3D Time of Flight Sensors for Robot Navigation

Mohammed R Adil, Chidambaram Alagappan, and Swathi D Basaveswara

Abstract—Robot navigation has been gaining a lot of importance in recent times with robots dominating major fields of study. Image detection has become a prerequisite for effective navigation. Conventional 2D cameras detect the brightness of the environment but the depth information of the scene is lost. This is where the Time-of-Flight 3D cameras come into play. In addition to the brightness, they also detect spatial information of the scene. The depth information is depicted using color codes. In comparison with stereo vision systems and laser range scanners, they combine the advantages of active sensors; providing accurate distance measurements, and camera based systems; recording a 2-D matrix at a high frame rate.

This paper focuses on the emphasis of using the new generation TOF cameras. Starting with the basic principle, we go on to calculate the depth resolution and also undertake a review for enhancing the images. We also check the latest developments and the errors that people of facing in the field of ToF sensors currently.

Finally, the paper discusses Photonic Mixer Devices (PMDs) and pseudo-four-phase-shift algorithm for the performance enhancement of 3D-TOF Vision systems.

Index Terms—ToF, 3D cameras, 2D cameras, perception stage in mobile robot design, image detection, PMD, pseudo four phase shift algorithm.

I. INTRODUCTION

For any autonomous mobile robot device, the ability to navigate is very important. Navigation is a complicated process and doesn’t just involve avoiding obstacles and staying in safe operating conditions. The four building blocks of navigation are perception-robot must be able to interpret meaningful data using the sensors, localization- the robot must be able to determine its position with regard to the environment, cognition- the robot must be able to determine its path and finally motion control- the mechanical traversal along the planned path [1]. Vision based navigation has been used widely in the perception stage which is traditionally done by extracting navigationally salient features from images and maps [2]. In most cases, the processes of exploring an unknown environment through maps and determining the relative position are performed simultaneously through a process known as Simultaneous Localization and Mapping (SLAM) [3].

Several methods have been employed to improve the navigation process and make it more suitable for real world use. An image from stereo vision camera which provides 3D details of an object can be fused with the measurements of a 2D laser range finder [4]. Stereo vision requires complicated algorithms and powerful sensors to construct its occupancy grid and despite all these, it is prone to error [5]. Another technique that is employed is the Structure from Motion technique or the Kinetic Depth technique. Motion of an object is regarded as cue for the shape of the object and the spatial trajectories of points are used to estimate the dimensions [6]. Additionally, this technique is likely to cause problem when dealing with static objects and with time varying phenomenon. In Stereo Vision, the image and the data from the laser range finders corresponding to the same time has to be overlapped to obtain a 3D vision. In Kinetic depth technique, the image of the same object has to be taken at two different time intervals—either ways, both techniques require data fusion which requires computing power. Yet another technique to build 3D images is the Laser Range Scanner which works on the principle of calculating the distance from the observer to a particular point. Laser Range Scanners provide sparse data sets, use mechanical components and do not provide a 3D image with one image capture [7]. In contrast to all these models, the time of flight cameras combine the features of active range sensors and camera based approaches and provide a complex image which contains both the intensities and also the distances of each and every point. Also, there is no fusion of data from two separate sources and the data is being gathered continuously [7]. All these features make Time of Flight cameras/sensors the best way to obtain 3D dynamic information of objects.

II. TIME OF FLIGHT CAMERAS

A. Principle behind the time of flight cameras

As the speed of light is constant, points that are distant from the camera will take greater time to reach it. Using this principle, we can write algorithms to generate a perfect 3D image. Different time of flight cameras have been proposed but the one containing a single CMOS chip appears most widely in the literature. The distance to the object us calculated using properties of light and phase shift of modulation envelope of the light source [8].
The phase and amplitude of the reflected light can be detected using various signal processing techniques. Usually, to get a high resolution CCD based sensors are employed [9]. An architecture diagram of a CMOS sensor is shown in Figure 2. CMOS sensors usually have 64x64 pixel array and are implemented on a single chip using ordinary, low cost CMOS process. It also needs to have ADC and also a mechanism to generate high speed modulation signals [8].

