Face recognition performance with superresolution

Shuowen Hu,1,* Robert Maschal,1 S. Susan Young,1 Tsai Hong Hong,2 and P. Jonathon Phillips2

1United States Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783, USA
2NIST, 100 Bureau Drive, Gaithersburg, Maryland 20899, USA
*Corresponding author: shuowen.hu.civ@mail.mil

Received 29 September 2011; revised 19 April 2012; accepted 24 April 2012; posted 30 April 2012 (Doc. ID 155384); published 20 June 2012

With the prevalence of surveillance systems, face recognition is crucial to aiding the law enforcement community and homeland security in identifying suspects and suspicious individuals on watch lists. However, face recognition performance is severely affected by the low face resolution of individuals in typical surveillance footage, oftentimes due to the distance of individuals from the cameras as well as the small pixel count of low-cost surveillance systems. Superresolution image reconstruction has the potential to improve face recognition performance by using a sequence of low-resolution images of an individual’s face in the same pose to reconstruct a more detailed high-resolution facial image. This work conducts an extensive performance evaluation of superresolution for a face recognition algorithm using a methodology and experimental setup consistent with real world settings at multiple subject-to-camera distances. Results show that superresolution image reconstruction improves face recognition performance considerably at the examined midrange and close range.

OCIS codes: 100.0100, 100.6640, 100.4995, 100.2980.

1. Introduction

The affordability of surveillance systems has led to their widespread usage on commercial properties and for residential monitoring. Consequently, video footage of criminal activity is often available to law enforcement to help identify suspects. Therefore, face recognition software is a crucial tool that the law enforcement community may use to search watch lists and criminal databases to identify the suspect acquired on video. However, typical low-cost surveillance systems have small pixel counts. Furthermore, the suspect could be far away from the camera, resulting in images with very limited number of pixels on the face (i.e., low face resolution).

Research studies have shown that while face recognition algorithm performance is dependent on face resolution, this dependence is highly nonlinear. Boom et al. [1] examined the effect of resolution on face recognition and observed that performance became severely degraded for face images with sizes less than 32 × 32 pixels. However, performance was observed to be fairly similar for face images with sizes ranging from 32 × 32 pixels to 128 × 128 pixels [1], substantiating the highly nonlinear nature of face recognition performance with respect to face resolution. The Facial Recognition Vendor Test 2000 [2] also observed that the evaluated face recognition systems yielded similar performance for face images with face resolutions of 30 to 60 pixels measured in terms of eye-to-eye distance, but that performance severely degraded for some algorithms at an eye-to-eye distance of 15 pixels.

