SICK AG WHITE PAPER

A COMPARISON OF WORKING PRINCIPLES FOR 3D TIME-OF-FLIGHT, STEREO AND ACTIVE STEREO

3D SNAPSHOT TECHNOLOGIOES - THE BASICS

AUTHORS

Dr. Nadja Nagel
Applications Engineer – 3D Compact Systems
SICK AG Waldkirch/Germany

Dr. Fabian Zimmer
Applications Engineer – 3D Compact Systems
SICK AG Waldkirch/Germany

Dr. Anatoly Sherman
Head of Product Management and Applications Engineering – 3D Compact Systems
SICK AG Waldkirch/Germany

SICK
Sensor Intelligence.

TABLE OF CONTENTS

Abstract

As digitalization is on the rise in practically every industry segment, sensors are ever-present in assisting people and machines in everyday working life. 3D snapshot cameras are vision sensors which perceive a detailed 3D image of their surrounding environment in a single shot. This rich 3D information enables a broad spectrum of automation and assistance tasks: Collision avoidance for manned and autonomous vehicles, automatic palletizing and goods dimensioning, or robot guidance in pick-and-place processes are only a few examples. However, the performance depends substantially on the choice of 3D snapshot camera as different snapshot technologies operate on fundamentally different physical principles. In addition, technical requirements and ambient conditions vary enormously between industries, sites, and tasks.

The SICK Visionary portfolio covers all three main technologies of 3D snapshot: 3D time-of-flight, stereo, and active stereo. With this white paper, we want to give the reader a basic technical understanding of these 3D snapshot technologies, their characteristics, and short examples where to use them.

3D Snapshot Technologies: Experts in their Specific Field

A robot arm sorts goods by shape and color. A forklift automatically picks up pallets and smartly avoids obstacles on its route: This impressive reality is part of Industry 4.0 and reflects the rapid increase of automation and digitalization in nearly all industry segments. The mentioned examples, amongst others, require intelligent 3D cameras that act as machine eyes, able to determine the position and shape of objects.

With 3D snapshot technology, 3D and 2D images are created within a single snapshot – much faster than the blink of an eye. 3D snapshot can thus be used to quickly solve a variety of automation tasks. With its help, automated guided vehicles (AGVs) or manned heavy-duty vehicles can, for example, capture their surrounding environment live in 3D, navigate, follow people, and bypass obstacles even under harshest outdoor conditions. Due to their speed and their ability to capture 3D images both from static and dynamic scenes, cameras based on 3D snapshot technologies are also perfectly suitable for, e.g., fast and accurate freight dimensioning in logistics hubs or on conveyor belts. Other automation tasks, for instance, those involving robots for depalletizing and automated milking also greatly benefit from these technologies.

In large part, this great flexibility in applications and environments is possible due to differing technologies behind the existing 3D snapshot camera types. With the broad Visionary portfolio by SICK that covers all three main technologies, the industrial partner can benefit from single sourcing.

The present white paper guides you through the basic technical principles of these three types of 3D snapshot technologies: 3D time-of-flight („3D ToF“), 3D stereovision („stereo“), and 3D structured light stereovision („active stereo“). This basic understanding of working principles can help you in finding the best fitting solution for your specific task of industrial 3D machine vision from a technology-neutral point of view.

3D snapshot cameras provide a detailed image of their environment. With their help, diverse automation tasks where 3D and 2D perception, data streaming, detection, positioning, guidance, inspection, and measuring is needed, are mastered. Compared to technologies, which obtain depth information by scanning the scene, 3D snapshot cameras needs no mechanically moving components inside the camera and no relative movement between the sensor and the scene. Thus, two great advantages of 3D snapshot are its acquisition speed and its suitability for dynamic as well as static scenes. A single image shot suffices to provide a rich 3D dataset, containing the distance image and – depending on the technology – an infrared intensity image (2D grayscale), a color image, and further information such as a confidence or state map. The intelligence of the sensors lies in the task specific processing of this data. Thus, if the user wishes so, the sensor directly provides a simple tailor-made answer to the given task.
Besides these common features, the three main types of 3D snapshot cameras differ substantially in their operating principle and therefore in their performance under different circumstances.
Each technology has its strengths in different applications. SICK has the unique advantage to offer products based on every 3D snapshot technology that can solve most of the currently known applications. This guarantees technically and economically neutral evaluation and therefore optimal application consulting. The direct comparison of these three different 3D snapshot technologies ensures that the customer always finds the technology best suited to the individual application.

