Comparative Evaluation of Wavelet-Based Super-Resolution from Video for Face Recognition at a Distance

E. Bilgazyev S. K. Shah I. A. Kakadiaris

***Abstract*— Face recognition is a challenging problem, especially when low resolution images or image sequences are used for the task. Many methods have been proposed that can combine multiple low resolution images to realize a higher resolution or super-resolved image. Nonetheless, their utility and limitations for use in face recognition are not well understood. In this paper, we present a quantitative and comparative evaluation of wavelet transform based methods for image super-resolution. We evaluate different basis functions, varying levels of decomposition, and multiple methods for coefficient fusion to maximize the benefit of the super-resolved image for the task of face recognition. We have used a Discrete Wavelet Transform and the shift-invariant Dual-Tree Complex Wavelet Transform. Results are reported across both manually generated datasets and data from a surveillance system.**

I. Introduction

Advances in electronics, sensors, and optics have led to ubiquitous availability of video-based surveillance systems. A multitude of applications are driving the need for recognition of individuals from surveillance videos [2]. Among various biometrics possible for ascertaining the identity of an individual, face is by far the most convenient and applicable in surveillance video. However, this modality poses additional challenges to the already difficult problem of face recognition. In typical surveillance scenarios, camera perspectives are distant and subjects to be recognized cover a small area of the image (resulting in poor face image resolutions) [2]. In addition, motion of the subjects may result in blurred and out of focus capture of faces [28]. Solving these issues is critical to building a robust face recognition system that can leverage video from surveillance systems, and in general, solve the problem of face recognition at a distance (FRAD) [2], [19].

The main objective of this paper is to investigate the benefits and limitations of wavelet-based super-resolution methods for the task of FRAD. Specifically, we perform a quantitative evaluation of wavelet-based super-resolution image reconstruction for the purpose of face recognition from surveillance video. We evaluate a variety of basis functions and their ability to maximize performance of a face recognition system along with an evaluation of the added computational cost. In addition, we also consider different fusion techniques [16], [21] to combine wavelet coefficients from multiple frames. The wavelet-based method is compared to other classical reconstruction methods. Additional parameters examined include the optimal number of frames to be used for reconstruction as well as the minimum image resolution of each frame needed in order the to obtain any benefit from super-resolution for the task of face recognition.

The rest of paper is organized as follows: Section II presents previous work, Section III describes our processing pipeline, multi-frame registration, the computation of image coefficients based on the Dual-Tree Complex Wavelet Transform and Discrete Wavelet Transform, and coefficient fusion techniques. A discussion of the experimental design and results are presented in Section IV. Finally, Section V summarizes our findings and offers conclusions.

II. previous work

For many 2D face recognition systems, resolution of the image dictates the ability to detect key descriptors of the facail anatomy (e.g., eyes, lip corners, face contour), and thereby the ability to perform face recognition (FR). As a result, when the resolution of a face region is lower than expected, the performance of the system would be unacceptable [28]. For example, for FaceIt™ [10], a commercial face recognition system, the minimum distance between pupils expected is 60 pixels [19]. In recent years, several methods have been proposed to address the limitation of image resolution [3], [28]. One proposed solution is to down-sample face images in the gallery database to the size of low resolution probe images captured for a subject [31]. Another proposed solution is based on extracting features from face regions in the probe image [6], [33] and matching them to features computed from the high resolution gallery images. However, this is dependent on the ability to extract reliable features and a degradation in resolution beyond a certain point will lead to ungraceful system failures. In addition, blurred probe images would also pose a challenge for this approach, which is common in the case of faces extracted from video. Example-

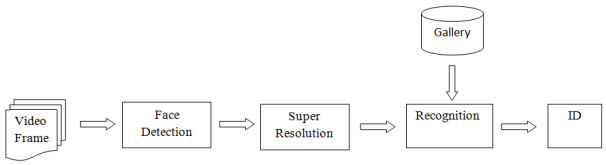


Fig. 1. Flow chart of FR using super-resolution reconstruction from low resolution (LR) images.

All authors are with the Computational Biomedicine Lab, University of Houston, TX, USA. This research was funded in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL) and by the University of Houston (UH) Eckhard Pfeiffer Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, the U.S. Government, or UH.

