

Implementation of a multiview passive-stereo-imaging system with SoC technology

Richard Fütterer*^a, Mathias Schellhorn^a, Gunther Notni^a

^aTechnische Universität Ilmenau, Gustav-Kirchhoff-Platz 2, 98693 Ilmenau, DE 03677693930;

ABSTRACT

Passive-stereo-imaging as a part of three-dimensional image processing can be applied for many use cases where a depth information is necessary. This is often the case when classical two dimensional image processing get to its borders. Passive-stereo needs at minimum two images of the same object of interest. An initial one-time system calibration is absolutely essential. The usage of more than two sensors offer more system stability and reduce hidden image elements. A variant of semi-global-matching algorithm is used to find the matching points in the image pair and is implemented in the FPGA. The lateral shift of two matching points is called disparity. All disparity values together result in a disparity image which is the basis for a point cloud.

By using embedded technologies, a compact and portable acquisition system could be realized. Programmable system-on-chip (PSoC) combining FPGA and ARM computing power in one chip design. This technology is an ideal solution to so acquire the images from the image sensor and calculate the disparity images.

Keywords: passive-stereo imaging, three dimensional image processing, portable 3D system, semi global matching, embedded FPGA ARM system, programmable system on a chip, SoC, PSoc

1. INTRODUCTION

Passive-stereo image processing is an important technology for recognition and measuring applications. Passive-stereo is a part of three dimensional image processing and needs no additional pattern or laser line projection. Beside classical two dimensional image processing, nearly for each pixel a depth information is given. Passive-stereo works under hard illumination conditions, where other methods deliver no results, i.e. because of direct sun illumination. The principle is very similar to the human visual perception and requires (at minimum) two image sensors. A passive-stereo system can be implemented in an embedded system, i.e. with SoC technology as shown here. This paper shows the implementation of a passive-stereo vision system in a Xilinx-Zynq-7000 7030 SoC system. The advantages of this sensor principle is the small size, the low power consumption (about 6W) and the single-shot principle, that means there is no image sequence necessary.

2. BASICS OF PASSIVE-STEREO IMAGING

To calculate passive-stereo data, there are two images acquired at the same time necessary. In general, there are no special requirements relating to the orientation of the image sensors. In this case, a parallel orientation is useful. Nearly parallel aligned image sensors have no significant angle between the optical axis and the projected images are rectangular and not trapezoid. Rectangular projected images need a less complex rectification process; this is beneficial for embedded systems. When using a standard epipolar (Figure 1) geometry, an object point p is mapped to both image sensors. Because of mechanical tolerances and lens distortions, a rectification of both images is necessary. When we get the rectified images, the point p is mapped in both images to the same image line numbers in v on different columns in u . The difference between the coordinates p_L and p_R is called disparity. Object distance and disparity value are in inverse proportion, so object points in far distance will result in low disparity value. This causes a low depth resolution for far distant object points and a higher resolution for object points near the lenses.

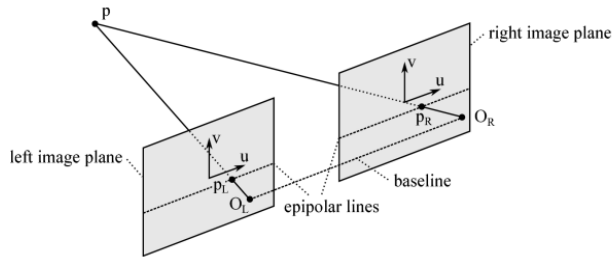


Figure 1: Epipolar geometry

Systems depth resolution and minimum working distance in optical axis direction is dependent to some elements. The optical elements are pixel size and lens focal length. The mechanical component is the base length between the two lenses. The maximum disparity in pixel which is calculated by the SoC limits the minimum working distance to the point p . Because of parallel aligned lenses, the images contain an area of rows on the image borders where no disparity could be found. In this area on outer left and right side are objects only for one image sensor visible. On the second sensor this object isn't visible. With increasing maximum disparity this area gets bigger. So a compromise between maximum disparity value and image useable image area has to be found.

