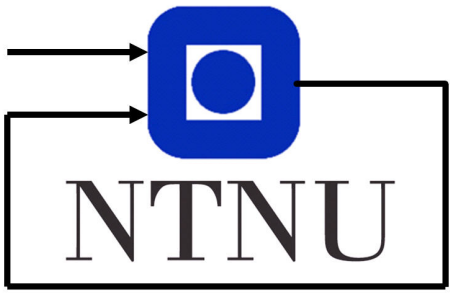
TOF cameras on UAVs for Collision Avoidance and Mapping

Chris André Brombach

Desember, 2018



Norwegian University of Science and Technology
Department of Engineering Cybernetics

TRONDHEIM

1 Abstract

Autonomous flight is highly dependent on visual sensor data to locate possible obstacles and to navigate in an unknown environment. The main purpose of this text is to argue whether time of flight sensors is a good alternative depth sensor to be used on multirotors for localization and mapping of the environment. This will be done by comparing the performance of the camera to an active infrared stereo camera, on a series of different tests.

2 Introduction

This text is written in cooperation with Scout Drone Inspection, a spin-off from the Norwegian University of Science and Technology (NTNU), where their main goal is to develop an autonomous drone to be used for indoor navigation and inspection. The purpose of the autonomous drone is to map the environment and locate regions of interest where there might be defects and signs of degradation [1]. The multirotor must be able to navigate in enclosed and poorly lit environments, while being lightweight, with a minimal use of sensors, to ensure longer flight time.

To safely navigate and map this kind of environment one might make the use of RGBD cameras, i.e. cameras which output a standard RGB picture, and a depth map, which usually is a greyscale picture where the intensity is proportional to the measured distance. Examples of such cameras are stereo cameras, structured light cameras and time of flight cameras. The main drawback with stereo cameras is that they are passive, and they wont work in poorly lit environments without an external light source. Structured light and ToF cameras are active sensors, i.e. they project light onto the environment and estimate depth based on the returning light. These kind of active sensor will be the main target of this paper.

3 Depth sensors

In general there are four different types of techniques used for 3D-imaging; laser scanners, stereo systems, encoded light and time of flight. These can be divided into two groups, active and passive sensors. The passive sensors only make use of the light already in the environment, and is highly dependent on ambient light to work properly. On active sensors, the light source is a part of sensor and light is projected in a specific way such that the sensor is able to generate depth data based on how the light is observed in the camera. Usually the projected light is in the infrared spectrum. A description of the sensors and their working principles will be given below.

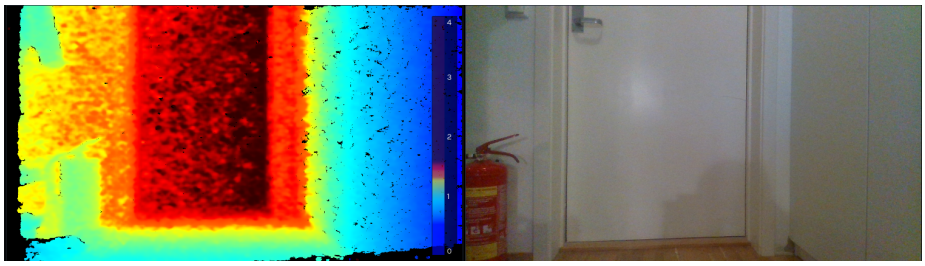


Figure 1: Typical RGB-D output. Depth map on the left and RGB picture to the right

3.1 The pinhole camera model

The pinhole model is based on that light is only able to go through a tiny hole, such that a point in the world only projects to a single point in the image plane, as seen in the figure **??**. This allows us by simple trigonometry to map the 3D points in the world to the image plane.

The common method is to use the virtual image in front of the pinhole instead of the image plane. This allows for a simple mathematical mapping between this image and the word. The relationship between the world and the image coordinates may be described as which will be further discussed in the next section.

3.2 Stereo cameras

Even though active sensors will be scope of interest in this paper, the stereo principle forms the basis for some of the techniques used in structured light systems.

The working principle a stereo cameras is simple in theory, but the generation of a depth map is computationally heavy and in need of advanced algorithms. The main idea is to use two cameras, where the transformation T_{LR} between the camera frames C_L and C_R is known. By finding the same point in both of the pictures, it is possible to calculate the disparity. The disparity is the distance a point have moved from one picture to the next d which is defined as $X_L - X_R$ where X_L is the x-coordinate of the point in the left camera image, and, X_R is in the right. The baseline b is the distance between the two cameras.

If a 3D point in the world is observed in the images of both of the cameras, the distance from the cameras to this point may be calculated. The calculation of depth is given in equation 1, where b is the baseline, f is the focal length of the cameras and d is the disparity.

$$z = \frac{b * f}{d} \tag{1}$$

The complexity of computing the depth lies in finding the disparity d. Both b and f are static parameters which is found by calibration of the cameras.

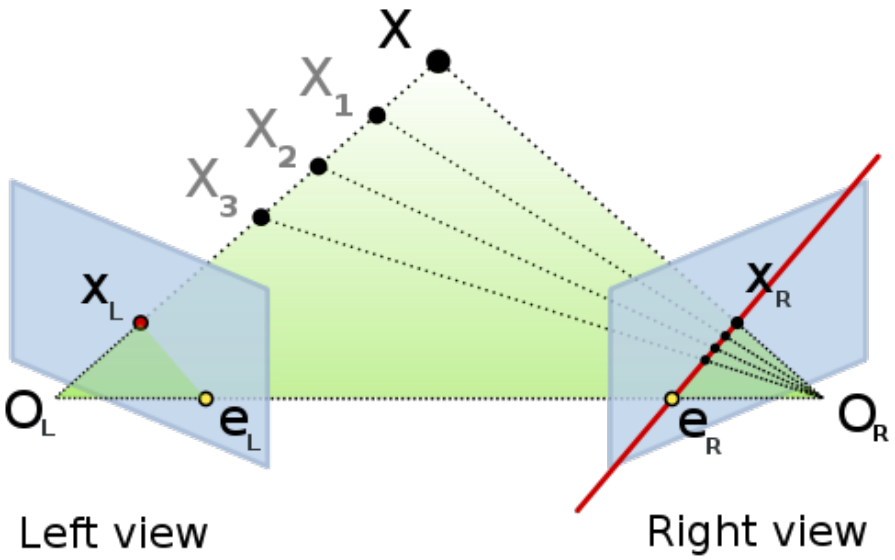


Figure 2: The camera centers and the point in the world form what is called the epipolar plane. As seen in the picture a point in the left image will be along straight line in the right image. Picture from [3]

