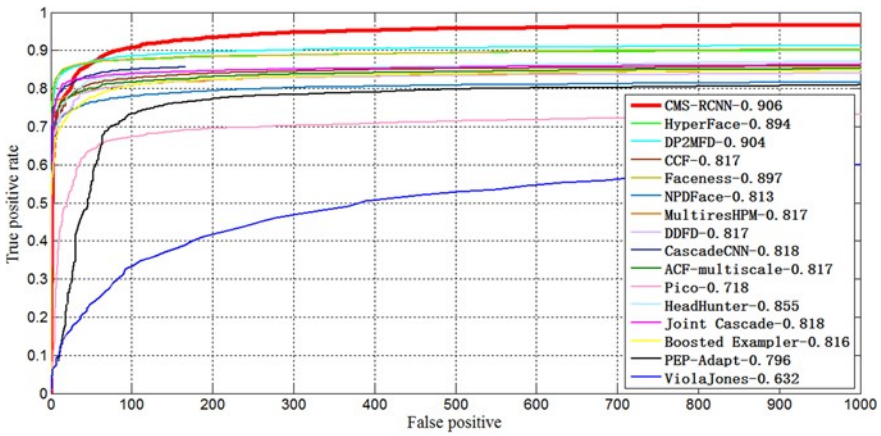


Figure 31: ROC curves of our proposed CMS-RCNN and the other published methods on FDDB database. Our method achieves the best recall rate on this database. Numbers in the legend show the average precision scores.

Figure 32: Some examples of face detection results using our proposed CMS-RCNN method on FDDB database.

ISIS Style Faces

This data set is collected for general precision and recall test on ISIS style faces. The videos are downloaded from YouTube searched by keyword like “Iraq war” and “ISIS”. The videos are then disintegrated to frames. Faces that meet our detection prerequisite were labeled with bounding boxes for each frame. The metrics for ISIS dataset are accuracy and recall rate. The accuracy describe the

percentage of detected faces being actual faces. The recall rate describe the ratio of detected actual faces vs. total number of actual faces. We randomly select 1920 of these frames containing faces and do precision and recall test on it. Our method detects 3729 faces with 88.56% accuracy and 83.05% recall. We only have 400 false alarms. Figure 33 shows some qualitative results

Figure 33: Some examples of face detection results using our proposed CMS-RCNN method on ISIS style faces.

9. Face Matching with Deep Learning

With the many advances being made in deep learning, we have focused some of our research effort on utilizing deep learning for face recognition in order to achieve better performance. One of the recent works that performs very well is using deep learning with Center Loss to encourage the network to extract discriminative features. We have been implementing this Center Loss network for benchmarking as well as performing research into improving such a network.

Approach to Face Matching: Deep Convolutional Neural Networks with Center Loss

Our approach to face matching will be built off of an approach to handle the fundamental problem in classification. For a good classification result, we need to minimize intra-class variation and maximize between-class variation. For maximization of separation between classes, we utilize the negative log-likelihood (NLL) loss. This loss implicitly tries to minimize intra-class variation as well. However, recently a loss called center loss to explicitly minimize the intra class variation between samples has been proposed which we benchmark in our study. An example of how the NLL loss acts on data is shown in **Error! Reference source not found.**.

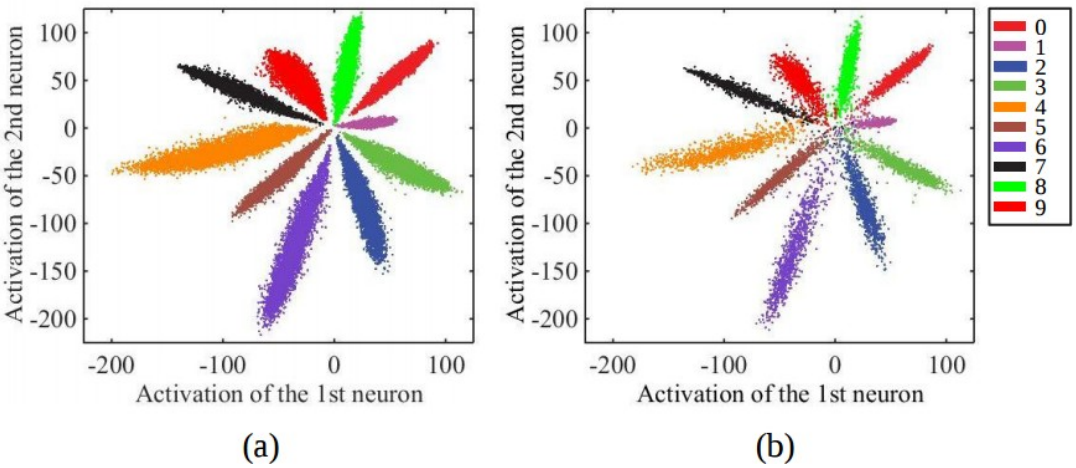


Figure 34: The activation of the 2D neuron at the last layer of a neural network of a model trained on the MNIST database of 10 handwritten digits. a) training set samples b) test set samples

