ACIP LAB - Case 4

Anna, Hamza, Kimia

May 23, 2024

Contents

1 Introduction

This lab addresses the challenges of multimodal image registration, particularly in aligning images captured using different spectral filters. In this assignment, we focus on registering two images captured with a dual-camera setup using RaspberryPi cameras, where one camera uses a filter to cut off the visible spectrum and the other uses a filter to cut off the NIR spectrum. The primary goal of this lab session is to evaluate the effectiveness of image registration techniques and to assess the quality of the registration through control point (CP) location accuracy.

In this session, we focus on:

- Analyzing the effectiveness of SIFT feature detection method for multimodal image registration.
- Evaluating the geometric transformation required to align the images, particularly focusing on projective transformations due to the potential perspective distortions.
- Comparing manually selected control points with automatically detected control points using alternative feature detection methods like KAZE to evaluate registration quality.
- Calculating the average Euclidean distance between corresponding control points to quantitatively assess the registration accuracy.

By systematically experimenting with these techniques, we aim to find an effective method for achieving precise and accurate image alignment, thereby enhancing the reliability of applications that depend on image registration.

2 Methodology

2.1 Image Preparation

Multimodal image registration involves aligning images captured using different spectral information, which can result in significant differences in image appearance. In this case, the images were taken with filters that cut off the visible and NIR portions of the spectrum. Therefore, images were loaded and preprocessed to focus on specific color channels. In this assignment we solely focused on the red channel of both images to continue the registration with.

The red channel often captures more information in both visible and NIR images. In visible light photography, the red channel captures a wide range of wavelengths that overlap with the start of the NIR spectrum. This overlap can provide more consistent features between the two modalities, making it easier to find correspondences.

2.2 Initial Alignment

Initial misalignment was visually assessed using MATLAB's `imshowpair` function, which provides a clear comparison between the fixed and moving images. By looking at the images at the start we can see the potential transformations that has happened to th eimage.

The right image appears to be translated relative to the left image. This is evident as corresponding objects (books, shelves) do not align horizontally or vertically. For example, the books and items on each shelf do not match up in both images.

There may be a slight rotational difference between the two images. This can be inferred from the slight angular displacement where vertical lines (such as the edges of the bookshelf) are not parallel between the two images.

There may be some perspective distortion, especially since the images were taken with different filters and possibly slightly different viewpoints. This is suggested by the slight differences in angles and alignment of lines that should be parallel.

Considering the observed misalignments, the primary transformation required to align these images appears to be a projective transformation which later we'll see. This conclusion is due to the combination of multiple transformations that are present in the images.

Figure 1: Initial misalignment of the red plane between reference and sensed images, highlighting the starting point of the misalignment to be corrected.

2.3 Feature Detection and Matching

SIFT features were detected and matched across the images to determine corresponding points crucial for the registration process. We can see that most of the CPs that were found are suitable for registration; however there is one CP that is matched uncorrectly. Therefore, in the next section we will try to tune the matches and find the best and most suitable ones.

Figure 2: Matched points before registration, showing the initial feature matches including potential outliers using SIFT.

2.4 Feature Detection and Matching - Tuning

In the previous section it can be seen that there are some SIFT matched CPs that are not suitable to be selected for registration. Therefore, here we used MaxRatio attribute for MatchFeatures function in MATLAB, with a value of 0.4, to eliminate the uncorrect match in the previous section.

