2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Multi RGB-D Camera Setup for generating large 3D Point Clouds

Wim Lemkens[1*], Prabhjot Kaur[1*], Koen Buys[2], Peter Slaets[1], Tinne Tuytelaars[3] and Joris De Schutter[2]

Abstract— **The advent of inexpensive RGB-D cameras brings new opportunities to capture a 3D environment. This paper presents a method to create a modular setup for generating a large 3D point cloud, with attention to the study of interference, the influence of a USB extension cable, and the calibration procedure. The study of interference includes the influence of the distance between the cameras, the orientation of the cameras, and the illumination. Furthermore, this paper proposes a number of evaluation metrics for similar setups.**

I. Introduction

Classical motion capture rooms like the Vicon [1] and the Optitrack system [2] are typically expensive to purchase or intrusive and also impose a vendor lock-in. These systems are designed for high accuracy applications, such as biomechanical gait analysis [3]. This accuracy is however not needed for many robotic applications. Therefore, other solutions using less accurate measurement systems seem promising.

Recently several of RGB-D cameras have been developed such as the Kinect [4] and the Xtion Pro Live [5]. As these devices provide depth data as well as correlated color images, these capturing environments can be used in a wide variety of applications such as motion capturing and 3D scanning.

Because of the properties of those cameras, setting up a capturing environment can be challenging however. An overview of the challenges and the steps needed to create a working setup are presented in this paper and can be used as a starting point for similar setups.

The purpose of this study is, to create a modular multiple RGB-D cameras setup, that allows to generate large 3D point clouds. The interference problem between the different cameras, the influence of the USB extension cables and the calibration procedure are also discussed.

This paper is divided into several sections. First, section II gives an overview of the related work. The proposed approach is described in section III. The performed experiments and their results are discussed in section IV and section V presents the achieved objectives.

II. Literature Study

Originally only a limited set of specifications were released for the PrimeSense-sensor [6] based devices (e.g. Kinect, Xtion Pro). Therefore a number of people started evaluating its performance, in both single RGB-D camera setup as well as in multiple RGB-D camera setups. However, to the best of our knowledge no study of an elaborate setup used for large room motion capture has been performed to date. In this section the existing multi-camera approaches, the preceding works related to the interference problem, the basics of the calibration procedure, and the camera specification are discussed.

Existing multi-camera approaches

Svoboda et al. [7] proposed an algorithm for the automatic calibration of multi-RGB-camera systems based on rank-four factorization and Euclidean stratification.

Another existing multi-camera approach consists of calibrating a robot's multiple actuators and its sensor suite. Pradeep et al. [8] proposed a flexible framework that allowed the calibration of multiple cameras at the same time. The framework is inspired by the bundle adjustment approach of Triggs et al. [9]. By the use of bundle adjustment, jointly optimal 3D structure and viewing parameters (like camera pose and calibration) estimates can be produced. Kelly et al. [10] introduced a general framework for temporal calibration of multiple sensors (proprioceptive and extroceptive sensors). They determined an accurate estimate of the relative time delay between sensor data streams.

Interference

An aspect that has to be taken into account when using a multiple camera setup is interference. When using multiple structured light depth cameras together, the light pattern of the different cameras can interfere with one another causing incorrect depth measurements for the commonly visible areas.

Sumar et al. [11] studied the effect of interference. The problem is the biggest if both cameras are pointed in the same direction. Interference is expected to be one of the main influences in the accuracy of the system.

A proposed solution by e.g. Schroder et al. [12] is putting a shutter in front of the projector to alternate the projected structured light between the different cameras. One disadvantage is that the shutters need to be synchronised. A harder problem however, is that the frame rate of the depth image becomes divided by the number of cameras that could cause interference. For a setup containing more than two RGB-D cameras, this is not a viable option.

As disabling the projector by software is slow and not available for all drivers, Faion et al. [13] proposed disabling the projector by modifying the electronics of the device. This solution is cleaner than the shutters, as it does not need any moving parts, but still has the disadvantage of reducing the frame rate of the depth map. Moreover due to the switching a dynamic offset in the depth image is introduced that is

*Equally contributed
[1]Department of Industrial Sciences Groep T, Leuven, Belgium
[2]Department of Mechanical Engineering, University of Leuven, Belgium
[3]Department of Electrical Engineering, University of Leuven, Belgium

978-1-4673-6357-0/13/$31.00 ©2013 European Union

Field of view	58°horizontal, 45°vertical, 70°diagonal
Frame rate	30 frames/second
Resolution	640x480
Colour value	32 bit
Depth value	16 bit (from which 11 bits are used)

TABLE I: Table with camera(Asus Xtion pro live) specifications

auto corrected by the camera. This auto correction takes a number of frames time, lowering the total framerate.

