

A compact 3D camera suited for mobile and embedded vision applications

Stefano Mattoccia, Ilario Marchio, Marco Casadio
Department of Computer Science and Engineering
School of Engineering and Architecture, University of Bologna
Bologna, Italy
stefano.mattoccia@unibo.it

Abstract—Recent years have seen the widespread diffusion of 3D sensors, mainly based on active technologies such as structured light and Time-of-Flight, enabling the development of very interesting 3D vision applications. This paper describes a compact 3D camera based on passive stereo vision technology suited for mobile/embedded vision applications. Our 3D camera is very compact, the overall area of the processing unit is smaller than a business card, lightweight, it weighs less than 100 g including lenses, has a reduced power consumption, about 2 Watt processing stereo pairs at 30+ fps, and can be easily configured with different baselines and processing units according to specific application requirements. The overall design is mapped on a low cost FPGA, making the hardware design easily portable to other reconfigurable devices, and allows us to obtain in real-time accurate and dense depth maps according to state-of-the-art stereo vision algorithms.

Keywords—3D; stereo vision; FPGA; RGBD; depth sensor

I. INTRODUCTION AND RELATED WORK

In recent years the widespread diffusion of accurate 3D sensors has greatly increased the interest for 3D vision leading to very interesting applications. Most of these RGBD (image + depth) sensors rely on active technologies that, by perturbing the environment according to different approaches, enable depth sensing. Well known examples of 3D sensors based on active technologies are the Kinect and time-of-flight sensors. A different well known technology, purely based on standard imaging devices, is passive stereo vision. This technology infers depth by triangulating corresponding points projected from the sensed scene to, at least, two imaging sensors with known relative position. Although pattern projection can be used to improve its effectiveness especially in untextured regions, compared to other 3D sensing devices, stereo vision is a passive technology and this fact has some notable advantages wrt active technologies (e.g., enabling simultaneous sensing of the same area with different sensors, suited to indoor and outdoor environment). Despite these positive aspects, stereo vision is computationally demanding and for this reason has been considered for a long time not suited to mobile/embedded applications due to the typical high energy requirements (and size) of the computing platforms, such as high end CPUs or GPUs, that allow to implement real-time stereo vision systems.

However, modern reconfigurable computing architectures and recent algorithms proposed in literature enable to design

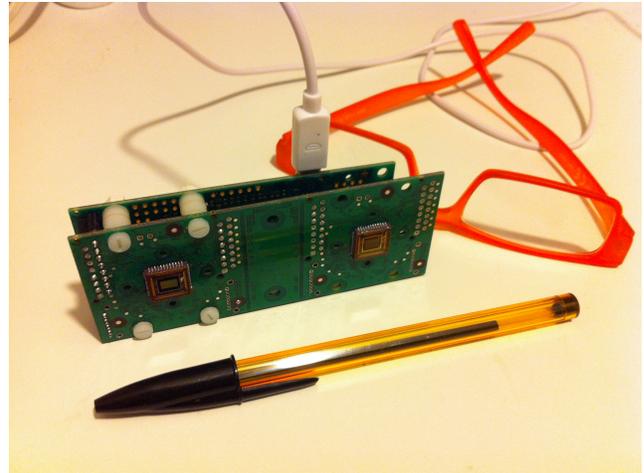


Figure 1. Stereo camera with onboard FPGA processing. In this setup the baseline is about 6 cm.

very compact, lightweight and accurate 3D sensors based on stereo vision that fit with the constrained energy requirements of typical embedded and mobile vision applications. Our proposal, compared to most existing stereo cameras with FPGA processing, allows us to obtain very accurate results with a processing pipeline self contained into a low cost FPGA as shown in Figure 1. An exhaustive review and evaluation of significant stereo vision algorithms is available in [3]. A more recent review of this research area, focused on computing architectures suited for real-time stereo vision systems, was proposed in [4]. Finally, a review of stereo vision algorithms suited to constrained FPGA architectures can be found in [2].

Since our target computing platform has constrained resources, especially in terms of available memory, *global algorithms* [3] seem not well suited for our purposes. However, in this class falls also algorithms based on simplified energy minimization methodologies that enforce a smoothness constraint on 1D domains. Algorithms that belong to this class are typically based on *dynamic programming* or *scanline optimization* [3]. In particular, the effective SGM algorithm [1] based on multiple independent scanline optimization has become very popular in recent years.

In the remainder we provide a brief description of our

3D camera and of its processing pipeline, entirely mapped on a low cost FPGA, based on a memory efficient version of the SGM algorithm. In the configuration reported in this paper, shown in Figure 1, the FPGA is a Xilinx Spartan 6. In this setup the camera can process two synchronized video streams provided by two (color or monochrome) global shutter imaging sensors with a maximum resolution of 752×480 pixels and a maximum frame rate of 60 fps.

