A logo with text on it

Description automatically generated

CEE 609 Environmental Data Science

**Finding safety margin of co-registering the optical and thermal camera field of view**

Provided by:

**Fatemeh Rezaei**

**Fall 2024**

# Introduction

Urban pluvial flooding occurs when local precipitation intensity exceeds the capacity of drainage systems. While stream gage networks and flood risk maps are crucial for federal management of coastal and riverine flood risk, there is a gap in real-time monitoring or flood risk assessment of urban pluvial flooding. Early detection and monitoring of flood events are vital for timely response and mitigation. This research proposal aims to fill this gap by developing a low-power distributed sensor network with a camera-based flood mapping platform (Urban Flood Observation Network or UFO Net) for real-time monitoring and automatic high-water mark mapping, with the potential to improve flood management significantly. The UFO-Net Mapping Node Flood Viz has two cameras: a multispectral camera (capturing red, green, blue, and near-infrared radiation) and a long-wave infrared thermal camera. Multispectral cameras differentiate land cover types and water bodies, while thermal cameras identify flood-related temperature changes. The flood mapping device then uses edge-AI to classify a 5-band image (RGB, NIR, LWIR) into a bitmap of inundated/not inundated pixels. This research develops a rapid, replicable workflow for co-registering the optical and thermal camera field of view. This step is critical because a well-calibrated and co-registered flood detection system enhances performance, reliability, and utility, allowing precise data interpretation and fusion to monitor and manage flood events effectively. Our final goal is to generate a combined five-band (RGB, NIR, LWIR) image from the two cameras rapidly on edge and find the RMSE of the coregistration algorithm when the camera is moving. We must identify how often we need to recalibrate the corregistration parameters for a camera in operational use. To achieve this, the thermal image must be statistically up-sampled to a higher resolution using linear interpolation, and key points (or common features in both images) must be automatically detected and aligned. Additionally, an IMU that would give us a reliable Euler angle would help us detect the movement of the camera. Consequently, an optimized registration method with a short computational time was explicitly developed for Raspberry Pi 4.

# Literature Review

There are many algorithms for registering images; in 2021 González-Pérez, Sara, et al. assessed some of them.[1] To register the images obtained with the visible and IR cameras, they need to calculate the calibration parameters of each sensor and the relative pose between them. For this matter, they used the Zhang method before starting to register. This prototype uses low-cost and off-the-shelf available sensors in thermal infrared and visible spectra. Four different techniques, including Geometric Optical Translation, Homography, Iterative Closest Point (ICP), and Affine transform with Gradient Descent, have been implemented and analyzed for the registration of images obtained from both sensors. All four algorithms’ performances were evaluated using the Simultaneous Truth and Performance Level Estimation (STAPLE) with several overlapping benchmarks, such as the Dice coefficient and the Jaccard index. As a result, Affine-ASGD performed the best in both evaluation tests.

Iteration helps us find the most accurate affine transformation. The Simple ITK toolkit provides us with a combination of Iterative Closest Point (ICP) and Affine transform. SimpleITK is a new interface to the Insight Segmentation and Registration Toolkit (ITK), designed to facilitate rapid prototyping, education, and scientific activities via high-level programming languages[2]. ITK is an open-source C++ toolkit actively developed over the past 18 years and is widely used by the medical image analysis community. It includes several hundred classes for image analysis, including a wide range of image input and output, filtering operations, and higher-level components for segmentation and registration. There are many options for choosing a transformation method, which is computed based on the image geometry, and there are also many similarity metrics for minimizing the difference between two images. The similarity metric reflects the relationship between the intensities of the images (identity, affine, stochastic…). The available metrics include Mean Squares, Demons, Correlation, ANTS Neighborhood Correlation, Joint Histogram Mutual Information, and Mattes Mutual Information[3].

One of the other popular methods for registration is the SIFT (Scale-Invariant Feature Transform) method. SIFT is a feature detection algorithm employed to detect key points or standard features in multispectral and thermal images. Turner, D. and other colleagues in 2014 [4] Wrote an article investigating the use of Unmanned Aerial Vehicles (UAVs) as tools for environmental remote sensing. They used multiple sensors for vegetation monitoring. Collecting data with numerous sensors requires an accurate spatial co-registration of the various UAV image datasets. In this study, they used an Oktokopter UAV to investigate the physiological state of Antarctic moss ecosystems using three sensors: (i) a visible camera (1 cm/pixel), (ii) a 6-band multispectral camera (3 cm/pixel), and (iii) a thermal infrared camera (10 cm/pixel). Imagery from each sensor was geo-referenced and mosaicked with a combination of commercially available software and algorithms based on the Scale Invariant Feature Transform (SIFT). The mosaic’s spatial co-registration validation revealed a mean root mean squared error (RMSE) of 1.78 pixels. Many false matches were often removed from the millions of matches with a Random Sample Consensus (RANSAC) algorithm. Fischler and Bolles first proposed the RANSAC algorithm as a resampling technique for estimating the parameters of a model and finding data that outliers may contaminate[5].