Butler et al. [14] tested another solution. To solve the interference problem, the RGB-D cameras were vibrated by fixing a motor on them, an unbalance is realized by an offset-weight. Every RGB-D camera has its own projector rigidly attached to it and thus its light pattern is moving in sync with its own IR camera. Because of the relative motion of the other RGB-D cameras however, the other light patterns will have motion blur and will not be detected as points in the pattern.

Calibration

The IR and the RGB camera of the RGB-D camera have to be calibrated separately. However, K. Berger et al. [15] proposed a method to calibrate them together by the use of a mirror checkerboard. Here, the black squares on the standard checkerboard are replaced by mirrors, which will be detected by the sensors as far away.

Chow et al. [16] suggested a calibration method for the Kinect with three orthogonal textured planes and compared the performance with a single textured plane.

Specifications

Sumar et al. [11] analysed quantitatively multiple Kinects on a single computer and their depth sensors capabilities. This study provides many camera specifications that can be seen in Table I.

Another aspect they investigated was the USB-connection between a camera and a computer. They found out that the number of RGB-D cameras that can be connected to a computer is equivalent to the number of USB host controllers available on the computer, since the datastream of a RGB-D camera almost completely uses all the bandwidth available on a USB controller. Although USB3.0 should be backward compatible with USB2.0, connecting it to USB3.0 will fail to read out the camera.

Furthermore, Khoshelham [17] and Kramer et al. [18] did some accuracy analysis of the depth data on a Kinect. They concluded that the density of points in the point cloud decreases with increasing distance to the IR camera because the depth resolution is low at large distances. It is seven centimetre at the advised maximum range of five meter with an additional four centimetre noise on the measurement. Olesen et al. [19] did a through analysis of the location of the noise within the depth map of a single Kinect and compared the quality of the depth map of a scene viewed by one, two and three Kinects simultaneously.

Primesense introduced its next generation embedded depth sensors, with the carmine 1.09 and the capri 1.25 sensor. It is used for improved algorithms like the multi-modal 3D sensing techniques [20].

Microsoft [4] has announced its new XBox One that will be shipped with a new Kinect. The Kinect will use a 1080p wide-angle camera but will have smaller range.

III. Material and Methods

The original aim of our setup was to evaluate if it is possible to track a quadrocopter drone [21] during flight and to evaluate if the data was sufficient to be incorporated in the control loop. This would allow avoiding expensive motion capture setups, because only centimeter accuracy is needed in the global coordinate frame and the local control loop would be solved with on-board visual servoing.

We believe however that the presented setup can be useful in other applications as well where a large area needs to be monitored and interpreted. This can be personal robotics, elderly monitoring, security, device tracking, and so on. In order to reassure the generality of the setup to accommodate other applications, broad requirements were set and defined here after. This section discusses the requirements to be met, the geometry of the setup, the scalability of the setup, the used USB-cables, the interference problem, and the calibration procedure.

Requirements

Several requirements have to be met to create a useful multiple camera setup:

1) *The interference* should be minimized to ensure the accuracy of the depth measurement.
2) The monitored *volume* should be as big as possible with as few cameras as possible.
3) The setup should be *modular* so it can be scaled.
4) Workload of the image processing should be *distributed* over multiple computers. If there is only one USB-controller per computer available. Only one camera can be connected per USB-controller [11].

Requirement 1 is addressed by putting a vibration on the camera [14] and by placing the cameras such that the camera planes are almost perpendicular. Requirements 2 and 3 are fulfilled by creating modules containing four cameras, each on a corner of a square as shown in Figure 1. Requirement 4 can be addressed by using one computer for each camera. This way, the work can be distributed. This causes a decrease in network load if the data can be reduced sufficiently before sensor fusion occurs.

Geometry

In this setup, the capturing environment is located in a room of ten meter on five meter, with the ceiling at three meter height as seen in Figure 3b. It contains twelve cameras divided in three modules. Each module contains four cameras on the corners of a square as shown in Figure 1. Figure 2 shows the frustums of the twelve cameras with every module represented in an other color. The cameras were mounted on the ceiling (as seen in Figure 6). All cameras in a module are approximately five meter apart. The computers for the data

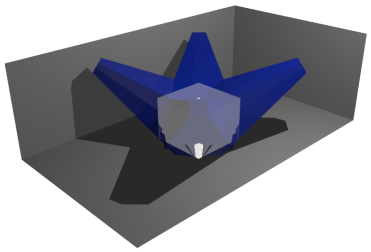


Fig. 1: Viewing frustums of the RGB-D cameras in a single module of the setup. The useful volume is in the center highlighted.

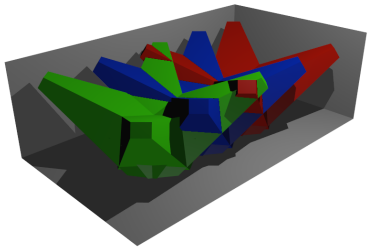


Fig. 2: Viewing frustums of the RGB-D cameras in a setup with three modules.

analysis are mounted in a rack in one corner of the room as seen in Figure 3a.

