73^{rd} International Astronautical Congress (IAC), Paris, France, 18-22 September 2022

IAC–22–A6.IP.39.x74064

VIS-TIR cameras data fusion to enhance relative navigation during In Orbit Servicing operations

Alessandro Colombo*, Gaia Letizia Civardi[1†], Michele Bechini[1‡], Matteo Quirino[§]and Michéle Lavagna[¶]

The paper proposes an effective and robust approach to cope with operating in proximity of non-cooperative orbiting objects for In Orbit Servicing. To improve vision-based relative navigation systems, the proposed approach exploits an image processing chain on visible (VIS) and thermal infrared (TIR) images, by firstly aligning the two source images and then fusing the data with pixel-wise image fusion methods to obtain complementary and more informative results. For each step of the image processing chain, multiple methods were investigated and evaluated both qualitatively and quantitatively. The lack of datasets for VIS and TIR images acquired on orbit under proximity manoeuvring scenarios, drove towards an in-house developed simulator, using VEga Secondary Payload Adapter (VESPA) as the target case study.

keywords: Sensor Fusion, Advanced Image Processing, Vision Based Navigation, Rendering Tool, Thermal Infrared Imaging.

1. Introduction

In the past decades, the interest in on-board autonomous relative navigation has grown intensively in the research community, with the focus of proposing effective approaches to cope with operating in proximity of uncooperative orbiting objects. Laser or camera-based sensors are a common choice for relative state reconstruction and navigation, and a mandatory one when one object might not be cooperative, as in the case of on-orbit servicing demonstrators (OOS), including active debris removal. Artificial uncooperative targets are here considered, being the most challenging scenario and constraining towards a robust solution leaning on the chaser capabilities only. In such instances, state-of-art onboard reconstruction of the chaser-target state vector relies mostly on images acquired in the visible spectra, and indeed solutions related to visible cameras (VIS) have been widely studied and practically applied in the context of uncooperative/cooperative rendezvous [1, 2]. However, visible imaging strongly depends on illumination conditions, consequently influencing the navigation solution accuracy and robustness and restricting operations planning. Nevertheless, OOS missions can be severely limited if illumination constraints for correct imaging are included in the close proximity operations design and definition: target orbit beta angle and attitude history, solar aspect angle provoked by the chaser feasible fly-around, and the camera axis might lead to limited opportunity to properly detect and track the target itself with unacceptable either mission length or risk increase.

This work hence proposes to introduce the use of thermal infrared images (TIR), reliant only on the emitted radiance by the target and thus insensitive to illumination conditions, avoiding the bottlenecks produced by the latter. Thermal infrared images however present a lower resolution and poorer contrast with respect to visible ones, which in turn negatively affects image processing algorithms, as highlighted in [3]. This work employs pixel-level image fusion to obtain a more informative image to be fed to the subsequent Image Processing (IP) step to retain the complementary advantages of the different spectral bands. The integration of multi-sensor images can provide a complex and detailed scene representation, thereby increasing the accuracy of decision-making in subsequent tasks. However, to obtain successful results through the use of pixel-wise image fusion techniques, the images to be fused are required to be strictly geometrically aligned [4].

To cope with the problem an analytic approach of image registration is derived and successively applied to the source images in order to align them. Here, a

*Politecnico di Milano, Italy, alessandro43.colombo@mail.polimi.it
†Politecnico di Milano, Italy, gaialetizia.civardi@polimi.it
‡Politecnico di Milano, Italy, michele.bechini@polimi.it
§Politecnico di Milano, Italy, matteo.quirino@polimi.it
¶Politecnico di Milano, Italy, michelle.lavagna@polimi.it
[1] Main authors, equal contribution.

projective transformation is determined considering the different viewpoints and Fields of View (FoVs) of the two sensors based on the relative camera parameters.

Steps of the image processing chain are extensively assessed through both subjective and objective criteria to identify the most efficient techniques to be adopted within a multispectral navigation chain. In addition, this work presents a specifically developed rendering pipeline, producing VIS-TIR datasets of spaceborne artificial targets to support the development and testing of a vision-based multispectral navigation chain. Figure 1 shows a schematisation of the work.

This paper is structured as follows. Sec. 2 presents a literature review about thermal infrared image rendering, image registration and multispectral image fusion techniques; Sec. 3 introduces the implemented rendering tool. In Sec. 4 and Sec. 5 the implemented techniques for image registration and image fusion are described together with the adopted performance metrics. Sec. 6 presents the results of the image processing chain, highlighting the best suited methods for onboard implementation. Conclusions and hints for future developments are finally reported in Sec. 7.

2. Literature Review

2.1 *Image rendering*

Synthetic VIS image rendering is a well-known task that is achieved via ray-tracing. Ray tracing is a rendering technique that relies on the concept of evaluating and simulating the path of view lines from the light sources to the virtual object. By tracking every ray from the light source and simulating the physics of the light, ray tracing techniques allow the computation of the colour intensity of the related pixels and can generate artificial images with a high degree of accuracy [5]. As most of the tracked rays never enter the eye/camera, some rendering engines save computational time by exploiting backwards ray-tracing techniques in which rays are instead traced from the camera into the scene, projected onto the object and lastly tracked to the nearest light source.

The open-source software, Blender, is used for this paper due to its high flexibility and high-quality outputs, employing the model of a pinhole camera for the entirety of the work. Concerning spaceborne synthetic yet realistic image datasets, the only currently available are the SPEED [6] and SPEED+ [7] datasets and the multi-purpose datasets by Bechini et al. [8–11], but an algorithm tailored to the VIS-TIR image fusion requirements is still needed, hence it has been decided to develop also the VIS image rendering tool together with the TIR image rendering one.

Thermal imaging is a process where a thermal camera captures and creates an image of an object by using infrared radiation emitted from the object. The amount of radiation emitted by an object increases with temperature; therefore, thermography allows one to see temperature variations and gradients. Thermography applications can be found in a variety of fields, from surveillance and military uses to scientific and medical purposes. Concerning the TIR images, TIR-based navigation is still an emerging topic for spaceborne applications, and thus thermal-infrared rendering has not been widely investigated within the research community. Few approaches exist, that tackle the problem in different ways, like [12, 13]. The approach exploited in this work and developed in [14] uses a high detail finite volume thermal model [15] of the object to then evaluate the radiative flux received by the sensor as to simulate the realistic output of a thermal imaging camera.

