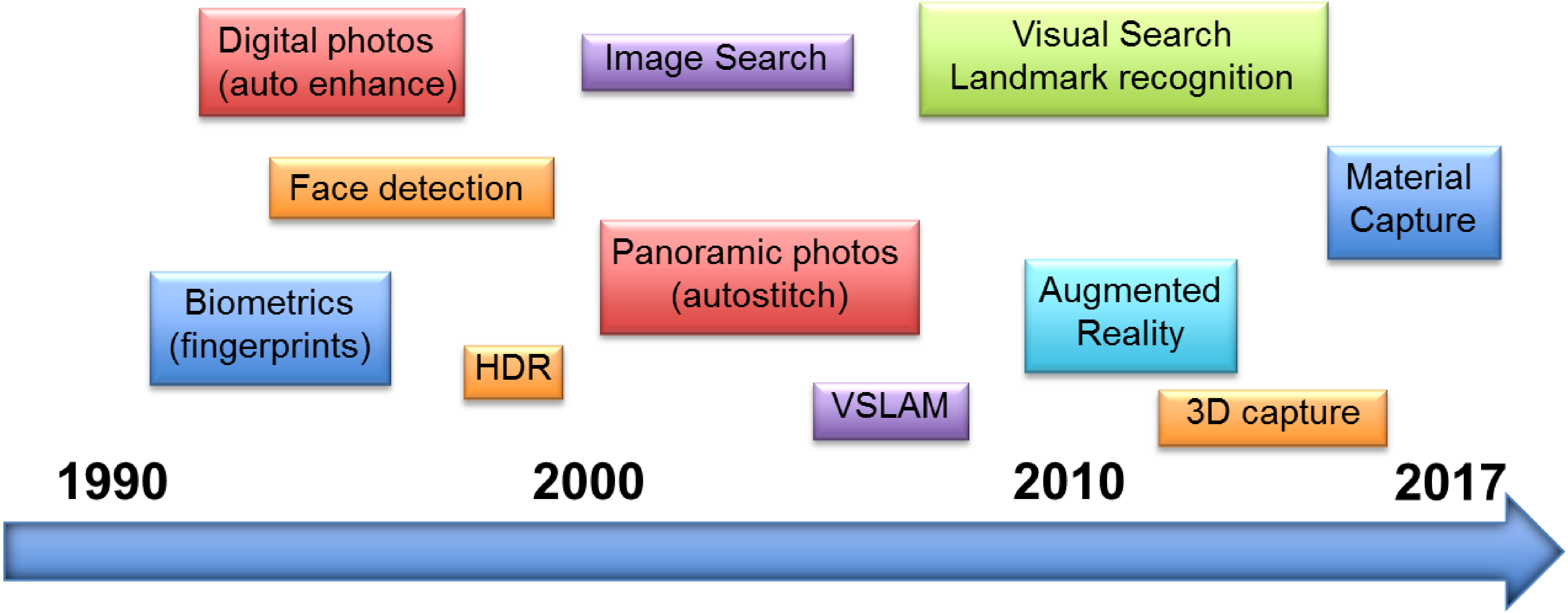
Visual Computing Group

Part 5.1

Mobile Metric Capture & Reconstruction: Introduction

Enrico Gobbetti, CRS4

Mobile applications: computer vision case

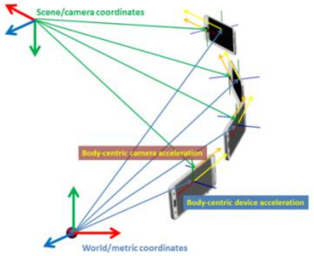


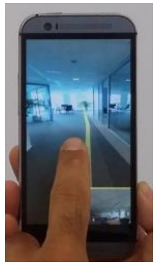
Mobile computer vision applications: trend