Junchi Bin and colleagues have discussed a comprehensive comparison of these different methods. They propose a new multi-objective optimization-based image registration (MOIR) algorithm to register the visible and thermal images at the pixel level. Specifically, the MOIR contains an object function, i.e., normalized gradient measurement (NGM), which enables difference measurement between VI–T images in both angular and linear perspectives. Meanwhile, the MOIR comprises a regularized stochastic gradient descent (RSGD) that ensembles advanced techniques in stochastic optimization to solve affine parameters for registration smoothly. Quantitative and qualitative results indicate MOIR’s accuracy and robustness in registering VI–T images from controlled and real-world datasets. Thus, the proposed method applies to industrial multimodal imaging systems such as autonomous driving and surveillance. Figure 1, clearly shows their results which is evaluated with three different method [6].

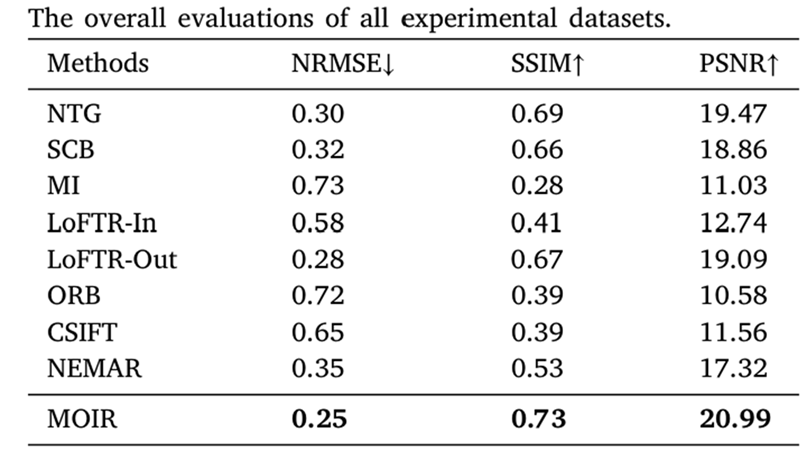


Figure -Comparison of different registration methods.

In this research, we also used an algorithm similar to MOIR and visually compared it with a combination of the SIFT method with RANSAC (Random Sample Consensus) results. We will also evaluate the movement of cameras on transformation with Root Mean Square Error or RMSE.

# Data and methods

Our raw data would be folders containing an optical image, a thermal one, and a separate NIR band image, and an Excel containing each pixel's temperature and IMU information of the Euler angle. For collecting our dataset and to capture the unique spectral signature of water and allow for day/night water detection, the Flood Viz node is equipped with two cameras: a Dorhea multispectral camera (capturing red, green, blue, and near-infrared radiation) and a FLIR Lepton long-wave infrared, or thermal, camera. A typical image type is a three-channel image where each channel has scalar values in [0,255]; often, people refer to such an image as an RGB image. This terminology implies that the three channels should be interpreted using RGB color space. We have two extra bands containing Near-Infrared (NIR) and Long-Wave Infrared (LWIR). A Near-Infrared (NIR) image band is a spectral region of the electromagnetic spectrum just outside the range of human vision. NIR bands are used in image analysis and change detection, especially for vegetation and wetlands. Long-Wave Infrared, or LWIR, is a subset of the infrared band of the electromagnetic spectrum, covering wavelengths ranging from 8[µm](https://www.infinitioptics.com/glossary/micron-um) to 14µm (8,000 to 14,000nm). This is the radiant heat that uncooled thermal imaging cameras see and detect distinct temperature differences. To gain data to start calibration, we will use a target to capture the image with our optical camera and some ice packs to better illustrate our target with our thermal camera. Then, we start taking pictures with both cameras and store our data for later. Later, we can store our data in NumPy arrays, a grid of values representing each pixel of images that are all the same type.

