Multi-Frame Super Resolution for Ocular Biometrics

Narsi Reddy*, Dewan Fahim Noor*, Zhu Li, Reza Derakhshani
Dept. of Computer Science and Electrical Engineering
University of Missouri at Kansas City, MO
{sdhy7, dfnrh2}@mail.umkc.edu, {lizhu, derakhshanir}@umkc.edu
*Authors contributed equally to this work

Abstract

Some biometrics methods, especially ocular, may use fine spatial information akin to level-3 features. Examples include fine vascular patterns visible in the white of the eyes in green and blue channels, iridial patterns in near infrared, or minute periocular features in visible light. In some mobile applications, an NIR or RGB camera is used to capture these ocular images in a "selfie" like manner. However, most of such ocular images captured under unconstrained environments are of lower quality due to spatial resolution, noise, and motion blur, affecting the performance of the ensuing biometric authentication. Here we propose a multi-frame super resolution (MFSR) pipeline to mitigate the problem, where a higher resolution image is generated from multiple lower resolution, noisy and blurry images.

We show that the proposed MFSR method at 2× *upscaling can improve the equal error rate (EER) by* 9.85% *compared to single frame bicubic upscaling in RGB ocular matching while being up to* 8.5× *faster than comparable state-of-the-art MFSR method.*

1. Introduction

Biometrics plays an important role in protecting data and physical access these days, especially for mobile devices which are misplaced or lost more often than other devices. Nowadays smart phones are the primary computing device for an exceedingly larger portion of the population, and in that context mobile biometrics are considered a safe and convenient password replacement to prevent unauthorized access to the device and its sensitive data. One of the popular biometrics traits which gained attention of research and industry alike is ocular biometrics in visible spectrum. This is because ocular images can be easily captured just using the selfie camera without the need for additional hardware. In mobile ocular biometrics, visible parts of ocular region such as the iris, vascular patterns seen over the sclera and periocular regions are used to authenticate users. Studies show that the methods using local features may outperform those based on global features for ocular and periocular biometrics in visible light [27]. These local feature methods mainly depend on fine spatial features such as vascular patterns over the white of the eye, and skin texture of the periocular region.

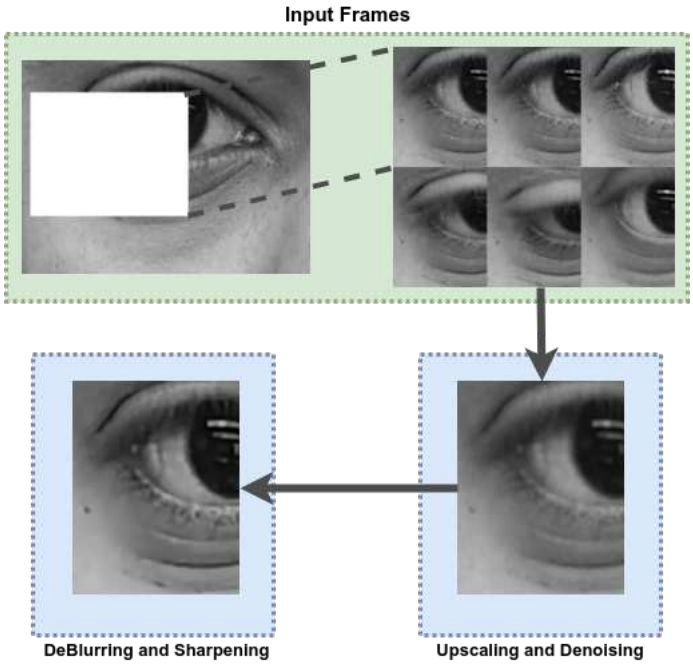

Figure 1. The block diagram of the proposed multi-frame super resolution method.

However, matching these local features via mobile device may cameras introduce some challenging problems. Since the images are captured using "selfie" cameras in unconstrained environment, they can have varying types of illumination, noise, blur, and user to camera distance, which will affect the performance of biometric matching methods. Even though there is a rather considerable body of research

in the area of multi-frame image enhancement and super resolution (SR), these techniques have not been successfully applied to visible light mobile ocular biometrics.

SR techniques have been used in other biometrics traits such as face [3, 31, 12], iris [29, 26] and fingerprint [19, 2]. Most of these methods are single image super resolution(SISR) techniques where the image enhancement process is done using the information available in a single capture. But in case of multiframe super resolution (MFSR), image enhancement process is done using information from more than one image, usually for a burst of successive captures, providing better results compared to SISR [13, 15]. Such bursts of captures can be easily obtained with modern camera modules, including those used on the front of mobile devices.

