3D Vision System Development

Mattias Johannesson
Expert, 3D Vision
Agenda

• Why do we need 3D Vision?
• Definitions in 2D & 3D Vision
• 3D Techniques and Applications
  – What fits where?
• Conclusions
Why 3D?

This IS a photo of a real object!
Why 3D?

This is a 3D camera image of the same object
Why 3D?

Different 2D view of same object...
3D Vision Use

- To locate
- To identify
- To inspect
- To measure
- To navigate

3D more difficult than 2D!
  - Get good “image”
    - Illumination more critical than in 2D
  - Use capable SW package
    - Avoid reinventing the wheel
Data Types

- **2D intensity**
  - 2D array of brightness/color pixels

- **2.5 D range**
  - 2D array of range/height pixels
  - Single view-point information
  - Depth Map / Distance Map

- **3D surface range data**
  - Surface coordinates [x,y,z]
  - Point cloud data

- **3D "voxel"**
  - A volume [x,y,z] of densities
  - e.g., CT scan
Data Types

- **2D intensity**
  - 2D array of brightness/color pixels

- **2.5D range**
  - 2D array of range/height pixels
  - Single view-point information
  - Depth Map / Distance Map

- **3D surface range data**
  - Surface coordinates [x,y,z]
  - Point cloud data

- **3D "voxel"**
  - A volume [x,y,z] of densities
  - e.g., CT scan
Map of 3D

Base Technologies:
- Triangulation
- Time-of-flight
- Interferometry
Map of 3D

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry
- Time-of-flight
  - Binary Coded
  - Phase Coded
  - CW
  - Pulsed
Map of 3D

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry
- Time-of-flight
  - CW
  - Pulsed

Binary Coded
Phase Coded
Map of 3D

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Structured Light
- Shading
- Laser Triangulation
- Binary Coded

Active
- Interferometry
- Time-of-flight
- Phase Coded
- CW
- Pulsed

Active Triangulation
- Where is the light
Map of 3D

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry
- Time-of-flight
  - CW
  - Pulsed

Shading
- Laser Triangulation
- Binary Coded
- Phase Coded

Time-of-flight
- When is the light
Map of 3D

- **3D imaging**
  - Passive
    - Focus
    - Lightfield
    - Stereo
  - Active
    - Structured Light
    - Interferometry
    - Time-of-flight
      - CW
      - Pulsed

Interferometry
- *How is the light*
Acquisition Speed

• Basic Acquisition strategies:
  • Snapshot
    – Stereo
    – Primesense / "Kinect 1"
    – Time-of-flight array camera
  • "Almost" snapshot
    – Coded light projection
    – Moving camera stereo
• 1D scanning
  – Laser triangulation + Linear movement
  – 1D scanning (depth from focus, interferometry)
• 2D scanning motion
  – 2D scanner
  – Linear movement of object + 1D scanning
Accuracy

• **Resolution**
  - Pixel size $\Delta X, \Delta Y$
    - Not Feature size!
  - $\Delta Z$ – depth resolution

• **Repeatability**
  - First step to accuracy

![Diagram showing low, mid, and high accuracy levels with corresponding repeatability.](image)
**Accuracy**

- **Resolution**
  - Pixel size $\Delta X$, $\Delta Y$
    - Not Feature size!
  - $\Delta Z$ – depth resolution

- **Repeatability**
  - First step to accuracy
Accuracy

• Resolution
  – Pixel size $\Delta X, \Delta Y$
    • Not Feature size!
  – $\Delta Z$ – depth resolution

• Repeatability
  – First step to accuracy

• Accuracy
  – If the system is repeatable then accuracy is “just” calibration
Calibration

- Map Relative image coordinates to World coordinates

![Diagram showing calibration process](image)
Calibration

- Map Relative image coordinates to World coordinates

\[ u' = u + u_0 \left( c_1 r^2 + c_2 r^4 \right) + 2c_3 u_0 v_0 + c_4 \left( r^2 + 2u_0^2 \right) \]

\[ v' = v + v_0 \left( c_1 r^2 + c_2 r^4 \right) + 2c_4 u_0 v_0 + c_3 \left( r^2 + 2v_0^2 \right) \]

\[ u_0 = u - u_c \]

\[ v_0 = v - v_c \]

\[ r = \sqrt{u_0^2 + v_0^2} \]
Calibration Procedure

• Measure a known target, let the SW crunch the data...
  – Many Software options for calibration available
Calibration – Rectification

