Object Recognition in Augmented Reality

Tomáš Fabián
VSB-TUO, FEECS
Department of Computer Science

2017 Czech-Austrian Summer School on "Deep Learning and Visual Data Analysis"

Our work presented here was partially supported by the
EU H2020 686782 PACMAN
project solved with Honeywell
http://mrl.cs.vsb.cz/h2020

Clean Sky 2
JOINT UNDERTAKING

Agenda

- Brief Introduction to AR
  - Terminology
  - Related devices
- Depth Sensors
  - Sampling rate issues
- Main Challenges of AR
  - Object pose estimation pipeline
- Example
  - A brief example of AR application

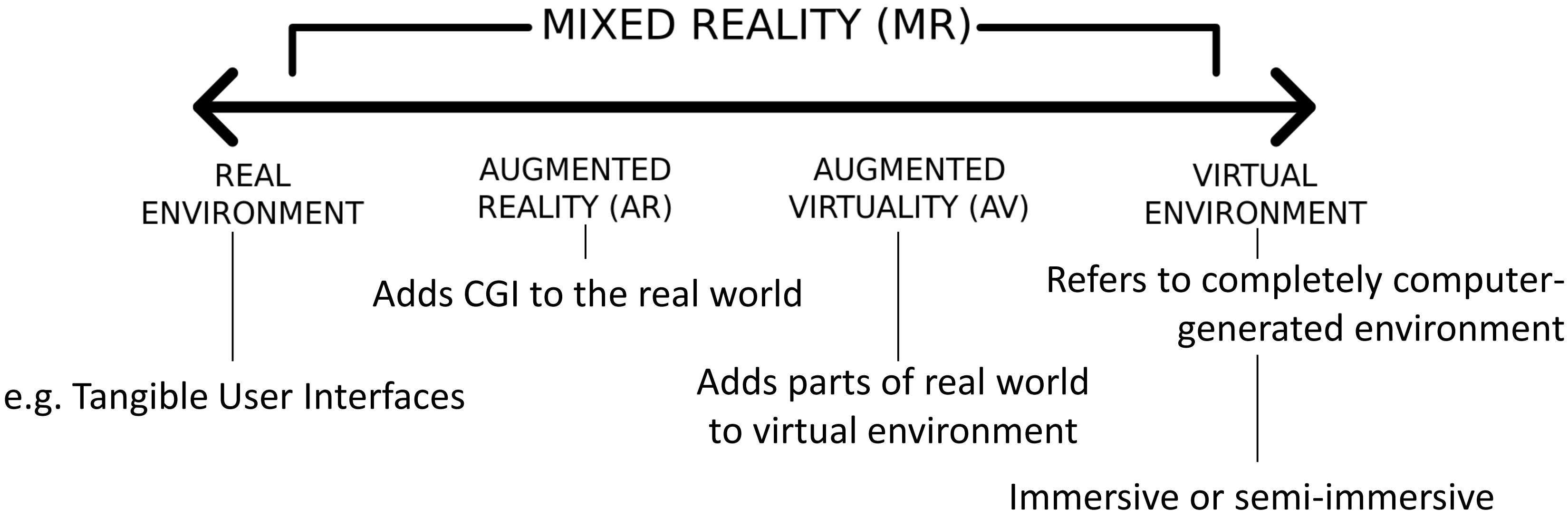
Buzzwords


computer-generated
environment
enhancements
perception
user's
understandings
direct
origin
Unlike
features
viewed
Virtuality
timely
differentiates
real-world
adds
current
procedure
accessible
text-notification
topical
provide
technology
create
interaction
relevant
Virtual
Offline
objects
surgical
simulation
touch
Augmented
business
physical
camera
simply
Merged
application
form
apps
instruction
smartphone
Augmediated
entirely
applications
Mixed
simple
certain
overlay
inhabit
Real
new
integrated
enhance
real
Cell
highlight
uses
add
assistance
reality
something
also
activities
created
one's
tool
thus
view
users
everyday
overlays
improved
point
many
used
top
image
existing
audio
key
highly
driving
live
data
device
worlds
want
case
harmoniously
integrates
version
use
can
complicated
graphics
Modified
virtual
Applications
means
Everyday
text
versus
value
coexist
Online
world
called
Environment
phones
feedback
natural
views
enhancing
development
enhanced
indirect
added
provided
word
Modulated
images
perform
experience
augment
digital
Mediated
user's
element
Dimished
real-time
life-threatening
sounds
environments
requires
superimposed

Ontology of AR

- Milgram's virtuality continuum (VC)

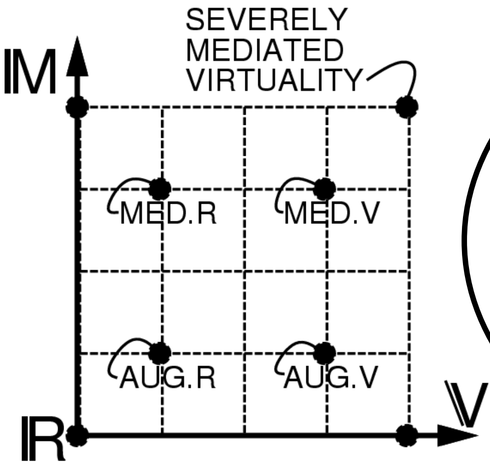
Milgram and Kishino (1994)



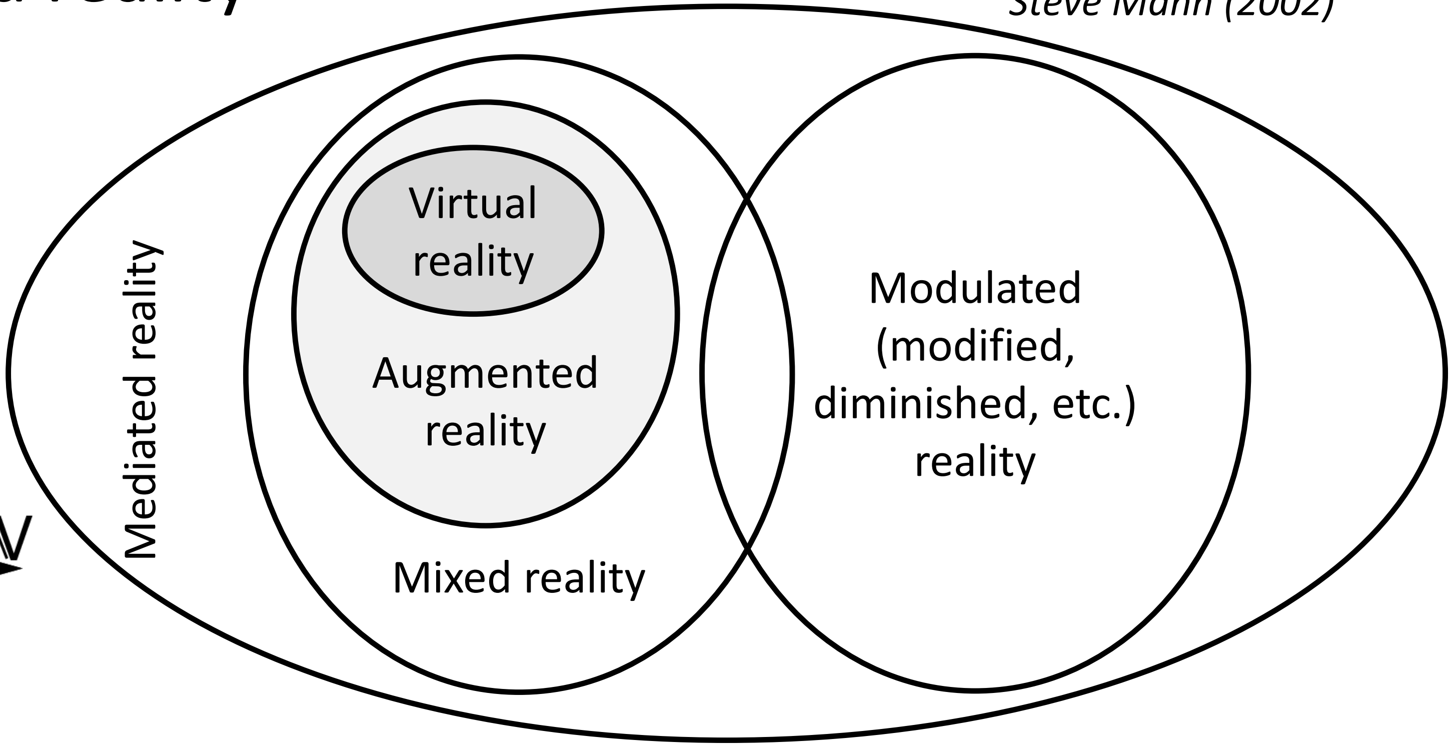
  - AR are technologies allowing users to operate in real world with additional information provided through artificial means

Ontology of AR

- Mediated reality



Steve Mann (2002)



  - Second axis (Mediality) denotes modification of reality or virtuality
  - Diminished reality (DR) is the direct opposite of AR

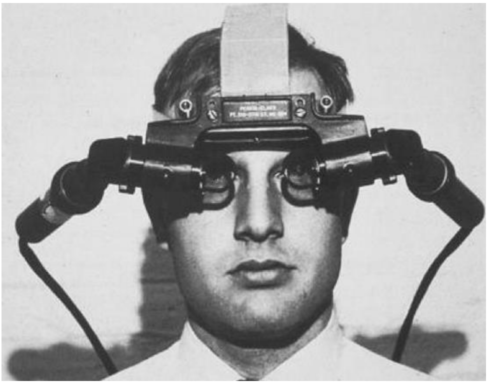
Quick History

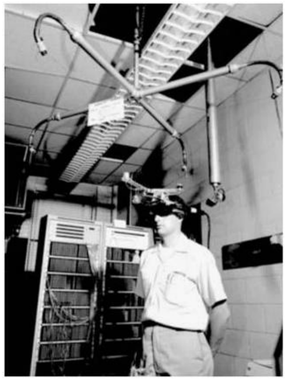
- Early Years of VR/AR: The Sword of Damocles