In this paper, we propose a MFSR pipeline for improving the performance of mobile ocular biometrics. The proposed MSFR method has two stages, (1) An upscaling and denoising stage using discrete cosine transform interpolation filter(DCTIF) based sub-pixel registration and block motion estimation followed by bilateral interpolation, and (2) a deep learning based deblurring model for removing defocus blur. We explain this process in more detail in section 2. We also compare our proposed MFSR method to state-of-the-art MFSR methods when applied to ocular biometric matching.

The rest of the paper is organized as follows. The proposed method is explained in section 2. The experimental setup is provided in section 3. Results are shown in section 4 and conclusion is drawn in section 5.

2. Proposed Method

Let y_k denote the k^{th} frame captured from the mobile RGB camera. Each frame is degraded with noise (η_k), blur (ψ_k) and projection (θ_k) a combination of rotation, translation and scale.

Assuming the multiple degraded frames are generated from one high resolution (HR) image, we can formulate the degraded frames captured from the mobile RGB camera as follows,

$$y_k = \psi_k \theta_k X + \eta_k, \quad where\ k = 1, 2, 3, ..., n \tag{1}$$

From equation 1, approximation for HR image can be written as,

$$\begin{aligned} X = \psi\, \theta\, Y + \eta, \quad where\ \psi = (\psi_1, ..., \psi_n)^{-1}, \\ \theta = (\theta_1, ..., \theta_n)^{-1}, \eta = (\eta_1, ..., \eta_n)^{-1}, Y = (y_1, ..., y_n) \end{aligned} \tag{2}$$

In this paper, we divided the HR image (X) approximation into two parts. First, approximation for projection (θ) and noise (η) using a multiframe upscaling technique where all the frames upscaled and registered with DCTIF based interpolation and subpixel motion registration. Second, a bilateral filtering method to combine these registered frames to generate upscaled and denoised image. Finally, approximation for blur and smoothness (ψ) is done on a single upscaled image using a deep learning method.

We introduce the proposed mutiframe upscaling using subpixel registration and bilateral interpolation for generating the HR images in section 2.1. This is followed by the proposed deep learning-based image deblurring and sharpening technique in section 2.2. Figure 1 shows the block diagram of the proposed method.

2.1. Multi-Frame Upscaling

2.1.1 Subpixel Registration

Subpixel registration is the first step in image enhancement, assuming that both enrollment and verification datasets have frame bursts available for each capture. For upsampling during the subpixel registration, an interpolation filter is needed. Here, we used a DCTIF filter. DCTIF is an interpolation filter based on DCT. It is a fractional pixel interpolator. Since DCT exhibits properties like the optimal Karhunen-Loeve Transform (KLT) [10], it is quite efficient. The equation of DCT -II (2D dimension) is as follows [23]:

$$X(K) = \sqrt{\frac{2}{N} \sum_{n=0}^{n-1} c_k x(n) cos\left(\frac{(n + 0.5)\pi k)}{N}\right)} \tag{3}$$

where *x(n)* contains the original pixel values, and *X(k)* contains the DCT domain values (*k=0,1,2…., N-1.*). The inverse DCT-II equation is as follows:

$$x(n) = \sqrt{\frac{2}{N} \sum_{k=0}^{N-1} c_k X(k) cos\left(\frac{(n + 0.5)\pi k)}{N}\right)} \tag{4}$$

$$c_k = \begin{cases} \frac{1}{\sqrt{2}}, k = 1 \\ 1, otherwise \end{cases} \tag{5}$$

If we substitute forward DCT-II from equation-3 into inverse DCT-II in equation-4, we get the interpolation formula as follows:

$$\begin{aligned} x\left(i + \frac{P}{L}\right) = \sum_{m=0}^{N-1} x(m) \frac{2}{N} \sum_{k=0}^{N-1} c_k^2 cos\left(\frac{(m + 0.5)\pi k}{N}\right) \\ cos\left(\frac{(i + \frac{P}{L} + 0.5)\pi k}{N}\right) \end{aligned} \tag{6}$$

where *i+p/L* is the new interpolated position, *i* is the previous position, *L* is the upsampling factor, and *x (m)* is the pixel values of the neighborhood.

From the subsequent low-resolution images, using subpixel motion estimation, a frame is predicted. To predict the frame, block motion estimation algorithm is used. In general, subpixel motion accuracy is achieved by searching the best matching block in an interpolated reference search area. The method integrates the block matching algorithm and optical flow method to estimate the motion. The DCTIF interpolation filter is used for the interpolation. For each matching block in a new predicted frame, the match is found when it has the lowest matching error in the reference. Error is computed as the sum of absolute errors between corresponding pels. Displacement between the current and previous matching block is the motion vector [25] which also denotes the direction. In an iterative process, current matching block is replaced by the best one.

