Augmented Reality :
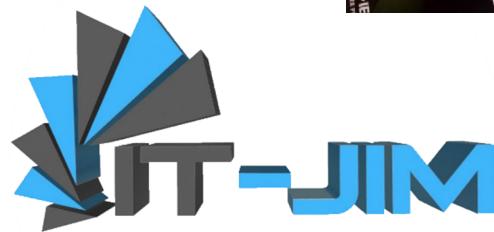# *Modern Trends and Technical Solutions*

Ievgen Gorovyi, PhD,
founder & CEO @ It-Jim,
signal and image processing

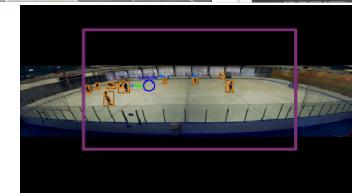# It-Jim: Key Interests

CEO & founder, PhD, signal and image processing
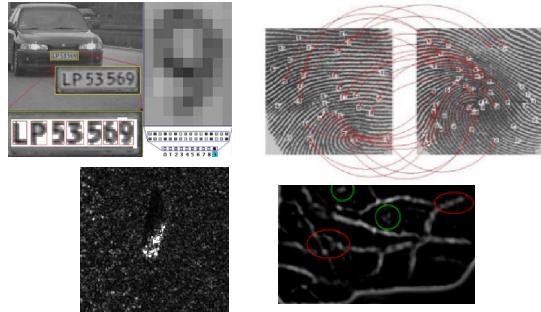author of >40 publications

Indoor Navigation

Augmented reality

Pattern recognition

Object tracking
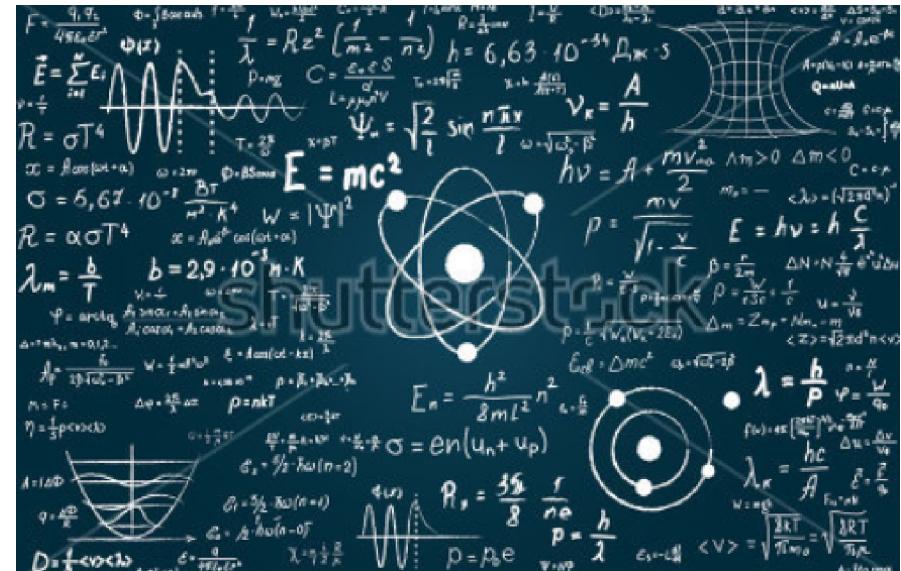
OUTLINE

- **Introduction**
  - ❑ AR history
  - ❑ Examples, applications, market
  - ❑ Software and hardware

- **AR: What's inside?**
  - ❑ AR and camera pose
  - ❑ AR markers
  - ❑ Detection
  - ❑ Tracking
  - ❑ Image retrieval
  - ❑ SLAM for AR

- **Mobile AR in Production**
  - ❑ AR user experience
  - ❑ SDK infrastructure
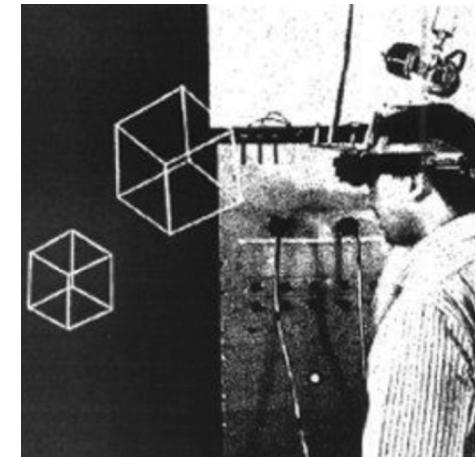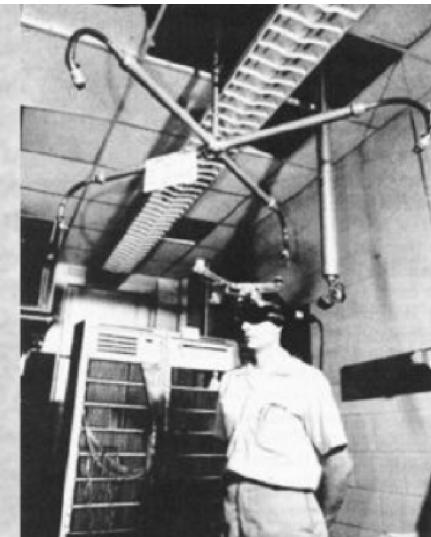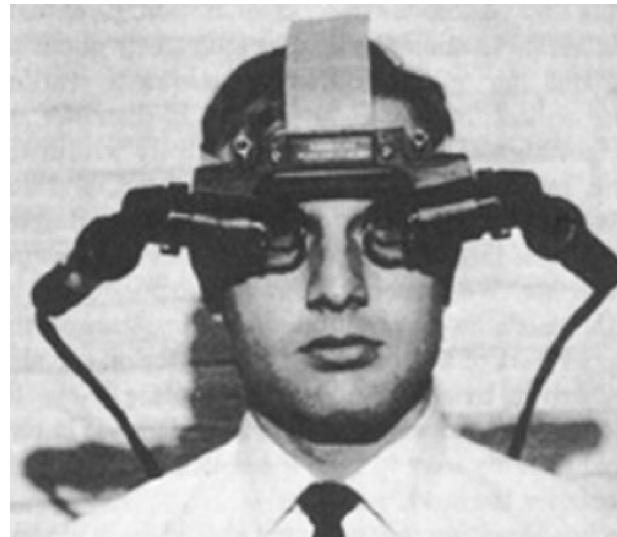 - ❑ Real examples

- **Conclusions**

Augmented Reality: Intro
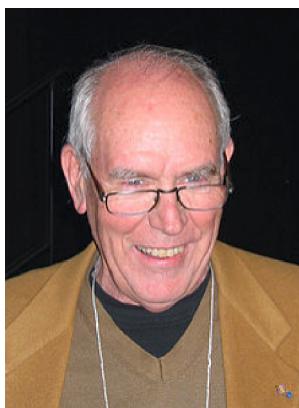
Augmented Reality: What's It?

AR History: First Appearance

1968 *1st head-mounted display – Sword of Damocles (Ivan Sutherland)*

Sutherland - a father of computer graphics

Sketchpad* (first light-pen)
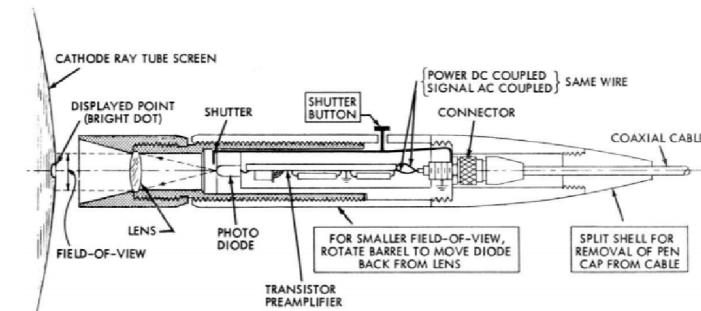
Sketchpad:A MAN-MACHINE GRAPHICAL COMMUNICATION SYSTEM, I. Sutherland, PhD thesis, MIT, 1963.

AR History: First Steps

| Real Environemnt |
| Augmented Reality (AR) |
| Augmented Virtuality (AV) |
| Virtual Environment |

1968
Sutherland

1978
VideoPlace
Myron Krueger

1980
EyeTap
(Steve Mann)

1990
Tom Caudell
(Boeing)

1994
Reality continuum
(Milgram)

1999
AR in NASA
ARToolKit

AR History: Modern Times

2000
AR Quake

2008
1st commercial AR app BMW mini

2010
Magic Leap

2015
MS Hololens

2016
Pokemon Go

2017
ARKit, ARCore

2018
ARKit V2.0
Magic Leap One

AR: Introduction

7
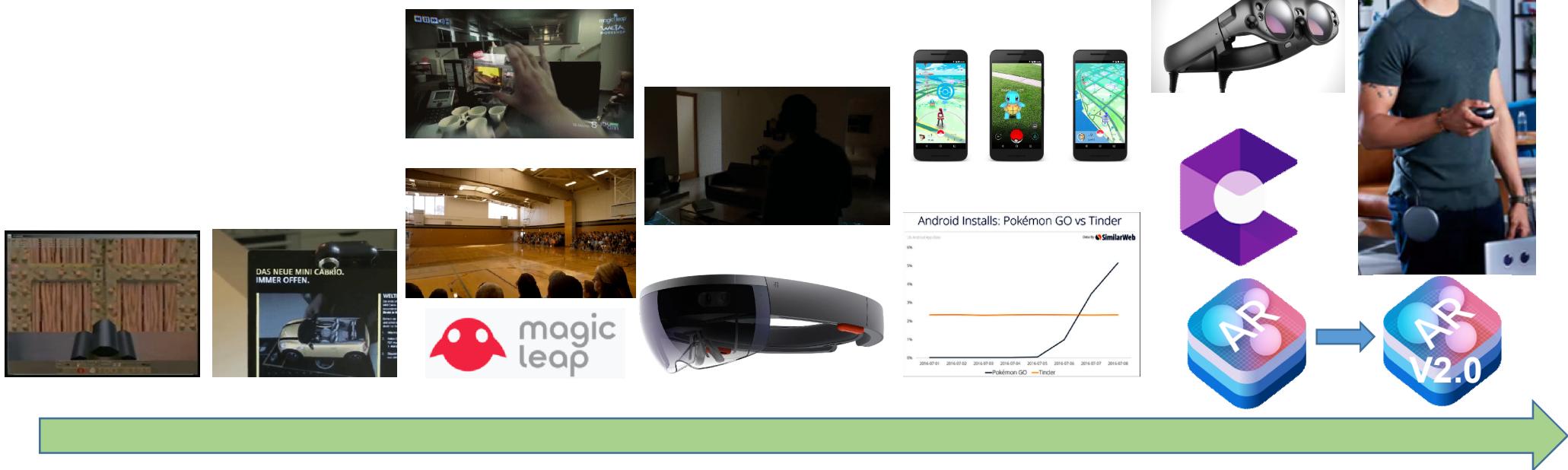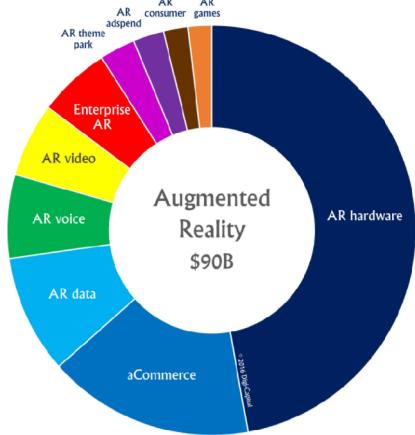
Mobile AR Market

Market share by 2020 (Digi-Capital)

AR/VR Platform Revenue

■ Mobile AR ■ Smartglasses ■ Premium/standalone VR ■ Mobile/standalone VR
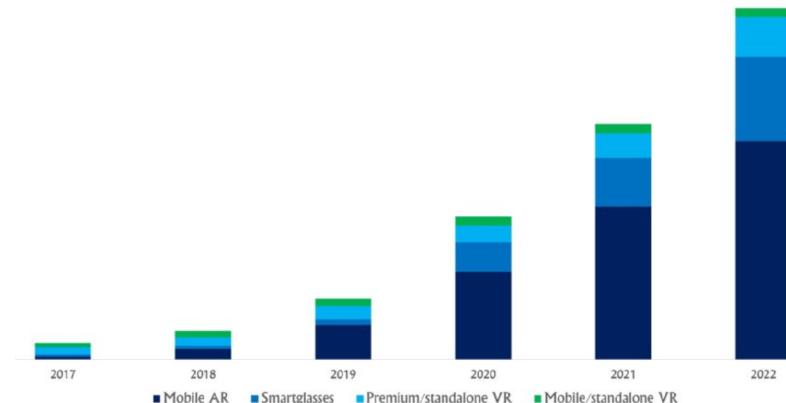
AR/VR country revenue vs growth rate (2017 to 2022)
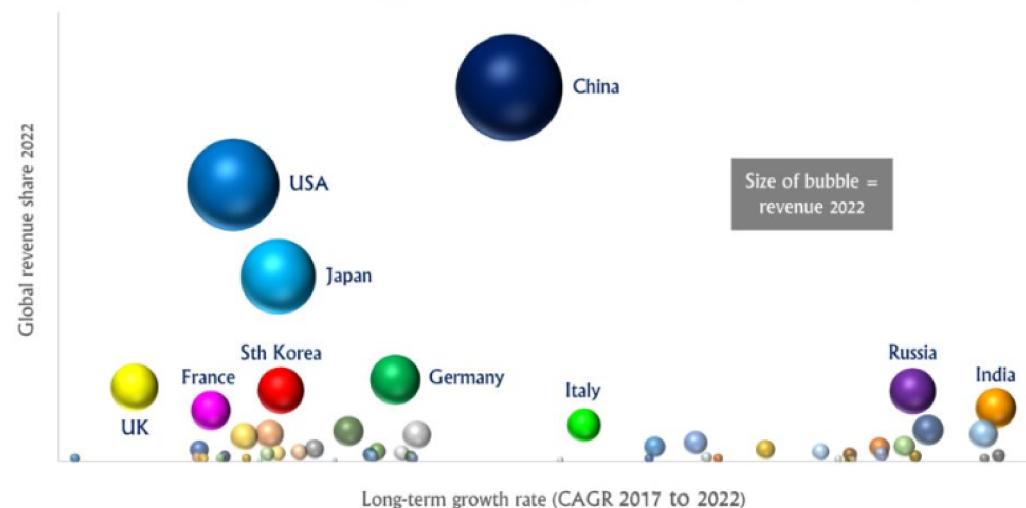
Mobile AR: Software Landscape

	country	platforms	features
vuforia	UK	Android iOS Windows	Complex object recognition, smart glasses support, 0.5M community!
kudan AUGMENTED REALITY	UK, Japan	Android iOS	Versatile infrastructure for mobile, drones and automotive
wikitude	Austria	Android iOS	Unity, smart glasses support
MAXST	South Korea	Android Apple Windows	Unity, Various trackers
(AR)	USA	iOS	Unity, VIO, Multiplayer
	USA	Android iOS	Motion tracking, environmental understanding, Light estimation

AR Use Cases

AR in sports (simplest)

AR in retail

Location-based AR

AR in real estate

Furniture AR

AR in Education

Basic education
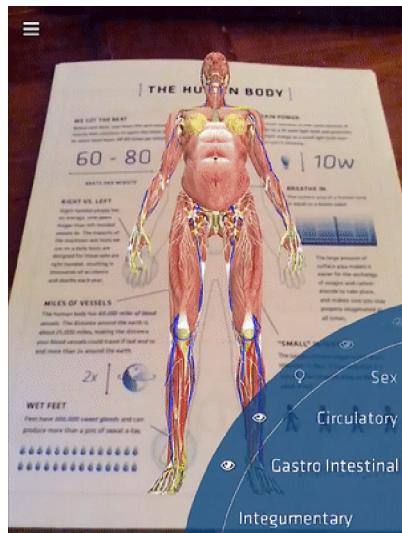
DAQRI anatomy

DAQRI elements 4D
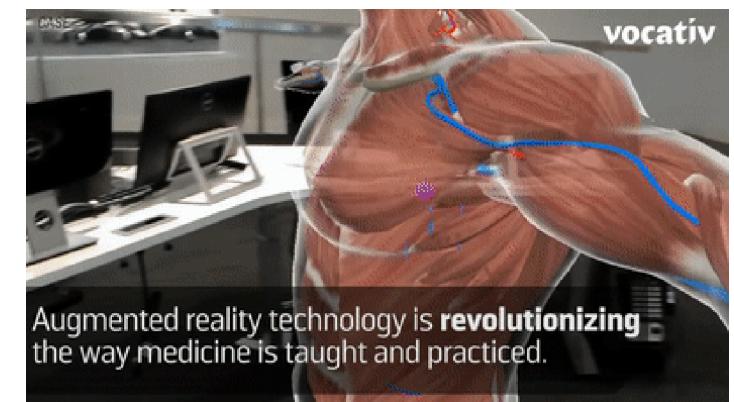
Medical education with Hololens

AR in Entertainment

AR highway (ARCore)

AR astronomy (ARKit)
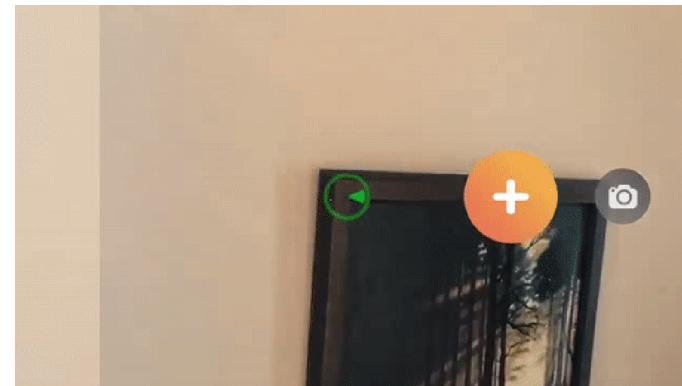
AR measure (ARKit)

AR portals (ARKit)

12

AR in Medicine
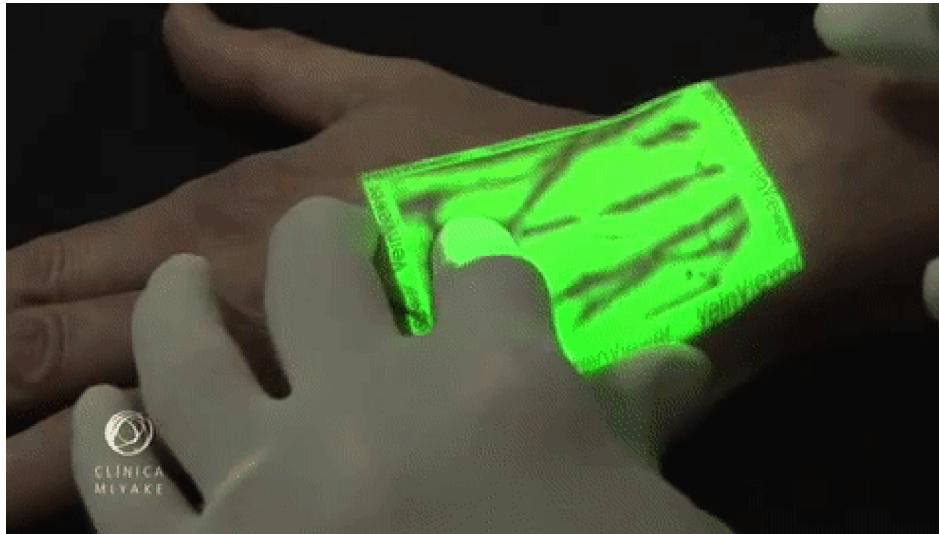
Vein viewer (NIR+Optical)
Clinica Miyake

than wearing some type of shield, which is often what surgeons do.

