

Make My Day - High-Fidelity Color Denoising with Near-Infrared

Hiroto Honda

Toshiba Corp., Kawasaki, Japan
ETH Zurich, Switzerland

hiroto.honda@toshiba.co.jp

Radu Timofte

Computer Vision Lab
ETH Zurich, Switzerland

timofter@ee.vision.ethz.ch

Luc Van Gool

KU Leuven, Belgium
ETH Zurich, Switzerland

vangool@ee.vision.ethz.ch

Abstract

We address the task of restoring RGB images taken under low illumination (e.g. night time), when an aligned near infrared (NIR or simply N) image taken under stronger NIR illumination is available. Such restoration holds the promise that algorithms designed to work under daylight conditions could be used around the clock. Increasingly, RGBN cameras are becoming available, as car cameras tend to include a Near-Infrared (N) band, next to R, G, and B bands, and NIR artificial lighting is applied. Under low lighting conditions, the NIR band is less noisy than the others and this is all the more the case if stronger illumination is only available in the NIR band. We address the task of restoring the R, G, and B bands on the basis of the NIR band in such cases. Even if the NIR band is less strongly correlated with the R, G, and B bands than these bands are mutually, there is sufficient such correlation to pick up important textural and gradient information in the NIR band and inject it into the others. The algorithm that we propose - coined ‘Make My Day’ or MMD for short - is akin to the previously published BM3D denoising algorithm. MMD denoises the three (visible - NIR) differential images to then add back the original NIR image. It not only effectively reduces the noise but also includes the texture and edge information in the high spatial frequency range. MMD outperforms other state-of-art denoising methods in terms of PSNR, texture quality, and color fidelity. We publish our codes and images.

1. Introduction

At night time - and without artificial light sources around - the light impinging on an image sensor obviously is faint. This low photon count emphasizes both the shot noise and read-out circuit noise. Hence, if we want to reconstruct an image that would come close to what is obtained under better lighting conditions, it will not suffice to boost contrast, as this would also increase the noise to intolerable levels. The noise has to be strongly reduced for this to work, *i.e.* an appropriate image denoising is required [1, 2, 3]. In this

paper, we focus on the task of trying to get RGB images taken under low-illumination conditions closer to images of the same scene as they would be acquired under stronger illumination.

An obvious way to reduce the noise in the RGB bands for a dark scene is to use strong enough visible illumination, *e.g.* a strobe flash. Apart from being possibly disturbing for people, such illumination causes unnatural color balance effects and/or undesirable reflections. Here we study the alternative of using a near-infrared (NIR) band together with NIR lighting. This is not as far-fetched as it may seem. First, regular silicon photo detectors are also sensitive to most NIR wavelengths, mainly around 900 nm but even up to 1200 nm. Commercial digital cameras use IR-cut filters, and so far only specific cameras like night vision cameras or the Kinect camera make use of this NIR capability. Yet, cameras in cars increasingly extract information from both the visible (RGB) and NIR domains [7]. With the growing popularity of driver assistance or even fully autonomous driving, the use of such cameras is bound to see a rampant increase. Hence, we can expect such RGBN cameras to soon enable further consumer applications. NIR imagery provides a unique perspective when looking at things such as plants, fabric, water, or blood vessels. Many interesting applications have been proposed such as eye tracking [18], dehazing [11], material classification [9] and so on [4, 5, 12, 14, 16, 17].

Make My Day. In this paper we propose a novel image denoising method, that enhances the RGB bands of images taken under weak visible illumination, based on the NIR band acquired under stronger NIR illumination. Thereby this denoising of the RGB bands creates visible images that come a lot closer to those seen under good white illumination. It is like turning a night time image into a day time image, hence the method is coined the ‘Make My Day’ or MMD algorithm.

Under conditions such as simultaneous exposure and sharing the field and focus, the R, G, B, and NIR bands are highly correlated, as we will empirically show. If one looks for the patches most similar to a given patch within one and

the same band, then the patches most similar to the corresponding patch but within a different band are likely found there at the same locations. The overlap between such most similar patches across bands follows from but also indicates their degree of correlation (as will be shown). Our strategy therefore is to copy structure that is better visible in the not so noisy NIR band into the other bands, based on the aforementioned observations (for which we will give a stronger underpinning soon).

Related Work. The NIR luminance information has been used for color image restoration previously [6, 8, 15, 19]. For example, Krishnan and Fergus [6] propose to use gradients in both the NIR and Ultra Violet (UV) bands to arrive at improved RGB denoising. Zhuo *et al.* [19] apply a weighted least squares smoothing method to the RGB bands while transferring details from the NIR band. Matsui *et al.* [8] use joint bilateral filtering to decompose the RGB image into a so-called large-scale image and a detail image. The detail image is denoised by for each patch weighting all the neighboring patches according to their similarities to the NIR band. The denoised detail image is then added to the large-scale image for the final output. This method is perhaps most like ours. While Matsui *et al.* use all the patches from a search window in all bands, we look for the most similar patches in the clean NIR image, thus tapping into information that does not depend on the noise in the R, G, or B bands. Moreover, we believe that joint bilateral filtering is not robust in the presence of strong noise and an erroneous image decomposition affects the final denoising result – the noisy large-scale image is preserved, and an erroneous detail image is denoised. MMD minimizes this risk as we decompose each of the bands that are to be denoised into the comparably noise-free detailed NIR band on the one hand, and the images obtained as the difference between the R, G, or B bands and a weighted amount of the NIR band on the other, with noise mainly concentrated in the latter. Recently, Yan *et al.* [15] proposed a method named CrossField for joint image restoration via scale maps which model derivative-level confidence for inferring commonly usable edges across bands.

