AI-Based Tracking of Fast-Moving Alpine Landforms Using High Frequency Monoscopic Time-Lapse Imagery

Hanne Hendrickx[1,2*], Melanie Elias[1*], Xabier Blanch[1,3], Reynald Delaloye[2], Anette Eltner[1]

[1]Institute of Photogrammetry and Remote Sensing, TUD Dresden University of Technology, 01062 Dresden, Germany
[2]Department of Geosciences, University of Fribourg, Fribourg, 1700, Switzerland
[3]Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain

Correspondence to: Hanne Hendrickx (hanne.hendrickx@tu-dresden.de), Melanie Elias (melanie.elias@tu-dresden.de)

* shared first authorship

Abstract: Active rock glaciers and landslides are dynamic landforms in high mountain environments, where their geomorphic activity can pose significant hazards, especially in densely populated regions like the European Alps. Moreover, active rock glaciers reflect the long-term thermal state of permafrost and respond sensitively to climate change. Traditional monitoring methods, such as in-situ differential Global Navigation Satellite System (GNSS) and georeferenced Total Station (TS) measurements, face challenges in measuring the rapid movements of these landforms due to environmental constraints and limited spatial coverage. Remote sensing techniques offer improved spatial resolution but often lack the necessary temporal resolution to capture sub-seasonal variations. In this study, we introduce a novel approach utilising monoscopic time-lapse image sequences and Artificial Intelligence (AI) for high-temporal-resolution velocity estimation, applied to two subsets of time-lapse datasets capturing a fast-moving landslide and rock glacier at the Grabengufer site (Swiss Alps). Specifically, we employed the Persistent Independent Particle tracking (PIPs++) model for 2D image point tracking and the image-to-geometry registration to transfer the measured 2D image points into 3D object space and further into velocity data. For the latter, we use an in-house tool called GIRAFFE, which employs the AI-based LightGlue matching algorithm. This methodology was validated against GNSS and TS surveys, demonstrating its capability to provide spatially and temporally detailed velocity information. Our findings highlight the potential of image-driven methodologies to enhance the understanding of dynamic landform processes, revealing spatio-temporal patterns previously unattainable with conventional monitoring techniques. By leveraging existing time-lapse data, our method offers a cost-effective solution for monitoring various geohazards, from rock glaciers to landslides, with implications for enhancing alpine safety. This study marks the pioneering application of AI-based methodologies in environmental monitoring using time-lapse image data, promising advancements in both research and practical applications within geomorphic studies.

1 Introduction

Active rock glaciers are creeping permafrost features (Barsch, 1996; Kääb and Reichmuth, 2005), serving as indicators of permafrost distribution in high mountain environments (Marcer et al., 2017; RGIK, 2023). Their velocity results from various

Style Definition: Bullets: Outline numbered + Level: 1 + Numbering Style: Bullet + Aligned at: 0.63 cm + Tab after cm + Indent at: 1.27 cm

Formatted

parameters such as topographic conditions, rock glacier material, and internal structure, and it reflects long-term, temperature driven changes in permafrost structure (Cicoira et al., 2021; Delaloye et al., 2010). Higher creep rates typically occur towards the lower permafrost limits, where mean annual air temperatures approach 0°C (Frauenfelder et al., 2003). They efficiently transport sediment (Delaloye et al., 2010; Kummert and Delaloye, 2018), and this becomes more pronounced as rock glacier creep rates increase in a warming climate (Delaloye et al., 2013; Pellet et al., 2023). ~~The same is true for certain types of landslides, where the primary driver of motion is permafrost creep.~~ Similarly, large volumes of sediment can be mobilised by permafrost-affected rock slope failures, such as deep-seated slides, topples, or deformations that involve in-situ bedrock (McColl and Draebing, 2019).This can pose significant geohazards when direct connections to downslope infrastructure exist. Precise monitoring of these fast-moving high-alpine landforms is thus essential for future alpine safety (Hermle et al., 2022), as it provides information about the environmental drivers and enhances process understanding.

The monitoring of fast creeping rock glaciers (>_3 m a^{-1}, Marcer et al. 2021) or landslides is particularly challenging. Traditional techniques that require frequent field access, such as in-situ differential Global Navigation Satellite System (GNSS) measurements, face environmental and logistical obstacles. Permanent GNSS installations can offer displacement observations with millimetre accuracy at a continuous temporal resolution, but they may not have the desired longevity on fast-moving landforms due to extreme cases of block sliding, rotation, and rockfall, necessitating re-levelling or instrument replacement (Cicoira et al., 2022). Both GNSS and total station (TS) measurements only measure discrete points, resulting in a limited spatial distribution. However, spatial heterogeneity of landform movement can be expected depending on internal rock glacier structure and terrain characteristics (RGIK, 2023). Improved spatial coverage can be achieved using remote sensing data, such as 3D point clouds derived from Uncrewed Aerial Vehicles (UAV) and Terrestrial or Airborne Laser Scanning (TLS and ALS). These techniques minimize the need for extensive field access, enabling operators to avoid in-person exposure to the fastest-moving areas while still capturing detailed data from these regions. However, these methods often lack the temporal resolution necessary to capture sub-seasonal variations of the landform to its environmental drivers, essential to increase process understanding. Time-lapse imagery or webcam data have the capability to capture the kinematics of alpine landforms with, e.g., hourly resolution. Fixed photogrammetric camera systems, which are increasingly implemented, involve multiple time-lapse cameras to reconstruct high-resolution 3D point clouds similar to those from UAVs or TLS, and at a fraction of the cost ~~(Blanch et al., 2023; Eltner et al., 2017; Ioli et al., 2024, Ulm et al., 2025)~~(Blanch et al., 2023; Eltner et al., 2017; Ioli et al., 2024, Ulm et al., 2025). Nonetheless, deploying multiple cameras in dynamic alpine environments can be challenging, often requiring wide baselines that complicate point cloud generation (Ioli et al., 2024). Recent studies have demonstrated that metric measurements can be obtained from 2D images alone when a 3D model is available, reducing the need for extensive camera arrays (Altmann et al., 2020; Elias et al., 2023; Wegner et al., 2023). Furthermore, monoscopic camera data, which is more readily available than stereo images, often spans a decade or more in the European Alps (Kummert et al., 2018). Generally installed in stable terrain, these cameras tend to have greater longevity than permanent GNSS installations. However, most time-lapse camera systems are highly weather-dependent. Changes in the camera's intrinsic and extrinsic properties caused by thermal variations (e.g., Elias et al., 2020) and external disturbances such as snow and wind are common, significantly affecting

Formatted: Font: Times New Roman

the image configuration. Additionally, maintaining a perfectly stable camera position over the long term can be very challenging in a dynamic mountain environment.

Measuring the velocity of boulders in the camera's field of view is possible by measuring homologous point correspondences between images in a series, a common challenge in computer vision called feature tracking using optical flow-based methods and feature matching (Fortun et al., 2015). Feature tracking is essentially an optimisation problem, where the location of highest similarity between a reference template and a template in the destination image is considered a match (Eltner et al., 2022). Common similarity measures include Normalised Cross-Correlation (NCC) (Heid and Kääb, 2012) and Least Square Matching (LSM, Schwalbe and Maas, 2017). While traditional motion estimation methods are commonly applied in natural hazard management, they face decorrelation challenges with large displacements, strong illumination changes, and occlusions. Image pre-selection is crucial to minimize illumination variations (e.g., selecting images from the same time of day), thereby limiting temporal resolution (How et al., 2020).

In the era of Artificial Intelligence (AI), traditional feature tracking can leverage the power of Deep Learning. Convolutional Neural Networks (CNNs), employed as feature extractors, replace handcrafted features or the use of image intensities or gradients (Hur and Roth, 2020). This approach offers the advantage of representing each pixel with a high-dimensional feature vector, blending distinctiveness and invariance to, for example, appearance changes, thereby ~~enhancing~~improving feature robustness over time. This enhances, on the one hand, image feature tracking (e.g., PIPs++ (Zheng et al., 2023), VideoFlow (Shi et al., 2023), Raft (Teed and Deng, 2020)), and, on the other hand, image matching algorithms, as seen with models like SuperGlue (Sarlin et al., 2020), LightGlue (Lindenberger et al., 2023) or LoFTR (Sun et al., 2021).

~~The goal of this research is~~This study aims to ~~obtain~~ derive high-resolution spatio-temporal ~~high-resolution~~ landform velocities from monoscopic time-lapse cameras, verified through ~~permanenting~~permanent in-situ GNSS and TS surveys. ~~To improve motion estimation in mountain time-lapse image sequences, we~~We tested the deep learning model called Persistent Independent Particle tracking (PIPs++) (Harley et al., 2022; Zheng et al., 2023) to track landslide and rock glacier movements, and the AI ~~matching algorithm~~-based image matcher LightGlue (Lindenberger et al., 2023) for ~~image~~2D-to-~~geometry~~3D registration ~~to transfer 2D image measurements to 3D information~~. This process is implemented in ~~an~~our in-house, open-source tool~~, designated~~ GIRAFFE (Geospatial Image Registration And reFErencing~~), GIRAFFE performs~~), whichrenders synthetic ~~image rendering~~ images from ~~a~~ scaled 3D point ~~cloud~~clouds and ~~approximates of~~estimates a linear camera model to ~~fit~~match the 2D image perspective ~~of the originally captured 2D image. The method was presented in~~. Originally introduced by Elias et al. (2019) ~~with a particular emphasis on image-based~~ for water level ~~measurement. The matching was based on handcrafted features utilising the Scale-Invariant Feature Transform (~~monitoring using SIFT~~) operator~~ (Lowe et al., 2004), ~~which is an effective method when the images to be matched have similar perspectives and lighting conditions. To enhance robustness against these conditions, the tool was further developed using~~ GIRAFFE was later improved with the AI-based SuperGlue matcher (Sarlin et al., 2020) ~~matcher (Elias et al., 2022) and demonstrated the potential~~ for multi-modal image-to-

geometry registration (Elias et al., 2022; 2023). In this study, we ~~employ~~use the stable and open-source LightGlue ~~image~~ matcher~~. We publish~~ and release the GIRAFFE source code ~~of GIRAFFE to carry out own~~to support flexible, time-lapse-based velocity ~~measurements with our presented approach.~~

analysis. We introduce a novel low-cost image-based remote measurement technology designed for ~~application in~~ challenging and hard-to-access terrains. The proposed method has been validated ~~by processing~~for two ~~subsets from two extensive~~ time-lapse image datasets, ~~converting them into~~by extracting image-based velocity information. Two rapidly moving alpine landforms, a landslide and a rock glacier, were observed at the Grabengufer site in Switzerland~~. We highlight the potential of these innovative methodologies~~ (Kenner et al., 2014). We highlight the potential our methodologiy to manage large datasets effectively and enhance the spatio-temporal understanding of landform dynamics. Our approach not only improves analytical capabilities but also facilitates the use of basic systems, such as monoscopic cameras, for automated, cost-effective monitoring and quantification of landform movements.

2. Study area ~~and dataset description~~

The Grabengufer study area features two fast-moving alpine landforms (Fig. 1a). The upper section (2700 – 2880 m a.s.l.) consists of an extensive deep-seated landslide moving up to 1.5 m a^{-1} (Fig. 1a blue polygon). ~~The landslide is made of in-situ bedrock, fractured by its motion. Its thickness is estimated to be around 30 to 40 m and up to 100 m in the most northern section where it is covered by debris fallen from the Grabenhorn.~~ According to installed Ground Surface Temperature (GST) loggers, its mean annual surface temperature is still below 0° despite recent warming. The southern part of the landslide is frozen and ice saturated (unpublished ~~data).~~geophysical data and local inspection of fresh scars after rockfalls). Maximal intra-annual velocities are reached around November and the velocities have multiplied by four (from 0.3 to 1.2 m a^{-1}) between 2009 and 2024. ~~Therefore, permafrost creep is expected to be one of the major drivers of the current motion of the landslide.~~ However, due to the absence of rock glacier morphology and the uncertainty of driving factors for the landform's motion, we use the generic term 'landslide'~~,~~. The frontal section of the southern part of this landslide is ~~very~~highly unstable and ~~the~~a frequent source of ~~frequent~~ rock falls. ~~It~~Downslope, it gradually ~~turns~~transitions into a zone of very rapid mass movement, referred to as the "(rock glacier) feeding section~~". Both .~~" The transition between these two areas is gradual, without a clear boundary: the frontal ~~and~~section marks the upper part of the landslide, where debris of various sizes detach, while the feeding ~~sections~~section moves significantly faster and shows no further detachment. Both areas are monitored by a permanently installed webcam (hereinafter ~~named~~ Wcam04, Fig. 1c, orange polygon). For the purpose of this study, and to simplify terminology, we do not distinguish between the two in the analysis and refer to them collectively as the "frontal feeding section." The fed rock glacier (2400 – 2600 m a.s.l., Wcam05, Fig. 1b red polygon) is very active and was considered destabilised during the 1940s and 2000s~~.~~ (Delaloye et al., 2013). As of summer 2023, in-situ measurements indicate that it moves again exceptionally fast at a rate of 91 to 255 m a^{-1} (0.25 to 0.70 m d^{-1}~~,~~), marking a third phase of destabilisation. The

Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: English (United Kingdom)

rock glacier tongue terminates on a very steep slope section leading into a gully prone to debris flows, which is observed by a separate webcam not included in this study.

