A Novel Wavelet Stereo Matching Method to Improve DEM Accuracy Generated from SPOT Stereo Image Pairs

Yan Xia Anthony Tung Shuen Ho YanWen Ji

School of Electrical and Electronic Engineering

Nanyang Technological University, Singapore 639798

Tel. (65)7905465, Fax: (65)7904161, E-mail address: eyxia@ntu.edu.sg

Abstract: This paper presents a novel stereo matching algorithm using wavelets to obtain a dense disparity model for DEM generation from complex SPOT stereo image pairs of a region in HongKong. The SPOT stereo images are first adjusted to derive the corresponding epipolar image pairs such that the matching points lie on the corresponding epipolar lines of the two images. A new edge detector based on bi-orthogonal wavelets is then used to extract the minima images from the epipolar image pairs. The edge features and the dense disparity estimate are derived from the minima images using a novel hybrid stereo matching (HSM) method that combines both feature-based and area-based approaches. Incorporating the dense disparity estimate, HSM then generates a final dense disparity map using the epipolar image pairs. The results obtained indicate that this algorithm can improve 3-D topographic accuracy better than four pixels with a fewer number of control points needed.

Keywords: Wavelets, DEM, SPOT, Area-based Matching, Feature-based Matching

INTRODUCTION

Passively sensing three-dimensional structure by means of computational stereo image has been addressed both from the biological and computational points of view. In 1982, Barnard and Fischler defined six necessary steps for stereo analysis: image acquisition, camera modeling, feature acquisition, image matching, depth reconstruction, and interpolation [1]. Of these steps, image matching is considered to be the most difficult to solve and is dependent on the choice of feature primitives. When the matched features are low level and dense, such as the intensity at each pixel, and the matching strategy is applied locally throughout the image at each pixel, we call the matching strategy “area-based.” When a few interesting (usually more abstract) features are first selected from the image and the matching strategy is applied to this subset, we use the term “feature-based.” [2] Both ways have advantages and disadvantages. We use a hybrid approach that integrates “area-based” and “feature-based” primitives. This approach allows it to take advantage of the unique attributes of each of these techniques. Moreover, the novel stereo matching algorithm combines the advantage of wavelets to improve DEM accuracy generated from SPOT stereo images pairs.

In the remainder of this paper, we demonstrate the usefulness of combining information from area-based and feature-based approaches for DEM generation from complex

SPOT stereo image pairs of a region in HongKong. Beginning with an adjustment for image pairs to derive the corresponding epipolar images pairs, an edge detector based on bi-orthogonal wavelets is used to extract the edges images from the epipolar image pairs. The wavelet transform is used because the evolution across scale of its magnitudes and orientation can be used to characterize localized signals like edges and corners. The dense disparity estimate is then derived from the edges images using a novel hybrid stereo matching (HSM) method that combines both feature-based and area-based approaches. Incorporating the dense disparity estimate, HSM generates a final dense disparity map using the epipolar image pairs. Result obtained from HSM indicates a better performance in terms of DEM accuracy and computational efficiency.

DESCRIPTION OF THE APPROACH

Figure 1 is the flow diagram of stereo matching for SPOT image pairs.

Initial Processing

In both area-based and featured-based technique, the epipolar constrain is used to simplify stereo matching to a one-dimensional problem – along the image row direction. But, we cannot assume that unrectified SPOT stereo images would satisfy the epipolar geometry constraint because SPOT platform would be subjected to nonlinear motions during the period of image acquisition.

Some initial processing steps are performed to form the

Figure 1 Flow Diagram of Stereo Matching

collinear images. First, the image pair is cropped to give a series of geometry rectifies based on the images acquisition parameters, e.g. satellite latitude, longitude, height, and velocity, off-near viewing. The second step is to provide an approximate epipolar alignment. This is done by providing matches for four points, where each point is in the neighborhood of each corner of the images. These are used to generate a translation and skewing of all points in the images.