In the authors’ experience of working with law enforcement agencies, it is not uncommon for faces in typical surveillance footage from residential and commercial properties to have resolutions less than 30 pixels in terms of eye-to-eye distance, especially when the suspect is far away from the camera. Therefore, the limited face resolution within surveillance footage is a major obstacle for face recognition software. Pennsylvania Justice Network (JNET) states that low resolution and distance are two of the main factors that limited face recognition effectiveness of its statewide implementation of a facial recognition search system for investigators [3].
Face recognition continues to be an active area of research focused on improving performance through the development of new feature transforms, classification techniques, and mathematical frameworks to handle the large variability of face imagery found in real life. Many factors contribute to this variability: illumination, pose, and scale/resolution are several of the main factors. While research has been predominantly focused on solving the pose and illumination challenges for face recognition, some efforts have been devoted to solving the face resolution problem through the use of superresolution image reconstruction, which utilizes a sequence of low-resolution (LR) images containing the face in the same pose to reconstruct a high-resolution face image with more details. Boult et al. [4] proposed a superresolution method via image warping for face recognition, and Baker and Kanade [5] proposed hallucinating faces through a Gaussian pyramid-based method; however, these works did not conduct a performance evaluation to assess the benefit of superresolution for face recognition. More recently, Wheeler et al. [6] developed a multiframe face superresolution method with an active appearance model for registration and evaluated the face recognition improvement using the Identix FaceIt software. However, only 138 images (split between six ranges in terms of eye-to-eye distance) from three test subjects were used in [6]. Due to the small sample size, the observed improvement with superresolution is unlikely to be statistically meaningful in [6]. Whereas [4–6] perform superresolution image enhancement in the pixel domain prior to the face recognition algorithm, Gunturk et al. [7] developed an eigenface-domain superresolution technique for face recognition. The algorithm of [7] performs superresolution reconstruction in a low-dimensional face space through principal component analysis (PCA)-based dimensionality reduction and showed an improvement in face recognition performance with a minimum distance classifier in the eigenspace. In contrast to [4–7], Hennings-Yeomans et al. [8] proposed an algorithm incorporating face features into superresolution as prior information, and Huang et al. [9] developed a superresolution approach based on correlated features and nonlinear mappings between low-resolution and high-resolution features. Fookes et al. [10] conducted the most recent work on superresolution for face recognition, evaluating the performance of two face recognition algorithms with three superresolution techniques. However, as in [7–9], Fookes et al. [10] also utilized synthetically generated LR face images by downsampling the original high-resolution imagery. Downsampled face imagery does not accurately depict real-world compressed surveillance face images at varying subject-to-camera distances, especially since compression is highly nonlinear with more pronounced effects on facial details for far subject-to-camera ranges. Although [10] also used a white Gaussian noise corrupted version of the downsampling set.

Specifically, the objectives of this work are to (a) assess the benefit of superresolution for face recognition with respect to subject-to-camera range, (b) assess face recognition performance using superresolved imagery reconstructed using varying numbers of LR frames, and (c) evaluate face recognition performance of individual frames within the LR sequence as well as the performance of a decision level fusion of the sequence to compare with superresolution face recognition results. The database of moving faces and people acquired by O'Toole et al. [14] was used for this study, specifically the parallel gait video datasets and close-up mug shots. Face recognition performance with the LR and superresolved imagery was assessed with the local region principal component analysis (LRPCA) face recognition algorithm [11,12] developed at Colorado State University. Correct verification rates are calculated and compared at three face resolutions/scales in terms of eye-to-eye distance corresponding to different subject-to-camera distances within the video footage. Results show that superresolution image reconstruction significantly improves face recognition verification rates at the examined mid- and close ranges, with some improvement at the far range.

2. Methodology

A. Database

Parallel gait videos and static mug shot images from the video database of moving face and people [14] are used for this work. The parallel gait video shows the subject moving towards the camera from 13.6 m away to approximately 1.5 m away, providing a large sequence of face imagery at different face resolutions from which query sets can be formed. A sample frame containing the subject at the far range is shown in Fig. 1. Since faces in the parallel gait videos are acquired from the frontal perspective, the corresponding frontal mug shots are used to form the gallery set. The resolution of the videos as well as of the frontal mug shot is $720 \times 480$ pixels (note that the corresponding pixel count is $345,600$ pixels, substantially less than even one megapixel). The videos were acquired with compression using a Canon Optura Pi digital video camera. Figure 2 shows (a) close range face image of a subject, (b) close range face image downsampled by a factor of 3 to simulate far range using procedure of [10], and (c) far range face image of the subject taken from the same video. The downsampling procedure of [10] involved convolving the close range face image with a Gaussian filter of $d/4$ and then downsampling by $d$, where $d$ is the...
downsampling factor. Note that while the compression artifacts are not obvious in the close range face image (simply because facial features consist of many pixels), the compression distortions are highly pronounced in the actual far range face image taken from the same video. Simulating the far range face image by downsampling, as past studies have done in examining superresolution for face recognition, does not closely resemble actual far range face imagery due to the effects of compression. The parallel gait videos used in this work emulate real-world compressed surveillance footage and enable a realistic assessment of superresolution benefit for face recognition.

