REAL-TIME DENSE STEREO MAPPING FOR MULTI-SENSOR NAVIGATION

D. Grießbach [a], A. Börner [a, *], I. Ernst [a], S. Zuev [a]

[a] German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Optical Information Systems,
Rutherfordstr. 2, 12489 Berlin - (denis.griessbach, anko.boerner, ines.ernst, sergey.zuev)@dlr.de

Commission V, WG V/5

KEY WORDS: image based navigation, dense stereo, stereo camera, calibration, data fusion

ABSTRACT:

Reliable estimation of position and orientation of the system in the 3D-world is normally the first and absolutely necessary requirement for the functional and operational capability of any unpiloted moving system for different real time applications.

A multi sensor approach for realisation of this task in combination with an optimal fusion of these measurements provides the best estimation accuracy of parameters. Depending on the applications the main inertial sensors will be combined with one or more other specific sensors like optical sensor (stereo cameras), GPS and others. The main investigation points are the development of the methods and algorithms for complex state estimation and their implementation in a real-time software and hardware solution, which serves as a base navigation system for different applications, for example for indoor navigation, driver assistance for vehicles and trains or the estimation of exterior orientation for airborne and space borne camera systems and attitude control for small satellites. An integral positioning system (IPS) was deployed and tested based on investigations.

The derivation of high quality products from the data of the optical stereo sensor can strongly improve the resulting navigation information. Additional to feature based stereo matching algorithms - which can be executed on a standard mobile computer in real-time - we use a GPU-implementation of the high quality Semi-Global Matching (SGM) algorithm. It can compete with the currently best global stereo methods in quality, at the same time it is much more efficient. Precondition for a real-time SGM approach is a special epipolar geometry of the input images: all epipolar lines must be parallel to the x-axis. Therefore the stereo camera system parameters of interior and exterior orientation have to be accurately determined. We have fully implemented the image rectification step and the SGM algorithm with several cost functions with OpenGL/Cg. This processing unit results in a real time 3D system called E3D and can be combined with the IPS sensor head.

The combination of both systems IPS and E3D allows improving the quality of data products. Disparity data integrity can be checked by controlling the orientation parameters of the stereo cameras, 3D points can be referenced in time and space, feature based matching can be improved and speeded up by using a priori knowledge from the dense disparity map, 3D points at infinity can be used for determining the rotation part of the ego motion. Such a multi sensor system is a perfect platform for SLAM applications.

1. INTRODUCTION

Real time navigation is a mandatory requirement for systems acting autonomously (e.g. robots, UAV's) or providing assistance information, like driver assistance systems for cars. Such navigation systems have to be reliable and robust. In most cases, they should work indoor and outdoor. Knowledge about data integrity is essential for acceptance.

Single sensor systems will not be able to fulfil these requirements due to their limited ability to measure defined physical parameters only. Consequently, DLR developed a multi-sensor approach (IPS – integral positioning system) allowing the integration of all kind of sensors providing position or attitude data or their derivates (e.g. speed, acceleration, angle speed). It has to be assumed that sensor data will be provided asynchronously.

The system introduced in this paper uses a stereo camera system as a main sensor. Navigation information retrieved by image processing is aided by other sensors.

The same sensor head can be used for dense stereo matching in a parallel processing chain resulting in a disparity map. This disparity map can be converted into depth maps or 3D environment models. The matching algorithm implemented is semi-global matching, a very powerful but time consuming method for finding conjugate pixels. In order to fulfil the real time requirements the algorithm was implemented on a GPU (graphical processing unit). This system is called E3D (Real time 3D).

Both systems can profiteer from each other. The paper describes IPS and E3D, their requirements, limitations and performances.

2. SENSOR HEAD

Both systems, IPS and E3D, use the same sensor head consisting of different sensors. Data processing chains are implemented in parallel. The results of the separated data processing streams are fused afterwards in real time.

2.1 Sensors

This chapter gives an overview about the applied sensors and their main parameters. The first three sensor systems are a default configuration. Other sensors can be included if needed, e.g. GPS receiver, barometer.