This setup was chosen to reduce the amount of interference. The tests in Section IV show that the interference is worst when two parallel cameras are at approximately the same distance. As shown in Figure 4, in this setup however, at no point in the frustums of two parallel RGB-D cameras, that point will be at the same distance from the two cameras. At the point where the difference in distance to the two cameras is the smallest, the difference is still about one meter (for a measurement range of five meters).

The overlapping frustums of one module creates a cube with a ground plane of 1.8 m x 1.8 m and a height of 1.5 m, it is presented as a white square in Figure 5. For more precise capturing, the cameras can be mounted closer together as the resolution decreases with the distance to the camera.

(a) The server rack containing the computers for data analysis.

(b) The room in which the setup is located.

Fig. 3: The location of the setup.

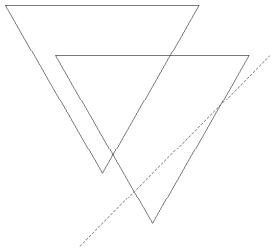


Fig. 4: The frustums of two parallel cameras in the setup. The dashed line marks the equidistant points to the cameras. It is clear that there will never be a point in both frustums that is equidistant to both cameras as the line diverges from one of the frustums.

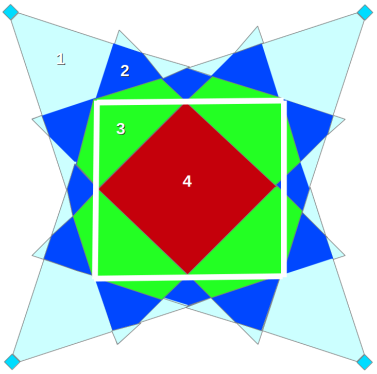


Fig. 5: Top view of one module of four cameras - The area covered in the white square displays the useful area of one segment. The areas coloured in dark blue (2) present the areas where the frustum of two cameras overlaps with each other, the areas in green (3) are presenting the overlapping of three cameras and the red (4) coloured square represents the overlap of the four cameras.

Scalability and USB-extension cable

The modularity of the setup allows for a scalable setup. The complexity of the setup scales with the captured volume. In order to monitor twice the volume, twice the hardware including computing power is needed. The volume can be extended by adding modules.

Because the motion capturing setup is in a room of ten meter by five meter and some RGB-D cameras are in the corners of this room, at least 15 m of USB cable is needed for the camera in the corner distant from the computer. All the computers are placed together in a rack in the corner of the room. Because the USB-cable of the camera has a length of approximately one meter, in our setup, the connection between a camera and a computer is made by the use of active USB extension cables. Every RGB-D camera has its dedicated computer to process its data. If every computer could be placed close to the camera, there is no need of an USB extension cable. In our setup however, it was impractical to have the computers spread across the room. As detailed in SectionIV, the cables used for this setup allow for a total cable length of 36 m ($3 \times 12m$), without including the cable of the RGB-D camera itself.

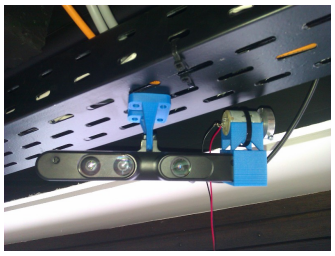


Fig. 6: The RGB-D camera and attached motor to reduce the interference.

Interference

To reduce the interference, the cameras were vibrated by fixing a motor with an unbalance realized by an offset weight [14]. To every device a DC motor was attached as shown in Figure 6. As Butler et al. showed, a vibration of more than 30 Hz was enough to significantly reduce the noise. No exact speed control was needed. Inexpensive DC motors provided enough speed and vibration to reduce the interference. All motors were connected to a single power supply and had a frequency of about 150 Hz.

The vibration should have an amplitude big enough to cause motion blur in the image of the other RGB-D cameras, but small enough to prevent motion blur in the image of the RGB-D camera itself. The amplitude is influenced by the weight, the rotation frequency of the motor, the mounting method and the rigidity of the surface the RGB-D camera is attached to.

The 3D-printable mount for the Xtion and for the motor can be found at the accompanying website [22].

Calibration

Two different kinds of calibration were applied: intrinsic calibration and extrinsic calibration. For both calibrations we used the standard calibration procedure.

1) The intrinsic calibration: The intrinsic calibration is done for every camera (RGB and IR) on every RGB-D camera. This calibration allows the focal length, principal point and the camera distortion (radial and tangential) to be estimated enabling a non-deformed image from the camera. A checkerboard is held at several positions in front of the camera. A Harris-Stephens corner detector [23] was used to detect the corners of the squares on the board. The intrinsic properties were calculated with an algorithm based on the work of Zhang [24]. In theory all cameras could be calibrated at the same time. However, simultaneous streaming of RGB and IR data from a PrimeSense device is not possible. Therefore, it is impossible to do a calibration of both RGB and IR cameras at the same time and the calibration has to be done in at least two stages. As most indoor lighting (LED, fluorescent,...) doesn't contain much IR light, it can be useful to add lighting for the IR calibration that introduces IR light like incandescent light to improve the image quality.

