Enhancing Face Recognition at a Distance using Super Resolution

Nadia AL-Hassan   Sabah A. Jassim   Harin Sellahewa

Department of Applied Computing, University of Buckingham

Buckingham, MK18 1EG, UK

{Nadia.al-hassan; sabah.jassim; harin.sellahewa}@buckingham.ac.uk

ABSTRACT

The characteristics of surveillance video generally include low-resolution images and blurred images. Decreases in image resolution leads to loss of high frequency facial components, which is expected to adversely affect recognition rates. Super resolution (SR) is a technique used to generate a higher resolution image from a given low-resolution, degraded image. In this paper, we investigate two existing super-resolution methods to reconstruct a high resolution from a single/ multiple low-resolution images. The first method is based on pairs of high and low quality dictionaries in the spatial domain and the other method is the so called Back-Project Iterative Interpolation for super resolution. Recent works have used SR as a pre-processing step to overcome the problem of low-resolution images in face recognition. However, super resolution reconstruction process being ill-posed results in visual artifacts that can be visually distracting to humans and/or effect machine feature extraction and face recognition algorithms. In this paper, we propose an alternative super resolution scheme based on dictionaries in high frequency wavelet subbands. The proposed method is evaluated on database of high and low-resolution images- Extended Yale B database and UBHSD database- and for face recognition at a distance. We shall demonstrate that the proposed approach at level 3 DWT resolution has superior performance in comparison to the other methods.

Keywords

Low Resolution, Face Recognition, Super Resolution (SR), Dictionary, and Wavelet Transform (WTs).

1. INTRODUCTION

Face recognition has been investigated extensively in recent decades [1], and has many applications in daily activities such as identity management (e.g., passports, driving license cards), video surveillance, and secure access control to sensitive physical as well as virtual facilities. For many face recognition algorithms, the recognition rate drops dramatically when the resolution of face images decreases since many details of discriminating facial features can only be captured in images of sufficient resolution. However, in some applications, like video surveillances, we may only have access to low-resolution (LR) and/or degraded images captured at a distance where the face is small relative to the field of view. Most areas are monitored by CCTV cameras that have limited spatial resolution, which hinders the fine details from being captured. Under these circumstances, a face recognition system would be severely degraded without enhancing the resolution of the captured still images or video sequences. In order to improve recognition accuracy, pre-processing to enhance contrast, resizing and multi-frames fusion techniques have become a necessity. For resizing and multi-frames fusion a variety of traditional non-adaptive interpolation techniques (e.g. bilinear, bi-cubic, wavelet) [2] as well as more recently developed adaptive interpolation techniques (e.g. edge-directed based and non-local) [3] can be used. Recently, the use of high definition cameras [4] and example-based approaches to super-resolution (e.g. the compressive sensing dictionary based) [5] have been investigated as an alternative to interpolation for improved resolution in relation to face recognition in unconstrained environments. This paper is concerned with these approaches.

A low-resolution image is a result of a combination of degradation (e.g. noise, blurring) and down sampling procedures applied to a hypothetical high-resolution image.

Super resolution (SR) is the process of obtaining a high-resolution (HR) image either from a single low resolution (LR) [6] or from multiple low-resolution images [7]. In other words, one can imagine that a low-resolution image has the same content of the hypothetical high-resolution image but with missing high frequency details. Super resolution of a single image refers to techniques that reconstruct a high resolution image from a given low resolution image by a procedure that involves up-sampling using a criteria for estimating the missing details followed by procedures to filter out distortions such as noise and blurring effects. Super resolution is expected to maximize Peak signal to noise ratio (PSNR) or minimize the mean squared error between the original hypothetical high-resolution image and the reconstructed super resolution. Super resolution is currently an active area of research with practical applications in medical imaging, target detection and recognition, satellite imaging and video processing.

Mathematically the super resolution problem can be modeled as a solution of:

$$y = DHX + n$$

where y is the known low-resolution image, D and H denote the decimation matrix and blurring matrix respectively; X represents the unknown HR image and n is additive noise.

Generally, reconstruction approach is a severely ill-posed problem because of the insufficient number of low-resolution pixels, ill-conditioned registration, unknown blurring operator, and the solution is not unique.

Over the last few years, numerous super resolution methods have been developed and applied to a variety of image classes with differing degrees of success on different image classes (e.g. [8], [9], [10], [11], [12], [13], [14], [15], [16]). In this paper, we investigate two existing spatial domain super resolution methods to reconstruct high-resolution face images from single/multiple

low-resolution images primarily to test their suitability for face recognition at a distance. We test the performance of these techniques and demonstrate that they could provide a cost-effective alternative to using high definition cameras, but the improvement in recognition, if any, is not significant. Performance of face recognition schemes relies on the ability to extract facial features corresponding to high frequency content. This is a motivation for our investigations of SR in the wavelet domain. In this paper, we propose a method for face recognition that reconstructs the high frequency subbands from low-resolution input images using dictionaries built on those subbands from a training high-resolution face images.

The rest of this paper is organized as follows. Existing work is discussed in section 2. Background material is reviewed in section 3.The proposed wavelet based SR approach is described in section 4. The biometric databases and experimental protocols adopted here are described in section 5. Recognition experiments are presented and discussed in section 6 and finally section 7 is devoted to conclusions and future work.

2. EXISTING WORK

Super-resolution techniques can be categorized into classes. Various optimization approaches were proposed to further stabilize the inversion of such an ill-posed problem, such as [8], [9], [11]. While, the performance of these algorithms degrades rapidly when the desired magnification factor is large or the number of input images is small. Another category of super resolution approach is based on interpolation [10], [9], [12]. However, interpolation methods such as bilinear or bi-cubic interpolation tend to generate jagged artifacts, ringing effect and overly smooth images, which lack high frequency components and thus blur the edge information of the reconstruction image. The third case of super resolution (SR) method is based on learning approaches. In [13] the author Freeman proposed an example based learning strategy that applied to images where the low-resolution to high-resolution prediction is learned via a Markov Random Field solved by belief propagation. In another case, Bilgazyev [14] proposed a super resolution algorithm based on sparse representations. They compute a sparse representation of the input low resolution (LR) image using the dictionary built for (LR) images and estimate the high-resolution (HR) high-frequency components using that sparse representation with respect to the HR dictionary. The estimated high frequency components of the high-resolution (HR) image are then added to the low-resolution (LR) input image to create a super resolution (SR) image. In this approach both high resolution and low-resolution dictionaries are trained using dual tree complex wavelet transform (DT-CWT) from a set of training images. Zeyde [15] has proposed another dictionary based technique for super resolution using sparse representation. The K-means single value decomposition (K-SVD) algorithm is also used for dictionary training.