  “The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.“

  Ivan E. Sutherland, The Ultimate Display, 1965

  – The first VR head-mounted display (HMD)



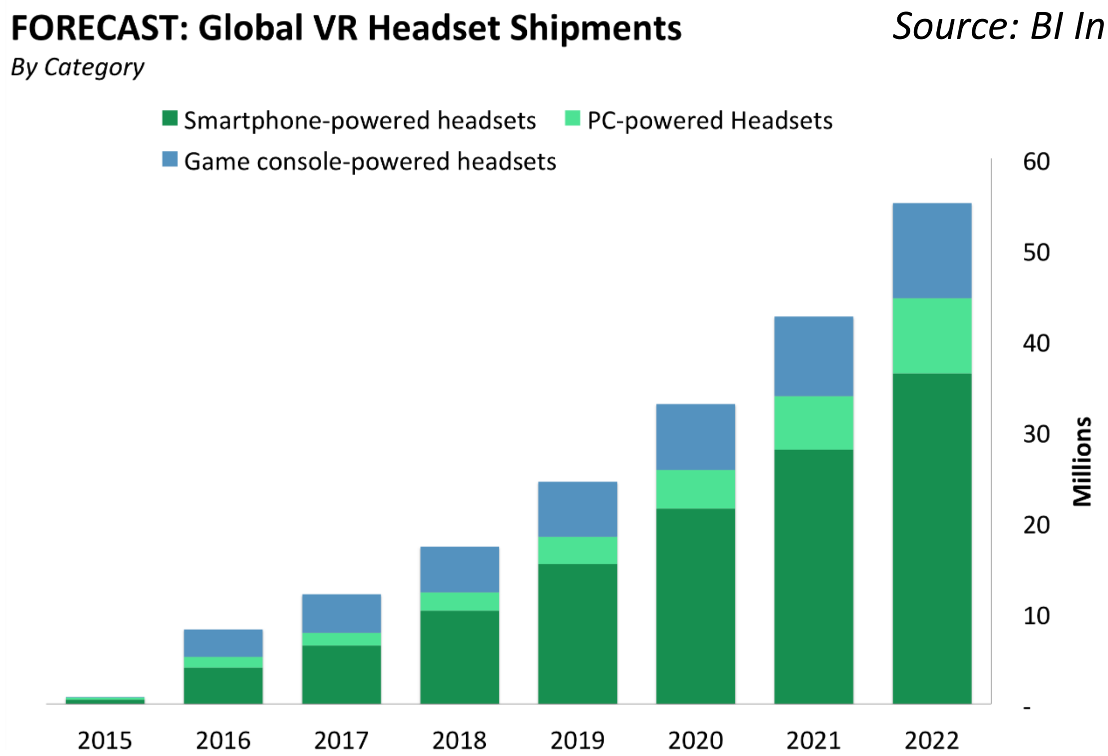


Quick History

- …and now, 50 years later?

  Yet still far from this ultimate goal of computer controlled matter but much has changed

Source: BI Intelligence Estimates



Head Mounted Display

- A variety of near-to-eye devices and head-mounted displays are widely available now…

Meta2 DK

Epson Moverio BT-300

MS HoloLens DE

Head Mounted Display

HMD

Display
LCD, OLED, refresh rate, optics FOV etc.

Head Tracking
Use multiple sensors for positional calculation

Eye Tracking
Interaction with UI, depth of field

Other Hardware
Chassis, head strap etc.

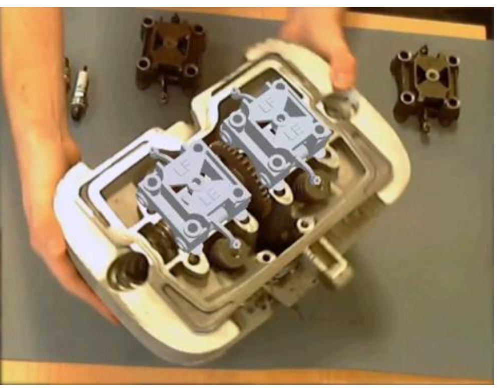
Audio Hardware
Multi-speaker audio

Computer Hardware
Self-contained vs. remote

Companion Input Devices
e.g. motion tracked controllers

Available Apps and SDKs

- Many SDKs and their application are around…



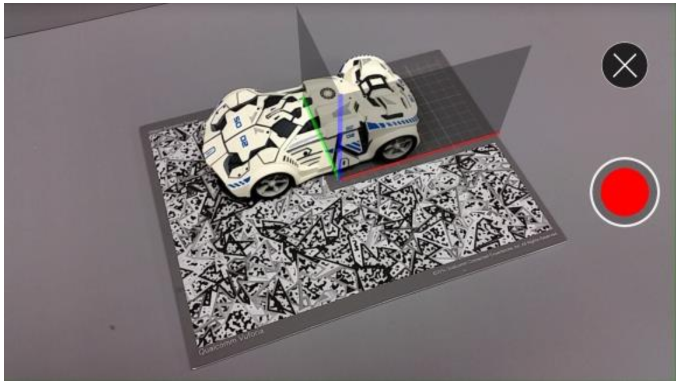
vicomtech
visual interaction & communication technologies

DiOTA
AUGMENTING INDUSTRIES

craftar
BY catchoom

Available Apps and SDKs

- Many SDKs and their application are around…



AR-media™ AUGMENTED REALITY MEDIA

wikitude See more.

vuforia™

Available Sensors

- Color Sensors
  - Widely available matured technology (CMOS prevails)
  - No geometric distortion, i.e. deviation from rectilinear projection (calibration required)
  - Low presence of image noise, i.e. random variation in brightness (well lit scene required)
  - Rolling shutter causes distortion of fast-moving scenes
- Depth Sensors
  - Provide valuable range information
  - Low cost consumer-grade devices are available
  - But…

Available Sensors

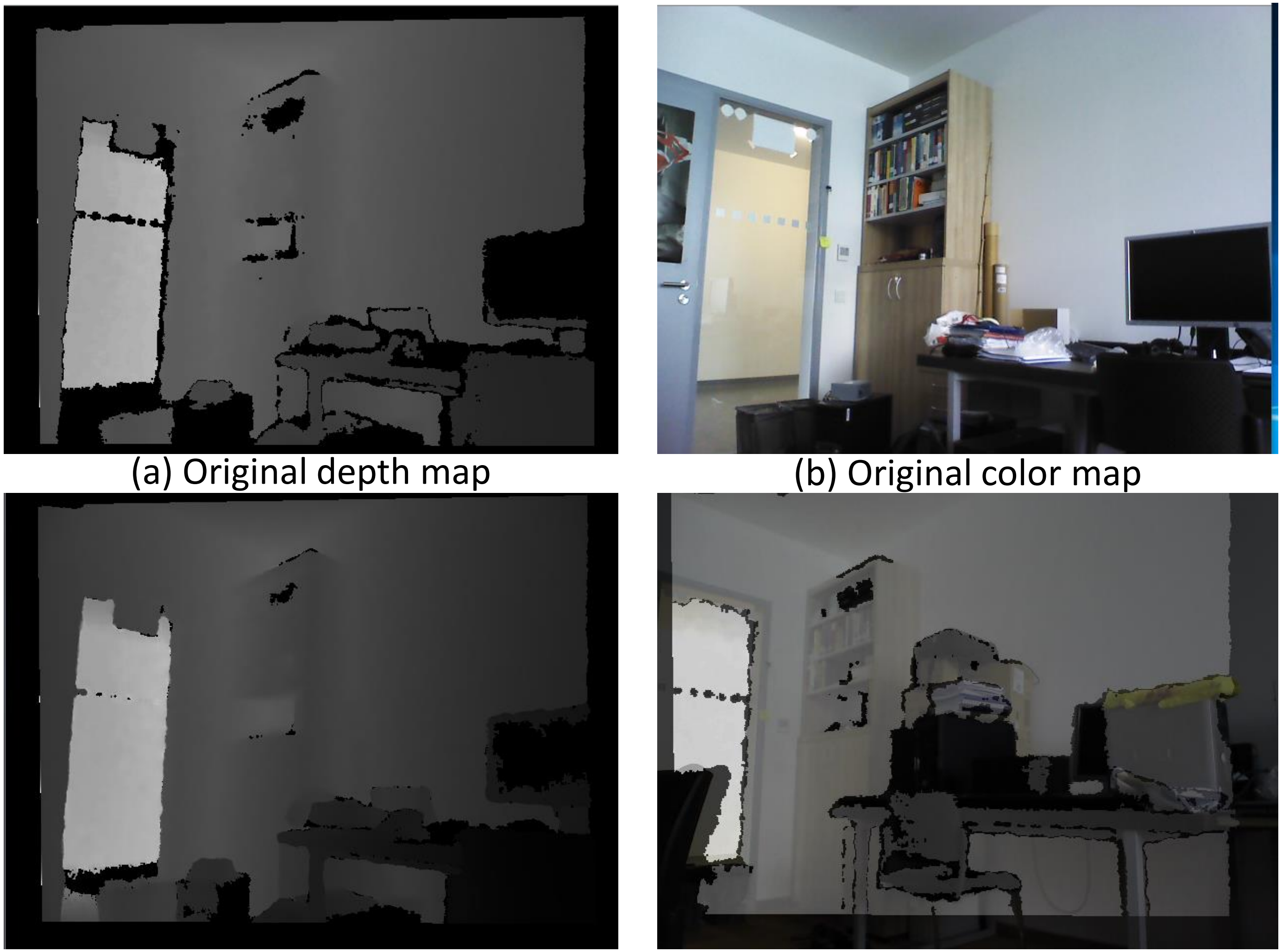
- Orbbec Astra

„Astra is developed to be highly compatible with existing OpenNI applications, making 3D camera ideal for pre-existing apps that were built with OpenNI."