Center loss is a loss that keeps track of the centers or the mean activation of each class. It then uses the means as targets towards which it forces the samples from each class to go to. For instance, if we have a supervised classification task of 10,000 classes, we would have to maintain 10,000 centers, one for each class. As and when samples are fed into the network for training, the MSE or mean squared error is computed between the features of the samples and the centers for the respective class. The error gradients are then back propagated after combining with the gradients from the NLL loss with a weight λ. This minimizes both losses for every mini batch and provides an explicit way to enforce within class similarity.

The ideal way to enforce center loss would be to compute the mean activations of every training sample in our training data. In other words, we need to take the entire training set into account and average the features of every class in each iteration, which is inefficient even impractical. Therefore, the center loss cannot be used directly. This is possibly the reason that such a center loss has never been used in CNNs until now. To address this problem, one can make two necessary modifications. First, instead of updating the centers with respect to the entire training set, one can perform the update based on mini-batch. In each iteration, the centers are computed by averaging the features of the corresponding classes (In this case, some of the centers may not update). Second, to avoid large perturbations caused by few mislabeled samples, one use a scalar to control the learning rate of the centers. Figure 36 illustrates how combining the gradients of the NLL and center loss differently result in feature embeddings for the handwritten digit recognition dataset MNIST.

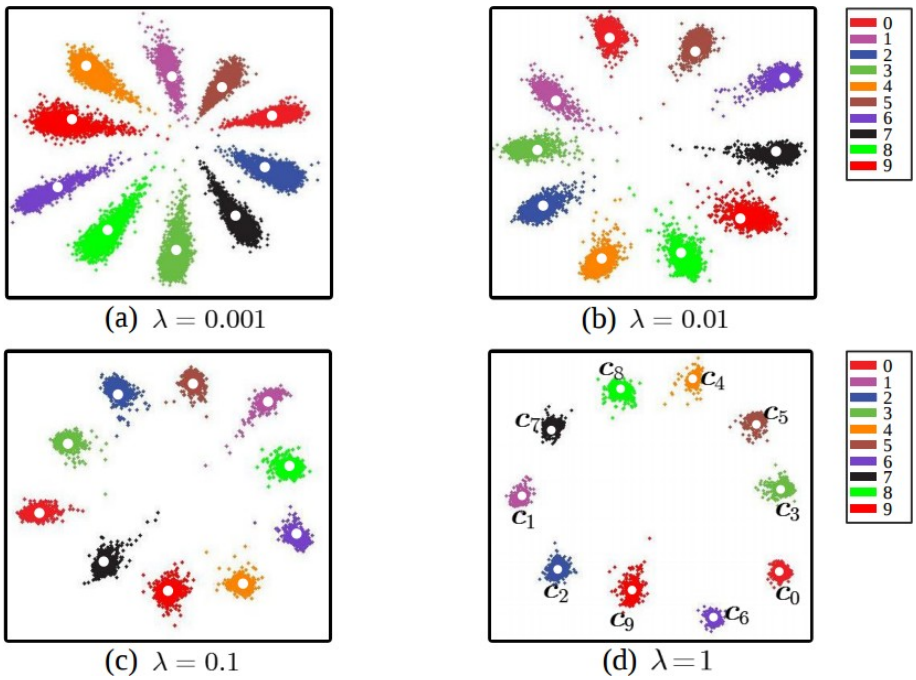


Figure 35: The distribution of deeply learned features under the joint supervision of softmax loss and center loss. The points with different colors denote features from different classes. Different λ lead to different deep feature distributions. The white dots (c_0,c_1,…c_9) denote 10 class centers of deep features.