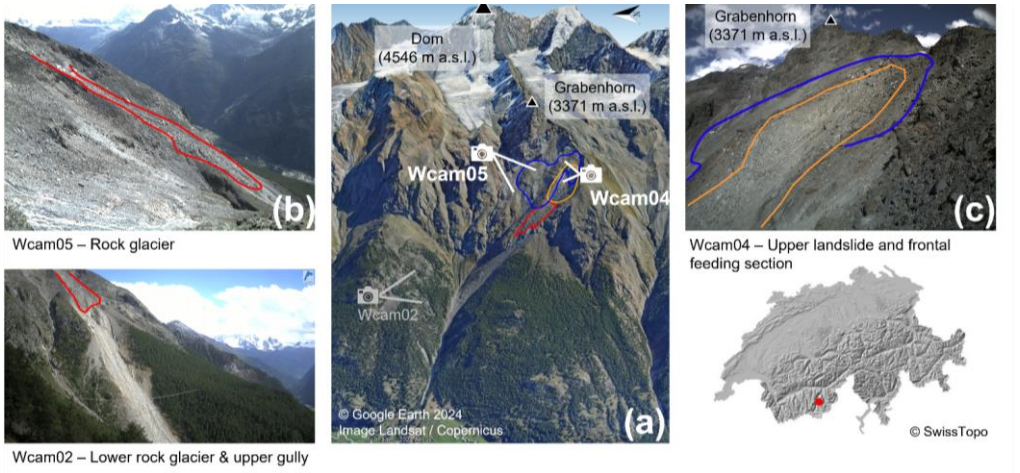
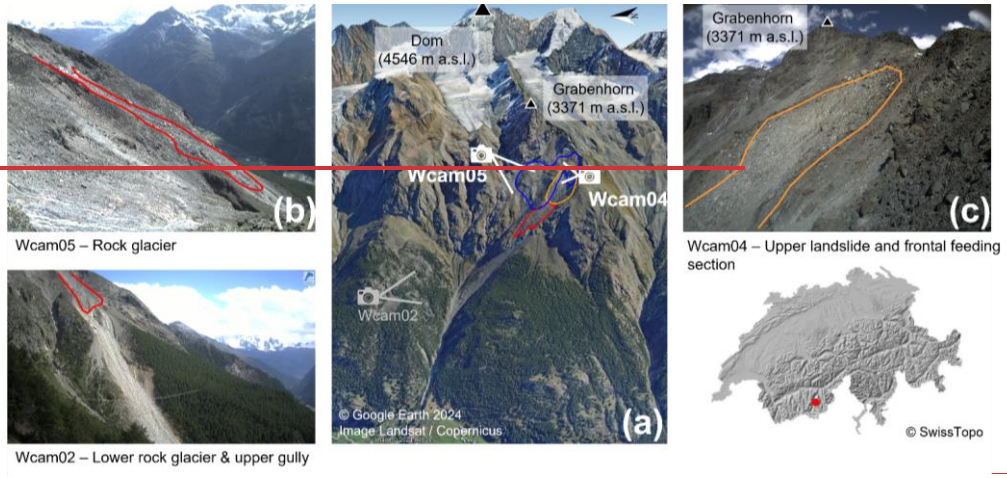


Figure 1: a) Overview of the Grabengufer study area in the Mattertal, Swiss Alps, showing the locations of the webcams used in this study and the studied landforms (blue – landslide extent, orange – frontal feeding section, red – rock glacier), b) Example image from webcam Wcam05, capturing the rock glacier, and c) example image from webcam Wcam04, focused on the frontal section of the landslide gradually feeding into the rock glacier—referred to as the frontal feeding section in this study.

3. Dataset description

The site is extensively monitored through bi-annual dGNSS surveys (conducted at the end of June and October), several permanent GNSS installations operated by PermaSense (Cicoira et al., 2022) and Canton of Valais ~~(providing continuous data saved as 2-hour means),~~ ten fixed reflectors for repeated TS measurements (approximately once a month), three time-lapse cameras operational since 2010/2013 (Wcam02, Wcam04, and Wcam05), and three GSTs. For this study, we used the monthly TS measurements and a permanent GNSS system installed and operated by the Canton of Valais (providing continuous data saved as 2-hour means) to validate the velocities derived from the time-lapse image sequences. This GNSS system uses a local fixed reference station in stable terrain located approximately 3 km away and 250 m lower in elevation, to provide accurate differential positioning. The image data analysed in this study was captured by the webcams Wcam04 and Wcam05, powered by solar panels. The characteristics of these webcams are summarised in Table 1.



~~**Figure 1: a) Overview of the Grabengufer study area in the Mattertal, Swiss Alps, showing the locations of the webcams used in this study and the studied landforms (blue – landslide extent, orange – feeding section, red – rock glacier), b) Example image from webcam Wcam05, capturing the rock glacier, and c) example image from webcam Wcam04, focused on the frontal section of the landslide feeding into the rock glacier – referred to as the feeding section in this study.**~~

The 3D reference data utilised for image-to-geometry registration was derived from UAV parallel-axis, nadir image flights conducted using a DJI Phantom 4 RTK. These flights were performed on July 3, 2023, and July 22, 2023, to reference images from Wcam04 and Wcam05, respectively. ~~The image blocks were processed using Structure-from-Motion (SfM~~The July 3 dataset consists of 1,000 images captured at 85 m above ground level, yielding a 2.5 cm Ground Sampling Distance (GSD). Four GCPs were used for georeferencing, with positional errors of 1 cm horizontally and vertically. The challenging terrain limited the inclusion of additional checkpoints. The July 22 dataset covers a larger area, extending to the valley, with 1,800 images taken at 88 m altitude, maintaining a similar GSD. Seven GCPs were distributed across the site, with two independent checkpoints. GCP residual errors averaged 4 cm, while checkpoint errors were higher (7.5 cm horizontal, 9.5 cm vertical). The georeferencing of these 3D point clouds employed an integrated approach, combining RTK measurements of image position and onboard IMU data. The image blocks were processed using Structure-from-Motion Multiview Stereo (SfM-MVS) techniques implemented in Agisoft Metashape v.1.7, resulting in high-resolution 3D point clouds enriched with RGB colour information. ~~The georeferencing of these 3D point clouds employed an integrated approach, combining RTK measurements of image poses and onboard IMU data. Additional ground control points (GCPs) were used to ensure a stable configuration (e.g., Elias et al., 2024). The dataset from July 3, 2023, comprises over 1,000 images captured at a constant flight altitude of approximately 85 m above ground level, yielding a ground sampling distance (GSD) of 2.5 cm/pixel. The overall reprojection error for the sparse points was 0.3 pixels. Four GCPs were incorporated for integrated georeferencing; however, the challenging~~

~~terrain limited the inclusion of additional checkpoints. Positional errors at the GCPs were observed to be 1 cm in both horizontal and vertical directions. It is important to interpret these errors cautiously, as they represent internal precision errors tied to the SfM-based 3D point cloud computation process. The dataset from July 22, 2023, covers an extensive area extending from the investigation site to the valley. Approximately 1,800 images were collected at a consistent flight altitude of around 88 m above ground level, resulting in a GSD comparable to that of the earlier dataset. Seven GCPs were strategically distributed across the area, from the valley floor to the upper slopes. Additionally, two independent checkpoints were established: one within the Wcam05 investigation area and another in the valley. The average residual error at the GCPs was approximately 4 cm in both horizontal and vertical components. However, deviations at the independent checkpoints were notably higher, with errors of 7.5 cm horizontally and 9.5 cm vertically. Both~~Additional GCPs were used to ensure a stable configuration (e.g., Elias et al., 2024). The resulting 3D point clouds were resampled to a point spacing of 5 cm. This resampling facilitated the management of the large volume of 3D data, comprising hundreds of millions of points, while preserving detailed topographic representation. ~~For further analysis, the point clouds were utilised to transfer the image measurements from the webcam to object space.~~

Table 1: Camera and image dataset specifications

		Wcam04	**Wcam05**
Camera specs	Camera model	Mobotix M12	Mobotix M15
	Sensor / pixel size (ps)	½" CMOS (6.4x4.8 mm) / 3.1µm	
	Image resolution	2048 x 1536 px (3.15 MPx)	
	Focal length (f)	5.4 mm (43 mm in 35-mm equivalent), crop factor 8	
Camera pose	Location in CH1903+*	X: 2628628 m, Y: 1104523 m, Z: 2790 m a.s.l.	X: 2629114 m, Y: 1105319 m, Z: 2870 m a.s.l.
	~~Distance to landform~~	~~100 - 350 m~~	~~700 – 1,100 m~~
	View direction (NESW)	100 °	235 °
Acquisition specs	Date of installation	~~01.09.2013~~20.20.2010	17.01.2011
	Sampling interval	hourly during daylight	
	Data transmission	GSM (Global System for Mobile Communication) internet connection (server storage)	
	# of images (start-2023)	48 118	51 919
	Distance to landform	100 - 350 m	700 – 1,100 m
	Approx. image GSD**	0.06 – 0.21 m	0.40 – 0.63 m
	Subsample for this study	Weekly images for summer 2022 (06.06 to 10.10.2022, 19 images)	

*CH1903+ is a commonly used Swiss coordinate system, **GSD = Ground Sampling Distance. Size of the pixel on the landform of interest

Formatted Table

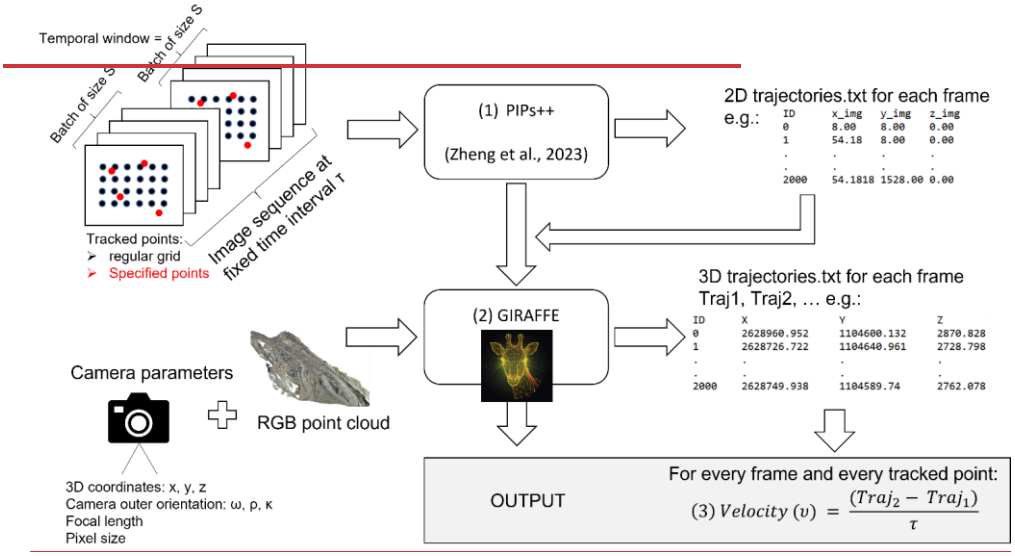
Formatted: Highlight

Formatted: Highlight

3. Methods

The workflow, visualised in ~~Figure~~Fig. 2, consists of ~~several sequential~~two main steps designed to efficiently and accurately derive 3D trajectories from time-lapse image sequences. First (1), translation vectors (pixel-based trajectories) for predefined tracking points, either arranged on a regular grid or at specified pixel coordinates, are extracted from the image sequence. In the next step (2), these 2D trajectories, estimated between consecutive images, are mapped to real-world coordinates using a 3D point cloud of the region of interest, which ~~serves as~~defines the reference system (see Fig. 2). This transformation enables the calculation of 3D coordinates for the tracked points to generate 3D trajectories. Finally, these trajectories are used to compute velocities, with units determined by the temporal resolution of the input data, such as the time-lapse interval (e.g., m d^{-1}).