In [16], Xiaoqing has developed a two stage SR technique to reconstruct a high-resolution image from sequence of low-resolution images. The first stage uses sparse representation to magnify each input low-resolution image based on learning dictionary. In the second stage, the output high-resolution image is obtained by fusing the intermediate high-resolution image sequence based on projection onto convex sets (POCS) method.

Most of the researchers they used super resolution as pre processing to obtain a high-resolution facial image. He, J [17] has developed a face recognition scheme by using SR technique that constructs a high-resolution image from a sequence of low-resolution images, which would be processed using Gabor feature based recognition. The performance has been evaluated for AR face database and reported (95%) recognition rate. Al-Azzeh [18] proposed method uses super resolution to generate a super resolved video sequences from a low-resolution video sequences and uses frames acquired from the high-resolution video sequences to train and test the performance of the principal component analysis based face recognition system. Simultaneous super resolution and feature extraction for face recognition was proposed by Hennings-Yeomans [19],[20] to estimate a super resolution image by searching for similar features of the LR image in the training set. However, this method uses the relationship between the HR and LR image, and its performance depends on the training database. Arachchige [21] used super-resolution techniques as a middle step to improve face recognition in low-resolution video sequences they used PCA for face matching.

3. BACKGROUND

Here we present some background on the two Super Resolution methods that we investigate there suitability for face recognition. These methods generate a high-resolution image from a single/multiple low-resolution image. We shall also describe briefly the multi-resolution (wavelet) transforms as our choice of face feature extraction scheme and we are proposing a dictionary-based SR in the wavelet domain.

3.1 Back Project Iterative Method

Super Resolution refers to creating a magnified high-resolution image from sequence of the low-resolution images. Irani and Peleg [22], [23] have formulated the Iterative Back-Project (IBP) super resolution reconstruction approach to obtain super resolve image from monochrome and color low resolution image sequences ,this approach is similar to the back project used in tomography .The algorithm starts with an initial guess $f^{(0)}$ for the output high resolution image and the imaging process is simulated to generate low-resolution images $\{g_k^{(0)}\}$ based on initial guess corresponding to the observed input images g_k. The simulated low-resolution images are then compared with the observed ones and used the error between them to improve the initial guess by back project each value. In this paper, we used the back-project iterative method to reconstruct a high resolution from only two low-resolution images. Each low-resolution image contains different information. These low-resolution images are created from the original image by taking pixels in an image with purely even, and purely odd valued indices after blurring the input image by Gaussian filter. The initial approximation $f^{(0)}$ for the high-resolution output image is estimated by combining the two low-resolution images over the finer grid and completing the high-resolution grid by interpolating the missing pixels using bi-cubic interpolation method to obtain a single blurred image of higher spatial sampling rate. The iterative step is then applied to remove possible artifacts from the initial approximation. The imaging process is simulated to obtain low-resolution images by take even and odd valued indices from the initial approximation after blurred the initial image, the difference images between input low resolution images and simulated low-resolution images are combined and leaving the other pixels equal to zero to obtain error image which is the same size of the original image or initial approximation. This image was then de blurred by convolution it with a restoration filter using a wiener filter. De blurred error image used to improve the initial image which yield a high resolution image $f^{(1)}$ .The new image is input for the next iteration cycle, a new set of low-resolution images are generated and subtract to form the error vector. The iterative update scheme of the high-resolution image can be expressed by

$$f^{(n+1)} = f^{(n)} + ((g_k - g_k^{(n)}) \uparrow s)*P \quad , n=0, 1, 2 \ldots t$$

where k is the number of the low-resolution images, $\uparrow$ is denote up sampling by a factor s and P is a back project restoration filter. The restoration wiener filter can be expressed by the following equation:

$$W = \frac{1}{H} \cdot \frac{|H|^2}{(|H|^2 + k)}$$

where H is the degradation function and k is the scalar constant.

3.2 Super Resolution by Dictionary (SR by D)

Yang , et al. [24], [5], [25] generated a super resolution image from a single low-resolution input image based on sparse signal representation from an appropriate chosen over-complete dictionary D. A dictionary [26] is a generalization of vector space basis and can be represented an over complete $m \times n$ a matrix whose columns from a pool of bases of $\mathbf{R}^m$ so that any vector in $\mathbf{R}^n$ can have multiple representations in terms of different bases in $\mathbf{R}^m$. Many Applications require representations that are optimal in accordance to a practical requirement, such as seeking the sparsest representation. This SR approach can be justified by the assumption that a low-resolution image, or an image patch, is a sparse representation of a high-resolution image/patch. The reconstruction of high resolution images benefits greatly perform the innovative paradigm of compressed sensing (CS) which relaxes the stringent bound imposed by Nyquist-Shannon sampling theory for acquiring and reconstructing signals that are supposed to be sparse or compressible. CS is based on L_1-minimisation solutions for underdetermined linear system, and many techniques have been developed for reconstructing sparse matrices.