2.2 *Image registration*

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns two images: the reference and sensed images. Image registration includes a variety of methods and has been widely used in many fields, including computer vision [16], medical image analysis [17], and remote sensing [18]. Specifically, the registration of infrared and visible images refers to the problem of multimodal registration. A broad overview of the registration methods can be found in the literature [19]. In general, registration methods can be classified into two categories, i.e., area-based and feature-based methods. Area-based methods deal directly with the intensity values of entire original images; for example, minimizing the total distance between the pixel correspondences under a certain metric. Since the information contained in infrared and visible images differs, area-based methods are usually not particularly suitable. Feature-based methods first extract two sets of salient structures (e.g., feature points) and then determine the correct correspondence between them and estimate the spatial transformation accordingly, which is further used to align the given image pair. Compared to area-based methods, feature-based methods are more robust against typical appearance changes and scene movements and are potentially faster if implemented

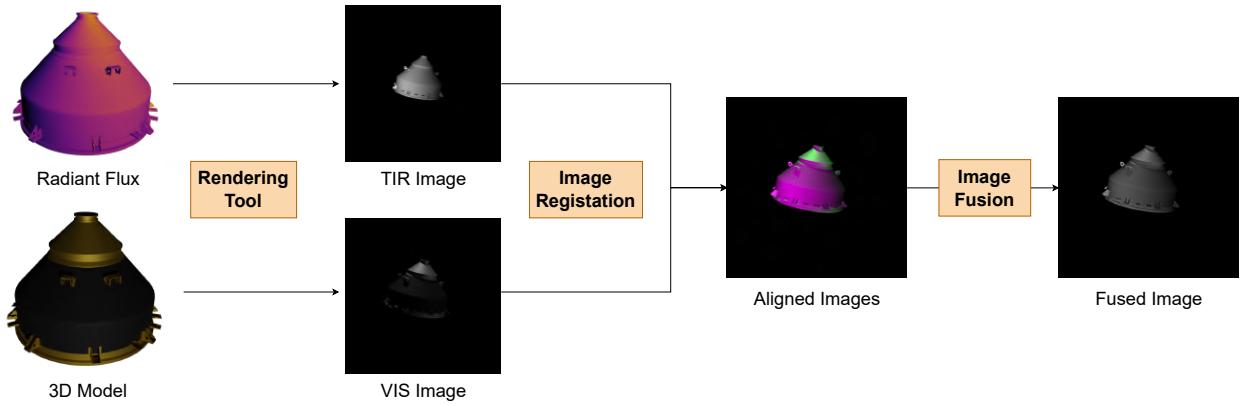


Fig. 1: Rendering pipeline and image processing chain workflow.

correctly. For infrared and visible images features that represent salient structures are preferred, like edge maps exploited in [20]. Most common image registration problems can be solved through the use of homography matrices. The latter relates the transformation between two planes (up to a scale factor). In the case of a planar scene, images will be related by homographies, no matter whether the views are obtained through a rotation or a translation. In the generic case of non-planar but rigid objects, a direct 2D transformation won't yield perfect results, as the issue of non-planarity can cause features to be present in one image while not being visible from the other view. For this reason, the problem is often tackled by exploiting complex optical flow algorithms or through 3D reconstruction and point-cloud matching. Since the aforementioned techniques would not be a viable strategy for current autonomous navigation systems, as they are computationally expensive, the problem is here investigated by working with the known information for the pair of calibrated sensors. The method exploits analytical expressions for the solutions to the problem, instead of the traditional numerical procedures, to derive a projective transformation to be applied to the VIS image in order to improve the alignment with the reference TIR image.

2.3 *Image fusion*

Image fusion is a technique whose aim is to exploit the strengths of sensors operating in different spectra to generate a robust and informative image that can ease the subsequent processing phase. Images of different types, such as visible, infrared, computed tomography (CT), and magnetic resonance imaging (MRI), are good source images for fusion. Among the combinations of these types, infrared and visible source images present characteristics that are inherent in nearly all objects and share complementary properties, thus producing robust and informative fused images. Fusion algorithms have been used in a wide range of application fields, such as object recognition [21], detection for surveillance [22] and remote sensing [23], yet they have never been applied in the context of spaceborne navigation. Different pixel-level image fusion algorithms exist and they can be grouped according to their baseline theory, as highlighted in [4]. The main categories considered for this paper are:

- **Multi-scale transform-based methods**: the source images are first decomposed into components at different scales, using methods such as pyramid transformation, wavelet transform, or edge-preserving filters. The multi-scale representations of the VIS and TIR images are then fused according to a given fusion rule. Lastly, the fused image is obtained using the inverse multi-scale transform on the fused representations.
- **Subspace-based methods**: these methods aim to project a high-dimensional input image into low-dimensional spaces or subspaces. Some of the most common techniques are Principal Component Analysis (PCA) and Independent Component Analysis (ICA), which transform correlated variables into uncorrelated ones called principal components. Other methods exist, such as Non-negative Matrix Factorization (NMF), however, it is time-consuming and has low computational efficiency, and thus it has been discarded.
- **Saliency-based methods**: visual saliency

is defined as the subjective perceptual quality which makes some pixels stand out from their neighbours and thus attract our attention. Within the field of multispectral image fusion, visual saliency can be used either to compute fusion weights or to extract salient objects from the background, for instance within the context of target detection and recognition while preserving the integrity of the salient object.

All the aforementioned methods present both strengths and weaknesses, and thus it is desirable to combine their advantages to improve image fusion performance. Different ways of combining existing principles exist, such as hybrid multi-scale transform and saliency or multi-scale transform and sparse representation, for example as in [24]. In addition to the described techniques, other types of pixel-wise infrared and visible image fusion methods exist, such as entropy, Markov random field, morphology, and infrared feature extraction and visual information preservation. With regards to this work, neural network and sparse representation-based methods have been discarded since they both require a large image database to be implemented, which is not currently available, and introduce in the whole navigation chain a huge computational overhead.

3. Image Rendering

The image generation algorithm has been executed entirely on Blender after importing the target into the built environment. Two major assumptions have been used in the rendering process of TIR and VIS datasets, with both being justifiable by literature review methods and personal reasonings. The images are in fact rendered noise-free and with a completely black background. The first is a viable way to increase the flexibility of the method, as noises can be easily added in post-processing to obtain realistic results like described in [14]. The absence of background has been instead used for different reasons. First of all, cases of images without a background can be found in real data. Furthermore, adding the Earth into the background would have resulted in an unreasonable increase in complexity for both the rendering pipeline and image fusion processes, when both are at an early stage of development. Cameras parameters are reported in Table 1. As to consider the sole effect on image misalignment due to differences in instrument positioning and FoVs, other TIR camera parameters were set to coincide with those of the VIS camera. Images for the VIS dataset are reproduced at different time steps of an entire orbit of the target so as to consider the full range of illumination conditions.

Table 1: Cameras characteristics.

VIS	
Array Size	1024 x 1024 px
FoV	35.45° x 35.45°
Focal Length	17.6 mm

TIR	
array size	1024 x 1024 px
FoV	45° x 45°
Focal Length	17.6 mm

Fig. 2: Example of rendered VIS image.