Approach to Face Matching: Deep Convolutional Neural Networks with Ring Loss

Our approach to face matching will be built off of an approach to handle the fundamental problem in classification. For a good classification result, we need to minimize intra-class variation and maximize between-class variation. There are multiple methods to do this in literature. Center loss defines a set of class centers estimated from the sample features. However, this method is complicated, requiring the class center to be stored during training. Our developed method Ring Loss, on the other hand is very simple and augment the negative log-likelihood (NLL) loss for classification. The NLL loss implicitly tries to minimize intra-class variation as well. An example of how the NLL loss acts on data is shown in Figure 34.

Our developed Ring Loss, was motivated from the study of learning robust features. Ring loss, a simple and elegant feature vector normalization approach for deep networks was designed to augment standard loss functions such as Softmax. We argue that feature vector normalization is an important aspect of supervised classification problems where we require the model to represent each class equally well. The direct approach to feature normalization through the hard normalization operation results in a non-convex formulation. Ring loss gradually learns to constrain the norm to the scaled unit circle while preserving convexity leading to more robust features. We apply Ring loss to large-scale face recognition problems and present results on LFW, the challenging protocols of IJB-A Janus, and MegaFace with 1 million distractors.

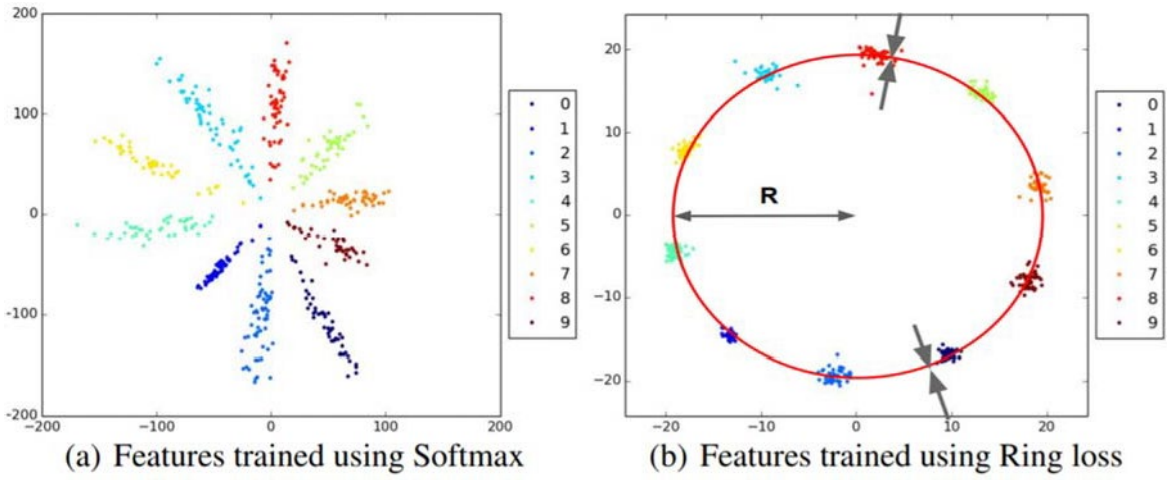


Figure 36: The distribution of deeply learned features under the joint supervision of vanilla Softmax loss and ring loss. The points with different colors denote features from different classes.

Evaluation Datasets

The databases that we test on for this component are recent challenging large scale face databases. Three of these are the MegaFace database, Labelled Faces in the Wild (LFW) and IARPA IJB-A Janus. We detail the three databases below.

MegaFace:
The MegaFace database is a standard face matching benchmark database. The testing database contains two sets of face images. The distractor set contain one million distractor subjects. The target set contain 100K images from 530 celebrities. Both are required for evaluating a proposed face matcher. The testing procedure is basically extracting features from both sets of face images and matching particular celebrities against the whole database. The evaluation scripts are all provided online.

