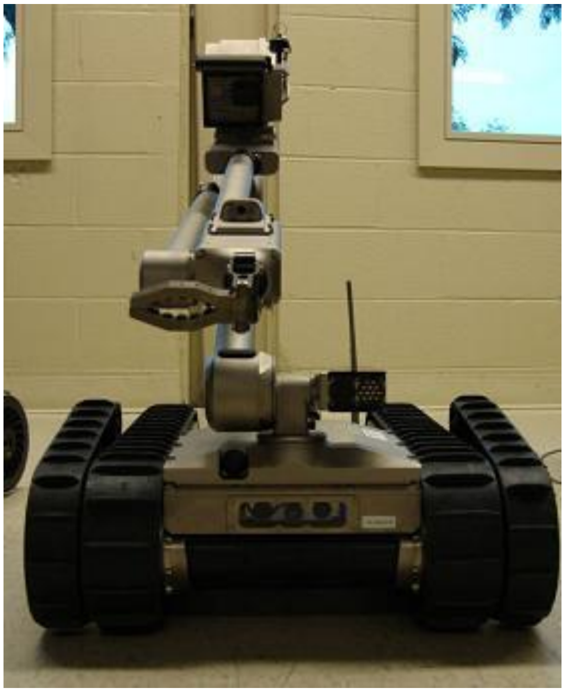
TYZX | systems that see

Giving Robots Real-time Vision and Depth Perception for Navigation, Person-Tracking, and Surveillance

Robots are proving themselves to be fast, efficient, and reliable at performing dangerous missions for military personnel and first responders. Autonomous and semi-autonomous robots are being deployed for missions such as:

- **Explosive ordnance disposal (EOD)**, including the detection and detonation of roadside bombs, car bombs, unexploded ordnance, and other hazardous explosives, while personnel remain at a safe standoff distance.
- **HazMat detection**, including the detection of chemical, nuclear, or biological hazards.
- **Route clearance**, such as removing debris or ordnance from the path of a convoy.
- **Structure mapping**, which involves surveying buildings, bunkers, and other structures for hazards before personnel encounter them.
- **Person-tracking and person-following**, either for surveillance, police actions, or troop maneuvers.
- **Long-term surveillance** (sometimes called “persistent stare”) of a fixed location.



An iRobot PackBot configured with a TYZX G2 Vision System in its payload for situational awareness and obstacle detection/obstacle avoidance.

All these jobs—from helping first responders in urban centers to conducting surveillance in mountain passes—require unwavering attention and precision, despite great physical risk.

That’s why they’re best performed by robots. Robots can move along roads, searching for bombs and defusing them. Using sophisticated treads and flippers, robots can clamber through streets, alleys, parking lots, stairwells, and hallways, in search of accident victims, suspicious persons, or bombs. Robots can also manipulate objects, lifting things, setting them down, or pushing them aside—skills useful for tasks as varied as delivering supplies in emergencies to removing hazardous objects.

Robots not only do these jobs; they do them very well. The U.S. military finds that robots detect and disable roadside bombs five times faster and with more accuracy than humans. Increasingly, semi-autonomous or autonomous operation to reduce the warfighter’s workload and distraction while in hazardous environments.

The Requirements for Real-time Vision

To bring autonomous operations to any of these missions—from EOD to surveillance to inspecting crushed buildings for survivors—robots need to be able to see what's in front of them and make intelligent decisions about their surroundings. This visual awareness is important in many contexts including:

- **Situational awareness:** detecting and interpreting the appearance of people, objects, and terrain, and projecting their status in the near-term future.
- **Object detection and object avoidance (ODOA):** detecting objects or people in a projected path and steering around them, while avoiding other objects or hazards.
- **Person-tracking and person-following:** detecting people standing or moving in a scene and tracking their movements and locations, as well as navigating through the environment in pursuit of a specific person.

Providing robots with real-time vision is challenging. To be able to robustly interact with their surroundings in dynamic situations, robots need a vision system that is:

- **Fast**, delivering the lowest possible latency between the viewing of the event and its interpretation and response. The frame rate of the vision system must be able to accommodate situations in which people, objects, and unmanned systems are moving quickly.
- **High resolution**, so that the robot can perceive detail with greater accuracy and at longer distances.
- **Robust,** so that it functions reliably in changing lighting conditions, providing accurate data at dawn, noon, and dusk, in rain and in sun, etc.
- **Compact and low-power**, so that the vision system does not become a significant drain on the battery life of a robot or other unmanned system. Small, lightweight components help keep smaller robots human-portable.
- **Passive**, reducing detectability on the battlefield.