B. Calculating the depth resolution
To do this, we assume the pixel output is single ended voltage even though it is differential. Let \( P_{\text{laser}} \) be the optical power of the light source, and \( A \) the total area illuminated. The amount of laser light received by the sensor depends on the reflectivity of the objects in the direction of the sensor. This reflectivity is denoted by \( r \). The total number of electrons generated in one pixel at the end of an integration (shutter) period of \( T \) can be written as:

\[
N_{\text{electrons}} = \frac{P_{\text{laser}} k_{\text{opt}} q_e r T}{A}
\]

Where \( q_e \) is the quantum efficiency and \( k_{\text{opt}} \) is a constant determined by the properties of the optical system including lenses, diffuser, pixel size, etc [8]. Essentially \( p \) represents the phase overlap between the reflected light and the reference signal. Hence the voltage across the storage capacitor at the end of integration is

\[
V_{\text{signal}} = \frac{q p N_{\text{electrons}}}{C} = \frac{q p P_{\text{laser}} k_{\text{opt}} q_e r T}{C A}
\]

The value of \( p \) at a particular pixel depends on the distance between this pixel and the target it is imaging. Depending on the distance, and hence \( p \), \( V_{\text{signal}} \) changes from \( V_{\text{signal}}(p=0) \) to \( V_{\text{signal}}(p=1) \). Since \( V_{\text{signal}}(p=0) = 0 \), the total voltage swing \( V_{\text{swing}} \) is equal to \( V_{\text{signal}}(p=1) \). One of the dominant components of noise on \( V_{\text{signal}} \) is the shot noise which is why we only consider this one loss and neglect all other losses like the ADC quantization noise, KT/C reset noise, thermal noise etc. The error in \( V_{\text{signal}} \) due to shot noise is calculated by projecting the uncertainty in the number of electrons to the voltage across the storage capacitor. In turn this results in an RMS (root mean square) voltage error of

\[
V_{\text{noise}} = q \sqrt{p N_{\text{electrons}}} = q \sqrt{\frac{P_{\text{laser}} k_{\text{opt}} q_e r T}{A}}
\]

The voltage resolution of the sensor is calculated by \( V_{\text{swing}}/V_{\text{noise}} \), which depends on \( p \) in \( V_{\text{noise}} \). For resolution analysis, we use \( V_{\text{noise}}(p=1) \) since it maximizes the magnitude of noise. Finally, the voltage resolution is given by [8]

\[
\text{resolution} = \frac{\text{Range}}{V_{\text{noise}}(p=1)} = \frac{c}{2 f_m} \sqrt{\frac{P_{\text{laser}} k_{\text{opt}} q_e r T}{A}}
\]

C. Enhancement of Depth Images
Once we have obtained the depth using the ToF cameras, we work to enhance these images. Optical noise existence, unmatched boundaries, and temporal inconsistency are the three critical problems which a ToF image suffers from [10]. Techniques like Gaussian smoothing and quadratic Bezier curve are used for static 3D images [10]. However, for enhancement of dynamic images, we use newly designed joint bilateral filtering, color segmentation based boundary refinement, and motion estimation based temporal consistency.

A bilateral filter is designed using both color and depth information at the same time. After color segmenting a color image, we extract the color segment set to detect object boundaries [10].
To minimize temporal depth flickering artifacts on stationary objects, we match previous and current frame color images. This predicts the movement of objects by estimating similarity between blocks in temporal domain [10].

D. Review of latest developments in Lock-in Time of Flight Sensors

Lock-in ToF cameras have been commercially available for more than five years and are based on analog phase demodulation [11]. These cameras are able to provide registered dense depth and intense images, complete image acquisition and high frame rate, small and compact design. Also, they don’t need any mobile parts and have auto-illumination [11].

The following table compares ToF cameras against other cameras:

<table>
<thead>
<tr>
<th>Differences</th>
<th>ToF cameras</th>
<th>Stereo vision</th>
<th>Structured light</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correspondence problem</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Extrinsic calibration</td>
<td>No, when used alone</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Auto illumination</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Untextured surfaces</td>
<td>Good</td>
<td>Poor</td>
<td>Good</td>
</tr>
<tr>
<td>Depth range</td>
<td>0.3 – 7.5 m.</td>
<td>Base-line dependent</td>
<td>Light-power dependent</td>
</tr>
<tr>
<td>Image resolution</td>
<td>Up to 2048x204</td>
<td>High resolution, Camera dependent</td>
<td></td>
</tr>
<tr>
<td>Frame rate</td>
<td>Up to 25 fps.</td>
<td>Typically 25 fps, Camera dependent</td>
<td></td>
</tr>
</tbody>
</table>

Non Systematic Errors:
Signal to Noise Ratio usually happens in scenes not uniformly illuminated. This is usually fixed by low amplitude filtering. Multiple light reception errors happen due to interference. This can be resolved by comparing angle of incidence of neighboring pixels.

