Direct alignment of narrow field-of-view hyperspectral data and full-view RGB image

Guillaume Caron, Suzan Joseph Kessy, Yasuhiro Mukaigawa, Takuya Funatomi

DIRECT ALIGNMENT OF NARROW FIELD-OF-VIEW HYPERSPECTRAL DATA AND FULL-VIEW RGB IMAGE

Guillaume Caron[1,2], *Suzan Joseph Kessy*[3], *Yasuhiro Mukaigawa*[3], *Takuya Funatomi*[3]

[1]CNRS-AIST JRL (Joint Robotics Laboratory), IRL, Tsukuba, Japan
[2]University of Picardie Jules Verne, MIS lab, Amiens, France
[3]Nara Institute of Science and Technology, Nara, Japan

ABSTRACT

The novelty of this paper is the alignment method of narrow field-of-view hyperspectral images to full-view RGB images. The interest is to locate hyperspectral measurements in an environment described by an equirectangular image. But the very different modalities (3 vs. hundreds of channels) and fields-of-view are challenges for accurate alignment. We solve these problems within a dense direct alignment framework that optimizes the warping parameters together with those of a global illumination difference model. Our alignment code is shared with an example dataset available at github.com/jrl-umi3218/hsrgbalign.

***Index Terms*—** Direct alignment, hyperspectral imaging

Fig. 1: Inside Amiens Cathedral, France: (left) 42m height at the center; (right) stained-glass III scanned by a hyperspectral head in the *triforium*, a corridor at 20m above the ground.

1. INTRODUCTION

The interest in hyperspectral imaging (HS imaging or HSI) is growing in various fields ranging from cultural heritage [1], to agriculture [2]. HSI provides a spectrum of each pixel in the image of a scene. Many of the applications require high *spectral*-resolution (2nm bins or even shorter [3]). Despite the long scanning time of whisk-broom HSI (*i.e.* one hour for 200×400 pixels of 2068 channels [4]), recent illumination variation compensation techniques [4] could avoid the usual need for light calibration [5]. This enables HSI's applicability to measurement of historic buildings and artifacts with natural lighting, particularly for challenging subjects, such as stained-glass windows [6] (Fig. 1).

Capturing such data "isolated" proves the capability of HSI but is not sufficient for the needs of archiving heritage buildings. Indeed, the last decade has seen large efforts to structure the latter tedious archiving works within Heritage Building Information Models (HBIM) [7]. To warrant exchange amongst surveyors and facility managers, HBIM integrates in one single model a large amount of multi-modal data of the building (*e.g.* blueprints, photographs, 3D geometric shape and full-view images). Hence, the accurate alignment of all data types is a central issue for HBIM. This is a challenge for new modalities as HSI of ultra high spectral resolution. Indeed, beyond the difference in data types, HSI's long scanning time implies either a low spatial density or a narrow Field-of-View (FoV) of the hyperspectral measurements.

There are hyperspectral image alignment algorithms with an RGB image within hyperspectral spatial super resolution methods [8, 9]. However, they only deal with very slight misalignments in carefully designed optical systems acquiring both types of images as if they were acquired from the same viewpoint at the same time.

Instead, this paper tackles the alignment of hyperspectral and full-view RGB images acquired from different viewpoints, at different times and with different camera FoV, all in addition to a very different number of channels. This issue is solved with a new alignment framework under the dense direct alignment paradigm, *i.e.* using all pixels of images to optimize the alignment parameters.

This paper is organized as follows. First, Section 2 reports related works on image alignment. Then, Section 3 reports the camera models considered. After that, the direct alignment algorithm is detailed in Section 4 and the results are reported in Section 5, before conclusion (Sec. 6).

Thanks to JSPS (Japan) and MAEI (France) agency for funding the Sakura project "FullScan".

