Author comment – General note

Dear reviewers,

We sincerely appreciate the time and effort you dedicated to reviewing our manuscript, as well as the valuable feedback provided on the previous submission. We have carefully addressed all the concerns raised and believe the revisions have significantly enhanced the quality of the manuscript. Below, we summarize the key changes and provide detailed, point-by-point responses to the reviewers' comments in the following sections.

After carefully discussing the reviewers' comments, we agreed that publishing the image-to-geometry tool would significantly benefit both the paper and the geoscience community, given the numerous questions raised about it by both reviewers. This decision required adapting the initial tool, developed by co-author Melanie Elias, to work specifically with georeferenced information from time-lapse images and preparing it for broader publication. Since the tool is Melanie Elias' intellectual property, and her contributions to the paper and in addressing the reviewers' comments have substantially increased, **we have revised the author order to include Melanie Elias as a joint first author.**

Besides this, the major changes in the manuscript are the following:

- Complete documentation of the image-to-geometry tool, now called GIRAFFE (see section '**3.2 From 2D to 3D: image-to-geometry registration using GIRAFFE'**), including source code on a separate GitHub repository (https://github.com/mel-ias/GIRAFFE).
- A new '**3.3 Combining PIPs++ and GIRAFFE'** section in the method section, specifically outlining the application of the introduced methods on our datasets, giving a detailed overview of the parameters used and how we validated the quality of our results.
- Under new section '**4.2. Error Assessment and Validation'** we enhanced our validation section, now including direct point-to-point comparison and a more thorough error assessment on two trajectories, considering the different error sources, including a new Figure 5 and Table 2.
- Under section '**5.1 Limitations'** we added a new figure, illustrating the limitations of the GIRAFFE method linked to the changing topography in the UAV data.

These changes implied following references to be added to the reference list:

- 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., 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.
- 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
- 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.

Below, you will find our responses to the reviewers' comments, with the reviewers' comments included in black.

With best regards,

Hanne Hendrickx and Melanie Elias

On behalf of all the co-authors

Answer to Reviewer 1 (Citation: https://doi.org/10.5194/egusphere-2024-2570-RC1)

This paper presents an interesting approach to monitoring fast-moving landforms, such as landslides and rock glaciers, using monoscopic time-lapse imagery and deep learning algorithms for feature matching and tracking. The authors applied their methodology to a case study at the Grabengufer site in the Swiss Alps, employing two low-cost monoscopic time-lapse cameras with relatively low image quality and significant camera-to-object distances. Utilizing the Persistent Independent Particle tracking (PIPs++) method, they effectively tracked distinctive features, overcoming challenges such as occlusions due to weather and object deformation. By tracking these features over time, they derived pixel displacement vectors illustrating movement in the rock glacier. Additionally, they used synthetic images generated from UAV point clouds with known extrinsic and intrinsic parameters to estimate the camera pose and interior orientation for the first image and scale the displacement vectors. This method, called the image-to-geometry approach, was previously published by the same group but only briefly detailed in this paper.

The code for feature tracking using PIPS++ and a sample dataset are provided as an open source to enhance reproducibility. However, the absence of code for the image-to-geometry method is a limitation.

Overall, this paper stands out as one of the first studies to apply state-of-the-art deep learning feature tracking algorithms from computer vision to geomorphology and geoscience. The image-to-geometry approach to georeferencing velocity fields is particularly relevant, presenting a viable alternative to traditional DEM back-projection, provided high-quality RGB point clouds are available.

Despite the paper's strong contributions, several areas could benefit from further clarification to enhance reader understanding and reproducibility in similar future studies. My primary concern is the image-to-geometry approach, which warrants a more thorough description. Given its importance in the proposed approach, I believe that the authors should summarize the main workflow with a step-by-step (synthetic) procedure, as it is really hard for a reader to understand the method without reading all the other papers.

Thank you for your kind words about the paper's scientific contributions. As stated in the general note, we now included the source code for image-to-geometry approach, now named GIRAFFE (Geospatial Image Registration And reFErencing). We also dedicated a lot more text for the explanation of the tool, both in the paper (incl. a new '**3.3 Combining PIPs++ and GIRAFFE'** section) and in the GitHub repository (https://github.com/mel-ias/GIRAFFE). Moreover, we dedicated special attention to the accuracy and precision of the tool, and discuss the potential source of error more precisely (under '**4.2. Error Assessment and Validation'** section).