II. OVERVIEW OF THE FPGA-BASED STEREO CAMERA

Our design strategy was aimed at obtaining a very compact, lightweight and energy efficient RGBD sensor based on passive stereo vision technology. For this purpose, after a deep analysis of advantages and drawbacks, we decided to follow a quite radical design strategy aimed at minimizing the overall hardware complexity. This choice leads us to define a memoryless computing architecture essentially based on a low cost FPGA and a communication controller that, in the specific case of the camera depicted in Figure 1, is compliant to the USB 2.0 standard. Our choice has several positive implications in terms of bill of material, portability, size, weight and power consumption. On the other hand this design strategy poses significant constraints concerning the computational structure of algorithms that can be actually implemented on this architecture. The main constraint is concerned with external memory (not available in our design), allowing the implementation of algorithms with a computational structure compatible with a *stream processing* approach. This computing strategy consists in minimizing the input buffering requirements for each module by processing incoming pixels as soon as they are made available from a previous module in the processing pipeline. Of course, this also means that the data produced by a module should be consumed by the successive module as soon as possible in order to minimize buffering requirements.

The overall processing pipeline consists of the following modules: image filtering and rectification, stereo correspondence based on a memory efficient version of the SGM algorithm, outliers detection and subpixel interpolation. The design also includes glue logic required to connect the imaging sensors to the FPGA and to send images and disparity maps to the external (i.e., outside the FPGA) USB controller. Appropriate algorithmic strategies, not reported in this paper due to space limitations, allowed us to implement the whole processing pipeline into the Spartan 6 FPGA without any external memory device. The overall power consumption, processing stereo pairs at more than 30 Hz and at 640×480 resolution, is about 2 Watt on a Xilinx Spartan 6 Model 75. Thank to its power requirement, the camera is self powered by the USB data cable.

Figure 2, reports experimental results concerned with the autonomous navigation of a small battery powered rover in a challenging indoor scenario. In this application, the disparity map computed in real-time by the stereo camera

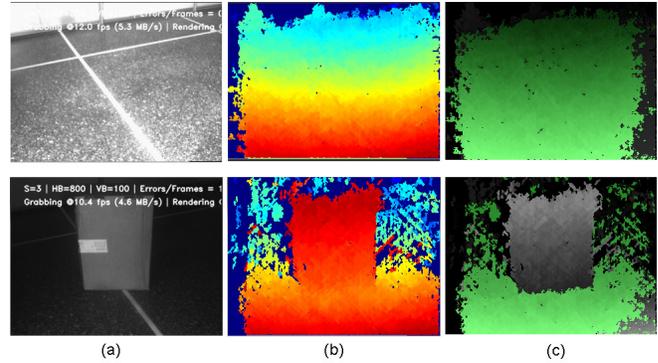


Figure 2. Experimental results concerned with autonomous navigation of a small battery powered rover: (a) Reference image, (b) Disparity map computed by the stereo camera shown in Figure 1, (c) Detected ground plane (green) and potential obstacles.

is processed to determine, according to a RANSAC based approach implemented on an embedded ARM board Odroid-U3 running Linux, the ground plane and potential obstacles in front of the rover. The whole navigation system (i.e., the proposed 3D camera and the Odroid-U3 embedded ARM board) used for these experimental results weights less than 150 g.

Additional details, experimental results and videos, not reported here due to space limitations, can be found at this link¹.

III. CONCLUSIONS

In this paper we have outlined an optimized computing architecture, based on a low cost FPGA, and the processing pipeline for a stereo camera suited for embedded/mobile vision applications. The proposed stereo camera is compact (the overall area of the processing unit is smaller than a business card), lightweight (less than 100 g with M12 lenses and holders) and allows us to obtain in real-time dense and accurate disparity maps based on a state-of-the-art stereo algorithms. The proposed optimized hardware design and processing pipeline enables a small power consumption (about 2 Watt processing stereo pairs at 640×480 and 30+ fps). These facts make the proposed 3D sensing device suited for applications characterized by strong energy constraints.

REFERENCES

- [1] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):328–341, 2008.
- [2] S. Mattocchia. Stereo vision algorithms on fpgas. In *9th IEEE Embedded Vision Workshop, CVPR 2013*, 2013.
- [3] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal Computer Vision*, 47(1/2/3):7–42, 2002.
- [4] B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald. Review of stereo vision algorithms and their suitability for resource-limited systems. *Journal of Real-Time Image Processing*, 2013.

¹<http://www.vision.deis.unibo.it/smatt>