Feature-Based Process

Feature-based analysis provides more precise positioning for the edges where there is an abrupt change in intensity. However, the feature-based processing provides only sparse matches (Figure 2-c) and requires interpolation as well as some method for modeling occlusion. It is hard to incorporate the smoothness assumption into the matching strategy. For the complex SPOT images, feature-based processing is not useful to obtain directly a dense disparity and a depth map -- DEM (Figure 2-d); but it is good idea to computer precise edge's positions and an estimate of local image texture for further area-based processing dense disparity generated.

An edge detector based on bi-orthogonal wavelets is used to extract the edges images--the minima images from the epipolar image pairs derived in initial processing. The edge detector incorporates the bi-orthogonal wavelet filter using B-spline designed by Yu and Ho[3]. The bi-orthogonal wavelet filter is used to produce edge maps with different degree of detail, Figure 2-b. Edge maps are then matched using adaptive matching windows that vary their shapes according to the directions of the edges. An adaptive searching range is applied because the searching range of each point may be different[4]. Then, the disparity estimate is derived from the minima images.

Area-Based Process

Area-based methods have been applied successfully to the analysis of aerial terrain images, where the surface varies smoothly and continuously[5]. They offer the advantage of directly generating a dense disparity map but are sensitive to noise and breakdown where there is a lack of texture or where depth discontinuities occur. In general, a smaller window is desirable to avoid unwanted smoothing, however, a larger window is needed so that the window contains intensity variation enough to achieve reliable matching in areas of low texture[6]. So, the selection of an appropriate window size is critical to achieving a smoothing and detailed disparity map.

In order to remedy this serious problem in window-based stereo methods, the window size and shape are adaptively adjusted based on the local amount variation in texture and current disparity estimates. The input to the area-based process is a stereo pair of epipolar images and a pair of disparity estimates from feature-based processing.

A. Local Variation. Equation (1) is used to compute the

Figure 2 HongKong SPOT Image Pair

local variation of the image texture to mark areas for which no match would be generated.

$$V_{x,y} = \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} 1 - \frac{\overline{x_j}}{\sqrt{\left(\frac{n-1}{n}\right) \sigma_j^2 + \overline{x_j}^2}} \quad (1)$$

Where the mean $\overline{x_j}$ and variance σ_j^2 are calculated along each row of an $n \times m$ window. $V_{x,y}$ becomes a minimum of zero when all of the values in each row of the window are the same, there are no features within the correlation window. $V_{x,y}$ reaches a maximum when variance of the row is large, and is indexed by the mean intensity so that any texture present in darker by regions is emphasized.

B Correlation: A normalized cross-correlation of the two intensity images is calculated, which creates a volume whose dimensions are the width of the left image by the width of the right image by the number of rows as given in equation (2).

$$C_{x,y,d} = \frac{(\sum_A Left_{x,y} Right_{x+d,y})^2}{\sum_A Left_{x,y}^2 \sum_A Right_{x+d,y}^2} \quad (2)$$

Where A is the correlation window centered at the pixels $(x, y : x + d, y)$, size changed between 11×11 and 19×19 based on disparity estimate.

C. Correct Match: Based on the best correlation and best match's rule, the correct matched point can be found by simply extracting the peaks in the array. Once this is done, we focus our attention on the problem areas only. The first

occurs in areas without measurable texture. For these, it is important not to generate matches from the cross-correlation information. Second, where multiple matches occur, we select the highest peak (best correlation) on which both views can agree, and assume that order reversals due to the spatial shift between viewpoints cannot occur within the relatively narrow fusion interval. Third, for incorrect areas representing points on which the two views cannot agree, these area may be replaced by the result of interpolation or feature process.

Moreover, when multiple matches occur, we iteratively apply the following constrains to identify the correct matched point and remove unlikely matches.

- a) Two-Views Constraint: If a pixel of the left image is “matched” to one in the right image that, in turn, has not been matched to that same pixel in the left image, either one or both matches are incorrect. Interpolation points do not need to obey the constraint.
- b) Order Constraint: With a defined order of points along a line in one image of the pair, the corresponding points have to occur in the same order in the other image.
- c) No Isolated Pixels Constraint: Isolated pixels are those pixels that have a disparity that differs more than 2.5 pixels from the average value in the surrounding 5*5 neighborhood. And, pixels for which fewer than six (of its 24) neighbors have nonnull values are also removed.