AR surgery (X-ray + optical)
Augmedics

FRONT-END

BACK-END

AR: Algorithms and Solutions

- **camera pose and marker detection**
- **image retrieval**
- **tracking**
- **markerless and SLAM**

AR Magic: What Do We Need?

6 DOF

Wrong augmentation

Correct augmentation
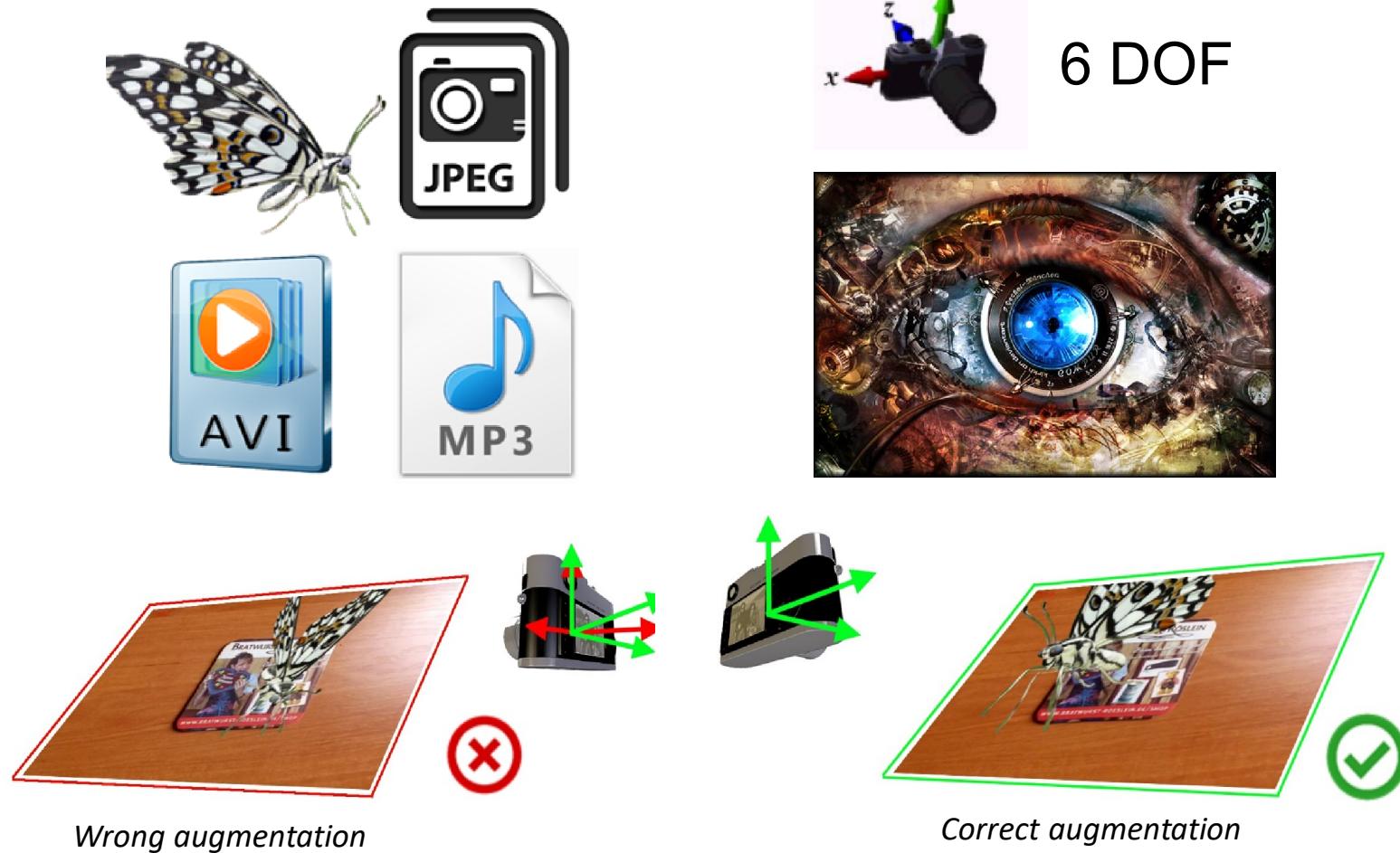
Marker-based vs Markerless

IT-JIM
signal and image processing

2 Markerless / Marker-based

$\triangle R|t$

6 DOF
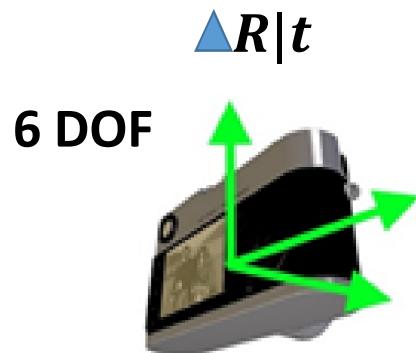
+ Easy to detect / No need for markers
Textureless scene is not a problem

− Occlusions / High complexity
Works for limited distances / Risk of textureless

15

Marker-based AR and Camera Pose

Rotation (α, β, γ)

Translation (x, y, z)

World coordinates

Two camera views
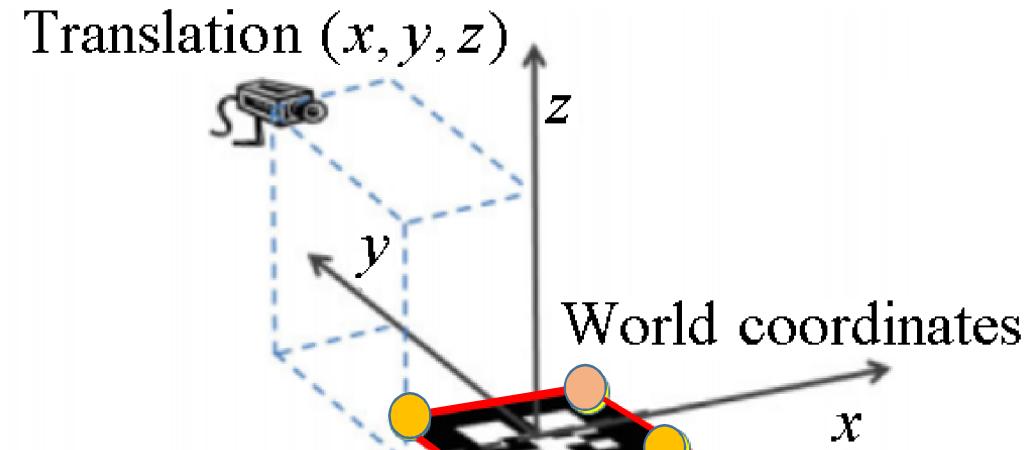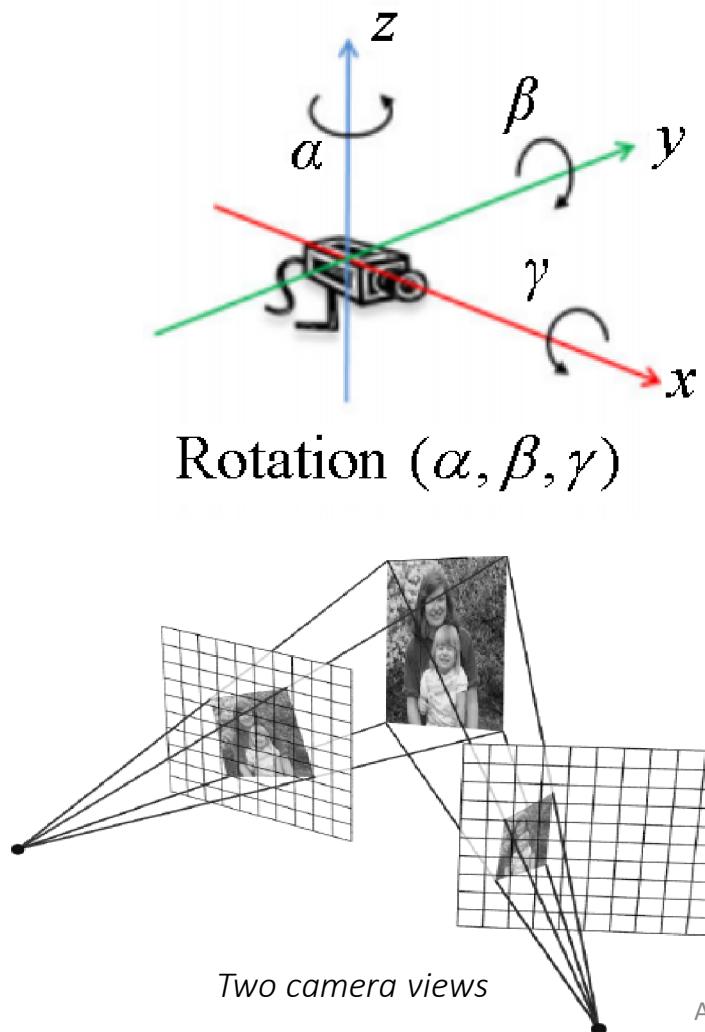
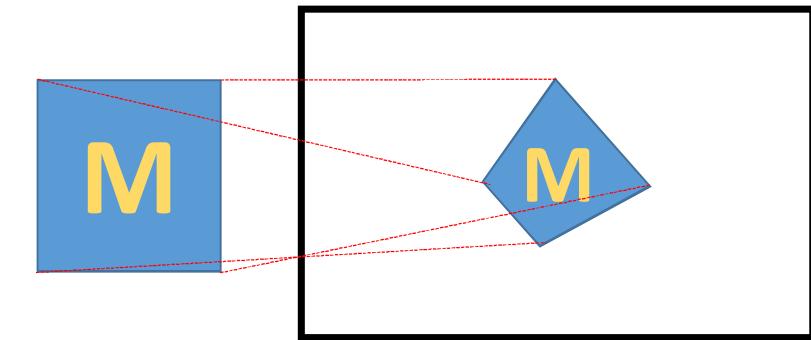$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \cong \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$
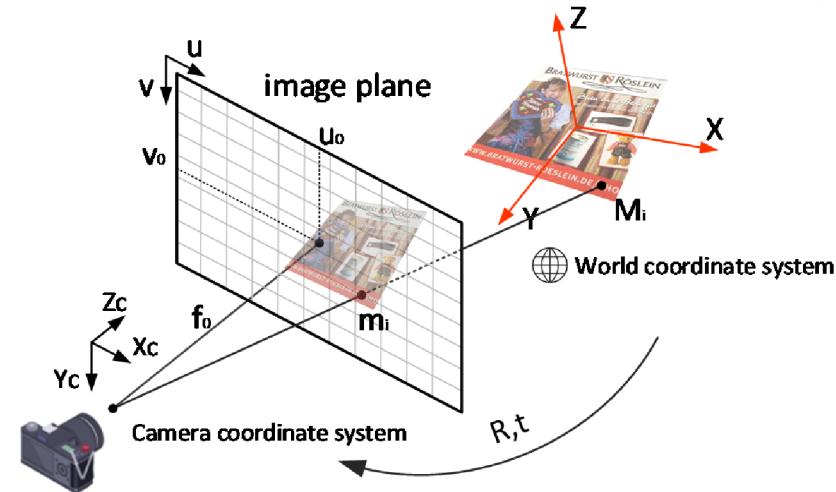
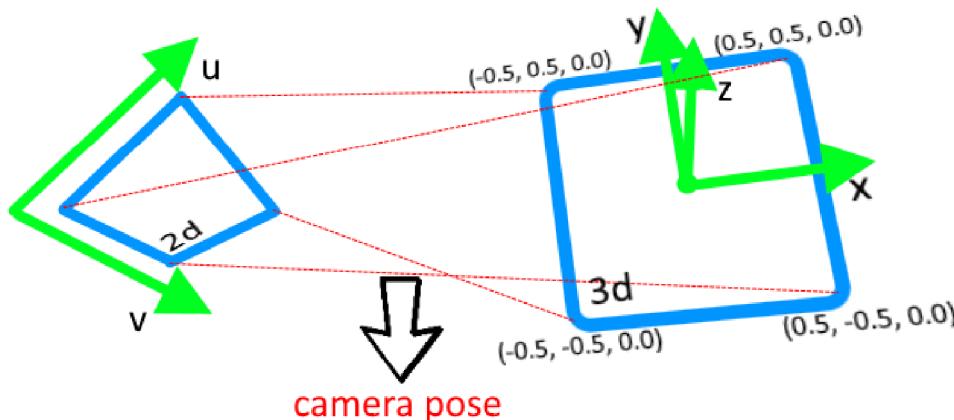*Homography*

AR: Algorithms and solutions

Camera Pose Estimation Steps

Homography estimation

Projective geometry

Camera pose extraction (2D-3D)

$$\begin{pmatrix} \\ \\ \end{pmatrix} = \begin{bmatrix} f_x & 0 & u_x \\ 0 & f_y & v_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{32} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} [\ ]$$

K $R|t$

AR: Algorithms and solutions

17

Fiducial Markers Diversity
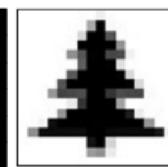
Template-based markers
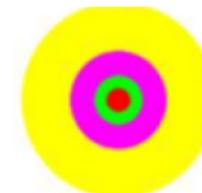

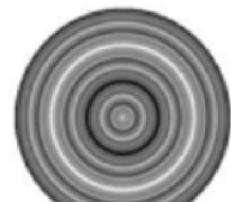ARToolKit ARToolKit Plus ARTag Intersense ARUCO Multiring color RuneTag FourierTag
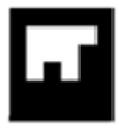
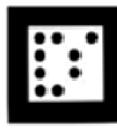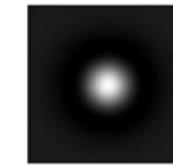Matrix BinARyID Cybercode Visualcode ARSTudio Spotcode CCTag CALTag
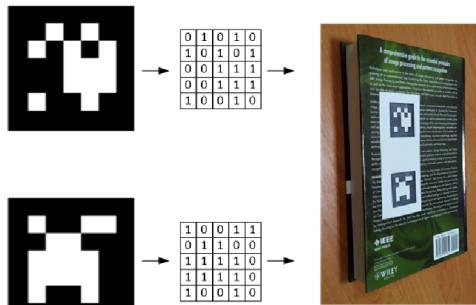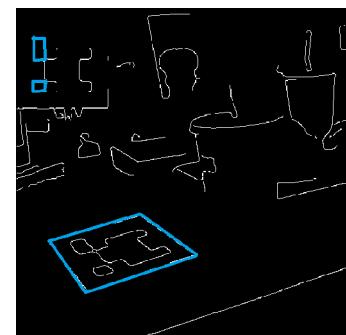
IGD SCV HOM ReacTIVision QRcode SIFTTag AprilTag
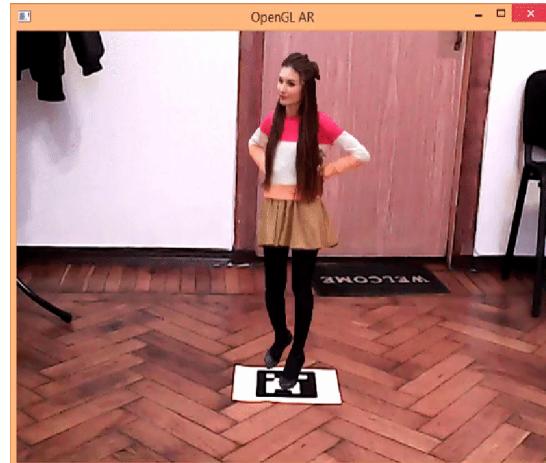
AR: Algorithms and solutions

18

ArUCO Markers

Binary marker detection

Binary marker recognition

Fiducial marker tracking
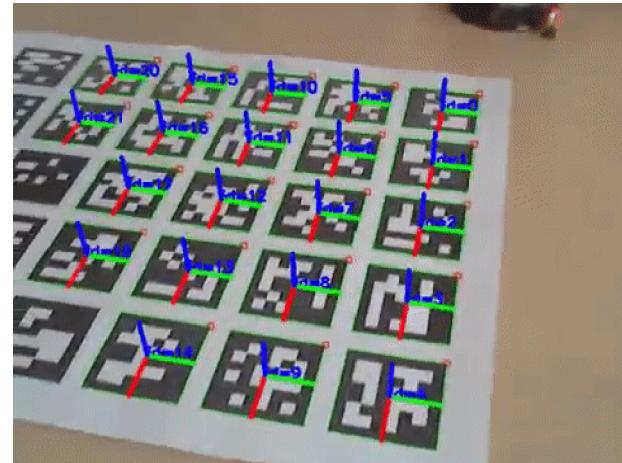
Multiple marker tracking

Fiducial Markers: Interesting Examples
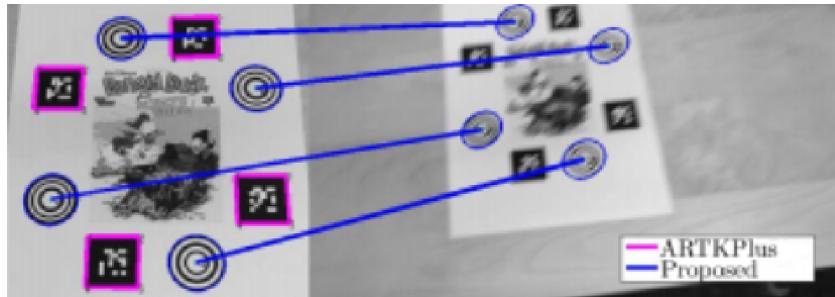
Resistance to motion blur
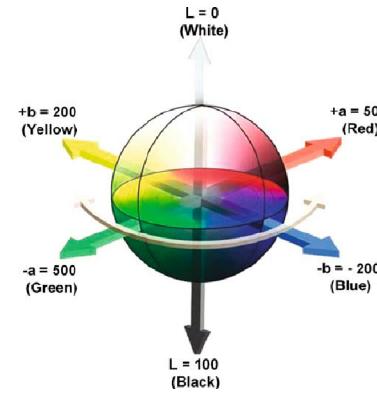
LAB color space

HSV color space

Initial scan   *Initial polygon*

Building polygon   *Convergence*

Refine corners   *Detection*

CCTag — 263 ms
RuneTag — 51 ms
AprilTag — 19 ms
ChromaTag — 1.4 ms

AR: Algorithms and solutions

L   a   b

R   G   B

H   S   V

20
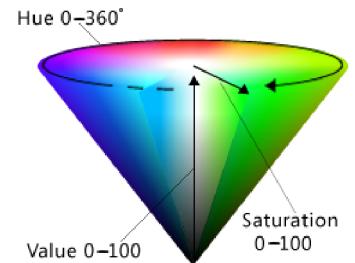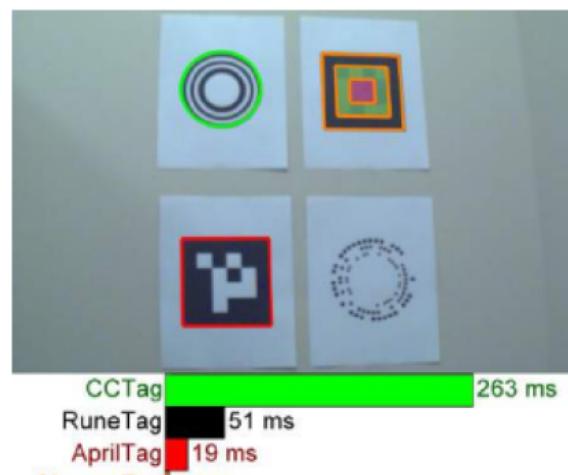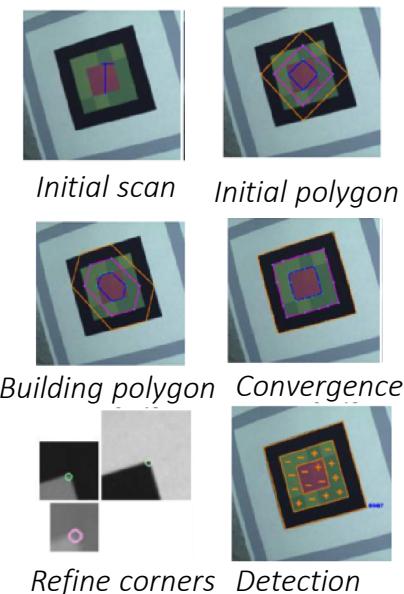
ChromaTag: A Colored Marker and Fast Detection Algorithm. Joseph DeGol,Timothy Bretl,Derek Hoiem, ICCV-2017, pp.1472-1481.