~~Our workflow requires minimal input data.~~ The primary requirement of our workflow is a set of time-lapse images from fixed or otherwise stabilized cameras, preferably captured at ~~fixed~~consistent intervals (e.g., hourly, daily, or weekly~~), preferably~~) with datetime information embedded in the filenames. Additionally, approximate values for the camera pose and intrinsic parameters, such as focal length and pixel size, are necessary. ~~No pre-filtering to remove images affected by fog, snow, or motion blur is needed.~~



Formatted: English (United States)

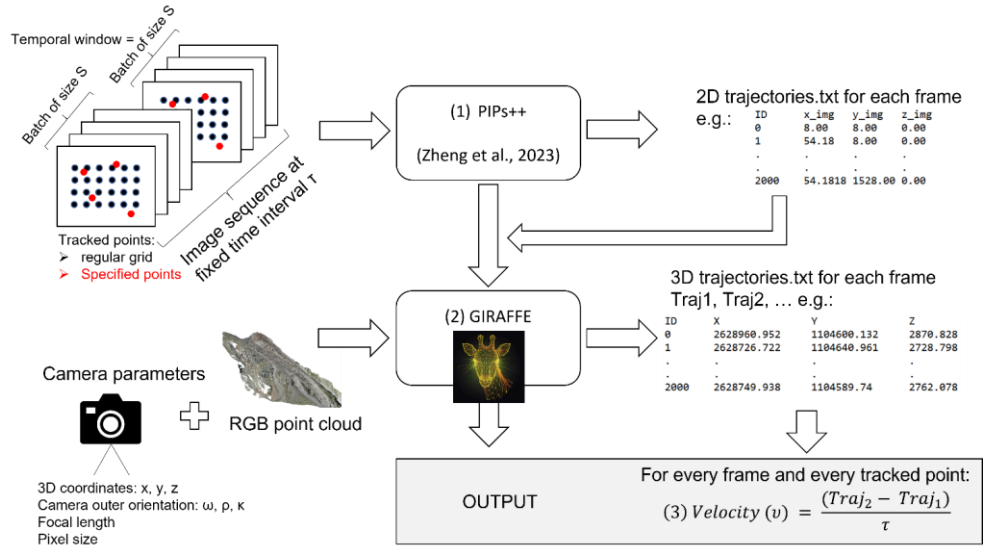
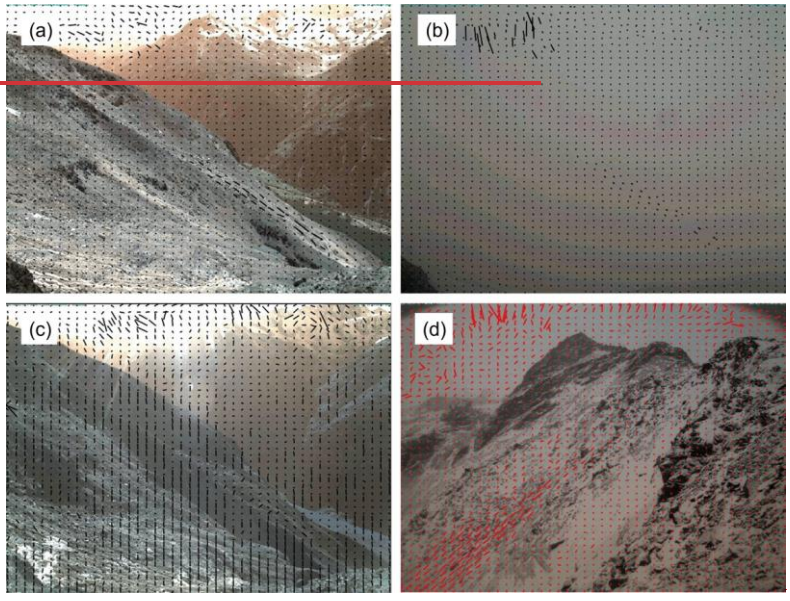


Figure 2: Schematic overview of the workflow presented in this study.

3.1 Application of the Persistent Independent Particle tracker (PIPs++)

The first step in our workflow involves feeding the image sequence into the Persistent Independent Particle tracker (PIPs++). ~~This~~The AI-based tracker, developed by Harley et al. (2022) and improved by Zheng et al. (2023), operates without retraining on our specific data. PIPs++ operates as a low-level tracker, relying on appearance-matching cues just like traditional methods~~, but leveraging the full information that is available from the image sequence within the temporal window.~~ (Eltner et al., 2020), but leveraging the full information that is available from the image sequence within the temporal window. It employs a 2D residual CNN for feature extraction in the initial step (He et al., 2016), generating a feature map for each frame independent of the image before or after. Following feature extraction, the algorithm feeds into a deep 1D CNN (ResNet) with fixed-length kernels applied to arbitrary temporal spans (Zheng et al., 2023). PIPs++ ~~first detects features and then~~ calculates descriptors ~~to estimate~~and then estimates the local similarity of each feature to match them over time. Using spatial pyramids, it iteratively refines the position of these matched features, similar to the RAFT model (Teed and Deng, 2020). Correlation matrices are calculated to align the feature templates as in already established approaches. PIPs++ operates within a multi-frame temporal context, considering all frames in a temporal window or batch of size S to search for the target. After a match is found, the trajectory of the feature within the batch is updated with one position per frame, resulting in a single, continuous trajectory per

batch for that feature. The correlation matrices are recalculated iteratively to refine the tracking process. If tracking fails in one frame due to occlusion (e.g., fog), the trajectory of the feature can still be estimated by interpolating the position, provided that the feature reappears in subsequent frames within the batch (as long as the fog does not last longer than the batch size S; see Fig. 3b). To increase robustness, the model computes similarity using multiple templates per feature. This means that the model does not only rely on the feature's initial appearance in the first frame, but adapts to changes in appearance along the trajectory, i.e., throughout the sequence. If the feature's appearance changes, due to snow cover (Fig. 3d), lighting shifts, or rotation, the model can update templates along the trajectory, storing multiple representations (templates) for the same feature. This multi-template approach allows tracking to continue even with appearance variations caused by environmental factors like light changes, snow cover, or self-occlusion (e.g., rolling boulders). Thus, the model handles appearance changes flexibly and tracks the feature over time using these multiple templates. ~~For further information about the model architecture, we refer to Harley et al. (2022) and Zheng et al. (2023). The output of the model is a simple trajectory .txt file per frame (see Fig. 2, lower right), with the pixel coordinates of each tracked point.~~The output of the model is a simple trajectory .txt file per frame (see Fig. 2, lower right), with the pixel coordinates of each tracked point. This file can be ~~visualized~~visualised on the respective image for visual inspection (Fig. 3). If a regular grid is used, it is reset each time a new batch starts.



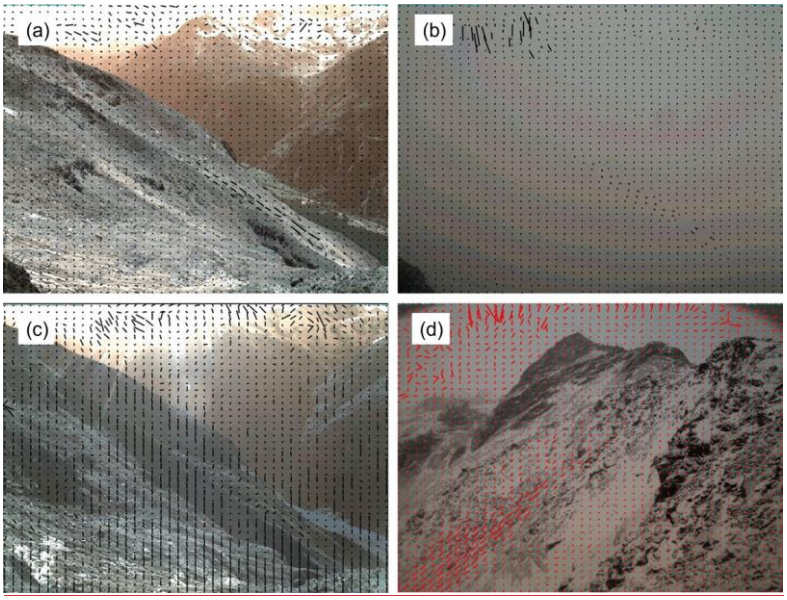


Figure 3: Some examples of velocity vectors as output from the PIPs++ model, a) view of Wcam05, clearly identifying the moving rock glacier body although its suboptimal viewing angle, b) same view occluded by fog but tracking is interfered from the image before and after, c) same view displaying a systematic shift in the stable areas due to camera movement, d) view of Wcam04 proving PIPs++ good performance in varying snow conditions.

3.2 From 2D to 3D: image-to-geometry registration using GIRAFFE

Image-to-geometry registration~~, implemented~~ in GIRAFFE~~, transfers 2D image data into 3D object space by aligning the camera image~~ aligns 2D images with a photorealistic synthetic image~~. This synthetic image is~~ rendered from a ~~coloured~~ 3D point cloud. This process enables the transfer of 2D image data into 3D object space by estimating the camera's pose within the point cloud's coordinate system. The following subsections aim to give a brief overview of the processing steps done in GIRAFFE, for more details we refer to the code documentation on GitHub.

3.2.1 Synthetic Image Generation

A synthetic image is rendered from an RGB-coloured 3D point cloud using an initial estimate of the ~~camera's original pose within the point cloud's 3D object coordinate system. The user must provide an approximation of the~~ camera's focal length and ~~the~~ pixel size ~~of the actual camera.~~. The ~~synthetic image will replicate the camera perspective of~~rendering simulates the real camera ~~image~~perspective while preserving ~~the~~ spatial information from the 3D point cloud. ~~Note, the colour information~~

Formatted: English (United States)

could also be inferred from intensity values, e.g., from a TLS/ALS point cloud, particularly when significant differences in intensity arise due to the varying reflectance properties of topographical features. However, in natural settings, such as the Grabengufer site, the intensity images typically lack sufficient unique features to facilitate alignment with camera images. Therefore, we utilise RGB-coloured 3D point clouds derived from UAV photogrammetry.

The rendering During the process involves projecting the , 3D points of the point cloud are projected onto a virtual image plane, assuming using a linear pinhole camera model and accounting for gaps and occlusions. Points closer to the camera contribute more information to the virtual image plane and are less likely to be occluded by other points along the depth direction. The . Frustum culling defines the view frustum, is determined using a method known as frustum culling (illustrated in Fig. 4). Despite variations in point size, gaps remain in the projected point cloud image. These gaps are addressed using a radiometric gap-filling algorithm based on 4), while a k-nearest neighbour (kNN) approach. Although the pixels generated addresses gaps —when no points fall into a pixel—through gap filling do not correspond to radiometric gap-filling. Note, in computer graphics, a frustum is the 3D volume visible through a camera or perspective projection, shaped like a pyramid with its top cut off, bounded by near and far clipping planes. Though these gap-filled pixels lack actual 3D information, they assist the matching algorithm in subsequent steps by facilitating correspondences image matching between the real and synthetic image.

Once the synthetic image is generated, it provides a photorealistic depiction of the 3D point cloud, matching approximately the perspective of the original true camera image, and retaining the original 3D information (except in areas filled radiometrically by the gap-filling process). .

3.2.2 Image Matching and Camera Pose Estimation

Using LightGlue, (Lindenberger et al., 2023), homologous points between the real and synthetic image are identified and matched, resulting in a list of corresponding 2D-3D image coordinates. AThereby, a kNN approach is applied again to the synthetic image to find the nearest neighbour, for each image coordinate from, the list that corresponds to anearest valid 3D point. By —using a default, a maximum distance criterion of 2.5 pixels is employed. This process produces a dataset of 2D image measurements from the real camera image, corresponding to 3D object points from the synthetic image. —.

These matched 3D object points can be treated as pseudo-control points to solve the camera pose, principal point location, and principal distance using an extended space resection approach formulated as a perspective-n-pose (PnP) problem. We employ the pinhole camera model and Levenberg-Marquardt optimisation implemented in OpenCV v4.10. When the distribution of 2D-3D point correspondences is uniform across the image (and sufficiently high to solve for more unknowns), this approach can also estimate lens distortion parameters. If the distribution is not suitable, lens distortion is neglected, which might lead to inaccuracies in referencing image measurements, particularly at the image margin. Pre-calibration of the camera may be performed because then images can be undistorted based on the calibrated interior orientation parameters before further processing in PIPs++. After determining the camera model of the real image - either linear (without distortion parameters) or non-linear (with distortion parameters) - including the intrinsic parameters and camera pose within the reference system of the

Formatted: English (United States)

~~3D point cloud, a connection is established between 3D points in the synthetic image and corresponding 2D points in the aligned real image. This connection can be used to identify discrete 3D coordinates for specific image measurements or to transfer thematic information from the image to the 3D points in the synthetic image, enabling additional data to be applied to the 3D point cloud for mapping purposes. This process can be considered a simplified form of back-projection.~~

~~The accuracy of the computed camera model, and consequently the precision of the 3D-referenced image points, is governed by two primary factors: the accuracy and density of the 3D point cloud, which serves as the source of pseudo-GCPs for space resection, and the spatial distribution of corresponding image observations, particularly when estimating the parameters of a non-linear camera model. To address this, we implemented an automatic distribution control mechanism, ensuring that the determination of the non-linear camera model is performed only if the pseudo-control point configuration meets the requirement of an adequate spatial distribution (i.e., points well-distributed across the entire image, including the edge regions). To achieve stable results for the exterior orientation parameters, the pseudo-control points must also be well-distributed in object space and should avoid collinear alignments or geometrically degenerate configurations, such as cylindrical or conical surfaces. In practice, when utilising a pseudo-control point field derived from a naturally occurring topography with thousands of potential 3D control points, the likelihood of such problematic constellations is negligible. Additionally, the distance criterion used to assign a 3D point to a 2D match with the real image in the synthetic image plays a critical role in the accuracy of the 2D-3D association. A smaller distance threshold results in higher precision for individual matches; however, it also reduces the number of available pseudo-control points, potentially affecting the robustness of subsequent space resection steps. Striking an appropriate balance between precision and the density of control points is thus essential for optimising the overall calibration process.~~

~~GIRAFFE enables the iterative execution of image-to-geometry registration and space resection, allowing for the gradual refinement of the (non-)linear camera model.~~These correspondences act as pseudo-control points to determine the camera pose and intrinsic parameters via an extended space resection approach, formulated as a perspective-n-pose (PnP) problem. Optimisation is performed using the Levenberg-Marquardt algorithm (implemented in OpenCV v4.10). If the spatial distribution of control points is sufficient, lens distortion parameters can also be estimated; otherwise, distortion is neglected.

