

PANSHARPENING ON THE NARROW VNIR AND SWIR SPECTRAL BANDS OF SENTINEL-2

A. D. Vaiopoulos and K. Karantzas

Remote Sensing Laboratory, National Technical University of Athens
HeronPolytechniou 9, Zographos, 15780, Greece -
arisvg@central.ntua.gr karank@central.ntua.gr

Commission VII, WG VII/6

KEY WORDS: Fusion, Benchmark, Image quality indexes, Validation, QNR, Q4, UIQI

ABSTRACT:

In this paper results from the evaluation of several state-of-the-art pansharpening techniques are presented for the VNIR and SWIR bands of Sentinel-2. A procedure for the pansharpening is also proposed which aims at respecting the closest spectral similarities between the higher and lower resolution bands. The evaluation included 21 different fusion algorithms and three evaluation frameworks based both on standard quantitative image similarity indexes and qualitative evaluation from remote sensing experts. The overall analysis of the evaluation results indicated that remote sensing experts disagreed with the outcomes and method ranking from the quantitative assessment. The employed image quality similarity indexes and quantitative evaluation framework based on both high and reduced resolution data from the literature didn't manage to highlight/evaluate mainly the spatial information that was injected to the lower resolution images. Regarding the SWIR bands none of the methods managed to deliver significantly better results than a standard bicubic interpolation on the original low resolution bands.

1. INTRODUCTION

Fusing effectively spatial and spectral information from different image modalities is a critical and valuable tool for numerous applications in geoscience, remote sensing, image analysis and computer vision. Among the several fusion techniques, pansharpening is a critical one, which focused on the injection of spatial information, extracted from a high resolution panchromatic (PAN) band, to other, with lower spatial resolution, multispectral (MS) ones.

The performance of pansharpening is of significant importance since currently most moderate to very high spatial resolution satellite sensors typically include at the same imaging system both higher and lower resolution spectral bands. Therefore, early research efforts which employed LANDSAT and SPOT satellite imagery focused on defining efficient quantitative evaluation tools towards deciding among several techniques for the optimal one [Gillespie et al., 1987, Chavez et al., 1991, Wald et al., 1997].

Most pansharpening methods can be classified into those which are based on (i) component substitution and (ii) multi-resolution analysis [Alparone et al., 2007, Vivone et al., 2015]. Methods based on component substitution [Gillespie et al., 1987, Garzelli et al., 2008, Choi et al., 2011, Zhang and Roy, 2016] try to decompose the spatial structure and spectral information through an efficient transformation. Methods based on the multi-resolution analysis are focusing on defining the optimal way the missing highpass information will be injected on the lower resolution image [Chavez et al., 1991, Otazu et al., 2005, Aiuzzi et al., 2006, Vivone et al., 2014].

A recent comprehensive evaluation [Vivone et al., 2015] among several state-of-the-art methods indicated that the same algorithms may score differently on different validation

frameworks. Two evaluation frameworks were considered: analysis (i) at reduced and (ii) at full resolutions. The first one employs the original image as a reference, whereas during the second one specialized indexes *e.g.*, QNR is employed. Component substitution methods can address aliasing problems and generally overcome misregistration problems. Methods based on multi-resolution analysis resulted into very good overall performances, while can be employed when multisensor data are considered due to their temporal coherence.

In this paper, the goal was to establish a framework and evaluate several methods for the pansharpening of the VNIR bands of Sentinel-2. The evaluation, also, include the SWIR bands and this was mainly because certain studies have also considered pansharpening multi-spectral bands that do not overlap spectrally with a panchromatic band [Vivone et al., 2015, Garzelli, 2015]. In contrast, what it is usually performed is that if the panchromatic band is spectrally overlapping with several of the multi-spectral bands then the multi-spectral bands may be pansharpened to provide a panchromatic spatial resolution equivalent.

2. MATERIALS AND METHODS

2.1 Description of Datasets

The Sentinel-2 raw datasets were collected at 2015/12/26. Based on the available Sentinel-2 Toolbox the Bottom-Of-Atmosphere (BOA) surface reflectance was computed. In particular, atmospheric corrections were applied to the Level-1C product (Top of Atmosphere, TOA) and consisted of two main parts: (i) Scene Classification which aims at providing a pixel classification map with classes like cloud, cloud shadows, vegetation, soils/deserts, water, snow, *etc.* and (ii) Atmospheric Correction aims at transforming TOA reflectance into BOA reflectance.

Three sub-regions were selected for the experiments as presented in Figure 1. The main objective for the selection was to contain a broad variety of land cover classes.