3.1 *Thermal-infrared image rendering*

The approach here presented starts from a high detail finite volume thermal model, provided following [15], characterized by a high level of detail in temperature field and geometry, represented in Figure 3. In order to replicate the thermal sensor output, the temperature field has to be converted into its corresponding infrared radiosity field, that is the actual energy received by the sensor. Neglecting all the reflections of the object, the expression for the radiant flux emitted by one face of the object mesh and received by one pixel of the thermal sensor reads:

$$Q_{f-p} = A_f \, F_{f-p} \, \varepsilon \, \sigma \, T_f^4 \qquad [1]$$

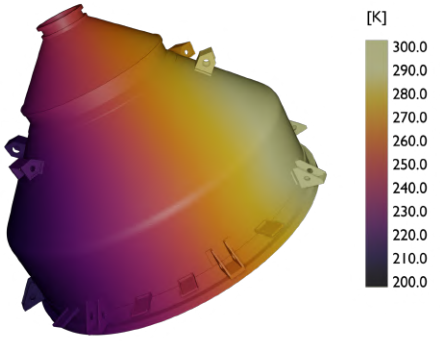


Fig. 3: VESPA temperature field.

Considering the radiant flux over the area of the respective face of the mesh, the expression can be rewritten as:

$$q_{f-p} = \frac{Q_{f-p}}{A_f} = F_{f-p}\,\varepsilon\,\sigma\,T_f^4 \qquad [2]$$

Where F_{f-p} is the view factor between a facet of the object mesh and a generic pixel of the camera. Assuming that the camera is far enough away from the object so that there is no difference in the view factor between a mesh face and different camera pixels, the view factors can be calculated for each face with the discrete form [15]:

$$F_{f-c} \approx \frac{(\underline{\hat{\mathbf{n}}}_c \cdot \underline{\mathbf{s}}_{cf})(\underline{\hat{\mathbf{n}}}_f \cdot \underline{\mathbf{s}}_{fc})}{\pi\, S^2} A_c \qquad [3]$$

Where A_c is the area of the thermal camera, $\underline{\hat{\mathbf{n}}}$ is the surface normal vector, $\underline{s}_{ij} = \underline{\boldsymbol{r}}_j - \underline{\boldsymbol{r}}_i$ represents the relative position vector between points belonging to the j-th and i-th surface respectively, while S is its magnitude. The computation must be performed for each external face of the mesh of the finite volume thermal model. Once the view factors are computed it is possible to take the temperature of each face along with the corresponding view factor and compute the radiance emitted by each face towards the thermal camera with Equation (2).

For TIR images generation, therefore, for each camera position the radiance field is mapped onto the mesh in Blender as a texture based on the model of a Lambertian emitter, eliminating the need to use light sources, which would imply the erroneous presence of visible features such as shadows. The same ray-tracing techniques exploited to obtain the synthetic VIS images are finally used for the conversion of the radiant flux field into the respective digital number [DN] in the rendering process, emulating the working principle of a real thermal camera.

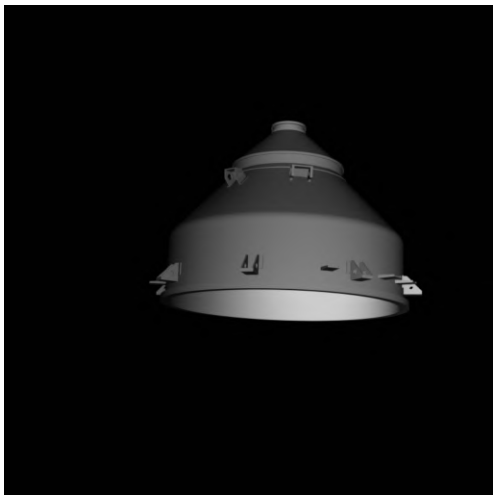


Fig. 4: Example of rendered TIR image.

4. Image Registration

To model a realistic scenario, the cameras have been assumed to be calibrated and with known relative position and orientation between the two, namely $\underline{\mathbf{t}}$ and $\mathbf{R}$.

Initial step of the registration process lies in the matching of the ratios and the visible portions of the images. This is done by firstly equalizing the FoVs of the images through resampling and successively by cropping the bigger frame in order to match the resolution of the images. Two strategies stem from here, either to upscale the frame with greater FoV, the TIR image in this case, or to downscale the VIS image with lower FoV. The first case guarantees to maintain a higher resolution at the cost of a slightly lower quality of the TIR image caused by the upscaling process. As this choice resulted in the most successful for further image processing steps [14], it was selected for this step of image registration. The scaling factor is obtained as ratio of the TIR and VIS camera FOVs:

$$f_{xy} = \frac{\tan(\mathrm{FOV}_{IR}/2)}{\tan(\mathrm{FOV}_V/2)} \qquad [4]$$

Bi-linear interpolation is exploited for the process, while the cropping is simply determined as the difference of the images dimensions after the resampling.

To comply with the effect of image misalignment, an analytic approach is then here investigated with the aim of determining a 2D perspective transformation induced by a plane between a set of corresponding image points $\underline{\mathbf{x}} \leftrightarrow \underline{\mathbf{x}}'$. This type of transformation, called homography, maps $\underline{\mathbf{x}} = \mathbf{H}\underline{\mathbf{x}}'$, with this

mapping being expressed by a general non-singular linear transformation:

$$\underline{\mathbf{x}}' = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \underline{\mathbf{x}} \quad [5]$$

Each view has an associated camera matrix, **P**, **P'**, combination of intrinsic and extrinsic parameters, mapping 3D point $\underline{\mathbf{X}}$ as $\underline{\mathbf{x}} = \mathbf{P}\underline{\mathbf{X}}$ in the first view, and $\underline{\mathbf{x}}' = \mathbf{P'}\underline{\mathbf{X}}$ in the second.