Specifications (just an excerpt, more on Orbbec web site)
Dimensions 160 × 30 × 40 mm
Weight 0.3 kg
Range 0.4 – 8 m, optimized 0.6 – 6 m (0.3 – 5.8 m Astra S)
Depth Image Size 640×480 (VGA) 16 bit @ 30 FPS
RGB Image Size 1280×960 @ 10 FPS
Field of View 60° horiz. × 49.5° vert. (73° diagonal)
Data Interface **USB 2.0**
Microphones 2
Operating Systems Windows, Linux, Android
Power USB 2.0 (Full Power 2.2 W, Standby Power 1.5 W)
Software Orbbec Astra SDK + OpenNI
Low HW requirements
Price 150 USD

Structured light
40 000 beams at 800 nm

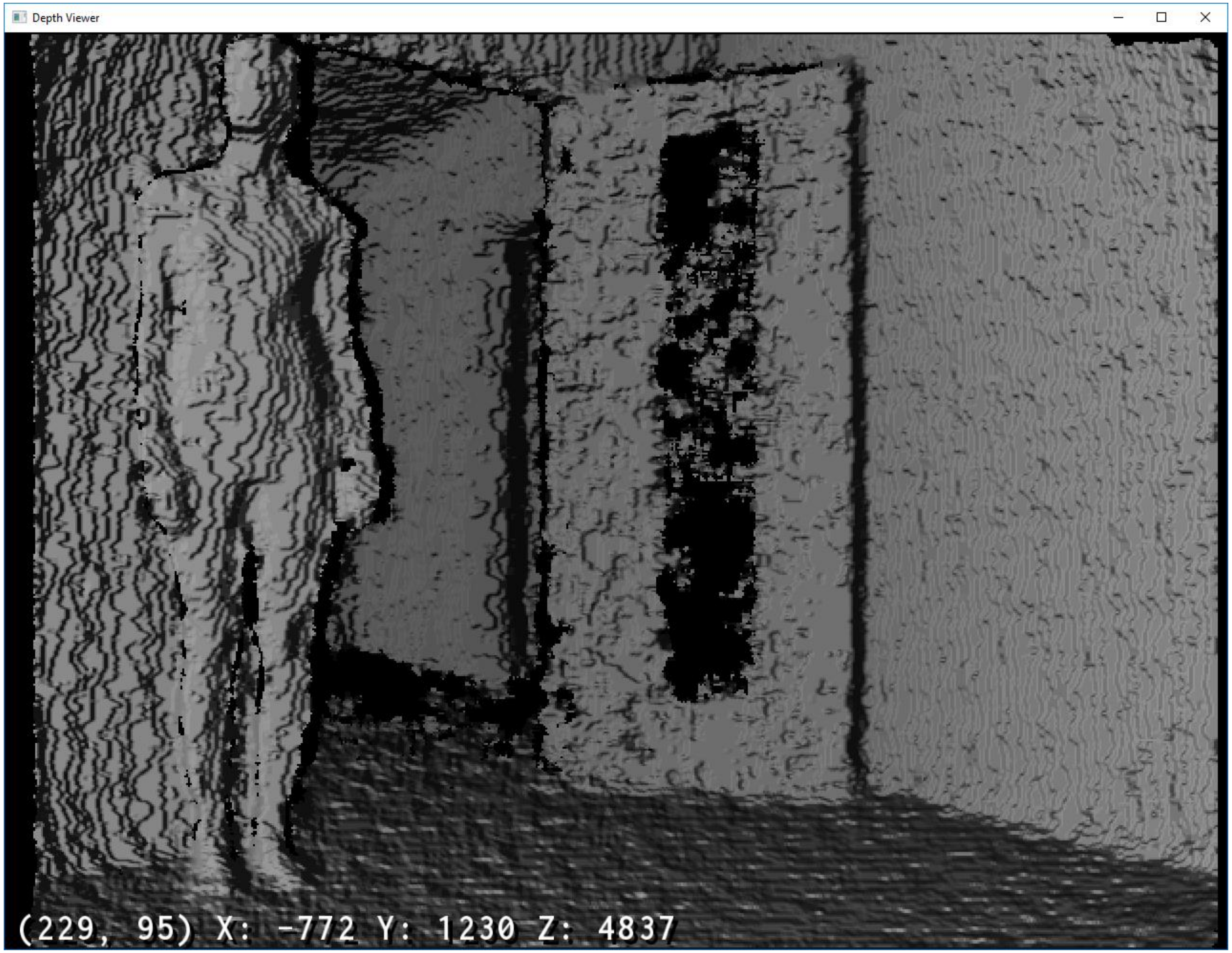
Color and Range Images



(a) Original depth map (b) Original color map

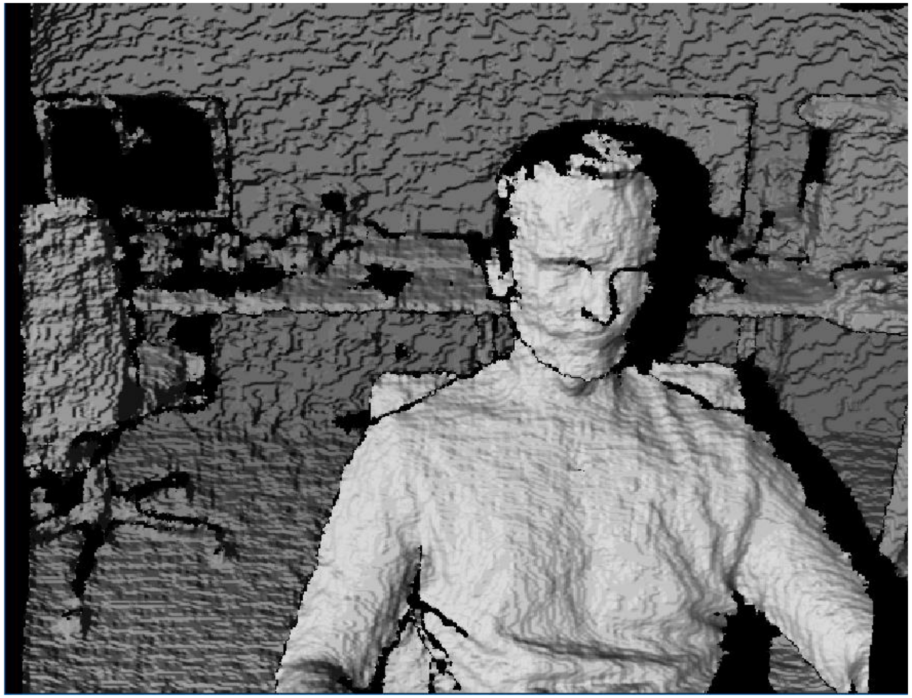
(c) Filtered depth map (d) Depth-color registration

Range Images

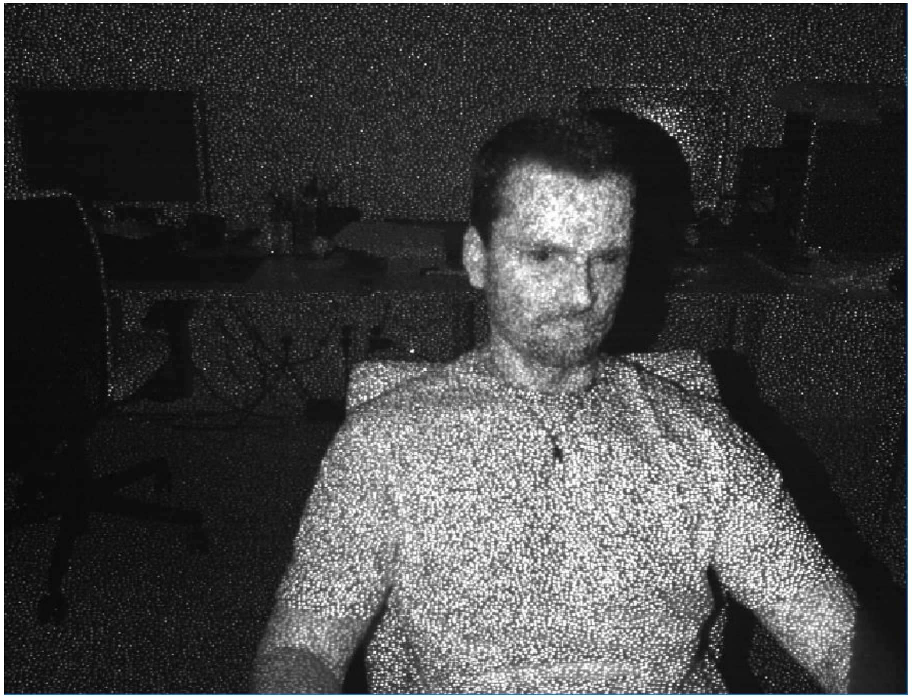


Z represents actual distance in mm (integers)