Specifically, I recommend the following:

1. Once features on the real images are matched with those in the synthetic images using Lightglue, how are the 3D coordinates of the points retrieved for space resection? Are you doing a back-projection/ray-tracing to the point cloud/DSM/mesh, or did you store the original 3D coordinates of the points generating each pixel of the synthetic image? A brief explanation here would be valuable.

Indeed, we store the original 3D coordinates of the reference point cloud within the view frustum and taking care of occlusions in a synthetic image. This synthetic image is matched to the true camera image using, in the current version of GIRAFFE, LightGlue. Basically, this gives us 2D-3D images. Since the synthetic image stores the original 3D information, we can retrieve the underlying 3D information from the reference point cloud to establish 2D-3D matches and run space resection to estimate the camera model using pseudo control points. In the revised version of the Manuscript, we detailed the entire procedure of GIRAFFE, in the section '**3.2 From 2D to 3D: image-to-geometry registration using GIRAFFE'**.

2. Are the authors using the image-to-geometry method to "scale" or to "georeference" the velocity vectors? Given the availability of full camera exterior and interior orientations, I suppose you are doing more than just "scaling". Please clarify this aspect.

This is now explained in the section '**3.2 From 2D to 3D: image-to-geometry registration using GIRAFFE.** We are indeed georeferencing the individually tracked image points (in pixels) to corresponding 3D information used to estimate the 3D trajectories and further on to the velocities, rather than 'scale' the velocity vectors. This has been clarified and changed throughout the text.

3. Please, add a few comments for comparing the image-to-geometry approach with a more traditional DEM-back projection, which is nowadays the most widely used approach to georeference velocity fields from monoscopic cameras.

This is now explained in the section '**3.2 From 2D to 3D: image-to-geometry registration using GIRAFFE'**, in the mid of the section:

> '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 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.'

Additionally:

4. The number of points used to compute the average of the image-based velocity in section 5 (e.g., Figure 6) is significantly different. Did the authors consider all features tracked in the two areas, regardless of their distance from the reference GNSS/theodolite measurement points? It may be more

reasonable to average points within a certain radius to account for spatial variability in the phenomena.

Indeed, for Figure 7 (previously figure 6) we considered all image-based velocities that we can track. For Wcam4, these are substantially more points because of the more favourable position of the camera (closer, a more perpendicular view on the area of interest), compared to Wcam5. We kept this figure as it is to show general landform movement, but indeed tracked only the 25 points within a 5x5 pixel grid around the total station/GNSS measurement points to do a more direct validation. These results are presented in a new Figure 6.

5. Did the authors assess the pixel-level noise of the PIPS++ tracking in stable areas? This assessment is similar to what is stated in the caption of Figure 5 for the "scaled" vectors and it would help distinguish noise from tracking and that introduced by vector scaling. Additionally, I suggest moving this information from the figure caption to the main body of the paper, as it is relevant.

Thank you for your suggestion. We moved this information to the section '**3.3 Combining PIPs++ and GIRAFFE'**, where we describe how we calculated the level of detection of our method. For this, we indeed assessed the noise in the assumed stable areas. This noise-level is clearly visible in the violin plot in Figure 5c, where we have a noise level of 0.7 cm/day velocity, thus on our weekly image this means that a point should at least move 5 cm in order that it is detectable for our method. These values are for Wcam4. For Wcam 5 the calculated noise level or level of detection was 14 cm. These values are also presented in results and discussion.

I have included additional minor comments directly in the PDF for further consideration. Overall, this paper presents a high-quality contribution to the field, and addressing these points would significantly enhance its clarity and impact.

