

ARCHITECTURAL SCENE RAPID RECONSTRUCTION BASED ON FEATURES

Y. Ding^{a,b,*}, J. Q. Zhang^b

^aInstitute of Earth Observation and Space System Delft University of Technology
Kluyverweg 1, 2629 HS, Delft, The Netherlands - Yi.ding@tudelft.nl

^bSchool of Remote Sensing Information Engineering Wuhan University
129 Luoyu Road, Wuhan 430079, China

KEY WORDS: 3D reconstruction, feature matching, differential evolution, visualization, robust estimation

ABSTRACT:

Retrieve the structure of model and the motion of camera is a classical and hot topic in computer vision and photogrammetry. A lot of automatic or semiautomatic techniques have been developed to optimize the retrieving processing from accuracy, stability and reality perspectives. These techniques are variant from data source, feature selection for matching, feature clustering and 3D model representation. The optimization algorithm and a completely automatic system are still under exploring. In this paper, we use some image-based algorithms for feature selection and matching of 3D man-made scene reconstruction. We present a robust point matching algorithm with RANSAC estimator, and compare two methods of line matching in a complex man-made environment. We point out the degeneracy when use epipolar line as a constraint to match line, instead use a global optimization method. Our experiments show that the proposed method is robust in a complex man-made scene.

1. INTRODUCTION

Observing the world through the eyes of machines or imitating the human ability of perception is the essence of computer vision and photogrammetry. Computer vision is an integrated subject of computer technology, image processing, pattern recognition, and computer graphics. There are many interesting research areas in computer vision, for example, robot navigation, traffic tracking, face recognition, recovering 3D structure of the environment, and so on. Among these research areas, recovering architectural 3D models from complex man-made environment has been a very hot one in recent years. Applications of architectural 3D models are, for example, tourist navigation and heritage protection. There exist a wide variety of methods related to model recovery (Baillard, 2000 and Cantzler, 2002). The differences between these methods are from the manner of getting data to the representation of the scene model. The most appropriate model recovery method depends on the type of a scene that is to be reconstructed, and the application requirements like real-time/unreal-time, detail representation degree, and etc..

Figure.1 is a general flow chart describing the processes of model reconstruction. Most of the existing image-based methods are based on three-stage process (Bartoli, A., 2007). First, sparse features (points, lines, ect.) are extracted and matched in multi-view and then 3D features and camera pose are reconstructed by using structure from motion techniques. The remaining two stages are scene model selection and parameter estimating. After these steps we can then represent the model in the form of a dense depth map, triangular mesh or a set of space planes.

When considering man-made environment, we choose distinct points to recover the camera motion and then to achieve plane-based model by clustering reconstructed line features. This plane-based model is motivated by reasons like very

constrained, compact representation and modify the reconstruction easily as described in (Bartoli, 2007; Hartley, 2000).

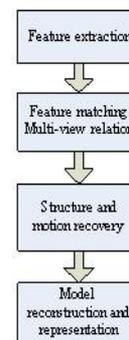


Figure.1 Flow chart of reconstruction.

Real man-made scenes are usually very complex with occlusion and noise. In this paper, we combine several algorithms to present a stable and automatic processing for recovery the structure of a scene and the motion of the camera with distinct features of point and line. We present a robust point matching algorithm with RANSAC estimator, and compare two methods of line matching. We point out the degeneracy when use epipolar line as a constraint to match line, instead use a global optimization method.

The rest of the paper is organized as follows. Point feature extracting and matching to recover the motion of the camera are described in section 2. The line extracting and matching approach with constraints are discussed in section 3. Conclusions and future work are presented in section 4. Each

* Corresponding author.

section is provided with experiments results of real man-made scene.

2. STRUCTURE FROM MOTION

As described in (Hartley, 2000) and (Marc, 2004), we can recover the structure of scene and the motion of camera from single or multi-view. In this paper we consider multi-view. The critical problem of reconstruction model in multi-view is to find corresponding features in the images. In a complex man-made scene, even advanced point extracting algorithm like Scale-invariant feature transform (SIFT) (Lowe, 2004) still induce a lot of wrong matches. In such case, a traditional least-squares based approach will fail to compute the fundamental matrix. Therefore a robust method is needed.

2.1 Feature extraction and matching

Typical point extraction and matching approaches make use of the Harris operator to extract corner points in multi-view separately and then compare them with an intensity constraint using dissimilar measurement, e.g. sum-of-square-differences (SSD) or zero-mean normalized cross-correlation (ZNCC). These measurements are invariant to image translation and are difficult to choose measuring window size especially in repeated or deficient texture region. Therefore we need an advanced approach like SIFT to cope with large variations in camera pose.



