

Dynamic Sensor Matching for Parallel Point Cloud Data Acquisition

Simone Müller
Leibniz Supercomputing Centre (LRZ),
Germany
Boltzmannstrasse 1
85748 Garching bei München
simone.mueller@lrz.de

Dieter Kranzlmüller
Ludwig-Maximilians-Universität (LMU)
MNM-Team
Oettingenstr. 67
80538 München
kranzlmuller@ifi.lmu.de

ABSTRACT

Based on depth perception of individual stereo cameras, spatial structures can be derived as point clouds. The quality of such three-dimensional data is technically restricted by sensor limitations, latency of recording, and insufficient object reconstructions caused by surface illustration. Additionally external physical effects like lighting conditions, material properties, and reflections can lead to deviations between real and virtual object perception. Such physical influences can be seen in rendered point clouds as geometrical imaging errors on surfaces and edges. We propose the simultaneous use of multiple and dynamically arranged cameras. The increased information density leads to more details in surrounding detection and object illustration. During a pre-processing phase the collected data are merged and prepared. Subsequently, a logical analysis part examines and allocates the captured images to three-dimensional space. For this purpose, it is necessary to create a new metadata set consisting of image and localisation data. The post-processing reworks and matches the locally assigned images. As a result, the dynamic moving images become comparable so that a more accurate point cloud can be generated. For evaluation and better comparability we decided to use synthetically generated data sets. Our approach builds the foundation for dynamic and real-time based generation of digital twins with the aid of real sensor data.

Keywords

Multi-Sensor, Dynamic Matching, Stereoscopy, Point Cloud, Real-Time, Data Acquisition, Computer Vision

1 INTRODUCTION

From the fields of autonomous driving [Tow19] and robotics [Liv12] to many other applications, a large amount of sensor data is needed to ensure the accurate and valid sensing of environmental space. Stereo cameras record distances through the synchronous recording of stereoscopic images, which allows to compute 3D points that form of a point cloud. Sensors like light detection and ranging (LiDAR) or stereoscopic systems record and process valid digital representations near real time through visual simultaneous localisation and mapping (SLAM), visual detection and tracking, as well as visual classification and recognition [You13]. The quality of such digitisa-

tions is severely restricted by sensor limitations such as range, depth resolution and image accuracy as well as different sampling rates of each sensor [Kad14]. These constraints lead to measurement fluctuations, outliers, asynchronous sensor adjustment and consequently imaging errors, which are particularly noticeable on inhomogeneous surfaces and edges. Figure 1 shows the result of a stereoscopic room acquisition by using ZED2 from Stereolabs [Ste20]. The manufacturer's software of the ZED2 already displays a real-time based point cloud as shown in the illustration. The flat walls and ceilings of digitised space seem uneven and distorted so that the entire point cloud appears inhomogeneous. In the frame on the left and right side of figure 1, the irregular arrangement of the coloured grey 3D points is clearly visible. The pattern of the orange line in the middle of the illustration shows how many measurement cycles and sequences are usually necessary to digitise a room in this form. In addition, the movement of the stereoscopic sensor leads to temporal latency and higher data processing as well as error rates so that localisation errors consequently occur [Sid03]. Regardless whether the sensor or

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

object is moving, the real-time based visualisation of environmental space is severely affected by movement, especially because further physical effects like lighting conditions, material properties and reflections have an influence.

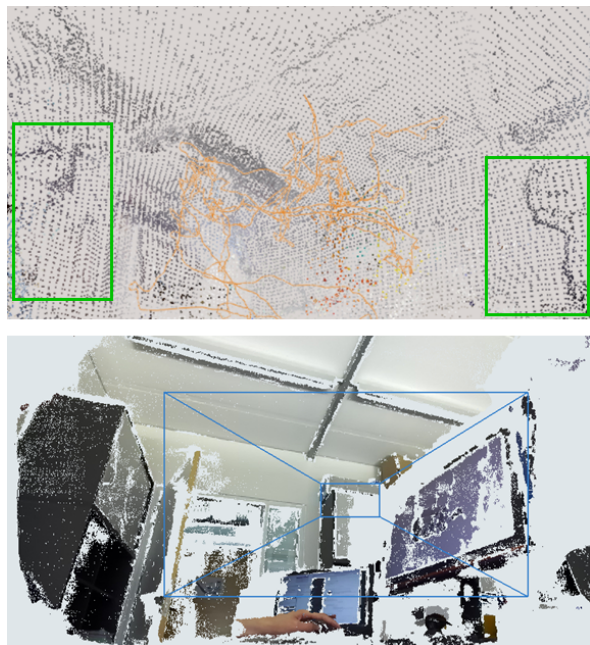


Figure 1: Point cloud visualisation captured by ZED2 stereo camera from Stereolabs [Ste20].