Range Images

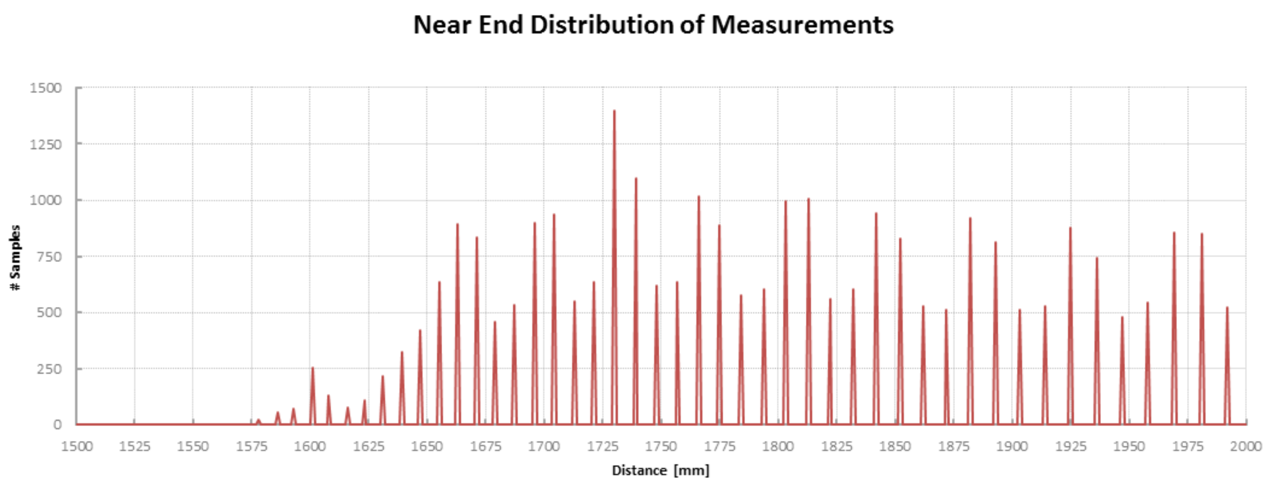


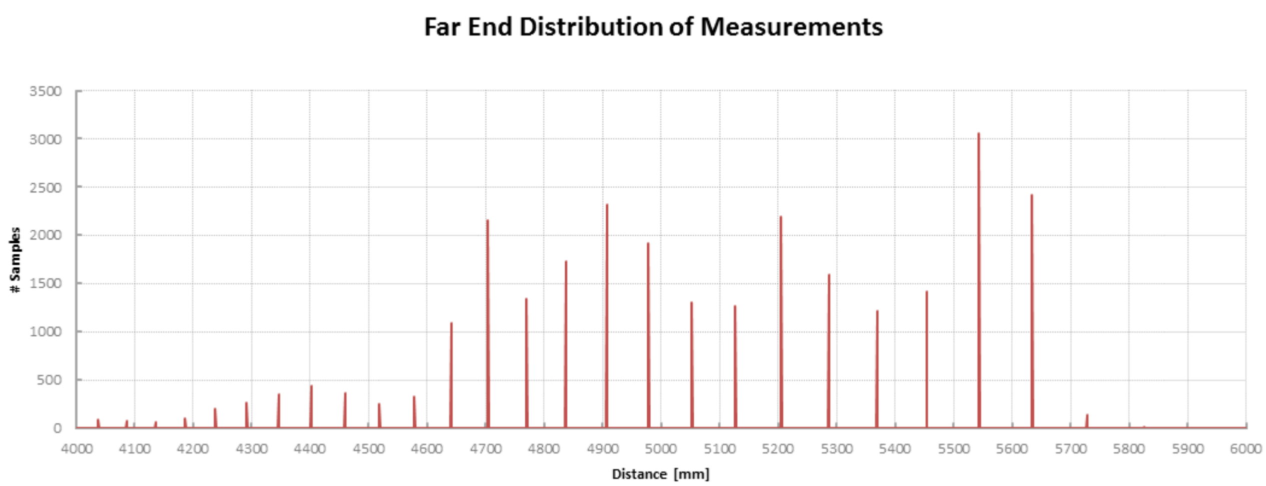
(a) Depth map



(b) Infrared coded structured light

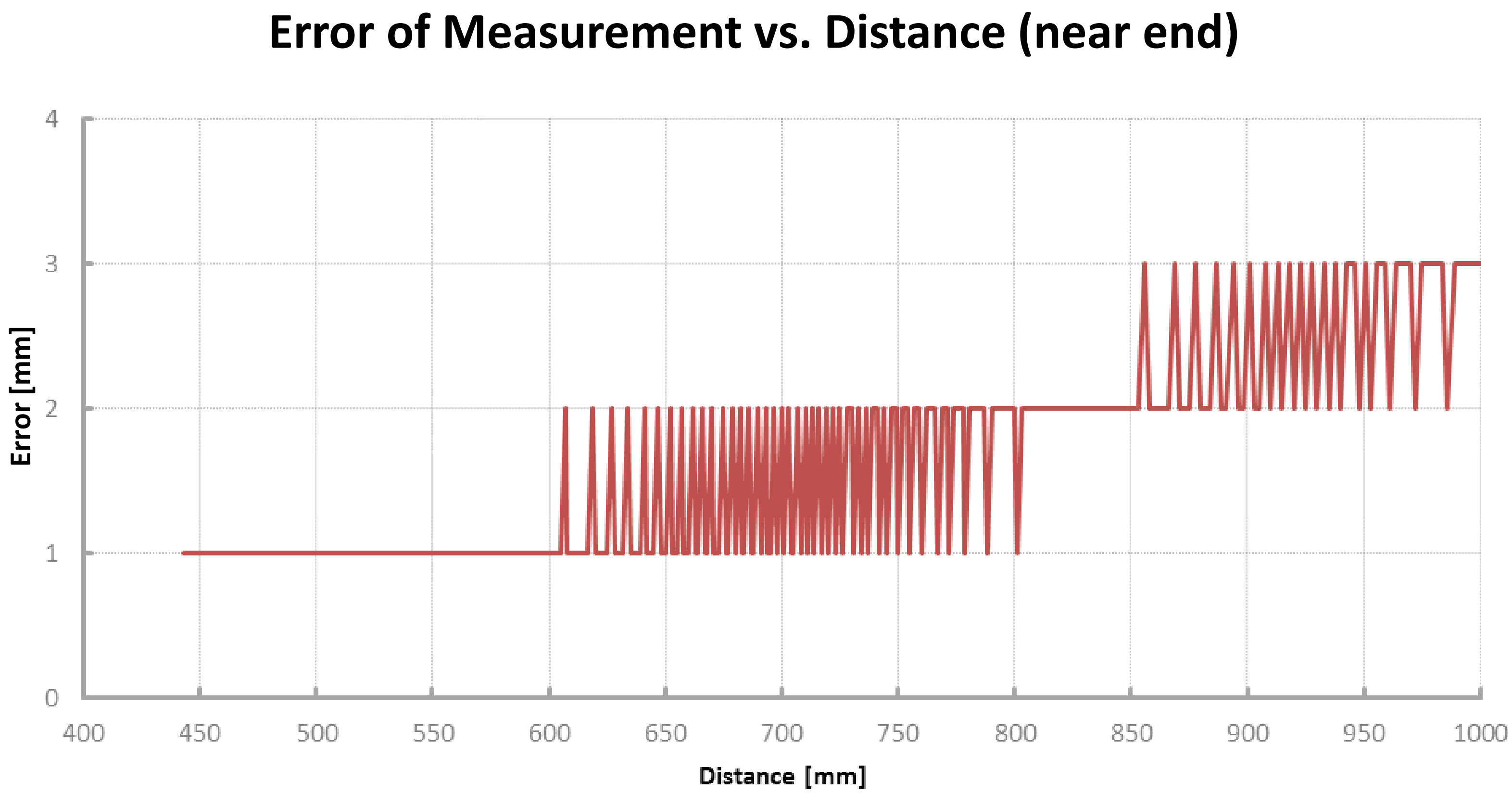
Range Data Sampling Rate





Range Data Sampling Rate

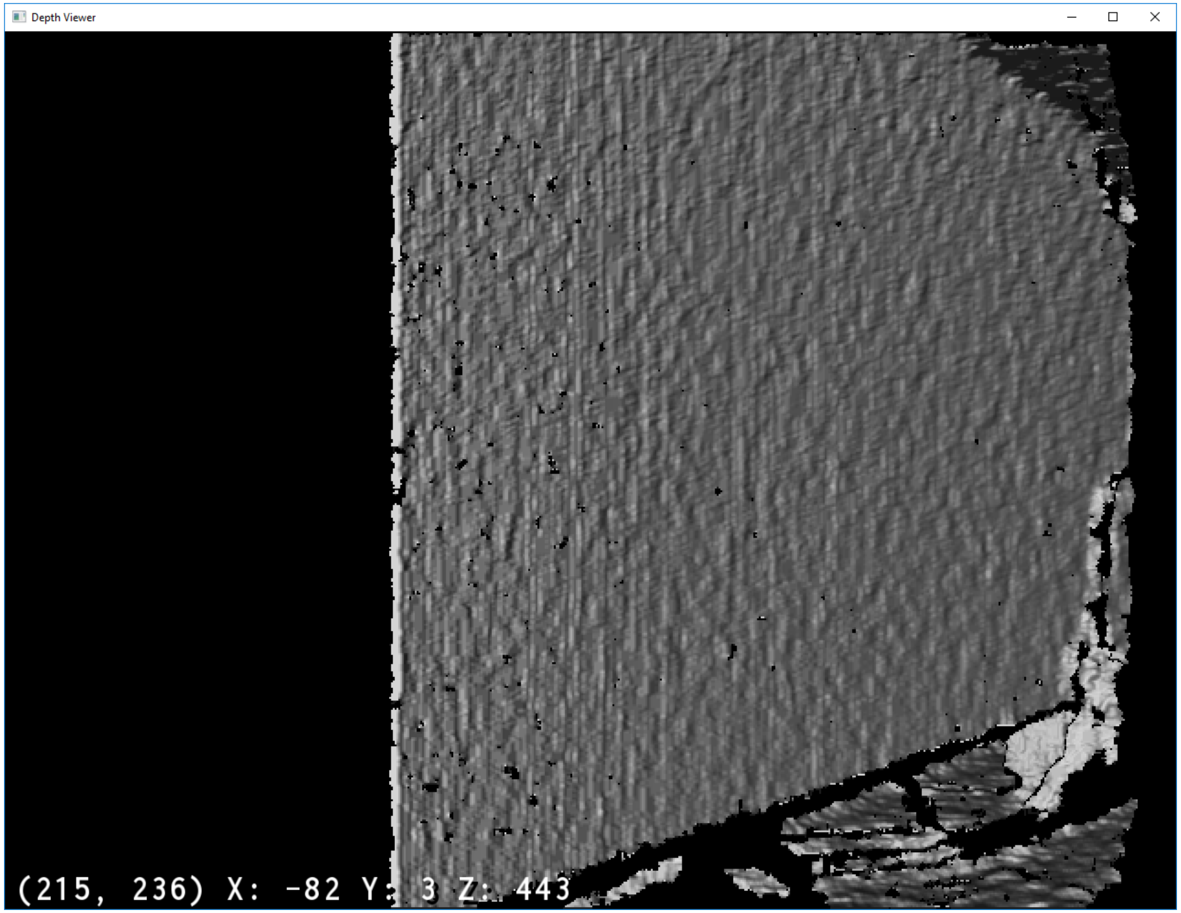
Range Data Sampling Rate



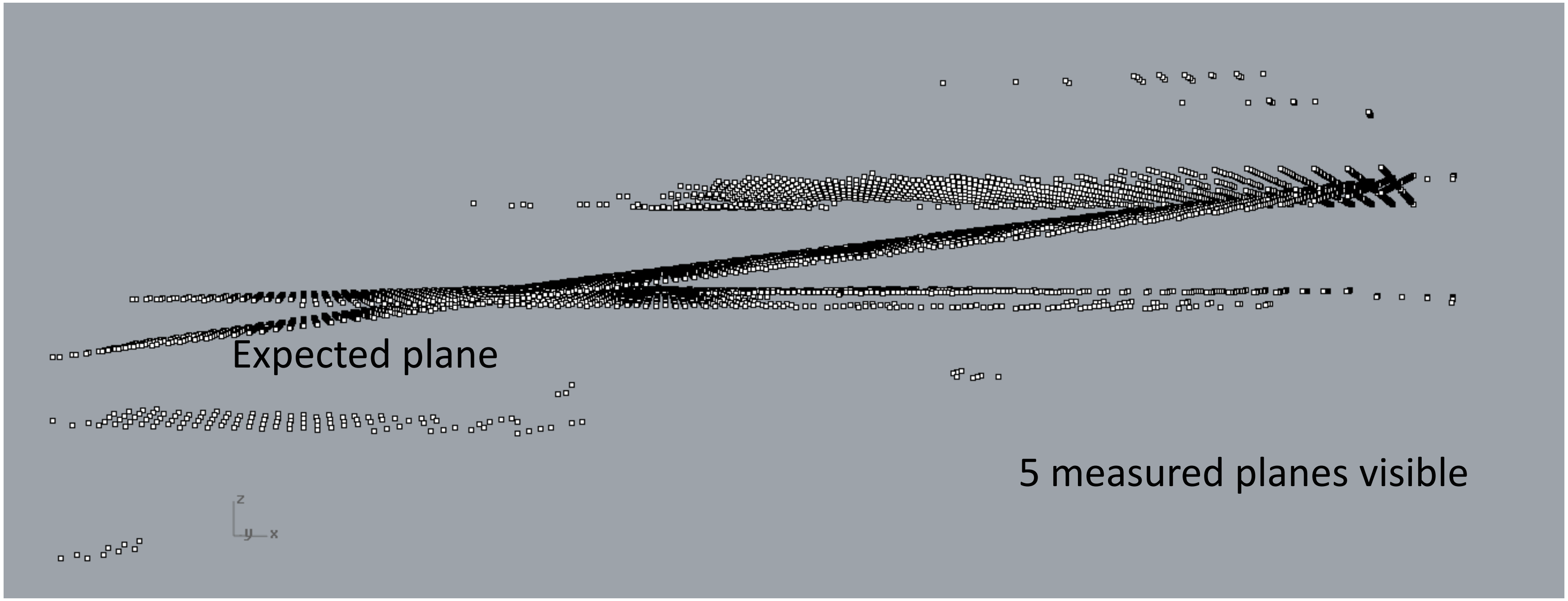
Range Data Sampling Rate

Plane Sampling

Perfectly flat plane at the distance of 44.3 - 100 cm

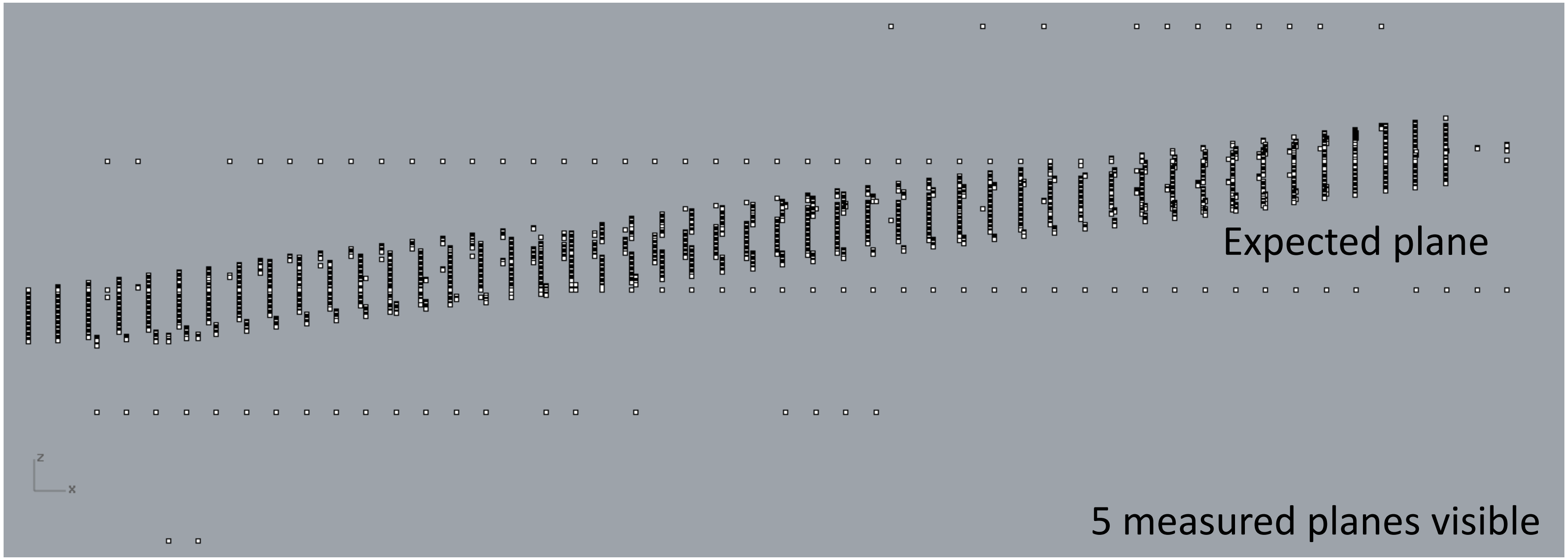


Plane Sampling cont.



Mean error 6.84 mm
