2017 2nd International Conference on Image, Vision and Computing

Power Line Detect System Based on Stereo Vision and FPGA

Xiao Zhou, Xiaoliang Zheng, Kejun Ou
School of Mechanical and Electronic Engineering
Wuhan University of Technology
Wuhan, China
e-mail: zhouxiao@whut.edu.cn, jayash@163.com, okj1993@163.com

Abstract—**In recent years, Unmanned Aerial Vehicle (UAV) is widely used for power lines patrol. While it's more efficient than manually patrolling, the UAV may easily collide with power lines. Therefore, measures must be taken to detect the distance between UAV and power lines. This paper presents a real-time power line detect system based on binocular vision techniques, which contains a hardware platform based on FPGA to calculate real time disparity map. We adopt a robust stereo matching algorithm based on cross-based arbitrary shape support region method to calculate the depth image. As power lines in captured images are always too slender to be recognized in depth map, we propose a matching cost initial method based on morphology and mini-census transform to preserve the edge of the power lines in depth images. Our tests show that this new method can properly preserves the power lines of texture-less images in depth map.**

Keywords-power line detect; FPGA; real time; stereo vision; morphology; census transform

I. INTRODUCTION

Power lines patrol is a very important part for the security and stability of a power system. At present, this work is mainly completed manually, which is low efficiency and wasteful of labor. In recent years, Unmanned Aerial Vehicle (UAV) is widely used in power patrol. However, serious hazards exist when aerial vehicles operated at low altitude in cluttered environments without constrained or controlled conditions of different obstacles, lighting effects, weather and so on. Specially, power lines are one of the most formidable hazards [1]. Therefore, power line detection is one of the critical technology of UAV.

Visual image technology has more broad detection range, higher accuracy than sonar technology and cost less than radar technology. With the decline cost and the improved computing ability of mobile chip, visual image technology becomes the first choice to achieve obstacle avoiding. The stereo system can infer the depth information by means of triangulation using two independent images captured by stereo cameras. Reconstruction involves determining the three-dimensional scene using the disparity of points that correspond in both images, based on known camera geometry.

The main challenge of binocular vision is stereo matching, which is also known as depth estimation. In the past several years, although numerous techniques have been suggested for depth estimation, all the methods are divided into two categories: 1) local methods; 2) global methods [2].

Local methods always use a local support window to realize cost aggregation step, while global methods try to compute depth based on a global cost optimization. While global methods can achieve a better accuracy than local methods in low-texture regions and occluded regions, they always cost more time than local methods. As a result, almost all the stereo systems based on embedded system use local methods as the main algorithm. Despite this, due to the complex calculation, almost all the state-of-art local methods are hard to realize the purpose of real time processing. Recent years, dedicated hardware platforms, including FPGAs and GPUs begin to play a vital role in real time stereo matching.

For UAV patrol tasks, the actual scenarios are always complicated, especially the low-texture skies and slender line information are two main challenges when trying to obtain the depth map. So how to preserves the line information and removes the unrelated information in depth map remains a challenging problem. At the same time, the UAV also needs obtain real time depth maps for upper decision-making layers.

This paper aims to meet these two challenges by providing a real time stereo system based on FPGA. We adopt a robust stereo matching algorithm based on cross-based arbitrary shape support region method to calculate the depth image. As power lines in captured images are always too slender to be recognized in depth map, we propose a matching cost initial method based on morphology and mini-census transform to preserve the edge of the power lines in depth images. Our tests show that this new method can properly preserves the power lines of textureless images in depth map.

II. RELATED WORKS

As stated in section I, the most important and difficult part of our system is to obtain real-time stereo depth map. Most of the existing works use GPU and FPGA as hardware accelerate platform. Zhang *et al.* [3] used a DSP-BILDER based FPGA system to implement stereo matching, achieving 30 f/ps at resolution of 1396*1110 with 128 disparity levels. Wang *et al.* [4] transplant AD-Census algorithm into FPGA with semi-global optimization after cost aggregation, which is able to process high definition images of 1600*1200 pixels at 42 frames per second. Abdulkadir *et al.* [5] proposed a new algorithm named AWDE, which is very suitable to be implement on hardware, they implemented their algorithm on Xilinx Vitex-5 and can achieve 60 f/ps at resolution of 1024*768 with 128 disparity levels. Zhang *et al.* [6] combine the mini-census transform and cross-based cost aggregation in their structure, which achieves 60 frames/s at 1024×768 pixel stereo images. Compared to GPU, though FPGA have little advantage over GPU on complex upper serial logic, it can handle more parallel structures and have lower power consumption, which is very attractive for UAV. We adopt FPGA as our hardware platform.