[*] Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

- Stereo camera
    - Type: Prosilica GC1380H, CMOS, panchromatic
    - Number of pixels: 1360 x 1024
    - Frame rate: up to 30Hz, typ. 8 Hz
    - Focal length: 4.8mm (indoor)/ 12mm (outdoor)
    - Base line: 0.2m
    - Field of view: 85° (indoor)/ 40° (outdoor)
    - Data interface: GigE
    - Data rate: 80MByte/s (uncompressed)
- Inertial measurement unit
    - Type: Systron Donner MMQ-50-200-400, MEMS
    - Angular rate range : 200°/deg
    - White Noise (angle random walk): 0.3 °/rt-hr (0.005 °sec/rt-Hz)
    - Acceleration range: ± 10g
    - White noise (velocity random walk) 0.5 mg/rt-Hz
    - Alignment: ≤5 mrad
    - Bandwidth: 50 Hz, nominal
- Tilt sensor
    - Measurement specialities, PEI-Z260-AL232
    - Measuring Range: ± 60 degree
    - Resolution: 0.1 degree
    - Repeatability: 0.3 degree @25°C
    - Accuracy (± 45°): < 0.5 degree RMS @25°C
    - Output Frequency: 14 Hz programmable

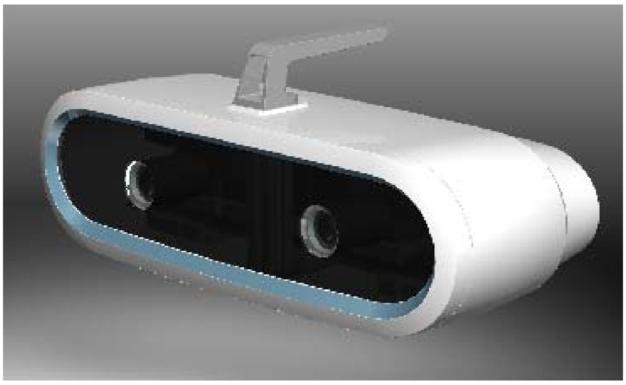

Figure 1. Sensor head for IPS and E3D

2.2 IPS box

As mentioned in the introduction, data from all single sensors can be generates asynchronously. For a data fusion it is absolutely necessary to provide a common time base for all sensor data. Precise time synchronization has to be realized whereby the chosen applications define real time requirements and acceptable synchronization errors.

In order to fulfil these requirements DLR designed and deployed a data synchronization box, called IPS box. It was developed to generate a time stamp for each incoming data set provided by a common clock. The IPS box consists of a FPGA card equipped with a Virtex IV.

An interface card provides CamLink, RS232, USB and Ethernet links.

It is planned to implement additional functions (e.g. image processing, state estimation) on the FPGA as well. In order to allow an easy implementation of different cores and functions, a hardware operation system is used.

The IPS box allows a precise synchronization of incoming sensor data for our applications. This approach relies on an constant time delay between generation of the sensor data and retrieval, which can be calibrated.

3. CALIBRATION

Referencing in space and time is one of the most important and one of the most underestimated tasks. Chapter 2 referred to time synchronization is, this chapter focuses on geometrical camera calibration and determination of mechanical alignments and system offsets of all aiding systems with respect the stereo cameras.

3.1 Geometric camera calibration

The objective of geometrical camera calibration is the determination of the line of sight of each detector pixel. Typically, a camera model describes the transformation from a pixel into a camera coordinate system which allows the application of a simple pinhole model afterwards. Within this step intrinsic (interior) camera parameters are needed. For any application using stereo or even multi-camera images an exact knowledge about the relative exterior camera calibration is crucial. It is used to apply the epipolar constraint or rectify images for calculating high dense disparity maps. Beyond classical methods (e.g. test pattern, goniometer/ collimator systems) DLR developed a novel method (Bauer et al., 2008) for geometrical camera calibration based on diffractive microstructures (DOE – diffractive optical elements). DLR has proven that this new method is able to determine interior camera calibration parameters at least at the same accuracy as known from classical approaches (rms < 1/100°). A high reproducibility and low durations of measurement are the big advantages of this method.

DLR's system based on DOE's is used as standard equipment for panchromatic cameras with aperture lower then 70mm and field of views between 30 and 120 degrees.

