Feature Matching Approach for 3D Reconstruction from Multiple Images

KhantKyawtKyawtTheint, MyintMyintSein
University of Computer Studies Yangon, Myanmar
khantkyawtkyawttheint@ucsy.edu.mm, myint.ucsy@gmail.com

Abstract

This paper handles a robust method to handle matching problem of the 3D building from remote sensed images, where two or more images are captured from the different views without any camera information. The system is based on the SIFT interest point detector on foreground object. The system provides a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different view of an object or scene. The features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. The system also describes the outlier removal and the homography matrix computation approach between two planes using corresponding points based on the RANSAC method. The proposed method gives a sufficient number of matches distributed on the image and is particularly suitable for 3D reconstruction from feature correspondence.

1. Introduction

Feature matching is a fundamental point of many problems in computer vision, including object or scene recognition, solving for 3D structure from multiple images, stereo correspondence, and motion tracking. One major challenge of the reconstruction problem is to find feature correspondences, that is, to locate the projections of the same three-dimensional geometrical or textural feature on two or more images. If the fundamental matrix is known, reliable and fast feature correspondence can be obtained in general situations. However, in order for the fundamental matrix to be computed one Procedure for Paper Submission needs a good initial set of feature correspondences. [1].

Feature-based image matching is one of the most fundamental issues in computer vision tasks. Feature based methods can be divided into four basic steps: (1) Feature Extraction, (2) Feature matching, (3) Outlier rejection, and (4) Homography fitting. A more robust matching of SIFT features for remote sensing images is proposed in [5] and SIFT descriptors was the most resistant to common image deformation and it is not only scale invariant, but also invariant to rotation, illumination and viewpoint changes [3]. According to their work, the invariant features extracted from images can be used to perform reliable matching between different views of an object or scene. The approach is efficient on feature extraction and has the ability to identify large numbers of features. For the three dimension reconstruction, there is needed to be the reliable feature correspondences are important. The scale-invariant features are efficiently identified by using a staged filtering approach. The proposed system aims to reconstruct 3D points of the scene from two view images. This paper provides the foreground extraction, feature detection and matching for 3D scene reconstruction from multiple images of unknown camera parameters. The method of foregroundextraction form the complex environment is based on RGB color space. Since the scaleinvariant feature transform (SIFT) feature points can be detected stably and relatively accurately, feature detection and matching techniques is based on SIFT and the matching points are refined by RANCS method for fundamental matrix computation in 3D reconstruction.

This paper is organized as follows: Section 2, we study the related work of correspondence feature matching approaches. In Section 3, step by step of the algorithm of feature matching for 3D reconstruction by multiple images is described. The matching algorithm based on the scale invariant feature transform (SIFT) providing the accurate matched points. Experiments and results of the system are described in Section 4. Finally, Section 6 provides some conclusions of the system and future works of research.

2. Related Work

The most widely used detector is the Harris corner detector, which is sensitive to the affine distortion of image. Therefore, they are not suitable to build feature sets in image that acquired by camera under various environments. As the number of features increases, the matching process rapidly becomes a bottleneck [2]. Multiview stereo algorithms have also

been extended to internet scale image collections with promising results [6]. However, these algorithms still lack detailed reconstruction in textureless regions. J.Ma et al. proposed a new vector field interpolation algorithm called vector field consensus (VFC) for establishing robust point correspondences between two sets of points [7]. Thismethod provides the feature correspondences based on the coherence of the underlying motion fields rather than geometric constraints and shows more effective than RANSAC, but less effective that VFC if there are many outliers.

3. Feature Matchingfor 3DReconsturction

3.1 Pre-processing Steps

The preprocessing steps contain resizing, color segmentation and grayscale converting have been done for feature extraction from acquired images. The system use two images grabbed from the Google Earth. For the background elimination, the detection of foreground information from images is one of the most basic in computer vision and 3D object reconstruction. The foreground information is considered as the interested object of the test area. The algorithm extract foreground object features in RGB color space. The input satellite images and foreground images are described in figure 1 (a) and figure 1 (b), respectively.

