



## AI-Based Extension of Rover Camera Stereo Range – Starting Validation on Mars 2020 Mastcam-Z Geologic Use Cases

Gerhard Paar<sup>1</sup>, Christoph Traxler<sup>2</sup>, Alexander Jones<sup>3</sup>, Thomas Ortner<sup>1</sup>, Jim Bell<sup>4</sup>, Sanjeev Gupta<sup>3</sup>, and Robert Barnes<sup>3</sup>

<sup>1</sup>JOANNEUM RESEARCH Forschungsgesellschaft mbH, DIGITAL - IVA, Graz, Austria (gerhard.paar@joanneum.at)

<sup>2</sup>VRVis GmbH, Vienna, Austria (traxler@vrvis.at)

<sup>3</sup>Imperial College London, London, United Kingdom (sanjeev.gupta@imperial.ac.uk)

<sup>4</sup>Arizona State University, Tempe, Arizona, USA (jim.bell@asu.edu)

### Introduction

Planetary rover camera stereoscopy has evolved within decades to serve engineering and science tasks from on-board navigation over mission and instrument planning to science operations and exploitation based on 3D data [1].

Mastcam-Z [2] is a zoomable multispectral stereoscopic camera on the Mars 2020 Perseverance rover. Its stereo range useful for geologic spatial investigations spans from very close range up until a few dozens of meters, with a quadratically increasing range error [3]. Whilst small deviations of geometric camera calibration from “truth” result in a predictable and correctable systematic error, actual stereo processing introduces 3D range noise in the named quadratically increasing form.

Recently available AI methods show promising solutions determining monoscopic depth (i.e. depth from single images in pixel resolution) [4] for a large range of applications. Such range maps lack true scale, as no artificial objects with known scale can be expected on planetary surfaces. Yet, pixel-resolution occlusion determination and micro-shape are well represented, as shown in Figure 1. This inspired the combination of monoscopic AI-based range determination with true scale as available from calibrated stereoscopy.

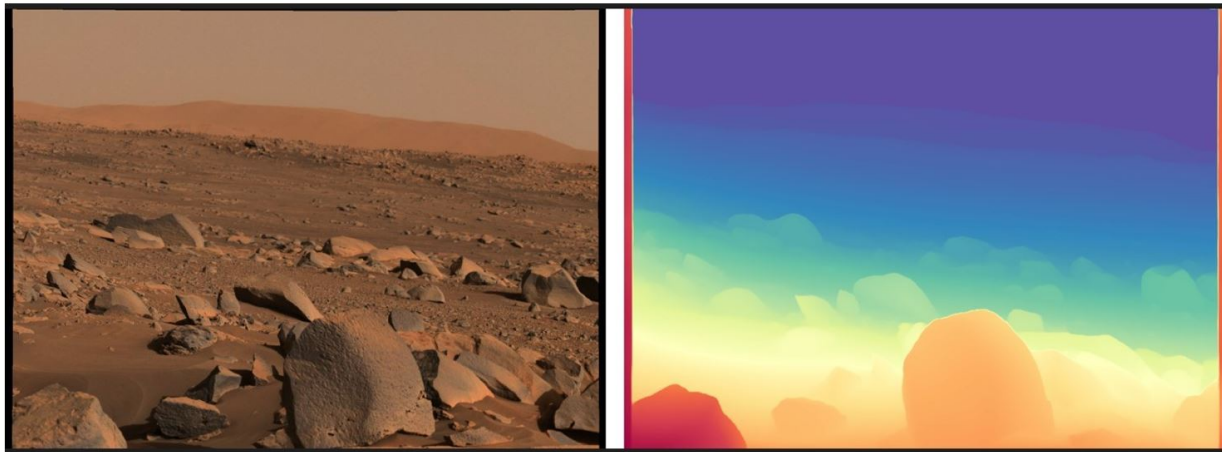


Figure 1: Prominent rocks in close vicinity of the rover as typical examples of occlusions in the context of Mars (left: camera image of Sol 168, Mastcam-Z sequence 8180 at a focal length of 48mm, right: color-coded monoscopic depth map from Depth-Anything Version 2, DA-V2) NASA/JPL-CalTech/MSSS/ASU