As the experiment section will show, our MMD algorithm substantially improves RGB denoising over the state-of-the-art methods, such as BM3D [2] and CrossField [15] for both simulated Gaussian noise conditions and real night.

The remainder of the paper is structured as follows. We introduce our Make My Day (MMD) approach for denoising in Section 2. There we provide the reasoning behind the approach and present empirical evidence for the assumptions made. The impact of the parameters used in the method are studied in Section 3. There we also report on its experimental performance. We refer to future work and potential applications in Section 4 and conclude in 5.

2. Make My Day Denoising

2.1. Denoising with Near Infrared

As input to the denoising system we have the noisy R, G, and B bands and a less noisy NIR band, all captured simultaneously. We will refer to the noisy Red, Green, Blue input bands as the x_R , x_G and x_B bands. The noise on these bands is supposed to be independent. The input NIR image x_{NIR} - supposed to be less noisy as NIR illumination is used - serves as a reference. Our goal is to denoise the RGB image and restore the original colors with high-fidelity.

MMD starts by calculating the difference between the R, G, and B images and the NIR image, resp.:

$$d_C = x_C - \alpha_C x_{NIR}, \quad C \in \{R, G, B\} \quad (1)$$

where α_C are the balance coefficients between each color band C and the NIR band. They allow us to control the influence of the NIR texture in each band. In our experiments α_R and α_G were empirically set to 0.4, while α_B to 0.2, since the camera sensor sensitivity for the B band is half that of the R and G bands. Ideally, α ought to be determined via histogram-matching between NIR and color channels for a camera condition.

Then we denoise the noisy difference images d_C , instead of the input images themselves, and finally add back the subtracted NIR image to obtain the final output (see Fig. 1). Filtering the difference images exploits the high correlation between the color and NIR bands.

For denoising we adopt the Non-local 3D Collaborative Filtering (3DCF) method, a process exploited in non-local denoising methods such as BM3D [2]. The 3DCF process (see Fig 1) goes as follows:

$$Z'_{ij} = T_{3D}^{-1} \Upsilon(T_{3D}(Z_{P_{ref}^{ij}})) \quad (2)$$

where T_{3D} and T_{3D}^{-1} stand for a 3D transform and its inverse, resp., Υ is a hard thresholding function, ij the label of patches we will describe later, $Z_{P_{ref}^{ij}}$ the 3D patch set formed from d_C , and Z'_{ij} the output 3D patch set.

The image is divided over a grid into N (overlapping) patches of size $b = h \times h$ ($h = 8$ in our experiments). Normally (*e.g.* in BM3D), in 3DCF for each patch p^i from the target channel, similar patches are obtained non-locally in a search window centered around the p^i patch of size $L \times L$ ($L = 40$ in our experiments) in its own channel. For MMD we assume the presence of a reference correlated channel from which the positions of the similar patches are extracted (see Fig. 1). For a p^i let p_{ref}^i be the patch in the reference channel at the same image position, and accordingly its corresponding search window. We retrieve k -nearest neighbors ($k = 32$ in our experiments) patches for p_{ref}^i in its search window. Let P_{ref}^{ij} be their image positions, where

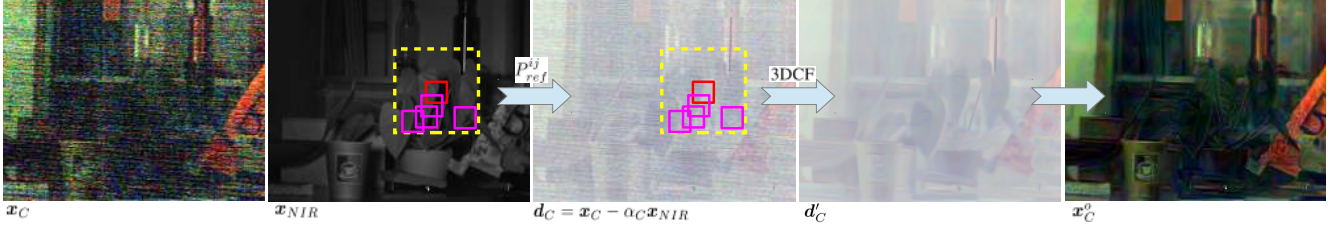


Figure 1: MMD steps: searching similar patches in the NIR band, forming 3D patches in the difference images at corresponding positions, and integrating the patches.

$j = 1, 2, \dots, k$. Because of the significant correlation between the channels (demonstrated in the next Section 2.2), we define the k NN patches of the p^i patch in the target channel based on the positions for the corresponding p_{ref}^i found in the reference channel. The k NN patches are stacked as $h^2 \times k$ vectors $Z_{P_{ref}^{ij}}$ at each patch, and transformed into the coefficient vector (which is of the same size). As 3D transform, we apply 2D Haar wavelet transform for the horizontal and vertical axes of the patch and an 1D Discrete Cosine transform (DCT) for the patch number axis. Then, hard thresholding Υ is performed on the 3D coefficient vectors as follows:

$$\Upsilon(y) = \begin{cases} y & \text{if } |y| > \rho; \\ 0 & \text{if } |y| \leq \rho. \end{cases} \quad (3)$$

where y denotes each coefficient from $T_{3D}(Z_{P_{ref}^{ij}})$, and ρ is the threshold value, which is set proportional to the noise level of the input image. Determining ρ is an important step in our method and is described in Section 2.3.