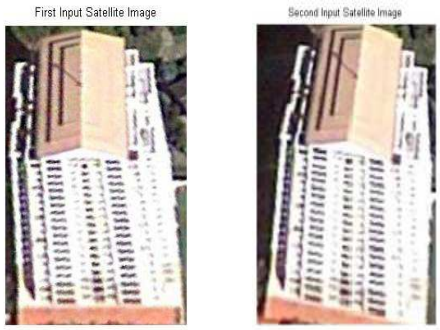

Figure1. (a) Input Satellite Images

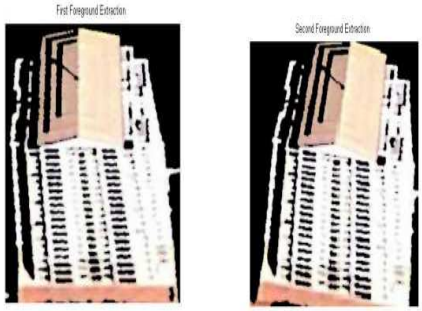

Figure1. (b) Foreground Extraction

3.2. Detecting Feature Point Correspondence

For the multi-view images or video set, the important step is to find the relevant feature points that correspond to the same 3D point in space. A restricted number of corresponding points, which spread over most regions of a scene, is sufficient to determine the geometric model. Many computer vision applications have feature extraction process as an intermediate step for locating particular elements on an image. While extracting features some of the important factors to be considered are invariance, delectability, interpretability and accuracy. Thus, the first step is to detect the suitable features points in the 2D multiple views and to match the selected feature points among different views. In the implemented system, a feature detection method called the scale invariant feature transform (SIFT) is applied to detect the feature point and generate feature descriptor for each feature point. The feature descriptors are used to match the feature point in different view.

3.3. Scale Invariant Feature Transform

The SIFT consists of five major steps:

(1) Scale-space extrema detection

The first stage of keypoint detection is to identify locations and scales that can be repeat ably assigned under differing views of the same object. Therefore, the scale space of an image is defined as a function, L(x; y; σ), that is produced from the convolution of a variable-scale Gaussian, G(x; y; σ), with an input image, I (x; y):

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y), \tag{1}$$

Where * is the convolution operation in x and y, and

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

To efficiently detect stable keypoint locations in scale space, we have proposed (Lowe, 1999) using scale-space extrema in the difference-of-Gaussian function convolved with the image, D(x; y; σ), which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma). \end{aligned} \tag{3}$$

(2) Keypoint localization

Keypoint localization performs a more accurate localization of the keypoint according to the nearby data, and eliminates points with low contrast. The

parameters of a 2×2 Hessian matrix are used to calculate the ratio of the curvature across the edge and the curvature in the perpendicular direction. If the ratio is larger than a threshold, the keypoint is discarded. The detected SIFT keypointswith the threshold level of 0.03 is as shown in Figure 2.

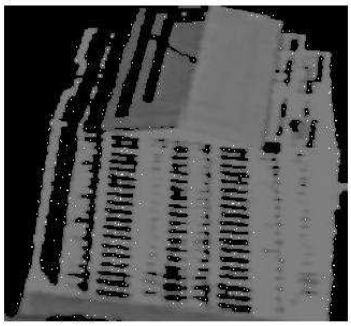

Figure 2.Detected SIFT Features

(3) Orientation assignment

By adding the orientation information of keypoints to the content of the descriptor, the matching process will be invariant to the rotation of objects in different views. Orientation assignment is performed at the detected keypoints by creating a gradient histogram multiplied by the gradient magnitude and a circular Gaussian window. An orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint. The orientation histogram has 36 bins covering the 360 degree range of orientations. For each image sample, L(x; y), the gradientmagnitude and m(x; y), and orientation, θ(x; y), are pre-computed using pixel differences:

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \quad (4)$$

$$\theta(x,y)=\tan^{-1}((L(x,y+1)-L(x,y-1))/(L(x=1,y)-L(x-1,y))) \quad (5)$$

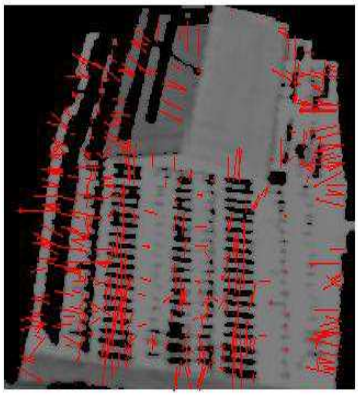

Figure3. SIFT Feature Orientations

(4) Keypoint descriptor

The previous operations have assigned an image location, scale, and orientation to each keypoint. These parameters impose a repeatable local 2D coordinate system in which to describe the local image region, and therefore provide in variance to these parameters. The next step is to compute a descriptor for the local image region that is highly distinctive yet is as in variant as possible to remaining variations, such as change in illumination or 3D viewpoint. The keypoint descriptors are formed to describe the features of the keypoints so that the corresponding keypoints can be tracked with respect to similar features in their descriptors. For a feature point, the 16×16 surrounding area of the keypoint is divided into 4×4 subregions array of histograms with 8 orientation bins in each and is used to calculate the descriptor.Therefore, the experiments in this paper use a 4x4x8 = 128 element feature vector for each keypoint.