3D Time-of-Flight (“3D ToF”): Versatile Use with Stable Precision over High Detection Range

A 3D ToF camera emits infrared light signals within a complete given field of view, which are reflected by all objects in its surrounding. The object distance is either determined via direct measurement of the “time-of-flight,” i.e., the runtime of short emitted light pulses (“pulsed 3D ToF”) or via the phase shift between the emitted and the reflected light of a continuously emitted amplitude modulated light wave (“phase-shift-based 3D ToF”). The two 3D ToF technologies show certain differences: For example, phase-shift-based 3D ToF works with several frequencies that easily avoid cross-talk between different cameras. In addition, further information like confidence values can be deduced. However, what kind of impact those differences will have on camera performance strongly depends on how well the camera has been designed and built. Hence, a fair comparison would require camera-by-camera discussion beyond the scope of this paper.

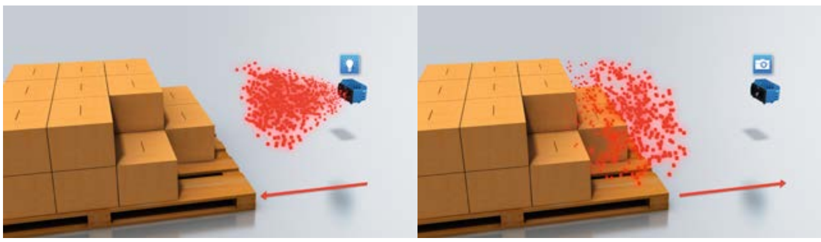

Figure1: In phase-shift-based ToF, modulated light is emitted and then captured by the camera after the reflection from the object.

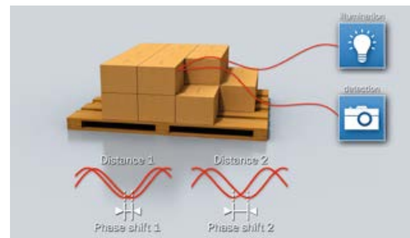

Figure 2: The distance is calculated by detecting the phase shift.

An important determinant of 3D ToF characteristics is its active light source. This ensures that 3D ToF cameras work even in scenes without contrast or without ambient light, for example on a flawless white wall or in a dark high-bay warehouse. Another advantage of an active illumination is that it allows for a large detection range.

Within the detection range, the accuracy and precision of 3D ToF distance measurements show only a weak dependency on distance (approximately linear behavior). This is a great advantage for tasks, which require a relatively large working range.
The power of the light that is emitted onto the scene decreases quadratically with its distance to the camera. Thus, the greater the distance of the target object, the more illumination power is needed. This power is limited by eye safety restrictions, which leads to a decrease of 3D data quality with increasing object distance.

The resolution of 3D ToF imagers has increased more and more over the last years, with one megapixel expectedly becoming a new standard in the near future. Yet, the lateral resolution is currently not as high as what is feasible with 2D imagers.

An active illumination implies interaction of the emitted infrared light with strong ambient light, which might result in a partial loss of depth information. The 3D snapshot camera Visionary-T Mini by SICK, for example, is equipped with a highly efficient laser illumination (VCSEL: vertical-cavity surface-emitting laser). For this reason the camera has a high ambient light robustness and at the same time an excellent detectability of dark objects. Further, the choice of the infrared spectrum of the laser emitter is crucial for a good ambient light robustness. Typical wavelengths are 850 nm or 940 nm, with the latter showing a more robust behavior against the sunlight spectrum.
As with any camera, the image acquisition time can be adapted to the conditions of the scene, i.e., to the remission properties and distance of its objects, as well as the scene‘s ambient light conditions. In dynamic scenes, the image becomes more prone to motion blur effects, the higher the image acquisition time is.