During the real-time acquisition of the environment, each captured frame is rendered individually. There the point cloud shows a rather high fluctuation of the 3D points due to aforementioned effects. Especially small and complex surfaces as well as moving objects can only be reproduced to a limited extent. The use of multiple stereoscopic sensors increases the density of surrounding spatial informations so that errors can be reduced by data matching. Often several statically arranged cameras are used to capture the environment [Tow19].

Our research employs the simultaneous and dynamic use of multiple camera systems. In addition to dynamic sensor implementation, we propose the real-time based generation of digital twins in three-dimensional space. We investigate the characteristics of such dynamic sensor systems and whether inhomogeneous errors can be reduced by this approach.

The paper is organised according to a fixed structure consisting of related work, fundamental concept of dynamic multiple sensors, methodology and experimental setup as well as finally results and discussion.

2 RELATED WORK

In this section we discuss the stereoscopic foundations and continue with details about the related works and

our resulting motivation.

Due to the practical features and the number of integrated sensors, stereo cameras are usually used for digital environmental sensing [Ste20]. Figure 2 describes the composition of such a stereo camera. This consists of two static cameras C_L and C_R whose image orientation \vec{v} is parallel to each other (for simplification of figure 2, the cameras are not drawn parallel). With the knowledge of defined distance D between the static fixed cameras C_R and C_L , it is possible to create synchronised image captures. The distance itself is a constant value that usually corresponds to the human inter-pupillary distance of approximately 6.5 cm [Ste20]. To avoid asynchronous sampling rates, the same settings and properties like image resolution, fps and sensitivity apply to all cameras.

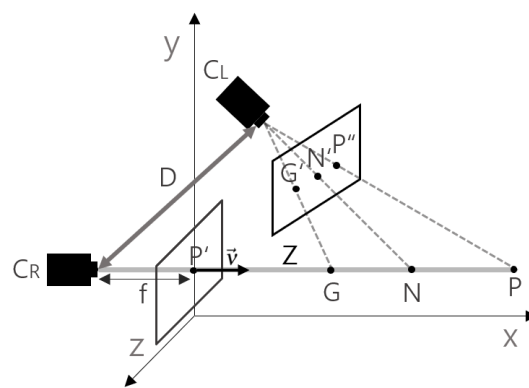


Figure 2: Schematic representation of stereoscopic depth perception. There are different pixels (G, N, P) on the optical beam Z . The three-dimensional point P and the focal points P' and P'' , which can be determined by the focal length f , span the epipolar plane [Zha16]. The distance D between right camera C_R and left camera C_L is a static value.

In three-dimensional space the movement of a stereo camera is defined by translation and intrinsic rotation. A basic distinction is made between the intrinsic sensor and image rotation. While sensor rotations refer to the position of the inertial measurement unit (IMU), the image rotations relate to the orientation of the camera picture. For figure 2, only the image orientation is relevant. The intrinsic rotation matrix is defined as $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ where the product of individual rotations is called Yaw(R_ψ), Roll(R_ϕ) and Pitch(R_θ). The rotations correspond to the orientation around their respective axes [Ayk18]:

Rotation around x-axis R_ϕ :

$$R_\phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix} \quad (1)$$

Rotation around y-axis R_ψ :

$$R_\psi = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix} \quad (2)$$

Rotation around z-axis R_θ :

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

The camera position is expressed by $T \in \mathbb{T}^{3 \times 3}$ where the translation directions are defined as T_X, T_Y, T_Z [Ike14], [Ska20]:

$$T = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (4)$$

To create a spatial image, depth perception D is required by using disparity d in three-dimensional space $\mathbf{X} \in \mathbb{X}^{3 \times 3}$. Since the pixel series of the camera pictures are identical, the distance between object P' and camera $|P' - P''|$ can be determined by using the epipolar geometry [Zha16].

$$d : |P' - P''| = \frac{D \cdot f}{Z} \quad (5)$$

The direction vector \vec{v} of the optical axis, which is perpendicular to the camera image, depends on rotation R_ϕ, R_ψ and focal length f .

$$\vec{v} = R_\phi \cdot R_\psi \cdot \begin{bmatrix} 0 \\ 0 \\ f \end{bmatrix} \quad (6)$$

In order to reproduce objects, the individual camera images must be matched. Since alignment and distance between the two cameras are static and defined, the individual rows of pixels can be compared with each other. The spherical structure of the front view of a camera creates an inhomogeneous and distorted grid. This geometric unevenness is illustrated in figure 3.