(5) Keypoint matching

Using the above steps, the keypoints and their descriptors in each view can be determined. The next step is to relate the keypoints in different views and to find the matching feature points between two different views. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the in variant descriptor vector. This algorithm is able to generate a large number of feature points that are densely distributed over a wide range of scales and most locations in the image, while being robust to scaling and rotation in 3D viewpoints, and to changes in illumination. In the implemented system, the corresponding features are tracked between the two view images, first and second images as shown in figure 1(b). The computed corresponding matching points are as shown in Figure 4.

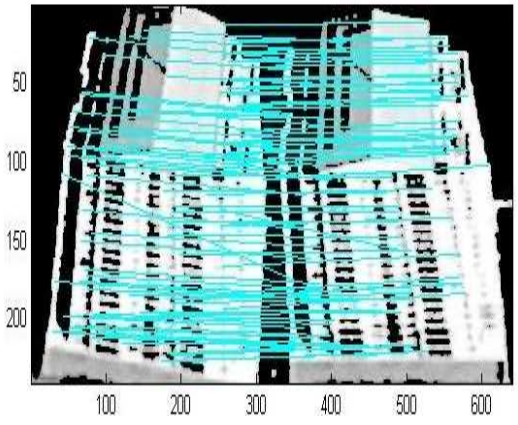

Figure4. SIFT Feature Correspondences

3.4.Random Sample Consensus (RANSAC)

In the feature detection and matching step, since the descriptors of feature points, which are used for feature matching in multiple views, are determined with respect to the local sub-region around the feature points, there may be some inaccurate matching between feature points in multiple views. These mismatching pairs of feature points are also called the outliers. An outlier removal and model estimation method called the (RANSAC) is used to remove the outliers and to estimate the two-view geometry using the inliers feature points. The refined matching points by RANCS between two images are shown in Figure 6. RANSAC is applying to estimate the two-view geometry by finding the fundamental matrix F and removing the outliers in the corresponding image feature points for 3D object reconstruction.

