

Leopold-Franzens Universität Innsbruck

Department of Computer Science



CTCS Report

## Image Processing Murtèl Rock Glacier

Noah Hurmer, Max Gallinat,  
Uzair Khan, Manuel Reichegger

**Date:** February 16, 2025  
**Project Partners:** Dr. Dominik Amschwand, Prof. Jan Beutel

## Abstract

Computer vision techniques are increasingly applied in environmental settings for monitoring and analysis including alpine applications. This report focuses on the Murtèl rock glacier in the Swiss Alps and explores methods to establish a workflow to filter, align, and perform cross-domain matching of visible-spectrum (RGB) and thermal infrared (TIR) images. The methodology proposed here incorporates an SVM-based image filter to discard unusable images, followed by SuperPoint and SuperGlue for robust RGB alignment across the time domain. A cross-correlation based template matching then integrates TIR data, aided by tuned image preprocessing. The resulting pipeline improves alignment consistency over different snow conditions and time intervals, enabling downstream analyses related to glacier movement, surface change, and thermal mapping. Although domain-specific cropping and masking were necessary to remove foreground obstructions unique to the Murtèl site, the overall approach remains broadly applicable to other alpine settings.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Project Overview . . . . .	3
1.2	Related Work . . . . .	3
1.3	Proposed Workflow . . . . .	4
<b>2</b>	<b>Filtering</b>	<b>5</b>
2.1	RGB filtering . . . . .	5
2.2	TIR filtering . . . . .	6
2.3	Summary of Filtering . . . . .	8
<b>3</b>	<b>Preprocessing</b>	<b>9</b>
3.1	Data Selection . . . . .	9
3.2	Ledge Masking . . . . .	9
3.3	Preprocessing for RGB–RGB Alignment . . . . .	10
3.4	Preprocessing for TIR–RGB Matching . . . . .	11
<b>4</b>	<b>Alignment</b>	<b>13</b>
4.1	SuperPoint and SuperGlue . . . . .	13
4.2	Results . . . . .	14
<b>5</b>	<b>Cross-Domain Matching</b>	<b>17</b>
5.1	Template Matching for RGB–TIR . . . . .	17
5.2	Results . . . . .	19
<b>6</b>	<b>Outlook</b>	<b>21</b>
6.1	Improvements . . . . .	21
6.2	Possible Downstream Analysis . . . . .	22
	<b>List of Abbreviations</b>	<b>23</b>
<b>A</b>	<b>Appendix</b>	<b>26</b>

# 1. Introduction

The Murtèl Rock Glacier, located in the Engadin valley, is a rock glacier (Krainer & Mostler 2000), characterised by its gradual movement and its dynamic surface, which includes snow patches, debris, vegetation, and exposed rock. The rock glacier's movement and surface changes are of interest in order to monitor its development and the factor influencing it.



Figure 1.1: Example images Murtèl Rock Glacier

The camera system used to monitor the Murtèl rock glacier consists of an RGB and a thermal imaging camera (TIR) in one casing. These cameras have recorded images of the surface of the rock glacier over a period of two years, capturing around 30,000 pairs of images. Each pair (Figure 1.2) consists of an RGB image and a corresponding TIR image taken at a short time interval.

The images have no clear key points, such as a mountain range or fixed elements in the background, to help identify features. This makes it difficult to align and track changes in the images. In addition to varying surface condition, snow coverage and

lighting conditions, the slight differences in camera position, introduced by wind effects, and resolution between the RGB and TIR cameras make image comparison more challenging. These factors make it difficult to maintain accurate and consistent monitoring of the rock glacier surface.

## 1.1 Project Overview

The goal of this project is to explore image processing techniques on the 30,000 image pairs dataset, including both RGB and thermal infrared (TIR) images, captured over all months and seasons. The main objectives are:

- To **filter** and classify images based on weather conditions.
- To **align** misaligned images and create a timeline of rock glacier imagery.
- To spatially **overlay** TIR image data onto RGB images for enhanced analysis.

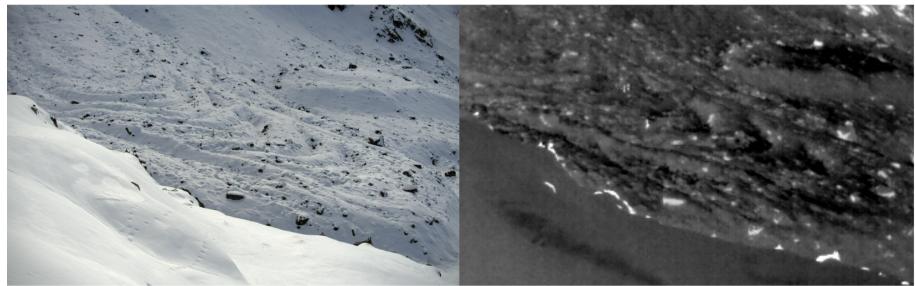


Figure 1.2: Example RGB and TIR image pair

## 1.2 Related Work

Several existing approaches inspire the development environmental image monitoring strategies, image quality assessment, and efficient matching techniques. In the context of imaging quality, Tsai et al. (2008) propose an entropy-based metric that characterizes the noise and blur within digital X-ray images using a single parameter. Focusing on real-time classification of environmental imagery, Ship et al. (2024) employ Support Vector Machines (SVM) to distinguish weather conditions. Ioli et al. (2024) present a deep learning-based multi-camera system for tracking glacier dynamics, successfully utilizing SuperGlue (Sarlin et al. 2020) to align different camera angles and performing change detection on the Belvedere Glacier in the Italian Alps. Keypoint extraction based models perform well in matching image frames and object detection (Rublee et al. 2011, DeTone et al. 2018). Other studies leverage template matching for alpine thermal infrared (TIR) image alignment (Arioli et al.

2024), using normalized cross-correlation for robust image matching (Zhao et al. 2006).

### 1.3 Proposed Workflow

The project employs various methods for classifying, aligning and cross-domain matching image data. Below is the complete workflow for this project comprised of the methods of highest performance in each domain:

1. **Image Filtering:** Support Vector Machines (SVM) (Mammone et al. 2009) is used for classifying RGB images based on clarity, and entropy measures (Tsai et al. 2008) are applied to filter out unusable TIR images.
2. **Image Preprocessing:** The RGB and TIR data are paired with each other based on the time of recording, whereby mismatched or filtered values are removed.
3. **Alignment:** The images are aligned to each other using SuperPoint (DeTone et al. 2018) and SuperGlue (Sarlin et al. 2020) for RGB-RGB image alignment.
4. **Cross-domain Matching:** Template matching with a cross-correlation (Brunelli 2008) measure is used for TIR-RGB overlay.

## 2. Filtering

Analysis of the Murtèl Rock Glacier images is complicated by varying environmental conditions such as snow cover, fog, changing light conditions and camera misalignment, all of which degrade image usability. As these images are to be aligned, it is essential to filter out those, that do not exhibit clear weather or are unusable, as shown in Figure 2.1. This section describes the process of removing poor-quality images from the dataset, including dataset creation, model evaluation and performance results.



Figure 2.1: Extreme Examples unusable RGB Images

### 2.1 RGB filtering

RGB filtering involves processing standard visible spectrum images, which are the primary source for image alignment in the temporal domain. To train machine learning models for this task, a subset of over 2000 images is manually labelled as either clear or unusable. The dataset for model training is then equalised to ensure that the distribution of usable and unusable images as well as seasonal representation is balanced.

Among tested methods, the Support Vector Machine (SVM) - a supervised learning algorithm that constructs a hyperplane to classify data points - with a Radial Basis

Function (RBF) kernel showed the best performance for filtering RGB images (Figure A.1). The features used to train the SVM were inspired by Ship et al. (2024) and include: Edge density, Brightness values, Color histograms (mean, variance, kurtosis, skewness) and additionally Metadata (Time of Day, Season).

Table 2.1 depicts feature importance, where higher values correspond to the more relevant features for decision by the model. This shows that metadata parameters are of negligible significance. The most important features are mainly image-related properties such as edge density and color distribution.