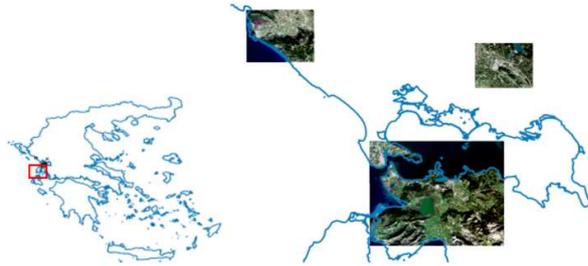


Figure 1. Relative location and size of the Sentinel-2 datasets of the three study areas. A natural colour composite is displayed (RGB B4-B3-B2).

The dimensions (in pixels) of the three images are the following ones for each area:

- a) Area 1: 1200x1200
- b) Area 2: 1024x1024
- c) Area 3: 2400x2400

The above image sizes refer to the size of the high resolution raw Sentinel-2 bands (10 m) which control, as well, the size for the Full Resolution experiment (section 2.1.2).

2.1.1 Benchmark structure and qualitative assessment

Two set of datasets were employed during our experiments: a) the initial ones (VNIR and SWIR bands at their original resolution) called from now on Full Resolution (FR) and b) a set of data which resulted after downsampling the original ones, called from now on Reduced Resolution (RR).

During the FR experiments the index QNR was calculated for evaluation purposes, while on the RR experiment the standard index Q [Wang and Bovik, 2002] was more suitable since the original multispectral image could serve as a reference image. Moreover, the Q4 index was calculated for the VNIR bands which forms a vector generalization of the standard Q accounting also for spectral distortion [Garzelli and Nencini, 2009, Vivone et al., 2015].

Figure 2 graphically describes the methodology followed in order to inject spatial information to the narrow VNIR spectral bands (B5-B8a) of Sentinel-2.

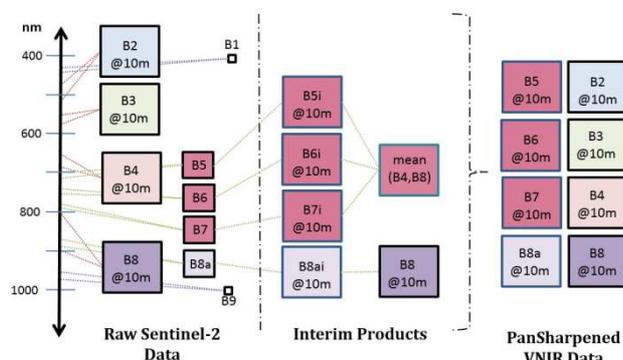


Figure 2. The structure of pansharpening procedure for the VNIR bands of Sentinel-2.

The first step of the procedure was to increase the resolution of the 20 m bands, using cubic interpolation. The second step was to prepare the most spectrally appropriate high resolution (10m)

bands to be used as the panchromatic one during pansharpening. To this end, for the case of Band 8a, the Band 8 was regarded directly as the panchromatic one. For the case of Bands 5, 6, 7 the average of Bands 4 and 8 was utilized. The third and final step was the application of the fusion algorithms on the computed intermediate products. This was the fusion process in Full Resolution.

For the Reduced Resolution experiment, the raw datasets were downsampled by a ratio of 2 and afterwards the aforementioned process was repeated, treating the downsampled imagery as new raw data. In this experiment, the resulting pansharpened bands have the same resolution with the original narrow VNIR bands. Therefore, the later were used for the quantitative assessment, by calculating the Q index.

The equivalent procedure is followed for the Sentinel-2 20 m SWIR bands B11 and B12. Spectrally, the closest candidate higher resolution band is B8 and thus this one was employed and regarded as the panchromatic one during pansharpening. In this case, however, the spectral sensitivity between the high resolution band (*i.e.*, B8) and the two SWIR bands was significant.

Regarding the employed fusion techniques (Table 1) that took part in our experiments, a description for the vast majority can be found in [Vivone et al., 2015 and the references therein], while details for the method #9 in [Padwick et al., 2010] and #13 and #14 in [Stanislas et al., 1998].

Fusion/Pansharpening Methods that Participated in all the Experiments		
#	Short Name	Full Method Name
1	ATWT	Additive A Trous Wavelet Transform.
2	ATWT-M2	A Trous Wavelet Transform (Model 2)
3	ATWT-M3	A Trous Wavelet Transform (Model 3)
4	AWLP	Additive W/let Luminance Proportional.
5	BDS	Band-Dependent Spatial-Detail.
6	Brovey	Brovey transform.
7	GS	Gram Schmidt (Mode 1).
8	GSA	Gram Schmidt Adaptive.
9	HCS	Hyperspherical Color Space.
10	HPF	High-Pass Filtering.
11	IHS	Fast Intensity-Hue-Saturation (GIHS) image fusion.
12	Indusion	Indusion: Decimated Wavelet Transform using an additive injection model.
13	LMM	Local Mean Matching.
14	LMVM	Local Mean and Variance Matching.
15	MTF-GLP	Generalized Laplacian Pyramid (GLP) with MTF-matched filter & unitary injection model.
16	MTF-GLP-CBD	GLP with MTF-matched filter and regression based injection model.
17	MTF-GLP-HPM	GLP with MTF-matched filter and multiplicative injection model.
18	MTF-GLP-HPM-PP	MTF-GLP-HPM with Post-Processing
19	PCA	Principal Component Analysis.
20	PRACS	Partial Replacement Adaptive Component Substitution.
21	SFIM	Smoothing Filter-based Intensity Modulation.