Suppose the camera matrices are those of a calibrated stereo rig with the world origin at the first camera.

$$\mathbf{P} = \mathbf{K}[\mathbf{I}|\underline{\mathbf{0}}] \qquad \mathbf{P}' = \mathbf{K}'[\mathbf{R}|\underline{\mathbf{t}}] \quad [6]$$

and the world plane π, with coordinates $\pi = (\underline{\mathbf{n}}^T, d)$ so that for points on the plane the relation $\underline{\mathbf{n}}^T\underline{\mathbf{X}}+d = 0$ holds. Then, following [25] formulation for a planar scene:

$$\mathbf{H} = \mathbf{K}'(\mathbf{R} - \frac{\underline{\mathbf{t}}\ \underline{\mathbf{n}}^T}{d})\mathbf{K} \quad [7]$$

where $\underline{\mathbf{n}}^T$ and d denote the plane normal and the distance from the camera respectively. The distance is roughly determined by finding the intersection point of the optical axes of the cameras.

$$d = \|\mathbf{R}^T\underline{\mathbf{t}}\| \tan\left(\arccos\left(\frac{\mathbf{R}^T\underline{\mathbf{t}} \cdot \hat{\underline{\mathbf{z}}}_{\mathrm{IR}}}{\|\mathbf{R}^T\underline{\mathbf{t}}\|\|\hat{\underline{\mathbf{z}}}_{\mathrm{IR}}\|}\right)\right) \quad [8]$$

Whereby the plane normal is assumed to be equal to the opposite of the VIS optical axis $\mathbf{n}^T = -\hat{\underline{\mathbf{z}}}_{\mathrm{V}}$ due to the target's surface information being unknown. Finally, Equation (7) can be applied to the VIS image in order to register it to the TIR one. A traditional features-based method is introduced to compare the results obtained for the registration. The following method is based on a simple brute-force ORB descriptors matching approach using Hamming distance as a measurement. To compensate for a large proportion of false matches in the data, the approach includes the RANdom SAmple Consensus (RANSAC) algorithm [26] to pick out matches which are true matches (inliers) versus false matches (outliers) and consequently increase the estimation accuracy of the homography.

4.1 *Performance Assessment*

Since the purpose of image registration is to align reference and sensed images, popular evaluation methods for the registration result are based on the quantification of the difference between the two source images, for example, depending on the displacement or the intensity level pixel-wise. However, infrared and visible images differ in the type of information retained leading to the choice of a pair of performance metrics usually exploited in the field of multi-modal image processing.

4.11 *Mutual Information*

Mutual information (MI) is a measure of image matching, that does not require the signal to be the same in the two images, as in the case of multi-modal image registration [27]. The MI between two random variables is computed through the Kullback-Leibler measure, defined as follows:

$$MI = \sum_{a,b} p_{A,B}(a,b) \log \frac{p_{A,B}(a,b)}{p_A(a)p_B(B)} \quad [9]$$

where $p_A(a)$ and $p_B(b)$ denote the marginal histograms of source images A and B, respectively. Similarly, $p_{A,B}(a,b)$ denotes the joint histogram of source image A with the second image B. A large MI metric means that considerable information is retained between the two images, i.e. that the two are more similar.

4.12 *Normalized Cross-correlation*

In signal processing, cross-correlation (CC) is a measure of similarity of two series as a function of the displacement of one relative to the other. In the field of image processing, in which the brightness of the image and template can vary due to lighting and exposure conditions, the images can be first normalized. For a given pair of images, NCC is defined following [27], as:

$$NCC = \sum_{x,y} \frac{(A(x,y) - \mu_A)(B(x,y) - \mu_B)}{\sigma_A \sigma_B} \quad [10]$$

where $A(x,y)$ and $B(x,y)$ are the pixels' intensities in images A and B at (x,y), respectively; μ_A and μ_B are their mean intensities, while σ_A and σ_B are the standard deviation intensities of A and B respectively. NCC(A,B) is larger when the two images are similar, with the maximum value achieved being equal to 1 in the case of NCC computed of a sample with itself.

5. Image Fusion Techniques

This section introduces the implemented image fusion techniques, which can be classified according to the criteria presented in Sec. 2. The presented methods are a subset of the selection of methods tested in [14] and all share the same assumption, i.e the source images should have the same resolution.

5.1 *Fusion Methods*

5.11 *Anisotropic diffusion-based fusion (ADF)*

The ADF algorithm can be regarded as a PCA-based technique. This implementation is largely based on the one described in [28]. Anisotropic diffusion is used to decompose images due to its capability of preserving edge information. Two layers are obtained, namely approximation and detail layer. The fused-based layers are obtained as a weighted superposition of the source images base layers, while detail layers are fused with the help of the Karhunen–Loeve (KL) transform, which is capable of transforming the correlated image components into uncorrelated ones. The KL transform can be practically implemented through the eigenvalue analysis of the two detail layers. Lastly, the fused image is reconstructed through a simple linear combination of fused approximation and detail layer.

5.12 *Image fusion using two-scale decomposition and saliency detection (TSFISD)*

The proposed implementation is inspired by the one presented in [29], with the main difference being the technique employed to compute the visual saliency maps. While in the original work median and mean image filters are employed, this version uses image convolution with a Scharr filter. The Scharr gradient reflects the significant structural features of an image, such as edges, outlines, and region boundaries and it is resilient with respect to image noise. A simple average rule is here used to perform base layer fusion.

5.13 *Image Fusion with Multi-scale Guided Filtering (MGFF)*

MGFF is a classic example of a hybrid multi-scale-based fusion method, developed on the basis of GFF [30]. Unlike its predecessor, in MGFF the guided filter is utilized in the decomposition process to obtain base and detail layers, taking advantage of its structure transferring property. Saliency and weight maps extraction is then performed with the latter being taken as the normalization of the first pixel-wise, saving computational effort. The whole process is iterated in a multi-scale decomposition and lastly, the fused image is reconstructed by combining base and detail layers with a weighted average.

5.14 *Infrared and visual image fusion through Infrared Feature Extraction and Visual Information Preservation(IFEVIP)*

The IFEVIP algorithm does not belong to any of the main categories described in Section 2 since it is not reliant on classic fusion methods. Its implementation, mostly based on [31], exploits quadtree decomposition [32] and Bézier interpolation [33] to firstly reconstruct the infrared background. The infrared bright features are extracted by subtracting the reconstructed background from the infrared image and then refined by reducing the redundant background information. To inhibit the over-exposure problem, the refined infrared features are adaptively suppressed and then added onto the visual image to achieve the final fusion image.

5.2 *Quality Metrics*

The performances of image processing algorithms for vision-based navigation strongly depend on the quality of the fused images, and thus the performance of the different fusion techniques should be evaluated both qualitatively and quantitatively. Subjective evaluation methods assess the quality of fused images according to the basis of human visual perception, such as artefacts or image distortion. Nevertheless, it is necessary to employ quantitative metrics to obtain a judging index that cannot be biased by observers or interpretation. It follows that reference-free criteria shall be adopted since it is not possible to compare the fused image with a reference ground truth image. The quality metrics used to evaluate the fusion algorithms are directly transposed from the work in [14]:

- Mutual Information (MI): measures the amount of information that is transferred to the fused output.
- Feature Mutual Information (FMI): measures the amount of feature information (edges, details) transferred to the fused image.
- Structural Similarity Index (SSIM): models loss of correlation, luminance and image distortion.
- Root Mean Square Error (RMSE): denotes the dissimilarity between the source images and the fused output; it is also the only metric for which a lower score equals a better result.
- Average Gradient (AG): quantifies the gradient information of the fused image, representing its detail and texture.

