Assignment 1 Report

Yusuf Demir
2210356074

1 Overview and Objectives

This assignment focuses on developing a document dewarping system using computer vision techniques. The main goal is to detect and correct geometric distortions in scanned documents to obtain a properly aligned, frontal view. To achieve this, a **custom Hough Transform** was implemented for edge and line detection, and a **RANSAC-based refinement** was applied to filter out noise and outliers. After obtaining the document's boundary, a **geometric transformation (homography)** was used to perform perspective correction.

2 Implementation Steps and Challenges

2.1 Data Preparation and Resizing

The dataset consists of six categories: *curved, fold, incomplete, perspective, random, rotate*, each containing distorted and ground-truth images. Many of these images had resolutions exceeding 3000 pixels on the longest side, making direct processing inefficient. To address this, images were resized while maintaining their aspect ratio, limiting the maximum dimension to **1500 pixels**.

Challenges:

- High-resolution images significantly increased computation time.
- Resizing had to be performed carefully to avoid losing important document details.

Solution: A **conditional resizing function** was implemented to ensure that image quality was preserved while optimizing performance.

2.2 Edge Detection, Hough Transform, and ROI Masking

After resizing, images were converted to grayscale and smoothed using a **Gaussian filter** before applying **Canny edge detection**. However, initial edge detection results were inconsistent due to varying lighting conditions, background clutter, and uneven document boundaries.

To improve detection:

- **Adaptive thresholding** was applied to enhance document edges.
- Morphological operations were used to remove small noise artifacts and fill gaps in broken edges.
- A **document mask (ROI)** was created by extracting the largest contour from the thresholded image. This helped isolate the document area from the background.

For line detection, OpenCV's **Hough Transform** was initially used and provided the most reliable results. However, when we implemented our own Hough Transform, we encountered difficulties in detecting structured document edges with the same accuracy. The custom Hough Transform struggled particularly with non-paper lines and background elements, often misidentifying or missing document boundaries.

Challenges:

- Our implementation of the Hough Transform was less effective than OpenCV's built-in function, leading to missing or incorrect line detections.
- We still have trouble filtering out unwanted lines outside the document region, as some edges in the background are mistakenly detected as part of the document.
- Selecting appropriate threshold values for line detection remains a difficult task, as different types of distortions (curved, folded, or rotated documents) require different tuning.

Unresolved Issues:

- Despite using an ROI mask, some unwanted background lines are still detected.
- Our method sometimes fails to detect full document edges, leading to incomplete quadrilateral detection.
- Further refinement is needed to handle highly distorted documents, particularly those with severe curvature or folds.

2.3 RANSAC Refinement

Since the detected lines still contained noise and did not always correspond to the document structure, we applied **RANSAC** to refine the segments and attempt to extract only the document's boundaries. The goal was to remove irrelevant lines and focus on the four edges of the paper.

Challenges:

- Despite refinement, some non-document lines are still mistakenly selected.
- The algorithm does not always detect all four document edges, leading to incomplete boundary extraction.
- Choosing appropriate thresholds for inlier selection remains difficult, as different distortions (e.g., curvature, folds) affect line detection.

Current Approach: A **RANSAC-based line fitting method** was used:

1. Randomly selected two points from detected edge pixels.
2. Computed a candidate line and counted how many points fit within a given threshold.
3. Repeated this process for multiple iterations, keeping the most reliable model.

Limitations and Future Work:

- The method still struggles to completely filter out background edges, leading to incorrect quadrilateral detections.
- Further improvements are needed to ensure that exactly four dominant lines corresponding to the document's boundaries are detected.

2.4 Quadrilateral Extraction and Perspective Correction

Once the refined lines were obtained, they were grouped based on orientation:

- **Horizontal lines** were selected as top and bottom edges.
- **Vertical lines** were used for left and right boundaries.

From each group, the most extreme lines were selected to form a quadrilateral representing the document's boundary.

Challenges:

- The **Hough Transform** often detected extra non-document lines, leading to incorrect quadrilateral formation.
- In many cases, **RANSAC failed to eliminate background edges**, resulting in inaccurate document boundaries.
- The algorithm struggled to correctly detect all four edges, especially for documents with folds or perspective distortions.
- The approach to selecting the **largest quadrilateral** sometimes led to selecting a region outside the document when larger non-document boundaries were detected. This caused the warped output to include parts of the background instead of the document. As seen in Figure 1, the selected quadrilateral encompasses areas beyond the document boundary, resulting in an inaccurate perspective correction.

