Analyzing Target-, Handcrafted- and Learning-Based Methods for Automated 3D Measurement and Modelling

Giulio Perda [1], Luca Morelli [1,2], Fabio Remondino [1], Clive Fraser [3], Thomas Luhmann [4]

[1] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: (gperda, lmorelli, remondino)@fbk.eu
[2] Dept. of Civil, Environmental and Mechanical Engineering (DICAM), University of Trento, Italy
[3] University of Melbourne, Australia - Email: c.fraser@unimelb.edu.au
[4] Institute for Applied Photogrammetry and Geoinformatics, Jade University of Applied Sciences, Germany
Email: luhmann@jade-hs.de

Technical Commission II

KEYWORDS: target, bundle adjustment, deep learning, 3D reconstruction

ABSTRACT
In industrial vision metrology, precise spatial measurement is vital for quality control and complex manufacturing, traditionally relying on target arrays for sub-pixel accuracy (0.05-0.1 pixels) and precision to beyond 1:200,000. However, target design, placement and measurement are often time-consuming and challenging for large-scale projects. Automated, markerless methods, generally called Structure-from-Motion (SfM), based on handcrafted algorithms or deep learning-based pipelines, offer greater flexibility but are not widely adopted due not only to concerns about reliability and precision, but also because in many industrial photogrammetry applications targets highlight particular feature points of interest, e.g. tooling points, holes and edges. This study reviews the differences between target-, handcrafted- and learning-based approaches, explores hybrid methods combining targets and natural features, and tests learning-based or handcrafted approaches against the traditional target-based method. Two end-to-end learning-based pipelines based on SuperPoint+LightGlue and KeyNet+AffNet+HardNet are evaluated. Results show that deep learning pipelines for tie point extraction provide enhanced automation but inferior triangulation precision, while being comparable to handcrafted methods.

1. INTRODUCTION

Image matching plays a pivotal role in conventional photogrammetric projects as well as in Structure-from-Motion (SfM), Visual Odometry (VO) and Simultaneous Localization And Mapping (SLAM) applications.
The utilisation of targets is near universal in high-precision industrial and engineering photogrammetry and camera calibration processes (Fraser, 1997; Remondino and Fraser, 2006): targets have a potential precision in image point measurement of 0.05-0.1 px and can yield measurement precision (RMS 1-sigma) in object space in the range of 1:100,000 to beyond 1:200,000 (Luhmann, 2010; Luhmann et al., 2016).
For some years now, the adoption of automated, targetless calibration, orientation and 3D object reconstruction methods (Barazzetti et al., 2011; Remondino et al., 2017) have shown flexibility and accuracy potential, becoming increasingly more widespread in particular for architectural and archaeological surveys as well as for drone-based photogrammetry. These methods employ handcrafted operators, such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004), Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011), Speeded-Up Robust Features (SURF) (Bay et al., 2006) or, more recently, deep learning-based approaches (Figure1). Learning methods for tie point extraction are specifically trained for challenging illumination, context, scenario variations and wide viewing angles and they have even more democratized automated image matching for image retrieval or 3D reconstruction purposes (Jin et al., 2021; Ruiz et al., 2023; Morelli et al., 2024a).
With these powerful alternatives available, two pressing questions arise: why have fully automated image triangulation methods, typically referred to as Structure-from-Motion (SfM), based on automated natural feature detection, not been more widely adopted in large-scale (industrial) vision metrology projects? And are physical targets, properly distributed in the scene, still the most effective way to achieve high-accuracy results?

1.1 Aim of the work

This work first reviews the principal distinctions between target-based, handcrafted and learning-based approaches for photogrammetric triangulation purposes, with the emphasis being upon measurement accuracy, reliability and practicability. The potential for an integrated approach employing both natural features and targets is then discussed, with project examples of where such an integrated network orientation and point determination approach could be optimal in terms of both accuracy and productivity. Finally, handcrafted or learning-based approaches are tested alone and compared to the solution employing targets to understand whether they can potentially substitute target measurement in metrology applications.
It is acknowledged that in many vision-based large-scale metrology applications, markers are necessary to signalise well-defined features (e.g. corners, edges, drill holes, tool inspection or adjustment points) through the use of point or adapter targets, or target clusters. This work considers more the case where targets are utilised as tie points in order to facilitate network exterior orientation; the focus is not upon potential reduction or replacement of targets that represent important physical features or properties.

2. RELATED WORK

In the domain of industrial metrology, feature matching is a fundamental technique for photogrammetric applications such as quality inspection, precise 3D coordinate measurement, and surface analysis, where accurate placement and automatic measurement of targets plays a key role. Both handcrafted and learning-based approaches have been applied to the task of feature correspondence determination, with each offering distinct advantages and limitations, particularly in terms of robustness, accuracy and processing efficiency.