2. RELATED WORKS

2.1. Dense direct image alignment approaches

Dense direct image alignment approaches use pixel brightness to express a cost (*e.g.* the Sum-of-Squared-Differences) evaluating the geometric alignment of a target image and a query image, warped toward the target image domain. Since the seminal work of Lucas and Kanade [10], *i.e.* warping a perspective image area with one Degree-of-Freedom (DoF), four decades of research have considered much more DoF [11], including exposure compensation [12], and many camera models, up to panoramic (hemispherical) [13]. Some approaches have expressed the alignment cost from RGB channels of pixels instead of their brightness [14]. But using grayscale [11] leads to the same subpixel accuracy.

2.2. Hyperspectral image alignment methods

Precise hyperspectral image alignment with RGB or panchromatic image is mandatory for hyperspectral super-resolution. Many studies [15, 16, 17] assume a precise alignment of the modalities and focus only on the integration. But perfect optical alignment is hard to guarantee, so recent works consider aligning modalities simultaneously to the super-resolution.

These works consider geometric alignment models from in-plane translation only [18, 19, 20] to rotation and projective homography [9]. But while successfully generating high resolution hyperspectral images, they are only robust to very slight misalignment of the input image pairs of similar viewpoint and FoV, *i.e.* 3–5 pixels according to their experimental results and none can treat large misalignment in different FoV.

3. CAMERA MODELS

This section recalls the basics of spherical and equirectangular camera projection models with unified notation. The coordinates $\mathbf{u} = (u, v)^\top \in \mathbb{R}^2$ of a point in the digital image plane are function of the line of sight of the 3D point observed in $\mathbf{u}$.

3.1. Spherical model for whisk-broom HS camera

With ${}^s_{\cdot l}$ and ${}^s_{\cdot u}$ noting lower and upper bounds of spherical coordinates, the camera samples in its local frame F_s the azimuth ${}^s\theta \in [{}^s\theta_l, {}^s\theta_u] \subseteq [-\pi, \pi]$, around axis Z_s and the elevation angle ${}^s\phi \in [{}^s\phi_l, {}^s\phi_u] \subseteq [-\pi/2, \pi/2]$, both with constant steps $\delta_\theta \in \mathbb{R}$ and $\delta_\phi \in \mathbb{R}$. Here, ${}^s\theta_l, {}^s\theta_u$ and ${}^s\phi_l, {}^s\phi_u$ denote the FoV of HSI whose dimensions are $N_s^r \times N_s^c$, with $N_s^r = \frac{{}^s\phi_u - {}^s\phi_l}{\delta_\phi}$ and $N_s^c = \frac{{}^s\theta_u - {}^s\theta_l}{\delta_\theta}$. Then, the mapping from $({}^s\theta, {}^s\phi)$ to ${}^s\mathbf{u}$ in the digital image plane is nothing but the affine transformation $\mathbf{K}_s \in \mathbb{R}^{3\times3}$:

$$\begin{bmatrix} {}^su \\ {}^sv \\ 1 \end{bmatrix} = \mathbf{K}_s \begin{bmatrix} {}^s\theta \\ {}^s\phi \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\delta_\theta} & 0 & -\frac{{}^s\theta_l}{\delta_\theta} \\ 0 & \frac{1}{\delta_\phi} & -\frac{{}^s\phi_l}{\delta_\phi} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {}^s\theta \\ {}^s\phi \\ 1 \end{bmatrix}. \quad (1)$$

3.2. Equirectangular

The equirectangular image format is a common output of full-view cameras. Many datasets feature such images in computer vision and robotics [21]. The equirectangular camera projection model is very close to the spherical model (Sec. 3.1). Azimuth[1] ${}^e\theta$ and elevation ${}^e\phi$ angles are expressed in the equirectangular camera coordinate system F_e with angle ${}^e\theta$ around the camera Z_e axis. However, as the equirectangular image format spans the whole full-view instead of a subpart with the spherical model (1), the affine transformation $\mathbf{K}_s$ of (1) is substituted with $\mathbf{K}_e \in \mathbb{R}^{3\times3}$. $\mathbf{K}_e$ is function of the output image dimensions $N_e^r \times N_e^c$, such that the camera axis X_e points toward ${}^e\theta = {}^e\phi = 0$:

$$\begin{bmatrix} {}^eu \\ {}^ev \\ 1 \end{bmatrix} = \mathbf{K}_e \begin{bmatrix} {}^e\theta \\ {}^e\phi \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{N_e^c}{2\pi} & 0 & \frac{N_e^c}{2} \\ 0 & \frac{N_e^r}{\pi} & \frac{N_e^r}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {}^e\theta \\ {}^e\phi \\ 1 \end{bmatrix}. \quad (2)$$

