Compression of Multispectral Images: Color (RGB) plus Near-Infrared (NIR)

Neda Salamati [#1], Zahra Sadeghipoor [#2], Sabine Süsstrunk [#3]

 # School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
[1] neda.salamati@epfl.ch
[2] zahra.sadeghipoor@epfl.ch
[3] sabine.susstrunk@epfl.ch

Abstract—We propose a compression framework for four-channel images, composed of color (RGB) and near-infrared (NIR) channels, which exploits the correlation between the visible and the NIR information. The high-frequency components of both visible and NIR scene representations are strongly correlated. By encoding only the DCT components that differ above a chosen threshold, we significantly improve compression ratios for a given quality level. To evaluate our proposed method, we compare our results with standard JPEG compression, as well as PCA-based approaches that are often employed to compress multispectral images. Our experiments show that applying our proposed method yields the same quality at a lower bit-rate, compared to conventional JPEG and PCA-based algorithms.

I. INTRODUCTION

Silicon-based camera sensors exhibit significant sensitivity beyond the visible spectrum (400-700 nm). They are able to capture wavelengths up to 1100 nm. Near-infrared (NIR) is the part of the radiation spectrum that ranges from 700 to 1100 nm. Even though this radiation can be captured by silicon, it is usually considered noise and is discarded by fixing a filter (hot-mirror) in front of the sensor. However, retaining instead of eliminating NIR information improves certain tasks in digital photography and computer vision, such as image enhancement [1], scene categorization [2], and illumination estimation [3]. Lu *et al.* in [4] propose a color filter array (CFA) design that can be employed to simultaneously capture NIR information in addition to red, green, and blue (RGB) channels in the visible part of the spectrum on a single sensor.

Compared to conventional color imaging, these emerging applications and this acquisition approach produce larger amounts of data to be transmitted, processed, and stored efficiently. Although many multispectral compression algorithms have been proposed [5], [6], [7], [8], to the best of our knowledge, the compression of RGB and NIR has not been specifically addressed so far. Therefore we propose a novel framework for RGB+NIR (RGBN) image compression.

There exist several compression algorithms for three-channel RGB images. Currently, one of the vastly employed lossy compression methods is the JPEG standard [9]. This method compresses images by quantizing the Discrete Cosine Transform (DCT) coefficients, and it yields acceptedly good results for visible images. The color images are first transformed into the YCbCr color space. The luminance component

Fig. 1. Top row: Color (RGB) images. Bottom row: Near-infrared (NIR) images of the same scene. While the image intensities between the two scene representations differ, we can immediately notice that the images are from the same scene due to the similar edge information.

(Y) is compressed with a quality better than the chrominance components, as the human visual system is less sensitive to high-frequency loss in chromatic components. In the case of four-channel (RGBN) images, JPEG can be used to compress RGB channels, and it can also be generalized to compress the fourth channel (N) (i.e., NIR as a one-channel gray-scale image is treated like Y in YCbCr and is encoded in the same way).

JPEG 2000 [10] differs from JPEG in the transform domain employed. Instead of DCT, multi-level discrete wavelet coefficients are computed for each channel. JPEG 2000 outperforms JPEG in terms of quality at very low bit-rates. Nevertheless, we chose to use JPEG instead of JPEG 2000 for two reasons. First, we are interested mostly in moderate compression, as used in most photographic applications. Second, most cameras still allow only JPEG compression due to the higher computational complexity that JPEG 2000 encoding entails. However, the proposed framework could easily be applied in the wavelet domain.