III. STEREO MATHING ALGORITHIMS

In this section, we describe in detail the proposed method of the stereo matching system. As stated in [7], the stereo system contains four main steps: cost initialization, cost aggregation, cost selection and cost refinement. Fig. 1 shows the block diagram of the proposed algorithm.

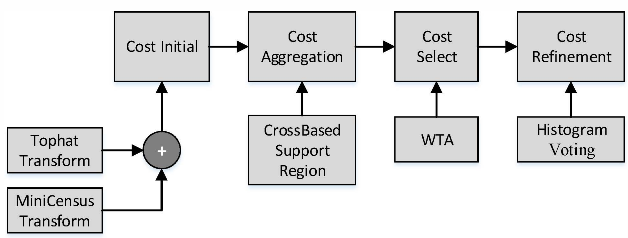

Figure 1. Block diagram of the proposed algorithm.

A. Support Region Constuct

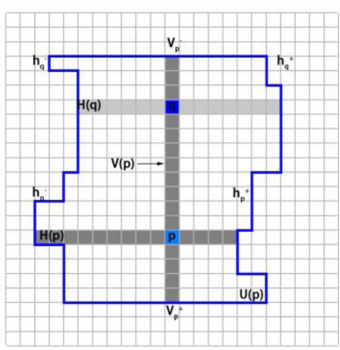

Figure 2. Cross-based support region, the region inside the blue envelope is the final adapted supported region.

For support region construction, we choose the adapt cross-based support region proposed in [8]. Although SAD matching method is much easier to implement and cost less time, in our tests, it brings more depth errors and the line information is easy to be drowned in other unrelated background. Cross-based support region method is proved to perform much better than a fixed support window method and is also friendly to hardware using image integration ideology [9]. The cross-based method builds support region according to the color similarity of a specific maximum support window and gets an arbitrary support area with a least 3*3 support window as shown in Fig. 2.

B. Cost Initialization Based on Mini-census and Morphology Filtering

Cost initialization step computes the initial matching cost volume for each disparity. Common cost measures include absolute differences (AD), Birchfield and Tomasi's sampling-insensitive measure (BT), gradient-based measures and non-parametric transforms such as rank and census [10].

Census transform [11] is independent to the luminance difference of the of two stereo cameras and performs better in real world stereo images. So, we choose census transform as the first component of the cost initialization.

MiniCensus transform is a reduced Census transform proposed in [12]. It is also hardware friendly and can significantly reduce the memory utilization. Compared to 5*5 standard Census transform which needs 24 bits for a pixel cost, MiniCensus transform only needs 6 bits. The MiniCensus transform translates the input pixel window information into a compare-vector, as shown in Fig. 3. The initial cost is defined as the Hamming distance between two compare-vectors input.

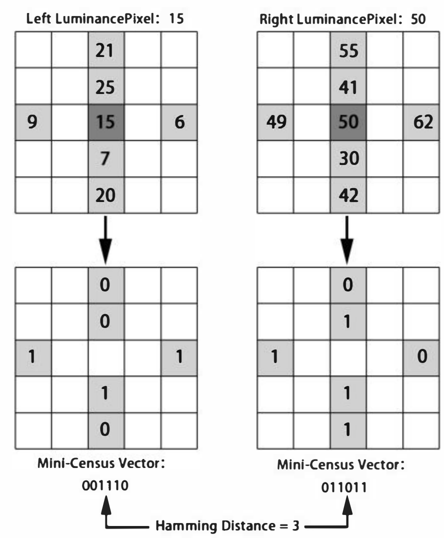

Figure 3. Mini-census transform and Hamming distance.

