



Generation of fused visible and thermal-infrared images for uncooperative spacecraft proximity navigation

Gaia Letizia Civardi¹, Michele Bechini¹, Matteo Quirino^{*,1}, Alessandro Colombo
Margherita Piccinin, Michèle Lavagna

Department of Aerospace Science and Technology, Politecnico di Milano, via La Masa, 34, Milan 20156, Italy

Received 15 November 2022; received in revised form 10 March 2023; accepted 15 March 2023

Abstract

State-of-the-art techniques for vision-based relative navigation rely on images acquired in the visible spectral band. Consequently, the accuracy and robustness of the navigation is strongly influenced by the illumination conditions. The exploitation of thermal-infrared images for navigation purposes is studied in the present paper, as a possible solution to improve navigation in close proximity with a target spacecraft. Thermal-infrared images depend on the thermal radiation emitted by the target, hence, they are independent from light conditions; however, they suffer from a poorer texture and a lower contrast with respect to visible ones. This paper proposes pixel-level image fusion to overcome the limitations of the two types of images. The two source images are merged into a more informative one, retaining the strengths of the distinguished sensing modalities. The contribution of this work is twofold: firstly, a realistic thermal infrared images rendering tool for artificial targets is implemented; secondly, different pixel-level visible-thermal infrared images fusion techniques are assessed through qualitative and quantitative performance metrics to ease and improve the subsequent image processing step. The work presents a comprehensive evaluation of the best fusion techniques for on-board implementation, paving the way to the development of a multispectral end-to-end navigation chain.

© 2023 COSPAR. Published by Elsevier B.V. All rights reserved.

Keywords: Sensor fusion; Advanced image processing; Vision based navigation; Thermal infrared imaging

1. Introduction

The possibility of operating in proximity of an uncooperative orbiting artificial object received great attention from researchers in the last few years, with a particular focus on on-board reconstruction of the chaser-target state vector through imaging. Such capability plays a crucial role for incoming missions as formation flying missions (FF)

with fractionated scientific payloads, on-orbit servicing demonstrators (OOS), and active debris removal, paving the way to regular in-orbit services, as reported by [Starek et al. \(2016\)](#). These are hot topics in our decade, which still need a significant technology development burst to become feasible. The close proximity manoeuvring requirement necessarily entails a guidance, navigation, and control chain solved autonomously onboard to ensure timeliness, reactivity, effectiveness, and robustness both in nominal and off-nominal operations. The first ring of that chain is indeed the relative state reconstruction and navigation. Artificial uncooperative targets are here considered, being the most challenging scenario and constraining towards a robust solution leaning on the chaser capabilities only, as

* Corresponding author.

E-mail addresses: gaialetizia.civardi@polimi.it (G.L. Civardi), michele.bechini@polimi.it (M. Bechini), matteo.quirino@polimi.it (M. Quirino), alessandro43.colombo@mail.polimi.it (A. Colombo), margherita.piccinin@polimi.it (M. Piccinin), michelle.lavagna@polimi.it (M. Lavagna).

¹ Equal contribution.

pointed out by [Opromolla et al. \(2017\)](#). In that operational context, imaging with passive sensors is the best option to collect meaningful measurements. Solutions related to visible (VIS) cameras have been widely studied, e.g. a comprehensive review is provided by [Sharma and D'Amico \(2016\)](#); for instance, VIS images-based navigation has been practically applied in the context of uncooperative/cooperative rendezvous, as in [Leinz et al. \(2008\)](#), [Castellini et al. \(2015\)](#). Instead, one of the first applications of a visual-based relative navigation system for uncooperative targets was during the Hubble Space Telescope Servicing Mission 4 as reported by [Naasz et al. \(2009\)](#).

However, visible imaging strongly depends on illumination conditions as it can be noticed in [Fig. 1](#) from [Bechini et al. \(2022a\)](#), where the Tango spacecraft from PRISMA mission (presented by [D'Amico et al. \(2013\)](#)) is only partially visible (inside the red circle) with low contrast with respect to the background, constraining the mission operation planning. Therefore, OOS missions are severely limited if illumination constraints for correct VIS 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. Illumination bottlenecks rise as significant for targets in LEO orbit, which experience long eclipses, as highlighted in [Fehse \(2014\)](#).

Degraded performances of VIS cameras in case of low-illumination conditions are well-known issues. At an early stage, this problem was solved by installing visible light emitters on spacecraft (e.g. during the docking in the 7th mission of the Engineering Test Satellite Program, as reported in [Oda and Inaba \(2000\)](#)). This solution is not suited for relative navigation scenarios due to the limited validity range. This research exploits thermal infrared

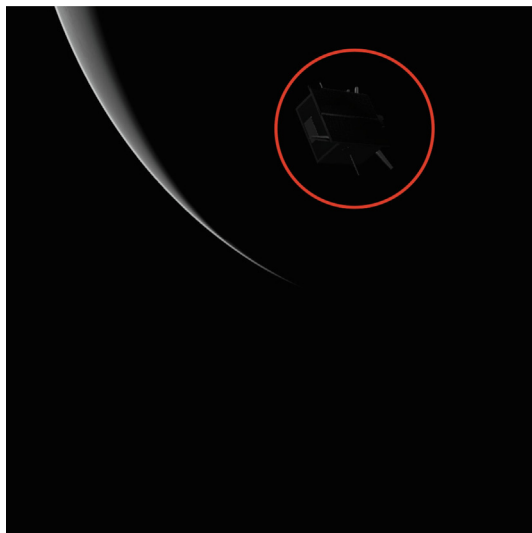


Fig. 1. Target low visibility scenario in synthetic noiseless validated VIS image.

(TIR) sensors leveraging their insensitivity to the illumination conditions, relying only on the emitted radiance, to overcome the limitations imposed by imaging sensors operating in the visible spectrum. The idea of using TIR images for highly challenging scenarios is not new and was first addressed in the context of active debris removal by [Yilmaz et al. \(2017b\)](#), while methods for features extraction from TIR images have been investigated by [Gansmann et al. \(2017\)](#) and [Yilmaz et al. \(2017a\)](#). For an extensive survey on monocular pose estimation architectures for noncooperative spacecraft also involving TIR images, the readers are referred to [Pasqualetto Cassinis et al. \(2019\)](#).

However, TIR sensors are characterized by a smaller array size if compared to visible ones, and TIR images present a lower resolution and poorer contrast with respect to VIS ones, which in turn negatively affects image processing algorithms, as highlighted in [Shi et al. \(2017\)](#). To retain the complementary advantages of the different spectral bands, [Deodshumukh et al. \(2003\)](#) proposed to apply tracking algorithms to both VIS and IR images and then fuse the result in a Kalman filter to achieve robustness in person motion tracking. The same concept was then applied to spacecraft tracking by [Palmerini \(2014\)](#) and tested on-ground for vision-based navigation for uncooperative targets rendezvous by [Schnitzer et al. \(2017\)](#). Subsequently, the idea of fusing the information extracted from VIS and TIR images was adopted for relative navigation and mapping of asteroids and unknown spacecraft by [Piccinin et al. \(2021\)](#) and [Civardi et al. \(2021\)](#), respectively. A different approach is proposed here since this work employs pixel-level image fusion to obtain a more informative image subsequently fed to the Image Processing (IP) step. The fusion methods are extensively assessed through both qualitative and quantitative criteria to identify the most efficient techniques to be adopted within a multispectral navigation chain. The pixel-level fusion was explored by [Jiang et al. \(2022\)](#) in the context of debris surveillance and identification. [Jiang et al. \(2022\)](#) proposes a new convolutional sparse representation-based image fusion algorithm and provides a comparison against deep-learning-based and hybrid multi-scale-based fusion methods in several scenarios. Pixel-level VIS-TIR image fusion, to the best of the authors' knowledge, has never been presented before within the context of spacecraft relative navigation techniques. The image fusion here is proposed to be performed before the IP. As a consequence, differently from the scenario in [Jiang et al. \(2022\)](#), the timeliness of the fusion step needs to be ensured to provide measurements to the navigation filter at a high rate. Some analyses regarding the execution time of a visual-odometry pipeline on space-qualified processors are reported in the work of [Lentaris et al. \(2018\)](#), in which the authors highlight the high computational burden of vision-based navigation algorithms. Since the image fusion process is an additional IP step at the beginning of the navigation chain, its computational cost shall be as low as possible. Furthermore, the research here presented delves into synthetic TIR image rendering

since no open-source tools are available for synthetic thermal infrared rendering purposes tailored to artificial targets.

The two major contributions of the paper can be then summarized as follows:

- Development of a flexible and accurate physically-based thermal infrared images rendering chain to test and develop TIR-based IP algorithms capable of rendering airless celestial bodies and artificial targets;
- Assessment and comparison of pixel-level VIS-TIR image fusion techniques to ease IP within the context of uncooperative spacecraft relative navigation. This includes also the assessment of pre-processing techniques needed to deal with noisy images at a different resolution, representing a real case scenario.

The paper is organized as follows: Section 2 presents an overview of already available image rendering tools and multispectral image fusion techniques; Section 3 details the new rendering chain implemented and the image noise models adopted; the selected image fusion algorithms are described in Section 4 together with the adopted performance metrics. Section 5 presents the results of the image fusion experiments, with applications to realistic noisy images at different resolutions. The metrics tailored to describe the efficiency of the fusion algorithms in case of noisy inputs are also reported. Conclusions resuming the main outcomes and the best-suited techniques for image preprocessing and fusion identified in this paper are reported in Section 6.

2. State of the art rendering tools and image fusion algorithms

2.1. Spacecraft images 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 which starts from the observer camera and ends on generic virtual objects that, together with the light rays simulated from the light sources to the virtual object, allow the computation of the color intensity of the related pixels. As pointed out in [Shirley and Morley \(2008\)](#), by simulating the physics of the light, ray tracing techniques can generate artificial images with a high degree of accuracy. Several available tools provide a user-friendly environment to develop 3D scenes to generate images via ray tracing, both commercial, as PANGU (Planet and Asteroid Natural Scene Generation Utility) presented by [Parkes et al. \(2004\)](#), and opensource, like POV-Ray (Persistence of Vision Raytracer) by [Plachetka \(1998\)](#) and Blender, developed by [Blender Online Community \(2018\)](#). Blender is the software used for this paper due to its high flexibility and high-quality outputs. Please notice that a pinhole camera model has been employed for the presented

work. Concerning spaceborne synthetic yet validated and realistic image datasets, they are constituted of only VIS images and the only currently publicly available (both qualitatively and quantitatively) are the SPEED (Spacecraft Pose Estimation Dataset) from [Kisantal et al. \(2019\)](#), its improved version, SPEED+, from [Park et al. \(2021\)](#) and the multi-purpose datasets in [Bechini et al. \(2022a,b,c\)](#) presented in the work from [Bechini et al. \(2023\)](#). Other spacecraft image datasets were released after SPEED and listed by [Musallam et al. \(2021\)](#), but among them, only the dataset published by [Proença and Gao \(2020\)](#) has been qualitatively validated against actual spacecraft images. However, an algorithm tailored to the VIS-TIR image fusion requirements is still needed, hence it has been decided to develop both the VIS image rendering tool together with the TIR image rendering one. 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. One of the simplest methods is to convert visible images into thermal images by simply scaling the image data into radiance data, similarly to the study presented by [Cosine Research for ESA \(2022\)](#). Such an approach has two main problems. The first one is that the actual temperature field is not computed, meaning that the object temperature features are neglected. Indeed, as reported in the executive summary from [Cosine Research for ESA \(2022\)](#), a dedicated texture must be applied to simulate the conduction effects and to eliminate all the visible features, such as shadows, that are not present in the thermal image. Secondly, to simulate the thermal inertia, multiple light sources are introduced to achieve a fake transient effect. Hence, due to the difficulties in creating representative TIR images from thermal mock-up highlighted by [Schnitzer et al. \(2017\)](#) and a valid thermal model for the target, there is a lack of publicly available TIR image datasets. It is acknowledged that [Avilés et al. \(2016\)](#) generated a synthetic dataset of both VIS and TIR images using a commercial software named ASTOS Camera Simulator without disclosing the images. On the contrary, [Jiang et al. \(2022\)](#) made the VIS and TIR dataset generated via the commercial software named Vega Prime software publicly available.