Another approach to compressing RGBN four-channel images is to employ a multispectral compression framework. Pennebaker *et al.* [11] and Abousleman *et al.* [7] propose to extend two-dimensional JPEG or JPEG 2000 into a three-dimensional version for multispectral compression. Fowler and Rucker [5] argue against the performance of these compression
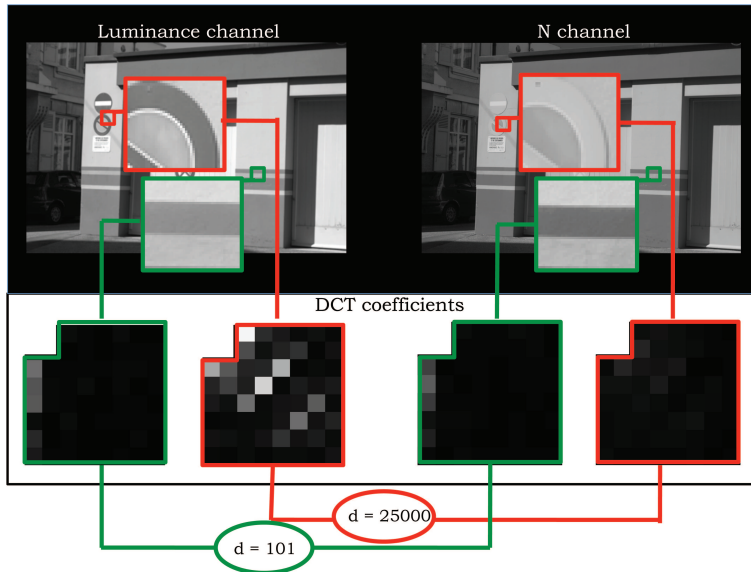
Fig. 2. An example of high frequency information of two different patches in the N (near-infrared) and the Y (luminance) channels. The squares with similar information in both channels have a green border and those with dissimilar information have a red border. The top row shows the squares in the spatial domain, and the bottom row the corresponding DCT coefficients.

approaches, because such methods do not take the specific characteristics of such data into account. To explore this potential, [5], [6] and [12] employ a two-stage process. First, the spectral correlation is removed, then the spatial correlation is exploited. In the first stage, a number of transforms are used to de-correlate the spectral data. Principal Component Analysis (PCA) [13] and Vector Quantization (VQ) [8] are the most commonly used methods in multispectral compression. One of the drawbacks of using a single transform matrix computed for a given database is that the compression performance depends on how well the database represents the world and how close the test image is to the database. For the second stage, in the spatial dimension, blocks of the transformed bands can be compressed by using a variety of well-known image compression methods such as DCT and discrete wavelet transform (DWT) quantization [5].

In this paper, we explore the potential of incorporating the specific characteristics of the NIR representation and their relations to the visible counterpart into the compression framework. As the NIR and RGB images represent the same scene, despite the many differences between them, there still exist many similarities in edge information and image details (see Figure 1 for illustration). Therefore, we propose an efficient compression framework that exploits the correlation and similarities and removes the redundancy between these channels.

Due to the specific nature of our multispectral images that are composed of three visible (RGB) channels and one invisible (N), we use an existing method, JPEG, to compress the RGB image. As a result, if only the visible information is needed, it can be decompressed without any additional computational overhead. If only the NIR image is desired, we decompress Y and N. This is another advantage of our method

over PCA-based compressions where all four channels always need to be decompressed.

In order to compress the N channel, we study the similarities between this channel and Y. We analyze the NIR and luminance correlation in the DCT domain, and we detect highly correlated patches in these channels. As the Y channel is already coded by normal JPEG, we completely remove the similar information in coding the N channel. When we reconstruct the N channel, the missing information is recovered from the decompressed Y channel.

We compare the performance of our algorithm to the results of conventional JPEG on four channels and PCA-based multispectral compression. We observe that our proposed framework achieves the same quality with a lower bit-rate. For the same bit-rate, our proposed method yields an improvement up to 5 dB in PSNR results compared to JPEG for NIR images.

In Section II, we study the correlation between the N and Y channels. We describe our compression framework in Section III. We compare our algorithm with PCA-based multispectral compression and conventional JPEG and present the results in Section IV. Section V concludes the article.

