# Performance Assessment of Face Recognition Using Super-Resolution


Shuowen Hu
U.S. Army Research Laboratory
2800 Powder Mill Rd.
Adelphi, MD 20783
(301)394-2526
shuowen.hu@arl.army.mil

Robert Maschal
U.S. Army Research Laboratory
2800 Powder Mill Rd.
Adelphi, MD 20783
(301)394-0437
robert.maschal@arl.army.mil

S. Susan Young
U.S. Army Research Laboratory
2800 Powder Mill Rd.
Adelphi, MD 20783
(301)394-0230
ssyoung@arl.army.mil

Tsai Hong Hong
National Institute of Standards and Technology
100 Bureau Dr.
Gaithersburg, MD 20899
(301)975-3444
hongt@nist.gov

Jonathon P. Phillips
National Institute of Standards and Technology
100 Bureau Dr.
Gaithersburg, MD 20899
(301)975-5348
jonathon@nist.gov



## ABSTRACT
The accuracy of face recognition algorithms is dependent on the resolution of the imagery, specifically the number of pixels contained within the face. Using a sequence of frames from low-resolution video, super-resolution image reconstruction can form a higher resolution image, aiding the face recognition stage for improved performance. In this work, images from a video database of moving faces and people are used to assess the performance improvement of face recognition using super-resolution.



## Categories and Subject Descriptors
D.3.3 [**Image Processing and Computer Vision**]: Enhancement.

## General Terms
Algorithms, Performance

## Keywords
Super-resolution image reconstruction, face recognition, image enhancement, video surveillance


## 1.1 INTRODUCTION
Performance of face recognition algorithms is dependent on a host of factors including resolution/scale, lighting conditions, facial expression, head orientation, occlusion, and image compression. In general surveillance video applications, the imagery is not only low-resolution, typically (760×480) pixels, but the person of interest may be at a far distance with respect to the camera, further compounding the problem of limited number of face pixels. The video database of moving faces and people [1] contains similar imagery as surveillance videos, and was used in this work. Using a sequence of frames from low-resolution video, super-resolution image reconstruction can form a higher resolution image, aiding face recognition systems.



*PerMIS'10*, September 28-30, 2010, Baltimore, MD, USA.


Super-resolution image reconstruction algorithms are generally composed of two stages: a registration stage and a reconstruction stage. During the registration stage, the shift of a given frame with respect to a reference frame is computed to subpixel accuracy, which is then utilized by the reconstruction stage to interpolate the low-resolution frames onto a higher-resolution grid. A necessary condition for super-resolution algorithms is the presence of differing subpixel shifts between frames within the low-resolution video sequence to provide distinct information from which the super-resolved image can be constructed. In this work, the super-resolution algorithm developed by Young and Driggers [2] is used to assess the performance improvement of face recognition with super-resolved imagery. For improved registration accuracy, the super-resolution algorithm separates the registration stage into a gross shift (i.e., integer pixel shift) estimation substage and a subpixel shift (i.e., decimal pixel shift) estimation substage; both substages are based on the correlation method in the frequency domain to estimate shifts. The reconstruction stage uses the error-energy reduction method with constraints in spatial and frequency domains to generate a super-resolved image with a resolution improvement factor of the square root of the number of frames.

Although face recognition performance improves with increased resolution, the improvement is highly nonlinear. Boom et al. [3] examined the effect of image resolution on performance of a face recognition system using face image resolutions of (8×8) pixels to (128×128) pixels. Their results indicated that the performance of the face recognition algorithm in terms of equal error rate exhibited large improvement from (8×8) pixels to (32×32) pixels, but only slight improvement from (32×32) pixels to (128×128) pixels; these results suggests that face recognition performance is highly dependent on resolution for lower resolutions, but ceases to depend on face resolution past some threshold resolution. The results from the Facial Recognition Vendor Test 2000 [4] confirm the findings of Boom et al. [3]. To assess the impact of resolution on face recognition performance, Blackburn et al. [4] used a high resolution gallery for training and varied the resolution (in terms of eye-to-eye distance) of the query set using a standard reduction algorithm. The eye-to-eye distance is defined as the number of pixels between the left and right eyes of the subject in the image. Results show that the tested face recognition systems yielded similar performance for query sets with eye-to-eye distance from 60 pixels to 30 pixels, but at an eye-to-eye distance of 15 pixels, performance becomes severely degraded for some algorithms. Wheeler et al. [5] examined the performance of face recognition for query images with and without super-resolution at varying resolutions in terms of eye-to-eye distance (note that this study used limited query sets containing on average only 23 subjects). Their results show that performance of face recognition improves with super-resolution compared to that without super-resolution once the eye-to-eye distance falls below 24 pixels. In general surveillance video applications, it is not uncommon for the width of the face captured by the camera to be less than 20 pixels with corresponding eye-to-eye separation of less than 11 pixels or 12 pixels. Therefore, super-resolution may aid face recognition systems to achieve improved recognition.