- Calibration gives world coordinate point cloud
  - Z image plane distorted
- Rectification gives image fit for "standard" processing
  - One Z value for each grid {X,Y} coordinate

Calibrated (x, r) points

Rectified profile

Uniform pixel grid
Calibration – Rectification

Uncalibrated, non-linear depth map
Calibration -> Point Cloud

Rectification
Resampling to grid
- uniform : \( \Delta X, \Delta Y \)
In depth map
3D Imaging Methods

- Triangulation
  - Stereo
  - Structured light
    - Sheet-of-light
    - Projected patterns

- Time-of-flight

- Misc.
  - Shading
  - Focus
  - Light field
  - Interferometry
Triangulation Methods

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry
- Time-of-flight

- Shading
- Laser Triangulation
- Binary Coded
- Phase Coded

- CW
- Pulsed
Triangulation Principle

\[ \gamma = 180 - \alpha - \beta \]

\[ L_1 = \frac{B \cdot \sin \beta}{\sin \gamma} \]

Robustness:
- Large B
- Large \( \gamma \)
Laser Triangulation

3D imaging

Passive
Focus  Lightfield  Stereo

Structured Light

Active
Interferometry  Time-of-flight

Shading  Laser Triangulation  Binary Coded  Phase Coded  CW  Pulsed
Laser Line Triangulation

- Camera view
- Sensor Image
- 3D profile

Range (mm)

Width (mm)
Laser Line Triangulation
Laser Line Profile Extraction

- Each 2D intensity image -> 1 3D profile
  - High frame rate needed
  - Sensor/camera processing
    -> early data reduction
- Find peak position / column
  - High sub-pixel resolution is possible,
    e.g. Center-Of-Gravity,
    Interpolated peak position, etc.
Geometry Options 1(2)

- Vertical laser gives “natural slicing”
- $\Delta z \sim \Delta x / \sin(\alpha)$  $\Delta x$ is pixel resolution in width
- $\Delta z > \Delta x$
Geometry Options 2(2)

- Vertical camera gives good 2D imaging options
  - can give very high Z resolution
- \( \Delta z \sim \Delta x / \tan(\beta) \)
  - \( \Delta x \) is pixel resolution in width
- \( \Delta z > \Delta x \) for \( \beta < 45 \)
- \( \Delta z < \Delta x \) for \( \beta > 45 \)

Uncalibrated depth map
Laser Line Width Considerations

- **Narrow line**
  - Poor sub-pixel resolution
  - Intensity modulation effects

- **Wide line**
  - High-resolution sub-pixeling
  - In good conditions ~1/10\(^{th}\) of a pixel reasonable
  - Wide line can give artifacts...

- ~5 pixel width @ 50% of peak “ok”
Wide Laser Line Observations

- The laser covers multiple pixels ... and can hit a distance transition or intensity modulation
- Laser speckles gives noise on the peak
Use Sharp Images!

- The plane in focus is parallel to the lens and sensor planes
- We want focus on laser plane
- Tilt sensor/lens to get plane in focus

Scheimpflug Principle!
Scheimpflug in use

- Example sensor image in laser triangulation
Laser Triangulation Products

• Product examples
  – Algorithm support in vision SW packages
  – SICK Ranger/Ruler/Trispector - Proprietary CMOS sensor, multi scanning/color
  – Automation Technology - Fast CMOS sensors and FPGA processing
  – Photonfocus - Fast CMOS sensors + Lin-Log response

Booth #1655

Booth #2552
Laser Triangulation Conclusions

• Benefits
  – “Micrometer to mm” resolution scalability
  – Fast and robust
  – With Moving objects -> No additional scanning needed

• Limitations
  – Occlusion (shadow effects)
  – Laser speckles
  – Not suitable for large outdoor applications (~ > 1 m FOV)
  – Not snapshot

• Typical applications have linear object motion :
  – Log/board/veneer wood inspection
  – Electrical components / solder paste
  – Food and packaging
Stereo Imaging

\[ B \]

\[ \alpha \]

\[ \beta \]

\[ L_1 \]

\[ L_2 \]

\[ P(x, y) \]
Stereo Imaging

- Stereo is based on (at least) 2 views of a scene
  - Human vision....

