IAC–22–A6.IP.11.x74248

Robust Monocular Pose Initialization via Visual and Thermal Image Fusion

Michele Bechini*, Gaia Letizia Civardi†, Matteo Quirino‡, Alessandro Colombo§, Michèle Lavagna¶

Monocular-based relative guidance navigation and control chain plays a crucial role in the new missions for proximity operations in orbit. To provide high-quality images in the widest range of scenarios as possible without overconstraining the mission analysis or the mission planning, it is here proposed to use as input for the image processing and pose estimation algorithm, multispectral images obtained by fusing at pixel level two images from two different monocular cameras operating in the visible and the infrared rage of the spectrum respectively. Since fused multispectral images have never been used for this purpose, the main objective of the work is to verify if the content of information that can be retrieved is enough to assess the relative pose between a chaser and an uncooperative known target and if they can be safely used as the primary input. The tool used to synthetically generate the images is here described as well as the pose estimation algorithm applied. During the reimplementation of the baseline algorithm, modifications and improvements have been introduced to make the edge detection less sensitive to variations in images and to better estimate a priori the size of the match matrix needed for the pose estimation in a more general framework with the proposed mathematical formulation. The tests performed with the pose estimation algorithm on the multispectral images revealed that they can be adopted as the primary source of measurement since their content of information is higher than the single visible or infrared images used alone, avoiding the problems that characterize both these spectral bands. This result is also confirmed by the outcomes of the relative pose estimation algorithm which shows impressive results in terms of accuracy for a feature-based algorithm.
keywords: Relative Navigation, Thermal Infrared Images, Monocular Images, Multispectral Images, Image Fusion, Pose Estimation

1. Introduction

New missions for proximity operations in orbit between artificial objects gained increasing attention during the last years with the aim of performing regular in-orbit services [1]. A particular interest is put on the relative state estimation of a chaser with respect to uncooperative targets (i.e. target spacecraft is not equipped with light-emitting markers nor is capable of communicating with the chaser). The relative pose (position and attitude) between spacecraft could be estimated in principle by using ground-based tracking approaches, but the main drawback is that the estimate is strongly affected by high uncertainty and would depend on the visibility of the spacecraft from the ground stations [2]. As a consequence, a ground-based approach is not well suited for scenarios like formation flying missions (FF) with fractionated scientific payloads, on-orbit servicing demonstrators (OOS), and active debris removal. Despite significant technology developments are still needed to make these missions feasible, the high level of reactivity of the chaser with respect to the target required for close proximity operations and maneuvering imposes the estimation of the relative pose directly onboard, relying only on the chaser capabilities. Hence, dealing with artificial uncooperative targets represents the most challenging scenario, requiring robustness in both nominal and off-nominal operations. Notice that the onboard pose estimation is only the first step towards a guidance, navigation, and control (GNC) chain solved autonomously directly on board ensuring timeliness, reactivity, effectiveness, and robustness.

Among the possible sensor suites, the ones that include monocular cameras for imaging are the most attractive solutions to collect meaningful measurements for the onboard GNC chain due to low power consumption, cost, and mass [2, 3]. More in detail, monocular cameras operating in the visible spectrum (VIS) have been widely studied and already applied

*Polytechnic University of Milan, Italy, michele.bechini@polimi.it
†Polytechnic University of Milan, Italy, gaialetizia.civardi@polimi.it
‡Polytechnic University of Milan, Italy, matteo.quirino@polimi.it
§Polytechnic University of Milan, Italy, alessandro43.colombo@mail.polimi.it
§Polytechnic University of Milan, Italy, michelle.lavagna@polimi.it

to both cooperative [4] and uncooperative [5] rendezvous missions. Despite the high-quality images that could be obtained by using VIS cameras, it must be noticed that the VIS images are strongly affected by the illumination conditions. In case of low illumination the target spacecraft can be partially visible or almost not visible (see Fig. 1), and in low contrast with respect to the background (being both a celestial body or the deep space) [6, 7], while in case of high direct illumination, depending on the camera, target, and Sun relative positions, saturations, flares, and stray lights noises may occur [8] (see Fig. 2). Both these conditions are highly challenging and can strongly affect the image quality and hence the output of an image processing chain tuned on nominal conditions, especially for LEO missions where the illumination conditions change abruptly [9]. To overcome this, the operation planning of a mission could be constrained to take into account these side effects and perform proximity operations only in optimal conditions but this can strongly limit OOS missions leading to limited opportunity to properly detect and track the target itself with unacceptable either mission length or risk increase. A valid alternative that can be adopted is to use another camera sensor operating in the thermal infrared (TIR) region of the spectrum, as highlighted in [6]. TIR cameras only rely on the radiance emitted by the target and collected by the sensor array of the camera itself hence, they are insensitive to the illumination conditions (as a consequence they do not suffer the issues shown in Figs. 1 - 2). However, despite the benefits of being insensitive to the illumination conditions, TIR cameras are characterized by smaller sensor sizes than VIS cameras hence TIR images have lower resolution, poorer textures, and higher noise levels. These drawbacks negatively affect the image processing algorithms [10]. The strategy of keeping the advantages of both VIS and TIR images has been explored in [6] by fusing at pixel level both the images into a single frame which retains more information than the single images. The fused of [6] are supposed to be fed to the image processing pipeline and to be adopted in a GNC chain.

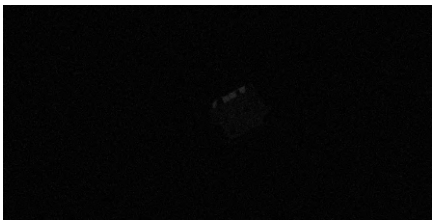


Fig. 1: Example of rendered VIS image in low illumination conditions.

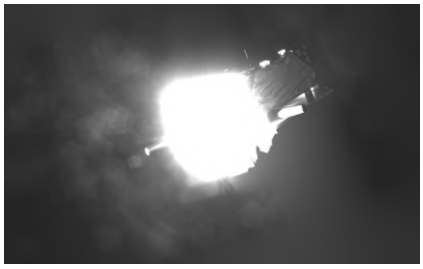


Fig. 2: Example of mockup VIS image in high illumination conditions [8].

The main contribution of this paper is the evaluation of the performances in the relative pose estimation that can be achieved by applying state-of-the-art algorithms for pose initialization already tested on VIS images on VIS-TIR fused images. In doing so, a new mathematical formulation to forecast the size of the match matrix in the modified version of the Sharma-Ventura-D'Amico (SVD) algorithm is proposed together with a new constrained way of organizing the features belonging to each perceptual group that allows the reduction of the generic search space for the pose estimation problem. By applying the SVD to fused VIS-TIR images, it has been verified that thanks to the high-quality information that can be extracted from fused images also in low light conditions, it is possible to retrieve the relative pose with low estimation errors. On the other side, it has been qualitatively proven that the Hough transform is not well suited for application in the wide range of scenarios that can be faced during a real mission hence, other feature detection methods should be preferred.