Table 1. The 21 fusion (pansharpening) methods which participated in this study.

3. EXPERIMENTAL RESULTS AND VALIDATION

The evaluation results of the pansharpened VNIR imagery for the Full Resolution (FR) and the Reduced Resolution (RR) experiments are presented in Table 1. For the FR experiment, QNR index was utilized, while the quality assessment for the RR experiment was carried out using the robust Q4 index [Garzelli and Nencini, 2009, Vivone et al., 2015].

In the left side of Table 2, the resulting average scoring of QNR and Q4 values are presented. Under this particular evaluation framework both FR and RR experiments took part at the final scoring for all 21 fusion methods. In the right side of Table 2, the ranking is based only on the evaluation outcome of the Q4 index, which was calculated during the RR experiment.

As one can observe there are several differences between the two evaluation frameworks. The most important are the following:

In (A), the *Interpolated Raw* image (just a cubic interpolation on the original low resolution bands) delivers the highest score. As expected, this indicates that the QNR index is scoring exclusively on coherence of the product, without taking into adequately account the spatial information, patterns, etc.

Moreover, the Indusion method ranked 4th in (A), whereas it ranked 14th place in part (B). This also indicated the important differences between the two evaluation frameworks. As expected, AWLP and ATWT were close in all experiments *i.e.*, close in (A), almost the same scoring in (B).

The MTF GLP HPM PP method was at the first place in both (A) and (B), if we ignore the simple upscaling. SFIM and HPF methods were ranked in the 5th and 6th place in (A), while in (B)

QNR and Q4			Only Q4		
Rank	Method	Score	Rank	Method	Score
1	Interpolated Raw	0.927	1	MTF GLP HPM PP	0.899
2	MTF GLP HPM PP	0.879	2	SFIM	0.893
3	PRACS	0.875	3	HPF	0.890
4	Indusion	0.865	4	MTF GLP CBD	0.885
5	SFIM	0.859	5	PRACS	0.881
6	HPF	0.856	6	Interpolated Raw	0.879
7	AWLP	0.856	7	MTF GLP HPM	0.875
8	MTF GLP CBD	0.853	8	MTF GLP	0.874
9	MTF GLP HPM	0.853	9	ATWT	0.873
10	MTF GLP	0.850	10	AWLP	0.871
11	ATWT	0.839	11	ATWTM2	0.851
12	BDS	0.825	12	ATWTM3	0.849
13	ATWTM3	0.807	13	LMM	0.841
14	GSA	0.805	14	Indusion	0.828
15	ATWTM2	0.804	15	GSA	0.818
16	HCS	0.800	16	HCS	0.804
17	LMM	0.785	17	GS	0.803
18	Brovey	0.783	18	BDS	0.803
19	GS	0.780	19	Brovey	0.800
20	PCA	0.766	20	LMVM	0.796
21	LMVM	0.762	21	PCA	0.787
22	IHS	0.762	22	IHS	0.780

Table 2. Quantitative results after the applications of several pansharpening methods on the narrow VNIR Sentinel-2 spectral bands using (A) the average of QNR & Q4 (left) and (B) only Q4 (right).

in 2nd and 3rd place, respectively. Similarly, the two methods MTF GLP HPM and MTF GLP took the 9th/10th and 7th/8th place.

Apart from the evaluation based on quantitative image similarity indexes, a qualitative one was also performed based on the scoring of two remote sensing experts who manually assessed the relative quality of the resulting output images. This qualitative assessment included the 10 methods that scored the highest values during the quantitative evaluation.

It should be noted that the resolution ratio of the Sentinel-2 datasets is 2/1 (10m/20m among VNIR bands) and thus RR experiments can be regarded as more reliable than the FR (with QNR) counterpart. The lower ratio than e.g., the case of very high resolution sensors (WorldView-2, IKONOS, etc.) provides a more accurate quality assessment (including both spectral and spatial components) during the RR experiments. In particular, the relatively small reduction to the resolution of the raw datasets, results in retaining more spatial information than in datasets with a higher ratio. Thus, the behavior of the pansharpening algorithms on the RR experiment can be related with more confidence with the behavior of pansharpening on the raw datasets.

Results from the performed qualitative evaluation from two photo-interpretation experts, after a thorough visual examination and comparison, are presented in Table 3. Again, the differences between the resulting overall ranking in relation with the two aforementioned quantitative frameworks (QNR and Q4, Q4) are significant.