- Key is matching between the images
  - But pixels are not at all unique so ...
    - Either (uniform) patches of pixels are matched or
    - Distinct features/landmarks are matched

- So, where do we match?
Where to Match?

• Lens centers and rays create a plane – Epipolar plane
  – Epipolar plane intersects sensor plane on a line
    • Match Along a line in a plane defined by Baseline & Ray
  – This is the Epipolar line
Epipolar Lines

- Unrectified
  - tilted/curved epipolar lines
Epipolar Lines

- **Unrectified**
  - tilted/curved epipolar lines

- **Rectified**
  - aligned epipolar lines

- **Find Disparity**
  - Difference in position on row
Disparity Matching

Image Patch: $f(u,v)$

Epipolar swath: $g(u\text{-disparity},v)$

Classical Matching Function Examples

Sum of Absolute Difference: $\text{SAD} = \sum(|f(u,v)-g(u\text{-disparity},v)|)$

Sum of Square Difference: $\text{SSD} = \sum(f(u,v)-g(u\text{-disparity},v))^2$
Disparity Matching

\[ f(u,v) \]

\[ g(u\text{-disparity},v) \]
Disparity Matching

Match

$g(u \text{- disparity}, v)$

Disparity

$\text{Match}$
Disparity Matching

\[ f(u,v) \]

\[ g(u - \text{disparity},v) \]

Match

Best Match
Disparity Matching

• Matching algorithm is key
  – SSD/SAD correlation are common
    • Brightness matching -> High Pass Filter
• “Coarse” pixel correlation positions
  – Interpolate to find sub-pixel matching position

• Feature matching algorithms gives sparse image data
  – High precision on found features
• Middelbury Stereo Vision Pages
  – Data sets & Comparisons – Academic...
No Structure – No 3D
Structure Comparison

No structure

Active structure
Stereo Products

- IDS - Ensenso with “noise” illumination
- Flir (Point Grey) - 2/3 cameras
- Chromasens – line scan color
- Most vision SW packages
- And many others...

Booth #2629
"One cam Stereo" - Primesense (Kinect)

- Projected pattern
  - Fixed “random” pattern
  - Pattern designed to be unambiguous
  - Pattern is “Reference Camera image”
  - IR laser diode
- Grayscale sensor for 3D triangulation
  - Generates $640 \times 480$ pixels image
  - 30 fps

- A few mm depth resolution
  - As stereo - not independent per pixel

- Heptagon Zora
  ~current version
Stereo Conclusions

• Benefits
  – Standard cameras
  – Can “freeze” a moving object/scene
  – Real snapshot
  – Good support in vision SW packages

• Limitations
  – No structure - no data -&gt; illumination constraints
  – Low detail level in X & Y – typically ~1:10 compared to pixels
  – Poor accuracy in Z
  – Limited Depth-of-field of camera

• Typical applications
  – Automotive safety/navigation
  – Traffic tolls – vehicle classification
  – Robot bin picking
Coded Structured Light

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Structured Light
- Binary Coded
- Phase Coded

Active
- Interferometry
- Time-of-flight

Shading
- Laser Triangulation

CW
- Pulsed
Coded Structured Light

- Generally called Digital Fringe projection or often “structured light”
- Light modulation:
  - Binary [Gray coded]
  - Continuous phase shift - “sinus”
Coded Structured Light Technology

• 2D Sensor to grab 3D “snapshot”
  – Pattern defines illumination angle beta

• For each pixel the illumination ray must be identified
  – With a single pattern this gives poor angular definition
    • Or usage of multiple pixels to define the illumination
  – Multiple patterns increase resolution depth dimension
Binary Coded Structured Light