Image Markers
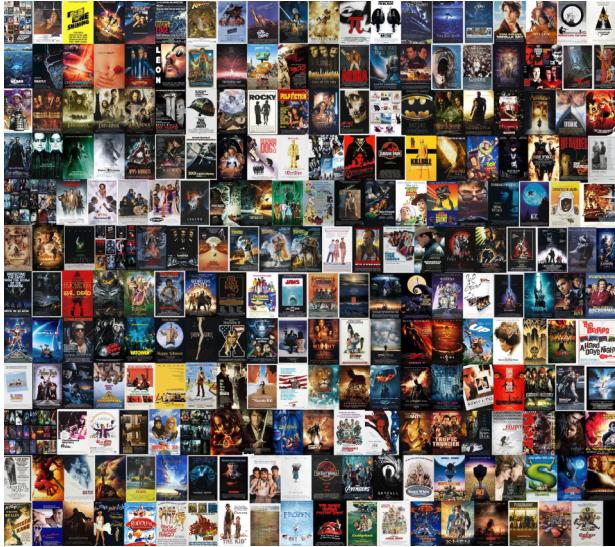
Film billboard
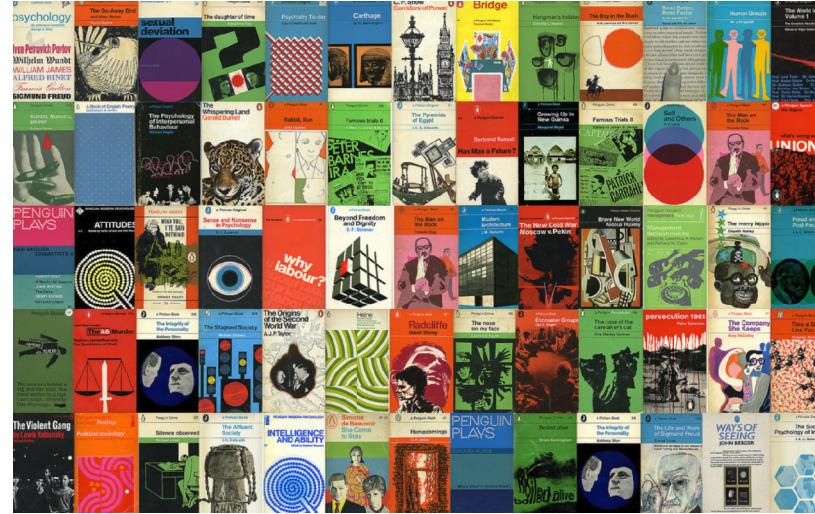
Book covers

Newspapers

Menu, food, etc.

Image Description: First Approaches

Image transformations

SIFT (1999)

scale space with box filters

SURF descriptor (2006)
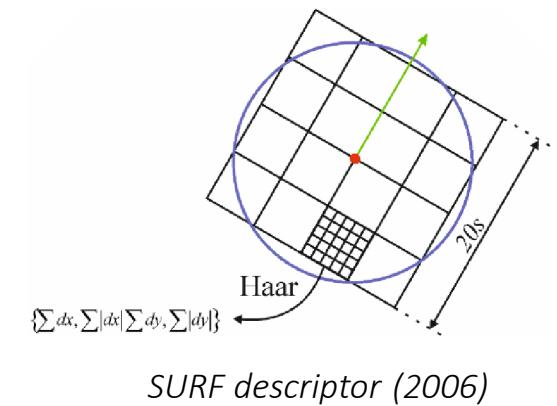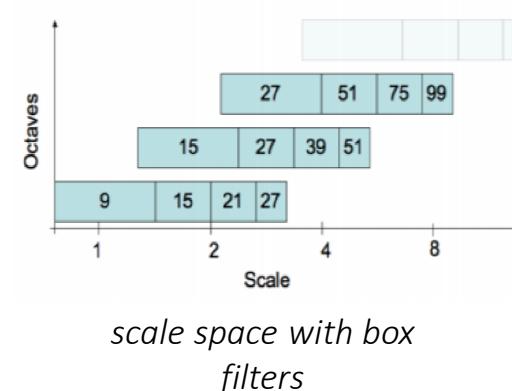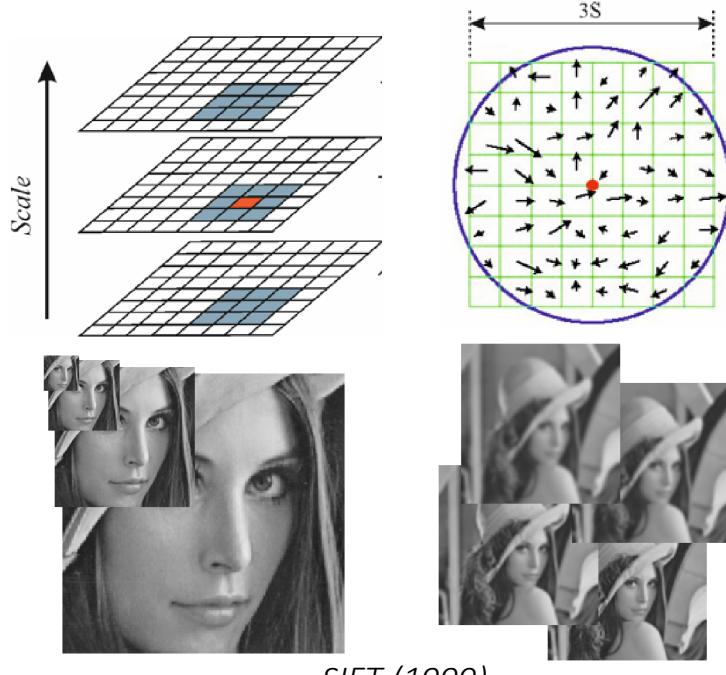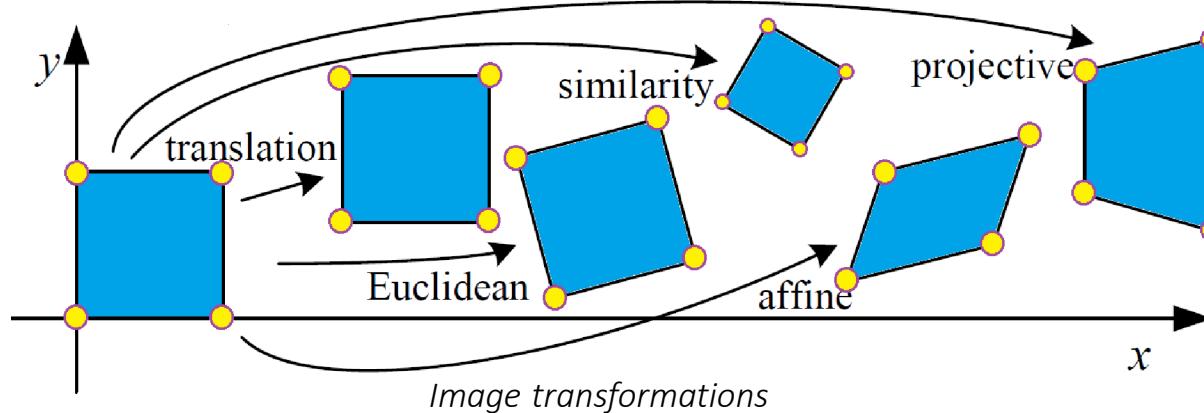
Local Feature Descriptors

Keypoint detection → Keypoint description → Matching
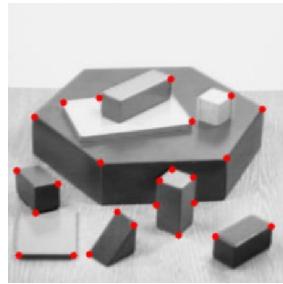
Corner detection
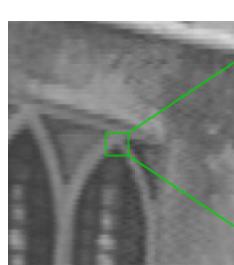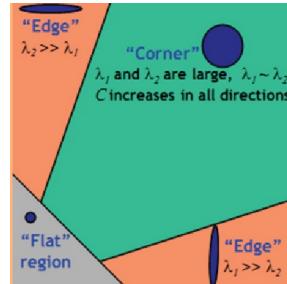
FAST

BRIEF

ORB

BRISK

FREAK

AR: Algorithms and solutions

23
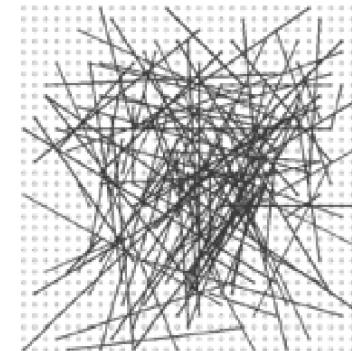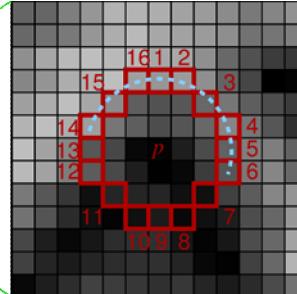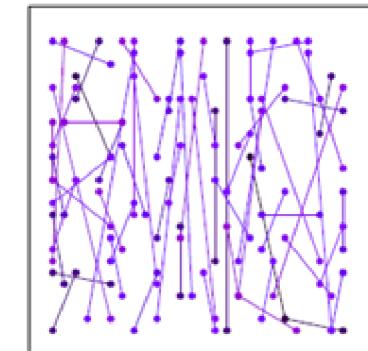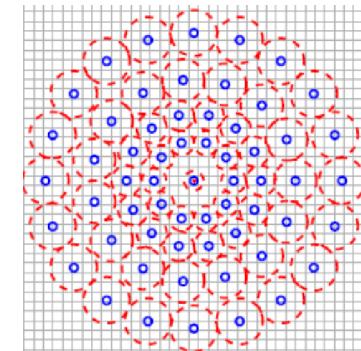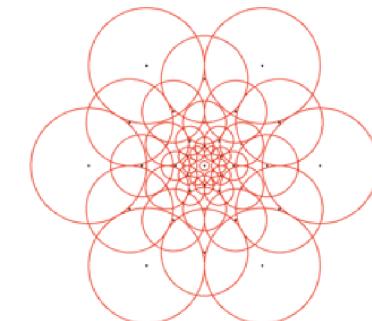
KAZE and A-KAZE

linear diffusion (SIFT) - slow

non-linear diffusion (KAZE)

$t_i = 5.12$ $t_i = 20.48$ $t_i = 81.92$ $t_i = 130.04$ $t_i = 206.42$

Scale-space comparison

M-SURF (KAZE)

A-KAZE (2013) vs ORB (2011)

M-LDB (A-KAZE)

24

MagicPoints (Magic Leap)

Synthetic data for training

Magic point detection

D. DeTonem T. Malisiewicz, A.Rabinovich - "Toward Geometric Deep SLAM" / https://arxiv.org/pdf/1707.07410.pdf
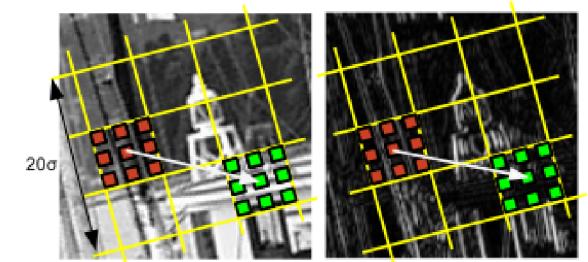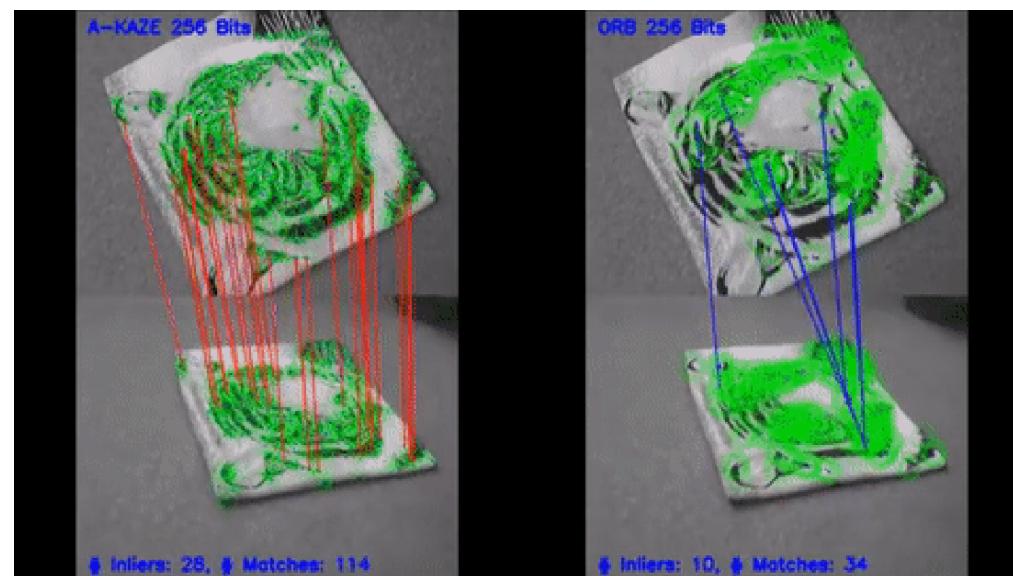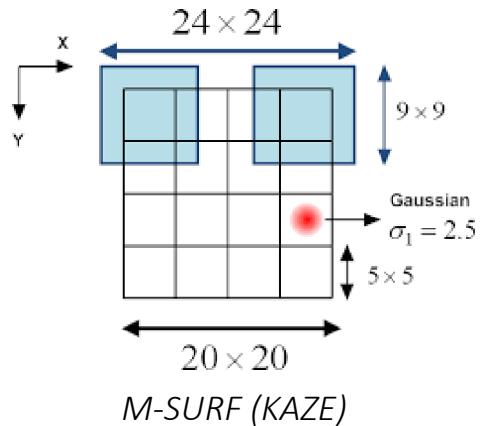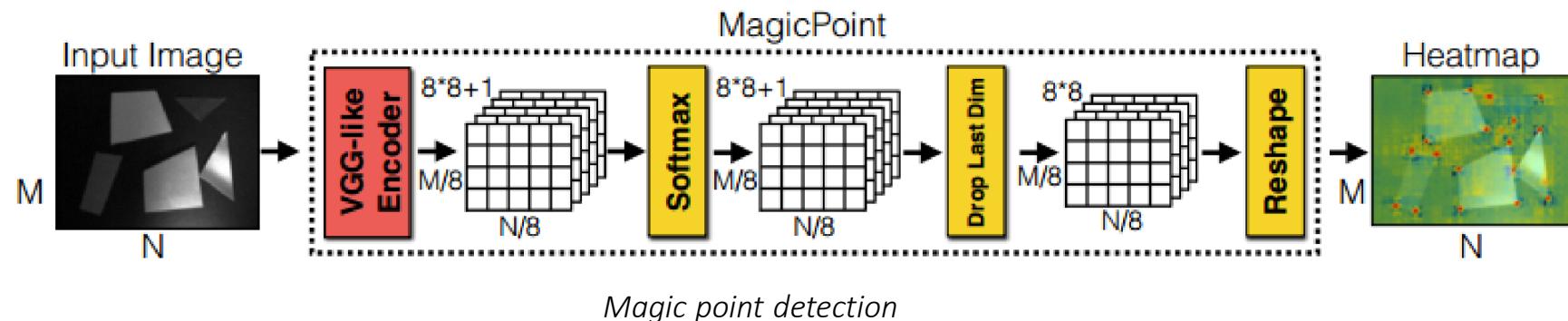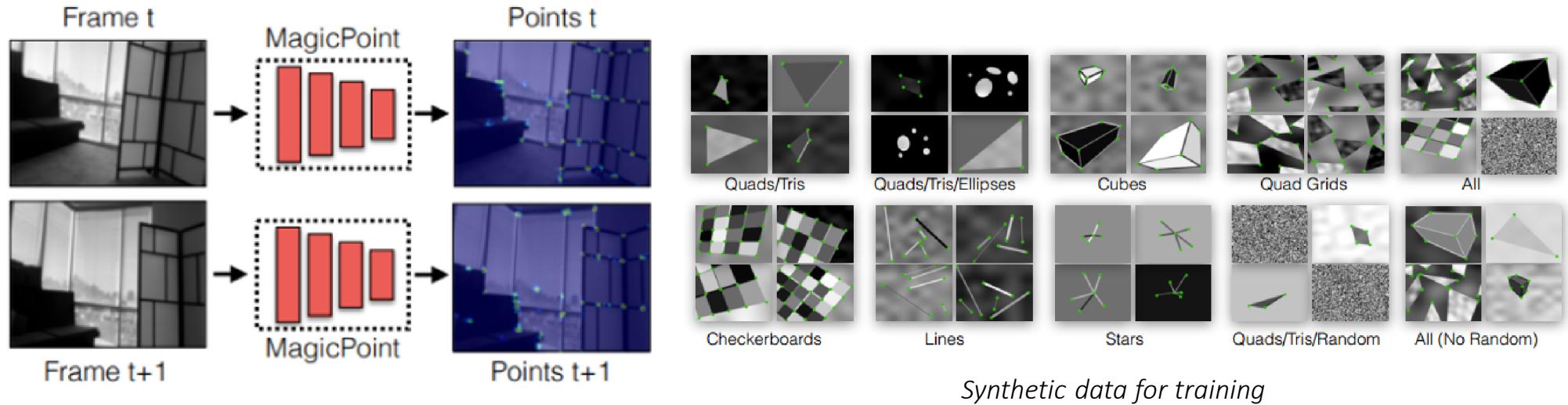
Matching

Matches before filtering

Incorrect pose

Matches after filtering

Correct pose estimation

✓ **Ratio test**

✓ **Keypoints spread**

✓ **Geometry validation**

Sample 4 random matches

Hi

Project points

Count points with projected distance

Extract solution with max amount of inliers

Hfinal

Random sample consensus scheme
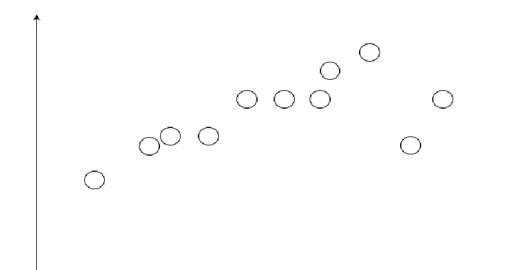
RANSAC for line fitting