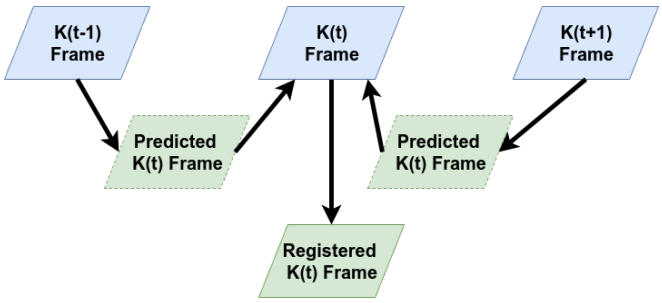

Figure 2. Predicting and registering frames $K(t-1)$ and $K(t+1)$ from $K(t)$ frame.

Figure 2 shows the general flow diagram of predicting a frame from consecutive frames. Once the predicted images are generated, they are registered to the image for which the are predicted took place. For registering images, an affine transform [21] is considered. Figure 3 shows the pixel intensities of a 3X3 image segment after the registration.

2.1.2 Applying Bilateral Filter in Upscaled Grid

After the registration, the LR images are registered into HR image grid. For the upsampling factor 2, having $k(t-2)$, $k(t-1)$, $k(t+1)$, $k(t+2)$, $k(t+3)$, frames for prediction of $k(t)$ and then registration, the HR grid is $2n \times 2n = 4n^2$. The registered frames have $n^2+n^2+n^2+n^2+n^2 = 5n^2$ grid. So, some of the pixels will be in same location and some of the locations will have missing pixels.

In Figure 4, the circles are the pixel values from 5 different frames which are registered (red, yellow, black, blue, green colors denote pixel values of 5 different frames). The crosses are the missing pixel locations. The positions where overlapped pixels are prevailing, pixel values are averaged. At the predict missing holes, bilateral filter [9] is used. Unlike other filter, bilateral filter not only takes care of the position of the pixel but also of the photometric distance of the

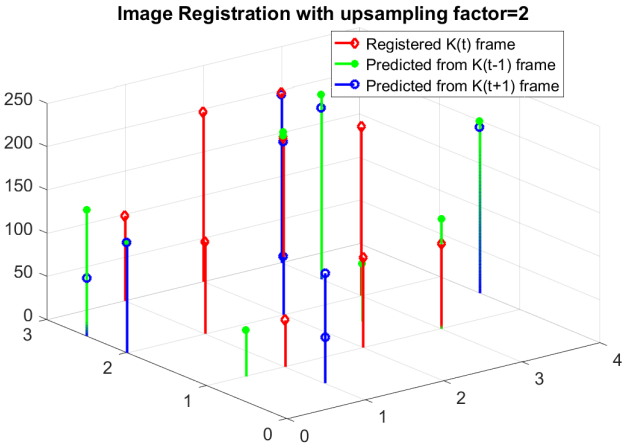

Figure 3. Pixel intensities of a 3×3 image segment after the sub pixel registration from two subsequent frames with affine transform. The pixel intensities from the predicted frames lie within half unit of the location axis of the resolution grid of registered k(t) frame due to half pixel accuracy.

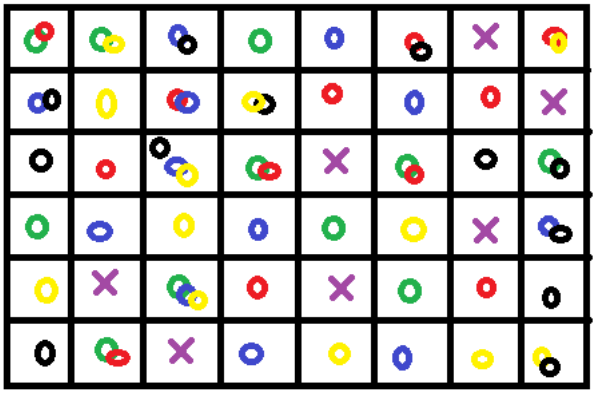

Figure 4. Overlapped and missing pixels after registration.

neighborhood pixels. Following are the equations for computing weights and predicted output using bilateral filter

$$w_{ij} = e^{-\frac{|p_i - p_j|^2}{2(\sigma_d)^2}} . e^{-\frac{|x_{in}[i]-x_{in}[j]|^2}{2(\sigma_r)^2}} \tag{7}$$

$$x_{out}[j] = \int_i^n \frac{w_{ij}}{\int_i^n w_{ij}} x_{in}[i] \tag{8}$$

where i is the pixel location of the input, j is the pixel location of the output, and x_{in} and x_{out} are the input and output pixel values respectively. w_{ij} is the weight, σ_d and σ_r are the heat kernel parameters.

