Journal of Imaging Science and Technology® 61(4): 040409-1–040409-9, 2017.

RGB–NIR Image Enhancement by Fusing Bilateral and Weighted Least Squares Filters

Vivek Sharma
CV:HCI, Karlsruher Institut für Technologie (KIT), Germany
ESAT-PSI, Katholieke Universiteit Leuven (KU Leuven), Belgium
E-mail: vivek.sharma@kit.edu

Jon Yngve Hardeberg▲ **and Sony George**
Norwegian University of Science & Technology (NTNU), Norway

Abstract. *Image enhancement using visible (RGB) and near-infrared (NIR) image data has been shown to enhance useful details of the image. While the enhanced images are commonly evaluated by observers' perception, in the present work, we rather evaluate it by quantitative feature evaluation. The proposed algorithm presents a new method to enhance the visible images using NIR information via edge-preserving filters, and also investigates which method performs best from an image features standpoint. In this work, we combine two edge-preserving filters: bilateral filter (BF) and weighted least squares optimization framework (WLS). To fuse the RGB and NIR images, we obtain the base and detail images for both filters. The NIR-detail images for both filters are simply fused by taking an average/maximum of both, which is then combined with the RGB-base image from the WLS filter to reconstruct the final enhanced RGB-NIR image. We then show that our proposed enhancement method produces more stable features than the existing state-of-the-art methods on RGB-NIR Scene Dataset. For feature matching, we use the SIFT features. As a use case, the proposed fusion method is tested on two challenging biometric verifications tasks using CMU hyperspectral face and CASIA multispectral palmprint databases. Our exhaustive experiments show that the proposed fusion method performs equally well in comparison to the existing biometric fusion methods.* © *2017 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2017.61.4.040409]

INTRODUCTION

Image enhancement or filtering using visible (RGB) and near-infrared (NIR) images has been used for several applications, such as aerial or landscape photography,[1] dehazing,[2] tone mapping,[3] biometrics,[4–6] image segmentation,[7] material classification[8] and more. Visible images enhancement using the near-infrared part of the electromagnetic spectrum enhances the contrast, details, and produces more vivid colors. Traditional approaches to enhancement have been tuned on the perception of quality from the perspective of human vision. However, with the advent of growing applications in image processing and computer vision, it is important to characterize the effect of filtering techniques on the machine vision system. As a consequence, it is important to understand the effects that these filters have on the image quality that can affect the image structures and cause performance reduction in computer vision applications.

Over the last two decades, several image enhancement approaches have been proposed. For combining NIR information with RGB images, the NIR channel is combined as either color, luminance or frequency counterpart. This combination is achieved using linear (Laplacian pyramid) or non-linear (anisotropic diffusion, robust smoothing, weighted least squares, and bilateral filtering) filters. Each of the filtering technique comes at some expense, such as high computational time, more artifacts, high noise level, inability to preserve edges and shape details, and more. All these shortcomings add up to undesired loss of image features; however, visually these images may appear very pleasant.

Motivated by the above observation, we propose to combine these filters in a meaningful way, such that a minimum loss of visual pleasantness or information content is attained. In this work, we combine two edge-preserving filters: bilateral filter (BF)[9,10] and weighted least squares optimization framework (WLS)[3] and show that the combination is much more interesting from image features standpoint. We propose to combine the base and detail layers (i.e., low- and high-frequency decompositions) for a pair of visible and NIR images with the BF and WLS filters. As we combine BF and WLS filters, the combination is denoted as BFWLS. We show the performance of method in Figure 1.

We evaluate the quality of BFWLS image features, and compare this with other fusion methods on RGB–NIR Scene Dataset.[11] In addition, we demonstrate the performance of the proposed fusion approach for face and palm-print verification tasks using CMU hyperspectral face[12] and CASIA multispectral palm-print[5] databases. Results show that the BFWLS based filtered images are reliable for biometric verification tasks.

The rest of the paper is organized as follows. First, we discuss the related work, and then we describe our proposed method. Following this, experimental results and analysis are given, and finally, the conclusions are drawn.

▲ IS&T Member.

Received Mar. 1, 2017; accepted for publication June 21, 2017; published online July 25, 2017. Associate Editor: Dileepan Joseph.
1062-3701/2017/61(4)/040409/9/$25.00

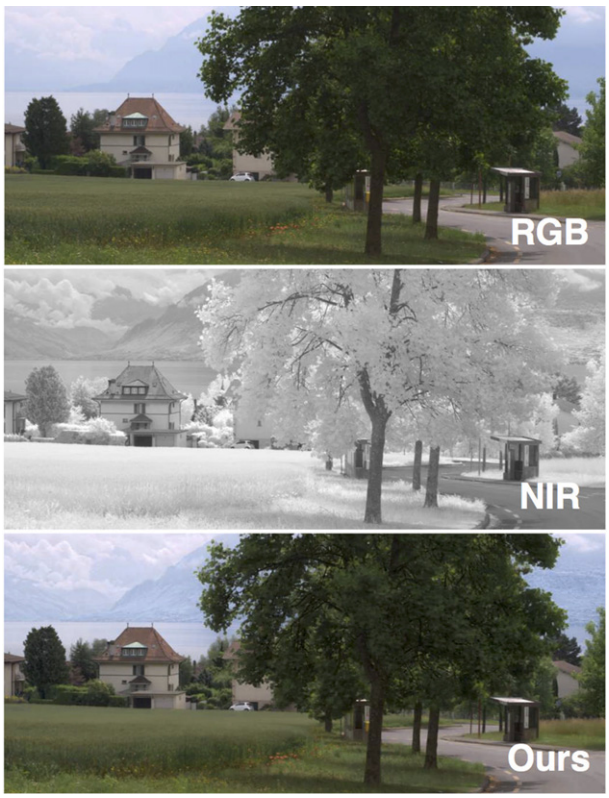


Figure 1. Example of scene enhancement by fusing visible (RGB) and near-infrared (NIR) images. Note that the enhanced image for the proposed method BFWLS-Avg (Ours) has enhanced image structures and details when compared to the visible image. Best viewed in color.