AR: Algorithms and solutions

Scale and Rotation Invariance

Problem: Local feature descriptors (SIFT, SURF, ORB, FREAK, etc.) do not handle extremal angles
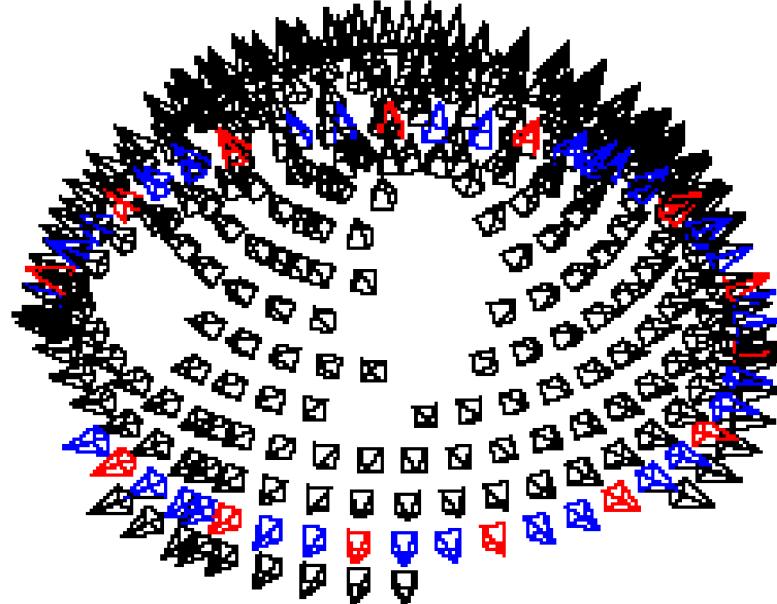
Synthetic views warping

Generated views

D-Nets
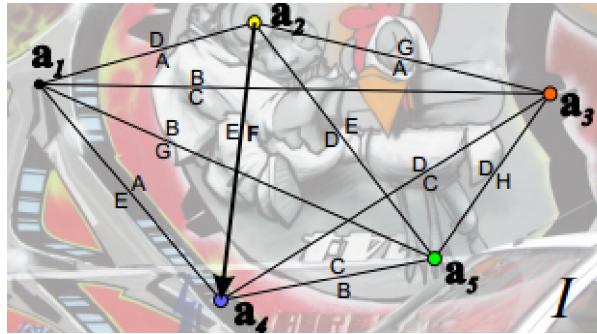
Geometric principles
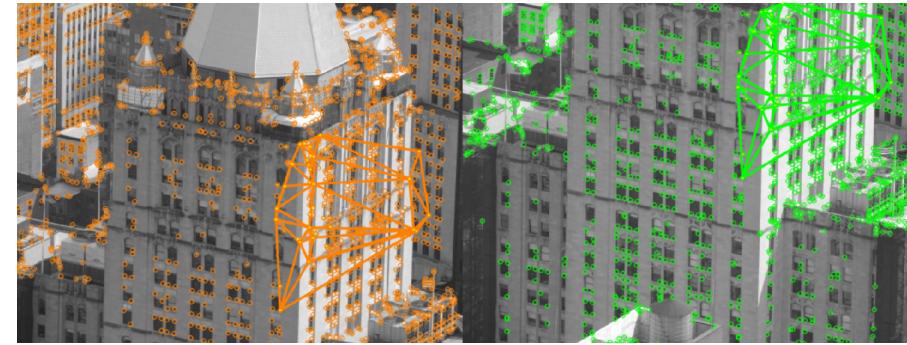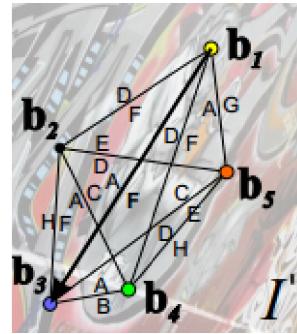
Two-view matching

Live example
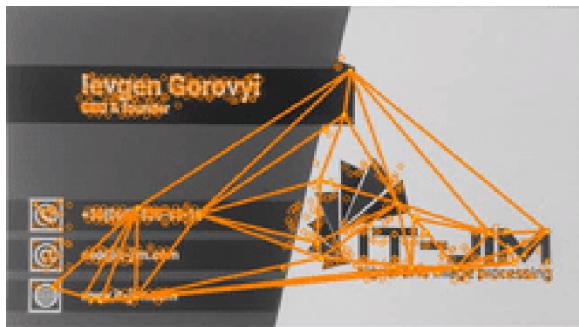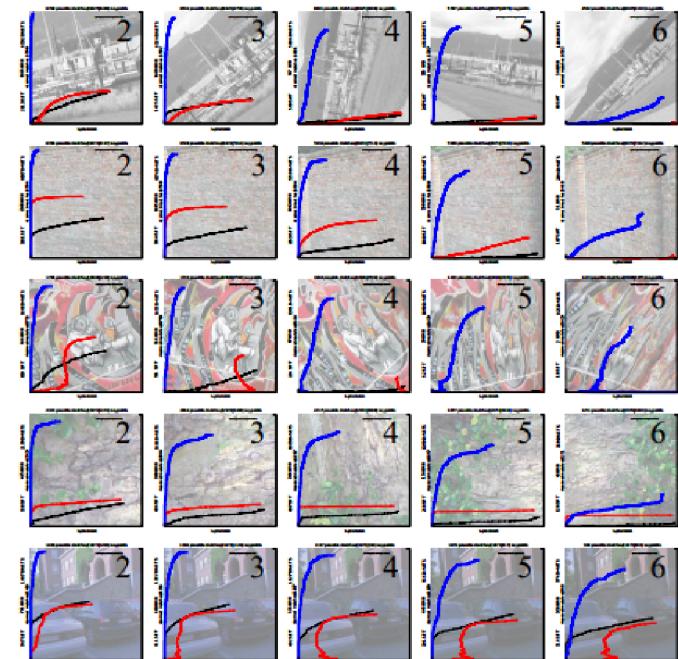
D-Nets vs ORB, SIFT

MagicWarp
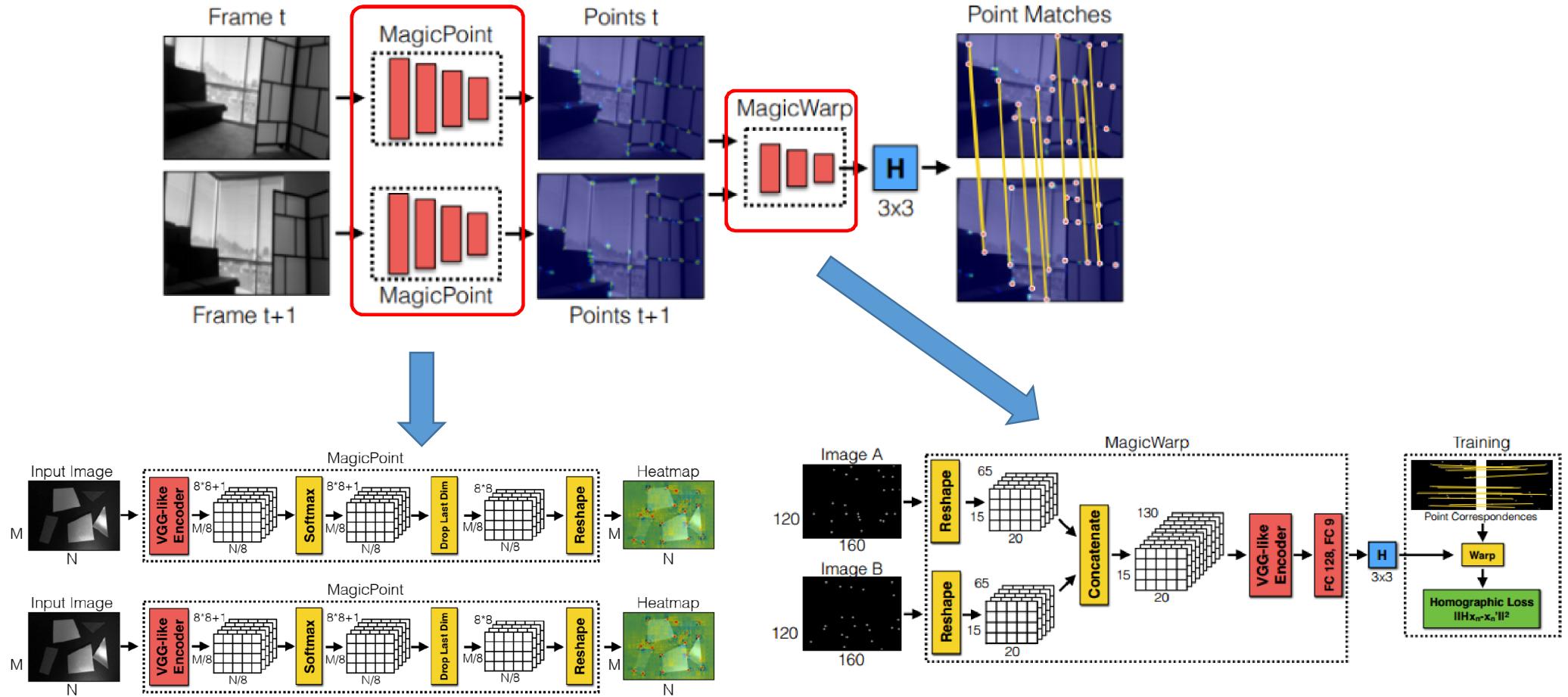
AR: Algorithms and solutions

D. DeTonem T. Malisiewicz, A.Rabinovich - "Toward Geometric Deep SLAM" / https://arxiv.org/pdf/1707.07410.pdf

MagicWarp Evaluation
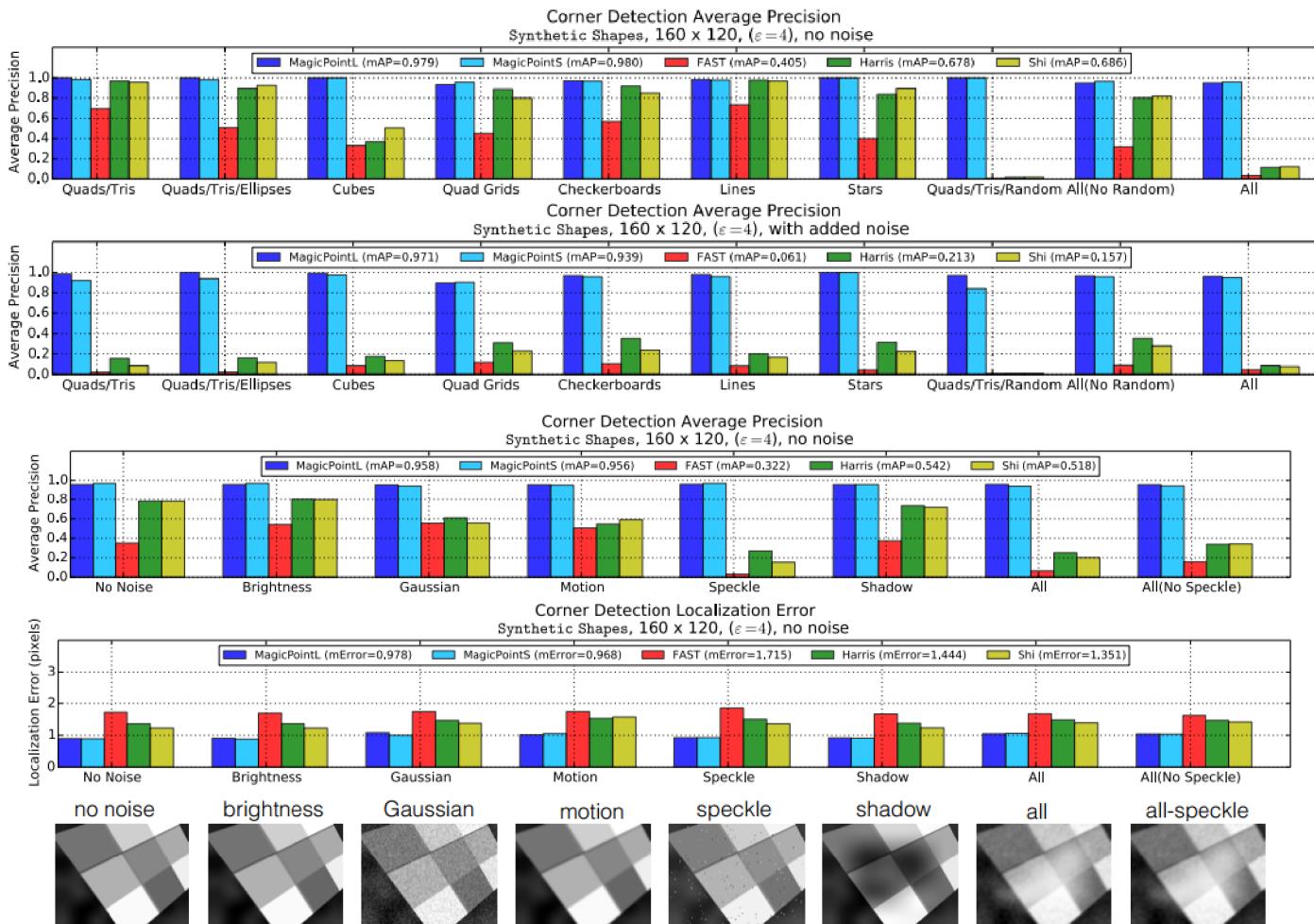
Relative Camera Pose using CNN
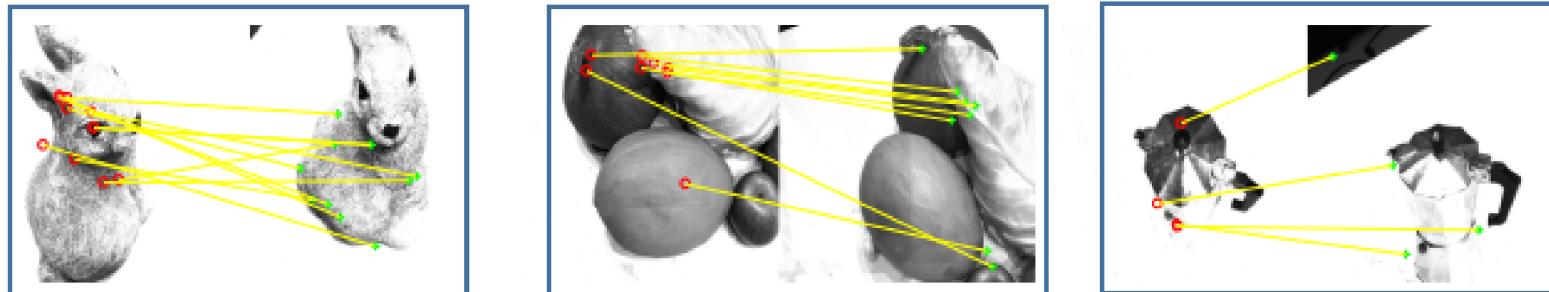
Challenging matching conditions

Network architecture

Δp

Representation part

Regression part
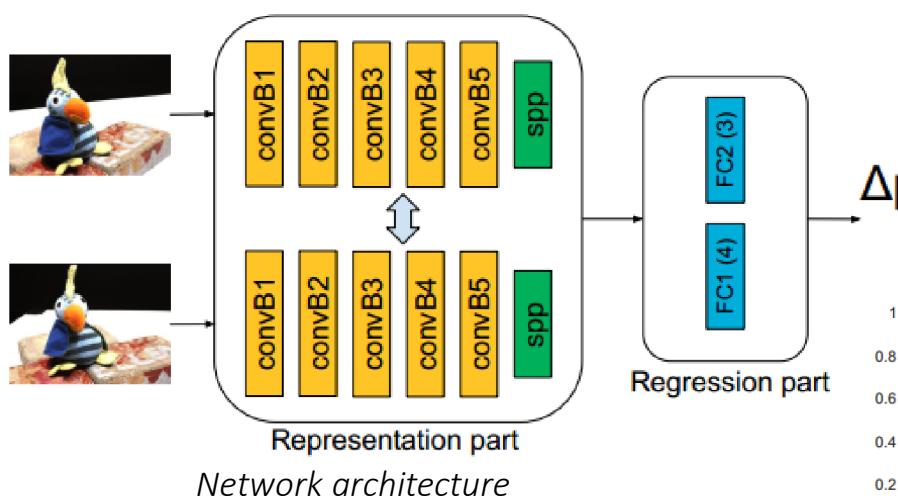
Textureless and light reflections
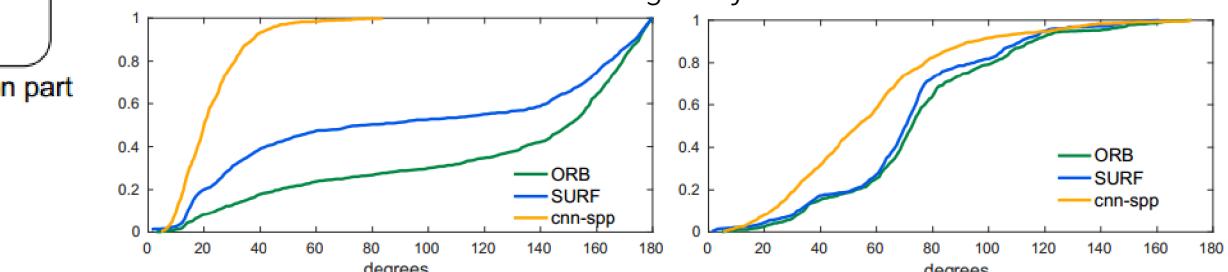
Cumulative hist

AR: Algorithms and solutions

Iaroslav Melekhov, Juha Ylioinas, Juho Kannala, Esa Rahtu
Relative Camera Pose Estimation Using Convolutional Neural Networks, International Conference on Advanced Concepts for Intelligent Vision Systems, 2017

AR: Algorithms and Solutions

- camera pose and marker detection
- image retrieval
- tracking engine
- markerless and SLAM

Image Retrieval: BoW

Question: How to track 1000 markers in real-time?

