MULTISPECTRAL IMAGE FUSION BY SUPER PIXEL STATISTICS

Nati Ofir

ABSTRACT
Multispectral image fusion is a fundamental problem of image processing and remote sensing. This problem is addressed by both classic and deep learning approaches. This paper is focused on the classic solutions that can work in real-time systems and introduces a new novel approach to this group of works. The proposed method carries out multispectral image fusion based on the content of the fused images. Furthermore, it relies on an analysis of the level of information of segmented superpixels in the fused inputs. Specifically, the proposed method addresses the task of visible color RGB to Near-Infrared (NIR) fusion. The RGB image captures the color of the scene while the NIR channel captures details and sees beyond haze and clouds. Since each channel senses different information of the scene, their multispectral fusion is challenging and interesting. Therefore, the proposed method is designed to produce a fusion that contains the relevant content of each spectra. The experiments of this manuscript show that the proposed method is visually informative with respect to other classic fusion methods. Moreover, it can be run fastly on embedded devices without heavy computation requirements.

Index Terms— Multispectral Images, Image Fusion, Near-Infrared.

1. INTRODUCTION
Image fusion is an important task of image processing with a plethora of related works. This work is introducing a new method to fuse multispectral images based on the image content inside its superpixel segmentation. I conducted this research based on the RGB-NIR dataset of [2]. Therefore, this paper is focused on the problem of RGB visible color $0.4 - 0.7 \mu m$ to NIR $0.7 - 2.5 \mu m$ fusion as was described in [14]. Different spectra capture different information of the scene and therefore their fusion is challenging and meaningful. The RGB channel captures the visible color of the image as seen by human eyes while the NIR channel can see beyond the haze, fog, and clouds, and thus captures far details in the image. See Figure 1 for example of the proposed multispectral image fusion. This fusion contains both color information and the structure of the far mountains, which are invisible in the input RGB image.

Modern cameras often have different sensors, thus their fusion is relevant and informative. A preprocessing step of fusion is sensor alignment as described in the multispectral image registration studies [15, 14]. This registration step is not trivial since single-channel alignment [12] methods do not work in the multi-modal case. After the alignment, a fusion of the input images can be carried out by $\alpha$-blending, Principal-Component-Analysis (PCA) blending [7] and spectral blending like Wavelet fusion [6]. An informative fusion can be carried out by Deep-Neural-Networks (DNN) as well [5]. In this paper, I focus on classic approaches for multispectral image fusion. I briefly introduce the theory behind PCA and spectral fusion. Then, I describe my approach which is spatial and based on superpixel segmentation of the input images. I apply a soft mask fusion such that the blending weight is changing between every pixel in the image. This soft mask emphasizes the details in the fused images and thus allows us to see beyond haze, fog, and cloud.

This paper is organized as follows, Section 2 covers the previous related works to this paper. In Section 3 I present a brief overview of existing approaches for classic image fusion that I compare the proposed method to them. In Section 4 I introduce the proposed spatial approach for image fusion based on superpixel content and statistics. Then, Section 5 introduces the proposed method visual results with respect to previous fusion approaches. Finally, I conclude the manuscript in Section 6.

2. PREVIOUS WORK
Multimodal image registration and fusion is a challenging task that was studied in the last decades. It can be addressed by classical computer vision [15] or by deep learning [14]. Each one of these approaches has its own strength and limitations as described in [13, 16]. Image fusion has an application in multispectral images [22] and in medical imaging [8]. Image fusion is also important in processing images from a single modality as in multi-focus [23]. This work is addressing specifically the problem of multispectral image fusion.

Image fusion can be addressed in various technical approaches. Early methods relied on global statistics [7] or on spectral properties like in wavelet fusion [17]. Advanced methods use guided filtering for fusion [11]. Toet et al [19] use a hierarchical approach of fusion. A novel group of works introduced fusion according the Principal Component Analysis (PCA) [10, 18].