BACKGROUND

Edge-preserving filtering is a technique to smooth an image, while preserving edges i.e. reducing the edge blurring effects across the edge like halos, phantom etc. The class of edge-preserving smoothing filters includes: Anisotropic Diffusion,[13] Laplacian pyramid decomposition,[14] the Weighted Least Squares framework,[3] Bilateral Filter,[9] the Edge-Avoiding Wavelets,[15] Geodesic editing,[16] Guided filtering,[17] and the Domain Transform framework.[18] These filters are very useful in reducing the noise in an image making it very demanding in computer vision and computational photography applications, such as automatic skin enhancement,[19] image deconvolution,[20] multiple illuminant and shadow detection,[21] realistic skin smoothing,[22] flash/no-flash denoising,[23] image upsampling,[24] transfer illumination from reference image to target image,[25] multimodal medical image fusion from MRI-CT, MRI-PRT and MRI-SPECT,[26] image restoration,[27] and so on. The literature on edge-preserving filtering is vast and we only summarize the BF and WLS techniques in this paper.

The bilateral filter (BF)[9] is a non-linear edge-preserving filter. The intensity value at each pixel in an image is a weighted mean of its neighboring pixels. Formally, we have:

$$g_p^{\text{Filtered}} = \frac{1}{W_p}\sum_{q\in S} G_{\sigma_s}(||p-q||)G_{\sigma_r}(||g_p-g_q||)g_q$$
$$W_p = \sum_{q} G_{\sigma_s}(||p-q||)G_{\sigma_r}(||g_p-g_q||) \quad (1)$$

where g and g^{Filtered} are the original input image and filtered image, respectively; p and q indicate spatial locations of pixels. G_{σ_s} and G_{σ_r} are kernel functions based on Gaussian distribution, where σ_s is the spatial kernel that controls the spatial weights and σ_r is the range kernel that controls the sensitivity of edges thus avoiding halo artifacts.

The weighted least squares optimization framework (WLS)[3] is a non-linear, edge-preserving smoothing approach to capture details at a variety of scales via multi-scale edge-preserving decompositions. The approach finds an approximate image g^{Filtered} that is as close as possible to the input image g, and, at the same time, is as smooth as possible along significant gradients in g. Formally, we have:

$$g^{\text{Filtered}} = F_\lambda(g) = (I + L_g)^{-1}g \quad (2)$$

where $L_g = D_x^T A_x D_x + D_y^T A_y D_y$ with D_x and D_y are discrete differentiation operators. A_x and A_y contain the smoothness weights, the smoothness requirement is enforced in a spatially varying manner which depend on g. λ is the balance factor between the data term and the smoothness term. λ controls the level of smoothing, increasing the value of λ results in progressively smoother images.

The BF could preserve well the structure of the information content in the image but may lose much shading distribution. On the other hand, WLS filter could preserve shading distribution of reference information well but may lose the edge structure of the information content in the image. In this way, the BF and WLS filters compliment each other.

Image Fusion: In this section, we review the previous work on image fusion using visible and near-infrared images, where they make use of BF and WLS filters to combine the luminance NIR image to complement the visible image for enhancing the visible images.[1] We divide the related literature into two categories that use: (i) bilateral filter, and (ii) weighted least squares filter.

– **Bilateral filter:** Fredembach et al. (Fre-BF)[22] make use of the BF proposed by Tomasi et al.[9] to decompose the RGB–NIR images into base (low-frequency component) and detail (high-frequency component) layers, and then swapped the detail layer of NIR with the ones of the RGB image for realistic skin smoothing. Similar to Ref. 22, instead of BF we employed WLS to in the same setup in order to check the performance of the WLS method, denoted as Viv-WLS. Bennett et al.[28] enhanced underexposed visible video footage by fusing it with simultaneously captured NIR video footage for noise removal by applying the dual bilateral filter.

Some other notable work where they use the same idea of image fusion but exploit the RGB channels only, such

