Hyperspectral and Multispectral Image Fusion Using 3D Wavelet Transforms

Eduardo Rittner Coelho, Ana Clara C. Silveira, Tarik P. e Sá, Diedre S. do Carmo, Paula D. P. Costa
Letícia Rittner, Denis Gustavo Fantinato
School of Electrical and Computer Engineering, Universidade Estadual de Campinas
Campinas, SP, Brazil

***Abstract*—The fusion of multispectral (MSI) and hyperspectral (HSI) images is a crucial technique in various fields such as remote sensing, medical imaging, and agricultural monitoring. MSI captures light across several specific spectral bands, while HSI provides detailed spectral information across contiguous bands. Combining these two types of images leverages the high spatial resolution of MSI and the rich spectral content of HSI, creating a single, high-resolution image that is both spatially and spectrally informative. Traditional wavelet based fusion methods often employ a single wavelet across all dimensions, which can result in suboptimal outcomes due to the different characteristics of spatial and spectral data. This paper explores the use of 3D wavelet transforms with varied wavelets across dimensions to improve the fusion process. Experiments conducted on the ICASSP HyperSkin Challenge dataset showed that a combination of Daubechies on the spatial dimensions and Coiflets on the spectral dimension obtained higher fidelity and SSIM when compared to simpler fusion methods.**

I. Introduction

Multispectral Images (MSI) devices capture light in a number of different frequencies across the electromagnetic spectrum, while Hyperspectral Images (HSI) devices capture light in a large number of equidistant frequencies across the eletromagnetic spectrum [1]. These images are already widespread throughout many applications, such as remote sensing [2], medical imaging [3] and crop quality assessment [4], and have been gaining traction in other contexts as well. This is mainly due to the fact that MSIs and HSIs are intrinsically more informative than RGB images, which contain only 3 bands in the specific wavelengths (associated to red, green and blue color), and the additional information they contain has proven to be highly valuable [1].

However, the higher spatial resolution provided by HSIs is also associated with elevated cost in acquisition cameras, which is higher when compared to their MSI counterparts [5]. A common strategy for cost reduction is to use a very low framerate (below 0.1 Hz) or to employ moving sensors (point and line-scan), which may introduce noise in images [6].

In this context, a common alternative to using a single high-cost high-resolution camera is to use a low spatial resolution HSI camera alongside a high resolution MSI camera to register a scene, and afterwards merge both images, ideally preserving all the relevant information. However, working with two distinct images of a single subject is not an easy task. Thus, MSI-HSI fusion is used to combine high quality spatial information from the MSI image with high quality spectral information from the HSI into one high resolution HSI, which can then be used, modified, analyzed and handled as one single image much more efficiently.

The process of fusing MSI and HSI images to create a single high-resolution image is a crucial step in maximizing the benefits of both techniques. This fusion process leverages the detailed spatial resolution of MSI with the rich spectral information provided by HSI. Several techniques have been developed to perform MSI/HSI fusion, ranging from simple interpolation methods to more sophisticated approaches involving machine learning and advanced signal processing [7].

One of the promising approaches in this domain is the use of wavelet transforms [8]. Wavelet-based fusion methods have been extensively studied due to their ability to decompose images into different frequency components, making it easier to merge detailed spatial and spectral information. Wavelet transforms can handle multi-resolution analysis, which is particularly useful in the context of image fusion, as they allow for the efficient combination of high-frequency details from MSI with the low-frequency spectral content from HSI. Despite the advantages, there are challenges associated with wavelet-based fusion techniques, particularly in the selection of appropriate wavelets that can effectively represent the information in different dimensions [9].

Most traditional approaches use a single wavelet for all dimensions, which may not capture the distinct characteristics of the spatial and spectral data adequately. This limitation can result in suboptimal fusion results, where critical details may be lost or improperly merged. In light of this, in this paper, we analyze the application of 3D wavelet transforms with the flexibility to vary wavelets across different dimensions. By combining the wavelets to better match the unique properties of the spatial and spectral information, higher-quality images can be achieved in the fusion process. Our experiments show this approach is able to improve the fidelity of the fused images under the perspective of quality measures, such as SSIM.