Feature	Value	Feature	Value
edge density	0.0799	red kurtosis	0.0575
green kurtosis	0.0763	blue skewness	0.0574
green skewness	0.0756	red variance	0.0567
blue variance	0.0673	contrast	0.0536
green variance	0.0627	blue kurtosis	0.0526
red skewness	0.0621	red mean	0.0522
green mean	0.0600	sharpness	0.0518
blue mean	0.0588	brightness	0.0396

Table 2.1: Feature importance SVM

The model was trained on 1,600 images and tested with 600. While other models tested achieve a higher accuracy, focus was declared on clear image precision, as ensuring a minimal number of unusable images are carried further into the downstream tasks has priority. The proposed model achieves a precision of 92.5% in correctly classifying high-quality images (Table 2.2), demonstrating the effectiveness of this filtering approach.

	Predicted Bad	Predicted Clear
Actual Bad	35.50%	2.00%
Actual Clear	11.17%	51.33%
<b>Precision: 92.50%</b>		

Table 2.2: Performance Measures SVM Filtering

## 2.2 TIR filtering

Thermal infrared (TIR) images offer complementary information to RGB images, particularly when monitoring environmental conditions such as temperature distri-

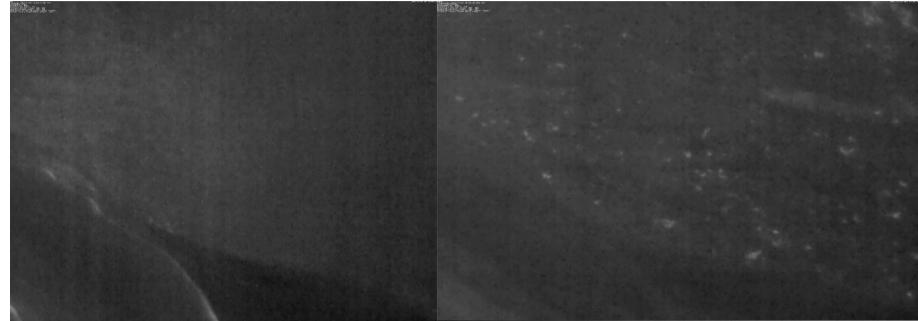


Figure 2.2: Example unusable TIR images

bution in rock glacier surfaces. However, conventional RGB-based feature extraction methods are not effective for classifying TIR images. Therefore an alternative approach was used, Shanon Entropy as a measure of information contained within an image, inspired by related work (Ioli et al. 2024).

**Shannon entropy:**

$$H(X) = - \sum_i p(x_i) \log_2 p(x_i)$$

where  $H(X)$  denotes the Shannon entropy and  $p(x_i)$  is the probability of the occurrence of intensity value  $x_i$  in the image.

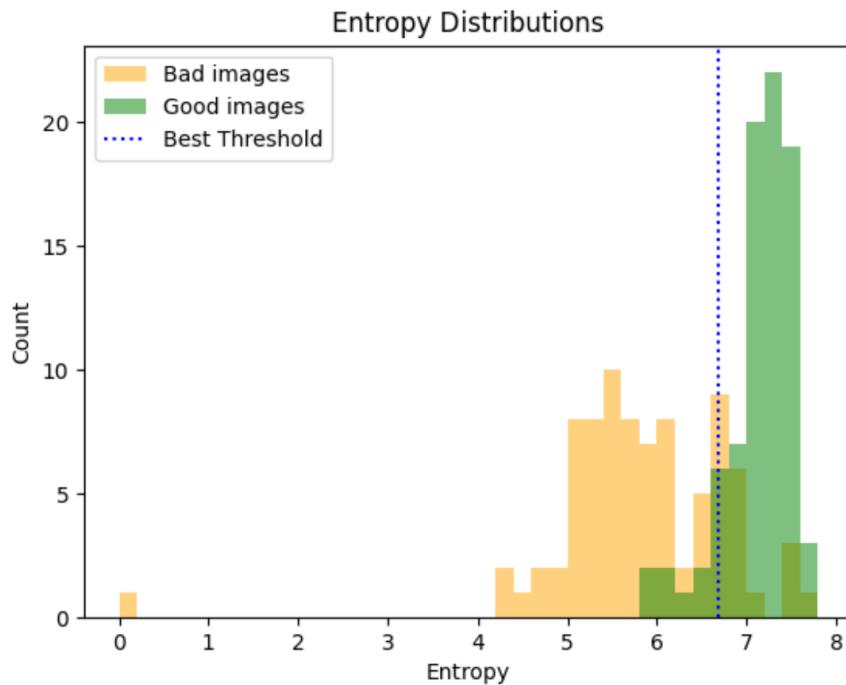


Figure 2.3: Entropy distributions of good/bad TIR images

The Shannon entropy is a quantitative measure of the information contained in the pixel intensity distribution (Tsai et al. 2008). Consequently, it can be employed as an indicator of image quality. Using 700 manually labelled TIR images, a subset was created to ensure an even distribution of clear, unusable and month-based variations.

On this set the Youden index method was utilised to determine a threshold for classification. It maximises the sum of sensitivity (true positive rate) and specificity (true negative rate), to find the optimal threshold in classification tasks. Setting the decision value in this case to 6,693 between high and low quality TIR images (Figure 2.3). With the set threshold, the precision of the model is 92.3%.

## 2.3 Summary of Filtering

The combination of RGB filtering using SVM and TIR filtering based on Shannon entropy forms a robust structure to efficiently select high-quality images for further alignment and analysis. The classification processes reaches a high precision in both cases, with the SVM model 92.5% precision for RGB images and with entropy-based filtering 92.3% precision for TIR images.

# 3. Preprocessing

Both visible-spectrum (RGB) and thermal infrared (TIR) data are collected under varying snow conditions, lighting changes, and seasonal shifts, which can introduce differences in appearance. By focusing on specific regions, enhancing feature similarity and suppressing transient elements, preprocessing and filtering images facilitates more accurate alignment and analysis for downstream tasks.

## 3.1 Data Selection

Smaller subsets of image pairs are assembled to systematically test the entire workflow from preprocessing to alignment and cross-domain matching. Specifically, 7 distinct timelines are created, ranging from all images taken on a single day to over 3 months, and include between 7 and 15 image pairs each. An RGB and a TIR image are paired up by closest time of capture to each other. The selected subsets cover different years, times of year, and snow conditions to capture a broad range of possible variations. Table 3.1 provides specific details of these datasets, illustrating differences.

## 3.2 Ledge Masking

Each camera frame contains a large ledge in the bottom-left corner, as this is the terrain on which the camera pylon is mounted, which is nearer in the foreground than the rock glacier. This poses issues for image alignment, as the closer nature combined with changing appearance of this ledge - i.e. snow level / covered rocks - may introduce errors in the form of incorrect distortions, in turn degrading the accuracy of alignment operations. This is true for both image modalities, TIR and RGB.

In order to minimize this issue, a portion of the bottom (24%) and left (18%) of each image is cropped to eliminate most of the ledge. Any remaining visible portion of the ledge is then masked so that it does not interfere with subsequent matching computations. A representative example of the cropped and masked image is shown in Figure 3.1. To reduce computational overhead and ensure compatibility with convolutional deep learning models used in the following, every image is uniformly downsampled to a resolution of  $1280 \times 790$ .

Dataset	Date(s)	Description
Maximum Snow Cover	March 2021	Dominated by heavy snow coverage, very limited visible rock surface.
Medium Snow Cover	October 2020	Snow coverage with some exposed rocks; variations in lighting and shadows.
Full Mix Snow-Rock	June–July 2021	Transitional period with melting snow over one month, leading to a gradual reveal of rock surfaces.
Summer First Snow	August–October 2021	Sequence captures the environment from fully exposed rock to the season’s first snowfall.
Single Day	November 2nd, 2020	Nearly complete snow coverage within a single day, featuring notable changes in lighting and image color conditions.
TIR Date Offset	09/28–10/06 (RGB); 10/14 (TIR) 2020	Approximately two-week offset between RGB and TIR image captures; Done in order to test robustness of models for less fitting cross-domain image pairs
Artificially Misaligned	June–July 2021	Full mix dataset deliberately translated, rotated, and warped to evaluate robustness against misalignment.