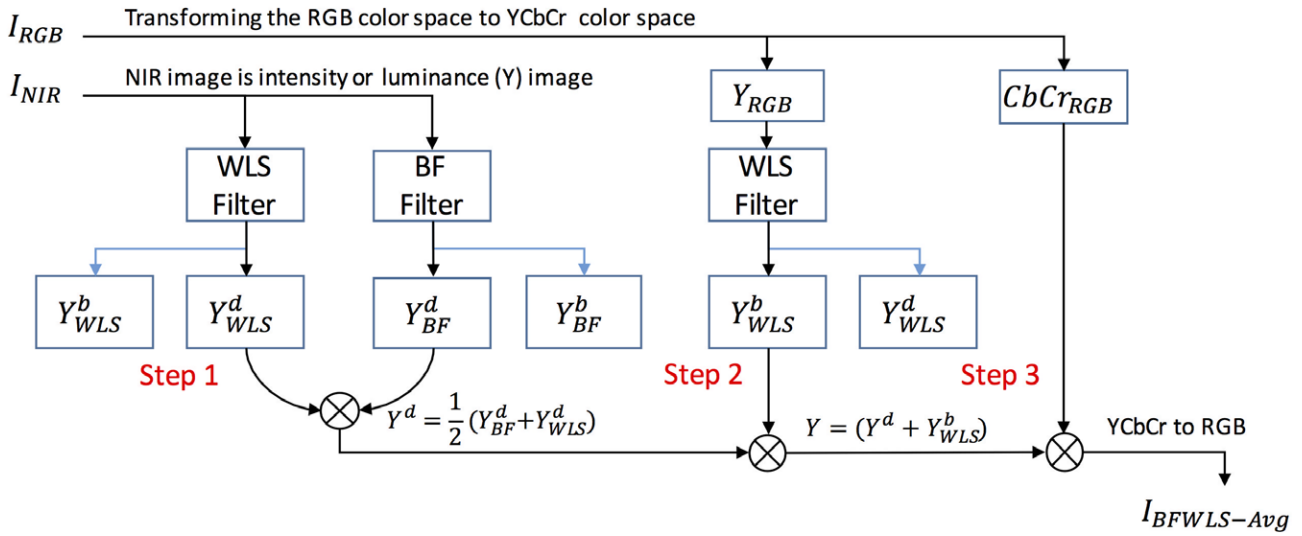


Figure 2. Overview of the full pipeline of our approach, BFWLS-Avg: For an input pair of images (I_{RGB}, I_{NIR}), the intermediate base (b) and detail (d) layers are obtained for both images using BF and WLS filters. In Steps 1–2 the new fused luminance images Y^d are obtained by simply taking mean of NIR: Y^d_{BF} and Y^d_{WLS} images, and which is then combined with base layer Y^b_{WLS} of RGB image resulting to enhanced luminance RGB–NIR image Y. And it is then combined with the chrominance of RGB image (Step 3), to construct the enhanced RGB–NIR new image.

as "Fast Bilateral Filtering" by Durand et al.[10] They reduce the contrast of the high-dynamic range images to display it on low-dynamic range media using BF. A major shortcoming of the seminal bilateral filter[9] decomposition is its speed, thus all the above mentioned papers use the fast bilateral filter[10] with no significant decrease in image quality.

– **Weighted least squares filter:** Schaul et al. (Sch-WLS)[2] improve the contrast of the haze-degraded color images by transforming the visible and NIR images via their multi-resolution decomposition algorithm using WLS filter. The authors fused the detail image of RGB and NIR pairs at each level of their multi-resolution decomposition. In comparison, in our method, we use the detail images of BF and WLS for fusion. Zhuo et al.[29] use an NIR image to enhance its corresponding noisy visible image using dual WLS smoothing filter.

Finally, it is worth noting the seminal work by Farbman et al.[3] who proposed the "Weighted Least Squares Filter" by combining detail layer of RGB channels at various scales using WLS multi-scale image decompositions for tone mapping and detail enhancement. We swapped the luminance channel of RGB image by the luminance channel of NIR image in the same setup in order to check the performance of[3] for RGB–NIR image fusion. We denote this method by Far-WLS.

To the best of our knowledge, our work is the first to combine the BF and WLS filters. We compare against various image fusion approaches in our experimental section.

PROPOSED ENHANCEMENT APPROACH

Figure 2 illustrates the entire procedure of our proposed approach. We denote the combined BF and WLS filtering algorithm as BFWLS. To fuse the visible and NIR images, we first transform the RGB color space into a luminance–chrominance color space, where the one channel NIR image contains intensity data or luminance only.[1] The chrominance is not used in the fusion algorithm, but simply re-combined in the final fused image.

Given an input image, the BF and WLS filters decompose an image into base and detail images. The detail images are obtained by simply subtracting the base image from the original image. The base image comprises low-frequency content with general appearance of the image over smooth areas, while the detail layer comprises of high-frequency contents with edges and sharp transition (e.g., noise).

In the first step, we apply BF-based and WLS-based decompositions of the NIR image for the extraction of base and detail images. We retain the average values of the detail-WLS and detail-BF images at each pixel. This fusion criterion is denoted as BFWLS-Avg.

The fusion criterion is based on the following observations: the BF filter preserves edges and can extract details at a fine spatial scale, but lacks the ability to extract details at arbitrary scales. Where as, WLS filter is very good at preserving fine and coarse details at arbitrary scales. Taking an average between two allows to retain the details from both and moderately boosts the details. We also found that the hidden details appeared in this way.

As a fusion criterion, we tried to retain the maximum values between the two to preserve the structure of the important information content from both, but may lose much shading distribution. We denote this fusion criterion as BFWLS-Max.

The base layer of RGB image contains low luminance information as perceived by humans visual system, thus the NIR base layer is discarded. In the second step (Step 2), we combine the fused detail layer of NIR image with the base layer of RGB image obtained using WLS, to obtain the new luminance image. This new luminance image enhances the original image's contrast and details, and it is then combined with chrominance of RGB image to reconstruct the final image in the third step (Step 3).

EXPERIMENTAL RESULTS & ANALYSIS

We perform two sets of experiments: First, we evaluate the image feature quality of BFWLS in comparison to other fusion methods via feature matching. Second, we test BFWLS on two challenging biometric face and palmprint verification tasks and compare it with other baseline biometric fusion methods. For the experimental evaluation, we used a desktop with Intel i7-2600K CPU at 3.40GHZ, 8GB RAM. All experiments were performed using the publicly available Vlfeat library.[30]

Evaluating the Quality of BFWLS

We evaluate our proposed method on 477 pairs of images from RGB–NIR Scene Dataset.[11] We evaluate the features quality by analyzing the amount of original features that remain after applying the transformation to an image. A more detailed evaluation to quantitatively assess the quality of features in terms of robustness and discriminability can be found in Refs. 31, 32. We repeat the procedure of feature assessment suggested in Ref. 31 and apply synthetic rotation transformation with (45°, 90° and 180°) to each image. Finally the feature matching is done between the original and transformed image pairs with *threshold* = 1.5 and produces a number of matches. We compare the matches obtained from the fusion algorithm against the matches extracted from RGB images, and report the relative change (in %) as an evaluation criterion. For removing the outliers, we apply RANSAC[33] with homography. By discarding the outliers, we focus on the features that are more likely to provide true matches in feature matching, and obtain a reduced set of candidate correspondences with fewer accurate data points. For feature matching, we use the SIFT implementation from the Vlfeat library.[30] For the SIFT descriptors, we use a bin size of 8 and step size of 4.