2.2. Deep Learning based deblurring

For deblurring process we incorporated DenseNet architecture [14], where all the features from previous layers is

propagated to the next layer to extract new features. Assuming k channels and a model depth or number of layers as d, at each layer the features from all the previous layers are concatenated ($k*(d-1)$ channels). A 1×1 kernel is applied to squeeze all the features back to k channels. To extract new features, a 3×3 convolution operation with k channels is applied. Unlike DenseNet, in our method we used scaled exponential linear units(SELU) activation [18] instead of batch normalization(BN) with rectified linear unit(ReLU). This is because SELU has self normalization properties which eliminates the BN operation and SELU is also computationally lighter than BN. The DenseNet model is also used to extract residual image which is subtracted from the input image as shown in Figure 5.

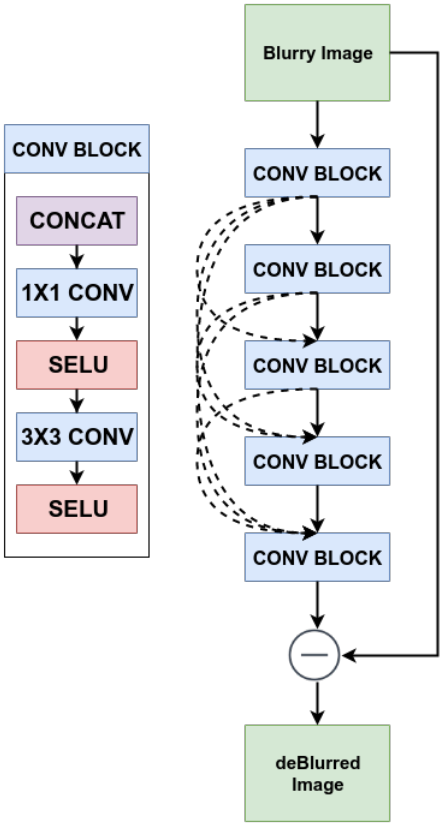

Figure 5. DenseNet-based architecture for deblurring.

As there are no pooling layers and only 3×3 kernels, the effective receptive window of the deep learning model is $(2d+1)\times(2d+1)$. So, each new pixel at the output of the model is generated by extracting features from $(2d+1)\times(2d+1)$ pixels in input image. In the proposed deep learning model, with a model depth of $d=5$ and number of channels $k=24$, the effective receptive window is 11×11 pixels.

Loss function: Mean square error(l_2) and mean absolute error (l_1) are the most commonly used loss functions in the super resolution literature [6, 16] and they achieved better PSNR values. From equation- 9 it can be seen that the error is calculated only between each pixels of the predicted image $\hat{x}_i$ and its target x_i. This tends to cause deformations in structure. In literature, it is also shown that MSE and consequently PSNR do not correlate with human image quality perception [32]. Especially in ocular biometrics, it is necessary to maintain the low level (level-3) texture information as most ocular biometric information is derived from them [11, 27].

$$Loss = \frac{1}{N}.\sum_{i}^{N}\|(x_i-\hat{x}_i)\|_n \tag{9}$$

In order to maintain these level-3 features, we incorporated structural similarity index(SSIM) [30] as loss function. As shown in equation 10, SSIM loss is calculated at each pixel i by considering structural information from the surrounding pixels window $K\times K$.

$$SSIM(i) = \frac{(2\mu_x\mu_{\hat{x}}+C_1)(2\sigma_{x\hat{x}}+C_2)}{(\mu_x^2+\mu_{\hat{x}}^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{10}$$

$$Loss = 1-SSIM \tag{11}$$

Where μ and σ^2 are mean and variance of pixels in a $K\times K$ window around predicted pixel $\hat{x}_i$ and its target x_i. $\sigma_{x\hat{x}}$ is the covariance.

Standard window size used in literature is 11×11. However, as features in ocular images are way smaller than the standard window size, we chose a 5×5 window to calculate the SSIM.

The proposed model is trained using stochastic gradient descent(SGD) optimizer[5] with starting learning rate of 0.01 and a 0.9 momentum. After every 20 epochs, the learning rate is reduced by $1/10^{th}$ of initial rate. This process is continued for 50 epochs.