All additional minor comments have been addressed; we give a brief summary here:

- Figure 1 was adapted.
- Information about the own base station for improving differential position of the permanent GNSS was added.
- More image specifications were added in table 1, making it possible to calculate a GSD depending on camera distance.
- Assumptions about stability of the cameras over time have been clarified.
- Temporal prior was explained differently for clarification.
- The text about the PIPs++ model was clarified on the points addressed by the reviewer.
- The image-to-geometry approach was explained more in-depth.
- We explain the iterative procedure of GIRAFFE more in-depth, with time indication and additional benefit of ever iteration: 'By default, the software is configured to perform three iterations, a value determined empirically as a balance between computational efficiency and the likelihood of achieving convergence even with suboptimal initial parameter estimates (one iteration takes between a few seconds and a minute, depending on the point number of the used 3D point cloud). With each iteration, the synthetic image generated by GIRAFFE increasingly aligns with the perspective of the real camera, improving both the number and spatial distribution of pseudo-control points. This iterative approach enhances the stability

and robustness of the camera model adjustment procedure while accommodating coarse initial guesses with higher reliability.'

- Figure 5 caption was clarified (meaning of the points labelled in black and the dotted line)
- Figure 6 (now 7) has been updated with a more precise version of the image-to-geometry approach, being less prone to error. This changed the look of the velocity profile of Cam4, with no acceleration detected for the entire landform (roughly 100 points more included than in the previous version) but when the same boulder of the fixed GNSS was tracked, the same acceleration could be detected to a lesser extent – but without apparent delay.
- Figure 7 indeed summarizes all the points of the landform that our method detects, and they differ substantially from cam4 vs cam5 because of the very oblique viewing angle of cam5 and the greater object-to-camera distance. More points could be tracked with cam4 because the camera is more perpendicular to the study object and also a lot closer (100 m vs 700 m). However, for validation we now use indeed only the points around the measured GNSS/Total station point (5x5 pixel around the approx. location) to account for spatial variability in the landform. Note that spatial variability in the rock glacier seems to be quite limited, while in the landslide it is substantially larger, with large blocks moving faster than the overall landform.

Answer to Reviewer 2 (Citation: https://doi.org/10.5194/egusphere-2024-2570-RC2)

This paper introduces some useful improvements for a well-known method in environmental monitoring. The increased degree of automatization and the AI supported multi frame approach are a valuable contributions, making the approach more robust and user friendly. In widest parts it is well written and good to follow.

Nevertheless, major changes are necessary on the manuscript. The study claims pioneering and considerable improvements in certain other aspects but remains completely without proof. The validation of the data is very weak. The accuracy of the method (also under different setups / image geometries) remains widely open. The landform-wide mean values for displacement velocities are rather meaningless for this. I sorely missed a point-to-point comparison of individual displacements measured with in situ systems (ground truth) and image tracking at one certain location. Validation includes not only the displacement magnitude but also the displacement direction, especially if the independent detection of displacement direction is highlighted as a benefit of the method. This must be made up for and an honest accuracy analysis of the measured displacements (stable areas are not that relevant because the key problem is scaling!) has to be carried out, discussing different setup scenarios of camera / pointclouds and monitoring object. See also all the detailed comments. That will reveal a lot of error sources and problems which are not discussed here at all. (E.g. topography and local horizons will change over time but you use a single pointcloud to claibrate your model and many other problems)

We added a point-to-point comparison of individual displacements measured with in-situ GNSS or Total Station (ground truth) and image tracking at three specific locations, presented in a Table 2. This was also explained in detail in the new section '**3.3 Combining PIPs++ and GIRAFFE'**. Unfortunately, only two total station points were suitable for direct comparison for Wcam05 (new Figure 6). Most ground truth points are either not in view of the cameras, or they move to slow to be able to yield good results with our method (below the level of the detection). Nevertheless, we did a complete error assessment and a direct comparison of the two points that were available, across 10 timestamps. For Wcam04, only one point (the in-situ GNSS) was suitable for validation, although it also moves relatively slow and close to the level of detection. This point was evaluated across 7 timestamps (Table2).

You might notice that I was a bit annoyed in the beginning by the many references to climate change. Of course, this is an important topic and of course you can refer to it but not at any cost when it misses a clear connection. Although the overall language is good you must be careful in using the right terminology and to formulate precisely.

We agree that the references to climate change were not always in place. We adapted this in the abstract and introduction, but kept some references to climate change where it was more fitting, since we consider this a relevant thematic when talking about landform movement in permafrost areas.

