3D Vision System Development

Mattias Johannesson
Expert, 3D Vision
SICK
Agenda

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

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

• **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
Ranging Techniques

• Passive range imaging
  – No dedicated light source
  – Very, very difficult in a real world...

• Active range imaging
  – Dedicated illumination
  – Typically defined spatially Where
    • Triangulation
  – Or defined temporally When
    • Time of flight principles
Acquisition Speed

• Range imaging is typically not snapshot as 2D cameras

• Snapshot \( O(1) \)
  – Stereo
  – Primesense / "Kinect"
  – Time-of-flight array camera

• "Almost" snapshot \( O(\log(N)) \)
  – Structured light projection
  – Moving camera stereo

• 1D scanning \( O(N) \)
  – Linear movement of object / illumination
  – 1D scanning (depth from focus, interferometry)

• 2D scanning motion \( O(N \times M) \)
  – 2D scanner
  – Linear movement of object + 1D scanning
Accuracy

- **Resolution**
  - Pixel size $\Delta X$, $\Delta Y$
  - The min $\Delta Z$

- **Repeatability**
  - First step to accuracy

![Accuracy Diagram](image)
Accuracy

- **Resolution**
  - Pixel size $\Delta X$, $\Delta Y$
  - The min $\Delta Z$

- **Repeatability**
  - First step to accuracy

![Accuracy Diagram](image-url)
• **Resolution**
  – Pixel size $\Delta X$, $\Delta Y$
  – The min $\Delta Z$

• **Repeatability**
  – First step to accuracy

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

• Map Relative image coordinates to World coordinates
  – Lookup Table or Model

\[
\begin{align*}
    u' &= u + u_0 \left( c_1 r^2 + c_2 r^4 \right) + 2 c_3 u_0 v_0 + c_4 \left( r^2 + 2 u_0^2 \right) \\
    v' &= v + v_0 \left( c_1 r^2 + c_3 r^4 \right) + 2 c_4 u_0 v_0 + c_3 \left( r^2 + 2 v_0^2 \right) \\
    u_0 &= u - u_c \\
    v_0 &= v - v_c \\
    r &= \sqrt{u_0^2 + v_0^2}
\end{align*}
\]

\[
\begin{pmatrix}
X \\
Z \\
s
\end{pmatrix} = H \begin{pmatrix}
    u' \\
    v'
\end{pmatrix} =
\begin{pmatrix}
    b_{11} & b_{12} & b_{13} \\
    b_{21} & b_{22} & b_{23} \\
    b_{31} & b_{32} & b_{33}
\end{pmatrix} \begin{pmatrix}
    u' \\
    v' \\
    1
\end{pmatrix}
\]
Calibration Procedure

• Measure a known target, let the SW crunch the data...
 Calibration – Rectification

• Calibration gives world coordinate point cloud
  – Z image plane distorted

• Rectification gives image fit for "standard" processing
  – 1 Z value for each “integer” {X,Y} coordinate
Calibration – Rectification

Uncalibrated, non-linear

Calibration -> Point Cloud

Rectification
Resampling to grid
- uniform: $\Delta X$, $\Delta Y$
3D Imaging Methods

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

- Time-of-flight

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

\[ \gamma = 180 - \alpha - \beta \]
\[ L_1 = \frac{B \sin \beta}{\sin \gamma} \]

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

• "Simple" special case of "structured light"
Laser Line Triangulation 1

- 2D sensor + laser line
- Each 2D intensity image -> 1 3D profile
  - 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.
Laser Line Width

• **Wide line - high-resolution sub-pixeling**
  – Typical at least 3-5 pixels needed
  – Wide line can give edge artifacts

• **Narrow line – less edge artifacts on range**
  – Intensity modulation effects
  – Poor sub-pixel resolution

\[ \Delta z = 1 \text{ pixel} \]

Intensity image from laser line
Peak Finding Issues

- The laser covers multiple pixels ... and can hit a distance transition or intensity modulation
- Laser speckle gives noise on the peak
Ambient Handling

