



TYZX | *systems that see*

## White Paper

# Sight in an Unstructured World – 3D Stereo Vision and Systems that See

*Ron Buck, President and CEO, Tyzx*

### Abstract

*For decades, the business world has sought a practical way to add ‘computer vision’ capabilities to applications and products. Industries understand the windfall of new business benefits that computer vision could introduce – from more sophisticated homeland defense capabilities, to improvements in automotive safety, to smarter robots, to more interactive video games, and more.*

*This paper describes the challenges for prior approaches to computer vision, and highlights 3D stereo vision’s emergence as the preferred method of enabling sight in machines.*

### Introduction

Of the five human senses, sight is generally regarded as the most important for exploring and understanding the world around us. So it’s no surprise that the business world is excited about using computer vision to enable systems to better see, interpret and respond to real world events, in real-time.

However, despite the business community’s strong desire to find practical ways to incorporate computer vision into commercial offerings, early computer vision approaches fell short. The few examples that worked well were limited to tightly controlled research lab and manufacturing environments. They have not proven practical for real-world deployment.

This paper takes a closer look at some of the previous approaches to realizing commercially-viable ‘systems that see,’ and explains why 3D stereo vision is the approach promising the price and performance necessary to usher computer vision into the consumer world.

## Using Cameras to See – Historic Limitations

The scientific community has always used cameras as eyes for machine vision. However, for a variety of technical reasons, it is a challenge for computers to interpret camera-generated images in the same manner as humans would. The real world is in constant motion, full of shifting backgrounds and objects of different shapes and sizes. The ability to interpret this dynamic landscape is computationally and theoretically problematic. When a computer looks at the world through a lens it just sees pixels. A pixel doesn't make sense on its own. While image sensors (driven by Moore's Law and the digital camera market) continue to make dramatic improvements in cost and resolution, the abilities to interpret visual data and to do so in real-time have been the missing links.

Computer vision engineers working to interpret an image typically begin by "segmenting" an image – basically breaking the image up into discrete objects. Their software searches an image for pixel groups with certain characteristics (lines, similar colors, etc), which may indicate separation from other groups of pixels. Unfortunately, most of the historic efforts to achieve systems that see have revolved around 2D imaging, rather than 3D imaging. 2D images are notoriously difficult for computers to interpret and respond to, for a variety of reasons such as lighting, shadows, and partially obscured objects, just to name a few.

Consider the following 2D image:



If one asks a computer (based on this 2D image) where the two people are in this scene, and which person is taller, it's an overwhelmingly difficult challenge. In 2D, the woman in the back (standing on the lawn) appears to be the same height as the child in the front (standing on the

table) – and the same distance from the camera. The woman's pants are a similar shade as the background wall, making it difficult for a computer to identify her as a unique object separate from that wall. The ability to identify unique objects and assign attributes to them (i.e. which one is closer or farther, which one is larger or smaller, which one is moving faster or slower, etc.) requires extraordinarily complex algorithms, error-prone heuristics and heavy computation.

Another challenge with 2D image interpretation is the dynamic nature of the world. Lighting conditions, background and other

factors change constantly in the real world. So a computer algorithm that finds a red dress in one 2D image may not recognize the dramatically changed shade of red in the next image caused by the shadow of a passing cloud. 2D can be used effectively only in controlled environments. In changing conditions, 2D doesn't work for demanding, real-time applications.

## Why 3D Vision is Better than 2D

A major breakthrough in computer vision occurred when researchers turned to biology and the study of stereo vision to turn 2D images into 3D volumes. This approach better represents distance, size and spatial relationships between different objects in the camera's field of view.

For example, the image of the woman and the child changes dramatically when viewed in 3D from stereo vision (darker pixels are closer):

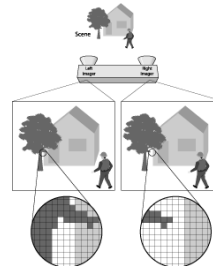


Now the computer can see the woman standing out from the background (clearly delineated from the wall). The computer also sees that the woman is further from the camera than the girl in the front. 3D sensing

facilitates easier and more reliable segmentation – and provides absolute size and shape information. By transforming 2D images into 3D images – image interpretation is simplified, and the results are more accurate.

As in human vision, stereo vision uses two eyes/imagers instead of one. It determines the distance to each pixel by measuring parallax shift – the apparent shift of a point when viewed from two pre-set locations (such as the left and right eye of a human).

The parallax shift is relative to the geometry (i.e. baseline of the cameras and other parameters). A pixel covers a certain area based on the object's distance from the camera. With that data, one can determine the size of objects, their respec-



tive distance from one another, as well as the distance from the camera. Further, recent technological advances make it possible to determine the parallax shift at fractions of a pixel – enabling even higher precision down to 1mm at camera frame rates of more than 30 per second.

3D vision also makes it possible to build products and applications that can react intelligently in real-time despite constantly-changing external environments. 2D vision fails in dynamic environments whereas 3D vision relies on size, shape and distance which are invariant under changes in lighting. For example, a 5-foot-tall person standing 10 feet from the camera wearing a red dress will still be 5-foot-tall and stand 10 feet from the camera when a cloud passes overhead, even though the dress will no longer be the same shade of red.

### **3D Interpretation Is Better – But What About the Cost and Complexity Issues?**

While three-dimensional images are more easily and reliably interpreted, and parallax enables generation of 3D images, the amount of computation required to perform the task has historically made 3D stereo vision prohibitively slow, costly and impractical.

Moore's Law, which stipulates that the number of transistors on a chip double every 18 months, applies to the challenges of 3D vision. In stereo computation, there is Woodfill's Law (named after Tyzx CTO, Dr. John Woodfill) – which stipulates that the computational complexity of computing a 3D image is cubed relative to the size of an image's edge. As camera technology improves on the scale of Moore's Law, digital images become increasingly higher resolution – and in the process are becoming correspondingly more difficult to convert into 3D.

