Research Article

Stereo Matching Algorithm Based on 2D Delaunay Triangulation

Xue-he Zhang, Ge Li, Chang-le Li, He Zhang, Jie Zhao, and Zhen-xiu Hou

State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150080, China

Correspondence should be addressed to Ge Li; lige@hit.edu.cn

Received 5 January 2015; Accepted 26 May 2015

Academic Editor: Simon X. Yang

Copyright © 2015 Xue-he Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To fulfill the applications on robot vision, the commonly used stereo matching method for depth estimation is supposed to be efficient in terms of running speed and disparity accuracy. Based on this requirement, Delaunay-based stereo matching method is proposed to achieve the aforementioned standards in this paper. First, a Canny edge operator is used to detect the edge points of an image as supporting points. Those points are then processed using a Delaunay triangulation algorithm to divide the whole image into a series of linked triangular facets. A proposed module composed of these facets performs a rude estimation of image disparity. According to the triangular property of shared vertices, the estimated disparity is then refined to generate the disparity map. The method is tested on Middlebury stereo pairs. The running time of the proposed method is about 1 s and the matching accuracy is 93%. Experimental results show that the proposed method improves both running speed and disparity accuracy, which forms a steady foundation and good application prospect for a robot's path planning system with stereo camera devices.

1. Introduction

In recent years, mobile robot vision navigation research has mainly focused on obtaining three-dimensional information of the robot surroundings accurately and in real time. Stereo matching algorithm is a key issue in three-dimensional scene reconstruction. Running time and matching accuracy play a vital role in mobile robot visual navigation and autonomous positioning.

Scharstein and Szeliski [1] presented a taxonomy of dense, two-frame stereo methods to assess the different components and design decisions made in individual stereo algorithms. According to whether or not the algorithm includes a global optimization function, stereo matching algorithms are divided into local and global methods. Typically, local stereo matching algorithms [2] use a support window of fixed shape and size to calculate the matching cost. The disparity is then obtained after aggregating the matching cost by summing or averaging over a support region. Stereo algorithms based on local correspondence are typically fast. Nevertheless, an adequate choice of window shape and size is necessary, as it leads to a trade-off between low matching

ratios for small window sizes and border bleeding artifacts for larger ones. As a consequence, poorly textured or ambiguous surfaces cannot be matched consistently. Meeting the needs of practical applications is difficult. Algorithms based on global correspondences overcome some of the aforementioned problems by imposing smoothness constraints on the disparities in the form of regularized energy functions. Given that optimization methods such as MRF-based energy functions are generally NP-hard, various approximation algorithms have been proposed, for example, graph cuts [3], belief propagation [4], and max-flow or simulated annealing. Although global optimization methods can acquire high accuracy disparity space map (DSI), the modeling process is complex that they generally require large computational efforts and high memory capacity even on low-resolution imagery. For example, the running time of the graph cut method is nearly 20 s or longer. Therefore, further study on local methods according to the needs of practical applications is valuable.

To solve the mutual constraints between running time and matching accuracy, much research effort has been done and numerous optimum methods have been presented.

Zhou et al. [5] presented a fast stereo matching algorithm based on support point expansion and used Gibbs Random Field to describe the energy function. Compared with traditional belief propagation or graph cut algorithm, the algorithm has the advantage of good matching accuracy and running speed. Mei et al. [6] used AD-census method to initialize the matching cost and improved the matching accuracy by adding the smoothing constraints along the scan line after aggregation. Efficient large-scale stereo matching algorithm (ELAS) [7] builds a priori on the disparities by forming a triangulation on a set of support points which can be robustly matched. Computing the left and right disparity maps for a one megapixel image pair takes about one second on a single CPU core. However, the accuracy of disparity map should be improved when it is applied in the three-dimensional reconstruction or target recognition. The basic idea of the literature [8, 9] is to apply the triangulation method to accomplish disparity estimation. The triangle area is used as the matching unit, and all pixels within the triangle are assumed to have the same disparity. Even though it can obtain a dense literature [10], the running time is slower and is likely to cause excessive smoothness in some triangle areas.