Passive-stereo imaging gets to its borders, when surfaces with no or less texture are in the images. Such surfaces result in image areas with nearly the same grey value. Here is it difficult to get disparity points.

3. IMPLEMENTATION ON THE SOC

The calculation of the disparity points is done by a variant of a semi-global-matching algorithm. The SGM algorithm works line by line, pixel by pixel through the whole image with the goal, to label each image point of the left image with a corresponding point of the right image. The original SGM (as introduced by Hirschmüller (2005)) algorithm computes along several paths in two dimensions through the image, symmetrically from all directions up to a certain maximum disparity. The cost trough each path is stored and compounded by image similarity and smoothness (penalizing small disparity steps). The disparity with the lowest cost is chosen. The variant of SGM used here differs from original SGM in the number of paths. Here only paths in horizontal and vertical downwards directions are calculated. The SGM algorithm is being contained in a IP core which handles the whole image processing steps till the calculation of the disparity map. The following steps are done by the IP core:

- Reading images from Memory and image filtering (i.e. speckle, illumination)
- Image rectification and pre-processing
- Semi-global-matching, cost volume post-processing and disparity map post-processing [3]
- Buffering and writing the disparity map and rectified images
- Post processing and write the results to memory

For rectifying the images, first an initial calibration of the whole system is necessary. This calibration procedure contains additionally the correction of lens distortion and deliver a rectification map which is stored in the system DDR ram and is accessible by the stereo core. In calibration process several images with a calibration pattern are acquired and saved. An offline calculation on PC deliver a rectification map, which is stored in system flash or SD-Card. The camera parameter (Q-matrix) is stored in a separate file. While booting the system, all relevant data is transferred to system RAM and is accessible for FPGA-fabric and ARM device.

The whole image processing is spitted into software parts, these running in the ARM cores of the SoC. The hardware parts are implemented in the FPGA fabric. Figure 2 show the necessary steps in hardware and software to get disparity data. All FPGA cores for image acquisition and calculation (including the stereo core) are controlled by a bare metal application running on the ARM core. The second ARM core controls the Gigabit-Ethernet FPGA-IP-Core. A shared memory area makes the data exchange between both ARM cores possible.

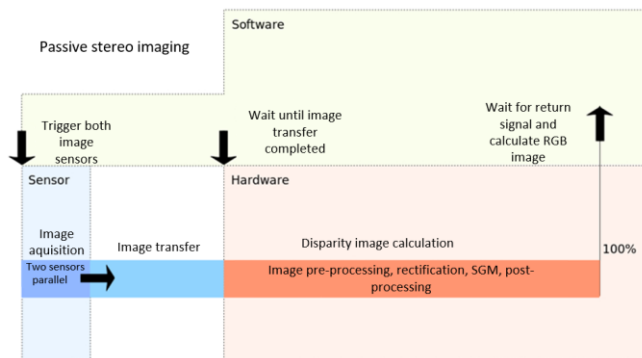


Figure 2: image calculation steps

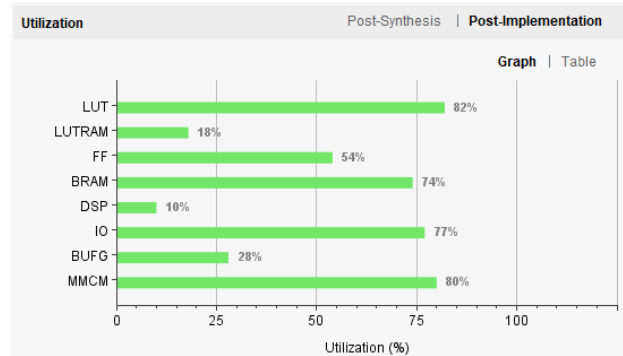


Figure 3: FPGA fabric utilization

Depending on the maximum disparity value which has to be configured, the calculation process takes a lot of iterations. To speed up the system to an acceptable calculation time, the calculation could be parallelized. The maximum factor of parallelization which could be implemented is a number with power of two and depends to the available FPGA hardware resources. In this case the factor of 32 fills the programmable FPGA logic nearly to 80 percent. The FPGA fabric utilization is shown in Figure 3. Another factor which increases the demand of logic resources is the image resolution, which is in this case 640 by 480. When the disparity image is calculated, there are two ways for data transfer. The image is shown on a display with a HDMI interface and/or transferred to a PC. As a physical interface a Gigabit-Ethernet interface is implemented with GigE-Vision and GeniCam compatibility. The transferred data is represented as 8-bit gray value, so nearly all GigE-SDK software can acquire this data. The stereo imaging system acts as a camera and the image data contain the left image, the disparity map and calibration data. The acquired 8-bit data has to be reinterpreted and parted to image data and disparity map.