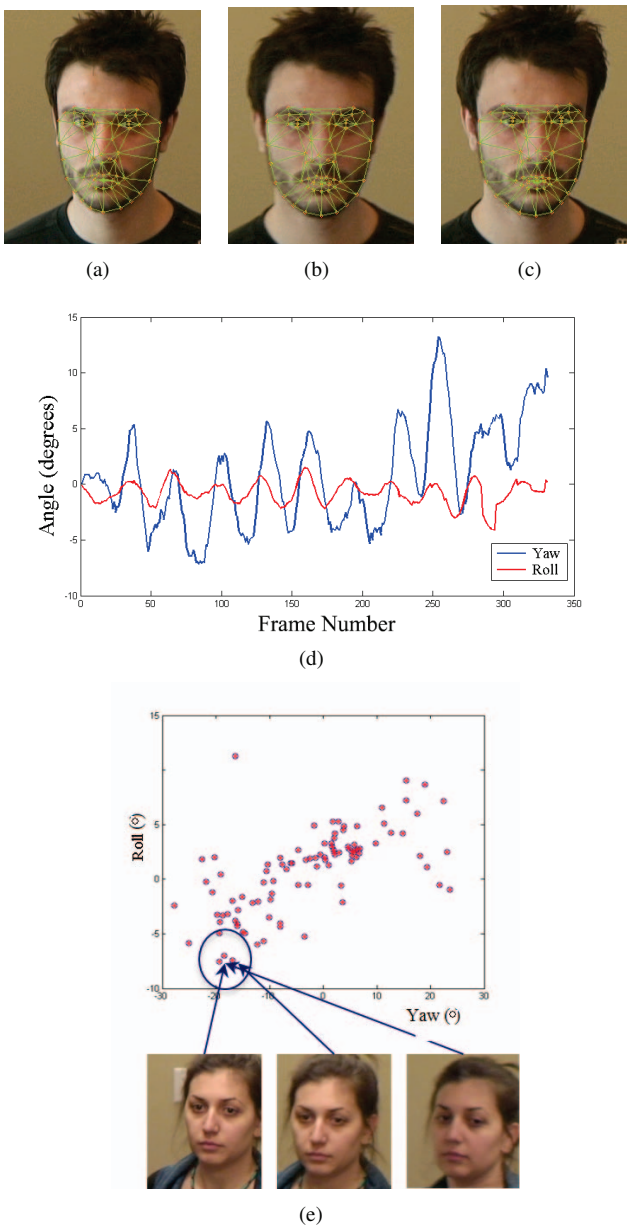


Fig. 2. Output of ASM for (a) -8°, (b) 0°, (c) -16° respectively. Poses are estimated using the PittPatt SDK [18]. (d) Pose estimated by PittPatt SDK. The line in blue shows the variation of yaw, while the red line shows the variation of roll across time. Notice that the changes are periodic over time. (e) By plotting yaw vs roll we can find the nearest neighbors to form a group of similar poses for reconstruction of a super-resolved image.

based high resolution image reconstruction from a given low resolution image was proposed by Kim and Kwon in [9]. However, this approach relies on having a large database of high resolution images under varying pose and illumination conditions. This may not be feasible for all application scenarios.

Another class of methods is based on the reconstruction of a super-resolution face image from multiple low resolution images [3], [28]. Given a surveillance system's video sequence with subjects moving in the scene, face regions can be extracted and combined to obtain a higher resolution image, which in turn can be post-processed for recognition. This approach by far holds the most promise and is known to be a scalable solution for a multitude of FR scenarios. Fig. 1 depicts a typical framework for multi-frame based super-resolution fused for face recognition. First, face regions are extracted from low resolution video sequences and registered. Many super-resolution methods have been proposed [9], [15], [25], [26], [28], [32] to combine the registered images at the second stage. Classical methods range from projection onto convex sets (POCS) [8], [24], non-uniform interpolation [26], and iterative back projection [20]. These methods have high computational cost that may be unacceptable for online application scenarios. Recently, Farisu *et al.* [5] proposed a fast and robust super-resolution (FRSR) algorithm. However, the algorithm output is dependent and sensitive to its parameters (alpha, beta, error, iterations) [5]. Wavelet analysis has been shown to be well suited for reconstruction, denoising, and deblurring, providing accurate and sparse representation of images consisting of smooth regions with isolated abrupt changes [15], [17]. Many investigators have used wavelet-based super-resolution in a variety of application domains including biomedical [8], biometrics [11], and astronomy [30]. Key benefits of wavelet-based methods include lower computational cost [17], minimal memory requirements during reconstruction [30], and high signal-to-noise-ratio (SNR) [8] after reconstruction. Hence, wavelet-based super-resolution reconstruction can be beneficial in addressing the image resolution challenges in FRAD.

III. Methods

A. Multi-Frame Registration

For reconstruction-based super-resolution algorithms, registration is a crucial step. Without properly registering multiple frames, super-resolution algorithms will result in a poor output. In this work, we use two types of registration: landmark and intensity-based registration. An Active Shape Model (ASM) can be used [4], [22] to detect key anatomical landmarks in each image from a video sequence. However, ASM may not result in promising results for low resolution facial images. Thus, in addition to ASM, intensity values are also used for registration. Specifically, the Fourier-Mellin Transformation (FMT) can be used for registering two images [27]. The scale, rotation and translation ($[\mathbf{s,R,T}]_{FMT}$) parameters are obtained by using FMT. Using the feature points detected by ASM, we register all frames to a reference frame. Let f_i be a reference frame's anatomical landmark coordinates ($i \in R^N$ and N is the number of total frames) and let f_j be an input frame's landmark coordinates which will be registered to reference frame ($j \in R^N$). To register f_j to f_i the scale, rotation and translation parameters are estimated to satisfy $f_i \cong W(f_j, \mathbf{c})$, where :

$$W(f_j, \mathbf{c}) = \begin{bmatrix} c_1 & c_2 \\ c_4 & c_5 \end{bmatrix} f_j + \begin{bmatrix} c_3 \\ c_6 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \end{bmatrix} \begin{bmatrix} f_j \\ 1 \end{bmatrix}. \quad (1)$$