The main issue is, given a point in one image, how to find the corresponding point in the other image. This problem is called the correspondence problem. A common way to solve this problem is by feature detection and based on the area around the pixel a descriptor is made. Corresponding points may then be found by comparing descriptors and if they are similar a match has been found [2].

To reduce processing time, corresponding points may be found by using epipolar geometry. The focal points in each camera and the point in the world forms what is called the epipolar plane, (FIG). The image points will be along the intersection between the image plane and the epipolar plane. This reduces the problem considerably since a corresponding point will be along a single line in the image plane, i.e. the intersection between the image plane and the epipolar plane.

The major drawback with stereo vision is its dependency on feature points. If the area of interest is without texture, like a single color wall with no distinct elements, the system will have a hard time finding feature points and this will effect the depth estimation. The depth resolution error of stereo systems are proportional to the square of the distance to the target, given in eq. 2 [4]. These kind of systems performs best when the objects are closer to the cameras and

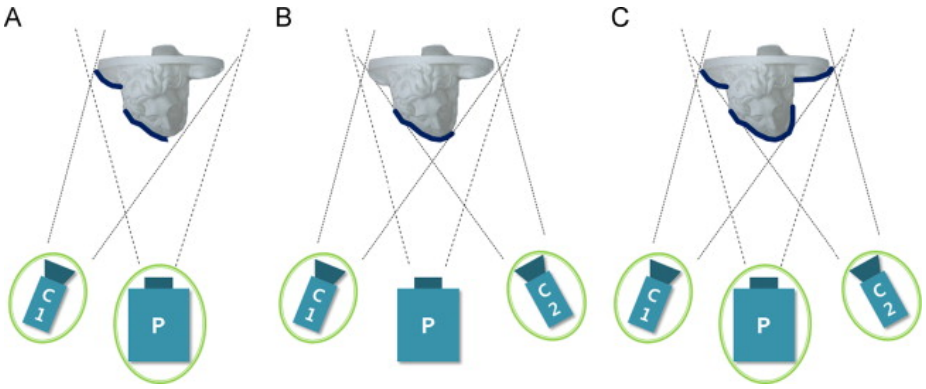


Figure 3: A, B and C is the SLV, ASV and SLS systems respectively, where the blue area is the area where depth information may be extracted for each of the systems.

when the baseline and focal length are large.

$$\Delta z = \frac{z^2}{bf}\Delta d \tag{2}$$

3.3 Structured light

Like in stereo camera systems, the depth information is found by triangulation. To give texture to the scene, structured light systems projects known patterns of light onto the object, and captures the pattern with a camera close to the projector. For standard structured light systems one of the cameras in a stereo system is switched out with an projector. The objects in the scene will distort the projected light pattern, and based on this distortion the depth can be estimated. If several cameras are used, the pattern of the projected light can be used to find correspondences of points in the images, and the depth can be estimated from the disparity just like in a standard stereo system.

Structured light systems are divided into three different systems, structured light vision (SLV), active stereo vision (ASV), and structured light stereo (SLS), [5], [6]. The SLV system is the system described above where correspondences is found between camera and projector. Active stereo vision systems are stereo systems as described in section 3.2, where the projector only adds texture to the scene of interest, and the correspondences between the cameras are found. Structured light stereo systems are the hybrid method where both camera-projector and camera-camera correspondences are used. All three systems are pictured in 3.

There are several different kinds of patterns that might be projected, and there are different methods of how to project them. For high depth resolution sequential projections is used, i.e. several of different patterns of the same pattern type are projected onto the same object. Some examples of patterns

used is binary, gray and phase shift, seen in fig **??**. These kinds of methods have great depth accuracy but only work on static objects, because several pictures are taken to estimate a single depth map, [7].

(a) Binary coded (b) Grayscale (c) Phase coded

Figure 4: Example of projection patterns, pictures from [8]

The structured light technology used in this paper is based on the active stereo vision system where a static pattern of infrared dots, as in fig 5, are projected onto the scene. This kind of system is for all practical purposes a stereo systems as described in section 3.2. This means that all strengths and weaknesses are inherited and the depth resolution of the system is based on the pixel resolution of the camera and the textures found in the picture. The major advantage is that these systems work in poorly lit environments and in scenes with minimal texture.