Yang , et al. super resolution scheme process patches from the input low-resolution image and recovers the sparse linear representation of a high-resolution image patch from these low-resolution image patches using a pair of coupled dictionaries D_h for high-resolution patches and D_L for low-resolution patches. The compact dictionaries D_h and D_L are created from the large dictionaries X_h and X_l respectively where X_h are the set of sampled high-resolution image patches and X_l are the corresponding low-resolution image patches. The coupled of dictionaries D_h and D_L are generated/trained by (1) building X_h for a sufficiently large random raw patches carefully selected from high resolution images in a training set, (2) synthetically generating the low resolution image patches in X_l by blurring and down sampling the high resolution training database, and for each patch subtract the mean pixel value, so that the dictionary represents image textures (3) Combine the two large dictionaries by $X = \begin{bmatrix} \frac{1}{\sqrt{N}} X_h \\ \frac{1}{\sqrt{M}} X_L \end{bmatrix}$ where M and N are the dimensions of the high resolution and low resolution image patches in vector form. Moreover, (4) use the same learning strategy in the single dictionary, which is shown in the Algorithm1, to generate the couple of dictionaries D_h and D_L.

The SR scheme works as follows: (1) for each input low-resolution patch, find a sparse representation with respect to low dictionary D_L and (2) use the coefficients of this representation with the corresponding patches in D_h to generate the output high-resolution patch. The computed sparse representation adaptively selects the most relevant patch bases in the dictionary to best represent each patch of the given low-resolution image. For global reconstruction, constrain used back project iterative method for a single image to remove possible artifact from the local sparse and

to eliminate the reconstruction errors in the high-resolution image. To turn image patches to vectors for the dictionary D_L a feature extraction is applied to the low-resolution patches consisting of four 1D filters chosen as first and second order gradient filters. The resulting four feature vectors as turned into one vector by concatenation as the final representation of the low-resolution patch. The high frequency components of the low-resolution image are the most important for predicting the lost high frequency content in the target high-resolution image. The size of the dictionary will be very large if all the patches are used and hence require expensive computation, and hence, a smaller manageable random set of patches are used in the construction of these dictionaries. In our implementation, they used the same learning strategy in the single dictionary [24], [27] shown below in the Algorithm 1.

Algorithm1

---

1. Start with initial D with Gaussian random matrix
2. Fix D, find Z
$$Z = \min\|X - DZ\|_2^2 + \lambda\|Z\|_1$$
3. Fix Z, update
$$D = \min\|X - DZ\|_2^2$$
4. Iterate 2, 3 until converge

5. Find the D_h and D_L from $D = \begin{bmatrix} \frac{1}{\sqrt{N}} D_h \\ \frac{1}{\sqrt{M}} D_L \end{bmatrix}$

3.3 Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is a multi-resolution signal analysis tool that hierarchically decomposes a signal into its low and high frequency components allowing one to view the signal's regular patterns as well as its anomalies. At a resolution level of k, the pyramid scheme decomposes an image into 3k+1 subbands $(LL_k, LH_k, HL_k, HH_k, \ldots \ldots LL_1, LH_1, HL_1, HH_1)$ at a resolution level of k where the LL_k is considered as the k-level approximation of image I, while LH_k, HL_k, HH_k captures horizontal, vertical and diagonal features of the image. There are numerous wavelet filter banks that can be implemented in a variety of decomposition schemes .Here we use the Haar wavelet filter in a pyramid decomposition scheme. The coefficient in each subband of wavelet transformed face image (the super resolved image) is used as a feature vector representation of the face for recognition purposes [28], [29]. In the recognation stage, the city Block distance is used to classify the unknown face images. All recognition experiments use level 3 subbands, and our proposed super resolution method builds the dictionary in these high frequency subbands.

4. PROPOSED METHOD
4.1 Super Resolution based on dictionary in the wavelet domain (SR by D-LH3).

Performance of face recognition schemes are adversely affected by image degradation such as the blurring of the edges that represent the most discriminating facial features. The fact that such facial features are associated with high frequency content suggests that building dictionaries in high frequency wavelet subbands and recovering such subbands of the hypothetical high-resolution face images could result in improved face recognition. Hence, we propose a method that reconstructs high frequency subbands from low-resolution input images using dictionaries

built in those subbands from a training set of high-resolution face images. For simplicity, and due to the fact that most face features are associated with high frequencies in the horizontal directions, we shall describe our proposal for the LH_3 subband.

In a similar manner to SR in the spatial domain, we build the over-complete dictionaries D_h from randomly selected patches from LH_3 decomposed high-resolution images in a training set. To create the corresponding D_L we synthetically generate the LR patches by blurring the training images with Gaussian filter and down sampling to half the size of the original HR patches. The super resolution reconstructing of the high frequency LH_3 subband of a given low-resolution face image is achieved by first dividing the LH_3 subband into patches with an overlap of one pixel to avoid discontinuities during reconstruction. We then compute a sparse representation to these patches using D_L and the corresponding high patch bases D_h will be combined according to these coefficients to generate the output high frequency LH_3 components. Finally, the Back project iterative method is used to reconstruct the high frequency LH_3 subband. The sparse representation is obtained by minimizing

$$\alpha = \min \lambda \|\alpha\|_1 + \frac{1}{2}\|F D_l \alpha - Fy\|_2^2$$

Where the parameter $\lambda = 0.15$ balances sparsity of the solution and fidelity of the approximation to y. This is essentially a linear regression regularized with L_1 norm on the coefficients and F is formed by the concatenation of four 1D filters:

$$f_1 = [-1,0,1] , f_2 = f_1^T , f_3 = [1,0,-2,0,1] , f_4 = f_3^T$$

Chosen as first and second order gradient filter to the high frequency patch from LR image .The learned dictionary pair is built with the high frequency patches, which are randomly selected from each of the database; both high resolution and low resolution and we select approximately 100,000 patch pairs sampled from the LH_3 subband of the images. For each patch of size 5x5, we subtract the mean pixel value. In all experiment, we fix the dictionaries to the size 512 and we select 114 face images as training images from Extended Yale-B database described in section 5. The steps of this method are described below in the Algorithm 2. We normalize the coefficients of the reconstructed (super-resolved) high frequency LH_3 subband for the face recognition system by Z-score normalization (ZN).