Training dataset: We generated training data from Describable Textures Dataset(DTD) [4] which contains image patches of various textures. Figure 6 shows some samples from DTD dataset we handpicked for training the deblurring model. The training dataset of $100K$ samples is created by randomly cropping 64×64 regions from the DTD dataset. During training, with batch size of 32, randomly Gaussian blur with standard deviation varying from $\sigma = 0.5$ *to* 2.0 and small amount of Gaussian noise is applied to each sample.

3. Experimental Setup

3.1. Ocular Biometrics Datasets

To test the ocular biometrics performance, we used original raw data form the VISOB dataset [28] collection to generate multi-frame eye crops. VISOB dataset was collected using three different mobile devices (iPhone 5, Oppo N1,

Figure 6. Deblurring model training dataset samples from describable texture dataset.

Lighting	Session - I	Session - II	images/subject
Dim Light	1407	1423	28.3
Daylight	1461	1323	27.84
Office	1237	1881	31.18
Total	**4105**	**4627**	**29.11**

Table 1. Characteristic of the multi-frame ocular biometrics dataset generated for performance evaluation.

and Samsung Note 4) in three different lighting conditions (normal office, daylight and dim office). All the images were captured in burst mode in two visits separated by 2 to 4 weeks apart and in each visit images were collected in two sessions each 10 to 15 minutes apart. The volunteers held devices in selfie mode and the software on the device captured multiple images in burst mode.

Figure 7. Samples from dataset exhibiting noise, lighting and pose variations. All the generated samples have 6 consecutive frames.

For experiments in this paper we randomly choose 50 subjects out of 550 from OPPO device. To generate samples, we selected 6 consecutive frames which are at least 96% spatially correlated. Then, using dlib [17] face detector and face landmark localization, we found the eye region in the first of 6 frames and then same region is cropped from all the other frames. All the cropped images are resized to 100X135 pixels. Table 1 shows number of samples per lighting condition per session. For evaluation, Session-I and session-II are considered as enrollment and verification sets respectively. Figure 7 shows some samples from the dataset displaying variations in noise, blur, lighting and eye movements.

3.2. Biometric Matcher

Our goal is to design an MSFR method which increases the performance of a level-3 based ocular biometrics matching. To that end, we chose Speed Up Robust Features(SURF) point detector and descriptor [1]. SURF is partially based on Scale Invariant Feature Transform(SIFT) descriptor [22], but is faster. SURF feature points are detected using an approximated Gaussian blob detector and their orientation is calculated using Haar-wavelet responses. Gaussian smoothening is approximated using a square shaped filter for faster processing. These detected feature points are localized in scale and space. For each feature point, a 128 dimensional extended feature descriptor is extracted from a square window centered on the feature point and using its orientation. In our case, all the feature points are detected only in the first 2 octaves with 4 scales at each octave given the resolution of our source data.

For feature matcher, we incorporated the technique used in [11]. We first generated matched point pairs between enrollment image and verification image using nearest neighbor symmetric match (NNS) criteria [24]. Then, to remove the outliers from the matched point pairs, RANSAC algorithm [8] was used. Finally, inlier count was considered as the match score. Equal error rate(EER) is considered as the scalar performance metric.

3.3. Comparison With State-of-the-art Methods

To compare the biometrics matching performance and computation time of the proposed method, we chose six notable super resolution techniques. Out of the six, three are single frame methods and the other three multi-frame methods. As the baseline for both single frame and multi-frame SR methods, bicubic interpolation was chosen. In case of multi-frame bicubic upscaling, we averaged all the frames. For single frame, we chose SRCNN [6] and VDSR [16], as they are most commonly cited and benchmarked methods in SISR. In MFSR, we choose Maximum a-posteriori (MAP) based method [7] and iteratively re-weighted minimization(IRWSR) super resolution methods [20].

4. Experimental Results

All the experiments were conducted on a PC equipped with Intel i7-6700K at 4.00GHz with 32GB of memory. To training the deep learning models, Nvidia GTX 1080TI graphic card used with PyTorch framework on python. For fair comparison, we used original code and models provided by the state-of-the-art methods. During testing stage, all the experiments were conducted only on CPU using MATLAB 2016b and PyTorch.