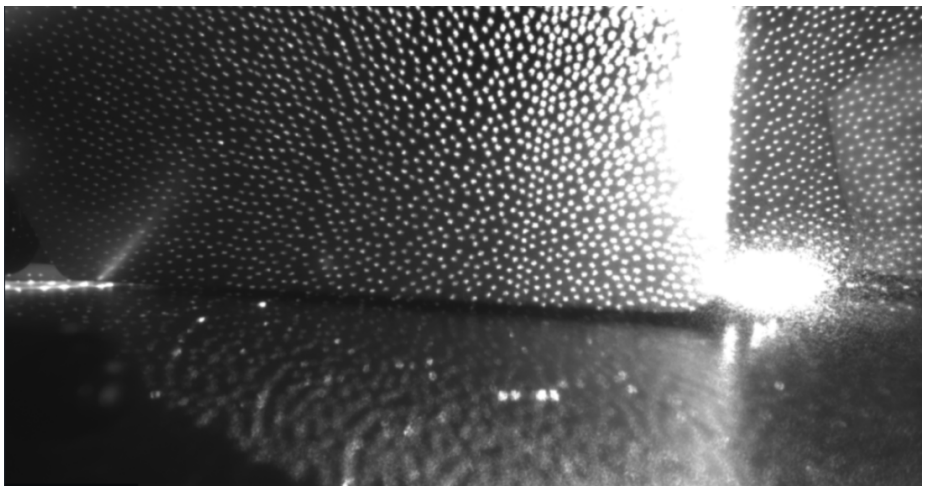


Figure 5: Active stereo IR dot pattern

3.4 Time of flight cameras

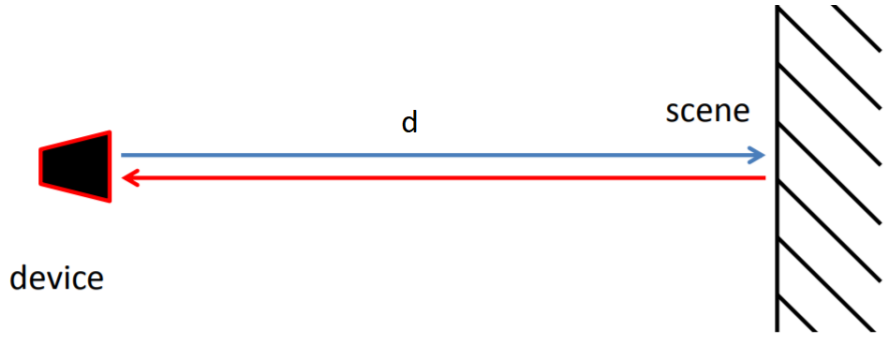


Figure 6: The time of flight principle, [9]

Time of flight (ToF) cameras are based on an entirely different technology than stereo systems and structured light. The main idea is to project infrared light onto the scene and calculate the time it takes for the light to return, hence the name, time of flight. Time of flight systems either use pulse modulation or continuous wave modulation, [10].

In pulse modulation ToF systems the light source is pulsed and the time of flight is measured directly. Since the speed of light is ≈ 0.3 meters/nanosecond the hardware has to be fast if a high resolution depth estimate even should be feasible. It is also important that the arrival time is detected very precisely, just a small offset will have a major impact on the depth measurement. Typically short high powered light pulses are used, with fast rise and fall times. The depth is then given by 3 where τ is the time measured and c is the speed of light.

$$d = \frac{1}{2}c\tau \tag{3}$$

In continuous wave modulation the method to calculate distance is to measure the phase shift of a modulated infrared signal. The infrared light is modulated according to eq. 4, where A_E is the amplitude of the signal, and f_{mod} is the modulation frequency. The reflected signal is given by eq. 5a, where '$\Delta\phi$ is the phase shift of the signal and B_R is the interference from the ambient light at the infrared wavelength. This equation is simplified into eq. 5b where $B = B_R + A_R$ and $A = A_R$. A is called the amplitude and B is called the offset of the signal.

$$S_E(t) = A_E(1 + sin(2\pi f_{mod}t)) \tag{4}$$

$$S_R(t) = A_R(1 + sin(2\pi f_{mod}t + \Delta\phi)) + B_R \tag{5a}$$

$$S_R(t) = Asin(2\pi f_{mod}t + \Delta\phi)) + B \tag{5b}$$

In equation 5b the variables A, B and $\Delta\phi$ are to be estimated. This is done by sampling the signal of the reflected modulated signal four times per

period. Each sample is taken at a phase step by 90 degrees, i.e. for samples Q_0 - Q_3, $Q_0 = S_R(t = 0)$, $Q_1 = S_R(t = \frac{1}{4}f_{mod})$, $Q_2 = S_R(t = \frac{2}{4}f_{mod})$, $Q_3 = S_R(t = \frac{3}{4}f_{mod})$. The variables is found by minimizing the square error between estimated and predicted measurement, given in eq. 6, and the solution is given in equations 7a, 7b and 7c, [9], [11].

$$(A, B, \Delta\phi) = argmin_{A,B,\Delta\phi} \sum_{n=0}^{3} (Q_n - Asin(\frac{\pi}{2}n + \Delta\phi) + B)^2 \tag{6}$$

$$A = \frac{\sqrt{(Q_0 - Q_2)^2 + (Q_1 - Q_3)^2}}{2} \tag{7a}$$

$$B = \frac{Q_0 + Q_1 + Q_2 + Q_3}{4} \tag{7b}$$

$$\Delta\phi = arctan2(Q_0 - Q_2, Q_1 - Q_3) \tag{7c}$$

The distance is then given by eq.8, where c is the speed of light.

$$d = \frac{c}{4\pi\ f_{mod}} \Delta\phi \tag{8}$$

According to [11] the depth measurement variance can be approximated by where c_d is the modulation which is a sensor parameter.

$$\sigma = \frac{c}{4\sqrt{2}\pi\ f_{mod}} \frac{\sqrt{A + B}}{c_d A} \tag{9}$$

Since the phase wraps around every 2π, i.e $sin(t) = sin(t+2\pi)$, the distance will have an aliasing distance, and this ambiguity distance is given by eq. 10. This is also the maximum distance the ToF cameras are able to measure. If the maximal measurement is to be increased, the modulation frequency has to be decreased. However the modulation frequency should be chosen as large as possible since the variance is inversely proportional to f_{mod} as seen in eq. 9.

$$d_{amb} = \frac{c}{2f_{mod}} \tag{10}$$

To reduce the effect of various ToF noise sources, the average of the distance measurements is taken over several measurement periods. The averaging intervals is called the integration time, and is a tuning parameter when using ToF cameras. In practice the interval is between [1, 100] ms, and a high integration time will in general give good ToF distance repeatability [9]. Long integration times may lead to undesirable side effects, like motion blur and saturation.