Figure.2 Distinct points extracted by Harris (top) and SIFT (bottom).

Figure.2 shows a comparison of results from Harris and SIFT extracting operators. The top image uses Harris operator, and the extracted corner points are of high precision. However, some points on the tree are also extracted. These points are not needed, because they can not be differentiated from their neighbourhood region, and then induce wrong matching. We need to delete these points before further processing, otherwise,

this set of points will result a lot of outliers. The bottom image uses SIFT operator to extract points which are almost on the main building. We don't need a pre processing step compared with using Harris operator. Points extracted by SIFT are highly distinctive. This is because the SIFT operator takes advantage of scale-space extrema detection, and detected points are local extrema with respect to both space and scale.

2.2 Computation of fundamental matrix

The fundamental matrix expresses the geometry structure between two views. The general method needs at least 8 corresponding points, $m_i \leftrightarrow m'_i$, to solve linearly matrix F which satisfies the condition $m'_i F m_i = 0$. With more than 8 pairs of points, a least-squares approach minimizes the cost function in equation (1)

$$C(F) = \sum (d(m', Fm)^2 + d(m, F^T m')^2) \quad (1)$$

When the outliers are more than 50%, the least-squares approach will fail. We use a well-developed estimation method, RANdom Sample Consensus (RANSAC) (Fischler, 1981), to detect outliers. The results before and after outliers deleting are shown in

Figure.3. Each figure is superimposed by two views, and the two end points of each red line in the figure denote a corresponding point pair.

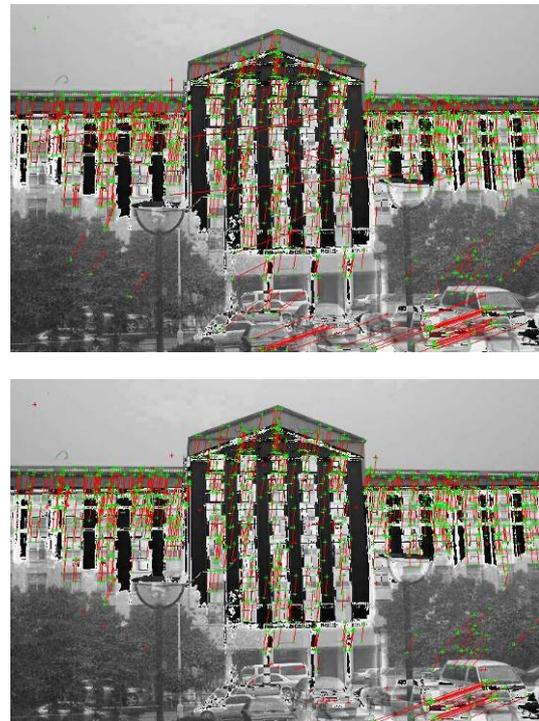


Figure.3 The top image with outliers and the bottom image without.

2.3 Computation of the camera matrices and computation of points in space

A perspective camera is modeled through the projection equation $m \sim PM$. The camera matrices P corresponding to a

fundamental matrix F may be chosen as $P = [I | 0]$ and $P' = [[e']_x F | e']$ (Hartley, 2000). The reconstructed points will be shown in Figure.8 in the next section.

3. LINE MATCHING AND RECONSTRUCTION

When line features are extracted by a detection operator like Canny, many variant methods can be used for matching. A fast and stable matching method should satisfy two critical requirements: appropriate search range and distinct dissimilar measurement. Usually, epipolar line geometry and intensity information are two effective constraints. So first we try to use these two constraints to match lines.

3.1 Epipolar line constraint and degeneracy

Figure.4 shows the geometric constraint described as epipolar line. The two end-points of a line segment generate two epipolar lines in the other image. These two lines intersect at epipole e' . The corresponding line segment should be necessarily intersected or contained in the shadow range of the right figure in Figure.4. Perhaps more than one line are contained in this range. These lines are all regarded as candidate corresponding lines. Then we compare the similarity of the intensity neighbourhood of the each candidate line with respect to the intensity neighbourhood of the original line to match them uniquely.

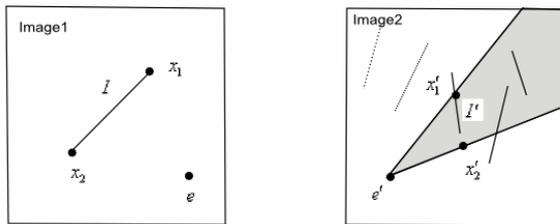


Figure.4 Applying the epipolar line to reduce the search space of candidate lines.