In this paper, we propose an effective local stereo matching method based on Delaunay triangulation for stereo matching, which allows dense matching with small aggregation windows by reducing ambiguities on the correspondences. Our approach builds the disparity model over the disparity space by forming a triangulation on a set of robustly matched correspondences, named support points, which are detected by a Canny edge operator. The efficient algorithm reduces the search space and can be easily parallelized. As triangles share a joint vertex, the initial disparity of each vertex is refined. As demonstrated in our experiments, our method achieved good performance when compared with prevalent approaches.

2. Support Points Generation

As support points, we denote pixels that can be robustly matched due to their texture and uniqueness. We are inspired by the characteristic that an image edge contains most of the image information and widely exists in objects and backgrounds or between objects.

The common edge detection operators are Prewitt operator, Sobel operator, Canny operator, LOG operator, Roberts operator, and so on. Prewitt operator and Sobel operator, as first-order differential operators, are average filters and weighted average filters, respectively. LOG operator first uses a Gaussian function to smoothen images and then uses Laplace transform to process images. Sobel, Robert, and Prewitt methods are sensitive to noise and easily form a nonclosed edge area [11], and their edge detection effects are often unsatisfactory. Canny method can be applied to different occasions for the advantages of low missing detection rate, low error detection rate, and high accuracy of edge positioning.

However, the parameters of a (high and low threshold) Canny operator were artificially set and were not adaptive for

different images. A good edge detection effect is difficult to obtain when the threshold is manually set. Thus, selecting the appropriate threshold for image edge detection is very important. The method based on gradient magnitude histogram and intraclass variance minimization [12] is adopted to determine the adaptive threshold. This method does not require artificially setting the thresholds and can automatically obtain its own threshold according to different images, excluding the influence of human factors.

The number of support points is larger and affects the running time of the subsequent triangulation algorithm, as well as the entire running time of stereo matching. Under the premise of preserving the edge details of the reference image, we selected the support points every other line after edge detection. Experimental results also provide good verification with the support point images shown in Figure 1. We can see that all support points are distributed along the image edges uniformly.

3. Support Point Triangulation

The 2D triangulation of the reference image aims to represent the entire image with a set of triangular mesh. The disparity map is described as a set of triangular areas with same or similar disparity. The triangular meshes reflect the topological relation between a pixel and its neighboring pixels. Given the premise of preserving the disparity discontinuities and edge details, the triangulation in homogeneous areas should be large enough to reduce the matching ambiguity. In areas where the depth is homogeneous, the density of points should be small, and a higher number of points must exist near depth discontinuities to correctly preserve the object details.

Many 2D triangulation methods exist, and the representative method is Delaunay triangulation. The Delaunay triangular mesh is the most regularization triangular mesh [13]. The most commonly used Delaunay triangulation algorithms include insertion methods, incremental method, and divide and conquer method. Insertion method is simple and efficient and takes up less memory, but its time complexity is poor. Incremental triangulation method is not commonly used because of its low efficiency. Meanwhile, divide and conquer method has been shown to be the fastest Delaunay triangulation generation technique. Considering running time, we use the divide and conquer method to triangulate the initial set of support points. Triangulation results of higher-resolution images from the Middlebury website (Cones, Teddy, Aloe, and Venus) are shown in Figure 2.

4. Disparity Estimation

4.1. Initial Disparity Estimation. After the 2D triangulation of the reference image, each triangle is initially assumed to present uniform depth. The initial estimation step is assigning a unique depth value to each triangle.

For each triangle t_j in the reference image, we assume that $M_{j,r}(d)$ is the matching function of t_j with respect to image R at disparity d . The matching function is chosen based

FIGURE 1: Support points images detected by the Canny operator.

on the histogram of pixel gains and its ability to deal with illumination variations in an image. Considering a triangle t_j in the reference image, for each pixel x in t_j , the ratio can be calculated as

$$r_c^d(x) = \frac{I^c(x)}{I'^c(x+d)}, \quad (1)$$

where I' is the image that needs to be matched in the stereo pair, x is pixel coordinates, and d is the disparity value. The ratios between corresponding pixels at each color channel $c \in \{R, G, B\}$ are computed, and a histogram of the ratios considering all color channels is obtained. An ideal match at the correct disparity value should lead to similar ratios at all pixels and color channels.

