HYPERSPECTRAL VNIR - SWIR IMAGE REGISTRATION: DO NOT THROW AWAY THOSE OVERLAPPING LOW SNR BANDS

Federico Grillini, Jean-Baptiste Thomas, Sony George

Colourlab, Department of Computer Science, NTNU, Norway

ABSTRACT

We study the influence from a series of factors on the quality of the registration of VNIR and SWIR hyperspectral images. We specifically consider the registration of VNIR and SWIR images of different spatial resolutions acquired in controlled laboratory conditions on different historical artefacts. This registration problem is defined by a large scaling difference and small translation and rotation. We compare four methods and demonstrate that the largest effects on the quality metrics are due to the image contents and to the deployment of specific ranges of spectral bands. In particular, we demonstrate that using bands of similar nominal wavelengths gives significant advantage. While those bands have typically low values of Signal-to-Noise ratio and are frequently discarded, for this application, they can be treated as a valuable asset.

Index Terms— Hyperspectral imaging, Registration, VNIR-SWIR, Cultural Heritage

1. INTRODUCTION

In the field of Cultural Heritage, imaging spectroscopy is encountering a growing trend in popularity which makes the technique more accessible to many facilities and laboratories around the world. As a non-invasive, non-destructive technique, it represents a reliable tool for documentation, monitoring, and analysis of historical artefacts [1]. Another trend that this field of research is experiencing is the deployment of spectral data in different modalities and ranges of the electromagnetic spectrum. However, such modalities are usually analyzed and processed independently, with conclusions and findings that are eventually inferred by combining the individual results.

Image registration is a preliminary step that allows the combined analysis of separate datasets that refer to the same scene. Data alignment is a compulsory requirement for pixel-based analysis such as pigment mapping [2, 3] and spectral unmixing [4] or further fusion processing like pansharpening and hypersharpening [5, 6]. Therefore, registered images should be as close as possible to sub-pixel precision, in order to provide accurate inferences regarding the material properties. The literature in remote sensing offers a plethora of articles that address the problem of hyperspectral images registration, also in the case of different hyperspectral imaging modalities.

Image registration techniques can be broadly divided into two families: feature-based and intensity-based approaches [7]. Feature-based methods require the detection of control points in both fixed (or reference) and moving images; which at a later stage are matched according to the similarity of their constructed feature vectors. Examples of this family are the SIFT [8] and SURF [9]. Intensity-based methods rely on the optimization of a cost function that reflects the final quality of alignment between fixed and moving images. This family is often adopted to solve multimodal registration problems maximizing metrics such as correlation and Mutual Information (MI). Normally, both approaches are performed on a predefined fixed-moving image pair. However, in the case of hyperspectral imaging in two different modalities, these *roles* are not defined, i.e., the decision of which spectral set is the reference depends on the application. In addition, both fixed and moving datasets are represented by a series of images, which means that the spectral information can be exploited to build richer feature vectors, as proposed in several registration attempts [10, 11, 12]. The abundance of images can also be considered an asset in generating the most suitable image pair that leads the registration, in the same fashion of the band selection and band synthesis approaches adopted in hypersharpening [6]. In the field of Cultural Heritage imaging, multimodal registration represents a crucial task, since historical artefacts are often studied in cross-disciplinary frameworks that require the capturing of images carrying different information (topography maps, XRF, imaging spectroscopy, FTIR, etc.). A popular approach was proposed in [13], where the maximum of the modulus of the wavelet transform is deployed to identify control points and then the local maximum of normalized cross correlation between phase images helps in identifying potential matched pairs between fixed and moving images. The method, referred here as to MWTPXC (Max Wavelet Transform Phase Cross-Correlation), was developed mainly for applications on historical paintings.

In this article we tackle the specific problem of registering spectral data with two hyperspectral imagers working in complementary ranges of the electromagnetic spectrum: Visible-Near-Infrared (VNIR) and Shortwave-Infrared (SWIR) by evaluating the influence of selected factors on the quality of

the final aligned result. The capturing of hyperspectral data for historical artefacts usually takes place in laboratory or in-situ, with the latter being more challenging as several aspects - such as the illumination - cannot be controlled as well as in laboratory conditions. In this work, we adapt and compare several registration methods to the case of VNIR-SWIR images. We perform an experiment on images of objects that mock-up cultural heritage artefacts (paintings, textiles, drawings), which images were acquired in controlled laboratory conditions. We show that the methods benefiting from prior scaling and overlapping bands perform the best.