II. RGB+NIR Image Attributes

Studying RGB and NIR representations of different natural scenes, we observe that although a large portion of blocks look different in the NIR and RGB images, a significant amount of these blocks contain similar details in the N channel as they do in Y; and the main dissimilarity is the difference in their average pixel intensities. For instance, the blocks marked with green borders in Figure 2 show very similar edges, whereas the block in the N channel has pixel intensities different from Y. As edges mainly contribute to high-frequency information, the high-frequency coefficients of the Y channel are strongly
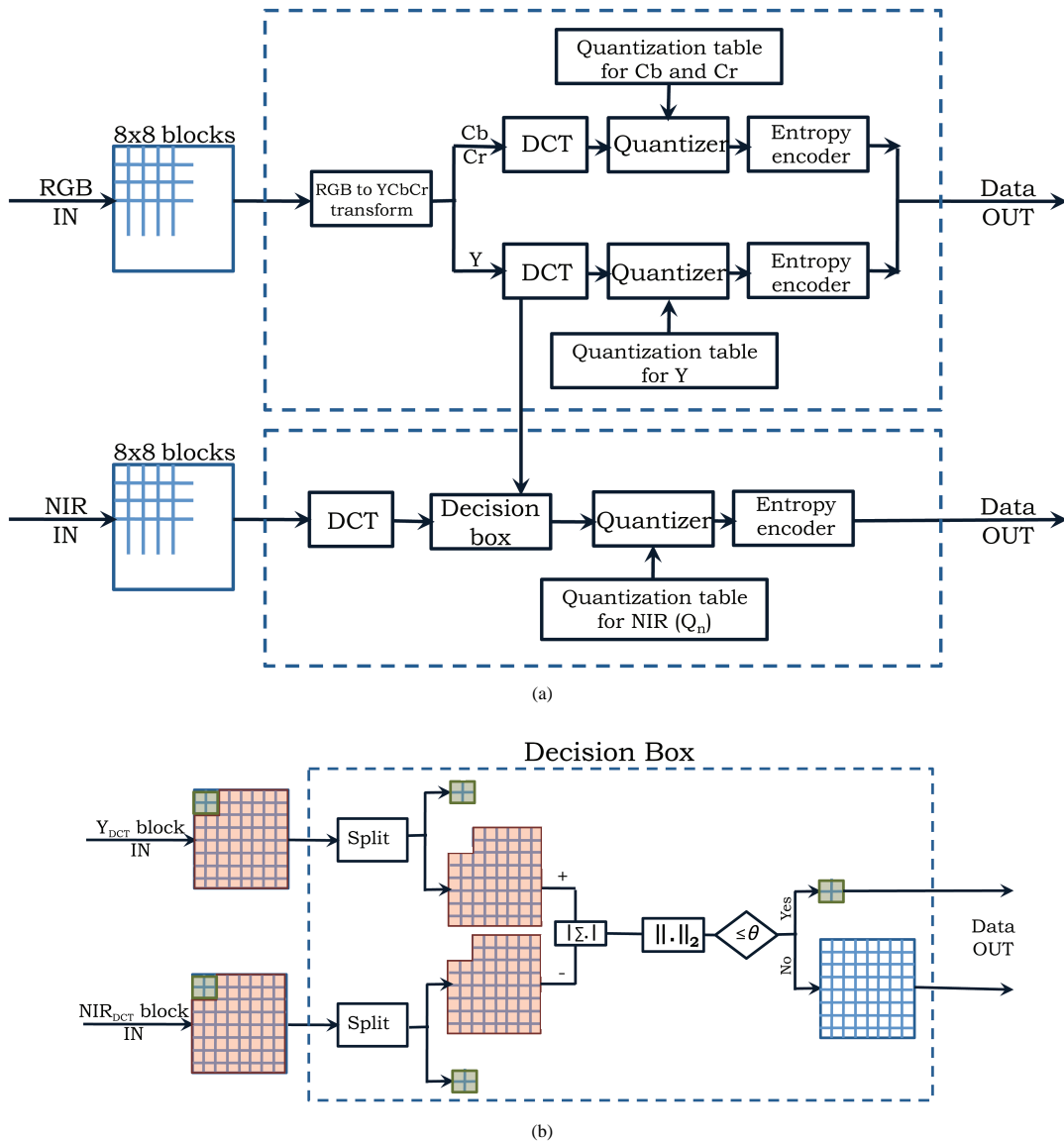
Fig. 3.   The block-diagram of our proposed compression framework.

correlated to those of N. The difference in DCT coefficients is very small in the high-frequency part of the spectrum.

The second group of blocks are those that look significantly different in N and Y channels. In these blocks, both texture and pixel intensities differ in the visible image and the NIR representation, due to different material and illuminant characteristics in the different wavelength bands. An example of this kind of image block and the difference between Y and NIR coefficients in the DCT domain are shown in Figure 2 in red.