The objective of this work is to comprehensively assess the performance improvement of face recognition with super-resolution using the video database of moving faces and people [1] at three subject-to-camera ranges in terms of eye-to-eye pixel distances for the query sets: (a) (5 to 10) pixels eye-to-eye distance, (b) (15 to 20) pixels, and (c) (25 to 30) pixels. Receiver operating characteristic curves are generated for the low-resolution original query sets and the super-resolved query sets at each range.

# 2. METHODOLOGY

## 2.1 Database

The video database of moving faces and people [1], containing both static images as well as videos, is used for this study. Specifically, the following components of the database are used: facial mug shot still images and parallel gait video. The facial mug shot still images are close-up camera acquisitions of the subjects in the database, and therefore can be used to form the high resolution (in terms of eye-to-eye pixel distance) target set for the face recognition algorithm. Only the frontal mug shots are used to form the target set because the face recognition algorithm used in this study was designed for frontal face images. The parallel gait video shows the subject moving towards the camera, enabling several different frame sequences to be selected for super-resolution that have different face resolutions (in terms of eye-to-eye distances).

Note that in order to quantify the improvement in performance achieved with super-resolution, the baseline performance must be also established by using the original low-resolution imagery as the query set. To form the baseline query set, the first frame of each sequence is used.

## 2.2 Super-resolved Query Sets

Face recognition performance is assessed at three ranges in terms of eye-to-eye distance: (a) (5 to 10) pixels, (b) (15 to 20) pixels, and (c) (25 to 30) pixels. Note that the pixel uncertainty in the eye-to-eye distance measurement for each subject is $\pm 1$ pixel. For each range, three different query sets are formed: (i) original low-resolution (LR) imagery, (ii) super-resolved (SR) imagery using four consecutive frames from the video (producing a resolution improvement factor of two in the x- and y-directions and improving the high frequency content of the imagery), and (iii) super-resolved imagery using eight consecutive frames from the video (producing a resolution improvement factor of ≈2.8 in the x- and y-directions and improving the high frequency content of the imagery). Note that (iii) contained the four frames from (ii) and four additional consecutive frames. Super-resolution image reconstruction generates an image with resolution increased in the x- and y-directions by a factor of the square root of the number of frames [2], which is the resolution improvement factor given above. A total of nine different query sets (nomenclature in Table 1) each containing 80 subjects with one image per subject are generated to assess performance. These query sets enable a performance assessment of face recognition as a function of subject range and image resolution.

Table 1. Query set nomenclature. Top row represents subject range from camera in terms of eye-to-eye distance.

| | **(5 to 10) Pixels** | **(15 to 20) Pixels** | **(25 to 30) Pixels** |
|---|---|---|---|
| **Low-Resolution** | $LR_{(5 \text{ to } 10)}$ | $LR_{(15 \text{ to } 20)}$ | $LR_{(25 \text{ to } 30)}$ |
| **Super-Resolved 4 Frames** | $SR4_{(5 \text{ to } 10)}$ | $SR4_{(15 \text{ to } 20)}$ | $SR4_{(25 \text{ to } 30)}$ |
| **Super-Resolved 8 Frames** | $SR8_{(5 \text{ to } 10)}$ | $SR8_{(15 \text{ to } 20)}$ | $SR8_{(25 \text{ to } 30)}$ |

## 2.3 Face Recognition Algorithm

The local region principal component analysis (LRPCA) face recognition algorithm developed by Colorado State University [6] and based on the principal component analysis (PCA) method of Bolme et al. [7] is used for this work. PCA-based methods are expected to benefit from the increased high frequency information using super-resolution image reconstruction. The LRPCA algorithm was first trained on imagery from "The Good, The Bad, and The Ugly" (GBU) subset of the Multiple Biometric Grand Challenge consisting of 522 subjects. Note that to avoid biasing the algorithm, training on a separate dataset (with respect to the query and target sets) is preferable. The parallel gait videos in the video database of moving faces and people [1] contained imagery acquired under different lighting conditions and backgrounds than the facial mug shots, so the use of the GBU for training is appropriate as the GBU database also contains imagery acquired under varying conditions, but of different subjects. Each image in the query and target datasets is first normalized to a resolution of (512×512) pixels using manually determined eye coordinates. The query datasets are listed in Table 1, and the target dataset for each query set consists of frontal mug shots of the query subjects.