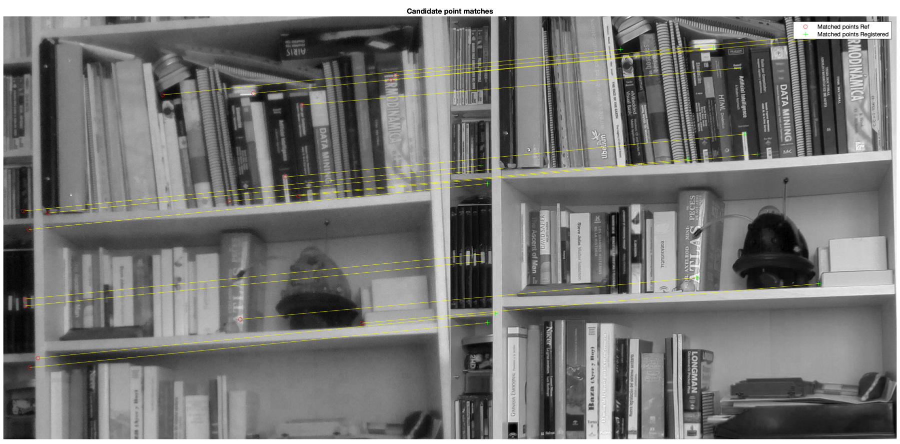

Figure 3: Tuned matched points with MaxRatio before registration.

3 Experiments and Results: Case 4: Hard Difficulty

3.1 SIFT, Manual and KAZE - Red Channel

3.1.1 Feature-Based Registration

The registration was attempted using a projective transformation based on matched SIFT features. We see that the the structural alignment of the two images appears to be accurate. The objects, such as books and shelves, are correctly aligned, confirming that the geometric registration process was successful. However, we also notice some regions in the image with color of green which indicates significant difference between the two images, although we see that the images are aligned. This is probably due to the intensity differences of the two images as they are captured with different filters. If we go back to the original images, we will notice that these green areas were very different in intensity.

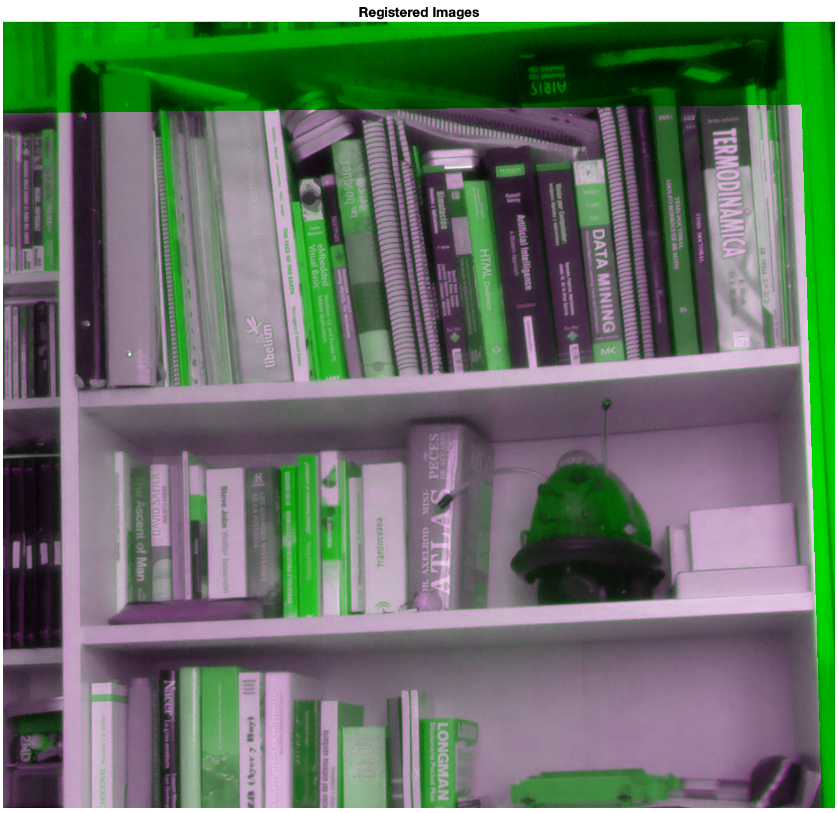

Figure 4: Registration output, done on the red channel.