The remainder of this paper is organized as follows: the available literature on the topics related to this article is reviewed in Sec. 2.3, while in Sec. 3.3 the tool for VIS and TIR images generation used in this work is briefly described. The SVD implementation details as well as the modifications and improvements introduced are reported in Sec. 4.2, while the results of the modified SVD applied to the fused VIS-TIR image are shown in Sec. 5. The main conclusions and some hints for possible future developments are listed in Sec. 6.2.

2. Literature Review

2.1 *Thermal Rendering*

Thermal rendering for space objects is still an emerging topic and therefore, there are few approaches available in the literature. One approach is to convert the visible image into an infrared one by simply scaling the digital number of the pixels, as done in [11]. Unfortunately, avoiding the computation of the temperature field means that there are shadows in the resulting image, which do not exist in real infrared images as everything emits radiations. Furthermore, the thermal inertia of the object is not modelled, thus to simulate transient effects multiple varying light sources must be introduced. A step forward in the workflow is to have a simplified thermal model of the object and the model of the thermal camera such as proposed in [12]. The method adopted in this article increase the accuracy of the latter method by the computation of a high fidelity finite volume thermal model of the object [13]. Having such a high fidelity model enables the computation of the view factors between each mesh face of the object and the camera, allowing the computation of the radiant flux received by the camera thus producing high-quality infrared images.

2.2 *Image Fusion*

To overcome the inherent limitations of the distinguished spectral bands, pixel-level image fusion is proposed in this work. 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, such as object recognition [14], detection for surveillance [15] and remote sensing [16], yet they have never been applied in the context of spaceborne navigation. Several pixel-level image fusion algorithms exist and they can be grouped according to their baseline theory, as described in [17]. The main categories are multi-scale transform, sparse representation, neural networks, subspace and saliency-based methods, hybrid models, and other methods. Building on the outcome of the image fusion techniques comparison presented in [6], only multi-scale algorithms have been considered for this work. These methods are characterized by three common steps: the two source images are first decomposed into components at different scales using techniques such as pyramid transformation of edge-preserving filters. Then, the multi-scale representations of the VIS and TIR images are fused according to a given fusion rule. Lastly, the fused image is obtained through the inverse multi-scale transformation. The two fusion methods considered in our research are here briefly outlined.

2.2.1 *Anisotropic diffusion-based fusion (ADF)*

The implementation is based on the one described in [18]. 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. Lastly, the fused image is reconstructed through a simple linear combination of fused approximation and detail layer.

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

The algorithm employed in this study is inspired by the one presented in [19] 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, 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 to image noises. Base layer fusion is achieved as a weighted summation.

2.3 *Pose Initialization*

Fused VIS-TIR images have never been used before as direct input for the image processing step to retrieve the relative pose between two spacecraft. Despite that, the problem of estimating the relative pose of a known uncooperative target with respect to a monocular camera has been widely studied in the last few years. There are three main approaches for pose estimation that are: pure feature-based estimation, pure deep learning-based estimation, and hybrid methods. Feature-based methods are the most "classical" ones and the first that have been proposed to deal with relative pose estimation in space. These methods are based on the extraction of hand-crafted features (like corner [20] or edges [21]) of the target from a single image that are used, together with the features information retrieved from the a priori knowledge of a 3D CAD model of the target, to solve the Perspective-n-Points (PnP) problem and get an estimate of the relative pose between the camera and

the target reference frames. In 2014, the first algorithm to enable proximity operations among uncooperative spacecraft was proposed [22] and subsequently tested on true spaceborne images from the Prototype Research Instruments and Space Mission technology Advancement (PRISMA) mission [23]. In 2018, the Sharma-Ventura-D'Amico (SVD) algorithm [24] was proposed as an advancement of the algorithm in [22]. The SVD achieved state-of-the-art (SOTA) performances improving both the robustness and the efficiency achieved in [22]. A further improvement upon the SVD was proposed in 2019 in [25], where a three-streams image processing was adopted. The three parallel streams independently extract corners or edges, but only the features common to all the streams are used to solve the PnP, compensating for possible false detections in one of the streams. Despite the slight improvements in the robustness achieved, the computational cost of the algorithm in [25] is strongly increased with respect to the SVD due to the three parallel streams, hence it could not be applied in a real case scenario. Concerning the pure deep learning-based methods, the relative pose is directly regressed by using convolutional neural networks (CNNs) trained on this specific task. Depending on the formulation used, the relative pose can be estimated through CNNs by solving a direct regression problem [26, 27], or a pure classification problem (even if a refinement of the regressed pose is usually required) [28], or as a hybrid regression-classification problem [29]. The last approach is a hybrid between pure feature-based and deep learning-based methods where the CNN is used to regress landmarks or feature points that are then coupled with the features extracted from an available 3D CAD model and used to solve the PnP problem and obtain the relative pose [2, 30]. From the results of the 2019 ESA's Kelvins Pose Estimation Challenge [31] it can be concluded that the most effective approach to deal with relative pose estimation via monocular images is using hybrid methods, hence CNNs aided PnP solvers.

It must be noticed that the impressive robustness and effectiveness of CNN-aided methods are due to the training phase which is highly time-demanding. To learn proper weights and achieve high performances, the training phase must be carried out by using properly labeled image datasets. Despite that, nowadays only few spaceborne datasets are publicly available: URSO [26], SPEED [32], SPEED+ [33], and the *Multi-Purpose Labeled Spacecraft Dataset* [7] (which comprises also relative pose labeled images [34]). Due to this lack of datasets and also because all the images of the available datasets are VIS images, it is not possible to train CNNs to handle TIR or fused VIS-TIR images. Hence, in this work, only feature-based algorithms have been considered. In particular, the SVD has been selected as the baseline algorithm due to its robustness and, mostly, due to its high efficiency with respect to all the other feature-based pose initialization algorithms proposed.

3. VIS-TIR Fused Images Generation

The following subsections briefly describe the steps needed to obtain the VIS-TIR fused images starting from the rendered VIS and TIR source images. Both image types are obtained using Blender [35] as the rendering engine inside the image generation tool described in [6]. For a more detailed description of the tool and all the design choices adopted for the tool itself, please refer to [6].

3.1 *VIS Image Generation*

Currently, photorealistic VIS images can be rendered by using several tools, most of which are based on OpenGL (which allows the quick generation of scenes via rasterization) or on ray-tracing (slower but more accurate and physically based). By considering only open-source software, POV-Ray [36] has been already successfully adopted to generate photorealistic spaceborne validated VIS images in [7], while it has never been adopted to generate TIR images. In this paper, Blender has been preferred as the main software for image rendering since both VIS and TIR images can be generated [6] by using "Cycles" as the rendering engine. Cycles is a rendering engine that uses backward path tracing, a process similar to backward ray tracing, where paths are scattered from each pixel of the camera through the scene and propagated until they hit a light source. An example of a VIS image obtained by using Blender's Cycles is reported in Fig. 3. The image reported shows the render obtained by using as target a simplified Tango model as in [7]. The simplified Tango model is always employed for all the images in this work. Also, the camera parameters given in Tab. 1 will be left unchanged throughout the paper.