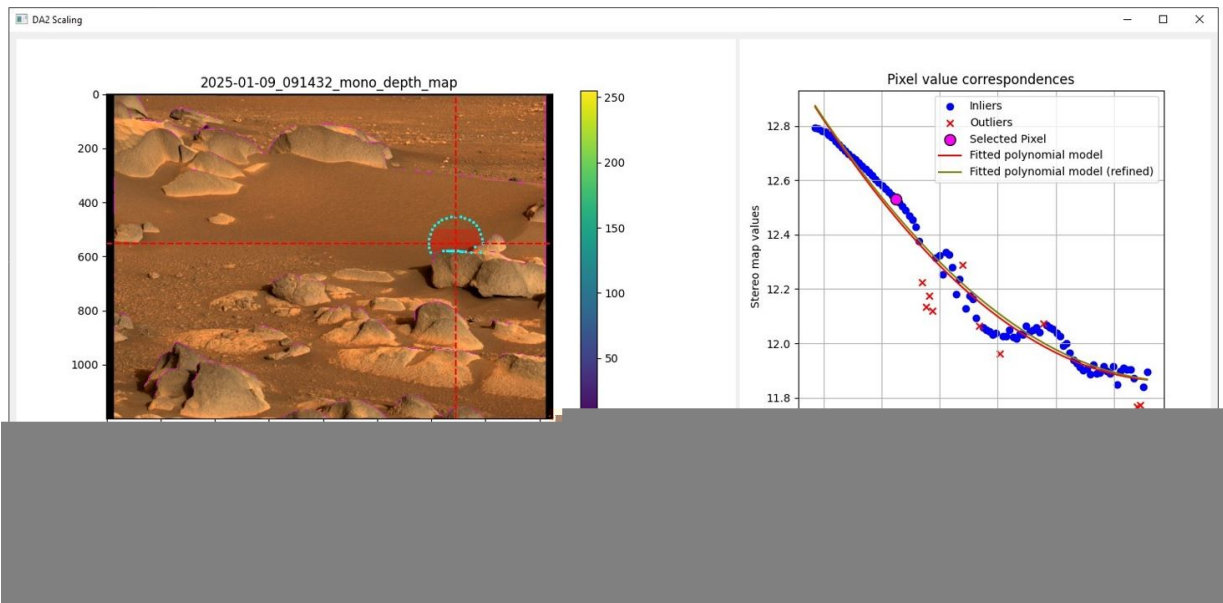
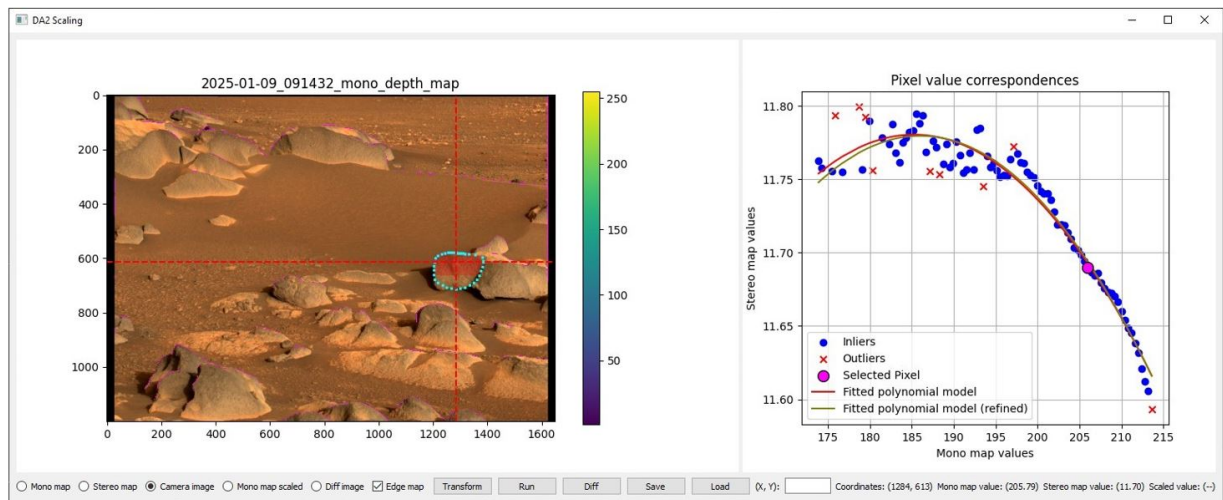
### AI-Based Enhancement of Stereoscopic Range Products