II. HSI/MSI Fusion

Image fusion is generally defined as collecting all the important information from multiple input images and combining it into one single output image. The objective is to obtain one single image, more informational than each of the input images, preserving as much as possible all the relevant information from all the input (Fig. 1).

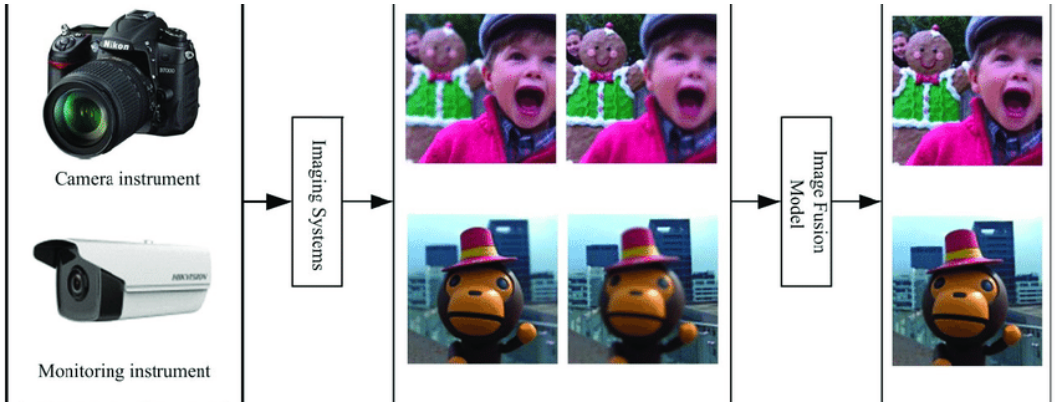

Fig. 1. Different examples of image fusion. Adapted from [10]

In the context of fusing MSI and HSI images, the resulting HSI should have complimentary spatial and spectral resolution characteristics. The resulting HSI ideally retains all the spatial information contained in the MSI image and all the spectral information contained in the HSI, as well as discard any redundant information (Fig. 2). There are many methods currently in use for MSI-HSI fusion, such as high-pass filtering, IHS, PCA and Discrete Wavelet Transform (DWT) methods [11]. In this work, we focus on the last one.

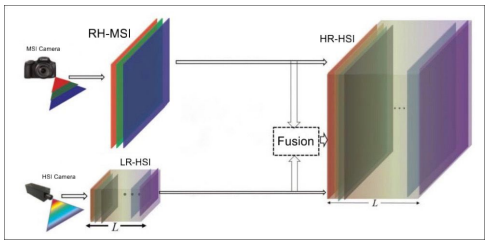

Fig. 2. MSI-HSI Image fusion visualization

III. Wavelet-Based Fusion

Wavelets are mathematical functions that contain certain characteristics which make them generally useful in signal processing applications via the Wavelet Transform [8]. It can be thought of as a more general application of the Fourier transform, wherein any arbitrary wavelet can be used to transform the incoming signal, instead of only sine and cosine waves. More specifically, wavelets are localized wave-like functions that exist for a finite duration, this property allows them to decompose signals at different scales, revealing both local and global level details, in other words, both frequency and time domain information.

The DWT operates on a finite quantity of scales and translations of a wavelet function and can be applied in any number of dimensions, and more importantly, can be applied with different wavelets and levels of decomposition in each dimension, which can be very useful for data where different dimensions have different overall characteristics. The most common multilevel DWT approach is to decompose each axis with a high-pass (H) and low-pass (L) wavelet, and then only decompose the approximation subband at each subsequent level (Fig. 3). The approximation subband is the subband comprised of only low-pass decomposed axes.