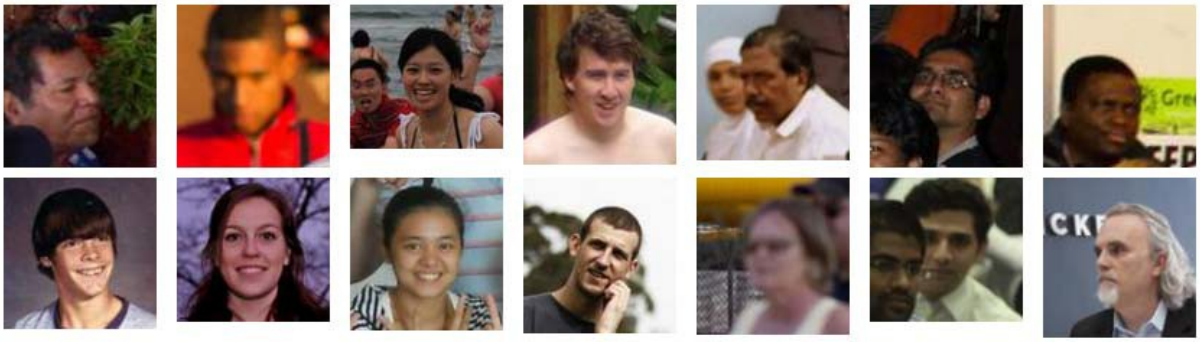


Figure 37: Example face images from MegaFace

Labelled Faces in the Wild (LFW):
Labeled Faces in the Wild (LFW), a database of face photographs designed for studying the problem of unconstrained face recognition. The data set contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured

have two or more distinct photos in the data set. The only constraint on these faces is that they were detected by the Viola-Jones face detector.

Figure 38: Example face images from LFW

IARPA IJB-A Janus:

IJB-A is a new publicly available challenge proposed by IARPA and spread by NIST to push frontiers of face recognition in the wild since lately LFW performance saturated. IJB-A consists of 500 subjects under extreme conditions regarding pose, expression and illuminations with a total of 5712 images. The IJB-A evaluation protocol mainly consists of face verification (1:1) and face identification (1:N). The interesting thing about this dataset is that each subject is described by a template containing a set of images or frames extracted from videos.

Figure 39: Example face images from IJB-A Janus

Benchmarking center loss on LFW

We have implemented center loss from scratch and have benchmarked the loss on the LFW database. Figure 41 shows the ROC curve for our model trained on a database of 0.4 million images (CASIA-WebFace). The model was a 34 layer Residual Network.

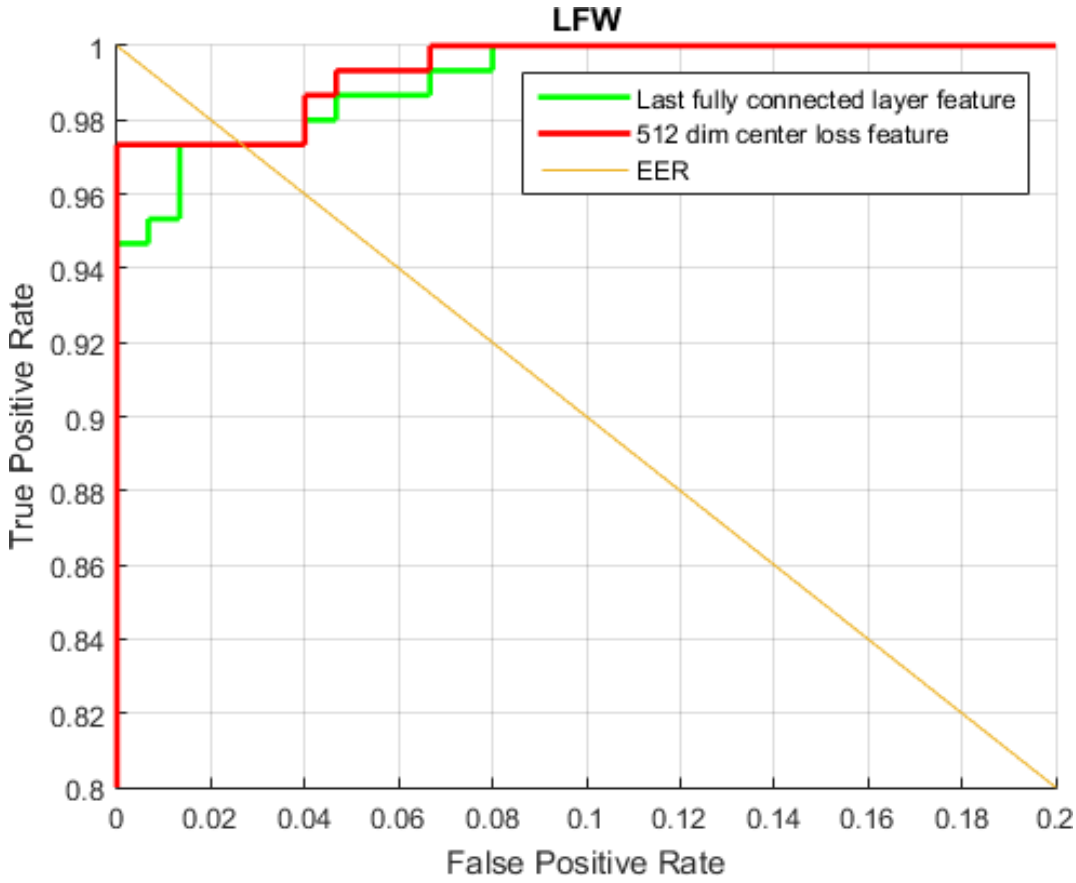


Figure 40: ROC curve of our model trained from scratch with center loss on LFW. The red curve is the curve obtained using the center loss features and the blue curve was obtained using the final layer features. The center loss features performed better than the final layer features.