Table 1: VIS Camera characteristics.

array size	1024×1024 px
FoV	$35.45° \times 35.45°$
Focal Length	17.6 mm

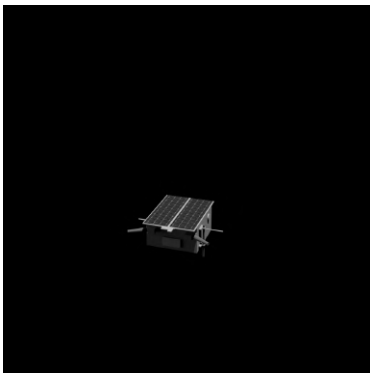


Fig. 3: Example of rendered VIS image.

VIS images are mostly affected by electronic noise and blurring due to the fixed depth of field of real cameras. To replicate real spaceborne images, the noiseless VIS images obtained by the rendering chain (as the one reported in Fig. 3) are postprocessed by adding a white Gaussian noise with $\sigma^2 = 0.0022$ and blurred with a Gaussian blurring characterized by $\sigma^2 = 1$ and zero mean. The noise levels have been selected equal to [6, 7, 31]. An example of the VIS image reported in Fig. 3 after the noise postprocessing is shown in Fig. 4.

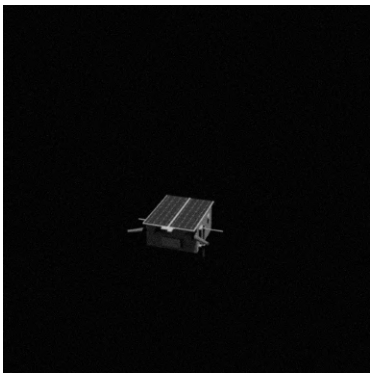


Fig. 4: Example of rendered VIS image noised.

3.2 *TIR Image Generation*

The synthetic generation of infrared images consists in computing the temperature field of the object and, given the position of the camera, computing the view factors between each face of the object mesh and the camera itself. Using the view factors it is possible to compute the radiance emitted by each face in the direction of the camera. The radiance field is then applied as a texture on top of a Lambertian emitter and then, using the same ray tracing technique used for the VIS image generation, it is possible to convert the radiance field received by each pixel into the respective digital number (DN). In this way, the principle of a thermal camera is emulated. For the sake of clarity, it is here reported a picture of the temperature field under exam Fig. 5 and a picture of the respective radiance field Fig. 6. For more details on the actual computation and the temperature field and the radiance field, please refer to [6]. It is important to point out the effect of the view factors on the radiance field. As it is highlighted by the red arrows in Fig. 6, as the view factors approach zero value in the directions tangent to the curved surfaces, the radiance field goes to zero as well. This is a qualitative effect that is well caught by the TIR rendering tool and it is fundamental for future validations. To generate the actual noiseless TIR image, the radiance field is mapped on top of a Lambertian emitter and the rendering of the scene is performed through Cycles directly in Blender. The camera settings for the TIR image are reported in Tab. 2 and an example of the final TIR picture is reported in Fig. 7.

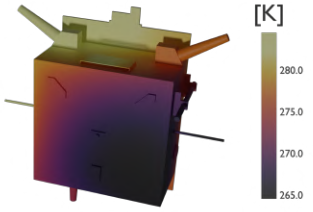


Fig. 5: Tango temperature field rendering.

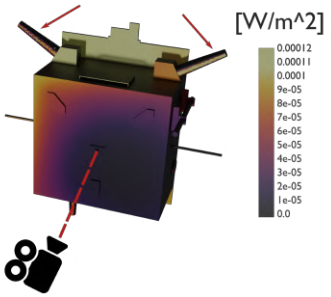


Fig. 6: Tango radiance field rendering.

To increase the photorealism of TIR images, characteristic noises are applied in a postprocessing phase, as done for VIS images. With regards to thermal imaging sensors, microbolometers are mostly affected by two sources of noise: the thermal noise and

Table 2: TIR Camera characteristics.

Array Size	512×512 px
FoV	$35.45° \times 35.45°$
Focal Length	17.6 mm

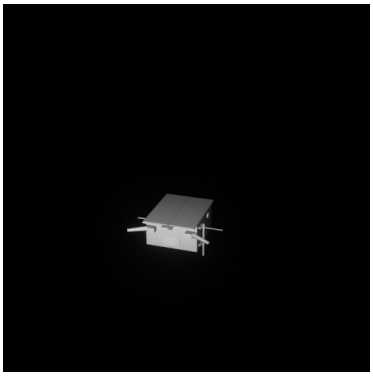


Fig. 7: Example of rendered noiseless TIR image.

the $1/f$ (or flicker) noise. The former is a characteristic of all electronic devices and it is modeled as a white noise, assuming the same characteristics adopted for VIS images [6]. The flicker noise is obtained by applying a suitably shaped low-pass filter to an additive white gaussian noise characterized by $\sigma^2 = 0.0022$ and zero mean. An example of the TIR image reported in Fig. 7 after the noise postprocessing is reported in Fig. 8.

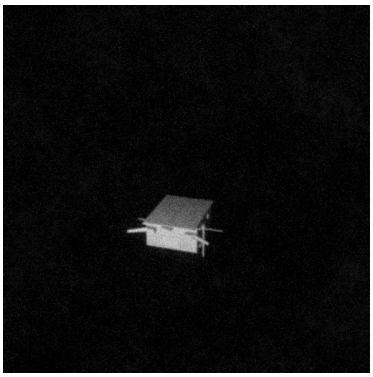


Fig. 8: Example of rendered TIR image noised.

3.3 *VIS-TIR Image Fusion*

All the fusion techniques previously described, share the assumption that the two source images must have the same resolution. To handle the resolution mismatch between the source VIS and TIR images in real case scenario, the latter are upscaled using the bicubic interpolation method, since in [6] it has been pointed out to be the best-suited method. An example of the fused VIS-TIR image obtained by applying the ADF method to Fig. 4 and the upscaled Fig. 8 is shown in Fig. 9. It can be noticed that the fused image retains the pixel brightness of the source image while preserving most of the details and texture information of the VIS source image.

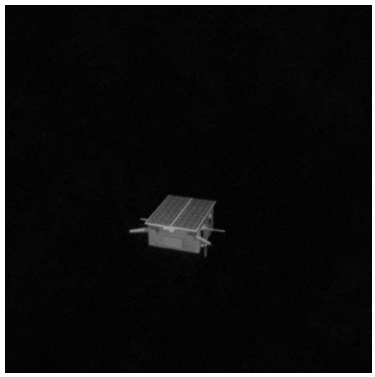


Fig. 9: Example of fused image obtained with ADF.

4. Modified SVD Algorithm

To better assess the possibility of using the VIS-TIR fused images previously described, they have been labeled with the relative pose during the rendering process and fed as input to a pose initialization algorithm. As said before, since there are no datasets of TIR or VIS-TIR fused images and since the computational time to generate a TIR image is not compatible with the generation of a dataset of proper size, it has been decided to use a feature-based algorithm. The SVD algorithm has been selected due to its high accuracy and relatively low computational cost with respect to other algorithms. The original SVD algorithm proposed by Sharma [24] was fully reimplemented on *Python 3.10* by applying some modifications (see Sec. 4.1) and by proposing a mathematical formulation as well as a more detailed description of the constraint adopted to reduce the search space during the definition of the match matrix between 3D and 2D perceptual groups (see Sec. 4.2) that was not given in the original paper. The interested reader is referred to the original paper [24] for a more detailed description.