2. IMAGING SET-UP AND PROBLEM DESCRIPTION

The capturing set-up is illustrated in Fig.1 and it shows the two HySpex hyperspectral imagers manufactured by NEO (Norsk Elektro Optikk) - VNIR1800 and SWIR384 - in the pushbroom configuration. In this system, the translational stage shifts the scene across the fields of view of the cameras which synchronously acquire the full spectrum of a single spatial line per time of exposure. The VNIR camera deploys 186 spectral bands in the range 400-1000 nm, with 1800 pixels on the acquisition line. On the other hand, the SWIR camera has 288 bands in the range 950-2500 nm, with only 384 pixels on the acquisition line. Thus, a scaling spatial resolution ratio of approximately 4.6 exists between the two datasets, while the difference in field of view for the two imagers - 17° for VNIR and 16° for SWIR - introduces shifts in the registration problem. The illumination geometry is designed in a way that the irradiance impinging the target is not harmful and does not exceed the light dosage limits for historical artefacts [14]. The relative angles between camera, target, and lights are the same for both VNIR and SWIR subsetups in order to avoid significant differences in intensity due to the Bidirectional Reflectance Distribution Function (BRDF) of the object surface. The deployed halogen lights emit continuously in the range 400-2500 nm, but although they are of the same model, the reading of illuminance on the target in lux does not match. For this reason, working with spectral reflectance factors instead of radiance data will facilitate the registration, especially for feature-based methods.

The registration of VNIR and SWIR spectral data can be regarded as a *mild* multimodal registration problem, with the multimodality that arises from the fact that the information acquired comes from different regions of the electromagnetic spectrum. However, the two imagers capture the same type of information (spectral reflectance), and the multimodality becomes less accentuated if we consider that there exists a narrow interval (950-1000 nm) in which the two datasets share overlapping nominal wavelengths. At a closer inspection of the data in the narrow shared region it is possible to observe that the spectral values do not concatenate accurately. This is due to a series of factors such as the decrease in signal-to-noise ratio (SNR) at the extremities of the two spectral ranges [15] (as illustrated in Fig.2), the change in bandwidth between the two sensors (CMOS and MCT), and the differences in BRDF that exist at the pixel level, due to spurious differences in angles between target, camera, and illumination.

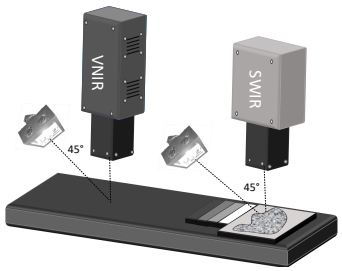

Fig. 1. Schematic representation of the dual camera imaging set-up.

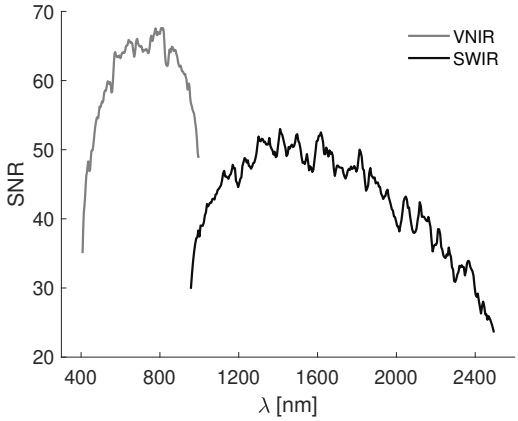

Fig. 2. Signal-to-Noise Ratio of the two hyperspectral imagers as a function of wavelength. The curves are obtained capturing a standardized uniform 99% reflective Spectralon tile and applying the ratio between the mean and standard deviation for each spectral band.

Since the spectral sets are internally co-registered, it is assumed that is necessary to learn only one homography between a fixed-moving image pair that can then be applied to the rest of the moving bands.

3. MATERIAL AND METHODS

In this section we describe the selected factors that potentially affect the results of the registration performances. In order to evaluate the quality of registration, five metrics are selected. Three of them are commonly used to evaluate image similarity: normalized Mutual Information (nMI) [16], Peak Signal to Noise Ratio (PSNR), and Relative Dimensionless Global Error (ERGAS) [17]. Two quality metrics are adapted to suit the evaluation of VNIR-SWIR alignment: the Structural

Similarity Index Measure (SSIM) [18] is modified to account only for its contrast and structure components and renamed truncated-SSIM (tSSIM):

$$tSSIM(X,Y) = \frac{2\sigma_X\sigma_Y + c}{\sigma_X^2 + \sigma_Y^2 + c} \cdot \frac{\sigma_{XY} + c/2}{\sigma_X\sigma_Y + c/2} \quad (1)$$