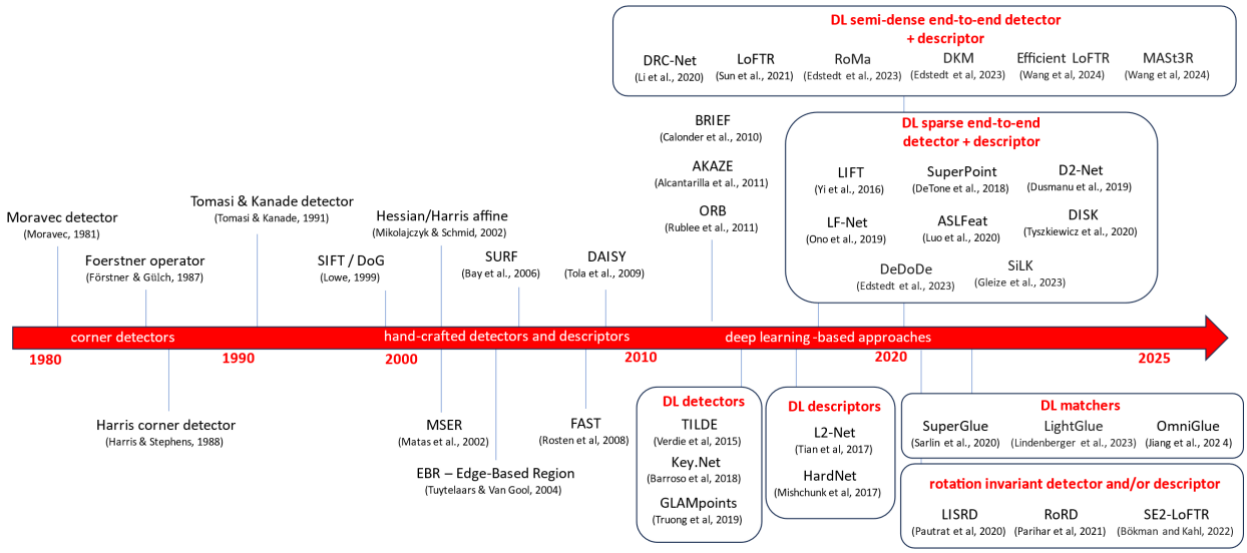

Figure 1: Some milestone methods within the evolution of image matching for tie point extraction

2.1 Handcrafted Feature Matching

Handcrafted feature matching methods have been widely adopted due to their reliability, interpretability and robustness under controlled conditions (Ma et al., 2021). Tie points are typically identified by combining hand-crafted detectors and descriptors with feature matching techniques. Most popular feature detectors include corner (Harris et al., 1988; Rosten and Drummond 2006; Rublee et al. 2011) and blob detectors (Lowe, 2004; Bay et al. 2006; Yiet al. 2016). For accurate pairing, detected and described keypoints must exhibit high repeatability, strong discriminative power, geometric invariance, low sensitivity to changes in scene brightness, and sparsity to minimize memory usage (Remondino et al., 2021).

Techniques such as SIFT (Lowe, 2004) have been adopted in applications requiring precise measurement and alignment, such as automated target placement and 3D coordinate measurement for manufacturing and inspection processes (Gonzales et al., 2015; Loch et al., 2013). SIFT's robustness to scale, rotation, and illumination changes makes it ideal for setups where targets need to be placed and measured on objects of varying shapes and sizes. To address the need for faster computation in real-time applications, SURF (Speeded-Up Robust Features) (Bay et al., 2006) was proposed to maintain robustness while allowing quicker processing. SURF is thus suited to high-speed inspections that require accurate placement and matching of fiducial targets, as seen in automated assembly lines or in-line quality control, where parts must be quickly verified without interrupting the production flow (Ren et al., 2022).

For highly efficient target matching in real-time environments, ORB (Oriented FAST and Rotated BRIEF) (Rublee et al., 2011) offers a balance of speed and precision by using binary descriptors. ORB is also the key method in various SLAM-based approaches (Mur-Artal et al., 2015; Campos et al., 2021). Also, AKAZE (Accelerated-KAZE) (Alcantarilla et al., 2013) has been proposed for applications involving textured surfaces, since it offers fast, reliable feature matching to support precise target localization in texture-rich metrology tasks (Zheng et al., 2022).

2.2 Learning-based Feature Matching

In recent years, learning-based feature matching methods, especially those leveraging convolutional neural networks (CNNs), have introduced new capabilities for handling more complex and variable imaging conditions. These models, available in the DIM toolbox[1] (Morelli et al., 2024a), can automatically learn feature representations optimized for specific measurement tasks, making them highly suitable for applications requiring precise feature detection and measurement in environments where lighting, perspective, or surface texture may vary. CNN-based methods for keypoint detection and description can either separate or merge the two tasks. Yi et al. (2016), Ono et al. (2019) and Revaud et al. (2019) have suggested that training them together leads to more reliable keypoints for matching. Loss functions, including pairwise and triplet losses, are used in applications such as image retrieval and 3D reconstruction. Keypoint detection methods include TILDE (Verdie et al., 2015), Quad-Net (Savinov et al., 2017), and Key.Net (Barroso et al., 2019), which aim for repeatable points for matching but may face challenges with reliability. End-to-end methods such as LIFT (Yi et al., 2016), SuperPoint (DeTone et al., 2018), and D2-Net (Dusmanu et al., 2019) train both detection and description together to enhance matching accuracy and reliability. Additionally, detect-and-describe methods such as ASLFeat (Luo et al., 2020) and R2D2 (Revaud et al., 2019) fully integrate both steps into one network. More recent developments in semi-dense end-to-end pipelines, such as Efficient LoFTR (Wang et al., 2024) and MASt3R (Leroy al., 2024) have shown considerable potential in providing complete and accurate 3D reconstructions at even higher processing speeds.

Learning-based methods have shown good performances for well designed networks (Remondino et al., 2021) and have been utilized for applications outside vision metrology, such as matching of historical images (Zhang et al., 2021; Maiwald et al., 2021), vision-based navigation (Morelli et al., 2023) and terrestrial 3D photogrammetric reconstruction (Markiewicz et al, 2021). Critical scenarios with wide-baselines (Jin et al., 2020; Bellavia et al, 2022) or crowdsourced data (Morelli et al., 2024b) have also attracted utilization of learning-based pipelines.

3. METHODOLOGY

Here, traditional target array positioning and measurement is compared to both handcrafted and learning-based feature extraction and matching pipelines. Tie points are extracted by means of three methodologies:

[1] https://github.com/3DOM-FBK/deep-image-matching

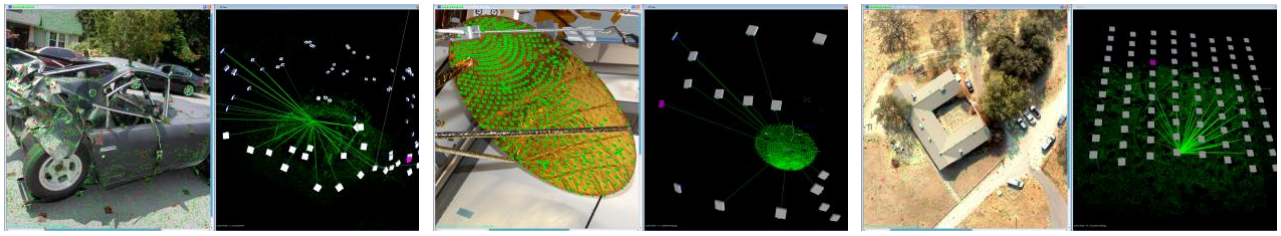

Figure 2: Image samples and camera networks for the considered datasets: car (right), antenna (centre) and UAV (right).