Table 3.1: Overview of the different dataset selections and their key characteristics

### 3.3 Preprocessing for RGB–RGB Alignment

Several filters and transformations were trialed in order to unify image characteristics and highlight key features that aid in matching and alignment. The motivation here is to remove any brightness or color difference between images that are solely introduced by light, shadow, time of day and year or weather variations.

Most of these variations are dealt with by grayscaling and histogram equalization, which improves contrast and makes subtle features more prominent. Other tested filters include edge sharpening or Gaussian blurring to control noise and emphasize boundaries; color morphology operations such as erosion and dilation; and edge extraction to simplify each image to a more abstract, boundary-focused representation.

Experimental evaluations of the downstream alignment operations indicate that a

combination of grayscaling, histogram equalization, and masking provides effective results for RGB–RGB alignment. But since various filter combinations with multiple possible alignment algorithms exist, this was not an exhaustive search and therefore may not represent the most optimal preprocessing. This strategy combined with SuperGlue (Sarlin et al. (2020), see section 4.1) alignment was simply the most performant of the ones we measured. This preprocessing is visible in figure 3.1.

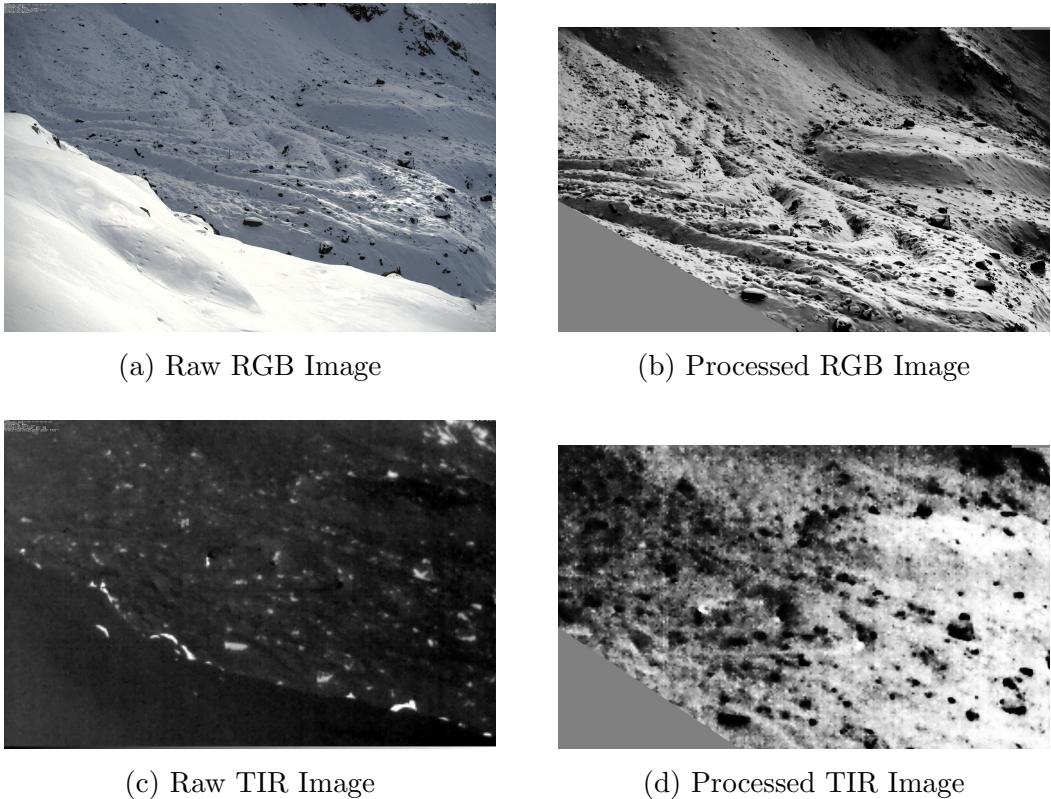


Figure 3.1: Example of the entire preprocess workflow for an image used in alignment and cross-domain matching tasks. (a) shows the raw RGB image taken on November 2, 2020, and (b) shows the cropped, resized, filtered, and masked output of that image. Similarly, (c) and (d) show the corresponding processing for the TIR image in the same pair.

### 3.4 Preprocessing for TIR–RGB Matching

For this operation, multiple filtering strategies for images of both modalities were also conducted. Color images were converted to grayscale, as thermal infrared (TIR) images inherently do not contain color information. Matching two images of different modalities using extracted edges with convolutional kernels did not result in

accurate matching throughout various alignment strategies tested, against expectations. Neither did morphology operations on the rgb images to make *colors* more diluted, as would be expected in surface temperature data. Gaussian blurring of the RGB images, however, did enhance compatibility between the two modalities.

Images captured in the TIR domain also benefit from selective filtering. Cross-correlation analysis demonstrates that downscaling TIR images to approximately 82% of their original size aligns their focal length with that of the RGB camera. In addition, TIR images show improved alignment accuracy after histogram equalization and downstream template matching operations decide whether to invert the picture or not. The motivation here is explained in more detail in section 5.1. Example of TIR preprocessing to be seen in figure 3.1.

# 4. Alignment

Ensuring that sequential RGB images are geometrically aligned is an essential step for several downstream applications. Accurate image alignment facilitates simpler fusion of thermal infrared data and support analyses of changes over time, such as variations in snow coverage or the detection of events such as rock falls and avalanches. Furthermore, potential downstream tasks such as mapping image pixels onto digital elevation models, require a well-defined and stable camera orientation.

Although the camera is fixed at one position, mounted to a pylon, wind-induced movements or possible manual adjustments result in small but visually perceivable shifts in camera orientation. The absence of a well-defined horizon line and permanent ground control points (GCP) further complicates the process of aligning multiple images. As the observable area is subject to snow accumulation and melt, varying snow coverage can obscure features that would otherwise serve as reference points, increasing the difficulty of pixel-value-based alignment methods such as cross-correlation (xcorr). Consequently, more robust alignment strategies, that extract important keypoints of images are desired.

## 4.1 SuperPoint and SuperGlue

SuperPoint (DeTone et al. 2018) is a convolutional encoder-decoder architecture designed for interest point detection and point descriptor extraction in images. It identifies visually significant points (keypoints) in an image and generates a descriptor for each point.

SuperGlue (Sarlin et al. 2020) is a attention based Graph Neural Network which can then compare these keypoint maps and descriptors between two images to establish point correspondences, validating or rejecting keypoint matches. This procedure yields a set of matching keypoints that serve as anchor points for estimating a geometric transformation. Using this to estimate an affine transformation matrix, we can align one image to another.

Both models are available pretrained by *magic leap* (Sarlin et al. 2020) and SuperGlue is also available with pretrained *outdoor* weights, which likely improves performance on this task compared to other pretrained models.

In the context of image alignment for this project, the described method provides keypoints of images that are less susceptible to changes in lighting or disappearing

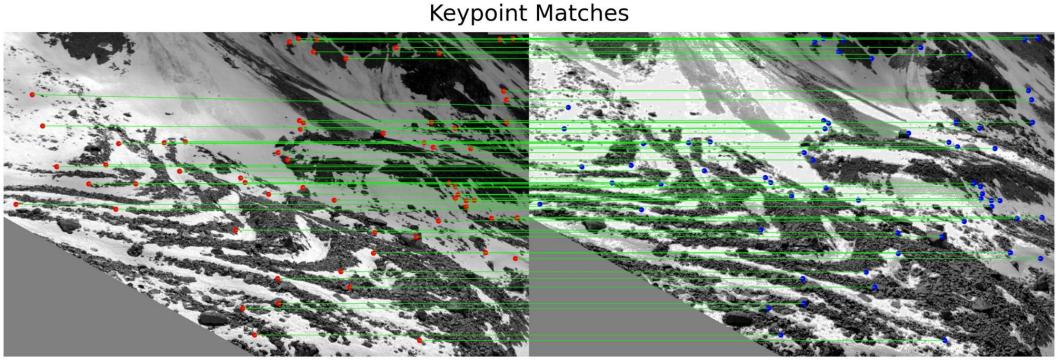


Figure 4.1: Keypoints matched by SuperGlue between 2 RGB images. Number of keypoints created by SuperPoint downsampled to 200 here for the this visualisation.

or emerging features that arise from variable snow coverage or surface changes such as avalanche debris. Figure 4.1 provides an illustration of the keypoints detected by SuperPoint and matched by SuperGlue between two sequential images.