Image curvature is particularly noticeable at the edges. By using special computer vision algorithms, such as semi-global matching (SGM) [Hir11, Rob20], the pixels can be matched despite curved image edges. SGM estimates the density of the disparity map from the rectified stereo image pair. For this purpose, the maximum allowed disparity shift ($C(p, D_p)$) and the regularisation cost $PT[|D_p - D| \geq 1]$ are migrated by

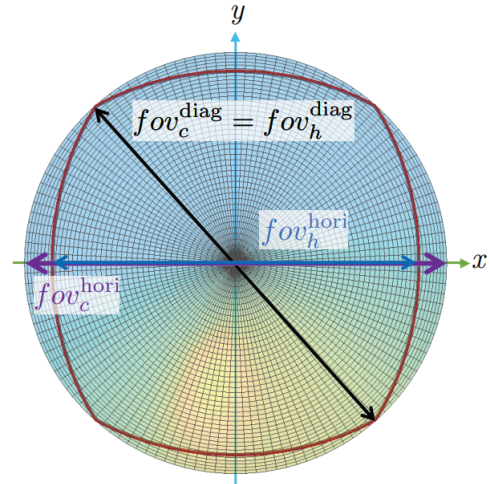


Figure 3: Front view of the camera image surface and the borders of the image content [Ayk18].

summation.

$$E_{(D)} = \sum(C(p, D_p)) + \sum(PT[|D_p - D| \geq 1]) \quad (7)$$

Depth mapping allows us to visualise the disparity as a point cloud in which the individual pixels are in temporal relation to the stereo camera. The resulting point cloud is considered as one object. There is no delimitation or allocation that would directly help to identify objects and to perceive the occluded back side of the object. In addition, only a part of the object front is visible, which allows a spatial contour. Another aspect is the quality of such point clouds.

With O'Riordan's study of colour analysis for reflective surface recognition, challenges such as image matching, false boundary as well as problems due to surface texture and obstacle detection were named [Ori18]. Marton et al. also recognised these problems and worked on a rapid surface reconstruction methodology [Mar09]. They analysed large and noisy point cloud data sets. Although they achieved good results, they also saw a need for further optimisation with regard to the high level of noise.

Chang et al. named the sparse stereo correspondences as well as their disparity limitations and have presented an efficient warping-based method for stereoscopic image retargeting [Cha11]. They have formulated these constraints as an energy minimisation problem for obtaining optimal warping fields for the images and have extended interactive stereoscopic image processing with their approach.

Another approach is provided by Zienkiewicz et al. with the method for incremental surface reconstruction from a moving camera [Zie16]. They use the concept of dynamic level where they adaptively select the best resolution of the model and fuse measurements in an

efficient multi-scale mesh representation. As a result, they declared obvious limitations in the use of height map. By optimising the three-dimensional settings and developing a more flexible multi-scale fusion method, they saw the possibility to reduce the limitations.

Due to the limited field of view afforded by a single stereo camera, synchronous and real-time 360 degrees sensing are difficult to implement. Therefore, multiple sensors like LiDAR or stereoscopic systems are usually used to perceive 360 degrees.

With an autonomous vehicle platform, Siddharth et al. worked on temporally synchronised perception and navigation data [Sid03]. The stereo cameras and LiDAR systems were statically mounted on the vehicle. From the sensor data, the environment was captured into a 3D point cloud, ground reflection map and ground truth pose of the hosted vehicle. The data sets recorded by using SLAM algorithms were also the first dynamically recorded multi data. Siddharth showed that synchronous 360 degree sensing is promising but the sensors are highly dependent on weather and light conditions.

Piatkowska et al. worked on the problem of asynchronous stereo vision for dynamic image sensors [Pia13]. They extended existing methods for event-based processing through the cooperative approach, which enables spatio-temporal and asynchronous three-dimensional reconstruction. They proved that dynamic, asynchronous and cooperative implementation is possible using a specific algorithm.

We question how the named challenges of faulty surface textures, physical influences such as light conditions, reflection, surface textures and also weather conditions can be improved. We come to the conclusion that multiple sensors can not only provide sensing but also the possibility of different perspective views. Therefore, we propose the hypothesis that synchronous dynamic multi-sensors can be used to improve real-time based point cloud visualisation. With perspective change only the relevant images would be used for evaluation. In this way, influences such as direct sunlight or limitations of depth perception could be reduced as far as possible.

3 FUNDAMENTAL CONCEPT OF DYNAMIC MULTIPLE SENSORS

This section introduces the underlying concept of multiple cameras, which are able to move dynamically through three-dimensional space and perceive the environment in the process.

Complex multi-sensor systems like IMU [Spa17] allow motion profiles of cameras to be recorded and used as a direct reference for determining position. Through detection of the sensor orientation

and spatial position, images can be aligned accordingly.