An advantage of 3D ToF is that independent of contrast or texture in the scene, a detailed coverage of the 3D environment (i.e., a high data density) is achieved. Yet, different remission properties due to different object surfaces and materials will have an impact on the depth measurement. The depth calculation depends on the light, which is reflected from the objects. For objects with strong unwanted light reflections, like mirrors or safety vests, so-called “multipath effects“ might occur: Reflections lead to a longer time-of-flight and therefore falsely increase the measured depth. Unwanted reflections can also arise inside the camera: Light might be reflected by the optics of the camera itself, again decreasing the quality of the 3D measurement.
Due to 3D ToF physics, the accuracy of a 3D ToF measurement is temperature-dependent. This effect, however, can be compensated by a good camera thermal design and calibration within the common operating temperatures.
As the 3D ToF technology operates in the near infrared spectrum, 2D infrared grayscale images can be acquired. These infrared images look very much like a normal 2D grayscale image of the scene and can therefore be used for further image processing, e.g., segmentation algorithms.

3D Stereo Vision (“Stereo”): Two Eyes for Object Classification and Harsh Outdoor Environments

Stereo cameras acquire 3D images by making use of the stereoscopic principle, the “two-eyes principle.” Similar to the vision of human eyes, stereo cameras derive a 3D image from two synchronized cameras that view the same scene under a slightly different angle. The distance values are calculated from the differences between the images captured by each eye. Due to the slightly different viewing angle, identical objects have marginally different positions in the two images. It is the same effect that causes a thumb to “jump“ when held in front of the face and alternately closing the left and right eye. To derive the disparity, both images have to be matched, which means that an algorithm searches for corresponding points in the left and right image to superimpose the two. The calculation effort that is necessary to generate a 3D image from two stereo images is higher compared to a 3D ToF dataset. Also, the fact that identical objects cast a different shadow for each imager, can, for example, lead to a shadow on both sides of central objects. A clear plus of stereovision is the simultaneous acquisition of a depth image and a 2D image – often a color image – in the visible spectrum, as compared to the infrared spectrum used for the 3D ToF camera. This offers more options and data for image processing especially when co-registering the 2D and 3D data. This allows for, e.g., robust object classification algorithms. The possibility to display a live 2D video stream is a clear benefit, e.g., for collision warning systems in manned vehicles or different quality inspection tasks.

This kind of stereo technology is also called “passive” stereo because ambient light is sufficient to gather 3D images and no active internal light source is used. The calculation of the depth information is based solely on the contrast present in the scene. Hence, the data density of stereo cameras depends on the scene’s texture. As a result, the disparity algorithms of stereo cameras are challenged by homogeneous surfaces, as it is difficult to identify corresponding points in the two images due to a lack of outstanding features and landmarks. Hence, surfaces with low or no contrast such as white walls, can lead to partial data loss. Such kind of low-texture scenes are less common in outdoor environments, but can rather be present indoors, for example, in the form of a brand new flawless floor in a factory hall.

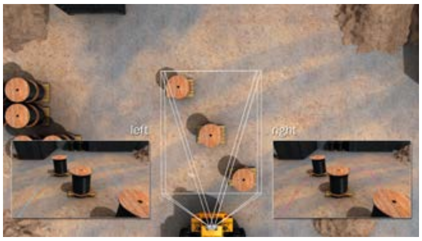

Figure 3: Step one: Acquisition of two images from slightly different perspectives.

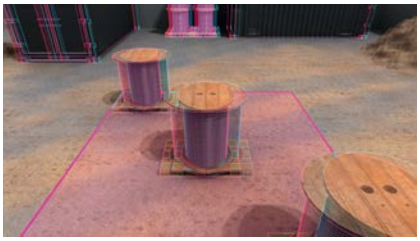

Figure 4: Step two: The two images are superimposed on one another.