In key vision applications such as object tracking, where a computer must interpret high resolution images, generate 3D, and compare objects against a range of characteristics in a matter of milliseconds, Woodfill's Law posed an insurmountable hurdle to the widespread adoption of 3D vision. Indeed, at Xerox PARC in the late 1980s, when an object (a cat) was first successfully tracked in real-time by a computer as it moved through an unstructured environment, it literally required a supercomputer to meet the computational demands.

Researchers use the term Pixel Disparities per Second (PDS) to evaluate the performance of stereo vision systems. This term measures the total number of pixel to pixel comparisons made per second. It is computed from the area of the image, the width of the disparity search window in pixels, and the camera frame rate. Tyzx has developed a patented architecture for stereo depth computation and implemented it on an application-specific integrated circuit (ASIC) called the DeepSea Processor. This chip enables the computation of 3D images with very high frame rates (200 fps for 512x480

images) and low power requirements (<1 watt) – properties that are critical in for low-cost, volume applications in commercial markets. The DeepSea Processor is capable of 2.6 billion PDS, which today is more than 10 times faster than any other stereo vision system available.

### **What about Lidar and Radar?**

In addition to 2D vision, there have also been attempts to realize the promise of systems that see by using Radar and Lidar.

Radar is the method of detecting distant objects and determining their position, velocity, or other characteristics by analysis of very high frequency radio waves reflected from their surfaces. Lidar is the method of detecting distant objects and determining their position, velocity, or other characteristics by analysis of pulsed laser light reflected from their surfaces.

Radar and Lidar have shortcomings when it comes to the industry's need for systems that see. Both approaches are 'active' – meaning they rely on broadcasting and returning echoes (RF and light, respectively). In both cases, this need to send/receive a signal tends to have a very negative impact on the accuracy and resolution of the 3D results generated. They tend to be very coarse measurements (particularly as one gets further away from the sensor),

Further these methods tend to be intrusive, making them impractical for use in settings where there are humans (as is the case with most commercial settings). Lidar, for example, relies on the transmission of infrared light, which can be dangerous to human eye health beyond nominal power settings.

Stereo vision is not intrusive. It relies on ambient light. Stereo vision also produces a standard color 2D image for conventional image processing as well as a 3D image. Further, stereo vision takes advantage of the effects of Moore's Law on improvements in the quality and cost of commodity imagers. Radar and Lidar do not.

## Systems that See – Criteria for Real World Application

The historic limitations of sensing through Radar, Lidar and 2D vision have helped researchers define pre-requisites for high-performance, low-cost, high-volume, commercially-viable computer vision systems. These characteristics include:

- ❖ *It must be fast* – There must be low latency (lag time) between the event being viewed, and the interpretation/response of the system designed to “see.” The frame rate needs to be high enough to follow fast moving objects. The vision capability must be fast enough that the machine that is using it can incorporate the data and respond in real-time.
- ❖ *It must have high resolution* – The preferred method for systems that see is to interpret a high degree of detail. Higher resolution yields more accurate results, and allows cameras to survey large areas, further reducing cost.
- ❖ *It must be small, inexpensive, and require little power* – For mainstream use, a sensor must be small enough, cheap enough, and have low enough power requirements to work inside of high-volume, commodity products.
- ❖ *It must be passive* – There are systems (known as “active”) that push energy into an area as a means for image interpretation. In some cases, these methods can be dangerous to humans (as is the case with Lidar). In other cases, the problem with active methods (as in the case with Homeland Security and Defense applications) is that they are detectable. For a variety of reasons, the most desirable systems are passive systems that perform 3D sensing without making their presence known, or infringing on human safety.
- ❖ *It must have a broad useful range* – One system should serve applications in environments at both close range (centimeters away) as well as long range (hundreds of meters) without requiring a change in technology.

## Tyxx 3D Stereo Vision Meets all of these Criteria

The Tyxx DeepSea Stereo Vision System’s computer vision capabilities already meet the criteria for commercially-viable systems that see, and the price-performance curve of the technology is improving rapidly:

- ❖ *Low latency* – The DeepSea Development system is a half-length PCI card that connects directly to the Tyxx Stereo Camera. It provides depth and color information to a conventional personal computer. At its core is the high performance DeepSea chip that performs Tyxx’s patented, pipelined, implementation of the Tyxx invented Census matching algorithm with low latency and high frame rates.
- ❖ *Capable of interpreting a high degree of detail* – The DeepSea chip performs nearly 50 GigaOps/sec. providing dense, 16-bit depth estimates for every pixel in an image.
- ❖ *Small, inexpensive, and with low power requirements* – The Tyxx Stereo Cameras are lightweight and low power, employing inexpensive off-the-shelf digital CMOS imagers and miniature lenses for low-cost deployment. The DeepSea card – though powerful – requires very little power (under 5 watts) and contains the complete stereo computation engine. This design frees the personal computer’s processor and memory for application processing tasks.
- ❖ *Passive* – 3D stereo is a passive sensing method.
- ❖ *Excellent range* – With Tyxx, almost any operating parameters are possible given an appropriate camera configuration, without requiring any changes to the underlying stereo computation engine.

## About Tyxx

Tyxx is a 3D vision company providing a platform of hardware, software and services for building products that see and interact with the world in three dimensions. Tyxx products and services are used by industry leaders in automotive, consumer electronics, robotics and security markets. Founded in 2002 and based in Menlo Park, California, Tyxx is privately funded. For more information, visit [www.tyxx.com](http://www.tyxx.com) or email [info@tyxx.com](mailto:info@tyxx.com)