Benchmarking Ring Loss on LFW, IJB-A Janus and Megaface

We have developed and implemented Ring Loss from scratch and have benchmarked the loss on the LFW and IJB-A Janus database while training on the MS-Celeb database with close to 3.2 million imahges. Figure 41 shows the ROC curve for our model on IJB-A Janus trained on this database. The model was a 64 layer residual network. For LFW, our model achieved a high accuracy of **99.50 %.** For Megaface our model was trained on the CASIA-WebFace dataset with 0.4 million images to adhere to the "small" dataset protocol defined. We achieve an accuracy of **68.62 %** with a 34 layer Resnet model.

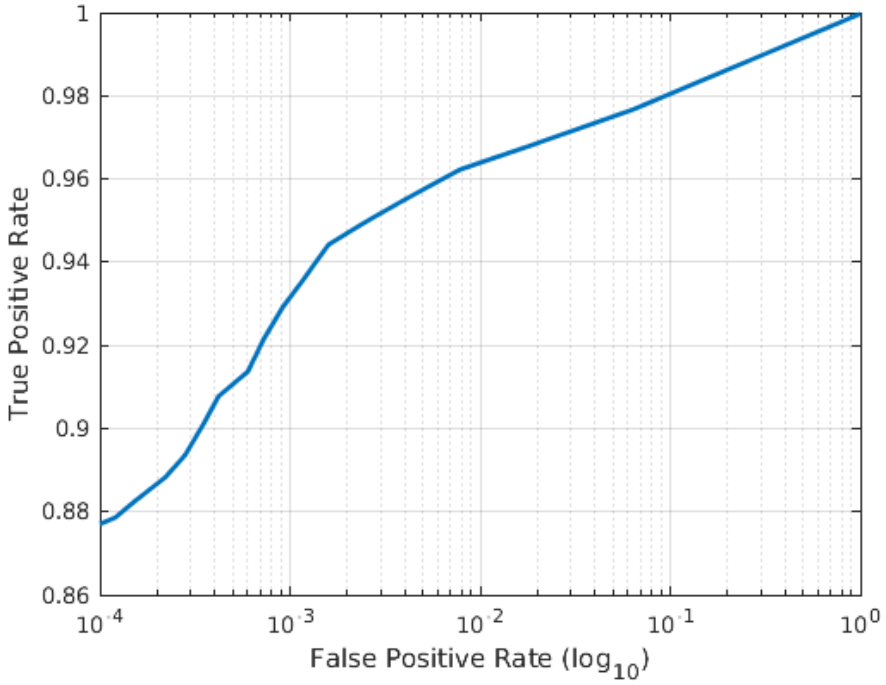


Figure 41: ROC curve of our model trained from scratch with ring a combination of ring loss and sphereface on the IJB-A Janus test dataset (x axis is the False Accept Rate or FAR and the y axis is the True Positive Rate or TPR).

Inventions, patent applications, and/or licenses

Invention disclosure and provisional patent submitted: Marios Savvides, Juefei Xu, Dipan Kumar Pal, “Reconstruction the full face image from the periocular region”, submitted June 13, 2014

Other products

IRB no. IRB00000603 (IRB Protocol number HS13-717) “Acquisition of Images and Videos of Faces” was approved by the CMU IRB Office on Jan 29th 2014. This IRB approval is essential towards data collection tasks which are integral to this project.

The purpose of this study is to explore the degradation of biometric matching capabilities of faces in images and video under degradation in acquisition conditions. The collected data will be used to analyze and evaluate the efficacy of the proposed researched solution on this real-world data. A session will comprise of a maximum of 15 minutes which involves the time taken to explain the procedure to the participant and the time taken to acquire the required images and/or video. Images and/or videos may be captured under a variety of conditions, including use of multiple imaging devices, under different natural and artificial illuminations, varying poses, occlusions, expressions, etc. The participants will be asked to sign a consent form to indicate his/her willingness to participate. This request may come from the PI or any of the graduate students and post-doctoral staff working with the PI involved in this research. The risk to participants is negligible. The study involves acquisition of images and video of the subject from one or more camera devices, which is a common occurrence in most participants' daily lives. The information that links the participant's name and contact information will be stored in a secure location inside the research lab. The PI will be point of contact to gain access to any of this information.