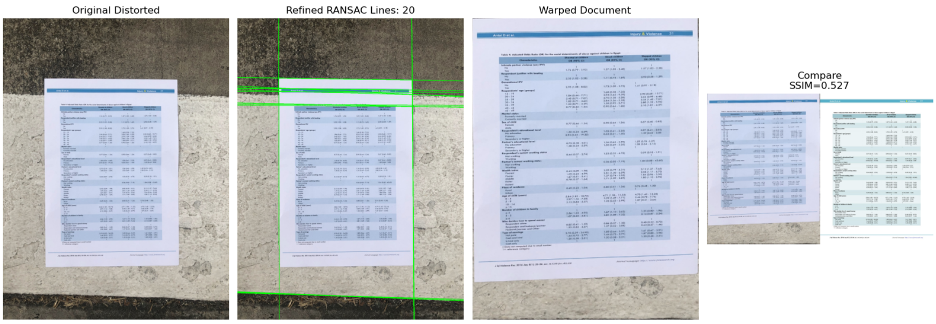


Figure 1: Example of incorrect quadrilateral selection where the largest detected quadrilateral exceeds the document boundary.

Current Limitations:

- Some images were left with incomplete or misaligned quadrilateral boundaries.
- When corner detection failed, the warped output was distorted and unusable.
- Large quadrilateral selection sometimes resulted in boundary overflow, especially when background edges were mistakenly considered as part of the document.

Attempts to Improve:

- Several **angle and distance thresholds** were tested to refine detected edges.
- A more **strict filtering mechanism** was introduced to discard false edges.
- If an edge was missing, an **approximation technique** was used to infer its location.
- Additional checks were implemented to reduce the selection of non-document lines by prioritizing regions with higher text density and consistent brightness.

Despite these refinements, our approach still produced inconsistent results in challenging cases, requiring further optimization.

2.5 SSIM Evaluation and Future Improvements

The quality of the dewarped document was assessed using the **Structural Similarity Index (SSIM)**, comparing the warped output with the ground-truth image.

Challenges:

- **SSIM scores were highly inconsistent**, as even minor errors in corner detection caused large variations in similarity.

Inconsistent SSIM Results: Despite achieving seemingly reasonable results in some cases, the SSIM metric often failed to accurately reflect the quality of the dewarping process. Figure 2 and Figure 3 demonstrate two cases where the SSIM score is inconsistent.

- **Good Example:** In Figure 2, the document is correctly dewarped, with straight text lines and clear readability.
- **Bad Example:** In Figure 3, the document is poorly warped, with text distorted and misaligned, but the SSIM score is bigger.