While a further discussion on the qualitative assessment follows the regarding results presented in Figures 3, 4, 5 and 6, it is clear that the ranking after an attentive visual inspection and the one from the quantitative similarity indexes significantly differ. This fact primarily identifies a need for novel quantitative frameworks that can take into account more effectively both the spatial and spectral information towards closing the gap with the expert-based assessments.

If one compares the output ranking between the QNR and Q4 alone then Q4 is more close to what the experts indicated. However, Q4 still lacks on assessing crucial qualitative parameters that can combine image sharpness and spectral fidelity. Note that all images were observed and plotted in the following figures based on exactly the same parameters regarding histogram min/max values, enhancement and color rendering.

Scoring from Experts (1-10)				
Rank	Method	Exp.#1	Exp.#2	Score
1	MTF GLP CBD	9.7	9.0	9.35
2	ATWT	9.6	9.0	9.30
3	AWLP	9.6	9.0	9.30
4	HPF	9.1	9.0	9.05
5	SFIM	9.0	7.5	8.25
6	BDS	8.0	8.5	8.25
7	PRACS	6.0	6.5	6.25
8	MTF GLP HPM PP	6.0	6.0	6.00
9	MTF GLP	6.0	6.0	6.00
10	Indusion	1.0	6.0	3.50

Table 3. Quantitative evaluation of the different methods based on the assessment from remote sensing experts.

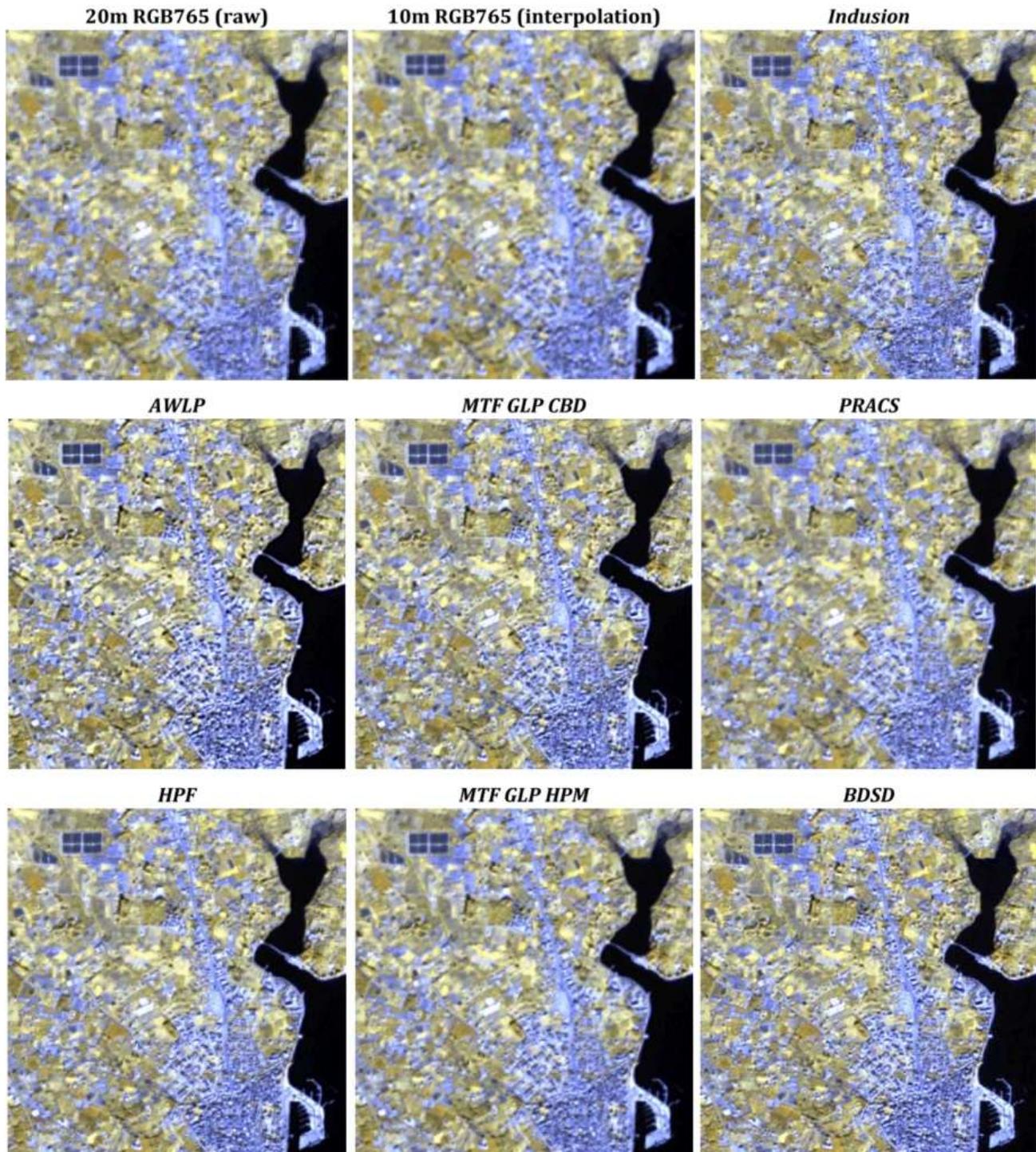


Figure 3. Results after the application of different pansharpening methods on the narrow VNIR Sentinel-2 spectral bands