Abstract

Line 8, First Sentence in Abstract: A vague and phrase-like statement. Personally, I do not like the term permafrost dynamics in this context, what exactly is this? Dynamics in rock glaciers are different then in landslides (perhaps you want to specify which type of landslides you are talking about?). Active rock glaciers are indicators of permafrost dynamics. Why are they critical indicators? Landslides in permafrost are mostly affected by several different processes and there are only a few documented

cases, which show a direct link between landslide activity and permafrost characteristics e.g. rock temperature or permeability. Filtering the permafrost signal out of landslide kinematics is in most cases too difficult to call them a "critical indicator". Yourself write later that you don't know what drives the upper landslide. The same if you claim these landforms represent the thermal state of permafrost. On a decadal scale, most of the rock glaciers accelerated (Some of them also decelerated) and do thus roughly reflect the long-term thermal state of permafrost. However, on a shorter time scale it is impossible the establish a correlation between permafrost thermal state and rock glacier velocity, e.g. because water plays such an important role. For landslides it becomes even more difficult, I do not see how we can substantially conclude from landslide activity on the thermal state of permafrost. There might be a political incentive to write "climate change" in the first sentence of each paper but there is no scientific need.

We have removed the vague 'permafrost dynamics' from the abstract and split the sentences so that active rock glaciers and landslides are not described together, since their different characteristics. The word critical has been removed. Rock glaciers have been described as indicators for mountain permafrost by the RGIK working group, the base document generally accepted by the rock glacier community and referred to in the text.

Line 22: Sorry to come with this again, but why do you need a super high spatiotemporal resolution to reveal climate change impacts? Don't you need particular long time series to show the effect of climate change?

I think both high spatiotemporal resolution datasets and long time series are needed to reveal climate change impacts. Considering that some time-lapse datasets span over a decade, this data might reveal correlations between changing environmental drivers and changing landform activity.

Intro

Introduction is rather long with some repetitions, particularly towards the end. It should be shortened.

Repetitions at the end of the introductions were removed.

Feature tracking in time lapse imaging is meanwhile a widespread, operationally applied method in natural hazard management. Many engineering offices are using it. You should mention this.

This has been added in Line 71:

> 'While traditional motion estimation methods are commonly applied in natural hazard management, they face decorrelation challenges with large displacements, strong illumination changes, and occlusions.'

Line 28/29: better here in this context but I would rather write: "… internal structure, and it reflects long term, temperature driven changes in permafrost structure."

Adapted.

Line 29: “Creep rates/velocity” instead of “flow speeds”

Adapted

Line 29: “occur towards the lower permafrost limit…”

Adapted

Line 30: “the acceleration of these processes becomes more pronounced as rock glacier creep rates increase in a warming climate”

Which acceleration? Strange sentence makes no sense. Language.

This sentence has been adapted – sediment transfer becomes more pronounced with rock glacier creep rates increase.

Line 32/33: What are permafrost related creep features? Landslides do not creep per definition. Be careful to use the correct terminology.

This was changed to ‘The same is true for certain types of landslides, where the primary driver of the motion is permafrost creep’.

Line 35/36: As written before, this does not make sense to me. First and foremost, monitoring is important for the safety because it helps to prevent hazards. Moreover, your high-resolution data is interesting for research because it helps process understanding. This is the core of your study! Impacts of climate change become evident when you monitor over more than 3 decades and the connection to the thermal state is complex. This is rather far from your study and there is no need to link it at all costs here. Use creep only if you mean creep.

Agreed. This has been adapted to ‘…as it provides information about the environmental drivers and enhances process understanding.’

Line 37: The same is true for very slow rock glaciers…

I did not say that monitoring rock glaciers in general is easy, but fast-moving landforms are particularly challenging because bi-annually measured GNSS/Total Station points are hard to find again, the terrain is unstable and dangerous, permanent GNSS stations need re-levelling or instrument replacement, tracking methods on bi-annual UAV data might face decorrelation because of landform speed… This is all elaborated in the lines following line 40.