## 4.2 Results

**Model Configuration.** For the RGB alignment process, the following SuperPoint configuration provided the best results in preliminary testing: a non-maximum suppression radius (*nms\_radius*) of 4, a keypoint threshold (*keypoint\_threshold*) of 0.1, a maximum of 600 keypoints (*max\_keypoints*), and discarding 5 pixels around each image border (*remove\_borders*). And for SuperGlue, the *outdoor* weights were used.

**Threshold Exemption.** Any estimated transformations that surpass a set threshold for any of the matrix values is seen as a failure and the image discarded in the alignment process.

**Note on Model Performance Measure.** It is difficult here to judge performance on the actual misalignment of images, as this can only be done visually by eye. This preprocess workflow and model choice combination was chosen based on visual inspection alone. While it seems like an inaccurate measure, by creating a side by side gif of unaligned and aligned images through the data subset timelines (created in section 3.1) and focusing on the position of some distinct boulders, an approximation of model alignment performance can be gained, reasonable enough to compare models to each other. This is especially apparent for the artificially misaligned set of images, as the misalignments are greatest here, see figure 4.2. Also directly measurable is the number of extreme transformations, i.e. when a model estimates an extreme realignment, such that it would be sorted out in the full workflow as described in *Threshold Exemption*.

An artificial performance metric can be defined by comparing the model's estimated affine transform to the known ground-truth transform used to misalign an image to itself. An alignment error  $E$  measures the average Euclidean distance between the transformed corner points, where a lower value of  $E$  indicates that the estimated transformation more closely matches the known ground-truth transform, reference section A.

**Results.** While it is difficult to visually depict the performance of the model on the actual image datasets in paper form, the exaggerated effect of the alignment can be seen in figure 4.2, where a few of those artificially misaligned images and their respective model-aligned counterparts are shown. This represents a simulated extreme effect of camera pylon movement, where an actual timeline of images is taken (where different lighting conditions as well as snow melt is encountered), subsequently randomly translated, rotated and warped into the artificially misaligned set (see section 3.1). Warping is hard to identify, but the black borders give an indication of the translation and the rotation performed on the images.

A further indication of the success of the proposed method is the fact that out of all 75 total images over every image subset (see section 3.1) - out of which 68 are to be aligned, as the start images of each set represent the initial reference - only 2 are visually obviously not correctly aligned to the previous and under 5 further images are visually incorrectly warped to some minor extent upon close inspection but match alignment in a section of the image. Both sets are following images in respectively the same set (full failures in *full mix*, see also figure A.2 and most partials in *almost max snowcover*).

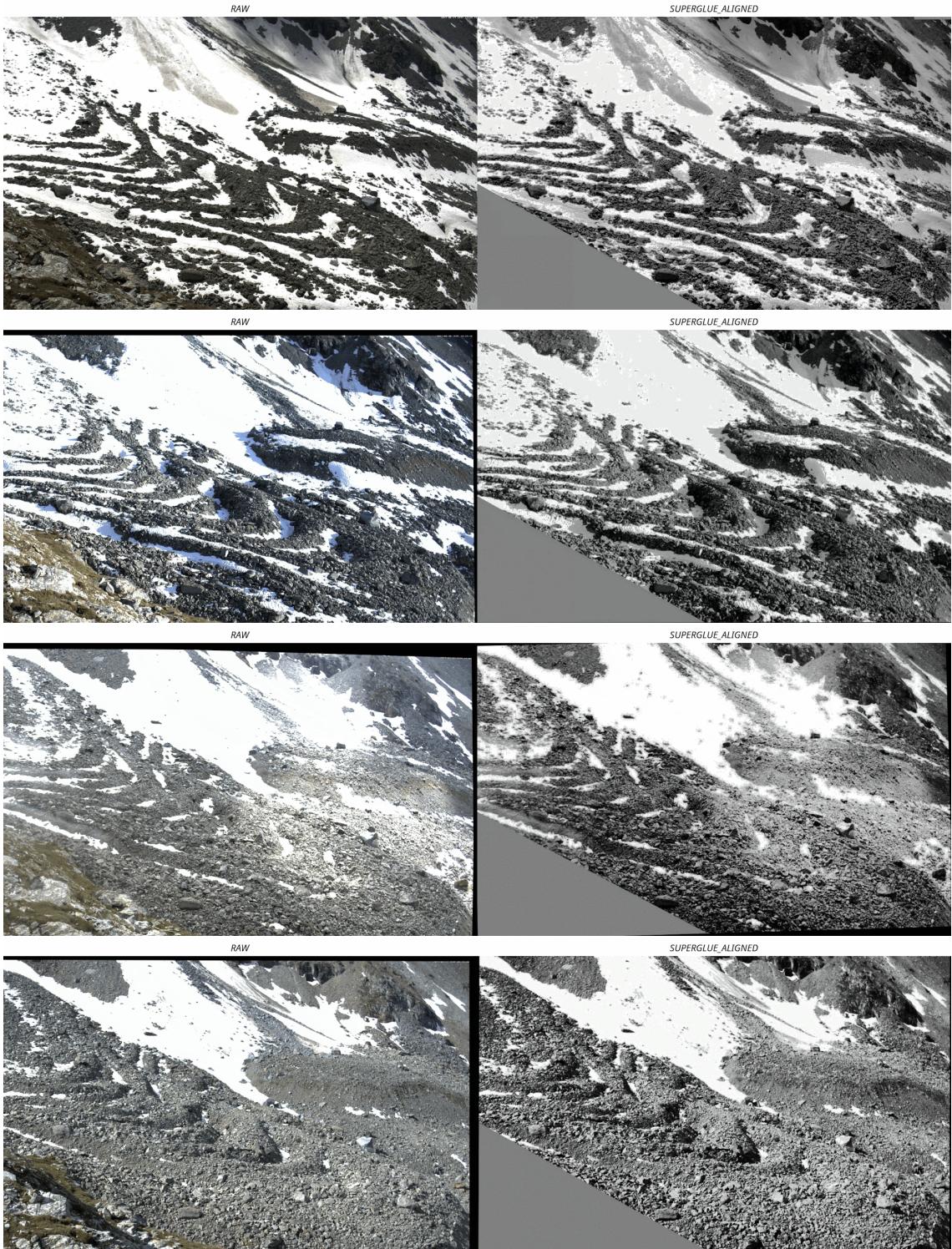


Figure 4.2: Example SuperGlue Alignment of some of the images of the artificially misaligned dataset. Images on the left represent a timeline of cropped and downsized raw RGB images (before filtering and masking) as inputs into SuperGlue, and the images on the right the applied proposed transformations.

# 5. Cross-Domain Matching

Accurate spatial alignment of TIR and RGB imagery underpins many applications, including temperature analysis of terrain features and integrated visualization of environmental changes. Although TIR and RGB images are paired up during a previous step, the alignment between the two different modalities cannot simply be performed once manually, as the pair of images are not captured simultaneously, but often minutes apart. Wind-driven pylon movements can therefore introduce misalignment for each frame. Consequently, a robust matching strategy between these images of different modalities that accounts for these ongoing shifts is necessary.

## 5.1 Template Matching for RGB–TIR

Although a purely pixel-value-based approach, template matching (Brunelli 2008) unexpectedly yielded the most reliable performance in conducted tests over keypoint based methods. Typically, direct intensity comparisons are susceptible to large variations in brightness and color. However, the tuned image preprocessing described in section 3.4 appear to mitigate these differences sufficiently to allow template matching to succeed.