Extrinsic calibration will estimate the relative positions and orientations between cameras. In this case two kinds of extrinsic calibration are needed; the extrinsic calibration of each RGB-D camera on its own and the extrinsic calibration between the different cameras.

2) A. The extrinsic calibration for each RGB-D camera on its own: For each RGB-D camera on its own, the RGB image should be calibrated with the IR image in order to overlay the color image on the depth image that is generated by the device. First, the corners on a checkerboard are detected with a Harris Stephens corner detector [23] followed by the usage of sparse bundle adjustment as described by Triggs et al. [25].

As the RGB and IR images can not be streamed at the same time, a recording has to be made for one stream and has to be played back together with the live stream of the other image stream. As the checkerboard needs to be in the same position for the recorded stream as well as the live stream and the checkerboard cannot be visible in all the RGB-D camera views it is not possible to calibrate all twelve RGB-D cameras at the same time. The position of the RGB camera relative to the IR camera will be fixed as both are contained in the same device.

B. The extrinsic calibration between the different RGB-D cameras Each RGB-D camera also needs to be calibrated relative to the other RGB-D cameras, to be able to create one complete point cloud. A coloured point cloud can be generated from the depth map and the RGB image.

The used software [26], [27] only allows for calibration of two cameras at a time. As there will be one reference camera, and the relative position of every other camera has to be determined, for a total of n cameras, at least $n-1$ calibrations will be needed. Similar to SLAM [28] loop closure also is a problem here. The scene itself is not large, but the distance between different camera views is large. Therefore the calibration can be improved by capturing more relations between the cameras and combining them for instance by averaging the position coordinates and using linear spherical interpolation on the orientation.

Cost

An important factor for the choice of this setup was to reduce the cost of the system. A comparison with the Vicon system that is currently in use is made. The Vicon setup consists of 8 T10S cameras and a computer controlling the setup. The RGB-D setup based on PrimeSense devices in this comparison also consists of 8 cameras. As noted before, one computer is used per camera. But these computers can be relatively small. Table II shows that if the limitations of the RGB-D setup are within the requirements of the application the setup can cost up to 25 times less then using a traditional high-end setup.

IV. Experiments

To determine if countermeasures for e.g. the interference are needed, some tests were conducted prior to the creation of the setup. First, the USB-cable influence and interference are discussed. This is followed by tests on calibration and finally the obtained result is discussed.

	Vicon	PrimeSense
8 cameras	137 000	1200
8 mounts	2240	160
computer	4500	4000
total	143740	5360

TABLE II: Cost comparison of the important hardware between a Vicon system and a PrimeSense based system. Prices are in Euro. Network is not compared as it depends on the application.

USB extension cable

As the USB standard does not officially allow for extensions longer than five meters, we evaluated the influence of the USB cable in the bandwidth and delay of the signal. The cables that were tested were all twelve meter Roline USB 2.0 Active Extension cables.

First the limit of the number of cables in series was tested. The number of cables was increased one by one and the signal was tested. According to the specifications of Roline it would be possible to cascade up to five cables. During the experiments however, only up to three cables of this type could be cascaded before the signal was lost.

Next, the bandwidth was tested. Three cascaded cables didn't give any notable reduction in frame-rate.

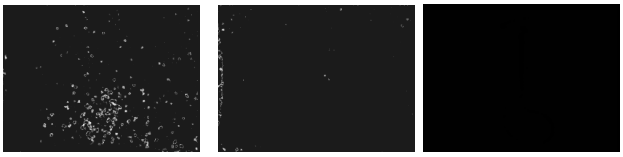
Finally the delay was tested. The clocks of two computers where synchronised. On those computers the stream of two cameras were recorded and timestamped. After a flash was shown in the field of view of both cameras, the recoding was checked for the flash. There was less than one frame difference in the point where the flash became visible. This means that the delay will be less than 1/30th of a second as there were 30 frames per second. This delay however is not necessarily caused by the cable. It can also be explained by the way the operating system kernel handles the USB interrupts or by a small delay in the start of the recording. For the purpose of the motion capture setup this delay is acceptable. A total length of 36 m is available to connect the camera that is the most distant from the computers.

Interference

To quantify the interference, we used an algorithm based on the sensor selection algorithm of Faion et al. [13]. This algorithm expects no changes in the viewed scene and calculates the average variance of the pixels over several frames (60 frames). Eq. (1) shows the calculation of the variance that has to be done for every pixel in the depth map that is received from the RGB-D camera. The average for all the pixels as shown in Eq. (2) is taken as a measurement for the quality of the depth measurements. A larger value indicates more interference.

$$s_{x,y} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(p_{x,y,i} - \bar{p}_{x,y})} \tag{1}$$

$$\bar{s} = \frac{1}{width \times height}\sum_{x=1}^{width}\sum_{y=1}^{height} s_{x,y} \tag{2}$$



(a) Cameras are 0 cm from each other (b) Cameras are 50 cm from each other (c) The right image is a reference with no interference.