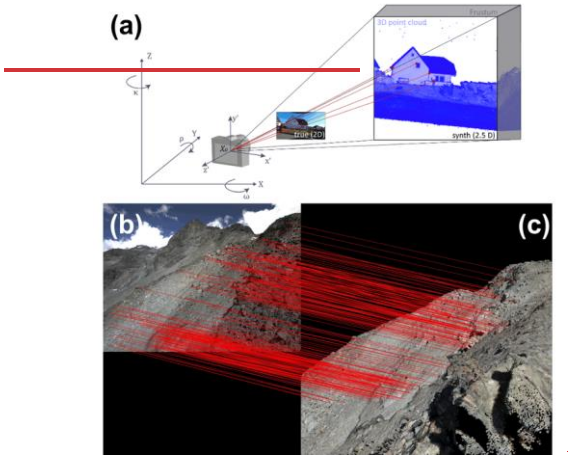
3.2.3 Accuracy Considerations

The accuracy of image-to-geometry registration depends on:

- **Point Cloud Quality:** Higher density and accuracy improve precision in space resection.
- **Spatial Distribution of Matches:** Uniformly distributed 2D-3D correspondences across the whole image enhance model stability, while poorly distributed points (e.g., along collinear alignments) can degrade results.
- **Distance Threshold for Matching:** A lower threshold (here set to 2.5 pixels) improves precision but reduces the number of valid matches, affecting robustness.

An automatic distribution control mechanism implemented in GIRAFFE ensures an adequate pseudo-control point configuration before estimating non-linear camera models. To achieve this, the image is divided into quadrants, ensuring a

balanced distribution of pseudo-control points across the image domain. Striking an appropriate balance between precision and the density of control points is also essential for optimising the overall calibration process. Moreover, GIRAFFE executes image-to-geometry registration iteratively, refining the camera model over multiple iterations. By default, ~~the software is configured to perform~~ three iterations~~, a value determined empirically as a balance between~~ are performed, balancing computational efficiency and ~~the likelihood of achieving~~ convergence ~~even with suboptimal initial parameter estimates (one~~reliability. Each iteration ~~takes between a few seconds and a minute, depending on the point number~~improves alignment, of the ~~used 3D point cloud). With each iteration, the~~ synthetic image ~~generated by GIRAFFE increasingly aligns with the perspective of~~ with the real camera image, improving both the number and spatial distribution of pseudo-control points. ~~This iterative approach enhances the stability and robustness of the~~ This leads to a more stable camera model adjustment ~~procedure while accommodating coarse initial guesses with higher reliability~~.



Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

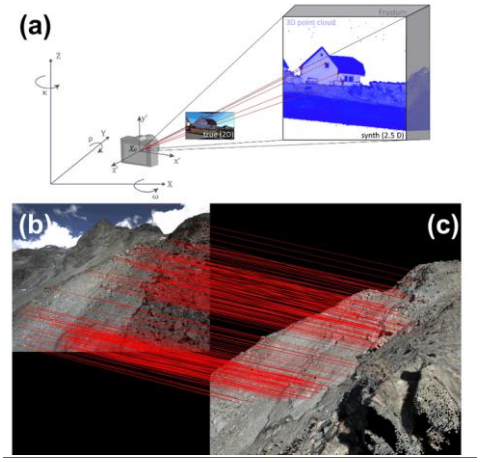


Figure 4: a) Visualisation of the image-to-geometry scaling (Elias et al., 2023), b) real 2D image from Wcam04 being matched (red lines) to c) the rendered synthetic image derived from the UAV data~~, displaying a good distribution.~~. Matches focus on the image centre (region of interest), ensuring strong alignment there, while lens distortions remain unmodeled since no matches are available at the edges, where distortion is expected to be greater.

3.3 Combining PIPs++ and GIRAFFE

3.3.1 Image Pre-Processing

We assumed that the poses of both webcams remained stable over the 19-week observation period. However, we noticed slight movements of Wcam04 caused by wind, which we corrected using a homography-based transformation. Thereby, the first image of the time-lapse sequence is considered the reference to which all subsequent images are registered using SIFT to estimate the points for homography calculation. This calculation is done using the RANdom SAmple Consensus (RANSAC) method to account for potential outliers, such as moving image features in the area of the rock glacier or landslide, which could otherwise introduce errors. The derived homography is then applied to perform a projective transformation of all consecutive images. In this way, the camera model ~~could~~can be considered consistent ~~for~~across all images ~~of~~in the analysed time-lapse sequence, i.e., the images appear as if they were captured from the same perspective. Accordingly, we employed GIRAFFE only to the reference image to ascertain the configuration of the image geometry and utilised the estimated camera model to reference all pixels tracked by PIPs++ throughout the sequence in object space.

In contrast to this image-based stabilisation, the camera model of each image of the time-lapse sequence could ~~also~~ have been calculated independently of the others via GIRAFFE and then applied to the respective PIPs++ image measurements. However,

this method was not employed due to the superior robustness of image-based stabilisation, particularly under poor image quality conditions, such as those caused by fog. Note that this type of image pre-processing is only necessary if the camera's extrinsic geometry appears to change. Otherwise, AI-based feature tracking can proceed directly from the initial images.

3.3.2 Image feature tracking and 3D referencing

We used a NVIDIA RTX A6000 GPU with 48 GB of memory to track 2000 points across a batch size of 19 weekly frames on 3MPx-large images before hitting computational limits. For a temporal window of 19 frames, the same 2000 features arewere tracked continuously before the grid was reset, after which tracking restarted with the reset grid as new input. Processing 400 images with this setup took approx. two minutes. For validation, specific points of interest, such as boulders with available GNSS/TS data, were tracked in addition to the grid points entered into PIPs++. These specific points were inserted into the script as pixel coordinates.

Initial camera specifications for the time-lapse webcams were available from EXIF metadata (i.e., focal length) and GNSS-measurements of the camera housings, which was used as approximation for the 3D coordinates of the camera projection centres. It isFor the orientation approximation, it was assumed that the camera iswas not tiled or rotated along the lateral axes of the reference system. The viewing angle related to the geographic view direction (NESW) was determined by compass. The pitch and roll angles were assumed to be 90 degrees (looking parallel to the XY-plane) and zero degrees (no rotation around the viewing axis). Although our initial estimates for the camera model from GNSS, compass and image EXIF data were already sufficient, we used three iterations during the optimisation. The image points tracked by PIPs++ were referenced registering the reference image of the time lapse sequence to the 3D object space. The camera distortion coefficients could not be determined due to an unsuitable distribution of pseudo-control point observations as large portions of the image were covered by sky and topographyterrain not representedcaptured in the 3D UAV data. (e.g. Fig. 4).

The resulting lists of 3D coordinates per timestamp matched, matching the PIPs++ feature points and, were accompanied by an identifier that allowed the 3D coordinates to be uniquely assigned across the timestamps to estimate, enabling the estimation of the per-camera 3D trajectories. Finally, mean velocities in metres per day were calculated dividing the displacement between two consecutive 3D points by the time interval, (in this case, one week), resulting in a total of 18 weekly velocity measures per tracked image point.

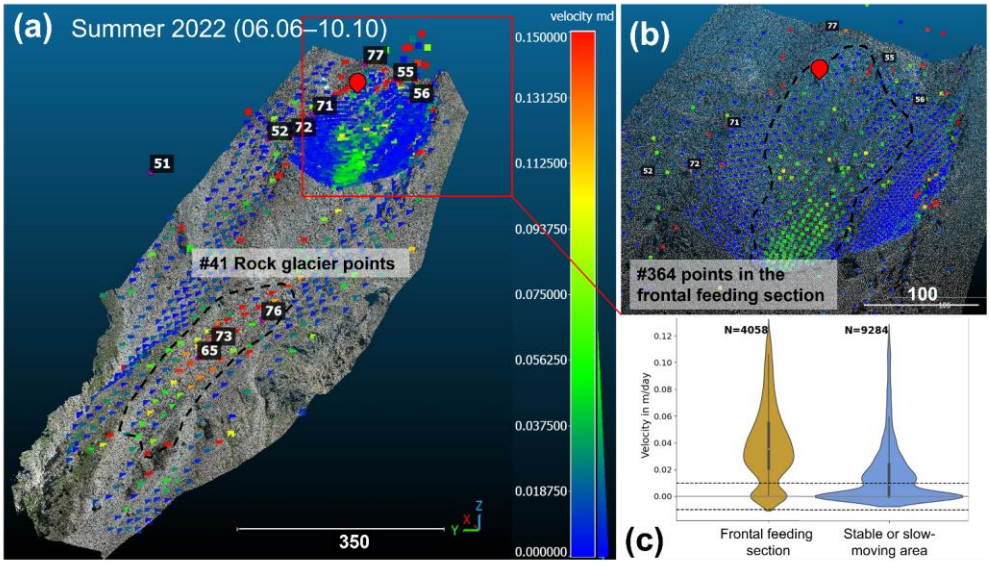
4. Results

4.1 Geomorphic results

Our image-based 3D trajectory and velocity measurements were based on two subsets from time-lapse cameras of a permafrost-affected landslide and its frontal section feeding into a rock glacier. The results for both investigated region of interests for the entire summer period (06.06. until 10.10.2022, 19 weekly images) are shown in FigureFig. 5a. As expected from landform displacements, the results show a largely homogenous flow velocity field with faster velocities in the upper area ofon the rock

glacier and in the lowermost part of the investigated feeding area ~~of~~from the landslide. The feeding area continues downward and is not captured by neither Wcam04 or Wcam05 as evidenced by Fig. 5a. An average movement of 0.03 m d^{-1} (0.21 m of absolute movement between weekly images) was observed across the entire investigated feeding section, while in stable and slow-moving areas, movement was typically below 0.01 m d^{-1} (Fig. 5c). For the rock glacier, the average movement during the summer period was 0.10 m d^{-1} (0.~~21~~70 m of absolute movement between weekly images), with surrounding stable or slow-moving areas showing ~~movement~~displacement rate generally below 0.05 m $d^{~~-1~~.-1}$ (0.35 m of absolute weekly movement). The feeding section exhibits significant spatial variability~~, with~~: the lowermost part moves fastest (Fig. 5a, 5b), and, in addition, certain large boulders ~~moving faster~~travel more quickly than their immediate surroundings~~, ~~unlike the rock glacier, where movement is more uniform.

This revealed variability must be considered when using discrete points from GNSS and TS measurements to characterize landform movement. While the two TS points, used to validate the Wcam05 velocity estimates, show a similar trend to the overall movement of the rock glacier ~~using~~ (Fig. 7b), the permanent GNSS system reveals a distinct different movement pattern, with a marked acceleration in velocity at the end of July that is not reflected in the feeding section's overall movement (Fig. 7a).



~~The uppermost landslide area and the lower rock glacier are presumably driven primarily by permafrost creep. However, the driving mechanisms of the frontal feeding section observed by Wcam04 remain unknown. To better understand the factors controlling velocity changes and the overall behaviour of these landforms, further investigation into environmental drivers such as rainfall and temperature is necessary – especially considering that the summer of 2022 was exceptionally warm and dry, with consecutive heat waves. However, this is beyond the scope of this study.~~

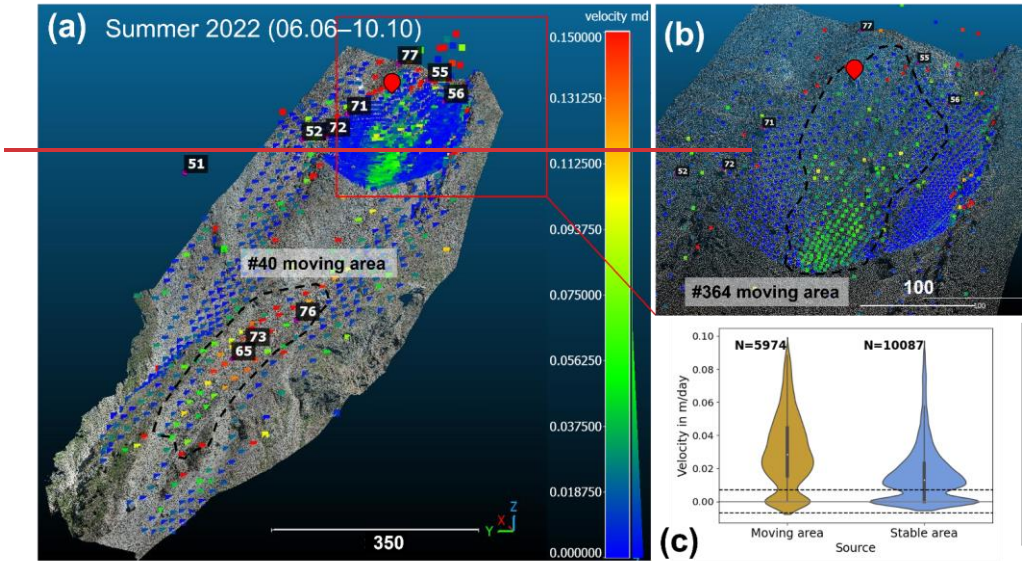


Figure 5: a) Velocity results (in m d^{-1}) over the entire summer period (127 days) for both the frontal feeding section (upper area, observed by Wcam04) and the rock glacier (lower area, observed by Wcam05), displayed on UAV point cloud data, including total station-measured points (black and white labels) and the permanent GNSS installation (marked with a red location pin), b) close-up of the frontal feeding section (Wcam04), c) plot of the weekly velocity distribution for both the ~~moving~~frontal feeding section and the surrounding stable or slow-moving areas (data from Wcam04), showing a clear noise level between ±0.007 m d^{-1} in both regions. Dotted lines indicate the moving areas in this study. Points in blue display ~~stable~~ areas moving less than about 0.02 m d^{-1} for the considered timeframe of 127 days.