4.1 *SVD Feature Detection*

The feature detection subsystem of the SVD is the image processing (IP) step of the algorithm. This block receives as input the 2D image and gives as

output the detected features collected in perceptual groups that are then fed to the subsequent block, which is the pose determination subsystem, described in Sec. 4.2. The image fed to the IP step is preprocessed with a Gaussian filter with standard deviation $\sigma = 1.15$ to reduce the magnitude of the noise. During some tests performed it has been highlighted that the SVD is highly sensitive to the selection of this hyperparameter and that a bad selection can completely jeopardize the feature detection step, hence the entire algorithm [37]. The IP of the SVD is then split into two branches, the weak gradient elimination (WGE) and the Sobel branches, which are then merged again after the feature detection.

The first task that is accomplished in the WGE branch is the computation of the image gradient of the blurred input image. This is performed by using the intermediate gradient algorithm (instead of the Prewitt filter originally proposed) since it resulted to be less sensitive to background and noises. The WGE is then applied to the image, performing also the detection of the spacecraft in the image. The WGE is particularly useful in locating the spacecraft in the image also in presence of the background, by canceling out the weak image gradients that belong to the background itself. The WGE is performed by normalizing, sorting in 100 uniform bins, and fitting the distribution by an exponential probability distribution function. All the pixels that correspond to a cumulative distribution lower than a threshold are classified as "weak" and set to zero. For the case analyzed here, the threshold value for the cumulative distribution is selected as equal to 0.998. After the WGE, the Region of Interest (ROI) can be defined by computing the Cumulative Distribution Function (CDF) of the gradients along the x and y axis in the image reference frame. Assuming a normal distribution for the gradients, by limiting the CDF to the values between 0.05 and 0.95, only the central 90% is retained, which corresponds to the ROI. The ROI detected in this process is here enlarged by 10% to avoid cutting out the ROI portions of the target spacecraft.

Concerning the Sobel branch, the original image is processed by using the Sobel operator to identify the edges in the image. In the implementation here presented the intensity image is thresholded by setting to zero all the pixels with an intensity magnitude below 0.9. Then, to improve the performances of the Hough transform, the image is processed by performing an orientated non-maximal suppression and by removing all the pixel chunks that are formed by less than a threshold amount of pixels [37] and that may act as a noise for the line detector. The pixel threshold value is selected as equal to 1% the diagonal size of the ROI.

The final part of both the WGE and the Sobel branches is the features extraction and merging. The original SVD implementation adopted the Hough transform [38] as a feature extractor, applied to the full-scale gradient image previously computed. To apply the Hough transform, the following hyperparameters must be defined: the definition of the resolution of the evaluated line distances ρ_{res} and angles θ_{res}, the number of peaks to be identified in the Hough transform matrix, the expected minimum length of the line segments L_{min}, and the maximum gap between two points to be considered in the same line segment λ_{max}. The effectiveness of the Hough transform is strongly linked to these hyperparameters and, depending on the position and orientation of the target in the image, they should be tuned by hand per each image. To avoid this, in [24] it is suggested to scale both L_{min} and λ_{max} linearly with the diagonal size in pixel of the ROI through the constants k_1 and k_2 respectively. Some of the test performed during this study highlighted that the linear scaling is not sufficient and that the Hough transform still perform badly for most of the VIS images from the SPEED dataset that we used as the initial benchmark for the SVD. As a consequence, our implementation of the SVD performs the Hough transform only on the image cropped to the ROI detected. This results in a reduction of the sensitivity of the hyperparameters with respect to the position of the target in the image. Once the line segments have been detected, they are merged to avoid duplicates and/or truncated or segmented edges that can jeopardize the identification of the perceptual groups. After that, the line segments of the two separated branches are merged all together to create the final set of features. Again, the merging process is tuned to avoid duplicated and/or truncated line segments. For more details on both the merging needed, the interested reader is referred to [24, 37].

4.2 *Match Matrix Definition and Pose Estimation*

The main innovation introduced by the SVD is the feature synthesis, which refers to the organization of the simple line segments into high-level geometrical groups named "perceptual groups" to reduce the search space of the correspondence problem to be solved with the Efficient PnP (EPnP) algorithm. Notice that a minimum of six correspondences from the n 2D features to the m 3D points of the available 3D CAD model are needed to solve the EPnP, leading to

a number of correspondences equal to [24]:

$$\binom{m}{6}\binom{n}{6}6! \quad [1]$$

The intuition presented in [22] and better developed in [24] was that it was possible to build more complex high-level feature groups that reduce the dimensionality of the correspondence problem by introducing some geometrical constraints between the feature points. The perceptual groups used also in this paper are the parallel pairs, proximity pairs, parallel triads, proximity triads (or open tetrads), and closed tetrads. The antennas are treated as a separate perceptual group. Increasing the level of complexity of the perceptual group reduces the probability of false detection. The original paper proposes some geometrical constraints that must be verified to properly identify the perceptual groups. These constraints can be tailored to the examined case by tuning some threshold values and, among them, one of the most important is the parameter τ that is multiplied by the diagonal size of the ROI to get the threshold length below which each line segment is categorized as antennas. The same geometrical constraints are applied to the line segments detected both in the 2D image and in the 3D CAD model. From the 3D CAD model used in this work [24], it is possible to extract 18 parallel pairs, 16 proximity pairs, 12 parallel triads, 12 proximity triads, 2 closed tetrads, and 5 antennas. For more details on the geometrical constraints to identify the perceptual groups, refer to [24, 37].

Once both the 2D and 3D perceptual groups are available, the original paper proposed to just combine the endpoints of corresponding feature groups through simple combinations to build a "match matrix". Because at least 6 correspondences are needed for the EPnP, all the perceptual groups except the parallel triads must be combined with other feature groups. The antennas are used as an additional feature to reach the six correspondences needed and, for this reason, the correct classification is of paramount importance. How the feature groups are combined and a general mathematical formulation for the size of the match matrix is not reported in the original paper. Hence, in this paper, we present a novel way of constraining the feature groups by including rules on how the vectors for each line segment are stored during the execution of the algorithm. By using these additional constraints, it is possible to derive mathematical expressions to evaluate the number of rows in the match matrix needed for each combination of perceptual groups proposed in [24] and hence, to know the dimensionality of the problem a priori.