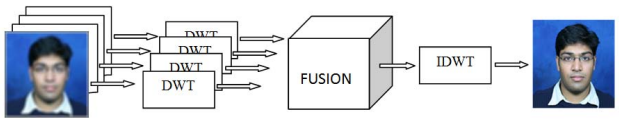


Fig. 3. Flow chart of obtaining a super-resolved image from LR frames/images using a wavelet transform and fusion. After faces are detected from N LR frames and registered to a reference frame, for each face image (on the left). The wavelet transform is applied to extract image information. Then, all information from N face images is fused. After fusion, the inverse wavelet transform is applied.

To estimate the **c** parameters, we use Procrustes Analysis [12]. Estimated parameters by FMT and Procrustes Analysis can be combined using 2, where α is weight parameter:

$$[\mathbf{s},\mathbf{R},\mathbf{T}] = \alpha[\mathbf{s},\mathbf{R},\mathbf{T}]_{ASM} + (1-\alpha)[\mathbf{s},\mathbf{R},\mathbf{T}]_{FMT}. \quad (2)$$

Two ASM models are trained for standard and non-standard pose (Figs.2(a-c)). The PittPatt SDK is used to estimate head pose and to combine similar poses into one group to create a super-resolved image. Fig. 2(d) depicts the pose output of PittPatt for one video sequence obtained by a surveillance camera, where the subject is walking towards the camera. Fig. 2(e) depicts yaw vs. roll angles. The selection of frames to be used for super-resolution reconstruction is based on local grouping of yaw and roll estimates. Let $\Omega(f_i)$ be a function that returns a head pose vector of frame f_i. To form a group of similar poses to f_i 3 is used:

$$\|\Omega(f_i) - \Omega(f_j)\|_2^2 \leq \varepsilon_0, \quad (3)$$

where ε_0 is a threshold. Fig. 2(e) depicts one local group of selected and corresponding frames.

B. Super-Resolution Reconstruction

After having registered a local group of frames, then a wavelet-based super-resolution is used to reconstruct a high resolution image. The general processing flow is as depicted in Fig. 3. In this section, we briefly describe each step of the wavelet-based reconstruction.

1) Dual-Tree Complex Wavelet Transform: The dual-tree complex wavelet transform (DTCWT) is an enhancement to the Discrete Wavelet Transform (DWT), which generates complex coefficients by using a dual tree of wavelet filters to obtain their real and imaginary parts [23]. Selesnick and Kingsbury [23] introduced DTCWT, which has the following important additional properties: shift invariance, good directional selectivity in 2D, perfect reconstruction, and efficient computation compared to general complex wavelets. In DTCWT, to achieve perfect reconstruction and good frequency characteristics, two parallel fully decimated trees with real filter coefficients are used [23]. The 2D DTCWT decomposes an image $f(x,y)$ in terms of a complex shifted and dilated mother wavelet $\psi(x,y)$ and a scaling function $\phi(x,y)$:

$$f(x,y) = \sum_{k,l\in Z^2} a_{k,l}\phi_{k,l}(x,y) + \sum_{\theta}\sum_{k,l\in Z^2} b_{k,l}^{\theta}\psi_{k,l}^{\theta}(x,y), \quad (4)$$

where $a_{k,l}$ is (k,l) translated complex scaling coefficients, $b_{k,l}$ is (k,l) translated complex wavelet coefficients, and $\theta \in (\pm 15^\circ, \pm 45^\circ, \pm 75^\circ)$. The (k,l) translated complex scaling and wavelet functions $\phi_{k,l}$ and $\psi_{k,l}$ are given by:

$$\phi_{k,l}(x,y) = \phi_{k,l}^{re}(x,y) + i\phi_{k,l}^{im}(x,y) \quad (5)$$

$$\psi_{k,l}(x,y) = \psi_{k,l}^{re}(x,y) + i\psi_{k,l}^{im}(x,y), \quad (6)$$

where $i = \sqrt{-1}$. The real and imaginary parts of the DTCWT are computed using separate filter bank structures with wavelet filters h_0, h_1 for the real part and g_0, g_1 for the imaginary part, respectively. The DTCWT is implemented using separable transforms and by combining subband signals appropriately [18]. In 2D, DTCWT produces six wavelet sub-bands of different orientations (-75°, -45°, -15°, 15°, 45°, 75°) and captures image information in that direction. The properties of DTCWT can be shown through a simple synthetic experiment (Fig. 4(a)). The dual-tree complex wavelet transform (Fig. 4(b)) can discriminate between features at positive and negative orientations. Hence, there are six sub-bands capturing features along lines at orientations of (-75°, -45°, -15°, 15°, 45°, 75°), while the ordinary DWT (Fig. 4(c)) cannot discriminate features between -45° and 45°. If we use DWT to extract high frequency components, in the fusion step we may fuse the coefficients at different orientations, when we want to fuse only similar high frequency coefficients.