Algorithm 2: Super Resolution based on dictionary in the LH_3 sub band

---

1. **Input**: high resolution training images database and a low resolution image Y.
2. **Prepare the Dictionaries D_h and D_L**: randomly selected raw patches from LH_3 high frequency subband to the images.
3. **For** each 5x5 patch y of LH_3 high frequency subband to Y, taken starting from the upper-left corner with 1 pixel overlap in each direction,
   - Find the sparse representation to y patch $y = D_L \alpha_0$.
   - Generate the high frequency x patch $x = D_h \alpha_0$. Put the patch x into an output high frequency X_0.
4. Using back project method to reconstruct the high frequency by :
   $X_{n+1} = X_n + ((Y - DHX_n) \uparrow s)*P$ n=0,1,2,.......k
5. **Output**: reconstruction high frequency LH_3 subband.

---

5. EXPERIMENTAL PROTOCOL AND DATASET.

This section gives a brief description of the two-benchmark face biometric databases used in our experiments as well as the adopted experimental protocols used to evaluate the performance of the existing and the proposed super-resolution methods for face recognition.

5.1 Extended Yale B face database.

The cropped frontal face images of the Extended Yale B database provide an excellent testing platform for extreme variation in illumination. It consists of still images and has 38 subjects, each having 64 images captured under different illumination conditions. The total number of images in the database is 2414 images and size of each image is 192 x 168 pixels. In addition to these frontal pose images, the database includes an ambient illumination image for each subject but these were not used in our experiments. The images in the Extended Yale B are divided into five illumination subsets according to the angle θ of the light-source with respect to the optical axis of the camera as shown in Table1.

Table 1. Illumination sets in the Extended Yale B database

Subsets	Angles	Number of images
1	$\theta < 12$	263
2	$20 < \theta < 25$	456
3	$35 < \theta < 50$	455
4	$60 < \theta < 77$	526
5	$85 < \theta < 130$	714

We selected three images per subject from subset 1, which contain images of good quality as high-resolution images to build the high dictionary D_h for super resolution method. The total number of high-resolution images in the training set for the dictionary is 114 and includes images of both genders. Figure 1 shows example of training face images. In all the experiments, we selected the P00A+000E+00 image of each subject from subset 1 for the gallery set. Note that this image is not part of the dictionary training set. All the remaining images are included in the probe set, which are blurred and down sampled to the half.
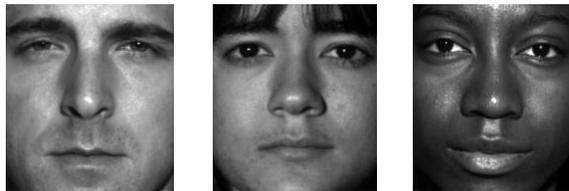
Figure 1. Example of three out of 114 training face images.

5.2 UBHSD databases

The UBHSD video database contains 160 videos of 20 subjects [4]. The videos of each subject were recorded in two sessions with a gap of at least two days between the recording sessions. Each recording session includes videos captured in an indoor and outdoor location. Two video recordings one high-definition (HD) and one standard-definition (SD) of subject were captured at each location. During a recording, a subject walks a distance of either 4 meters (in indoor recordings) or 5 meters (in outdoor recordings) toward the camera, from a start-point to a stop-point, providing

face data at different distances. The minimum distance between the camera and the subject (stop-point) is a meter. The subjects walk in a natural way, which included head movements and facial expressions. Twelve frames - three frames per each distance range - from each video were selected in a systematic way to capture the subject at four distance ranges from the camera position. The frames in the first range, *Range 1* (R1), are nearest to the camera, while the frames in the fourth range, *Range 4* (R4), are the farthest away from the camera. The database consists of blurred face images, faces with eyes closed and slightly varying poses. Each subject has 96 face images, thus the total number of face images in the database is 1920. The face region in each frame was manually cropped at the top or middle of the forehead, bottom of the chin, and at the base of the ears. Then, all face images were converted to gray scale and rescaled to size of 128×128 pixels by using MATLAB code 'imresize' with the default bi-cubic interpolation. Figure 2 shows examples of cropped and rescaled face images extracted from the respective HD and SD videos frames.

High Definition, Indoor, Range 1- 4

Standard Definition, Indoor, Range 1- 4

High Definition, Outdoor, Range 1- 4

Standard Definition, Outdoor, Range 1- 4

Figure 2. Examples of cropped and rescaled face images from HD and SD videos captured in indoor and outdoor conditions.

The evaluation protocol of UBHSD database involves four configurations for each video resolution; we work with two configurations Matching Indoor and Unmatched Indoor. Table 2 shows the gallery and probe sets for these two configurations. Each configuration has four test cases (e.g. MI$_1$ ...MI$_4$). The gallery set G of test case i ($i = 1 ... ... .4$) consists only ranges R$_i$ face images in Session 1. For each test case, images from all four ranges, in both indoor and outdoor videos in Session 2 are used as probe images. There is no overlap between the gallery and probe sets. In Matched configurations, both the gallery and probe images come from the same video resolution. In Unmatched configurations, gallery and probe images are from different video resolutions. For each test case, the gallery set consists of 60

images (3 images per subject) and the probe set consists of 480 images (24 images per subject).

**Table 2. The configuration for the UBHSD video database.**

Configuration		Session 1		Session 2			
		HD	SD	HD		SD	
		Location		Location			
		Indoor		In	Out	In	Out
HD	MI$_i$	G, R$_i$		P,R$_{1-4}$	P,R$_{1-4}$		
	UI$_i$	G, R$_i$				P,R$_{1-4}$	P,R$_{1-4}$
SD	MI$_i$		G, R$_i$			P,R$_{1-4}$	P,R$_{1-4}$
	UI$_i$		G, R$_i$	P,R$_{1-4}$	P,R$_{1-4}$		

6. RESULTS AND ANALYSIS.

We conducted two sets of experiments: - the first set aims to show comparisons between the super resolution methods to enhance the resolution of single/multiple low-resolution images and the second set to compare the performance of SR algorithms in terms of face recognition rates.