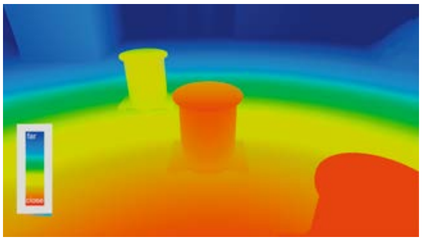

Figure 5: Step three: An image with depth information is created.

Another benefit of this technology, which makes stereo a perfect candidate for outdoor use, is its high dynamic range regarding ambient light. Changing lighting situations occurring, for example, in 24/7 operation outdoors with both strong direct sunlight and very low lighting conditions are no impediment for stereo technology. While active light sources can be impaired by sunlight, stereo technology shows a greater robustness to sunlight. On the other hand, this means that an operation in complete darkness is inhibited due to the absolute dependence of the system on external light. Another advantage of the stereo technology is its hardware robustness against harshest environmental influences by design. Taking Visionary-B product line from SICK as an example, with no active lighting source within the sensor, the stereo camera is suitable for large ambient temperate ranges and very robust against shock and vibrations as an active illumination and therefore an aditional heat source and mechanical part within the device is avoided. The fact that it only uses ambient light for image acquisition makes the data less sensitive towards outdoor conditions including fog, rain, and dust, typically found in mining, construction, or agriculture.

The flexibility to build optics with different baselines (distance between the two lenses) and focal lengths (which determines the distance of the camera‘s focus), allows the design of stereo cameras that achieve a high data quality at a specific working distance. Yet, up from that sweet spot, an increasing distance to the camera results in a lower depth resolution as the disparities become smaller. Thereby, the depth accuracy decreases exponentially. Yet, by solely relying on ambient light, the working distance and range of stereo cameras is commonly quite large compared to 3D ToF or active stereo.
The availability of cutting-edge, high-resolution 2D imagers allows building stereo cameras with high-resolution 3D images. However, to achieve a certain lateral resolution in the depth image, one must take into account that the disparities and hence the depth information can only be assessed where the two fields of view of the eyes overlap. Furthermore, it has to be considered that the disparity algorithms inherently tend to blur object edges. To obtain the desired lateral depth resolution these two effects have to be compensated for by an adequately high physical resolution of the imagers. The increasing calculation effort combined with the costs for such a pair of cutting-edge imagers might be translated into higher costs per performance for a stereo system.

3D Structured Light Stereovision (“Active Stereo”): Highest Resolution and Precision in Near Range

Active stereo cameras are based on the stereoscopic principle described above, combined with an active infrared light source that emits structured light, e.g., a point pattern. The reason for this structured light becomes clear when imagining looking at a wide snowfield. Such a large surface with no clearly visible contrast or structure impairs the 3D vision ability of human vision. As explained above, the algorithms of the stereo camera calculating the disparities face the same challenges. The clever part in active stereo vision is that instead of a homogeneous illumination of the scene, structured light is projected onto the scene. The used pattern or structure is designed such that the disparities between the two images can be identified more reliably and precisely, independently of the scene’s texture. Hence, the projected structured light increases the accuracy of the depth measurement and enables the reception of depth values even from homogeneous surfaces. Furthermore, active stereo cameras can, as opposed to stereo cameras, be operated in complete darkness, as they have an integrated light source. However, as for 3D ToF, the infrared structured light emitted by active stereo systems can be influenced by strong ambient light (e.g., ambient light robustness up to 40 klux for SICK’s Visionary-S). But there is an important difference: Unless the scene is completely homogeneous, disparities can still be deduced from the contrast in the surrounding environment even if the infrared light is no longer visible to the camera. The system is then acting as a stereo system and the measurement of distances is still possible.
Compared to stereo, the additional costs for the active illumination have to be taken into account. Compared to a 3D ToF camera, in applications with short working distances (a few meters), an active stereo camera will convince with its excellent depth resolution.