Among commercial software, PANGU v6 includes a tool for TIR image rendering through a lookup table-based thermal image rendering model for natural scenarios, including physics-based features such as thermal lag and local variations in emissivity and absorptivity. Instead, for artificial bodies the thermal rendering is equation-based with a model that accounts for thermal energy from solar, planetary reflectance, planetary emission, background radiation, and internal heat sources as reported by [Martin et al. \(2021\)](#) and [STAR-Dundee \(2022\)](#). The methodology adopted in PANGU consists of using thermal contributions to calculate temperatures at the pixel level and can generate thermal radiance images, using

Planck's law to convert temperatures to radiance image pixel intensities. A different approach for rendering thermal images of natural bodies is adopted in the work of [Piccinin et al. \(2021\)](#), in which a simplified thermal model of asteroids is used to perform a thermal simulation of the body. The output of the thermal simulation is then processed to produce a radiance image and converted into a thermal image employing the model of an uncooled microbolometer. Concerning natural bodies, a comparison between PANGU and the work in [Piccinin et al. \(2021\)](#) can be made. The former tool is capable to generate almost real-time thermal images, which is a powerful capability to perform closed-loop tests of vision-based GNC algorithms; moreover, by employing mission data fitting, very accurate images can be obtained for similar conditions. On the other hand, the latter approach has a broader generality, as it is not limited to already explored natural bodies and the proposed rendering chain is end-to-end, starting from the thermal simulation up to the radiance image and sensor's model.

Building on the approach in [Piccinin et al. \(2021\)](#), the new image generation pipeline introduced in this paper focuses on artificial spacecraft and improves the accuracy of the thermal model of the object using the high detail finite volume thermal model presented in [Quirino et al. \(2021\)](#). Furthermore, the new method introduces the detailed view factors between the camera and the object in order to compute the actual radiative flux received by the camera sensor, thus providing realistic features as output. Thus, in the context of spacecraft thermal images, the differences between our tool and PANGU are again the computational time, which is almost real-time in PANGU, and the thermal analysis, which in the here proposed approach relies on accurate finite volume thermal models. The here employed rendering engine is Blender.

2.2. Image fusion

Visible images are typically characterized by high texture detail, while they suffer from overexposure or challenging illumination conditions. On the other hand, thermal infrared images are insensitive to such disturbances but typically have poor texture and low resolution. 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. Fusion algorithms have been used in a wide range of application fields: [Singh et al. \(2008\)](#) deal with object recognition, [Kumar et al. \(2006\)](#) use fused images for surveillance purposes, [Simone et al. \(2002\)](#) apply image fusion to remote sensing. However, to the best of the authors' knowledge, image fusion has 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 by [Ma et al. \(2019\)](#). The main categories are multi-scale transform, sparse representation, neural network, subspace

and saliency-based methods, hybrid models, and other methods. 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. The most promising methodologies for an online space application, are here presented.

2.2.1. Multi-scale transform-based methods

As outlined in [Dogra et al. \(2017\)](#), these methods typically comprise three steps: 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. These methods can be extremely versatile according to the selected decomposition technique and fusion rule. On the other hand, the computational time associated to multi-scale methods rapidly increases with the number of decomposition levels.

2.2.2. Subspace-based methods

Subspace-based methods aim to project a high-dimensional input image into low-dimensional spaces or subspaces. Images are often composed of redundant information, and thus a low-dimensional subspace can be used to capture the intrinsic structure of an image, with the benefit of a high computational efficiency. 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 a low computational efficiency, and thus it has been discarded. For the interested reader, a comprehensive classification of subspace-based methods is presented in [Mitchell](#).

2.2.3. Saliency-based methods

According to [Toet \(2011\)](#), visual saliency is defined as the subjective perceptual quality which makes some pixels stand out from their neighbors and thus attract our attention. According to the working principles of human visual perception, saliency-based fusion methods can preserve the integrity of the salient object region and thus they can effectively extract any bright regions from the thermal-infrared image. 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 for target detection and recognition purposes. Methods based on image saliency are able to retain a high level of detail in the fused image, however, all these techniques tend to be sensitive to the amount of noise present in the source images.

2.2.4. Hybrid methods

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, [Liu et al. \(2015\)](#) proposed a fusion framework in which multi-scale transform is used to decompose the source images, while sparse representation is exploited to obtain the fusion coefficients.

3. Thermal infrared rendering

The goal of the paper concerning the generation of thermal images is to introduce a novel approach that can provide a high level of geometrical accuracy and at the same time high accuracy in the generated temperature field. As stated in Section 2, when converting a visible image to obtain an infrared one, the geometrical accuracy is preserved but the temperature field accuracy is degraded.

Thus the newly presented approach uses the high-detail finite volume thermal model by [Quirino et al. \(2021\)](#) to provide the details needed concerning both the temperature field and the geometry. The approach is foreseen to overcome the limits of the current methods available in the literature for thermal image generation, taking into account the relative position of the thermal sensor receiving the radiation. That is a difficult task, which is not present in the current literature as the view factors of the scene must be computed and used to compute the actual radiation received by the thermal sensor.

The high-detail finite volume thermal model of the object is reported in [Fig. 2](#). In order to replicate the thermal sensor output, the temperature field must 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:

$$Q_{f-p} = A_f F_{f-p} \varepsilon \sigma T_f^4 \quad (1)$$

where A_f is the area of a generic mesh facet, F_{f-p} is the view factor between a facet of the object mesh and a generic pixel of the camera, T_f is the temperature at the centre of the mesh facet, ε is the infrared emissivity of the object and σ is the Stephan–Boltzmann constant. Considering the radiant flux over the area of the respective face of the mesh, the expression can be rewritten as:

$$q_{f-p} = \frac{Q_{f-p}}{A_f} = F_{f-p} \varepsilon \sigma T_f^4 \quad (2)$$

The view factor F_{f-p} can be computed through the numerical evaluation of the following double integral:

$$F_{f-p} = \frac{1}{A_f} \int_{A_p} \int_{A_f} \frac{(\hat{\mathbf{n}}_f \cdot \mathbf{s}_{fp})(\hat{\mathbf{n}}_p \cdot \mathbf{s}_{pf})}{\pi S^2} dA_f dA_p \quad (3)$$

In which $\hat{\mathbf{n}}$ is the surface normal vector, $\mathbf{s}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ 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. It is assumed the camera is far enough from the object so there is no difference in the view factor between one face of the mesh and the different pixels of the camera. Thus it is only necessary to iterate over the number of the object mesh faces.

All the mesh face normal unit vectors and areas are computed, thus with the thermal camera position and orientation the view factors are calculated for each face with the discrete form of Eq. 3:

$$F_{f-c} \approx \frac{(\hat{\mathbf{n}}_c \cdot \mathbf{s}_{cf})(\hat{\mathbf{n}}_f \cdot \mathbf{s}_{fc})}{\pi S^2} A_c \quad (4)$$

where A_c is the area of the thermal camera and the rest of the terms as described for Eq. 3. The p subscript is substituted by c as the thermal camera is considered one whole surface (i.e. all pixels have the same view factor for the single mesh face in exam). The resulting view factor field computed for a camera position facing the back panel of Tango (i.e. the one opposite to the solar panels) is reported in [Fig. 3](#).

Once the view factors are computed it is possible to take the temperature of each face (from the temperature field) along with the corresponding view factor and compute the radiance emitted by each face towards the thermal camera with Eq. 2. The radiance field is then rendered in Blender and the results are reported in [Fig. 4](#) for the same camera position used to retrieve the view factors reported in [Fig. 3](#). The output is qualitatively correct as it is possible to see on the antennas (i.e. cylindrical shapes at the corners) that the radiant flux decreases on the edges as the view factor values drop for the considered camera pose.

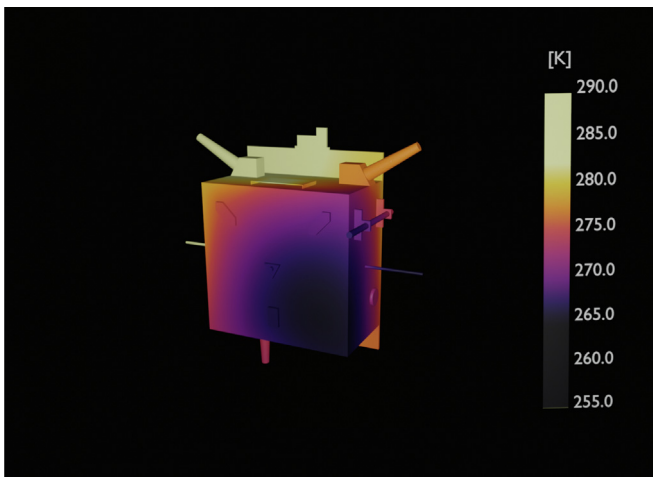


Fig. 2. Example of temperature field.

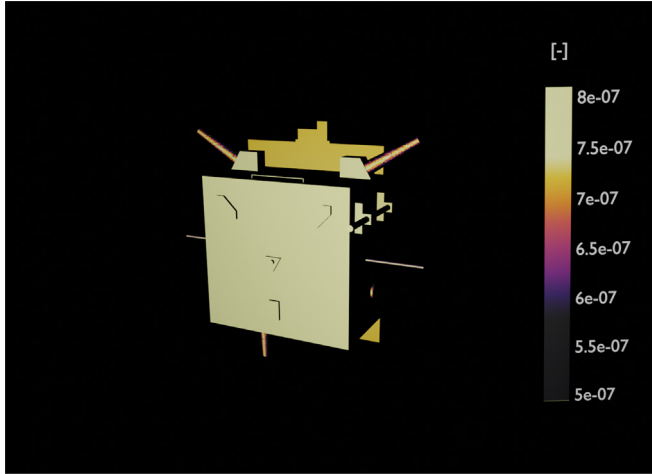


Fig. 3. View factors for camera position normal to back panel.

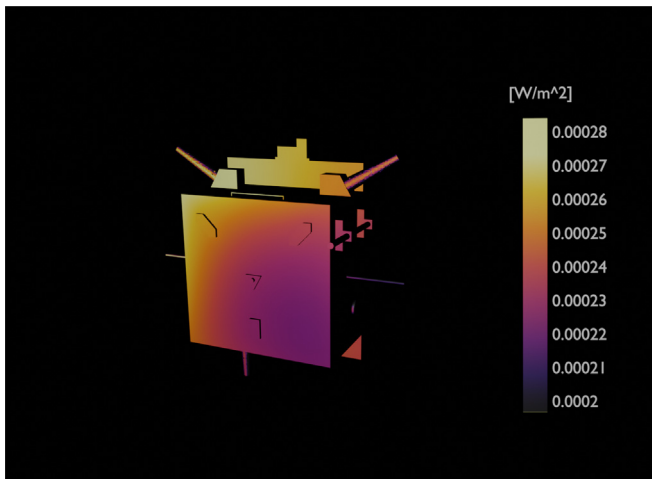


Fig. 4. Radiance field for camera normal to back panel.

It must be pointed out that Fig. 4 is not the image that is provided to the fusion algorithms but it is a rendered image with proper lighting. Indeed, for the generation of TIR images, the radiance field is mapped over the object as a texture based on the model of a Lambertian emitter, getting rid of the need for light sources, that 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. An accurate calibration of the thermal sensor gain and offsets based on the expected temperature field will be included in the image generation pipeline to accurately set the conversion to DN, which is for now done by tuning the flux scale. Nonetheless, the output of the workflow is a valid input for the image fusion techniques that are assessed in Section 4. An example of the final result used for the fusion algorithms is reported in Fig. 7.

Since one of the goals of this paper is to prove the feasibility of the approach, the object is assumed to be gray and diffuse, thus the emissivity is constant for the whole object. However, thanks to the full accessibility to the Finite Volume Method (FVM) code, it is possible to introduce different values of emissivity for the different regions of the object. Introducing different materials is not expected to negatively affect the performance of the fusion algorithms since this should allow having more details in the final image. It is expected that regions with different materials have a significant difference in temperature ranges as the contact resistance in vacuum is very high. In the image, such phenomena should create more contrast between the regions with different materials. Nevertheless, further assessments through dedicated analyses need to be performed.