- *targets*: in terrestrial datasets, target centroids are automatically detected in the images at sub-pixel accuracy by intensity-weighted centroiding. Often, targets which fail to be identified but are nevertheless deemed good for visual centroiding are manually measured. For the UAV dataset, GNSS target positions were acquired by means of Real-Time Kinematic (RTK) positioning with centimeter accuracy. They were manually measured in the images, the emphasis being on direct measurement of 3D coordinates of the targets and direct scaling of the reconstructed model.
- *handcrafted*: SIFT-like methods are used to extract natural features in the images with the classical detector/descriptor approach. To further strengthen the image network, target measurements are also coupled with handcrafted methods in a combined approach, with the emphasis being on both direct measurements of 3D coordinates of points of interest and full 3D scene reconstruction.
- *learning-based*: automatic feature extraction and matching is carried out with learning-based end-to-end pipelines; we used a rotation-dependent pipeline with SuperPoint (DeTone et al., 2018) and LightGlue (Lindenberger et al., 2023) - with a priori upright rotation (SP+LG upright) and a rotation-invariant approach using KeyNet (Barroso et al., 2018) coupled with AffNet (Mishkin et al., 2018) and HardNet (Mishchunk et al., 2017) - available in the Kornia library (Riba et al., 2020). No targets are measured. It is worth noting that these learning-based methods work at pixel level, therefore they lack sub-pixel accuracy for keypoint detection (Morelli et al., 2024a). The emphasis is therefore on model building and indirect feature/tie point measurement.

The first two approaches are executed entirely inside Australis (2024), while learning-based features are extracted and matched externally by means of the DIM toolbox (Morelli et al., 2024a) and their image coordinates are then imported into Australis for orientation and bundle adjustment. A maximum of 20,000 keypoints per image are extracted and a self-calibrating bundle adjustment is run. A minimum of 4 intersection rays and a 3 deg intersection angle per point are set in the bundle adjustment.

Name	*# img*	*Type*	*Target type*	*# targets*	*Image rotation*
Car	89	convergent	Coded	232	90-270
Antenna	17	convergent	Single dot	677	0-90
UAV	90	near nadir	GNSS-measured	65	0-180

Table 1. Overview of the tested datasets.

3.1 Evaluation datasets

We conducted our research on one near-nadir aerial, and two convergent, non-aerial datasets, one being ground based and the other being on the International Space Station, as shown in Figure 2 and summarized in Table 1.

The **Car dataset** contains 89 images of a crashed car. Red on black targets were placed inside and onto the car. The challenge with this dataset is presented by the 90 deg to 270 deg camera rotations, which make rotation-invariant feature matching methods necessary for all images to be oriented with good redundancy. Four images of the interior of the car share a limited overlap with the exterior.

The **Antenna dataset** features a 60cm-diameter inflatable antenna measured on the International Space Station. Some 670 white dot targets were placed over its entire surface, at a separation of 1 to 4 cm. Four targets located on a plate ca 40 cm from the surface defined the coordinate system. Images were acquired to accurately track surface deformation (up to 0.2 mm) over time. The network comprised images recorded from one side only of the centre of the face of the antenna, making the geometry unbalanced. Moreover, the surface of the antenna was highly reflective, making it unsuitable for automatic feature matching, due to the ambiguity stemming from target similarity. Some targets were even difficult to measure manually with sufficient accuracy, so they deemed non usable. The extraction of points in the background of the scene could help in strengthening the network. This dataset is used for more in-depth analysis on the influence of bundle adjustment parameters on result metrics.

The **UAV dataset** comprised 8 parallel nadir image strips over a somewhat flat and vegetated landscape. Rotation between consecutive strips was 180 deg. The scale within the images was very similar and no orthogonal camera rolls were present. The dataset included 65 GNSS-surveyed ground targets. Five of these, arranged in a cross configuration with one at the centre of the scene, were used as bundle adjustment constraints (GCPs), with a priori sigma of 0.005 m in all axes. The remaining 60, which served as checkpoints, were loosely weighted at 0.05m. GNSS-measured camera station positions were assigned a priori sigma values of 0.025m.

Evaluation metrics included the number of oriented images, the number of 3D points referenced, the RMS value of the reprojection error for all image points (RMS Vxy), the estimated accuracy of 3D point coordinates (RMS 1-sigma), the minimum number of points on an image, the minimum intersection angle for all image pairs which see a point (for all object points in the dataset), the average intersection angle, and the maximum and average number of intersecting rays per point for all points of the dataset The intersection angle metrics provide an indication of the robustness of reconstruction, with a lower intersection angle for a point indicating a less favourable geometry. The maximum and average rays per point provides a measure of internal reliability, a higher average value indicating a better ability to detect xy image coordinate observation blunders. Finally, due to the unavailability of real word coordinates for targets, we calculated the RMSE as the residual of a 7-parameter shape-invariant transformation on the coordinates estimated by the targets-only solution. For the UAV dataset, instead, the object point RMSE values were computed using the available 60 checkpoints.

4. RESULTS AND DISCUSSION

For all three datasets, metrics are given in Table 2. Image correspondences based on targets show a higher centroiding accuracy and reliability, resulting in lower reprojection errors (RMS Vxy) compared to the other approaches, which translates

to a higher precision in object space. Evidence of this can be seen by looking at corresponding features around target centroids for the targets and SP+LG solutions (Figure 3). Typically, an automatic feature extractor does not guarantee that it will consistently identify the centroid of a target. As a result, these automatic methods may yield inconsistent correspondences for the same image point. Additionally, the chosen learning-based extractors are limited from the start as they operate at the pixel-level and lack sub-pixel precision. These two factors significantly increase the mean reprojection error. The RMS Vxy values vary significantly from point to point, depending mainly on the multiplicity of intersecting rays and the intersection angle. A higher number of points on an image of measured natural features did not directly translate to a more geometrically robust network.