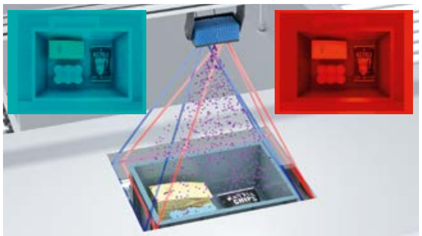

Figure 6: Step one: Acquisition of two images from slightly different perspectives. The scene is illuminated by the integrated structured light.

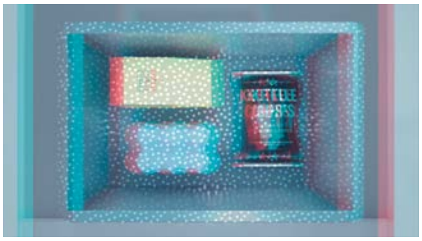

Figure 7: Step two: Overlay and correlation of left and right image enhanced by using the pattern of the non-visible structured infrared illumination.

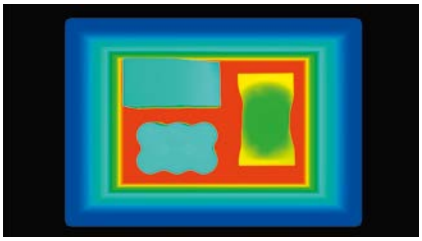

Figure 8: Step three: An image with depth information is created.

Software Defined Sensors

With the three different 3D snapshot technologies 3D ToF, stereo, and active stereo, a large range of possible applications is already successfully covered. Even though the working principle of each described technology is different, a common feature of all three technologies is the ability to generate high density 2D and 3D information of the environment. The data analysis and processing, which is characterized by the software, defines the sensor properties. That‘s why the ultimate flexibility and customization can only be achieved with flexible and adaptable camera software. 3D snapshot cameras image processing capabilities on board like SICK’s Visionary-T AP, Visionary-S AP or upcoming Visionary-T Mini AP can be programmed by the user to run stand-alone applications on the sensor and provide smart 3D data processing on the edge tailored to specific application needs. Furthermore, it offers the option to equip cameras with off-the-shelf logistics software app solutions, fulfilling today’s needs of flexibility and simplicity at the same time. We call it software defined sensor.
The table below shows a qualitative summary of the main differences between the three explained technologies.

3D Time-of-flight-technology, e.g., Visionary-T and Visionary-T Mini 	**Stereo technology, e.g., Visionary-B** 	**Active stereo technology, e.g., Visionary-S** 
⊕ Balanced depth accuracy over the whole operating range	⊕ Good robustness against harsh outdoor conditions by design	⊕ Excellent depth accuracy in near range, also for homogeneous surfaces
⊕ Good price-performance ratio with high density depth values, also for homogeneous surfaces	⊕ Wide field of view and large working range	⊕ Color (RGB) image in visible light possible
⊕ Operation in complete darkness possible	⊕ Built-in natural 2D image in visible light	⊕ Texture and remission independency
⊕ Sharp object edges	⊕ Robust against extreme ambient light conditions and their rapid changes	⊕ Scalability for higher resolution and data density
⊖ Dependency on object remission	⊖ Moderate accuracy, drops exponentially when moving away from sweet spot	⊖ Blurred edges due to stereo correlation algorithm
⊖ Susceptibility to multipath effects	⊖ Dependency on ambient light and scene texture	⊖ Accuracy drops exponentially when moving away from sweet spot
Common software platform (AppSpace): Empowering the user to develop software applications (apps) and deploy solutions that can be executed on the 3D camera itself. This enables easy configuration and adaption, for example, to process evaluated 3D data in the cloud.		