With this model, the Y_e axis of the camera frame F_e is parallel to the horizontal axis of the equirectangular image.

3.3. Spherical to equirectangular transformation

For the alignment algorithm hereafter, it is clear to set both types of images in the same space. Without lack of generality, we chose to transform the spherical hyperspectral data to the equirectangular image plane. Basically, we apply (2) to ${}^s\theta$ and ${}^s\phi$ and we obtain ${}^e\mathbf{u}^*$:

$$[{}^eu^*, {}^ev^*, 1]^\top = \mathbf{K}_e\, [{}^s\theta, {}^s\phi, 1]^\top. \quad (3)$$

4. DIRECT ALIGNMENT

The hyperspectral and RGB equirectangular data are aligned with a projective transformation: the homography on the unit sphere (Sec. 4.1). To accurately align hyperspectral and RGB data, they are first converted to grayscale as $I_S({}^e\mathbf{u}^*)$ and $I_C({}^e\mathbf{u})$, respectively (Sec. 4.2.1), before being used in a direct alignment scheme densely aligning the whole set of pixels of $I_S({}^e\mathbf{u}^*)$ to $I_C({}^e\mathbf{u})$ ($.^*$ is for reference coordinates). To account for different exposures of $I_S({}^e\mathbf{u}^*)$ and $I_C({}^e\mathbf{u})$, a global exposure affine transformation model is considered too (Sec. 4.2). In the rest of the paper, exponents to indicate a quantity is expressed in a frame (mainly F_e) are omitted.

4.1. Geometric model: homography on unit sphere

Homography is a well known projective transformation between two sets of corresponding image points [22]. For equirectangular images, it requires first to map image coordinates $\mathbf{u}^*$ and $\mathbf{u}$ to azimuth and elevation by inverting (2), *i.e.* for $\mathbf{u}$ (and similarly for $\mathbf{u}^*$):

$$[\theta, \phi, 1]^\top = \mathbf{K}_e^{-1}\, [u, v, 1]^\top. \quad (4)$$

[1] e refers to equirectangular image quantities.

The homography $\mathbf{H} \in \mathbb{P}^2$ is defined for Cartesian homogeneous coordinates. We transform (θ, ϕ) to Cartesian coordinates $\mathbf{X} = [X, Y, Z]^\top \in \mathbb{R}^3$ with $||[X, Y, Z]|| = 1$ with:

$$[X, Y, Z] = [\cos\phi \ \cos\theta, \ \ \cos\phi \ \sin\theta, \ \ \sin\phi], \quad (5)$$

and similarly for (θ^*, ϕ^*) to $\mathbf{X}^*$. Then, the homography $\mathbf{H}$ relates Cartesian spherical coordinates $\mathbf{X}$ and $\mathbf{X}^*$ under the collinearity constraint ($\times$: cross product):

$$\mathbf{X} \times \mathbf{H}\mathbf{X}^* = \mathbf{0}. \quad (6)$$

Developing and rearranging (6) for $M \in \mathbb{N}$ pairs of corresponding points in a single system leads to a linear matrix equation of the form $\mathbf{A}\mathbf{h} = 0$, with $\mathbf{h} \in \mathbb{R}^9$, the vector of the nine elements of $\mathbf{H}$, and $\mathbf{A} \in \mathbb{R}^{3M\times 9}$, classically solved by computing the estimate $\mathbf{h}_0$ of $\mathbf{h}$ as the kernel of $\mathbf{A}$ using the Singular Value Decomposition (SVD) [22]. Then, $\mathbf{H}_0$, made of the 9 elements of $\mathbf{h}_0$, serves as an initial guess for the homography optimization, minimizing the direct cost (Sec. 4.3).