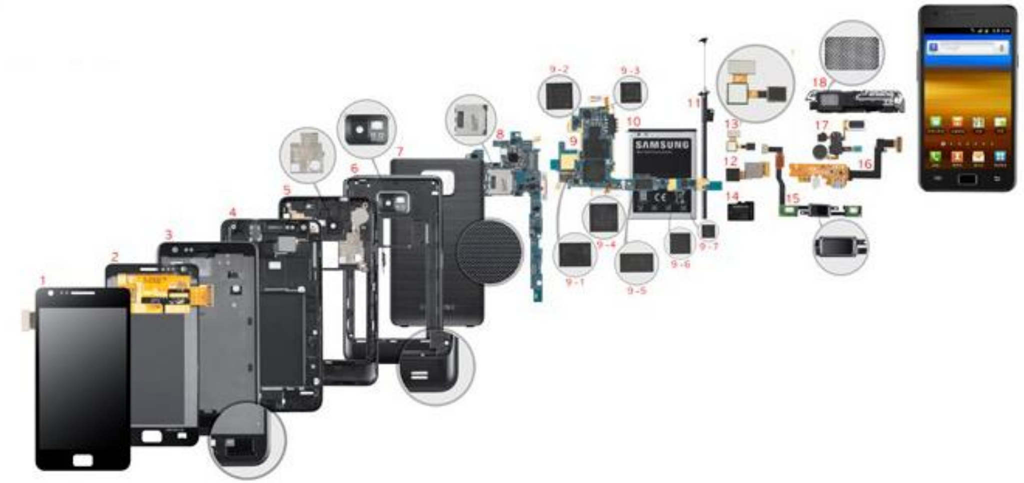
- **Mostly 2D**
  - Image enhancement
  - Image stitching
  - Image matching
  - Object detection
  - Texture classification
  - Activity recognition
  - …
- **Mostly 3D**
  - Camera localization
  - Pose estimation
  - 3D shape recovery
  - 3D scene reconstruction
  - Material/appearance recovery
  - Augmented reality
  - …

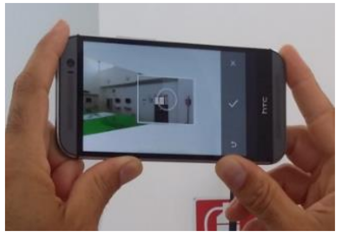
Now a mobile device
is identified by many specific features!

- **Features**
  1. Mobility
  2. Camera
  3. Non-visual sensors
  4. Processing power
  5. Connectivity
  6. Display
  7. Active light









Features (1/7): Mobility

- **Consumer/common tools**
  - Smartphones
  - Tablets

- **Embedded solutions**
  - Autonomous driving
  - Assistive technologies

- **Specific setups**
  - Drones
  - Robots

Features (1/7): Mobility

- **Consumer/common tools**
  - Smartphones
  - Tablets

- **Embedded s**
  - Autonomous
  - Assistive tec

On-site applications
Personal applications
Embedded systems

- **Specific setups**
  - Drones
  - Robots

Features (2/7): High-res/flexible camera

- **Impressive features**
  - High resolution and good color range (>12 MP, HDR)
  - Small sensors (similar to point and shoot cameras – approx. 1/3”) or even **double sensor**
  - High video resolution and frame rate (4K at 30fps)
- **Wide variety of field of views**
  - standard, fisheye, spherical
- **Specialized embedded cameras…**
  - Better lenses and sensors…
    - Modern SPC

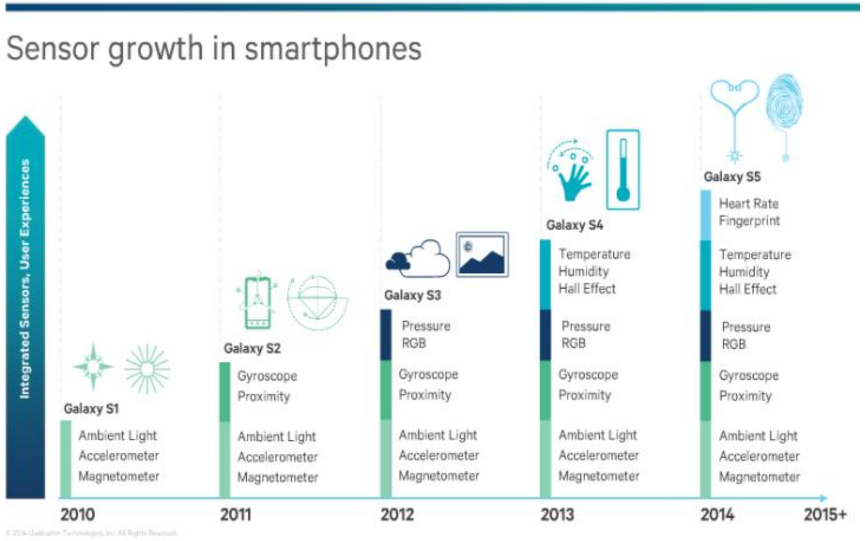
Features (2/7): High-res/flexible camera