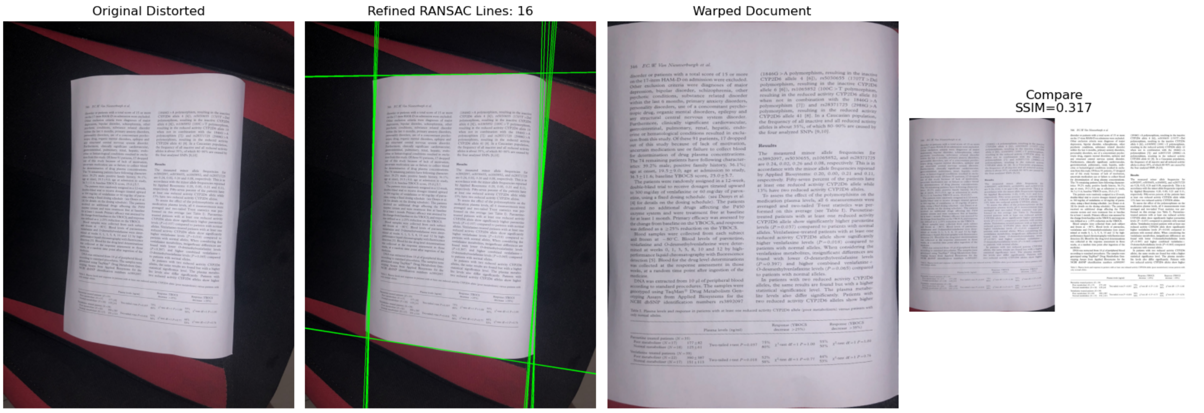


Figure 2: Good Example: Successfully dewarped document with SSIM = 0.317

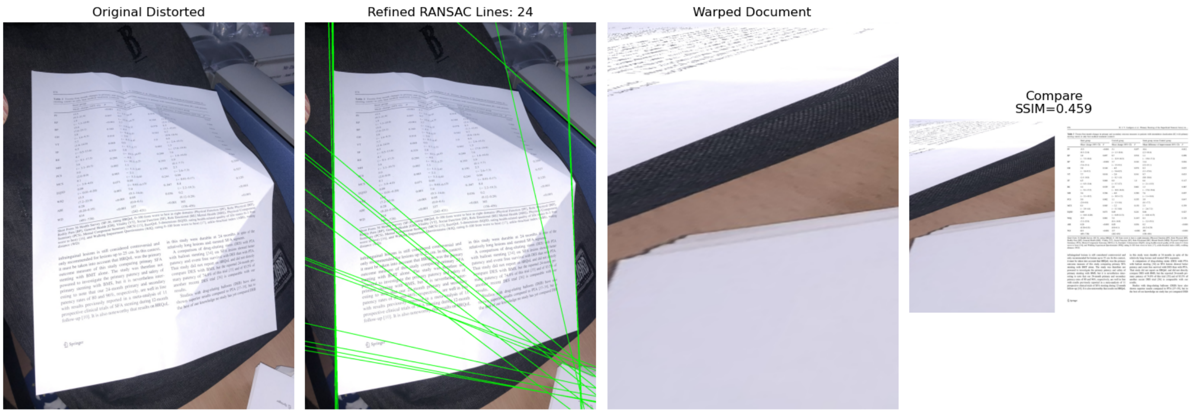


Figure 3: Bad Example: Poorly dewarped document with SSIM = 0.459

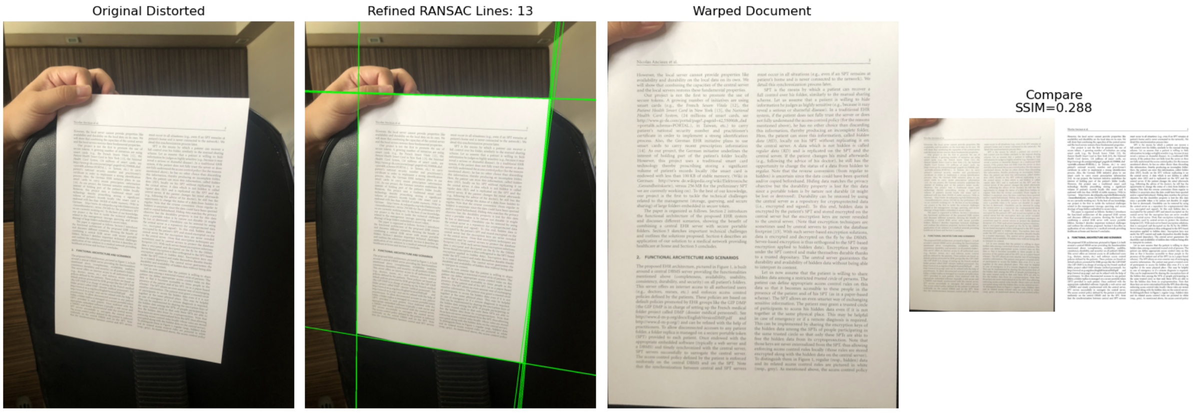
While the current approach provides a baseline for document dewarping, significant improvements are needed to achieve reliable results across different distortions.