4.2. Intensity model: global affine transformation

4.2.1. Multi-channel to gray level

An image maps pixel coordinates $\mathbf{u}$ to a vector of $N \in \mathbb{N}$ channels sampling the spectrum. Below, we recall the data conversions considered for alignment.

RGB to Gray For a grayscale image, $N = 1$. Considering grayscales are classically coded with 8 bits, we note the grayscale image mapping $I(\mathbf{u}) : \mathbb{R}^2 \longmapsto [\![0, 255]\!]^1$, the latter being a discrete interval. For an RGB image, $N = 3$ and we note the RGB image mapping $C(\mathbf{u}) : \mathbb{R}^2 \longmapsto [\![0, 255]\!]^3$.

Classically, the grayscale image $I_C(\mathbf{u}) \in \mathbb{R}$ is computed as the weighted sum of the three channels of $C(\mathbf{u})$. The weights are defined as $\mathbf{w}_C = [0.299, 0.587, 0.114]^\top$, following standards [23]. Thus, $I(\mathbf{u})$ is computed from $C(\mathbf{u})$ with:

$$I_C(\mathbf{u}) = \mathbf{w}_C^\top \, C(\mathbf{u}). \quad (7)$$

Hyperspectral to Gray Hyperspectral images offer a much sharper capture of the spectrum (high spectral resolution) than RGB images, leading to thousands of channels [4], *i.e.* $N \gg 3$. To simplify, we assume the hyperspectral image is represented as the mapping $S(\mathbf{u}) : \mathbb{R}^2 \longmapsto [0, 255]^N$ (directly or after pre-processing).

Then, $I_S(\mathbf{u}) \in \mathbb{R}$, the conversion of $S(\mathbf{u})$ to grayscale, is also a weighted sum of channels, but of uniform weight $w_S = 1/N$. So, with $\mathbf{1} \in \mathbb{R}^N$ (N-vector of ones) we define:

$$I_S(\mathbf{u}) = w_S \, \mathbf{1}^\top \, S(\mathbf{u}). \quad (8)$$

4.2.2. A linear transformation for exposure difference

The different camera hardware used at different capture time make intensities $I_S(\mathbf{u}^*)$ rather different from $I_C(\mathbf{u})$, even for corresponding $\mathbf{u}^*$ and $\mathbf{u}$. Assuming a linear transformation between both exposures modeled by an intensity scale factor $m \in \mathbb{R}$ and an intensity offset $s \in \mathbb{R}$, we express $\overline{I}_C(\mathbf{u})$, the transformed intensity $I_C(\mathbf{u})$, as:

$$\overline{I}_C(\mathbf{u}) = m \ I_C(\mathbf{u}) + s. \quad (9)$$

4.3. Direct cost and optimization

The direct alignment algorithm is designed to optimize for parameters of both geometric (Sec. 4.1) and intensity (Sec. 4.2) models, minimizing the cost expressed as the difference between intensities $\overline{I}_C(\mathbf{u})$ and $I_S(\mathbf{u}^*)$ of all pairs $(\mathbf{u}, \mathbf{u}^*)$ of corresponding pixels, *i.e.* satisfying the homography parameters $\mathbf{h}$. We write the dependence of considered pixels in I_C to the homography and source pixel $\mathbf{u}^*$ as $I_C(\mathbf{h}, \mathbf{u}^*)$. Then, stacking the intensities of each image in vectors $\mathbf{I_C}(\mathbf{h}) \in \mathbb{R}^{N_s^r N_s^c}$ and $\mathbf{I_S}^* \in \mathbb{R}^{N_s^r N_s^c}$, respectively, the optimization problem is:

$$\left[\widehat{\mathbf{h}}, \widehat{m}, \widehat{s}\right]^\top = \underset{\mathbf{h},m,s}{\arg\min} \ \frac{1}{2} \left\| m \ \mathbf{I_C}(\mathbf{h}) + s - \mathbf{I_S}^* \right\|^2. \quad (10)$$