The workflow chain for the TIR image rendering is depicted in Fig. 5. In order to follow such chain it is mandatory to have full access to the mesh topology and temperature field. For this reason, the FVM open-source code OpenFOAM has been used as it guarantees the accessibility to the case and the flexibility needed.

It is important to highlight that the rendering chain is very flexible and can be adopted for any type of spacecraft, space debris, and celestial object. The only limitation to the presented rendering chain is the computational power available to face the problem, especially for the celestial objects where the number of cells in the simulation is very high in order to grasp the texture of the terrain and have realistic images. For artificial targets (e.g. spacecraft and space debris), the computational power requested for the thermal simulation is reduced. On the other hand, they are usually made of multiple parts and materials, thus it requires a bigger effort in the simulation setup with respect to the celestial objects which are made of one material with averaged thermophysical properties.

The adopted TIR and VIS camera characteristics used to render the images in Figs. 6,7, are reported in Table 1. The two cameras have the same Field of View (FoV), due to the fusion requirement of image alignment. The other parameters have been obtained by merging the parameters given in Kisantal et al. (2019) and cropped to the array size considered in Bechini et al. (2023). The TIR array size has been assumed to be half of the VIS one.

3.1. Noise modeling

Images obtained via VIS monocular cameras are mostly affected by electronic noise (similarly to all electronic devices) and blurring due to the fixed depth of field of real cameras, as reported in Boie and Cox (1992). To increase the accuracy in reproducing real spaceborne images, the noise-free VIS images obtained from the previously outlined pipeline are processed by adding a white Gaussian noise with $\sigma^2 = 0.0022$ and blurred with a Gaussian blur characterized by $\sigma = 1$ and zero mean. These values

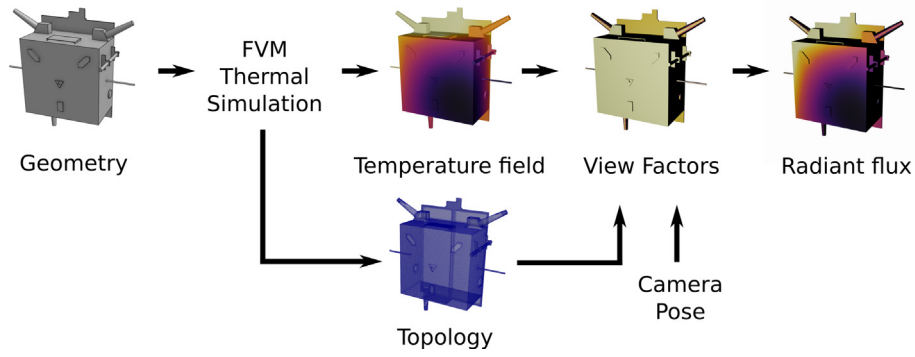


Fig. 5. TIR rendering workflow.

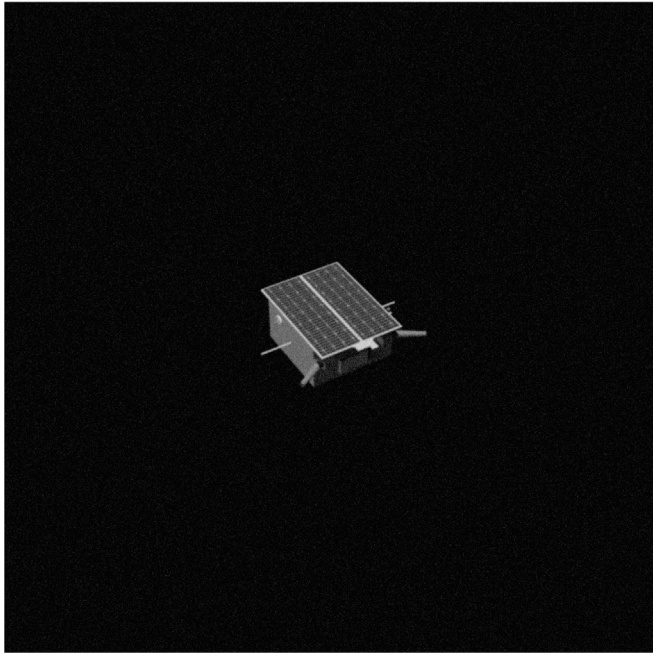


Fig. 6. Example of rendered VIS image.

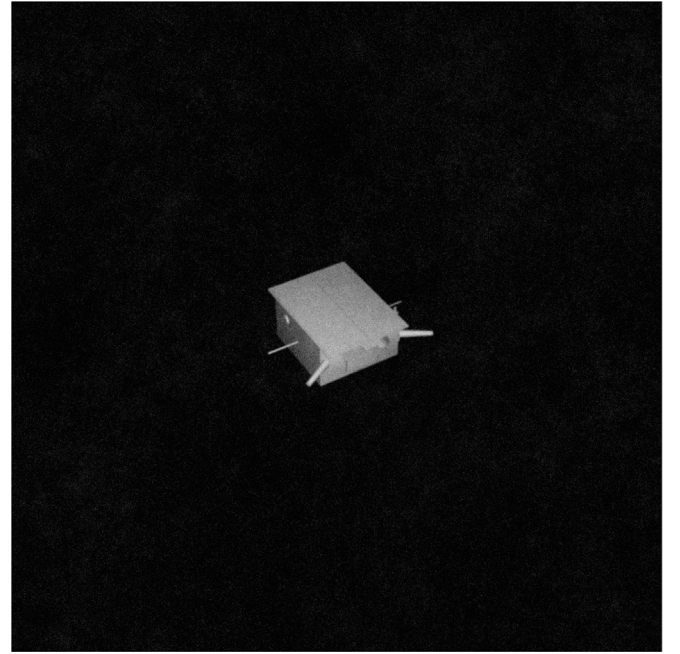


Fig. 7. Example of rendered TIR image.

Table 1

Cameras characteristics.

	VIS	TIR
Array Size	1024 × 1024 px	512 × 512 px
FoV	35.45 × 35.45	35.45° × 35.45°
Focal Length	17.6 mm	17.6 mm

have been obtained from the validation of the SPEED dataset images against real Tango images from Prisma mission as described in [Kisantal et al. \(2020\)](#), and adopted also for the validation in [Bechini et al. \(2023\)](#). With regards to thermal imaging sensors, [Gao et al. \(2011\)](#) demonstrated that microbolometers are mostly affected by two sources of noise: the thermal noise and the $1/f$ noise. The former is a characteristic of all electronic devices and it is modeled as an additive white Gaussian noise, assuming the same characteristics adopted for VIS images. The $1/f$ noise, which is also referred to as flicker noise or pink noise, is instead dominant at low frequencies, as demonstrated by [Brageot et al. \(2014\)](#). An additive pink noise can be numerically obtained by applying a suitably shaped low-pass filter to a white Gaussian noise. A two-dimensional Fourier transform is used to decompose white noise into the fre-

quency domain. The amplitude (A) of each frequency is then scaled such that the higher the frequency, the lower the amplitude using the following relationship:

$$A' = A \frac{1}{(f_x^2 + f_y^2)^{\alpha/2}} \quad (5)$$

where f_x and f_y are the spatial frequencies and α is an exponent which determines the spectral slope ($\alpha = 1$ for flicker noise). The inverse Fourier transform is then applied to convert the filtered result back to the spatial domain.

The variance of the White Noise to be filtered is here assumed to be $\sigma^2 = 0.0022$. Similar to VIS images, also the TIR images are blurred with a Gaussian blurring characterized by $\sigma = 1$ and zero mean. Notice that, as already done in [Bechini et al. \(2023\)](#), it has been decided to add the noise after the image rendering and to do not include the noise generation directly in the rendering process in order to maintain the high flexibility of the rendering pipeline. Hence, the noises are imposed by applying first the Gaussian Blur to the noiseless images. The additive Gaussian and the flicker noise are obtained by computing pixel-intensity noise maps of the same size as the original image. The additive white Gaussian noise map is obtained by randomly sampling pixel intensity values by following Gaussian distribution with the prescribed mean and standard deviation. Then, the computed noise map is added to the blurred image. The same holds for the flicker noise, but, as described above, the noise map is filtered (see Eq. 5) before being added to the blurred image. After that, the noised images are clipped to the correct range of values ($[0.0, 1.0]$ for normalized images, $[0, 255]$ otherwise). The most dominant noises in both VIS and TIR images can be characterized for the sensor actually adopted during a validation campaign and then added to the noiseless synthetic images generated by the proposed pipeline as discussed above, tailoring those images to the mission scenario of interest.

4. Image fusion techniques

This section is devoted to the description of the implemented image fusion techniques, which can be classified according to the criteria presented in Section 2. All the presented methods but the last one share the same assumption, i.e. the source images should have the same resolution. All the presented fusion methods require that the source images are perfectly aligned.

4.1. Fusion methods

4.1.1. Fusion through Fast Global Smoothing Decomposition and Target Enhanced Parallel Gaussian Fuzzy Logic (TEPGFL)

A representative multi-scale-based fusion method is here considered. The method employs multi-scale image decomposition, as described in Section 2.2.1, and saliency detection to determine the fusion weights. In particular, the work presented in [Duan et al. \(2022\)](#) is taken as a reference. An edge-preserving smoothing method, namely the fast global smoother (FGS) of [Min et al. \(2014\)](#), is employed as the multi-scale decomposition tool in this technique. Base layers are often fused according to a simple average rule. However, this approach suffers from a negative effect, that is, the loss of contrast in the fused image. The selected algorithm performs base layers fusion through a Gaussian fuzzy logic-based weighting rule that is based on the concept of the Gaussian membership function.

To perform detail layer fusion, the visual saliency map is obtained using the Scharr gradient algorithm, due to its robustness to noise, and a choose-max coefficient rule is employed to select the fusion weights.

4.1.2. Anisotropic diffusion-based fusion (ADF)

The ADF algorithm can be regarded as a PCA-based technique, which falls into the broad class of subspace-based methods outlined in Section 2.2.2. This implementation is largely based on the one described in [Bavirisetti and Dhuli \(2016a\)](#). 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. The method is expected to have a high computational efficiency due to the fast image decomposition process.

4.1.3. Image fusion using two-scale decomposition and saliency detection (TSFISD)

This algorithm is a hybrid method that builds on both multi-scale decomposition and the concept of image saliency, which was described in Section 2.2.3. Our implementation is inspired by the one presented in [Bavirisetti and Dhuli \(2016b\)](#), 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, our 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.

4.1.4. Image Fusion with Guided Filtering (GFF & MGFF)

The following algorithms are another example of hybrid multi-scale-based fusion methods, as pointed out in Section 2.2.4. The GFF method introduced by [Li et al. \(2013\)](#), proposes an approach based on a two-scale decomposition of the images into base and detail layers with an average filter. Saliency maps and corresponding initial weight maps are calculated using Laplacian and Gaussian operators. A guided-filtering technique is employed to refine the initial weight maps, which are finally combined with the respective layers to yield the fused image.

Two of the main problems of the GFF, namely the loss of image features due to the Laplacian operator and not exploiting the advantages of a multi-scale decomposition, are addressed in [Bavirisetti et al. \(2019\)](#), which introduces

the MGFF algorithm. 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, reducing the 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.

4.1.5. Infrared and visual image fusion through Infrared Feature Extraction and Visual Information Preservation (IFEVIP)

The IFEVIP algorithm does not belong to any of the aforementioned categories since it is not reliant on classic fusion methods. Its implementation, mostly based on the work of [Zhang et al. \(2017\)](#), exploits quadtree decomposition and Bézier interpolation 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 to the visual image to achieve the final fusion image.

4.1.6. Different Resolution Image Fusion (DRIF)

The last algorithm here considered is the one developed by [Du et al. \(2018\)](#) and it allows one to directly fuse images that have different resolutions. This method formulates the fusion problem as a total variation minimization problem. The cost function is composed of a data fidelity term that constrains the pixel intensity similarity of the downsampled fused image with respect to the source TIR image, while a regularization term forces the gradient similarity of the fused image with respect to the VIS one. To relieve the computational cost, the fast iterative shrinkage-thresholding algorithm (FISTA) framework is applied. It is worth underlying that the resulting fused image has the same resolution as the source VIS.