4.2 Error Assessment and Validation

The accuracy of the 3D trajectories and velocity derivations essentially depends on three factors: The reliability of the PIPs++ feature tracking, the determination of the camera model by image-to-geometry registration, and the resolution and accuracy of the 3D point cloud. To assess the accuracy potential, a comprehensive analysis of all potential sources of error was necessary. However, due to the numerous algorithms, a classic error analysis is difficult, especially because AI-based tools such as PIPs++ often do not provide any information on reliability. We have included the following methods in our evaluations:

- Level of Detection (LoD): Assessing pseudo-movements in stable areas resulting from an accumulated error of PIPs++ measurements and its translation into object space define a so-called LoD, describing the threshold of significant movements that could be detected with our method.
- Theoretical accuracy analysis based on the reprojection results of GIRAFFE.
- Validation with ground truth data by comparing our ~~measured~~ 3D trajectories based on time-lapse images with GNSS and TS data serving as ground truth at selected 3D objects.

Formatted: Outline numbered + Level: 1 + Numbering St Bullet + Aligned at: 0 cm + Tab after: 0 cm + Indent at: 0.63 cm

4.2.1 Level of Detection

The LoD was experimentally determined analysing the detected 3D movements in presumably stable or slow-moving areas (Fig. 5c). For Wcam04, the designated ~~stable~~ area lies within the broader zone of the deep-seated landslide (Fig. 1a, blue polygon). While this area moves significantly slower than most of the frontal feeding section, its stability is relative and only valid when considered over a short timeframe. Here we detected a LoD of about 0.007 m d^{-1} or an absolute movement of +/- 5 cm between weekly time-lapse images. For Wcam05, the LoD was around 0.02 m d^{-1} or an absolute movement of +/- 14 cm between weekly image frames, which is consistent with the larger camera-to-object distance (Table 1).

4.2.2 Theoretical accuracy analysis

For both the stable (Wcam05) and software-stabilised (Wcam04) time-lapse image series, the GIRAFFE software was employed once to estimate the linear camera model of the reference image. To assess the precision of the image-to-geometry registration, GIRAFFE calculates the overall root mean square error (RMSE). This is achieved by projecting the 3D pseudo-control points used for space resection back into the image using the estimated camera model and measuring the mean distances between the original image observations and their projected positions. The RMSE for Wcam04 and Wcam05 was 4.8 pixels and 4.5 pixels, respectively.

~~Based on the initial parameters of Wcam04 and Wcam05, we perform a~~A theoretical accuracy analysis ~~considering the RMSE values. For~~shows that for Wcam04, ~~the~~with a camera-to-object distance ~~ranges between~~ of 100–350 m and ~~350 m, resulting in an image scale ranging from 1:18,000 to 1:65,000. An error~~ GSDs of ~~4.8 pixels in the image corresponds, given a webcam~~ 0.06–0.21 m (see Table 2), a 4.8-pixel ~~pitch of 3.1 µm,~~error leads to ~~a~~ positional ~~error in object space between 0.3 cm and 1.0 m, respectively. In contrast, for Wcam05, the camera-to-object distance is between 700 m and 1,100 m, leading to a varying~~

~~image scale of 1:130,000 to 1:200,000. Under the same pixel size assumption,~~ errors of 0.3–1.0 m. For Wcam05, at 700–1,100 m with GSDs of 0.40–0.62 m, a 4.5-pixel error ~~translates to an object space error of 1.8 m to 2.8 m, respectively. These~~ results ~~highlight that as the camera~~in 1.8–2.8 m errors. Increased distance ~~increases, minor uncertainties in~~ amplifies errors from camera geometry ~~modelling can lead to significant errors in the measurements in object space. Consequently,~~ uncertainties, leading to higher inaccuracies ~~are expected~~ for Wcam05.

~~To mitigate errors resulting from the estimated camera model, particularly at smaller scales,~~ GIRAFFE ~~by default applies an~~mitigates this with outlier filtering ~~of the registered image measurements in object space. Only 3D coordinates with a~~, retaining 3D points with reprojection ~~error below 2.0 pixels were retained to ensure the inclusion of features reliably located in the 3D object space based on the estimated camera model. A translation of the maximum permitted reprojection error of~~errors under 2.~~0 pixels into object space reveals the maximum expected error of the 3D points, providing the base of the 3D trajectories, which varies in accordance with the image scale, i.e. between~~ 5 pixels, corresponding to object space errors of 0.10 ~~m and~~ –0.40 m for Wcam04 and 1.0 ~~m and~~ –1.2 m for Wcam05.

4.2.3 Validation with ground truth data

To validate our method ~~with~~, we used GNSS and TS ground truth data ~~from GNSS and TS, we~~ for selected ~~individual~~ 3D objects visible in the camera's field of view ~~and excluding those with movements close to or~~. Objects with movement below the LoD were excluded. Due to logistical and safety constraints, the ground truth data does not cover the fast-moving areas of the frontal feeding section. However, ~~they are~~ground truth data is available in the upper landslide area (Fig. 5b). ~~At least~~ For Wcam05, two moving objects on the rock glacier were suitable for direct validation, ~~both located on the rock glacier observed by Wcam05. And~~while for Wcam04 ~~one~~, a permanently installed GNSS antenna ~~in the view was available for validation. However, no further~~provided a reference. Other potential objects were ~~available because the GNSS and TS data was~~either outside the camera's field of view or ~~the image-based results were~~had movements below the LoD. ~~For the two object points visible in Wcam05, 3D trajectories across five timestamps were reconstructed, resulting in eight measurements suitable to validate the 3D trajectories measurements from PIPs++ and GIRAFFE.~~ Given the challenge of precisely identifying the measured ground truth point in the image data, we tracked 5x5 patches, i.e. 25 pixels, around the presumably image position of the respective object~~, i.e. 25 pixels~~ (Fig. 6a).

~~We fitted 3D lines as~~ For the two objects in Wcam05, 3D trajectories ~~instead of using individual 3D points for~~were reconstructed across five monthly timestamps, yielding eight validation ~~because this offers the advantage that we can reduce the impact of~~measurements. To improve robustness against noise and outliers, we validated against fitted 3D trajectory lines using the detected positions in the ~~data to investigate systematic errors. Thereby, we need to assume a linear movement, which however was apparent from~~five timestamps rather than the ~~reference data provided by TS. The~~ individual 3D point coordinates. This approach assumes linear motion, which was supported by TS reference data.

We quantified alignment quality by calculating the standard deviation of ~~the~~ perpendicular distances between ~~the~~measured points and the fitted ~~3D line was calculated. These distances quantify how well the points align with the fitted line for both the~~

Formatted: English (United Kingdom)
Formatted: English (United Kingdom)
Formatted: English (United Kingdom)
Formatted: English (United Kingdom)
Formatted: English (United States)
Formatted: English (United States)

reference measurements and the trajectory lines. The PIPs++/GIRAFFE measurements. The measurements from PIPs++/GIRAFFE for Wcam05 (Fig. 6b) exhibit a considerably higher level ofshowed greater noise compared to those fromthan the TS reference data (Fig. 6c,d7). However, if we consider the spatial (XYZ) differences of the vectors and the Euclidean distancedistances between the vectors, we see that the total discrepancy of the directions and thus of the approximated 3D trajectories in 3D space is of the order ofvectors indicated only minor discrepancies, within a few decimetres. We conclude that the measured 3D trajectories obtained using PIPs++ and image-to-geometry registration are subject to random decimeters. This suggests that fluctuations, ruling out the presence of in measured 3D trajectories are random rather than systematic errors. Furthermore, the directionalDirectional differences between reference and measured trajectories and our measurements indicate a variation ofvaried by only a few degrees (Table 2). These discrepancies are largely attributablelikely due to the difficult image-to-object imaging geometry challenges (linked to camera configuration and the resulting difficult grinding cut geometries.

detailed in Table 1) and occlusion effects caused by the camera perspective, particularly in the presence of boulders. For Wcam04, the points exhibited minimal movement during the observation period, which of tracked points limited the ability to fit straight lines to the 3D trajectories. However, the basic distance between the reference and image measurements was determinedtrajectory fitting. Instead, accuracy was estimated using the centre of gravity of the respective 3D point sets. This approach enabled an estimation of the general accuracy of fit based on the 3D coordinates. A relatively large A 1.3 m offset of 1.3 m was observed (Table 2for Wcam04 (Table 2), attributed towhich is consistent with the GNSS antenna being positioned behind a boulder and therefore not visible in the image. Consequently, we had to track the antenna's 1.5–2 m elevation above the surface and its position on the opposite side of the boulder for validation instead of the antenna's actual position. Given that. Because the boulder is approximatelyroughly the size of a small car, this offset is likely due to the discrepancy in tracked position. However, the sametruck, we performed validation on the opposite side rather than at the true antenna location. Despite this, velocity pattern could be observed in both the patterns in GNSS and PIPs++/GIRAFFE data and the tracked point in PIPs++/GIRAFFE, though to a minor extentremained consistent (Fig. 78a).

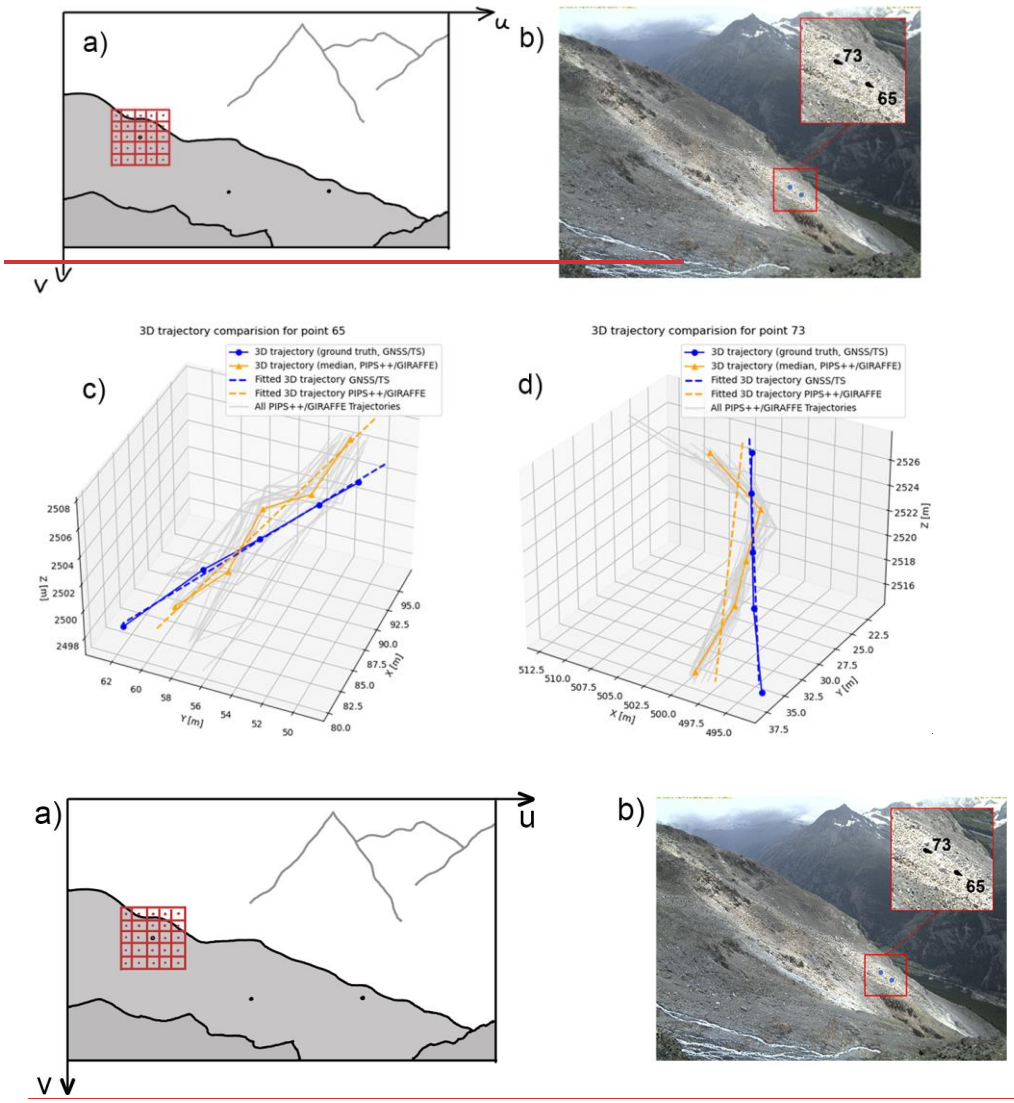
Table 2. Calculation of fitted 3D trajectories of significant objects tracked over 109 days with GNSS (one point, visible in Wcam04) and over 118 days with TS (two points, visible in Wcam05) as well as with PIPs++ and GIRAFFE in the associated images, taken on the same days as the GNSS/TS ground truth measurements. SD = Standard deviation. Diff = Difference

