Stitching technique based on SURF for Hyperspectral Pushbroom Linescan Cameras

Manuel Villa*, Jaime Sancho*, Marta Villanueva*, Gemma Urbanos*, Pallab Sutradhar*,
Gonzalo Rosa*, Guillermo Vazquez*, Alberto Martin*, Miguel Chavarrias*, Luis Perez†,
Angel Nuñez†, Alfonso Lagares†, Eduardo Juarez*, Cesar Sanz*
* Research Center on Software Technologies and Multimedia Systems. Universidad Politécnica de Madrid (UPM)
† Instituto de Investigación Sanitaria Hospital 12 de Octubre (IMAS12), Madrid, España
{manuel.villa.romero, jaime.sancho, gemma.urbanos, marta.villanueva.torres, pallab.sutradhar, gonzalo.rosa.olmeda,
a.martinp, guillermo.vazquez.valle, miguel.chavarrias, eduardo.juarez, cesar.sanz}@upm.es
{apnunez,ljimenezr,alfonso.lagares}@salud.madrid.org

2021 XXXVI Conference on Design of Circuits and Integrated Systems (DCIS) | 978-1-6654-2116-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/DCIS53048.2021.9666155

***Abstract*—Hyperspectral imaging (HSI) has a crucial role in material classification tasks. The primary type of sensors in this field are linescan sensors, offering a high spatial and spectral resolution and improving the quality of the captured hypercubes. However, these cameras have inherent problems such as the need of dedicated hardware to move the camera over the scene and the necessity of pre-processing the raw captures in order to obtain the actual hypercube.**

This work makes an analysis and an adaptation of stitching techniques to apply them to the medical area. Also, it includes different methods to post-process the hypercubes obtained and it analyzes the SURF algorithm to improve processing times. The results obtained show that a stitching technique based on SURF is suitable for the medical area.

***Index Terms*—HSI, stitching, line-scan, SURF**

I. Introduction and motivation

Hyperspectral imaging (HSI) is becoming a vital factor in the field of material classification. This technology is based on the acquisition of images in slices of the wavelength spectrum, ranging from ultraviolet (UV) to infrared (IR). The result of this spectral sampling on each pixel of the image is called spectral signature. Therefore, the so-called hypercube is composed of spatial (x,y) and spectral (λ) information.

HSI was initially conceived for remote sensing applications such as agriculture [1], forestry planning [2], security and defence [3]. Recent developments in HSI have led to extending its uses to food quality [4] or medicine [5] areas.

The sensor technology to capture this type of images has evolved over the last decade where the most important ones are point-scanning, pushbroom, snapshot and wavelength scanning as shown in Figure 1.

The two main differences between these sensor technologies are (1) the need of performing spatial scanning as well as (2) the maximum number of achievable wavelengths. Point-scanning, linescan and pushbroom (referring to pushbroom wedge-linescan cameras) obtain a higher spatial resolution from a scene with hundreds of wavelengths. In contrast, snapshot cameras have a reduced spatial and spectral resolution but allow capturing hypercubes in a reduced time without any scanning process.

This work was supported by the Regional Government of Madrid (Spain) through NEMESIS-3D-CM project (Y2018/BIO-4826).

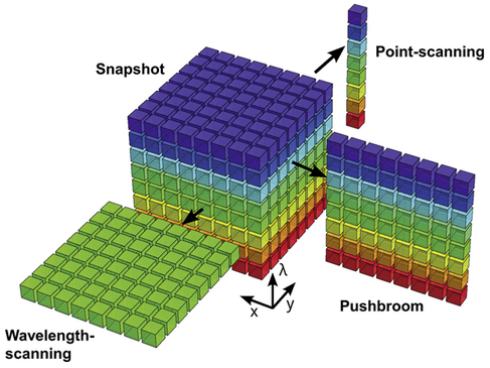


Fig. 1: Sensors types for HSI.

The necessity of creating relative movement between the scene and the sensors entails two problems: (i) the movement must be precise to avoid motion artefacts in the hypercube, and (ii) a post-processing step is necessary to obtain the hypercube.

The post-processing stage is different in the linescan and pushbroom sensors. A scanning process with linescan sensors originates a set of spectral lines with all the spectral information contained in the scene. Then, only the alignment of these strips is needed, as the spectral information is already aligned. However, when using pushbroom sensors, both spatial and spectral information need to be aligned.