III. PROPOSED RGB+NIR COMPRESSION SCHEME

Many transform-based compression algorithms [10], [9] take advantage of the fact that natural images can be sparsely represented in the frequency domain. In this paper, we follow the same approach to efficiently code the frequency information of RGB and NIR images.

In RGB image compression, the most commonly used color encoding is YCbCr, where Y stands for luminance and Cb and Cr for chrominance (blue-yellow and red-green, respectively). The reason is that the spatial and chromatic information of natural RGB images are well separated in the YCbCr space. The different channels are de-correlated, and little spectral redundancy exists between the different channels. Moreover, the human visual system is much less sensitive to distortions in high-frequency chromatic information, which allows us to strongly compress the chrominance channels without affecting image quality.

Considering the advantages of YCbCr, we first transform the RGB image to YCbCr and then compress them according to the JPEG standard. This approach includes computing the representation of $N \times N$ blocks in the DCT domain, quantizing the DCT coefficients, and then applying an entropy coder to

the quantized coefficients. We use the 8-bit Huffman coding to encode the data.

To exploit the spectral redundancy between the Y channel and the NIR image, we consider the frequency information of the NIR block and its visible counterpart. As can be seen in Figure 2, many of the regions in these two representations share almost the same "texture" information. Hence, we need to code the texture information of these blocks only once and then use it to reconstruct both representations.

We assume that blocks with similar textures in N and Y behave similarly in high-frequency bands. Let us refer to the DCT coefficients of a given block as $(B_{DCT}|B \in \{Y, N\})$. We consider the first $L \times L$ block of coefficients as low-frequency (L_B) and the rest are considered as high-frequency components (H_B). For each block, the energy of difference between N and Y high-frequency components is computed as follows:

$$d = \sum_{i=1}^{V} (H_N(i) - H_Y(i))^2, \tag{1}$$

where V is the number of coefficients in H_B.

For a given block, if d is smaller than a predefined threshold (θ), then this block is considered to contain the same texture in Y and N (Figure 2, the green block). In this case, compressing the high frequency of both N and Y leads to the transfer of redundant information. Thus, we remove the high frequencies and quantize only the low-frequency coefficients for N, while we keep and quantize all DCT coefficients of the Y channel. At the decoder, the missing high-frequency coefficients of NIR is estimated using the corresponding information in the luminance.

However, if the energy of difference (d) exceeds the threshold, there is no redundancy between Y and N in that block (Figure 2, the red block). Thus, all DCT coefficients for both of these channels have to be coded and stored. In all cases, DCT coefficients are quantized and then coded using the Huffman entropy coder.

The quantization table for an NIR block can be written as follows:

$$Q_n = \begin{cases} Q_L & \text{if } d \leq \theta \\ Q_Y & \text{if } d > \theta, \end{cases} \tag{2}$$

where Q_Y is the quantization table proposed in the JPEG standard for Y, and Q_L is the first $L \times L$ sub-matrix of Q_Y. The schematic of the proposed framework is illustrated in Figure 3.

The important parameter of our compression framework is θ. The threshold defines how close H_N and H_Y need to be so that we can consider the corresponding blocks to be similar. Clearly, choosing a lower threshold means less blocks are counted as similar in both N and Y channels. Thus, less high-frequency information is removed from N, which results in a lower compression ratio and higher reconstruction quality of NIR. Hence, to achieve the best performance, we propose to set θ as a function of the decompressed image quality. In the next section, we explain how we derive θ based on our training set.

IV. Experiment

We use a dataset of 227 images [2]. Each image is composed of four channels, R, G, B, and N. More details on capturing RGBN images with current cameras can be found in [14]. Examples of the RGB and NIR channels are shown as pairs in Figure 1.

We randomly separate our dataset into 5 sets of images, and we define 5 sets of experiments accordingly. For each experiment, one fold is used as the testing set and the remaining images are used for training the model. We repeat this process 5 times and report the mean and standard deviation of the bit-rate and peak signal to noise ratio (PSNR).

Our frequency-based (FB) method is applied to each fold test set, where the parameters of the method are learned from the corresponding training data. Like JPEG, the DCT coefficient blocks are 8×8. For the N channel, the first 2×2 components are always coded and the rest of the energy in the block is compared with the energy of the corresponding components in the Y counterpart.