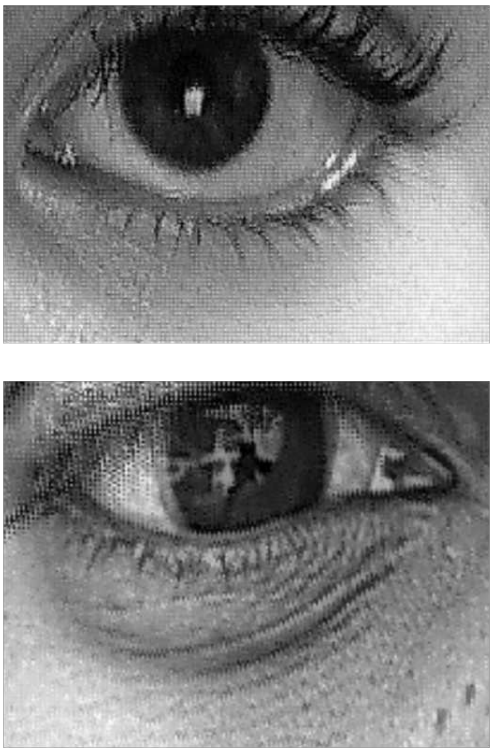

Figure 8. Checkered patterns appearing on IRWSR super resolved images.

All the images from our biometrics dataset were enhanced with the proposed MFSR and also the other selected SR methods for comparison. For each lighting condition (normal office, daylight and dim office) we evaluated the EER(%) using the matcher introduced in section 3.2. Also, we evaluated the execution time for each SR method. All the results are reported in Table 2. Super resolution reconstructions of the tested method are shown in Figure 10.

From the matching results in Table 2, it can be seen that proposed method is the second best performing compared to all the MFSR and SFSR techniques tested. Compared to IRWSR with 17.45% EER in daylight condition, the proposed method EER is decreased by 2.1% while only taking 1.32sec execution time compared to IRWSR at 11.5sec, a 8.5× speed up. MAP is the closest to the proposed method in both speed and matching performance with only 0.66%

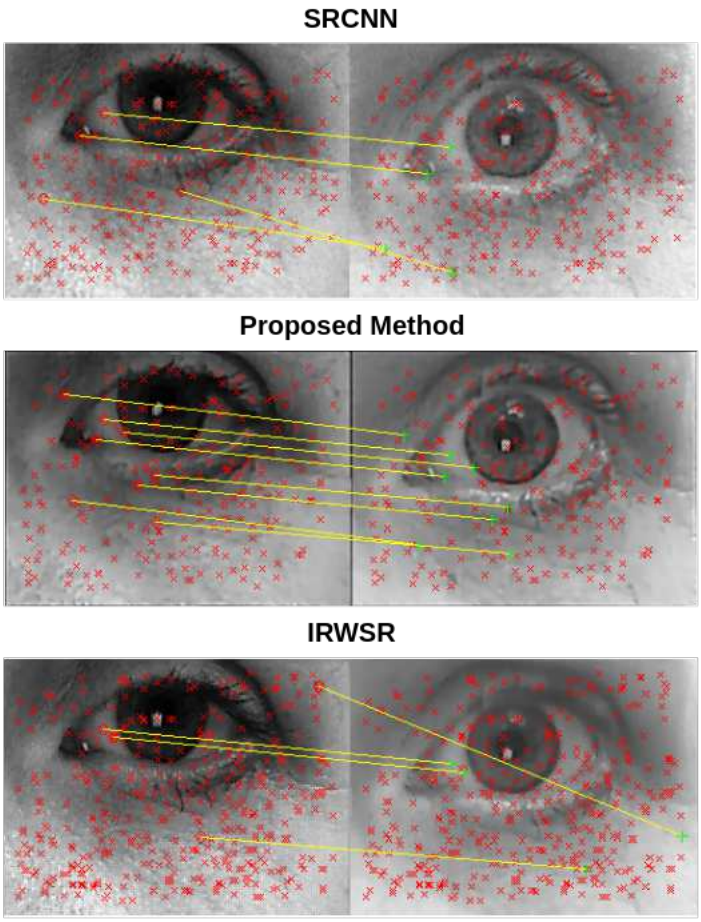

Figure 9. A comparison showing number of feature points detected and number of feature points matched after RANSAC. Both IRWSR and SRCNN faltered during matching due to presence of noise and checkered micro patterns on the super resolved images.

increase in EER for dim light condition. The SFSR method, SRCNN, is competitively faster than proposed method in execution time. However, both SRCNN and VDSR lags in matching performance by at least 2%. It can also be noticed that the over-all results of office lighting is lower than other lightings is because, for many samples dlib [17] algorithm eye region correctly as shown in Figure 7.