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry
- Time-of-flight

Structured Light
- Binary Coded
- Phase Coded

Interferometry

Time-of-flight
- CW
- Pulsed
Binary Coded

3 patterns – 8 directions

Gray code minimizes error impact
Illustration of depth uncertainty at “110”
Phase Coded Structured Light

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Structured Light
- Shading
- Laser Triangulation
- Binary Coded

Active
- Interferometry
- Time-of-flight

Phase Coded
- CW
- Pulsed
But, for the camera it is just an intensity...
Phase Coded 2a

3 unknown:
\[ I(x,y,t) = I(x,y) + I'(x,y) \cos(\varphi(x,y,t)) \]

\[ t = 0 \]
Phase Coded 2b

3 unknown:
\[ I(x,y,t) = I(x,y) + I'(x,y) \cos(\phi(x,y,t)) \]

\[ t = 1 \]
Phase Coded 2c

3 unknown:
\[ I(x,y,t) = I(x,y) + I'(x,y) \cos(\varphi(x,y,t)) \]

\[ t = 2 \]
Phase Coded 2d

3 unknown:
I(x,y,t) = I(x,y) + I'(x,y)\cdot\cos(\varphi(x,y,t))

Analytical expression in each pixel
-> range, modulation, background

More common:
4 patterns with 90 degree separation
-> Simpler math & more robust
Phase Unwrapping

- High frequency -> High accuracy – Ambiguous
- Low frequency – Low accuracy – Unambiguous
- Combine results to unwrap
- In theory 2 frequencies are typically enough
- Typically 4-5 frequencies -> ~ 15-20 images / “snap”
- Coarse binary patterns + high frequency phase coded common
Conclusions Coded Structured Light

• Commercial system examples
  – ViaLux Z-snapper
  – LMI Gocator
  – Shape Drive

• Benefits
  – Very good 3D measurements, with quality measure
  – Independent measurement in each sensor pixel
  – Fast – “almost snapshot”

• Limitations
  – Needs static scene during multiple projection capture
  – The dynamic range in each pixel must be enough to make the phase calculation
    • Ambient, low/high reflection and specularities limit
      – 2 cameras common to overcome this
    • Large FOV difficult to realize.

• Typical applications
  – Reverse engineering shape capture
  – Medical imaging
  – Electronics inspection
High-speed / Hybrid

- **Fraunhofer**
  - Gobo-projector
    - Rotates fixed patterns
  - 360 Hz patterns
    - 36 Hz 3D
  - over 1 KHz 3D presented
- **Numetrix**
  - Reduced #exposures via beamsplitters/color separation
- **IDS X-series**
  - Pattern combines random dots and high frequency sinus, few shifts -> coarse and fine
Baseline vs Accuracy

- Baseline is distance between sensor and illumination or between cameras
- A larger baseline gives larger displacement on the sensor per $\Delta z$
  - Better resolution / accuracy
- A larger baseline gives more differences between the “views”
  - More areas not seen by both cameras - occlusion
  - Less accurate matching, especially for rounded structures and tilted surfaces
Occlusion Illustration

Intensity

Range

Camera Occlusion

Illumination Occlusion
Ambient Handling

• Ambient light not good!
Ambient Handling

- Ambient light not good
  - Interference filter on camera
Wavelength

- Focussing limits proportional to wavelength
  - Speckle size too
- IR: Invisible, but poor focussing, Eye safety issues
- Red: Cheap lasers, high CMOS/CCD sensitivity, high ambient
- Blue: Good focussing, less ambient, expensive

Comparison laser triangulation:

405 nm, 20 micron width
635 nm, 25 micron width
General Conclusions Triangulation

• Most common 3D principle
  – "Simple" methods
  – Robust if active
  – Reasonably fast
  – Reasonably accurate

• Difficult to scale to distances more than a meter or two
  ... which leads us to
Time-of-flight

