

A NOVEL HARDWARE ARCHITECTURE FOR AN EMBEDDED STEREO VISION SENSOR

AMBROSCH, Kristian; HUMENBERGER, Martin & KUBINGER, Wilfried

Abstract: *This Paper describes our current work on a novel hardware architecture for an embedded stereo vision sensor that is suitable for robotics applications. The architecture is based on two Digital Signal Processors (DSPs) for the pre- and post-processing as well as a Field Programmable Gate Array (FPGA) for stereo matching.*

Key words: *Stereo Vision, Embedded Systems, DSP, FPGA, Robotics.*

1. INTRODUCTION

In the field of robotic applications there is a high demand for sensors that can create a three dimensional (3D) view of the surrounding environment. Today laser range scanners are being used for this kind of application, which have a limited operating range when operating at high frame rates.

Another method to create a 3D depth map is to use stereo vision. Stereo vision algorithms compare the images of two cameras and find the displacement of the objects in those images using area or feature based matching. The displacement of the objects is also called disparity and measured in pixels. Using the disparity and the a priori knowledge of the distance between the cameras, the 3D depth map can be computed using triangulation.

Stereo vision algorithms are computationally extremely expensive and therefore currently not widely used for robotic applications because of the resulting high system costs. Therefore we propose a novel hardware architecture that is based on the use of a Field Programmable Gate Array (FPGA) for the stereo matching. This enables the design of an embedded stereo vision sensor that can be produced with minimum costs in series production, when taking into account that the FPGA design can be used for the production of an Application Specific Integrated Circuit (ASIC) which has very low costs per piece when produced in high numbers.

2. RELATED WORK

Various examples of stereo vision algorithms using FPGAs exist in the literature.

Some of these implementations use more than one FPGA like (Corke et al, 1997), (Miyajima et al., 2003) and (Niitsuma et al., 2005). Since the assembly costs raise with the use of multiple chips, an architecture using more than one FPGA would be far too expensive.

(Woodfill et al, 2006) have developed an embedded stereo vision sensor called G2 Vision System that is based on an ASIC for the stereo matching, an FPGA for the pre-processing as well as an Analog Devices Blackfin Digital Signal Processor and a PowerPC. The system is capable to calculate a depth map from two 512x480 images at 60fps but only with a maximum disparity of 52 pixels. Thus the systems suitability for robotic applications is very limited, because here objects at a very small distance have to be detected which enforces much higher disparities.

Other works that are using only one FPGA but have a too small disparity are (Yi et al., 2004), (Niitsuma et al, 2004) and (Lee et al., 2005).

3. ARCHITECTURE

3.1 Cameras

A very important decision is whether to take analog or digital cameras. Analog cameras have the advantage of a very robust data transmission over long distances. Also a lot of multimedia Digital Signal Processors (DSPs) already have analog video capturing implemented on the silicon.

But analog cameras usually provide interlaced pictures, which not only increases the latency of the system (one also has to wait for the second frame to create a non interlaced picture) but also creates image blur when objects are moving. This is because the object has moved from one frame to another, but this information is only captured by 50% of the image lines. Of course also non interlaced cameras exist, but digital cameras are about to replace the analog ones. Thus only the most common types of analog cameras will be available in the future, which will definitely be the interlaced ones. For digital cameras three very common interfaces exist. The first one is USB, which takes a lot of computational resources for the handling of the communication. Therefore it is not advisable to use it in a system with limited resources. The other ones are IEEE1394 according to the IIDC standard and CameraLink. Both are suitable for our application and we recommend using a system that can switch between both standards (e.g. by having a daughter interface board).

3.2 Pre-Processing

The task of the pre-processing stage is to compute images that are rectified to fulfill the epipolar geometry (Zhang, 1998). This is a task that has to be calibrated for each single stereo sensor. Thus it is not advisable to implement it in hardware, because a reconfiguration would always need a complete re-synthesis of the design. A DSP based implementation is much more advisable, because it is much more flexible.

3.3 Stereo Vision Matching

To evaluate the best way to implement the stereo vision matching, we implemented a software as well as a hardware solution for a small stereo vision algorithm.

The algorithm was implemented in VHDL and synthesized for an Altera EP2S60 using Quartus II. The implemented algorithm was an Sum of Absolute Differences (SAD) with a block size of 3x3 pixels using 320x240 input images in 8 bit grayscale. The designed used 19520 Logic Elements (LEs) which is about 30% of the chip surface. The computation time was 2.3ms. The software implementation was run on an Intel Pentium 4 with 3 GHz clock frequency and 1 GB memory. For the optimization of the software we used Intel's Open Source Computer Vision Library. In this case the computation time was 391ms, which is 166 times slower than the FPGA implementation. This example shows that an FPGA solution is much more suitable for stereo matching than a purely processor based system.