4.2. Performance 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. In our case, reference-free criteria are adopted, since it is not possible to compare the fused image with a reference ground truth image. The performance metrics used to evaluate the fusion algorithms are here briefly described.

4.2.1. Mutual Information (MI)

The mutual information (MI) index measures the amount of information that is transferred from the source images to the fused output. MI measures the dependence of two random variables. The fusion MI metric is defined as a simple summation of the MI between the source VIS and the fused image and the MI between the source TIR and the fused image, respectively:

$$MI = MI_{VIS,F} + MI_{TIR,F} \quad (6)$$

being $MI_{VIS,F}$, $MI_{TIR,F}$ the amount of transferred information from the source VIS and TIR images to the fused image, respectively. Please notice that no weights are added to the linear summation due to the fact that, in principle, it is impossible to know apriori the relative contribution of each source image to the final fused image. The MI between two random variables is computed through the Kullback–Leibler measure. Specifically, the MI of between the fused image F and source image X is defined as follows:

$$MI_{X,F} = \sum_{x,f} p_{X,F}(x,f) \log \frac{p_{X,F}(x,f)}{p_X(x)p_F(f)} \quad (7)$$

where $p_X(x)$ and $p_F(f)$ are the marginal histograms of source image X and fused image F, respectively, $p_{X,F}(x,f)$ is the joint histogram of source image X and fused image F. Further details regarding the theoretical background of this quality index are available in the work of [Qu et al. \(2002\)](#) for the interested reader.

4.2.2. Feature Mutual Information (FMI)

The performance metric presented in [Haghighat et al. \(2011\)](#) is related to image processing tasks. Images are often represented by their features, such as edges, details, and contrast, therefore measuring the amount of feature information that is transferred from source images to the fused image is directly linked to our visual navigation purposes. FMI is based on MI and feature information and it can be defined as follows:

$$FMI = MI_{\hat{VIS},\hat{F}} + MI_{\hat{TIR},\hat{F}} \quad (8)$$

where \hat{F} , \hat{VIS} , \hat{TIR} denote the feature maps of fused, visible, and infrared, images, respectively. A large FMI metric generally indicates that a considerable amount of feature information has been transferred from the source images to the fused one, and thus, the output of the fusion is suitable for image processing purposes. In particular, we consider edge information for this metric, due to the importance of edge strength in the subsequent image processing phase.

4.2.3. Structural Similarity Index (SSIM)

As pointed out in [Wang et al. \(2004\)](#), SSIM models image loss and distortion. It consists of three contributions, namely loss of correlation, luminance, and contrast distortion. SSIM is computed as the product of three parts:

$$SSIM_{X,F} = \frac{2\mu_x\mu_f + C_1}{\mu_x^2} \cdot \frac{2\sigma_x\sigma_f + C_2}{\sigma_x^2\sigma_f^2 + C_2} \cdot \frac{\sigma_{xf} + C_3}{\sigma_x\sigma_f + C_3} \quad (9)$$

In which $SSIM_{X,F}$ denotes the structural similarity between the source image X and the fused image F ; x and f denote the image patches of source and fused images in a sliding window; μ_x, μ_f indicate the mean intensity values of source and fused image, respectively; σ_x and σ_f are the image intensity standard deviation (SD); σ_{xf} denotes the covariance of source and fused images and it is computed as:

$$\sigma_{xf} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(f_i - \mu_f) \quad (10)$$

The three constants C_1, C_2, C_3 are numerical stability parameters. Finally, the structural similarity between the source images and the resulting fused image is computed as follows:

$$SSIM = SSIM_{VIS,F} + SSIM_{TIR,F} \quad (11)$$

A high SSIM value indicates a good quality fused image.

4.2.4. Root Mean Squared Error (RMSE)

The root mean squared error (RMSE) metric for image fusion is evaluated as:

$$RMSE = \frac{RMSE_{VIS,F} + RMSE_{TIR,F}}{2} \quad (12)$$

and it denotes the dissimilarity between the source images and the fused image. The RMSE between the source image X and the fused image F is defined as:

$$RMSE_{X,F} = \sqrt{\frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M (X(i,j) - F(i,j))^2} \quad (13)$$

where $M \times N$ denotes the size of the images and (i,j) are the image coordinates. A low RMSE metric hints that the fused image has a small amount of error and distortion.

4.2.5. Average Gradient (AG)

The average gradient (AG) metric quantifies the gradient information of the fused image, which in turn represents its detail and texture. For an image f with the size $M \times N$, the AG metric can be defined as follows:

$$AG = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}} \quad (14)$$

where (i,j) are the image coordinates, ∇F_x and ∇F_y denote the horizontal and vertical intensity gradient values, respectively. The larger the AG index, the more gradient information the fused image contains.

5. Results

This section is devoted to the analysis of the image fusion results. First of all, the problem of image rescaling is addressed, selecting the best alternative for either upscal-

ing TIR images or downscaling VIS ones. Subsequently, the image fusion techniques are evaluated according to the metrics discussed in Section 4.2 and according to their computational cost.

5.1. Selection of image rescaling methods

As already noticed, TIR images present a lower resolution and poorer contrast with respect to VIS ones, as well as a higher level of noise; this affects the quality of the fused image. To select the best image downscaling or upscaling techniques, the results obtained from fusing directly the VIS and TIR images at the same resolution (ideal scenario) against the fused images obtained by preprocessing the TIR images (e.g. upscaling) or the VIS ones (e.g. downscaling) are compared in this section. To assess the effects of upscaling and downscaling on noisy images, the additional evaluation metrics considered are the (peak signal-to-noise) ratio ($PSNR_S$), the mean square error MSE_S , and structural similarity index ($SSIM_S$). The $PSNR_S$ is evaluated between fused images obtained from noiseless VIS-TIR couples both at the same resolution and the ones obtained by considering the difference in resolution previously discussed. The $PSNR_S$ is a standard metric widely used to evaluate the effects of IP chains on the noise in images. To provide also a metric that can represent the differences in the structure of the obtained images, which can be high due to the difference in the resolution, the $SSIM_S$ is also taken into account. Then, as pixel-to-pixel comparison, also the MSE_S is considered. As for the $PSNR_S$, also the $SSIM_S$ and the MSE_S are evaluated between fused images from noiseless high-resolution VIS-TIR couples and noised mixed resolutions VIS-TIR couples. Please notice that in the remainder of this paper the subscript “S” is used (for standard) to differentiate the metrics used to compare an image with respect to a single source image (1:1 comparison), from the metrics adopted to evaluate the fused images with respect to their VIS and TIR sources (1:2 comparison) presented in Section 4.2.

Before delving into the analysis of the fusion techniques for noised VIS-TIR images at a different resolution, an evaluation of the interpolation methods to downscale/up-scale images is here proposed. The methods considered in this analysis are: area interpolation from Wong and Herley (1997), bicubic interpolation as in Keys (1981), and Lanczos resampling from Duchon (1979). The metrics (MSE_S , $PSNR_S$, and $SSIM_S$) have been evaluated over three datasets (namely TIR, VIS “In Eclipse”, VIS “In Light”) of 20 images each. Please notice that for the analyses that are performed, the TIR images have been rendered with resolution of both 512×512 px and 1024×1024 px. The datasets mentioned have been created with the previously described TIR-VIS image generation tool. A single sun-target angle is employed in the dataset, therefore TIR images have been rendered assuming the same temperature field for Tango, only changing the chaser-target relative pose. Similarly, for each VIS images

Table 2
Upscaling of TIR images evaluation metrics.

Method	MSE _S	PSNR _S	SSIM _S
Area	0.0065	22.13	0.48
Bicubic	0.0063	22.29	0.49
Lanczos	0.0063	22.24	0.48

dataset only the relative pose is changed and, as a consequence, images correspond to different sun-target-camera phase angles. The VIS and TIR datasets are generated such that the camera locations with respect to the target are uni-

formly distributed over an ellipse with major and minor axis equal to 14.5 m and 8 m respectively. The ellipse is centered in geometrical center of the target, with the camera constrained to point toward the target. The ellipse is tilted by 45° every 5 frames to have a full coverage of the target. In such a way, by keeping constant the sun-target angle, the target illumination conditions result to be well distributed, as it can be noticed in Fig. 8.

Please notice that the source VIS and TIR images are perfectly aligned. Moreover, the relative distances has been kept constrained to low values in order to ease the detec-

Table 3
Downscaling of VIS images evaluation metrics for in light and in eclipse illumination conditions.

Method	In Eclipse		
	MSE _S	PSNR _S	SSIM _S
Area	0.00075	31.34	0.57
Bicubic	0.00094	30.34	0.54
Lanczos	0.00104	29.87	0.51
Method	In Light		
	MSE _S	PSNR _S	SSIM _S
Area	0.0022	27.19	0.56
Bicubic	0.0024	26.70	0.53
Lanczos	0.0025	26.48	0.50

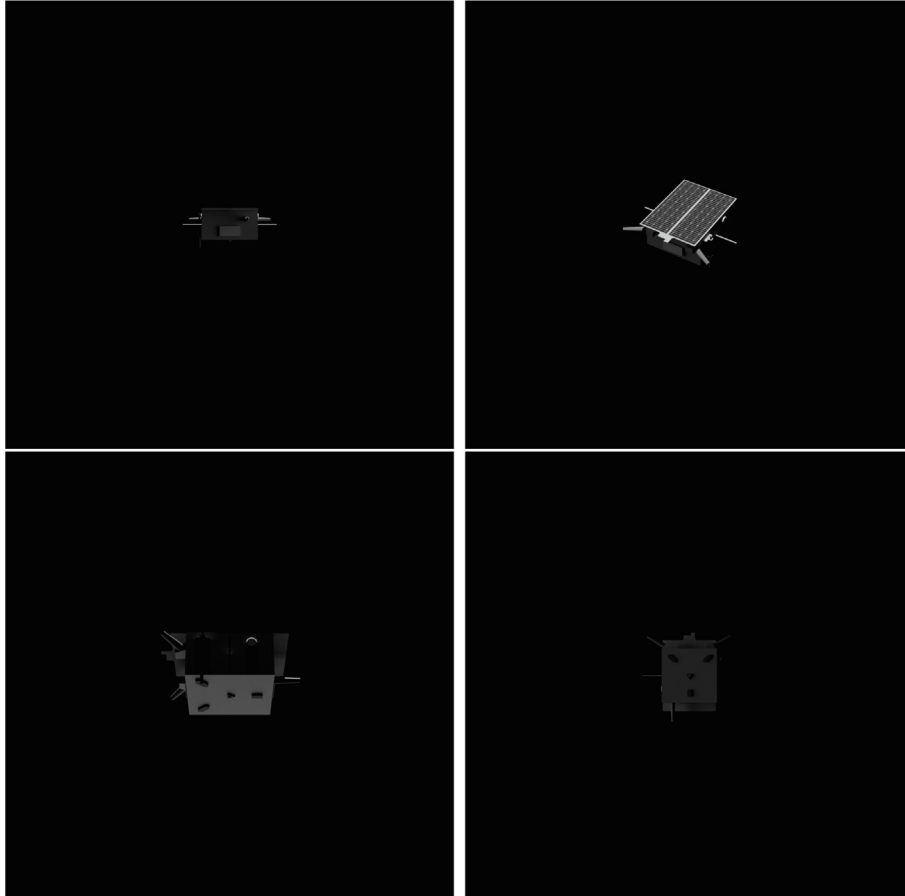


Fig. 8. Examples of illumination conditions of noiseless rendered VIS images included in the VIS datasets.

tion of possible artifacts due to the tested rescaling and fusion techniques by a qualitative comparison of the images.

The rescaling algorithms are applied considering an upscaling factor of 2 and a downscaling factor of 0.5, and then averaged. The results are reported in Table 2 for TIR upscaling, and in Table 3 for VIS downscaling applied on both VIS datasets. In Table 2 and Table 3 the optimum values for each of the three metrics used have been written in bold. From the results obtained, the bicubic interpolation is used to rescale VIS images in all the illumination conditions from 1024×1024 px to 512×512 px, while for rescaling the TIR images from 512×512 px to 1024×1024 px, the area interpolation method is selected. By comparing the results reported in Table 2 and Table 3, it can be noticed that all the interpolations methods used for rescaling have a noise-reduction effect, with the maximum $PSNR_S$ gain achieved in the downscaling process. Both these outcomes are expected since the upscaling/downscaling procedure is carried out by interpolating adjacent pixel values, that has been proven to reduce the noise. Moreover, the downscaling is performed also by averaging the pixel intensities, that further increase the PSNR.