From visual inspection of the super resolved images in Figure 10, it can be seen that all the single frame methods, bicubic, SRCNN and VDSR, have more noise compared to the multi-frame methods. Both MAP and bicubic with averaging have issues when registering images, which is causing ghosting effect in the output images as shown in Figure 10(b, d). Another thing to be noticed is that some samples generated using IRWSR based MFSR method have line patterns appearing on the images as shown in Figure 8. It might be that these patterns are helping SURF to detect more points. However, as shown in the Figure 9, the final

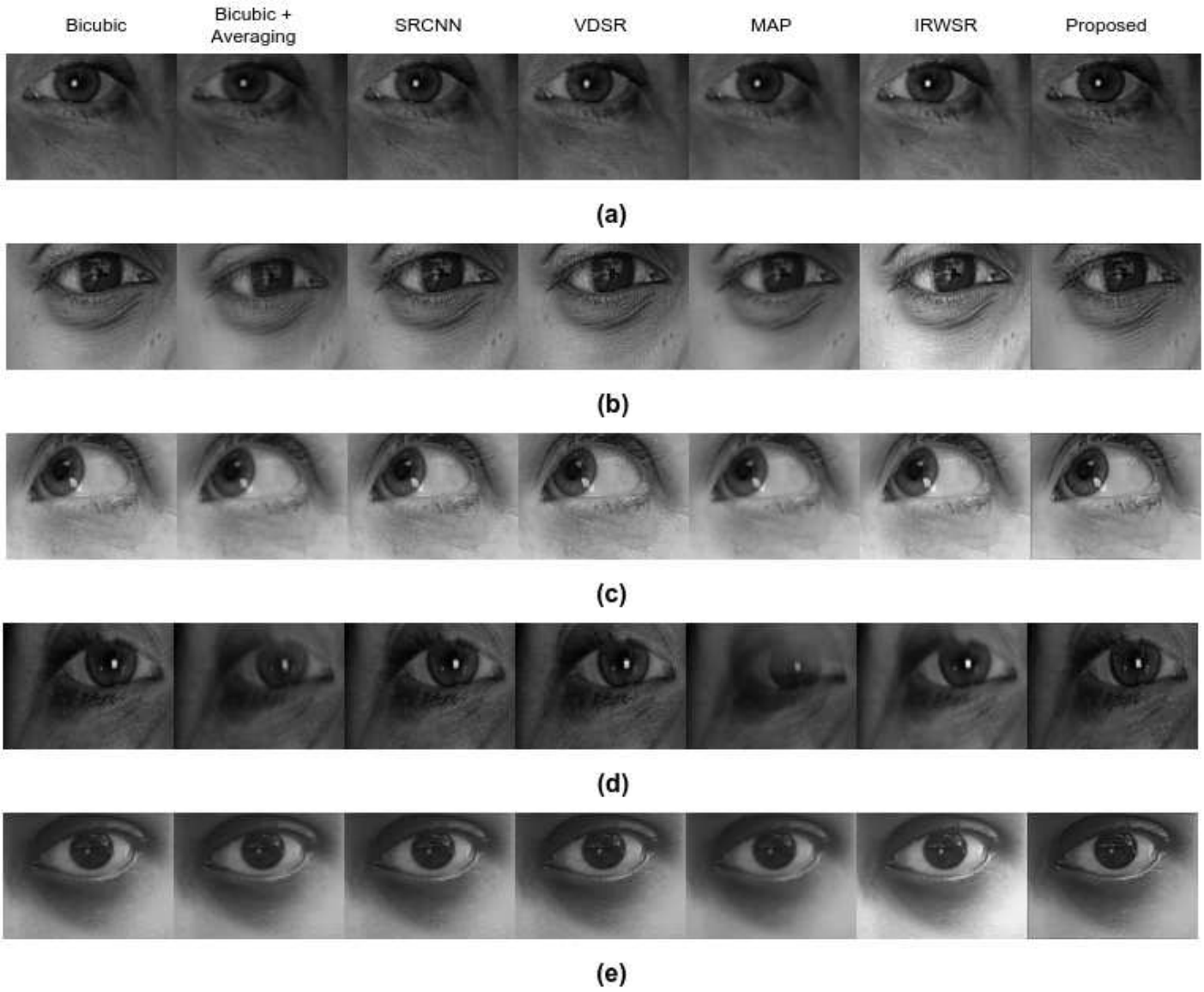

Figure 10. Super resolved images using various super resolution methods.

Methods	Dim light (EER)	Daylight (EER)	Office (EER)	Time(sec.)
Single Image Super Resolution (SISR) Methods				
Bicubic	32.08%	29.39%	38.83%	0.0025
VDSR [16]	22.13%	21.16%	31.38%	1.56
SRCNN [6]	22.46%	21.59%	31.22%	0.33
Multi-Frame Super Resolution (MFSR) Methods				
Bicubic + Averaging	26.32%	23.45%	32.02%	0.0068
MAP [7]	21.61%	20.20%	28.99%	1.28
IRWSR [20]	17.84%	17.45%	26.03%	11.5
Proposed	20.79%	19.54%	28.06%	1.34

Table 2. Comparison of EER(%) and execution time(sec.) for evaluated super resolution methods. All the results are generated for 2× upscaling.

matched points for IRWSR and noisy SISR methods are not robust compared to the proposed method.