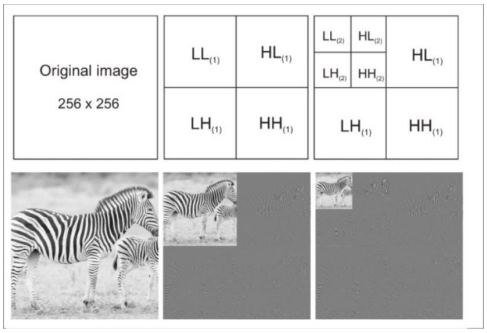

Fig. 3. 2 level 2D DWT on an image

A. Classic Wavelets

The classical wavelets are split into different families, each with its own distinct characteristics, shapes and applications. The simplest wavelet is the Haar [12], which resembles a step function, being very good at detecting edges, but not much else. The Daubechies are a family of wavelets where each wavelet is denoted dbN and N is the number of vanishing moments ($db1$ is exactly equal to the Haar wavelet). They are versatile and suited to decompose both edges and continuous signals, making them a good candidate for visual applications. Similarly, the coiflets are a family of numbered wavelets with $N/3$ vanishing moments and have shown to be better than the Daubechies in some specific image applications [13], and due to their smoother curve may be better at capturing signals without edges [13]. While there are other wavelet families, the latter two were chosen based on their prevalence in MSI/HSI literature. These wavelet families can be used in 2D DWTs as well as higher dimensions.

B. 3D Wavelets

The 3D DWT is an extension of the 2D DWT to one more dimension, which in our case would be the spectral axis.

The n level 3D DWT produces $7n + 1$ subbands, which all combined have the same shape as the original image (Fig. 4). Although usual, 3D DWT does not necessarily decompose all 3 axes with the same wavelet or level, and this can potentially be exploited to produce better results, consume less computational power, or discard unnecessary information. For example, different wavelets can be chosen for different axes whose data is previously known to differ in some way, so for axis 1 we may choose a wavelet that is good at detecting edges, while for axis 2 we may choose a wavelet that does better at continuous signals. Similarly, different level decompositions can be applied to different axes, so if axis 1 is generally smaller and has less information than axis 2, we may choose to decompose it at a lower level, since a higher level decomposition in this case would not necessarily obtain more information.

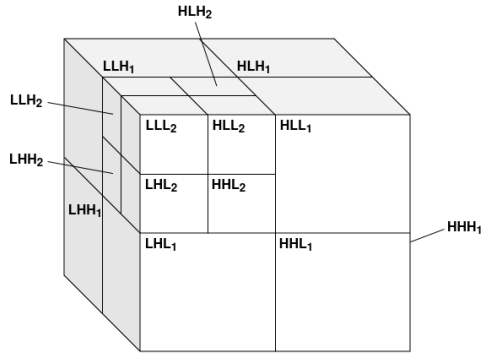

Fig. 4. Iluustration of a 2 level 3D DWT subbands

C. 3D Heterogeneous Wavelet-Based Fusion

Wavelet-based MSI Fusion takes advantage of the flexibility provided by the DWT to extract information specific to the spatial domain and frequency domain. The idea is that by decomposing the 2 input images via DWT, we can pick and choose only the relevant information from each image [9]. So, from the input image with more bands and lesser resolution we extract the high-level frequency detail, and from the input image with less bands and higher resolution we extract the high-level spatial detail. Thus, the resulting fused image should retain both the high-level frequency information and high-level spatial information. Another benefit of using the DWT to fuse HSIs is that different axes can be decomposed by distinct wavelets, which means that it is possible to pick a wavelet function more suited to the specific domain it's being used on. Therefore, it may be beneficial to choose a wavelet that is able to retain more edges, and sharper information for the spatial domain, and a smoother wavelet for the spectral domain.