Figure.5 Top row pictures show the epipolar line constraint for a vertical line, while bottom row pictures show this constraint for a horizontal line. In four pictures, red lines represent extracted feature lines and blue lines represent epipolar lines.

However, there might be degeneracy, that as described in (Hartley, 2000), lines in 3-space lying on epipolar planes cannot be determined from their images in two views. The degeneracy

usually occurs when we get images what are almost parallel to the space object surface. For example, in the first row of Figure.5, the search range is clear and doubtless for the vertical lines, but for the horizontal lines in the second row, the search range becomes narrower and is difficult to confirm. This is a big problem when use eoipolar line as constraint condition for line matching. Therefore, we need to find a better solution.

3.2 Homography constraint

The projective geometry of two cameras is described as $m' \sim Hm$, where H is the homography plane. Although the object building is not a plane, when compared to the distance between the camera center and the object, we can regard a facade as an approximate plane. Since we just need homography condition to restrict the matching search range, we don't need very high precision. Figure.6 illustrates the relation between homography and epipolar geometry. From the figure, we can see that, if the space point is out of the homography plane π , then $m' \neq Hm$, where m and m' are a pair of points corresponding to a same space point M . We try to determine H in two different ways. The first way is using redundant corresponding points to find an optimal solution with a least-squares approach. The second one is featureless based on a global optimization method inspired by the differential evolution (DE) algorithm (Price, 2005). This method has been successfully used for image registration (Karimi, 2008).

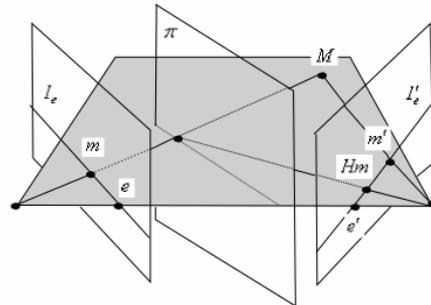


Figure.6 Relation between homography and epipolar geometry. Any space point M mapped by the homography plane π lies on its corresponding epipolar line l'_e .

Using redundant corresponding points, a least-squares approach minimizes the energy function in equation (2)

$$e = \min \sum d(Hm, m')^2 \quad (2)$$

The DE algorithm performs a global search in the parameter space, using N vectors $\{x_{i,G} | i = 0, 1, 2, \dots, N-1\}$ as a population for each generation G , where $x = [x_0, x_1, x_2, \dots, x_{D-1}]^T$ is a D -dimensional parameter vector. The initial population of DE is chosen randomly. DE generates new parameter vectors by adding the weighted difference between two population vectors to a third vector according to

$$v_{i,G+1} = x_{r_1,G} + F(x_{r_2,G} - x_{r_3,G}) \quad (3)$$

where $r_1, r_2, r_3 \in [0, N - 1]$ are randomly chosen integers. A given energy function can determine whether the old vector should be replaced by the new vector. When processing terminates, the final parameters correspond to a global minimum energy value.

In the case of H matrix, we give an energy function as follows:

$$e = \min \sum_x \sum_y (I_i(x, y) - I_{i+1}(x', y'))^2 / n \quad (4)$$

where $I_i(x, y)$ is the pixel value of point m in the image i and $I_{i+1}(x', y')$ is the pixel value of point m' in the image $i + 1$. m' is calculated from $m' \sim Hm$.

Figure.7 shows the difference between the original and the transformed images. The transformed image satisfies the constraint condition $m' \sim Hm$. Image (a) shows the difference between two original images. In image (b), the transformed image comes from H matrix which is calculated using redundant corresponding points. These point pairs are obtained from SIFT extraction operator described in section 2.1. Since the points congregate at the top of the image, the bottom part of the image is out of control. Result using DE is shown in image (c). The bottom part still has corresponding problem because it has comparative depth different with the surface of the object building. But for the top part, the result is good enough for the reason of constraint matching search. Image (d) shows that, in DE algorithm, the energy value concentrates to a global minimum with enough generation.