2) Fusion: In this section, we describe fusion techniques to fuse detail coefficients before applying the inverse transform. After decomposing the image into sub-images using DWT or DTCWT, for each of N images to be fused, we have the approximate and detail coefficients. Many fusion algorithms have been proposed including EM [30], POCS [8], Weighting [17], Maximum Coefficients [17], and region based [13]. In this work, we have used maximum of coefficients (MAX) and general weighted (GW) fusion schemes.

a) Maximum of Coefficients: The most commonly used fusion method is to choose the wavelet coefficients using the maximum absolute value of the sub-band coefficients of the two images to be fused [17]. In this method, the magnitude of the coefficients of images is computed as:

$$M(x,y) = \|D^{re}(x,y)\|_2^2 + \left\|D^{im}(x,y)\right\|_2^2, \quad (7)$$

where $D^{re}(x,y)$ is the real part of the high frequency coefficients of a sub-band, and $D^{im}(x,y)$ is the imaginary part of the high frequency coefficients of the same sub-band. The magnitude of the coefficients is denoted as $M(x,y)$. For the two images to be fused, the new coefficient *D(x,y)* at position *(x,y)* in the fused image can be obtained by:

$$D(x,y) = \begin{cases} D_1(x,y) & \text{if } M_1(x,y) > M_2(x,y) \\ D_2(x,y) & \text{if } M_1(x,y) \leq M_2(x,y) \end{cases}, \quad (8)$$

where D_1 and D_2 are the high frequency complex coefficients of the first and second image.

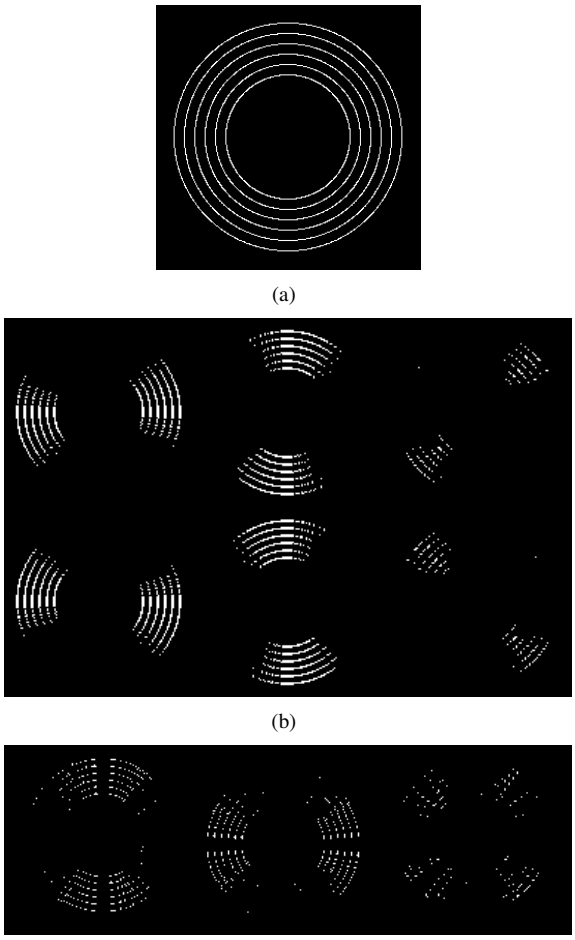


Fig. 4. Binary image (256x256) depicting circles with different radii. (a) Synthetically generated circles with different radii. (b) Output of 2D DTCWT. One can observe that each wavelet subband responds for particular orientation, and 2D DTCWT can discriminate between positive and negative orientations. (c) Output of ordinary DWT. One can observe that DWT can not distinguish orientations -45° and 45°.

b) General Weighted Average: The General Weighted Average fusion technique assigns a weight to each sub-band to be fused:

$$F(x,y) = \sum_{i=0}^{N-1} w_i D_i(x,y), \tag{9}$$

where w_i is the weight for fusion of high frequency coefficients of image I_i. One way to obtain the weights is suggested by Pajares and Cruz [17]. Given two images I_1 and I_2 and corresponding magnitude high frequency coefficients M_1 and M_2, weighted fusion would be given by:

$$D = w_1 D_1 + w_2 D_2, \tag{10}$$

where w_1 and w_2 are the weights. Let m be the normalized correlation average over a neighborhood between $M_1(x,y)$ and $M_2(x,y)$ magnitude of sub-images, and let α be a threshold. Then, w_1 and w_2 can be given by:

$$\begin{cases} w_1 = 0, & w_2 = 1 & m < \alpha \\ w_1 = \frac{1}{2}\left(\frac{m-\alpha}{1-\alpha}\right), & w_2 = 1 - w_1 & \text{otherwise} \end{cases} \tag{11}$$