B. Superresolution

This study used the reconstruction-based superresolution algorithm of Young and Driggers [13], which utilizes a series of undersampled/aliased LR images to reconstruct an alias-free high-resolution image. This reconstruction-based superresolution algorithm consists of a registration stage and a reconstruction stage. The registration stage computes the gross shift and subpixel shift of each frame in the sequence with respect to the reference frame using the correlation method in the frequency domain. The reconstruction stage uses the error-energy reduction method with constraints in both spatial and frequency domains, generating a superresolved image that improves the high-frequency content that was lost or corrupted due to the undersampling/aliasing of the sensor. The resolution improvement factor of the superresolved image is the square root of the number of frames used to reconstruct the superresolved image. A necessary condition for superresolution benefit is the presence of different subpixels shifts between frames to provide distinct information from which to reconstruct a high-resolution image. The natural movement of the subject in the parallel gait video provided this necessary subpixel shift.

C. Query Sets

Frame sequences at three different subject-to-camera distances are extracted from each subject’s parallel gait video: far range (∼13 m), midrange (∼9 m), and close range (∼5 m). The face resolutions (in terms of eye-to-eye distances) corresponding to the far, mid-, and close ranges are 5–10, 15–20, and 25–30 pixels, respectively. Three query sets are constructed for each range: (a) original LR imagery (taken as the first frame within the sequence), (b) superresolved imagery using four consecutive LR frames (SR4), and (c) superresolved imagery using eight consecutive LR frames (SR8). SR4 and SR8 enable an assessment of the impact of the number of frames used for superresolution on face recognition performance. The resolution improvement factor in the x and y directions is 2 and 2.8 for SR4 and SR8, respectively. Consequently, the size of the SR4 face image is a factor of 2 larger in the x and y dimensions than the corresponding LR face image; the size of the SR8 face image is a factor of 2.8 larger in the x and y dimensions than the LR face image. A total of nine different query sets (Table 1) are generated to evaluate the improvement in face recognition with superresolution; each query set contains 200 subjects with one image per subject.

D. Face Recognition

This study used the state-of-the-art baseline LRPCA face recognition algorithm developed by Bolme et al.

Table 1. Query Set Nomenclature

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Range</th>
<th>Number of Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-resolution</td>
<td>5–10 Pixels</td>
<td>4 frames</td>
</tr>
<tr>
<td>Superresolved</td>
<td>15–20 Pixels</td>
<td>8 frames</td>
</tr>
<tr>
<td>LR5–10</td>
<td>LR15–20</td>
<td>LR25–30</td>
</tr>
<tr>
<td>SR85–10</td>
<td>SR815–20</td>
<td>SR825–30</td>
</tr>
</tbody>
</table>

Top row represents subject-to-camera range in terms of eye-to-eye distance.
The high-frequency content recovered in the superresolved imagery is expected to aid principal component analysis (PCA)-based methods, since current PCA-based algorithms often employ a large number of basis vectors (on the order of thousands for this study). As a preprocessing step, all query and gallery images are cropped and normalized to 256 × 256 pixels through bilinear interpolation using manually defined eye coordinates. The LRPCA algorithm was trained using the “The Good, The Bad, and The Ugly” (GBU) subset of the Multiple Biometric Grand Challenge, containing a total of 522 subjects. Training on a separate dataset distinct from the query and gallery sets avoids biasing the performance of the algorithm. The gallery set corresponding to each query set consists of one frontal mug shot for each subject.

E. Performance Measurement

For each query set and gallery, the output of the LRPCA face recognition algorithm is a similarity matrix $S$ containing the similarity measure between every probe in the query set and every gallery image. Note that for this work, both the query and gallery sets contain a single image of each subject ($N = 200$ subjects total); therefore, the similarity matrix is a $N \times N$ square matrix with the diagonal elements containing the $N$ match scores and $N(N - 1)$ off-diagonal elements containing the nonmatch scores.