Light scattering arises due to multiple light reflections between the camera lens and the sensor. This can be minimized by using a filter based on the combination of amplitude and intensity values.

Motion blurring can happen both along lateral and axial axes. Several methods have been proposed to resolve this error.

E. Errors and Compensations for ToF cameras

The two types of errors with ToF cameras are systematic-which can be managed by calibration and nonsystematic-which can be managed by filtering [11].

Systematic Errors:

Depth Distortion happens because the IR light cannot be generated as a sinusoidal because of irregularities in the modulation process. This type of error produces an offset that depends only on the measured depth for each pixel. As this error depends upon the measured depth distance, it can be resolved using a reference truth distance or by modeling the error from multiple relative measurements.

Integrated time related error is when you get different depth values for the same scene at different times and the cause for this error is still under investigation. Built in pixel related errors arise either due to material properties of the CMOS or due to capacitor charge time delay during signal correlation process.

Amplitude related errors occur due to low or overexposed reflected amplitudes. This can be caused by non uniform NIR LEDs which cause depth misreading at pixels distant from image center, low illumination for scenes with objects at different distances, difference in object reflectivity.

Temperature related errors can happen because internal camera can impact depth processing [11].

F. Photonic Mixer Devices (PMD)

Photonic Mixer Devices are also based on ToF principle and can realize a 3D image without complex electronics similar to a CMOS device [12].
systems do not require an optical reference channel for most applications [12].

Components of a PMD system: The PMD chip is surrounded by the peripheral electronics. The modulation driver defines modulation frequency and signal characteristics. The field-of-view is defined by illumination source in consideration of maximal optical power and best optical efficiency. Development Tools and Software are complementing the modularity of the design by providing fast digital interfaces and a flexible driver software [12].

By using the four phase shift algorithm, we double the frame rate thereby enhancing the performance of PMD sensors. To get a brief idea of this algorithm, please refer the following diagram [13].

Only two image captures instead of four are required to calculate the phase difference $\phi$. The frame rate of PMD TOF sensors is doubled without changing the integration time $T_{int}$. A typical frame rate of TOF PMD cameras is 50 Hz. The proposed approach increases this frame rate to 100 Hz. Furthermore, only two phases ($\omega \tau = 0^\circ$ and $\omega \tau = 90^\circ$ ) instead of four have to be generated by a PLD. This will decrease the software complexity of the system. The proposed algorithm is better suited for the implementation into a low-cost FPGA as the maximum clock frequencies are always limited [13].

III. CONCLUSION

This paper is a literature survey paper on how robot navigation can be made easier by using the 3D time-of-flight cameras as they have the advantage over the conventional camera of showing us the missed parts of the image or a obstacle that a conventional camera cannot. They also use the color codes to depict the distance of the obstacle from the robot, therefore providing an altogether clear picture of the situation and the dynamics of the image. We have also explained a particular type of a 3D time-of-flight camera called the Photonic Mixer Devices or PMD sensors and how these particular sensors will help make robot navigation better. A pseudo four phase shift algorithm has also been proposed which is a TOF enhancement algorithm.

REFERENCES


doi: 10.1109/ICPR.2010.577
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5597001&isnumber=5595735

doi: 10.1109/JSEN.2010.210106
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5686908&isnumber=5961208


doi: 10.1109/TIM.2010.204088
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5409568&isnumber=5443470

Rizwan Adil is a Graduate Student pursuing his Masters in Science degree for Electrical and Computer Engineering from Auburn University. He is currently working under Dr Mao for his thesis.

Chidambaram Alagappan is a graduate student pursuing his Masters in Science degree for Electrical and Computer Engineering from Auburn University. He is currently working on his thesis under Dr Vishwani Agrawal.

Swathi Dumpala B is a graduate student pursuing his Masters in Science degree for Electrical and Computer Engineering from Auburn University. She is currently working on her thesis under Dr Thaddeus Roppel.