This proposed process relies on the OpenCV *templateMatch* function (Itseez 2015), with cross-correlation (xcorr) as measure, which computes similarity between a designated template and a broader search area in the target image. The cross-correlation measure is given by:

$$R(x, y) = \frac{\sum_{x',y'} T(x', y') \cdot I(x + x', y + y')}{\sqrt{\sum_{x',y'} T(x', y')^2 \cdot \sum_{x',y'} I(x + x', y + y')^2}}$$

with:

- $T(x', y')$  := template function or image patch at coordinates  $(x', y')$ ,
- $I(x + x', y + y')$  := image function at shifted coordinates  $(x + x', y + y')$ ,
- $(x, y)$  := horizontal and vertical offsets of the template in the image,
- $(x', y')$  := local coordinates within the template,

For image pairs where the maximum correlation value of the best match is below

0.8, they will be removed as part of the workflow. Similarly, for matches with translations of above threshold translation compared to others.

**Rotation Invariance.** A slight rotation adjustment was introduced to also accommodate the rotational dimension (possible camera tilt). Specifically, the TIR image is rotated from  $-1^\circ$  to  $+1^\circ$  in increments of  $0.2^\circ$ , and template matching is performed for each rotation. Among these attempts, the rotation with the highest correlation score is selected as the final match.

**Automatic Inversion.** An additional complexity arises from the thermal image's brightness scale compared to its counterparts in the RGB image. Regarding a TIR image scaled from black to white as cold to hot, in winter, cold snow appears darker in TIR images, whereas warm rocks appear lighter. But on an RGB image, snow is white and hot rock features are dark. This relationship is inverted during summer conditions, where heat-absorbing rocks grow brighter in TIR imagery and cold shadows become darker. Therefore, the TIR image is sometimes inverted for better value match. Rather than manually toggling between these two states dependent on season, the model automatically chooses between the original and an inverted version of the TIR image during template matching (see figure 5.1) based on correlation of the match.

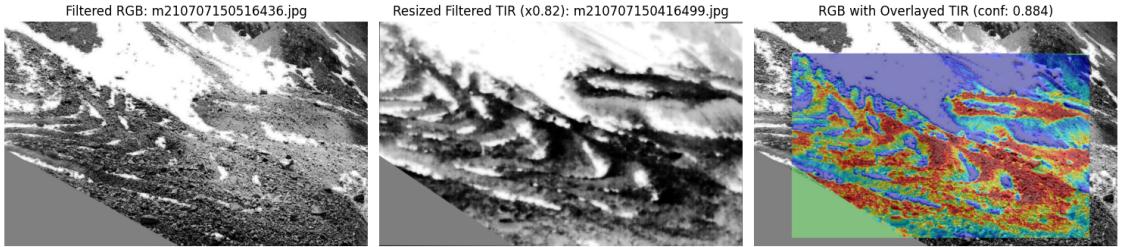


Figure 5.1: An example of the cross-domain matching, where the processed TIR image is inverted for the highest template matching confidence. Opposed to the middle row TIR images in figure 5.3, the TIR image in the middle here is depicted inverted and represents the template with which  $\text{xcorr}$  is calculated.

The image is part of the *Fullmix* dataset. (see 3.1)

Figure 5.2 illustrates why an automated decision is preferable to a fixed season-based assumption, as situations can arise (e.g., near-maximum snow cover) where the nominally wintertime condition more closely resembles the thermal patterns of summer.

**Scaling Trials** Because the TIR and RGB sensors have different focal lengths, a set of scaling experiments was conducted to find the most accurate relative scale

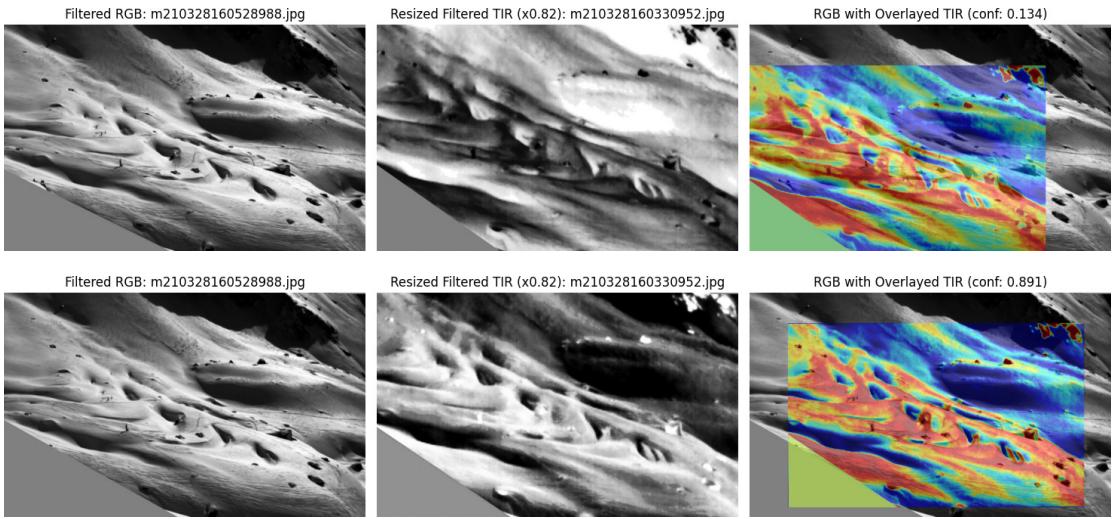


Figure 5.2: Example of TIR inversion Importance for Cross Domain Matching. The rows are the same image from the *Max Snowcover* (see 3.1) set, matched once with inverted TIR template (top) and non-inverted (bottom). The bottom match shows correct TIR–RGB alignment.

for TIR imagery. A range from 0.75 to 0.90 was sampled, and scaling by a factor of 0.82 consistently produced the best correlation of the templates. This factor was verified through visual inspection, reflecting the closest overlay of TIR features on the RGB frames. To account for possible non-uniform scaling along different axes, independent scaling of the  $x$ - and  $y$ -dimensions was also tested. However, a uniform scaling factor was sufficient to handle slight focal length discrepancies and enable robust template matching.

## 5.2 Results

Applying the proposed rotation, inversion, and scaling strategy to the prefiltered image pairs produced consistently high correlation scores. Chosen rotations ranging from  $-0.4^\circ$  to  $+0.6^\circ$  were selected by the matching routine in practice, indicating that the additional rotation degree of freedom was relevant but remained within a narrow range. All tested images achieved a correlation coefficient of at least 0.82, with an image pair within the *Full Mix* set reaching a maximum correlation of 0.96. Notably, the *Max Snowcover* set yielded the highest average correlation. Examples of TIR matching results are visible in figure 5.3. Lower performance was observed in scenarios featuring extensive shadows or noisier TIR images, particularly during lower-light conditions.

Since there is no direct ground-truth for TIR overlay, evaluation depended on visual