Concerning the antennas, the possible correspondences between 3D and 2D antennas can be halved by knowing which endpoint is the tip and which is the root. The tip in 3D and 2D can be identified as the further endpoint from the origin of the body reference frame and the center of the ROI respectively. The antenna information are saved in an ordered list as: $Antenna = [\mathbf{P}_{root}; \mathbf{P}_{tip}]$. In the same way, the list containing the information of a 2D or 3D closed tetrad can be ordered as: $Tetrad = [\mathbf{P}_{TL}, \mathbf{P}_{TR}, \mathbf{P}_{BR}, \mathbf{P}_{BL}]$, where the terms are respectively the location of the top-left, top-right, bottom-right, and bottom-left endpoints of the closed tetrad. By maintaining for each 2D and 3D tetrad this "clockwise" ordering of the points, it is possible to reduce the correspondences between 2D and 3D tetrad to 8 without violating the geometry of the tetrad itself. Regarding the proximity triads (or the open tetrads), it is possible to give an ordering by writing the list starting from an end (one of the two "free" endpoints) and then by writing the endpoints up to the other end, hence: $ProxTriad = [\mathbf{P}_{1,1}, \mathbf{P}_{1,2}, \mathbf{P}_{2,1}, \mathbf{P}_{2,2}]$ where $\mathbf{P}_{1,1}$ is the starting point (a free end of the open triad), $\mathbf{P}_{1,2}$ is the other endpoint belonging to the same line segment of $\mathbf{P}_{1,1}$, $\mathbf{P}_{2,1}$ is the endpoint linked to $\mathbf{P}_{1,1}$ by another line segment, and $\mathbf{P}_{2,2}$ is the other free-end of the open tetrad. By using the aforementioned ordering it is possible to constraint the number of admissible combinations to 2. In the same way, it is possible to provide an ordering also for the proximity pairs, by starting from one of the two free ends. In particular, the list can be written as: $ProxPair = [\mathbf{P}_1, \mathbf{P}_{common}, \mathbf{P}_2]$ where $\mathbf{P}_1$ and $\mathbf{P}_2$ are the two free ends, while $\mathbf{P}_{common}$ is the common point that is shared by the two line segments that form the proximity pair. As for the case of proximity triads, by using the ordering here proposed, there are only 2 admissible combinations of 2D and 3D proximity pairs. Moving to the parallel triads, it must be noticed that in the perspective geometry the positional order of the three line segments that are involved can be changed by simply changing the observation point, hence it can not be constrained. As a consequence, the ordering can only be given by writing one segment per time (hence two endpoints) ordered such that the unit vectors written from one endpoint to the other of each line segment are in agreement (i.e. the cosine of the angle comprised between each pair of unit vectors is higher than zero). Thus, the list can be written in an ordered way as:

$ParTriad = [\mathbf{P}_{1,1}, \mathbf{P}_{1,2}, \mathbf{P}_{2,1}, \mathbf{P}_{2,2}, \mathbf{P}_{3,1}, \mathbf{P}_{3,2}]$, where $\mathbf{P}_{i,j}$ with $i = 1, ..., 3$ and $j = 1, 2$ are the endpoints of the line segments, written such that the unit vectors from $\mathbf{P}_{i,1}$ to $\mathbf{P}_{i,2}$ are in agreement. In this case, since in the projective geometry it is not possible to flip the pointing direction of one of the unit vectors without changing the pointing direction of all of them, the ordering introduced limits the possible combinations of 2D and 3D parallel triads to 12. The same ordering can be adopted also for the parallel pairs hence, by using the same notation used above, it is possible to write the list as: $ParPair = [\mathbf{P}_{1,1}, \mathbf{P}_{1,2}, \mathbf{P}_{2,1}, \mathbf{P}_{2,2}]$. For the parallel pairs, by using the same properties of the projective geometry used before, it is possible to constrain the feasible combinations to 4.

The constraints firstly introduced here above and resumed in the scheme in Fig. 10 strongly reduce the dimensionality of the problem with respect to the generic solution reported in Eq. 1. To build the match matrix, it must be noticed that the parallel triads already have 6 features and can be used alone, while all the other perceptual groups must be combined with the antennas. In particular, it should be noticed that the tip of a visible detected antenna is always in view, this does not happen for the root. Hence, here only the tip of the antennas are used to reach the 6 features required by the EPnP, increasing the accuracy of the algorithm. Due to this, 3 2D antennas are required for the proximity pairs, while for all the other perceptual groups considered here 2 2D antennas are needed. By defining as $\phi_{a,2D}$ the number of 2D antennas detected in the image, and as $\phi_{a,3D}$ the number of 3D antennas in the 3D CAD model, it is possible to evaluate the number of combinations needed to build the match matrix as a function of k, the number of antennas needed to obtain a complete set of 6 features per each perceptual group, as:

$$\frac{1}{k!}\frac{\phi_{a,2D}!}{(\phi_{a,2D}-k)!}\frac{\phi_{a,3D}!}{(\phi_{a,3D}-k)!} = \frac{1}{k!}D_{a_{2D},k}D_{a_{3D},k} \quad [2]$$

By using the possible combinations of 2D-3D perceptual groups reported above with Eq. 2, the general formulation to forecast the rows occupied by the combinations of each feature group (see Tab. 3) in the match matrix can be derived mathematically. It must be noticed that the equations reported in Tab. 3 are general and are valid for each case since they are not tailored to a well-defined scenario as the ones reported in [24]. Notice that it is extremely important to properly tune the parameter τ to correctly classify the antennas due to their importance in the disambiguation of the EPnP solutions, but also since the dimensions of the match matrix explode if there is a high number of line segments classified as antennas.

After the definition of the match matrix as reported above, all the combinations of 2D-3D features identified are fed to the EPnP solver to define a pose for each of them. Then the five best poses in terms of reprojection error are further optimized by using a Newton-Raphson optimization. After this process, the reprojection error is evaluated once again and the best final pose is selected [24].

5. Fused VIS-TIR Pose Initialization

The SVD with all the modifications described above has been applied to retrieve the relative pose of a simplified Tango spacecraft model (as the one described in [7]) by using only fused VIS-TIR images. The objective was to assess if the information that can be retrieved from a fused VIS-TIR image are such that the pose can be estimated correctly, as for VIS images in nominal illumination conditions. The SVD is based on the evaluation of image gradients (i.e. the contrast between different objects) hence, by applying it, it is possible to assess also if the fusion can correctly keep an acceptable contrast level by averaging between the VIS (high contrast) and TIR (low contrast) images. To better verify the improvement offered by using fused images during the pose estimation, the benchmark case has been selected such that the VIS image is taken in low light conditions (see Fig. 11a), when almost no information can be retrieved from the VIS image. The TIR image associated with the VIS image is reported in Fig. 11b where it can be noticed that the contrast of the target with the background is high, but the contrast between the surface of the target is almost null. The fused VIS-TIR image used as a benchmark is given in Fig. 11c and it has been obtained by using the ADF algorithm.