After the denoising step (see Fig. 1), the output 3D patches are put back onto their original locations, similar to what happens in BM3D. Like this, the denoised difference image d'_C is formed. The NIR image is then added back to d'_C to obtain the denoised final output:

$$\mathbf{x}_C^o = \mathbf{d}'_C + \alpha_C \mathbf{x}_{NIR}, \quad C \in \{R, G, B\} \quad (4)$$

The similarity search is performed only once at the first step, and the position information P_{ref}^{ij} is used for denoising d_C for all color bands $C \in \{R, G, B\}$. The denoised difference images d'_C are blurred at the edges as a side effect of the 3DCF. On the other hand the NIR image contains the high frequency texture. Therefore, adding the denoised difference images to the NIR image results in a high-quality output image which combines the RGB color information with the luminance texture from the NIR image.

Algorithm 1 and Fig. 1 summarize our MMD method.

2.2. Correlation between Color Channels

The bands of a color image are correlated, supposing they capture the same scene at the same moment. As a matter of fact, the closer the channels are in the color spectrum (see Fig. 2) the more correlated they are. This correlation

Algorithm 1. The Make-My-Day (MMD) algorithm.

Input:

\mathbf{x}_C - the observed image bands, $C \in \{R, G, B\}$,
 \mathbf{x}_{NIR} - the reference image band (NIR),
 α_C - balance coefficient, a - coefficient

Patch location formation:

Obtain the similar patch positions P_{ref}^{ij} from \mathbf{x}_{NIR}
 Estimate noise σ
 Calculate $\rho = a\sigma$
 Compute $\mathbf{d}_C = \mathbf{x}_C - \alpha_C \mathbf{x}_{NIR}$

For each patch i :

Form the 3D patch set $Z_{P_{ref}^{ij}}$ from \mathbf{d}_C
 Perform 3DCF:
 $Z'_{ij} = T_{3D}^{-1} \Upsilon(T_{3D}(Z_{P_{ref}^{ij}}))$.

Pixel-wise average output patches Z'_{ij} to form \mathbf{d}'_C

Output:

image $\mathbf{x}_C^o = \mathbf{d}'_C + \alpha_C \mathbf{x}_{NIR}$

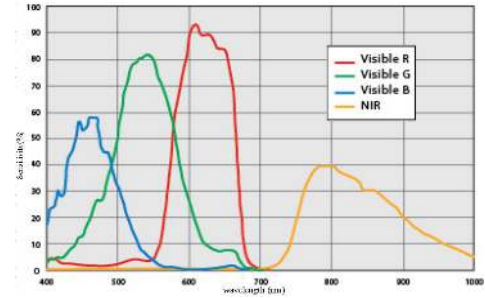


Figure 2: Spectral sensitivity of JAI AD130GE camera.

also implies that local self-similarities within one channel are found with high probability at corresponding positions in another channel, as we empirically show next. This ‘non-local means principle’ allows MMD to combine the most similar patches (aka nearest neighbors) for noise reduction. In order to denoise a channel, named ‘target channel’, one can use the self-similarity relations found in a different correlated channel with less noise, named ‘reference channel’.

We investigate the correlation between bands, using the RGB-NIR channels of Image 1 and Image 2 (Fig. 6). Firstly, the k nearest neighbors (k NN) of the target patch (here a block of 8×8 pixels) is obtained in the target channel. ¹

¹We bounded the search of nearest neighbors to a 40×40 surrounding search window.

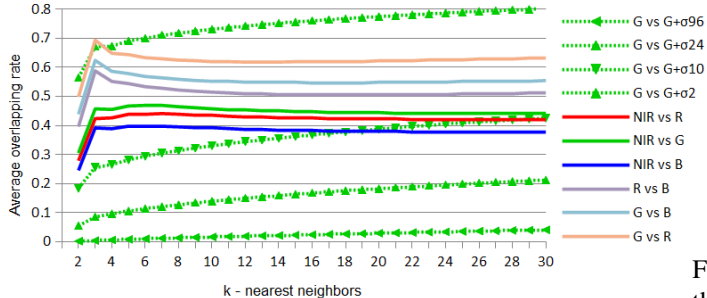


Figure 3: Correlation between original RGB-NIR channels of Images 1 and 2 as an indication of the overlap between the positions of the k NN in a channel for a patch and the positions of the k NN in another channel for a patch at the same position.

Secondly, the k nearest neighbors (k NN) are obtained for the corresponding patch at the same location in the reference channel. Thirdly, we count the number of positions from the k NN found in the target channel for a patch that are among the positions from k NN found in the reference channel for a corresponding patch at the same image location. By dividing by the number k of nearest neighbors we measure the overlap between the bands as an indication of the structural similarities. By averaging over all the image target band patches we obtain (as shown in Fig. 3) average estimated overlapping rates. Thus, a value of 0.4 means that, on average, 40% from the positions of the k NN for any target patch are found among the positions of the k NN for the reference band patches. For MMD to work, this value should be high.

The correlation between the color bands varies from above 0.6 for G and R and slightly below 0.4 for NIR and B in terms of average overlapping rates. The correlation is stronger between the R, G, and B bands and weaker between these and the NIR band (see Fig. 2).

Apart from these cases where the noise on the bands was low, we also report the correlations (overlapping rates) between the G channel without added noise and the same G channel with different amounts of Gaussian noise added ($\sigma \in \{2, 10, 24, 96\}$). As expected, the correlation decreases as more noise gets added. What is essential for our method is that the correlation between the G band without added noise and the NIR channel is significantly higher than the one between the same G channel and the G channel damaged by medium or high amounts of noise ($\sigma > 10$).