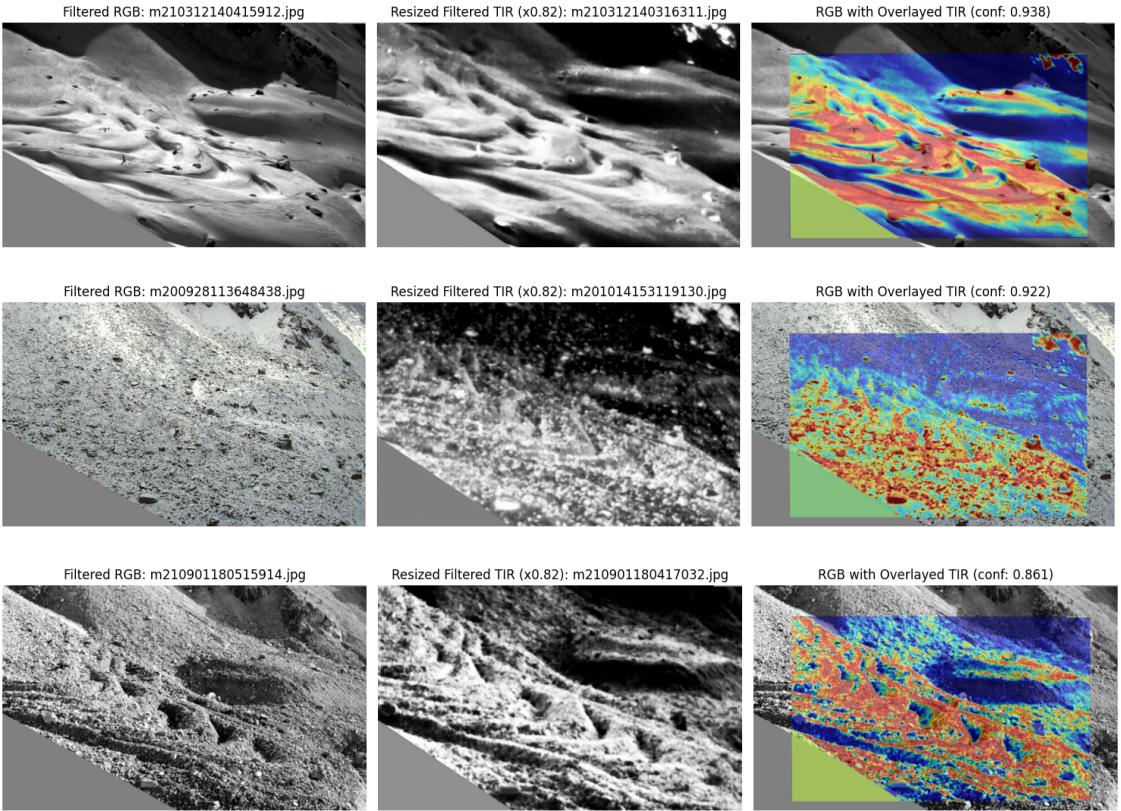


Figure 5.3: A infrared image overlaid in blue-red colorscheme over the RGB image with the proposed template matching workflow on the right column of images. The left column of images are the preprocessed RGB images, with the middle column the processed (not showing the automatic conversion here if it would apply). The rows represent example image pairs from the datasets *Max Snowcover*, *TIR Date Offset* and *Summer First Snow* from top to bottom respectively (see 3.1).

inspection. Out of 65 natural images of the test sets (excluding the *Artificially Misaligned* set), only 2 exhibited a small but noticeable misalignment. Figure A.3 illustrates one such failure, where the final TIR overlay is shifted slightly relative to the expected position.

# 6. Outlook

The workflow described in this study transforms raw RGB and TIR images into a cross-domain- and geospatially-aligned timeline of image data, as illustrated in figure 6.1. By filtering out non-viable frames using an SVM-based approach, aligning RGB images with SuperGlue, and matching those images to TIR data via template matching, the pipeline demonstrated reliable performance on the selected subsets of the Murtèl rock glacier dataset. Although only imagery from this particular site was employed in the analyses, the only domain-specific component is the manually defined cropping and masking step to remove foreground elements. Consequently, the method could be generalizable to other locations.

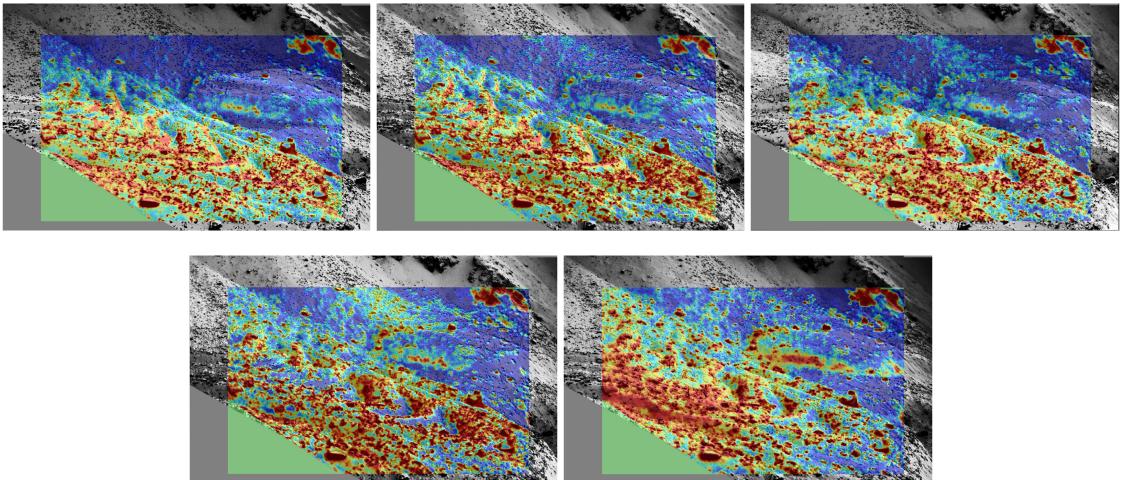


Figure 6.1: Example of the entire workflow including filtering, pairing, alignment and cross-domain matching on a subset of the *Medium Snowcover* set (see 3.1).

## 6.1 Improvements

**Filtering.** For both RGB and TIR filtering, it is apparent that some of the images falsely identified as clear just have a clear ledge in the foreground, yet the actual rock glacier behind it is obscured by fog. This issue should be mitigated with a similar masking or cropping + masking preprocessing as done in section 3.2 of the ledge in front before the filter operation, which in turn likely improves the precision of this task.

**RGB Alignment.** While the SuperGlue model has yielded successful alignment results in this project, its capabilities may exceed the requirements imposed by the relatively small movements observed in the images. Consequently, simpler approaches, such as direct feature-based methods, might be more efficient for this use case. On a hardware level, (artificial) GCP, clear horizon line or static features within the camera frame likely also simplifies the alignment process. In addition, aligning images over longer timescales introduces the risk of incremental errors accumulating across consecutive frames. Without intervention, these small distortions may compound into noticeable misalignments. To mitigate this issue, a systematic strategy, such as periodic re-initialization, should be incorporated into the workflow.

**Cross-domain Matching.** The issue with pixel based matching such as xcorr based template matching is that it relies on appearance/intensity similarity of the different image modalities. While this apparently can be achieved to a sufficient degree here using applied filters, it poses the question of robustness especially for seasonal transitional periods. An example of this failing can be seen in figure A.4, where theoretically incorrect inversion achieves a higher confidence, whilst misaligning the TIR image. A similar approach to *threshold exemption* as described in section 4.2 can obviously handle these extreme misplacements by removing the associated images from the timeline. Nonetheless this represents an open issue for this task.

## 6.2 Possible Downstream Analysis

The established workflow enables a range of extended applications. With consistently aligned 4 dimensional (RGB and TIR) imagery, it becomes feasible to map digital elevation models (DEMs) by projecting the captured scenes onto known topographic data, as this can now be done once and subsequently surface features as well as temperature can be analyzed over time directly on a height map. This framework also allows monitoring snow coverage over seasonal cycles, detecting sudden events such as rock falls or avalanches through changes in the image sequence, and tracking surface deformation indicative of rock glacier movement, see examples in figure A.6.

It was discovered, that the proposed RGB alignment can also directly be applied to estimate surface movement using images further apart in time, see section A.

## Acknowledgements and Reproducibility

This project was part of the CTCS lecture as a module taught at the Universität Innsbruck.

The images were acquired in the framework of the PERMA-XT project run by University of Fribourg/GEOTEST/PERMOS.

Special thanks goes out to the Project Partners Dr. Dominik Amschwand and Prof. Jan Beutel for providing insights into the domain as well as their constructive cooperation.

All code required to reproduce the experiments conducted can be found on github at [https://github.com/minimops/murtel\\_ctcs](https://github.com/minimops/murtel_ctcs).