One of the advantages with time of flight sensors is the intensity image, which is the set of A values from eq. 7a. As seen from time of flight variance eq. 9, the variance of the depth will decrease if A increases. In other words, the intensity image is telling which of the depth values in the depth map that are trustworthy, and which measurements that can be discarded.

3.5 Strengths and weaknesses

In the previous subsections the working principles of stereo systems, structured light and ToF were presented as well as some of strengths and weaknesses were discussed. Here a summary will be given, and form the basis for the experiments in this paper. The table below is based on [11, 12].

Considerations	**Stereo systems**	**Structured light**	**Time of flight**
Range	Medium to far	Short to medium	Short to far
Resolution	Medium	Medium	Medium
Depth accuracy	Medium	High	Medium
Software complexity	High	Medium	Low
Response time	Medium	Slow	Fast
Compactness	Low	Medium	Low
Low light performance	Weak	Good	Good
Bright light performance	Good	Weak	Good

Stereo systems works well with ambient light and may be used both outdoors as well as indoors. The system has descent depth accuracy, especially close to the camera system, but further away the accuracy will decrease with the square of the distance, in accordance with eq. 2. The system is dependent on ambient light, and wont work at all in low light environments without some sort of a light source. It is also dependent on a scene with texture, or the estimated depth map will be quite spare or even nonexistent.

Structured light has the highest depth accuracy and yields a great depth estimate at a given range. There are different types of structured light systems and the performance between them may vary. Active stereo systems, combine the technology of structured light and stereo systems. These systems work indoors and have decent performance outdoors, but the sunlight might be so strong that the projected pattern wont show in the picture. Another important strength is that it works well in low light conditions. Some of drawback are transparent objects, surfaces that scatter the projected light and absorbent surfaces where the pattern is not reflected, which is typical for active sensors. The accuracy of the measurement is also effected by the distance, similarly to stereo systems.

Time of flight cameras will in theory perform well both indoors, outdoors, and in low light conditions. It will be effected by ambient light, where the sensor may saturate and in that case depth estimation becomes impossible. A simple solution is to reduce the integration time, but this will reduce the accuracy of the depth estimate. The system will work in outdoor conditions, but in strong sunlight the accuracy will be effected. ToF systems are fast, and the software is simplistic compared to the other technologies. The modules can be built small and light, and are ideal for systems where weight reduction is important. The range is lower compared to the other systems described, and like structured light it is effected by transparent object, light scattering and absorbent surfaces.

4 Visual sensors on autonomous vehicles

Visual sensors on autonomous vehicles is primarily used for one thing, mapping the environment.

4.1 Obstacle avoidance

Obstacle avoidance is the problem of detecting objects in the environments and in some cases classify the object such that the appropriate action may be taken. To detect objects a monocular camera is necessary, but if the position of the object is to be determined some kind of depth measurement has to be taken. Detection of objects in a camera can be done in many ways, but a simple approach is segmentation by thresholding. The idea is to find the area in an image that contains the object based on the intensity of the object and the surrounding area, .e.g a red balloon on a dark background. A potential problem is that several balloons are side by side, or the contrast to the background is small, then it will be harder segmenting the objects in picture. A possible solution with RGB-D cameras is using the depth map. Since objects typically can be assumed to be a constant distance away from the camera, the color of the object will be uniform as for the objects in fig 11c. This contrast allows for easy segmentation as in the example with the red balloon. By combining thresholding in the RGB image and the depth map the system should be able to separate an object from the rest of the environment without much trouble.

4.2 SLAM

Simultaneous localization and mapping, or better know as SLAM is the problem of mapping an unknown environment as well as estimating the localization of the camera and its path through the environment. The SLAM algorithms can in general divided into two categories, direct and indirect, and these can as well be divided into dense and sparse mapping algorithms, yielding a total of four categories. Dense algorithms use moste of the pixels in the image, while sparse algorithms only use a subset of the pixels. The most typical is direct dense/sparse and indirect sparse. The main difference between a direct and indirect SLAM algorithm is that in the indirect algorithms the image is preprocessed by finding feature points in the image, and descriptors to these point. Based on the movement of these point in one frame to another it is possible to find the transformation between these frames, and thus the movement of the camera.

Indirect SLAM algorithms preprocess the image by finding feature points and corresponding descriptors for these points. Based on the movement of these points in one frame to another, it is possible to find the transformation between these frames, and as following the movement of the camera. The problem is solved as a minimization problem, as described in eq. 11, where $u_i - \pi(T_{cw}X_i^W$ is the geometric error and $\pi()$ is the projection of a point in camera frame to the image like in section 3.1 . The variables to be found is T_{cw}, the transformation

between world frame and camera frame, and X_i^*, which are the feature points found in the picture in the world coordinates. This problem is called full bundle adjustment and with some clever changes it can be solved as a least squares problem and matrix manipulations [2].

$$T_{cw}^*, x_i^* = argmin_{T_{cw}, x_i^*} \sum_i \left\| u_i - \pi(T_{cw} X_i^W) \right\| \tag{11}$$

Direct SLAM algorithms does the processing directly on the intensity of the images, and do not find any feature points or such. The transformation between world and camera is calculated based on the photometric error, which is the error between the current image and the transformed image. The minimization function is given in eq. 12. The typical way to generate the map, as in LSD-SLAM, [13], a depth map is estimated. The map is refined by a high number of baseline stereo comparisons in image regions where the expected stereo accuracy is large

$$T_{cw}^* = argmin_{T_{cw}} \sum_i \left\| (I_i - I_c(\pi(T_{cw} X_i^W))) \right\|^2 \tag{12}$$

RGB-D cameras are typically used with dense SLAM algorithms. The reason for this, as mention in the paragraph above, is that dense algorithms estimate a depth map between frames. The idea is to use the depth map from the RGB-D camera together with the depth map estimate. These two are fused together into a less uncertain depth map of the frame.