Fig. 7: The interference of cameras at different distances. The white spots mark unstable depth readings. More interference can be seen when the cameras are close together (left) then when they are far apart (center). All intensities are on the same scale.

Where:
x and y are the coordinates of the pixel in the image.
n is the number of frames over which the noise is calculated.
$p_{x,y}$ is the pixel value.
$\bar{p}_{x,y}$ is the average value of that pixel over the frames the noise is calculated.
$s_{x,y}$ is the noise of a specific pixel.
$\bar{s}$ is the average noise in the image.
$width$ and $height$ are the width and height of the image.

Figure 7 gives a visual representation of the standard deviations calculated by Eq. (1). The lighter the pixel, the more noise that pixel contains. For every image, the average of the pixels in that image is taken as the measure of the total noise on the depth map.

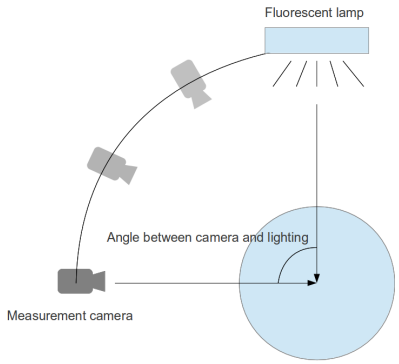
This evaluation metric was used in several tests to determine the factors having the most influence on the interference. The influence of lighting, vibration procedure, distance and angle between the cameras is investigated.

Lighting: The influence on the noise in the depth measurements of the lighting in the room where the setup would be placed was evaluated. The lighting consisted of several warm fluorescent lamps. A camera was put one meter away from a sphere pointing at the sphere. Figure 8a shows the setup for this test where measurements calculated by Eq. (2) were taken with the camera at different angles.

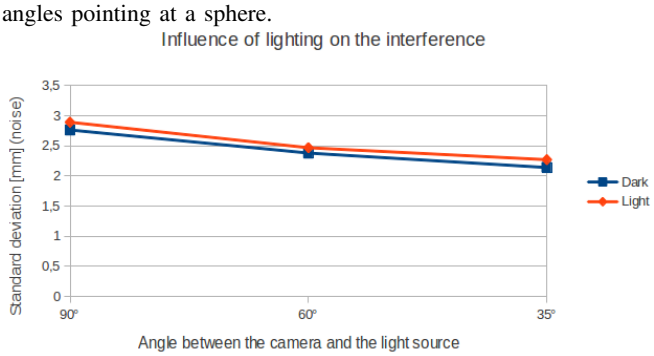
As seen in Figure 8b, there was at no point more then an increase of 1.2 mm in the noise. As further tests point to much greater sources of interference than lighting, lighting is not considered as a problem for the capturing environment.

Distance: In this test the influence of the baseline was tested.

Two RGB-D cameras were mounted above each other pointing in the same direction at a flat surface parallel to the camera plane as shown in Figure 10. One RGB-D camera was doing the acquisitions and another providing the interfering light. By measuring the interference caused by the interfering camera at several positions along a line perpendicular to the surface, the influence of the distance between the RGB-D cameras was measured. Figure 9b shows that the distance strongly affects the interference when the RGB-D cameras are close together. The noise increases if the interfering camera comes closer to the surface.

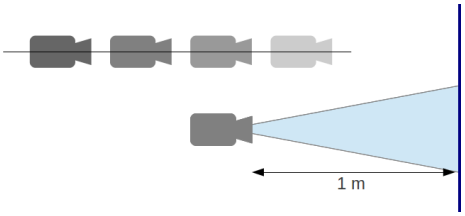


(a) Setup for testing the influence of the lighting on the depth measurements. Measurements were taken with the camera at different angles pointing at a sphere.

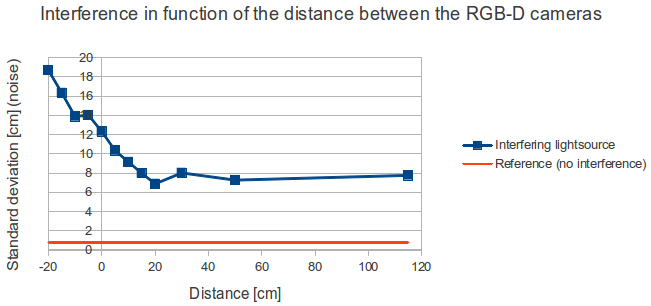


(b) The average noise in distance measurements increases 1.2mm under fluorescent light.