Combining equation 10 and 11:

$$D = \begin{cases} D_2 & m < \alpha \\ D_2 - \beta(D_1 - D_2) & \text{otherwise} \end{cases} \tag{12}$$

where

$$\beta = \frac{1}{2}\left(\frac{m-\alpha}{1-\alpha}\right). \tag{13}$$

IV. Experiments

Several experiments were performed to evaluate the performance of wavelet-based super-resolution (WBSR). We compare WBSR (in terms of performance gain and reconstruction time) with other super-resolution algorithms. In addition, we assess the minimum image resolution (in terms of pupillary distance) and the number of frames required for WBSR reconstruction to affect the performance of a face recognition system. Our experiments were designed to investigate the limitations of WBSR for use in face recognition from surveillance video.

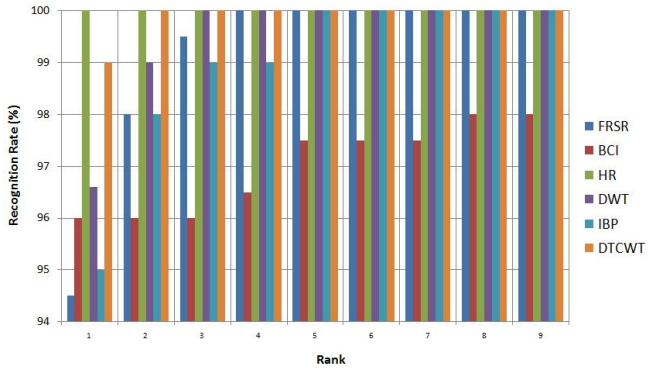
The first set of experiments involves the use of synthetic data derived from a single HR image. For the second test, we used LR video sequences obtained from a surveillance system installed at the University of Houston. In both tests, we have used PittPatt FTR SDK 4.0 for face detection and recognition, and the parameters were $\alpha = 0.5$ for a GW fusion scheme.

A. Synthetic Data

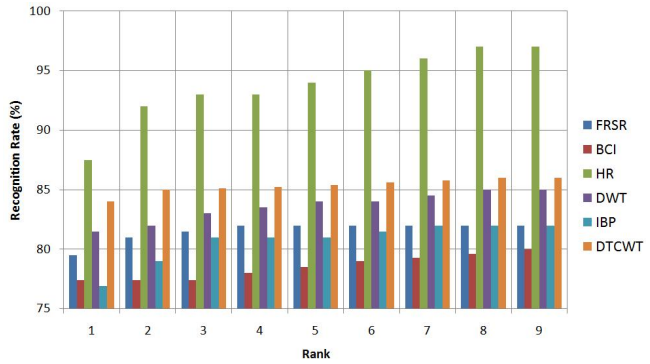
The first set of experiments involves the use of synthetic data derived from a single HR image. We compare the recognition rate of high resolution, synthetic low resolution, and reconstructed images. We have used Pupillary Distance (PD) (distance between two eye centers) to define the resolution of the face. Two datasets were used: XM2VTS [14] and CMU MultiPIE [7]. We have used 193 subjects and 202 subjects from XM2VTS and MultiPIE, respectively. The gallery from the XM2VTS dataset was based on images from one session with no side illumination effects, and the probe set was constructed from images in a different session (CD008) where subjects were illuminated from the right or left. For the CMU MultiPIE dataset, two different sessions (first and fourth) were used with no side illumination, one for gallery and the other for probe. Each HR image in the probe set was rotated, translated and down-sampled to create 20 LR images with PD of about 10-12 pixels. For wavelet-based super-resolution, we used different fusion algorithms, decomposition levels and wavelet basis functions for DWT (Table I). For DTCWT we used nearly shift invariant wavelet

TABLE I
DIFFERENT FUSION SCHEMES, DECOMPOSITION LEVELS AND BASIS FUNCTIONS WERE USED FOR DWT-BASED SR RECONSTRUCTION.