In which X and Y are the examined images, σ^2 the variance of the image, σ_{XY} the covariance, and $c = 0.03L^2$ (with L being the dynamic range of the images). In order to compare pixel distances across images with significantly different spatial resolution, the commonly used pixel displacement is modified to account for the real-world dimension of a pixel, in a measure here called Ground Sampling Distance Error (GSDE):

$$GSDE(X,Y) = \Delta p(X,Y) \cdot g(X) \ \ [\mu m] \quad (2)$$

In which Δp is the classic pixel displacement computed with the euclidean distance between two matched points in images X and Y, and g is the ground sampling distance of the fixed image X (g is approximately 50 μm for VNIR and 200 μm for SWIR, computed from a 30 cm distance between camera and target).

The selected factors are of three types: registration methods, image contents, and related to decisions. The selected registration methods include feature-based methods like SIFT [8] and spatio-spectral SIFT [12] (SS SIFT), and a combination of feature and area based methods such as SIFT followed by maximization of nMI (SIFT OPT) and MWTPXC [13]. The image contents, which mock-up historical artefacts and are illustrated in Fig.3, are a drawing on paper (D), a flat painting (FP), a painting with textural reliefs (RP), and a piece of textile (T). The factors related to decisions regard: selection of reference image (VNIR band or SWIR band), considered spectral range (full or overlap), modality of fixed and moving images generation (band selection or band synthesis), scaling (a priori or instrinsic in the homography matrix), and type of homography (affine or projective).

In this study we perform the image registration task covering all the combinations of the reported factors. In total, 352 tasks are performed. Some of the factors do not apply to all the methods, such as SS SIFT that does not require the generation of a single band and operates only on same spectral ranges. Similarly, MWTPXC requires the two images to be always scaled to the same size a priori.

When the overlap range is considered, only the last 13 bands of VNIR and first 8 bands of SWIR constitute the available spectral sets. For the generation of single fixed and moving image pairs from the spectral sets it was decided to choose the approaches of band selection and band synthesis. Band selection picks the two most correlated spectral bands between VNIR and SWIR sets, while band synthesis generates the bands using the first component of Principal Component Analysis (PCA) performed on the images.

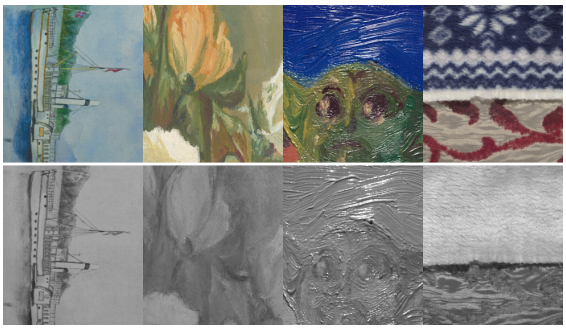

Fig. 3. Image contents in sRGB and IR (band-averaging). From left to right: Drawing on paper (D), Flat painting (FP), Painting with textural reliefs (RP), Textile (T).

4. RESULTS

The single-image quality metrics are computed after registering a pair of test bands that are located in the overlap range of the two spectral imagers, while ERGAS is computed on the spectral datasets of the overlapping range, up-sampling the SWIR data with linear interpolation to match the nominal wavelengths of VNIR.

Since the highlighted potential factors are several, a Principal Component Analysis was performed to inspect possible correlations and interactions. Fig.4 reports the loadings plot of the evaluated conditions along the two first components, which roughly explain 30% of the total variance. It is worth to point out that this percentage is typically a low value. However, by inspecting the next components (not shown) we can corroborate similar conclusions. The performance metric scores are highly correlated, taking into account that ERGAS and GSDE good performance exhibit a low score and that nMI, tSSIM and PSNR good performance exhibit a high score.

From this plot it appears that the most influencing factors are the selected methods, two of the image contents (D and RP), and factors like scaling, selected wavelength range, and band generation. The positions of the loadings with respect to the metrics suggest that points that lie close to the metric cluster (tSSIM, nMI, PSNR) positively influences the results, and conversely what lies further away provides a negative contribution. We also conclude that for further examination (and for simplicity) it is possible to refer to a single quality metric.