6.1 Experiments: Reconstructed super resolution image.

Super resolution approaches are expected to produce better quality images than interpolation. Our experiments test this hypothesis by calculating the PSNR (peak signal to noise ratio) values between the original high-resolution image and reconstructed super resolution images obtained from single/two low resolved versions of these images using the spatial domain techniques and compare the results with the well-known bi-cubic interpolation approach. Prior to downsampling, the original images are blurred with a Gaussian filter of level 7. The results in figure 3 and figure 4 are based on experimental conducted on the Extended Yale B database and UBHSD database respectively. Figure 3 shows PSNR comparison results where LR1 and LR2 are the low resolution images generated from the blurred input images by taking even and odd indexed pixels respectively. It can be seen from Figure 3 that the back-project iterative interpolation super resolution method (IISR) slightly outperforms the super resolution by dictionary in the spatial domain (SR by D) at each subset of the Extended Yale B database and both methods have far better PSNR values than the standard bi-cubic interpolation method. Moreover, the implementation of the IISR algorithm is faster than the SR by D algorithm.
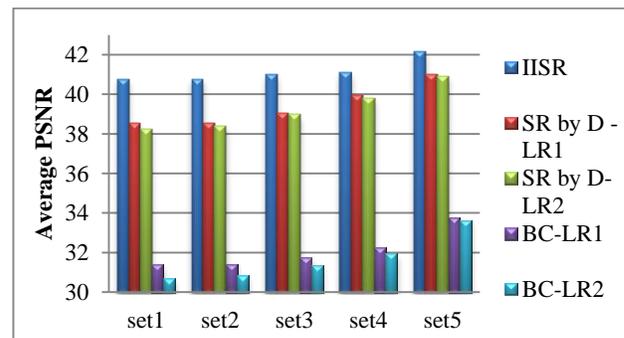

Figure 3. Average of PSNR values between the original high-resolution and the super-resolved images in the Extended Yale B database.

Figure 4 illustrates PSNR results for the UBHSD database. The results show that, the standard bi-cubic interpolation method produced lower quality high-resolution images compared to those produced by the super-resolution methods in the spatial domain at each distance ranges and for each sessions and different conditions (indoor and outdoor). In addition, the iterative interpolation method produce better quality images compared to the dictionary based method.
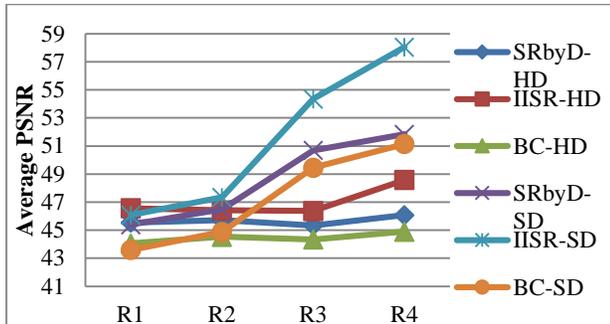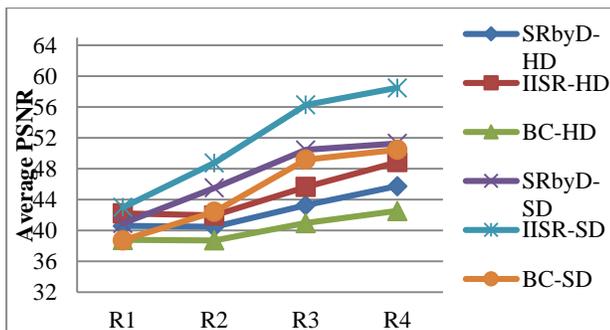

Indoor condition with session 1
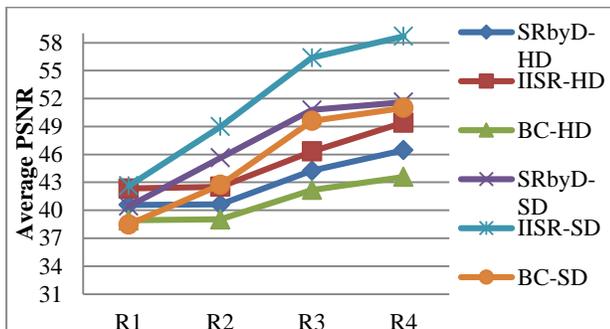
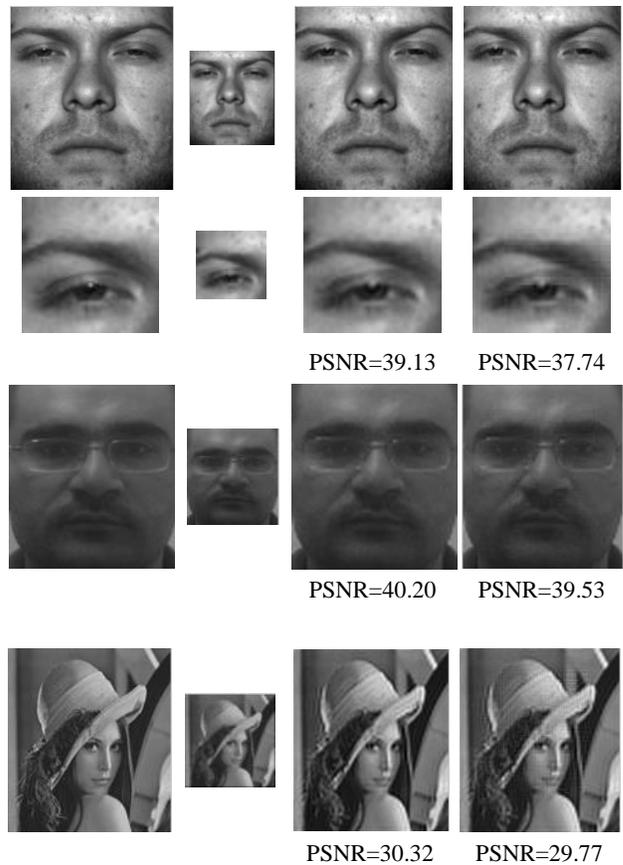Outdoor condition with session 1
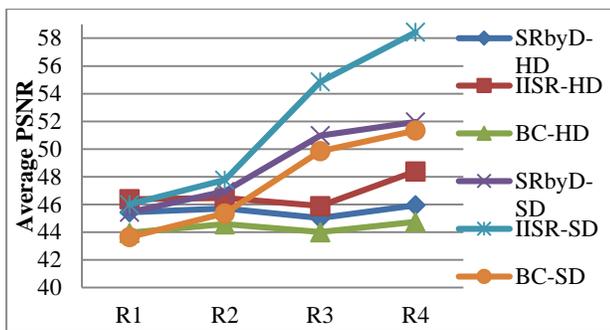
Indoor condition with session 2

Outdoor condition with session 2

Figure 4. PSNR for super resolution HD and SD video data in the UBHSD database and the average of all PSNR results to each distance ranges are found.