3.1.2 Evaluation Metrics

The quality of registration was assessed using the following metrics. We also have to mention that since the images are very different in intensity it wouldn't be suitable to use RMSE, an intensity-based metric, to evaluate the registration, as we know we won't get good results and it won't reflect the performance of the registration:

- Euclidean errors for manually and automatically detected control points to evaluate the geometric alignment.

Bellow you can see the CPs that were manually selected with MATLAB. As you can see in the image we took care to select distinct features in critical regions of the image, so they are meaningful and can help us further in line.

Figure 5: Key points manually selected on the original and registered images, plotted on the fixed image to evaluate registration accuracy.

Next, you can see the matched CPs that were extracted by KAZE method on both the original and registered image. And also the matching of those extracted features can be seen in Figure 7. However, we still see a lot of outliers and incorrect matches, therefore we also used MaxRatio of 0.3, to eliminate uncorrect matches which can be seen in Figure 8.

Figure 6: Original and registered control points detected automatically using KAZE, displayed on the fixed image for further evaluation of registration accuracy.

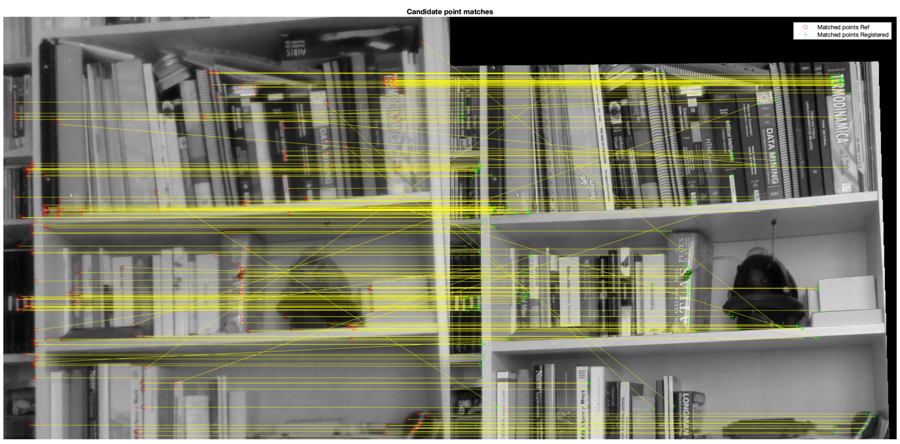

Figure 7: KAZE-based matched points showing feature matches between the fixed and registered images using KAZE, distinct from the SIFT features used for registration.

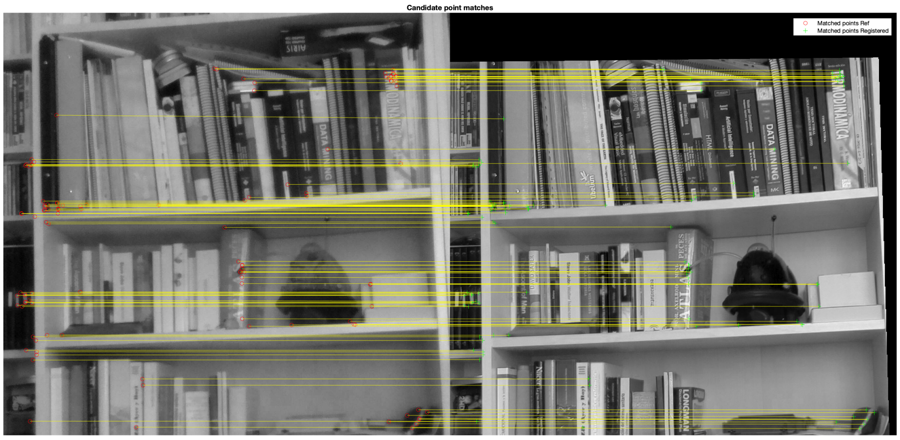

Figure 8: Tuned KAZE-based matched points using MaxRatio of 0.3.