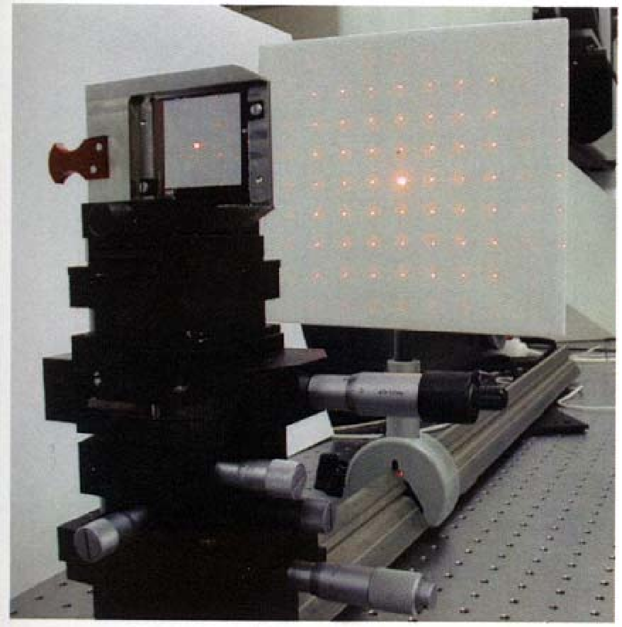

Figure 2. Projection of patterns generated by a diffractive optical element to a screen

3.2 Alignment and offsets

Another challenge is the complex alignment procedure for the whole sensor system. After calibrating the stereo camera system as shown in (Grießbach et al., 2010a) it is now feasible to calculate the rotation between cameras and tilt-sensor. Therefore the orientation of the camera and the tilt angles are measured at a few static positions in front of a calibration chart which allows the determination of two tilt angles between stereo system and tilt angle.

Finding out about the alignment between camera and IMU is slightly more ambitious. For the IMU which is measuring accelerations and angular velocities only a dynamic calibration is needed. Comparing both, measured IMU angle velocities and calculated angular velocities from the camera the rotational alignment can be estimated (angle alignment between IMU and camera). The translation alignments of all single sensors with respect to a reference coordinate system can be measured mechanically in most cases. If the reached accuracy is not enough, translations have to be estimated with a Kalman filter extended by additionally states for the alignment parameters.

4. INTEGRAL POSITIONG SYSTEM (IPS)

DLR developed a multi-sensor navigation system called IPS (integral positioning system) for the determination of position and attitude of mobile devices in indoor and outdoor environment. All sensors providing position or attitude or their derivates can contribute to a common state estimation.

4.1 Stereo vision

Stereo cameras are the main sensor. They are used for the determination of changes in position and attitude, which corresponds to a speed and an angular speed. This information is fed into the complex state estimation unit.

Assuming a precise geometric camera calibration allows us to apply a simple pinhole model for the projection of any 3D object point into a 2D image point and vice versa (at a certain distance).

For image based pose estimation features in image space are detected and tracked over consecutive frames. Natural landmarks such as corners, isolated points or line endings are suited to be used for pose estimation. Several algorithms were implemented and evaluated. We decided for Harris corner detector (Harris and Stephens, 1988). Different quantities can be used for the description of the strength of a corner (e.g. trace or minimal eigenvalue of the autocorrelation matrix).

Assuming a sequence of synchronously acquired stereo images, we apply two different matching steps: intra-matching (features in the left image and in the right image at time t_N) and inter-matching (left image at time t_N and left image at time t_{N+1}, right image at time t_N and right image at time t_{N+1}). Normalized cross correlation is applied as matching algorithm. Epipolar constraints allow a reduction of the size of the search window for intra-matching, the estimation of position and attitude change with the help of aiding sensor systems allow a prediction of the position of the search window for inter-matching.

After extracting of object points and performing the intra-matching 3D coordinates of object points can be reconstructed in a local coordinate system by knowing the interior orientation and the relative exterior orientation of the cameras.

The relative change in pose from time t_N to time t_{N+1} is estimated by considering all 3D points which could be reconstructed after inter-matching successfully. The transformation of point cloud N to point cloud N+1 is estimated minimizing the residual errors. A RANSAC approach stabilizes the solution.

4.2 Inertial Navigation

Inertial navigation systems (INS) consists of an inertial measurement unit (IMU) containing 3-axis- gyroscopes and –acceleration sensors as well as a computing unit to integrate the IMU signals to a navigation solution applying the known strapdown mechanization.

The state vector contains 6 parameters (position, velocity, acceleration, acceleration bias, angular velocity and angular velocity bias) with three dimensions and the quaternion with four components. This results in a 22 element state vector.

Four state estimation several filters were analyzed. Classical Extended Kalman filter (EKF) and scaled unscented Kalman (sUKF) filter were implemented.

Besides from being computational more costly than the EKF the sUKF has many advantages. By being accurate to the third order with gaussian inputs it produces more stable estimates of the true mean and covariance. Furthermore it behaves better in case of unknown initialization values and analytic derivations which can be quite complex are not needed. With highly non-linear observation equations of the used sensors just like parts of the state equations the sUKF ensures a reliable navigation solution.