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS: Who has been involved?

What individuals have worked on the project?

Name:	Marios Savvides
Project Role:	Principal Investigator
Nearest Person month worked:	14.49 Months
Contribution to Project:	Project oversight & high-level research decisions
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Vijayakumar Bhagavatula
Project Role:	Co-Principal Investigator
Nearest Person month worked:	2 Months
Contribution to Project:	Project oversight & high-level research decisions
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Khoa Luu
Project Role:	Research Programmer
Nearest Person month worked:	16.38 Months
Contribution to Project:	Research, code development.
Funding Support:	
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Vishnu Bodetti
Project Role:	Research Programmer
Nearest Person month worked:	4 months
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A

Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Utsav Prabhu
Project Role:	Graduate Student
Nearest Person month worked:	19 Months
Contribution to Project:	Research, code development.
Funding Support:	
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Keshav Seshadri
Project Role:	Graduate Student
Nearest Person month worked:	16 Months
Contribution to Project:	Research, code development.
Funding Support:	
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Juefei Xu
Project Role:	Graduate Student
Nearest Person month worked:	21 months
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Chandrasekhar Bhagavatula
Project Role:	Graduate Student
Nearest Person month worked:	11 months
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Dipan Pal
Project Role:	Graduate Student
Nearest Person month worked:	7 months
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Thi Hoang Ngan Le
Project Role:	Graduate Student
Nearest Person month worked:	7 months
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No
If travelled to foreign country(ies), duration of stay:	N/A

Name:	Karanhaar Singh
Project Role:	Graduate Student
Nearest Person month worked:	1 month
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No

Name:	Shreyas Venugopalan
Project Role:	Graduate Student
Nearest Person month worked:	2 months
Contribution to Project:	Research, code development.
Funding Support:	no change
Collaborated with individual in foreign country:	No
Country(ies) of foreign collaborator:	N/A
Travelled to foreign country:	No

What other organizations have been involved as partners?

Nothing to Report.

Have other collaborators or contacts been involved?

Nothing to Report.

IMPACT: What is the impact of the project? How has it contributed?

What is the impact on the development of the principal discipline(s) of the project? The representation model which has been designed has a potential for widespread impact in the face recognition and analysis community. Similar representation models can be constructed and used to overcome a variety of impediments in the field.

The representation model which has been designed for Occlusion Recovery has a potential for widespread impact in the face recognition and analysis community. Similar representation models can be constructed and used to overcome a variety of impediments in the field. The occlusion removal tool can be used in conjunction with other face pre-processing technologies to significantly expand the capabilities of face recognition systems, thereby making them more usable for real-world application.

The capability of hallucinating the full face from just the periocular region has greatly extended the usability of face images with masks, enabling law-enforcement to match such highly-occluded faces with high confidence.

What is the impact on other disciplines?

The novel learning algorithms developed in this project have a potential for widespread impact in the larger machine learning community. Principal subspace constructions as well as sparsity-inducing dictionaries are primarily used in many pattern recognition and machine learning tasks for data modeling. The corresponding algorithms which have been developed allow one to learn such linear representation models from extremely large amounts of incomplete or corrupted training data, which is a crucial requirement in many tasks.

What is the impact on the development of human resources?

Nothing to Report.

What is the impact on physical, institutional, and information resources that form infrastructure?

Nothing to Report.

What is the impact on technology transfer?

Nothing to Report.

What is the impact on society beyond science and technology?

Nothing to Report.

What dollar amount of the award's budget is being spent in foreign country(ies)?

None.

CHANGES/PROBLEMS

Changes in approach and reasons for change

Nothing to Report.

Actual or anticipated problems or delays and actions or plans to resolve them

Nothing to Report.

Changes that have a significant impact on expenditures

Nothing to Report.

Significant changes in use or care of human subjects, vertebrate animals, and/or biohazards

Nothing to Report.

Change of primary performance site location from that originally proposed

Nothing to Report.

SPECIAL REPORTING REQUIREMENTS

BUDGETARY INFORMATION

No changes to budget, nothing to report.