3.1.3 Results

The image registration process was evaluated using location-based metrics, producing the following results:

- **Manual Control Point Euclidean Error:** The average Euclidean distance between manually selected control points was 1.8933 pixels. This moderate value indicates some discrepancies in the geometric alignment of specific features or regions within the images, pointing to areas where registration could be improved.
- **Automatic Control Point Euclidean Error:** An error of 2.3313 pixels was observed for the automatically detected control points. This value higher value than the manually selected CPs, which can insicate a couple of things:

  1. Manual selection of control points benefits from human expertise and intuition. We can visually identify and select the most accurate corresponding points in both images, ensuring that they are indeed correct matches. In this case, we chose 31 matches carefully, which is 3 times more in number than the 10 points where selected for case3.

  2. Manual selection allows for precise placement of control points exactly at corresponding locations. This precision can lead to a more accurate transformation and better registration results.

  3. We, as humans can understand the context and semantics of the scene. This understanding helps in selecting control points that are more robust and less ambiguous.

  The previous points, are some possibilities that the manually selected CPs have done better than the KAZE features. However, we should take into account some disadvantages of this kind of CP selection:

  1. Manual selection is labor-intensive and time-consuming, especially for a large number of images or points.

  2. There is a level of subjectivity involved, which might lead to variability in the points selected by different individuals.

These metrics demonstrate a successful registration in terms of geometric precision, particularly in the manual alignment of control points.

3.1.4 Discussion

The registration process for the multimodal images, captured using different spectral filters, demonstrated both the strengths and limitations of feature-based registration techniques. The projective transformation employed in this process effectively corrected the perspective distortions and other geometric misalignments, resulting in a well-aligned composite image. The visual inspection of the registered image shows that key features, such as the books and shelves, are well-aligned, indicating a successful transformation. However, the overlay reveals slight misalignments in some areas, suggesting that while the transformation was broadly effective, there are still minor discrepancies that could potentially be refined further. A key takeaway from this exercise is the difference in registration accuracy between manually selected control points and automatically extracted features using KAZE. The manually selected CPs provided a much lower average Euclidean distance error, underscoring the precision that human expertise can bring to such tasks. This suggests that, despite the efficiency and speed of automated methods, they may not always capture the most reliable correspondences in multimodal contexts where spectral differences are pronounced. The higher registration errors observed with KAZE features point to the challenges inherent in automated feature matching in multimodal images.

3.2 Experiment on Green and Blue Channel

The registration process was repeated using the green channel and also the blue channel of the images to assess if the color channel impacts the registration quality. In both of the cases, using SIFT, only around four control points could be detected and matched. This inadequacy resulted in very poor registration quality for the green and blue channels. We can say that using the red channel had superior results, which can indicate that the blue and green channel suffer from inadecuate features in our case, which are multimodal images with significant spectral differences. This highlights the importance of channel selection in multimodal image registration and underscores the need for robust feature detection algorithms that can operate effectively across all channels.

3.3 Experiment on greyscale images

3.3.1 SIFT feature detection and extraction for registration

The registration process was repeated, this time on the greyscale of both images. The SIFT feature extraction gave very poor results as again not enough CPs were extracted from the two images. Bellow you can see the matched CPs and the output of registration assuming a projective transformation.