The hyperparameters discussed in Sec. 4.2 have been tuned to perform correctly on the benchmark image and to retrieve significant edges. The values of the hyperparameters used are reported in Tab. 4. Notice that the same values have been used for the parameters of the Hough transform for both the streams allowing to have some redundancies in the features detected. This is in contrast to what has been done in the reference paper [24], where the Hough parameters are different to detect distinct features in the two streams. The error in the estimation of the relative pose $(\hat{\mathbf{r}}, \hat{\mathbf{q}})$ with respect to the ground truth $(\mathbf{r}, \mathbf{q})$ have been evaluated in terms of the error in the estimation of the relative translational position along

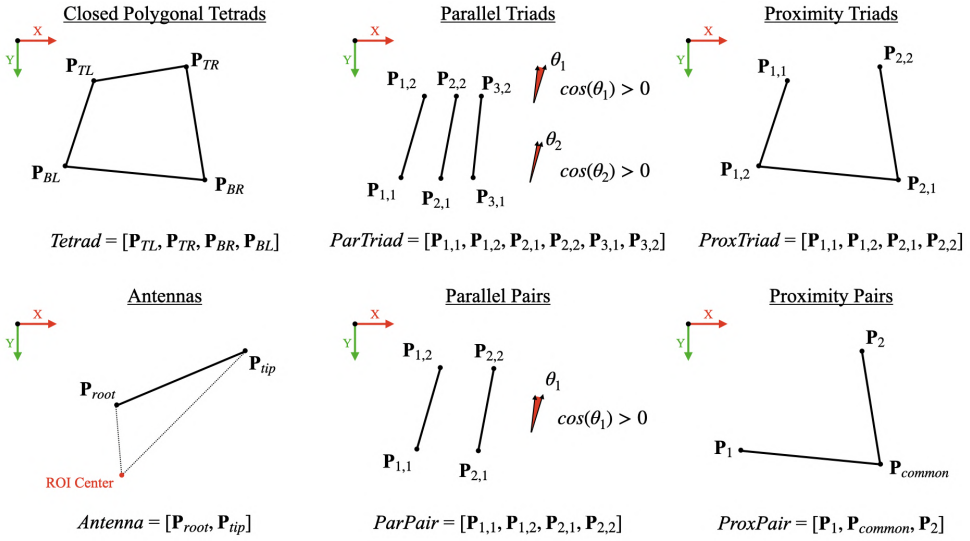
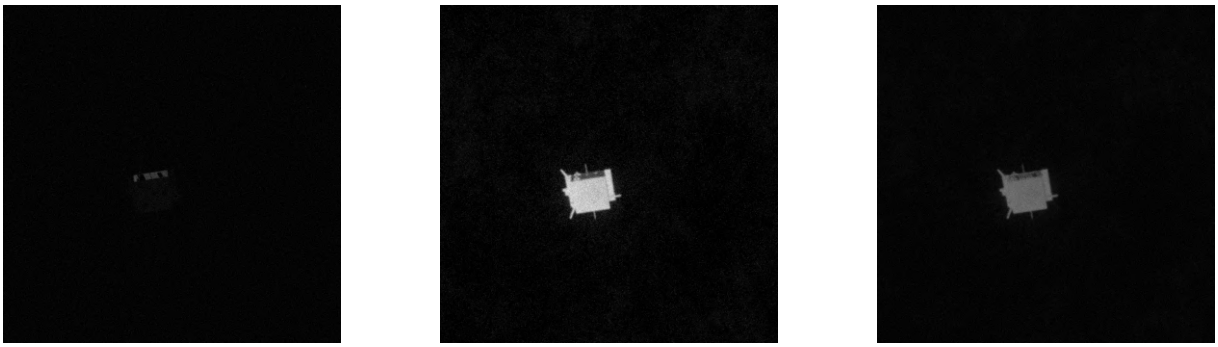


Fig. 10: Schematic of the constrained ordering of the perceptual groups.

Table 3: Rows of each perceptual group in the match matrix.

Feature Group	Points per feature group	Number of 2D feature groups	Number of 3D feature groups	Number of rows in match matrix
Antenna	1 (tip only)	ϕ_a	$\phi_a^{'}$	– –
Closed Tetrad	4	ϕ_b	$\phi_b^{'}$	$8\phi_b\phi_b^{'} \cdot \frac{1}{2} D_{a_{2D},2} D_{a_{3D},2}$
Proximity Triad	4	ϕ_c	$\phi_c^{'}$	$2\phi_c\phi_c^{'} \cdot \frac{1}{2} D_{a_{2D},2} D_{a_{3D},2}$
Parallel Triad	6	ϕ_d	$\phi_d^{'}$	$12\phi_d\phi_d^{'}$
Parallel Pair	4	ϕ_e	$\phi_e^{'}$	$4\phi_e\phi_e^{'} \cdot \frac{1}{2} D_{a_{2D},2} D_{a_{3D},2}$
Proximity Pair	3	ϕ_f	$\phi_f^{'}$	$2\phi_f\phi_f^{'} \cdot \frac{1}{3!} D_{a_{2D},3} D_{a_{3D},3}$



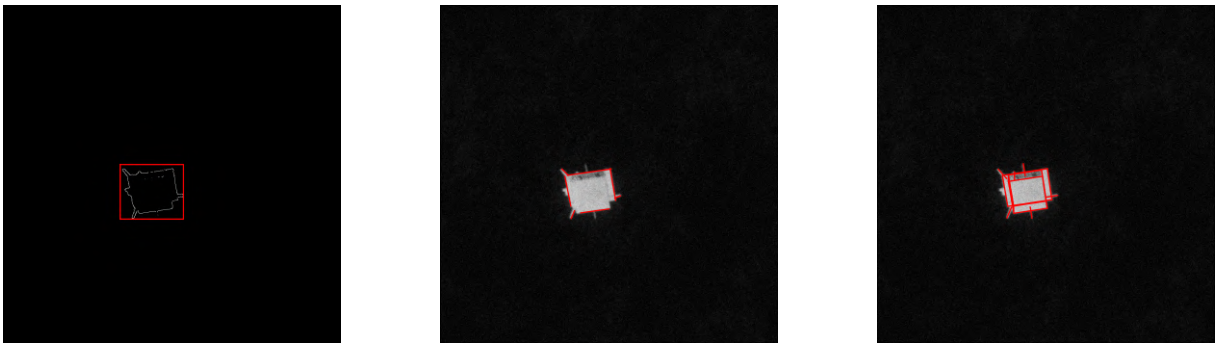
(a) Noised VIS image. (b) Noised TIR image. (c) Fused VIS-TIR image.

Fig. 11: Benchmark image used to apply the modified SVD algorithm.

each axis of the camera reference frame $\mathbf{E}_t = |\mathbf{r} - \hat{\mathbf{r}}|$, the error in the relative attitude $\mathbf{E}_R$ expressed as rotational error with respect to each axis of the camera reference frame, and as the "SLAB Score" [31], that is evaluated as:

$$e_{SLAB} = \frac{||\mathbf{r} - \hat{\mathbf{r}}||}{||\mathbf{r}||} + 2 \cdot arccos|\mathbf{q} \cdot \hat{\mathbf{q}}| \qquad [3]$$

In Fig. 12 are reported the ROI detected, the line



(a) ROI detected in VIS-TIR fused test image.

(b) Merged line segments detected in VIS-TIR fused test image.

(c) Reprojected wireframe model with estimated pose.

