Image Stitching

Shruti Murarka
ECE Department
VNIT
Nagpur, India
shrutim@students.vnit.ac.in

Rajesh Nagula
ECE Department
VNIT
Nagpur, India
rajeshn0811@students.vnit.ac.in

Dhruvi Shah
ECE Department
VNIT
Nagpur, India
dhruvi29shah@students.vnit.ac.in

***Abstract*—This paper presents a detailed explanation of the Harris Corner Detection as a method for detecting the key points and implementing the Scale Invariant Feature Transform algorithm. A comparison between the above two methods is discussed.**

***Index Terms*—SIFT, Harris Corner, RANSAC**

I. Introduction

Panorama is a well-known image-based rendering method. When an input set of images is stitched together, the result is a composite image representing a collection of scenes. To stitch the images together, there must be an overlap between them. By properly aligning this overlapping area between them, the images can be stitched together.

Recent research has taken into account various features such as Harris, SIFT, ORB, SURF, and so on [2], [3]. In this paper, we found certain types of features using a different keypoint extraction technique and compared the results. After computing the Harris corner features and SIFT, the correspondence points matching will be identified. Image stitching is accomplished by comparing these types of features and checking for precise point matching.

One of the most critical processes in implementing virtual environments, robot navigation, and object recognition is the recreation of 3D models. These methods are helpful for 3D object reconstruction, especially for building reconstruction in an autonomous mobile robot's outdoor scene.

II. Algorithm

This paper aims to generate an output image obtained by the stitching of the multiple input images. Following is the basic algorithm followed:

- Identify the points of the images which can identify the images uniquely. Such points are known as the " key points" of the image.
- The next step is to find the descriptors of these keypoints; when we use SIFT for feature extraction, it gives a 128-dimensional feature vectors output for each of the keypoint..
- The corresponding matches for the given keypoints in the images to be stitched are to be mapped.
- Next we identify the outliers and inliers of the image and eliminate the outliers.
- Homography matrix is found for two images and it is warp accordingly.
- Lastly, image is resized and excess black portion is removed.

A. Feature Detection

Image mosaicing relies on determining the spatial relationships between two images. Features are the useful keypoints that can be well localized and well aligned in both the images despite of geometric and photometric transformation. Corners are usually the keypoints we require since they change significantly in all directions.

1) Harris Corner: By making use of the Harris corner, we can get the location/coordinates of all the corners in the given image. The Harris corner detector algorithm is based on a fundamental principle: at a corner, the image amplitude changes significantly in several directions.

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u,y+v) - I(x,y)]^2 \quad (1)$$

Here, w(x,y) denotes weighted/window function which is usually Gaussian or rectangular function.

For corner detection, we must maximize E(u,v). Using Taylor Expansion and some mathematical steps to solve (1), we get,

$$\begin{gathered} E(u,v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} \\ M = \sum_{x,y} w(x,y) \begin{bmatrix} (I_x)^2 & I_x I_y \\ I_x I_y & (I_y)^2 \end{bmatrix} \end{gathered} \quad (2)$$

Here, I_x,I_y represents image derivatives in horizontal and vertical directions respectively.

They then created a score, which is simply an equation, that determines whether or not a window can have a corner.

$$R = det(M) - k(trace(M)) \quad (3)$$

Here, k = constant
det(M) =$\lambda_1 \lambda_2$
trace(M) =$\lambda_1 + \lambda_2$
All windows with a score R greater than a certain amount are corners.