We conclude that there is a significant correlation between the ‘noise-free’ color bands (see Fig. 3). Moreover, the similarities computed in such NIR band are closer to those in these color (RGB) bands than the similarities found in color (RGB) bands affected by medium to high noise. This result is important as it implies that we cannot rely on accurate self-similarities/ k NN retrieval within a channel affected by noise. As a result, using a correlated, less noisy

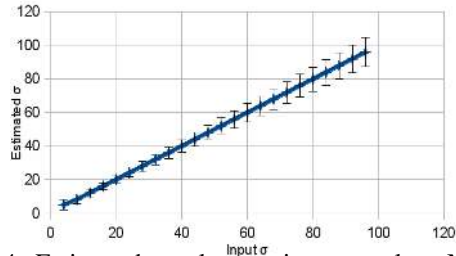


Figure 4: Estimated σ value vs. input σ value. Marks are the average and error bars are the standard deviation of estimated σ for thousands of different patches.

band is an appealing alternative. We exploit this observation in our MMD method.

2.3. Noise Estimation

In order to estimate the noise we use the correlation between the color bands and consider the NIR image as a noise-free reference for the noisy channels. First, we look for ‘flat’ patches (with small standard deviation) in the NIR image and assume that the corresponding patch in the target color channel image ought to be flat as well. Then, we calculate the difference in variance between the target color patch $\mathbf{x}_{i,C}$ and the NIR patch $\mathbf{x}_{i,NIR}$ as follows:

$$\sigma^2 = \text{Var}(\mathbf{x}_{i,C})^2 - \text{Var}(\mathbf{x}_{i,NIR})^2 \quad (5)$$

As a result, the random noise level σ (without structural signals) is calculated as the standard deviation for each acquired patch.

The relation between the actual σ value and the estimated σ value is shown in Fig. 4. We plot the average and standard deviation for estimated σ values for thousands of different patches. If we assume that the amount of noise is homogeneous throughout the image (the case of low-illumination where the sensor read-out circuit noise is dominant, as in the bulk of our experiments) then we take the averaged σ . Otherwise, when the noise is not homogeneous, the estimated σ can be memorized for each patch.

We set the threshold value ρ based on σ . In order to investigate the relation between the amount of noise and the best threshold value, we created a test image with Gaussian noise on a uniform offset, and performed the 3DCF process on it. For real camera noise, we used our JAI AD130GE RGB-NIR camera whose spectral sensitivity is in Fig. 2. The same 3DCF process has been carried out on the JAI camera output noise of one frame without incident light and with maximum analog gain. We found that the amount of noise and the threshold capable of denoising are proportional. This is because the 3D transform is a linear process. The sensor output images are considered to be result from the addition of pure signal and noise, so the threshold to reduce a certain amount of noise can be uniquely determined.

Fig. 5 shows the noise removal characteristics against ideal Gaussian noise and JAI camera dark noise. The hard

threshold ρ which is required for the suppression of a certain amount of noise differs between Gaussian and JAI noise. This is because of the low frequency spatial noise in JAI camera noise. Low frequency noise is correlated in the search window of 3DCF, and therefore difficult to remove. Here we set the coefficient for Gaussian noise to 4, which can suppress the noise by 30 dB, and the coefficient for JAI camera noise to 14, thus expecting 20 dB of noise removal.

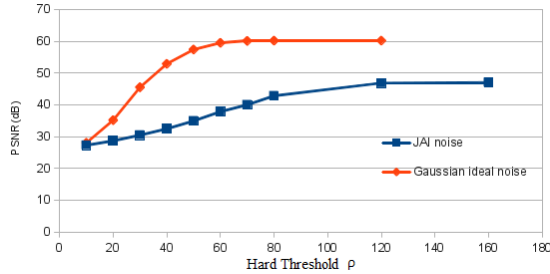


Figure 5: PSNR vs. hard threshold value (ρ) of the uniform image with Gaussian artificial noise and JAI dark noise. Both have standard deviation (σ) of 12.

3. Experiments

In this section we present the experimental setup (images, settings), analyze the effect of the MMD parameters, and report on our results for simulated Gaussian noise as well as for natural noise in comparison with state-of-the-art methods. Our codes and images are available at: <http://www.vision.ee.ethz.ch/~timofter/>

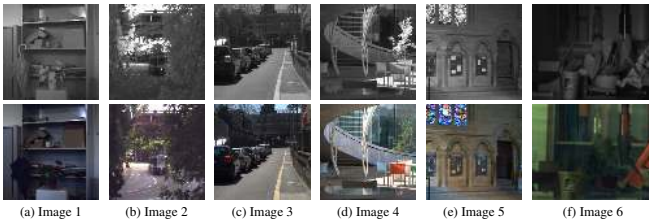


Figure 6: NIR and RGB test images 1-6; 7-9 are in Fig. 12.

3.1. Benchmark

Datasets. We conduct denoising experiments on the 9 RGB-NIR image sets shown in Figs. 6 and 12.² Image 1 through 5 were recorded under daylight illumination while the Images 6 through 9 have been taken in the dark (faint visible illumination) but with NIR illumination. Images 3, 4, and 5 are borrowed from the IVRG dataset [13], and the remaining ones have been recorded by us. We employed the JAI-AD130GE camera with two CCD sensors - an RGB sensor (Bayer layout) and an NIR sensor. Incident light passes through a C-mount lens in front of the prism which separates the light into RGB and NIR bands. Hence, the camera is capable of recording synchronized images of all

²Most of the RGB-NIR from the existing datasets and literature are not well-aligned at pixel level and/or do not provide reliable ground truth.

four R-G-B-N bands. There is some chromatic aberration between RGB and NIR due to the optical system, but we correct it by shrinking the NIR image digitally after recording. We use the NIR LED SAL30 for CCTV applications as light source. The RGB and NIR channels are obtained simultaneously, with the same exposure, but the analog gain (highest) and digital gain ($\times 4$) are applied only for the RGB channels to obtain sufficient output values. The RGB images contain the regular noise - a mixture of sensor read-out circuit noise, shot noise, and extra unknown noise with various frequency ranges and spatial correlation.