## List of Abbreviations

Expression	Representing
MRG	Murtèl Rock Glacier
RF	Random Forest
SVM	Support Vector Machine
RBF	Radial Basis Function
RBF	Radial Basis Function
RGB	Red Green Blue (Color Image)
TIR	Thermal Infrared Image
DL	Deep Learning
CNN	Convolutional Neural Network
xcorr	Cross-Correlation
GCP	Ground Control Points
DEM	Digital Elevation Map

# List of Figures

1.1	Example images Murtèl Rock Glacier . . . . .	2
1.2	Example RGB and TIR image pair . . . . .	3
2.1	Extreme Examples unusable RGB Images . . . . .	5
2.2	Example unusable TIR images . . . . .	7
2.3	Entropy distributions of good/bad TIR images . . . . .	7
3.1	Example of the entire preprocess workflow . . . . .	11
4.1	Keypoints matched by SuperGlue between 2 RGB images . . . . .	14
4.2	SuperGlue Alignment of artificially misaligned set . . . . .	16
5.1	Example of TIR inversion for Cross Domain Matching . . . . .	18
5.2	Example of TIR inversion Importance for Cross Domain Matching . .	19
5.3	Cross Domain Matching Examples . . . . .	20
6.1	Full Workflow Example on Medium Snowcover Image Set . . . . .	21
A.1	Comparison of SVM Kernel Performance . . . . .	26
A.2	Obvious Alignment Process Failures . . . . .	27
A.3	Example of Cross Domain Matching Failure . . . . .	28
A.4	Misalignment of Cross-Domain Matching due to Inversion . . . . .	29
A.5	RGB-TIR Trials with SuperGlue . . . . .	30
A.6	Rock Glacier Movement over 2 Years . . . . .	30
A.7	Movement estimation of Murtèl over 2 Years . . . . .	31

# List of Tables

2.1	Feature importance SVM . . . . .	6
2.2	Performance Measures SVM Filtering . . . . .	6
3.1	Overview of the different dataset selections and their key characteristics	10
A.1	Precision RF . . . . .	26

# Bibliography

- Arioli, S., Picard, G., Arnaud, L., Gascoin, S., Alonso-González, E., Poizat, M. & Irvine, M. (2024), ‘Time series of alpine snow surface radiative temperature maps from high precision thermal infrared imaging’.
- Breiman, L. (2001), ‘Random forests’, *Machine Learning* **45**(1), 5–32.  
**URL:** <https://doi.org/10.1023/A:1010933404324>
- Brunelli, R. (2008), ‘Template matching techniques in computer vision’.
- DeTone, D., Malisiewicz, T. & Rabinovich, A. (2018), Superpoint: Self-supervised interest point detection and description, in ‘2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)’, pp. 337–33712.
- Ioli, F., Dematteis, N., Giordan, D., Nex, F. & Pinto, L. (2024), ‘Deep learning low-cost photogrammetry for 4d short-term glacier dynamics monitoring’, *PFG-92* p. 657–678.
- Itseez (2015), ‘Open source computer vision library’, <https://github.com/itseez/opencv>.
- Krainer, K. & Mostler, W. (2000), ‘Rock glacier research initiative’. Accessed: 15 Feb. 2025.  
**URL:** [https://www.uibk.ac.at/projects/rockglacier/rockglacier\\_intro.html](https://www.uibk.ac.at/projects/rockglacier/rockglacier_intro.html)
- Mammone, A., Turchi, M. & Cristianini, N. (2009), ‘Support vector machines’, *Wiley Interdisciplinary Reviews: Computational Statistics* **1**(3), 283–289.
- Rublee, E., Rabaud, V., Konolige, K. & Bradski, G. (2011), Orb: An efficient alternative to sift or surf, in ‘2011 International Conference on Computer Vision’, pp. 2564–2571.
- Sarlin, P.-E., DeTone, D., Malisiewicz, T. & Rabinovich, A. (2020), SuperGlue: Learning feature matching with graph neural networks, in ‘CVPR’.
- Ship, E., Spivak, E., Agarwal, S., Birman, R. & Hadar, O. (2024), ‘Real-time weather image classification with svm’, *arXiv preprint arXiv:2409.00821* .
- Tsai, D.-Y., Lee, Y. & Matsuyama, E. (2008), ‘Information entropy measure for evaluation of image quality’, *Journal of digital imaging* **21**, 338–347.
- Zhao, F., Huang, Q. & Gao, W. (2006), Image matching by normalized cross-correlation, in ‘2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings’, Vol. 2, pp. II–II.

# A. Appendix

## RGB–Filtering using different SVM Kernels



Figure A.1: Comparison of SVM Kernel Performance

## Filtering RGB Images with RF

A random forest (Breiman 2001) is a decision tree-based learning method known for its ability to process high-dimensional data and perform well in classification tasks. This model was also tested for RGB image filtering and performed with 88.04% precision (Table A.1), not as well as SVM (Table 2.2), yet with a higher accuracy.

	Predicted Bad	Predicted Clear
Actual Bad	33.83%	3.67%
Actual Clear	4.83%	57.67%
<b>Precision: 88.04%</b>		

Table A.1: Precision RF

## Proposed Alignment Error Metric

Alignment Error  $E$  to be a measure of how well a model aligns an artificially misaligned image to its normal self.

$$E = \frac{1}{4} \sum_{k=1}^4 \left\| M_{\text{gt}} c_k - M_{\text{est}} c_k \right\|_2 \quad (\text{A.1})$$

normalized to [0,1] via:  $E' = \min(1, \frac{E}{E_{\max}})$

with:  $M_{\text{gt}}$  := ground-truth affine transform,

$M_{\text{est}}$  := estimated affine transform,

$\{c_k\}_{k=1}^4$  := corner points in homogeneous coordinates.

$E_{\max}$  := Diagonal length of image in pixels.

By applying multiple artificial misalignments (translation, rotation, warping, or their combination) and averaging  $E$  across these scenarios, a single performance metric can be produced for the alignment process.

## SuperGlue Alignment Failures

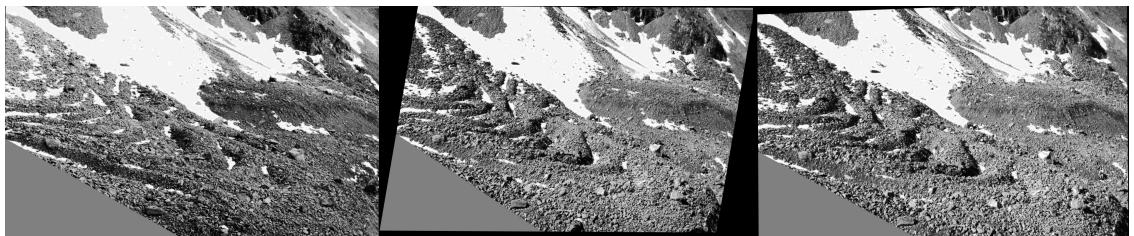


Figure A.2: Obvious Alignment Process Failures encountered while aligning the *full mix* image set with the above described SuperGlue process. The leftmost image is still correctly aligned to its predecessor, but the middle and right images are visually obviously misaligned.

## Other RGB–RGB Alignment Methods

Xcorr template matching was initially trialed for the color image alignment task, where a manually selected subsection of an image was mapped to the next. This resulted in a visually measured failure rate of .25, especially when conditions such

as snow coverage changed between the two images. This led to investigations into keypoint based methods for more robust alignment.

Some other keypoint extraction models were also inspected, mainly ORB (Rublee et al. 2011). First results were not promising and clearly inferior to SuperGlue, leading to the abandonment of further testing with this method.

## TIR Alignment Failures

For some select images, the TIR overlay method proposed, mismatches the RGB image and/or the alignment to the previous image. This can be seen in figure A.3, where the second row is visibly shifted to the right compared to the previous (top row) image pair and likely to the right of where it should be, even though the matching confidence is high. This is therefore a direct failure of the method that cannot be automatically caught by the full workflow, as the shift is neither great nor is the confidence low. Important to mention is however, that firstly, the input RGB images to the cross-domain matching method of this example are not previously aligned here and that secondly, the automatic TIR inversion was not yet introduced, when this example was made.