1. Receiver Operating Characteristic Curves

The similarity matrix is used to compute the correct verification rates as well as the corresponding false accept rates (FARs). In the verification model, the face recognition system is tasked with deciding whether the person in the probe image $p_i$ is the same as the person in the gallery image $g_j$ [15]. The decision is made based on the Neyman–Pearson theorem, testing whether the similarity score between $p_i$ and $g_j$ exceeds a given threshold $t_0$. The correct verification rate is computed by tallying the number of diagonal elements (match scores) that exceed $t_0$, and the FAR is computed by tallying the number of off-diagonal elements (nonmatch scores) that exceed $t_0$ [15]. Receiver operating characteristic (ROC) curves were generated by thresholding the similarity matrix $S$ at various thresholds across the range from $S_{\text{min}}$ to $S_{\text{max}}$. For each of the nine query sets listed in Table 1, a ROC curve was constructed in this manner.

2. Performance with Respect to Range

To visualize face recognition performance with respect to subject-to-camera range, the correct verification rates are plotted with respect to range at commonly used FARs of 0.01 and 0.05 for LR, SR4, and SR8. Confidence intervals are calculated and overlaid onto the plots to assess the statistical reliability of the performance improvement achieved with superresolution image reconstruction.

3. Confidence Intervals

To indicate the reliability of the calculated correct verification rates, 95% confidence intervals are determined using the bootstrap method, specifically following the procedure for biometrics detailed in [16]. The bootstrap is a nonparametric approach that makes no assumptions about the error distribution and is preferable to parametric techniques when the underlying distribution is unknown, as is the case for biometrics. Bootstrap involves resampling the available data (match scores for this study) many times with replacement to generate confidence intervals. For this work, the probe set contained one image per subject and the gallery contained one image per subject for LR, SR4, and SR8 at each range, satisfying the independent and identically distributed (i.i.d.) requirement of the bootstrap.

Recall that the output of the LRPCA algorithm is a similarity matrix $S$ containing scores of the similarity between a probe and all gallery images. Also recall that $S$ contains $M = N$ match scores along the diagonal and $N(N - 1)$ mismatch scores, where $N = 200$ is the number of subjects. For a given $t_0$, let the verification rate estimate be defined by the equation

$$\hat{F}(t_0) = \frac{1}{M} \sum_{i=1}^{M} 1(X_i \geq t_0).$$

where $X$ denotes the set of $M$ match scores and $1$ is the indicator function [16]. The bootstrap generates $X^* = \{X_1^*, \ldots, X_M^*\}$ by resampling with replacement, and then calculates $\hat{F}^*(t_0)$. This resampling procedure is repeated $B$ times ($B = 10,000$ for this work), generating bootstrap estimates $\hat{F}^* = (\hat{F}^*_1, \hat{F}^*_2, \ldots, \hat{F}^*_B)$. The lower and upper bounds of the 95% confidence interval is determined as values corresponding to the 2.5th and 97.5th percentile of the histogram of the $B$ bootstrap estimates $\hat{F}^*$.

4. Face Recognition Performance of Individual Frames and Decision Level Fusion

To address the question of how face recognition performance with superresolution compares to face recognition performance of individual frames within the LR sequence as well as to the performance of a decision level fusion scheme, further analysis was conducted. Superresolution exploits the additional spatial information contained in the temporal dimension (i.e., multiple frames) to reconstruct a more detailed face image for recognition, and therefore is expected to exceed the face recognition performance of any single frame within the LR sequence. To validate this expectation, face recognition performance was computed for each of the eight LR frames used to reconstruct SR8 and compared to face recognition performance of SR8. Furthermore, a simple fusion scheme for the LR frame sequence was implemented by averaging the similarity matrices from the eight LR frames. Fusion by averaging of similarity

\[20\text{ June 2012 / Vol. 51, No. 18 / APPLIED OPTICS 4253}\]
matrices exploits the spatial information in the temporal domain at the decision level and is expected to be an upper bound on the face recognition performance of any individual frame. Face recognition performance with superresolved imagery is then compared to the performance of this decision level fusion scheme.