The image sets (Figs. 6 and 12) cover various aspects such as natural and handmade objects and colors, indoor and outdoor scenes, daylight and night conditions. For example, in the indoor Image 1 the fabric of the umbrella and the leaves of the plant have a high NIR reflectance and therefore look bright in the NIR image. In the outdoor Image 2 the leaves occupy a large area of the image, resulting in a poor correlation between NIR and RGB images.

As for Image 6 (night), the ‘ground truth’ (GT) is obtained by averaging 1024 frames (30 sec) with the camera and the objects kept stable. The reason why we used frame averaging instead of long exposure is that long exposure changes the color balance due to the different offsets on the channels and the obtained image would therefore not be appropriate as the GT image.

Methods. We compare our MMD method with state-of-the-art methods. BM3D [2] is a well established state-of-the-art method for image denoising (enhancement) with publicly available code and is related to our MMD at the level of its filtering operations. BM3D does not use the NIR image. 3DCF+NIR is our MMD method where the color balance coefficients (α_C) are set to 0, and the denoising is thus directly applied on the input color images. By including 3DCF+NIR we show how critical the use of difference images is for denoising. CrossField [15] is a very recent method with top results using NIR images and a more intensive computation. We are grateful to the authors for running their method on our test images. The method of Matsui *et al.* [8] is unavailable, as we were told by the authors, and our implementation performs worse than BM3D. Unfortunately, there is a scarcity of both publicly available code sets and benchmarks for the task at hand.

Evaluation. We consider real night conditions as well as simulated noise conditions. For the case of real noise in dark conditions we test on 4 image sets. For simulated noise we add Gaussian noise to the RGB channels of Images 1 to 5, independently. All the evaluated methods share the conditions, and are used with the same input images. We report both quantitative results using Peak Signal to Noise Ratio (PSNR) and qualitative results. The recovered image is compared with the ground truth, the original image without noise or with a reference image captured in day-

light conditions.

3.2. Influence of parameters

We investigate the MMD parameters k and L on Image 1 to 5 with added Gaussian noise ($\sigma = 96$). The average PSNR saturates for k above 40 as shown in Fig. 7, where $L = 40$. We pick $k = 32$ for MMD as a trade-off between performance and computation time. For the same images and noise levels, we investigate the relation between the search window size L and the PSNR (Fig. 7). The MMD performance peaks for $L = 40 - 80$ and decays above 80. We suppose this is because of the RGB-NIR correlation getting weaker. The larger the search window gets, the more likely it is to search on different objects, which results in a poorer patch correlation between channels. Considering also the computation time, we set $L = 40$. The best L and k depend on the image content and the amount of noise. They could be optimized for each image separately.

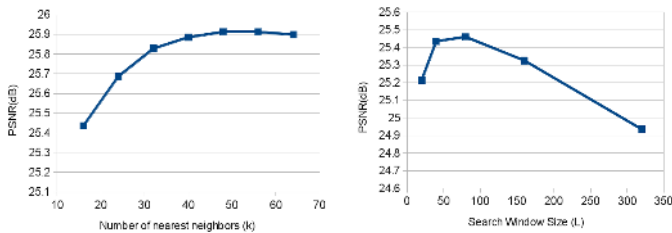


Figure 7: Parameters versus MMD performance obtained by averaging denoising results on Images 1 to 5. (left) k vs. PSNR for $L = 40$; (right) L vs. PSNR for $k = 32$.

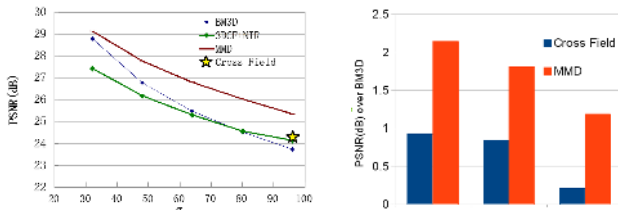


Figure 8: Method comparison. (left) Performance vs. noise level. Each point is the average of the RGB results on the Images 1-5. (right) Performance gain over BM3D vs. color channel. Each bar is the avg. of the Images 1-5 with $\sigma = 96$.

3.3. Simulated noise results

We report results on the RGB-NIR Images 1 to 5 with simulated Gaussian noise. Fig. 9 shows the noisy input images ($\sigma = 96$) and the BM3D and MMD denoising results. Compared to BM3D, MMD restores much more details. A PSNR comparison between BM3D [2], CrossField [15], our 3DCF+NIR, and MMD is shown in Table 1 and Fig. 8. MMD is more effective in the case of strong noise as shown in Fig. 8, outperforming BM3D by 1.7 dB in PSNR. The Red channel is better restored than the Blue channel, and this because the NIR correlates better with the Red than the Blue channel. The CrossField method also exploits NIR,

but MMD outperforms it in terms of RGB image restoration with higher color fidelity. This is because MMD is not based on the NIR intensity but on the texture information.