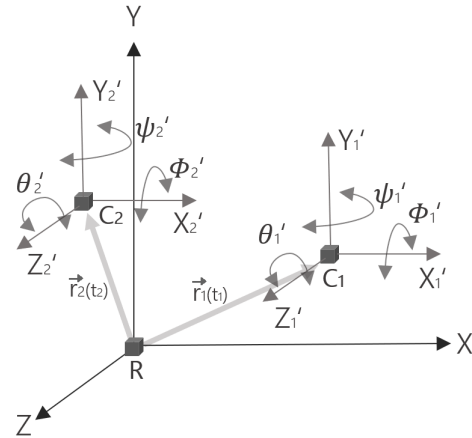


Figure 4: Schematic representation of the spatial coordinates of the stereoscopic sensors (C_1, C_2) in relation to an arbitrary reference point R : Translation directions (X, Y, Z) and intrinsic rotation (R_ψ, R_ϕ, R_θ)

In contrast to a single stereo camera, multisensor systems require an additional reference. Figure 4 illustrates the dynamic relation of the cameras to the reference, which are moving in a time-dependent vector field $\vec{r}_{1(t1)}, \vec{r}_{2(t2)}$. The velocity v_1, v_2 and acceleration a_1, a_2 of the respective position vectors can be used to determine the position at which the individual images were taken in space [Chr16].

$$v_t = \frac{\partial r_t}{\partial t} = \left(\frac{\partial X_t}{\partial t}, \frac{\partial Y_t}{\partial t}, \frac{\partial Z_t}{\partial t} \right) \quad (8)$$

$$a_t = \frac{\partial v(t)}{\partial t} = \left(\frac{\partial^2 X_t}{\partial t^2}, \frac{\partial^2 Y_t}{\partial t^2}, \frac{\partial^2 Z_t}{\partial t^2} \right) \quad (9)$$

Figure 5 shows the schematic reference related movement of a stereo camera that corresponds to the position recording of an IMU. In addition to the progression line (dotted line) on which the stereo camera moves, the vectorial viewing direction (arrows) can also be seen.

4 METHODOLOGY AND EXPERIMENTAL SETUP

The methodological approach consists of data initialisation and object recognition in three-dimensional space, generation and allocation of metadata records, point cloud matching, as well as rendering.

To ensure that the collected image data is correctly mapped in the point cloud, the references and stereo cameras must be initialised. Reference R is usually an object that forms the origin of the coordinate system with the parameters $T_{Ref}(0, 0, 0), R_{Ref}(0, 0, 0)$. The parameters can change as soon as the object is fixed

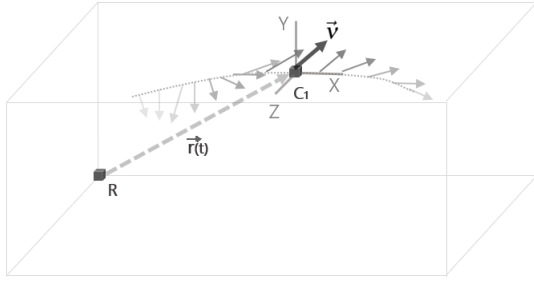


Figure 5: Dynamic movement of a vector field $r(t)$ based camera C_1 . The direction vector \vec{v} shows the orientation based on camera C_1 . The arrows show the time-dependent snapshots of the gaze alignment. The darker the arrows, the closer is the alignment to the present view. While the lighter arrows on the left side of the illustration belong to the past, the lighter arrows on the right side represent the future.

to a human being, for example, who is also moving. The generated points of the cloud are static if no adjustment is necessary. The offset position of the coordinate system corresponds to the distance between the rendered 3D point cloud position of a single frame and the current position of the capturing stereo camera. If the relations between the reference and cameras are not initialised, the newly created point cloud of individual stereo cameras will find itself at a different location. This shifts the already recorded point clouds of the individual stereo cameras by the delta contribution of the position coordinates Δ_1, Δ_2 .

$$\Delta_1 = T_{\text{Ref}(X,Y,Z)} - C_1(X,Y,Z) \quad (10)$$

$$\Delta_2 = T_{\text{Ref}(X,Y,Z)} - C_2(X,Y,Z) \quad (11)$$

To ensure that there is no shift in sensor position, the stereo cameras have to initialise themselves on known geometries. Initialisation always takes place as soon as a recognised geometry is recognised. To avoid this, a continuous position initialisation is necessary. One possible approach is to initialise geometries with an Artificial Intelligence (AI) based object recognition. A previously trained AI contains a three-dimensional geometric understanding of objects. If the AI recognises a known object, markers are set and stored in a map. These markings comprise localisation information of the own position and other marked geometries. A correction factor can be determined by comparing the position data of the map and the stereo cameras.