After a close look at Figure 3, one can divide the pansharpened images to two major groups, based on the spatial enhancement criterion. The first group, let A), would contain five methods: a) Indusion, b) AWLP, c) MTF GLP CBD d) HPF and e) BDSD. The second group, let B), consists of the remaining two methods: f) PRACS and g) MTF GLP HPM. The methods of the first group delivered superior performance in the spatial domain. It is evident that the PRACS and MTF GLP HPM algorithms produced somewhat less sharp images than those of the first group. This observation is best established through the

comparison and examination of the seaport objects/details (at the SE part of image) as well as the urban fabric between the different images. The better performance on retaining the spatial information of the AWLP, HPF and MTF GLP CBD methods (contrary to PRACS) is better observed in Figure 4. While images of group B) are indeed blurrier than those of group A), they demonstrate better performance in terms of spectral fidelity than some members of the group A). In particular, PRACS and MTF GLP HPM seem to preserve better the color information than AWLP, BDSD and HPF methods of the first group.

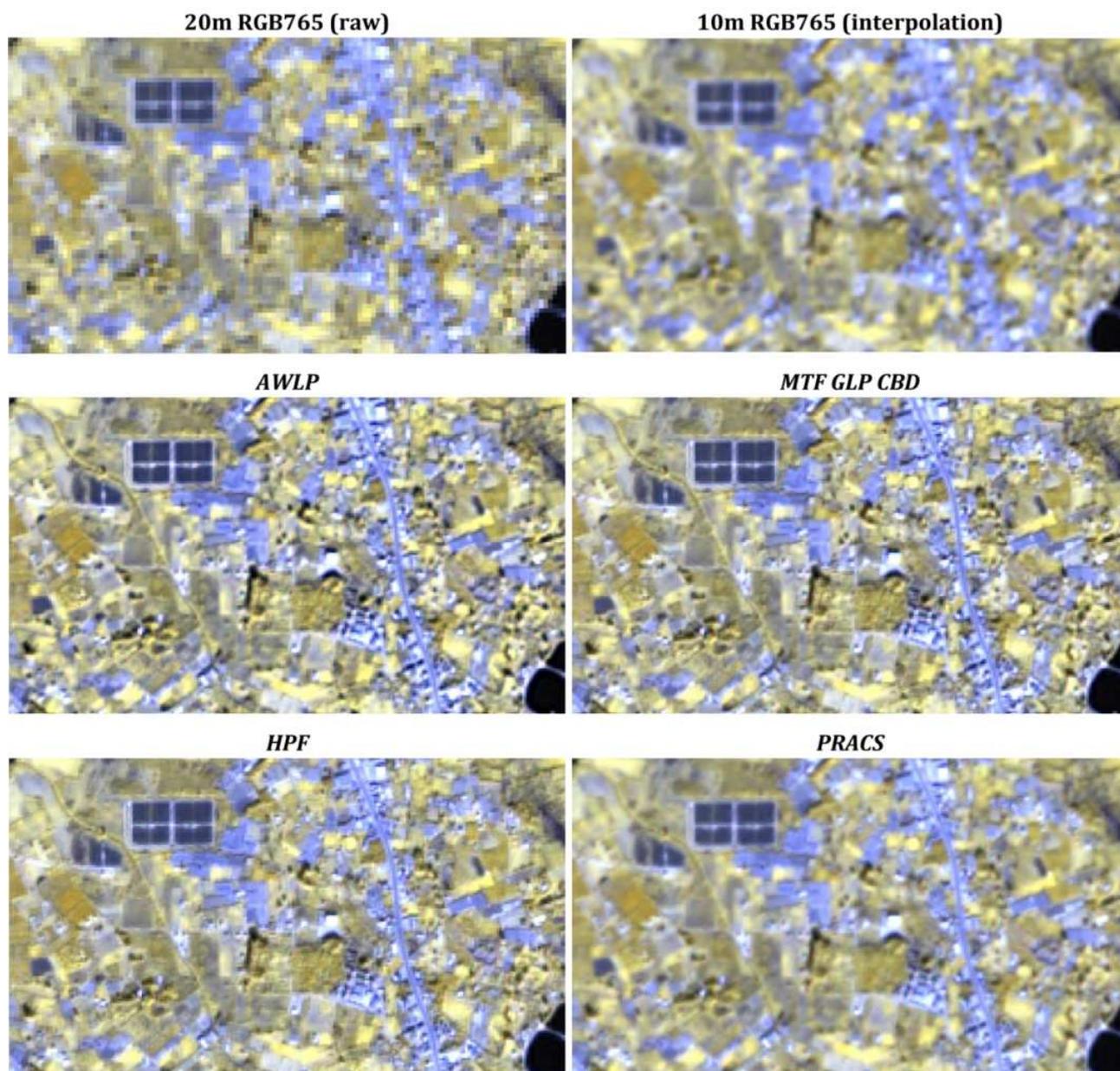


Figure 4. Results after the application of different pansharpener methods on the narrow VNIR Sentinel-2 spectral bands (zoom-in to a smaller region)