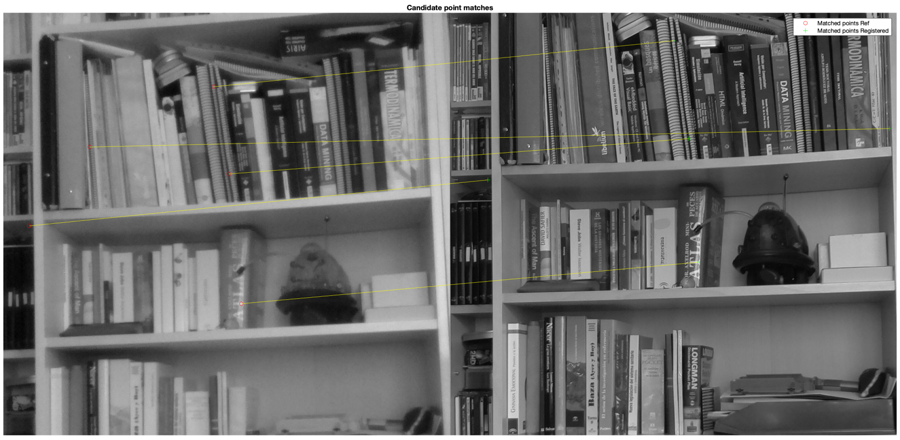

Figure 9: Matched points before registration, showing the initial feature matches using SIFT on greyscale images.

Bellow you can see the result of registration with SIFT feature extraction on greyscale images.

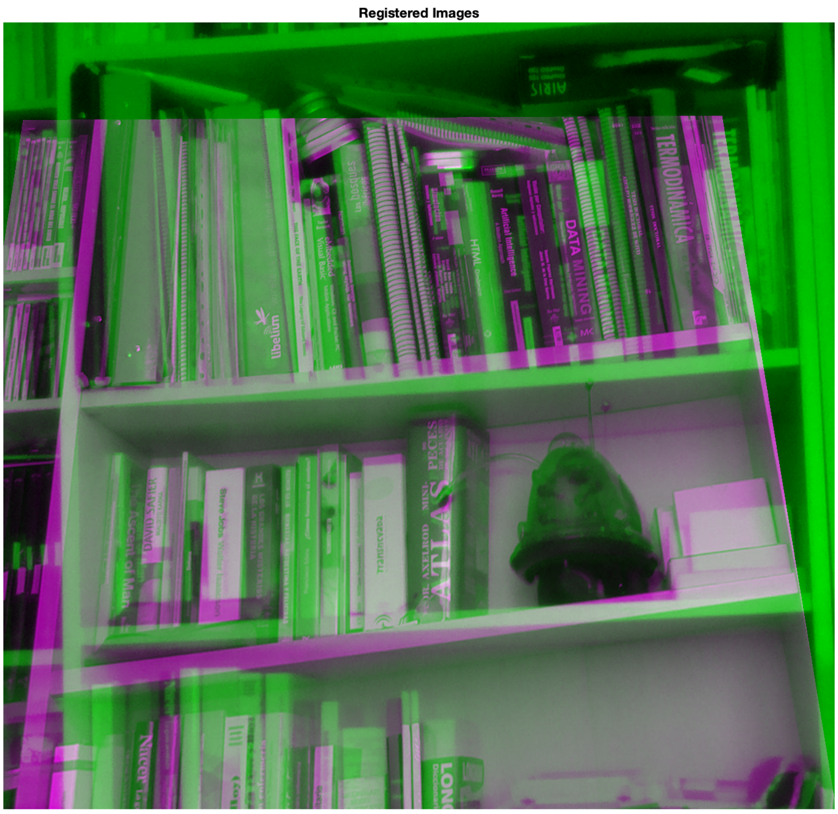

Figure 10: Registration result of greyscale images with SIFT features.

We can see that registration with SIFT features on greyscale images leads to worse results than using the red channel. This suggests that different channels or intensity representations may require specific feature detectors to achieve optimal results. The grayscale representation combines information from all color channels, which can sometimes dilute distinct features present in individual channels, especially in multimodal contexts where spectral differences are pronounced. The red channel, having a better overlap with the NIR spectrum, provided more robust and distinct features suitable for matching. This experiment indicates that the effectiveness of feature detection and matching can be highly dependent on the spectral characteristics of the channels used, and certain feature detectors may perform better with specific channels or intensity representations. Therefore, selecting the appropriate channel and feature detection method is crucial for achieving high registration accuracy in multimodal image registration.