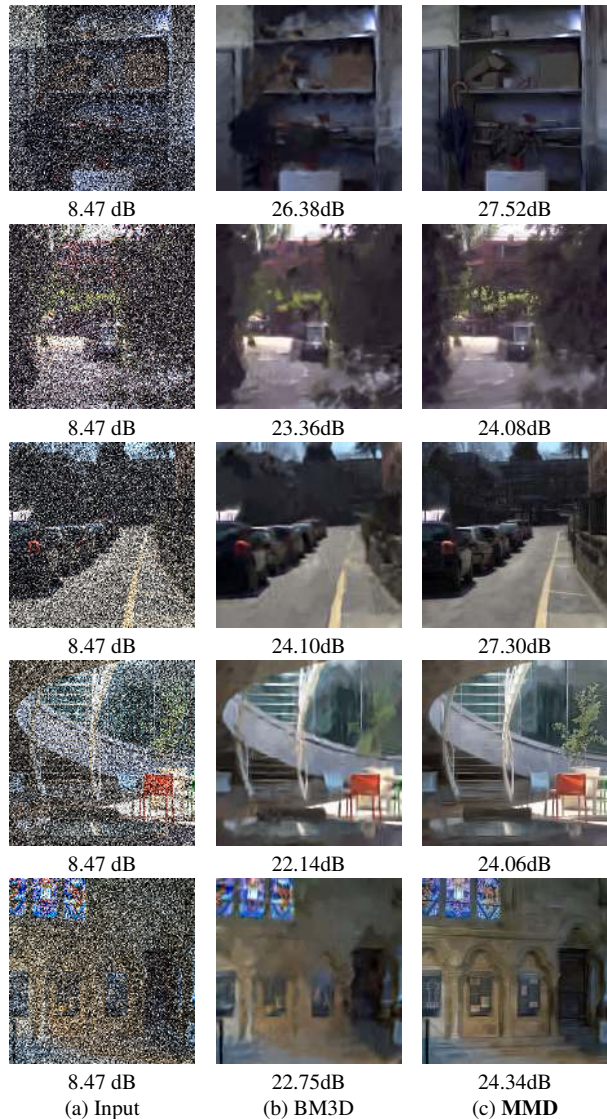


Figure 9: Denoising BM3D and MMD results for images 1 to 5 with added Gaussian noise ($\sigma = 96$). Some dots of the noisy input images are saturated and appear white here.

Table 1: Method comparison in average PSNR [dB] of R, G, and B channels over input images with high Gaussian noise ($\sigma = 96$).

Input \ Method	BM3D [2]	3DCF+NIR	CrossField [15]	MMD
Image 1	26.38	27.36	26.80	27.52
Image 2	23.36	23.13	23.52	24.08
Image 3	24.10	24.80	25.67	27.30
Image 4	22.14	22.96	22.20	24.06
Image 5	22.74	22.41	23.84	24.34
Average	23.74	24.13	24.41	25.46

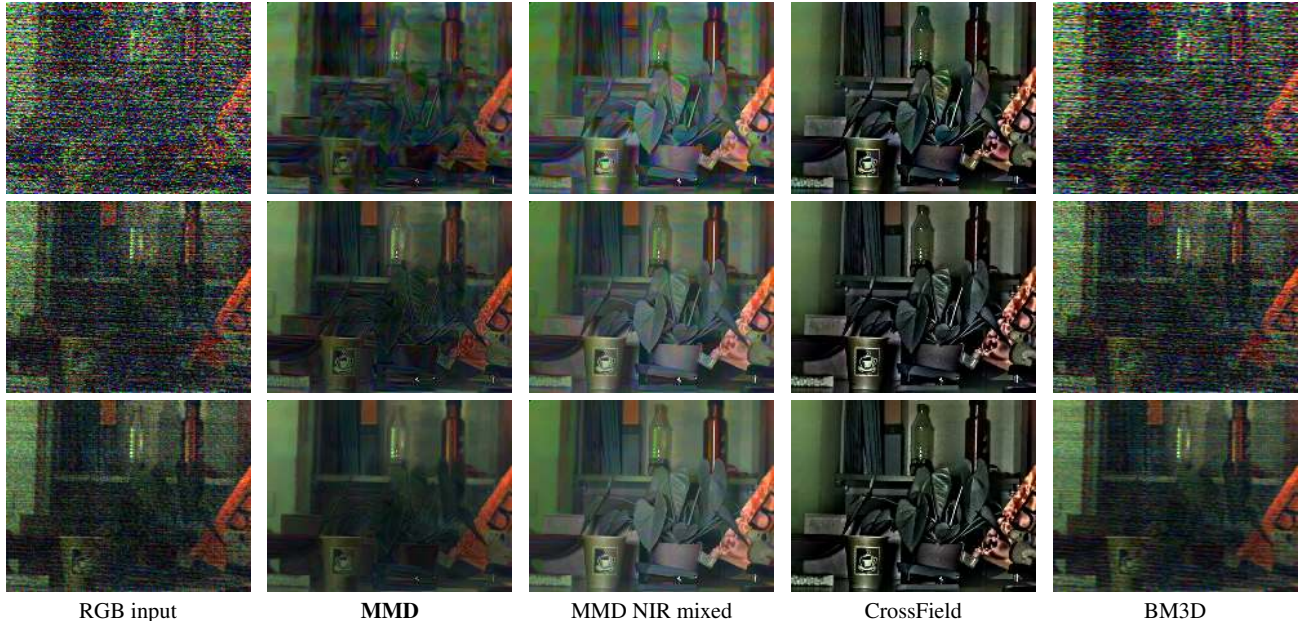


Figure 10: Denoising results of natural images under NIR illumination, recorded at night in a room with the windows covered. From top to bottom: images of a static scene averaged over 1 (30 msec), 4 (120 msec), and 16 frames (480 msec). The NIR input and ground truth are shown in Fig. 6. False color patterns result from the low frequency noise component.