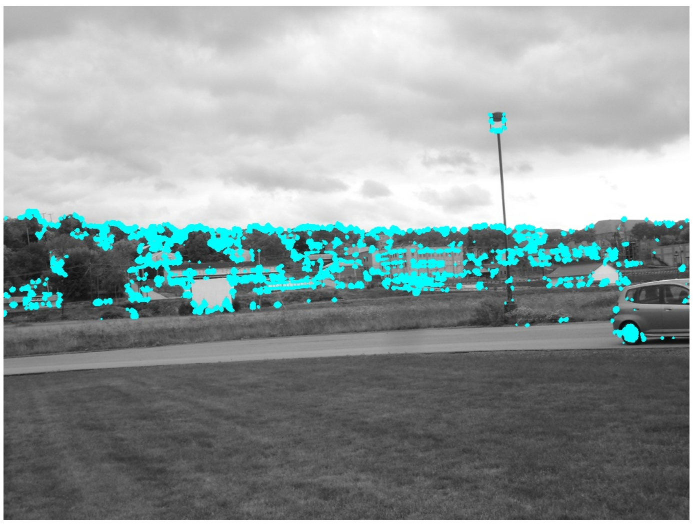


Fig. 1: Self Implemented Harris Corner Output for image1

Fig. 2: Self Implemented Harris Corner Output for image2

2) Scale-Invariant Feature Transform(SIFT): SIFT uses DoG (Difference of Gaussian) to extract features. We build the image pyramid using a continuous smooth and a Gaussian mask. The image's DoG pyramid can be obtained by subtracting neighboring smoothed images. We will find the maximum or minimum value of them by comparing pixel value of the current scale with the upper and lower scales in the area 3x3, i.e. 9 pixels. These points are also candidates for keypoints. To obtain image rotation invariance, each keypoint now has an orientation assigned to it.

The Gaussian function is given below:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \quad (4)$$

The Difference of Gaussian is given by:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (5)$$

Here, * denotes convolution.
Once possible keypoint sites are identified, they must be optimized to provide more reliable outcomes. Using Taylor series approximation and threshold, more accurate location of keypoint is obtained. Edges have a higher reaction in DoG, so they must be eliminated. These are removed using Harris corner detection. It generates keypoints of the same scale and position but different directions. It contributes to the stability of matching.

Fig. 3: SIFT keypoints for image1

B. Keypoints Descriptors

Descriptors are description of neighbourhood of keypoints. A 16x16 neighborhood is drawn around the keypoint. It's decomposed into 16 4x4 sub-blocks. An

Fig. 4: SIFT keypoints for image2

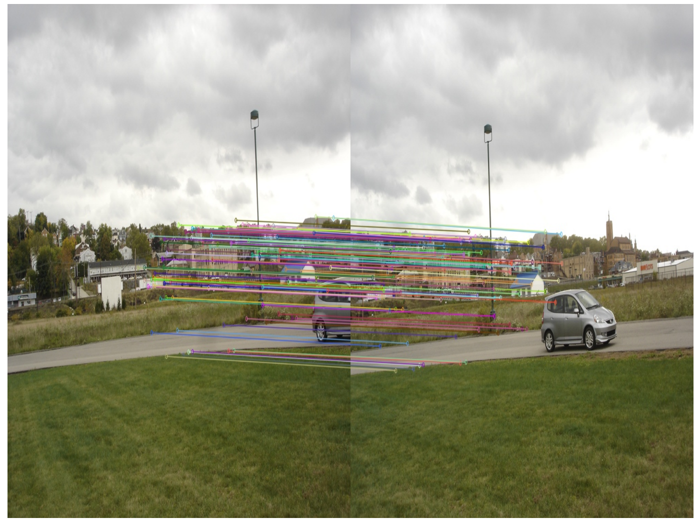


Fig. 6: Keypoints matching of SIFT

8-bin orientation histogram is generated for each sub-block. As a result, there are a total of 128 bin values available. To make a keypoint descriptor, it's defined as a vector.

C. Keypoint-Matching

We identified points of interest and extracted a vector function descriptor around each one. We would now determine the descriptor's correspondence. The Euclidean distance between two feature vectors is used by BFMatcher as the key point similarity criterion and matches them using the nearest neighbor algorithm. For a given threshold, if the distance ratio between the closest neighbor and the second-closest neighbor is less than the threshold, we have a precise match.

