SEQUENTIAL CORRESPONDENCE OF MULTIPLE VISUAL CUES FOR 3-D BUILDING RECONSTRUCTION

Guo Dong, Yan Chye Hwang
DSO National Laboratories,
20 Science Park Drive,
Singapore 118230

Liew Soo Chin, Kwoh Leong Keong
Centre for Remote Imaging, Sensing and Processing,
National University of Singapore,
Singapore 119260

1. INTRODUCTION

Stereophotogrammetry, using an overlapping stereo pair of images, has been a major tool for generating the 3-D information of buildings in earth maps. With the stereoscopy, a 3-D stereo model is formed in the brain of human operator by the process of binocular vision. The 3-D depth of the building of interest is perceived using two fundamental cues: double images and inward convergence of optical axes. To estimate the height of the building top, human operators have to manually drive the 3-D mouse (i.e. a mouse with height control) to delineate the building footprints and roof structure under a stereo view environment. As one of the most robust solutions in 3-D depth acquisition, stereoscopic viewing is very labor-intensive and inefficient in the stereophotogrammetry. This is due to the fact that the human brain can only accommodate a 15% scale difference between two images. Stereoscopy becomes difficult beyond it.

With recent advancements in computer vision and image processing, automatic building reconstruction from an IKONOS stereo pair has received a fair amount of attention. A recent survey of 3-D building reconstruction can be found in [1], including data-driven, model-driven and hybrid approaches etc. All approaches considered so far try to solve the problem of 3-D building reconstruction using some simple cues (corners, edges, simple building models etc.). However, they ignore the high-level information present in the image intensity structure. In this work, we propose a subpixel estimation of 3-D building height from an IKONOS stereo pair. Given the user-defined outline of building top in the left image, stereo correspondence is automatically established through a multi-scale feature (coarse-to-fine) matching. The height of the building top is robustly estimated with the help of RPC models. Experimental results show that the proposed method is accurate and robust. It produces the results that are comparable to those obtained from the stereophotogrammetry system.

2. METHODOLOGY

Using stereo vision techniques, the 3-D building height can be estimated from an IKONOS stereo pair. With an accurate calibration of camera intrinsic and extrinsic parameters, the 3-D building height is calculated by the triangulation of building tops from the two images. The main challenge lies in the stereo correspondence, i.e. computing a pair of building tops in two image spaces that result from the same building in the object space. In addition to the feature similarity, epipolar geometry provides another strong constraint of projective geometry for the stereo correspondence. Given a point in one image space, its possible matches in the other image space all lie on an epipolar line. The epipolar geometry is independent of scene structure, and only depends on the cameras’ internal parameters and relative pose. In an IKONOS stereo pair, camera intrinsic, extrinsic parameters and epipolar geometry are implicitly formulated by a simple, secure and accurate representation - RPC.

2.1. Stereo Correspondence with RPCs

Given a 3-D point \((X, Y, Z)\) in the object space, RPC model defines a polynomial relation that projects the 3-D point to a corresponding 2-D point in the image space. The stereo correspondence can be performed by the following approximation. 1) select a 2-D point \((r_l, c_l)\) in the left image space. 2) set the \(Z\) coordinate of \((r_l, c_l)\) in the object space. 3) compute the initial approximate coordinates \(X\) and \(Y\) using the first order term of left RPCs. 4) obtain the error equations in terms of \((X, Y)\) by the first order Taylor expansion. 5) calculate \((\Delta X, \Delta Y)\) by the least square and add them to \(X\) and \(Y\). 6) repeat 4 and 5 for the maximum number of iterations. 7) calculate the 2-D point \((r_r, c_r)\) in the right image space based on estimated \(X\), \(Y\) and \(Z\) using right RPCs. The accuracy of stereo correspondence is sufficient for the estimation of building height.
2.2. Estimation of 3-D Building Height

A semi-automatic optimization is used for the estimation of 3-D building height in this paper. The outline of building top in the left image is initialized by a user-defined polygon. The salient features of the building top are those pixels with strong gradients, which can be simply detected from the image. A multiscale feature matching (coarse to fine) is used to search for the optimum height of 3-D building, which is illustrated in Figure 1.

2.2.1. Similarity Measure

The similarity measure of building top faces the challenges of strong pixel variations, occlusions, shadows and repeating patterns in the IKONOS stereo pair. We employ the measure of Implicit Similarity [2] rather than NCC or MI etc. It is a measure of gradient energy that is invariant to intensity dissimilarities between two images. Only pixel positions in the left image are used for the measure. The adaptive weights are combined in it for the robust estimation against the outliers and noisy pixels.

2.2.2. Multiscale Feature Matching

The 3-D building height is iteratively estimated by a coarse-to-fine feature matching within the range of height \( H = [H_{\text{min}}, H_{\text{max}}] \), which is shown in Figure 2. Give an estimated height \( H \), each salient point in the left image is corresponded to the right image based on a subpixel RPC right projection. The pixel value of a corresponding point and its 8 neighbors in the right image are estimated by a bilinear interpolation. The coarse matching locates the local maxima of corresponding costs in terms of building height. The optimal matching is obtained by the filtration of false maxima based on cost values and peak shapes. It is further refined by the Mean Shift in the fine matching.

3. RESULTS AND COMPARATIVE STUDIES

The proposed approach has been evaluated using a typical database of 50 buildings generated by SOCET SET. It is able to produce a comparable result with an accuracy of 5 meters. Experimental results show that RPC right projection is a very effective solution to reduce the matching region in the right image. The coarse-to-fine feature matching is a practical solution for accurate and efficient computation. The robust estimation removes the outliers and noisy pixels, it increases the reliability of similarity measure.

4. REFERENCES