3.4. Real noise results

Our MMD method aims at restoring reasonable quality RGB images at night conditions, when those channels tend to get extremely noisy. Whereas we had artificial Gaussian noise added to Images 1 to 5 captured under daylight conditions, for the images captured in the dark the noise is a mixture of pixel-wise and horizontal stripe noise with various frequency ranges. Moreover, the noise is periodic, has spatial correlation. The issue is that there is low-frequency noise of a size larger than the 3DCF search window size. First, for Image 6 we perform noise estimation and obtain $(\sigma, \rho) = (13, 182)$ for 1 frame input, $(6.4, 90)$ for 4 frames average input, and $(3.2, 45)$ for 16 frames (see Fig. 10). We average the input frames as a straightforward way to reduce the noise and to compare the denoising methods.

We compare our MMD method with BM3D [2] and CrossField [15] in Fig. 10. The details in the RGB image are hardly distinguishable because of the strong camera noise. MMD restores the details in the input RGB reasonably well, guided by the NIR image. The plant, the flag, and the picture on the paper cup are well restored. The residual color noise is the low-frequency correlated noise. This could be reduced more provided the 3DCF search window size L is set larger, but at the expense of computation time. The CrossField method [15] also recovers the texture well, but, since the method relies more on the NIR intensity, it generates unnatural colors like pale colors on the plant, the flag or the bottle behind. Furthermore, we mix the MMD output RGB image and the NIR input in the YCbCr color

space as shown in Fig. 10 for a reference. 50% of the MMD output Y image and 50% of the NIR input are mixed to form the combined output’s Y, and the MMD Cb and Cr outputs are used directly as the combined outputs Cb and Cr, respectively. This result has less color fidelity but good visual quality. It shows that using more of the NIR channel improves the perceived vividness / quality (artificially (!)) at the expense of losing color fidelity. The CrossField results miss color fidelity and some details (*e.g.* the flower-pot color is different, the brown strip on the left is missing). The color artifacts in the CrossField result are hidden by the strong use of the NIR input, nonetheless at the expense of color fidelity. MMD exploits the NIR texture information and is capable of maintaining the original color balance. In PSNR terms, MMD largely outperforms CrossField [15] by 6.4 dB (1 frame input), 10.7 dB (4 frames averaged input), and 11.9 dB (16 frames averaged input). In Fig. 11, for Image 6, we go further and compare the denoised image results after (the same) brightness-contrast adjustment with the reference image taken under proper lighting conditions. MMD clearly has the highest color fidelity.

In Fig. 12 we report our Make-My-Day results for 3 other image sets (7, 8, and 9) that were captured at night time and we compare their results with their reference daylight images. We use the same automatic contrast enhancement operation (‘Equalize’ from the GIMP 2.8 software) for both the RGB input and MMD output. The first row is a typical setup and MMD performs very well given the single frame input. The second image set is taken under darker conditions with a faint light source remaining, yet the



Figure 11: Denoising results for Image 6 (Figs. 6 & 10) in comparison with a reference ‘daylight’ image after applying the same brightness-contrast adjustment operation to each image. The MMD colors come closest to those of the reference image.

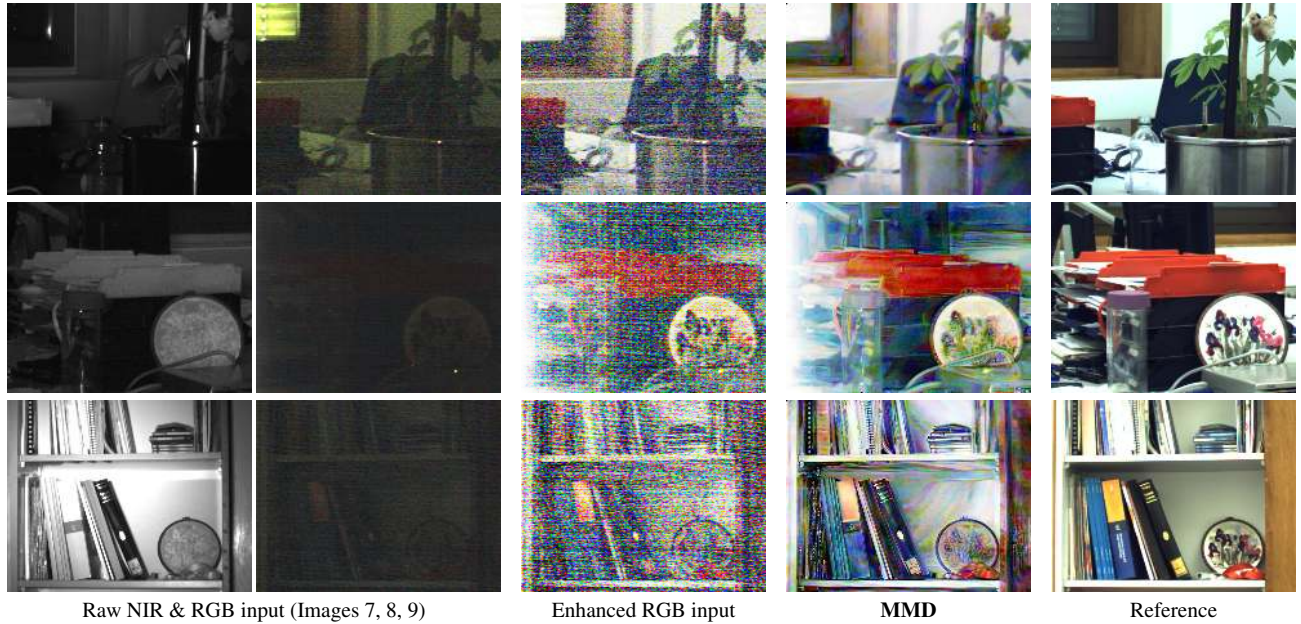


Figure 12: Make-My-Day results for NIR & RGB input images recorded in the night. The camera settings were set to for the best gain. The third row is a case of NIR saturation due to a short distance to the NIR illumination. Note the color similarity between the MMD results and the reference images taken in daylight.