3 Results and Discussion

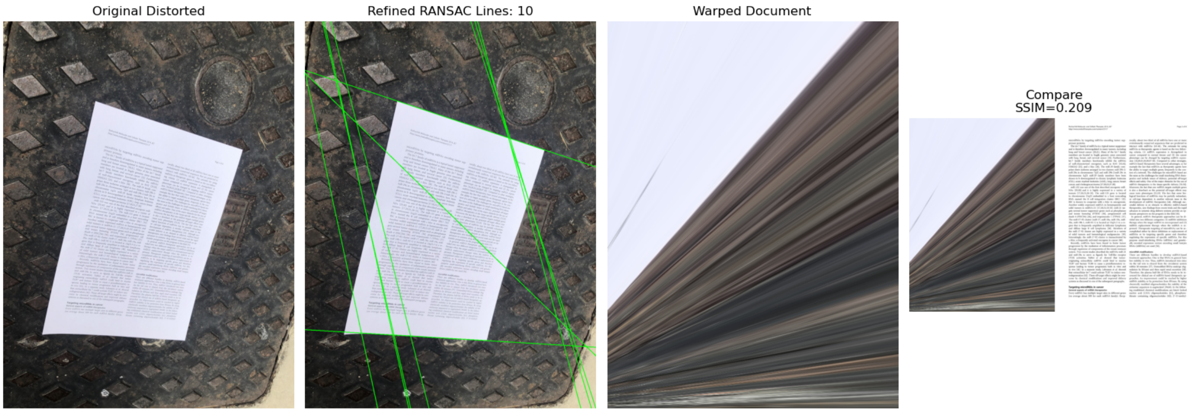
To evaluate the effectiveness of our document dewarping approach, we processed images from six different categories. Below, we present both successful and unsuccessful cases for each category, highlighting the strengths and limitations of our current method.

Additionally, we compared the **SSIM values** across all six categories to assess the performance of our approach. However, as discussed earlier, these SSIM scores are not entirely reliable indicators of the quality of the dewarping process. Due to inconsistencies in how SSIM measures structural similarity, well-dewarped documents and poorly dewarped ones may sometimes yield similar scores. Therefore, while the SSIM comparison provides a general sense of performance across categories, it should not be considered a definitive metric for evaluating our method.

3.1 Perspective Category



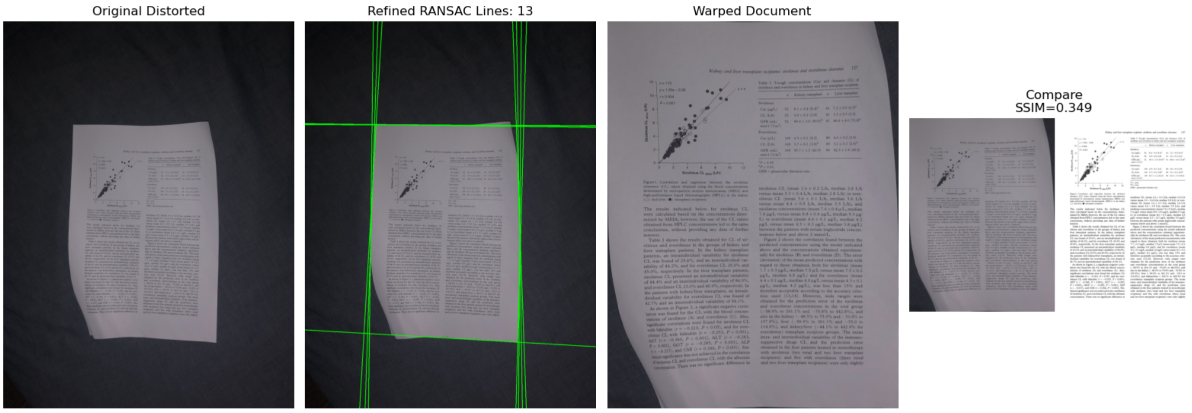
(a) Successful case: Perspective distortion corrected well.