Few studies have been published on how to address spectral and spatial alignment problems on pushbroom sensors. Feng [6] presents a solution in the remote sensing field, describing a procedure for building a hypercube with a vast amount of captures of a cotton crop using a drone. Nevertheless, this proposal raises many questions about whether the method proposed could be employed in other fields. This paper seeks to address how to reduce spectral and spatial misalignments on pushbroom cameras. To do so, the method proposed in this paper is based on a previous work of Feng [6], adding some modifications to fulfill the constraints set by the application addressed in this work, namely cancer tissue classification

during neurosurgical procedures.

The main limitation in this field is the scanning and post processing time. On the one hand, the scanning time is limited to less than a minute to avoid the presence of blood in the craniotomy scene. On the other hand, the brain classification needs to be as quickly as possible, in order to not interrupt the surgical operation. A consequence of this limitation a need of minimizing the number of captured images arises, also gaining importance the optimization of the whole procedure. However, the quality of the final hypercube is needed to be good enough to generate a proper classification. Therefore the main contributions of this paper are the following:

- The analysis and modifications of a stitching technique for hyperspectral pushbroom cameras in the medical area.
- The proposal of post-processing alternatives to improve the quality of the obtained hypercube while optimizing computing times.
- The study of the proposals in terms of speed and quality trade-off.

This paper is organized as follows: Section II introduces the related works found in the state-of-the-art (SoA). Section III and Section IV explain how the utilized linescan camera works and the algorithm to extract the necessary descriptors of the images to align it. Section V describes the proposed method for HSI cube stitching. Then, in Section VI the conducted experiments and employed materials are described, leading to the Section VII where the results are presented and analyzed. Finally, Section VIII presents the main conclusions of the work.

II. Related work

There is a considerable amount of literature on how to build panoramic images from a set of individual captures. These works are usually focused on resolving the problem with RGB cameras. Chen et al. [7] proposed a robust method for automatic panoramic images obtained from an RGB camera mounted on a UAV. In their analysis of panoramic images, a nonrigid matching algorithm based on motion field interpolation is introduced. This work produces a natural panorama without visible parallax in the overlapping regions and reduces the ghost effect produced by mosaicing.

In other work, L. Xia et al. [8] applies a similar system based on homographic techniques for HS cameras in UAV images. They present a method to reduce the number of bands in the processing procedure to increase the speed of the system and also to reduce the size of the hypercube. It is important to remark that the authors do not offer any solution to the gaps that appear in the reconstruction of the bands, only focusing on reducing the total size of the hypercube.

Another work that makes use of HS cameras mounted over a UAV is the one proposed by Feng [6]. The authors develop a novel image alignment and stitching method for processing HS cube images. Unfortunately, it does not detail how to merge the images resulting from the proposed alignment procedure.

As can be seen, research has tended to focus on HS UAV imaging rather than HS medical imaging. The existing algorithm for HS images stitching proposed by Yuan et al. [9] is only applicable to medical HS images without parallax and with low complexity on the scene.

This paper is the continuation of a previous work [10], where a simple stitching procedure was employed to generate hypercubes from a sequence of raw captures. In that work, the images were not aligned, only overlapping them using a theoretical displacement, which in practice was not completely correct. This entailed faster processing times, at the cost of inter-band displacement.

III. Background: HS pushbroom linescan camera

The work described in this paper is based on pushbroom cameras. Then, it is crucial to understand its operation correctly. The linescan filter layout has a wedge design. The filters in the linescan layout are organized in bands of fixed height (i.e. 5 pixels in this work) over the entire width of the active area of the sensor (i.e. 2045 pixels). Note that the number of available bands does not coincide with the actual number of bands on the sensor. The previous idea is depicted in Figure 2.

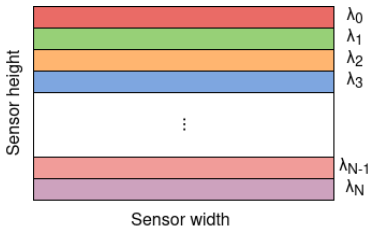


Fig. 2: Layout of filters.