If a match is good, the distribution of the histogram has few bins with large values and the rest are small, whereas a poor match has a more even distribution of the histogram. To find the large values of the histogram distribution, we use several methods such as image mean-square error estimate or entropy. We find that the following method is efficient and obtains good matching results. The matching function

$M_{jr}(d)$ for T_j with respect to image r at disparity d is given by

$$M_{jr}(d) = \frac{1}{3A_j} \max_l \left\{ \sum_{c \in \{R, G, B\}} \beta_{l-1}^c(d) + \beta_l^c(d) + \beta_{l+1}^c(d) \right\}, \quad (2)$$

where A_j is the area of triangle t_j and l is the index of histogram bin. We compute this histogram for each color channel using 20 bins $\beta_l^c(d)$ ranging from 0.7 to 1.1 and ignore the values outside this interval. We choose maximum sum of three adjacent bins as a good match because it is close to the total number of pixels. For any disparity value d , the value $M_{jr}(d)$ is within 0 and 1. A better matching is obtained when the value of $M_{jr}(d)$ is closer to 1.

The advantage of using a triangle as a matching unit is that each triangle shares edges with exactly three other triangles. This property allows a very straightforward implementation of the aggregation step, popular in pixel-based approaches but not common in region-based methods. In the aggregation step, the cost of adjacent triangle regions is also considered before selecting the best disparity.

FIGURE 2: Triangulation results of standard test images.

Denote Δ_j as the number of adjacent triangles of t_j , and the aggregated matching function M'_j is given by

$$M'_j(d) = M_j(d) + \sum_{i \in \Delta_j} \omega_{ij} M_i(d), \quad \omega_{ij} = e^{-B_{ij}^2 / (2\sigma^2)}, \quad (3)$$

where ω_{ij} is the measured function of the color similarity between triangles t_i and t_j based on the Bhattacharyya distance B_{ij} which is computed using the RGB values of both triangles. Parameter σ is used to control the attenuation degree of the exponential function ω_{ij} . The support weight of adjacent triangle is larger with increasing σ but possibly blurring the image edge. The weight was set experimentally to 0.16 as a compromise between image smoothing and edge blurring. The initial disparity value of each triangle t_j is given by $D_j = \text{argmax}(M'_j(d))$. We can obtain a piece-wise constant disparity map at the end of this step. Next, we will smoothen and refine the disparity in discontinuous areas.

4.2. Disparity Refinement. In the stage of disparity refinement, the disparity value of each vertex should be refined according to the similarity to its neighboring triangles so as to ensure that vertices related to similar triangles have similar depth values. Given that all vertices of each triangle

are potentially refined, the final result is a piece-wise linear representation of the depth map.

Considering that vertex V is shared by N triangles, D_j ($j = 1, \dots, N$) is the disparity values of this vertex in each triangle. We aim to find a refined disparity value for vertex V to ensure that the disparity value difference is reduced when the triangles are similar and kept unchanged when the triangles are dissimilar. The refinement step is formulated as a minimization problem, and the objective function is given by

$$E = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (d_i - d_j)^2 + \sum_{j=1}^N a_j (d_j - D_j)^2, \quad (4)$$

where w_{ij} is the similar weight between adjacent triangles i and j . a_j is the confidence value when we obtain the initial disparity D_j . The first term of (4) may be a regularization term, and it is minimized when all disparity values d_j are the same. The second term is minimized when $d_j = D_j$.

To ensure the accuracy and smoothing of the disparity value, the key issue is the selection of the weight w_{ij} and a_j . If a_j is large, the disparity accuracy increases, whereas the disparity map is smooth when w_{ij} is large. Therefore, a_j

FIGURE 3: Flowchart of the proposed algorithm.

should be selected based on the initial matching algorithm a_j given by

$$a_j = \frac{\sum M_{jR}}{CP_j}. \quad (5)$$

M_{jR} is the matching function similar to (2). $C = 3$ is the number of color channels. P_j is the number of pixels included in triangle t_j . The value of a_j represents the percentage of pixels inside the triangle that are present in the three largest contiguous bins considering three color channels. This value is close to one when good matches are obtained and decreases as the quality of the match gets worse. When triangles i and j are similar, the value of w_{ij} should be large so that the corresponding disparities of d_i and d_j are similar.