Based on this, the proposed method can be stated as follows: given two input images of varying shapes, we first resample them naively to the same shape, by upsampling the number of bands of the MSI image to the same number of bands in the HSI, and upsampling the resolution of the HSI to the same resolution of the MSI image. This resampling is done by simply repeating every pixel to achieve the desired resolution, and repeating every band to achieve the desired number. When both images have the same shape, the DWT is applied on both images with a given set of wavelet functions. For each pair of DWT coefficient subbands, we either choose one and discard the other, or get their average. Given n-level 3D DWT coefficients, let HHH_j be the wavelet coefficient subbands relative to the j-level high-pass (H) or low-pass (L) decomposition in all axes. The 3D wavelet fusion method is described as follows: For every subband with high frequency spatial information and low frequency spectral information (HHL_j, LHL_j, HLL_j), take the corresponding MSI subbands and discard the HSI ones. For every subband with high frequency spectral information and low frequency spatial information (LLH_j), take the corresponding HSI subbands and discard the MSI ones. For every other subbands which contain equal levels of spatial and spectral information, take the average between the MSI and HSI subbands. In doing so, the resulting image will have the high frequency spatial information from the MSI image and the high frequency spectral information from the HSI, while simultaneously discarding the high frequency spatial information in the HSI and high frequency spectral information in the MSI, which should in theory have a net positive effect on the fused image.

IV. HSI/MSI Dataset

The dataset [14] was originally developed for the ICASSP Hyperskin Challenge, featuring 264 HSIs and MSIs of 44 human faces in 6 orientations each. HSIs were obtained using a Specim FX10 camera covering 448 spectral bands from 400nm to 1000nm. The Specim FX10 is a pushbroom camera and had to be moved using a customized rig, scanning each line took approximately 22ms, and a whole 1024x1024 image took about 22.7 seconds. Along with the full 62 band HSI, the dataset comes with a 4 band MSI (RGB + NIR), used as the input for the challenge the dataset was designed for. The NIR band was captured at 960nm and RGB bands were obtained converting the 62 MSI channels into RGB using the RGB camera wavelength response [14].

The dataset utilized for testing the fusion methods described in this paper has some distinctions to the original dataset. The dataset was formatted in .hdf format, split into training and validation images, each image having 66 bands (61 HSI bands + 4 MSI bands). The training and validation distinction was not relevant to the fusion method, so they were abolished, and from each original image three images were derived: a 1024x1024 MSI (input), a 1024x1024 HSI (target) and a 256x256 HSI (input, obtained by skipping 3 pixels out of every 4). The 35_{th} band was discarded since it was a duplicate of the 36_{th}.

V. Experiments

All the experiments were completed in a virtual environment, and the code is available online[1]. All of them used

[1]https://github.com/eduardorittner/hsi_fusion

resampled input images so they would have the same shape, and the resulting fused image was then compared against the original high-resolution HSI to compute the metrics. Two baseline methods were considered: The first took a simple average between both input images, and the second consisted of a 2D-DWT applied individually to each band, the coefficients were then fused with the following fusion rules: a simple average between the approximation coefficients, and on all the other subbands, the HSI coefficients were discarded, and only the MSI coefficients were considered. Other than that, all other experiments were done with the 3D-DWT and same fusion rules and decomposition level, only changing the wavelets used in the DWT.

A. Metrics

Fusion results were evaluated based solely on the Structural Similarity Index (SSIM) from Skimage [2], which measures the similarity bewteen 2 images taking into account luminosity, contrast and structure [15]. Since the images in the dataset contained a substantial amount of background noise, the SSIM was evaluated only on the face of the subjects, by using the masks produced by the ICASSP challenge RainbowAI team [16] using a Segment Anything Model (SAM). Therefore, reported results do not reflect the actual value of the metrics, and they are only useful in relation to one another. In other words, only the difference between reported metrics is important.

B. Results

Results show that the 3D-DWT based fusion methods always produce higher SSIM values than the baseline methods, which indicates an overall higher level of overall image fidelity and structural similarity (Tab. I). The overall best scores were from DWTs performed with Daubechies on the spatial dimensions and Coiflets on the spectral dimension, which corroborates the initial hypothesis that Daubechies wavelets are better suited for spatial information, which generally contains sharp edges and high-frequency information, while Coiflets do better on spectral information, which has a smoother profile overall.

VI. Conclusion

The proposed 3D-DWT fusion scheme for MSI-HSI fusion shows promising results when compared to other simple methods. Wavelets are a natural fit for HSI images, specifically the pairing of Daubechies for the spatial domain and Coiflets for the spectral domain. It is also of interest for future work to ascertain whether the DWT can be used alongside more sophisticated deep-learning based methods to provide even better results.