A screenshot of a video

Description automatically generated

Figure - process of whole water Cam project

Then, we can co-register and match the images using Python software, a toolkit called SimpleITK, and the OpenCV library. SimpleITK is a simplified programming interface to the algorithms and data structures of the [Insight Toolkit](https://github.com/InsightSoftwareConsortium/ITK) (ITK). It supports bindings for multiple programming languages, including C++, Python, and R. These bindings enable us to develop image analysis workflows in Python. The toolkit supports over 15 different image file formats, provides over 280 image analysis filters, and implements a unified interface to the ITK intensity-based registration framework. The fundamental tenet of an image in ITK and consequentially in SimpleITK is that an image is defined by a set of points on a grid occupying a physical region in space. This significantly differs from many other image analysis libraries that treat an image as an array, which has two implications: (1) pixel/voxel spacing is assumed to be isotropic, and (2) there is no notion of an image’s location in physical space. SimpleITK images are multi-dimensional (the default configuration includes images from 2D up to 5D) and can be a scalar, label map (scalar with run length encoding), complex value, or have an arbitrary number of scalar channels (also known as a vector image). The region in physical space that an image occupies is defined by the image’s:

1. Origin (vector-like type) - location in the world coordinate system of the voxel with all zero indexes.
2. Spacing (vector-like type) – the distance between pixels along each dimension.
3. Size (vector-like type) - number of pixels in each dimension.
4. Intensities in all channels or direction cosine matrix (vector-like type representing matrix in row-major order) - direction of each axes corresponding to the matrix columns.

Since an image is also defined by its spatial location, two images with the same pixel data and spacing but different alignments are not considered equivalent in SimpleITK because they occupy different spatial locations. In our case, the goal is to find points with the same pixel data in the two other images and try to align them so we also get the same location and spacing for the points. That way, we could merge the two images and read all necessary data simultaneously from a combined 5-band image. For this matter, we should find the transformation of the optical image and then rotate the thermal image accordingly. Points in SimpleITK are mapped by the transform using the Transform Point method. Various global 2D and 3D transformations are available (translation, rotation, rigid, similarity, affine…).

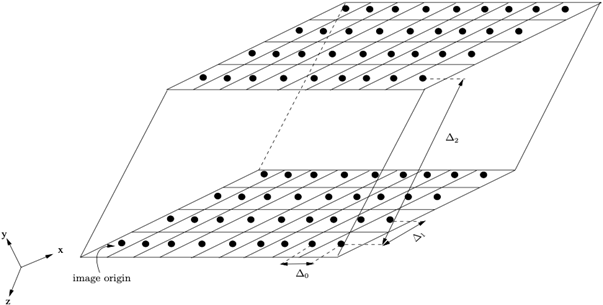


Figure - finding tie points in different photos[3]

Our Co-registering process has two major parts: first, finding the appropriate type of transformation between optical and thermal images, and second, applying the transformation to our thermal image, which also contains merging two images and creating a 5-band image. The figure below shows a flow chart of each step in this study.

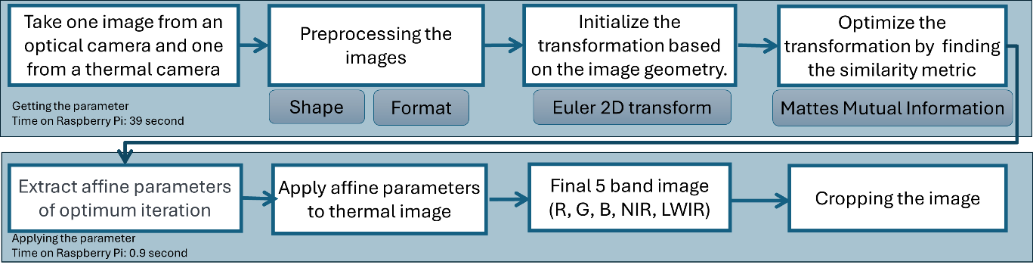


Figure -Flow chart showing research steps

## Finding transformation

We convert both images into SimpleITK format and set the optical image as fixed and the thermal image as moving. If the thermal image has a lower pixel resolution than the multispectral image, we apply an up-sampling technique to increase its resolution. Since the thermal image is inherently continuous and lacks color information, linear interpolation methods can be used for up-sampling. This process increases the pixel density of the thermal image to match the resolution of the multispectral image.