A total of 24 query sets (eight query sets per range corresponding to each of the eight frames in the LR sequence used to reconstruct SR8; Table 2) was generated to assess the variation in face recognition performance with respect to individual frames. Note that the LR sequence is an eight frame clip of the subject walking towards the camera. Due to the fast frame rate (30 Hz) and relatively slow speed of the subjects (walking speed), the change in pose is insignificant across the eight frames in the sequence. At the far range, since the change in face size across the eight frames does not exceed a single pixel, the same eye coordinates in terms of \((x, y)\) pixel locations are used for all eight frames. At the mid- and close ranges, face size does enlarge by a few pixels across the frames; therefore, eye coordinates are manually picked for all eight frames instead of for just the first frame as in the far range. Once the similarity matrix for each query set is computed with the LRPCA algorithm, ROC curves of face recognition performance with respect to individual frames can be generated. The decision level fusion method (denoted LR\textsuperscript{ave}) averages the similarity matrices across the eight frames at each range and generates the ROC curve using the averaged similarity matrix for comparison with face recognition using superresolved imagery.

3. Results and Discussion

A. Superresolved Imagery

Superresolved face imagery and original low resolution face imagery are shown in Fig. 3 at different ranges. At the far range, the LR image is heavily pixilated and distorted by compression, yielding a coarse facial outline and few facial features. Super-resolution with four and eight frames enhances the facial outline and some facial details, but compression artifacts have almost completely eliminated facial details in the low resolution frames, preventing significant facial feature enhancement.

As range decreases, the camera captures finer details and the detrimental impact of compression on facial features lessens because the size of these features is now larger. At the midrange, SR4 and SR8 produce considerable enhancement of the subject’s facial details. As range continues to decrease to the close range, superresolution benefit decreases as facial features become more and more defined in the low resolution imagery. Although the close range SR images may not appear significantly enhanced visually, facial recognition algorithms may still benefit from superresolution as these algorithms operate on different principles than the human visual system.

To provide a more objective assessment of the increase in high-frequency content with superresolution, spectral analysis was conducted using LR and superresolved face imagery. Let the following equations denote the cumulative power spectrum in wavenumber \(k_x\) and \(k_y\), respectively, where \(F(k_x, k_y)\) is the Fourier transform of the considered image:

\[
S_1(k_x) = \sum_{k_y} |F(k_x, k_y)|^2, \quad (2)
\]

\[
S_2(k_y) = \sum_{k_x} |F(k_x, k_y)|^2. \quad (3)
\]

For this study, the cumulative power spectrum is computed over the spatial region consisting of the eyes, which is a critical area for face recognition algorithms. Figure 4 shows the computed \(k_x\)-domain spectrums for LR and SR8 eye region at the midrange. The circled part of the plot in Fig. 4 represents the

![Table 2. Query Set Nomenclature for Evaluation of Face Recognition with Respect to Low-Resolution Frame](image-url)

<table>
<thead>
<tr>
<th>Frame 1</th>
<th>Frame 2</th>
<th>Frame 3</th>
<th>Frame 4</th>
<th>Frame 5</th>
<th>Frame 6</th>
<th>Frame 7</th>
<th>Frame 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Far range</td>
<td>LR\textsuperscript{1-10}</td>
<td>LR\textsuperscript{1-10}</td>
<td>LR\textsuperscript{1-10}</td>
<td>LR\textsuperscript{1-10}</td>
<td>LR\textsuperscript{1-10}</td>
<td>LR\textsuperscript{1-10}</td>
<td>LR\textsuperscript{1-10}</td>
</tr>
<tr>
<td>Midrange</td>
<td>LR\textsuperscript{1-20}</td>
<td>LR\textsuperscript{1-20}</td>
<td>LR\textsuperscript{1-20}</td>
<td>LR\textsuperscript{1-20}</td>
<td>LR\textsuperscript{1-20}</td>
<td>LR\textsuperscript{1-20}</td>
<td>LR\textsuperscript{1-20}</td>
</tr>
<tr>
<td>Close range</td>
<td>LR\textsuperscript{1-30}</td>
<td>LR\textsuperscript{1-30}</td>
<td>LR\textsuperscript{1-30}</td>
<td>LR\textsuperscript{1-30}</td>
<td>LR\textsuperscript{1-30}</td>
<td>LR\textsuperscript{1-30}</td>
<td>LR\textsuperscript{1-30}</td>
</tr>
</tbody>
</table>

*Superscript denotes frame number, while subscript denotes range in terms of eye-to-eye distance in pixels.*
high-frequency band recovered with superresolution using a sequence of eight aliased low-resolution frames.