Separation of measurements ~ 19 mm at the distance of 2.5 m

Plane Sampling cont.



Mean error 6.84 mm
Separation of measurements ~ 19 mm at the distance of 2.5 m

Sensors Capabilities

- Observations
  - At close distances (0.6 m) Astra produces accurate depth map

    Resolution close to 1 mm is achievable as the sampling rate allows for that level of precision
  - With increasing distance things are getting worse

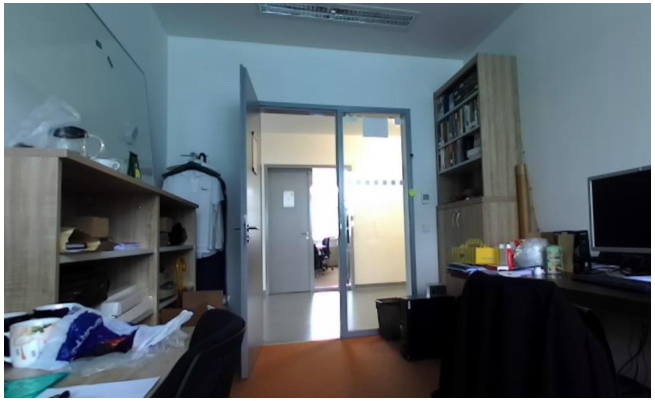
    At the far end (8 m) we can expect the resolution of roughly 20 cm