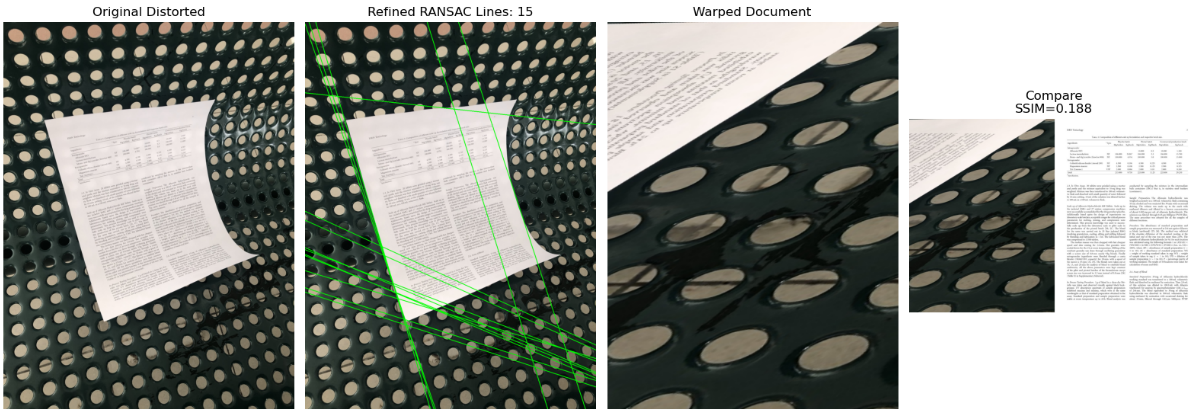
(b) Failure case: Incorrect boundary detection led to a distorted warp.

Figure 4: Comparison of successful and failed cases for the perspective category.

3.2 Curved Category



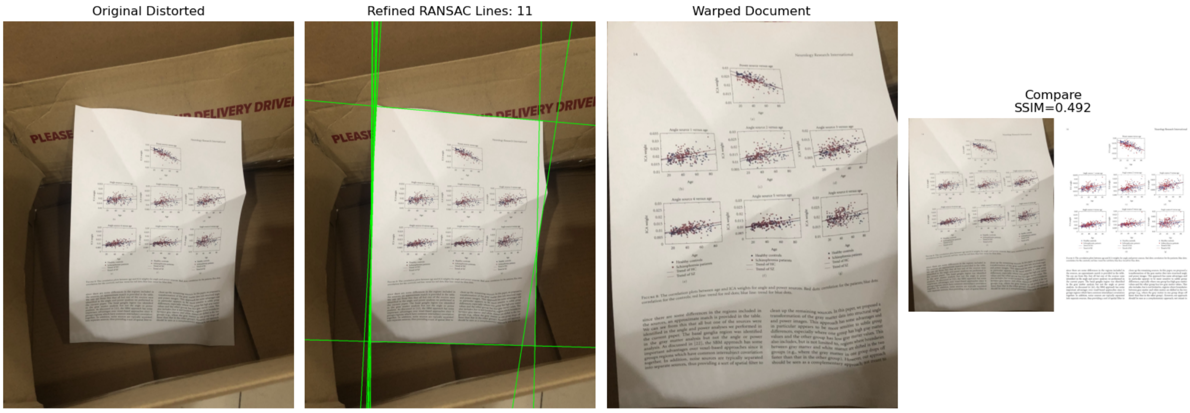
(a) Successful case: Minor curvature corrected.



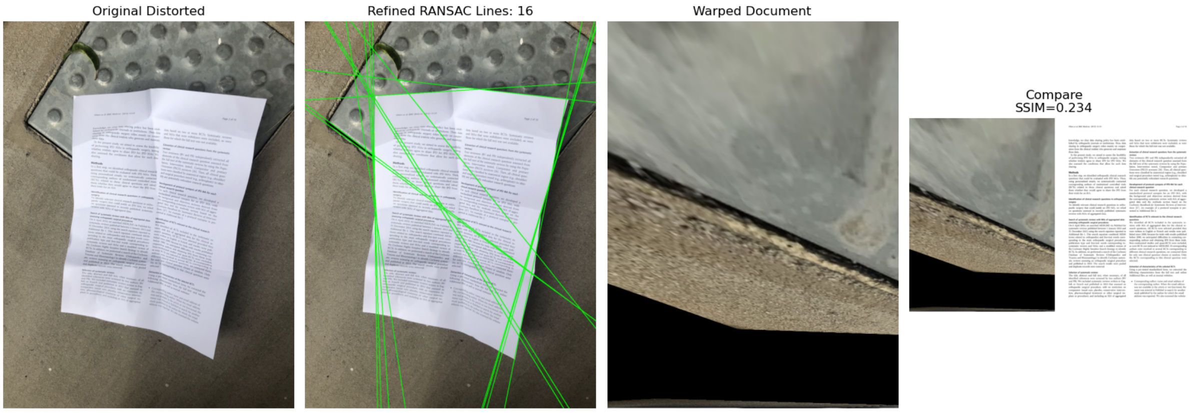
(b) Failure case: Warping failed due to extreme curvature.