6. Image Processing Chain Applications

This section is dedicated to representing the results of the proposed image processing chain, firstly for image registration and successively for the applications of image fusion with the described methods. Given the emphasis put on a possible future implementation of an autonomous navigation model, all algorithms are also evaluated according to their computational cost. Image registration methods are developed on PYTHON while all the fusion techniques are implemented on MATLAB. All runs are made on a Intel® CoreTM i7-1185G7 CPU, with clock frequency of 3 GHz and 16GB RAM memory.

6.1 *Image Registration*

The image registration processes have been tested on the whole dataset and compared with non-registered images to better evaluate the performance results of the two algorithms. Quantitative results, depicted in Figure 5 and Figure 6, show almost break-even values between the analytically co-registered images and the original source ones, while also highlighting an abnormal behaviour of the features-based method performance, which achieves much lower scores than the other two curves for the majority of the frames. It is important to remember that the performance of image processing algorithms strongly depends on the quality of the results, and thus the quantitative scores of the different methods techniques should be paired with a qualitative investigation of the obtained images, assessing the quality of the outputs according to the basis of the human visual perception, such as artefacts or image distortion. Indeed, by observing the frames under investigation it is noticeable that those are cases of registered images obtained through the use of features matching which suffered from distortions caused by an erroneous estimation of the homography matrix. The algorithm requires at least four point correspondences in order to estimate the homography matrix. Even though the RANSAC [26] contributes to reducing the number of false matches, it may happen that the correspondences found belong to different surfaces of the object, leading to the estimation of a degenerate transformation matrix as shown in Figure 7. As already introduced, the VIS dataset is reproduced by simulating the lighting conditions over an entire orbit of the target. Frames for which illumination conditions are limited, the rendering tool generates more flatter

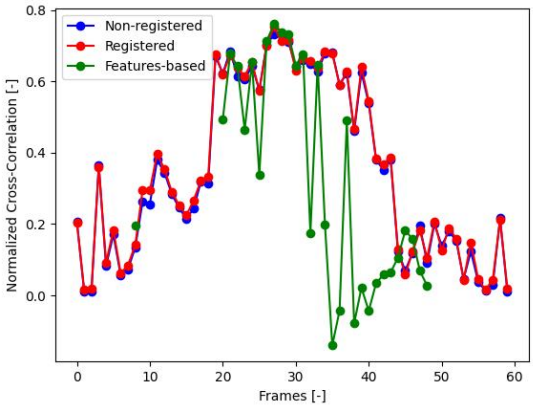


Fig. 5: Normalized Cross-Correlation.

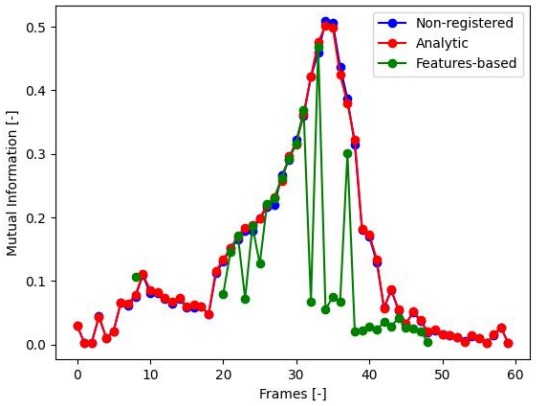


Fig. 6: Mutual Information.

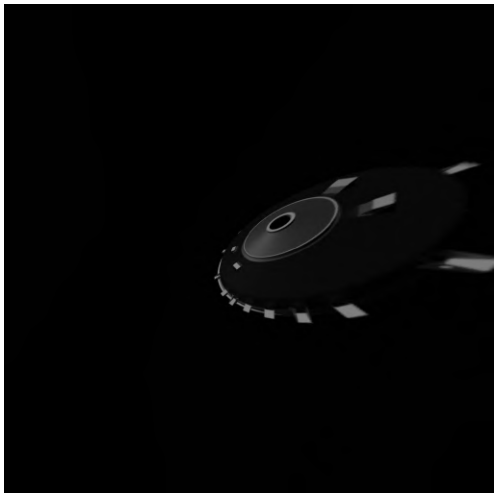


Fig. 7: Example of image distortion produced by the feature-based algorithm.

visible images in terms of contrast and visual features. Due to the high dependence on these factors, features-based algorithms, as the method proposed, are found to be incapable of producing a result for several images, due to the algorithm being unable of finding enough point correspondences to estimate the homography as it can be observed at the ends of graphs of the two metrics.

In contrast, the effectiveness of the analytic method remains unchanged depending on the lighting, as it does not operate on data obtained from the images to determine the projective transformation. Nevertheless, although the analytical method proved capable of reducing the misalignment effect, it did not substantially improve the results compared to the original source images due to the strong assumptions at the basis of the development of the approach. In particular, the analytic method proposed, results in a typical over-shooting of alignment transformation for part of the frames. This phenomenon is more evident in images in which the target distance with respect to the camera is lower. Indeed this is the case in which major differences are present between the two images as these frames also correspond to the cases in which multiple surfaces of VESPA are visible, as observed in Figure 8, which is typically in contrast with the exploitation of homographies for the process of image registration. On the other hand, the approximations

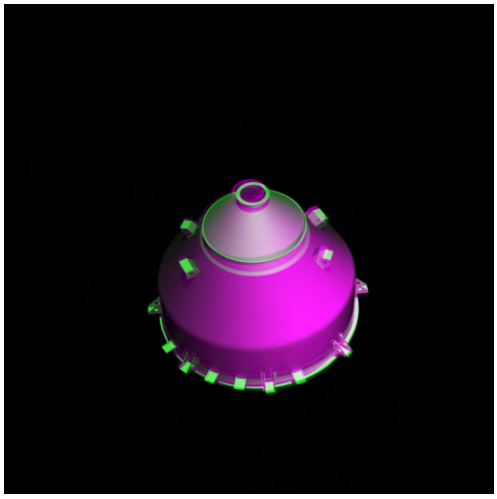


Fig. 8: Example of inaccurate image registration output.

used by forcing the planarity of the scene prove to be well-founded for cases where VESPA is seen in the distance, resulting in transformations closer to the 'exact' solution, giving the method higher scores.

Table 2 groups the average scores and computational time for both methods and the case where no transformation was applied to the source VIS images. In addition to the evident weaker overall performance of the features-based method, the computational time required to perform the registration is an order of magnitude higher than that of the proposed analytic approach, making it less suitable for onboard implementations.

Table 2: Image registration processes average scores.