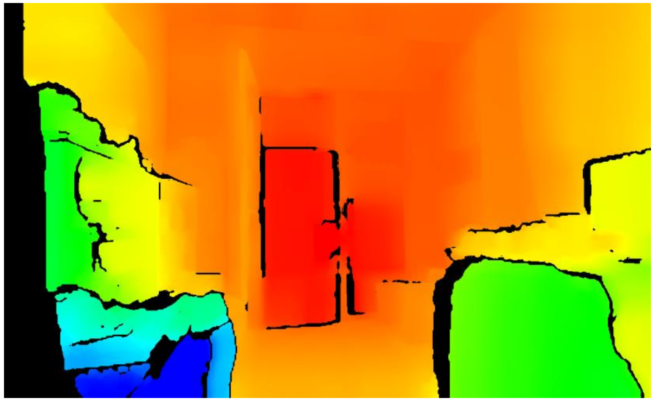
Sensors Capabilities

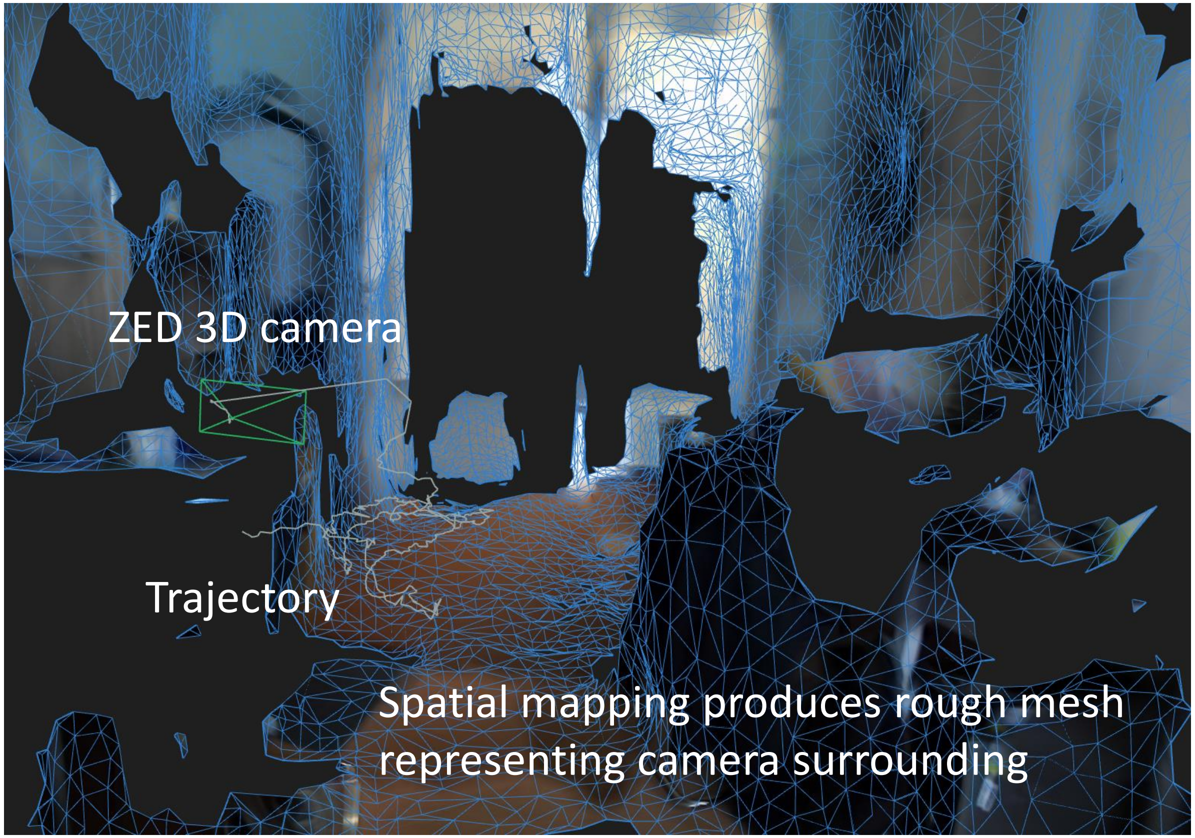
- Conclusion
  - We decided to pursuit these two paths independently
  - Approach A: RGB images with optional range data
  - Approach B: Depth images with possible use of color channels

Main Challenges

- Where is the camera?
  - Spatial mapping, 3D SLAM (e.g. HoloLens, ZED)



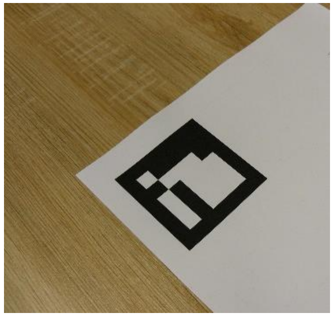




mrl

Main Challenges

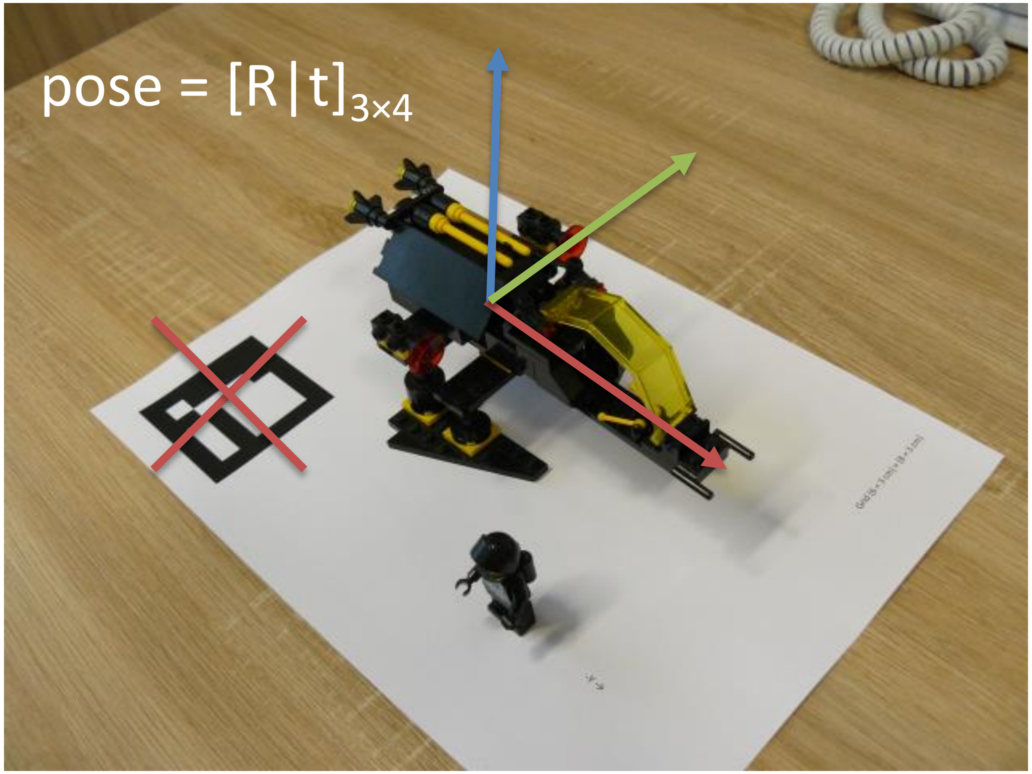
- Where is the sought object?
  - SDKs commonly recognize planar (fiducial) markers (e.g. VuMark, QR code)

  - Support for direct recognition and tracking of texture-less 3D objects is lacking
  - Algorithms for 6-DOF object pose estimation in RGB(-D) images are actual research topic

Main Challenges

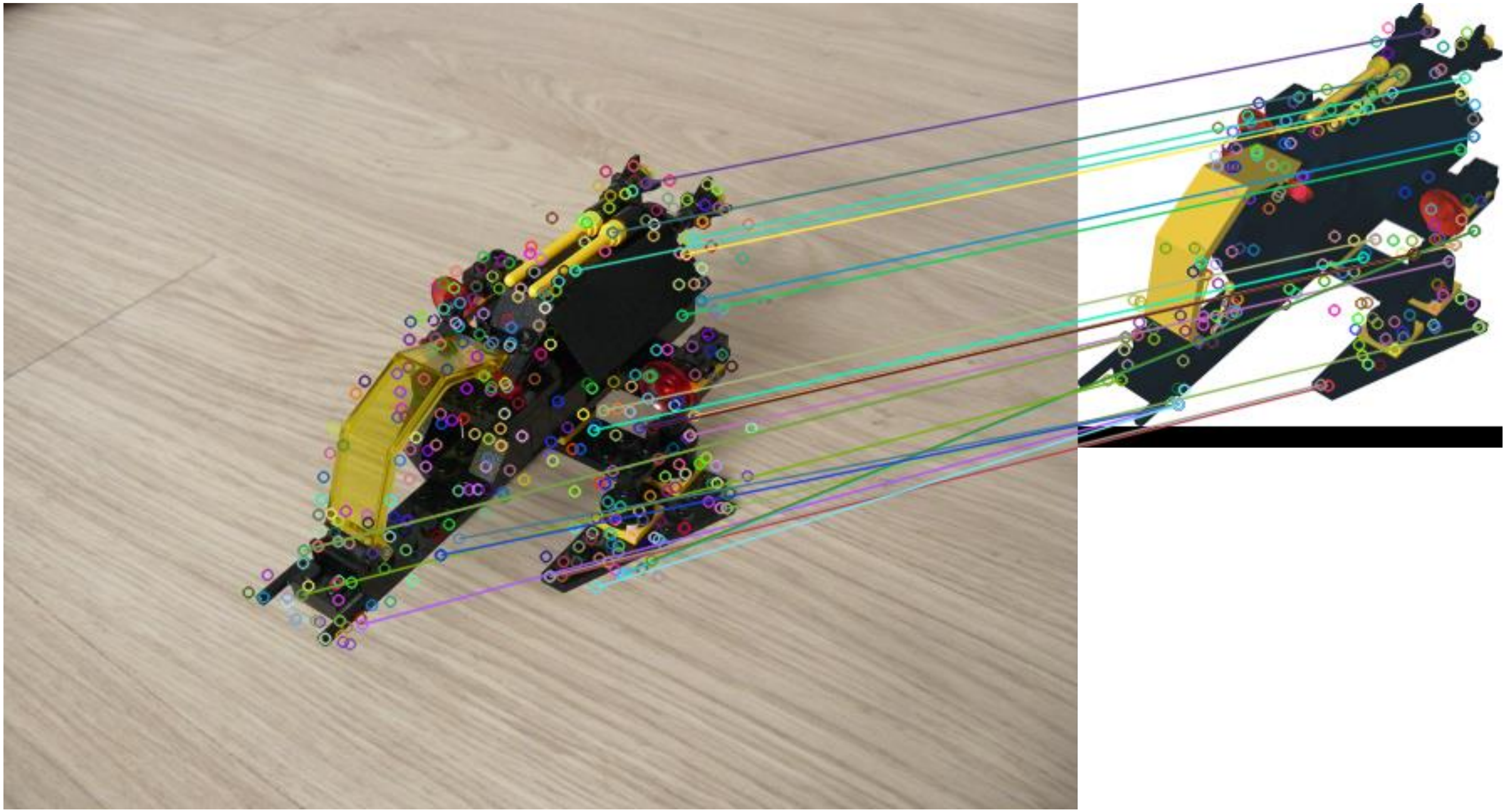
- How to estimate the relative pose of the object w.r.t. the camera without markers?