Line 37-58: All true but mention also disadvantages of the time lapse method, E.g. weather dependence, constantly changing image extents/complex distortions in many existing time series. Depending on the setup very limited accuracy when transforming displacements in the metric 3D space….

We agree that disadvantages of the time lapse method should be mentioned, this is now done in the introduction, line 61:

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

Line 53-54: The cited cases are very different from the current one. It is possible to measure distances in images when projecting them on a 3D model, however dependent on the setup (camera calibration and orientation, line of sight, object geometry, resolution of 3D model and so on) the accuracy is very limited and a reasonable accurate scaling of displacements from image to world coordinates can fail. Especially in the close range, at local horizons, in front or behind terrain which is shadowed in the image. You didn't prove the opposite in your study…

Thank you for your insightful comment. We appreciate the opportunity to clarify our methodology in more detail.

In GIRAFFE, the view frustum can be explicitly defined along the optical axis, allowing us to exclude occlusions caused by objects close to the projection center. This ensures that potential inaccuracies stemming from nearby obstructions are mitigated. In our study, we configured the frustum to extend from 1 meter to 2000 meters along the optical axis, as no occlusive objects were present in front of the camera. Furthermore, GIRAFFE offers a 'location_accuracy' parameter that expands the frustum when pose parameters are uncertain. This feature ensures that sufficient image content is available for accurate assignment of the camera image, even when the camera position is initially only approximately known. As the detailed functionalities and settings of GIRAFFE have already been comprehensively described in Elias (2019), we have referred readers to that publication for an in-depth explanation. Additionally, the README.md file and documentation on bounding box calculations in the GitHub repo of GIRAFFE provide further guidance on these parameters.

Our iterative approach facilitates the selection of the optimal camera model for each scenario, including cases such as calibrating lens distortion based on the pseudo-control point distribution. This methodology is elaborated upon in the extended section of our paper (see above).

Line 57: I don't think that is true. How do you know? Cameras just have a much longer history of widespread application than permanent GPS devices, which came up just about 10 -15 years ago.

In line 41-43 ', 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).' I explain that permanent GNSS often get destroyed on fast-moving landforms. Cameras that are directed at fast moving landforms are generally on (more) stable terrain, and will therefore require less maintenance and have a longer longevity. But even considering this, on the long term it might be difficult to find a suitable stable spot for installing a permanent camera in a dynamic mountain environment. This has been added to the disadvantages of the time lapse method in line 61 (see above).

Site description

What makes the upper half a landslide? Its surface structure could tempt you to call it a rock glacier too, doesn't it? Perhaps you can explain the characteristics of the two landforms in more detail. Is the landslide ice saturated?

Thank you for your suggestion. We have provided additional details on both landforms in the revised manuscript. There is a substantial amount of data collected from this site that is yet to be published, and we believe that this paper may not be the most appropriate place to present all of it. As such, we have included brief statements on this topic (lines 105-120) to give an overview while reserving the more comprehensive analysis for future publications:

> '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 temperature is still below 0° despite recent warming. The southern part of the landslide is frozen and ice saturated (unpublished data). 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, we use the generic term 'landslide'. The frontal section of the southern part of this landslide is very unstable and the source of frequent rock falls. It gradually turns into a very rapid mass movement, referred to as the "(rock glacier) feeding section". Both frontal and feeding sections are monitored by a permanently installed webcam (hereinafter named Wcam04, Fig. 1c orange polygon). 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. As of summer 2023, in-situ measurements indicate that it moves again exceptionally fast at a rate of 0.25 to 0.70 m d^{-1}, marking a third phase of destabilisation. The 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.'

Methods

Line 127: mm per day is more common.

This has been adapted to unit of choice – in our case m per day

Line 143: "independent of temporal prior" ??

'Independent from the image before or after.' This has been clarified.

3.1 I found it hard to understand what kind of trajectories are calculated for the multi-frame batches. Is there one trajectory per batch and one Trajectory position per frame? And how exactly are the trajectories "estimated" in case of fog? Is there kind of an interpolation or just a gap in the trajectory for the foggy frame? Or a calculation over a longer period? Please explain a bit more detailed.