A. Parameter Study

To find the relation between the target quality and the best value for θ, we start by forming pairs of Y and N blocks. The difference between N and Y high-frequency coefficients is first computed from the difference energy:

$$\Delta_H = d_{max} - d_{min}, \tag{3}$$

where d_{max} is the highest and d_{min} is the lowest energy in the dataset. The threshold is then obtained as follows:

$$\theta(\Delta_H) = \alpha \times \Delta_H, \tag{4}$$

where α is the portion of d for similar patches to Δ_H for all patches in the training set.

We varied α in the above equation from 0.01 to 0.075. The values obtained for θ are consistent in different folds. The results of our framework for NIR images with a number of thresholds are presented in Figure 4.

It can be observed that for large thresholds and bit-rates higher than 0.2 bpp, as significant amount of information is removed from NIR, our algorithm does not achieve a PSNR larger than 36 dB. These results suggest that, for low bit-rate compression, our algorithm performs best with larger thresholds. However, if the goal is to compress images for high quality, smaller thresholds achieve better results. Hence, we propose to set the threshold inversely proportional to the target bit-rate. For each bit-rate, we find the threshold that results in the highest PSNR for our training set. The results obtained with $\theta(\Delta_H)$ suggest that an exponential curve represents this relation well. Based on our training set, we obtain:

$$\theta(BR) = a \exp(b \times BR), \tag{5}$$

where $a = 992.3$ and $b = -8.5$ and BR is the target bit-rate.

V. Results

To assess the performance of the proposed algorithm, we present the results of two other methods. The first algorithm
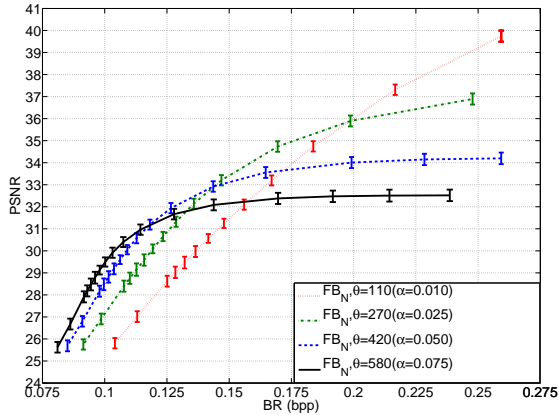
Fig. 4. PSNR results of our method (FB$_N$) with different thresholds ($\theta(\Delta_H)$) versus bit-rate for NIR images.

Fig. 5. PSNR results of JPEG$_N$ and our method (FB$_N$) versus bit-rate for NIR images.

Fig. 6. PCA components of one image in our dataset.

to which we compare our results is the conventional JPEG. The RGB image is encoded with JPEG standard (JPEG$_{RGB}$). The N channel is also encoded with JPEG standard (JPEG$_N$), by using the same quantization table as the Y channel of the RGB image.

Figure 5 lists the PSNR results of two methods for NIR images: JPEG$_N$ and our frequency-based (FB$_N$). The results are reported against various bit-rates from 0.075 bpp to 0.275 bpp. We observe that for the same bit-rate, our proposed method yields significantly better PSNR (up to 5 dB) compared to JPEG$_N$.

Moreover, we compare the performance of our proposed method with a PCA-based compression (PCA$_{RGBN}$). In PCA$_{RGBN}$, we first transform the data into the de-correlating PCA space. The basis vectors of this space are computed from the training data. Each of the transformed channels (see Figure 6) obtained by the PCA is then encoded using the JPEG standard.

The performance of our framework is compared with PCA$_{RGBN}$, as well as JPEG$_{RGBN}$, in Figure 7. This figure shows the error of reconstruction for all four channels in our dataset.

Our algorithm significantly outperforms JPEG$_{RGBN}$. It also achieves significantly better results compared to the PCA$_{RGBN}$ method for compression ratios from 42 to 64 (bitrates from 0.5 to 0.75). This proves our hypothesis that there exist some redundancies between NIR and RGB images, our framework efficiently removes these redundancies and results in the same PSNR with a lower bit-rate.

The improvement achieved by our method can be explained by the amount of information that is correctly removed. In our dataset, on average in ($70\% \pm 5$) of the blocks, the high-frequency of NIR information are removed.