A white-black calibration needs to be applied to the raw images. With the white-black calibration, the raw sensor values are converted into reflectance ones, expressed as percentages. This process is used to allow the comparison of different scenes with other light conditions. In addition, a spectral correction in the stitched image is needed to obtain the correct number of bands from the available sensor filters [11]. The line scan capture process is shown in Figure 3.

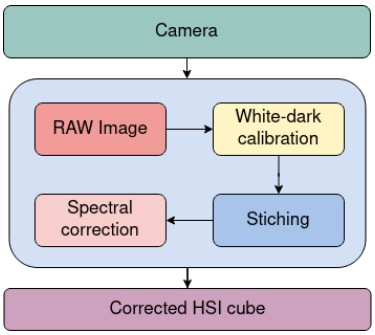


Fig. 3: Linescan pipeline.

As explained in Section I, to produce relative movement the camera needs to be moved over the scene. In order to capture the spectra of every point of an object, every point

of the subject must passes over each individual band. The implemented system consists of a precise linear actuator to create that relative movement. This actuator shall perform movements on the range of micrometres due to the height of bands.

IV. Background: SURF

The application of the SURF algorithm [12] is to find the correspondence between two images of the same scene or object. The algorithm extracts key points on the image at notable locations. Then, the region of each key point is represented as a descriptor. The descriptors must be invariant to color and noise. The descriptors extracted by the SURF algorithm allow estimating affine transformations, which are a combination of translation, scale and rotation transformations.

Scale transformations are usually implemented as image pyramids. The images are smoothed with a Gaussian filter and sub-sampled to achieve a higher level of the pyramid. The level of the pyramid and the size of the filter is configurable and has an expensive computational cost. The sensitivity of detecting key points and their rotation is also a parameter of the algorithm that may affect the quality and speed of the algorithm. A subsequent analysis of these parameters and their impact on the quality of the hypercube obtained after the stitching procedure will be detailed in Subsection V-D.

V. Analysis and Modifications to a Method for HSI Cube Stitching

In this work, the method proposed by Feng [6] is analysed and modified to fulfill the operation room constraints that are introduced in Section I. Taking this into account, the stitching procedure consists of the following steps: (i) estimating relative movement between consecutive images, (ii) merging and stitching of the captures that was customized to suit the medical requirements and (iii) post-processing the hypercube.

A. Estimating relative movement between images

This stage scheme is depicted in Figure 4.

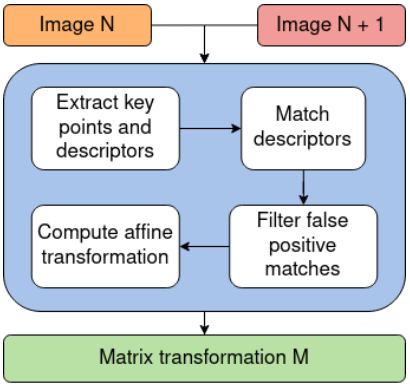


Fig. 4: Estimation of relative movement pipeline.

As can be seen in the Figure, consecutive images are the starting point. From each image are extracted several key points converted to descriptors with the help of the SURF algorithm. These descriptors are matched using a K-NN matcher (K = 2) meaning that two descriptors are proposed as possible matches for each one.

The matches obtained by the described method generate a high rate of false-positive matches. To reduce their presence, a filter based on the match length is needed. The length of the correct match between two descriptors shall be similar to the height of the spectral bands. Then if the length of a match is greater than the mean of the rest, it means that two diagonally arranged descriptors have been joined together. Hence, the shorter match proposed by the K-NN matcher is selected.

Once the rate of false-positive matches has been reduced, the affine transformation (i.e. Matrix M) that will align the two captures can be computed. The method used to compute the transformation is the RANSAC-based robust method [13].

To carry out this part of the algorithm, OpenCV [14] in C++ has been used to apply the SURF algorithm, extract the descriptors for the images, and estimate the affine transformation matrix. Nevertheless, to the best of authors knowledge, there is in the literature no process to join the captures as needed for this kind of pushbroom linescan cameras. The following section describes this process.

B. Merging and stitching procedure

Once all the M matrix have been computed, the alignment of the captures can be done. Figure 5 explains the process followed with each capture in order to build each band panoramic.