In order to evaluate the performance of our proposed method, we compare it with other fusion methods. As the codes for the fusion methods[2,22] were not available to the authors, we implemented all these techniques using fast BF filter from Durand et al.[10] and WLS filter from Farbman et al.[3] for image fusion. For fair comparison, we compare all the methods under the same evaluation protocol discussed above with the same parameters. The default parameters for all the methods are shown in Table I. For comprehensive discussion, we refer the readers to Refs. 3, 10. The source codes for fast BF,[10] WLS[3] are publicly available.

In addition, we also report the mean squared error (MSE), peak signal-to-noise ratio (PSNR), and time to fuse an RGB–NIR image pair as an evaluation criterion for each method. The PSNR and MSE are computed between the fused and the original RGB image. The metrics shown in Table II are mean values computed across 477 image pairs.

Table I. Parameters of BF and WLS filters for RGB–NIR image fusion. For fast BF,[10] the parameters $edge_{min}$, $edge_{max}$, σ_s and σ_r of the method are adapted to each image, thus requires no parameter setting. n is the total number of levels for decomposition.

Method	Parameters
Fre-BF[22]	fast BP:[10] parameters adapted for each image.
Sch-WLS[2]	$\lambda = 0.1$, $c = 2$, and $n = 6$
Far-WLS[3]	$\lambda_1 = 0.125$, $\lambda_2 = 0.50$, $c = 1.2$, and $n = 1$
BFWLS	fast BF:[10] parameters adapted for each image, and WLS: $\lambda = 0.125$, $c = 1.2$, and $n = 1$
Viv-WLS	$\lambda = 0.125$, $c = 1.2$, and $n = 1$
DWT[4]	haar mother wavelet, and $n = 9$
CVT[5]	#levels in the wavelet pyramid: 4 #scales in the local ridgelets: {3,4,4,5}

We can see in Table II that our method has more stable feature matches over state-of-the-art methods. BFWLS-Avg performs the best among all methods. BFWLS-Avg obtains improved feature matches over RGB images by 8.78% (Rel. Change: SIFT) and 6.27% (Rel. Change: SIFT+RANSAC), respectively. The performance gap of BFWLS-Avg is 5.49/5.45% better when compared to BFWLS-Max using Rel. Change: SIFT/SIFT+RANSAC criterion. BFWLS-Avg is 4.6%, 9.67%, and 3.75% better than the Fre-BF.,[22] Sch-WLS[2] and Far-WLS[3] using Rel. Change: SIFT+RANSAC criterion. Also, BFWLS-Avg is 4.6%, and 4.83% better than the Fre-BF.,[22] Viv-WLS using Rel. Change: SIFT+RANSAC criterion, which clearly shows that simply not averaging the detail layers i.e., using either the BF (Fre-BF) or WLS (Viv-WLS) detail layer alone gave worse results than with the fusing of the two methods. Note that Sch-WLS[2] feature matches degrade by 1.11/3.49% against RGB images after the image fusion. Our enhanced images have better high-frequency details, and further improve the ability to preserve edges because our approach can extract details at fine spatial at arbitrary scales due to combined fusion from BF and WLS filters, shown in Figure 4. As an additional example, in Figure 3 we illustrate the impact of feature matching using enhanced image versus non-enhanced image. The source code for all the fusion methods will be made publicly available.

Application to Biometric Verification Tasks

We test our proposed fusion approach for face and palmprint verification tasks using CMU hyperspectral face[12] and CASIA multispectral palm-print[5] databases in biometric settings. In order to evaluate the performance of our proposed method, we compare it with traditional biometric fusion methods and other baseline methods. For this evaluation, we generate an RGB image and panchromatic-NIR image for both datasets using standard image conversion of

Table II. Comparison of BFWLS with the other methods. The metrics shown are mean values computed across 477 image pairs. The images are of resolution 1024 × 680 pixels.

Metric (Mean)	Fre-BF[22]	Sch-WLS[2]	Far-WLS[3]	BFWLS-Max (ours)	BFWLS-Avg (ours)	Viv-WLS ours
Rel. Change: SIFT (%)	4.31	−1.11	4.92	3.29	**8.78**	3.83
Rel. Change: SIFT+Ransac (%)	1.67	−3.49	2.52	0.82	**6.27**	1.44
Time (Sec)	**0.20**	38.20	6.65	6.39	6.59	6.54
PSNR	32.36	22.83	13.28	30.28	32.69	**33.45**
MSE (10^{-4})	6.76	61	637	10	6.45	**6.42**

Table III. Details of the training and test set for CMU-HSFD[12] and CASIA[5] datasets.