The upscaling and downscaling methods, previously selected as best suited, have been applied to resize the images to the correct resolution, before performing the fusion test with all the methods discussed in Section 4 between noisy VIS and TIR images originally at a different resolution.

5.2. Fusion performances analysis

The last part of the pipeline to generate VIS-TIR fused images, starting from the clean images at a different resolution, involves then the application of the noise to the VIS and TIR source images separately, the application of VIS

Table 4

Averaged fusion metrics for upscaled TIR images, “In-Eclipse” dataset.

	AG	MI	FMI	SSIM	RMSE
ADF	14.66	0.98	0.848	1.96	12.31
GFF	15.35	1.52	0.878	1.94	11.76
IFEVIP	20.35	1.50	0.818	1.26	15.52
MGFF	16.13	1.01	0.837	1.94	12.14
TSIFSD	13.10	0.84	0.835	1.95	11.97
TEPGFL	14.97	0.94	0.784	1.12	18.63

downscaling or TIR upscaling by using the methods identified here as optimal, and finally the image fusion. A schematic outline of the different image fusion strategies is reported in Fig. 9

All the presented fusion techniques have been applied to the datasets previously discussed (see Section 5.1), with two the different alternatives for scaling: VIS downscaling and TIR upscaling. For each fused image, the evaluation of the performance metrics has been carried out. The performance metrics described in Section 4 have been considered together with the averaged MSE_S , $PSNR_S$, and $SSIM_S$ over the dataset to better understand the behavior of the proposed image fusion algorithms in presence of noises. To evaluate the metrics related to noise, fused images generated with noiseless VIS-TIR images at the same resolution by using the same method under analysis have been considered as references. Table 4 and Fig. 10 show the performance metrics average values and for each image pair respectively, obtained from VIS and upscaled TIR source images for the “In-Eclipse” dataset. From a qualitative inspection of the fusion results (an example for a single image is reported in Fig. 11), it is evident that the GFF algorithm assigns a low fusion weight to TIR images, thus producing a very dark image as output. The outliers in the MI metrics for GFF are related to images which are almost identical to the source VIS image; in fact, by analyzing the

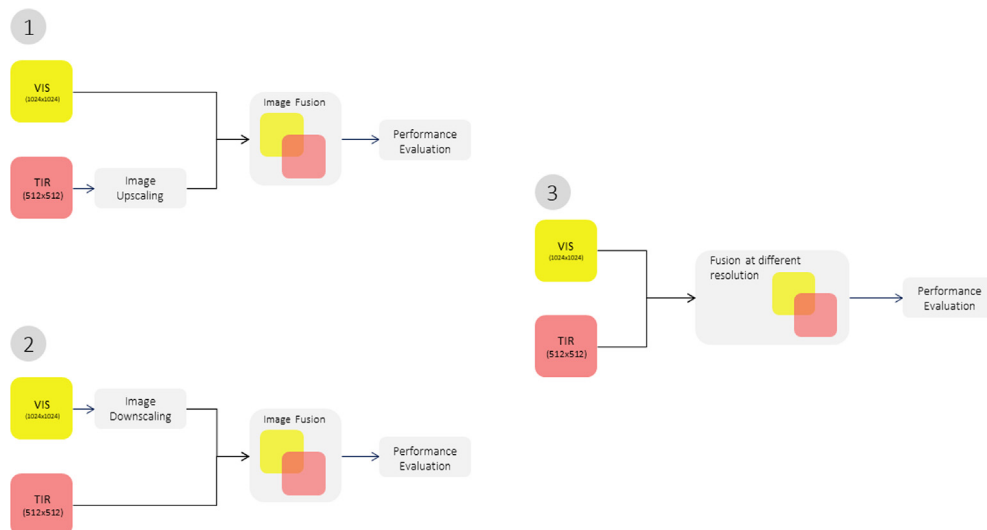


Fig. 9. Illustration of the image fusion schemes investigated.

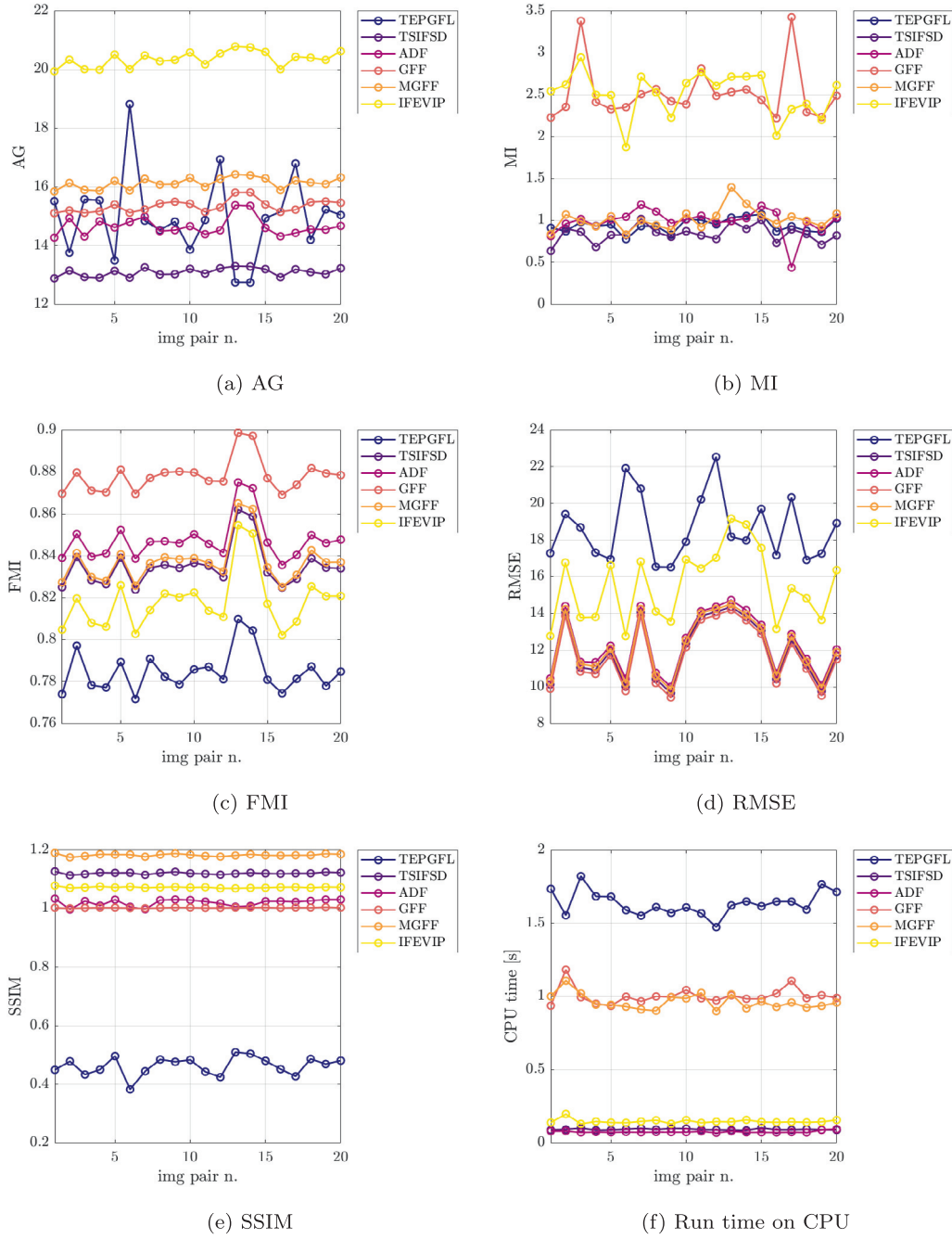


Fig. 10. Quantitative comparison of the metrics and run time on the “In-Eclipse” dataset, with TIR upscaling.

MI values in detail, it can be noticed that the $MI_{VIS,F}$ is very high, while the $MI_{TIR,F}$ is close to zero. This effect is instead not present when an intermediate score is achieved, as the total MI between the source VIS and TIR images and the fused image is lower in absolute value but made up of two similar contribution. On the other hand, MGFF only presents one outlier value, that is linked to a very high weight assigned to the TIR image during the fusion process. As for TEPGFL, it can be noticed that the method introduces halo effects and artefacts around the edges, and thus it should be discarded. This effect is also mirrored in the AG metrics, which presents outliers whenever halo and

artefacts are predominant. The other methods are instead capable of preserving the texture information of the solar panels from the VIS image while retaining the higher pixel intensity of the source TIR .

Table 5 and Fig. 12 show the performance metrics average values for each image pair respectively, obtained from TIR and downsampled VIS source images. Once again, it is possible to notice from Fig. 13 the halo and artefacts introduced by TEPGFL. In this case, the GFF algorithm is capable of preserving the pixel intensity of the TIR image, meaning that this method is sensitive to the upscaling process. Nevertheless, GFF cannot retain the texture informa-

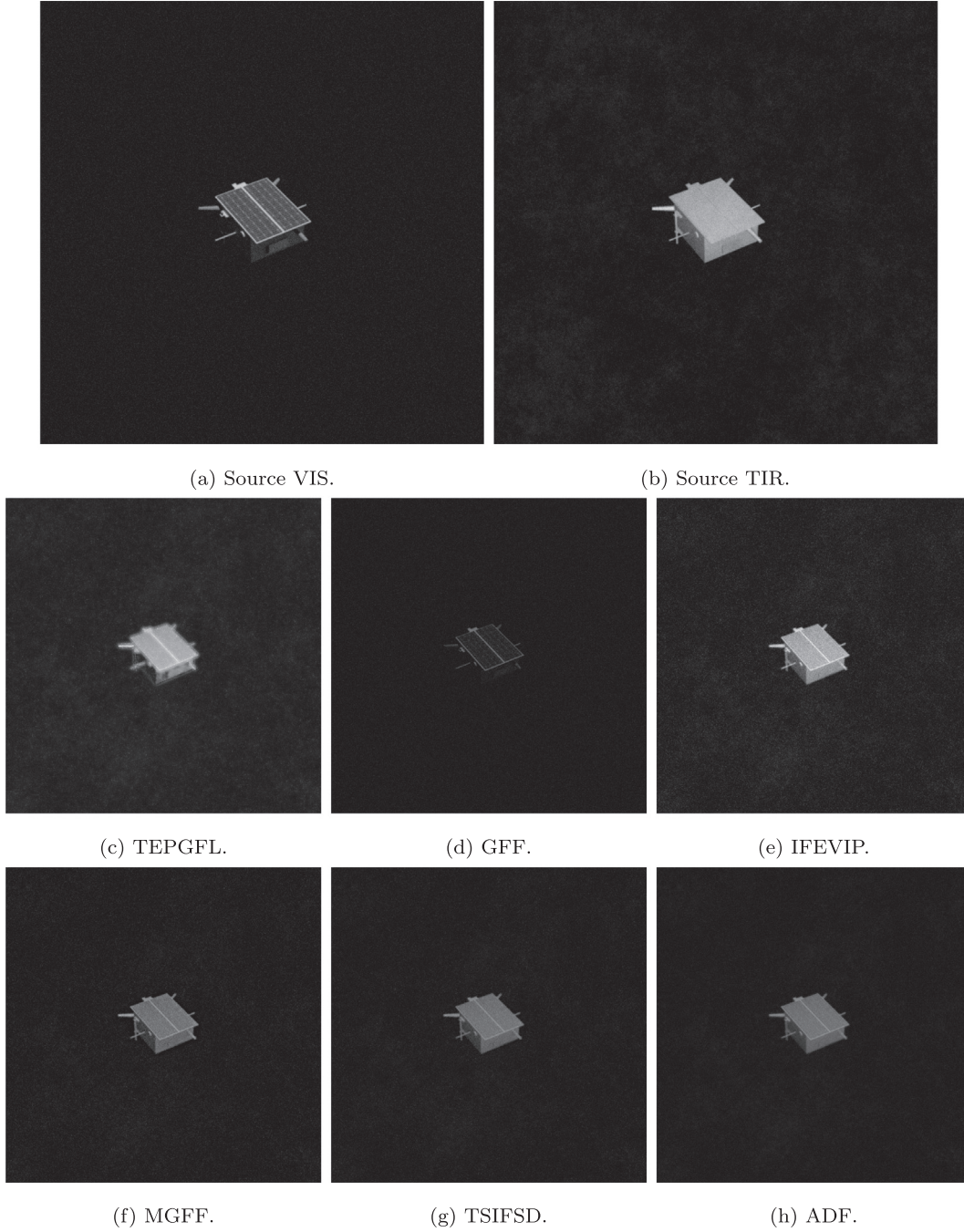


Fig. 11. Fusion results of representative algorithms on the “In-Eclipse” dataset, with TIR upscaling.