Recent work relies on deep neural networks. Chen et
Fig. 1. Example of the proposed fusion method results. From left to right: input RGB image, the proposed fusion result, input NIR image. The proposed fusion contains both the color information of the RGB and the far details captured by the NIR image.

al [5] use deep fusion to improve classification accuracy. Multispectral fusion was used to detect pedestrians with convolutional-neural-networks [20]. The idea of using superpixel for hyperspectral fusion is introduced in the recent work of [9]. The proposed method of this paper introduces a method that incorporates low-level statistics of image superpixels. Due to its simplicity, it can run fast on simple devices and produces meaningful multispectral image fusion. In contrast to previous works, I am focused on the multispectral case, and limit the solution to a real-time consumption given standard compute capabilities.

3. REVIEW OF CLASSIC FUSION METHODS

3.1. The $\alpha$ Blending Fusion

In this section, I will explain the theory behind several classic image fusion which forms a foundation for the proposed approach of this paper. The most basic image fusion technique is the global $\alpha$ blending as follows. Given multispectral image $NIR, V$ that are NIR and visible color images respectively I first convert them to grayscale such that:

$$I_1 = NIR, I_2 = G = \text{gray}(V).$$

(1)

Then given a constant $\alpha \in [0, 1]$ the $\alpha$ blending fusion is:

$$F_{\text{gray}} = \alpha \cdot I_1 + (1 - \alpha) \cdot I_2.$$  

(2)

Finally, the colored fusion will use a color preservation formula as follows:

$$F = \frac{F_{\text{gray}}}{G} \cdot V.$$  

(3)

3.2. PCA Fusion

The PCA fusion computes $\alpha$ according to the joint statistics of the fused images as follows. Given images $I_1, I_2$ as before I define the variance for every image,

$$V_j = \text{Var}(I_j(x)),$$  

(4)

And the joint covariance,

$$C = \text{cov}(I_1(x), I_2(x))$$  

(5)

The PCA decomposition is applied by computing the eigenvalues and eigenvectors of their covariance matrix:

$$|\lambda I - \begin{pmatrix} V_1 & C \\ C & V_2 \end{pmatrix}| = (\lambda - V_1)(\lambda - V_2) - C^2$$  

(6)

Then by demanding that this equation equals zero: $\lambda^2 - \lambda(V_1 + V_2) + V_1V_2 - C^2 = 0$ I can compute the discriminant:

$$\Sigma = \sqrt{(V_1 + V_2)^2 - 4 \cdot V_1V_2 - 4C^2},$$  

(7)

i.e. $\Sigma = \sqrt{V_1^2 + V_2^2 - 2 \cdot V_1V_2 - 4C^2}$. Then, the eigenvalues are:

$$\lambda_{1,2} = \frac{V_1 + V_2 \pm \Sigma}{2},$$  

(8)

The eigenvector is:

$$v_{\lambda_1} = \left[1, \frac{\lambda_1 - V_1}{C}\right]^T.$$  

(9)

And finally the $\alpha_{PCA}$ blending factor is:

$$\alpha_{PCA} = ||v_{\lambda_1}||^{-1}.$$  

(10)

To conclude, I proved how to perform multispectral image fusion based on global PCA coefficients. In the experiment Section 5 I compare the region content proposed method to this approach as a baseline for image fusion.

3.3. Spectral Fusion

Commonly used methods for fusing images by a local approach are spectral methods. Examples can be seen in [13] or in the wavelets fusion [17]. These methods can be expressed in a very general way as follows. The low pass filter of an image is computed by:

$$LP_j = I_j \ast \text{Gaussian}(x, y).$$  

(11)

Then, the high pass filter is calculated as follows:

$$HP_j = I_j - LP_j.$$  

(12)
Finally, I treat every band pass by different fusion mechanisms, for the low frequencies I apply $\alpha$ blending:

$$F_{LP} = \alpha \cdot LP_1 + (1 - \alpha) \cdot LP_2.$$  \hspace{1cm} (13)

For the high frequencies a maximum criterion is applied:

$$F_{HP}(x) = \max(HP_1(x), HP_2(x)).$$  \hspace{1cm} (14)

Finally, I sum up the different frequencies to derive the overall fusion:

$$F_{gray} = F_{LP} + F_{HP}.$$  \hspace{1cm} (15)

To conclude, I showed how to apply a simple spectral fusion. This idea forms the basics for the proposed content-based fusion. The region content fusion is also emphasizing high frequencies like in spectral fusion, however, it computes the gain according to local superpixel, and it does so by pixel-wise soft maps.