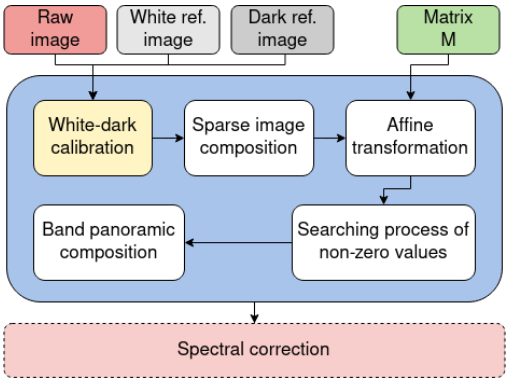


Fig. 5: Merging and stitching pipeline.

The first step, as was explained in Section III is the white-dark calibration, which needs a white and dark reference for each capture. In the second step, all the bands of the same capture are extracted to a black image with only a band, called sparse images. These images are necessary to apply the affine transformation, in particular, to allow the rotation over the images.

Due to the behaviour of the scanning process, the sparse image can be omitted in some of the bands. The exceptions depend on the capture number, i.e. the sparse image composition only is applied to the first band on the first capture, just as only the last band is used for the last capture. The resulting functioning is depicted in Figure 6.

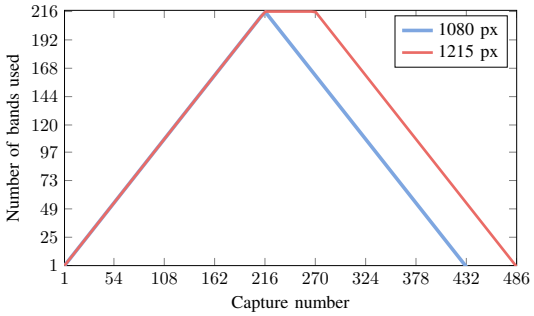


Fig. 6: Number of bands used in each capture.

Figure 6 differentiate two main parts. On the one hand, the part of the sequence where the object enters the scene (positive slope). On the other hand, the part of the sequence in which the object leaves the scene (negative slope). The number of bands changes if it is needed a larger image (red plot). Then appears an intermediate part where the whole object is present on the scene (non-slope). The height of the panoramic bands on the blue case is 1080 pixels for the camera used in this work (i.e. 216 captures of 5 pixels of height). On the red case, the result size will be 1215 pixels.

Then, the affine transformation is applied for each capture, using the matrix transformation M computed in Section V-A. Later, a scanning process of non-zero values in the sparse images is needed to locate the pixels which shall be copied to the panoramic band. This process is added if the band has suffered a rotation transformation due to the affine transformation. The searching step has been tuned to reduce the time execution, taking into account the expected position of the spectral band and a certain margin to start and end the search, as shown in Equation 1.

$$row = C_n * H_{band} \pm \epsilon \tag{1}$$

On Equation 1 are defined the starting and stopping row for the searching process, presented in Figure 5. To the starting row on the searching process, the ϵ factor is subtracted. Whereas to the stopping row, ϵ is added. This creates a margin over the position of the band to reduce the time of the searching process. Also, on the equation C_n refers to captures number, and H_{band} to the height of the band on sensors.

The last step for merging and stitching the HS bands is the most important one. This step raises the different bands and copies each translated band to its corresponding position on the panoramic band. The transformation between the translated and panoramic images depends on the capture number and the band being processed, as shown in Equation 2.

$$i_{pano} = H_{band} * C_n + i_{tra} - V_{size} + \lambda + 1 \tag{2}$$

The result of the equation is the row index in which the translated band has to be copied. H_{band} represents the height of the band on sensors, and the total number of rows are presented in the sensor is V_{size}. The equation depends on the capture number - C_n -, on the bands built in each iteration - λ - and the pixel position in the translated image - i_{tra} -.

$$i_{pano} = i_{tra} - V_{size} + H_{band} * C_n + (\frac{N_{imgs}}{2} - \lambda - 1) + 1 \tag{3}$$

Equation 3 applies the same transformation that the other one but to the set of captures where the object starts to leave the scene. As mentioned before, the number of bands used are reduced per iteration in this part of the algorithm. As a consequence of this change in the algorithm, and on the way that the captures are used, the λ of the Equation 2 changes to 3. N_{imgs} represents the total number of captures made in building 1080 pixels length bands.