		# oriented imgs	# referenced points	RMS Vxy (px)	RMS 1-sigma	Min. # points on image	Min. int. angle (deg)	Avg. int. angle (deg)	Max. rays intersection	Avg. rays intersection	RMSE (mm)
Car	Targets	89/89	627	**0.24**	**1:29300**	30	**6**	**76**	**55**	**13**	-
	Handcrafted	85/89	16258	0.27	1:21600	**721**	5	23	14	5	1.99
	Combined	89/89	28690	0.44	1:3600	77	5	35	**55**	6	**1.18**
	SP+LG	86/89	24467	0.81	1:4200	44	3	57	52	8	1.72
	SP+LG upright	89/89	**44603**	0.81	1:9800	**95**	3	48	53	8	3.32
	KN+AN+HN	89/89	41181	0.81	1:7200	31	3	44	29	6	2.88
Antenna	Targets	17/17	677	**0.31**	**1:14800**	43	**22**	**63**	**14**	**8**	-
	Handcrafted	17/17	1359	0.68	1:5200	133	10	34	**14**	6	0.84
	Combined	17/17	2224	0.43	1:8400	**681**	11	44	**14**	**8**	**0.09**
	SP+LG	16/17	5370	0.99	1:3900	239	9	36	11	6	0.91
	SP+LG upright	17/17	**5782**	1.05	1:4800	229	9	36	13	6	1.01
	KN+AN+HN	17/17	2274	1.16	1:3200	135	10	31	12	5	0.80
											RMSE (m)
UAV	Targets	85/90	65	**0.27**	**1:76800**	9	24	59	**29**	**15**	**0.020**
	Handcrafted	90/90	36158	0.37	1:30900	4985	15	40	23	8	0.286
	Combined	90/90	**47287**	0.38	1:37800	**5110**	18	37	**29**	8	0.022
	SP+LG*	45+45/90	6908	1.05	1:5700	872	**33**	54	17	6	0.215
	SP+LG upright	89/90	19463	1.15	1:7200	909	18	**62**	25	8	0.516
	KN+AN+HN	90/90	54259	0.94	1:6600	3632	19	47	27	7	0.891

Table 2. Metrics over the tested datasets. *The reconstruction of UAV SP+LG is split into two 45-images blocks because of 180° rotations in consecutive strips. Only the metrics for the first block are reported.

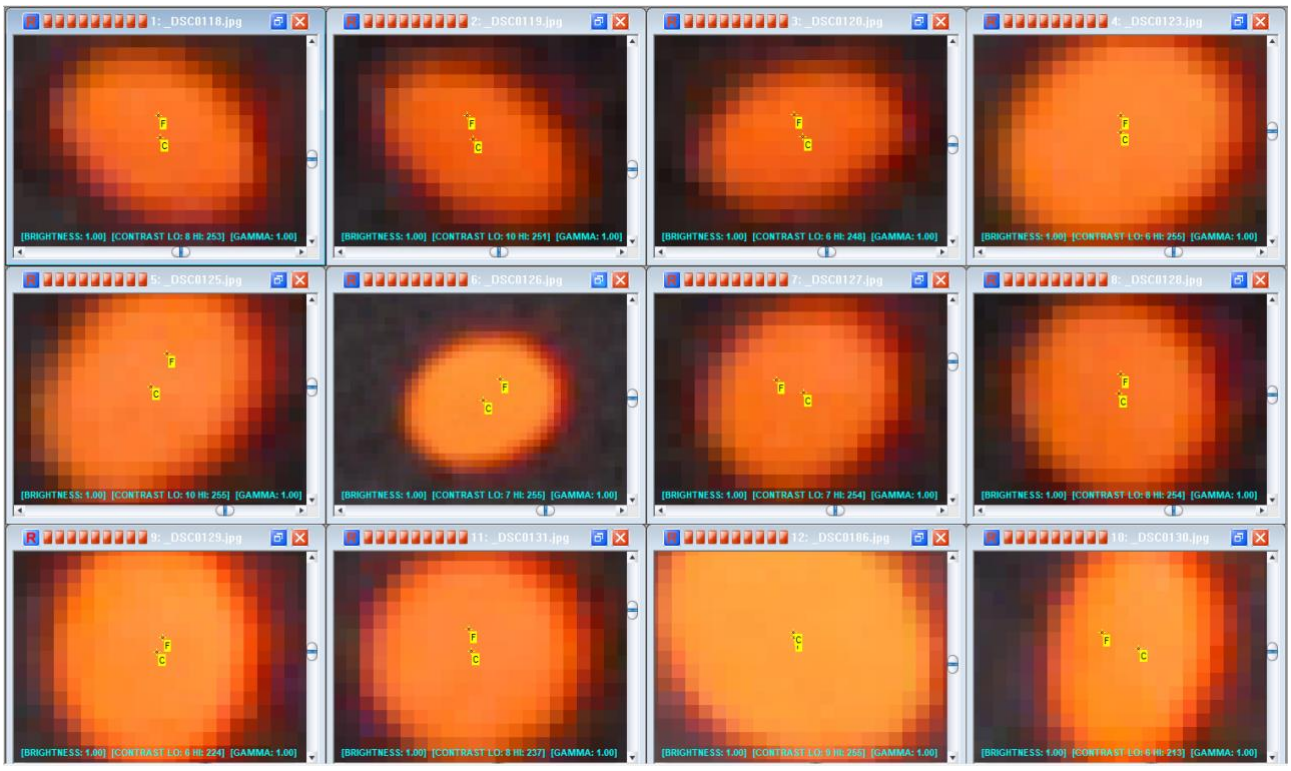

Figure 3. Automatic centroid C and learning-based tie point F detected over a target with SP+LG in the Car dataset. SP+LG does not recover the feature in a consistent manner. RMS Vxy for this point is 1.1px compared to 0.23px of the targets solution.