The above optimization problem is classically solved with the Levenberg-Marquardt algorithm [24]. Jacobians are computed with auto-differentiation, image gradients and interpolation at non-integer coordinates are computed with a bi-cubic interpolation of I_C. Parameters are initialized at iteration 0 as $m^{(0)} = 1$, $s^{(0)} = 0$ and $\mathbf{h}^{(0)} = \mathbf{h}_0$ (Sec. 4.1 and see Sec. 5 for the source of corresponding pairs of points).

5. RESULTS

5.1. Experiment setup

We applied the alignment algorithm of this paper on data acquired inside the Gothic cathedral of Amiens, France (Fig. 1). Hyperspectral and RGB images of tall and large (several meters) stained-glasses were captured in a narrow corridor. Alignment results are reported for three stained glasses: III (St Peter), XVIII (St Bishop) and South Portal (back).

We use the hyperspectral data of 2068 channels acquired by an imaging system comprising a single-point spectrometer (Ocean Optics Maya2000 Pro) connected with an optical fiber to a scanning head (Ocular Robotics RobotEye REHS25). This system was set on a tripod to keep a static base during the scanning. The inevitable variation of natural illumination during such a time is eliminated thanks to a quick extra scanned line [4]. In the experiments, we've selected $N = 663$ channels (among 2068) of the spectrum ranging from 400nm to 700nm approximately corresponding to the visible light also captured by the RGB full-view camera.

The RGB camera is an off-the-shelf Ricoh Theta V dual-fisheye camera used with its built-in High Dynamic Range (HDR) mode to avoid exposure issues. It outputs an equirectangular RGB image of 2688×5376 pixels. To ensure

the image sharpness, this camera is also used on a tripod but at a different time than the hyperspectral imaging system.

The alignment algorithm is implemented with C++ language using the OpenCV library [25] to detect and match keypoints for the linear estimation of the homography (Sec. 4.1) which matrix operations are handled with the Eigen library. The Ceres solver library [26] is used for the Levenberg-Marquardt optimization of the direct cost (10). Our implementation is shared (see abstract).

5.2. Point correspondences for initial H

We applied descriptor-based keypoint matching algorithms to automatically find the set of point correspondences between I_S and I_C, allowing the estimation of an initial homography $\mathbf{H}_0$ (see end of Sec. 4.1). Several keypoint matching algorithms have been considered among which AKAZE [27] and SuperGlue [28]. The most reliable matching sets have been obtained by the AKAZE features with upright descriptors matched by L2-norm. This version of AKAZE has shown to be more robust than the others to distortions of equirectangular images. However, there is still a minority of outliers that have classically been discarded by encapsulating the homography estimation within a RANSAC [22] process (Fig. 2).

5.3. Alignment results

An initial homography $\mathbf{H}_0$ warps the gray version I_S of an hyperspectral image (see an example with South Portal on Fig. 3a toward Fig. 3b) to be geometrically close to the gray version I_C the target RGB image (Fig. 3d). The remaining misalignment is minimized by the dense direct optimization described in Section 4.3 such that the optimal homography $\hat{\mathbf{H}}$ warps I_S with pixel accuracy to I_C (Fig. 3c). From $\mathbf{H}_0$ to the $\hat{\mathbf{H}}$, the cost (10) is divided by 6.8 for stained-glass III, by 6.1 for South Portal and by 6.4 for XVIII.