4.3 Data fusion

Figure 3 shows the combination of Kalman filter and opical system (tracker) which provides incremental attitude- and position updates. Receiving IMU- or inclinometer-measurements the full filter cycle is completed including a check for feasibility of the data. For incoming stereo images first the time-update is done. The a priori estimate enables the tracker to perform a very fast and reliable inter-frame matching. After triangulation the calculated move estimate is used for the measurement-update within the Kalman filter (Grießbach et al., 2010b).

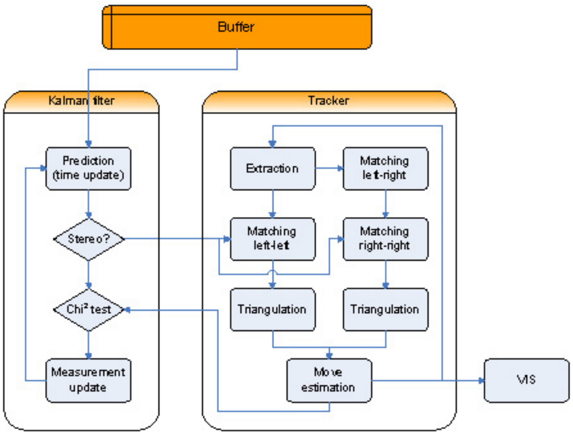

Figure 3. Figure placement and numbering

4.4 Results

Different experiments were performed in order to validate the IPS system and to estimate its performance. IPS was mounted on a robot arm and moved in a defined way. The comparison between robot trajectory and processed IPS navigation allowed a first analysis of the over all quality of the IPS system. IPS was mounted on several vehicles (cars, railway) in order to test the dependency on the motion model. The most difficult experiment was done by applying IPS as an indoor-navigation system carried by a person. Without any external reference data and covering a distance of about 90m, which took about 2min, a position error of less then 50cm was achieved.

5. REALTIME 3D SYSTEM (E3D)

The derivation of high quality 3D products from the data of the optical stereo sensor can strongly improve the resulting navigation information. Additional to feature based stereo matching algorithms described in section 4.1, which is executed on a standard mobile computer in real-time, we use a GPU-implementation of the high quality Semi-Global Matching (SGM) algorithm (Hirschmüller, 2008; Hirschmüller, 2005).

5.1 Semi-Global Matching Algorithm (SGM)

We assume a rectified (in means of parallel epipolar lines), binocular stereo pair as input and refer to the 8 bit intensity values of the left and right image. The SGM describes a very efficient approximation method to find the minimum of a global energy function E(D) based on the input images which consists of two terms – one for matching costs for all pixels and their disparities, the second for punishing disparity changes. The approximation is realized by a path wise cost summation. The matching costs are aggregated in a three dimensional cost volume (x dimension, y dimension, disparity range). The path directions are defined by a 8-neigbourhood.

Based on this cost volume S a solution for the minimization problem and hence the disparity image can be determined. The disparity of each pixel is given as the index of the minimal cost value. Subpixel interpolation can be done by consideration of adjacent cost values.

A consistency check can be realized in two different ways: a) after a left-right match a right-left match has to be executed, b) a diagonal search in S is applied.

For the suppression of outliers both disparity images are filtered with a 3×3-median filter. The result is used in a consistency test between left and right disparity image.

The costs depending only on the values of corresponding pixels in the energy function can be determined by different methods, e.g. as a simple grey value difference, as a Census cost function or an sum of absolute differences (SAD), or it can be based on Mutual Information (MI), which uses the joint entropy of both images and the individual entropies of the left and right image (Hirschmüller, 2008; Ernst and Hirschmüller, 2008).

5.2 GPU Implementation

Our development is based on NVIDIA's G8-series with underlying CUDA architecture (NVIDIA, 2007). We started our work with the OpenGL/Cg (Segal and Akeley, 2009; NVIDIA, 2009) programming technique. Our current implementation of SGM is based on this proven technique, but is also intended as a reference for a future migration to OpenCL (Khronos, 2008), which will overcome the OpenGL/Cg programming limitations.

Our implementation is based on the usage of at least three render buffers of a frame buffer object. These RGBA-buffers with a 16-bit-float data type (i.e. 1 sign bit, 5 exponent bits, and 10 mantissa bits) are used in a ping pong technique for keeping the data while the arithmetic operations are normally done in 32-bit-float precision. All work is carried out through the execution of OpenGL rendering commands and several specialized fragment and vertex shader programs.