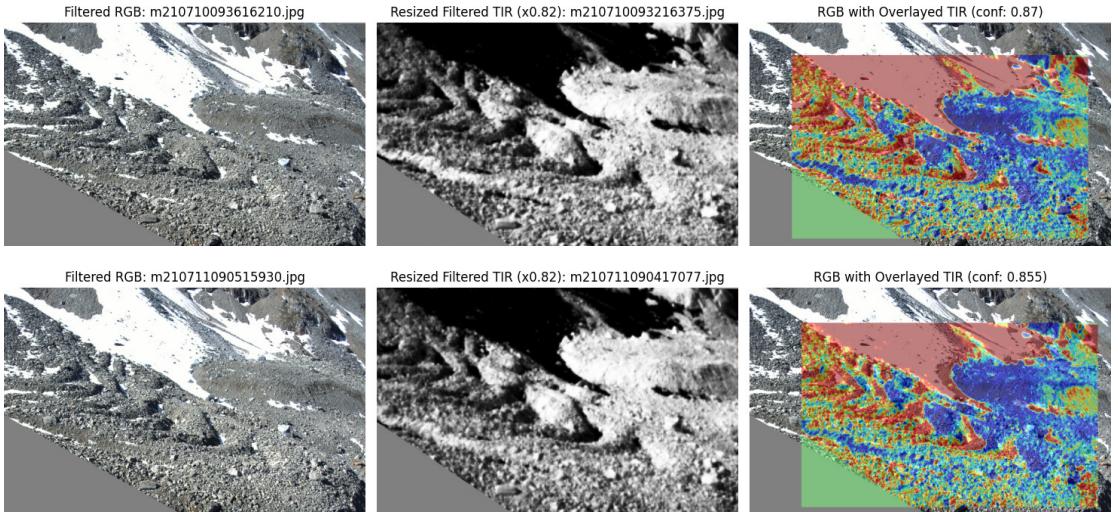


Figure A.3: Two subsequent image pairs, part of the *Fullmix* dataset (see 3.1), where a visible incorrect shift of the overlaid TIR data is noticeable.

Figure A.4 highlights the problem of automatic TIR template inversion choice for template matching. To note here is however, that this is an outlier in the *Summer First Snow* dataset and a possible result of the light fog present in the top right, which observably has an impact on the TIR image. This image may likely have been removed by the filtering workflow, which this example has not undergone, removing

this specific encountered problem.

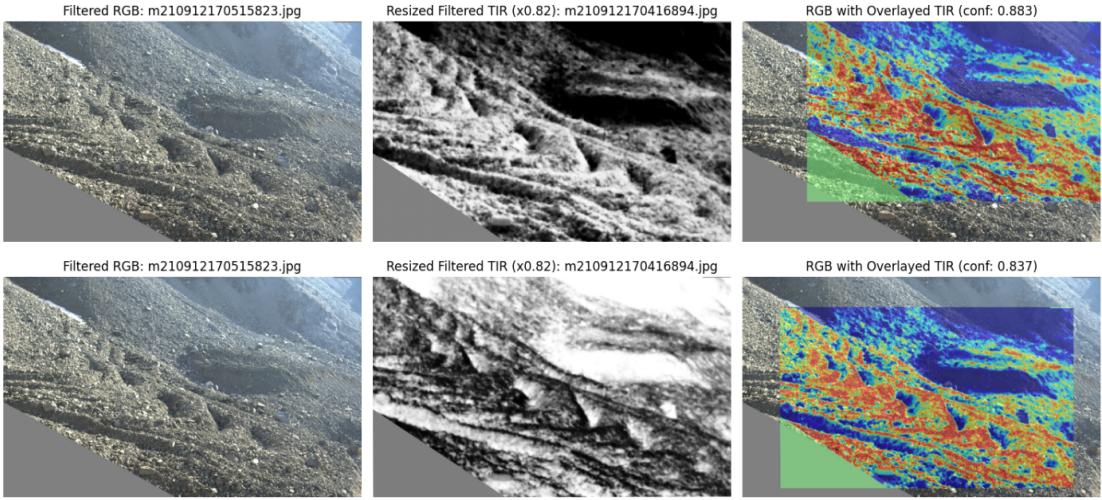


Figure A.4: These 2 rows as part of the *Summer First Snow* set represent the same image, matched with the non-inverted (top) and inverted (bottom) TIR. Without *Threshold Exemption*, the proposed model would prefer the top row, as it has a higher confidence, which is however observably incorrectly matched. The inverted TIR handles matching correctly.

## SuperGlue for TIR–RGB Matching

While SuperGlue seemed to additionally create satisfactory alignment between TIR images, initial testing did not yield promising results for cross-domain matching. Different preprocessing strategies were trialed here, matching edge extracted images, different sharpening or blurring convolutions. Even though for single select image pairs the alignment seemed visually matched, none of which showed robustness over multiple images. See figure A.5.

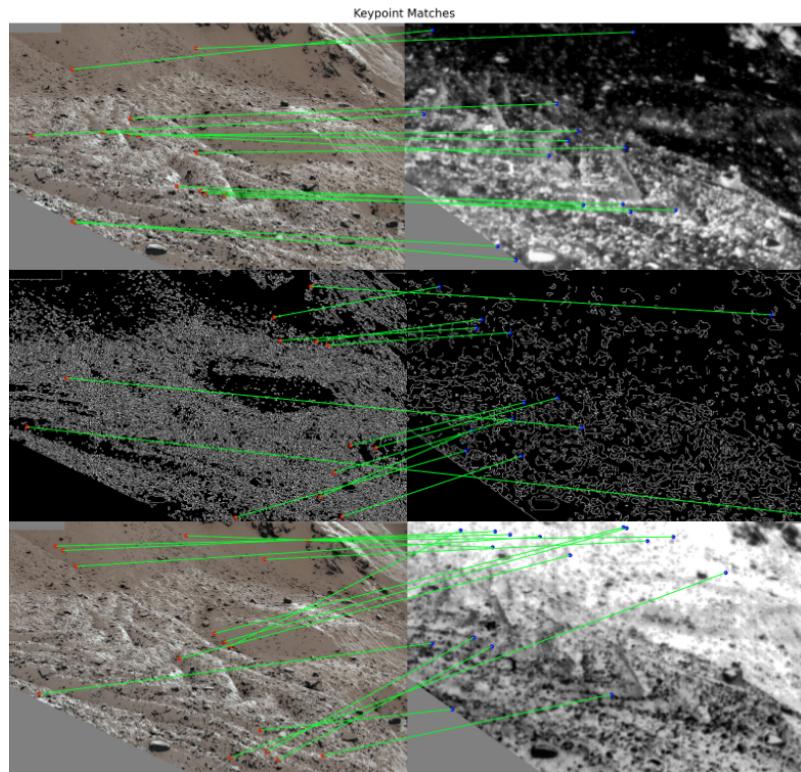


Figure A.5: Examples of Keypoint matches between RGB and TIR images using different preprocessing filters. The top row pair shows a working example with blurring & sharpening filters, the middle row edge-extraction matching, which fails and the bottom an inverted TIR matching which also visibly fails.

## Rock Glacier Movement Estimation



Figure A.6: These 2 images show the rock glacier 2 years apart (October of 2019 and November of 2021) in raw RGB, aligned by the small rock face on the top right of the image only. The right image shows new boulders that were not present two years before, encircled red.

The RGB alignment process proposed in this project can be hijacked to estimate the rock glacier movement over time. This is done by first masking all non static portions of both images, aligning the latter with SuperGlue and applying the transformation, then masking the specific section of interest and realigning the second image with superglue. The resulting affine transformation matrix of the second alignment is then an estimation of movement between those two images. This was trialed for Murtèl on two images 2 years apart as seen in figure A.6. Initial masking reveals the top right corner only and secondary masks hide all but the rock glacier section visible in figure A.7. The quiver arrows visualize the estimated transformation matrix in that area.

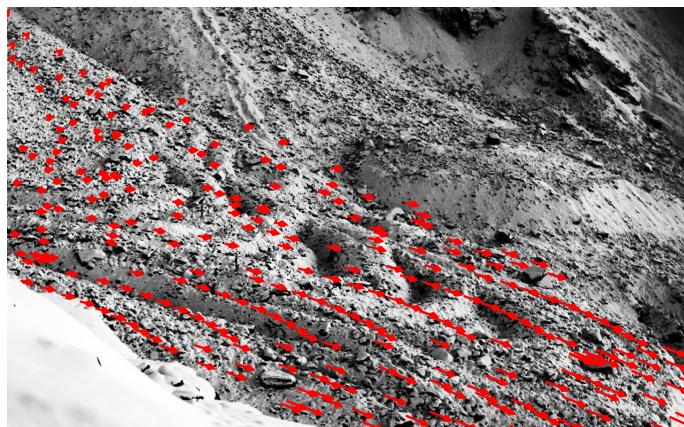


Figure A.7: Secondary transformation matrix visualized as quivers to indicate rock glacier movement over two years. Scaled up by a factor of 5 for visibility.