## 2.4 Measurements and Metrics

The LRPCA face recognition algorithm generates a similarity matrix which contains the distances between the query and target images, with a larger similarity score indicating higher similarity. The algorithm then scores the similarity matrices to generate receiver operating characteristic (ROC) curves, plotting the correct verification rate as a function of the false accept rate (also known as the false alarm rate, FAR). In the verification model, the face recognition system attempts to make the determination whether a person $q_i$ in the query set $Q$ matches a person $t_j$ in the target set $T$ at some threshold $c$ using the similarity matrix [9]. By varying the threshold $c$ and using a round robin evaluation procedure for the entire target set, probabilities of verification with corresponding false alarm rates are produced, enabling an ROC curve to be plotted. Note that since nine query sets are used in this work, nine ROC curves are generated and the area under the ROC curves are tabulated to assess overall performance. To visualize performance with respect to range, the correct verification rate is also plotted as a function of range at false alarm rate of 0.01, 0.05, and 0.10.

# 3. RESULTS AND DISCUSSION

## 3.1 Super-Resolved Imagery

Super-resolved imagery using four and eight frames are shown in Figure 1, along with original low-resolution imagery for three subjects at (5 to 10) pixels, (15 to 20) pixels, and (25 to 30) pixels eye-to-eye distances. Note that the native sizes of the LR, SR4, and SR8 images are different, but have been normalized for display and comparison purposes in Figure 1. The (5 to 10) pixels distance will also be referred to as the far range, (15 to 20) pixels as the mid-range, and (25 to 30) pixels as the close range.

For the far range, the LR image is highly pixilated and distorted by JPEG compression, yielding a coarse facial outline and few facial features. Super-resolution with four frames enhances the facial outline significantly along with some facial details, and super-resolution with eight frames produces a further improvement in terms of visible quality at the far range. For some subjects at the far range, the long subject-to-camera distance and compression artifacts almost completely eliminated facial details in LR imagery so that super-resolution yield little benefit; super-resolution does not create information where there was none to begin with. For the mid-range, super-resolution consistently produced improved imagery with finer facial features as seen in Figure 1 (outline of glasses is even visible in the SR8 image).