The priority objective for designing the memory buffer partitioning is to provide the input data in a form that minimizes the number of memory accesses in all computations and that allows the optimal use of the hardware and associated data structures designed for fast 3D rendering of textured objects. Initially, the left and right original images are considered as textures and loaded directly and rotated with all necessary levels of detail to the colour channels of a render buffer. According to Section 5.1, a cost volume S is calculated.

The path costs for one pixel in one path direction are dependent on only the path costs of the predecessor pixel in this direction, not on the costs of the neighbouring pixels perpendicular to the path. Therefore all path costs e.g. in horizontal direction for an entire image column can be calculated in parallel. The calculation of path costs for vertical or diagonal directions r is done analogously, respecting the corresponding predecessor dependencies. Four of the eight path directions can be computed at the same time using the four colour channels.

All path cost values depend not only on their predecessor values, but also on the minima of the accumulated path costs for all disparities for the previous pixel in the current path. For finding the minima for all disparities, a composite procedure of some comparisons in the fragment shader and application of OpenGL blending equation GL MIN turned out to be the fastest. When the path costs for one column and one row are available they are added by rendering of one rectangle respectively single rows to the cost volume S.

The next step is the disparity selection. The determination of the indices of the minimum values of the accumulated path cost values in S is done by a reduce method. In the first reduce step explicit indices are generated and stored together with the values. After the last reduce step, the remaining columns of the rectangles of S contain the minimal path costs and the corresponding raw disparity columns.

A sub-pixel interpolation can be integrated into the first reduce step with marginal computational effort.

As part of the GPU implementation different cost functions have been implemented and tested, among them the MI cost function and a Census cost function for a 5×5 environment window. The computation of disparity images with MI cost function requires a special hierarchical approach, it starts on a coarse level of detail. The disparity image that is found in this step is up-scaled and a new matching table for the next level of detail is calculated. This iteration terminates when a final disparity image for the original image resolution has been reached.

In order to avoid unnecessary data transfer to and from the CPU, the MI matching table is calculated on the GPU, too.

Special epipolar geometry of the input images is a precondition for a real-time SGM approach: all epipolar lines must be parallel to the x-axis. Therefore the stereo camera system parameters of interior and exterior orientation have to be accurately determined. The input images are rectified in a pre-processing GPU step before SGM calculation can be executed.

5.3 Results

A couple of experiments were carried out on different indoor and outdoor platforms.

We have tested the implementation on several NVIDIA GPU boards, e.g. on a GeForce 8800 ULTRA, a GeForce 260 und a GeForce 280. Disparity images that were computed on the graphics card are shown in Figure 4. The interesting aspect of the GPU implementation is the run time on different image sizes and disparity ranges. Figure 5 shows the results.

Our current implementation reaches e.g. 5.7 fps on a NVIDIA GeForce 280 using an image size of 640 × 480 pixel with 128 pixel disparity range, computing 5 hierarchical levels with MI cost function. With just one hierarchical level for a 5 × 5 Census cost function we reach 8.6 fps for the same size and disparity range on the GeForce 280, or 27 fps for 320 × 240 pixel with 128 pixel disparity range. All values are based on a calculation of the left disparity image and a subsequent fast consistency test. It turns out that the parallelization capabilities of the GPU for the SGM algorithm give biggest advantages for images with higher resolutions and many disparity steps.

Exemplary, figure 4 illustrates a measurement campaign on a railroad engine. The stereo camera system was mounted on the head of the vehicle as a forward looking sensor.