5 Depth cameras specification

A short specification of the different cameras used in this paper will be given in the following subsections. The information in this section is based on the official datasheets [14, 15, 16].

5.1 Intel Realsense D435

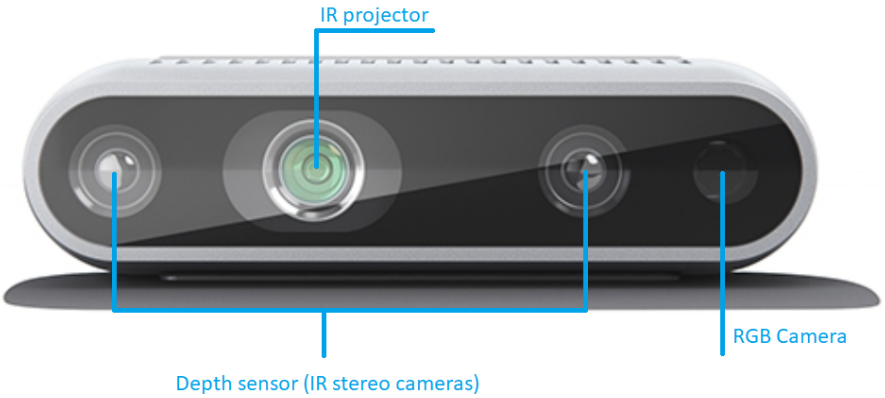


Figure 7: Intel Realsense D435, picture from [17]

The Intel Realsense D435 camera is a active infrared stereo camera. It projects a static pattern in the infrared spectrum to create texture for the IR stereo cameras, as described in section 3.3. The camera can generate a depth map without projecting the pattern onto the scene, but the estimate is improved significantly when it is done. The output from the camera is two IR images (1280x800), RGB image (1920x1080) and a depth map(1280x720). The frame rate of the camera varies with the depth map resolution. With a depth map resolution of (1280x720) the maximal FPS is 30, while the resolution has to be reduced to (480x270) to get 90 FPS. The depth map from D435 can be view in fig 1.

Parameters	**Camera Properties**
Dimensions	90 mm x 25 mm x 25 mm
Weight	72g
Resolution stereo	1280 x 800
Field of view (H x V x D)	91.2° x 65.5° x 100.6°
Resolution RGB	1920 x 1080
Field of view (H x V x D)	69.4° x 42.5° x 77°
Measurement range	0.2 - 10 m
Frame rate	Up to 90 FPS
Illumination	850 nm ±10 nm IR
Depth resolution	$\leq$ 1.5% of distance
Output	Depth map and grayscale intensity

5.2 DepthEye 3D

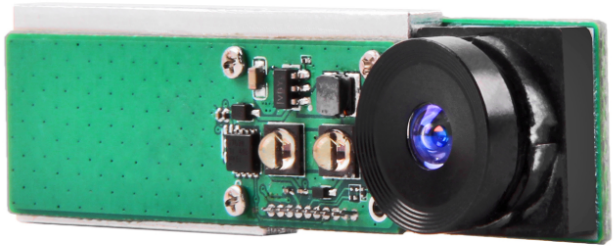


Figure 8: DepthEye 3D module, picture from [14]

The DepthEye 3D is a Time of flight camera based on the continuous wave modulation method described in section 3.4. The output from the camera is a raw depth map, point cloud and the intensity image.The sensor has a IR band-pass filter (820 nm to 865 nm), such that most of the ambient light that might lead to saturation is filtered out. Light with wavelength between 820 nm and 865 nm will have most effect on the sensor, since this spectrum is not filtered out. For indoor use this shouldn'tt be a problem since there are rarely any infrared light sources, but sunlight is expected to have an effect.

Parameter	**Camera Properties**
Dimensions	60 mm x 17 mm x 12 mm
Weight	20g
Camera Resolution	80 x 60
FOV (H x V X D)	75.3° x 59.7° x 90°
Measurement range	0.2 - 2 m
Frame rate	30 fps
Illumination	850 nm IR
Depth resolution	1.5 of distance
Output	Depth map and grayscale intensity

5.3 PMD Picoflexx

Parameter	Camera Properties
Dimensions	68 mm x 17 mm x 7.35 mm
Weight	8g
Camera Resolution	224 x 171
Viewing angle (H x V)	62° x 45°
Measurement range	0.1 - 4 m
Frame rate	5 - 45 fps
Illumination	850 nm IR
Depth resolution	$1\% - 2\%$ of distance
Output	Depth map and grayscale intensity

The PMD Pico flexx is Time of flight camera just like the DepthEye 3D. The output is a depth map, point cloud and intensity map, but the resolution of the camera is significantly larger. The pico flexx has a series of different modes where the frame rate varies from 5 fps to 45 fps, and the corresponding maximum range from 4 meters to 1 meter. According to (picoflexx website) the camera is more robust to ambient light since they use a combination of optical filters, and what is called Suppression of Background Illumination (SBI). The SBI prevents the sensor from early saturation by subtracting the ambient light, which makes the sensor more tolerant to all kinds of ambient light.

6 Experiments

The following subsections will discuss the various experiments and results with the depth sensors. The idea is to test the cameras in a series of smaller tests to map the performance of the sensors in different scenarios. Most of the experiments will be conducted by finding the distance to a uniform white wall.

6.1 Motivation

As described in (the slam section) visual sensors are used to map the area such that a autonomous vehicle can navigate safely in an unknown environment. The motivation for the following experiments is to find areas where time of flight sensors perform well, and it what kind of scenarios the data may be less trustworthy. Mono and stereo cameras are extensively used in the SLAM problems [13], and the idea is to find out how time of flight sensors compare to these kinds of sensors in different scenarios.