ID	DateValidation period	SD Linear line fit GNSS/ TS [m]	SD linear line fit PIPs++/-++/GIRAFFE [m]	Diff XOffset of the fitted lines [m]	Diff. Y [m]	Diff. Z [m]	Diff total movement [m]	DiffOffset in line direction [°]
GNSS	18 timestamps (End of June – Mid of October 2022)	-	-		1.28			-

Formatted: Tab stops: Not at 7.25 cm
Formatted: English (United Kingdom)
Formatted: English (United Kingdom)
Formatted: Font color: Auto
Formatted: Font color: Auto
Formatted: Font color: Auto
Formatted: Font color: Auto
Formatted: Font color: Auto
Formatted Table
Formatted: Centered
Deleted Cells
Deleted Cells
Deleted Cells
Formatted: Centered

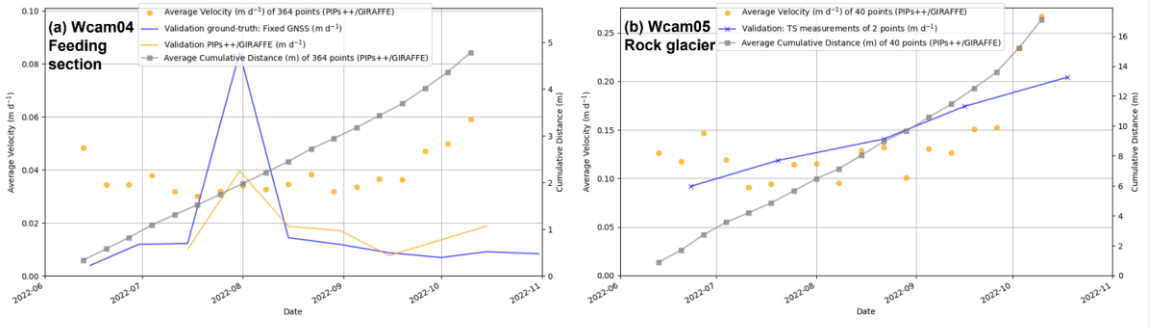
TS_65	~~23/06/2022~~ ~~20/07/2022~~ ~~22/08/2022~~ ~~16/09/2022~~ ~~18/10/2022~~Five monthly timestamps (Mid-June – Mid October 2022)	0.08	0.29	~~0.14~~	~~0.19~~	~~0.05~~			0.25		14.38
TS_73	~~23/06/2022~~ ~~20/07/2022~~ ~~22/08/2022~~ ~~16/09/2022~~ ~~18/10/2022~~	0.05	0.59	~~-0.10~~	~~-0.09~~	~~-0.03~~			0.14		8.23
~~**WCam05 - Absolute Mean Error (ME) [m]**~~		~~-~~	~~-~~				~~**0.12**~~	~~**0.14**~~	~~**0.04**~~	~~**0.19**~~	~~**11.3**~~
~~GNSS~~	~~30/06/2022~~ ~~15/07/2022~~ ~~31/07/2022~~ ~~15/08/2022~~ ~~31/08/2022~~ ~~15/09/2022~~ ~~01/10/2022~~ ~~15/10/2022~~ ~~31/10/2022~~	~~-~~	~~-~~				~~0.86~~	~~-0.94~~	~~0.01~~	~~1.28~~	~~-~~

Formatted: Left
Deleted Cells
Formatted: Centered
Formatted Table
Deleted Cells
Merged Cells
Deleted Cells
Formatted: Centered
Formatted: Centered
Formatted: Centered
Formatted: English (United Kingdom)



505 **Figure 6: Validation principle illustrated through direct comparison of two measured point trajectories (ground truth) and their corresponding trajectories derived from time-lapse images. Illustration of the tracking principle used for 3D trajectory validation.**

a) A 5x5 pixel grid around each ground truth ~~point~~ captures local movement. b) Location of the two validation points on the rock glacier observed by Wcam05. ~~**c) Trajectory of point 65: ground truth (blue), time-lapse-derived trajectory (median of 25 trajectories, orange), and individual trajectories from the 5x5 grid (grey) as well as the fitted 3D trajectories (dashed lines). d) Same data presented for point 73. For visualisation purposes, we applied an XY-offset in metres of [2.628*10^6, 1.105*10^6].**~~



Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

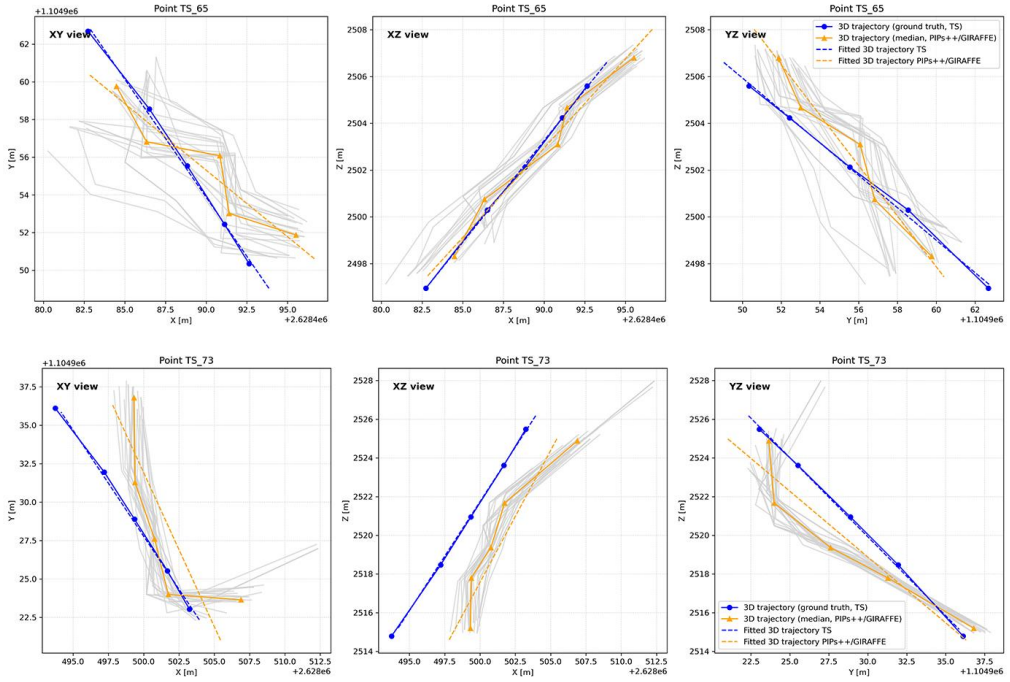


Figure 7: Comparison of 3D trajectory estimations from theodolite measurements (TS) and tracking data (PIPs++/GIRAFFE) across different 2D projections for two points (named 65 and 73). The XY, XZ, and YZ views show individual point trajectories with their respectively fitted 3D lines. Light grey lines represent all time-lapse trajectories, while blue and orange markers correspond to theodolite and median of the tracking data, respectively. Dashed lines indicate the best-fit linear 3D trajectory for each method.

Formatted: Strong, Font: 10 pt, Not Bold

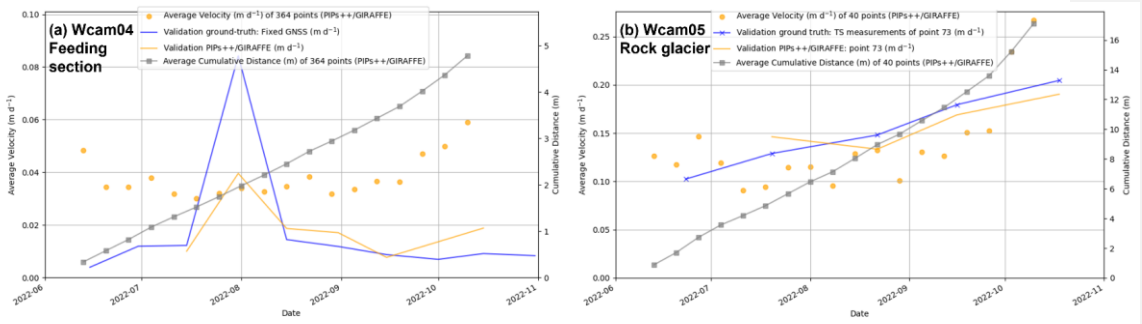


Figure 8: Graphs showing weekly velocities and cumulative distances~~. Panel~~ of tracked points measured using the PIPs++/GIRAFFE workflow, including validation against ground truth data. a) ~~displays data for the feeding~~Feeding section (Wcam04) and ~~panel~~ b) ~~for the rock~~Rock glacier (Wcam05) during ~~the~~ summer ~~of~~ 2022, derived from time-lapse imagery. Note that the data from October 26 is excluded from ~~graph~~ a) due to foggy conditions and was interpolated by the model. Validation points, marked in ~~red, include~~a solid blue line, represent measurements from ~~the~~a permanent GNSS antenna installed on a large boulder in the upper feeding area (a) and ~~two~~one theodolite ~~points~~point on the rock glacier (b). The corresponding tracked points by PIPs++/GIRAFFE is displayed in a solid orange line.

5. Discussion and Limitations

5.1 General performance of the proposed workflow

The above-described methodology has the potential to tackle entire time-lapse image datasets to determine landform velocities, in this case creeping permafrost landforms. The results of our workflow show a good agreement with GNSS and TS theodolite measurements considering the low-grade camera's used (Table 2, ~~Figure 6~~Fig. 7), proving our method to be reliable, robust and fast for creating a better spatial (Fig. 5) and temporal coverage (Fig. ~~7~~8) of the landform's displacement. Our pilot study demonstrates that significant motion and velocity information can be rapidly ~~and easily~~ extracted using AI-based methods from a basic, cost-effective device like a single webcam, greatly enhancing temporal acquisition frequency without the need for a camera array. Spatial heterogeneity of landform movement is evident as well, including instances where larger boulders move faster, seemingly 'surfing' on the main landslide body. This phenomenon, illustrated in ~~Figure 7~~Fig. 8a, is supported by in-situ GNSS data, which indicates faster movement compared to the overall landform. If this pattern were seen across the entire landform, it could suggest motion linked to permafrost creep or specific movement in the active layer, often triggered by significant water input from snowmelt or rainfall (RGIK, 2023). However, the movement ~~seems~~is mostly restricted to large

boulders, possibly pointing to a gravitational origin due to the steep terrain. The uppermost landslide area and the lower rock glacier are presumably driven primarily by permafrost creep. However, the driving mechanisms of the frontal feeding section observed by Wcam04 remain unknown. To better understand the factors controlling velocity changes and the overall behaviour of these landforms, further investigation into environmental drivers such as rainfall and temperature is necessary – especially considering the exceptionally warm and dry summer of 2022, which featured consecutive heat waves. Our developed method provides a framework for such future research, enabling the analysis needed to address these open questions.

Although our proof-of-concept did not achieve the millimetre to centimetre ~~to millimetre~~ accuracy of GNSS and TS measurements, we were still able to detect absolute displacements of 5 cm and 14 cm between consecutive frames (in our case weekly) using the setups of Wcam04 and Wcam05, respectively (Table 1). The measurements at the rock glacier (Wcam05) have a significantly lower spatial resolution (~~40~~41 vs. 364 points tracked in the feeding area) due to the greater distance of the webcam from the area of interest and its suboptimal viewing angle (Fig. 1b). ~~This~~The increased distance also made the measurements more sensitive to slight camera movements, which, along with the lower LoD, likely contributed to some of the higher variation in ~~measured~~computed rock glacier velocity (Fig. ~~6b~~7b). The output of our workflow ~~in a sense~~ yields a sparse point cloud for every image frame in the sequence with absolute distance and velocity information as a scalar field, as visualised in ~~Figure~~Fig. 5. This provides a great amount of spatial and temporal data in a manageable file format suitable for big data.

The superior performance of the PIPs++ model stems from tracking multiple timesteps jointly instead of frame-by-frame, enhancing temporal smoothness, coherency, and improving flow estimation accuracy (Hur and Roth, 2020). This makes PIPs++ especially suitable for environmental applications, where for example changes in light conditions are a common problem. Moreover, PIPs++ is trained on a very large and diverse artificial dataset PointOdyssey (Zheng et al., 2023), including rendered dynamic fog to account for (partial) occlusion and realistic in- and outdoor scenes. This is entirely absent from other synthetic datasets like the FlyingChairs dataset, which was utilised to train models such as FlowFormer (Huang et al., 2022) or GMFlow (Xu et al., 2022), i.e. transformer-based models, and justifies our use of PIPs++ as a tracker in this proof-of-concept without the need of re-training the model with a sample of our own data. While previous research, using monoscopic images to track a landslide (Travelletti et al., 2012) and a rock glacier (Kenner et al., 2018), was prone to mismatches because of its frame-by-frame strategy, our approach surpasses this limitation and at the same time makes it possible to process longer time periods and handle big data collected by hourly webcams more adequately. By tracking points in stable areas using either hand-crafted point operators like SIFT or AI-based methods, shifts in camera position, a common problem in long time-lapse imagery sequences, can be corrected to some degree to stabilise the image sequence by software. When using PIPs++ and GIRAFFE, two options exist to deal with camera movements: a) image stabilisation by software or b) individual referencing of each time-lapse image, i.e. run GIRAFFE image-per-image. The correction based on image pixels is frequently accomplished through the calculation of a homography to match a reference image. However, ~~this~~because the method ~~is~~assumes a planar scene, it can only ~~capable of compensating to a limited extent~~partially compensate for perspective

distortions ~~resulting from changes in perspective due to~~caused by camera movements. Consequently, it is ~~most~~ suitable for smaller movements. In cases of stronger movements, we recommend calculating the camera model ~~individually~~ for each individual image of the image sequence~~, whereby the relationship to the point cloud is calculated individually~~. As a result, even major changes in perspective should be handled adequately when translating the image measurement into object space.