Monocular depth models such as Depth Anything V2 (DA-V2, a deep learning-based system for estimating depth from a single camera image [4]) have emerged to powerful tools in image simulation, segmentation and scene understanding, however do not provide exact scale. Relative depth estimates produced by such monocular depth estimation need to be turned into physically meaningful metric depth values. Calibrated stereo camera configurations such as Mastcam-Z provide true-scale depth maps. Fusing them works by establishing local correspondences between the two depth maps and fitting range transformation functions that can convert pixel values from the monocular (unscaled) domain into the stereo (scaled for true distances) domain. These transforms are calculated and applied locally, either per pixel or per block of pixels, to account for spatial variations in the relationship between the two depth sources. Edge-aware processing ensures that depth discontinuities are handled more precisely on occlusions (such as rocks or boulders) to select search areas for which transformations are calculated. The result is a scaled depth map that combines the high-frequency detail and spatial completeness of DA-V2 with the geometric accuracy of stereo photogrammetry, enabling detailed yet quantitatively accurate depth maps suitable for scientific analysis and engineering applications, such as planetary rover navigation and terrain modeling.

The basic operation of the algorithm is as follows:

- Detect occlusion edges in mono depth map to form a binary "edge map"
- Iterate over mono depth map in blocks of configurable size
- For each mono depth map pixel, create a circular search area using the pixel as its center, taking into account occlusion edges
- Calculate correspondences between mono and stereo map depth values within the search area (Figure 2)
- Fit function to correspondence data using quadratic, polynomial, exponential, or linear approximation
- Transform mono map values in current pixel block to final product using fitted function.

Several enhancements are implemented – most notably a two-step function fitting procedure and per-pixel interpolation of transformations based on spatial distance – as well as more fine-grained options to control the behavior of the algorithm. A prototype version of this scaled DA-V2 algorithm has been integrated into the nominal Mastcam-Z 3D P<sub>Ro</sub>ViP vision processing [6].



## Qualitative Results

A series of preliminary tests indicate highly promising performance, leading to a substantial range extension of the nominal fixed-baseline stereo capabilities of stereoscopic imaging instruments, in particular for the planetary science case. From examples (see Figure 3) it is evident, that at least a range boost of 2 to 5 can be expected by using the described approach.