	CMU-HSFD[12]	CASIA[5]
Number of Subjects	29	32
Images/Subject	12–20	12
Training Images/Subject	8	6
Testing Images/Subject	4–12	6
Resolution (cropped)	132 × 132	172 × 172
Genuine Pairs	828	576
Impostor Pairs	11600	5760

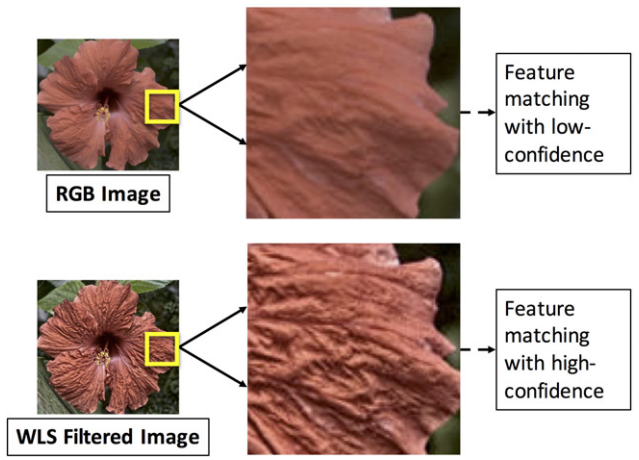


Figure 3. Example of a flower image taken from Farbman et al.[3] Note the higher contrast and sharpness in the WLS filtered image leads to feature matching with high confidence, in comparison to that of the RGB image. Best viewed in color.

hyperspectral (HSI) to RGB and panchromatic-NIR images (For transforming the HSI to RGB color space, we use (a) CIE 2006 tristimulus color matching functions, (b) CIE standard daylight illuminant (D65), (c) Silicon sensitivity of Hamamatsu camera, (d) RGB–NIR filters for modulating wavelengths in 400–1000 nm.). In this regard, we refer the reader to the book by Ohta and Robertson[34] for detailed steps.

As an evaluation criterion, in addition to True Positive Rate (TPR or Recall), we report the PSNR, MSE and processing fusion time for each fusion method. We now explain the experimental details: dataset, implementation details, reference/testing protocol, and baselines.

Experimental Details

Datasets

The CMU-HSFD face and CASIA palmprint datasets are publicly available datasets used in our experiments. Detailed specification of both databases is given in Table III.

Carnegie Mellon University Hyperspectral Face Dataset (CMU-HSFD)[12] (see Table III) is acquired using the CMU developed AOTF with three halogen light sources. Each hyperspectral cube contains 65 bands (450–1090 nm, with step size of 10 nm). The database contains data of 54 subjects acquired in 1–5 different sessions over a period of about two months. For each individual, frontal, left and right views with neutral expression were acquired. The database contains 4–20 cubes per subject over all sessions. We use 29 (i.e. $M = 29$) subjects in our experiments, for whom we have at least three sessions, and 12–20 cubes. The dataset suffers from shot noise. We apply a median filter of size 3 × 3 to remove the shot noise. We transform the hyperspectral cubes into an RGB and panchromatic-NIR image representations.[34]

CASIA Multi-Spectral Palmprint Image Database V1.0 (CASIA)[5] (see Table III) is acquired using five narrow band illuminator ({460,630,700,850,940}nm) and a white light. The database contains data of 100 subjects acquired in two sessions with the time interval of over one month. In each session, three samples were captured. Each sample contains six palm images for six illuminators respectively captured at the same time, leading to a total of 7,200 palmprint images for 100 different subjects over all sessions. The