Image Retrieval: d-BoW

BOW   K-means
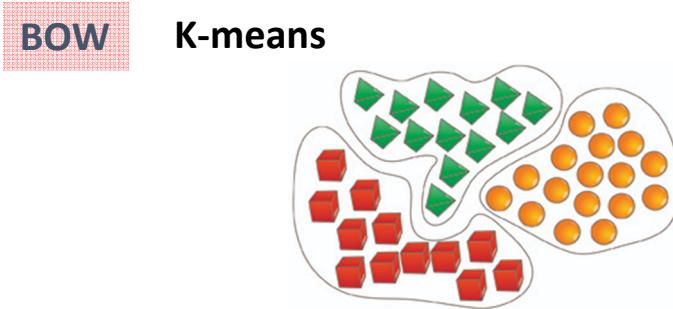
Algorithm:
1. Build feature -histogram
2. Match Histograms
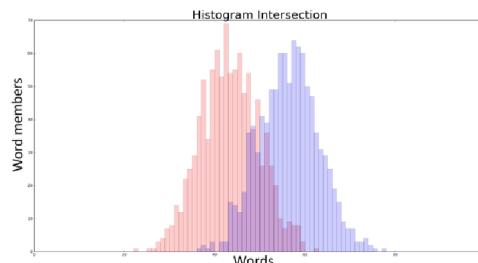
✓ Simple
✗ For small datasets
✗ Fixed feature vector size
✗ Computation time

d-BOW   Hierarchical K-means

Algorithm:
1. Define tree structure
2. Build Voc + DB
3. Match BoW feature vectors

✗ More complex solution
✓ For large datasets
✓ feature vector size < #clusters
✓ Computation time is low

33

Real-Time Marker Detection (1000 markers)

Image retrieval algorithm

#checks histogram

Retrieval time distribution
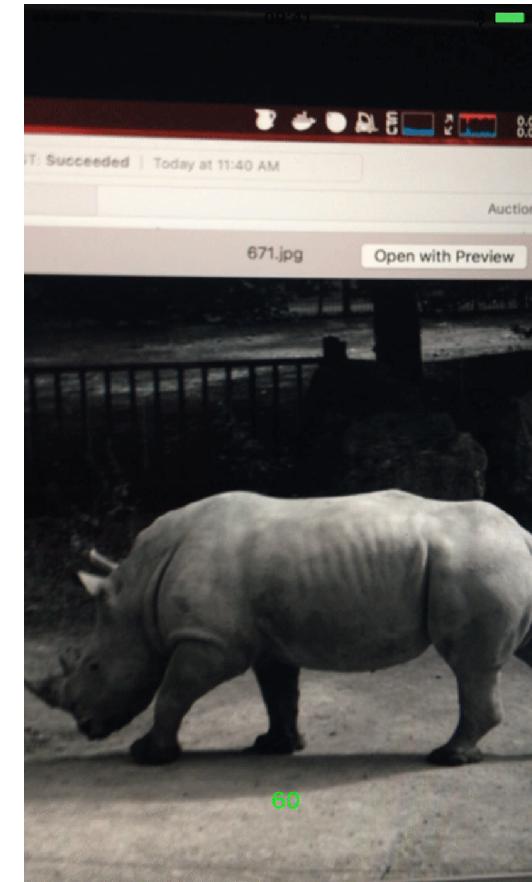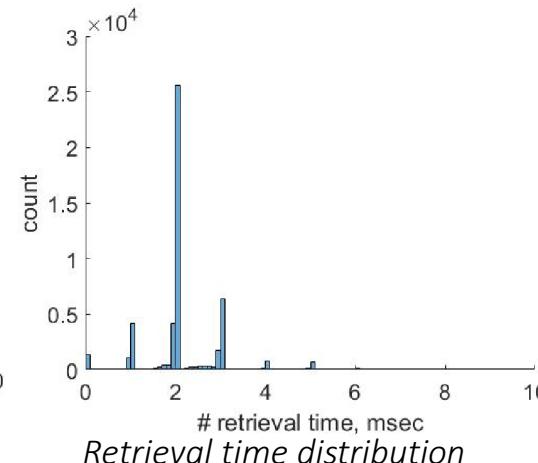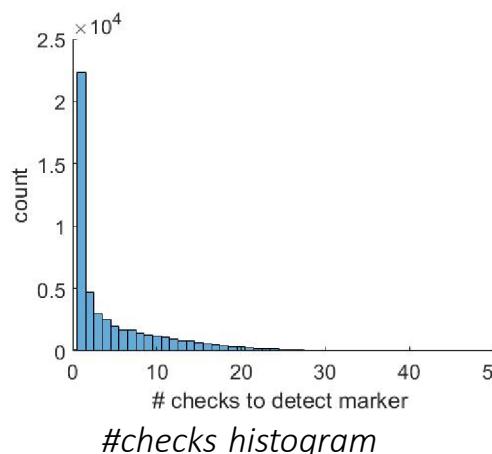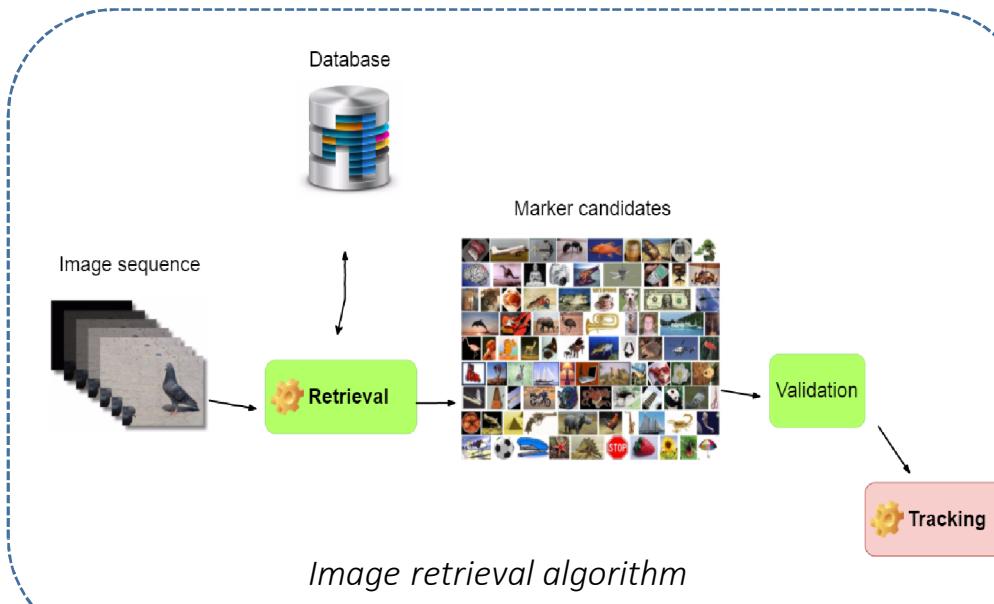
Retrieval example (iPhone 6)

AR: Algorithms and Solutions

- **camera pose and marker detection**
- **image retrieval**
- **tracking**
- **markerless and SLAM**

Detection vs Tracking

Problem: Tracking-by-detection is not good idea
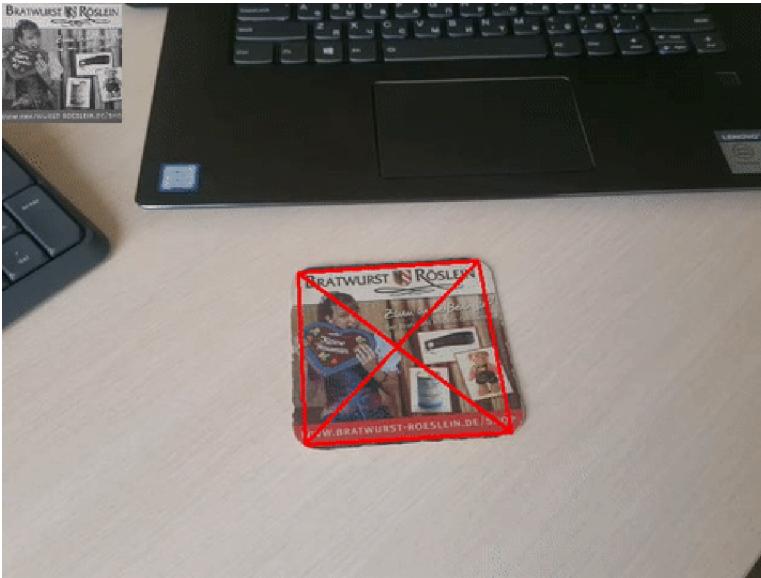
🚫 Slow

🚫 Fails for extremal angles

➕ Fast

➕ More robust to geometry

Marker Tracking: Optical Flow (LK)
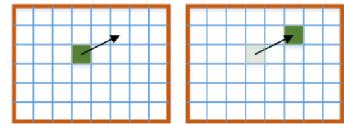
$$I_x u + I_y v + I_t = 0$$

$$\begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_y I_x & \sum I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}$$
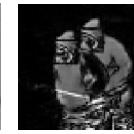
Ix, Iy, It

Image patch

467

Example of keypoints tracking

Video streamFrame 1

Optical flow magnitudeFrame 2

36

DeepFlow

P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid – "DeepFlow: Large displacement optical flow with deep matching" / ICCV-2013, Dec 2013, Sydney, Australia

DeepFlow Evaluation

Response maps

GT

dense HOG matching

DeepFlow

GT

dense HOG matching

DeepFlow

Matches   Optical Flow   Matches   Optical Flow

P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid – ” DeepFlow: Large displacement optical flow with deep matching ” / ICCV-2013, Dec 2013, Sydney, Australia

DL for Optical Flow: FlowNet

A. Dosovitskiy *et al* – "FlowNet: Learning Optical Flow with Convolutional Networks" / ICCV-2015, December 11-18, 2015, Santiago, Chile

DL for Optical Flow: FlowNet

A. Dosovitskiy *et al* – "FlowNet: Learning Optical Flow with Convolutional Networks" / ICCV-2015, December 11-18, 2015, Santiago, Chile

DL for Optical Flow: FlowNet 2.0

✓ Introduce a warping operation and stacking multiple networks

✓ Specialized network for subpixel motion

DL for Optical Flow: FlowNet 2.0

AR: Algorithms and solutions

42

DL for Optical Flow: FlowNet 2.0

E. Ilg et al – " FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks" / CVPR-2017, July 22-25, 2017, Honolulu, Hawaii
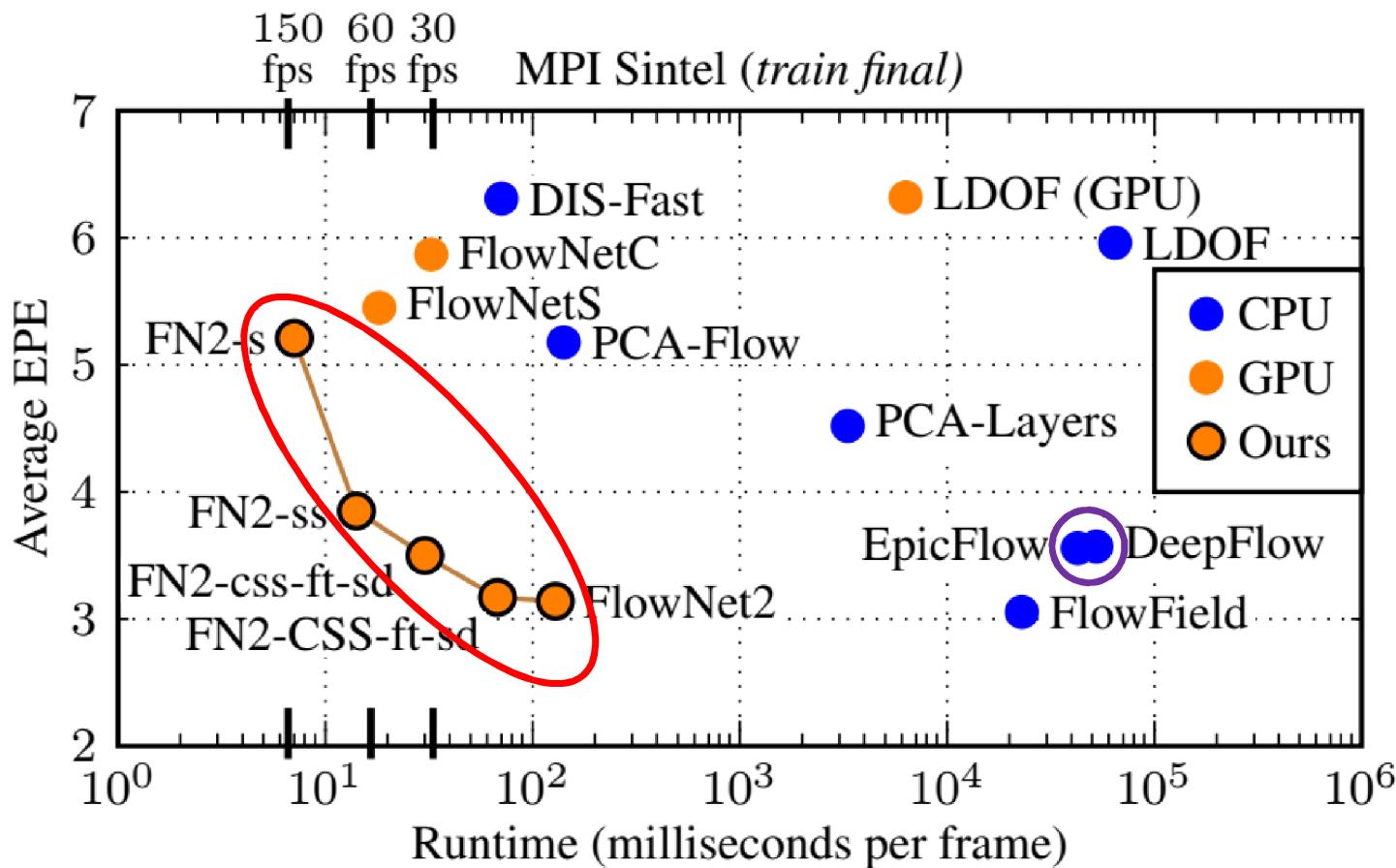
Template-Based Tracking: Lucas-Kanade
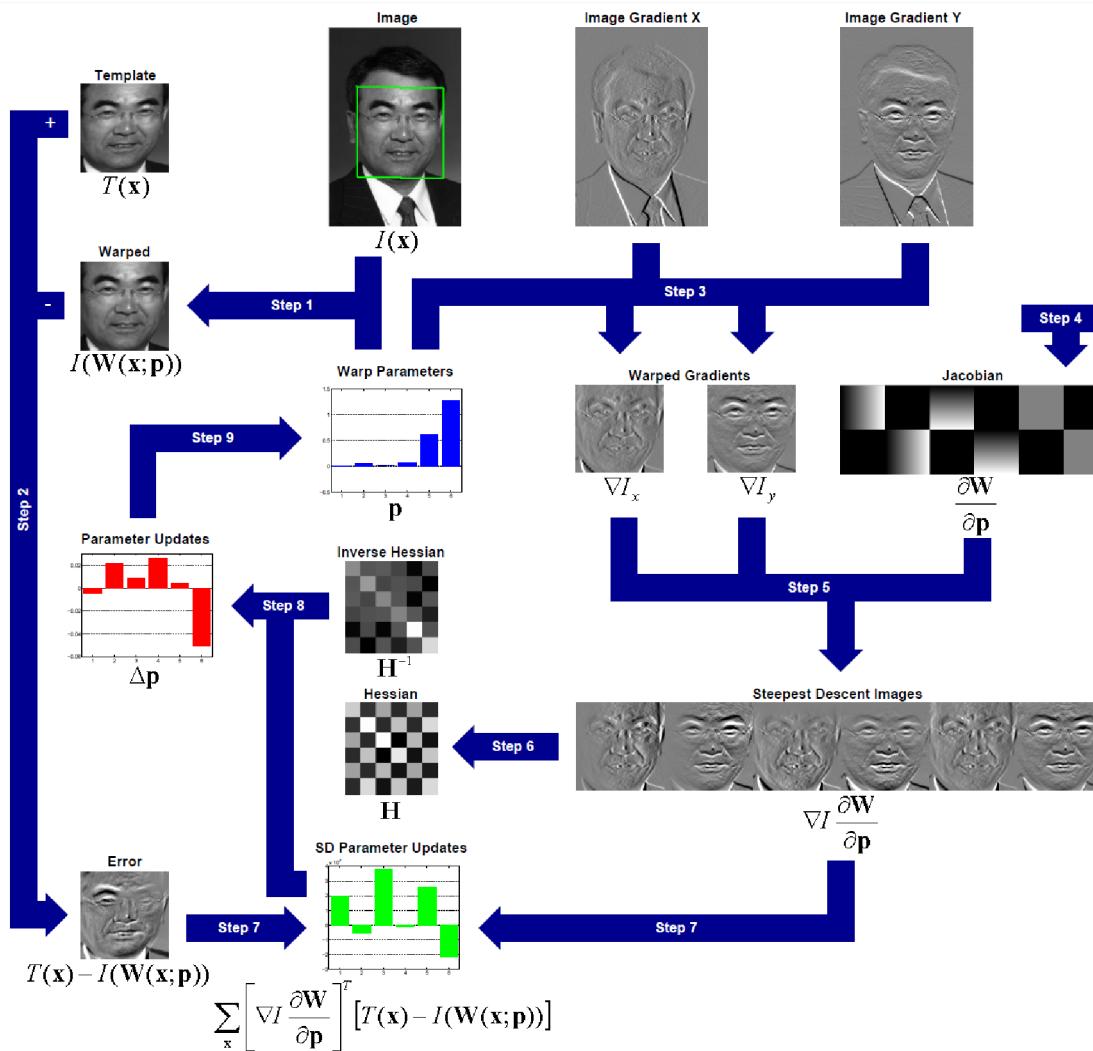
Template-based tracking (Gauss-Newton gradient descent)

44

Template-based tracking: Improvements

Efficient second-order minimization (ESM)
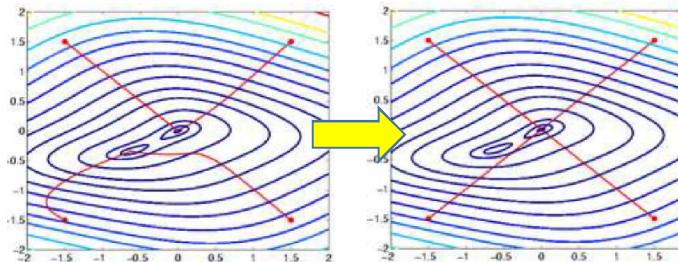
✓ less iterations needed

✓ larger area of convergence

✓ avoid local minima

✓ second-order Taylor approximation

✗ Jacobian is calculated each time

Sum of conditional variance (SCV)

ESM, but:

$$\triangle = \sum_{\mathbf{x}} \left(I(w(\mathbf{x}, \mathbf{p})) - T(\mathbf{x}) \right)^2$$

current image   reference image

$$\mathcal{E}(I(w(\mathbf{x}, \mathbf{p}))|T(\mathbf{x}))$$

expectation operator

SSD = 2.4e+9
SCV = 0

✓ more robust

ESM with gradient orientations (ESM-GO, 2017)
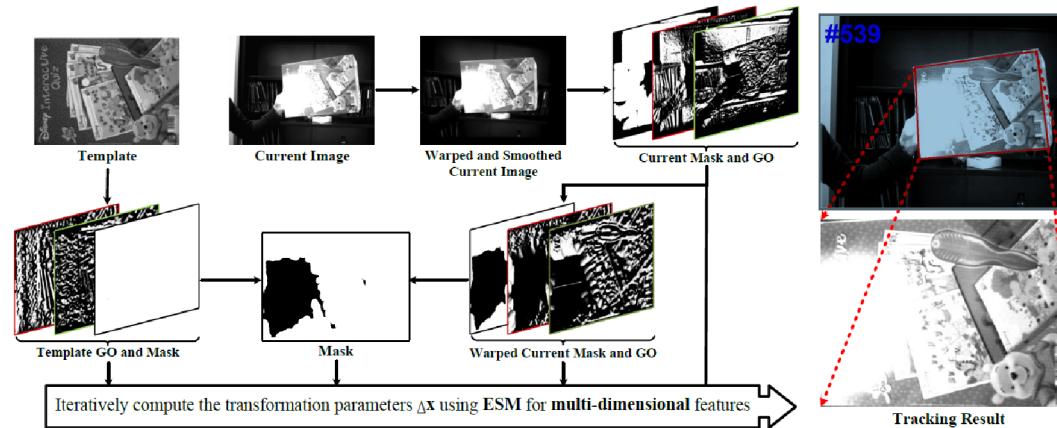
image gradients
⬇
illumination robustness

anisotropic diffusion
⬇
noise robustness

masking
⬇
low texture robustness

45

Fusion of Trackers

$$H_{cur} = H_{prev} \times \Delta H_{1..N}$$

Template

Frame

Hprev

ESM

Hprev*ΔH

Homography refinement

iteration: 1
esm error: 0.000000

Convergence steps
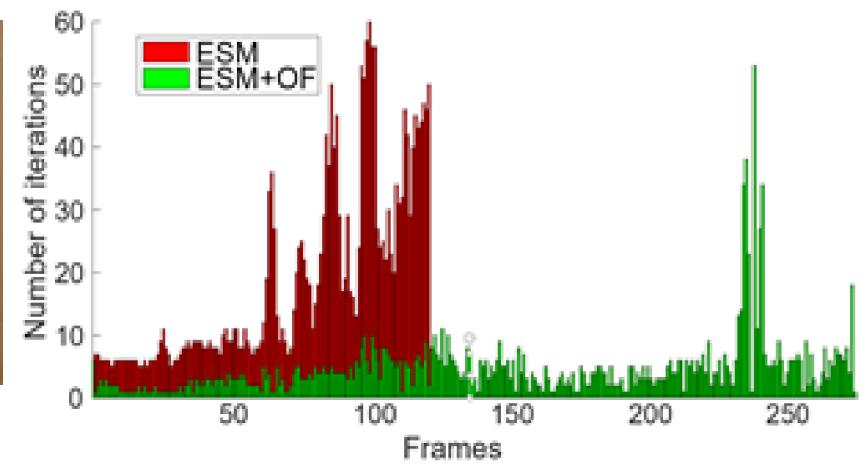
ESM tracking

Hybrid tracking

Hybrid Tracking*

✓ Drift-free

✓ Extremal angles

Hybrid tracking and camera pose estimation
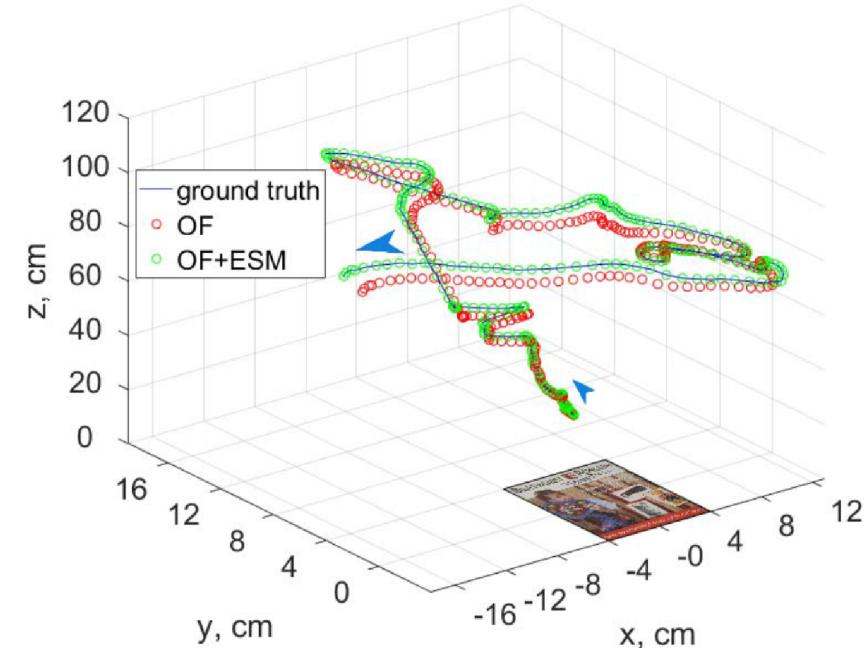
AR: Algorithms and Solutions