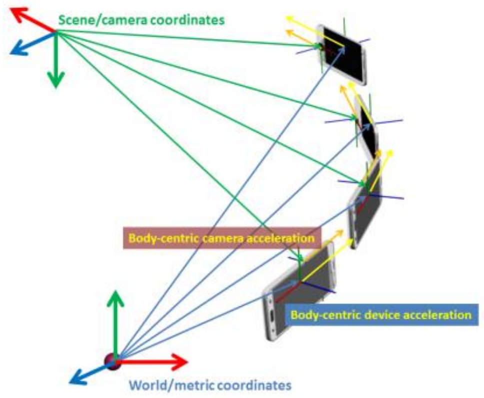
- **Impressive features**
  - High resolution and good color range (>12 MP, HDR)
  - Small sensors (similar to point and shoot cameras – approx. 1/3”) or even **double sensors**
  - High video resolution and frame rate (4K at 30fps)
- **Wide variety of f**
  - standard, fisheye
- **Specialized emb**
  - Better lenses and sensors…
    - Modern SPC

Computational photography
Visual capture
Augmented reality
Apps analyze/use snapshots or videos

Features (3/7): Non-visual sensors

- **Absolute reference instruments**
  - **GPS / A-GPS**
    - Mainly for outdoor applications
  - **Magnetometer**
    - Enable compass implementation
    - Often inaccurate for indoor
- **Relative reference instruments**
  - **Accelerometer**
    - Good metric information for small scale scene
    - Variable accuracy (sensitive to temperature)
  - **Gyroscope**
    - Very good accuracy for device relative orientation
- **Synced with camera!**