To evaluate the influence of the registration methods and image contents Fig.5 reports the statistics related to normalized Mutual Information in the registration tasks. SIFT OPT and MWTPXC come out as the best performing approaches, while SIFT and SS SIFT probably suffer from the intensity differences that exist between the images. Amongst the image contents, it is clear how the presence of textural reliefs and specularities negatively affects the results, as in the case of RP. Quite surprisingly, the textile image T did not pose a

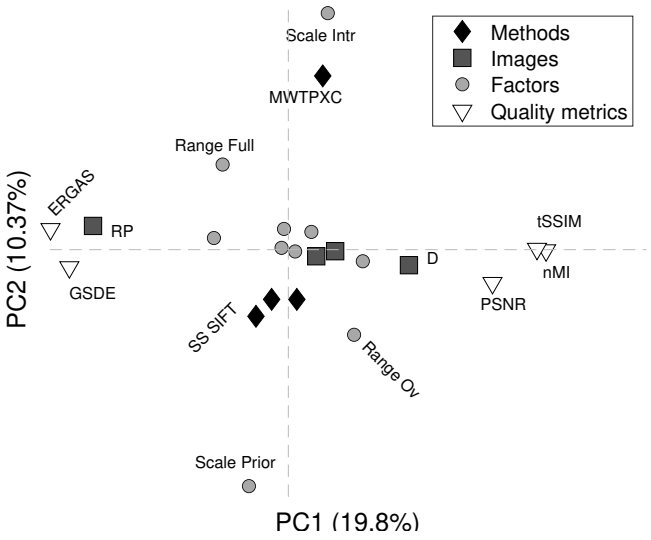

Fig. 4. Loadings of first 2 components of PCA. Only the names of the most influential variables reported for clarity.

big challenge, as consistently good results were achieved.

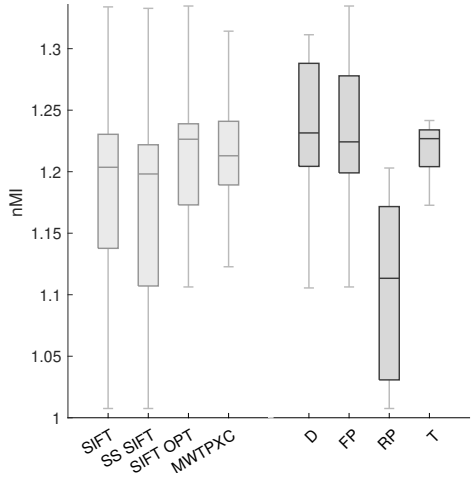

Fig. 5. Statistics of nMI for groups of registration runs divided into methods (left) and image contents (right).

To better appreciate the influence of the selected factors on the registration performances it is possible to examine the image contents individually. We quantified a factor influence by regressing a linear model that relates the conditions of observation with the normalized Mutual Information of the registered images. By using this analytical technique and appropriately normalizing the data, the sign of the regression coefficients will provide information regarding the nature of the influence: either positive or negative. By consequence, the absolute value of the coefficients will give indications on the degree of influence. A visualization of this analysis is provided in Fig.6 in which it is possible to observe that the exploitation of the overlapping range (950-1000 nm) brings a positive contribution to nMI. Other factors of positive contribution are the selection of the SWIR image as reference, and the prior scaling (a negative contribution from learning the scale intrinsically is highlighted). This analysis does not solve the question on whether it is suggested to proceed with band selection or band synthesis. Running the same analysis excluding the observations where SS SIFT is deployed (i.e., when band generation is not a factor) it would appear that band selection works better for the image contents of D, FP and T, while band synthesis is preferred for RP.

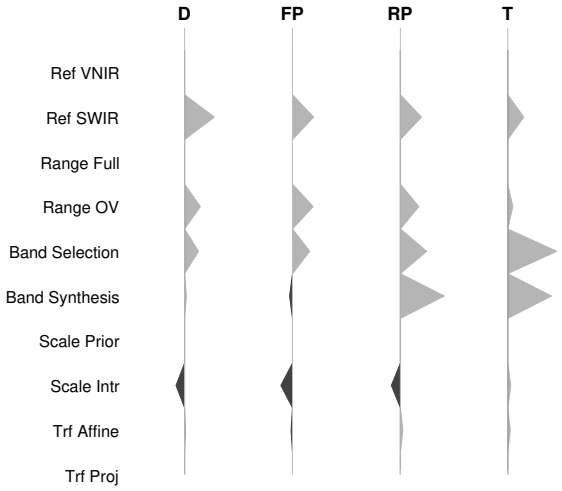

Fig. 6. Linear model coefficients to describe the factors influence on nMI. Light-gray triangles pointing to the right denote a positive influence on the registration result. Conversely, darker triangles pointing to the left highlight negative effects.