C. Post-processing

To end the construction of the hypercube it is necessary to perform a crop process on the resulting bands of the previous step. The applied transformations introduce padding in the upper part of the bands to align the bands between them. The beginning of the region of interest is set out on Equation 4 when the estimated movement mean is an integer.

$$\frac{V_{size}}{H_{band}} \tag{4}$$

In the case of a non integer displacement, Equation 5 defines the suitable transformation.

$$\frac{V_{size}}{H_{band}} + \frac{\lambda}{2} \tag{5}$$

Some problems appear due to the necessity of capturing as few images as possible to minimize the scanning time. When the movement of the actuator is not precise enough, black lines appear on the images. Two prosecutions are studied to delete the strips from bands panoramas: (i) remove these strips, reducing the total length of each band panoramic, (ii) apply a median filter to the image to reduce the impact of dark strips. These methods are chosen due to its low computational cost against other filter types such as an bicubic interpolation. The impact of dark stripes is shown in Figure 7.

D. Speed-up of SURF implementation

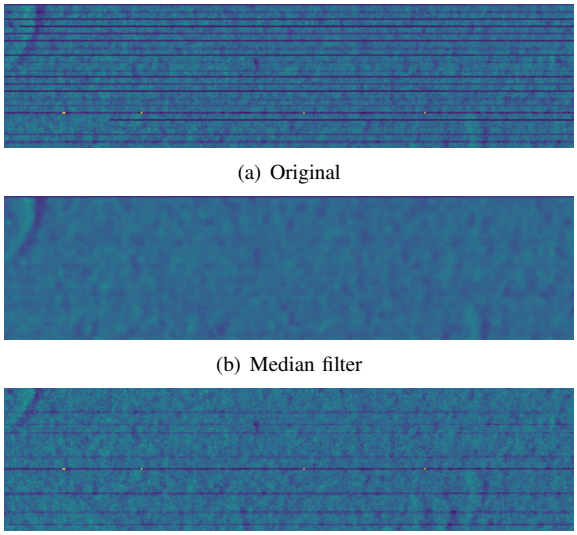
The aim of speeding-up the SURF usage is to reduce the processing time of the stitching procedure. Consequently, an analysis of the principal parameters of the SURF algorithm has been made. The analysis consists of searching the parameter values that maximize the algorithm performance.

The parameters analyzed are: (i) minHessian, which is related to the number of key points obtained with SURF, (ii) nOctaves, which influence the image pyramid used by the algorithm to estimate the scaling factor and (iii) nOctaves-Layers, which also affect to the scaling factor of the matrix transformation. The results obtained from this analysis will be detailed in Section VII.

VI. Test material

A. HS linescan camera

The Linescan camera employed in this work (XIMEA, MQ022HG-IM-LS150-VISNIR), features a spatial resolution



(a) Original

(b) Median filter

(c) Black stripes removed

Fig. 7: Comparative between post-processing methods.

of 2048×5 per band with 150 usable spectral bands, ranging from 470 to 900 nm with a spacing of 3 nm. This camera can perform up to 850 raw lines per seconds, depending on the exposition time set. The camera is mounted on a linear actuator (ZABER, X-LRQ300HL-DE51)with a resolution of $0.195\ \mu m$ and a whole range of $300\ mm$.

B. HS sequences

An external reference was used to determine whether the proposed method has fulfilled the requirements derived from its use in surgical environments. A PTFE-based polymer, with a known hyperspectral signature, has been employed. A single capture covering all the surface of the polymer was taken. The goal is to test the quality of the spectral signature obtained after the stitching method. Then, the spectral signature was extracted, taking a complete column of the image and applying the white-dark calibration and the spectral correction.

Other objects were scanned to test the inter-band displacement, the effect of post-processing methods and the speed-up achieved. The characteristics of the scenes are in Table I.