Figure 5: Comparison of successful and failed cases for the curved category.

3.3 Fold Category



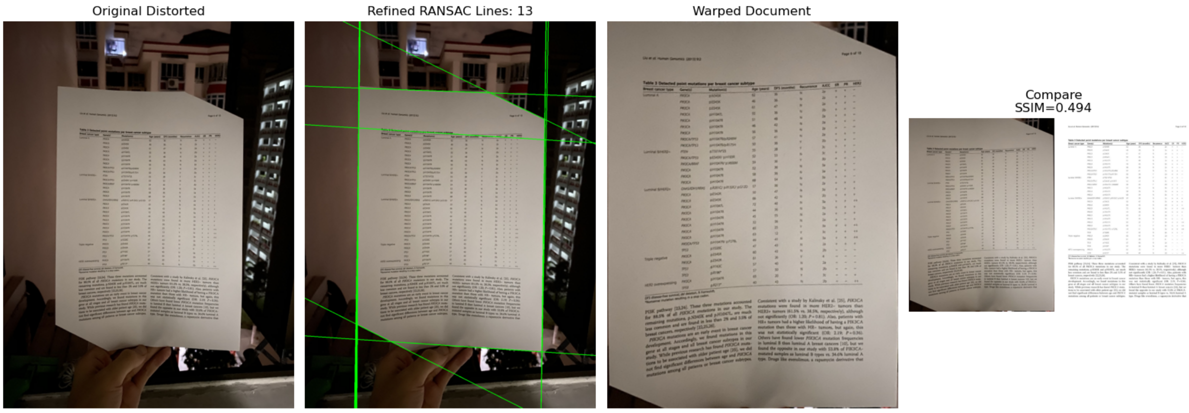
(a) Successful case: Good boundary detection and correction.



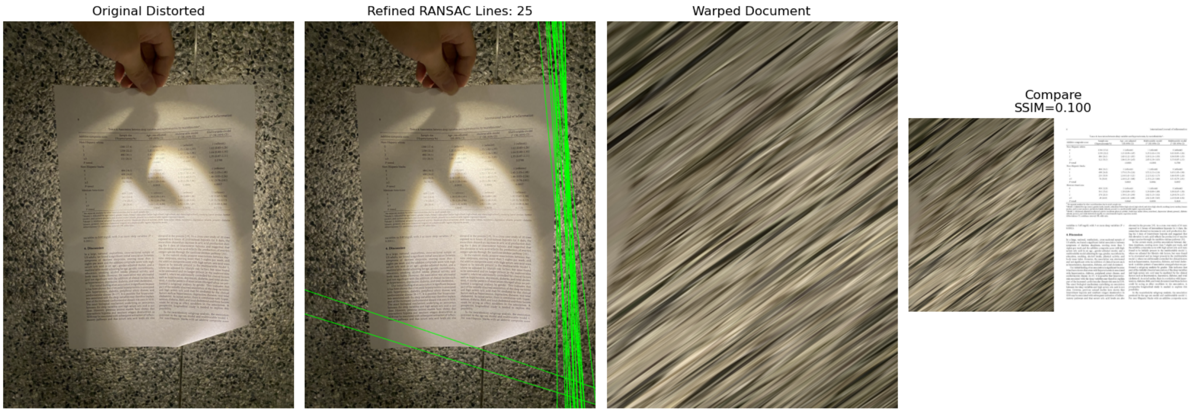
(b) Failure case: Background lines interfered with document edges.

Figure 6: Comparison of successful and failed cases for the fold category.