In order to test the viability of applying super-resolution scheme in the frequency domain, we conducted a subjective visual comparison between enhanced images by the dictionary based super resolution methods in the spatial domain (SR by D) and in the wavelet domain. In the wavelet domain, we super-resolved each subband according to Algorithm 2 and took the Inverse DWT to produce high-resolution image. For global reconstruction, we used Back-project iterative method after Inverse DWT to eliminate the reconstruction errors in the high-resolution image result. Four test images indicated in the first column of figure 5 are used for comparisons. As can be seen in figure 5, the wavelet image results are slightly smoother with jagged effects than enhanced image in the spatial domain (see the cropped region of the first image). The quality of the image based on super resolution method in the spatial domain is better than the one enhanced by super resolution in the wavelet domain. The visual results of the super resolution methods in figure 5 supported with the results of PSNR measured where super resolution method in the spatial domain has slightly higher PSNR than the by super resolution method in the wavelet domain; the original images of the database have been used as the ground truth to calculate the PSNR values.

PSNR=39.13   PSNR=37.74

PSNR=40.20   PSNR=39.53

PSNR=30.32   PSNR=29.77

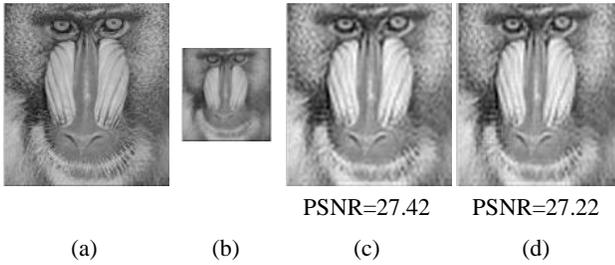| (a) | (b) | PSNR=27.42 (c) | PSNR=27.22 (d) |

Figure 5. Comparison between enhanced images. (a) First column is the original image. (b) The second column is the low-resolution image, which is generated by, blurred and down sampled the original image. (c) The third column result image magnified by factor 2 from super resolution method by dictionary in the spatial domain. (d) The last column is the super resolved image magnified by factor 2 from super resolution method in the wavelet domain.

6.2 Experiments: Face recognition.

6.2.1 Extended Yale B Database

Obtaining a higher PSNR does not necessarily contribute to a higher recognition rate since high fidelity reconstruction of low-frequency content may dominate the image. Face recognition degrades when probe faces are of significantly lower resolution than those in the gallery. We used super resolution methods in the spatial domain as proposed by [24] and the standard bi-cubic interpolation method to reconstruct a higher resolution version of the low-resolution probe and then perform matching in the usual way at higher resolution. Table 3 shows the results of identification accuracy for different illumination subsets of the Extended Yale B database. We used adaptive histogram equalization (AHE), developed in [28], to normalize variations in illumination of the super resolved images. The experimental results are mixed. While in some cases there is no significant difference between the SR methods (see results for set 2), in set 1 super resolution by dictionary method as pre-processing result in better accuracy of matching in the LL_3 subband. Rather surprisingly, when LH_3 subband is used for matching images in sets 3 and 4 and to some degree in set5, Bi-cubic interpolation has increased recognition accuracy as much as—if not more—by that achieved with the more complex SR methods. This could be due to the presence blocky artifacts in the non-interpolation based super resolution methods. Which result in degrading the face feature vectors specially in badly lit images.

6.2.2 UBHSD Database

Video signals captured by digital imaging devices are digitized at resolution levels lower than that of still images; hence, the quality of a frame extracted from a video sequence is lower than that obtained from a still imaging device. The images captured at a distance are usually of lower resolution, which leads to low recognition accuracy. The use of high-definition videos to overcome the problem of low-resolution in face recognition at a distance has recently been investigated by Al-Obaydy and Sellahewa, [4]. Here we usesuper resolution methods on the UBHSD video database to reconstruct a HR image from the low-resolution images (i.e. images from SD videos) as an alternative to the use of high-resolution images captured from HD cameras. Recognition experiments are conducted according to the same protocol adopted in [4]. In all experiments, LH_3 subband (with z-score normalization) was used as facial features [29]. The results in table 4 show that there is little difference in the identification accuracy rates between the: iterative interpolation method, super

resolution method by dictionary in the spatial domain and Bi-cubic interpolation method.

Table 3. Identification accuracy (%) on Extended Yale B database. Using SR methods in the spatial domain.

Subsets	Sub band	IISR	SR by D-LR1	SR by D-LR2	BC-LR1	BC-LR2
S1	LL_3	95.26	**98.68**	**98.68**	94.21	94.73
	LH_3	92.63	92.10	92.10	92.63	93.15
	HL_3	92.10	92.10	92.10	92.63	91.05
	HH_3	78.42	78.94	78.94	82.63	74.73
S2	LL_3	79.16	79.16	79.16	77.19	78.50
	LH_3	**100**	**100**	**100**	**100**	99.78
	HL_3	95.83	96.05	95.83	95.17	94.95
	HH_3	97.80	97.80	97.80	99.12	79.80
S3	LL_3	12.91	60.28	60.04	58.37	58.61
	LH_3	81.33	86.60	86.84	**88.51**	81.10
	HL_3	39.47	55.02	54.78	54.78	52.39
	HH_3	55.98	69.13	68.89	72.24	66.02
S4	LL_3	26.09	26.53	26.31	27.63	25.21
	LH_3	82.67	81.79	**83.33**	**83.11**	77.19
	HL_3	19.29	19.73	19.95	16.88	18.64
	HH_3	58.77	56.35	57.01	53.07	53.07
S5	LL_3	24.34	24.50	24.50	25.32	24.50
	LH_3	**91.28**	90.78	90.29	90.46	86.84
	HL_3	24.83	24.50	24.01	22.86	23.35
	HH_3	49.34	45.72	44.24	43.91	46.71

Table 4. Recognition accuracy (%) of Matched Indoor and Unmatched Indoor configuration. Used SR methods in the spatial domain to enhance the LR images of size 64x64.