TABLE I: HS sequences to test the system

Scene	Exposure time	Distance to the scene	Number of captures
Plain scene	70 ms	25 cm	432
Polymer scene	70 ms	50 cm	432
Flattened polymer scene	70 ms	52 cm	432
Operating theater scene	80 ms	52 cm	432

VII. Results and discussion

The results of the experiments used to test the precision of the system are summarized in Tables II and III. Table II shows the inter-band displacement achieved with the stitching procedure described in this work. It is compared with an initial implementation of the stitching problem described by Rosa [10]. The values presented in the table refers to pixel displacements between the first and last hypercube bands.

TABLE II: Inter-band displacement

Scene	Stitching with SURF	Stitching with [10]
Plain scene	5	> 100
Polymer scene	32	> 300
Flattened polymer scene	11	> 100

Due to the alignment achieved using the SURF algorithm, the inter-band displacement was reduced drastically, preserving 60 - 70% of correlation with the reference signature of the polymer as shown in Table II. Then, the utility of SURF is demonstrated.

TABLE III: Analysis of post-process methods

Method	Co-variance	Spectral angle	Correlation diff.	Spectral angle diff.
One capture sign	78.45%	4.29°	Ref.	Ref.
Median filter	68.67%	4.04°	-9.78%	+0.25°
Removing black stripes	63.13%	4.52°	-15.32%	-0.23°

An additional experiment to analyze the impact on the spectral signature of the post-process methods proposed is conducted. The results are presented in Table III. The correlation and spectral angle present on it are computed over the spectral signature of the polymer. In the first case, the reference, the differences in the correlation are attributed to problems in the calibration process.In the other cases, using post-processing methods, a discordance between 10 % and 15 % are present in the co-variance with respect to the reference. These differences are attributed entirely to the employed method, meaning that the median filter is considered a good option to post-process the hypercubes.

The results obtained modifying the leading parameters of the SURF algorithm are submitted in the table IV and Figure 9. As can be seen, the speed-up obtained adjusting the minHessian parameter is the highest one, with an additional increase of the correlation with the reference signature. Furthermore, to test these values, an actual surgery sequence was analyzed and presented in Figure 8.

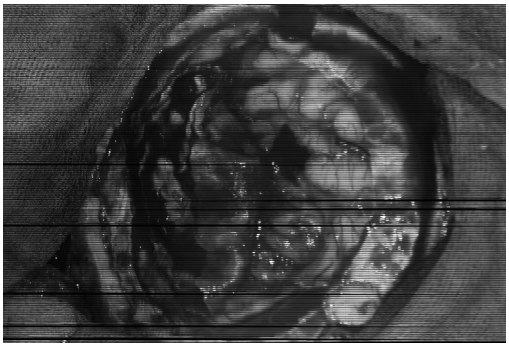


Fig. 8: Band after stitching procedure without post-processing.

The initial value of the minHessian parameter was 50 to ensure the correct functioning of the algorithm. However, the value 300 offers a pleasing result. with less than a 5% variation of system accuracy. The time showed in Figure 9 and Table IV refers to the total computational time of the first part of the algorithm by one core. Figure 9 detailed the impact over the execution time and the correlation with the reference spectral signature, varying the value of each parameter under study.