- Ambient light not good!
Ambient Handling

- Ambient light not good
  - Interference filter on camera
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
- \[ \Delta z \sim \frac{\Delta x}{\tan(\beta)} \]  
  \( \Delta x \) is pixel resolution in width
- \( \Delta z > \Delta x \) for \( \beta < 45 \)
- \( \Delta z < \Delta x \) for \( \beta > 45 \)
• Tilt sensor to focus laser plane
  – $\tan(\phi) \approx f \cdot \tan(\gamma) / a$
Laser Triangulation Products

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

Booth #706
Booth #500
Laser Triangulation Conclusions

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

• Limitations
  – Occlusion
  – Not suitable for large outdoor applications (\(\sim > 1 \text{ m FOV}\))
  – Not snapshot

• Typical applications
  – Log/board/veneer wood inspection
  – Electrical components / solder paste
  – Food and packaging
Stereo Imaging 2

- 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 patches of pixels are matched or
    - Distinct features/landmarks are matched

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

• Simple 1D case: Along a line in the plane
Epipolar Lines

- Unrectified
  - tilted/curved
  - epipolar lines
Epipolar Lines

• Unrectified
  – tilted/curved
  – epipolar lines

• Rectified
  - aligned epipolar and matching

• Find Disparity
  – Difference in position
Disparity Matching

Image Patch : $f(u,v)$

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

Matching Function : Sum of Absolute Difference, SAD

$$\sum(|f(u,v)-g(u\text{-disparity},v)|)$$
Disparity Matching

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

\[ \text{Match} \sim \text{Sum}(|f(u,v)-g(u\text{-disparity},v)|) \]
Disparity Matching

\[ f(u,v) \]

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

Disparity

Best Match

Match

disparity

disparity
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
No Structure – No 3D
Structure – 3D
No Structure – 3D

No structure           Active structure
Stereo Products

- IDS - Ensenso with “Kinect” illumination
- Point Grey - SAD correlation, 2/3 cameras
- Chromasens – line scan color
- Most vision SW packages
- And many others...

Booth #611
Booth #516
Booth #520
Stereo Conclusions

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

• **Limitations**
  – No structure - no data -> illumination constraints
  – Low detail level in X & Y – typically ~1:10 compared to pixels
  – Poor accuracy in Z

• **Typical applications**
  – Traffic tolls – vehicle classification
  – Robot bin picking
  – Automotive safety/navigation
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 in all dimensions
Fixed Pattern - Primesense (Kinect)

• Projected pattern
  – Fixed “random” pattern
  – Pattern designed to be unambiguous
  – IR laser diode

• Grayscale sensor for 3D triangulation
  – Generates 640 × 480 pixels image
  – 30 fps

• Additional standard RGB sensor
  – 640 × 480 pixels
  – For face recognition, etc.

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

• Other vendors, e.g. Asus
  – Future with Apple ??
Spatio-temporal Coded Light

• Generally called Digital Fringe projection or simply “structured light”

• Light modulation:
  – Binary [Gray coded]
  – Continuous phase shift - “sinus”
Binary Coded

3 patterns – 8 directions

Gray code minimizes error impact
Phase Coded 1

Projection direction

Sinus: phase is range

Intensity
Phase Coded 2

Intensity

Pattern 1 Pattern 3 Pattern 2

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

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

Pattern 1 : Shift 0
Pattern 2 : Shift 120 degrees
Pattern 3 : Shift 240 degrees

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
Examples Coded Light
Conclusions Coded Light

• Commercial system examples
  – ViaLux Z-snapper
  – LMI Gocator
  – Numetrix
  – 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
    • Large FOV difficult to realize.

• Typical applications
  – Reverse engineering shape capture
  – Medical imaging
  – Electronics inspection
Triangulation General

• Discussion on
  – Baseline - accuracy
  – Occlusion
  – Wavelength
Baseline vs Accuracy

• Baseline is distance between sensor and illumination or between cameras
• A larger baseline gives larger displacement 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
Wavelength

- Focussing limits proportional to wavelength
  - Speckle size too
- IR: Invisible, but poor focussing
- Red: Cheap lasers, high CMOS/CCD sensitivity, high ambient
- Blue: Good focussing, less ambient, expensive
- Comparison laser triangulation:
General Conclusions Triangulation