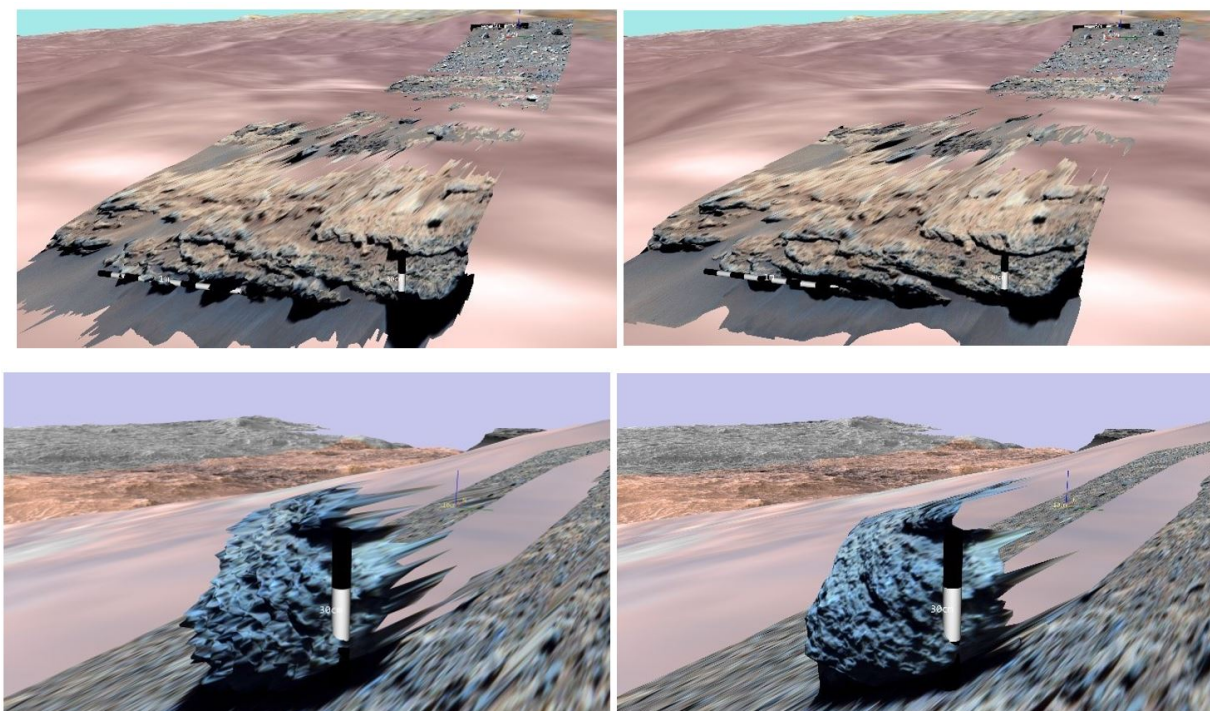
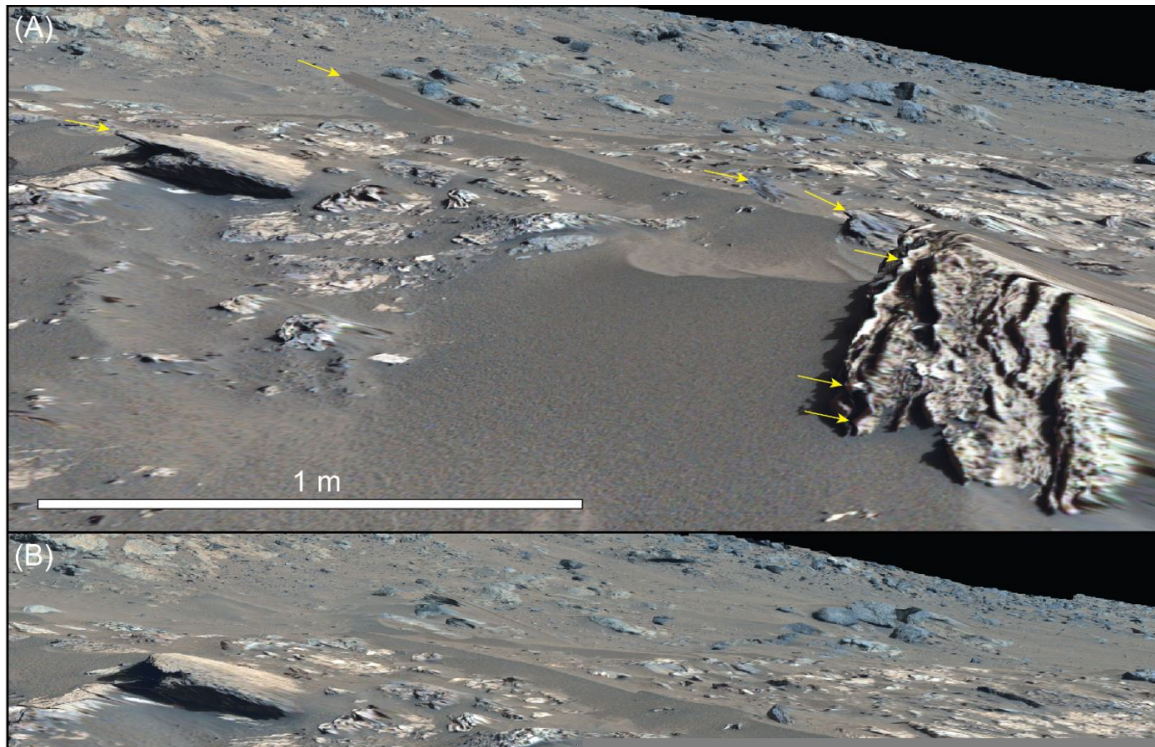


Figure 3: Examples of AI-based depth enhancement (PRo3D visualization of Mastcam-Z PProViP stereo product). Respective left: Standard stereo processing, right: AI-based depth enhanced true-scale version. Top: Sequence 3781 of Sol 928 at 110 focal length, captured at a distance of 31m. Bottom: Sequence 9503 of Sol 1448 at 110mm focal length, capturing distance 20m. NASA/JPL-CalTech/MSSS/ASU/USGS/JR/VRVis

A typical use case to both the original (conventional) stereo-derived 3D result and the AI-enhanced version in comparison (Figure 4) gains the following preliminary assessment:

- Prior to AI processing, the data from this area shows several “spikes” which preclude reliable structural measurements (e.g. strike/dip of veins in the foreground block) being acquired. In the case above (at 34 mm focal length), this occurs at distances of just  $\sim 6 - 8$  m from Mastcam-Z.
- Following AI enhancement, these “spikes” are removed, allowing more reliable measurements of Ca-sulfate vein orientations to be measured.