Table 5
Averaged fusion metrics for downscaled VIS images, “In-Eclipse” dataset.

	AG	MI	FMI	SSIM	RMSE
ADF	14.56	1.232	0.867	0.976	12.04
GFF	15.39	1.51	0.927	1.00	11.62
IFEVIP	16.67	1.91	0.906	1.03	14.72
MGFF	13.74	1.05	0.900	1.160	11.90
TSIFSD	10.89	0.842	0.893	1.076	11.7
TEPGFL	12.94	0.873	0.831	0.437	17.6

tion of the solar panels from the VIS images, thus reducing the overall detail level of the fused image. Qualitative performances of the other methods remain unchanged.

From the quantitative evaluation parameters, it appears that IFEVIP still retains the highest gradient information with respect to the other fusion methods. However, it tends to produce saturated images in the “In-Light” dataset as shown in the example reported in Fig. 14, and thus it is a misleading indicator in this case. By analyzing the performances of ADF and TSIFSD, it can be noticed that they

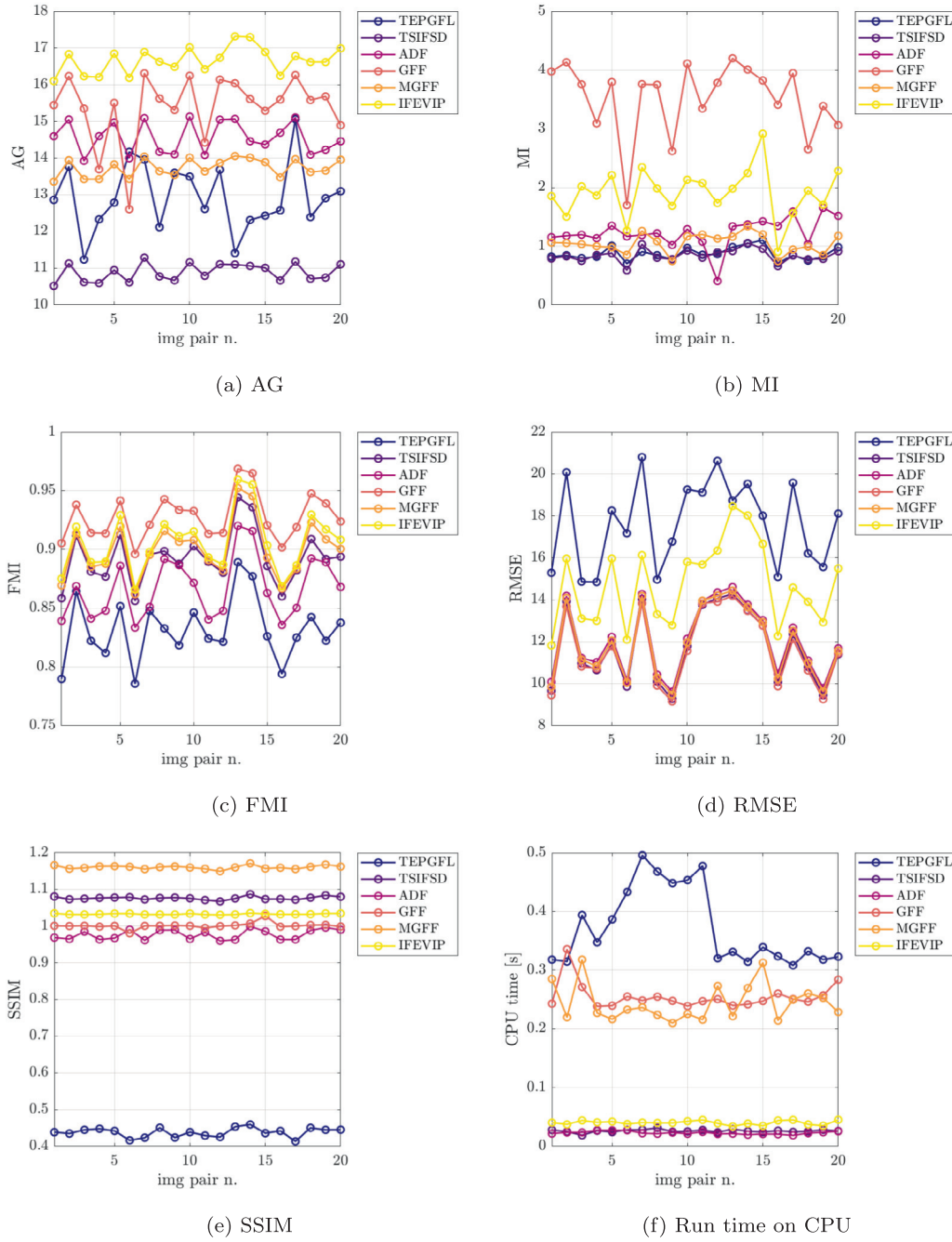


Fig. 12. Quantitative comparison of the metrics and run time on the “In-Eclipse” dataset, with VIS downscaling.

both yield a low RMSE for the fused image while having good performance scores in terms of FMI and MI. MGFF and GFF always have intermediate scores, however, as pointed out before, GFF is sensitive to image upscaling. Please notice that a qualitative analysis of the fusion results still plays a major role in the final decision, especially for a dataset of a limited size as the one we have at the moment. Indeed, if a larger dataset is employed, the presence of outliers would be mitigated, and thus it would be possible to rank the methods according to the total points scored for the metrics.

The metrics used to evaluate the effects of the fusion on the noise (averaged MSE_S , $PSNR_S$, and $SSIM_S$) are reported in Table 6, for all the cases tested.

Evaluating the effects of the fusion on a synthetic real case scenario with noisy images at different resolutions offers another perspective on the evaluation of the best-suited methods to fuse VIS-TIR spaceborne images. The results obtained by fusing noisy VIS-TIR images both at high resolution are also reported in Table 6 for comparison. The case of the same resolution is an ideal case but offers a good comparison since, as it can be noticed from

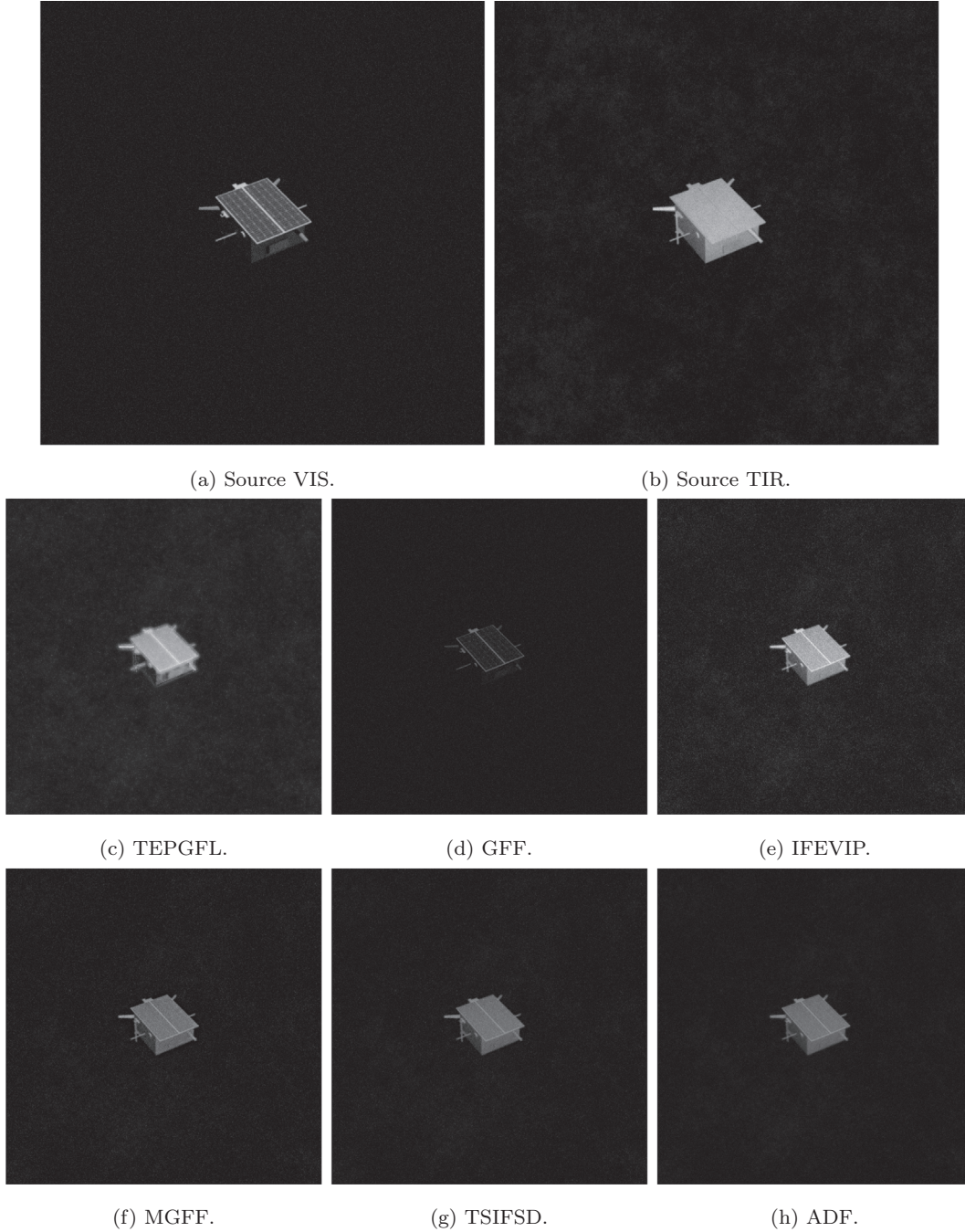


Fig. 13. Fusion results of representative algorithms on the “In-Eclipse” dataset, with VIS downscaling.

Table 6, it represents the optimal performance achievable as almost all the fusions applied with downscaling or upscaling one of the input images achieve lower performances on average.

With regards to the best way of dealing with inputs at different resolutions, from Table 6 it can be concluded that upscaling the TIR images offers the optimal performances for all the methods, achieving metric scores close to the optimal achieved for images at the same resolution. Hence it can be concluded that by upscaling the TIR images, the

fused image is less sensitive to input noise, as the output is closer to the ideal noiseless case (high performances in noise metrics). By averaging the scores obtained by each method, it can be also concluded that the TSIFSD and the ADF achieve by far the best performances among the methods considered here. By coupling these results with the insights given by the performance metrics previously discussed it can be concluded that both methods are valid and that they should be preferred among the others tested here. From the scores obtained and reported in Table 6, it

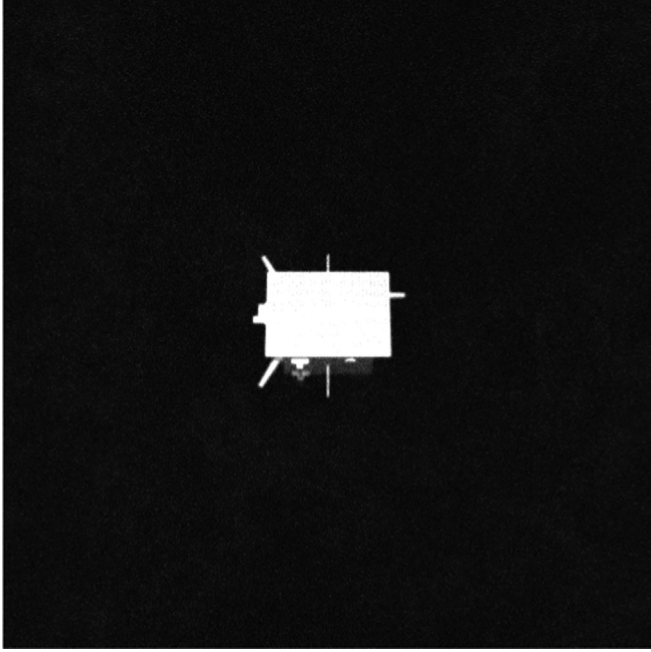


Fig. 14. Example of a saturated target in image fused with IFEVIP method, “In-Light” dataset.