Fig. 8: The influence of the lighting on the depth measurements.

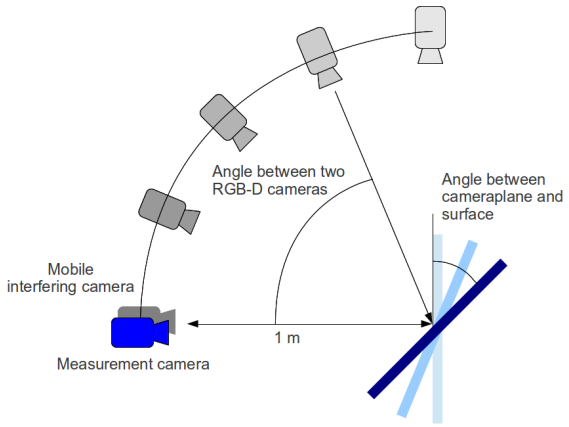


(a) Setup for testing the influence of the distance between the cameras on the interference. A fixed camera measured the noise, while a camera projecting an interfering pattern was moved along a straight line on the z-axis of the cameras.

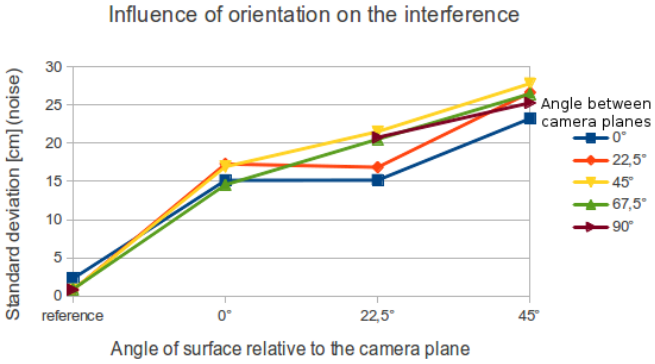


(b) The average depth noise in the image was plotted for different distances between the RGB-D cameras. The distance of the interfering camera to the viewed surface increases with increasing x-values. Above a certain distance, the noise level remains constant.

Fig. 9: The influence of the distance between the cameras on the interference.



(a) The setup for measuring the influence of the orientation of the RGB-D cameras on the interference. The measuring RGB-D camera remains fixed while the interfering RGB-D camera is moved over a semicircle. The surface that is viewed by the cameras is also measured in different positions.



(b) The reference value is measured without interfering RGB-D camera. The angle under which the surface is viewed has more influence over the amount of noise than the relative orientation of the RGB-D cameras.

Fig. 10: The influence of orientation of the RGB-D cameras and the viewed surface on the interference.

Angle between the cameras: The influence of the angle between the cameras was also investigated. The measuring camera and the interfering camera were set on a semicircle pointing at a flat surface at the center of that circle (See Figure 10a). The cameras were set at different positions on the circle while still being pointed at the center.

As seen in Figure 10b, the angle at which the RGB-D camera looks at the surface has more influence on the amount of noise than the angle between the camera-planes has. The reference values in the graph are the measurements without second RGB-D camera. If the surface on which the structured light pattern is projected is viewed under a steeper angle, the noise increases significantly. The angle between the cameras has less influence. Changing the orientation of the RGB-D cameras is not sufficient to prevent interference because for all tested orientations, interference remained an important factor. Eliminating the interference will be important for the capturing environment.

Vibration: To reduce the interference, vibration was applied to the RGB-D cameras [14]. To test the effectiveness

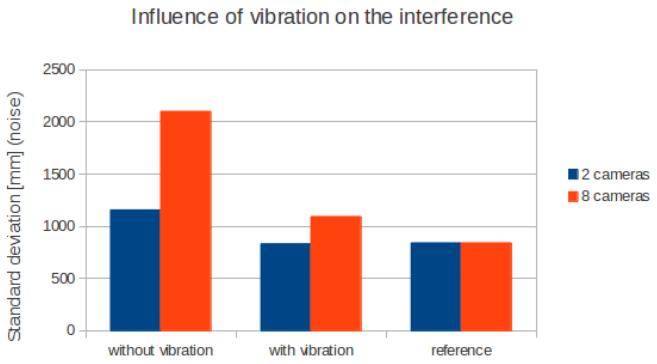


Fig. 11: The influence of vibration on the interference. The vibration caused by the motor reduces the noise caused by interference significantly.

of this method, two tests were performed. In all tests, one RGB-D camera was used to do the measurements. For one test one RGB-D camera was used for projecting the interfering structured light pattern. For the other test seven RGB-D cameras were used to test the effect in two complete modules. The reference measurement was taken without any interfering projectors. All cameras had a motor attached that was rotating at roughly 150Hz. For a quantitative study of the vibration frequency we refer to Butler et al. [14].