The Indusion method, although ranked high in the quantitative assessment based on the QNR and Q indexes, resulted into a relative blurry outcome with significant spatial discontinuities. Moreover it resulted into a spatial shift to a SE direction, which can be straightforwardly observed when overlaid with the original spectral bands #4 and #8. This spatial shift of the Indusion method was also presented during the SWIR experiment. Due to this defect and some minor artifacts (visible in larger scales) Indusion didn't scored high during the qualitative evaluation performed by the experts.

MTF GLP CBD, AWLP, HPF and BDSD performed remarkably well both in the spatial enhancement and spectral fidelity criteria. However, minor differences can be observed, when examined and compared in large scales. MTF GLP CBD provides an exceptionally well balanced image, with optimal trade-off between sharpness and original color preservation.

Thus, MTF GLP CBD was considered to produce the best overall result. AWLP closely follows next, with superior sharpness and increased local contrast which can be very useful for photo-interpretation tasks. Indeed, the geometry of the objects is better expressed in AWLP, with a relatively small, yet noticeable impact in the original spectral values. The HPF result is quite similar with that of MTF GLP CBD, although the later has a slightly more accurate and vivid color tonality. Next follows BDSD, which introduces a minor noise problem, mainly observed in homogenous areas (especially in the sea).

ATWT produces an almost identical image with that of AWLP. For this reason and due to the similarity of the algorithms as well as for space conservation, ATWT method is not displayed. In general, this elaboration justifies the visual scores presented in Table 3.