Figure 6 illustrates a possible geometric shape for initialisation. For this purpose, an identifiable geometry is provided with significant points by using computer vision algorithms. The retrievable spatial coordinates of the marker can be captured by both stereo cameras. The experimental set-up takes place in a games engine (Unreal Engine 4 [Ue21]). This has the advantage

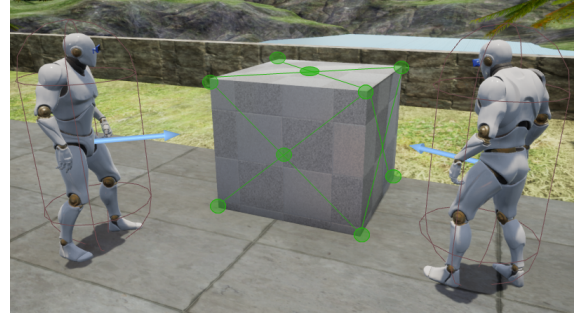


Figure 6: Initialisation of the coordinate systems of C_1 (figure left) and C_2 (figure right) by using geometrically arranged points on an object.

that a realistic environment can be created where all parameters and physical settings are traceable. In addition, static and dynamic objects can be used. The movement sequences of the stereo cameras can also be specified in the form of avatars. The user himself serves as a reference. Since the captured images of the cameras are position and time-dependent, a new metadata record must be created. Table 1 shows the relevant content of such metadata to reproduce the temporally related informations. Thus, the connected data are momentary images that help to reconstruct the point cloud depending on its position. In addition to sensor parameters, the already detected geometric initial points in the point cloud are also collected. The time stamp enables the synchronous and asynchronous use of the cameras. In the metadata set, a further distinction is made between the rotational orientation of the image and the rotation as well as translation of the entire stereo camera.

Image data	Sensor IMU	Point cloud
Timestamp	Timestamp	Timestamp
Resolution	Translation	Initial Points
Rotation	Rotation	3D Coordinates
-	Delta	Geometric Type

Table 1: Reference-dependent metadata set. Such a metadata record is created for each sensor and captured data. In addition to image and localisation data, it also contains information about the initialisation of the coordinates.

The images of the stereo camera itself are already aligned due to their static arrangement and therefore do not need to be taken into account. For images with dynamic spacing, the perspective of the image acquisition changes. As a result, the pixel rows are no longer comparable using the SGM algorithm. Therefore, a methodology must be considered for comparing the individual images.

Description	CPU(i5-7300HQ@2.5GHz)	GPU 0(31,9GB)	GPU 1(35,9GB)	RAM(64GB)
No-load Operation	4 % - 1.64 GHz	0 %	0 %	10 %
Simulation - Mean	44% - 2.86 GHz	1 %	38 %	12 %
Simulation - Maximum	75% - 2.86 GHz	11 %	41 %	13 %
Simulation - Minimum	9 % - 1.39 GHz	0 %	11 %	12 %
Real - Mean	54% - 3.14 GHz	0 %	12 %	13 %
Real - Maximum	78% - 3.09 GHz	7 %	12 %	13 %
Real - Minimum	54% - 3.14 GHz	0 %	12 %	13 %

Table 2: Performance comparison between real and simulated multisensor data during real time operation. The no-load operation of computer is shown as a reference.

Figure 7 illustrates the dynamic distance relationship between sensor C_1 , C_2 and reference R . It shows that depth perception can also be captured from stereo cameras where the field of view appears larger. Since the correct coordinates of the stereo cameras are known through the initialisation, the dynamic distance C_{12} between C_1 and C_2 can be calculated. If the image orientation of the vector field \vec{v} is corrected, the SGM algorithm can be applied again and a depth image can be generated. Using the generated position-fixed depth images, a point cloud can be reconstructed that has a wider field of view. By superimposing the depth images, the localisation errors are also superimposed.

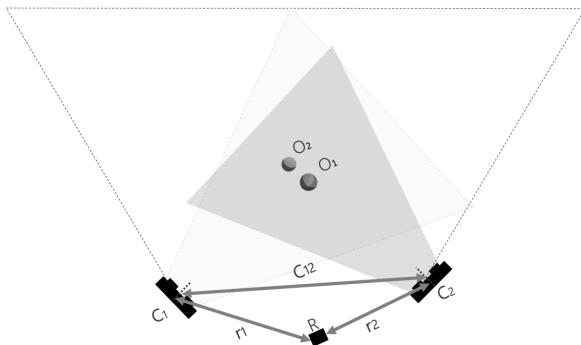


Figure 7: Two-dimensional representation of dynamically arranged stereo cameras C_1, C_2 in relation to reference R . While camera C_2 can only perceive object O_1 , camera C_1 captures both objects O_1, O_2 from its perspective. In contrast to a single stereo camera, the distance r_1, r_2 between reference R and cameras C_1, C_2 , respectively, are now non-static. The sensors move time-dependently through the environment. By matching C_1 and C_2 , the objects can be spatially captured in a larger area.