Contrary to previous tests, where the average noise was in the order of a couple of centimetres, these tests show an average noise of around a meter as can be seen in Figure 11. This is because the observed scene was a complete room and not a surface close to the cameras. Therefore, some areas viewed by the RGB-D camera were out of measurement range. Invalid measurements are registered as noise of five meter. This does not mean that the whole scene has a lot of noise. It only means that parts of the scene did not result in useful data. Those parts can be detected however and pose no problem for the setup.

For two cameras, the noise level with a running motor is almost the same as the reference, while the noise without a running motor is 35% higher than the reference. Therefore, the noise caused by interference is effectively removed. For eight cameras, not all the noise caused by interference is removed. However, while the noise level without running motor is 150% higher than the reference, the noise level when the motor is running is only 30% higher than the reference. From the experiments it is clear that the vibration causes a significant improvement of the measurements.

Calibration

In tests, the solution of a mirror checkerboard [15] did not work because the measured depth values at the edges of the mirrors were not stable enough for the calibration algorithm to do a proper calibration as seen in Figure 12. Therefore, two separate calibration steps were used for the calibration of the IR camera to the RGB camera as described earlier with the extrinsic calibration in Section III-.2.

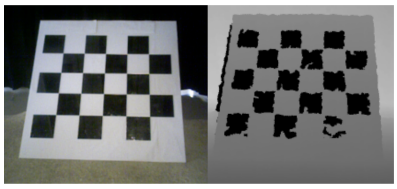
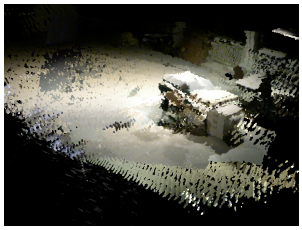
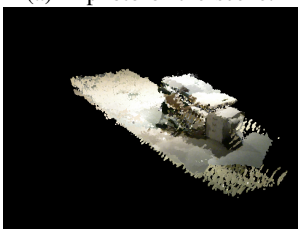


Fig. 12: RGB and depth image of mirror checkerboard used for calibration.

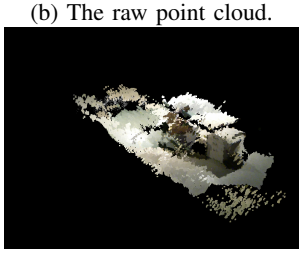
(a) A photo of the scene.



(b) The raw point cloud.



(c) A filtered point cloud of the volume of interest.



(d) A filtered point cloud of the volume of interest without interference reduction.

Fig. 13: The point cloud captured by the system with eight cameras.

Result

The scene shown in Figure 13a was captured. As can be seen in Figure 13b, the raw point cloud will be large and contain a lot of points laying outside the volume of interest. A simple pass-through filter is used to reduce the number of points by filtering out the points that are not in the volume of interest as seen in Figure 13c. The filtering can be done for each RGB-D camera separately on the computer belonging to that camera to reduce network load. As a reference a filtered point cloud is shown without the use of the vibration for noise reduction in Figure 13d.

V. Conclusion

In this report, a modular scalable setup for multiple RGB-D cameras was introduced.

Some tests were performed prior to the creation of the setup to check if the setup would be feasible. The tests showed interference between multiple RGB-D cameras being an important problem. To reduce the interference, each RGB-D camera is equipped with a small DC-motor with an unbalance, realised by an offset weight rigidly attached, to it to create vibration. Interference by indoor lighting is not a problem. The USB extension cables allow a sufficient increase in the distance between the RGB-D camera and the computer for a setup of this size. The capturing volume can be increased by adding modules. For doubling the volume

in the vertical direction, an extra RGB-D camera can be attached close to every standard RGB-D camera. Higher resolutions can be reached by placing the RGB-D cameras lower and closer together in a module.

More information (tutorial, designs, photos, software, videos, ...) can be found at the accompanying website [22].

VI. Future work

In a lot of applications, the point data can be processed and reduced for each camera. The reduced data can then be merged to a sparse dataset. For motion capturing applications in which all point data in the area of interest is needed however, efficient merging of the data should be investigated. As the generated data can be huge octree based algorithms like octomap [29] or megatree [30] will be needed.

No tests were performed on the long term effects of temperature and the vibration on the calibration. As these can be important for a permanent setup, these effects should be investigated.

The intrinsic camera calibration can be automated. A possible method is introduced by Pollefeys et al. [31]. This is a self calibration which can deal with varying types of constraints.

Periodically redoing the calibration of the different RGB-D cameras to each other may be necessary as the RGB-D cameras are placed on non-rigid mounts. The calibration procedure in general can be improved by the use of a laser pointer or a similar bright spot, as proposed by Svoboda et al. [7]. This will lead to a faster calibration procedure of multiple RGB-D cameras.

VII. Acknowledgements

The authors would like to acknowledge the Flemish government for financially supporting the authors.
Koen Buys is funded by KU Leuven's Concerted Research Action GOA/2010/011 Global real-time optimal control of autonomous robots and mechatronic systems, a PCL-Nvidia Code Sprint grant, an Amazon Web Services education and research grant.