### **Towards Validation of AI-Based Stereo Enhancement**

A highly comprehensible way for a *qualitative comparison* is to toggle between the surfaces to be compared. This allows to study structural deviations at different viewpoints. Changing into wireframe mode can further enhance the perception of topographic differences.

A series of *quantitative validation* approaches is under development. These include:

- 3D-comparison between AI-enhanced surface reconstruction using stereo pairs viewed from medium distance, with “conventional” stereo reconstructions of the same scene viewed from close-range (including co-registration to minimize systematic errors from localization and camera calibration) – see Figure 5, top and middle
- Color-coding the reference surface according to the distance to the other one using the distance between surface points by using the up-vector or surface normals (Figure 5, bottom)
- 3D-comparison of AI-enhanced surface reconstructions from images taken with the same imaging

- geometry under different illumination conditions
- Independent geologic analysis on AI-enhanced and conventional Digital Outcrop Models (DOMs) and results comparison of, e.g., dip-and-strike measurements
- Analysis of shadow shapes with obtained object outlines.

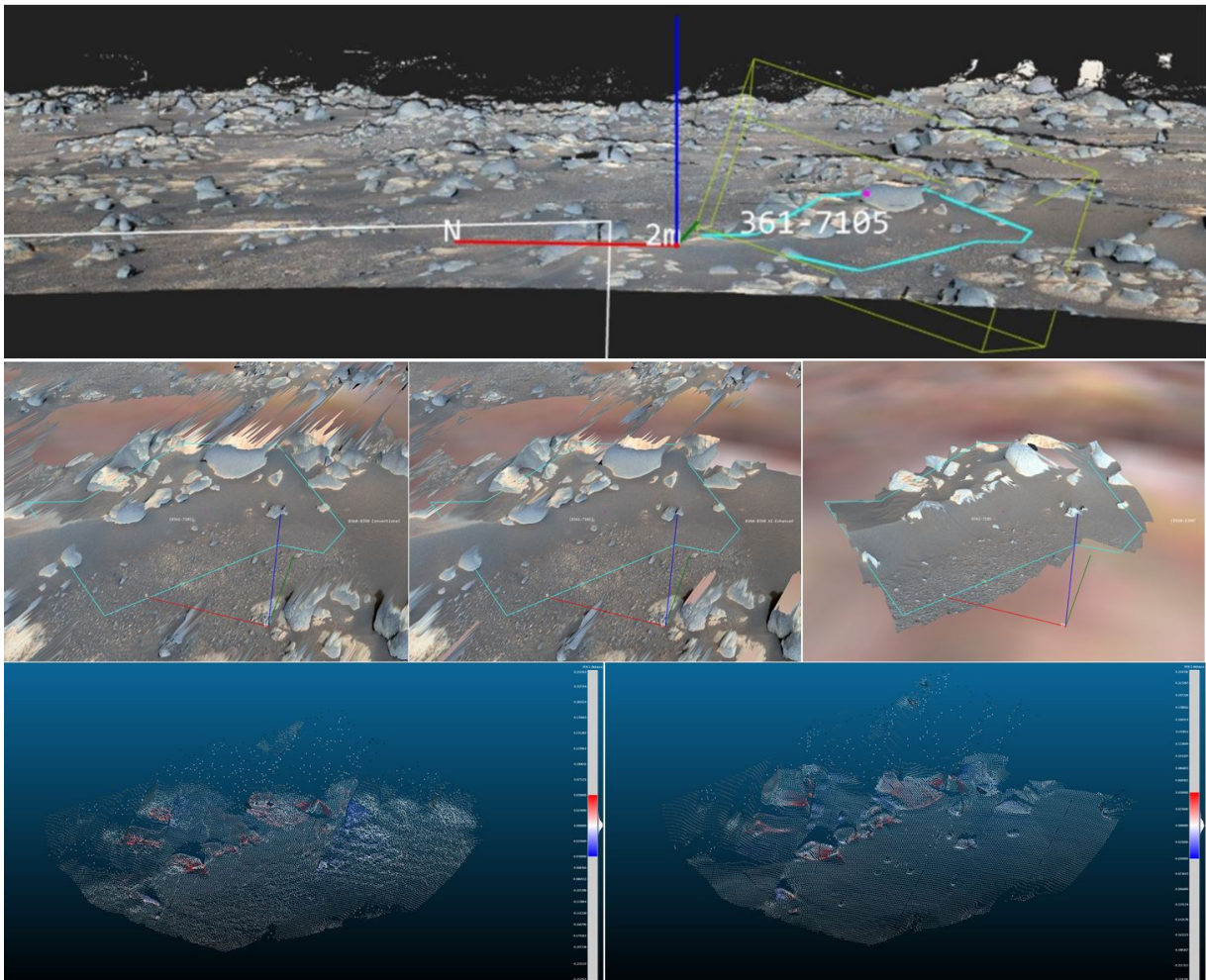


Figure 5: Top: View of Mars 2020 Mastcam-Z Sol 360 (Sequence 8390) stereo product from Rover position at Sol 360 (Distance ~20m) indicating footprint of Mastcam-Z Sol 361 (Sequence 7105) stereo product (taken at a distance of ~3m, the reference) inserted in blue. Middle left: Portion of Sol 360 conventional 3D reconstruction. Middle center: Portion of Sol 360 AI-enhanced 3D reconstruction Middle right: Sol 361 3D reconstruction co-registered to Sol 360 3D reconstruction. Bottom left: Sol 360 conventional 3D reconstruction color-coded with distances to Sol 361 3D reconstruction. Bottom right: Sol 360 AI-enhanced 3D reconstruction color-coded with distances to Sol 361 3D reconstruction. NASA/JPL-CalTech/MSSS/ASU/JR/VRVis/ICL

## Outlook

A statistically significant validation of the approach is presently in development and planned to be finalized before the joint EPSC/DPS Conference 2025.