G set	P set	SR methods	Gallery Image Range			
			R_1	R_2	R_3	R_4
SD_{128}	SD_{64}	SR by D	76.25	68.33	71.25	69.79
		IISR	75.78	68.05	71.60	69.51
		BC	76.66	68.75	71.66	69.79
SD_{64}	HD_{128}	SR by D	75.62	71.04	71.25	67.91
		IISR	75.62	70.83	71.45	68.33
		BC	75.62	71.25	71.25	68.75
HD_{128}	SD_{64}	SR by D	68.95	65.62	72.08	74.58
		IISR	68.47	65.97	72.02	74.94
		BC	68.75	66.66	72.50	75
SD_{64}	SD_{64}	SR by D	76.25	68.54	71.25	69.58
		IISR	75.78	67.84	71.60	69.31
		BC	76.04	68.95	71.45	69.79

The artifacts resulting from super resolution reconstruction process have an impact on the quality of the extracted wavelet face feature vectors and thereby have an adverse impact face recognition rate. In order to remedy this situation we avoid the full super resolution image reconstruction, and instead we propose an alternative approached for face recognition at a distance by super-resolving the horizontal high frequency (LH$_3$) subband of the input low-resolution image using dictionaries in the wavelet domain. In order to evaluate the performance of the proposed algorithm, the experiments are conducted on UBHSD video database, as before the low resolution images are generated by down sampling the images from SD video to half size and we use patches of the LH$_3$ as input to the dictionary based SR method to construct the LH$_3$ subband patch of the would be super-resolved image. The experimental results, table 5, demonstrate increased the identification accuracy rates at a distance by using Super resolution technique in the LH$_3$ wavelet subband when the gallery set contains the collection of standard definition camera and the perform of the proposed method is outperform than super resolution by dictionary in the spatial domain. However, in some cases the super resolution in the spatial domain achieve higher identification accuracy rates than super resolution method in the wavelet subband specially when the gallery set is a collection of high definition camera images. Al-Obaydy and Sellahewa, [4], found that the use of high definition cameras improved recognition accuracy rate at a distance than standard definition video data. Nevertheless super resolution method in the LH$_3$ wavelet subband increase the identification accuracy rates to the low resolution than high definition camera, (see table 6).

Table 5. Comparison between super resolution technique in the horizontal high frequency subband of the low-resolution image and super resolution method by dictionary in the spatial domain, reported results based on LH$_3$ subband.

G set	P set	SR methods	Gallery Image Range			
			R$_1$	R$_2$	R$_3$	R$_4$
SD$_{128}$	SD$_{64}$	SR by D-LH$_3$	69.79	73.95	73.33	74.37
		SR by D	76.25	68.33	71.25	69.79
SD$_{64}$	HD$_{128}$	SR by D-LH$_3$	76.87	71.25	71.45	72.91
		SR by D	75.62	71.04	71.25	67.91
HD$_{128}$	SD$_{64}$	SR by D-LH$_3$	65.83	64.79	70.62	70.83
		SR by D	68.95	65.62	72.08	74.58
SD$_{64}$	SD$_{64}$	SR by D-LH$_3$	77.91	77.29	74.37	73.75
		SR by D	76.25	68.54	71.25	69.58

Table 6. Rank one Recognition rates by Al-Obaydy and Sellahewa in [4]

G set	P set	Gallery Image Range			
		R$_1$	R$_2$	R$_3$	R$_4$
SD	SD	76.04	68.12	71.46	69.38
SD	HD	75.83	71.04	71.25	68.33
HD	SD	68.75	65.83	72.08	75.00
HD	HD	68.75	70.62	73.54	72.29

We also tested our proposed approach on the Extended Yale B database and, as before we used adaptive histogram equalization technique to normalize the illumination variation of the low-resolution images. The experimental results in figure 6 shows that the super resolution method in the LH$_3$ wavelet subband gave slightly higher identification accuracy rate than super resolution method by dictionary in the spatial domain in set 1 only (set 1 with well-lit face images). This can be attributed to the problem of sever variation in illumination which seem to compound the existing challenges of super-resolution that primarily deal with low resolved blurred images.
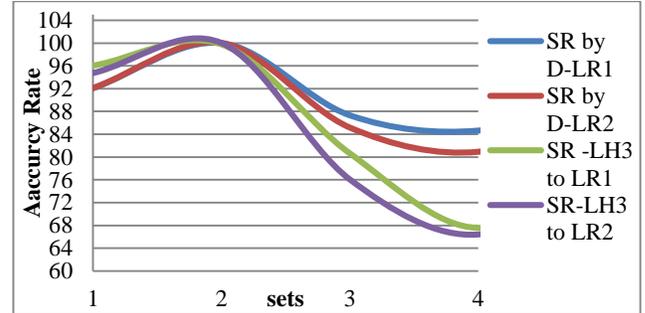
Figure 6. Identification accuracy rate on Extended Yale B database. Using SR by D in the spatial domain and SR by D in the wavelet domain.

7. CONCLUSIONS AND FUTURE WORK.

We investigated and tested the performance of various super-resolution techniques for face recognition at a distance, We also proposed a new Super Resolution method for face recognition based on building dictionaries of high frequency components in the facial images. The proposed framework combines two powerful tools: wavelet decomposition and sparse representations via dictionary learning to super resolve a single low-resolution image. This method outperforms previous super resolution method by dictionary in the spatial domain in terms of face recognition accuracy, but illumination seem to remain an added challenge. Experiments also demonstrate the proposed super resolution method could obviate the need for costly high-definition cameras for face recognition at a distance by super resolving images captured by standard-definition cameras.