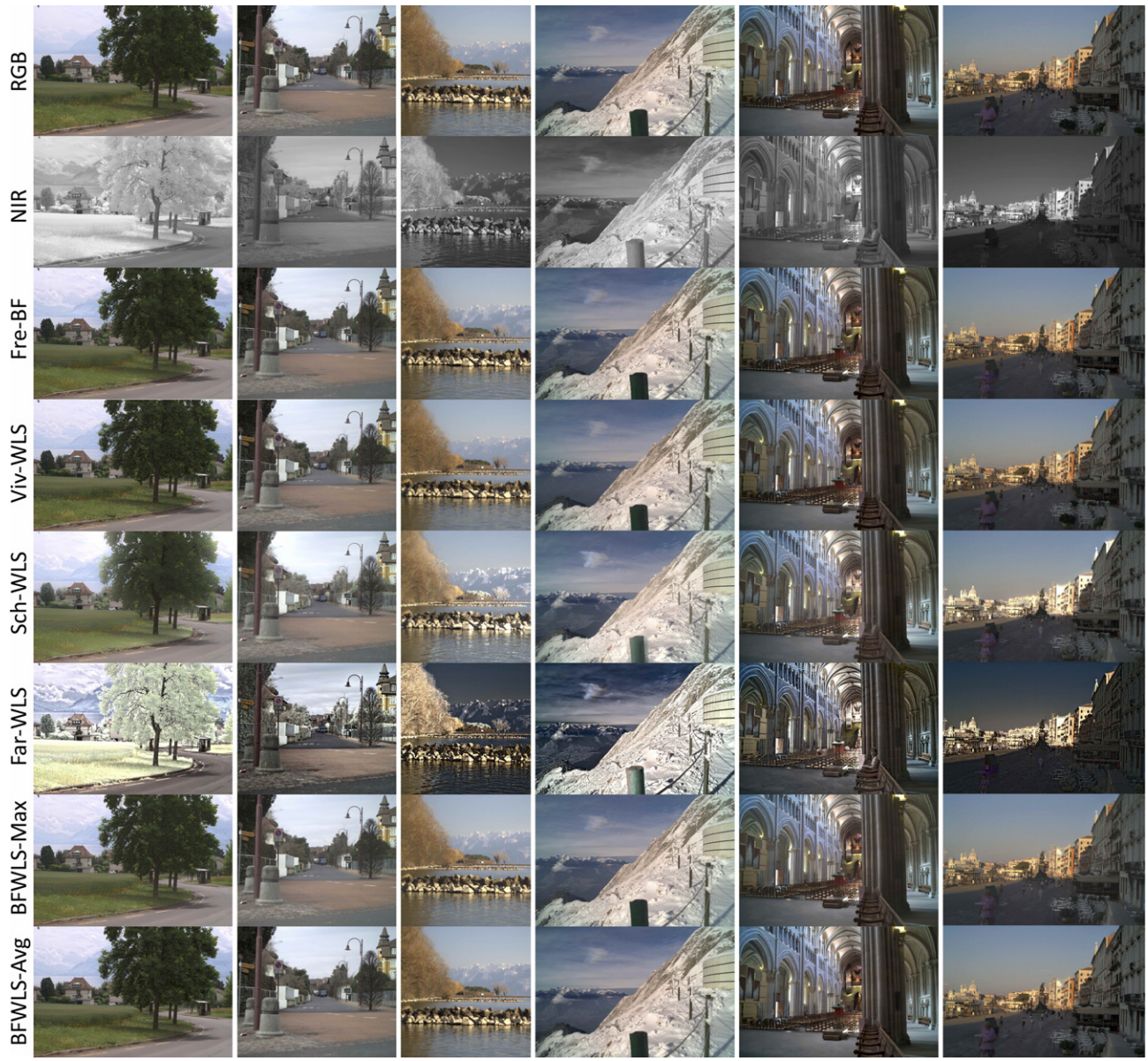


Figure 4. Visual comparison of BFWLS against state-of-the-art algorithms for RGB–NIR image fusion. In the zoomed-in view, note how BFWLS-Avg adapts well to the local structures and details when compared to the other methods. Best viewed in color.

ROIs were extracted according to the technique proposed in Ref. 35. Only 32 subjects palmprint were detected accurately using the extraction technique.[35] Therefore, in our experiments only 32 (i.e., $M = 32$) subjects were considered for the experimental evaluation. We use {460,630,700}nm as an RGB image and {850,940}nm were transformed into panchromatic-NIR image (by taking a simple mean of two images), while the image acquired with a white light is discarded.

Protocols

For face verification using CMU-HSFD, the reference samples of each subject are taken from the first session, while test samples are taken from all other sessions. For palmprint verification using CASIA, the reference samples are taken from the first session, while the test samples are taken from the second session.

Implementation Details

To extract the SIFT features, we use a bin size of 4, step size of 8, then the extracted SIFT features are Fisher encoded. To compute Fisher encoding, we build a visual dictionary using GMM with 100 clusters for CMU-HSFD and also 100 clusters for CASIA. We normalize the features using ℓ_2-*norm*. We denote dense SIFT Fisher vectors by DSIFT-FVs. These parameters are fixed for all descriptors. For all fusion methods, we keep these parameters fixed.

Verification System

The verification system is evaluated by computing similarity of the features for genuine and impostor pairs via

Table IV. Comparative performance on the CMU-HSFD[12] dataset.

Metric (Mean)	Fre-BF[22]	Sch-WLS[2]	RGB	Raghu-DWT[4]	Hao-CVT[5]	BFWLS-Avg (ours)	BFWLS-Max (ours)	Viv-WLS (ours)
PSNR	36.04	19.78	—	**47.79**	14.00	38.47	34.37	39.36
MSE (10^{-5})	25	1123	—	**1.86**	4722	14.73	37.27	12.16
Time (Sec)	**0.02**	1.69	—	0.05	26.15	0.29	0.3	0.29
DSIFT-FVs + Euclidean Distance:								
TPR(%)@EER	57.61	60.87	61.22	62.68	60.95	**65.42**	59.78	59.78
DSIFT + Cosine Similarity:								
TPR(%)@EER	**94.57**	92.03	94.2	94.2	93.48	94.2	**94.57**	94.2

Table V. Comparative performance on the CASIA[5] dataset.

Metric (Mean)	Fre-BF[22]	Sch-WLS[2]	RGB	Raghu-DWT[4]	Hao-CVT[5]	BFWLS-Avg (ours)	BFWLS-Max (ours)	Viv-WLS (ours)
PSNR	46.55	37.41	—	**57.55**	26.29	48.49	43.94	43.94
MSE (10^{-5})	2.36	19	—	**0.18**	348	1.73	4.19	4.24
Time (Sec)	**0.02**	0.63	—	0.05	25.35	0.13	0.13	0.1
DSIFT-FVs + Euclidean Distance:								
TPR(%)@EER	69.62	66.61	66.15	**76.01**	67.19	67.6	70.31	71.86
DSIFT + Cosine Similarity:								
TPR(%)@EER	83.33	83.8	82.6	82.69	84.31	83.8	**84.38**	83.59

Euclidean/L2 distance. The Euclidean/L2 distance is given by $||y - \hat{y}||_2$, where y and $\hat{y}$ are the DSIFT-FVs features for genuine/impostor pairs. The performance of the system is measured by calculating the Equal Error Rate (ERR), which is defined as a point when the rate of impostor pairs accepted (FAR) is equal to the rate of genuine pairs rejected (FRR). The lower the EER, the better is the biometric system. We report the true positive rate (TPR or Recall in %) at the EER. The approach is denoted as DSIFT-FVs + Euclidean Distance.