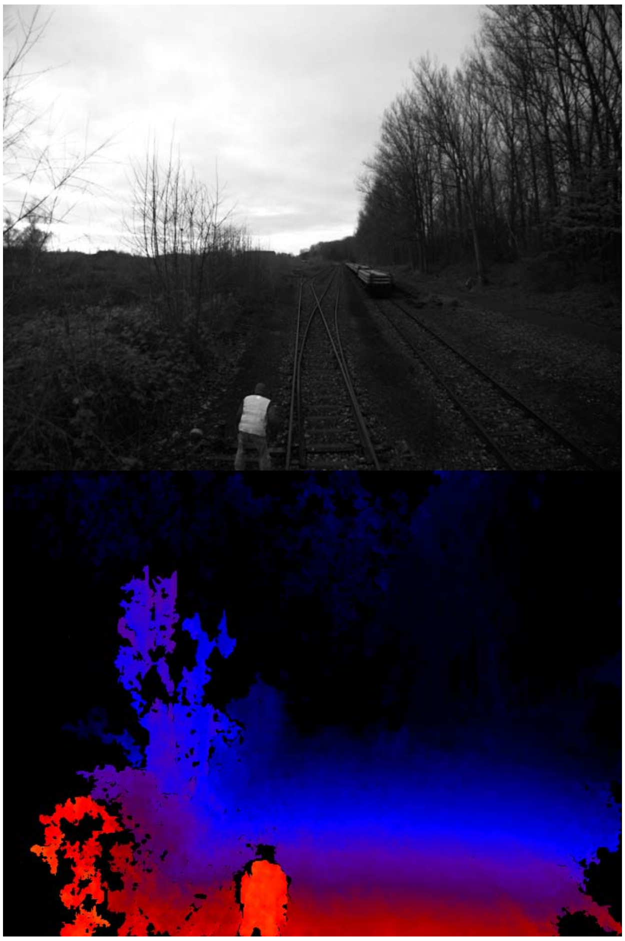

Figure 4. E3D experiment on a railroad engine, one image of the stereo camera system (top), disparity image (bottom)

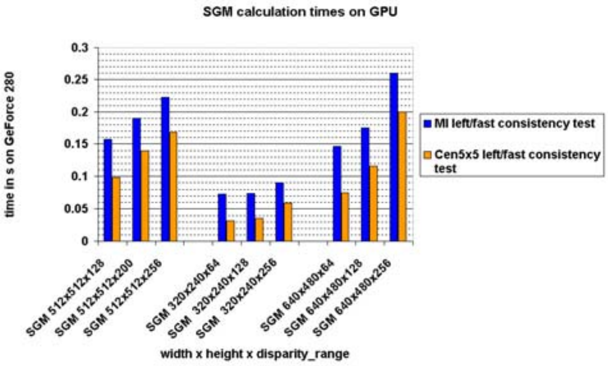

Figure 5. Computing times for different GPU's, image sizes and cost functions

6. APPLICATIONS

In this section we would like to focus on the simultaneous application of IPS and E3D systems. The sensor head is identically, both processing paths are implemented in parallel. It can be done easily due to a parallel GPU and CPU/FPGA usage.

- 3D environmental models retrieved by E3D can be referenced spatially by applying navigation data from IPS.
- Intra-matching for IPS can be taken over by E3D completely.
- Rotation and translation can be separated easily be selecting the zero-disparity pixels provided by E3D only for rotation.
- Navigation can be supported using a dense 3D point cloud provided by E3D.
- IPS intra-matching can be used for verifying epipolar geometry being necessary for E3D and providing self-calibration.
- A precise measurement of the exterior orientation of the cameras allows the calculation of a disparity map from the data of one camera only.

7. OUTLOOK

Verification of the combined system will be the next step. It will be used for different applications with defined user requirements.

Increasing the frame rate is an important task for both sub-systems, E3D and IPS. FPGA solutions can be helpful in the near future.

8. REFERENCES

Bauer, M., Grießbach, D., Hermerschmidt, A., Krüger, S., Scheele, M. and Schischmanow, A., 2008. Geometrical camera calibration with diffractive optical elements. *Opt. Express*, 16(25), pp. 20241–20248

Ernst, I., Hirschmüller, H., 2008. Mutual information based semi-global stereo matching on the gpu. *ISVC (1), volume 5358 of Lecture Notes in Computer Science*, pp. 228-239. Springer, 2008.

Grießbach, D., Bauer, M., Scheele, M. Hermerschmidt, A., Krüger, S., 2010a. Geometric stereo camera calibration with diffractive optical elements. *EuroCOW 2010*, Castelldefels, Spain.

Grießbach, D., Baumbach, D., Zuev, S., 2010b. Vision aided inertial navigation. *EuroCOW 2010*, Castelldefels, Spain.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. *Proc. of the 4th ALVEY Vision Conference*, pp. 147–151.

Hirschmüller, H., 2005. Accurate and effcient stereo processing by semi-global matching and mutual information. *CVPR (2), IEEE Computer Society*, pp. 807-814.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2)*, pp. 328-341

NVIDIA, 2007. *NVIDIA CUDA Compute Unified Device Architecture - Programming Guide*.

NVIDIA, 2009. *CG Reference Manual, 2.2 edition*.

Segal, M. and Akeley, K., 2009. *The OpenGL Graphics System: A Specification*.

Khronos OpenCL Working Group, 2008. *The OpenCL Specification, version 1.0.29*.