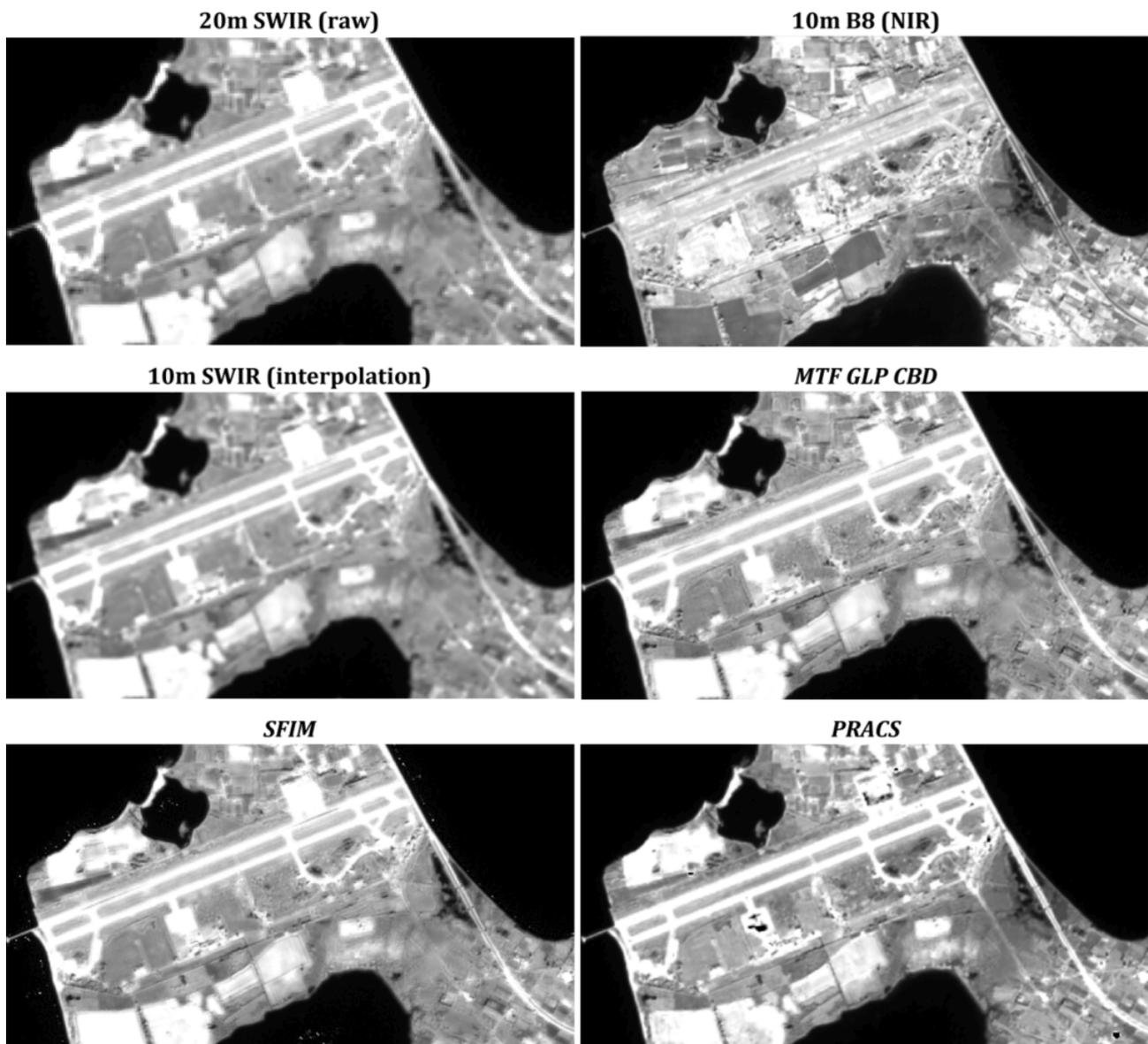


Figure 5. Results after the application of different pansharpening methods on the SWIR Sentinel-2 spectral bands

In Figure 5, three pansharpened SWIR bands are presented, along with both the raw SWIR band and the raw band 8 (NIR) of Sentinel-2 which was utilized as the panchromatic. Band 8 is presented here in order to illustrate the reflectance spectra of the various image objects. These spectral discrepancies hinder the performance of the pansharpening algorithms, which were natively designed to fuse images with a relatively high correlation between the spectral content. Some notable examples of the different reflectance spectra are:

- The airport. It appears almost completely white in the SWIR band, whereas in NIR has medium intensity values.
- Buildings and nearby airport objects. Some of them are very bright in the NIR band and dark in the SWIR (south of the airport). The opposite happens with the features located NW of the airport.
- Non-vegetated fields. These appear very dark in the NIR image and very bright in the SWIR (among others, the four large fields south and south-west of the airport).

Despite the aforementioned spectral discrepancies, most pansharpening methods managed to spatially enhance the lower resolution data, while preserving at a certain extent the spectral behavior of the SWIR bands.

After an attentive visual inspection one can observe that all methods modified the reflectance spectra at a certain extent, in regions that ingested spatial details in the SWIR images. Thus, all results present problematic areas in term of presenting the correct reflectance values of the particular image objects.

If one ignores these critical artifacts then the evaluation indicates that the MTF GLP CBD method produces a better result than the SFIM and PRACS methods. Both SFIM and PRACS suffer from several artifacts and burnt pixels. These problematic regions are more effectively observed in Figure 6. Here, the crucial alternation of reflectance spectra in various regions is more than apparent. All methods while ingesting spatial information due to the spectral dissimilarities with the reference higher resolution image, produced significant spectral