Fig. 15. Example of fused image obtained by applying DRIF method, “In-Light” dataset.

appears that also the direct fusion at different resolutions can be an applicable method but the obtained images have poor quality in terms of accuracy and fidelity of the textures (see example in Fig. 15), hence it should be discarded from the possible options.

5.3. Computational time analysis and final methods selection

A summary of run time analysis of the different fusion methods already shown in Fig. 10 and Fig. 12 for all the

Table 7
Averaged run time of selected fusion methods.

Method	Average run time [s]	
	Downscaled VIS	Upscaled TIR
ADF	0.012	0.045
GFF	0.153	0.549
IFEVIP	0.024	0.094
MGFF	0.136	0.536
TSIFSD	0.013	0.056
TEPGFL	0.241	1.040

Table 6
Averaged metrics for all fusion methods applied on noisy inputs.

Fusion Method	Input Type	MSE _S		PSNR _S		SSIM _S	
		Eclipse	Light	Eclipse	Light	Eclipse	Light
ADF	VIS Downscaled	0.0028	0.0038	25.58	24.35	0.48	0.48
	TIR Upscaled	0.0012	0.0012	29.20	29.12	0.46	0.46
	Same Resolution	0.0012	0.0012	29.36	29.32	0.47	0.47
GFF	VIS Downscaled	0.0064	0.0055	22.22	22.82	0.48	0.48
	TIR Upscaled	0.0036	0.0023	24.91	26.67	0.49	0.50
	Same Resolution	0.0015	0.0013	28.56	29.13	0.50	0.50
IFEVIP	VIS Downscaled	0.0078	0.0105	21.21	20.00	0.32	0.32
	TIR Upscaled	0.0025	0.0025	26.10	26.06	0.31	0.31
	Same Resolution	0.0026	0.0026	25.86	25.84	0.29	0.29
MGFF	VIS Downscaled	0.0033	0.0042	24.96	23.89	0.47	0.46
	TIR Upscaled	0.0013	0.0013	28.84	28.73	0.44	0.44
	Same Resolution	0.0014	0.0014	28.58	28.45	0.41	0.41
TSIFSD	VIS Downscaled	0.0028	0.0037	25.69	24.50	0.52	0.52
	TIR Upscaled	0.0010	0.0010	29.97	29.85	0.48	0.47
	Same Resolution	0.0009	0.0009	30.44	30.26	0.51	0.51
TEPGFL	VIS Downscaled	0.0080	0.0069	21.14	21.81	0.35	0.36
	TIR Upscaled	0.0041	0.0025	23.97	26.21	0.23	0.29
	Same Resolution	0.0035	0.0023	24.64	26.55	0.27	0.31
DRIF	Different Resolution	0.0019	0.0024	27.12	26.30	0.44	0.42

tested image pairs, is reported by considering the average values for all the runs performed in Table 7 considering upscaled TIR and downscaled VIS images respectively. All the algorithms mentioned above are implemented in MATLAB and run on an Intel® Core™ i7-8750H CPU, clocked at 2.20 GHz. As expected, the average run time of the fusion algorithms is reduced by almost one order of magnitude when the resolution of the source images is halved. The most computationally efficient methods are IFEVIP, TSIFSD, and ADF, which always have a run time below 0.1s. It can be concluded that all three methods are good candidates for on-board implementation (even if tests with hardware in the loop should be performed to properly assess the feasibility), while the others might be too computationally intensive. As for DRIF, with an average computational time of 6s, it can be concluded that the method is too computationally intensive for spaceborne applications, and thus it will not be considered any further.

Following the outcomes of the analyses discussed before and the computational time on CPU detailed above, it can be concluded that the best way of performing pixel level image fusion with source VIS and TIR images at different resolution, is to upscale the TIR images by using bicubic interpolation, and then fusing the source images by using TSIFSD or ADF.

6. Conclusions

Relying on images acquired in the visible band of the spectrum for spacecraft relative navigation purposes is a widely explored possibility. Despite the strong advantages with respect to more complex and expensive navigation suites, it has been highlighted a strong dependency from the target illumination conditions. In case of low illumination, the target can be almost non visible in the images, jeopardizing the VIS-based navigation chain. To deal with that, this paper proposes to fuse at pixel level two source images of the same scene acquired in different spectral bands (i.e. the visible and the thermal-infrared), and to use the fused VIS-TIR image as input of the image processing and navigation algorithms.

A lack of tools capable of generating synthetic but accurate TIR spaceborne images has been identified hence, in order to test the feasibility of the proposed approach, a new physically-based TIR images rendering method has been developed and included in a tool capable of generating both VIS and TIR images. Subsequently, the synthetic images generated by the tool have been used as input to perform an extensive analysis of VIS-TIR image fusion techniques in a realistic representative scenario. To ensure the high fidelity of the rendering tool with actual thermal-infrared images, it has been developed by taking as input both a thermal simulation of the scene to be captured and also the relative pose (attitude and position) between the camera and the target. These inputs are used to get the radiant flux scaled by using the camera to target facets view factors. The TIR images are rendered by using the

radiant flux map as emitting texture. Then, the VIS-TIR image generator is used to assess the performances of image fusion methods for close proximity navigation, hence it has been used to generate images of a simplified model of Tango from PRISMA mission. It is here acknowledged that the simulation of a proper model of actual thermal-infrared camera sensors is still missing in the image generator and should be included in the next steps. It is also acknowledged that the images do not have Earth in the background. Whilst the VIS image generation is already capable to include Earth in the background, the TIR image generation tool it is cable to do that as well but further analyses on this point need to be performed.

Concerning the image fusion methods, the analyses presented comprise the most widely used techniques, based on several different architectures, hence representative of a wide range of methodologies. The performances has been evaluated by using both qualitative and quantitative metrics, that are state-of-the-art in the computer vision field. Analysis have been conducted, comparing the fusion methods in different illumination conditions and for different relative poses between target and camera, addressing the problem of dealing with noisy input images at different resolution, which is representative of a real-case application scenario with TIR cameras resolution usually lower than VIS cameras. The performed analyses have a major focus on the comparison of a wide variety of fusion methods. For this reason, the analyses present some limitations regarding the used dataset, which has a reduced size. In particular, they are considered only one spacecraft target, a single thermal simulation and the number of different illumination conditions present in the dataset is reduced. Future work can be done to investigate such aspects through an image dataset representative of a wider range of scenarios, restricting to the most promising methods found here. From the outputs of the analyses performed, it has been pointed out the sensitivity of few fusion methods to both illumination conditions of VIS images and image upscaling/downscaling, resulting in both halos and white-color saturation in the fused images, hence in extremely poor textures. On the other side, both the performance metrics, the noise rejection metrics and also the preliminary evaluations of the CPU wall-clock time for the image fusion step, highlight that TSIFSD and ADF are the best methods, due to the extremely low computational time, optimal performance metrics, high tolerance to noisy inputs, and high visual quality of produced images both in light and in eclipse conditions. On the contrary, it has been demonstrated here that directly fusing noisy images at different resolutions performs poorly for spacecraft navigation purposes, resulting in almost textureless targets in the fused image. The outcomes of this paper represent the first step towards the development and testing of innovative multispectral image-based navigation techniques. As future work, further analyses shall deeply investigate the influence of the sun-target-camera phase angle over the fusion methods, verifying their robustness to illu-

mination changes. Moreover, the computational effort of applying fusion methods shall be evaluated using representative hardware.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Avilés, M., Mora, D., Canetri, M., Colmenarejo, P., 2016. A complete ip-based navigation solution for the approach and capture of active debris. In: 67th International Astronautical Congress, Guadalajara, Mexico, pp. 1–8.
- Bavirisetti, D., Xiao, G., Zhao, J., Dhuli, R., Liu, G., 2019. Multi-scale guided image and video fusion: A fast and efficient approach. *Circ. Syst. Signal Process.* 38, 5576–5605. <https://doi.org/10.1007/s00034-019-01131-z>.
- Bavirisetti, D.P., Dhuli, R., 2016a. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform. *IEEE Sens. J.* 16, 203–209. <https://doi.org/10.1109/JSEN.2015.2478655>.
- Bavirisetti, D.P., Dhuli, R., 2016b. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Phys. Technol.* 76, 52–64. <https://doi.org/10.1016/j.infrared.2016.01.009>.
- Bechini, M., Lavagna, M., Lunghi, P., 2023. Dataset generation and validation for spacecraft pose estimation via monocular images processing. *Acta Astronaut.* 204, 358–369. <https://doi.org/10.1016/j.actaastro.2023.01.012>.
- Bechini, M., Lunghi, P., Lavagna, M., 2022a. Tango spacecraft dataset for monocular pose estimation. Dataset Zenodo. <https://doi.org/10.5281/zenodo.6499007>.
- Bechini, M., Lunghi, P., Lavagna, M., 2022b. Tango spacecraft dataset for region of interest estimation and semantic segmentation. Dataset Zenodo. <https://doi.org/10.5281/zenodo.6507863>.
- Bechini, M., Lunghi, P., Lavagna, M., 2022c. Tango spacecraft wireframe dataset model for line segments detection. Dataset Zenodo. <https://doi.org/10.5281/zenodo.6372848>.
- Blender Online Community, 2018. Blender - a 3D modelling and rendering package. Blender Foundation. Stichting Blender Foundation, Amsterdam. URL: <http://www.blender.org>.
- Boie, R.A., Cox, I.J., 1992. An analysis of camera noise. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 671–674. <https://doi.org/10.1109/34.141557>.
- Braeot, E., Groussin, O., Lamy, P., Reynaud, J.L., 2014. Experimental study of an uncooled microbolometer array for thermal mapping and spectroscopy of asteroids. *Exp. Astron.* 38, 381–400. <https://doi.org/10.1007/s10686-014-9417-y>.
- Castellini, F., Antal-Wokes, D., Pardo de Santayana, R., Vantournhout, K., 2015. Far approach optical navigation and comet photometry for the rosetta mission. In: Proceedings of 25th International Symposium on Space Flight Dynamics, Munich, Germany, pp. 1–8.
- Civardi, G.L., Piccinin, M., Lavagna, M., 2021. Small bodies ir imaging for relative navigation and mapping enhancement. In: Presentation at 7th IAA Planetary Defense Conference, Wien, Austria, pp. 1–6, URL: <https://re.public.polimi.it/handle/11311/1231286>.
- Cosine Research for ESA, 2022. Thermal Infrared Imager Assessment Study for the Asteroid Impact Mission [Nebula Public Library. URL: <https://nebula.esa.int/content/thermal-infrared-imager-assessment-study-asteroid-impact-mission>. [Executive Summary Online; accessed 30. Apr. 2022].
- D'Amico, S., Bodin, P., Delpéch, M., Noteborn, R., 2013. Prisma. In: D'Errico, M. (Ed.), Distributed space missions for earth system monitoring. Springer, New York, New York, NY. chapter 21, p. 599–637. <https://doi.org/10.1007/978-1-4614-4541-8-21>.
- Deodshumukh, V., Chaudhuri, S., Roy, S.D., 2003. Cooperative infrared and visible band tracking. In: Proceedings of the International Conference on Applied Pattern Recognition, pp. 402–405.
- Dogra, A., Goyal, B., Agrawal, S., 2017. From multi-scale decomposition to non-multi-scale decomposition methods: A comprehensive survey of image fusion techniques and its applications. *IEEE Access* 5, 16040–16067. <https://doi.org/10.1109/ACCESS.2017.2735865>.
- Du, Q., Xu, H., Ma, Y., Huang, J., Fan, F., 2018. Fusing infrared and visible images of different resolutions via total variation model. *Sensors* 18, 3287. <https://doi.org/10.3390/s18113827>.
- Duan, C., Xing, C., Liu, Y., Wang, Z., 2022. Fusion of infrared and visible images using fast global smoothing decomposition and target-enhanced parallel gaussian fuzzy logic. *Sensors* 22, 40. <https://doi.org/10.3390/s22010040>.
- Duchon, C.E., 1979. Lanczos filtering in one and two dimensions. *J. Appl. Meteorol. Climatol.* 18, 1016–1022. [https://doi.org/10.1175/1520-0450\(1979\)018<1016:LFIOAT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2).
- Fehse, W., 2014. Rendezvous with and capture/ removal of non-cooperative bodies in orbit: The technical challenges. *J. Space Saf. Eng.* 1, 17–27. [https://doi.org/10.1016/S2468-8967\(16\)30068-4](https://doi.org/10.1016/S2468-8967(16)30068-4).
- Gansmann, M., Mongrard, O., Ankersen, F., 2017. 3d model-based relative pose estimation for rendezvous and docking using edge features. In: Presentation 10th International ESA Conference on Guidance, Navigation and Control Systems, pp. 1–15.
- Gao, Y.t., Chen, H.m., Xu, Y., Sun, X.n., Chang, B.k., 2011. Noise research of microbolometer array under temperature environment. In: International Symposium on Photoelectronic Detection and Imaging 2011: Advances in Infrared Imaging and Applications, International Society for Optics and Photonics. SPIE. pp. 222–228. <https://doi.org/10.1117/12.899421>.
- Haghighat, M.B.A., Aghagolzadeh, A., Seyedarabi, H., 2011. A non-reference image fusion metric based on mutual information of image features. *Comput. Electr. Eng.* 37, 744–756. <https://doi.org/10.1016/j.compeleceng.2011.07.012>, special Issue on Image Processing.
- Jiang, T., Yunfeng, C., Meng, D., Zhouyu, Z., 2022. Visible and infrared image fusion-based image quality enhancement with applications to space debris on-orbit surveillance. *Int. J. Aerospace Eng.* 2022, 1–21. <https://doi.org/10.1155/2022/6300437>.
- Keys, R., 1981. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* 29, 1153–1160. <https://doi.org/10.1109/TASSP.1981.1163711>.
- Kisantal, M., Sharma, S., Park, T.H., Izzo, D., Märtens, M., D'Amico, S., 2020. Satellite pose estimation challenge: Dataset, competition design, and results. *IEEE Trans. Aerosp. Electron. Syst.* 56, 4083–4098. <https://doi.org/10.1109/TAES.2020.2989063>.
- Kisantal, M., Sharma, S., Park, T.H., Izzo, D., Märtens, M., D'Amico, S., 2019. Spacecraft pose estimation dataset (speed). Dataset Zenodo. <https://doi.org/10.5281/zenodo.6327546>.
- Kumar, P., Mittal, A., Kumar, P., 2006. Fusion of thermal infrared and visible spectrum video for robust surveillance. In: Kalra, P.K., Peleg, S. (Eds.), Computer Vision, Graphics and Image Processing. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 528–539. https://doi.org/10.1007/11949619_47.
- Leinz, M.R., Chen, C.T., Beaven, M.W., Weismuller, T.P., Caballero, D. L., Gaumer, W.B., Sabastanski, P.W., Scott, P.A., Lundgren, M.A., 2008. Orbital Express Autonomous Rendezvous and Capture Sensor System (ARCSS) flight test results. In: Howard, R.T., Motaghedi, P. (Eds.), Sensors and Systems for Space Applications II, International Society for Optics and Photonics. SPIE. p. 69580A. <https://doi.org/10.1117/12.779595>.
- Lentaris, G., Maragos, K., Stratakis, I., Papadopoulos, L., Papanikolaou, O., Soudris, D., Lourakis, M., Zabulis, X., Gonzalez-Arjona, D., Furano, G., 2018. High-performance embedded computing in space: Evaluation of platforms for vision-based navigation. *J. Aerospace Informat. Syst.* 15, 178–192. <https://doi.org/10.2514/1.1010555>.