Range data—determining the distance between the robot and a person, object, or feature of the terrain—is vital to all three visual categories: situational awareness, ODOA, and person-tracking and person-following. To determine range data in real time, robots use either active or passive sensing technology.

Active sensing technology sends energy—such as radio waves or laser light—out into the robot's environment, then determines distance by performing calculations on the energy that returns. RADAR is perhaps the best known example of active-sensing technology. LADAR, which transmits beams of laser light, is also used for calculating range data. Both RADAR and LADAR have several shortcomings, especially in military operations, however. RADAR is often heavy and demanding of power. It typically has low angular resolution and unless the system is very sophisticated, signal returns from large objects and overwhelm closer, smaller objects making them difficult to detect. Being an active sensor, RADAR is readily detectable by hostile forces. LADAR's operating range is usually limited in distance and field of view particularly in daylight operation. Operation is typically limited to one or two dimensions, making it difficult for robots to detect people or to avoid obstacles appearing at unexpected heights. Scanning systems introduce rotating mechanical assemblies and slower update rates to the equation. Similar to RADAR, LADAR is readily detectable.

Passive technology interprets data that is detected by sensors without any prior emission of signals. Traditional video is a passive vision technology, but 3D situational awareness is difficult to achieve from a single video feed, even if that video feed is high resolution. Situational awareness provided strictly from 2D images, especially when contending with lighting changes and shadows, is not robust enough to support autonomous or semi-autonomous operation.

Stereo vision is also a passive technology. Just as the human brain can quickly calculate distances by comparing the similarities and differences between two images from adjacent sensors (our eyes), a stereo vision sensor can be used to calculate 3D images, giving robots the situational awareness they require. Stereovision can provide accurate distance measurements over a wide field of view under a wide range of illumination, but it also has limitations. First, the process of stereo correlation (turning the 2 – 2D images into 3D) is computationally expensive and difficult for conventional processors to do in real time. Stereo assumes some natural texture in the images to support correlation; fortunately there is an abundance of texture in nature apparent to imagers of sufficient resolution and dynamic range. Finally, the benefit of passivity can become a limitation in dark operations compared to active systems. Stereo can be used with IR sensors to support passive operation at night, and have the added advantage of not blinding night vision systems.

The Solution: TYZX 3D Vision Systems

Building on over a decade of research and development in computer vision, TYZX, the 3D vision company, offers compact vision systems for unmanned systems that allow them to achieve situational awareness in support of autonomous and semi-autonomous operation.

The TYZX G2 Embedded Vision System (EVS) is a small embedded vision system with low power requirements that delivers high performance visual analysis (up to 30 fps) in variable lighting conditions. Designed for widespread deployment in applications such as security and navigation, the TYZX G2 EVS provides real-time 3D vision processing without the need for active sensing technology or off-board computing resources.

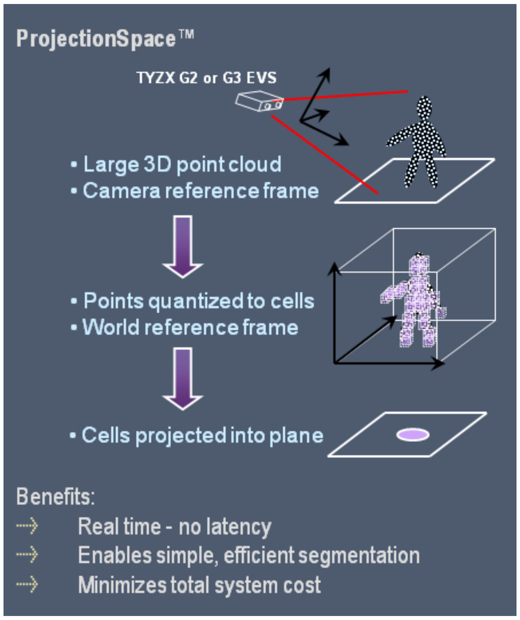
About the size of a hardcover book, the TYZX G2 EVS is small enough to fit easily in robot payload. It draws little power and uses passive cooling. Performing advanced 3D calculations in silicon, the G2 EVS gives robots accurate range data in all three dimensions.

TYZX ProjectionSpace

The TYZX G2 EVS uses TYZX ProjectionSpace™ primitives to transform point-cloud data into efficient 2D or 3D geometric representations and rapidly segment a scene into relevant objects. Dedicated TYZX hardware performs ProjectionSpace computations in real time, eliminating latency and processing burden. With TYZX, high-level applications get immediately useful 3D data, so they can work faster and more productively.