MMD result stays close to the reference. The third image set shows how much MMD is influenced by the NIR band. The saturation of the NIR band induces a ripple-noise effect in the flat regions of the MMD result. There is a trade-off to be noticed. The poorer the SNR of a color channel is, the stronger is MMD’s suppression of the color signal and more false patterns near edges appear, because the difference images d_C get blurred.

4. Future Work

The proposed strategy can be used for any image set that consists of simultaneously captured multi channel images, where one or more channels are more noisy than the others. **Multi Spectral Images.** Mid and long-wave infrared imaging techniques are popular in night vision applications. They can often do without artificial illumination. The correlation between those wavelength ranges and RGB is weaker than between NIR and RGB, however. Also under those circumstance, MMD will probably be able to yield color-restored images with hot objects highlighted, for instance.

Shadow Removal/Enhancement. MMD can be combined with the shadow detection technique described in [10]. The detection is carried out by the multiplication of NIR and RGB images. Low values indicate shadows. MMD could be used to restore the colors in the shadow area while maintaining the color balance.

5. Conclusions

We proposed the ‘Make My Day’ (MMD) method, a novel approach to turn night images into daylight-like images. MMD is a high-fidelity color denoising method using NIR illumination and sensing. MMD exploits the correlation between the color and NIR bands and uses the similarity information from the less noisy NIR band in a 3D collaborative filtering framework. The filtering works on the difference image between visible and NIR, to then add back the NIR image. Thus, MMD restores the texture in the high spatial frequency range while maintaining the color balance. As a result, it outperforms other state-of-art methods in terms of PSNR, texture quality, and color fidelity.

Acknowledgments. This work was supported by the ETH Zurich, Switzerland, and Toshiba Corp, Japan.

References

- [1] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65, 2005. [1](#)
- [2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. [1](#), [2](#), [5](#), [6](#), [7](#)
- [3] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, 2006. [1](#)
- [4] D. Firmenichy, M. Brown, and S. Susstrunk. Multispectral interest points for RGB-NIR image registration. In *Proc. International Conference on Image Processing (ICIP)*, pages 181–184, 2011. [1](#)
- [5] M. N. A. Khan, G. Fan, D. R. Heisterkamp, and L. Yu. Automatic target recognition in infrared imagery using dense hog features and relevance grouping of vocabulary. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 293–298. IEEE, 2014. [1](#)
- [6] D. Krishnan and R. Fergus. Dark flash photography. *ACM Transactions on Graphics, SIGGRAPH*, pages 1–11, 2009. [2](#)
- [7] Y. Luo, J. Remillard, and D. Hoetzer. Pedestrian detection in near-infrared night vision system. In *2010 IEEE Intelligent Vehicles Symposium*, pages 51–58, 2010. [1](#)
- [8] S. Matsui, T. Okabe, M. Shimano, and Y. Sato. Image enhancement of low-light scenes with near-infrared flash images. In *ACCV*, 2009. [2](#), [5](#)
- [9] N. Salamati, C. Fredembach, and S. Süssstrunk. Material classification using color and nir images. *Proc. IST/SID 17th Color Imaging Conference*, 2009. [1](#)
- [10] N. Salamati, A. Germain, and S. Susstrunk. Removing shadows from images using color and near-infrared. In *International Conference on Image Processing (ICIP)*, pages 1713–1716, 2011. [8](#)
- [11] L. Schaul, C. Fredembach, and S. Süssstrunk. Color image dehazing using the near-infrared. *Proc. IEEE International Conference on Image Processing*, pages 1629–1632, 2009. [1](#)
- [12] D. Socolinsky, L. Wolff, and A. Lundberg. Image intensification for low-light face recognition. In *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, pages 41–41, 2006. [1](#)
- [13] S. Süssstrunk. RGB-NIR scene dataset. http://ivrg.epfl.ch/supplementary_material/cvpr11/. [5](#)
- [14] A. Toet. Color the night: applying daytime colors to night-time imagery. In *Proc. SPIE*, volume 5081, pages 168–178, 2003. [1](#)
- [15] Q. Yan, X. Shen, L. Xy, S. Zhuo, X. Zhand, L. Shen, and J. Jia. Cross-field joint image restoration via scale map. In *Proc. ICCV*, 2013. [2](#), [5](#), [6](#), [7](#)
- [16] X. Zhang, T. Sim, and X. Miao. Enhancing photographs with near infra-red images. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. [1](#)
- [17] Y. Zheng. An overview of night vision colorization techniques using multispectral images: From color fusion to color mapping. In *Proc. International Conference on Audio, Language and Image Processing (ICALIP)*, pages 134–143, 2012. [1](#)
- [18] Z. Zhu and Q. Ji. Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Comput. Vis. Image Underst.*, 98(1):124–154, 2005. [1](#)
- [19] S. Zhuo, X. Zhang, X. Miao, and T. Sim. Enhancing low light images using near infrared flash images. In *Proc. International Conference on Image Processing (ICIP)*, pages 2537–2540, 2010. [2](#)