5 RESULTS AND DISCUSSION

For the evaluation, we delineate an area on which two avatars moved randomly. The periodic vibrations during walking and randomly looking around in the environment are also simulated. Both avatars are equipped with a virtual stereo camera, which made video recordings during the entire measurement.

During this time the user acts as the reference and could move freely around the terrain. Figure 8 shows the marked movement area. It can also be seen that various objects have been installed at different locations around the simulation.

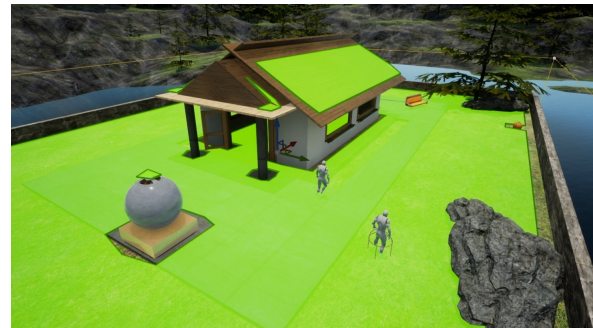


Figure 8: Marked green movement area of the avatars

Care was taken to use different sizes, different textures and reflective, transparent and rough surfaces. For the simulation we used an Intel Core i5 processor (@2.5GHz), 64GB RAM, external GPU1 Nvidia GeForce GTX 1050 Ti and the onboard GPU0 of Intel HD Graphics 630. Since real sensor data differs from synthetic data in its performance, we characterised the acquisition of real-time and dynamic based sensor data and arrived the measurement result in table 2. For measuring the real data we used two stereo cameras of the type ZED2 from Stereolabs and the manufacturer software of ZED2. For the simulation, we recorded test images of both stereo cameras, which contained position data and image data. During the process, the depth images and 3D points were calculated. In comparison to the simulation, we recorded the image and position data for the both ZED2 and also calculated the depth perception and 3D points. On average, the performance between real and synthetic data was not far apart. While the simulated data required a higher GPU (38%), the CPU load for real sensor data was somewhat higher (54%). It was noticeable that real multisensors have a higher sensitivity. In the static state, there is only a slight fluctuation in power. As soon as the stereo cameras were moved or rapid object movement occurred, the CPU load increased up to 78%.

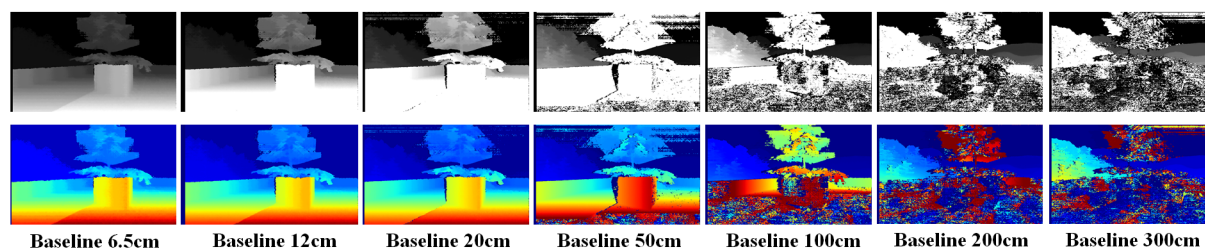


Figure 9: Display of the disparity by increasing the distance between the cameras.

Simulating the motion sequences of synthetic data has shown that the jerky rotations and oscillations caused by the up and down movement influence depth perception.

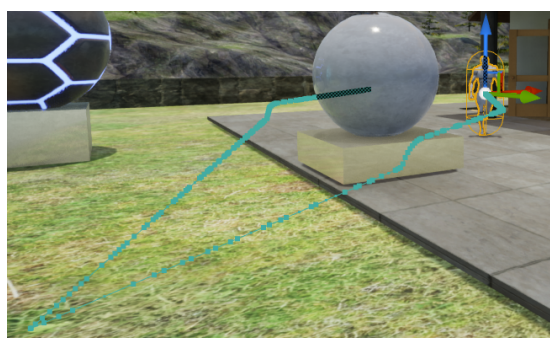


Figure 10: Movement sequence of the stereo camera C_1 .

This showed a speed-dependent change in performance. In contrast, the performance of the synthetic data was not influenced by movement sequences, but there was a strong fluctuation in performance due to several processes running simultaneously.

Basically, the technology works in soft real time. The more visualisations are based on real time, the further one moves away from real time. However, there are ways to achieve real-time processing. During the measurements, noticed that the real-time processing of dynamic sensor data depends on a number of factors. The speed of processing depends on the platform. A scientific platform like Python is significantly slower than a low level based platform like C/C++. Furthermore, a better result can be achieved by using other hardware resources. When using embedded systems, this means that a master-slave system could be essential. In the case of a computer, a GPU would be a good choice. Software technical adaptations, such as parallelising work processes as well as compressing and timing algorithms, produce better results in real time processing. In future Work, we want use the methodologies to reach hard real time.