There is one trajectory position per frame, leading to one trajectory per batch for this specific feature of the tracking. If the feature changes appearance slightly, due to snow or rotational or lightning changes, the model is built in a way that it can update templates along the trajectory. It basically saves multiple templates for one feature, meaning: the feature could also look like this or like this. Making the model way more robust for these typical environmental changes that cause traditional tracking methods to fail. When no match is found in one frame due to fog, the model will simply interpolate the trajectory as long as the feature is still found in the next frame, still allowing the feature to be tracked. Traditional methods also fail if occlusion due to fog will not yield a result. Some clarifying statements has been added to Line 185-195.

Line 160: Distortions due to camera shifts can be much more complex than simple offsets. Due to rotations or depending on the camera sometimes even slight changes in focus, the distortions are often spatially differential and only a part of it can be corrected by simple adjustments.

Thank you for your valuable feedback. We have addressed this point in the revised version of our paper. Instead of referring to a 'simple' shift, we now emphasize the importance of stabilizing the image sequence beforehand and provide both, an algorithmic description on an option how to do a software-side stabilization and provide a python script in Zenodo. Alternatively, when the movements of the camera are to be expected too extensive, we suggest using each individual image for georeferencing within GIRAFFE. This approach accounts for more complex distortions resulting from camera rotations or focus changes, ensuring more accurate correction and alignment.

Line 196: What means directly linked? There are of course other error sources than just the accuracy of the point cloud.

Thank you for your comment. We have extensively revised the error assessment section in the paper to address this concern. Our updated analysis now considers not only the potential errors in the point cloud but also the inherent error sources within GIRAFFE.

To provide a clearer overview, we have added a new Figure 8, which illustrates various error sources associated with the camera view and the changing UAV topography. Additionally, we present a comprehensive theoretical accuracy assessment, examining all relevant error factors in detail. This thorough evaluation ensures a more holistic understanding of the potential inaccuracies and their impacts on the results.

Line 215: Why only in the stable areas? The deciding thing would be to analyze accuracy in the moving part with ground truth!?

The stable areas were used to assess the level of detection of our method. This has been outlined in the new section '**3.3 Combining PIPs++ and GIRAFFE'**. We did use three points in the moving area to do a more thorough error assessment with ground truth available, as already explained above. Information about this direct validation is presented in Table 2 and a new Figure 6.

Results

Figure 5: What means accumulated velocity? Do you mean mean velocity? Theodolit points are the black white labeled points? How was the noise level defined? In Figure 5c there is still a lot of movement above the noise level in the stable areas.

Figure 5 is showing the total velocity for the entire summer period, for all the points tracked. The black and white labeled points are the ground truth, both by total station or by GNNS. This has been clarified. We agree that in figure 5c there is still movement above the noise level. This is because there is actually no stable area in the image view of Wcam04, since it is entirely in the deep-seated landslide area (Fig 1 blue polygon). However, these areas are relatively stable on a short time frame.

Line 238: I do not consider that as more interesting. Geoferencing of locations in an image is one thing. However, a crucial step is the scaling of displacements from pixel to global coordinates. If you do not detect change in image coordinates (or barely any change) there is nothing to scale and this major error source does not become influential. Accuracy analysis of significant displacements is thus the deciding and most interesting validation step of your method.

We agree a that it is very important for the accuracy analysis to include significant displacements. A direct comparison between ground truth and our results is done on three points in the moving area, each for different time stamps. See also answers above.

Validation part in general

This part is very weak. What you show in Figure 6 is not a solid validation of your work as you compare spatial averages with single point GPS measurements or conspicuous average values from multiple TPS points. To proof your concept, you have to compare single in situ measurements with local trajectories from your images in the near vicinity of this in situ measurement! The current validation isn't convincing at all.

We agree that a direct comparison between in situ measurements and our images is very important for validating the method. Only three points with each several timestamps were available for this assessment (see also explanation above). These trajectories are used for a thorough error assessment (Table 2, Figure 6). We dedicated now an entire new section '**4.2. Error Assessment and Validation'** to this topic.

Discussion

Line 261 ff: "The results of our workflow show a good agreement with dGNSS, theodolite and permanent GNSS measurements, proving our method to be reliable, robust and fast for creating a better spatial (Fig. 5) and temporal coverage (Fig. 6) of the landform's displacement"

This subjective statement isn't supported by the presented data.