5.2 Limitations

As PIPs++ still relies on appearance-matching cues similar to traditional methods, it remains sensitive to abrupt changes in appearance, such as substantial variations in snow cover or alterations in surface morphology due to e.g., extensive rockfalls. This issue was also highlighted as a major limitation by Kenner et al. (2018). However, because the algorithm can quickly process large amounts of data, we can leverage the full temporal resolution, allowing tracking to succeed as long as snow cover changes gradually, as shown in ~~Figure~~Fig. 3d.

Another important limitation of our approach is that PIPs++ does not perform well in detecting movements for every pixel in the image due to the use of a regular grid compared to other frame-to-frame approaches. Currently, a regular grid size of 2000 points per image is used, limited by computational power. While PIPs++ works well with the low-resolution images in this study, higher-resolution images quickly reach the limits of our available computational resources. One major limitation is the need for specific and expensive computational setups (e.g., an NVIDIA RTX A6000 GPU with 48GB). Even with these resources, increasing the temporal window, image resolution, and tracked points can quickly hit the limits. Considering appropriate GPU and CPU computational power, around 15 min was needed to process the example dataset of this paper in an end-to-end fashion, from raw time-lapse images to velocity graphs. The majority of this time is needed to scale the output of the PIPs++ model with GIRAFFE, which is highly dependent on the size of the 3D point cloud.

Ideally, the 3D point cloud used in GIRAFFE for image-to-geometry registration reflects the imaging situation at the time of the time-lapse image measurements, thereby avoiding intersection errors resulting from significant changes in surface topography ~~between the image acquisition and the measurement of the reference data set (Fig. 8~~(Fig. 9). Despite the temporal gap of approximately one year between the time-lapse image subsets and the 3D data, we assume that the overall topography has remained largely unchanged. Moreover, significant errors arising from potential discrepancies between the image content and the 3D surface are expected to be rejected in the outlier analyses in GIRAFFE. Such mismatches are unlikely to align with the estimated camera model, as the model calculation relies on pseudo-control points that are more evenly distributed and exhibit strong agreement between the image content and the object surface. Nevertheless, minor inaccuracies caused by isolated structural changes (e.g., displacement of boulders) cannot be entirely ruled out. Such discrepancies can only be identified when analysing the movement trajectories as a whole. Furthermore, the quality of the image-based 3D trajectories is heavily correlated to the quality of the ~~UAV~~point cloud data.

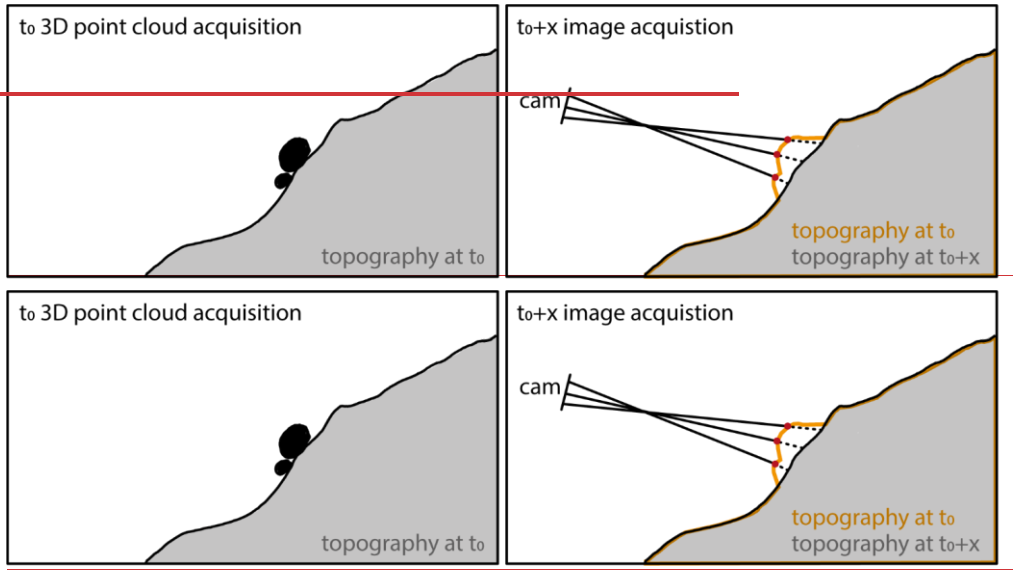


Figure ~~8~~9: Illustration of the georeferencing validity issue when using a 3D point cloud acquired at time t_0 to transfer image measurements from a time-lapse image captured after significant topographic changes. In this example, boulders present during the initial 3D data acquisition are since been gone. Consequently, image rays intersect the outdated modelled topography, resulting in erroneous coordinates. These errors propagate into derived 3D trajectories and velocities. The greater the topographic changes between the reference data and affected image measure, the larger the error. Therefore, the reference topography should ideally reflect the conditions at the time of image acquisition.

Additionally, a suboptimal oblique viewing angle, such as with Wcam05 (Fig. 1b), can complicate the matching process, as the rendered view will be limited to a narrow strip. This can affect the distribution of matches and compromise the estimation of the camera's interior orientation and lens distortions (Elias et al., 2019).

There is still potential to optimise the proposed workflow to achieve full automation for processing entire image sequences. This includes automatically detecting moving and stable areas, which currently relies on user defined criteria and thus depends on prior site knowledge. Additionally, we need to conduct further investigations into the in-field calibration of intrinsic camera parameters, especially the image distortion, and their stability over time (Elias et al., 2020). ~~This analysis aims to determine if it significantly enhances the quality of scaled results, similar to the benefits that are expected when using multiple epochs of UAV data for updating the camera calibration.~~ Improvements in 2D to 3D tools (like GIRAFFE) ~~could~~ also ~~benefit~~benefits

other disciplines, such as the quantitative analysis of historical terrestrial photographs for mapping historical ~~rockfall~~rockfalls (Wegner et al., 2023) and changes in glacier forefields (Altmann et al., 2020).

~~Moreover, additional research is necessary to evaluate whether AI-based tracking algorithms like PIPs++ could effectively track landform velocities using aerial datasets such as consecutive UAV orthophotos or satellite imagery. Furthermore, the~~The developed workflow could prove valuable for analysing monoscopic time-lapse image sequences of other dynamic processes, such as lava flows (James and Robson, 2014), solifluction and gelifluction movements (Matsuoka, 2014), and flow velocities in rivers (Eltner et al., 2020; Stumpf et al., 2016). Given the prevalence of time-lapse camera data collection, a rapid and efficient method for automatically processing such extensive datasets holds significant scientific relevance. Furthermore, the fast and robust processing of the time-lapse imagery makes it possible to function as an early-warning system when processing can be carried out in near real-time, as indicated by Kenner et al. (2018).

6 Conclusion

This proof-of-concept demonstrates the potential of AI-based algorithms for tracking and matching points to improve motion estimations in time-lapse image sequences of a mountain landscape. Two fast-moving alpine landforms - a landslide and a rock glacier at the Grabengufer site in Switzerland - were selected as the pilot study area. The initial results presented in this paper show that robust and reliable velocity information can be quickly derived with minimal input data and user intervention. Our pilot study opens the door to processing entire image datasets to reveal spatio-temporal patterns that traditional monitoring methods have previously overlooked, due to their limited spatial or temporal resolution and the inadequate computational and algorithmic power to handle large image datasets.

The PIPs++ model, used for tracking features in image sequences, excels in widening the temporal window and includes a template-update mechanism that allows for changes in feature appearance. Its key advantage is its ability to accurately estimate occluded trajectories within the temporal frame, avoiding suboptimal matches and enhancing tracking accuracy, making it especially robust for environmental applications and eliminates the need for filtering blurry or foggy images as a pre-processing step. Additionally, the model's rapid performance, processing 400 images in two minutes to track features through a temporal window of 19 frames, is promising for handling large datasets and developing early-warning systems.

The image-to-geometry approach, implemented in GIRAFFE, provides an accurate way to scale the 2D image data into 3D object space, even under suboptimal camera viewing angles and distance to the area of interest. ~~Future research will focus on in-field camera calibration and evaluating the potential benefits of using multiple epochs of UAV data.~~

This paper represents an important step forward in using monoscopic cameras and leveraging previously captured data that have not been processed automatically with metric values before. By significantly enhancing temporal acquisition frequency using basic time-lapse imagery ~~with zero cost~~, we can achieve a level of data resolution that would be expensive with in-situ differential ~~Global Navigation Satellite System (~~GNSS~~)~~ and georeferenced ~~Total Station (~~TS~~)~~ measurements or ~~Uncrewed Aerial Vehicles~~UAV and ~~Terrestrial~~TLS or ~~Airborne Laser Scanning~~ALS methods. ~~Validated by discrete TS and in-situ GNSS~~

~~measurements, our~~Our approach provides a spatially continuous understanding of landform movement. It allows data acquisition in areas where in-situ measurements are impractical due to logistical and safety constraints, and where other remote sensing techniques fail due to high landform displacements. Furthermore, depending on the image resolution, distance to the landform, and its velocity, our approach can achieve a sub-seasonal resolution of velocity information with an accuracy of several cm.

This study introduces a new open-source tool for scientists to automatically extract metric information from existing webcam datasets, extending its applicability to ~~various phenomena, including lava flows, water flows, and creeping landforms such as landslides, rock glaciers, and solifluction lobes~~further environmental process observations.

Code availability

Part of this work is based on existing algorithms, available under https://github.com/aharley/pips2 for PIPs++ tracker and https://github.com/cvg/LightGlue for LightGlue matching algorithm. GIRAFFE is available under https://github.com/mel-ias/GIRAFFE. The code used for this proof-of-concept is partly available under https://github.com/hannehendrickx/pips_env. Please follow this repository to receive further updates. All code is available upon reasonable request to the corresponding author. We encourage people with similar monoscopic time-lapse image datasets to reach out to the authors.

Data availability

The sample images and corresponding trajectory files used in this study are available at: https://github.com/hannehendrickx/pips_env and at Zenodo (doi:10.5281/zenodo.14260180) along with sample files that need to be set up according to the GIRAFFE documentation in order to run the tool**.** For an up-to-date overview of the sites, including the latest webcam images, please visit the following pages: Grabengufer Landslide and Grabengufer Rock Glacier.

Author contribution

HH, XB, ME and AE designed the research. HH, XB developed the Python code for running PIPs++ on the time-lapse images. ME developed GIRAFFE for image-to-geometry registration and the python code for stabilising image sequences. HH, ME prepared and processed data, made all the figures and wrote the initial version of the paper. RD conducted field investigations and provided field specific knowledge to aid the writing process. All authors improved and revised the initial manuscript.

Competing interests

The authors declare no conflicts of interests.

Acknowledgments

The authors thank Sebastian Pfaehler (University of Fribourg, CH) for capturing and processing the UAV data, and Sebastian Summermatter and Martin Volken (Geomatik AG / Planax AG) for the collection and processing of the total station data. We acknowledge the support of the municipality of Randa and the Canton of Valais (Service for Natural Hazards) in facilitating the field investigations. Geosat SA is acknowledged for setting up the in-situ GNSS devices and handling data processing, and we thank the Canton of Valais for providing access to the in-situ GNSS data. We also acknowledge Idelec SA for their assistance with webcam maintenance. We would like to thank the numerous collaborators at the University of Fribourg for their contributions to the data processing, webcam maintenance, and image transfer and archiving. Air Zermatt is also thanked for ensuring safe transport to and from the site.

ChatGPT was utilised as a tool in this study to assist with coding and to enhance the English language quality of the manuscript through text input. It was not employed to generate original text during the research process.

References

Altmann, M., Piermattei, L., Haas, F., Heckmann, T., Fleischer, F., Rom, J., Betz-Nutz, S., Knoflach, B., Müller, S., Ramskogler, K., Pfeiffer, M., Hofmeister, F., Ressl, C., and Becht, M.: Long-Term Changes of Morphodynamics on Little Ice Age Lateral Moraines and the Resulting Sediment Transfer into Mountain Streams in the Upper Kauner Valley, Austria, Water 2020, Vol. 12, Page 3375, 12, 3375, https://doi.org/10.3390/W12123375, 2020.

~~Glaciorisk - L'innovation digitale au service des dangers naturels: https://www.glaciorisk.ch/, last access: 17 July 2024.~~

Barsch, D.: Rockglaciers, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-80093-1, 1996.

Blanch, X., Guinau, M., Eltner, A., and Abellan, A.: Fixed photogrammetric systems for natural hazard monitoring with high spatio-temporal resolution, Natural Hazards and Earth System Sciences, 23, 3285–3303, https://doi.org/10.5194/nhess-23-3285-2023, 2023.

Cicoira, A., Marcer, M., Gärtner-Roer, I., Bodin, X., Arenson, L. U., and Vieli, A.: A general theory of rock glacier creep based on in-situ and remote sensing observations, Permafrost and Periglacial Processes, 32, 139–153, https://doi.org/10.1002/ppp.2090, 2021.