Many of these points have low intersection angles and fewer intersecting rays per point compared to the solution based on targets. As shown in Figure 4, the ray intersection distributions for learning-based methods clearly indicate that a more balanced distribution is achieved in the solution using targets. Nevertheless, learning-based methods tended to score higher intersecting rays per point compared to the handcrafted method, demonstrating the superior ability of recognizing the same point on images with different conditions and under larger perspective distortions.

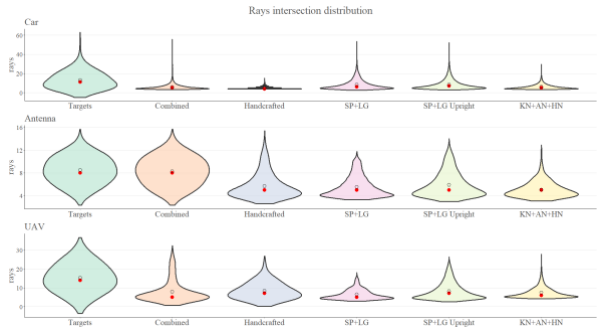

Figure 4: Distribution of intersecting rays per point, for the three datasets. A hollow point represents the median, a red point the average.

Car dataset.
The handcrafted method achieves a similar RMS Vxy to that of the solution using targets; it performs better in this regard than learning-based approaches. On the other hand, the maximum and average intersecting rays per point values are significantly lower than for all other methods. This could be explained by the fact that most features are extracted and matched on the ground, instead of on the car itself, this being visible in Figure 5. Apparently, natural features on the car are deemed too similar for matching. As the camera moves around the car, these points become progressively hidden and so are seen only on a limited subset of images. Learning-based methods, on the other hand, manage to extract points which are visible from more viewpoints, hence scoring more intersections. In object space the handcrafted method performs slightly better than the learning-based alternatives. Surprisingly, despite the similarity in RMS Vxy values, number of rays and intersection angle distributions, SP+LG Upright achieves close to double the RMSE of SP+LG. The addition of ~20000 extra correspondences and three oriented images likely has a negative effect on camera calibration. Resulting dense clouds between all methods show no significant differences.

Antenna dataset.
A large difference in the number of referenced points between learning-based and handcrafted methods is visible. Figure 6 indeed confirms how both handcrafted and learning-based methods cannot match extracted features on the difficult surface of the antenna, likely because of their high similarity and ambiguity. This results in fewer oriented images and a 3 to 4 times higher RMS reprojection error when compared to the solution with targets. In theory, a handcrafted solution offers the possibility of strengthening the network by referencing points beyond the object of interest, i.e. on background surfaces. In practice, these points further increase the average reprojection error because of their similar appearance. The influence of the minimum intersection angle was investigated. A bundle adjustment was run - with the learning-based methods providing half the average intersection angle of the solution based on targets - by increasing the minimum intersection angle between points in 9 deg steps (minimum intersection angle for all methods). Figure 7 shows that the number of oriented images stays constant for the targets and handcrafted pipelines, whereas the number decreases for learning-based approaches. Despite a somewhat similar intersection angle distribution and average value to that of the handcrafted method, the number of oriented images drops when learning-based points are restricted with a value higher than their average (>36). This information, combined with their lower number of points with respect to learning-based approaches, highlights the higher geometric quality of target correspondences and handcrafted features. As expected, by excluding geometrically unfavorable points, the RMS Vxy value decreases slightly for all methods. Restricting the bundle adjustment to points with a higher number of intersecting rays achieves similar RMS Vxy values, however for the learning-based methods, less oriented images result, as reported in Table 3. Although highly redundant, the quality of these learning-based features is not sufficient to orient all images of the network, likely because of their similarity. Learning-based and handcrafted methods yield 10 times higher RMSE compared to the solution based on targets, with comparable values.

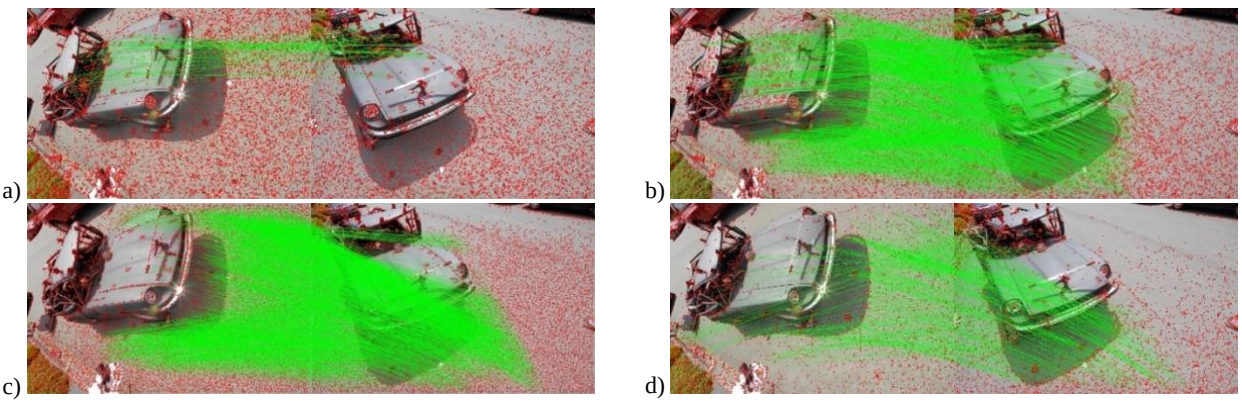

Figure 5. Learning-based, rotation-dependent method SuperPoint+LightGlue (a) struggles finding image correspondences in the case of rotated images, leading to very few correspondences and non-oriented cameras. Manually forcing the rotation (b) or using rotation-invariant descriptors-matchers like KeyNet + AffNet + HardNet (c) and SIFT-like (d) results in more matches, although few are on the object of interest.