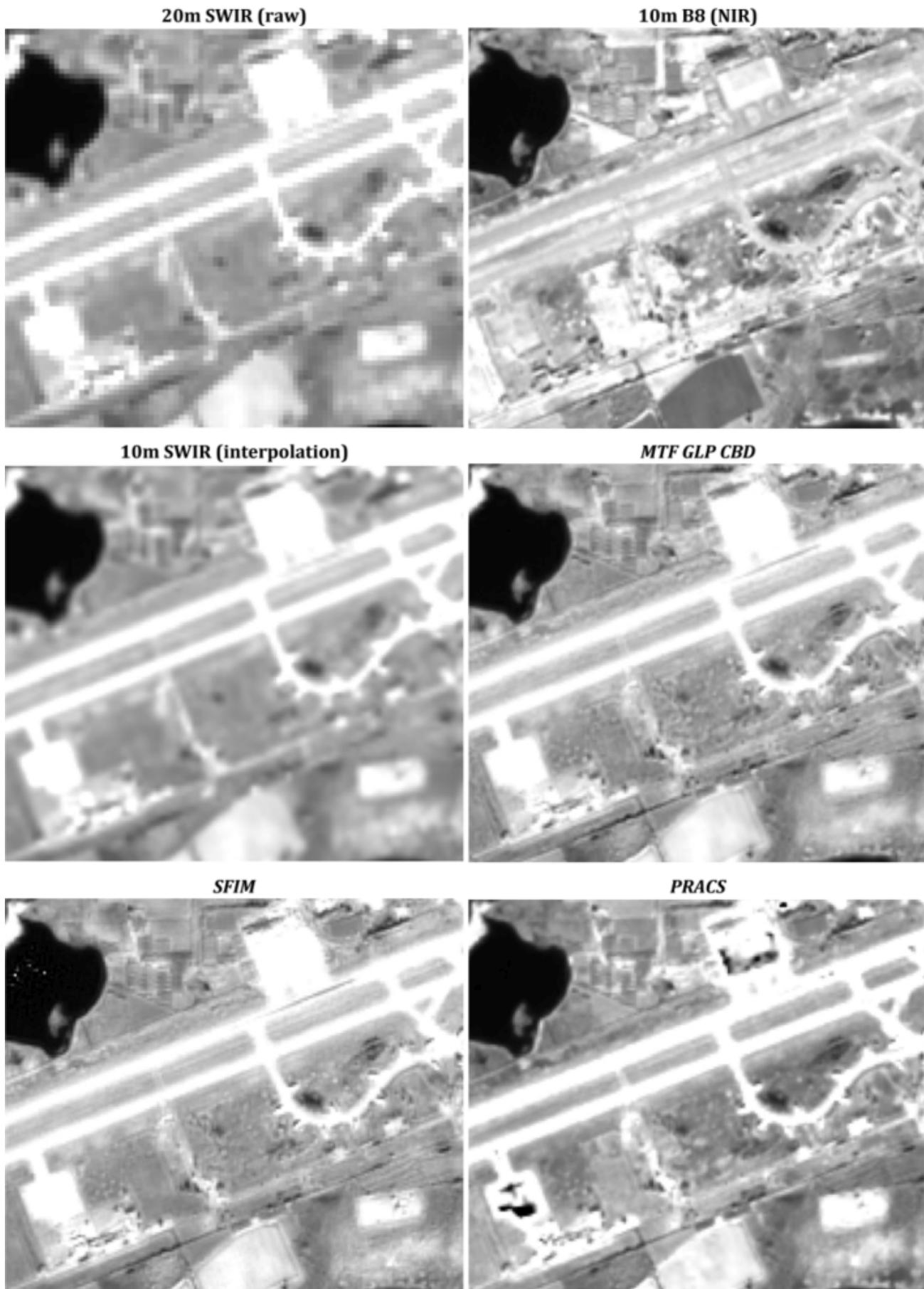


Figure 6. Results after the application of different pansharpener methods on the SWIR Sentinel-2 spectral bands

alteration, which may be crucial during any further image analysis or classification tasks.

4. CONCLUSIONS

In conclusion, this paper considered 21 pansharpening algorithms in order to spatially enhance the narrow 20 m VNIR and SWIR bands of the Sentinel-2 satellite. The fusion techniques were evaluated in Full Resolution by measuring the QNR index and in Reduced Resolution by calculating the Q4 and Q indexes. Additionally, a qualitative scoring assessed through visual inspection was carried out, for the best-performing methods index-wise.

Although the implemented index evaluation framework provided a starting base to separate poor performing methods from methods producing high-quality results, there were significant differences between the index results and the assessments from the evaluation of photo-interpretation experts. The main problem of the current index evaluation framework seems to be that the methods performing well in spectral fidelity are favoured excessively over high-performing methods in the spatial domain. Moreover, the introduction of small artifacts and burnt pixels in the resulting fused imagery is not properly penalized in neither evaluation framework. This fact highlights the need for more robust index validation frameworks, which would close the gap between manual and automated image quality estimation.

The joint overall evaluation results indicate that the method MTF-GLP-CBD delivered consistently higher quality products. AWLP and ATWT methods closely follow next and as a third choice SFIM or HPF could be used. However, a comprehensive evaluation over more study areas and under additional evaluation frameworks should be performed which will include also the rest of the Sentinel-2 spectral bands.

ACKNOWLEDGEMENTS

Part of this research was funded by the 'ELKE' PhD Scholarship of the National Technical University of Athens.

REFERENCES

Aiazzi B., L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; Bruce, L.M. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Trans. Geosci. Remote Sens.* 2007, 45, 3012–3021

Chavez P. S. Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic,"

Photogramm. Eng. Remote Sens., vol. 57, no. 3, pp. 295–303, Mar. 1991.

Choi J., K. Yu, and Y. Kim, "A new adaptive component-substitution based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.

Garzelli A. and F. Nencini, "Hypercomplex quality assessment of multi-/hyper-spectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.

Garzelli, A. Pansharpening of multispectral images based on nonlocal parameter optimization. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 2096–2107.

Garzelli A., F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.

Gillespie, A.R.; Kahle, A.B.; Walker, R.E. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. *Remote Sens. Environ.* 1987, 22, 343–365.

Otazu X., M. González-Audicana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.

Padwick, C., Deskevich, M., Pacifici F., Smallwood, S., 2010. "WorldView-2 pan-sharpening". Proceedings of the ASPRS Annual Conference, San Diego, California.

Stanislas D.B., Muller F., Donnay J., 1998. Fusion Of Multispectral And Panchromatic Images By Local Mean And Variance Matching Filtering Techniques, *Fusion of Earth Data*, Sophia Antipolis, France, 28-30.

Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 2565–2586.

Vivone G., R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.

Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* 1997, 63, 691–699.

Wang Z. and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

Zhang, H.K.; Roy, D.P., 2016. Computationally Inexpensive Landsat 8 Operational Land Imager (OLI) Pansharpening. *Remote Sens.* 2016, 8, 180.