Summary: Enabling a Large Spectrum of 3D Vision Tasks with 3D Snapshot Cameras

This white paper reviewed the technology basics of three main 3D snapshot technologies: 3D time-of-flight („3D ToF“), 3D stereovision („stereo“), and 3D structured light stereovision („active stereo“). However, beyond the mere technology level, an application-based comparison is necessary to find the best-fitting 3D technology for specific automation needs. Amongst others, important application-specific factors to consider are mounting position and space, distance to the target objects, required resolution, ambient conditions (light, vibrations, precipitation, etc.), object surface properties, desired data size and format, complexity of task, but also the available budget.

3D ToF, with its stable precision over the whole detection range and the active illumination is a versatile candidate for various applications such as palletizing and depalletizing, object detection, level measurements of containers, dimensioning tasks, or intrusion detection. The 3D ToF product family Visionary-T from SICK, for example, is the suitable technology for AGV applications such as collision avoidance. It offers AGV-specific features like individual 3D detection volumes within the field of view of the AGV that are easily adaptable to the AGV`s driving state during operation (e.g., different zones for slow and high speeds and for turning left or right). Its industrial robustness makes this technology the candidate of choice not only for indoor, but also for “semi”-outdoor operation, for example, in rather shaded areas between storehouses.

In very harsh, dirty environments with difficult ambient conditions, stereo sensors will usually have a more robust performance than 3D ToF. Typical for such harsh outdoor environments are heavy industrial vehicles with large blind spots, posing a high hazard for both workers and infrastructure. Therefore, collision-warning systems covering large 3D blind zones of vehicles, as realized by the SICK Visionary-B are crucial to help ensure work safety and reduce costs caused by damages. Its robustness combined with intelligent object recognition enable an active warning to the driver, tackling false alarms. Other typical stereo applications are object classification and line guidance in agriculture, level monitoring, and automated collision avoidance on construction sites and in mining or guided container picking in ports.

The active stereo technology, is especially useful for applications with high accuracy and precision requirements at short distances. Its structured light allows for excellent depth values even on homogeneous surfaces. A typical automation task is to distinguish objects such as consumer goods or boxes by their dimensions and their color. The active stereo camera Visionary-S from SICK already merges this depth information with RGB color images on the sensor. For tasks like quality control or dimensioning, the 3D snapshot technology enables 3D information without scanning, both in static and dynamic scenes. Besides, active stereo is typically used for tasks where a high depth resolution at short and intermediate distances is required. Common examples are robot guidance applications such as bin picking, positioning, palletizing and depalletizing.

Exemplary applications for the three 3D snapshot technologies:

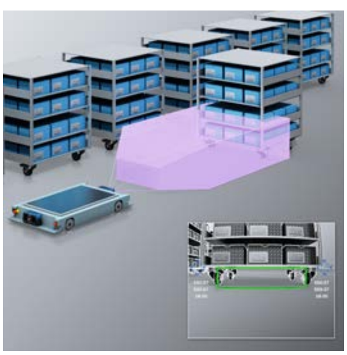

Figure 9: 3D ToF for the precise positioning of a platform AGV under a dolly

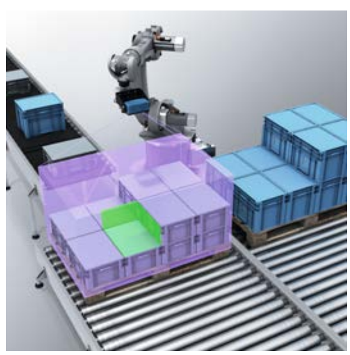

Figure 10: Active stereo for dimensioning and depalletizing

Figure 11: Stereo for active collision-warning in harsh outdoor environments

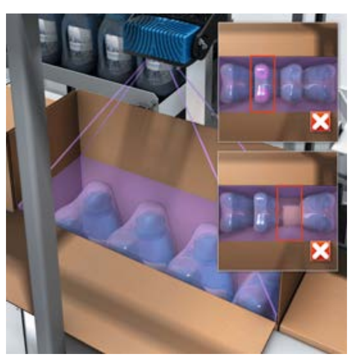

Figure 12: Active stereo for high-resolution quality control at close range.

Sensor Intelligence.