Using homography and intensity constraints, it is easy to match corresponding lines in two adjacent images. Using the camera matrices got from section 2.3, we can get the reconstructed 3D lines from the matched corresponding lines. The extracted 2D lines and the reconstructed 3D lines compared with the reconstructed 3D points are shown in Figure.8

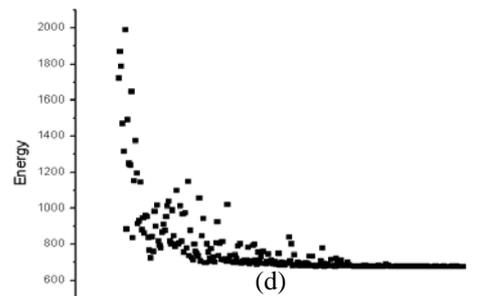


Figure.7 Homography matrix got from redundant corresponding points and differential evolution.

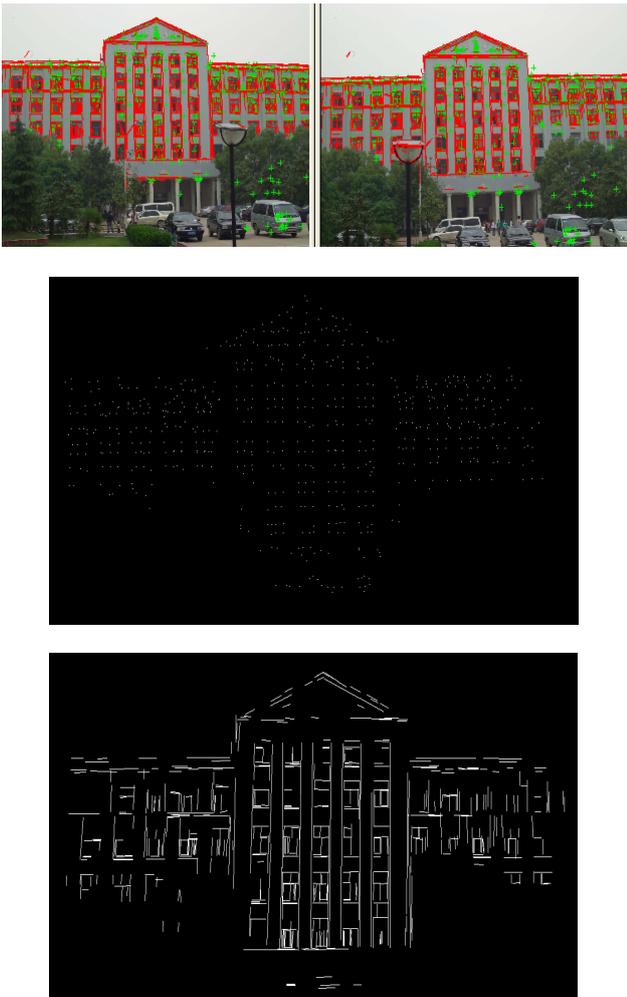


Figure.8 Extracted lines in adjacent images and reconstructed 3D lines compared with reconstructed 3D points.

4. CONCLUSION AND EXTENDED WORK

In this paper, we described robust points matching algorithm with RANSAC estimator, and compared two methods of line matching in a complex man-made environment. Through comparison, we conclude that, differential evolution is a global optimization method and it tolerates complex man-made scene which usually has a mass of noise. To represent reconstructed model with planes, further work includes clustering reconstructed 3D lines into planes and evaluating model parameters.

REFERENCES

- Baillard, C., Schmid, C., Zisserman, A., Fitzgibbon, A., 1999. Automatic line matching and 3D reconstruction of buildings from multiple views. *Proc. ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, 32(PART 3-2W5), pp.69-80.
- Baillard, C. and Zisserman, A., 2000. A plane-sweep strategy for the 3d reconstruction of building from multiple images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 33(B2), pp. 56-62.
- Bartoli, A., 2007. A random sampling strategy for piecewise planar scene segmentation. *Computer Vision and Image Understanding*, 105(1), pp. 42-59.
- Cantzler, H., Fisher, R.B., 2002. Improving architectural 3D reconstruction by plane and edge constraining. *In Proc. of the British Machine Vision Conf. (BMVC '02)*, Cardiff, U.K., pp. 43-52.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp.381—395.
- Hartley, R. and Zisserman, A. 2000. *Multiple view geometry in computer vision*, Cambridge University Press: Cambridge, UK.
- Karimi Nejadasl, F., Gorte, B.G.H., 2008. Optimization based image registration in the presence of moving objects. *International Calibration and Orientation Workshop*, Castelldefels.
- Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2)s pp. 91-110.
- Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, Reinhard Koch, 2004. Visual Modeling with a Hand-Held Camera. *International Journal of Computer Vision*, 59(3), pp.207-232.
- Price, K. V., Storn, R. M. and Lampinen, J.A., 2005. *Differential Evolution: A Practical Approach to Global Optimization*. First edn, Springer.