Fig. 12: Modified SVD performances on benchmark VIS-TIR fused image.

segments detected after merging the two streams of the SVD, and the wireframe model of Tango reprojected on the image by using the estimated pose. The results showed in Fig. 12 and reported in terms of errors in Tab. 5 are achieved by using the values given in Tab. 4.

Table 4: Modified SVD hyperparameters values.

k_1 (WGE & Sobel)	0.07
k_2 (WGE & Sobel)	0.0235
ρ_{res} (WGE & Sobel)	1 [pix]
θ_{res} (WGE & Sobel)	0.1 [deg]
τ (antenna threshold)	0.3

Table 5: Relative pose estimation errors.

$\mathbf{E}_t$ [m]	[0.004; 0.001; 0.047]
$\mathbf{E}_R$ [deg]	[0.72; 2.05; -0.36]
e_{SLAB} [-]	0.04401

From the results reported in Tab. 5 it is confirmed that fused VIS-TIR images can be exploited during proximity operations to provide high-quality images to be used for relative pose estimation. In particular, the errors in the estimation are extremely low, with a relative translational error well below 1% of the true relative distance between target and chaser, angular error $\leq$2 deg on all axes, and a SLAB score that, if confirmed on the entire SPEED dataset, could be in the top-3 scores ever.

To expand the results shown above, the same method has been applied to a small dataset of 20 fused VIS-TIR images. The SVD algorithm performed poorly on the dataset, giving an accurate pose estimation only for a single image. The merged line segments and the error in the pose estimation for that case are reported in Fig. 13 and Tab. 6 respectively, confirming the good edge detection performances of the SVD in nominal conditions in fused VIS-TIR images and the high accuracy of the pose estimation results.

Fig. 13: Merged line segments successfully detected in VIS-TIR fused image from 20 images dataset.

Table 6: Successful relative pose estimation errors on 20 images dataset.

$\mathbf{E}_t$ [m]	[0.001; 0.003; 0.038]
$\mathbf{E}_R$ [deg]	[2.01; 0.15; -0.20]
e_{SLAB} [-]	0.04200

Concerning the poor results on the dataset, it has been noticed that in most of the cases the impossibility of estimating a pose with high accuracy or even solving the EPnP problem is caused by a poor detection of line segments during the Hough transform. In

particular, concerning the antennas, even if they are clearly visible in the gradient-intensity image computed (as it can be noticed in Fig. 14) and even if the pose is quite close to the one in Fig.12, where the SVD converged to a high fidelity solution, they have not been detected.

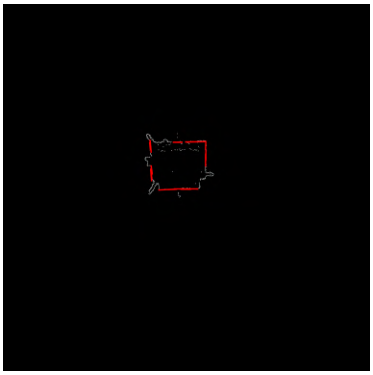


Fig. 14: Wrong edge detection in fused VIS-TIR image.

Despite the success rate of 5% for all the images in the small dataset considered here agrees with the success rate measured on average on the SPEED dataset in [39] and although also in the original paper the SVD algorithm provides accurate solutions only for few VIS images [24], other tests have been performed to exclude the possibility that the failure of the SVD in most of the cases is due to the fact that the inputs are fused VIS-TIR images. Hence, the SVD has been applied to the same image in Fig. 12 but obtained by considering a VIS image acquired in high illumination conditions (Fig. 15a), in low illumination conditions and fused by using the TSIFSD algorithm (Fig. 15b), and directly to a TIR image (Fig. 15c). The relative pose and the hyperparameters have been kept constant for all the tests and equal to the ones reported in Tab. 4.

Despite the line detection works well for the case of VIS-TIR fused images with ADF in low light conditions, by maintaining the same hyperparameters also for the other cases shown in Fig. 15, the line detection is completely jeopardized leading to antennas not detected and most of the edges not or partially detected. A bad tuning of the hyperparameters can be excluded because the intensity images on which the lines detected are signed are close to each other, but the performances are strongly different. Moreover, it can also be excluded the fact that fused images do not contain enough information to retrieve the relative pose because by looking at the images in Fig. 15 it can be noticed once again that the antennas and most of the edges are clearly detectable, but they have not been identified by the Hough transform.

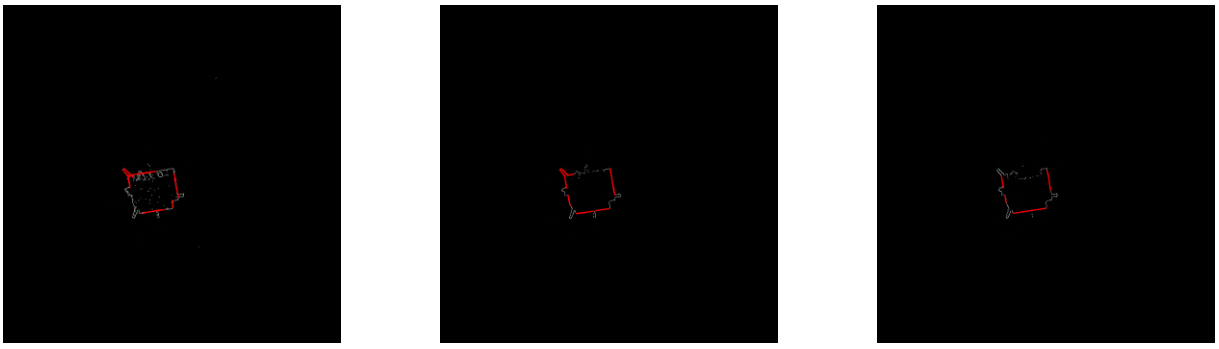
As a consequence, from all the tests reported above, it can be concluded that despite the accuracy that can be achieved in nominal conditions, due to the sensitivity of the Hough transform to the input image but also the hyperparameters, the Hough transform is not suited to be adopted in a wide range of scenarios as the ones that can be faced during in-orbit relative navigation, since most of them will fall in the off-nominal conditions for the Hough transform, where the SVD algorithm fails due to wrong or poor edge detection, even if the hyperparameters are linearly scaled with the ROI diagonal length, hence with respect to the relative distance between the target and the chaser. Despite that, it is still remarked that, by looking also at the images in Fig.15, both ADF and TSIFSD methods lead to VIS-TIR fused images of high quality that can be safely adopted as the primary input for relative navigation algorithms.

6. Conclusions and Future Works

6.1 *Conclusions*

The paper here presented aims to provide evidence that fused VIS-TIR images can be safely adopted as the primary source of measurements in relative GNC algorithms. VIS and TIR images have been generated by using a dedicated tool and then have been fused at pixel level by using both ADF and TSIFSD methods. The fused images obtained have been fed to a pose initialization algorithm. For this purpose the SVD algorithm has been adopted since it is a feature-based method, hence it does not require huge image datasets to be trained as the deep learning-based methods.