3.4 Incomplete Category



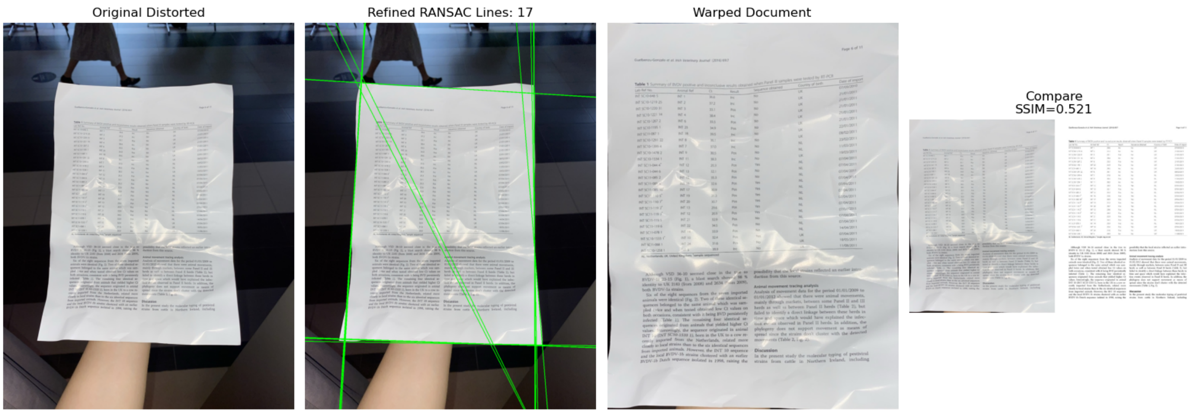
(a) Successful case: Minor curvature corrected.



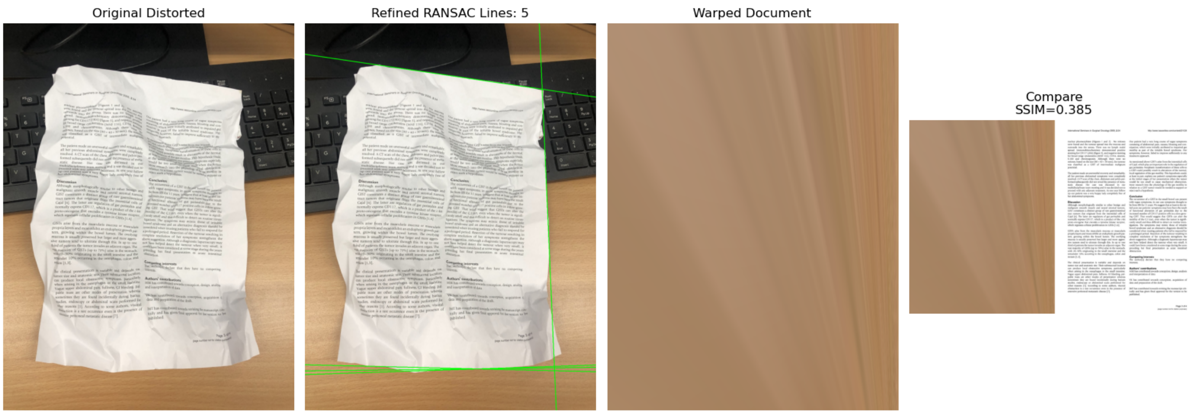
(b) Failure case: Warping failed due to extreme curvature.

Figure 7: Comparison of successful and failed cases for the curved category.

3.5 Random Category



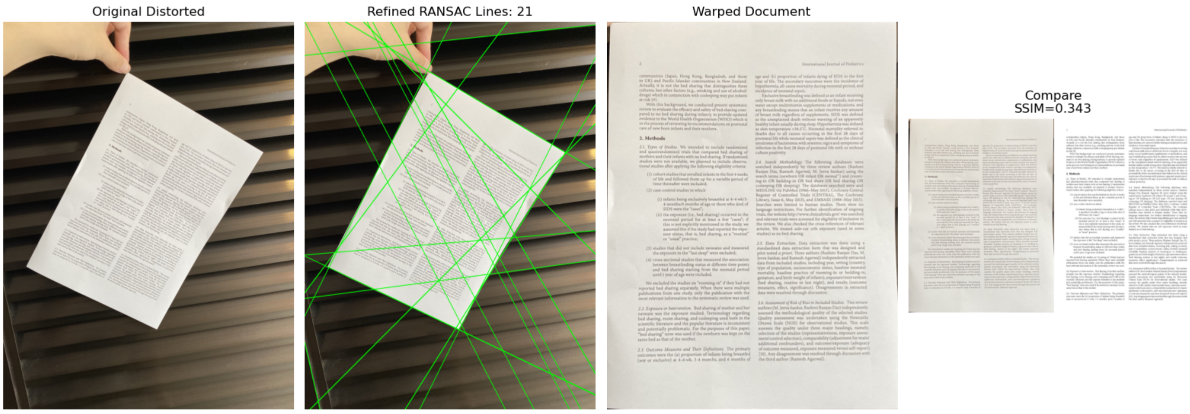
(a) Successful case: Minor curvature corrected.