Fig. 5: Keypoints matching of harris

D. Homography & Removing outliers

Homography is a mapping between two planar projections of an image. When there are four or more corresponding points in two images, we can compute a homography. It is 3x3 matrix with 8 DoF. It is generally normalized with h_{33} = 1

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{6}$$

The algorithm then attempts to construct a homography matrix between the two images using four of the corresponding keypoints after the matches have been determined. Based on the error between the projection of a point in image 1 and the corresponding point in image 2, this homography is then used to decide which keypoint matches are correct and which are not. These correct matches are given to RASAC.

The data includes inliers or data whose distribution can be described by a collection of model parameters, as well as outliers or data that does not match the model. RANdom SAmple Consensus (RANSAC) is a method for estimating the statistical model parameters of data from a collection of reference data sets with inappropriate data to obtain accurate sample data. RANSAC uses the voting scheme to find the most fitting answer for a dataset with both inliers and outliers. The dataset's data elements are used to vote on one or more versions. This voting scheme is based on two assumptions: that the noisy features would not vote uniformly for any single model and that there are sufficient features to agree on a successful model.

We are able to take the homography resulting in the most inliers as our best approximation of the true rotation between images after randomly sampling 4 points n times, measuring the homography of each, and estimating the number of inliers each time.

E. Warping

One image is warped to fit in the same frame as the other using homography, and a new image of all black pixels is formed that will fit both images in the new frame. After the images have been correctly aligned, they can be merged to create the stitched image.

F. Resizing Image

Now, the obtained stitched figure is resized. The stitched image is blurred and binary threshold is applied, so that it results binary image. Contours for This thresholded image are found. Boundary coordinates are extracted and imaged is cropped accordingly.

After combining the images, the procedure was replicated with the new image and one of the other images in the dataset. By repeating this process, each image is added to the panorama one by one till all of the images have been merged.

III. Output

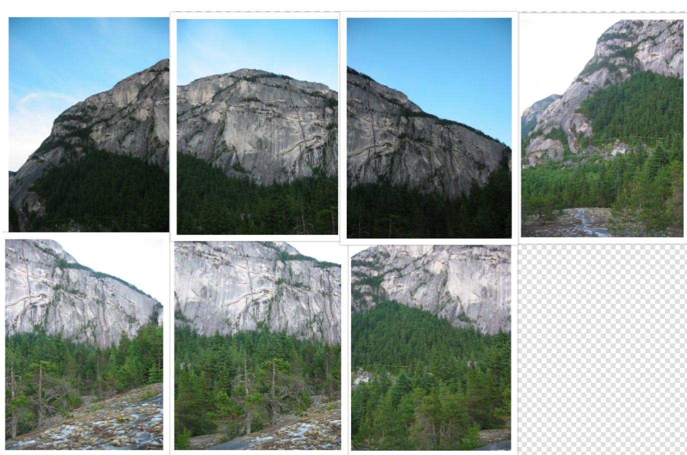


Fig. 7: Sample Dataset for MOUNTAIN Dataset

IV. Conclusion

This paper applied the Harris and SIFTS features of the correspondence matching problem for image stitching. Harris features fail for scaled images , since for scaled images corners appears to be the edges or vice versa depending on scaling factor. Meanwhile SIFT uses DoG which considers scale along with rotation. Based on the outputs, we may conclude that SIFT features are