Figure 8 shows two NIR scenes decompressed by FB$_{RGBN}$, JPEG$_{RGBN}$, and PCA$_{RGBN}$. This figure also shows that we achieve the same quality with lower bit-rates.
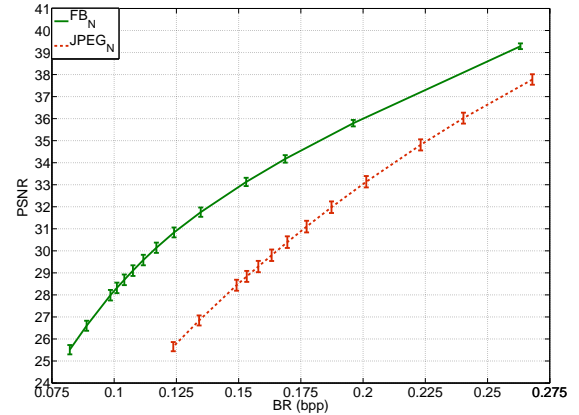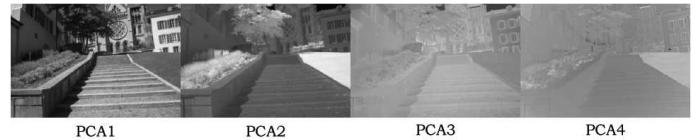
VI. CONCLUSION

We present a framework for compressing four-channel multispectral images composed of RGB and NIR information. We exploit the specific nature of our multispectral images: three channels (RGB) represent the color image and one channel (NIR) represents extra invisible spatial information. The performance of our method is compared to other multispectral compression approaches that first remove spectral redundancy by a PCA transform and then apply the spatial encoding in a transform domain like DCT or DWT.

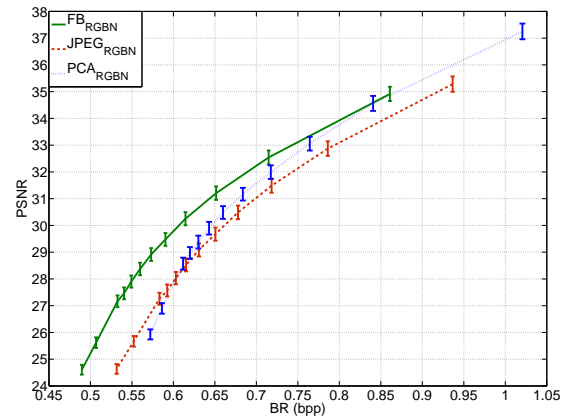Our compression scheme first uses standard JPEG for

Fig. 7. PSNR results of JPEG$_{RGBN}$, our method (FB$_{RGBN}$), and PCA$_{RGBN}$ versus bit-rate for four-channel (RGBN) images.
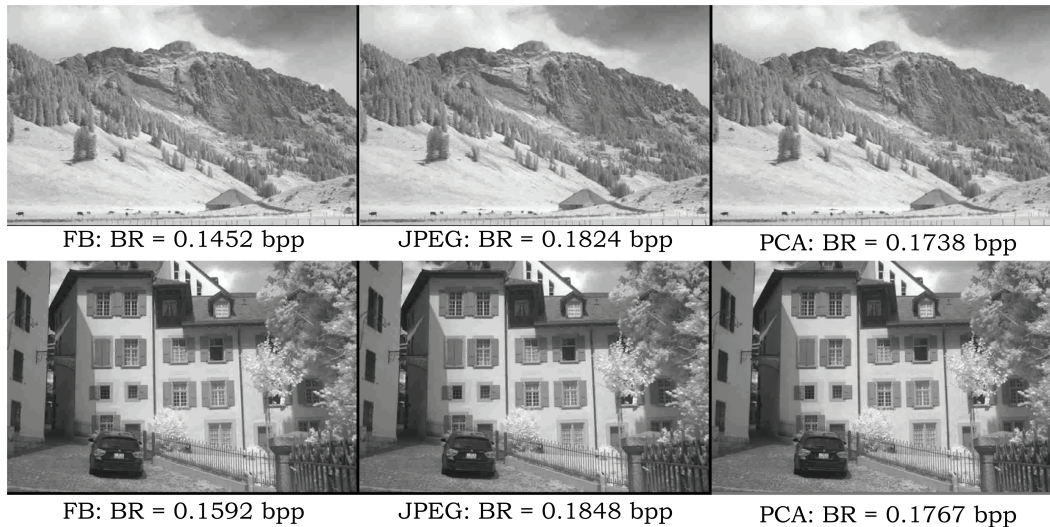
Fig. 8. NIR images decompressed by (from left to right) our method (FB$_{RGBN}$), JPEG$_{RGBN}$, and PCA$_{RGBN}$. The PSNR of all images are the same, while their bit-rates are different. The PSNR is 32.43 dB and 33.21 dB, respectively, for the first and second row.