The first outcome of the work presented is achieved in the re-implementation of the SVD algorithm. Some minor modifications have been introduced, such as the usage of cropped images fed to the line segment detection in the two parallel streams to improve the detection performances. Moreover, the most substantial improvement to the baseline is given by the clear definition of geometrical constraints applied when each perceptual group is detected and stored. These newly introduced constraints allow storing each feature group as a list in an ordered way, reducing the possible combinations between 2D and 3D corresponding perceptual groups. From this, it has been possible to derive a mathematical formulation that can be applied in a general case to forecast



(a) Lines detected in ADF VIS-TIR image in high light conditions.

(b) Lines detected in TSIFSD VIS-TIR image in low light conditions.

(c) Lines detected in TIR noised upscaled image.

Fig. 15: Results of line segment detection on the same image obtained with different methods.

a priori the dimensions of the match matrix.

By performing the pose estimation via SVD on a small dataset it has been assessed that fused VIS-TIR images can be safely adopted as the primary source of measurement since their content of information is higher than the single VIS or TIR images, avoiding the problems that characterize both the spectral bands. Concerning the pose estimation results, if the SVD converges, the estimation errors are extremely low, meaning that also a high degree of accuracy is preserved by using images fused at the pixel level. Notice that the best results have been achieved in low illumination conditions, where the target is almost not visible in the VIS image and hence where the vast majority of the pose initialization algorithms that leverage on VIS images would have failed in retrieving the correct relative pose. Despite that, it is worth highlighting that it has been confirmed also by the tests performed that the Hough transform does not guarantee high accuracy in line detection for a wide range of scenarios without re-tuning the hyperparameters, hence other methods should be identified to correctly detect complex features in a real case scenario.

6.2 *Future works*

The outcomes of the presented work resumed above paved the way for new developments in the relative navigation and image generation fields. Concerning the relative navigation, new algorithms with improved performances and more robustness than the SVD should be tested with fused VIS-TIR images to properly assess the possibilities and the limits of using multispectral images as the primary input in real case scenarios, leading also to the necessity of generating a wide dataset of fused images to test also deep learning-based algorithms.

Concerning image generation, in this work, it has been assumed that the VIS and TIR cameras are aligned on the same axis, with equal focal lengths but different resolutions. In a real case scenario, this assumption does not hold and the images must be also registered before being fused. This process can introduce artifacts in the image that could eventually affect both the feature detection and the pose estimation performances, hence this aspect should be taken into account for further analysis. For what concerns the TIR images generation, the most important future work is to validate the tool. Unfortunately, there are no datasets of spaceborne thermal images of artificial objects thus one way to validate the tool is to use the telemetry data. Nonetheless, the TIR image rendering tool used in the article is highly flexible and can be used for celestial objects such as asteroids for which the validation should be possible thanks to images from past missions such as Hayabusa.

References

[1] David K. Geller. Orbital rendezvous: When is autonomy required? *Journal of Guidance, Control, and Dynamics*, 30(4):974–981, 2007.

[2] Massimo Piazza, Michele Maestrini, and Pierluigi Di Lizia. Monocular relative pose estimation pipeline for uncooperative resident space objects. *Journal of Aerospace Information Systems*, 0(0):1–20, 2021.

[3] Roberto Opromolla, Giancarmine Fasano, Giancarlo Rufino, and Michele Grassi. A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity op-

erations. *Progress in Aerospace Sciences*, 93:53–72, 2017.

[4] 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.

[5] Francesco Castellini, David Antal-Wokes, Ramon Pardo de Santayana, Klaas, and Vantournhout. Far approach optical navigation and comet photometry for the rosetta mission. *Proceedings of 25th International Symposium on Space Flight Dynamics, 25th ISSFD*, 2015.

[6] Gaia Letizia Civardi, Michele Bechini, Alessandro Colombo, Matteo Quirino, Margherita Piccinin, and Michéle 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), Milan, Italy*, 6 2022.

[7] 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.

[8] Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. Speed+: Next-generation dataset for spacecraft pose estimation across domain gap. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–15. IEEE, 2022.

[9] Wigbert Fehse. Rendezvous with and capture / removal of non-cooperative bodies in orbit: The technical challenges. *Journal of Space Safety Engineering*, 1(1):17–27, 2014.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International con-*

ference on computer vision, pages 2564–2571. IEEE, 2011.

[21] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[22] Simone D'Amico, Mathias Benn, and John L. Jørgensen. Pose estimation of an uncooperative spacecraft from actual space imagery. *International Journal of Space Science and Engineering*, 2(2):171–189, 2014.

[23] Simone D'Amico, Per Bodin, M. Delpech, and Ron Noteborn. Prisma. In Marco D'Errico, editor, *Distributed space missions for earth system monitoring*, chapter 21, page 599–637. Springer Science & Business Media, 2013.

[24] Sumant Sharma, Jacopo Ventura, and Simone D'Amico. Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous. *Journal of Spacecraft and Rockets*, 55(6):1414–1429, 2018.

[25] Vincenzo Capuano, Shahrouz Ryan Alimo, Andrew Q. Ho, and Soon-Jo Chung. *Robust Features Extraction for On-board Monocular-based Spacecraft Pose Acquisition.* 2019.

[26] Pedro F. Proença and Yang Gao. Deep learning for spacecraft pose estimation from photorealistic rendering. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6007–6013. IEEE, 2020.

[27] Tae Ha Park and Simone D'Amico. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. In *11th International Workshop on Satellite Constellations & Formation Flying (IWSCFF), Milan, Italy*, 6 2022.

[28] Sumant Sharma, Connor Beierle, and Simone D'Amico. Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks. In *2018 IEEE Aerospace Conference*, pages 1–12, 2018.

[29] Sumant Sharma and Simone D'Amico. Neural network-based pose estimation for noncooperative spacecraft rendezvous. *IEEE Transactions on Aerospace and Electronic Systems*, 56(6):4638–4658, 2020.

[30] Stefano Silvestrini, Margherita Piccinin, Giovanni Zanotti, Andrea Brandonisio, Ilaria Bloise, Lorenzo Feruglio, Paolo Lunghi, Michèle Lavagna, and Mattia Varile. Optical navigation for lunar landing based on convolutional neural network crater detector. *Aerospace Science and Technology*, 123:107503, 03 2022.

[31] Mate Kisantal, Sumant Sharma, Tae Ha Park, Dario Izzo, Marcus Märtens, and Simone D'Amico. Satellite pose estimation challenge: Dataset, competition design, and results. *IEEE Transactions on Aerospace and Electronic Systems*, 56(5):4083–4098, 2020.

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

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

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

[35] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[36] Tomas Plachetka. Pov ray: persistence of vision parallel raytracer. In *Proc. of Spring Conf. on Computer Graphics, Budmerice, Slovakia*, volume 123, page 129, 1998.

[37] Francescodario Cuzzocrea. Analysis and validation of spaceborne synthetic imagery using a vision-based pose initialization algorithm for non-cooperative spacecrafts. *MSc Thesis. Politecnico di Milano*, 2020.

[38] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, jan 1972.

[39] Pierdomenico Fracchiolla. Analysis and validation of a vision-based pose initialization algorithm for non-cooperative spacecrafts. *MSc Thesis. Università degli Studi di Padova*, 2019.