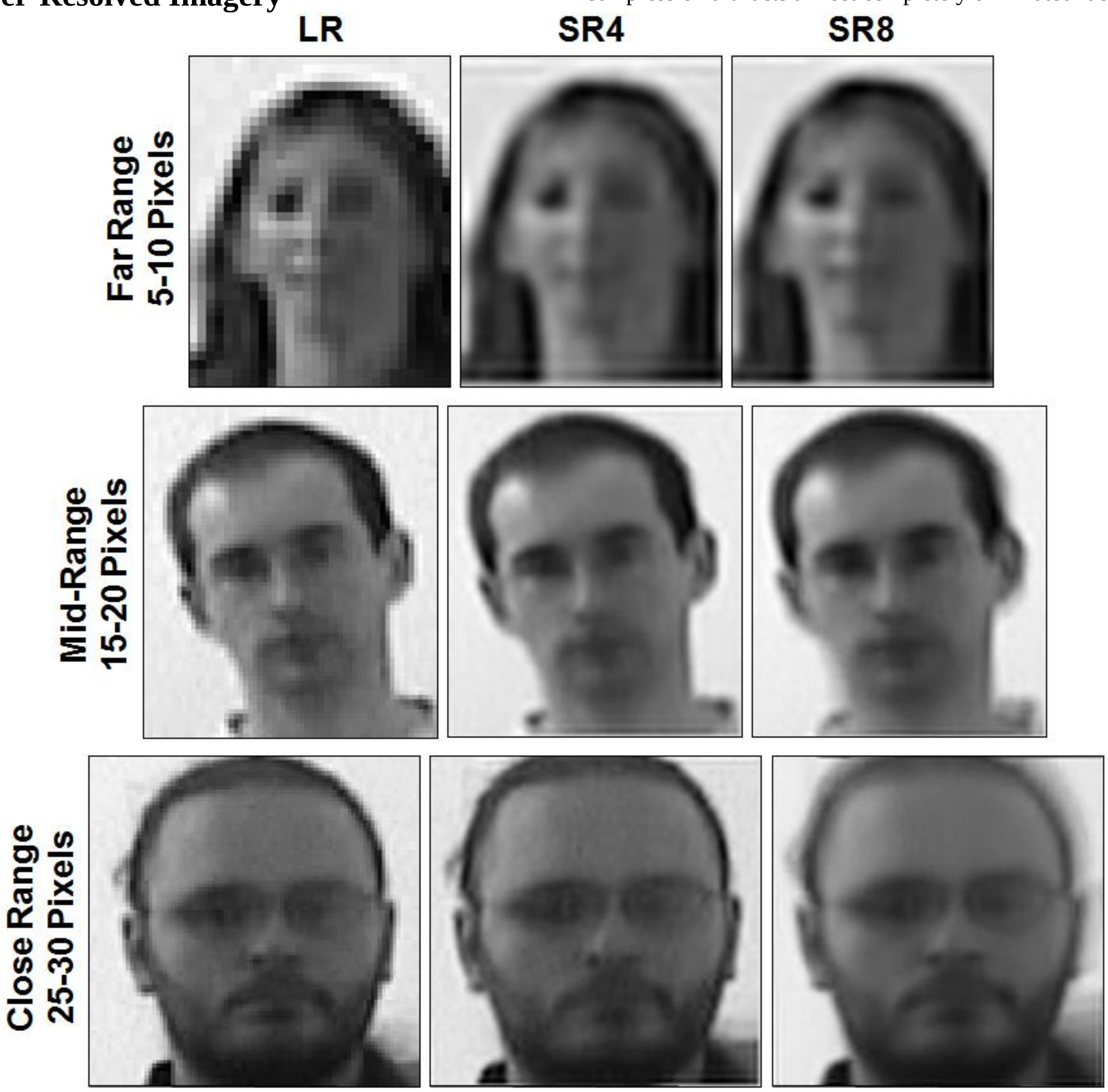


Figure 1. Original low-resolution (LR) imagery and super-resolved imagery (4 frames – SR4, 8 frames – SR8) for three subjects at three eye-to-eye distances: (5 to 10) pixels, (15 to 20) pixels, and (25 to 30) pixels. The database is obtained through http://bbs.utdallas.edu/facelab/otoole/database.htm, and all subjects who appear in this figure have consented to having their images published.

As range decreases, the camera can capture finer details; decreasing range also lessens the detrimental impact of compression on facial features. Therefore, the LR imagery already contains finer details at closer ranges, and super-resolution benefit decreases causing the close range LR image in Figure 1 to appear visually comparable to the close range SR images. Note that the SR8 close range image exhibits blurring, since scale changes are much more pronounced between frames at close ranges. Although the close range SR images may not appear visually enhanced, facial recognition algorithms may still benefit from super-resolution as these algorithms operate on different principles than the human visual system.

## 3.2 Receiver Operating Characteristic Curves

Receiver operating characteristic curves at the (5 to 10) pixel eye-to-eye distance (red), (15 to 20) pixel eye-to-eye distance (green), and (25 to 30) pixel eye-to-eye distance (blue) are shown in Figures 2, 3, and 4, respectively. Each figure contains three ROC curves corresponding to the low-resolution imagery and super-resolved imagery using four and eight frames. The area under the curve (AUC) for each ROC curve is tabulated in Table 2. At the far range, $SR4_{5\text{-}10}$ exhibits a slightly but consistently higher performance than LR, resulting in AUC of 0.6230 compared to AUC of 0.6028 for $LR_{5\text{-}10}$. Interestingly, $SR8_{5\text{-}10}$ has a slightly lower AUC than $LR_{5\text{-}10}$. At the mid-range, face recognition using the super-resolved imagery resulted in significantly and consistently higher performance than using the low-resolution imagery, yielding AUC improvement of 6.71% for $SR4_{15\text{-}20}$, and 10.66% for $SR8_{15\text{-}20}$. At the close range, although the benefit provided by super-resolution was not as great as at the mid-range, face recognition using the super-resolved imagery still yielded AUC improvement of 3.73% for $SR4_{25\text{-}30}$, and 4.64% for $SR8_{25\text{-}30}$. Note that AUC is a measure of the overall performance; performance is also commonly given for low false alarm rates, as addressed in Section 3.3.