Method	**NCC** [-]	**MI** [-]	**CPU Time** [s]
Non-registered	0.348	0.130	-
Analytic	0.355	0.131	0.0028
Features-based	0.321	0.129	0.033

6.2 *Image Fusion*

Image fusion methods are then tested on source images that have been co-registered analytically, with a batch of fused images with each algorithm shown in Figure 10. The quality metrics evaluated for the presented fusion algorithms are reported in Figure 9 in addition to the stated computational time required by each algorithm to perform the VIS-TIR image fusion. From a qualitative evaluation of the fusion outputs, satisfactory results were obtained over the whole dataset by all four algorithms. In particular, in frames where the image registration process did not retain the desired results, the effect of geometric misalignment of source images has been mitigated with success by all methods with the only minor exception being TSIFSD, which produces images with more evident visual differences. The poorer performance of TSIFSD is also confirmed by the scores obtained by the method in metrics indicating image similarity and correlation like MI, FMI and SSIM Table 3. From a qualitative point of view, the most robust method to the effect of misalignment resulted being ADF as it tends to typically merge the images more evenly and smooth out the visual differences in the fused output. This process comes with the drawback of producing images that are more flattened, hence losing details and texture in the fused result explaining why the method achieves the lowest score in AG in Table 3 among all the implemented methods. In Figure 9, an erratic behaviour can be observed in the SSIM values obtained from ADF, however, a qualitative inspection of the fused images revealed these values to be unjustified. In general, all algorithms perform with similar results across the majority of

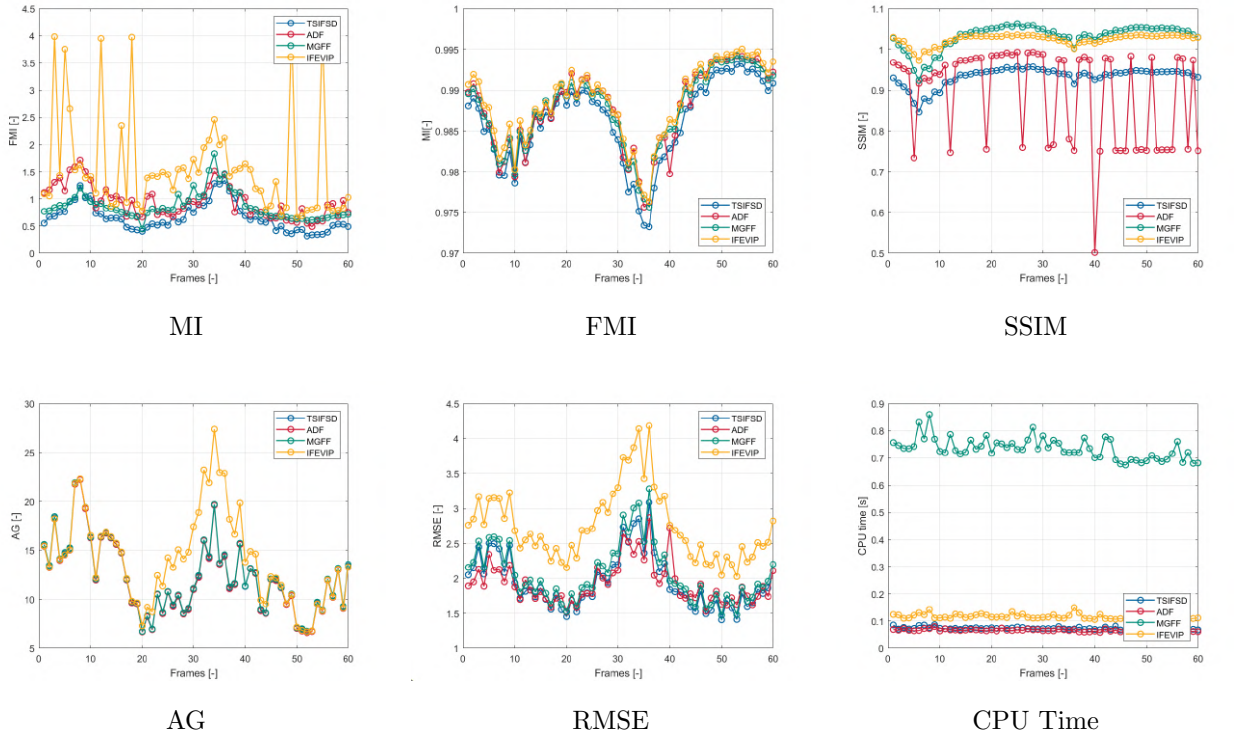


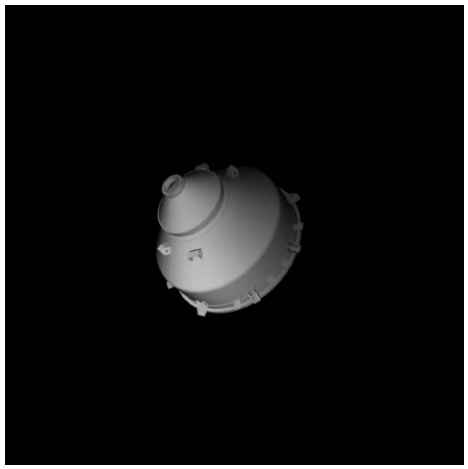
Fig. 9: Quality metrics and run time performance on the whole dataset.

quality metrics, with the only exception being IFEVIP. In fact, the latter is the only one among the listed image fusion methods that does not split the source images into different layers, while it operates directly only on TIR images to extract the bright feature information and superimpose it on the VIS images, resulting in higher average gradient values on the entire dataset. While this is a good indication in most cases, it comes with the issue of obtaining saturated images when the VIS source image is itself characterized by high intensity levels, which can occur in the presence of highly reflective surfaces when directly hit by the Sun rays, as stated in [14]. Simultaneously poorer performances are obtained for what concerns structural similarity since some visual features are lost in the fusion. Note however that these values are not fully representative of the fusion result similarity, as part of the TIR image is discarded in the fusion process, i.e. the information contained in the background. Whereas the problem of image saturation needs to be carefully considered, the method needs to be further investigated before it is discarded, as one cannot help but highlight the fact that the method gives valid results in the majority of cases. Indeed, a review of the bright features adaptive suppression step upstream of the process of adding them to the VIS image could further inhibit the overexposure effect and consequently reduce the risk of producing saturated fused images. As previously introduced, the computational efficiency is a parameter of equal importance with respect to quality metrics when evaluating the performance of the image fusion methods and a possible suitability for onboard implementations. The most computationally efficient methods are IFEVIP, TSIFSD and ADF. Although MGFF shows promising results in terms of the quality of the fused images obtained and ranking second on average on quality metrics, it might be too computationally intensive and hence considered not suitable for further investigations. Based on the previous observations and on the results obtained also from the quality metrics, the ADF and IFEVIP methods can be selected as the most promising options for future applications.