- **marker detection**
- **image retrieval**
- **tracking engine**
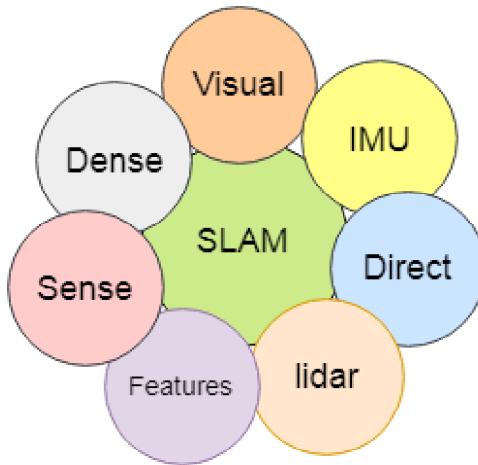- **markerless and SLAM**

Markerless: VIO and SLAM

VO: Visual Odometry.
Goal: camera tracking using the scene. Local map optimization

SLAM: Simultaneous Localization and Mapping.
Goal: persistent 3d and camera location simultaneously. Global map optimization
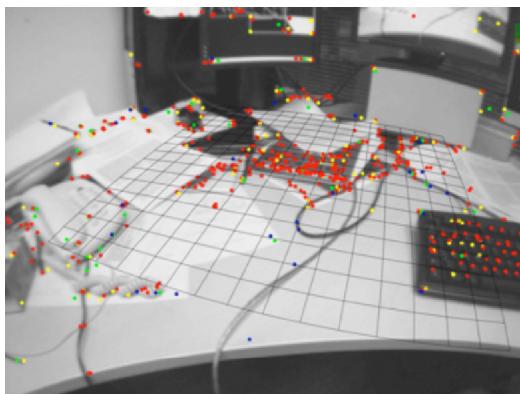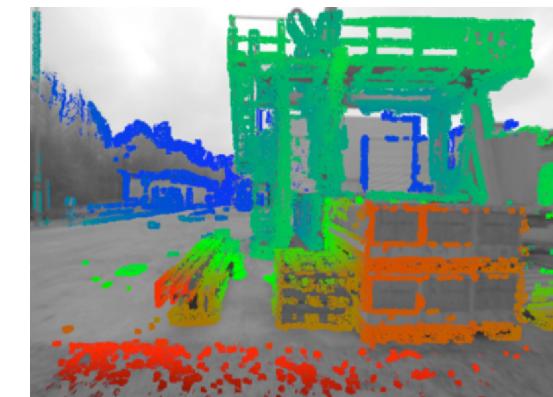

SLAM for AR
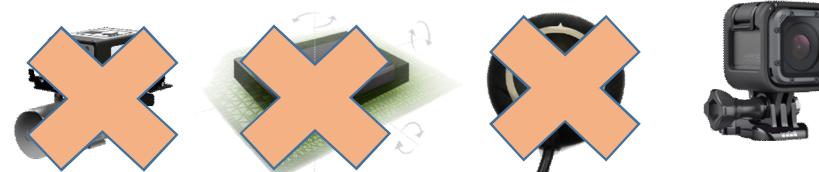
SLAM: Direct/Indirect, Sparse/Dense

Local features

Sparse map

Dense map

49
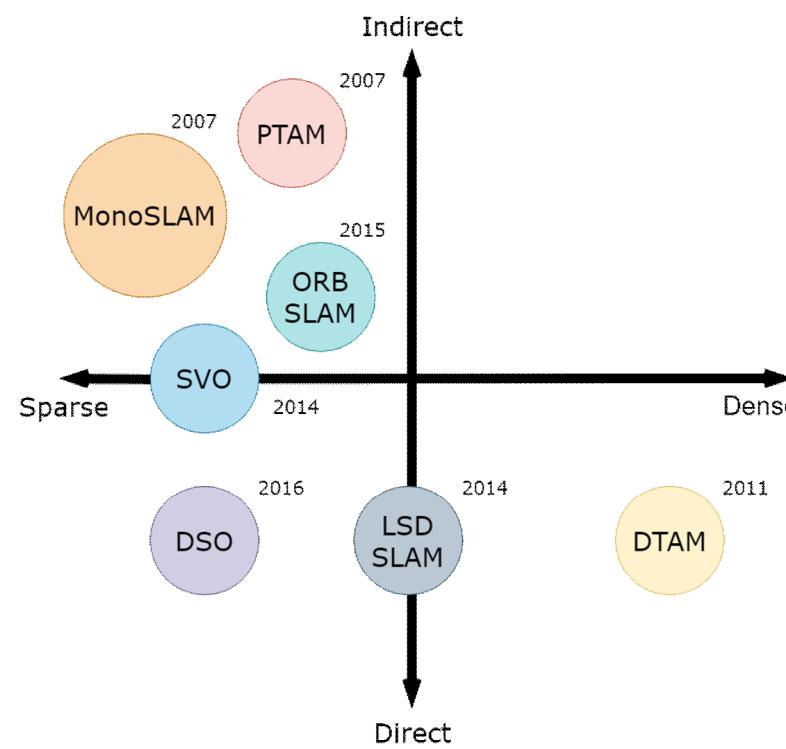
Slam System Components

01 Initialization

02 Tracking & mapping

03 Optimization

04 AR interaction

50
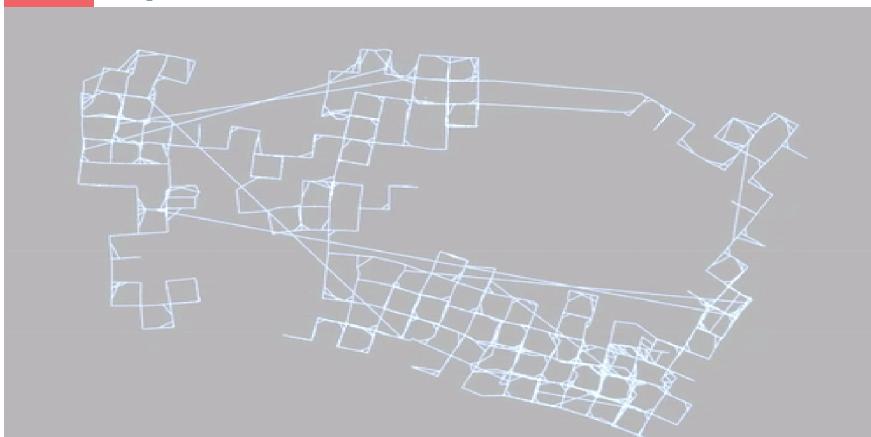
Parallel Tracking and Mapping (PTAM)

PTAM workflow
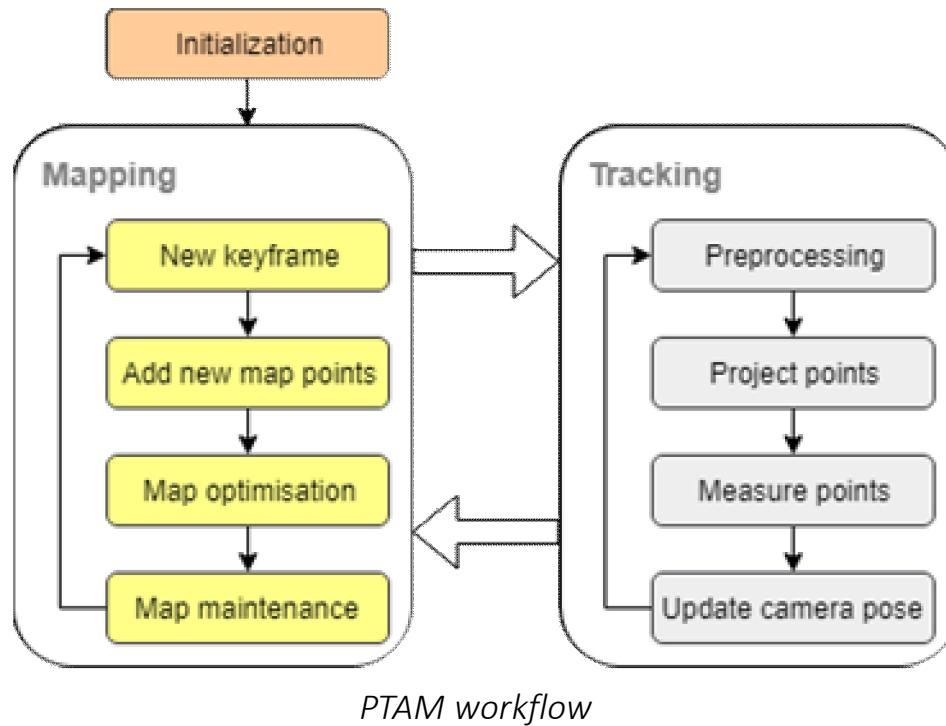
Parallel Tracking and Mapping
for Small AR Workspaces

Extra video results made for
ISMAR 2007 conference

Georg Klein and David Murray
Active Vision Laboratory
University of Oxford

Demo (iPhone 3!)

Parallel tracking and mapping for small AR workspaces, Klein G, Murray D, ISMAR 2007
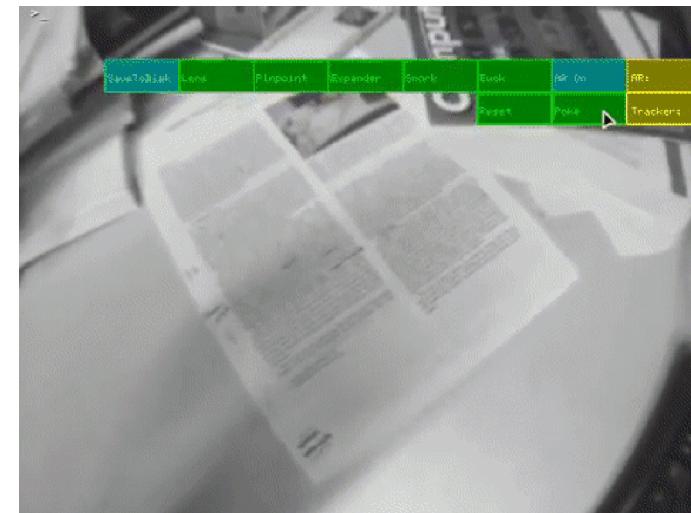
ORB-SLAM
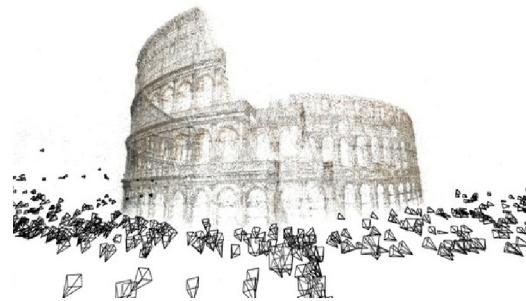
Tracking ✓   Map ✓   Matching ✓
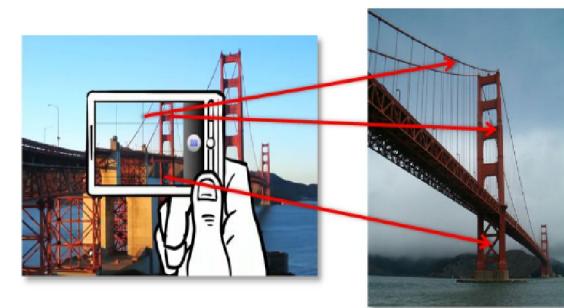
Feature tracking   Camera tracking

Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, 2015.
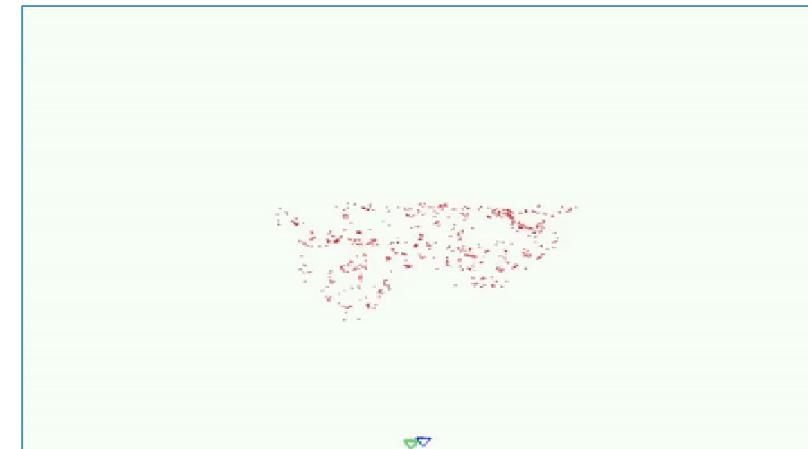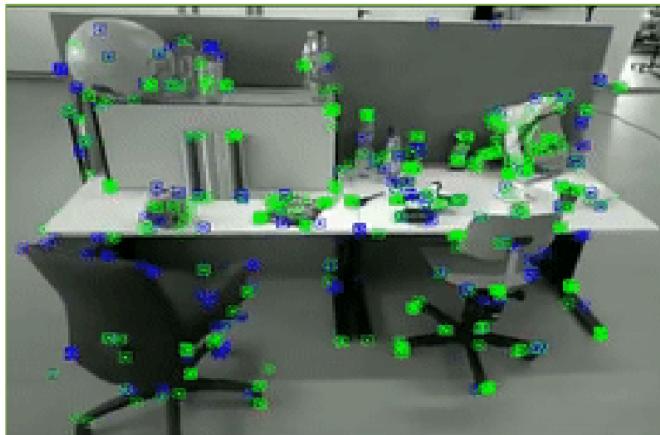
LSD SLAM

Tracking ✓   High gradients ✓   Map ✓
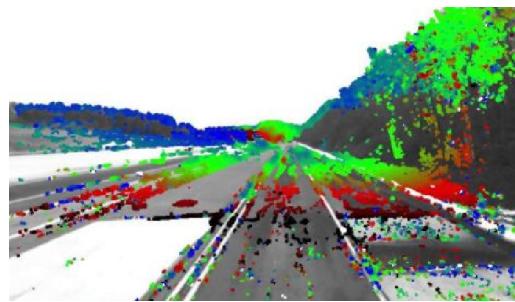
Image tracking

Mapping

LSD-SLAM: Large-Scale Direct Monocular SLAM, J. Engel, T. Schöps, D. Cremers, ECCV '14
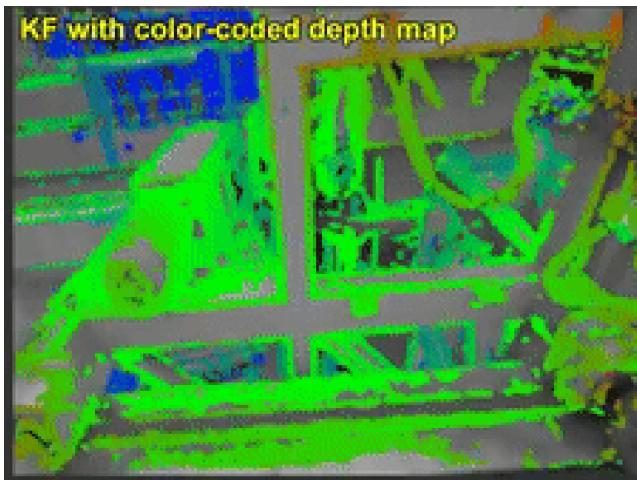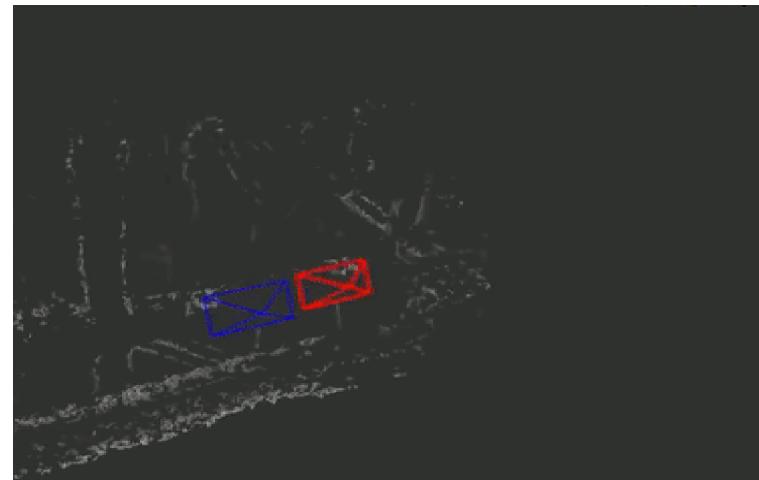
Deep Virtual Stereo Odometry