6.2 Indoors with and without ambient light

The sensors are primarily constructed for indoor use, and as such the performance is to be tested in this environment. A thorough test will be executed, where the distance error and standard deviation will be measured for different distances from the wall; 0.5, 1.5, 2.0, 3.0 and 4.0 meters. The experiment will be conducted with normal indoor lighting, and without any ambient light.

6.2.1 Results

Ambient light:

Distance	Pico error	D435 error	D435 Std. Dev.	Pico Std. Dev.
0.5	0.00 m	0.00 m	0.004 m	0.001 m
1.0	0.01 m	0.01 m	0.009 m	0.001 m
1.5	0.02 m	0.02 m	0.015 m	0.001 m
2.0	0.02 m	0.03 m	0.028 m	0.002 m
3.0	0.03 m	0.06 m	0.101 m	0.003 m
4.0	0.05 m	0.07 m	0.115 m	0.006 m

No ambient light:

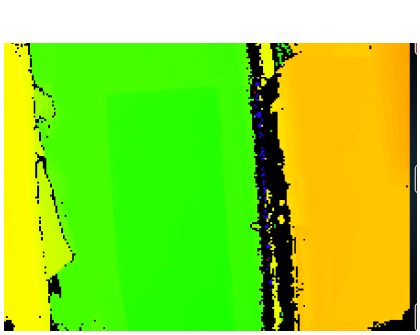
Distance	Pico error	D435 error	D435 Std. Dev.	Pico Std. Dev.
0.5	0.00 m	0.00 m	0.003 m	0.001 m
1.0	0.01 m	0.01 m	0.007 m	0.001 m
1.5	0.02 m	0.02 m	0.017 m	0.001 m
2.0	0.02 m	0.03 m	0.035 m	0.002 m
3.0	0.03 m	0.06 m	0.111 m	0.003 m
4.0	0.05 m	0.07 m	0.125 m	0.006 m

6.3 Pico flexx modes

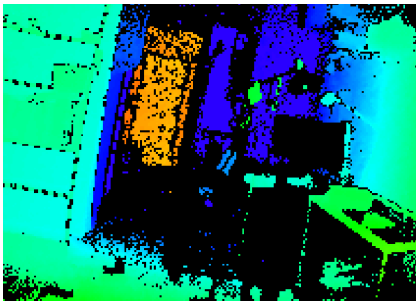
The pico flexx camera has several different modes presented in the table below. It is of interest to find how well the sensors performs in the different modes, the maximal and minimal measurement distance will be found for each case, as well as the standard deviation at each min/max distance. Since the background is suspected to have an effect on the range, two set of backgrounds are used, a white reflective one, and a dark matte background.

Mode	Range(m)	Frame rate	Exposure Time (μs)
1	1.0 - 4.0	5 fps	2000
2	1.0 - 4.0	10 fps	1000
3	0.5 - 1.5	15 fps	700
4	0.3 - 2.0	25 fps	450
5	0.3 - 2.0	35 fps	600
6	0.1 - 1.0	45 fps	500

6.3.1 Results



(a) Minimal salt and pepper noise



(b) High level of salt and pepper noise in the background.

Figure 9: Significant increase in invalid measurements when the camera is close to its maximum range

In the table below the min and max distances are given when the background is white, and the corresponding standard deviations at each point. The min and max values are the distances where the sensors is not able to gather distance data, or when it is a combination of valid and false data, fig 9b.

Mode	Max	Min	Std. dev. max	Std. dev min
1	6.9 m	0.47 m	0.016 m	0.001 m
2	5.5 m	0.35 m	0.017 m	0.001 m
3	4.5 m	0.30 m	0.017 m	0.000 m
4	3.5 m	0.25 m	0.009 m	0.000 m
5	2.4 m	0.25 m	0.008 m	0.001 m
6	2.3 m	0.22 m	0.009 m	0.000 m

The min and max distances with a dark background.

Mode	Max	Min	Std. dev. max	Std. dev min
1	2.9 m	0.1 m	0.09 m	0.001 m
2	2.2 m	0.1 m	0.008 m	0.001 m
3	1.75 m	0.1 m	0.004 m	0.000 m
4	1.5 m	0.1 m	0.005 m	0.000 m
5	1.2 m	0.1 m	0.002 m	0.001 m
6	1.0 m	0.1 m	0.001 m	0.000 m

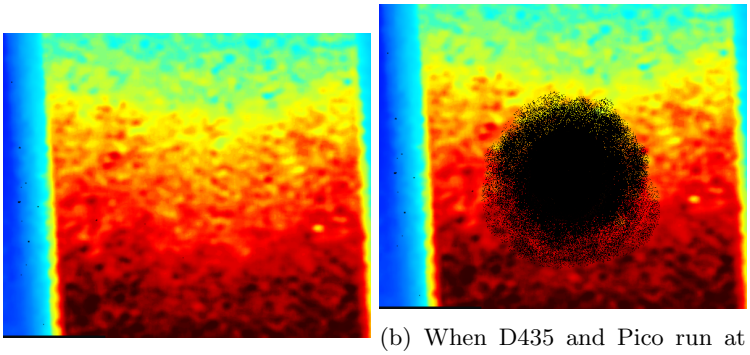
6.4 Several sensors working together

Sometimes it might be useful to have several sensors to measure depth in different directions. Since the sensors are active they may influence each other and

effect the measurement. To test this the three sensors will be run simultaneously, and the data is compared to the case where each sensor is running by itself.