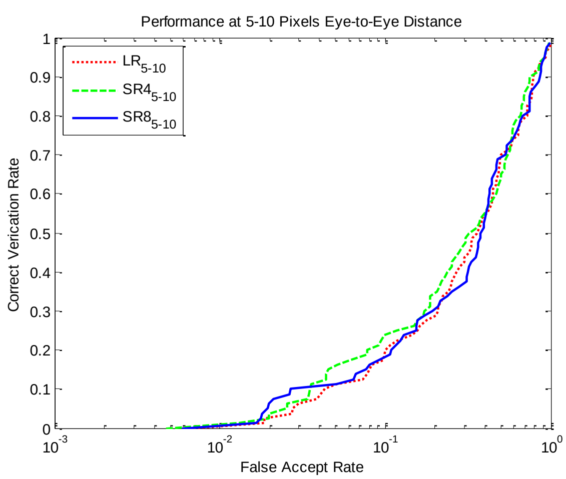


Figure 2. ROC curves at (5 to 10) pixels eye-to-eye distance for the original low-resolution (LR) query set and the corresponding super-resolved query sets using four frames ($SR4_{5\text{-}10}$) and eight frames ($SR8_{5\text{-}10)}$).

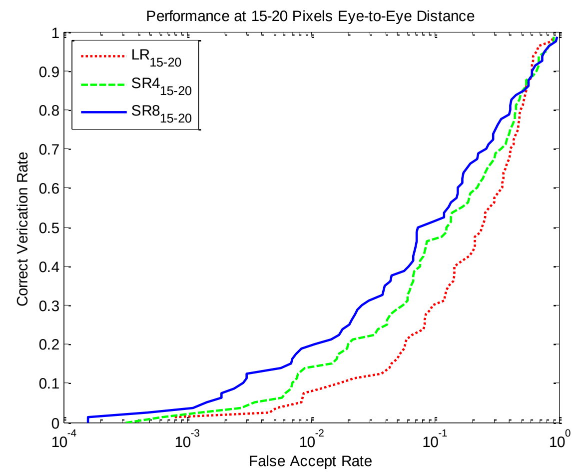


Figure 3. ROC curves at (15 to 20) pixels eye-to-eye distance for the original low-resolution (LR) query set and the corresponding super-resolved query sets using four frames ($SR4_{15\text{-}20}$) and eight frames ($SR8_{15\text{-}20}$).

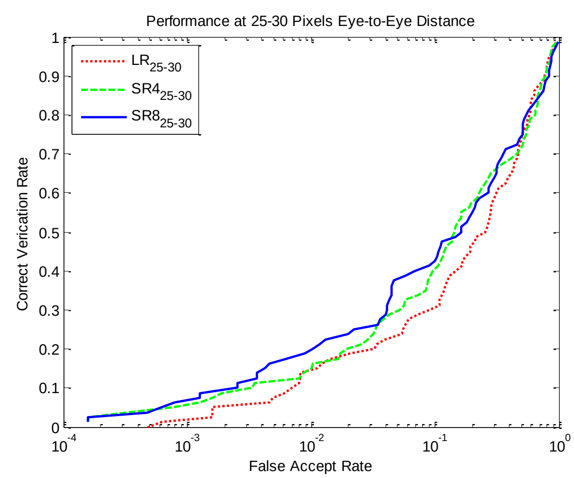


Figure 4. ROC curves at (25 to 30) pixels eye-to-eye distance for the original low-resolution (LR) query set and the corresponding super-resolved query sets using four frames ($SR4_{25\text{-}30}$) and eight frames ($SR8_{25\text{-}30}$).