After resizing them to the same size, we initialize the transformation based on the image geometry. Among all transformation options, we used the Euler 2D transform with rotation in radians around a fixed center with translation. To find the best transformation alignment, we should iterate the process and choose a proper similarity metric to calculate our accuracy until we reach the defined threshold. The similarity metric method reflects the relationship between the intensities of the images (identity, affine, stochastic…). The similarity metric used here is “Mattes Mutual Information” to align the thermal image to the optical one. Gradient descent optimization is employed to minimize the difference between the images. (Computes the mutual information between two images to be registered using the method of Mattes). The final transformation is computed after 100 iterations. We could save the transformation as an h5 (HDF5) file to use later.

## Apply transformation

The thermal image is resampled using the computed transformation to align with the optical image. The final combined image (optical + thermal) is generated using cv2.addWeighted, which overlays the two images. The NIR band stored with the optical image is extracted and combined with the RGB + LWIR thermal bands to form the final 5-band image. The image is then cropped to remove unnecessary regions based on a bounding box around the thermal image. Finally, the function returns the transformation, the final combined image, and the cropped 5-band image output. Our project’s final goal is to use the 5-band image for the segmentation process and train it to detect water. When a flood occurs and water levels rise, our sensor can detect it and call the authorities.

Generally speaking, repeated registrations of the same datasets will yield slightly different results. This is associated with the registration framework’s similarity metric computation implementation, which utilizes randomization and multi-threading.

# Results

The key result of this project is finding tie point between images and creating five-band images containing RGB, NIR, and LWIR that are used as inputs for the segmentation process for detecting the surface inundation location.

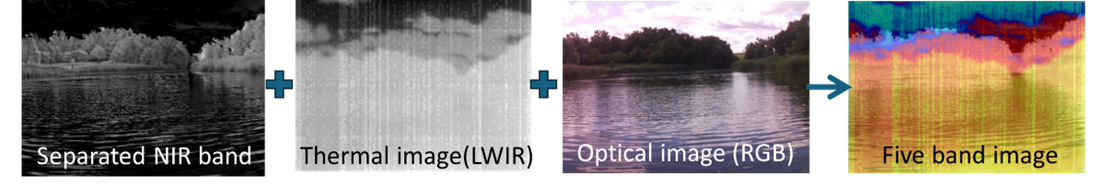


Figure - 5 different bands of images

The final images show better results than other algorithms, such as ORB (Oriented FAST and Rotated BRIEF) and SIFT (Scale-Invariant Feature Transform). Figure 4 vividly draw shifting in the ORB algorithm and how it has been fixed using the ITK tool and iteration method.

A comparison of a infrared image of a park and a building

Description automatically generated

Figure -comparison of algorithms in detecting tie points

As mentioned before, we used a pre-trained super-resolution model, called EDSR, for up-sampling our thermal image, and its result has been shown in Figure 7.