Integrated and, Interpolation and Surface Generation

Final dense disparity map is the edge-feature disparity result associated with the disparity obtained in the area-based process. Since depth and orientation discontinuities usually give rise to these intensity edges, some of the detected edges

Table 1

North (m)	Calc-North	East (m)	Calc-East	Elevation m	Calc-Elv
2491703.786	2491682.337	801646.126	801695.919	4.051	-22.785
2501096.186	2501084.58	837148.837	837143.378	5.698	-8.323
2460389.027	2460361.729	798676.642	798639.219	58.775	20.562
2471026.887	2471049.354	830358.202	830366.677	5.418	1.205
2480929.790	2480945.087	799368.802	799383.904	11.413	-6.193
2474672.167	2474669.936	817531.151	817529.975	16.228	7.972
2480111.842	2480102.066	799785.566	799794.950	4.111	-8.980
2476089.280	2476109.104	846210.698	846246.760	68.620	25.487
2486528.231	2486520.405	834189.351	834185.143	4.402	5.619
2497461.386	2497437.727	822666.389	822671.387	11.357	-12.998
2485803.665	2485812.327	826738.677	826742.282	6.601	-2.005
2476821.932	2476810.952	800328.062	800314.251	4.157	-0.312
2463012.047	2462987.752	794891.513	794826.729	3.616	27.248
2467876.852	2467850.662	811866.845	811857.368	2.740	15.058
2471876.921	2471875.575	820115.735	820114.885	5.070	12.993
2479328.896	2479323.271	829556.953	829556.86	0.206	9.790

should correspond to discontinuities. Therefore, we smooth the disparity map, keeping the disparity at the edges fixed. As for unmatched area points, we employ an interpolation algorithm to produce a rough disparity.

The approach goal is not to produce a dense disparity image of the scene but also eventually to calculate the disparity to generate a description of the scene entity – DEM. We use 3-D Direct Linear Transformation (DLT) Model to calculate entities depth. 21 ground control coordinates are measured from large scale HongKong’s topographic maps. They are shifted to the center of the area covered by the scene. A least squares adjustment is applied to the DLT model to determine the 16 liner orientation parameters of the left and right images. Then stereo intersection is implemented independently to calculate the ground coordinates of conjugate points-- produce the DEM of the HongKong SPOT image. The results indicate that DEM accuracies can be achieved within four pixels, Table 1.

CONCLUSION

We develop a complete and working implementation of a stereo algorithm and have presented some ideas whose use in stereo correspondence provides some impressive results on DEM generated from SPOT image pair. In particular, we provide one use for those edges; serve as boundaries and disparity estimate to limit the search for discontinuities in depth or orientation. And, we have also applied a number of constraints from area-based approaches to stereo matching into the selecting a match from the correlation volume to resolve the ambiguity and verify the matches. When there are no edge features, it acts as a good area-based process, and when there is no texture, it acts as a feature-based system. Of course, we have only used the simplest integration of area and feature-based methods to demonstrate the feasibility of a more complete system.

REFERENCE

- [1]. S. Barnard and M. Fischler, “Computational stereo”, ACM Comput. Surveys, vol. 14, no. 4, pp. 553-572, 1982
- [2]. Steven D. Cochran and Gerard Medioni, 3-D Surface Description from Binocular Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 10, October, 1992
- [3]. T.Yu, A. T.S.Ho, “ A Novel hybrid Bi-orthogonal Wavelets/ADPCM Algorithm For Very Low Bit Rate Sattelite Image Compression”, IGRASS 99’, Hamburg, Germany
- [4]. Yanwen Ji, A. T. S. Ho, et el, An Efficient Novel Multi-resolution Hybrid Matching Method, SPIE’s Aerospace/Defense Sensing, Simulation, and Control, 16 - 20 April, 2001, at Orlando, FL,USA
- [5]. K. Mori, M. Kidode, and H. Asada, An iterative prediction and correction method for automatic stereo comparison, Comput. Graphics Image Processing , vol.2, pp. 393-401, 1973
- [6]. C. Lawrence Zitnick, Takeo Kanade, A Volumetric Iterative Approach to Stereo Matching and Occlusion Detection, 1998 Carnegie Mellon University