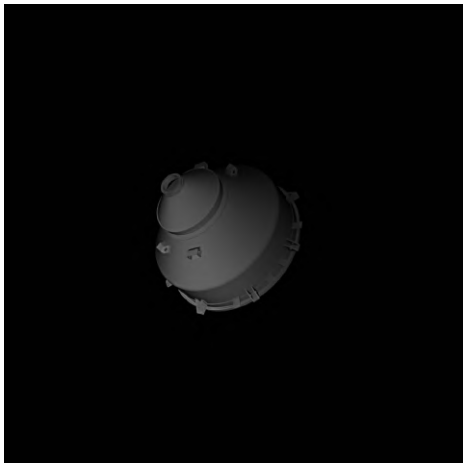
6.3 *Final Remarks*

For what concerns image registration, the results confirm the validity of the proposed analytic ap-

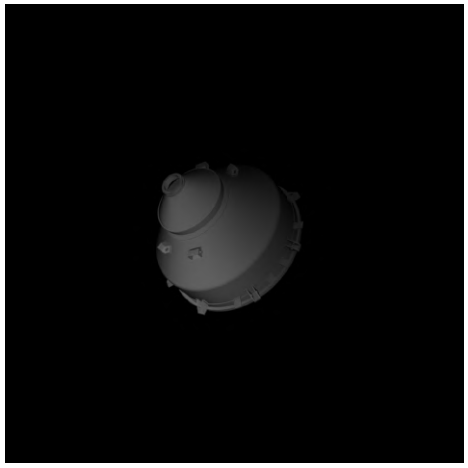
VIS source image



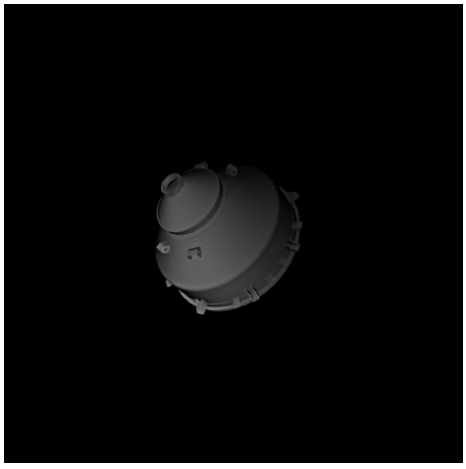
TIR source image



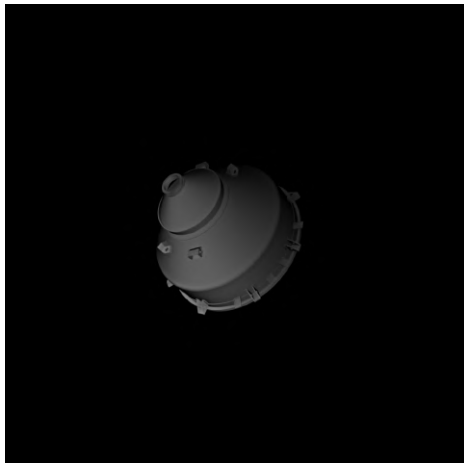
TSIFSD



ADF



MGFF



IFEVIP

Fig. 10: Fusion representative results of the implemented algorithms.

Table 3: Image fusion quality metrics and run time average scores.

Method	MI [-]	FMI [-]	SSIM [-]	AG [-]	RMSE [-]	CPU Time [s]
TSIFSD	0.6717	0.9862	0.9334	1.9548	12.2852	0.07356
ADF	0.9534	0.9877	0.8872	1.9319	12.2207	0.0661
MGFF	0.8696	0.9877	1.0314	2.0607	12.3493	0.7346
IFEVIP	1.5725	0.9885	1.0241	2.7532	14.1234	0.1163

proach, although the assumptions introduced lead to degenerate results in the case where the three-dimensionality of the object in view is evident. The features-based algorithm results to be beaten on the whole scale, both in terms of quantitative performance and qualitative evaluation of the co-registered images. It must be noted however that no attempt was made to try to optimize the algorithm for the specific case under consideration, and yet, the method did not considerably affect the computation cost of the entire image processing chain, and in the rare cases where a high number of co-planar key-points matching was achieved, the method proved capable of determining a better estimate of the projective transformation than that obtained analytically. This calls for a due further investigation of the numerical method, with a possible pairing with the analytic formulation to increase the robustness of the approach.

With regard to pixel-level VIS-TIR image fusion, all proposed methods were found capable of achieving the desired results of fused images both in the case of VIS images with good illumination and vice versa. Of the four, IFEVIP and ADF appear to be the best performing methods overall, as they also seem to be the most robust to misalignment effects. However, the two methods are not without their flaws, with the main problems of both being related to the intensity gradient information obtained in the fused image. While IFEVIP occasionally risks producing saturated images as the output of the fusion process, the fused images generated through the use of ADF, on the other hand, tend to be characterized by low contrast. Both issues are not unsolvable. In IFEVIP, the superimposition step of bright features on VIS images should be reviewed to make it more robust to the effect of over-exposure and consequently limit the risk of producing saturated fused images. Common contrast enhancement techniques with low computational impact could instead be applied to ADF fusion results so as to enrich the intensity gradient contained in the images and facilitate the operation of subsequent image processing algorithms.

7. Conclusions and Future Works

7.1 *Conclusions*

This paper presented the development and analysis of an image processing chain aimed to provide a foundation for the study of relative navigation about uncooperative artificial targets based on multispectral imaging sensors. In the absence of adequate databases in the literature on which to conduct the study, the problem of generating synthetic visible and thermal infrared images for artificial space-related targets is firstly tackled. The outputs produced for the rendering chain proved to be satisfactory, with the developed tool capable of recreating realistic scenarios for both the VIS and TIR datasets. These datasets have been exploited for the testing of the proposed image processing chain, characterised by an initial registration of the source images for each frame, in order to have aligned images to feed to the image fusion methods. Although for the case under consideration, an exact solution for image registration by direct application of projective transformations is not obtainable, this work has shown that the approximations obtained through the proposed approaches, in particular validating the analytic derivation, achieve acceptable results for subsequent implementation of further image processing steps. The application of pixel-wise image fusion proved suitable for the case at hand, producing promising results, especially in the case of poorly lit scenarios, which could be able to overcome the respective flaws of visible and thermal-infrared images. All four implemented algorithms have proved capable of achieving fusion results on the entirety of the dataset, with results indicating IFEVIP and ADF as the most promising methods, due to the low computational time, the optimal performance metrics and the high quality of produced images independently from the lighting conditions. Furthermore, despite the fact that in

this work the algorithms were only tested for noise-free images and without considering the effect of the smaller thermal-infrared sensor array size, the issues mentioned have already been addressed in a separate paper [14]. Based on these analyses, it is clear that infrared-visible image fusion has proven to be a promising approach for generating informative and detailed data on which subsequent vision-based navigation algorithms can be implemented.