3.3.2 SURF feature detection and extraction for registration

As we saw, in contrast to using the red channel, SIFT feature detector didn't perform well for the greyscale image. Therefore, we also experimented with using SURF feature detector in order to carry out the registration to see the importance of the feature detection method. Bellow you can see the CPs extracted from the greyscale images using SURF; it can be seen that it's higher in number than the SIFT CPs. Also we have to mention that the matches are tuned with MaxRatio of 0.55.

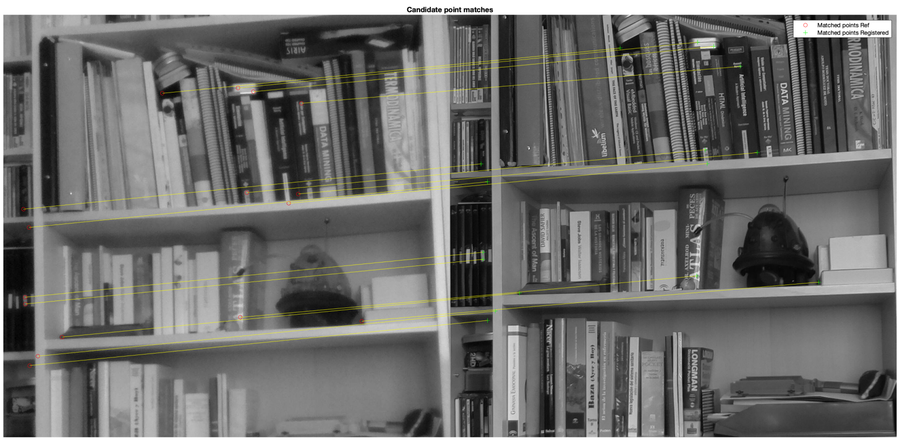

Figure 11: Tuned matched points before registration using SURF on greyscale images.

Below you can see the result of the registration of two greyscale images using SURF features.