3.4 Post-Processing

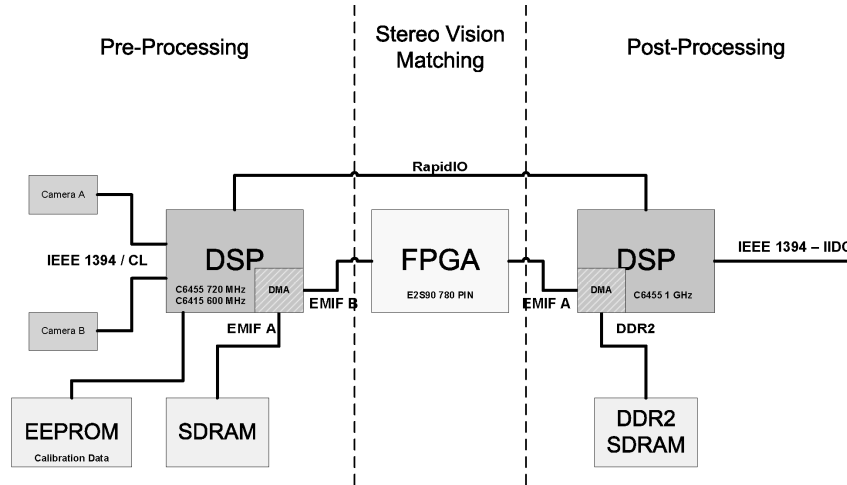


Fig.1: Hardware Architecture

The task of the post-processing stage is to refine the output of the stereo matching. Since the post-processing is very frame dependent it cannot easily be implemented in hardware. This is the reason why we didn't evaluate the use of an FPGA for this task and took a DSP.

3.5 Communication

For the communication between DSPs and the FPGA, we evaluated five different interfaces.

DSPs from Texas Instruments (TI) support serial RapidIO. This communication interface offers speeds up to 1 Gbit/s. It gets along without a separate clock line, because the interfaces synchronize via the communication line. The main problem is that an Intellectual Property (IP) core from Altera is extremely expensive. Thus it is not advisable to use this interface in our hardware architecture. HyperTransport is a parallel high speed interface which is easily implementable in hardware. Furthermore there are public licence IP cores available from OpenCores. But TI doesn't support HyperTransport and therefore it is not suitable for our purpose. PCI is a standard interface for connecting hardware. It offers a data transfer rate of 1 Gbit/s while operating at 33 MHz and there are public licence IP cores available at OpenCores. But it needs 47 bus lines and is only rarely supported by DSPs. Thus we decided not to take PCI. Another method for inter processor communication is true dual ported SRAM. Here chips can exchange data without the need for synchronization, but SRAM is very expensive and the power consumption is definitely too high for our purpose.

DSPs from Texas Instruments have a memory interface called EMIF. This memory interface can be used to access registers on the FPGA offering another method of transferring data between DSPs and FPGAs. Using the Direct Memory Access (DMA) controller to transfer the data, it is possible to transfer the data from the FPGA to the DSPs main memory without interrupting the DSPs core. This is the main reason why we decided to take this interface for our architecture.

3.6 Results

Figure 1 shows our hardware concept. It contains two digital video cameras connected to a fixed point DSP from TI via IEEE1394a or CameraLink. This DSP is designated to pre-process the image data. Then the data is fed into a Stratix II EP2S90 FPGA using the DSPs EMIF port. In the EP2S90 the stereo matching is performed. Afterwards the disparity map is read by the post-processing DSP using its EMIF port and transferred into its main memory. When the post-processing is finished, the final depth map is sent to the receiver via the IEEE1394a port. Furthermore the post-processing DSP can detect deviations in the vertical alignment of the camera system and send new calibration data to the pre-processing DSP using their RapidIO connection. To store the calibration data, the pre-processing DSP is connected to an EEPROM.

4. CONCLUSION

We proposed a novel hardware architecture that is based on two DSPs for pre- and post-processing as well as on an FPGA for stereo matching. Furthermore, we compared the implementations of a stereo matching algorithm on an FPGA and a PC with the result that the FPGA is 166 times faster. This shows that FPGAs are an excellent choice to guarantee real-time behavior on an embedded stereo vision sensor.

5. ACKNOWLEDGEMENTS

This research has been supported by the European Union project ROBOTS@HOME under grant FP6-2006-IST-6-045350.

6. REFERENCES

- Corce, B. & Dunn, P. (1997). Real-Time Stereopsis Using FPGAs, *Proceedings of the IEEE Conference on Speech and Image Technologies for Computing and Telecommunications*.
- Lee, S.; Yi, J. & Kim, J. (2005). Real-Time Stereo Vision on a Reconfigurable System, *Lecture Notes in Computer Science*, Vol. 3553, pp 299-307.
- Miyajima, Y. & Maruyama, T. (2003). A Real-Time Stereo Vision System with FPGA, *Lecture Notes in Computer Science*, Vol. 2778, pp 448-457.
- Niitsuma, H. and Maruyama, T. (2004). Real-Time Detection of Moving Objects, *Lecture Notes in Computer Science*, Vol. 3203, pp 1155-1157.
- Niitsuma, H. and Maruyama, T. (2005). High Speed Computation of the Optical Flow, *Lecture Notes in Computer Science*, Vol. 3617, pp 287-295.
- Woodfill, J.; Gordon, G.; Jurasek, D.; Brown, Te. & Buck, R. (2006), The Tyzx DeepSea G2 Vision System, A Taskable, Embedded Stereo Camera, *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition - Workshop on Embedded Computer Vision*.
- Yi, J.; Kim, J.; Li, L.; Morris, J.; Lee, G. & Leclercq, P. (2004), Real-Time Three Dimensional Vision, *Lecture Notes in Computer Science*, Vol. 3189, pp 309-320.
- Zhang Z. (1998), Determining the Epipolar Geometry and its Uncertainty: A Review, *International Journal of Computer Vision*, Vol. 27, No. 2, pp 161-195.