4. THE PROPOSED METHOD: SUPER PIXEL REGION FUSION

In this section, I introduce the proposed approach of multispectral images fusion by region content analysis. It utilizes inspiration from the two previous methods of PCA fusion and spectral fusion. The approach is to fuse the two input images based on a soft smooth pixel map:

$$F_{gray}(x) = mask(x) \cdot I_1(x) + (1 - mask(x)) \cdot I_2(x).$$  \hspace{1cm} (16)

This fusion is done in a hierarchical method using the Laplacian Pyramids. In the next steps, I will explain how to find a smooth $mask(x, y)$ that emphasizes the content in the fused images according to the superpixel region content.

Given an input image $I$ such that pixels of $I$ are $\{x_1, ..., x_N\}$.

The proposed method finds a superpixel decomposition of the pixels such that:

$$X = \{x_1, ..., x_N\} = \bigcup_{j} X_j.$$  \hspace{1cm} (17)

For every super pixel $X_j$, I compute the simple statistics in the input image as follows:

$$grade(X_j) = std(X_j) = \sqrt{E_{x \in X_j} [I^2(x)] - E_{x \in X_j} [I(x)]^2}.$$  \hspace{1cm} (18)

This standard deviation statistic has a significant correlation with the level of content in the image superpixel. Note that other statistics can be incorporated easily to the grade of the proposed method like Entropy and Variance. Then the grade of a pixel in one of the input fused images is the grade of its superpixel:

$$\forall x \in X_j : grade(x) = grade(X_j).$$  \hspace{1cm} (19)

See Figure 2 for illustration of the pixel grades for the NIR and Gray visible images. Finally I compute the pixel level mask of the fusion by the following formula:

$$mask(x) = \sigma[grade(I_1(x)) - grade(I_2(x))].$$  \hspace{1cm} (20)

The sigmoid $\sigma(x) = \frac{e^x}{1 + e^x}$ is used to emphasize the information of where in the fused images the content is. Finally, I normalize $mask(x)$ linearly such that it lies in the range of $[0, 1]$ and apply a Gaussian smoothing on its values. See Figure 3 for an example of a fusion mask of the proposed method of superpixel fusion.

Fig. 2. The grades that are computed by the proposed method on the input NIR and RGB images in Figure 1. Dark indicates super pixels with low level of information.

Fig. 3. The mask of the pixel blending that is computed by the proposed method for the input NIR and RGB images in Figure 1. Dark indicate high weight for the color visible RGB.

To conclude this section, I explained the theory and algorithmic steps behind the main contribution of the paper which is a multispectral image fusion based on the superpixel statistic content.
5. RESULTS

The proposed approach is tested and evaluated on the NIR-RGB multispectral dataset of [2]. The proposed approach produces image fusion that is meaningful, i.e., contains the color information of the RGB together with the far details of the NIR images. Therefore, I evaluate the proposed method by visual examples and I compare it to the well-known PCA fusion [21] and to spectral fusion [15]. Moreover, I measured its performance by the edge preservation property of the multispectral image fusion. It seems to produce the best edge preserving fusion out of all the classic approaches for multispectral fusion that are discussed in this manuscript.

![Image](image_url)

**Fig. 4.** Examples of the proposed multispectral image fusion results. From left to right: input RGB, the proposed fusion, input NIR.

| Image | SuperPixel | PCA   | Spectral |
|-------|------------|-------|----------|
| 1     | 35.91      | 15.75 | 24.10    |
| 2     | 39.67      | 34.38 | 33.99    |
| 3     | 21.61      | 20.72 | 19.78    |
| 4     | 32.41      | 22.43 | 21.68    |
| Average | 32.40     | 23.32 | 24.89    |

**Table 1.** Percentage of Canny [4] edges from the input images that remain in their fusion of Figure 5. Note that the proposed super-pixel method preserves the edges in the highest level among the other classic approaches for image fusion.

Figure 4 shows examples of real multispectral images in the proposed fusion results. It can be seen that the fused images show details behind the haze, fog, and clouds that are forming obstacles in the RGB input image. It can be seen that although this fusion mask is local, I achieve a smooth and neat fused image. Figure 5 compares the local approach to the global PCA fusion. The proposed fusion is more informative and combines information from both spectral channels. The far mountains can be seen only by this method, and not by the PCA approach. That forms proof that local weighting of pixel fusion, is an advantage when it is done correctly. Table 1 shows that the proposed method achieves the highest percentage of Canny [4] edge preservation in the fusion image. This is another indication that the super-pixel approach emphasizes the details in the input multispectral images.