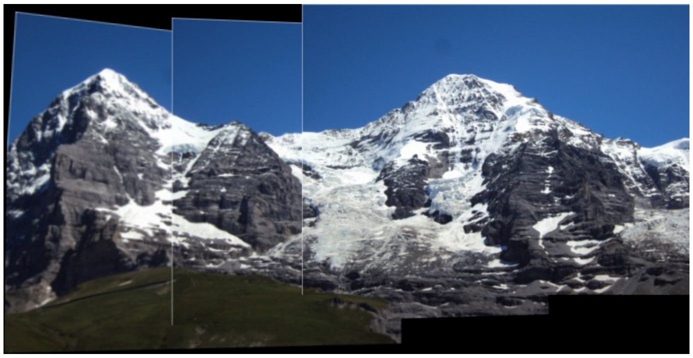
tbp

(a) Self Implemented Harris corner detection results

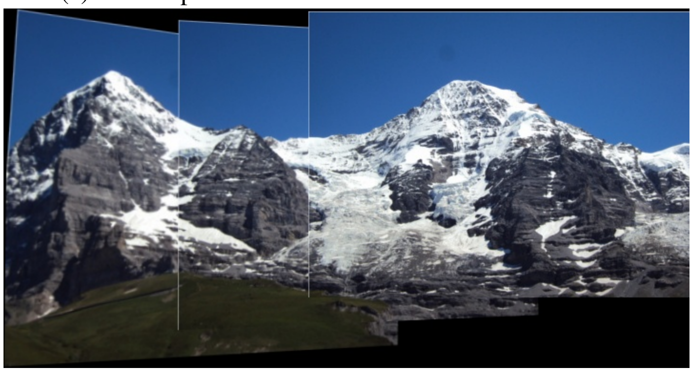
(b) SIFT results

(c) cv2.stitcher results

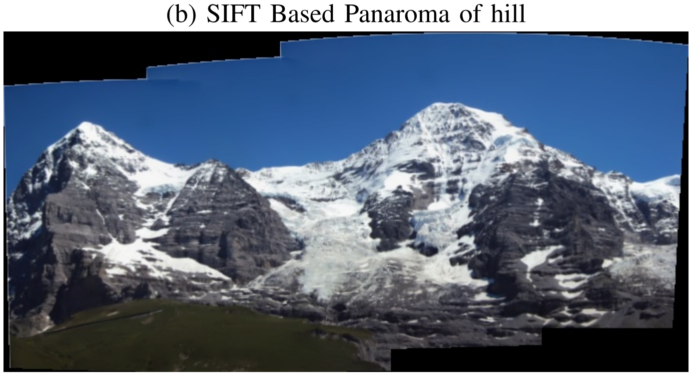
Fig. 8: Output for CAR Dataset



(a) Self Implemented Harris Based Panaroma of hill

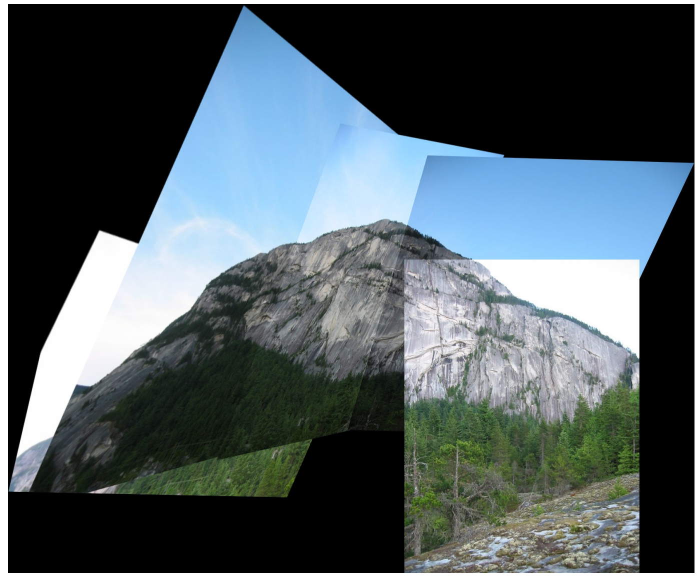


(b) SIFT Based Panaroma of hill

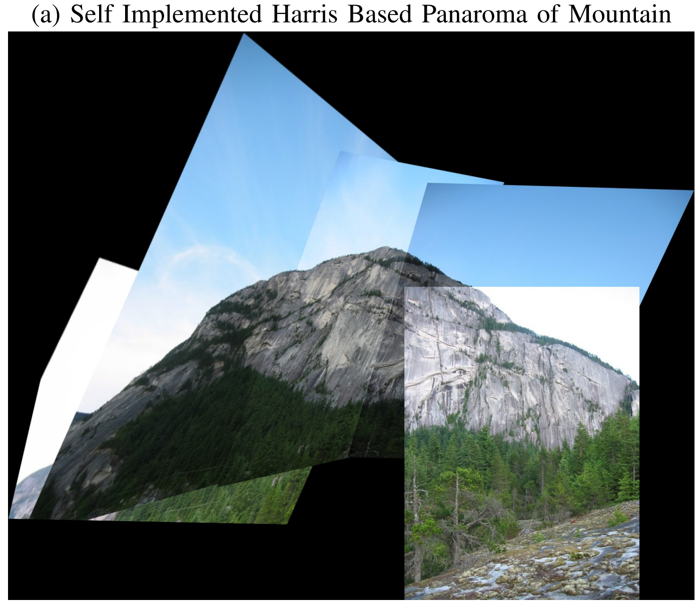


(c) cv2.stitcher results

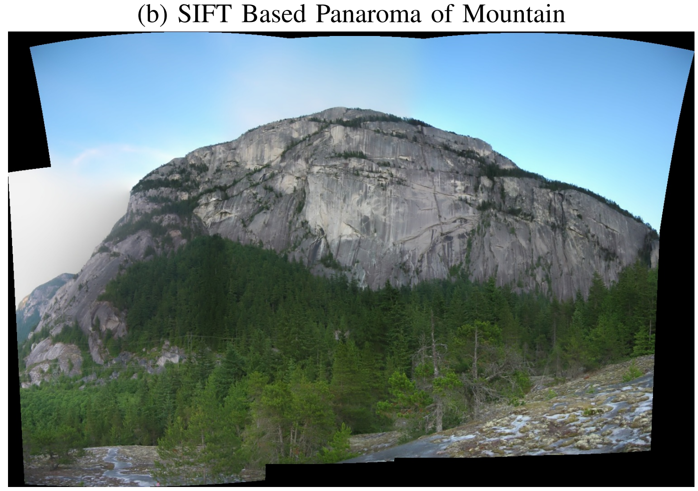
Fig. 9: Output for HILL Dataset



(a) Self Implemented Harris Based Panaroma of Mountain



(b) SIFT Based Panaroma of Mountain



(c) cv2.stitcher results

Fig. 10: Output for MOUNTAIN Dataset

both stable and appropriate for this function. The output result for Mountain Dataset are bad since it includes right left & up stitch. Our algorithm is limited to right & left stitching.

Image stitching is a very useful tool when photos with a very wide angle or higher resolution are required. It is extremely useful in space explorations. It is used for mapping the bodies of the space bodies like the Earth, Mars and the Moon from the satellite images. It is also a very common feature in most digital cameras and smartphones and is also becoming an important technology for the virtual and augmented reality.

ACKNOWLEDGMENT

This project is guided by Ms. Snigdha Bhagat, Assistant Professor of Electronics & Communication Engineering Department, Visvesvaraya National Institute of Technology, Nagpur, India.

REFERENCES

[1] My-Ha Le, Byung-Seok Woo and Kang-Hyun Jo ”A Comparison of SIFT and Harris Conner Features for Correspondence Points Matching” unpublished.

[2] Luo Juan and Oubong Gwun, “A Comparison of SIFT, PCA-SIFT and SURF” International Journal of Image Processing, Volume 3, Issue 5, 2010

[3] Herbert Bay, Tinne Tuytelaars, Luc Van Gool, “SUFT: speeched up robust features”, ECCV, Vol. 3951, pp. 404-417, 2006

[4] Harrison Chau & Robert Karol ”Robust Panoramic Image Stitching” unpublished.

[5] Min, Z., Jiguo, Z., & Xusheng, X. (2012). Panorama Stitching Based on SIFT Algorithm and Levenberg-Marquardt Optimization. Physics Procedia, 33, 811–818.