A planned improvement of the visual comparison method between surfaces is to show color-coded deviation vectors between vertices. This works best in wireframe mode or with semi-transparent surfaces. Furthermore it should be possible to query numerical deviation values by clicking on a surface point.

## Acknowledgement

This work was funded by the ASAP Project AI-Mars-3D (FFG-911920). We thank Komyo Furuya of JR for operational end-to-end implementation and documentation of the re-scaling algorithm.

## References

- [1] Paar, G., Deen, R. G., Muller, J. P., Silva, N., Iles, P., Shaukat, A., & Gao, Y. (2016). Vision and image processing. *Contemporary planetary robotics: An approach toward autonomous Systems*, 105-179.
- [2] Bell, J. F., Maki, J. N., Mehall, G. L., Ravine, M. A., Caplinger, M. A., Bailey, Z. J., ... & Wolff, M. J. (2021). The Mars 2020 perseverance rover mast camera zoom (Mastcam-Z) multispectral, stereoscopic imaging investigation. *Space science reviews*, 217, 1-40.
- [3] Ruoff, N.A., Deen, R.G., Pariser, O., Mars 2020 Project Software Interface Specification (SIS), Camera Data Products. NASA/JPL, 2025, [https://pds-geosciences.wustl.edu/m2020/urn-nasa-pds-mars2020\\_mission/document\\_camera/Mars2020\\_Camera\\_SIS.pdf](https://pds-geosciences.wustl.edu/m2020/urn-nasa-pds-mars2020_mission/document_camera/Mars2020_Camera_SIS.pdf)
- [4] Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., & Zhao, H. (2024). Depth anything v2. *Advances in Neural Information Processing Systems*, 37, 21875-21911.
- [5] Gerhard Paar, Thomas Ortner, Christoph Traxler, Robert Barnes, Matt Balme, Christian Schröder, Steven G. Banham (2022). Preparing 3D vision and visualization for ExoMars. ASTRA 2022: 16<sup>th</sup> Symposium on Advanced Space Technologies in Robotics and Automation. <https://www.vrvis.at/publications/pdfs/PB-VRVis-2022-017.pdf>
- [6] Gerhard Paar , Thomas Ortner , Christian Tate , Robert G. Deen , Parker Abercrombie, et al (2023). Three-dimensional data preparation and immersive mission-spanning visualization and analysis of Mars 2020 Mastcam-Z stereo image sequences. *Earth and Space Science* 10.3 (2023). <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029%2F2022EA002532>