The choice of a single detail high frequency subband (LH$_3$) was motivated bt the fact such subbands are highly sparse. In the future our work will focus on increase the sparse vectors in the dictionary by using the advantageous properties of the laplace distribution of the wavelet coefficients in these subbands, and we shall also investigate fusion strategy for using several wavelet subbands.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Zhao, W., et al. 2003. Face recognition: A literature survey. Acm Computing Surveys (CSUR), ACM, Issue 4, Vol. 35, pp. 399-458.

[2] Tsai, P.S. and Acharya, T. 2006. Image up-sampling using discrete wavelet transform. Joint Conf. Information Services, Advances in Intelligent Systems Research, (Oct.2006).

[3] Luong, H.Q., Ledda, A. and Philips, W. 2006. Non-local image interpolation. Image Processing, IEEE International conference on, pp. 693-696 (Oct.2006).

[4] Al-Obaydy, W. and Sellahewa, H. 2011. On using high-definition body worn cameras for face recognition from a distance. Biometrics and ID Management, Springer-Verlag, pp. 193-204.

[5] Yang, J., et al. 2008. Image super-resolution as sparse representation of raw image patches. In Proc.IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[6] Anbarjafari, G. and Demirel, H. 2010. Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image. ETRI Journal , Issue 3, Vol. 32, pp. 390-394.

[7] Bannore, V. and Swierkowski, L. 2007. Fast Iterative Super-Resolution for Image Sequences.Digital Image Computing Techniques and Application , 9th Biennial Conference of the Australian Pattern Recognition Society on. PP.286-293 (Dec 3-5 2007).

[8] Fanaswala,M.H. 2009. *Regularized Super-Resolution of Multi-View Images.*Master Thesis.Applied Science in Electrical Engineering . Carleton University (August 2009).

[9] Bannore,V.2009. Iterative-Interpolation Super-Resolution Image Reconstruction. Springer, pp. 19-50.

[10] Gilman, A. 2009. *Least-squares optimal interpolation for direct image super-resolution.* Doctoral Thesis, Massey University, Palmerston North New Zealand.

[11] Zibetti, M.V.W., Bazan, F.S.V. and Mayer, J. 2008. Determining the regularization parameters for super-resolution problems. Signal Processing, Issue 12, Vol. 88, pp. 2890-2901.

[12] Sun, J., Xu, Z. and Shum, H.Y. 2008. Image super-resolution using gradient profile prior.Computer Vision and Pattern Recognition .IEEE Conference on . pp. 1-8 (June 23-28, 2008).

[13] Freeman, W.T. and Pasztor, E.C. 1999. Learning low-level vision. Computer Vision,.The Proceedings of the Seventh IEEE International Conference on, Vol. 2, pp. 1182-1189.

[14] Bilgazyev, E., et al. 2011. Sparse Representation-based Super-Resolution for Face Recognition At a Distance. Proceedings of the British Machine Vision Conference. pp. 52.1-52.11.

[15] Zeyde, R., Protter, M. and Elad, M. 2010. On Single Image Scale-Up using Sparse-Representation. Springer, Curves and Surfaces. pp. 711–730.

[16] Xiaoqing, Shutao Li. 2011. Multi-frame Image Super-resolution Reconstruction based on Sparse Representation and POCS. International Journal of Digital Content Technology and its Applications, (August 2011) , Issue Number 8, Vol.5.

[17] He, J.and Zhang,D. 2010. Face super-resolution reconstruction and recognition from low-resolution image sequences. Computer Engineering and Technology (ICCET),2nd International Conference on , Vol. 2, pp. V2-620 (April 16-18, 2010).

[18] Al-Azzeh, M., Eleyan, A. and Demirel, H. 2008. PCA-based face recognition from video using super-resolution.Computer and Information Sciences.ISCIS 08 .23rd International Symposium on, pp. 1-4 (Oct. 2008).

[19] Hennings-Yeomans, P.H., Baker, S. and Kumar, B.V. 2008. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1-8.

[20] Hennings-Yeomans, P.H., Baker, S. and Kumar, B.V.K.V. 2008. Recognition of Low-Resolution Faces Using Multiple Still Images and Multiple Cameras.Biometrics:Theory ,Application and Systems. 2nd IEEE International Conference on , pp. 1-6 (Sept 2008).

[21] Arachchige, S.R.B. 2008. *Face Recognition in Low Resolution Video Sequences using Super Resolution.*Master Thesis. Rochester Institute of Technology. (August 2008).

[22] Irani, M and Peleg,S. 1991. Improving resolution by image registration. CVGIP: Graphical Models and Image Processing, Issue 3, Vol. 53, pp. 231-239.

[23] Peleg, S., Keren, D. and Schweitzer, L. 1987. Improving image resolution using subpixel motion. Pattern Recognition Letters, Issue 3, Vol. 5, pp. 223-226.

[24] Yang, J., et al. 2010. Image super-resolution via sparse representation. IEEE Transactions on Image Processing, Issue 11, Vol. 19, pp. 2861-2873.

[25] Yang, J., et al. 2008. Face hallucination via sparse coding. Image Processing,15th IEEE International Conference on . pp. 1264-1267 (Oct. 12-15,2008).

[26] Rubinstein, R., Bruckstein, A.M. and Elad, M. 2010. Dictionaries for sparse representation modeling. Proceedings of the IEEE, Issue 6, Vol. 98, pp. 1045-1057.

[27] Lee, H., et al. 2007. Efficient sparse coding algorithms. Advances in neural information processing systems. Vol. 19, p. 801.

[28] Sellahewa, H. and Jassim, S.A. 2009. Image quality-based adaptive illumination normalisation for face recognition. Proceeding of SPIE , Vol. 7306, p. 73061V.

[29] Sellahewa, H. and Jassim, S.A. 2010. Image-Quality-Based Adaptive Face Recognition. Instrumentation and Measurement, IEEE Transactions on, Issue 4, Vol. 59, pp. 805-813 (April 2010).