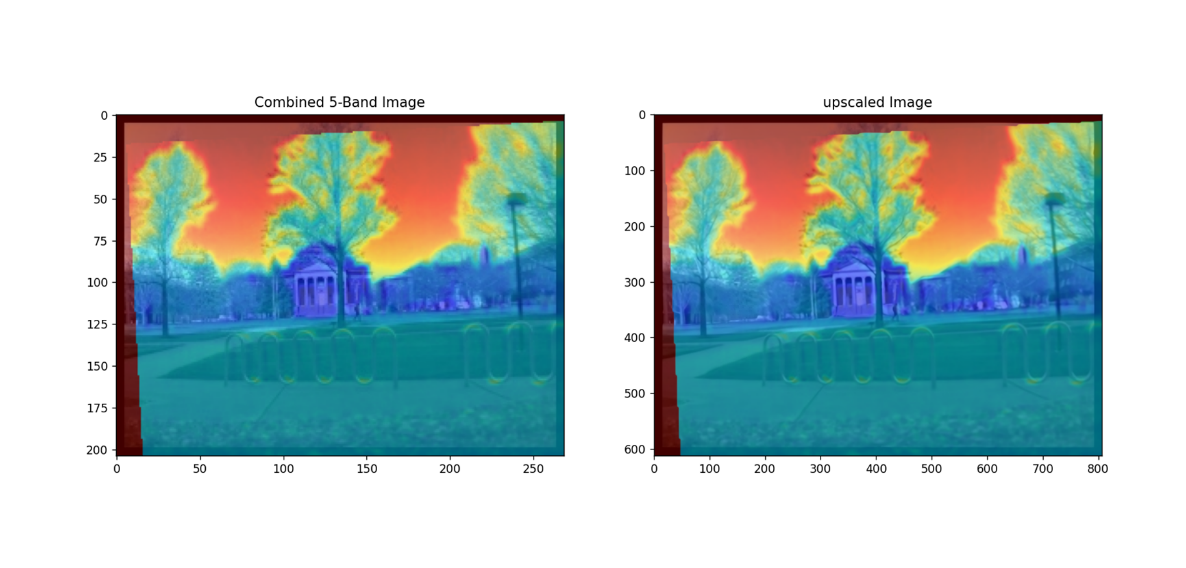


Figure - Increase thermal image quality with pre-trained algorithm

Mounting our sensor at high elevation has the challenge of unwanted movement of cameras. Running registration algorithms are either power-efficient or time-efficient. The result has been shown that moving the camera box by 10 degrees in each direction would not distort the transformation, and we could still use the central transformation for other images and not start the iteration process over.

A collage of a room with a door

Description automatically generated

Figure -Applying the Center image’s affine transformation to moved images in different directions

# Further discussion

The next step in this project is finding the exact allowable movement of the camera and comparing the RMSE to determine the threshold at which we should rerun the algorithm to find the best transformation for our picture.

In this experiment, the implemented data-driven techniques effectively address perspective transformation, but they lack the precision of affine transformations, a key advantage of the proposed MOIR method. In future developments, MOIR will be extended by integrating with data-driven models like LoFTR-Out to achieve a balanced combination of generalization and efficiency for real-world scenarios.

Convergence speed can still be improved. Future developments will explore the application of advanced optimizers such as the Augmented Lagrangian Method (ALM) and Alternating Direction Method of Multipliers (ADMM).

Wavelet and spectral transforms effectively extract image gradients and edges from multiple scales. Consequently, future development will investigate the applications of these transformations in the context of our research.

# Conclusion

In conclusion, developing a flood detection system using multispectral and thermal cameras represents a significant advancement in remote sensing and disaster management. Through this project, we have addressed key challenges and contributed to advancing flood detection technology by focusing on the co-registration of images from these sensors in a power-constrained environment. Our research has demonstrated the importance of these fundamental steps in ensuring the flood detection system's accuracy, reliability, and effectiveness. By combining calibrated and co-registered images, our system can provide real-time monitoring and early warning of flood events, empowering decision-makers with timely and actionable information to support disaster preparedness, response, and recovery efforts.

# Instructions using the repository

There is a code for single image registration, which you can use by editing the file path. Another code set contains the two steps of getting transformation and applying them for multiple images, which you should run respectively after modifying the code for the saved path you used. Please email me to submit merge requests.

**References**

[1] S. González-Pérez *et al.*, “Assessment of registration methods for thermal infrared and visible images for diabetic foot monitoring,” *mdpi.comS González-Pérez, D Perea Ström, N Arteaga-Marrero, C Luque, I Sidrach-Cardona, E VillaSensors, 2021•mdpi.com*, 2021, doi: 10.3390/s21072264.

[2] R. Beare, B. Lowekamp, and Z. Yaniv, “Image Segmentation, Registration and Characterization in R with SimpleITK.,” *J. Stat. Softw.*, vol. 86, Aug. 2018, doi: 10.18637/jss.v086.i08.

[3] “Registration Overview — SimpleITK 2.4.0 documentation.” Accessed: Sep. 26, 2024. [Online]. Available: https://simpleitk.readthedocs.io/en/master/registrationOverview.html

[4] D. Turner, A. Lucieer, Z. Malenovský, D. H. King, and S. A. Robinson, “Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic moss beds,” *mdpi.comD Turner, A Lucieer, Z Malenovský, DH King, SA RobinsonRemote Sensing, 2014•mdpi.com*, vol. 6, p. 6, 2014, doi: 10.3390/rs6054003.

[5] M. A. Fischler and R. C. Bolles, “Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.

[6] J. Bin *et al.*, “The registration of visible and thermal images through multi-objective optimization,” *ElsevierJ Bin, H Zhang, Z Bahrami, R Zhang, H Liu, E Blasch, Z LiuInformation Fusion, 2023•Elsevier*, Accessed: Dec. 15, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523000623?casa\_token=PKNRU4Ld0DAAAAAA:963GUW8P0haIayKTzXC7QHLXykEXEmzE3x8BJ0KHxnsAKlP23cEBeqrv\_ZVs\_osmFXZ2ruDXqSig