As we state in section I, power lines in captured images are always too slender to be recognized in depth map. Some image enhancement work should be done before stereo matching. In our scene, the background information of the target images captured by UAV when patrolling is basically sky or trees, which should be weaken when operating enhancement.

In this paper, we propose morphology filtering to accomplish this task. The Tophat transform [13], which is widely used in small target detection, can preserve edge information of the input images. We utilize this information and add the census transform result as the final initial cost.

As the power lines in stereo images are in vertical direction to calculate the depth information, a linear shape structure element will be select in Tophat transform.

The final initial cost is calculated as follows:

$$e(x,y,d) = ham(f_C^L(x,y), f_C^R(x-d,y)) + ham(g_C^L(x,y), g_C^R(x-d,y)) \quad (1)$$

The ham stands for standard hamming distance, and the function f stands for the MiniCensus transform, the function g stands for the Tophat transform.

C. Cost Aggregation

To accelerate cost aggregation step, the author in [8] using a technique naming “OII” (Orthogonal Integral Images). This technique decomposes cost aggregation step into two orthogonal 1-D aggregations. The cost will be aggregated in horizontal direction and then aggregated in vertical direction by methods of integration respectively. The vertical first method is proved to be efficient in software, but will cost considerable resources in FPGA. A one direction fixed support arm construction is proposed in [x] to reduce the design complexity for FPGA. YI at [9] proposed a vertical aggregation first method which is proved to have the ability to reduce considerable Block RAM resources. We adopt this method as the cost aggregation method.

The vertical aggregation first method firstly computes the left and right arm span of the current pixel, then computes the vertical arms in each column as shown in Fig. 4.

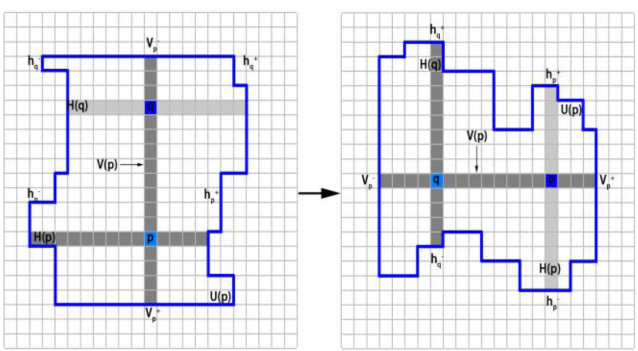

Figure 4. Left : horizon first aggregation, right: vertical first aggregation.

D. Cost Selection

Like most local stereo matching method, we take a Winner Takes All (WTA) method to get the preliminary estimated value. WTA method takes the minimum value of the aggregate result as the estimated value, which is:

$$d_p^0 = \min_d E_d(p), \ d \in [0, d_{\max}] \quad (2)$$

E. Cost Refinement

The preliminary estimated value can be further refined using a local high-confidence voting scheme proposed in [14]. The final disparity of the pixel p is decided as:

$$d_p^* = \max_d \alpha_p(d), \ d \in [0, d_{\max}] \quad (3)$$

IV. HARDWARE IMPLEMENTATION

A. System Overview

The aim of our work is to detect power lines in real time. A binocular vision system must be provided in our system. Considering that the system use FPGA to calculate depth image, the binocular cameras should have an interface that is easy to control. We select camera-link cameras in our system. The cameras connect to the FPGA by LVDS interface with a USART configuration interface. Offline calibration should be done with the help of MATLAB calibration toolbox to compensate the distortion of the cameras. Then the corrected image streams are connected to the stereo modules to get a real-time depth image.

Considering that some upper logic and decision-making layers should be done according to the depth map, a better solution is using a DSP to complete these tasks. A dual port ram is used for the DSP and FPGA to communicate with each other. Also, the DPRAM stores the final result and send it to the video output controller to generate output video stream. The overall structure of the system is as follows:

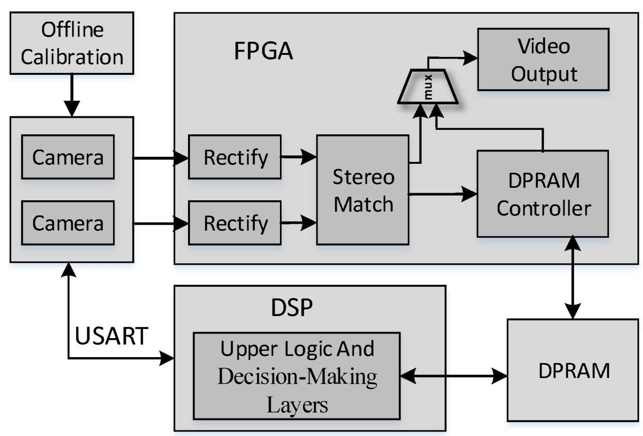

Figure 5. Overall structure of the system.