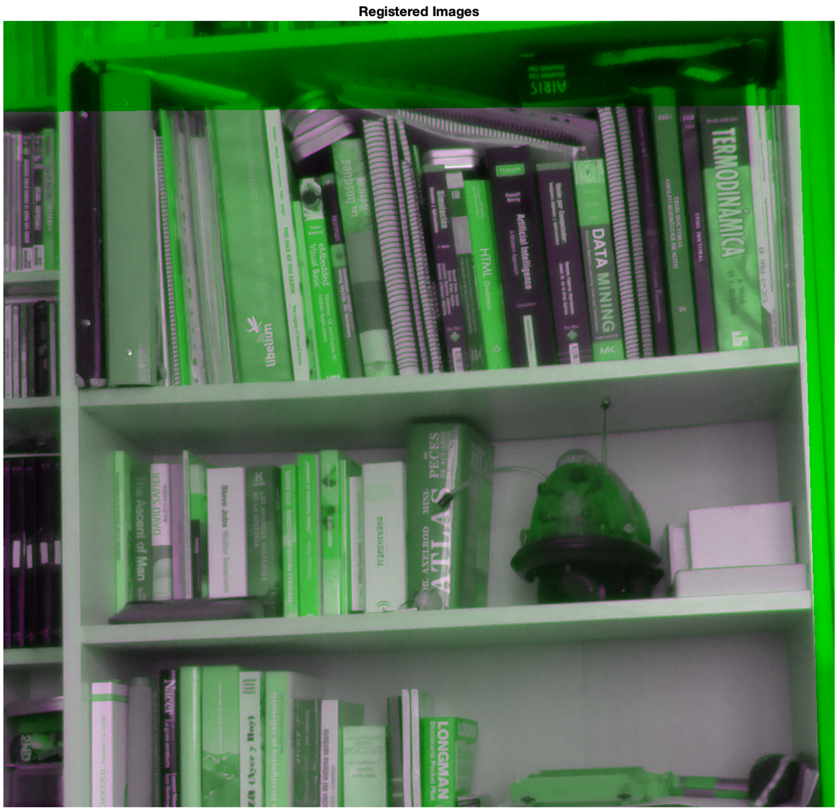

Figure 12: Registration result of greyscale images with SIFT features

3.3.3 Discussion

Using SURF on grayscale images provided better results compared to SIFT on grayscale. SURF's efficiency and robustness to various transformations make it more suitable for detecting features in grayscale representations. However, even with SURF, the registration results were still not as good as those obtained using the red channel. While both SIFT and SURF are robust feature detectors, their performance can vary depending on the image representation. SURF's better performance on grayscale images suggests that it might be more adaptable to different intensity representations, but it still falls short of the results obtained with the red channel.

3.4 Experiment with Affine Transformations

Affine transformations were tested to evaluate their effectiveness in registering images with potential to handle rotations, translations, scaling, and shearing more flexibly than projective transformations. In this section the red channles of the images were used for the registration.

3.4.1 Results for Affine Transformation

The registration using affine transformation yielded the following metrics:

- **Manual Control Point Euclidean Error:** The value was 1.89, the same as the projective transformation.
- **Manual Control Point Euclidean Error:** This value is again calculated using the KAZE feature detector, the same as before when we registered with red channel. THe value this time is 5.57 which is higher than using projective transformation

Bellow you can see the result of registration using affine transformations. you can see that there are misalignment in edges of books.

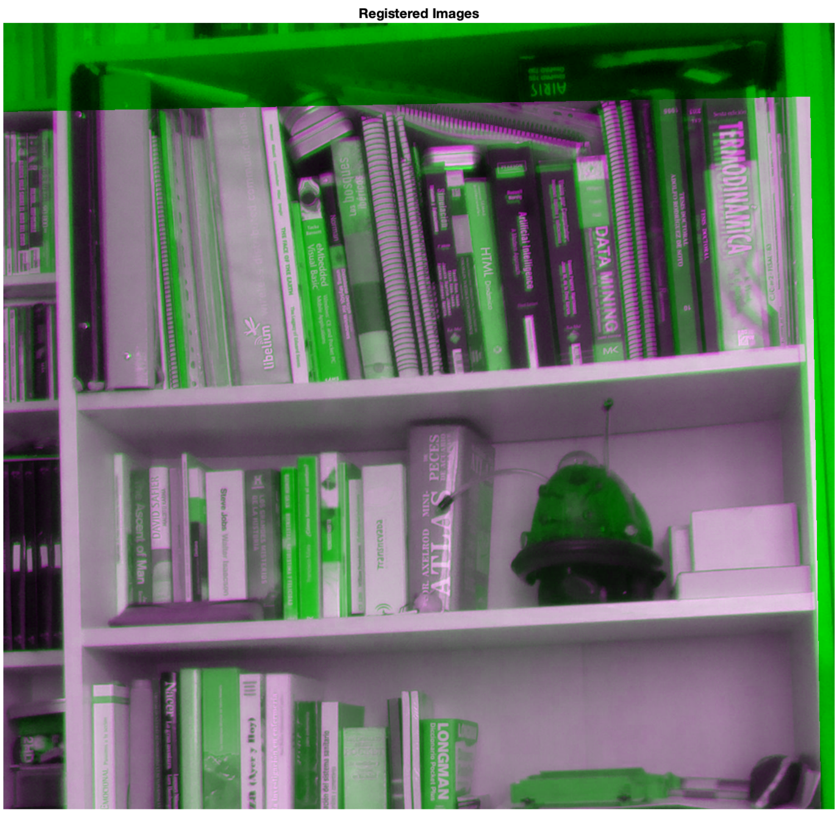

Figure 13: Registration result of red channels of images, using affine transformation

3.4.2 Discussion

The comparison between projective and affine transformations for the registration of multimodal images reveals important insights into the effectiveness of these transformations in handling different types of misalignments. Projective transformation is capable of handling complex perspective distortions. It maps points from one plane to another using a homography, which can accommodate the converging or diverging lines typically seen in images taken from different viewpoints or with different lenses. The ability to handle these distortions makes projective transformation particularly suitable for multimodal image registration where perspective differences are significant. While affine transformations can handle linear transformations like translation, rotation, scaling, and shearing, they fall short when it comes to correcting perspective distortions.