	DWT	DTCWT
Fusion	MAX, GW	MAX, GW
Level	2,3,4	2,3,4
Basis	haar, db8, db16, sym4, sym8, sym16 coif2, coif4, bior1.3, bior1.5, bior3.1	Abdelnour *et.al.* [1]

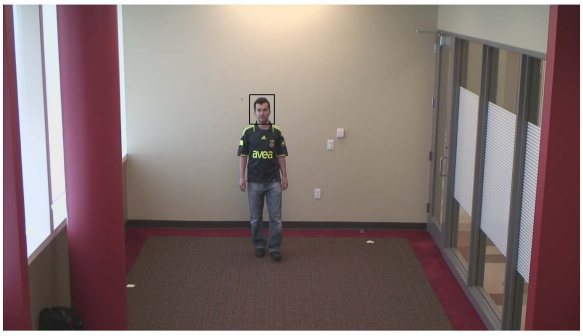


(a)

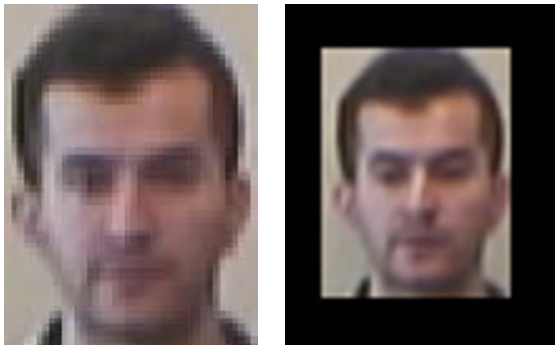


(b)

Fig. 5. Recognition rate for the CMU MultiPIE and XM2VTS datasets. (a) For the CMU MultiPIE dataset, the sym16 wavelet family filter (two level decomposition) and DTCWT with weighting-based fusion resulted in the best rank-one recognition rate of 99%. There is almost no difference between HR and WBSR curves. (b) For the XM2VTS dataset, the coif2 wavelet family filter (two level decomposition) with weighting-based fusion resulted in 0.81 rank-one recognition rate with small computation time, and with an additional computation time DTCWT resulted in 0.84 rank-one recognition rate (HR rank-one is 87% and FRSR is 76%).



(a)



(b) (c)

Fig. 6. Illustration of the surveillance camera system (installed at University of Houston) output and super-resolved image. (a) A frame acquired by the surveillance camera. The output of PittPatt face detection is marked by a bounding box. (b) Magnification of bounding box (PD=11 pixels). (c) SR image obtained using wavelet-based decomposition with 'db16' and two level decomposition.

proposed by Abdelnour *et al.* [1]. In all, this resulted in 438 different experiments across the two datasets. Fig. 5 depicts the recognition rates for the WBSR with the optimal combination of fusion algorithm, decomposition level, and the basis function that resulted in the maximum face recognition performance.

Next, we compared DWT- and DTCWT-based super-resolution algorithms with Fast Robust Super-Resolution (FRSR) [5], Iterated Back Projection (IBP) [20], and Bicubic Interpolation (BCI) on both the datasets. For FRSR, we used the same parameters as Wheeler *et al.* [28]. Table II summarizes rank-one recognition rates and reconstruction times for the different super-resolution methods. Note that, WBSR achieves almost the same recognition rate as the original HR image with relatively small additional computational cost.

B. Video Sequences: UHBD14

For the second set of experiments, we used LR video sequences obtained using a version of the Biometric Surveillance System (BSS) [29] installed at the University of Houston. Persons are detected and tracked using a stationary wide-field-of-view camera and a near-field-of-view pan-tilt-zoom camera is automatically controlled to collect high resolution facial images. Persons are detected using background subtraction and tracked in the ground-plane with a Kalman filter. The system repeatedly selects a subject for imaging to capture a facial image. Facial images can be captured at distances up to 2550m. HR images of 32 subjects were obtained using a DSLR camera in a controlled environment and used as a gallery. Three months later, 32 subjects (34 videos: 32 subjects without glasses and 2 subjects with glasses) were asked to walk through the surveilled environment, and LR video sequences were captured and used as probes. Each video has 350-400 frames with a pupillary distance of 8-12 pixels. Fig. 6 depicts the environment, an example of a detected face, and the corresponding super-resolved image. Once again, we compared different basis functions, decomposition levels, and fusion methods for WBSR to maximize recognition performance. Since each video has a large number of frames, we initially form local groups based on pose estimates and each group is combined to generate a single super-resolution image. This results in several super-resolution images per video. Face recognition is performed across all SR images and the highest of the scores is selected. Our results indicate that DTCWT with two level decomposition and fusion based on general weighted average results in the highest rank-one recognition.

TABLE II

RANK-ONE RECOGNITION RATES AND RECONSTRUCTION TIME ('-' REPRESENTS THAT THE PITTPATT FACE DETECTOR DIDN'T DETECT FACES IN ALL LR FRAMES). NOTE THAT THE BICUBIC INTERPOLATION REQUIRES LESS RECONSTRUCTION TIME, BUT RESULTS IN LOW RECOGNITION RATE. HOWEVER, DWT (SYM16, 2 LEVEL DECOMPOSITION) RESULTS IN HIGH RECOGNITION RATE AND IN FAIRLY SHORT RECONSTRUCTION TIME. WHEN ALLOWING ADDITIONAL RECONSTRUCTION TIME BETTER RECOGNITION RATE CAN BE OBTAINED USING DTCWT FOR BOTH SYNTHETIC AND REAL DATA.