6.4.1 Results

From the test data, time of flight sensors do not seem to be effected by each other or by the infrared pattern projected from the D435. There is no change in the depth map or in the intensity image. The active stereo cameras is greatly effected by the time of flight cameras, 10b, where some of the depth data in the middle of the picture is corrupted. This effect only occurs when the captured frame is taken at the same time as the ToF cameras project light onto the scene.



(a) Normal visual data

(b) When D435 and Pico run at the same time

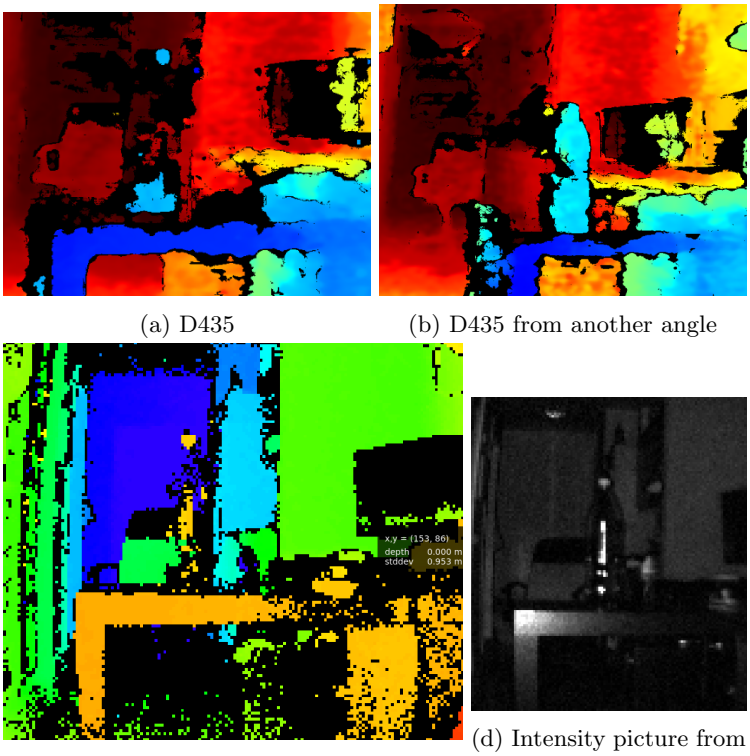
Figure 10: Effect of ToF light in depth map of D435

6.5 Transparent objects

Since the cameras in question is active, one would suspect that they have some trouble detecting transparent object. The purpose of this test is to check if this statement is true, and if there is a distance where the object can be detected.

6.5.1 Results

A simple transparent bottle was used as the test object. The general result is that transparent object is hard to detect in both time of flight cameras and active stereo cameras. The D435 camera do perform a bit better than the time of flight camera, but it highly depends on the point of view, as seen in fig 11a and fig 11b. In fig 11c the bottle is hardly visible, only a vertical line at the center of the bottle is reflected back to the ToF camera as seen in intensity image, fig 11d. Most of the bottle in the intensity image is dark, but a single white line, and this is the area where depth data from fig 11c is any good.



(a) D435

(b) D435 from another angle

(c) Depth map from pico flexx

(d) Intensity picture from Pico flexx

Figure 11: Depth maps of a transparent bottle.

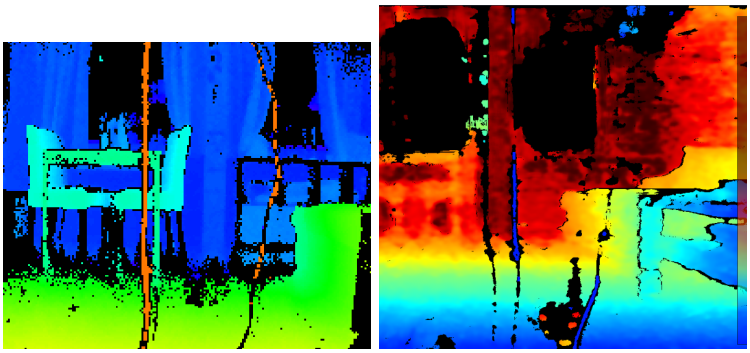
6.6 Thin objects

Some objects might be so thin that they are not detected at a certain distance. To test this two thin wires (3 mm) with different colors will be placed in front of the cameras at different distances.

6.6.1 Results

The distances where the wires disappear from the images from the three sensors are found in the table below. The black wire is not detected at all in the image of DepthEye 3D, and it has the shortest range of them all. This might be expected because of the low resolution of the camera. Dark object are apparently the weakness of time of flight cameras, since the detection range of the black wire is less than half of the range of the white wire. The stereo camera D435 does generally perform better on detecting thin objects, and the color of the object has less influence. The Pico flexx camera is however able to detect the white wire at the largest range.

Camera	Black disappear	White disappear
DepthEye	NaN	0.5 m
Pico Flexx	0.55 m	1.4 m
D435	0.8 m	1.1 m



(a) Pico flexx depth map of the wires. The white wire (left) is clearly the most visible

(b) D435 of the wires. A shadow effect is visible in this picture as both of the wires are duplicated

Figure 12: Depth maps where the wires are visible

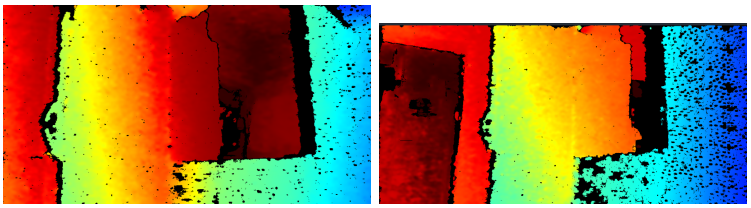
6.7 Sharp angles

If the camera has a sharp angle to a highly reflective surface, the projected light might not return to the camera, but just bounce of the surface. To test and find the angle where this might happen, a mirror is used as the reflective surface, and a set of sharp angles is tested.