B. FPGA Implementation

Firstly, FPGA buffers the LVDS data using a LVDS buffer; then the video data is pushed into a cross clock FIFO to sync to local clock domain. After that, the video streams are calibrated by calibration circuits to compensate for camera distortion.

The next step is stereo matching. The rectified images will be distributed into three roads parallelly. A MiniCensus transform, Tophat transform and adapted cross region build circuit will be performed. After that, calculate the initial cost by adding the result of Tophat transform and Hamming distance for the MiniCensus results of the stereo images. Then Cost aggregation step is done by doing vertical aggregation firstly. The final step of the algorithm is WTA module and refinement module. The overflow of the FPGA implementation is shown in Fig. 6.

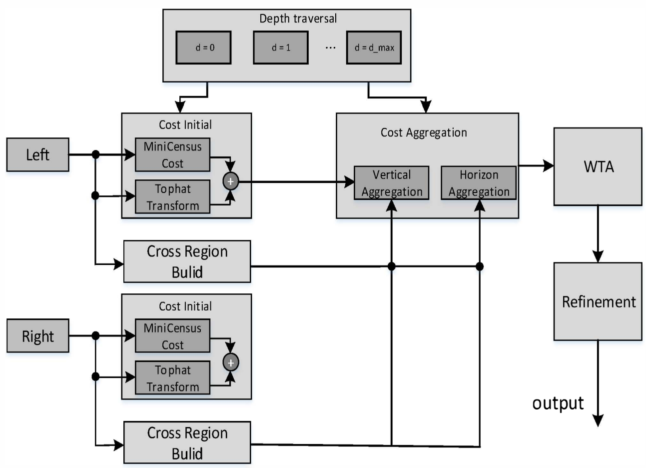

Figure 6. FPGA implementation overflow.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we evaluated the performance of the proposed stereo matching method using the real world stereo images. Fig. 7 and Fig. 8 presents the test results for the real-world images.

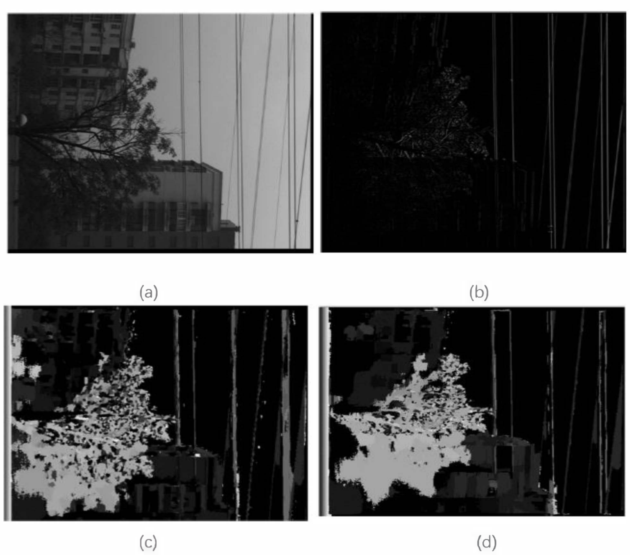

Figure 7. Test result of real world stereo image.

In Fig. 7, (a) is the one of the input stereo image; (b) is the Tophat transfer result of (a) with a process radius of 3; (c) is the result of only cross-based support region method; (d) is the final result of the method this paper proposed. Notice that the depth map with Tophat transform result can bahave better with slender lines. The main reason is that we add the edge information when we try to find the same area in the target picture, and it gives us more information to do stereo matching.