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

• Difficult to scale to distances more than a meter or 2
  ... which leads us to
**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
  - ~Millimeter resolution in laser scanners
  - Shutter-based used in imagers
- **CW - Continuous Wave**
  - Modulated continuous illumination
  - Phase shift ~distance
  - Used in most TOF imager arrays
  - Low resolution due to complex pixels
  - Similar ambiguity problem as phase coded structured light
Pulsed TOF Shutter Principle

Emitted pulses

Reflected pulses

Shutter

Near

Far

Intensity

“Early”

“Early”

Relationship between “Early” and Intensity gives range
Pulsed TOF Camera

- Odos imaging has new 4 Mpixel pulsed TOF camera
  - Z accuracy 2-12 cm
  - up to 470 fps

Booth #202
Kinect One

- Now with TOF!
- ~500x400 pixels
- CW / pulsed?
TOF Array Conclusions

• Pulsed 2D array : Odos, 4 Mpixel
• CW 2D array “3D cameras”
  – Mesa Swiss Ranger ~150x150 pixels
  – IFM –”smart camera” ~60x50 pixels

• Benefit
  – Snapshot

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

• Typical applications :
  – People counting
  – Automatic milking machine
Technology Compare 1

Laser triangulation

Scanned TOF (Phase)

Passive stereo

Laser intensity
Technology Compare 2

- Laser triangulation
- Scanned TOF (phase)
- Active stereo

Triangulation, TOF, Stereo
Misc. 3D Methods

- Less common
  - Interesting theory
  - Special cases
  - On Bonus Slides!
3D Camera Standard?

- Explicit 3D support in vision standards underway!
- Companies working include:
  - Cameras:
    SICK, LMI, Point Grey, Automation Technology, Odos
  - SW (Cameras):
    MVTec, Matrox, Teledyne DALSA, STEMMER
3D Applications

- Packaging
- Electronics
- Wood
- Robotics
- Printing
- Pharmaceutical
- Logistics
- Food
- Automotive
Robot Vision and 3D

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

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

- Scanning ruler 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
3D OCR / Code Reading

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

Small FOV TOF 3D
- Milking Robot (LMI)

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

• 3D laser line triangulation + line scan intensity/color
Logistics with TOF

- Measure volume and position on pallet or conveyor
3D Technology Overview

- Interferometry
- Phase coded Structured Light
- Laser Triangulation
- Triangulation
- Stereo
- Time Of Flight
Application Discussion

• 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
  – Resolution X-Y-Z and accuracy requirements
    • Sampling theorem: at least (defect size) / 2 pixel size
  – Acquisition speed / object movement
  – Error tolerance: False Positives / False Negatives?
  – Environment – ambient and size limitations, laser class limitations

• 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?
Finally

- Any questions ??

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

3D data from scanning ruler – laser triangulation
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
Bonus slides

• Collected here to show if time allows
• Printed also “in logical place” in the handouts
Kinect Application Sample

- MvTec Halcon Image Processing SW example – Finding Box
Binary Coded

Standard binary code
- Many bits switch at same position
TOF With CW Modulated Light Source

- Modulate the light source
  - e.g. $f = 30 \text{ MHz} \Rightarrow 5 \text{ m range ambiguity limit}$
- 4 capacitors per pixel
  - Integrating one $90^\circ$ phase interval each
- Integrate for many cycles
  - e.g. $20 \text{ ms} \Rightarrow 5 \text{ ms/capacitor}$
- Find phase $\phi$ from the 4 values
  - $\phi_0$ from pixel with direct optical feedback
- Wrapping problem for distances larger than 5 m
- Problems with ambient, reflections, saturation, motion ...

\[
d = c \frac{\phi - \phi_0}{4\pi f}
\]
Misc. 3D Methods

• Less common
  – Interesting theory
  – Special cases
  – On Bonus Slides!
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"

Image taken by Todor Georgiev, Adobe Systems, with his plenoptic camera.
Light-Field 3D 2

- Processing of light-field image
  - Refocussing
  - 3D calculation

- Cameras – Raytrix (Lytro)

- Features
  - "No occlusion"

- Limitations
  - Depth accuracy
    "lens aperture triangulation"
  - Special cameras
  - Complex processing
  - Lytro : no 3D, not industrial
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
Interferometry 2

• Features
  – Sub-micron accuracy

• Limitations
  – Complicated scanning mechanics
  – Static scene needed during scan