Subsequently, the key issue is the selection of w_{ij} . As in [14], we note that color similarity and proximity are two main concepts in classic Gestalt theory for visual grouping. The more similar the color of a pixel is, the larger its support weight is. Assuming that similarity and proximity can be regarded as independent events, w_{ij} is given by

$$w_{ij} = \frac{1}{A_i} \exp\left(\frac{\delta_{ij}}{\gamma_c} + \frac{d_{ij}}{\gamma_d}\right), \quad (6)$$

where δ_{ij} is the similarity distance defined as the Euclidean distance between the mean RGB values of t_i and t_j . The distance can be calculated by the following equation:

$$\delta_{ij} = \|I(i) - I(j)\|_2. \quad (7)$$

$\|\cdot\|_2$ is second-order norm form and $I(\cdot)$ represents the pixel gray level.

d_{ij} is the spatial distance, defined as the Euclidean distance between the centroids of t_i and t_j , given by

$$d_{ij} = \|i(x, y) - j(x', y')\|_2, \quad (8)$$

where $i(x, y)$ and $j(x', y')$ are the centroid of t_i and t_j , respectively. (x, y) and (x', y') are the centroid coordinates of triangles t_i and t_j , respectively.

Parameters γ_c and γ_d are thresholds that control the decay of the support weight. We have fixed $\gamma_c = 10$, $\gamma_d = 12$ based on previous experiments and obtained good matching results.

5. Experiment Results and Analysis

An overview of the proposed method is shown in Figure 3. First, we obtained the image edge information using the optimal edge operator Canny. The Delaunay triangulation method was then used to divide the entire image that needs to be matched into a series of two-dimensional triangles according to the edge points. Second, we formulated the matching model to accomplish initial disparity estimation according to the characteristics that each triangle shared edges with another to achieve cost aggregation. Finally, we refined the initial disparity according to the characteristics that triangles shared vertices and obtained the final disparity map.

To verify the effectiveness of the proposed approach, we tested the method on higher-resolution images from the Middlebury website. We present four images here, namely, Cones, Teddy, Aloe, and Venus, with different resolutions. We implement the method on a PC with a single CPU of 2.79 GHz and 3 G memory; the program language is C++. The calculated disparity maps were evaluated by measuring the percent of bad matching pixels. A comparison of results of disparity maps is shown in Figure 4. The black areas are occluded regions. A comparison of results of matching accuracy and running time of nonoccluded regions is shown in Table 1. As shown in Figure 4, the proposed method can obtain a clear outline disparity map. In occluded areas (Aloe left side) and disparity discontinuous areas (newspaper edge area of Venus) the proposed method can also obtain good matching results. To verify the effectiveness of the proposed algorithm, we compared our approach with ELAS [8] methods in terms of matching accuracy and running time. As shown in Table 1, the average running times of the proposed algorithm and ELAS were 1.043 s and 1.045 s, respectively. The average error matching results in nonoccluded areas of our method and ELAS were 6.75% and 7.83%, respectively. The running time of the proposed algorithm was close to that of the ELAS algorithm. However, the error matching ratio of our method was lower than that of the ELAS algorithm. Therefore, the running time and matching accuracy of the proposed method are able to meet the needs of practical applications.

We also used the stereo vision system (as shown in Figure 5(a)) based on D-H coordinates captured in the real-world images to verify the effectiveness of the proposed algorithm. The transmission agents of the stereo vision

TABLE 1: Comparison of results of different algorithms.

	Cones	Teddy	Aloe	Venus
Image resolution	900 × 750	900 × 750	1282 × 1110	950 × 750
Error matching ratio of ELAS (%)	5.7	15.5	6.6	3.52
Running time of ELAS (ms)	1021	1280	996	876
Error matching ratio of our method (%)	5.4	13.2	5.2	3.2
Running time of our method (ms)	1078	1224	1025	854

FIGURE 4: Compare results of disparity maps.

system are constituted by rotary joint and pitch joint. The angle range of rotary joint and pitch joint synchronously determines the scope of the scene, and the rotation velocity and acceleration of the joint determine the responsiveness of the stereo vision system. The rotary accuracy of the joints is more relevant to the positioning accuracy degree of the stereo vision system. Considering the aforementioned features, the angle ranges of the rotary joint and pitch joint are within -60° to 60° , whereas angular velocity and the positioning accuracy are $90^\circ/\text{s}$ and 0.8° , respectively. We used a pair of fixed focus

camera WA-922H to capture visual information. As shown in Figures 5(d) and 5(e), in the real-world scene, the proposed method can also obtain a good disparity map, which further verifies the effectiveness of the proposed method.