Cicoira, A., Weber, S., Biri, A., Buchli, B., Delaloye, R., Da Forno, R., Gärtner-Roer, I., Gruber, S., Gsell, T., Hasler, A., Lim, R., Limpach, P., Mayoraz, R., Meyer, M., Noetzli, J., Phillips, M., Pointner, E., Raetzo, H., Scapozza, C., Strozzi, T., Thiele, L., Vieli, A., Vonder Mühll, D., Wirz, V., and Beutel, J.: In situ observations of the Swiss periglacial environment using GNSS instruments, Earth System Science Data, 14, 5061–5091, https://doi.org/10.5194/essd-14-5061-2022, 2022.

Delaloye, R. and Staub, B.: Seasonal variations of rock glacier creep: Time series observations from the Western Swiss Alps, in: XI. International Conference On Permafrost – Book of Abstracts, Bibliothek Wissenschaftspark Albert Einstein, 24 June 2016, Potsdam, Germany, 2016.

Delaloye, R., Lambiel, C., and Gärtner-Roer, I.: Overview of rock glacier kinematics research in the Swiss Alps: seasonal rhythm, interannual variations and trends over several decades, Geographica Helvetica, 65, 135–145, https://doi.org/10.5194/gh-65-135-2010, 2010.

Delaloye, R., Morard, S., Barboux, C., Abbet, D., Gruber, V., Riedo, M., and Gachet, S.: Rapidly moving rock glaciers in Mattertal, Jahrestagung der Schweizerischen Geomorphologischen Gesellschaft, 21–31, 2013.

Elias, M. and Maas, H.-G.: Measuring Water Levels by Handheld Smartphones: A contribution to exploit crowdsourcing in the spatio-temporal densification of water gauging networks, The International Hydrographic Review, 9–22, 2022.

Elias, M., Kehl, C., and Schneider, D.: Photogrammetric water level determination using smartphone technology, The Photogrammetric Record, 34, 198–223, https://doi.org/10.1111/phor.12280, 2019.

Elias, M., Eltner, A., Liebold, F., and Maas, H.-G.: Assessing the Influence of Temperature Changes on the Geometric Stability of Smartphone- and Raspberry Pi Cameras, Sensors, 20, 643, https://doi.org/10.3390/s20030643, 2020.

Elias, M., and Maas, H. G.: Measuring Water Levels by Handheld Smartphones – A contribution to exploit crowdsourcing in the spatio temporal densification of water gauging networks, The International Hydrographic Review, 27, 9–22, https://doi.org/10.58440/ihr-27-a01, 2022.

Elias, M., Weitkamp, A., and Eltner, A.: Multi-modal image matching to colorize a SLAM based point cloud with arbitrary data from a thermal camera, ISPRS Open Journal of Photogrammetry and Remote Sensing, 9, 100041, https://doi.org/10.1016/j.ophoto.2023.100041, 2023.

Elias, M., Isfort, S., Eltner, A., and Maas, H.-G.: UAS Photogrammetry for Precise Digital Elevation Models of Complex Topography: A Strategy Guide, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-2-2024, 57–64, https://doi.org/10.5194/isprs-annals-X-2-2024-57-2024, 2024.

Eltner, A., Kaiser, A., Abellan, A., and Schindewolf, M.: Time lapse structure-from-motion photogrammetry for continuous geomorphic monitoring, Earth Surface Processes and Landforms, 42, 2240–2253, https://doi.org/10.1002/esp.4178, 2017.

Eltner, A., Sardemann, H., and Grundmann, J.: Technical Note: Flow velocity and discharge measurement in rivers using terrestrial and unmanned-aerial-vehicle imagery, Hydrology and Earth System Sciences, 24, 1429–1445, https://doi.org/10.5194/hess-24-1429-2020, 2020.

Eltner, A., Hoffmeiste, D., Kaiser, A., Karrasch, P., Klingbeil, L., Stöcker, C., and Rovere, A.: UAVs for the Environmental Sciences, 494 pp., 2022.

Fortun, D., Bouthemy, P., and Kervrann, C.: Optical flow modeling and computation: A survey, Computer Vision and Image Understanding, 134, 1–21, https://doi.org/10.1016/j.cviu.2015.02.008, 2015.

Frauenfelder, R., Haeberli, W., and Hoelzle, M.: Rockglacier occurrence and related terrain parameters in a study area of the Eastern Swiss Alps, Permafrost`, 253–258, 2003.

Harley, A. W., Fang, Z., and Fragkiadaki, K.: Particle Video Revisited: Tracking Through Occlusions Using Point Trajectories, 2022.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016.

Heid, T. and Kääb, A.: Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery, Remote Sensing of Environment, 118, 339–355, https://doi.org/10.1016/j.rse.2011.11.024, 2012.

Hermle, D., Gaeta, M., Krautblatter, M., Mazzanti, P., and Keuschnig, M.: Performance Testing of Optical Flow Time Series Analyses Based on a Fast, High-Alpine Landslide, Remote Sensing, 14, https://doi.org/10.3390/rs14030455, 2022.

How, P., Hulton, N. R. J., Buie, L., and Benn, D. I.: PyTrx: A Python-Based Monoscopic Terrestrial Photogrammetry Toolset for Glaciology, Frontiers in Earth Science, 8, 2020.

Huang, Z., Shi, X., Zhang, C., Wang, Q., Cheung, K. C., Qin, H., Dai, J., and Li, H.: FlowFormer: A Transformer Architecture for Optical Flow, in: Computer Vision – ECCV 2022, Cham, 668–685, https://doi.org/10.1007/978-3-031-19790-1_40, 2022.

Hur, J. and Roth, S.: Optical Flow Estimation in the Deep Learning Age, in: Modelling Human Motion: From Human Perception to Robot Design, edited by: Noceti, N., Sciutti, A., and Rea, F., Springer International Publishing, Cham, 119–140, https://doi.org/10.1007/978-3-030-46732-6_7, 2020.

Ioli, F., Dematteis, N., Giordan, D., Nex, F., and Pinto, L.: Deep Learning Low-cost Photogrammetry for 4D Short-term Glacier Dynamics Monitoring, PFG, https://doi.org/10.1007/s41064-023-00272-w, 2024.

James, M. R. and Robson, S.: Sequential digital elevation models of active lava flows from ground-based stereo time-lapse imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 97, 160–170, https://doi.org/10.1016/j.isprsjprs.2014.08.011, 2014.

Kääb, A. and Reichmuth, T.: Advance mechanisms of rock glaciers, Permafrost and Periglacial Processes, 16, 187–193, https://doi.org/10.1002/ppp.507, 2005.

Kääb, A. and Vollmer, M.: Surface Geometry, Thickness Changes and Flow Fields on Creeping Mountain Permafrost: Automatic Extraction by Digital Image Analysis, Permafrost and Periglacial Processes, 11, 315–326, https://doi.org/10.1002/1099-1530(200012)11:4<315::AID-PPP365>3.0.CO;2-J, 2000.

Kenner, R., Bühler, Y., Delaloye, R., Ginzler, C., and Phillips, M.: Monitoring of high alpine mass movements combining laser scanning with digital airborne photogrammetry, Geomorphology, 206, 492–504, https://doi.org/10.1016/j.geomorph.2013.10.020, 2014.

Kenner, R., Phillips, M., Limpach, P., Beutel, J., and Hiller, M.: Monitoring mass movements using georeferenced time-lapse photography: Ritigraben rock glacier, western Swiss Alps, Cold Regions Science and Technology, 145, 127–134, https://doi.org/10.1016/j.coldregions.2017.10.018, 2018.

Kummert, M. and Delaloye, R.: Mapping and quantifying sediment transfer between the front of rapidly moving rock glaciers and torrential gullies, Geomorphology, 309, 60–76, https://doi.org/10.1016/j.geomorph.2018.02.021, 2018.

Kummert, M., Delaloye, R., and Braillard, L.: Erosion and sediment transfer processes at the front of rapidly moving rock glaciers: Systematic observations with automatic cameras in the western Swiss Alps, Permafrost and Periglacial Processes, 29, 21–33, https://doi.org/10.1002/ppp.1960, 2018.

Lindenberger, P., Sarlin, P.-E., and Pollefeys, M.: LightGlue: Local Feature Matching at Light Speed, in: ICCV, 2023.

~~Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60, 2, 91–110, https://doi.org/10.1023/B:VISI.0000029664.99615.94~~

Marcer, M., Bodin, X., Brenning, A., Schoeneich, P., Charvet, R., and Gottardi, F.: Permafrost Favorability Index: Spatial Modeling in the French Alps Using a Rock Glacier Inventory, Frontiers in Earth Science, 5, 105, https://doi.org/10.3389/feart.2017.00105, 2017.

Marcer, M., Cicoira, A., Cusicanqui, D., Bodin, X., Echelard, T., Obregon, R., and Schoeneich, P.: Rock glaciers throughout the French Alps accelerated and destabilised since 1990 as air temperatures increased, Communications Earth & Environment, 2, 1–11, https://doi.org/10.1038/s43247-021-00150-6, 2021.

Matsuoka, N.: Combining Time-Lapse Photography and Multisensor Monitoring to Understand Frost Creep Dynamics in the Japanese Alps, Permafrost and Periglacial Processes, 25, 94–106, https://doi.org/10.1002/ppp.1806, 2014.

McColl, S. T. and Draebing, D.: Rock Slope Instability in the Proglacial Zone: State of the Art, Springer, Cham, 119–141, https://doi.org/10.1007/978-3-319-94184-4_8, 2019.

Pellet, C., Bodin, X., Cusicanqui, D., Delaloye, R., Kaab, A., Kaufmann, V., Noetzli, J., Thibert, E., Vivero, S., and Kellerer-Pirklbauer, A.: State of the climate in 2022: Rock glaciers velocity, Bulletin of the American Meteorological Society, 104, 41–42, https://doi.org/10.1175/2023BAMSStateoftheClimate.1, 2023.

RGIK, R. glacier inventories and kinematics: Guidelines for inventorying rock glaciers: baseline and practical concepts (Version 1.0), https://doi.org/10.51363/unifr.srr.2023.002, 2023.

Sarlin, P. E., Detone, D., Malisiewicz, T., and Rabinovich, A.: SuperGlue: Learning Feature Matching with Graph Neural Networks, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 4937–4946, https://doi.org/10.1109/CVPR42600.2020.00499, 2020.

Schwalbe, E. and Maas, H.-G.: The determination of high-resolution spatio-temporal glacier motion fields from time-lapse sequences, Earth Surface Dynamics, 5, 861–879, https://doi.org/10.5194/esurf-5-861-2017, 2017.

Shi, X., Huang, Z., Bian, W., Li, D., Zhang, M., Cheung, K. C., See, S., Qin, H., Dai, J., and Li, H.: VideoFlow: Exploiting Temporal Cues for Multi-frame Optical Flow Estimation, https://doi.org/10.48550/arXiv.2303.08340, 20 August 2023.

Stumpf, A., Augereau, E., Delacourt, C., and Bonnier, J.: Photogrammetric discharge monitoring of small tropical mountain rivers: A case study at Rivière des Pluies, Réunion Island, Water Resources Research, 52, 4550–4570, https://doi.org/10.1002/2015WR018292, 2016.

Sun, J., Shen, Z., Wang, Y., Bao, H., and Zhou, X.: LoFTR: Detector-Free Local Feature Matching with Transformers, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 4, 8918–8927, https://doi.org/10.1109/CVPR46437.2021.00881, 2021.

Teed, Z. and Deng, J.: RAFT: Recurrent All-Pairs Field Transforms for Optical Flow, in: Computer Vision – ECCV 2020, Cham, 402–419, https://doi.org/10.1007/978-3-030-58536-5_24, 2020.

Travelletti, J., Delacourt, C., Allemand, P., Malet, J.-P., Schmittbuhl, J., Toussaint, R., and Bastard, M.: Correlation of multi-temporal ground-based optical images for landslide monitoring: Application, potential and limitations, ISPRS Journal of Photogrammetry and Remote Sensing, 70, 39–55, https://doi.org/10.1016/j.isprsjprs.2012.03.007, 2012.

~~Ulm, M., Elias, M., Eltner, A., Lotsari, E., and Anders, A.: Automated change detection in photogrammetric 4D point clouds – Transferability and extension of 4D objects by change for monitoring riverbank dynamics using low-cost cameras, Proceedings of the 6th Joint International Symposium on Deformation Monitoring (JIDSM), 2025–April, 1–8, in press, 2025.~~

Wegner, K., Stark, M., Haas, F., and Becht, M.: Suitability of terrestrial archival imagery for SfM-MVS based surface reconstruction of steep rock walls for the detection of rockfalls, Journal of Geomorphology, https://doi.org/10.1127/jgeomorphology/2023/0775, 2023.

Xu, H., Zhang, J., Cai, J., Rezatofighi, H., and Tao, D.: GMFlow: Learning Optical Flow via Global Matching, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2022-June, 8111–8120, https://doi.org/10.1109/CVPR52688.2022.00795, 2022.

Zheng, Y., Harley, A. W., Shen, B., Wetzstein, G., and Guibas, L. J.: PointOdyssey: A Large-Scale Synthetic Dataset for Long-Term Point Tracking, 2023.