Database SR Name	LR	HR	FRSR	IBP	BCI	DWT	DTCWT
MultiPIE	66.9%	100.0%	94.5%	94.9%	96.0%	99.0%	99.0%
XM2VTS	-	87.5%	79.5%	76.9%	77.4%	81.5%	84.0%
Real Data	14.5%	-	51.6%	46.8%	40.6%	51.6%	56.2%
Recon Time (s)	-	-	38	71	2.6	13	21

Furthermore, we systematically evaluated the best number of frames to be used for WBSR so as to maximize the gain in face recognition accuracy. We observed that the recognition rate is maximized when 8-12 consecutive LR frames are used as input to WBSR (Fig. 7(a)). Given our surveillance system, which captures approximately 16.5 frames/sec, this amounts to video frames acquired for 0.5-0.75 seconds. We also evaluated the lower bound of image resolution for which WBSR would generate a super-resolved image that could be used for face recognition. Fig. 7(b) indicates that there is a significant decrease in recognition rate if the SR frames have a PD less than 12.

V. CONCLUSION

In this paper, we have presented a comparative analysis of wavelet-based super-resolution using different fusion algorithms for the purpose of face recognition from surveillance video. The wavelet-based methods were compared to common super-resolution algorithms and the benefits and limitations were evaluated in terms of the optimal number of frames to be used for reconstruction as well as the minimum image resolution of each frame needed in order the obtain any benefit from super-resolution for the purpose of face recognition. In general, we found WBSR to perform better than popular algorithms such as Fast and Robust Super-Resolution and with significantly lower computational cost.

REFERENCES

[1] A. F. Abdelnour and I. W. Selesnick. Symmetric nearly shift-invariant tight frame wavelets. *IEEE Transactions on Signal Processing*, 53(1):231–239, 2005.

[2] M. Ao, D. Yi, Z. Lei, and S. Z. Li. *Handbook of remote biometrics*, chapter Face Recognition at a Distance: System Issues, pages 155–167. Springer London, 2009.

[3] O. Arandjelovic and R. Cipolla. A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In *Proc.* 11^{th} *IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, Oct. 14-20 2007.

[4] T. Cootes and C. Taylor. Active shape models: Smart snakes. In *Proc. British Machine Vision Conference*, pages 266–275, Leeds, UK, Sep. 22-24 1992.

[5] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004.

[6] S. Gillan, P. Agathoklis, and M. Yasein. A feature based technique for face recognition using mexican hat wavelets. In *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pages 792–797, Victoria, B.C., Canada, Aug. 2009.

[7] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28:807–813, 2010.

[8] J. T. Hsu, C. C. Yen, C. C. Li, M. Sun, B. Tian, and M. Kaygusuz. Application of wavelet-based POCS superresolution for cardiovascular MRI image enhancement. In *Proc.* 3^{rd} *International Conference on Image and Graphics*, pages 572 – 575, Hong Kong, China, Dec. 2004.

[9] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1127–1133, 2010.

[10] L1 Identity Solutions. L1 faceit SDK.

[11] C. Liu and D. Dai. Face recognition using dual tree complex wavelet features. *IEEE Transactions on Image Processing*, 18, 2006.

[12] L.Wang, H.Ning, W.Hu, and T.Tan. Gait recognition based on procrustes shape analysis. In *Proc. IEEE International Conference on Image Processing*, volume 3, pages 433–436, Rochester, New York, 2002.

[13] B. J. Matuszewski, L. K. Shark, J. P. Smith, and M. R. Varley. Automatic fusion of multiple non-destructive testing images and cad models for industrial inspection. In *Proc.* 7^{th} *International Conference on Image Processing and Its Applications*, volume 2, pages 661–665, Kobe, Japan, 1999.

[14] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Proc. Audio- and Video-Based Biometric Person Authentication*, pages 72–77, Washington, DC, Mar. 22-23 1999.

[15] N. Nguyen and P. Milanfar. A wavelet based interpolation restoration method for superresolution. *Circuits Systems Signal Processing*, 19:321–338, 2000.

[16] P. H. Nishan, P. Hill, N. Canagarajah, and D. Bull. Image fusion using complex wavelets. In *Proc.* 13^{th} *British Machine Vision Conference*, pages 487–496, Cardiff,UK, Sep. 2-5 2002.

[17] G. Pajares and J. M. Cruz. A wavelet based image fusion tutorial. *Pattern Recognition*, 37(9):1855–1872, Sep. 2004.

[18] Pittsburgh Pattern Recognition. PittPatt face tracking & recognition software development kit, 2009.

[19] S. Prince, J. Elder, Y. Hou, M. Sizinstev, and E. Olevsky. Towards face recognition at a distance. In *Proc. Institution of Engineering and Technology Conference on Crime and Security*, pages 570 –575, London,UK, Jun. 2006.