Table 2. Area under the ROC curve. Top row represents range in terms of eye-to-eye distance.

| | **(5 to 10) Pixels** | **(15 to 20) Pixels** | **(25 to 30) Pixels** |
|---|---|---|---|
| **Low-Resolution** | 0.6028 | 0.7150 | 0.6914 |
| **Super-Resolved 4 Frames** | 0.6230 | 0.7630 | 0.7172 |
| **Super-Resolved 8 Frames** | 0.6002 | 0.7912 | 0.7235 |

## 3.3 Performance versus Range

To visualize how performance varies with respect to distance, the correct verification rate is plotted in Figure 5 as a function of range for three false alarm rates: (a) 0.01, (b) 0.05, and (c) 0.10. Note that performance results at FAR of 0.001 are often reported for face recognition algorithms, but FAR of 0.001 was not used here due to the smaller subject sample size (80 subjects). At FAR = 0.01, although the change in performance is roughly linear for LR, a sharp knee is observed for SR4 and SR8, signifying that performance is flattening for ranges closer than the mid-range (15 to 20) pixels eye-to-eye distance. At the far range, using more frames may accentuate facial artifacts from compression effects and negatively impact the PCA-based algorithm; therefore, SR8 is observed to produce a lower verification rate than SR4, especially for higher FAR. At the mid-range where the knee occurs, the correct verification rate is 13.99% for $SR4_{15\text{-}20}$ and 19.60% for $SR8_{15\text{-}20}$ compared to 8.01% for $LR_{15\text{-}20}$, resulting in an improvement by 74.66% and 144.69%, respectively. A similar trend is observed at FAR = 0.05 with the knee also occurring at the mid-range where the correct verification rate is 28.97% for $SR4_{15\text{-}20}$ and 38.04% for $SR8_{15\text{-}20}$ compared to 16.25% for $LR_{15\text{-}20}$, resulting in an improvement by 78.28% and 134.09%, respectively. These results are consistent with the findings of [3-5]. For (5 to 10) pixels, (15 to 20) pixels, (25 to 30) pixels pixel eye-to-eye distances, SR4 produced "effective" eye-to-eye distances of (10 to 20) pixels, (30 to 40) pixels, and (50 to 60) pixels, respectively (since the resolution improvement factor for SR4 is two in the x- and y-directions). Blackburn et al. [4] observed that face recognition performance is similar for eye-to-eye distances from (30 to 60) pixels, which would correspond to the effective eye-to-eye distances of the mid and close ranges of this study for SR4. Once the effective eye-to-eye distances falls below 30 pixels, performance changes drastically as observed in this study for SR4 as well as SR8 and.

Interestingly, at FAR = 0.10, performance of SR4 and SR8 at the close range is lower than for the mid-range, suggesting that super-resolution image reconstruction is most effective for the mid-range where eye-to-eye distance is between (15 to 20) pixels. At the close range, the correct verification rate is 40.46% for $SR4_{25\text{-}30}$ and 42.63% for $SR8_{25\text{-}30}$ compared to 30.71% for $LR_{25\text{-}30}$, resulting in an improvement by 31.75% and 38.81%, respectively. For the investigated mid and close ranges of this study, super-resolution produced a significant improvement in the correct verification rate over the original low-resolution imagery at all false alarm rates below 0.40.

## 3.4 Absolute Verification Rates

Although super-resolution produced significant benefits for the LRPCA face recognition algorithm at the mid and close ranges relative to the performance using low-resolution imagery, the maximum achieved correct verification was below 20.00% at FAR of 0.01. This should not be surprising, as the verification rate achieved by an advanced fusion algorithm for face recognition on the Ugly partition of the MBGC/GBU data was slightly above 45% at a false alarm rate of 0.01 (http://face.nist.gov/mbgc/MBGCFuture/FutureChallengesMBGC.html). The target and query sets extracted from the video database [1] used in this study exhibited differences in lighting conditions, facial expressions, distance of subject to camera as well as slight differences in facial aspect angle. The variations exhibited between the query and target sets of this study are expected to be possibly more extreme than the variations in the Ugly partition of the GBU database. In addition, whereas the GBU images are high resolution (3008 × 2000) pixels, imagery from the video database is (720 × 480) pixels. The imagery used in this work is therefore a close approximation to typical surveillance imagery, and the benefits of super-resolution for face recognition shown here suggest that super-resolution image reconstruction will be effective in practical applications.