Acknowledgment

This research was funded by the ELDORADO Institute and by grants 2020/09838-0 and 2024/08388-2, São Paulo Research Foundation (FAPESP).

[2]https://scikit-image.org/

TABLE I
Fusion Results

	Method	SSIM
Baseline	Average	0.87 ± 0.02
	2D-DWT	0.89 ± 0.05
3D Wavelets	Haar	0.91 ± 0.02
	Db2	0.93 ± 0.05
	Db4	0.92 ± 0.05
	Db8	0.92 ± 0.05
	Coif1	0.92 ± 0.05
	Coif2	0.92 ± 0.05
	Db1-Db1-Coif4	0.93 ± 0.05
	Db2-Db2-Coif2	$\mathbf{0.94 \pm 0.05}$
	Db4-Db4-Coif4	0.93 ± 0.05
	Coif2-Coif2-Db2	0.92 ± 0.05

References

[1] J. M. Amigo, "Chapter 1.1 - Hyperspectral and multispectral imaging: Setting the cene," in *Hyperspectral Imaging*, ser. Data Handling in Science and Technology, J. M. Amigo, Ed. Elsevier, 2019, vol. 32, pp. 3–16.

[2] J. Gonzalez-Ibarzabal, M. Franquesa, A. Rodriguez-Montellano, and A. Bastarrika, "Sentinel-2 reference fire perimeters for the assessment of burned area products over latin america and the caribbean for the year 2019," *Remote Sensing*, vol. 16, no. 7, 2024.

[3] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, Jan. 2014.

[4] F. Liu, R. Yang, R. Chen, M. Guindo, Y. He, J. Zhou, X. Lu, M. Chen, Y. Yang, and W. Kong, "Digital techniques and trends for seed phenotyping using optical sensors," *Journal of Advanced Research*, 2023.

[5] J. Y. H. Raju Shrestha, "Evaluation and comparison of multispectral imaging systems," in *22nd Color Imag. Conf.*, 2014, pp. 107–112.

[6] S. Zhang, X. Kang, Y. Mo, and S. Li, "Noise analysis of hyperspectral images captured by different sensors," in *2020 IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS 2020)*, 2020, pp. 2707–2710.

[7] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Information Fusion*, vol. 69, pp. 40–51, 2021.

[8] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques — an introduction, review and comparison," *ISPRS Journal of Photog. and Remote Sensing*, vol. 62, no. 4, pp. 249–263, 2007.

[9] Y. Zhang and M. He, "3d wavelet transform and its application in multispectral and hyperspectral image fusion," in *2009 4th IEEE Conference on Industrial Electronics and Applications*, 2009, pp. 3643–3647.

[10] L. Li and H. Ma, "Pulse coupled neural network-based multimodal medical image fusion via guided filtering and wseml in nsct domain," *Entropy*, vol. 23, p. 591, 05 2021.

[11] S. Karim, G. Tong, J. Li, A. Qadir, U. Farooq, and Y. Yu, "Current advances and future perspectives of image fusion: A comprehensive review," *Information Fusion*, vol. 90, pp. 185–217, 2023.

[12] C. K. Chui, *An introduction to wavelets*. Elsevier, 2014.

[13] M. Srivastava, Y. Yashu, S. K. Singh, and P. K. Panigrahi, "Multisegmentation through wavelets: Comparing the efficacy of daubechies vs coiflets," *arXiv preprint arXiv:1207.5007*, 2012.

[14] B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat, Y. Cai, and et al., "NTIRE 2022 spectral recovery challenge and data set," in *2022 IEEE Conf. Comp. Vis. Patt. Recog. Workshops (CVPRW)*, 2022, pp. 862–880.

[15] 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*, vol. 26, no. 1, pp. 98–117, 2009.

[16] A. C. C. Silveira, D. S. do Carmo, L. H. Ueda, D. G. Fantinato, P. D. P. Costa, and L. Rittner, "Vision transformer MST++: Efficient hyperspectral skin reconstruction," in *IEEE ICASSP 2024*, 2024.