With this information a point cloud calculation is done by multiplying the camera Q-matrix (calibration data) with the disparity map. In a python sample application, the data from stereo system is acquired via GigE interface. Adding the gray value from the left image, which is the reference image if the system, each point of the cloud gets the equivalent gray value. By usage of the visualization toolkit 'VTK' of python, the calculated point cloud is visualized (Figure 4 and Figure 5).

The stereo system could also calculate the point cloud data by itself. An implemented IP-Core created in Xilinx Vivado HLS is in parallel connection to the disparity data memory interface. In the PC image acquisition application there is the possibility to choose between disparity map or point cloud data transfer.

The system could be used in a standalone application, because of the possibility to save data on SD Card.

4. INCREASING SYSTEMS POTENTIAL – A THIRD SENSOR SLOT

Because of problems with poor results on less textured areas and reflection areas a place for a third image sensor was implemented. There are several options for using the third sensor slot. When using the same sensor from the stereo pair, now three stereo pairs could be calculated. Because of an equidistant arrangement, left-middle and right middle combination could be calculated. The results are verified or missing points are complemented. By using the outer left and right image pair, the resolution in depth direction is doubled but the minimum working distance increases.

In another task the third slot is used by a laser dot projector. The infrared works with the Microsoft Kinect pattern projector delivers an artificial texture on difficult surfaces. So more valid disparity points could be detected. Figure 4 and Figure 5 show the same scene under the same conditions, in the left without laser dot projection and in the right with enabled projection. In this case with laser projection are twenty percent more valid points found. There are less defects in the point cloud especially on low textured even areas. Laser dot projection has big advantages on technical surface where formation of laser speckle is insignificant. Brilliant surfaces could be a big problem just like in classical image processing too.

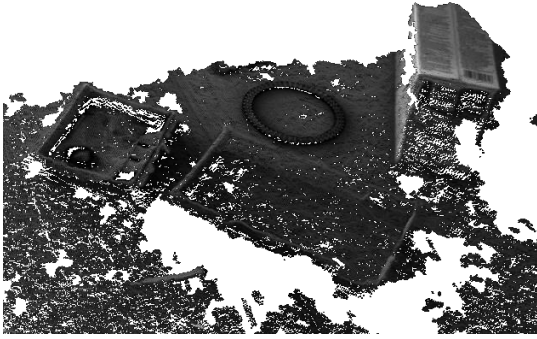


Figure 4: live point cloud without laser dots

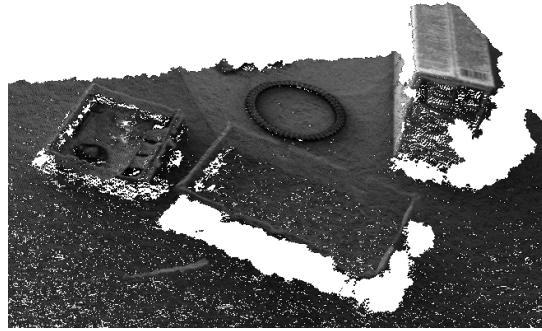


Figure 5: live point cloud with projected infrared laser dots

Further options to use the third sensor slot is the possibility to combine the point cloud with a RGB or multispectral or a high-resolution image, depended of the used image sensor. With a RGB sensor the point cloud could be colorized. A multispectral image sensor, i.e. based on filter-on-chip technology can deliver a survey of different materials which are contained in the point cloud. By using a high-resolution image-sensor the accuracy of 2D-image-processing could be combined with the knowledge of the third dimension. This combination is a fast alternative to 3D-systems which use stripe projection in combination with a sequence of acquired images.