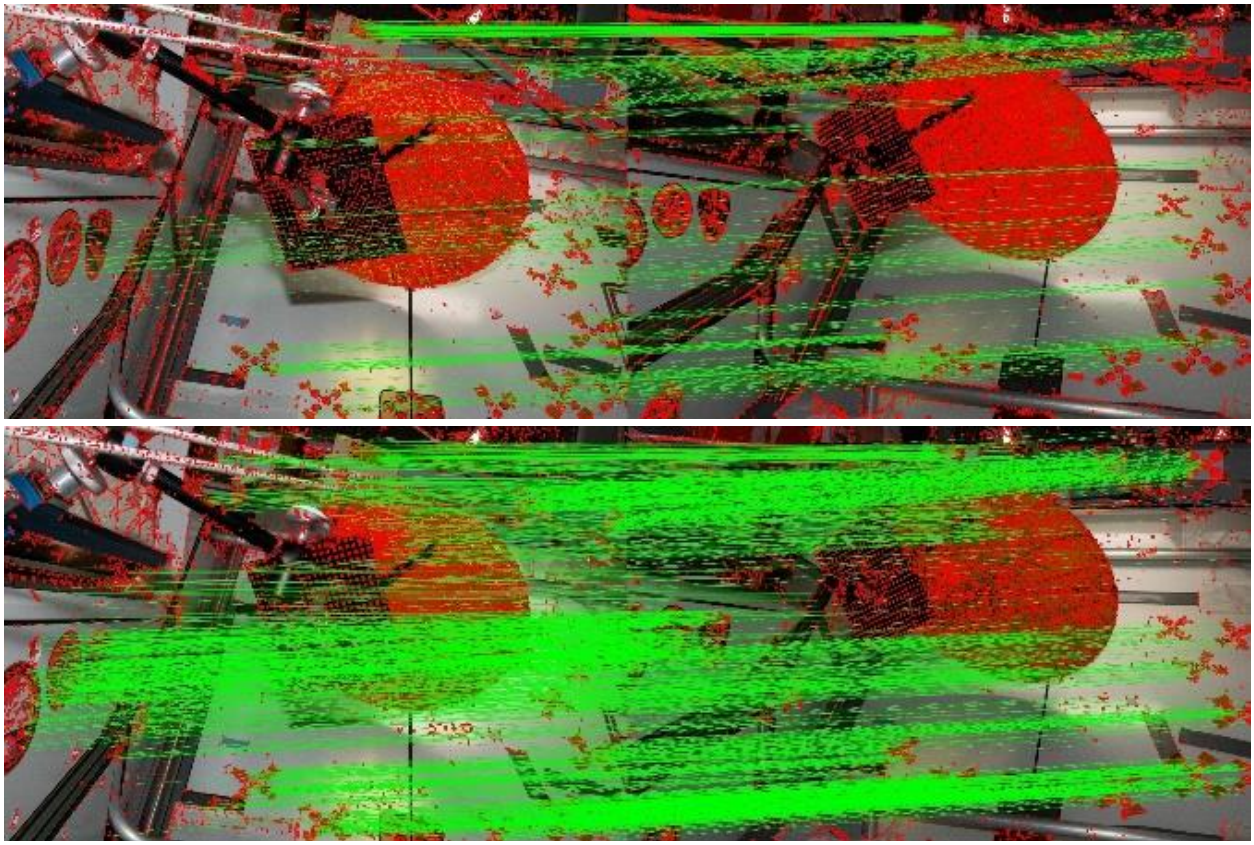

Figure 6. Feature matching for two pairs in the antenna dataset. Although many features are extracted, handcrafted (top) and SP+LG (bottom) find few correspondences over the reflective surface of the antenna, but more on the distinctive background.

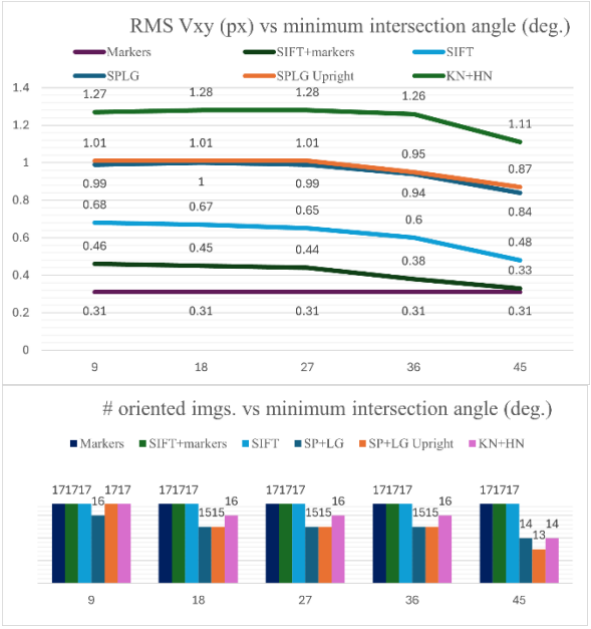

Figure 7: RMSE Vxy (top) and number of oriented images vs minimum intersection angle (bottom) for the Antenna dataset.

Applications requiring precise surface computation should still depend on careful placement and measurement of target arrays. However, learning-based methods can provide a useful starting point by offering an initial estimate of network relative orientation and camera calibration. This can streamline subsequent manual measurements, potentially allowing for the removal of all initial handcrafted features.

	# oriented images		# triangulated points		RMSE Vxy	
	≥ 4 rays	≥ 6 rays	≥ 4 rays	≥ 6 rays	≥ 4 rays	≥ 6 rays
Targets	17/17	17/17	677	600	0.31	0.31
Handcrafted	17/17	17/17	2224	516	0.43	0.7
Combined	17/17	17/17	1359	1138	0.68	0.38
SP+LG	16/17	11/17	5370	2071	0.99	1.04
SP+LG upright	17/17	13/17	5780	2609	1.01	1.07
KN+AN+HN	17/17	16/17	849	238	1.27	1.3

Table 3: Point redundancy effects in the bundle adjustment for the Antenna dataset. Learning-based methods oriented less images when a tighter threshold is set. The RMSE Vxy can be lowered by increasing the number of observations.