Furthermore, the enhancement in edge contrast is demonstrated in Fig. 5, which plots the intensity values of LR and SR8 at the midrange along a horizontal profile across the eyes. The improvement in edge contrast is especially noticeable across the pupils (horizontal axis = ±10) in Fig. 5.

B. Receiver Operating Characteristic Curves

ROC curves at the 5–10, 15–20, and 25–30 pixel eye-to-eye distance are shown in Figs. 6, 7, and 8, respectively. Each figure contains three ROC curves corresponding to the LR (red dotted line), superresolved using four frames (SR4; dashed green line), and superresolved using eight frames (SR8; solid blue line) imagery. At the far range, the ROC curves for SR45–10 and SR85–10 lay slightly but consistently above the ROC curve for LR5–10, suggesting that face imagery at the far range possessed too few details for superresolution to provide any substantial enhancement to aid the LRPCA face recognition algorithm. At the midrange, SR815–20 outperformed SR415–20, which in turn outperformed LR15–20 across the FARs from FAR = 0.001 to 0.6; superresolution effectively enhances facial details at the midrange to yield a large improvement in face recognition performance over the baseline LR imagery. At the close range, while both SR425–30 and SR825–30 produced higher face recognition performance than LR25–30 at all FARs, the improvement is not as large as
achieved at the midrange since the original imagery already contains detailed facial features.

C. Performance with Respect to Range

For practical applications, performance at low FARs is of particular interest; therefore, in verification rate as a function of range is examined at FAR = 0.01 and 0.05 in Fig. 9. At all FARs, the LR curve exhibits a slight knee at the midrange; the knee is more pronounced for the SR4 and SR8 curves, signifying that the change in performance with respect to range is more nonlinear for SR imagery.

At the far range, the already limited facial details are distorted by compression, preventing substantial enhancement by superresolution image reconstruction. At the midrange where the greatest benefit from superresolution is observed, the correct verification rate is 21.0% for SR4_{15–20} and 27.0% for SR8_{15–20} compared to 14.5% for LR_{15–20}, resulting in an improvement of 31.6% and 57.9%, respectively.

A large improvement of the verification rate occurs from the far range to the midrange, but the improvement is visibly smaller from the midrange to the close range. For SR8, which produced effective eye-to-eye distances ~2.8 times the original size; the verification rate exhibited only a small improvement from the midrange to the close range. These results are consistent with the findings of [1,2] that showed the improvement in face recognition performance slowed considerably once the eye-to-eye distance surpassed approximately 30 pixels.

To generate the confidence intervals shown in Fig. 9, the procedure described in Subsection 2.E.3 was performed using the similarity matrix $S$ for LR, SR4, and SR8 at each of the three ranges. For the far range, although face recognition improves with superresolution, the confidence intervals overlap for LR, SR4, and SR8, suggesting that no significant benefit is achieved with superresolution. At the close range, the confidence interval for SR4 exhibits a partial overlap with that of LR, and the confidence interval for SR8 exhibits only a slight overlap with that of LR. The small overlaps suggest that superresolution improves recognition rates for face recognition with high reliability, especially when using eight frames. At the midrange, the confidence interval for SR4 partially overlaps with that of LR, and the confidence interval for SR8 exhibits no overlap at all with that of LR. The lack of any overlap demonstrates that the face recognition performance improvement achieved with superresolution using eight frames is not only highly reliable, but is also significant for the midrange where eye-to-eye distance is between 15–20 pixels.