We believe that this statement holds with our added validation data, using direct comparison.

Line 266: This conclusion is not new but was made by many authors before.

Thank you for your feedback. We have revised the statement as follows: 'Our pilot study demonstrates that significant 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.' While it is true that time-lapse images could previously be used to improve the temporal and spatial resolution of velocity data, our approach offers a faster, easier, and more robust solution through AI-based methods. We believe that our methodology represents an advancement over traditional methods and enables the processing of larger datasets.

Line 269: You do not know if where this discrepancy originates from. The displacement is not uniform over space and perhaps the GPS is located in a faster area but not on a surfing boulder as you say. As said above, comparing a single GPS with the average velocity of the entire landform is not a purposeful approach.

This has been solved by using a more direct comparison for validation. Indeed, the GNSS antenna is located on a faster moving block and the speed up cannot be seen in the overall landform displacement (updated Figure 7). We therefore tracked the same block with our PIPs++/GIRAFFE workflow and could detect the same speed up, although in a lesser extent. Note that tracking the exact location of the GNSS antenna was not possible because it is located at the back side of the boulder, not directly visible by the camera. This have all been updated in Figure 7.

Line 294/295: This statement is not supported by your data and most likely wrong. In theory your method might be able to calculate 3D trajectories in world coordinates, but I assume strongly that these trajectories deviate considerably from trajectories measured by GNSS or Total station at the same location. Be aware that displacement directions are even more prone to error influences than displacement magnitudes. You do not validate your displacement directions from image tracking with ground truth at all, how can you make such a conclusion then? You not even compared displacement magnitudes on a specific location between image tracking and ground truth…

We believe that this statement is true with our additional validation data obtained by direct comparison. The trajectories calculated for two validation points (new Figure 6 and Table 2) show a good agreement in terms of the main direction of motion with the Total Station reference measurements. The main directions of movement were determined by fitting linear 3D trajectories (linearity assumption chosen after evaluation of TS based trajectories). The variation / angular difference of the direction between 3D trajectory from TS measurements and PIPs++ / GIRAFFE measurements was determined by direction vector calculation. This information was also added to the revised version of the Manuscript in section 4.2.3.

We also adapted the section about stabilizing images using the offset to:

> '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 method

is only capable of compensating to a limited extent for perspective distortions resulting from changes in perspective due to 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. ‘

Line 291: only to a certain degree as mentioned above.

We removed this statement, as we propose to either stabilize the image sequence beforehand or to estimate the (non-) camera model of each image of the time-lapse sequence using GIRAFFE to deal with camera movement. This way, the 3D points are georeferenced using individual camera models. We outlined the pros and cons of both options in line 280-285.

Limitations:

Line 311: True, moreover the RGB information for the point clouds might be difficult to acquire in very steep terrain (Rock walls). This is not discussed at all.

Thank you for raising this point. With the capabilities of current flight planning software, obtaining high-quality RGB point clouds in steep terrain has become more feasible. For example, UAV data can be effectively acquired through ‘façade scanning,’ where drone flights are conducted parallel to the steep terrain, allowing the camera to face the area of interest perpendicularly. Additionally, modern terrestrial laser devices, often used for rockwall monitoring, are equipped with RGB cameras to colorize the acquired point clouds. Given these advancements, we believe a detailed discussion of this topic may not be necessary within the scope of this paper.

General: You whole methods depends on one point cloud. If surface geometry is changing, what is obviously the case here. Your displacement scaling and the calculation of displacement directions is distorted or will fail, depending on the size of terrain change.

Thank you for highlighting this important point. To address your concern, we have included a new Figure 8, which illustrates the potential impact of changing surface geometry on the accuracy of our method. Additionally, we have provided a detailed explanation of this limitation in the ‘Limitations’ section of the discussion to ensure clarity and transparency regarding the challenges associated with terrain changes:

‘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). 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 data.'

355: Accurate? How do you know?

We believe that this statement holds with our added validation data, using direct comparison.

362: Not really validated.

We believe that this statement holds with our added validation data, using direct comparison.