How ProjectionSpace Works

In most 3D vision applications, raw 3D-sensor point clouds are most often represented as an image with row and column locations containing metric-distance measurements. To represent an object or obstacle relative to a camera, each point first needs to be transformed into a 3D metric location; that is, each point needs to be assigned X, Y and Z coordinates represented by 3 floating-point numbers. This transformation of every point in a space seen by a camera can be computationally expensive.



Such a transformation produces coordinates that are relative to the sensor's coordinate system, rather than a useful coordinate system from the real world, such as the dimensions of a room. Or, to take an example from a camera mounted in a vehicle, sensors may be pointed slightly toward the ground or looking to the left or right of a vehicle axis. A common, useful next step then is to perform a rigid 3x3 transform, taking the points from the sensor's coordinate system and placing them in a world coordinate system such as that of a vehicle.

When it comes to applications making sense of a scene and interacting in real time, 3D points positioned in a preferred coordinate system is vastly more useful than a raw point cloud, but the data representation of 3 floating-point values per point is computationally expensive to operate on, especially with limited resources.

A more useful representation would be to assign or "project" the points into cells in a regular 2D or 3D grid – a Euclidean 3D quantized volume. If a point's 3D floating-point coordinates lie within a particular 2D or 3D cell's volume, that cell's count can be incremented. Incrementing a cell's account accumulates evidence that the cell is occupied. A high count provides high confidence that the cell is occupied. Applications can also set a minimum count threshold on cells thereby eliminating spurious "sensor noise" and further reducing application workload.

By performing these operations, TYZX ProjectionSpace changes a point-cloud represented as a large floating-point data structure into an easily searched and segmented array, in the preferred coordinate system of the application. While valuable in their own right for most applications, perhaps the best part is that all of these operations are computed in parallel at 30 frames per second directly in hardware. This means no additional latency for results, no application CPU or memory bandwidth burden and of course power savings.

Applying ProjectionSpace to Robot Navigation

Obstacle Detection and Obstacle Avoidance (ODOA)

Obstacle detection algorithms attempt to identify objects that will block the passage of a vehicle (such as an autonomous robot), since these objects could likely result in the vehicle needing to steer, slow, or stop. 3D images are critical for object detection. To take the appropriate action, a navigation system needs to know an object's distance from the vehicle, as well as the true size of the object. But searching full distance images for regions containing an obstacle can be time-consuming. In real-time applications such as vehicle navigation, time is a precious commodity.

ProjectionSpace images greatly assist with the speed and accuracy of obstacle detection. When configured with appropriate parameters, a ProjectionSpace image can summarize a scene in terms that are useful for identifying potential objects – reducing the size of the overall image to be analyzed, reducing noise, and reducing clutter that's irrelevant to the navigation of the robot.

Obstacles with a large surface area perpendicular to the ground will stand out with high pixel counts in a top-down ProjectionSpace image, even for vehicles traveling over rough (and non-planar) ground surfaces. If necessary, an application can perform further computation to verify the presence and characteristics of a detected obstacle, but this computation can be limited to just the portion of the image containing the obstacle. Reducing computation reduces latency, and improves the responsiveness of the application overall.

An application using ProjectionSpace for obstacle detection typically takes into account vehicle-specific parameters, such as the location where the sensor is mounted, the orientation of the sensor relative to the ground, and the size of the vehicle. Details about the sensor position and orientation are used to provide the 3D rigid transform applied to each range pixel as part of the ProjectionSpace transform.

A vehicle's sensor is often mounted off the ground and tilted downward to obtain a good view of the scene directly in front of the vehicle. Unfortunately for the application guiding the vehicle, tilting the sensor this way creates a distorted view of obstacles close by, because pixels at the top of an object will appear closer to the sensor than pixels on that object near the ground, even if the obstacle is a vertical obstacle such as a pole. It would more convenient for obstacle-detection algorithms if the application measured the distance to all parts of the object along a vector parallel to the ground—in other words, if the obstacle-detection algorithm evaluated obstacles within the context of the vehicle's path. Then, if a pole were 3 feet away from the vehicle when measured along the ground, there would be many counts in a cell—in this case, a ProjectionSpace cell—at 3 feet of distance in the virtual view, and distance to the object would be easy to determine. To further aid with the 3D representation, the size of the vehicle can be used to set the bounds of the ProjectionSpace data along each axis.

Setting clipping planes to monitor only data within a certain distance from the ground eliminates clutter from the ProjectionSpace image. A vehicle that's low to the ground does not really need to know if there is a topological feature, such as a high tree branch, well out of the vehicle's path. Eliminating such data from transforms accelerates ProjectionSpace calculations.