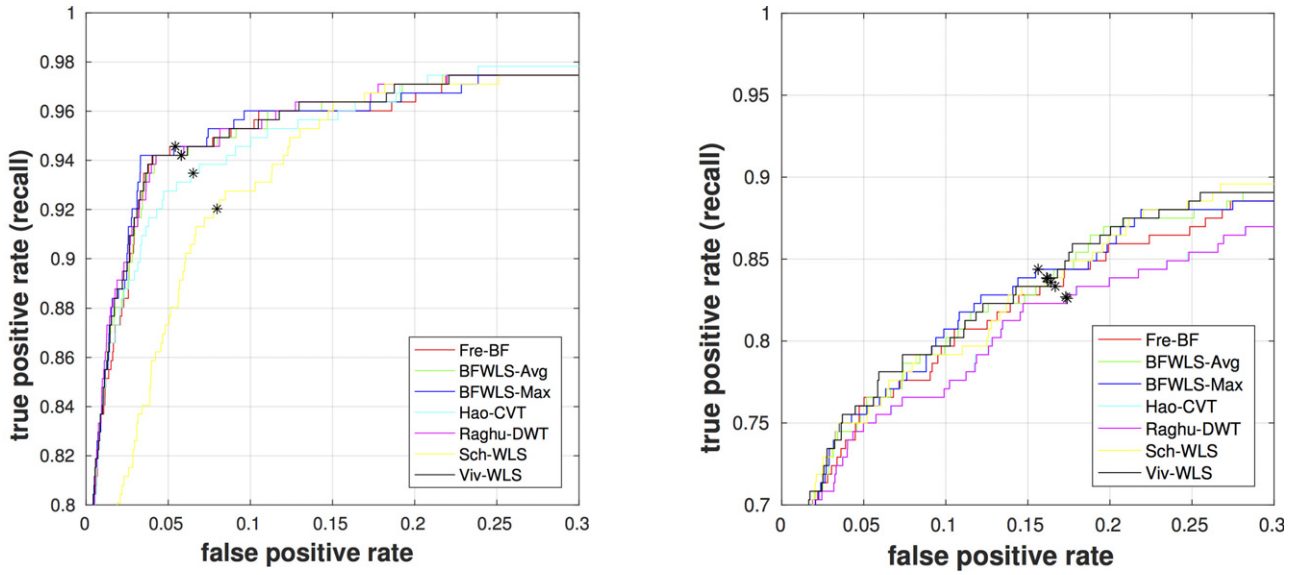
To compute DSIFT+FVs, we need to build a visual dictionary using GMM with a given pre-defined number of clusters. The performance of the system is influenced by the number of clusters, increasing the number of clusters leads to better performance. Optimizing the number of clusters and thus improving the performance is not the goal of this paper. Thus, for a better assessment to evaluate the feature quality we present another technique. Given that we have a pair of images I and $I^{'}$, we extract a dense set of SIFT features: $y \in \mathbb{R}^{D\times K}$ and $\hat{y} \in \mathbb{R}^{D\times K}$, where D denotes feature dimensions and K is the number of features extracted. Then we compute the cosine similarity between the descriptors given as: $\text{score}_k = < y_k, \hat{y}_k > /||y_k||\ ||\hat{y}_k||,\ k \in [1, \ldots, K]$. The final similarity score is computed by taking mean of all scores given as: $Similarity(I, I^{'}) = 1/K \sum_{k=1}^{K} \text{score}_k$. The approach is denoted as DSIFT + Cosine Similarity.

Baseline Fusion Methods

We compare BFWLS with a few baselines: the traditional fusion approaches from biometrics and image processing community, Raghu-DWT (Raghavendra et al.[4]), Hao-CVT (Hao et al.[5]), Fre-BF (Fredembach et al.[22]), Sch-WLS (Schaul et al.[2]), and Viv-WLS. In Raghu-DWT,[4] the images are the first multi-scale decomposed in n levels, then at each nth level weighted fusion rule is applied to combine the decomposed wavelet coefficients. And finally, the fused image is constructed by inverse DWT. Similar is the case with Hao-CVT,[5] where the authors multi-scale decompose the RGB–NIR image pairs, and fuse them at each nth level of wavelet pyramid. For a fair comparison, we use an average weighting as a fusion rule for both Raghu-DWT and Hao-CVT to combine the RGB and NIR pair images. As the codes for the fusion methods[2,4,5,22] were not available by the authors, we implemented all these techniques using fast BF filter from Durand et al.,[10] WLS filter from Farbman et al.,[3] Curvelet Transform (CVT) from Starck et al.,[36] and Discrete Wavelet Transform (DWT) for image fusion. The default parameters for all the methods are shown in Table I. The source code for all the fusion methods will be made publicly available.

Results

In Tables IV–V, we quantitatively evaluate the TPR(%)@EER of our proposed method and compare with tradition



(a) CMU-HSFD (b) CASIA

Figure 5. The ROC curves for the CMU-HSFD and CASIA datasets using DSIFT + Cosine Similarity method for verification tasks.* denotes the EER when the false accept rate is equal to the false reject rate. Best viewed in color.

biometric fusion methods and other baseline fusion methods for face and palmprint verification tasks. For fair comparison, we compare all methods under the same evaluation protocols discussed above.

It is evident from Tables IV–V that BFWLS shows promising results in comparison to other fusion methods on CMU-HSFD and CASIA datasets using DSIFT + Cosine Similarity method. Note that the TPR of BFWLS is better for RGB images on both datasets. Figure 5 shows the ROC curve for CMU-HSFD and CASIA datasets.

In case of DSIFT-FVs + Euclidean Distance method, we can see that for CMU-HSFD our method performs the best, while Raghu-DWT performs the best on CASIA. As previously mentioned, we give a pre-defined number of GMM clusters; hence the performance is influenced by the number of clusters. Thus, DSIFT + Cosine Similarity is a better evaluation criterion for feature quality assessment.

CONCLUSION

In this paper, we present a method to combine visible and near-infrared images using edge-preserving filters: bilateral filter and weighted least square filter. Our method successfully enhances the visible images using near-infrared information, to attain a result image richer in information for computer vision applications. To illustrate the effectiveness of the proposed method, experiments are performed to evaluate the image feature quality on RGB–NIR Scene dataset. Our proposed fusion method is also tested on two challenging biometric face and palmprint verification tasks. The proposed method not only improves the image feature quality for recognition tasks, but also the resulting fused image has more detail information and high image contrast that makes visually these images appear very pleasant.