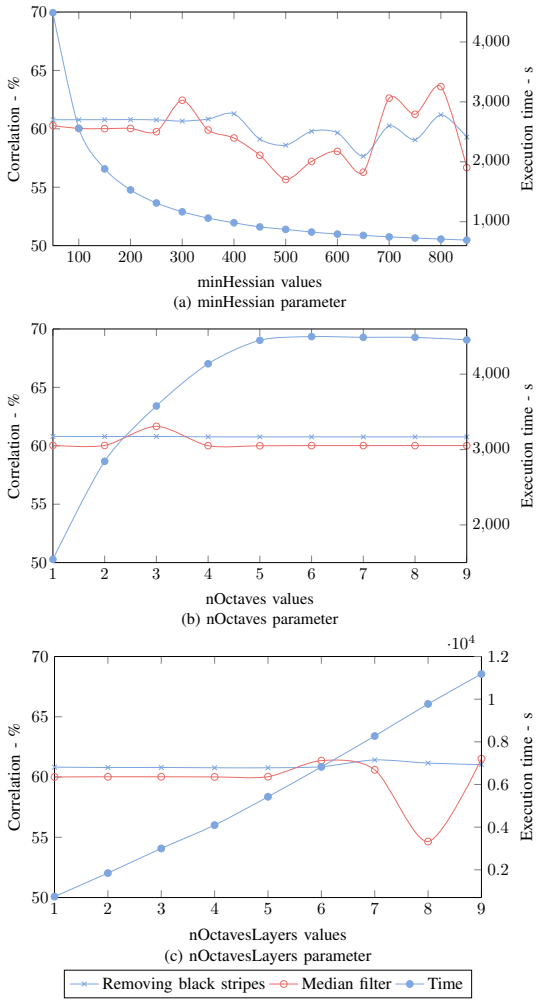


Fig. 9: Influence of SURF parameters over correlation and execution time per post-processing method.

Initially, the parameters of nOctaves and nOctavesLayers were set to 4 to test if it is necessary to apply scaling to the captures. Given the results shown in Table IV these parameters do not offer any advantage.Moreover, they are computationally expensive, so it was decided that the optimal parameters for these variables would be 1. Then, the speed-up achieved is $\times 12.1$ without affecting the performance of the system.

VIII. Concluding Remarks and Future Lines

The evidence from this study suggests that the adaptation of the stitching procedure suits the operating theatre requirements. However, further data collection is required to precisely determine the performance of the system in the clinical area. Despite all the dataset used in this works, was obtained simulating the surgeries conditions.

TABLE IV: Speed-up achieved on SURF algorithm per parameter

Parameter	Original time	Final time	Speed-up	Correlation difference
minHessian	4488.7 s	1164.9 s	$\times$3.9	+ 4.59%
nOctaves	4138.1 s	1543.4 s	$\times$2.7	+ 0.26%
nOctavesLayers	4096.2 s	744.8 s	$\times$5.5	- 0.12%

Considerable progress has been made concerning the inter-band displacement problem. In which the proposed implementation acquires a very low displacement, allowing to obtain good quality HS signatures. The displacement has been reduced on a magnitude order.

Also, the selection of the parameters to reduce the scanning process allows completing the composition of the hypercube at the same time that the actuator is ready for another sequence. The total speed-up is $\times$ 12, reducing the processing time in a multi-core system to less than a minute. Then, this will allow performing continuous scanning during surgery without processing limitations. However, research into reducing the processing time, making use of GPU acceleration is already underway. Regarding the post-process methods proposed in this works, the results show that the best method to perform it is the median filter. Further studies, which extend the methods to post-process and remove the black stripes from the hypercubes will need to be performed.

References

[1] D. J. Mulla, *Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps*. 2013.

[2] J. C. White and et. al., "Remote Sensing Technologies for Enhancing Forest Inventories: A Review," 2016.

[3] F. Dolce and et.al., "Earth observation for security and defense," *Handbook of Space Security: Policies, Applications and Programs*, pp. 705–731, 2020.

[4] D.-W. Sun, *HS imaging for food quality analysis and control*. Elsevier, 2010.

[5] M. A. Calin and et.al., "HS imaging in the medical field: Present and future," *Applied Spectroscopy Reviews*, vol. 49, 08 2014.

[6] A. Feng and et. al., "Evaluation of cotton emergence using uav-based narrow-band spectral imagery with customized image alignment and stitching algorithms," *Remote Sensing*, vol. 12, no. 11, 2020.

[7] J. Chen, Q. Xu, L. Luo, Y. Wang, and S. Wang, "A robust method for automatic panoramic uav image mosaic," *Sensors*, vol. 19, no. 8, 2019.

[8] L. Xia and et. al., "Stitching of hyper-spectral uav images based on feature bands selection," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 1–4, 2016.

[9] C. Yuan and et. al., "A multi-dimensional HS image mosaic method and its acquisition system," pp. 1–6, 2018.

[10] G. Rosa and et. al., "HS images acquisition: an efficient capture and processing stitching procedure for medical environments," pp. 1–6, 11 2020.

[11] XIMEA, "HS imaging data correction (Online resource)." Accessed 20-5-2021 https://www.ximea.com/support/attachments/5981/SpectroNet-2016-03-Ximea-V02.pdf.

[12] H. Bay and et. al., "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008. Similarity Matching in Computer Vision and Multimedia.

[13] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, 1981.

[14] Itseez, "Computer vision library."