We are also grateful to T. De Laet, B. Van Soom and JP. Merckx.

References

[1] Vicon, "Motion capture systems from vicon." http://www.vicon.com/, March 2013.

[2] OptiTrack, "Optitrack - optical motion capture systems and tracking software." http://www.naturalpoint.com/optitrack/, March 2013.

[3] S. Corazza, L. Mündermann, A. Chaudhari, T. Demattio, C. Cobelli, and T. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach," *Annals of Biomedical Engineering*, vol. 34, no. 6, pp. 1019–1029, 2006.

[4] Microsoft XBOX Kinect. xbox.com, 2010.

[5] ASUS Xtion PRO. http://www.asus.com, 2011.

[6] Primesense.LTD, "What is it all about?." http://www.primesense.com/, March 2013.

[7] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *Teleoperators and Virtual Environments*, vol. 14, p. 4, 2005.

[8] V. Pradeep, K. Konolige, and E. Berger, "Calibrating a multi-arm multi-sensor robot: A bundle adjustment approach," in *Int. Symp. on Experimental Robotics (ISER) New Delhi, India*, 2010.

[9] B. Triggs, A. Zisserman, and R. Szeliski, *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21-22, 1999 Proceedings*, vol. 1883. Springer, 2000.

[10] J. Kelly and G. S. Sukhatme, "A general framework for temporal calibration of multiple proprioceptive and exteroceptive sensors," in *Proc. International Symposium on Experimental Robotics, New Delhi, India*, 2010.

[11] L. Sumar and A. Bainbridge-Smith, *Feasability of Fast Image Processing Using Multiple Kinect Cameras on a Portable Platform*. PhD thesis, Department of Electrical and Computer Engineering University of Canterbury Christchurch, New Zealand, .

[12] Y. Schröder, A. Scholz, K. Berger, K. Ruhl, S. Guthe, and M. Magnor, "Multiple kinect studies," *Computer Graphics*, 2011.

[13] F. Faion, S. Friedberger, A. Zea, and U. D. Hanebeck, "Intelligent sensor-scheduling for multi-kinect-tracking," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pp. 3993–3999, IEEE, 2012.

[14] D. Butler, S. Izadi, O. Hilliges, D. Molyneaux, S. Hodges, and D. Kim, "Shake'n'sense: Reducing interference for overlapping structured light depth cameras," in *ACM annual conference on Human Factors in Computing Systems*, 2012.

[15] K. Berger, K. Ruhl, C. Brümmer, Y. Schröder, A. Scholz, and M. Magnor, "Markerless motion capture using multiple color-depth sensors," in *Proc. Vision, Modeling and Visualization (VMV)*, vol. 2011, p. 3, 2011.

[16] J. Chow, K. Ang, D. Lichti, and W. Teskey, "Performance analysis of a low-cost triangulation-based 3d camera: Microsoft kinect system," in *Int. Soc. for Photogrammetry and Remote Sensing Congress (ISPRS)*, vol. 39, p. B5, 2012.

[17] K. Khoshelham, "Accuracy and resolution of kinect depth data," 2011.

[18] *Hacking the Kinect*. Technology in action, 2012.

[19] S. M. Olesen, S. Lyder, D. Kraft, N. Krüger, and J. B. Jessen, "Real-time extraction of surface patches with associated uncertainties by means of kinect cameras," *Journal of Real-Time Image Processing*, pp. 1–14, 2012.

[20] P. Alexandra Crabb, "Primesense unveils capri, worlds smallest 3d sensing device at ces 2013." http://www.primesense.com/news/primesense-unveils-capri/, December 2012.

[21] A. Techonolgies, "Ascending techonolgies quadrotors." http://www.asctec.de/uav-applications/research/products/, March 2013.

[22] W. Lemkens and P. Kaur, "Ros multi rgb-d camera setup." http://rosmultirgbd.wordpress.com/, March 2013.

[23] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. Alvey Vision Conf.*, pp. 147–151, 1988.

[24] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.

[25] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustmenta modern synthesis," *Vision algorithms: theory and practice*, pp. 153–177, 2000.

[26] Willow Garage, "Robot Operating System (ROS)." http://www.ros.org, 2008. Last visited 2012.

[27] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. B. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA Workshop on Open Source Software*, 2009.

[28] J.-S. Gutmann and K. Konolige, "Incremental mapping of large cyclic environments," in *Computational Intelligence in Robotics and Automation, 1999. CIRA'99. Proceedings. 1999 IEEE International Symposium on*, pp. 318–325, IEEE, 1999.

[29] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013. Software available at http://octomap.github.com.

[30] Willow Garage. http://ros.org/wiki/megatree, 2013.

[31] M. Pollefeys, R. Koch, and L. V. Gool, "Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 7–25, 1999.