Figure 10 shows a random movement pattern of the stereo cameras where a pointed inlet of the movement line is visible.

We found that the simulated movements were too jerky. It leads to distorted images and consequently to reduced depth perception. Fast movements and natural vibration are also present in real stereo cameras, but the images are already stabilised by firmware. Thus, for the simulation we need to stabilise the cameras or dampen the movements. Due to the random movements, the cameras were only a few moments in right position for the matching.

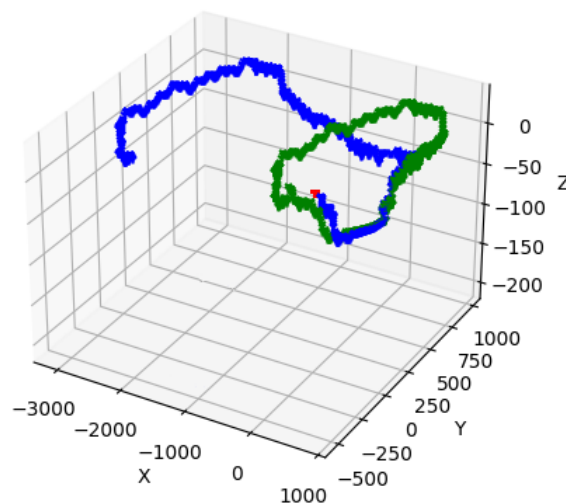


Figure 11: Movement sequence of the stereo cameras (C_1 - Blue) and (C_2 - Green) in dependency of reference (R - Red).

In addition, we analysed the disparity when the distance between the cameras was increased. Figure 9 shows the results of the changed baseline between the cameras. For better illustration we show the disparity in grey (upper row) as well as in colour (bottom row). We increased the distance from 6.5 cm to 300 cm. To ensure comparability, the cameras were arranged parallel to each other. The figure shows that the increase in distance leads to a higher dispersion. From a baseline of 50 cm, a scattering in the near range becomes visible. With the further increase in distance, the dispersion increases from near to far range. Since we kept the parameters constant, it is unclear whether parameter adjustment would lead to a better result.

It can also be seen that the far range in the disparity becomes more visible with increasing distance. There is a proportionality between the baseline and depth visibility of the 3D points.

We conclude that the position matching worked well where no incorrect positions were transmitted. The result of the movement pattern of the both stereo cameras and the reference is shown in figure 11. The illustration shows that the stereo cameras are only initially close to each other. Since there is no equally identifiable geometry with increasing distance, we come to the following consideration:

The synchronous implementation of a dynamic multi-sensor systems is only possible for a limited distance. As soon as the sensors are too far apart, we lose the connection to the object. An expanding approach is to implement the asynchronous matching of dynamic multi-systems. For this implementation it is necessary to store relevant images with the timestamp and position in a map. By using an AI, the best positioned images could be retrieved and matched with the images from the stereo camera.

6 CONCLUSIONS AND FUTURE WORK

In this paper we presented the concept of the synchronous and dynamic use of multiple stereo cameras. Our motivation was to improve challenges such as faulty surfaces and textures, physical influences such as light conditions, reflection, surface textures and also weather conditions with the help of the expansion of sensors. For this purpose, we use a simulation environment that contained both static and dynamic geometries as well as different textures and physical properties. We used stereo cameras as test sensors, each fixed to avatars from the simulation environment.

Dynamic matching of sensors is a promising technique that can be used to optimise point clouds. We have discovered that the synchronous implementation can only be realised for limited distances because the objects are no longer visible as the distance increases. Since we consider the approach of asynchronous dynamic sensor matching, we will continue our research work in this direction. It had already been shown that there is a working range in which the distances can be varied without getting a worse result. However, the working range can be adjusted by modifying the parameters of point cloud processing. In our future work we will build on this knowledge and carry out the parameter modification. For this purpose, we will use an AI that recognises geometries for initialisation as well as stored images of different positions. We will further optimise the simulation and make adjustments in the area of data storage and data processing. In

addition to AI, we will use an intelligent database to map images and geometries. We want to identify the minimum and maximum dynamic distance for sensory perception of objects. Our measurement has already shown that objects become smaller or no longer visible as the distance increases. From this conclusion, we want to research the dynamic systems behave with overproportionally large objects.