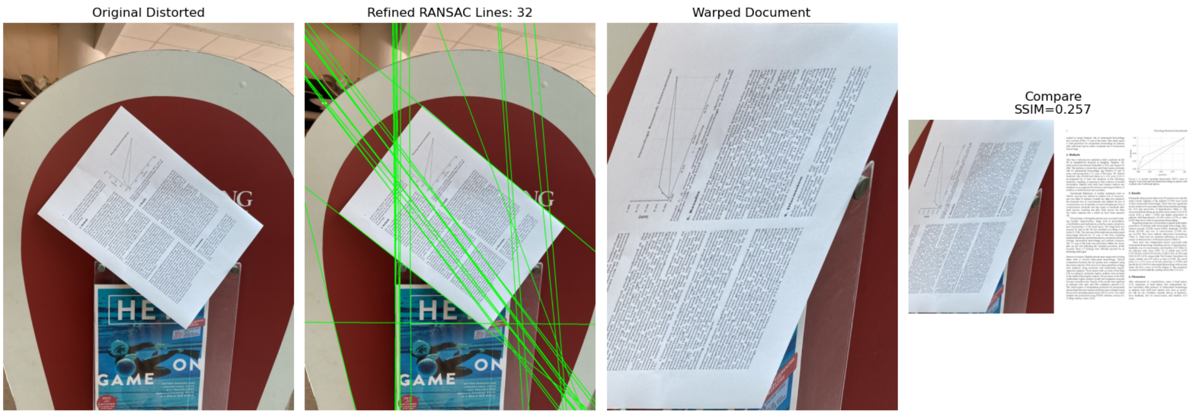
(b) Failure case: Warping failed due to extreme curvature.

Figure 8: Comparison of successful and failed cases for the curved category.

3.6 Rotate Category



(a) Successful case: Minor curvature corrected.



(b) Failure case: Warping failed due to extreme curvature.

Figure 9: Comparison of successful and failed cases for the curved category.

3.7 Analysis of SSIM Scores

Although some of our results visually appear satisfactory, the **SSIM scores are lower than expected**. This discrepancy arises due to the following reasons:

- **Misalignment in warping:** Even slight deviations in corner detection can cause noticeable errors in SSIM measurements.
- **Edge misclassification:** Non-document edges (e.g., background structures) interfere with Hough and RANSAC processing, leading to inaccurate warps.
- **Curved distortions:** Our method is not optimized for handling complex 3D deformations, significantly reducing SSIM performance.
- **Inconsistent SSIM Scores:** SSIM can sometimes provide unreliable results, as shown in Figure 3. In certain cases, well-dewarped images and poorly dewarped images may yield similar SSIM values, which indicates that SSIM alone is not a reliable indicator of warping quality. The SSIM score is highly sensitive to brightness, contrast, and minor distortions that may not visually affect the quality of the output document.

Future improvements should focus on refining edge selection, integrating **curvature correction models**, and developing more robust similarity metrics to enhance document alignment.

4 Conclusion

Our document dewarping approach demonstrates **promising results in structured cases**, particularly when documents are primarily affected by perspective distortions. However, handling **severe curvatures, folds, and noisy backgrounds remains a major challenge**.

Category	Average SSIM Score
Perspective	0.312
Curved	0.303
Fold	0.336
Incomplete	0.337
Random	0.322
Rotate	0.342

Table 1: Comparison of Average SSIM Scores Across Different Categories

Key Takeaways:

- Perspective distortions can be **partially corrected** using Hough Transform and RANSAC, but boundary detection is still unreliable.
- The **SSIM metric is sensitive** to misalignments, requiring more robust feature matching for fair evaluation.
- Background clutter significantly impacts line detection, leading to **false quadrilateral extractions**.

Future Work:

- **Improve line filtering techniques** to ensure only document boundaries are selected.
- **Enhance warping algorithms** to better handle curved and folded documents.