StackNet: generation of stereo from mono

N.Yang, R.Wang, J.Stuckler and D.Cremers – "Deep Virtual Stereo Odometry: Leveraging Deep Depth Prediction for Monocular Direct Sparse Odometry" / ECCV-2018, September 8-14 2018, Munich, Germany
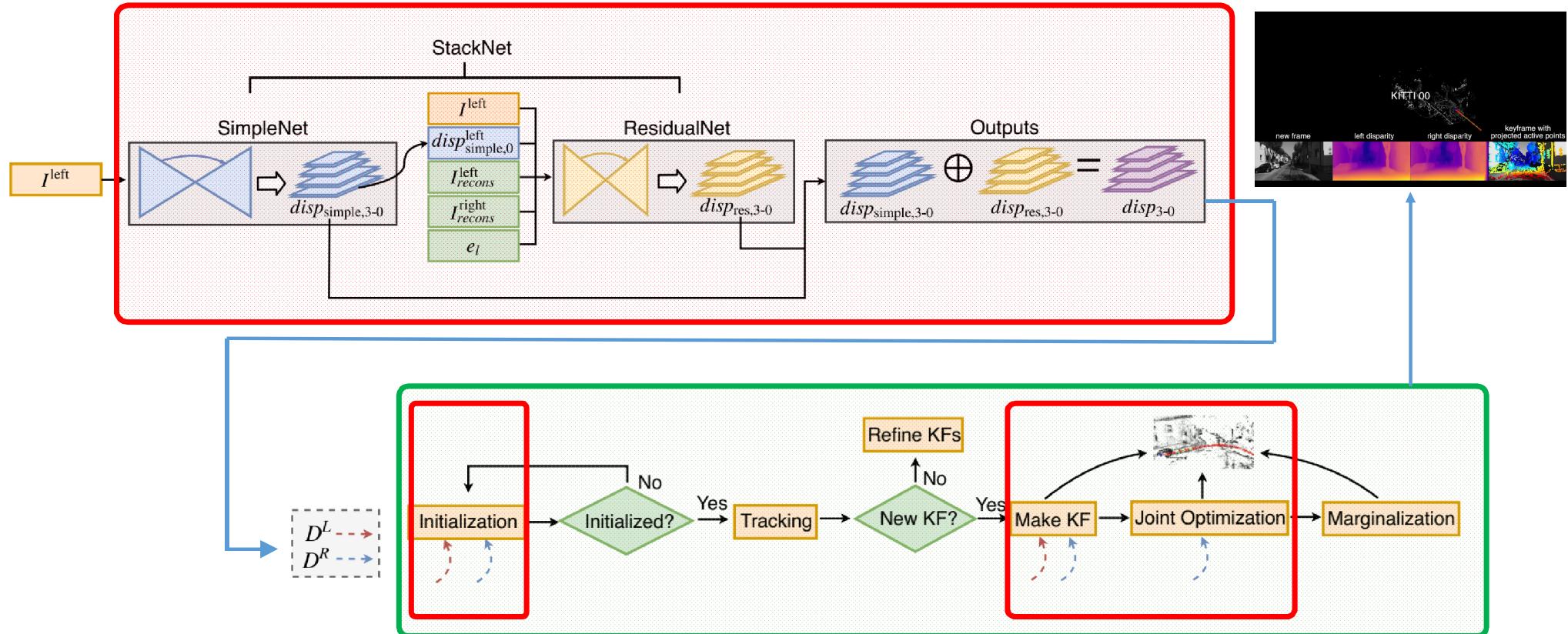
Deep Virtual Stereo Odometry
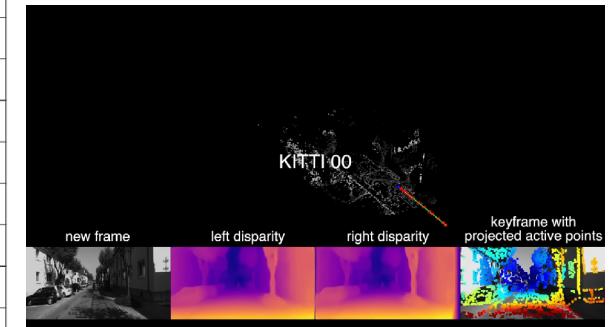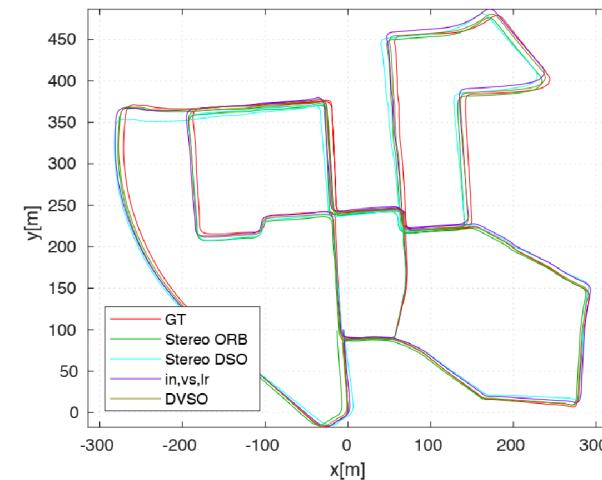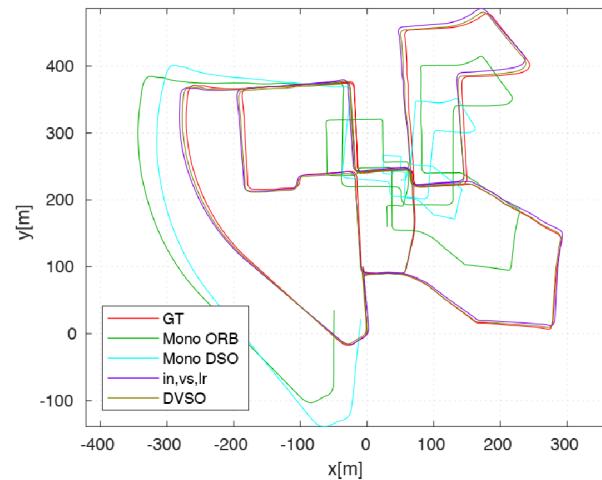
Image size: 512×256

Inference time: 40 ms
(Titan X Pascal GPU)
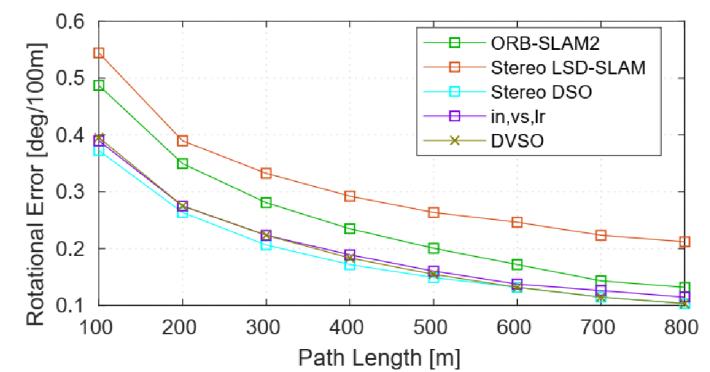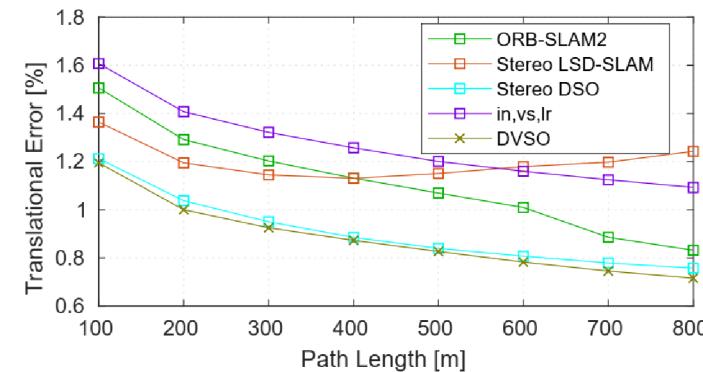
N.Yang, R.Wang, J.Stuckler and D.Cremers – "Deep Virtual Stereo Odometry: Leveraging Deep Depth Prediction for Monocular Direct Sparse Odometry" / ECCV-2018, September 8-14 2018, Munich, Germany

Deep Depth Densification (D³)

Method idea

Subsampled depths and distances

Z.Chen, V.Badrinarayanan, G.Drozdov, A.Rabinovich – "Estimating Depth from RGB and Space Sensing" / ECCV-2018, September 8-14 2018, Munich, Germany
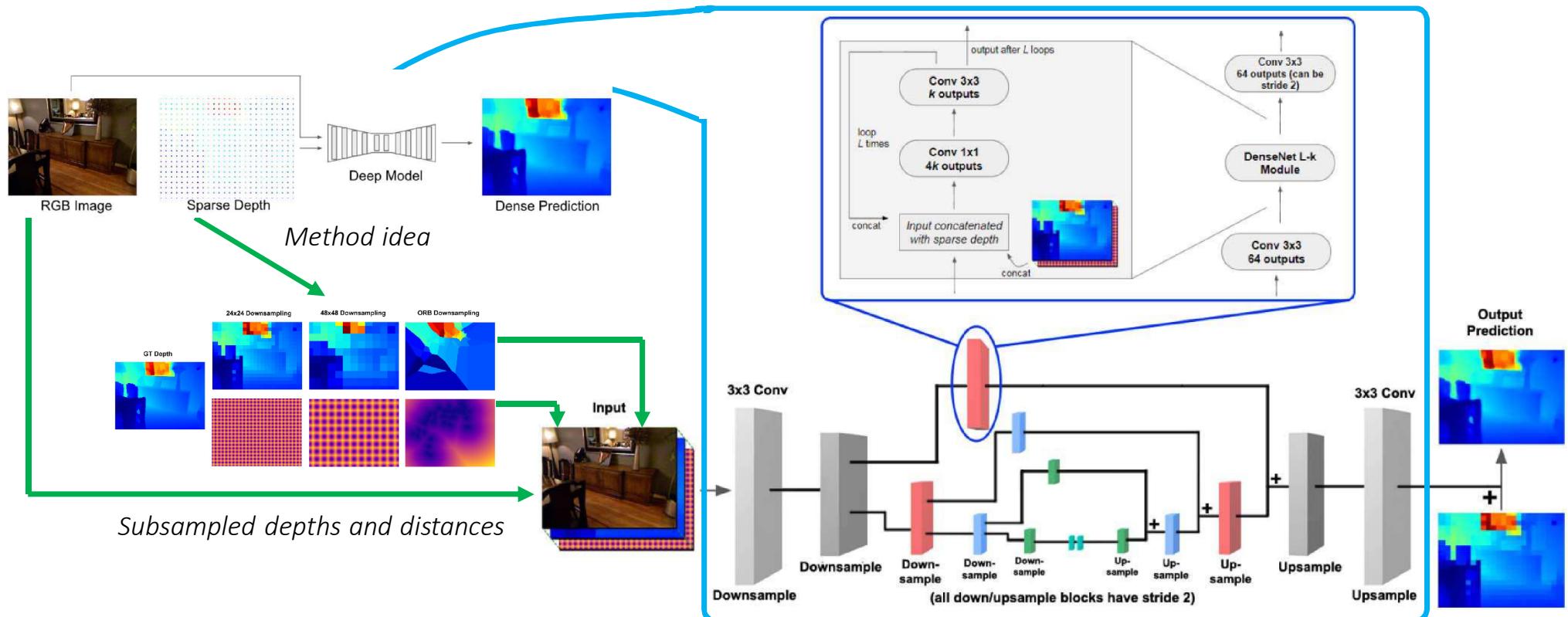
Deep Depth Densification (D³)
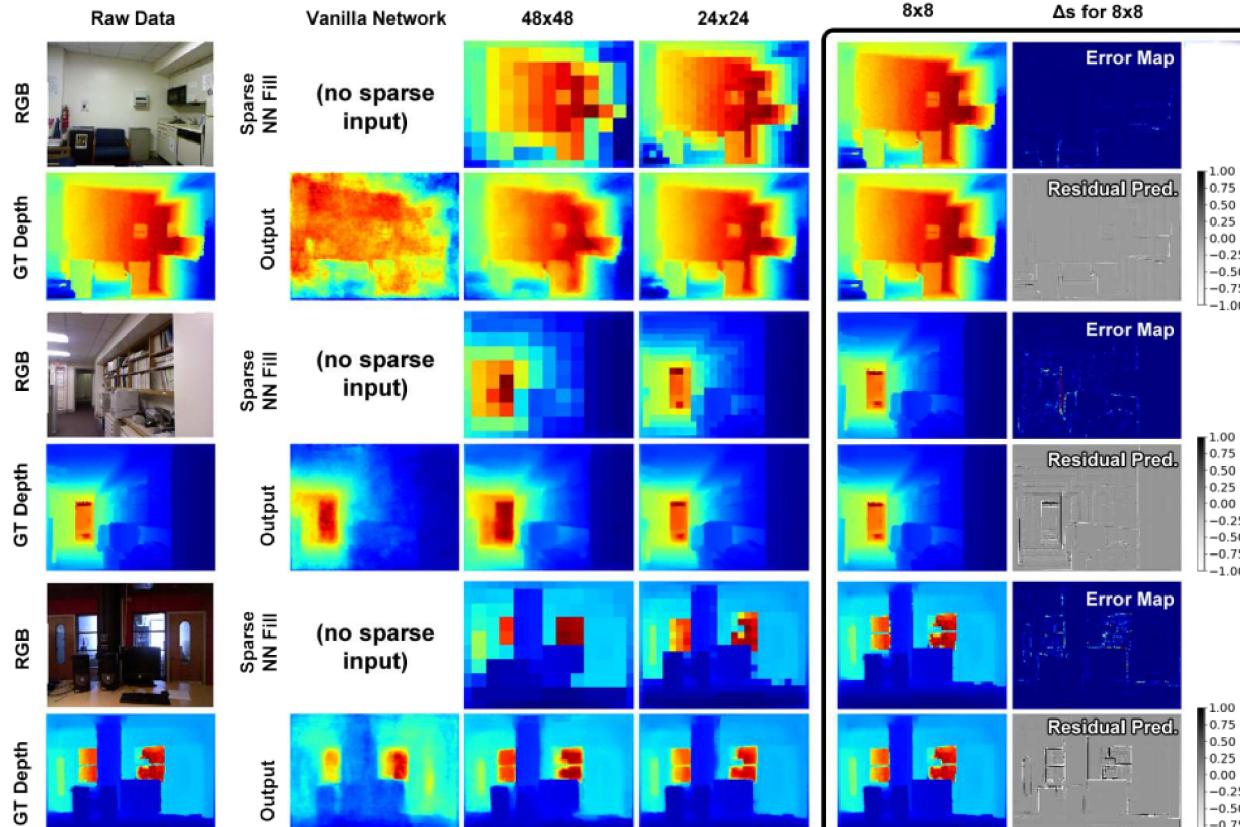
Comparative results

RMS errors

AR: Algorithms and solutions

Z.Chen, V.Badrinarayanan, G.Drozdov, A.Rabinovich – "Estimating Depth from RGB and Space Sensing" / ECCV-2018, September 8-14 2018, Munich, Germany
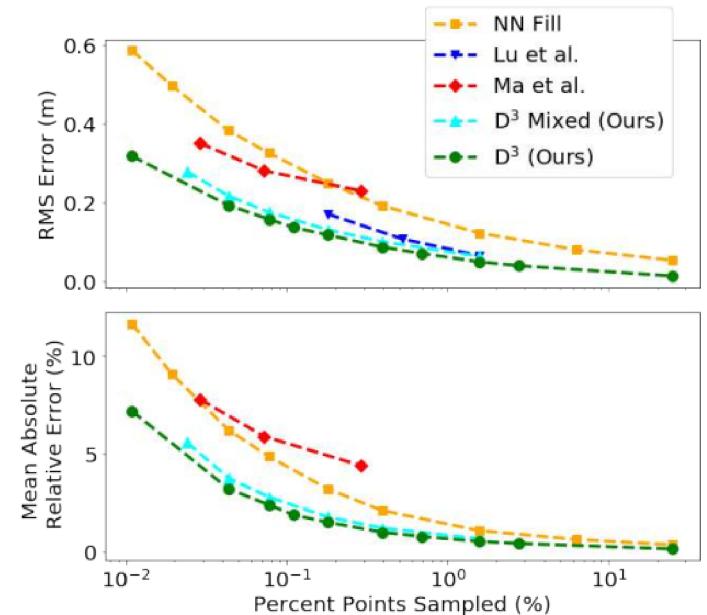
Mobile AR in Production

AR Data Flow: User's Side

AR Marker Quality

Quality check list:

✓ Image contrast
✓ Image details
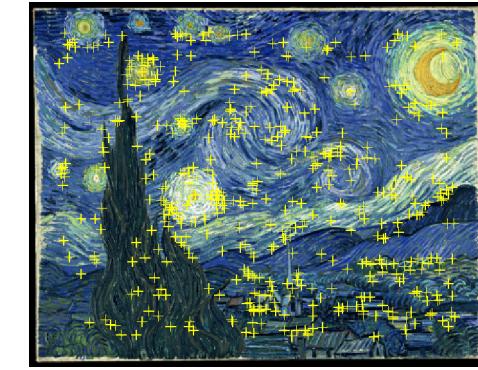✓ Features distribution
✓ Features repeatability and uiqueness
✓ Image viewing angles and scale

Best ROI selection:

Criteria:
1. Number of interest points in window
2. Interest points distribution in window

Original image   Sliding window processing   Marker

Mobile AR in production

59

Preprocessing: Text Extraction

Three main concepts:

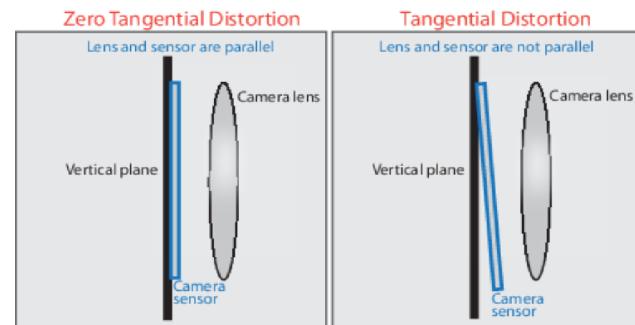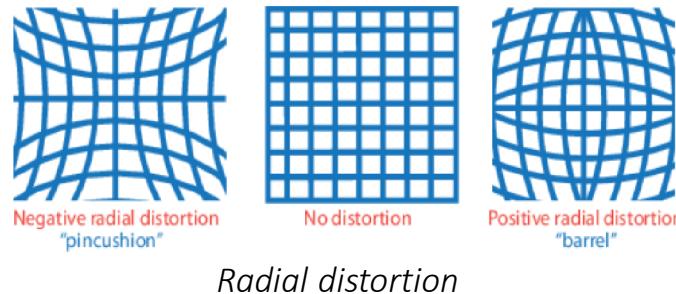| High-Pass filtering	Laplacian of Gaussian	Laplacian

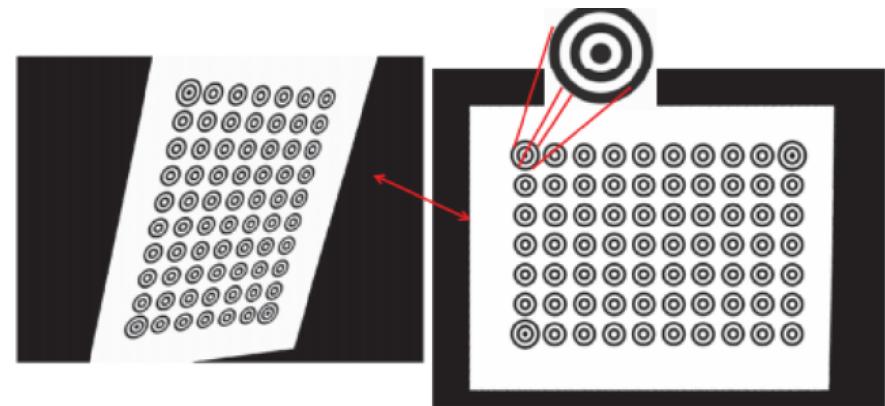Initial Grayscale Image   Initial Grayscale Image   Initial Grayscale Image