6.7.1 Results

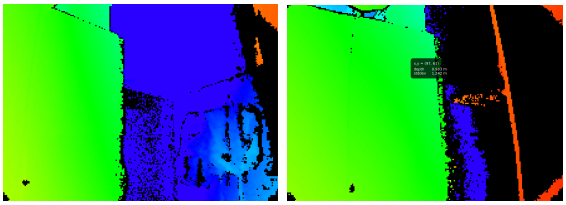
In the images ,fig 13a and fig 13c,the cameras are looking in the direction of the mirror. In fig 13a it is clear that the camera is not able to estimate the correct distance to the mirror, since the camera finds and uses texture that is behind the camera. The same result is seen in fig 13c. where the the green area is the actual depth, and the blue area is the mirror reflecting infrared light from behind the scene.

The depth maps of a sharper viewing angle are shown in fig 13b and fig 13d. There is little change between the cases in fig 13a and fig 13b, and as such the D435 is able to estimate depth at sharp angles. The Pico flexx on the other hand not able to estimate depth at angles $\leq 20°$, that is the projected light is not reflected back.



(a) Frontal view with D435

(b) Sharp angle with D435



(c) Frontal view with Pico flexx

(d) Sharp angle with pico flexx

Figure 13: Depth map of a transparent bottle.

6.8 Discussion

Both of the cameras D435 and Pico Flexx are quite robust in different kind of environments. There is little evidence that ambient light effect these cameras in any major way, and the performance is equally good in no light environments. The major difference between the cameras are the range, where the D435 comes out on top, maximum 10 meters, but the Pico Flexx has significant range too, a maximum distance of seven meters, however this range is dependent on how reflective the scene is. The depth data from the time of flight camera is quite stable with a maximum standard deviation of 0.017 m, where D435 had standard deviations up to 0.125 m. Even though the difference in noise in the data is significant, the mean of the depth data is well within 1 – 1.5 % of groundtruth.

One of the major drawbacks with the time of flight camera is it sensitivity to background color and the amount of reflection from the objects in the scene. As seen in section 6.3 the range of the sensors was cut with a third by changing the background from a reflective white surface to a matte dark one, while the active stereo camera D435 had no issues with this change. In general do the ToF camera perform quite well when the objects has a high degree of reflection, as seen in (wire test) where a thin white 3 mm wire was observable in the image from 1.4 meter.

The time of flight camera is quite robust to ambient light, and no notable negative effect was observed when running the sensors side by side. The D435 has some issues where the IR-camera would detect light flashes from the ToF cameras and in some frames the ir-pattern would become unobservable, effecting the depth estimation.

6.9 Conclusion

Time of flight cameras are performance is on par with established depth sensors. The cameras are quite robust to ambient light and the depth data is stable within given bounds. Based of the test conducted in this paper there is evidence pointing to that time of flight cameras could be well suited as visual sensors for indoor navigation and mapping.

References

[1] S. D. Inspection. Scout drone inspection, front page. [Online]. Available: https://www.scoutdi.com/

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[3] Epipolar geometry. [Online]. Available: https://en.wikipedia.org/wiki/Epipolar$_g eometry$

[4] N. Instruments. (2013) What to expect from a stereo vision system. [Online]. Available: http://zone.ni.com/reference/en-XX/help/372916P-01/nivisionconceptsdita/guid-10d358bd-3dcd-4ccd-a73c-672e48aed39a/

[5] C. Je, S. W. Lee, and R.-H. Park, "High-contrast color-stripe pattern for rapid structured-light range imaging," in *European Conference on Computer Vision*. Springer, 2004, pp. 95–107.

[6] W. Jang, C. Je, Y. Seo, and S. W. Lee, "Structured-light stereo: Comparative analysis and integration of structured-light and active stereo for measuring dynamic shape," *Optics and Lasers in Engineering*, vol. 51, no. 11, pp. 1255–1264, 2013.

[7] H. G. Jung. Structured light projection. [Online]. Available: http://web.yonsei.ac.kr/hgjung/Lectures/AUE859/7.%20Structured%20Light%20Projection.pdf

[8] J. Geng, "Structured-light 3d surface imaging: a tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[9] C. D. Mutto, P. Zanuttigh, and G. M. Cortelazzo. (2013) Time-of-flight cameras and microsoft kinecttm. [Online]. Available: http://lttm.dei.unipd.it/nuovo/Papers/ToF-Kinect-book.pdf

[10] R. Lange. (2006) 3d time-of-flight distance measurement with custom solid-state image sensors in cmos/ccd-technology. [Online]. Available: https://dokumentix.ub.uni-siegen.de/opus/volltexte/2006/178/pdf/lange.pdf

[11] L. Li, "Time-of-flight camera–an introduction," *Technical white paper*, no. SLOA190B, 2014.

[12] T. E2V. (2018) 3d imaging technology - time of flight. [Online]. Available: https://www.azom.com/article.aspx?ArticleID=16003

[13] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European Conference on Computer Vision*. Springer, 2014, pp. 834–849.

[14] S. T. Co. (2018) Deptheye 3d visual tof depth camera. [Online]. Available: https://www.seeedstudio.com/DepthEye-3D-visual-TOF-Depth-Camera-p-3025.html

[15] pmdtechnologies. (2018) Development kit brief camboard pico flexx. [Online]. Available: https://pmdtec.com/picofamily/wp-content/uploads/2018/03/PMD$_D$$ev$$K$$it_B$$rief_C$$B_p$$ico_f$$lexx_C$$E_V$0218 − 1.*pdf*

[16] Intel. (2018) Intel® realsense™ depth camera d400-series. [Online]. Available: https://www.mouser.com/pdfdocs/Intel$_D$400$_S$$eries_D$$atasheet.pdf*

[17] ——. (2018) Intel® realsense™ depth camera d435. [Online]. Available: https://click.intel.com/intelr-realsensetm-depth-camera-d435.html