6. CONCLUSIONS

I introduced a new method for multispectral fusion which applies a spatial soft map based on input image superpixel segmentation. This method shows advantages over existing approaches such that the details in the input images are preserved better in the fusion result, and still, the information of the color remains valid. In addition, I explained the theory behind the PCA and spectral fusion techniques and compared their principles to the proposed approach. As a whole, this paper produces an informative research work on the interesting problem of multispectral image fusion, in the color RGB to NIR domain.

7. REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
[2] M. Brown and S. Süssstrunk. Multispectral SIFT for scene category recognition. In Computer Vision and Pattern Recognition (CVPR11), pages 177–184, Colorado Springs, June 2011.

[3] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. In Readings in computer vision, pages 671–679. Elsevier, 1987.

[4] J. Canny. A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6):679–698, 1986.

[5] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu. Deep fusion of remote sensing data for accurate classification. IEEE Geoscience and Remote Sensing Letters, 14(8):1253–1257, 2017.

[6] L. J. Chipman, T. M. Orr, and L. N. Graham. Wavelets and image fusion. In Proceedings., International Conference on Image Processing, volume 3, pages 248–251. IEEE, 1995.

[7] C. He, Q. Liu, H. Li, and H. Wang. Multimodal medical image fusion based on ihs and pca. Procedia Engineering, 7:280–285, 2010.

[8] A. P. James and B. V. Dasarathy. Medical image fusion: A survey of the state of the art. Information fusion, 19:4–19, 2014.

[9] S. Jia, X. Deng, J. Zhu, M. Xu, J. Zhou, and X. Jia. Collaborative representation-based multiscale superpixel fusion for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 57(10):7770–7784, 2019.

[10] S. S. Kumar and S. Muttan. Pca-based image fusion. In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII, volume 6233, page 62331T. International Society for Optics and Photonics, 2006.

[11] S. Li, X. Kang, and J. Hu. Image fusion with guided filtering. IEEE Transactions on Image processing, 22(7):2864–2875, 2013.

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

[13] N. Ofir and J.-C. Nebel. Classic versus deep approaches to address computer vision challenges. arXiv preprint arXiv:2101.09744, 2021.

[14] N. Ofir, S. Silberstein, H. Levi, D. Rozenbaum, Y. Keller, and S. D. Bar. Deep multi-spectal registration using invariant descriptor learning. In 2018 25th IEEE International Conference on Image Processing (ICIP), pages 1238–1242. IEEE, 2018.

[15] N. Ofir, S. Silberstein, D. Rozenbaum, Y. Keller, and S. D. Bar. Registration and fusion of multi-spectral images using a novel edge descriptor. In 2018 25th IEEE International Conference on Image Processing (ICIP), pages 1857–1861. IEEE, 2018.

[16] Y. N. Ofir. Classic versus deep learning approaches to address computer vision challenges: a study of faint edge detection and multispectral image registration. PhD thesis, Kingston University, 2021.

[17] G. Pajares and J. M. De La Cruz. A wavelet-based image fusion tutorial. Pattern recognition, 37(9):1855–1872, 2004.

[18] A. Srivastava, V. Bhateja, and A. Moin. Combination of pca and contourlets for multispectral image fusion. In Proceedings of the international conference on data engineering and communication technology, pages 577–585. Springer, 2017.

[19] A. Toet. Hierarchical image fusion. Machine Vision and Applications, 3(1):1–11, 1990.

[20] J. Wagner, V. Fischer, M. Herman, S. Behnke, et al. Multispectral pedestrian detection using deep fusion convolutional neural networks. In ESANN, volume 587, pages 509–514, 2016.

[21] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. IEEE signal processing magazine, 26(1):98–117, 2009.

[22] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 53(7):3658–3668, 2015.

[23] W. Zhao, D. Wang, and H. Lu. Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network. IEEE Transactions on Circuits and Systems for Video Technology, 29(4):1102–1115, 2018.