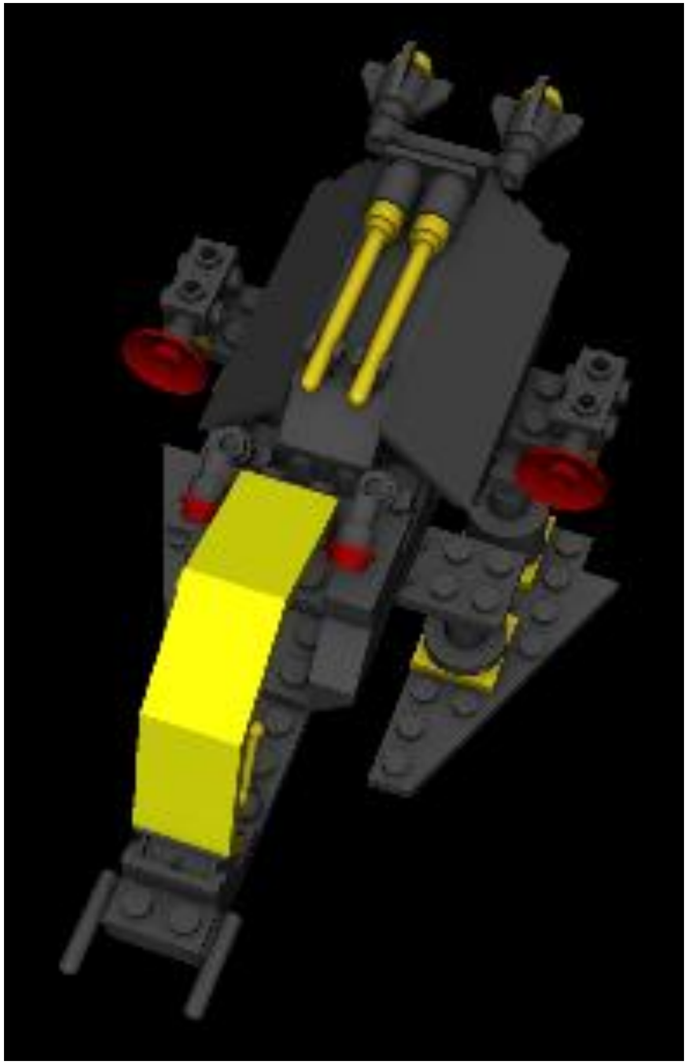
6-DOF Pose Estimation in RGB Images

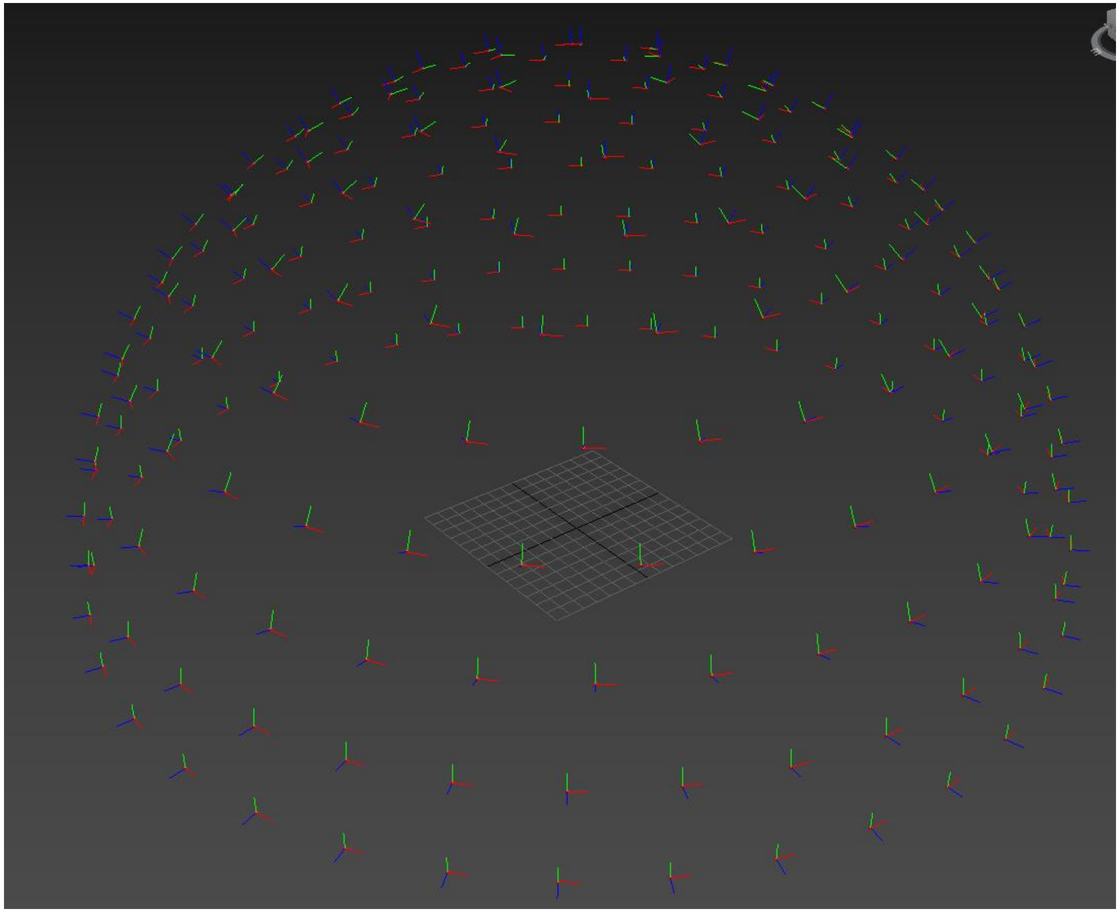
- Object instance detection and pose estimation methods
  - Template based matching
    - Edge based
    - Gradient based
  - Local descriptors (scalable, robust to object pose changes)
    - SIFT, SURF, ORB, BRIEF…
  - Voting based approaches
    - Excessive quantization and post-processing required
  - CNN based approaches
    - Rather bit slow