7 REFERENCES

- [Ayk18] Aykut T., Burgmair C., Karimi M., Xu J., Steinbach E., Delay Compensation for Actuated Stereoscopic 360 Degree Telepresence Systems with Probabilistic Head Motion Prediction. 2018 IEEE Workshop on Applications of Computer Vision (WACV), <https://ieeexplore.ieee.org/document/8354326>
- [Cha11] Chang C.H., Liang, C.K. and Chuang, Y.Y., Content-aware display adaptation and interactive editing for stereoscopic images. 2011 IEEE Transactions on Multimedia, 13(4) pp. 589-601., DOI: 10.1109/TMM.2011.2116775
- [Chr16] Christian B., Pucker N., Mathematische Methoden in der Physik. 2016 Springer, DOI: 10.1007/978-3-662-49313-7
- [Hir11] Hirschmüller H., Semi-Global Matching Motivation, Developments and Applications. 2011 Photogrammetric Week 11 pp. 173-184., <https://elib.dlr.de/73119/>
- [Ike14] Ikeuchi K., Computer Vision. 2014 Springer. p 249., DOI: 978-0-387-30771-8, <https://www.springer.com/de/book/9780387307718>
- [Kad14] Kadambi A., Bhandari A., Raskar R., 3d depth cameras in vision: Benefits and Limitations of the Hardware With an Emphasis on the First and Second Generation Kinect Models. 2014 Springer pp. 3-26, https://web.media.mit.edu/~achoo/tr/3d_benefits_limits.pdf
- [Liv12] Livatino S., Banno F., Muscato G., 3-D Integration of Robot Vision and Laser Data With Semiautomatic Calibration in Augmented Reality Stereoscopic Visual Interface. 2012 IEEE Trans. on Industrial Informatics, <https://ieeexplore.ieee.org/abstract/document/6062673>
- [Mar09] Marton Z.C., Rusu R.B., Beetz M., On Fast Surface Reconstruction Methods for Large and Noisy Point Clouds. 2009 IEEE international conference on robotics and automation pp. 3218-3223, <https://ieeexplore.ieee.org/document/5152628>

- [Ori18] O’Riordan A., Neue T., Dooly G., Stereo Vision Sensing: Review of existing systems. 2018 12th International Conference on Sensing Technology (ICST), <https://ieeexplore.ieee.org/document/8603605>
- [Pia13] Piatkowska E., Belbachir A.N., Gelautz M., Asynchronous stereo vision for event-driven dynamic stereo sensor using an adaptive cooperative approach. 2013 International Conference on Computer Vision Workshops (ICCV Workshops) pp. 45-50, <https://ieeexplore.ieee.org/document/6755878>
- [Rob20] Roboception GmbH, Stereo-Matching. <https://doc.rc-visard.com> <https://roboception.com/de>, last visited January 13th 2021
- [Sid03] Siddharth A., Ankit V., Gaurav P., Wayne W., Kourous H., McBride J., Ford Multi-AV Seasonal Dataset. 2003 Cornell University, <http://arxiv.org/pdf/2003.07969v1>
- [Ska20] Skala V., Karim S.A.A., Kadir E.A., Scientific Computing and Computer Graphics with GPU: Application of Projective Geometry and Principle of Duality, International Journal of Mathematics and Computer Science. 2020 International Journal of Mathematics and Computer Science, Vol.15, No.3, pp.769-777, <http://afrodita.zcu.cz/skala/publications.htm>
- [Spa17] You Z. - Space Microsystems, Micro/nano Satellites, Inertial Measurement. 2017 Sciencedirect, <https://www.sciencedirect.com/topics/engineering/inertial-measurement>, last visited January 13th 2021
- [Ste20] Stereolabs - ZED2, <https://www.stereolabs.com/zed-2>, last visited January 13th 2021
- [Tow19] Fernandes A., Maheshkumar H., Yang K., Sijo V.M., Vinayagam T., 3D Object Detection for Autonomous Vehicles. 2019, <https://towardsdatascience.com/3d-object-detection-for-autonomous-vehicles-b5f480e40856>, last visited January 13th 2021
- [Ue21] Epic Games Inc. - Unreal Engine, <https://www.unrealengine.com/en-US>, last visited January 13th 2021
- [You13] You L., Ruichek Y., Cappelle C., Optimal Extrinsic Calibration Between a Stereoscopic System and a LIDAR. 2019 IEEE Trans. on Instrumentation and Measurement 62(8) pp. 2258-2269., DOI: 10.1109/TIM.2013.2258241
- [Zha16] Zhang Z., Epipolar Geometry. 2016 Springer, https://doi.org/10.1007/978-0-387-31439-6_128
- [Zie16] Zienkiewicz J., Tsiotsios A., Leutenegger S., Davison A., Monocular, Real-Time Surface Reconstruction using Dynamic Level of Detail. 2016 IEEE Fourth International Conference on 3D Vision (3DV) pp. 37-46. (3DIM-PVT), <https://ieeexplore.ieee.org/document/7785075>