UAV dataset.

The utilization of targets shows its clear advantage in the aerial dataset: residuals against checkpoints are one order of magnitude lower than for methods which do not use targets. Handcrafted and learning-based methods do not produce acceptable results and would require tighter a priori weighting of GCPs. Despite this, the handcrafted and rotation-invariant learning-based methods close the gap left by the targets-only solution, orienting five more images and providing a more complete although less geometrically accurate 3D reconstruction. The non-rotation invariant SP+LG pipeline, though capable of orienting only half of the images, achieves smaller object space residuals, most likely because noisy correspondences between consecutive strips are not utilized. Comparing the focal length estimates from the UAV network solution using targets with the self-calibration estimate from other methods, we find biases of 0.023, 0.018, 0.029 and 0.042mm for handcrafted, SP+LG, SP+LG upright and KNHN, respectively. These translate to object-space shape

distortions with errors up to 2m in the vertical direction, an example being given in Figure 8 for SP+LG Upright.
Once again, weak geometry highlighted by a low number of intersecting rays per point, constant scale within images and the absence of camera orthogonal rolls - and changes in adjusted mean image coordinate scale - are likely the reason behind these mismatches. GCP residuals can be arbitrarily reduced by tighter a priori weighting although this can result in this instance in even more significant variations in interior orientation parameter estimates.

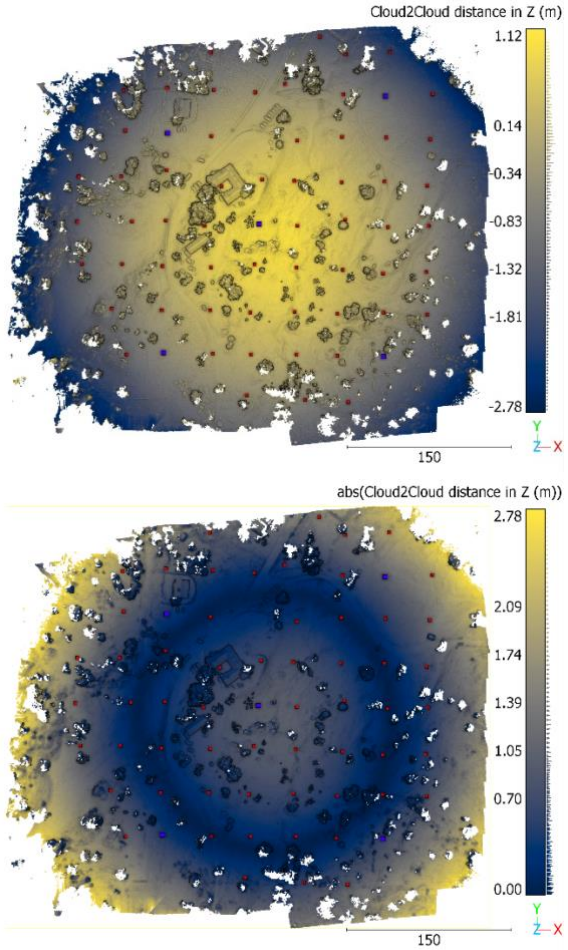

Figure 8. Signed (top) and absolute (bottom) height (Z) discrepancies for the dense point clouds generated in the UAV network from SP+LG Upright and targets approaches. GCPs are shown in blue and checkpoints in red. Shape distortion induced by wrongly estimated focal lengths is more noticeable at the periphery where constraints are less tight.

5. CONCLUSIONS

An analysis of photogrammetric image orientation and 3D reconstruction methods based on targets, conventional operators (handcrafted) and learning-based approaches has been presented. Traditional use of targets was found – unsurprisingly – to result in superior image measurement accuracy and a higher geometric strength of the reconstruction, while providing, at the same time, a direct way of scaling the result. The combined utilization of handcrafted features and targets generally inherits the object point accuracy from the targets and allows for a more complete 3D reconstruction, but still carries the burden of target array positioning and measuring. Learning-based methods, on the other hand, can speed up the process of 3D reconstruction but offer measurement accuracies which are not comparable to target-based methods and are presently far from suitable for precise vision-based industrial metrology applications. A possible future scenario may see an initial interior orientation parameter estimation with learning-based features followed by a refinement using targets. Handcrafted and learning-based feature extraction and matching approaches provide similar results, although the latter invariably yield a much larger number of sub-optimal feature matches which are in turn more susceptible to bad network geometries.

REFERENCES

Australis, 2024: https://www.photometrix.com.au/australis/

Alcantarilla, P.F., Solutions, T., 2011. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7), pp.1281-1298.

Barazzetti, L., Mussio, L., Remondino, F., Scaioni, M., 2011. Targetless camera calibration. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. 38(5/W16).

Barroso-Laguna, A., Riba, E., Ponsa, D. and Mikolajczyk, K., 2019. Key. net: Keypoint detection by handcrafted and learned CNN filters. *Proc. CVPR*, pp. 5836-5844.

Bay, H., Tuytelaars, T. and Van Gool, L., 2006. Surf: Speeded up robust features. *Proc. ECCV*, part I 9 pp. 404-417.

Bellavia, F., Morelli, L., Menna, F., Remondino, F., 2022: Image orientation with a hybrid pipeline robust to rotations and wide-baselines. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XLVI-2/W1-2022, 73–80

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., Tardós, J. D., 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, Vol. 37(6), pp. 1874-1890.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. *Proc. ECCV*, pp. 224-236.

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-net: A trainable cnn for joint description and detection of local features. *Proc. CVPR*, pp. 8092-8101.

Fraser, C., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52(4):149-159.