Programmers can configure the cell size used for aggregating ProjectionSpace data in proportion to the vehicle scale and the accuracy required for the ODOA application. For instance, obstacle location requirements might be on the order of 20 centimeters, so cell sizes could be set to 10 centimeters, accumulating many distance pixels in one cell in closer regions of the image. Making cell size configurable reduces the number of cells in the ProjectionSpace image overall. It also makes it straightforward for an application to automatically eliminate spurious erroneous pixels, which will be broadly distributed and produce very small cell pixel counts.

Minimal Latency

To appreciate the importance of low latency in robotic navigation, consider the example of a robot traveling 16 kph, which translates to just under 4.5 mps. If the frame rate of the vision system is 30 fps, that means that in the time the vision system has processed one frame of images, the robot has traveled 0.15 m. Clearly, the vision system must operate with great efficiency and minimal latency in order for the robot to detect and avoid hazards such as ordnance or even a pothole.

Similarly, consider the seemingly simple task of crossing a trail full of bicyclists and pedestrians. Low frame rates and high latency will provide poor results because the world will have changed before the robot could plan or move. Providing consistent, low-latency 30 fps updates turns static 3D images into real time velocity vectors allowing the robot to predict where obstacle *will be* when it moves instead of where it was.

The TYZX ProjectionSpace performs its operation for a scene in less time than it takes to process a single frame of imagery. This means that the robot has accurate range data within 1/30th of a second, enabling the robot to respond nimbly and effectively in its environment.

Meeting the Requirements for Robot Navigation

The table below lists the requirements for robot navigation and summarizes how the TYZX G2 EVS using ProjectionSpace calculations meets these requirements.

Requirement	**TYZX Feature**
Fast	Processes visual data to 30 fps with minimal latency, ensuring that robots can respond quickly to their environments.
High resolution	Works at resolutions of a few centimeters or even millimeters, as required.
Adaptable	Works in variable lighting conditions.
Compact and low-power	Fits in small spaces, and draws less than 15 Watts, preserving robot battery life.
Passive	Works indoors or outdoors without the need for external lights, RADAR, or lasers.

TYZX and iRobot

iRobot Corporation has chosen the TYZX G2 EVS for several military robotics research projects requiring real-time vision and depth-perception. iRobot has demonstrated the ability to integrate the TYZX G2 EVS onto its PackBot® and Warrior™ platforms: rugged tactical mobile robots designed to perform dangerous search, reconnaissance and bomb-disposal missions while keeping troops out of harm’s way.

Results

Using TYZX Embedded Vision Systems, iRobot military robots are able to achieve:

- **Enhanced Situational Awareness via 3D Visualization** – Standard monocular cameras provide video footage that is “flat,” sometimes making it difficult for a robot operator to judge distance. 3D visualization provides depth perception and a more detailed view of the environment. Using an operator control unit (OCU) integrated with TYZX stereo vision data, the robot operator can more easily manipulate objects such as unexploded ordnance.
- **Person Detection and Person Following Capabilities** – Using the TYZX system for person detection, iRobot researchers are developing advanced autonomous navigation algorithms to demonstrate person following capabilities. Using onboard sensing from the TYZX system, iRobot’s tactical mobile robots have demonstrated the ability to detect, recognize, track, and follow specific persons of interest.

- **Obstacle Detection and Obstacle Avoidance (ODOA) Capabilities for Increased Autonomy** – TYZX G2 technology has enabled iRobot's SEER payload for its PackBot and Warrior platforms to support autonomous ODOA for complex vertical structures. Whereas traditional planar LIDARs provide only a 1D horizontal sweep of obstacles, TYZX technology provides range details in 2D. This allows the robot to sense how high an obstacle is and to determine if it can overcome that obstacle.

Tom Wagner, Vice President and Technical Director for iRobot, summed up the TYZX G2 EVS's performance this way: "The TYZX G2 EVS provides our robots with the ability to sense and assess the surrounding environment. The G2's onboard processing capability, as well as its lack of moving parts, makes it a fitting sensor for our PackBot and Warrior platforms. Our robots are used in complex terrains and the G2 system provides sensor data needed to enable advanced capabilities on our platforms."

The iRobot example demonstrates the effectiveness of the TYZX approach to real-time 3D vision for robots performing hazardous missions.

TYZX | *systems that see*

TYZX, Inc.
3715 Haven Avenue, Suite 110
Menlo Park, California
Tel: +1 (650) 282-4500
Email: sales@tyzx.com
WWW: www.tyzx.com