compressing the three color channels, and then it exploits the strong correlation between high-frequency information in the NIR (N) and the luminance (Y) channel, by removing the spectral redundancy. Thus, one advantage of our approach is that if only the color image is needed, it can be decompressed using a standard JPEG decoder. If the NIR image is desired, only the Y and the N channel need to be decompressed. This is computationally more efficient than using PCA, where all channels always have to be decompressed.

In our algorithm, to compress the N channel that represents the NIR image, first the DCT coefficients of each block are compared to their Y counterparts. In the case of sufficient similarity, we encode only DCT coefficients that represent the low-frequency NIR information. At the decoder, the missing information for the N channel is replaced by the corresponding information in Y. If the information is not similar enough, all coefficients are encoded.

To decide whether or not, for a given block, the high-frequency information of the NIR and Y channels are correlated enough, we set a threshold based on a training dataset. We show that the threshold should be chosen to be inversely proportional to the target bit-rate, and we present a mapping function.

Our experiments show that the proposed compression framework achieves lower bit-rates at the same PSNR for medium and high compression ratios compared to both conventional JPEG and a PCA-based compression method.

REFERENCES

[1] S. Süsstrunk and C. Fredembach, "Enhancing the visible with the invisible: Exploiting near-infrared to advance computational photography and computer vision," in *SID 2010*, vol. 48, 2010.

[2] M. Brown and Süsstrunk, "Multispectral SIFT for scene category recognition," in *CVPR*, 2011.

[3] C. Fredembach and S. Süsstrunk, "Illuminant estimation and detection using near-infrared," in *Proc. of IS&T/SPIE EI: Digital Photography V*, 2009.

[4] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk, "Designing color filter arrays for the joint capture of visible and near-infrared images," in *ICIP*, 2009.

[5] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG 2000 and principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, 2007.

[6] A. Kaarna and J. Parkkinen, "Transform based lossy compression of multispectral images," *Pattern Analysis and Applications*, vol. 33, no. 50, pp. 4–39, 2001.

[7] G. P. Abousleman, M. W. Marcellin, and B. R. Hunt, "Compression of hyperspectral imagery using the 3-D DCT and hybird DPCM/DCT," *IEEE Geoscience and Remote Sensing Letters*, vol. 33, no. 1, pp. 26–34, 1995.

[8] S. E. Qian, A. B. Hollinger, S. Williams, and D. Manak, "Vector quantization using spectral index-based multiple subcodebooks for hyperspectral data compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1183–1190, 2000.

[9] W. B. Pennebaker and J. L. Mitchell, *JPEG still image data compression standard*. Springer, 1993.

[10] D. S. Taubman and M. W. Marcellin, *JPEG2000: image compression fundamentals, standards, and practice*. Kluwer Academic Publishers, 2002.

[11] J. T. Rucker, J. E. Fowler, and N. H. Younan, "JPEG2000 coding strategies for hyperspectral data," in *International Geoscience and Remote Sensing Symposium*, 2005.

[12] C. I. Chang, B. Ramakrishna, J. Wang, and A. Plaza, "Low bit-rate exploitation-based lossy hyperspectral image compression," *Journal of Applied Remote Sensing*, vol. 4, pp. 1–24, 2010.

[13] P. Ready and P. Wintz, "Information extraction, SNR improvement, and data compression in multispectral imagery," *IEEE Transactions on Communications*, vol. 21, no. 10, pp. 1123–1131, 1973.

[14] C. Fredembach and S. Süsstrunk, "Colouring the near-infrared," in *IS&T/SID Color Imaging Conference*, 2008.