- Li, S., Kang, X., Hu, J., 2013. Image fusion with guided filtering. *IEEE Trans. Image Process.* 22, 2864–2875. <https://doi.org/10.1109/TIP.2013.2244222>.
- Liu, Y., Liu, S., Wang, Z., 2015. A general framework for image fusion based on multi-scale transform and sparse representation. *Informat. Fusion* 24, 147–164. <https://doi.org/10.1016/j.inffus.2014.09.004>.
- Ma, J., Ma, Y., Li, C., 2019. Infrared and visible image fusion methods and applications: A survey. *Informat. Fusion* 45, 153–178. <https://doi.org/10.1016/j.inffus.2018.02.004>.
- Martin, I., Dunstan, M., Vural, D., Gestido, M., 2021. Simulating thermal infrared images of planets, asteroids and spacecraft. In: Presentation at 8th ICATT (International Conference on Astrodynamics Tools and Techniques) 2021 Conference Proceedings, European Space Agency, Netherlands. p. 259.
- Min, D., Choi, S., Lu, J., Ham, B., Sohn, K., Do, M.N., 2014. Fast global image smoothing based on weighted least squares. *IEEE Trans. Image Process.* 23, 5638–5653. <https://doi.org/10.1109/TIP.2014.2366600>.
- Mitchell, H.B., *Image Fusion*. Springer, Berlin, Germany. URL: <https://link.springer.com/book/10.1007/978-3-642-11216-4>.
- Musallam, M.A., Gaudilliere, V., Ghorbel, E., Al Ismael, K., Perez, M. D., Poucet, M., Aouada, D., 2021. Spacecraft Recognition Leveraging Knowledge of Space Environment: Simulator, Dataset, Competition Design and Analysis. In: 2021 IEEE International Conference on Image Processing Challenges (ICIPC), IEEE. pp. 11–15. <https://doi.org/10.1109/ICIPC53495.2021.9620184>.
- Naasz, B.J., Bums, R.D., Queen, S.Z., Van Eepoel, J., Hannah, J., Skelton, E., 2009. The hst sm4 relative navigation sensor system: Overview and preliminary testing results from the flight robotics lab. *J. Astronaut. Sci.* 57, 457–483. <https://doi.org/10.1007/BF03321512>.
- Oda, M., Inaba, N., 2000. Results of nasda's ets-vii robot mission and its applications. In: Hollerbach, J.M., Koditschek, D.E. (Eds.), *Robotics Research*. Springer London, London, pp. 431–438. https://doi.org/10.1007/978-1-4471-0765-1_52.
- Opromolla, R., Fasano, G., Rufino, G., Grassi, M., 2017. A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations. *Prog. Aerosp. Sci.* 93, 53–72. <https://doi.org/10.1016/j.paerosci.2017.07.001>.
- Palmerini, G.B., 2014. Combining thermal and visual imaging in spacecraft proximity operations. In: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), IEEE. pp. 383–388. <https://doi.org/10.1109/ICARCV.2014.7064336>.
- Park, T.H., Märtens, M., Lecuyer, G., Izzo, D., D'Amico, S., 2021. Next Generation Spacecraft Pose Estimation Dataset (SPEED+). Dataset Zenodo. <https://doi.org/10.25740/wv398fc4383>.
- Parkes, S., Martin, I., Dunstan, M., Matthews, D., 2004. Planet surface simulation with pangu. In: Presentation at Space OPS 2004 conference, p. 389. <https://doi.org/10.2514/6.2004-592-389>.
- Pasqualetto Cassinis, L., Fonod, R., Gill, E., 2019. Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft. *Prog. Aerosp. Sci.* 110, 100548. <https://doi.org/10.1016/j.paerosci.2019.05.008>.
- Piccinin, M., Civardi, G.L., Quirino, M., Lavagna, M., 2021. Multispectral Imaging Sensors for Asteroids Relative Navigation. In: Presentation at 72nd International Astronautical Congress (IAC 2021), pp. 1–12. URL: <https://re.public.polimi.it/handle/11311/1191261>.
- Plachetka, T., 1998. Pov ray: persistence of vision parallel raytracer. In: *Presentation at Proc. of Spring Conf. on Computer Graphics*, Budmerice, Slovakia, p. 129.
- Proença, P.F., Gao, Y., 2020. Deep Learning for Spacecraft Pose Estimation from Photorealistic Rendering. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 6007–6013. <https://doi.org/10.1109/ICRA40945.2020.9197244>.
- Qu, G., Zhang, D., Yan, P., 2002. Information measure for performance of image fusion. *Electron. Lett.* 38, 313–315. <https://doi.org/10.1049/el:20020212>.
- Quirino, M., Marocco, L., Guilizzoni, M., Lavagna, M., 2021. High energy rapid modular ensemble of satellites payload thermal analysis using openfoam. *J. Thermophys. Heat Transfer* 35, 715–725. <https://doi.org/10.2514/1.T6165>.
- Schnitzer, F., Sonnenburg, A., Janschek, K., Gestido, M., 2017. Lessons-learned from on-ground testing of image-based non-cooperative rendezvous navigation with visible-spectrum and thermal infrared cameras. In: Presentation at 10th International ESA Conference on Guidance, Navigation and Control Systems, pp. 1–15.
- Sharma, S., D'Amico, S., 2016. Comparative assessment of techniques for initial pose estimation using monocular vision. *Acta Astronaut.* 123, 435–445. <https://doi.org/10.1016/j.actaastro.2015.12.032>.
- Shi, J.F., Ulrich, S., Ruel, S., 2017. Spacecraft pose estimation using principal component analysis and a monocular camera. In: *AIAA Guidance, Navigation, and Control Conference*, pp. 1–24. <https://doi.org/10.2514/6.2017-1034>.
- Shirley, P., Morley, R.K., 2008. *Realistic Ray Tracing*. AK Peters Ltd, Natick, Massachusetts.
- Simone, G., Farina, A., Morabito, F., Serpico, S., Bruzzone, L., 2002. Image fusion techniques for remote sensing applications. *Informat. Fusion* 3, 3–15. [https://doi.org/10.1016/S1566-2535\(01\)00056-2](https://doi.org/10.1016/S1566-2535(01)00056-2).
- Singh, R., Vatsa, M., Noore, A., 2008. Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recogn.* 41, 880–893. <https://doi.org/10.1016/j.patcog.2007.06.022>.
- STAR-Dundee, 2022. PANGU - Planet and Asteroid Natural scene Generation Utility - v6 datasheet. URL: https://www.star-dundee.com/wp-content/star_uploads/product_resources/datasheets/PANGU-v6-Datasheet.pdf. [Online Datasheet; accessed 18. Jan. 2023].
- Starek, J.A., Açıkmeşe, B., Nesnas, I.A., Pavone, M., 2016. Spacecraft autonomy challenges for next-generation space missions. In: Feron, E. (Ed.), *Advances in Control System Technology for Aerospace Applications*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 1–48. https://doi.org/10.1007/978-3-662-47694-9_1.
- Toet, A., 2011. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2131–2146. <https://doi.org/10.1109/TPAMI.2011.53>.
- Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. <https://doi.org/10.1109/TIP.2003.819861>.
- Wong, P.W., Herley, C., 1997. Area-based interpolation for scaling of images from a ccd. In: *Proceedings of International Conference on Image Processing*, vol. 1. pp. 905–908. <https://doi.org/10.1109/ICIP.1997.648113>.
- Yilmaz, O., Aouf, N., Majewski, L., Sanchez-Gestido, M., 2017a. Evaluation of feature detectors for infrared imaging in view of active debris removal. In: Presentation at 7th European Conference on Space Debris, pp. 1–8. URL: <https://conference.sdo.esoc.esa.int/proceedings/sdc7/paper/486>.
- Yilmaz, O., Aouf, N., Majewski, L., Sanchez-Gestido, M., Ortega, G., 2017b. Using infrared based relative navigation for active debris removal. In: Presentation at the 10th International ESA Conference on Guidance, Navigation and Control Systems, pp. 1–16.
- Zhang, Y., Zhang, L., Bai, X., Zhang, L., 2017. Infrared and visual image fusion through infrared feature extraction and visual information preservation. *Infrared Phys. Technol.* 83, 227–237. <https://doi.org/10.1016/j.infrared.2017.05.007>.