To highlight the positive effects of considering only the overlapping range of spectral bands, Table 1 reports the values of the quality metrics of full range against overlap for two image contents - FP and D - when applying SIFT OPT, using SWIR as reference, with band synthesis generation, prior scaling, and affine transformation.

Image		**nMI**	**tSSIM**	**ERGAS**	**GSDE** (μm)
	Ideal	**2**	**1**	**0**	**0**
D	**Start**	1.13	0.70	1.97	380
	Full λ	1.19	0.85	1.66	150
	Ov λ	1.29	0.98	1.30	29.1
FP	**Start**	1.13	0.45	4.51	822
	Full λ	1.11	0.45	4.75	996
	Ov λ	1.31	0.92	4.58	55.0

Table 1. Registration of some of the quality metrics comparing cases of Full spectral range against Overlap.

5. CONCLUSION

Obtaining sub-pixel precision when registering hyperspectral images coming from different modalities enables the further processing and accurate study of historical artefacts. Knowing which factors improve or degrade the performances of alignment is crucial to obtain the best possible results. With the aid of multivariate techniques we highlighted that in the case of VNIR-SWIR image registration a prior scaling, the deployment of SWIR as reference, and the usage of the bands shared by both imagers contribute positively in achieving good results. In particular, we advise against discarding spectral bands in the overlapping wavelength range that upon a first inspection have a lower signal-to-noise ratio, since they can actually be used as a valuable asset.

6. REFERENCES

[1] Costanza Cucci, John K Delaney, and Marcello Picollo, "Reflectance hyperspectral imaging for investigation of works of art: old master paintings and illuminated manuscripts," *Accounts of chemical research*, vol. 49, no. 10, pp. 2070–2079, 2016.

[2] Hilda Deborah, Sony George, and Jon Yngve Hardeberg, "Pigment mapping of The Scream (1893) based on hyperspectral imaging," in *International Conference on Image and Signal processing*. Springer, 2014, pp. 247–256.

[3] John K Delaney, Kathryn A Dooley, Annelies Van Loon, and Abbie Vandivere, "Mapping the pigment distribution of Vermeer's Girl with a Pearl Earring," *Heritage Science*, vol. 8, no. 1, pp. 1–16, 2020.

[4] Chen Shi and Le Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sensing of Environment*, vol. 149, pp. 70–87, 2014.

[5] Gemine Vivone, Mauro Dalla Mura, Andrea Garzelli, and Fabio Pacifici, "A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6102–6118, 2021.

[6] Massimo Selva, Bruno Aiazzi, Francesco Butera, Leandro Chiarantini, and Stefano Baronti, "Hyper-sharpening of hyperspectral data: A first approach," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2014, pp. 1–4.

[7] Arthur Ardeshir Goshtasby, *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*, John Wiley & Sons, 2005.

[8] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[10] Suhad Lateef Al-Khafaji, Jun Zhou, Ali Zia, and Alan Wee-Chung Liew, "Spectral-spatial scale invariant feature transform for hyperspectral images," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 837–850, 2017.

[11] Yanshan Li, Qingteng Li, Yan Liu, and Weixin Xie, "A spatial-spectral SIFT for hyperspectral image matching and classification," *Pattern Recognition Letters*, vol. 127, pp. 18–26, 2019.

[12] Yang Yu, Yong Ma, Xiaoguang Mei, Fan Fan, Jun Huang, and Jiayi Ma, "A spatial-spectral feature descriptor for hyperspectral image matching," *Remote Sensing*, vol. 13, no. 23, pp. 4912, 2021.

[13] Damon M Conover, John K Delaney, and Murray H Loew, "Automatic registration and mosaicking of technical images of Old Master paintings," *Applied Physics A*, vol. 119, no. 4, pp. 1567–1575, 2015.

[14] Matija Strlič, David Thickett, Joel Taylor, and May Cassar, "Damage functions in heritage science," *Studies in Conservation*, vol. 58, no. 2, pp. 80–87, 2013.

[15] Unal Okyay and Shuhab D Khan, "Spatial co-registration and spectral concatenation of panoramic ground-based hyperspectral images," *Photogrammetric Engineering & Remote Sensing*, vol. 84, no. 12, pp. 781–790, 2018.

[16] Aaron F McDaid, Derek Greene, and Neil Hurley, "Normalized mutual information to evaluate overlapping community finding algorithms," *arXiv preprint arXiv:1110.2515*, 2011.

[17] Lucien Wald, *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*, Presses des MINES, 2002.

[18] Zhou Wang, Eero P Simoncelli, and Alan C Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Ieee, 2003, vol. 2, pp. 1398–1402.