D. Face Recognition Performance of Individual Frames and Decision Level Fusion

To examine the LRPCA face recognition performance of individual frames within the LR sequence, the ROC curves of each LR frame are computed and shown in Figs. 10–12. The ROC curve of the simple decision level fusion method derived from averaging the similarity matrices across the eight frames.
frames (LR\text{ave}) at each range is overlaid onto the plots. Averaging the similarity matrices exploits the spatial information across the eight frames at the decision level for face recognition and is compared against face recognition with superresolution (SR8).

Figures 10–12 show the ROC curves for each of the 8 LR query sets corresponding to different frames at the far, mid-, and close ranges, respectively. The ROC curve for the average similarity scores (LR\text{ave}) across frames is shown in bold red along with the SR8 ROC curve shown in bold blue. The ROC curves for the LR frames exhibit some variation, but tend to be clustered together and lay within the confidence intervals as computed in Subsection 3.C. Note that there is no observable ordering of the ROC curves for LR\textsuperscript{1} to LR\textsuperscript{8} from lowest to highest. This verifies that the scale change across the eight frames as subject walks towards the camera is minor and does not produce any patterns in the ordering of the ROC curves. At the far range, the ROC curve for the first frame (LR\textsuperscript{5-10}) interestingly lay above the seven other LR frames which closely overlap with each other. This may be due to the definition of the eye coordinates based on the first frame, which were then reused for the remaining seven frames at the far range. This eye coordinate definition procedure may have produced a slightly more accurate eye coordinate selection for the first frame than the other frames, even though the change in eye coordinates did not exceed an integer pixel across the frames in the sequence at the far range.

---

**Fig. 10.** (Color online) ROC curves at the far-range for each low resolution frame (superscript 1–8). The ROC curve for LR\text{ave} is generated by averaging similarity matrices of the eight individual frames and generating the ROC curve.

**Fig. 11.** (Color online) ROC curves at the midrange for each low resolution frame (superscript 1–8).
The ROC curves for LR\textsuperscript{ave} in general lay above the ROC curve of any individual LR frame. Since LR\textsuperscript{ave} is a decision level fusion in exploiting the spatial information across the eight frames, it is not unexpected that the ROC for LR\textsuperscript{ave} tends to be an upper bound for the ROC curve of any individual frame. However, the ROC curve for SR\textsuperscript{8} lay above the ROC curve of LR\textsuperscript{ave} at all three ranges, showing that super-resolution image reconstruction is a more effective method in exploiting the spatial information across the temporal dimension to improve face recognition performance.

To provide a quantitative measure of overall face recognition performance of the LRPCA face recognition algorithm using superresolved and LR imagery, the area under the curve (AUC) is computed for each ROC curve in Figs. 10–12 across FAR ∈ [0, 1] and tabulated in Table 3. Note that the maximum possible value for the AUC is 1. The AUCs for LR are gener-ally consistently close to each other across frames 1–8 at the three ranges. The “best frame” in terms of AUC is the 1st frame for the far range, 8th frame for the midrange, and 6th frame for the close range as underlined in Table 3. The AUC for SR\textsuperscript{8} is 5.1\% larger than the best frame at the far range, 7.65\% larger than the best frame at the midrange, and 5.04\% larger than the best frame at the close range. Furthermore, the AUC for SR\textsuperscript{8} is 14.36\% larger than LR\textsuperscript{ave} at the far range, 2.88\% larger than LR\textsuperscript{ave} at the midrange, and 1.38\% larger than LR\textsuperscript{ave} at the close range. Note that although the AUC for SR\textsuperscript{8} is only a few percent better than LR\textsuperscript{ave} at the mid- and close ranges, the increase is still substantial as the AUC was computed over the whole range of FARs (FAR ∈ [0, 1]); typically, ROC curves for detection/classification algorithms in a given experiment overlap at higher FARs (ex. FAR > 0.1). Therefore, the results of Table 3 show that super-resolution provides improvement in LRPCA face recognition performance compared to any frame as well as to the decision level fusion across the eight frames.