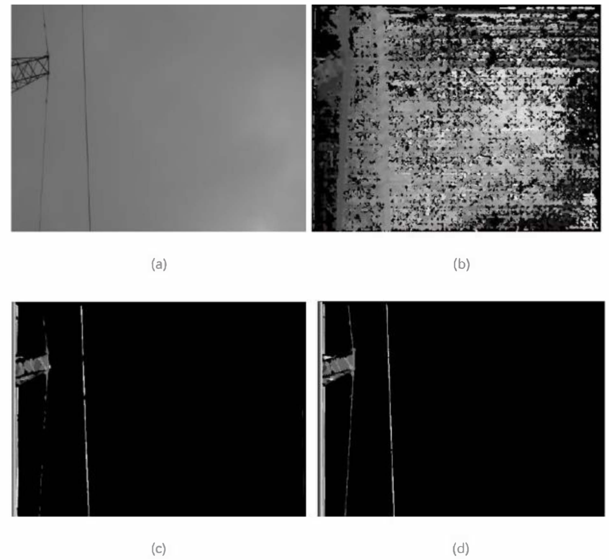

Figure 8. Test result of another real-world stereo image.

Fig. 8 is another test result of the real-world stereo images, (a) is the one of the input stereo image; (b) is the OpenCV result with SAD method and SGBM [15]; (c) is the result of only cross-based support region method; (d) is the final result of the method this paper propsed. Compared to (c) and (d), (b) fails to recognize the power lines in the input images, when the power line is very slender, the proportion of the power lines is very small in the fixed SAD support window, while the cross-based support region is very small and should have the same width with the power lines in (a). At the same time, the power line in the left side on depth (d) is more complete than (c), which also proves that the method in this paper can preserve the slender power lines in depth map better.

REFERENCES

[1] Song B, Li X. Power line detection from optical images[J]. Neurocomputing, 2014, 129(129):350–361.

[2] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense twoframe stereo correspondence algorithms,” Int. J. Comput. Vis., vol. 47, nos. 1–3, pp. 7–42, 2002.

[3] Zhang X, Chen Z W. A FPGA Stereo Matching Algorithm Modeled By DSP Builder[J]. Journal of Computers, 2014, 9(10).

[4] Wang, Wenqiang, *et al.* "Real-time high-quality stereo vision system in FPGA." International Conference on Field-Programmable Technology IEEE, 2013:358-361.

[5] Akin A, Baz I, Schmid A, *et al.* Dynamically adaptive real-time disparity estimation hardware using iterative refinement[J]. Integration the Vlsi Journal, 2013, 47(3):365-376.

[6] L. Zhang, K. Zhang, T. S. Chang, G. Lafruit, G. K. Kuzmanov, and D. Verkest, “Real-time high-definition stereo matching on FPGA,” in Proc. 19th ACM/SIGDA Int. Symp. Field Program. Gate Arrays, 2011, pp. 55–64.

[7] H. Hirschm¨uller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. IEEE TPAMI, 31(9):1582–1599, 2009.

[8] Zhang K, Lu J, Lafruit G. Cross-based local stereo matching using orthogonal integral images[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2009, 19(7):1073-1079.

[9] Wang W, Yan J, Xu N, *et al.* Real-time high-quality stereo vision system in FPGA[M], IEEE Transactions on Circuits and Systems for Video Technology. 2013:358-361.

[10] Mei, Xing, *et al.* "On building an accurate stereo matching system on graphics hardware." IEEE International Conference on Computer Vision Workshops, ICCV 2011 Workshops, Barcelona, Spain, November 2011:467-474.

[11] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In Proc. ECCV, pages 151–158, 1994.

[12] Chang Y C, Tsai T H, Hsu B H, *et al.* Algorithm and Architecture of Disparity Estimation With Mini-Census Adaptive Support Weight[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2010, 20(6):792-805.

[13] Jain A K. Fundamentals of digital image processing :[M]. Wiley-Blackwell, 2011.

[14] J. Lu, G. Lafruit, and F. Catthoor, "Anisotropic local high-confidence voting for accurate stereo correspondence," in Proc. SPIE-IS&T Electron. Imaging, vol. 6812, Jan. 2008, pp. 605822-1–605822-10.

[15] Hirschmuller H. Stereo Processing by Semiglobal Matching and Mutual Information[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 30(2):328-41.