5. Conclusion

In this paper, we explored the application of super resolution to mobile ocular biometrics. We proposed a two stage multi-frame super resolution pipeline. First, all the input frames are subpixel-registered using block motion estimation and affine transformation in DCTIF interpolated region. Then bilateral filter is used to fill up the missing pixel in upscaled region after combining all the frames. Finally, a deep learning based image sharpening and deblurring technique is used to increase the contrast and detail of the fine textures of ocular images. We show that the proposed method achieves EERs that are similar or better than the state-of-the-art super resolution methods while being computationally more efficient.

6. Acknowledgment

This work was made possible in part by a grant from ZOLOZ (www.zoloz.com).

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006. 5

[2] W. Bian, S. Ding, and Y. Xue. Fingerprint image super resolution using sparse representation with ridge pattern prior by classification coupled dictionaries. *IET Biometrics*, 6(5):342–350, 2017. 2

[3] E. Bilgazyev, B. Efraty, S. K. Shah, and I. A. Kakadiaris. Improved face recognition using super-resolution. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–7, Oct 2011. 2

[4] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014. 4

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. 4

[6] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. Springer, 2014. 4, 5, 7

[7] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE transactions on image processing*, 6(12):1646–1658, 1997. 5, 7

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Readings in computer vision*, pages 726–740. Elsevier, 1987. 5

[9] A. Gadde, S. K. Narang, and A. Ortega. Bilateral filter: Graph spectral interpretation and extensions. In *2013 IEEE International Conference on Image Processing*, pages 1222–1226, Sept 2013. 3

[10] M. Gastpar, P. L. Dragotti, and M. Vetterli. The distributed karhunen ndash;lo egrave;ve transform. *IEEE Transactions on Information Theory*, 52(12):5177–5196, Dec 2006. 2

[11] V. Gottemukkula, S. Saripalle, S. P. Tankasala, and R. Derakhshani. Method for using visible ocular vasculature for mobile biometrics. *IET Biometrics*, 5(1):3–12, 2016. 4, 5

[12] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, 12(5):597–606, May 2003. 2

[13] K. Hayat. Super-resolution via deep learning. *arXiv preprint arXiv:1706.09077*, 2017. 2

[14] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3

[15] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2(2):109–122, 2016. 2

[16] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 4, 5, 7

[17] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(Jul):1755–1758, 2009. 5, 6

[18] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-normalizing neural networks. In *Advances in Neural Information Processing Systems*, pages 972–981, 2017. 4

[19] R. K. Kuldeep Singh, Anubhav Gupta. Fingerprint image super-resolution via ridge orientation-based clustered coupled sparse dictionaries. *Journal of Electronic Imaging*, 24:24 – 24 – 10, 2015. 2

[20] T. Khler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger. Robust multiframe super-resolution employing iteratively re-weighted minimization. *IEEE Transactions on Computational Imaging*, 2(1):42–58, March 2016. 5, 7

[21] H. Lin, P. Du, W. Zhao, L. Zhang, and H. Sun. Image registration based on corner detection and affine transformation. In *2010 3rd International Congress on Image and Signal Processing*, volume 5, pages 2184–2188, Oct 2010. 3

[22] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 5

[23] H. Lv, R. Wang, X. Xie, H. Jia, and W. Gao. A comparison of fractional-pel interpolation filters in hevc and h.264/avc. In *2012 Visual Communications and Image Processing*, pages 1–6, Nov 2012. 2

[24] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, Oct 2005. 5

[25] P. Muralidhar and C. B. R. Rao. Analysis of block matching motion estimation algorithms. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pages 1–4, July 2013. 3

[26] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman. Feature-domain super-resolution for iris recognition. *Computer Vision and Image Understanding*, 117(10):1526–1535, 2013. 2

[27] U. Park, R. R. Jillela, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, March 2011. 1, 4

[28] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula. Icip 2016 competition on mobile ocular biometric recognition. In *IEEE International Conference on Image Processing (ICIP) 2016, Challenge Session on Mobile Ocular Biometric Recognition*, 2016. 4

[29] H. Ren, Y. He, J. Pan, and L. Li. Super resolution reconstruction and recognition for iris image sequence. In *Chinese Conference on Biometric Recognition*, pages 193–201. Springer, 2012. 2

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4

[31] F. W. Wheeler, X. Liu, and P. H. Tu. Multi-frame super-resolution for face recognition. In *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, Sept 2007. 2

[32] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017. 4