[20] F. Qin, X. He, W. Chen, X. Yang, and W. Wu. Video super-resolution reconstruction based on subpixel registration and iterative back projection. *Journal of Electronic Imaging*, 18(1):013007, 2009.

[21] D. Rajan and S. Chaudhuri. Data fusion techniques for super-resolution imaging. *Information Fusion*, 3(1):25–38, 2002.

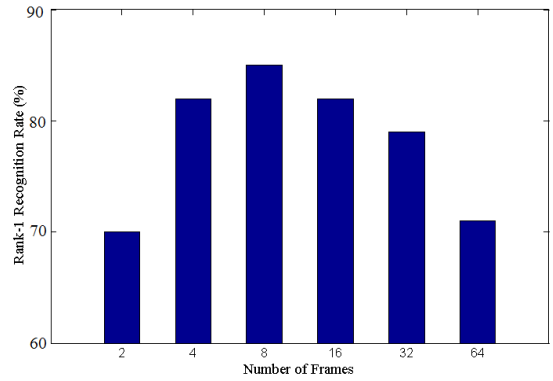
[22] M. Rogers and J. Graham. Robust active shape model search. In *Proc. European Conference on Computer Vision*, pages 517–530, London, UK, 2002.

[23] I. W. Selesnick, R. G. Baraniuk, and N. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazines*, 22(6):123–151, 2005.

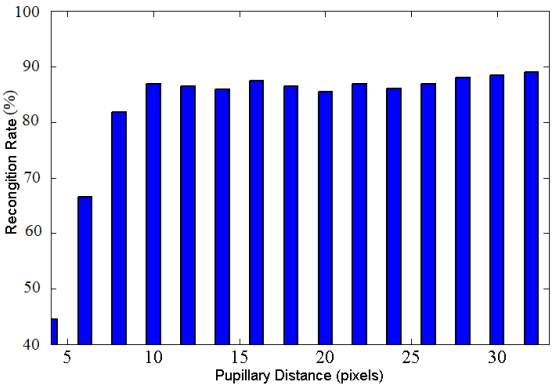
[24] M. I. Sezan. An overview of convex projections theory and its application to image recovery problems. *Ultramicroscopy*, 40(1):55–67, 1992.

[25] M. Shimizu, Y. Shin, M. Tanaka, and M. Okutomi. Super-resolution from image sequence under influence of hot-air optical turbulence. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, Anchorage, AK, Jun. 24-26 2008.

[26] J. Sun, Z. Xu, and H. Shum. Image super-resolution using gradient profile prior. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, Jun. 24-26 2008.



(a)



(b)

Fig. 7. Recognition Rate vs Number of frames and resolution required for WBSR (a) Number of frames for reconstruction vs. recognition rate. When two consecutive frames are used for reconstruction 94% of faces were detected, while when 4 or 64 consecutive frames were used for reconstruction 97% of faces were recognized. The best recognition rate was obtained when we used 10-12 consecutive frames (0.6-0.8s) for reconstruction. (b) Distance between pupils in pixels vs recognition rate. The minimum pupillary distance for PittPatt Face Recognition should be no less than 12 pixels. The performance of face recognition decreases with decreased PD.

[27] C. Wang, Y. Cheng, and C. Zhao. Robust subpixel registration for image mosaicing. In *Proc. Chineese Conference on Pattern Recognition*, pages 1–5, Beijing, China, Apr. 2009.

[28] F. Wheeler, X. Liu, and P. Tu. Multi-frame super-resolution for face recognition. In *Proc.* 1^{st} *International Conference on Biometrics Theory, Applications and Systems*, pages 1–6, Washington D.C, Sep. 27-29 2007.

[29] F. Wheeler, R. Weiss, and P. Tu. Face recognition at a distance system for surveillance application. In *Proc.* 4^{th} *International Conference on Biometrics Theory, Applications and Systems*, Washington D.C, Sep. 27 -29 2010.

[30] R. Willet, I. Jermyn, R. Nowak, and J. Zerubia. Wavelet based super resolution in astronomy. In *Proc. Astronomical Data Analysis Software and Systems*, volume 314, pages 107–116, Strasbourg, France, 2003.

[31] Y. Xu and Z. Jin. Down-sampling face images and low-resolution face recognition. In *Proc.* 3^{rd} *International Conference on Innovative Computing Information and Control*, pages 392–392, Dalian, China, Jun. 2008.

[32] D. Zhang, H. Li, and M. Du. Fast MAP based multiframe super resolution image reconstruction. *Image and Vision Computing*, 23(7):671–679, 2005.

[33] L. Zhuang, M. Wang, W. Yu, N. Yu, and Y. Qian. Low-resolution face recognition via sparse representation of patches. In *Proc.* 5^{th} *International Conference on Image and Graphics*, pages 200–204, Xi'an, China, Sep. 2009.