Example of Pose Estimation

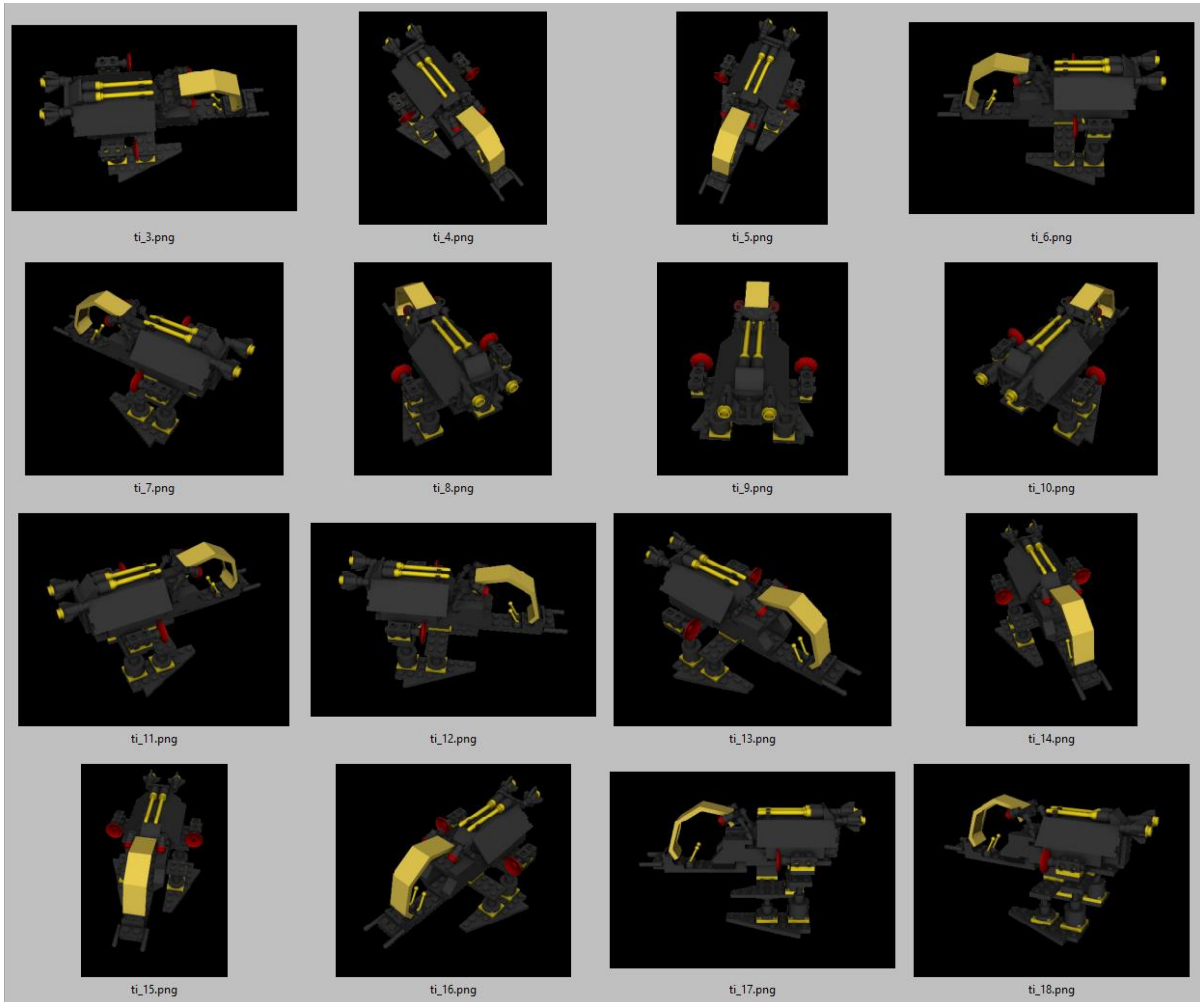


Example of Pose Estimation

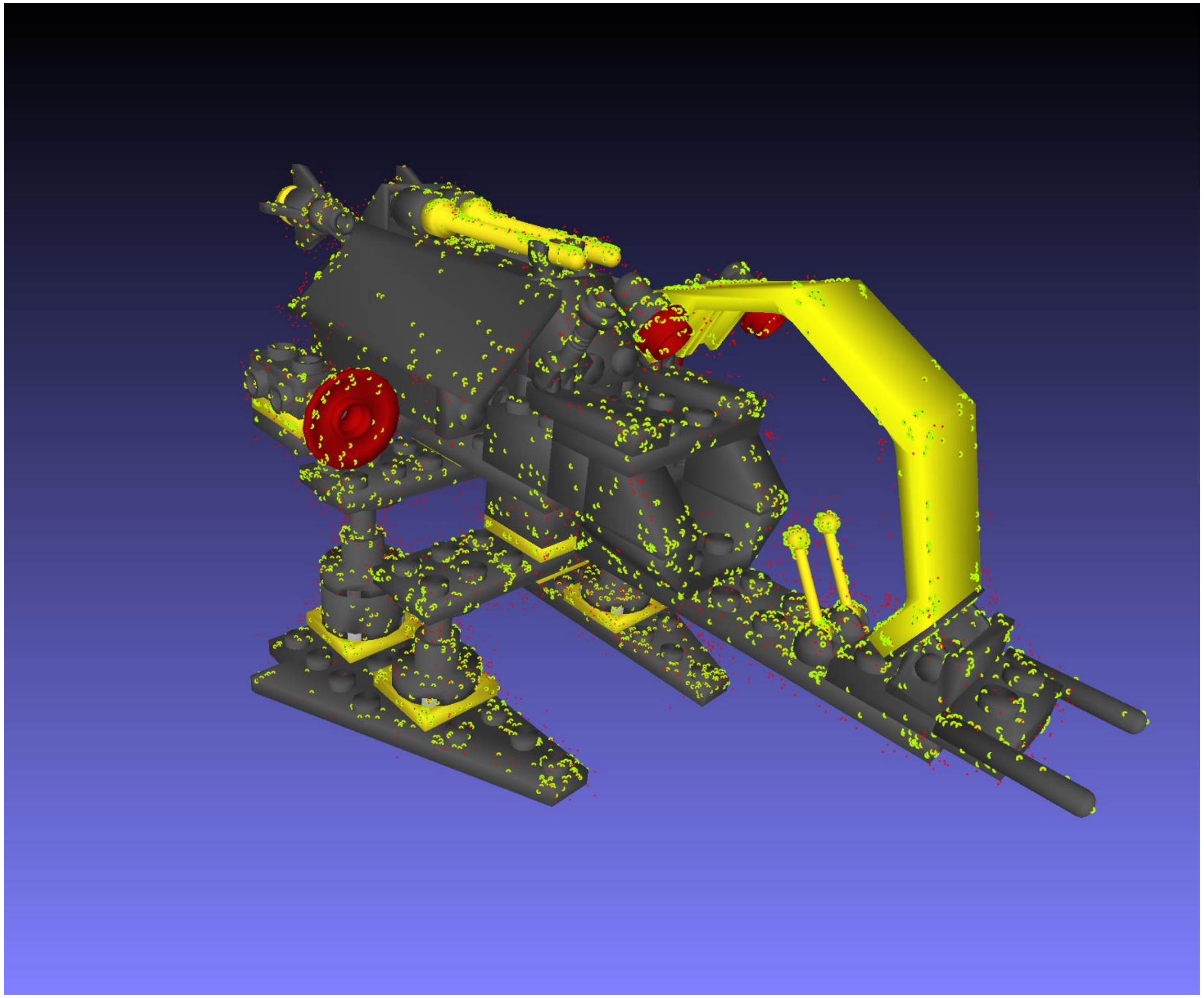




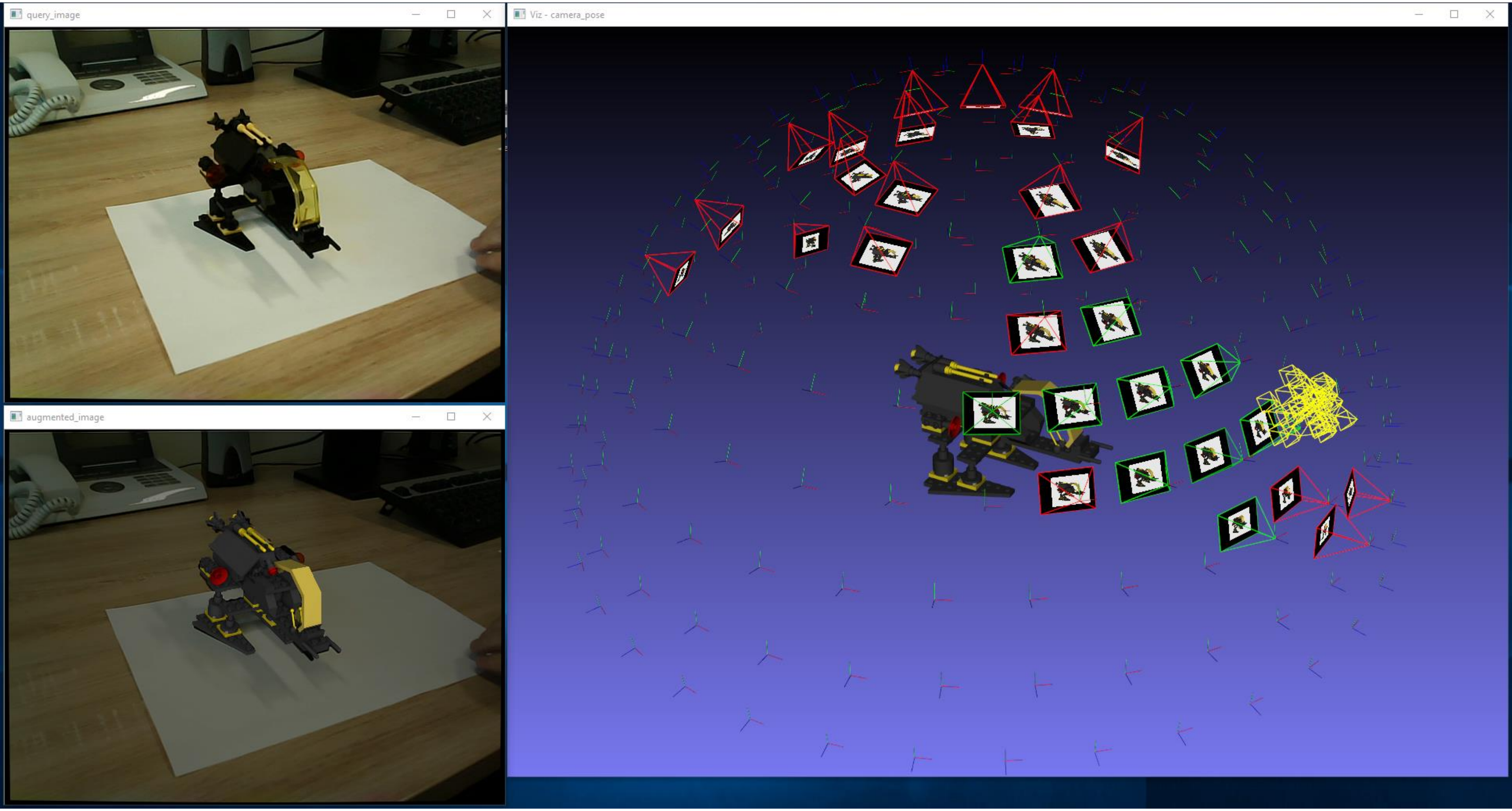
Example of Pose Estimation



Example of Pose Estimation



Example of Pose Estimation



Example of Pose Estimation

Thank you for your attention