6. Conclusions

This paper presented a stereo matching algorithm based on Delaunay triangulation. Considering that edge detection has

(a) Stereo vision system

(b) Left image of stairs

(c) Right image of stairs

(d) DSI of ELAS

(e) DSI of our method

FIGURE 5: Experiment results of real environment.

an important influence on image recognition, an adaptive Canny operator was applied to detect image edge. The operator has the advantage of high accuracy edge positioning and it can effectively reduce the error matching ratio. The running time of stereo matching can be accelerated by using a triangle mesh as the matching unit and the gray information of the image to accomplish initial disparity estimation. The method was tested on Middlebury stereo pairs. The running time of

the proposed method is about 1s and the matching accuracy is 93% compared with that of ground truth map. Experimental results showed that the proposed method improved both the running time and the matching accuracy. In our future research, we will design an integrated vision system for parallel image processing and then apply it to the study of binocular vision navigation and path planning for six-legged robot.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) “Environment Modeling and Autonomous Motion Planning of Six-Legged Robot” (no. 61473104) and National Magnetic Confinement Fusion Science Program “Multi-Purpose Remote Handling System with Large-Scale Heavy Load Arm” (2012GB102004).

References

- [1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 7–42, 2002.
- [2] R. Brockers, “Cooperative stereo matching with color-based adaptive local support,” in *Computer Analysis of Images and Patterns*, vol. 5702 of *Lecture Notes in Computer Science*, pp. 1019–1027, Springer, Berlin, Germany, 2009.
- [3] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.
- [4] K. P. Murphy, Y. Weiss, and M. I. Jordan, “Loopy belief propagation for approximate inference: an empirical study,” in *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI '99)*, pp. 467–475, Morgan Kaufmann Publishers, Stockholm, Sweden, July–August 1999.
- [5] Z. W. Zhou, J. Z. Fan, G. Li et al., “Design and application of fast matching method based on support point expansion,” *Optics and Precision Engineering*, vol. 21, no. 1, pp. 207–216, 2013.
- [6] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, “On building an accurate stereo matching system on graphics hardware,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV '11)*, pp. 467–474, November 2011.
- [7] A. Geiger, M. Roser, and R. Urtasun, “Efficient large-scale stereo matching,” in *Computer Vision—ACCV 2010*, vol. 6492, pp. 25–38, Springer, Berlin, Germany, 2011.
- [8] C. Xia, Y. Yang, R. Ju, and G. Wu, “Effective local stereo matching by extended triangular interpolation,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '13)*, pp. 1–6, July 2013.
- [9] G. P. Fickel, C. R. Jung, T. Malzbender, R. Samadani, and B. Culbertson, “Stereo matching and view interpolation based on image domain triangulation,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3353–3365, 2013.
- [10] G. P. Fickel, C. R. Jung, R. Samadani, and T. Malzbender, “Stereo matching based on image triangulation for view synthesis,” in *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP '12)*, pp. 2733–2736, IEEE, Orlando, Fla, USA, September–October 2012.
- [11] R. Maini and H. Aggarwal, “Study and comparison of various image edge detection techniques,” *International Journal of Image Processing*, vol. 3, no. 1, pp. 1–11, 2009.
- [12] M. Li, J. H. Yan, G. Li, and J. Zhao, “Self-adaptive Canny operator edge detection technique,” *Journal of Harbin Engineering University*, vol. 28, no. 9, pp. 1002–1007, 2007.
- [13] P. Maur, “Delaunay triangulation in 3D,” Tech. Rep., Department of Computer Science and Engineering, University of West Bohemia, Pilsen, Czech Republic, 2002.
- [14] K.-J. Yoon and I. S. Kweon, “Adaptive support-weight approach for correspondence search,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.