Despite the promising results, several future developments can be individuated to improve the accuracy of the rendering tool and the flexibility of the image processing chain:

7.2 *Future work*

The presented rendering tool is still at an early stage of development, with great potential for growth, starting with the improvement of the thermal camera model with the inclusion of realistic and non-uniform thermal sensor gains and offsets. In addition, future work will focus on validating the proposed rendering approach, with a quantitative evaluation that can be conducted on natural celestial objects, such as asteroids, thanks to thermal infrared images obtained in past missions, such as Hayabusa2.

Finally, the image processing chain shall be included in the pose estimation pipeline, so as to verify the actual possible use of the output images within a visually-based navigation chain.

References

[1] Stefano Silvestrini, Margherita Piccinin, Andrea Capannolo, Michèle Lavagna, and Jesus Gil Fernandez. Centralized Autonomous Relative Navigation of Multiple Cubesats around Didymos System. *Journal of the Astronautical Sciences*, 68(3):750–784, September 2021.

[2] Manny R. Leinz, Chih-Tsai Chen, Michael W. Beaven, Thomas P. Weismuller, David L. Caballero, William B. Gaumer, Peter W. Sabasteanski, Peter A. Scott, and Mark A. Lundgren. Orbital Express Autonomous Rendezvous and Capture Sensor System (ARCSS) flight test results. In Richard T. Howard and Pejmun Motaghedi, editors, *Sensors and Systems for Space Applications II*, volume 6958 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 69580A, April 2008.

[3] Gaia Letizia Civardi, Margherita Piccinin, and Michèle Lavagna. Small bodies ir imaging for relative navigation and mapping enhancement. In *7th IAA Planetary Defense Conference, Wien, Austria*, 04 2021.

[4] Jiayi Ma, Yong Ma, and Chang Li. Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, 45:153–178, 2019.

[5] Peter Shirley and R Keith Morley. *Realistic ray tracing*. AK Peters, Ltd., 2008.

[6] Mate Kisantal, Sumant Sharma, Tae Ha Park, Dario Izzo, Marcus Märtens, and Simone D'Amico. Spacecraft pose estimation dataset (speed). *Zenodo*, February 2019.

[7] Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. Next Generation Spacecraft Pose Estimation Dataset (SPEED+). *Zenodo*, October 2021.

[8] Michele Bechini, Paolo Lunghi, and Michéle Lavagna. Spacecraft pose estimation via monocular image processing: Dataset generation and validation. In *9th European Conference for Aeronautics and Aerospace Sciences (EUCASS), Lille, France*, 7 2022.

[9] Michele Bechini, Paolo Lunghi, and Michèle Lavagna. Tango Spacecraft Dataset for Monocular Pose Estimation. *Zenodo*, April 2022.

[10] Bechini Michele, Lunghi Paolo, and Lavagna Michèle. Tango Spacecraft Dataset for Region of Interest Estimation and Semantic Segmentation. *Zenodo*, April 2022.

[11] Michele Bechini, Paolo Lunghi, and Michèle Lavagna. Tango Spacecraft Wireframe Dataset Model for Line Segments Detection. *Zenodo*, March 2022.

[12] Thermal Infrared Imager Assessment Study for the Asteroid Impact Mission | Nebula Public Library, April 2022. [Online; accessed 30. Apr. 2022].

[13] Margherita Piccinin, Gaia Letizia Civardi, Matteo Quirino, and Michéle Lavagna. Multispectral imaging sensors for asteroids relative navigation. In *71st International Astronautical Congress (IAC 2021), International Astronautical Federation, IAF, Dubai, United Arab Emirates*, 10 2021.

[14] Gaia Letizia Civardi, Michele Bechini, Alessandro Colombo, Matteo Quirino, Margherita Piccinin, and Michelle Lavagna. Vis-tir imaging for uncooperative objects proximity navigation: a tool for development and testing. In *11th International Workshop on Satellite Constellations Formation Flying (IWSCFF 2022), International Astronautical Federation, IAF, Milan, Italy*, 06 2022.

[15] Matteo Quirino, Luca Marocco, Manfredo Guilizzoni, and Michèle Lavagna. High energy rapid modular ensemble of satellites payload thermal analysis using openfoam. *Journal of Thermophysics and Heat Transfer*, 35(4):715–725, 2021.

[16] Bruce D Lucas, Takeo Kanade, et al. *An iterative image registration technique with an application to stereo vision*, volume 81. Vancouver, 1981.

[17] Jiayi Ma, Junjun Jiang, Chengyin Liu, and Yansheng Li. Feature guided gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration. *Information Sciences*, 417:128–142, 2017.

[18] Alexander Wong and David A Clausi. Arrsi: Automatic registration of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1483–1493, 2007.

[19] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

[20] Jiayi Ma, Ji Zhao, Yong Ma, and Jinwen Tian. Non-rigid visible and infrared face registration via regularized gaussian fields criterion. *Pattern Recognit.*, 48:772–784, 2015.

[21] Richa Singh, Mayank Vatsa, and Afzel Noore. Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognition*, 41(3):880–893, 2008.

[22] Praveen Kumar, Ankush Mittal, and Padam Kumar. Fusion of thermal infrared and visible spectrum video for robust surveillance. In Prem K. Kalra and Shmuel Peleg, editors, *Computer Vision, Graphics and Image Processing*, pages 528–539, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[23] G. Simone, A. Farina, F.C. Morabito, S.B. Serpico, and L. Bruzzone. Image fusion techniques for remote sensing applications. *Information Fusion*, 3(1):3–15, 2002.

[24] Yu Liu, Shuping Liu, and Zengfu Wang. A general framework for image fusion based on multiscale transform and sparse representation. *Information Fusion*, 24:147–164, 2015.

[25] Ezio Malis and Manuel Vargas. Deeper understanding of the homography decomposition for vision-based control. 01 2007.

[26] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, 1981.

[27] Zhe Zhang, Deqiang Han, Jean Dezert, and Yi Yang. A new image registration algorithm based on evidential reasoning. *Sensors*, 19(5), 2019.

[28] Durga Prasad Bavirisetti and Ravindra Dhuli. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform. *IEEE Sensors Journal*, 16(1):203–209, 2016.

[29] Durga Prasad Bavirisetti and Ravindra Dhuli. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Physics & Technology*, 76:52–64, 2016.

[30] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7):2864–2875, 2013.

[31] Yu Zhang, Lijia Zhang, Xiangzhi Bai, and Li Zhang. Infrared and visual image fusion through infrared feature extraction and visual information preservation. *Infrared Physics & Technology*, 83:227–237, 2017.

[32] Xiangzhi Bai, Yu Zhang, Fugen Zhou, and Bindang Xue. Quadtree-based multi-focus image fusion using a weighted focus-measure. *Information Fusion*, 22:105–118, 2015.

[33] L. Zhang. In situ image segmentation using the convexity of illumination distribution of the light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1786–1799, 2008. Cited By :12.