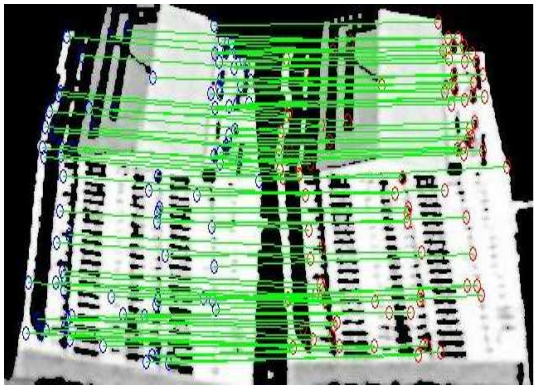

Figure5. Point Match after outlier were removed

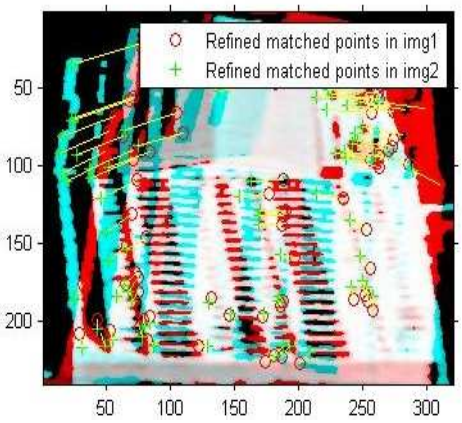

Figure 6. Refined Correspondence Points by RANSAC

4. Experimental Result

The system was tested with MATLAB 2014a software on Core i5 CPU with 2.3 GHz processor with 4 GB of RAM in 64 bit operating system. In order to verify the robust and fast method by SIFT for finding correspondence matching points, this paper conducted experiments with two satellite images from Google Earth from different view-points, and the size of the image of each view is 240×320. Figure 1(b) shows the result of foreground detection and Figure 2 shows the detected SIFT feature. The numbers of detected keypoints of each image is shown in Table.1. The number of matching is 113 matched points and the number of inliers points are 103 pairs after removing outlier points. According to the experiments, the presented SIFT method provides better accurate and more matched points than Harris corner matching method. The performance of the matching procedure is measured by execution time, which is the time taken from getting from each step of the presented procedure. These are shown in Table.1. The overall execution times of feature matching points for Harris corner matching method is taken for 0.531136 and 0.495345 seconds with SIFT. Therefore, presented SIFT is more matched points and faster execution time than other methods as shown in Figure 7and Figure 8 illustrates the comparison chart for four types of matching method based on numbers of match and execution time.

Table 1. Execution Time of each step

	Time taken(us)
Foreground Detection for First Image	0.935248
Foreground Detection for Second Image	0.569913
Feature point Extraction by SIFT	0.623549
Feature Matching	0.278707

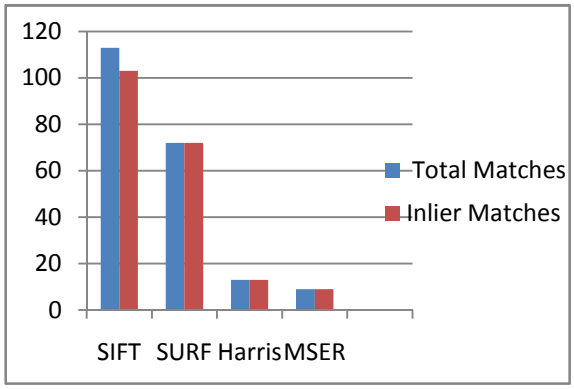

Figure7. Comparison of SIFT and other methods on total matches and Inlier matches

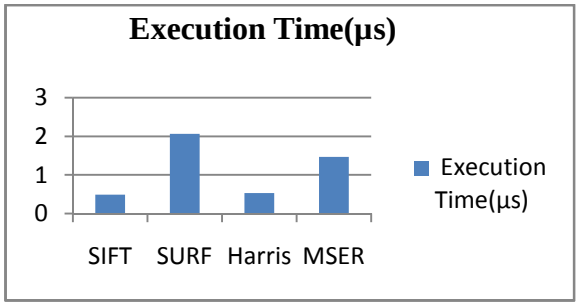

Figure 8. Comparison of Execution Time on Four Methods

5. Conclusion

In this paper, for the 3D modeling system implements feature detection and feature correspondences techniques to relate multiple views. The algorithm of the scale invariant feature transform (SIFT) is used to detect the feature and to match the features among different views. Outlier removal and fundamental matrix estimation algorithm called the RANCS is used to refine the feature matching between two views. The two-view geometry can be estimated using the inliers of feature points. The presented method gives efficiently accurate matched points and better execution time. In future, the correspondences of features points generated from the proposed method can be used as the input for triangulation computation to reconstruct 3D scene.

References

[1] M. Pilu. Uncalibrated Stereo Correspondence by Singular Value Decomposition. Digital Media Department HP Laboratories Bristol HPL-97-96 August, 1997 .

[2] C. Harris and M. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pages 147.151,1988.

[3] D. G. Lowe, ―Object recognition from local scale-invariant features, in Proceedings of the International Conference on Computer Vision, vol 2, (1999): 1150–1157.

[4] Li, Qiaoliang, et al. "Robust scale-invariant feature matching for remote sensing image registration." Geoscience and Remote Sensing Letters, IEEE 6.2 (2009): 287-291.

[5] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. “Towards internet scale multi-view stereo”. IEEE Conf. on Computer Vision and Pattern Recognition, 2010.

[6] J.Ma, J.Zhao, J.Tian, A.L. Yuille and Z. Tu. “Robust Point Matching via Vector Field Consensus”, IEEE Tansactions on Image Processing, vol. 23, No.4, Aprial 2014.

[7] J.Ma, J.Zhao, J.Tian, A.L. Yuille and Z. Tu. “Robust Point Matching via Vector Field Consensus”, IEEE Tansactions on Image Processing, vol. 23, No.4, Aprial 2014.

[8] K.k.k.Theint and M.M.Sein, “3D Building Reconstruction from Google Earth Images”, International Conference on Computer Applications (ICCA2014), Yangon, Myanmar, 2014.