5.4. Evaluation

The images of differences between target I_C and the warping of I_S thanks to either initial $\mathbf{H}_0$ or optimal $\hat{\mathbf{H}}$ are a good tool to quantitatively evaluate the alignment quality. Indeed, only warping with $\mathbf{H}_0$ leads to obvious remaining misalignment, particularly visible on the top and bottom parts of stained-glass XVIII (Fig. 4c) and the top of III (Fig. 4a). Though present, it is less obvious on South Portal (Fig. 4b). Actually,

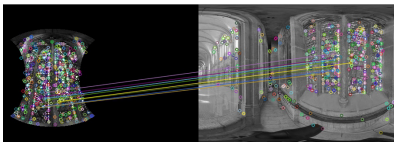


Fig. 2: RANSAC-based AKAZE feature matching for homography initialization (South Portal).

(a) I_S (b) $\mathbf{H}_0$ warps I_S (c) $\hat{\mathbf{H}}$ warps I_S (d) Target

Fig. 3: (a) Initial I_S of South Portal warped with (b) initial homography, (c) optimal homography solving (10). (d) Color version of the target image. Images are cropped afterwards.

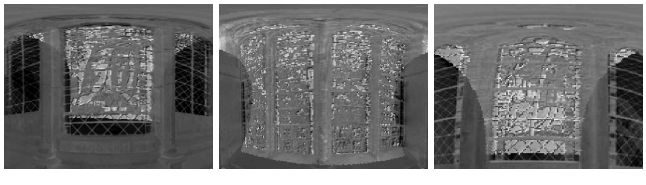
since the South Portal HSI to align has a larger FoV than those of III and XVIII (65 degrees vs. 41 for III and 39 for XVIII), $\mathbf{H}_0$ is estimated more reliably than for III and XVIII.

Inversely, the optimal $\hat{\mathbf{H}}$ corrects all the remaining misalignment encountered with $\mathbf{H}_0$. It is obvious for XVIII when looking at the top glass of trilobal shape, offset toward the bottom by several pixels with $\mathbf{H}_0$ (Fig. 4c) but almost perfectly aligned with $\hat{\mathbf{H}}$. Equivalent observations can be made for III and South Portal (Fig. 4d-4e vs. Fig. 4a-4b).

6. CONCLUSION

This paper has introduced the direct alignment of hyperspectral image of narrow FoV and full-view RGB image, *i.e.* optimizing a projective transformation minimizing the differences between pixel intensities. To make intensities comparable, RGB and hyperspectral channels have been converted to single channel first and a global exposure transformation model considered within the optimization process. Results show despite the very different FoVs and the nature of the data, AKAZE feature points can lead to a rough alignment, serving as a good initial guess for the direct optimization that reaches an accurate alignment. Our implementation is publicly shared to enable hyperspectral measurements integration in a HBIM.

(a) III (b) South Portal (c) XVIII



(d) III (e) South Portal (f) XVIII

Fig. 4: Alignment evaluation. Zoom on images of differences for all three stained-glasses: (a-c) with $\mathbf{H}_0$; (d-f) with $\hat{\mathbf{H}}$.

References

[1] Costanza Cucci and Andrea Casini, "Chapter 3.8 - hyperspectral imaging for artworks investigation," in *Hyperspectral Imaging*, vol. 32 of *Data Handling in Science and Technology*, pp. 583 – 604. Elsevier, 2020.

[2] Javier Tardaguila, Manfred Stoll, Salvador Gutiérrez, Tony Proffitt, and Maria P. Diago, "Smart applications and digital technologies in viticulture: A review," *Smart Agricultural Tech.*, vol. 1, pp. 100005, 2021.

[3] Costanza Cucci, Andrea Casini, Marcello Picollo, and Lorenzo Stefani, "Extending hyperspectral imaging from Vis to NIR spectral regions: a novel scanner for the in-depth analysis of polychrome surfaces," in *Optics for Arts, Architecture, and Archaeology IV*. 2013, vol. 8790, pp. 53 – 61, SPIE.

[4] T. Funatomi, T. Ogawa, K. Tanaka, H. Kubo, G. Caron, E. Mouaddib, Y. Matsushita, and Y. Mukaigawa, "Eliminating temporal illumination variations in whisk-broom hyperspectral imaging," *Int. J. of Computer Vision*, 2022, to appear.