Using the best frames in terms of AUC from Table 3 at each range (denoted LR\textsuperscript{*}), verification rates with respect to range are plotted in Fig. 13 at FAR of 0.01 and 0.05. SR\textsuperscript{8} outperforms LR\textsuperscript{*} as well as LR\textsuperscript{ave}, with small to no overlap of confidence intervals at the midand close ranges. At the midrange, the verification rate is 0.45 for SR\textsuperscript{8}, 0.37 for LR\textsuperscript{ave}, and 0.31 for LR\textsuperscript{*}, representing a 21.6\% improvement and a 45.2\% improvement over LR\textsuperscript{ave} and LR\textsuperscript{*} at FAR = 0.05, respectively. At the close range, although the performance improvement achieved with SR\textsuperscript{8} is not as significant as the midrange, the benefit of super-resolution is still substantial.

For surveillance systems on residential and commercial properties where low cost cameras are prevalent, faces of individuals captured on camera are commonly between 15 and 30 pixels across in terms of eye-to-eye distance which correspond to the examined mid- and close ranges. Superresolution

![Fig. 12. (Color online) ROC curves at the close-range for each low resolution frame (superscript 1–8).](image)

Table 3. Area under the Curves for LR\textsuperscript{i}, Where i ∈ [1,8] Denotes the Frame Number, AUC for LR\textsuperscript{ave} (Computed from the ROC of the Average Similarity Scores across the Eight Frames), and AUC for SR\textsuperscript{8} at Far, Mid-, and Close Ranges\textsuperscript{*}

<table>
<thead>
<tr>
<th></th>
<th>LR\textsuperscript{1}</th>
<th>LR\textsuperscript{2}</th>
<th>LR\textsuperscript{3}</th>
<th>LR\textsuperscript{4}</th>
<th>LR\textsuperscript{5}</th>
<th>LR\textsuperscript{6}</th>
<th>LR\textsuperscript{7}</th>
<th>LR\textsuperscript{8}</th>
<th>LR\textsuperscript{ave}</th>
<th>SR\textsuperscript{8}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Far range</td>
<td>0.6200</td>
<td>0.5553</td>
<td>0.5339</td>
<td>0.5655</td>
<td>0.5311</td>
<td>0.5474</td>
<td>0.5583</td>
<td>0.5666</td>
<td>0.5697</td>
<td>0.6516</td>
</tr>
<tr>
<td>Midrange</td>
<td>0.7529</td>
<td>0.7514</td>
<td>0.7557</td>
<td>0.7542</td>
<td>0.7467</td>
<td>0.7411</td>
<td>0.7625</td>
<td>0.7756</td>
<td>0.8115</td>
<td>0.8349</td>
</tr>
<tr>
<td>Close range</td>
<td>0.7559</td>
<td>0.7428</td>
<td>0.7503</td>
<td>0.7352</td>
<td>0.7584</td>
<td>0.7828</td>
<td>0.7805</td>
<td>0.7679</td>
<td>0.8105</td>
<td>0.8217</td>
</tr>
</tbody>
</table>

*Underlined numbers denote the best frame in terms of AUC at each range.
is expected to provide significant benefits in enhancing the LR face images and improving facial recognition performance.

4. Conclusion
Using a video database similar to real world surveillance footage, this study shows that superresolution provides considerable benefits for the state of the art baseline LRPCA face recognition algorithm at the examined mid- and close ranges. In surveillance applications, low-cost cameras and oftentimes the far distance of individuals result in a very limited number of face pixels, severely affecting face recognition performance. Superresolution image reconstruction can be used to enhance the high-frequency content of low resolution surveillance imagery, improving face recognition performance and potentially aiding the nation in law enforcement and homeland security applications.

The authors thank Professor Ross Beveridge, David Bolme, Stephen Won, and Martha Givan for their help, as well as the reviewers for their valuable comments and suggestions.

References