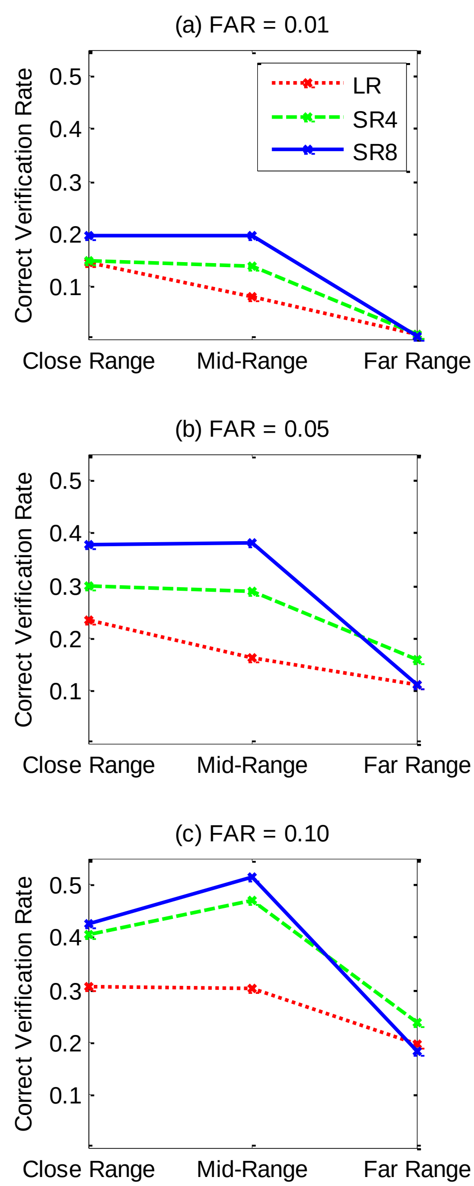


Figure 5. Performance as a function of range at three false alarm rates: (a) 0.01, (b) 0.05, and (c) 0.10.

## 4. CONCLUSION

Super-resolution has been shown to provide significant benefits for the LRPCA face recognition algorithm, especially for the mid-range where the eye-to-eye distance of the subjects is between (15 to 20) pixels. Performance of the LRPCA algorithm at the mid-range increased from a verification rate of 16.25% at FAR = 0.05 using the original low-resolution query set to 38.04% using the super-resolved imagery, showing that significant benefits can be achieved with super-resolution for face recognition.

In surveillance applications, lower resolution cameras are commonly used, typically acquiring images of individuals at a distance. The limited number of face pixels severely impact face recognition performance. Super-resolution image reconstruction improves the resolution and enhances the frequency content of low-resolution imagery, benefitting face recognition systems and potentially aiding the law enforcement community and homeland security.

## 5. REFERENCES


[1] O'Toole, A.J., Harms, J., Snow, S.L., Hurst, D.R., Pappas, M.R., Ayyad, J.H., and Abdi, H. 2005. A Video Database of Moving Faces and People. *IEEE Transactions on Pattern Analysis and Machine* Intelligence 27, 5 (May 2005).

[2] Young, S.S., and Driggers, R.G. 2006. Super-resolution image reconstruction from a sequence of aliased imagery. *Applied Optics* 45.

[3] Boom, B.J., Beumer, G.M., Spreeuwers, L.J., and Veldhuis, N.J. 2006. The effect of image resolution on the performance of face recognition system. In *Proceedings of the 7th International Conference on Control, Automation, Robotics, and Vision* (Singapore, December 05-08, 2006).

[4] Blackburn, D.M., Bone, M., and Phillips, P.J. 2000. Facial Recognition Vendor Test 2000. http://www.frvt.org/FRVT2000/

[5] Wheeler, F.W., Liu, X., and Tu, P.H. 2007. Multi-frame super-resolution for face recognition. In *Proc. of IEEE First International Conference on Biometrics: Theory, Applications and Systems* (Washington, D.C., USA, September 27-29, 2007).

[6] Bolme, D.S., Beveridge, J.R. 2010. CSU Face Recognition Resources. *http://www.cs.colostate.edu/facerec/algorithms/lrpca2010.php*.

[7] Bolme, D.S., Beveridge, J.R., Teixeria, M., Draper, B.A. 2003. The CSU face identification evaluation system: Its purpose, features, and structure. In *Proc. of 3rd Int. Conf. on Vision Systems* (Graz, Austria, April 1-3, 2003).

[8] Grother, O., Micheals, R., and Phillips, P.J. 2003. Face Recognition Vendor Test 2002 Performance Metrics. In *NIST IR* 6982.

[9] Rizvi, S.A., Phillips, J.P., Moon, H. 1998. The FERET Verification Testing Protocol for Face Recognition Algorithms. In *NIST IR* 6281.