ACKNOWLEDGMENTS

The authors would like to thank M. Saquib Sarfraz for his valuable suggestions on the experimental evaluation of biometric verification tasks. A partial amount of this work was carried out while Vivek Sharma was at ESAT-PSI, KU Leuven

REFERENCES

[1] C. Fredembach and S. Süsstrunk, "Colouring the near-infrared," *Proc. IS&T/SID CIC16: Sixteenth Color Imaging Conf.* (IS&T, Springfield, VA, 2008), pp. 176–182.

[2] L. Schaul, C. Fredembach, and S. Süsstrunk, "Color image dehazing using the near-infrared," *Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2009), pp. 1629–1632.

[3] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," ACM Trans. Graph. (TOG) **27**, 2205–2221 (2008).

[4] R. Raghavendra and C. Busch, "Novel image fusion scheme based on dependency measure for robust multispectral palmprint recognition," IEEE Pattern Recognit. **47**, 2205–2221 (2014).

[5] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," *Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2008), pp. 281–284.

[6] R. Raghavendra, B. Dorizzi, A. Rao, and G Kumar, "Particle swarm optimization based fusion of near infrared and visible images for improved face verification," IEEE Pattern Recognit. **44**, 401–411 (2011).

[7] N. Salamati and S. Süsstrunk, "Material-based object segmentation using near-infrared information," *Proc. IS&T/SID CIC18: Eighteenth Color Imaging Conf.* (IS&T, Springfield, VA, 2010), pp. 196–201.

[8] N. Salamati, C. Fredembach, and S. Süsstrunk, "Material classification using color and NIR images," *Proc. IS&T/SID CIC17: Seventeenth Color Imaging Conf.* (IS&T, Springfield, VA, 2009), pp. 116–222.

[9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *IEEE Int'l. Conf. on Computer Vision (ICCV)* (IEEE, Piscataway, NJ, 1998), pp. 839–846.

[10] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," ACM Trans. Graph. (TOG) **21**, 257–266 (2002).

[11] M. Brown and S. Süsstrunk, "Multi-spectral sift for scene recognition," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2011), pp. 177–184.
[12] L. J. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," MU Technical Report'02.
[13] K. N. Nordström, "Biased anisotropic diffusion: a unified regularization and diffusion approach to edge detection," Image Vis. Comput. **8**, 318–327 (1990).
[14] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **31**, 532–540 (1983).
[15] R. Fattal, "Edge-avoiding wavelets and their applications," ACM Trans. Graph. (TOG) **28** (2009).
[16] A. Criminisi, T. Sharp, C. Rother, and P. Pérez, "Geodesic image and video editing," ACM Trans. Graph. (TOG) **29** (2010).
[17] K. He, J. Sun, and X. Tang, "Guided image filtering," *European Conf. on Computer Vision (ECCV)* (Springer, Berlin-Heidelberg, 2010), pp. 1–14.
[18] E. Gastal and M. Oliveira, "Domain transform for edge-aware image and video processing," ACM Trans. Graph. (TOG) **30** (2011).
[19] S. Süsstrunk, C. Fredembach, and D. Tamburrino, "Automatic skin enhancement with visible and near-infrared image fusion," *ACM Multimedia Conf.* (ACM, NY, USA, 2010), pp. 1693–1696.
[20] L. Yuan, J. Sun, L. Quan, and H. Shum, "Progressive inter-scale and intra-scale non-blind image deconvolution," ACM Trans. Graph. (TOG) **27** (2008).
[21] C. Fredembach and S. Susstrunk, "Illuminant estimation and detection using near-infrared," Proc. SPIE **7250**, 72500E (2011).
[22] C. Fredembach, N. Barbuscia, and S. Süsstrunk, "Combining visible and near-infrared images for realistic skin smoothing," *Proc. IS&T/SID CIC17: Seventeenth Color Imaging Conf.* (IS&T, Springfield, VA, 2009), pp. 242–247.
[23] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," ACM Trans. Graph. (TOG) **23**, 664–672 (2004).
[24] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," ACM Trans. Graph. (TOG) **26** (2007).
[25] X. Chen, M. Chen, X. Jin, and Q. Zhao, "Face illumination transfer through edge-preserving filters," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2011), pp. 281–287.
[26] W. Li, Z. Zhao, J. Du, and Y. Wang, "Edge-preserve filter image enhancement with application to medical image fusion," J. Med. Imaging Health Inform. **7**, 16–24 (2017).
[27] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia, "Cross-field joint image restoration via scale map," *IEEE Int'l. Conf. on Computer Vision (ICCV)* (IEEE, Piscataway, NJ, 2013), pp. 1537–1544.
[28] E. Bennett, J. Mason, and L. McMillan, "Multispectral video fusion," *ACM SIGGRAPH 2006 Sketches* (ACM, NY, USA, 2006).
[29] S. Zhuo, X. Zhang, X. Miao, and T. Sim, "Enhancing low light images using near infrared flash images," *Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2010), pp. 2537–2540.
[30] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," *ACM Multimedia Conf.* (ACM, NY, USA, 2010), pp. 1469–1472.
[31] H. Su, W. Chuang, W. Lu, and M. Wu, "Evaluating the quality of individual SIFT features," *Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2012), pp. 2377–2380.
[32] D. Firmenichy, M. Brown, and S. Süsstrunk, "Multispectral interest points for RGB–NIR image registration," *Int'l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2011), pp. 181–184.
[33] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM **24** (1981).
[34] N. Ohta and A. Robertson, "Colorimetry: fundamentals and applications," *Chapter 3: CIE Standard Colorimetric System* (John Wiley & Sons, West Sussex, England, 2006).
[35] W. Li, B. Zhang, L. Zhang, and J. Yan, "Principal line-based alignment refinement for palmprint recognition," IEEE Trans. Syst. Man Cybern. **42**, 1491–1499 (2012).
[36] J. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," IEEE Trans. Image Process. **11**, 670–684 (2002).