5. SYSTEMS PARAMETERS AND ACHIEVED RESULTS

The passive-stereo system is constructed in two variants (Figure 6 and Figure 7); a third variant is planned. First there is a system with three equal image sensors using the same lenses. The second system combines two sensors with a Kinect dot projector. Both systems use the same hardware platform. Depending on the exposure time and used output mode, up to 15 frames per second are calculated. There are two modes for data output, via HDMI and GigE. In GigE mode disparity data or calculated 3D-points are transmitted.



Figure 6: three sensor stereo system



Figure 7: Stereo system with laser projector

The system accuracy could be calculated by the optical parameter and the estimated disparity error of 0.25 pixel. In a calibrated system with 96 mm baseline distance of the image sensors, like in Figure 7 and 10.6 μm pixel pitch with 8 mm focal length the calculated depth error is shown in Figure 8.

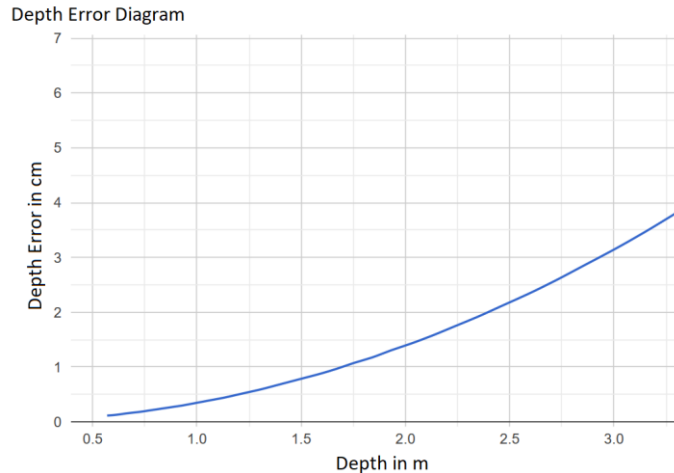


Figure 8: Depth error diagram [1]

To evaluate the system measurements with calibration spheres and planes where done. The sphere-spacing error was measured reclined to VDI guideline 2634 sheet 2 [2] and differs about 1.1 mm to 2.0 mm. The difference between the depth error in theory and measurements could be located mostly in two influences. First there is a significant random noise in the source images. This noise could generate deviations in the disparity values or defects. The second influence is the calibration sphere itself. The surface is white and has nearly no texture, so it's difficult to find disparity values. For additional measurements a special calibration object was constructed with wooden spheres and hidden mounting. Wooden CNC manufactured spheres have a natural textured surface and deliver more valid points. Another problem is the amount of valid 3D points of the sphere. In condition to the measuring principle in maximum less than fifty percent of a sphere is visible to the imaging system. In real condition the useable area is about thirty percent. In consequence the calculation of spheres center and diameter gets more inaccurate with less valid points.

The measured plane flatness is about 0.8 mm in the middle of the measuring field up to 1.8 mm on the borders. The whole measuring room in x-y is about 650 mm by 850 mm and 700 mm in height direction.

6. CONCLUSION

The two presented variants of a passive-stereo-system are implemented with Xilinx 7-Series SoC technology. The system accuracy is less than active-stereo-solutions, but the system is less complex, it has a better energy efficiency and works under hard conditions like direct sun illumination. The accuracy is too poor for measuring applications, but absolutely sufficient for object recognition and orientation tasks. The two transportable demonstrators are in a compact enclosure and could flexible used for many applications.

REFERENCES

- [1] Schauwecker, K., "Rechner für Objektiv-Brennweite und Stereo-Basisbreite", <<https://nerian.de/support/resources/calculator/>> (03. June 2019).
- [2] VDI Guide VDI/VDE 2634 Part 2: "Optical 3-D measuring systems - Optical systems based on area scanning", Beuth Verlag GmbH Berlin (2012-8).
- [3] Schauwecker, K., "Time_Stereo_Vision_on_FPGAs_with_SceneScan", https://www.researchgate.net/publication/327835490_Real-Time_Stereo_Vision_on_FPGAs_with_SceneScan (2019-6)
- [4] Hirschmüller, H., "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information", IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA (2005)