Camera Calibration

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \overbrace{\begin{bmatrix} f_x & 0 & u_x \\ 0 & f_y & v_y \\ 0 & 0 & 1 \end{bmatrix}}^{K} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{32} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$
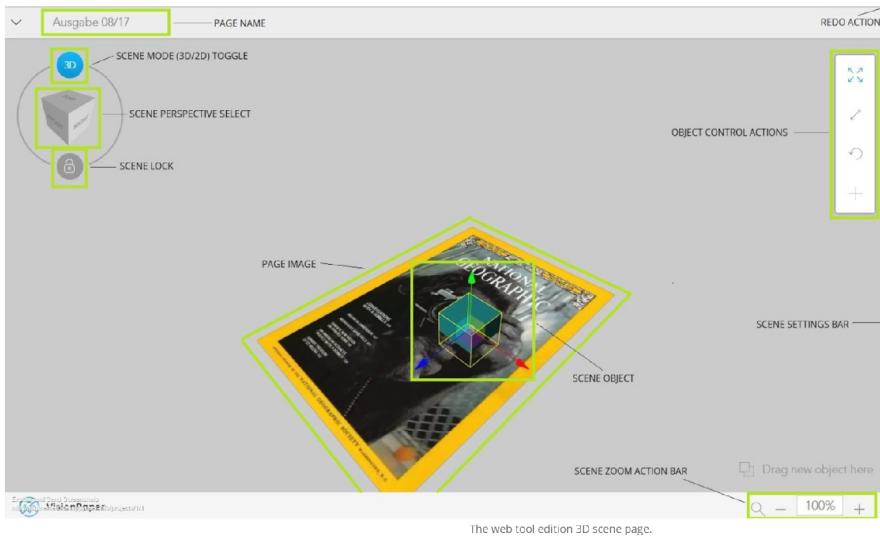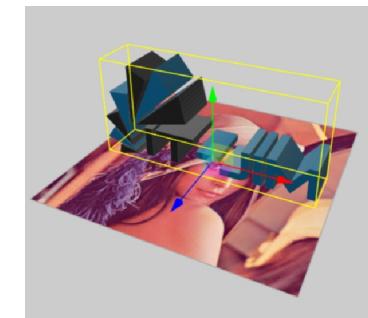

Negative radial distortion "pincushion"   No distortion   Positive radial distortion "barrel"

Radial distortion

Zero Tangential Distortion — Lens and sensor are parallel

Tangential Distortion — Lens and sensor are not parallel

Vertical plane   Camera lens   Camera sensor

Tangential distortion
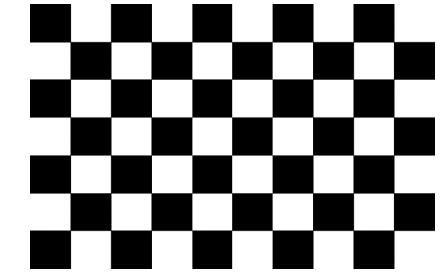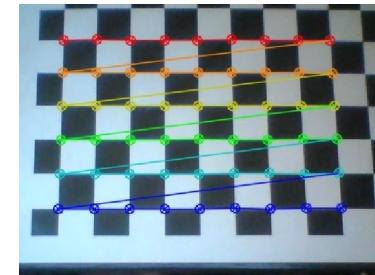
Checkboard

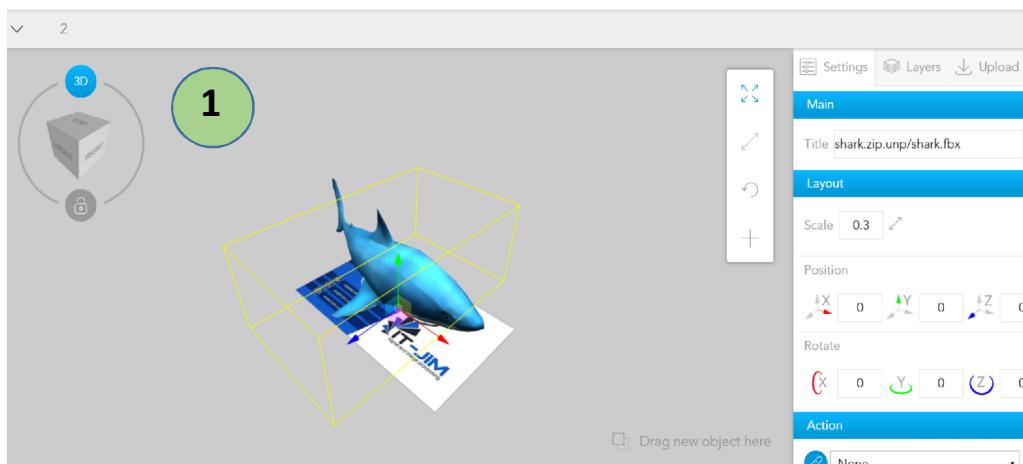*Calibration pattern example*

Web AR Tool

AR WebTool

2

1
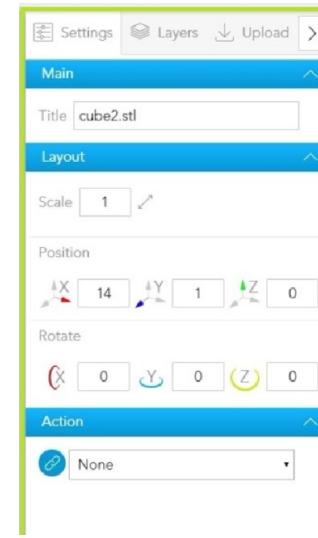
3D model in WebTool　　Mobile AR in production　　　　　*3D model in mobile APP*

Web AR Tool: How it Works?

AR Data Flow: SDK Side

AR tool (WEB or Local)

1

2

Select marker
(Image data)

AR core

Select data for augmentation
(3d model/image/video/audio)

JPEG

AVI   MP3

3

iOS&Android API

Tracking data
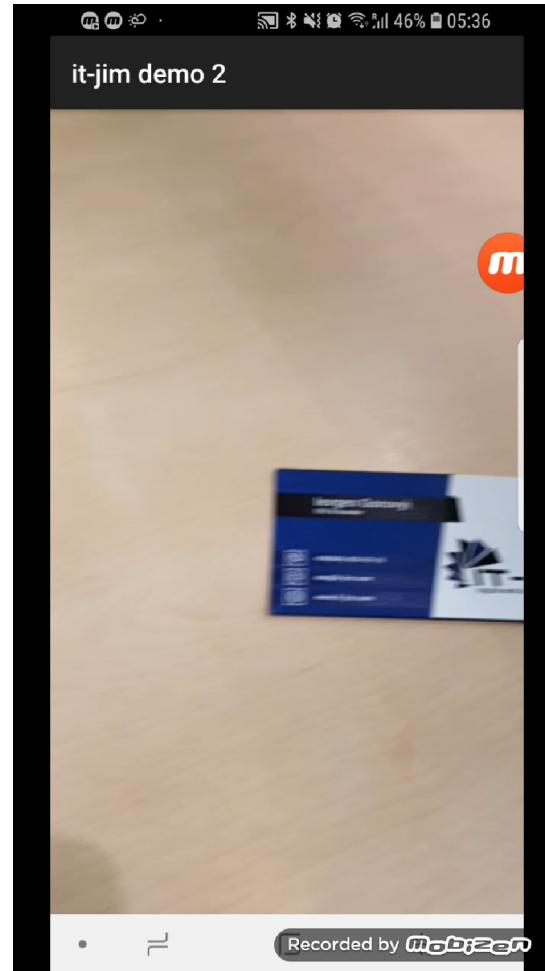
Client app
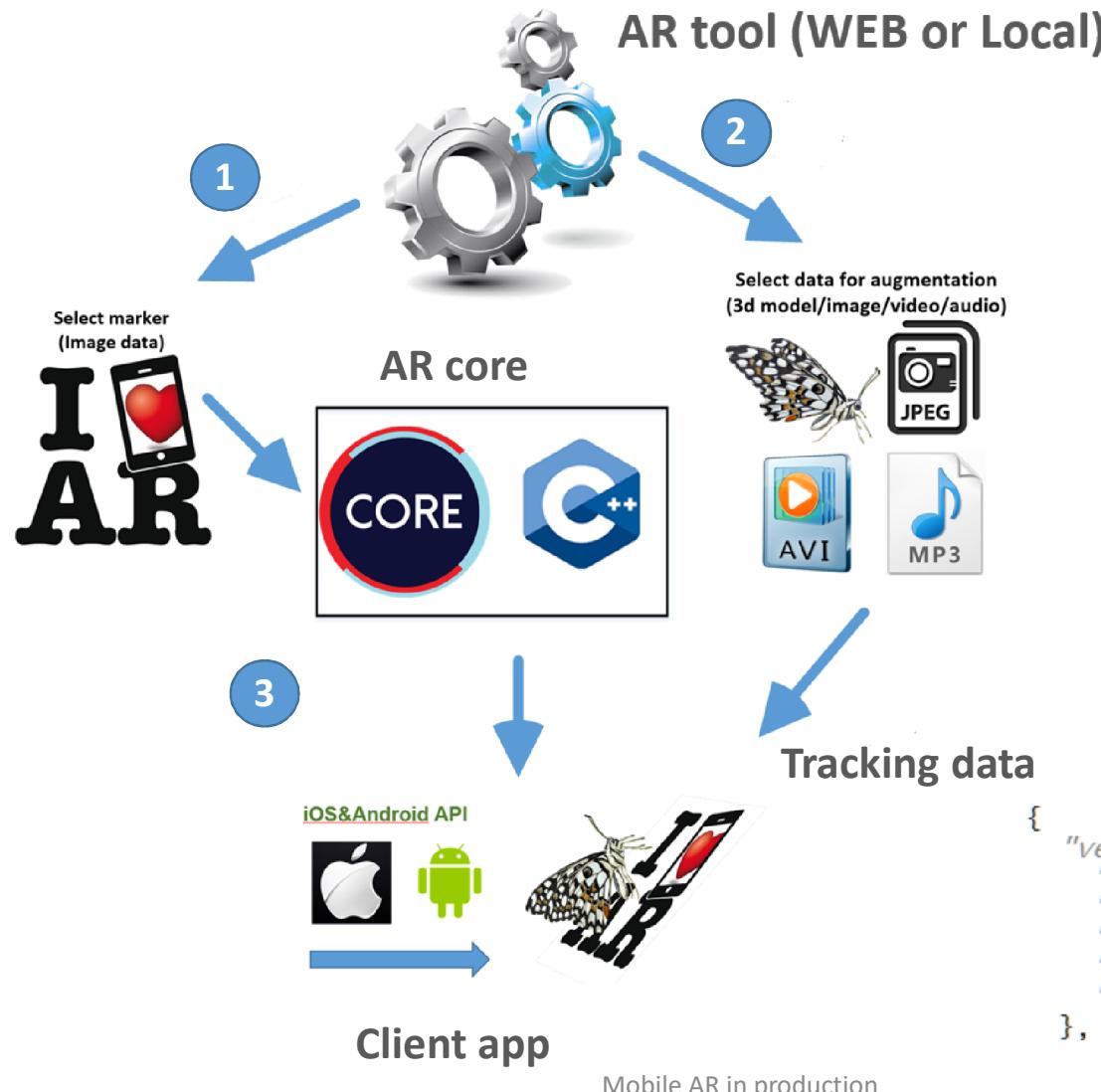
markers

models

Config.json

dbow.bin

pyramid.bin
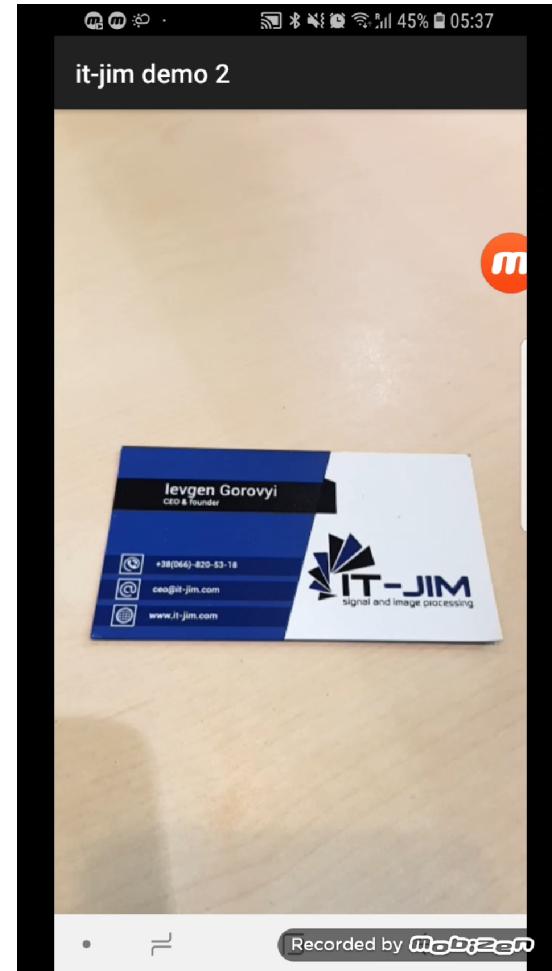
bundle_v2_example
  markers
  models
    0
    6
    10
    31
    32
    33
    34
    51
    52

```
{
  "version" : 1,
    "markerConfiguration" : {
    "marker_count" : 13,
    "markers_per_frame" : 1,
    "pyramid_filename" : "pyramid.bin",
    "database_filename" : "dbow.bin"
  },
```

AR It-Jim: Buttons demo

Buttons augmentation

Models augmentation

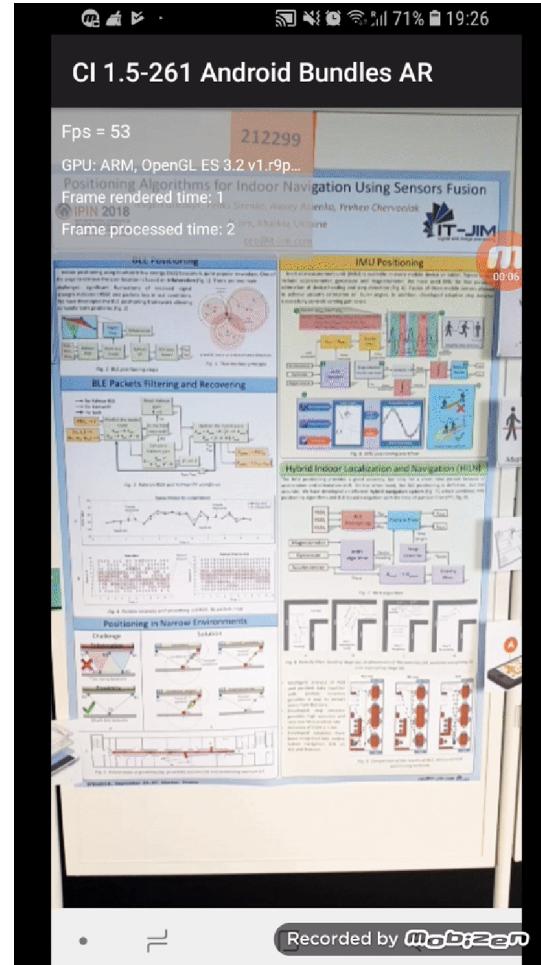Mobile AR in production

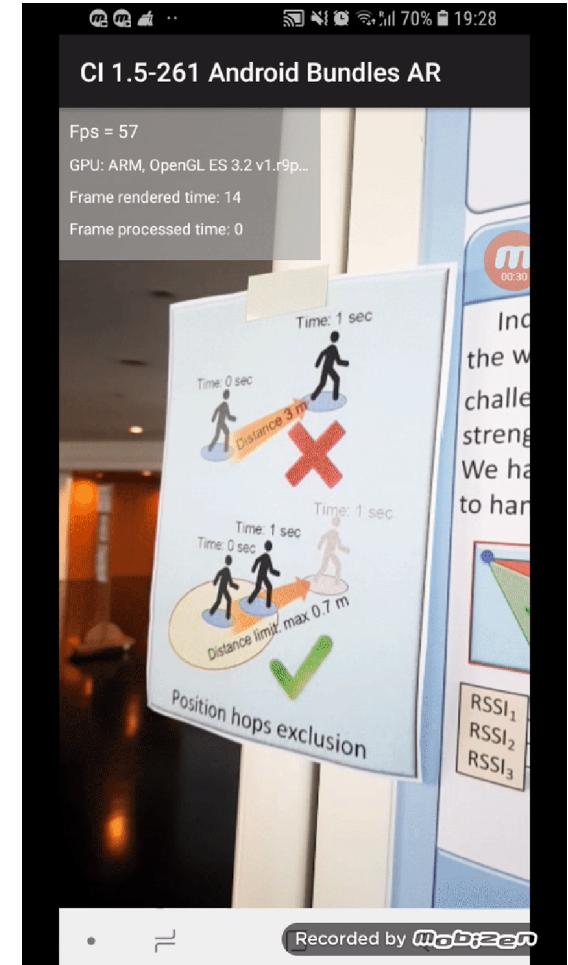# AR It-Jim: Cars Cube
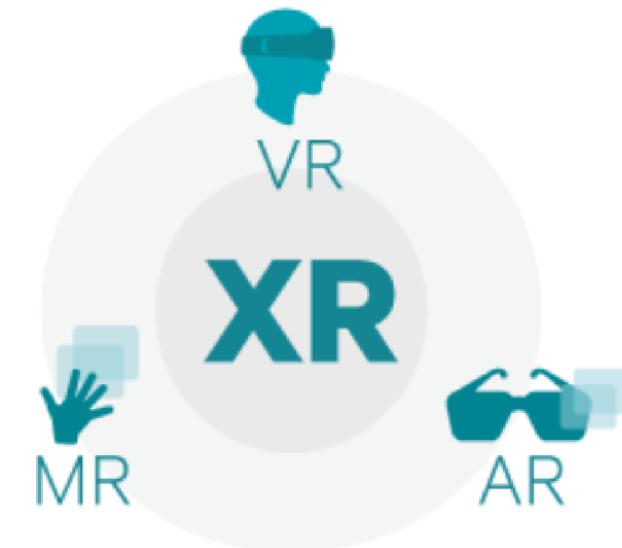

Screenshot

Cars cube example

AR Poster

IPIN 2018
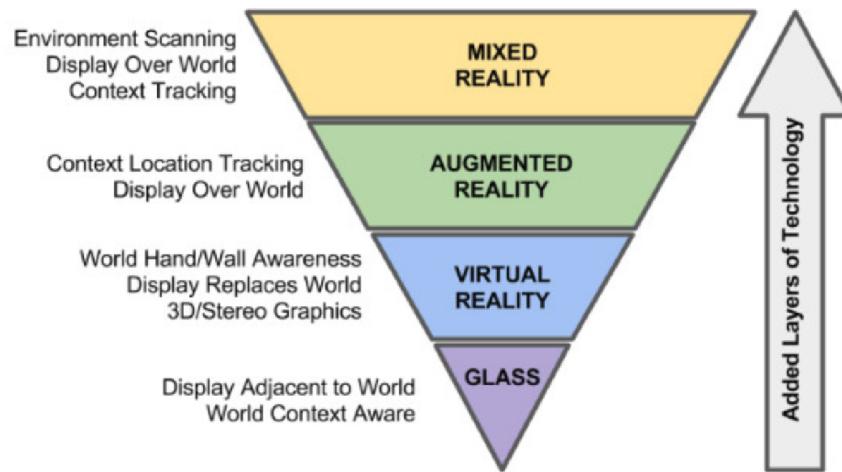
AR on poster

Toward XR

Robots vs Humans

Conclusions

- **Augmented reality is truly immersive field**

- **AR is not only about data, AR is about robust computer vision solutions**

- **Real-time performance on mobile is challenging**

- **AR is cool, but still not ready for a mass market**

THANK YOU!
Questions?

ceo@it-jim.com