The metric used for evaluation, the average Euclidean distance error, quantitatively supports the visual observations. A lower metric value for the projective transformation confirms its effectiveness in achieving better alignment.

4 Analysis of Results

This section provides a detailed analysis of the experimental results obtained from various registration methods using different feature detectors and color channels. The analysis aims to identify the most effective approach for image registration under the given conditions.

4.1 Comparative Analysis

A comparative review of the experiments suggests that the choice of feature detection method and color channel significantly impacts the registration accuracy. The use of the red channel, coupled with feature detection methods like SIFT and SURF, provided the most substantial improvements in registration accuracy, as indicated by the lowest average Euclidean distance errors.

4.1.1 Influence of Color Channels

The experiments with different color channels revealed that the red channel consistently outperformed the green and blue channels in terms of both intensity and geometric accuracy. This finding underscores the importance of channel selection in image registration tasks, particularly in scenarios involving varied spectral characteristics. The red channel showed better spectral overlap with the near-infrared spectrum, capturing more robust and distinctive features crucial for effective registration. Also comparing with registration on greyscale images which have information from all the color channels, we see that again, registration based on red channel outperforms registration with greyscale images.

4.1.2 Effectiveness of Feature Detection Methods

Among the feature detectors tested, SIFT and SURF proved to be the most reliable for capturing detailed features necessary for precise registration. SIFT on grayscale images provided worse results compared to using the red channel, indicating that specific channels may enhance the effectiveness of feature detection. While SURF on grayscale images improved performance compared to SIFT, it still did not surpass the accuracy achieved with the red channel.

4.1.3 Role of Transformation Methods

The comparison between projective and affine transformations showed that projective transformations generally provided more satisfactory results. The projective transformation effectively corrected perspective distortions and achieved better alignment, as evidenced by a lower average Euclidean distance error. In contrast, the affine transformation, despite its theoretical flexibility to handle translation, rotation, scaling, and shearing, did not adequately address the perspective differences, resulting in noticeable misalignments and higher errors. This suggests that projective transformations are more suitable for multimodal image registration, especially when dealing with significant perspective distortions.

4.1.4 Conclusion and Final decision

The detailed analysis highlights the critical role of channel selection and the choice of feature detection and transformation methods in multimodal image registration. The red channel emerged as the most effective for capturing robust features, while SIFT and SURF provided reliable feature detection. Projective transformations were superior to affine transformations in handling perspective distortions and achieving accurate alignment. These findings emphasize the importance of a strategic approach to selecting image channels and registration methods to optimize accuracy in multimodal contexts. Future work could explore adaptive techniques that combine information across multiple channels or leverage advanced feature extraction methods to further enhance registration performance.

References

[1] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," in International Journal of Computer Vision, 60, 91-110, 2004.

[2] Bay, H., Tuytelaars, T., Van Gool, L., "SURF: Speeded up robust features," in Computer Vision and Image Understanding, 110(3), 346-359, 2008.

[3] Alcantarilla, P. F., Bartoli, A., Davison, A. J., "KAZE Features," in European Conference on Computer Vision (ECCV), 2012.

[4] Rublee, E., Rabaud, V., Konolige, K., Bradski, G., "ORB: An efficient alternative to SIFT or SURF," in IEEE International Conference on Computer Vision (ICCV), 2011.

[5] Goshtasby, A. A., *Image Registration: Principles, Tools and Methods*, Springer-Verlag, 2012.

[6] MathWorks, "Image Registration"

[7] Thilina Sameera, "Translation Invariant Image Registration using Phase Correlation"

[8] MathWorks, "Find Image Rotation and Scale Using Automated Feature Matching"