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry

Time-of-flight
- CW
- Pulsed

Binary Coded
- Phase Coded
Time-of-flight

- Pulsed
  - Send a light pulse – measure the time until it comes back
  - Light speed 0.3 Gm/s ...
    at 1 m it comes back after ~7 ns
  - Measure “indirect” delay time
- CW - Continuous Wave
  - Modulated continuous illumination
    - Phase shift ~distance
  - Used in most TOF imager arrays
  - Low resolution due to complex pixels
- ~ a few mm-cm depth resolution
TOF CW

3D imaging

Passive
- Focus
- Lightfield
- Stereo
- Shading
- Laser Triangulation

Active
- Structured Light
- Interferometry
- Time-of-flight
- CW
- Pulsed
TOF with CW Modulated Light Source

• Modulate the light source intensity
  – Distance = Phase shift
  – e.g., \( f = 30 \text{ MHz} \Rightarrow 5 \text{ m} \)
    range ambiguity limit

• “4 capacitors per pixel”
  – one 90° phase interval each
  – Integrate for many periods
    • e.g., 20 ms \( \Rightarrow 5 \text{ ms/capacitor} \)
    – Find phase \( \varphi \) from the 4 values

• Wrapping problem for distances larger than e.g. 5 m

\[
d = c \frac{\varphi - \varphi_0}{4\pi f}
\]
Kinect One

- 512x424 @30 Hz
- Multi frequency CW
- Multi-exposure HDR
- SDK available
  - Not industrial...

TOF Pulsed

3D imaging

Passive
- Focus
- Lightfield
- Stereo

Active
- Structured Light
- Interferometry
- Time-of-flight

Structured Light
- Binary Coded
- Phase Coded

Interferometry
- CW
- Pulsed
Pulsed TOF Shutter Principle

Relationship between Gated and Full gives range

X-Y resolution today ~Megapixel, but Z resolution not as good as for CW.
TOF Array Conclusions

• Pulsed 2D array:
  – Basler, Odos & Fotonics VGA/XGA announced
    • 3D + color option

• CW 2D array
  – SICK 3vistor-T ~150x150 pixels
  – IFM Efector ~150x150 pixels

• Benefit
  – Snapshot

• Basic limitations
  – Z resolution > cm
  – X-Y resolution (CW)
  – Secondary reflections (CW)
  – Fast movements
  – Ambient light
  – Intra scene dynamic range

• Typical applications:
  – Gaming
  – People counting
  – Automatic milking machine
  – Navigation
Technology Comparison 1

- A test scene with a mix of objects & materials
  - ~1x1 m, cameras ~2 m away
Technology Comparison 2

TOF 3D
“Active”
Stereo 3D
Technology Comparison 4

Laser Triangulation 3D
Technology Comparison 6

Phase Coded 3D

3D TOF

Laser Triangulation 3D

Active Stereo 3D
Cross Section Boxes:
Phase pattern Projection,
Laser Triangulation
Stereo
TOF

~10 mm
Misc. 3D Methods

• Less common
  – Interesting theory
  – Special cases
Map of 3D

- 3D imaging
  - Passive
    - Focus
    - Lightfield
    - Stereo
  - Active
    - Structured Light
      - Laser Triangulation
      - Shading
    - Interferometry
      - Binary Coded
      - Phase Coded
    - Time-of-flight
      - CW
      - Pulsed
Shape from Shading

• Gives shape information, but not real distance
  – Shade from different directions of illumination gives surface orientation information
  – Integrating the orientation gives depth variations

• Limitations
  – Only surface orientation, no actual depth
  – No discontinuities allowed
Light-Field 3D 1

• Micro lens array used to create "4D" light-field image on standard image sensor
  – 2D direction "subpixels" in each 2D "pixel"
Light-Field 3D 2

- Processing of light-field image
  - Refocussing
  - 3D calculation
- Cameras – Raytrix
  - AIT Multi-line linescan
- Features
  - "No occlusion"
- Limitations
  - Depth accuracy "lens aperture triangulation"
  - Special cameras
  - Complex processing
Depth from Focus

• Grab a sequence of images focused from A to B
• Scan through the stack and find where local focus is maximized
  – That gives the range