Gonzáles, G. and Meggiolaro, M., 2015. Strain field measurements around notches using SIFT features and meshless methods. *Applied Optics*, 54(14), pp.4520-4528.

Harris, C. and Stephens, M., 1988, August. A combined corner and edge detector. *In Alvey vision conference*, Vol. 15, No. 50, pp. 10-5244.

Leroy, V., Cabon, Y. and Revaud, J., 2024. Grounding Image Matching in 3D with MASt3R. *arXiv preprint arXiv:2406.09756*.

Lindenberger, P., Sarlin, P.E., Pollefeys, M., 2023. Lightglue: Local feature matching at light speed. *Proc. ICCV*, pp. 17627-17638.

Loch, G.N., Szymanski, C., Stemmer, M.R., 2013. Evaluation of SIFT in machine vision applied to industrial automation. *Proc. 11th IEEE INDIN*, pp. 414-419.

Luhmann, T., 2010. Close range photogrammetry for industrial applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):558-569.

Luhmann, T., Fraser, C., Maas, H.G., 2016. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115: 37-46.

Luo, Z., Zhou, L., Bai, X., Chen, H., Zhang, J., Yao, Y., ... & Quan, L. (2020). Aslfeat: Learning local features of accurate shape and localization. *Proc. CVPR*, pp. 6589-6598.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M. and Trulls, E., 2021. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2): 517-547.

Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021. Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision*, 129(1): 23-79.

Maiwald, F., Lehmann, C. and Lazariv, T., 2021. Fully automated pose estimation of historical images in the context of 4D geographic information systems utilizing machine learning methods. *ISPRS Int. Journal of Geo-Information*, 10(11), p.748.

Markiewicz, J., Kot, P., Markiewicz, Ł., Muradov, M., 2023. The evaluation of hand-crafted and learned-based features in Terrestrial Laser Scanning-Structure-from-Motion (TLS-SfM) indoor point cloud registration: the case study of cultural heritage objects and public interiors. *Heritage Science*, 11(1), p.254

Mishchuk, A., Mishkin, D., Radenovic, F., Matas, J., 2017. Working hard to know your neighbor's margins: Local descriptor learning loss. *Advances in neural information processing systems*, *30*.

Mishkin, D., Radenovic, F., Matas, J., 2018. Repeatability is not enough: Learning affine regions via discriminability. *Proc. ECCV*, pp. 284-300.

Morelli, L., Ioli, F., Maiwald, F., Mazzacca, G., Menna, F., Remondino, F., 2024a. Deep-image-matching: a toolbox for multiview image matching of complex scenarios. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 309–316.

Morelli, L., Mazzacca, G., Trybała, P., Gaspari, F., Ioli, F., Ma, Z., Remondino, F., Challis, K., Poad, A., Turner, A., and Mills, J. P., 2024b. The Legacy of Sycamore Gap: The Potential of Photogrammetric AI for Reverse Engineering Lost Heritage with Crowdsourced Data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2-2024, 281–288

Morelli, L., Menna, F., Vitti, A., Remondino, F., Toth, C., 2023. Performance Evaluation of Image-Aided Navigation with Deep-Learning Features. *Proc. ION GNSS+ 2023*, pp. 2048-2056.

Mur-Artal, R., Montiel, J.M.M., Tardo, J.D., 2015. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, Col. 31, no. 5, pp. 1147-1163.

Ono, Y., Trulls, E., Fua, P., Yi, K.M., 2018. LF-Net: Learning local features from images. *Advances in neural information processing systems*, 31.

Ren, Z., Fang, F., Yan, N., Wu, Y., 2022. State of the art in defect detection based on machine vision. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 9(2), pp.661-691.

Remondino, F., Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol.36(5), 266-272.

Remondino, F., Nocerino, E., Toschi, I., Menna, F., 2017. A critical review of automated photogrammetric processing of large datasets. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-2/W5, pp. 591-599.

Remondino, F., Menna, F., Morelli, L., 2021. Evaluating hand-crafted and learning-based features for photogrammetric applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 43, pp.549-556.

Revaud, J., De Souza, C., Humenberger, M., Weinzaepfel, P., 2019. R2d2: Reliable and repeatable detector and descriptor. *Advances in neural information processing systems*, 32.

Riba, E., Mishkin, D., Ponsa, D., Rublee, E., Bradski, G., 2020. Kornia: an open source differentiable computer vision library for pytorch. *Proc. Winter Conference on Applications of Computer Vision* pp. 3674-3683.

Rosten, E. and Drummond, T., 2006. Machine learning for high-speed corner detection. *Proc. ECCV*, Part I 9 pp. 430-443.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision*, pp. 2564-2571.

Ruiz de Oña, E., Barbero-García, I., González-Aguilera, D., Remondino, F., Rodríguez-Gonzálvez, P., Hernández-López, D., 2023: PhotoMatch: An open-source tool for multi-view and multi-modal feature-based image matching. *Applied Sciences*, 13(9):5467.

Savinov, N., Seki, A., Ladicky, L., Sattler, T., Pollefeys, M., 2017. Quad-networks: unsupervised learning to rank for interest point detection. *Proc. CVPR*, pp. 1822-1830.

Verdie, Y., Yi, K., Fua, P., Lepetit, V., 2015. Tilde: A temporally invariant learned detector. *Proc. CVPR*, pp. 5279-5288.

Wang, Y., He, X., Peng, S., Tan, D., Zhou, X., 2024. Efficient LoFTR: Semi-dense local feature matching with sparse-like speed. *Proc. CVPR*, pp. 21666-21675.

Yi, K.M., Trulls, E., Lepetit, V., Fua, P., 2016. Lift: Learned invariant feature transform. *Proc. ECCV*, pp. 467-483.

Zhang, L., Rupnik, E., Pierrot-Deseilligny, M., 2021. Feature matching for multi-epoch historical aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 182, pp.176-189.

Zheng, J., Li, K., 2022, June. The logistics barcode id character recognition method based on akaze feature localization. *In 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Vol. 10, pp. 275-279.