[5] R. Pillay, J. Hardeberg, and S. George, "Hyperspectral imaging of art: Acquisition and calibration workflows," *J. of The American Institute for Conservation*, vol. 58, pp. 3–15, 02 2019.

[6] Agnese Babini, Sony George, Tiziana Lombardo, and Jon Yngve Hardeberg, "Potential and challenges of spectral imaging for documentation and analysis of stained-glass windows," *London Imaging Meeting*, vol. 2020, no. 1, pp. 109–113, 2020.

[7] F. Banfi, M. Previtali, C. Stanga, and R. Brumana, "A layered-web interface based on hbim and 360° panoramas for historical, material and geometric analysis," *The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W9, pp. 73–80, 2019.

[8] Ying Fu, Yongrong Zheng, Lin Zhang, Yinqiang Zheng, and Hua Huang, "Simultaneous hyperspectral image super-resolution and geometric alignment with a hybrid camera system," *Neurocomputing*, vol. 384, pp. 282–294, 2020.

[9] Y. Lin, Y. Zheng, Y. Fu, and H. Huang, "Hyperspectral image super-resolution under misaligned hybrid camera system," *IET Image Processing*, vol. 12, no. 10, pp. 1824–1831, 2018.

[10] Bruce D. Lucas and Takeo Kanade, "An iterative image registration technique with an application to stereo vision," in *Int. Joint Conf. on Artificial Intelligence*, 1981, vol. 2, p. 674–679.

[11] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2004, vol. 1, pp. 943–948.

[12] Pedro M.Q. Aguiar, "Unsupervised simultaneous registration and exposure correction," in *IEEE Int. Conf. on Image Processing*, 2006, pp. 361–364.

[13] G. Caron, E. Marchand, and E. Mouaddib, "Tracking planes in omnidirectional stereovision," in *IEEE Int. Conf. on Robotics and Automation*, May 2011, pp. 6306–6311.

[14] Christophe Collewet and Eric Marchand, "Colorimetry-based visual servoing," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Oct. 2009, pp. 5438–5443.

[15] Naveed Akhtar, Faisal Shafait, and Ajmal Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3631–3640.

[16] Rei Kawakami, Yasuyuki Matsushita, John Wright, Moshe Ben-Ezra, Yu-Wing Tai, and Katsushi Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *CVPR 2011*. IEEE, 2011, pp. 2329–2336.

[17] Hyeokhyen Kwon and Yu-Wing Tai, "Rgb-guided hyperspectral image upsampling," in *IEEE Int. Conf. on Computer Vision*, 2015, pp. 307–315.

[18] B. Aiazzi, L. Alparone, A. Garzelli, and L. Santurri, "Blind correction of local misalignments between multispectral and panchromatic images," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 1625–1629, 2018.

[19] Jaehyup Lee, Soomin Seo, and Munchurl Kim, "Sipsa-net: Shift-invariant pan sharpening with moving object alignment for satellite imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10166–10174.

[20] Soomin Seo, Jae-Seok Choi, Jaehyup Lee, Hyun-Ho Kim, Doochun Seo, Jaeheon Jeong, and Munchurl Kim, "Upsnet: Unsupervised pan-sharpening network with registration learning between panchromatic and multi-spectral images," *IEEE Access*, vol. 8, pp. 201199–201217, 2020.

[21] H.-E. Benseddik, F. Morbidi, and G. Caron, "PanoraMIS: An Ultra-wide Field of View Image Dataset for Vision-based Robot-Motion Estimation," *The Int. J. of Robotics Research*, vol. 39, no. 9, pp. 1037–1051, Aug. 2020.

[22] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, second edition, 2004.

[23] ITU, "Bt.601: Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios," 2011.

[24] Kenneth Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied math.*, vol. 2, no. 2, pp. 164–168, 1944.

[25] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[26] Sameer Agarwal, Keir Mierle, and Others, "Ceres solver," http://ceres-solver.org.

[27] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conf.*, 2013.

[28] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2020, pp. 4937–4946.