• Features
  – No occlusion
  – No structured illumination needed

• Limitations
  – Slow
  – Needs structure to estimate focus
  – Pixel regions needed to estimate focus
  – Poor accuracy
    • “Triangulation using lens aperture”
Interferometry 1

Coherent (Laser) Light
- Periodic Interference
-> Flatness measurements

Incoherent (White) Light
- Interference @ same distance
-> Shape measurements
• **Features**
  – Sub-micron accuracy

• **Limitations**
  – Complicated scanning mechanics
  – Static scene needed during scan
3D Applications

Packaging

Electronics

Wood

Robotics

Printing

Transport

Logistics

Food

Automotive
3D Technology Overview

- Interferometry
- Coded Structured Light
- Laser Triangulation
- Stereo
- Time Of Flight

Z Resolution / Accuracy vs Distance / FOV size
Application Discussion 1

• Application requirements complex
  – What are the requirements for example, for a “good cookie”?  
  – Iterative requirement work and testing a good way forward

• Basic requirements
  – Cost!
  – FOV size
  – Acquisition speed / object movement
  – Resolution X-Y-Z and accuracy requirements
    • Sampling theorem : at least (defect size) / 2 pixel size
  – Classification never 100% Detect Error – “Positive” :
    • Reject a Good Part : False Positives
    • Accept a Bad Part : False Negatives
  – Acceptance - define procedure, test objects and results.
  – Environment – ambient and size limitations, laser class limitations
Application Discussion 2

• Technology selection
  – Which technology would fit best?
    • Will the technology I have in my toolbox work?

• Early test
  – Try to get 3D data to prove/disprove visually the basic requirements
    • Can the defect be seen?
    • Can I see all aspects without occlusion?
    • Do I have enough signal without eye safety/cost issues?

• Don’t reinvent the wheel!
  – Buy the best subsystems for the application
Processing Software Options

- **MVTec Halcon**:  
  - Very complete SW library, good 3D camera drivers  
    - Booth #567
- **Matrox MIL**:  
  - Software, Cameras & Vision processors  
    - Booth #2424
- **AqSense SAL 3D**:  
  - Dedicated laser profiling SW & 3D shape matching (bought by Cognex)
- **Stemmer CVB**:  
  - A lot of tools
- **Open SW**:  
  - Point Cloud Library: Extensive “big data” processing  
    - OpenCV: Camera calibration, not much 3D
- **And many more...**
3D Camera Standard!

- Explicit 3D support in vision standards underway!
  - GenICam Feature definitions in place
  - GigE Vision support est Q2 2017

Companies using these standards include:

- MathWorks®
- STEMMER Imaging
- Active Silicon
- National Instruments
- Pleora Technologies

International Vision Standards
Booth #2921
A few App Examples
3D OCR / Code Reading

• VIN number stamped into car chassis
• Tire codes
"Backwards" Examples

Small FOV TOF 3D
- Milking Robots (LMI / Mesa)

Large FOV laser triangulation
- Timber truck load volume (SICK Ranger)
Road/Rail Inspection

- 3D laser line triangulation + line scan intensity/color
Train Inspection
Logistics with TOF

- Measure volume and size of box on pallet or conveyor

![Diagram showing volume and size measurement](image-url)
Robot Vision and 3D

• Random bin picking an old "Holy Grail"
• Overhead 3D vs "hand 3D"
• Main problems:
  – Object location / description
    • Geometrical primitives
    • CAD models
  – Finding pick point
  – Controlling robot

• ... Finally, general systems coming
Bin Picking in Action
3D Bin Picking System Example

• ScanningRuler sweeps laser over the scene
  – Complete 3D image

• Bin-picking application
  – Co-register coordinate system of camera system and robot
  – Estimate pose of picking candidates in 3D data
  – Ensure collision free gripping of the part
Finally

• Any questions ??

Mattias Johannesson
Expert 3D Vision, Core Design Identification and Measuring
Mattias Johannesson
Expert 3D Vision, Core Design Identification and Measuring

SICK IVP AB
Wallenbergs Gata 4
583 30 Linköping
Sweden

Phone: +46 13 362142
Email: mattias.johannesson@sick.com

www.sick.com