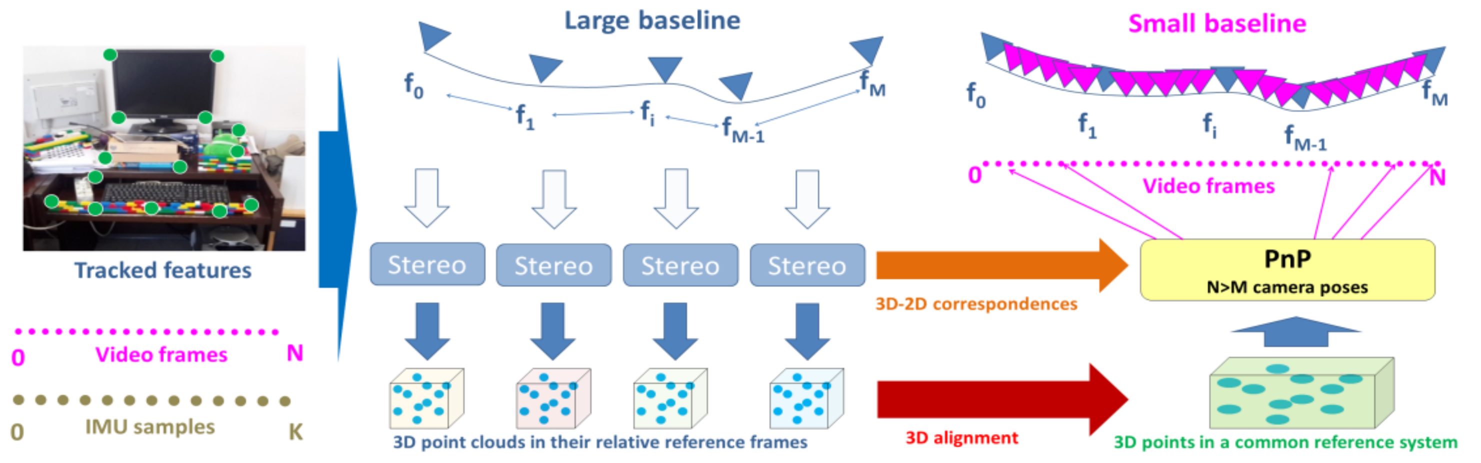


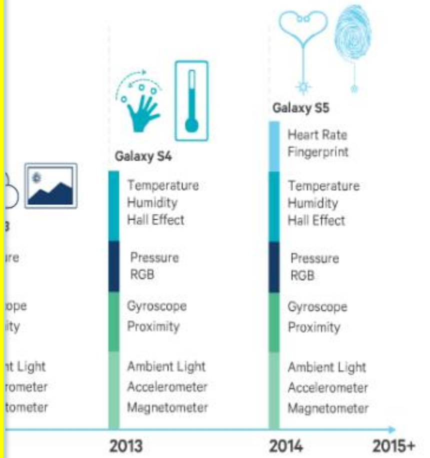


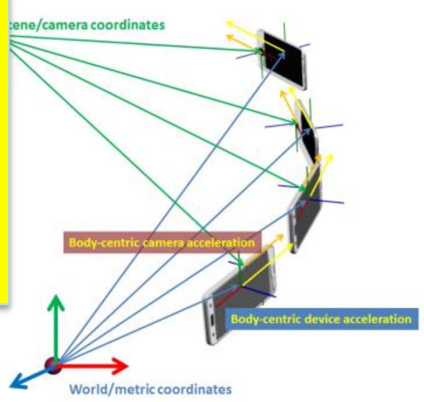
Features (3/7): Non-visual sensors

Data fusion!

Ex. Garro et al. **Fast Metric Acquisition with Mobile Devices.** VMV 2016



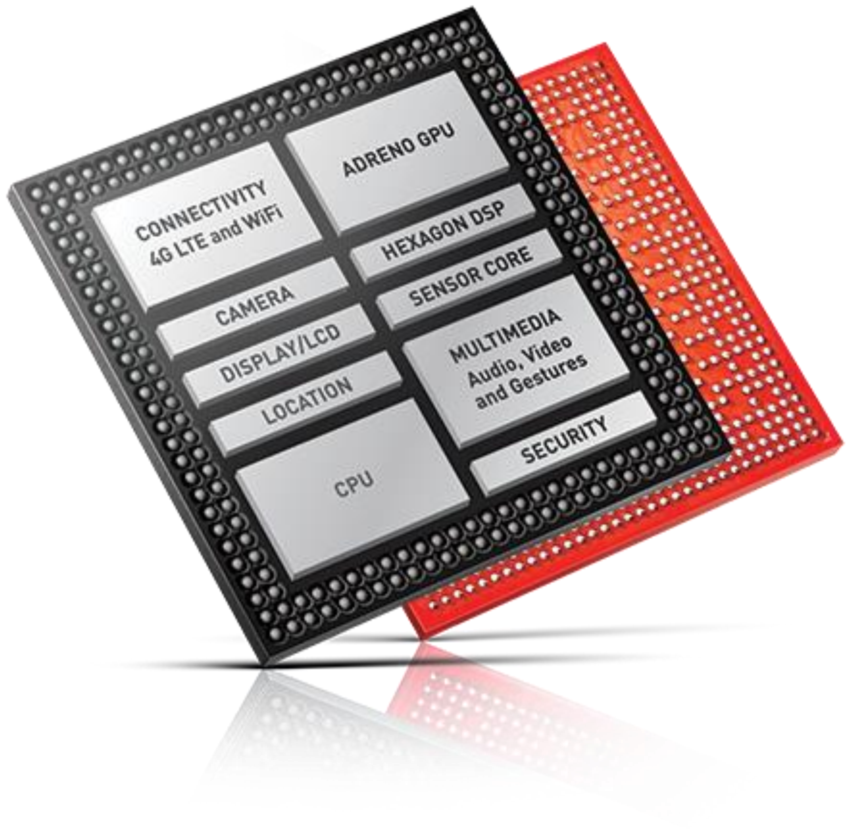




- Synced with camera!

Features (4/7): Processing power

- **Growing performance of mobile CPU+GPU**
  - (*see previous sections*)
- **Capable to run computer vision pipeline on mobile device**
  - i.e. *OpenCV* for Android
- **Main limitation: power consumption**



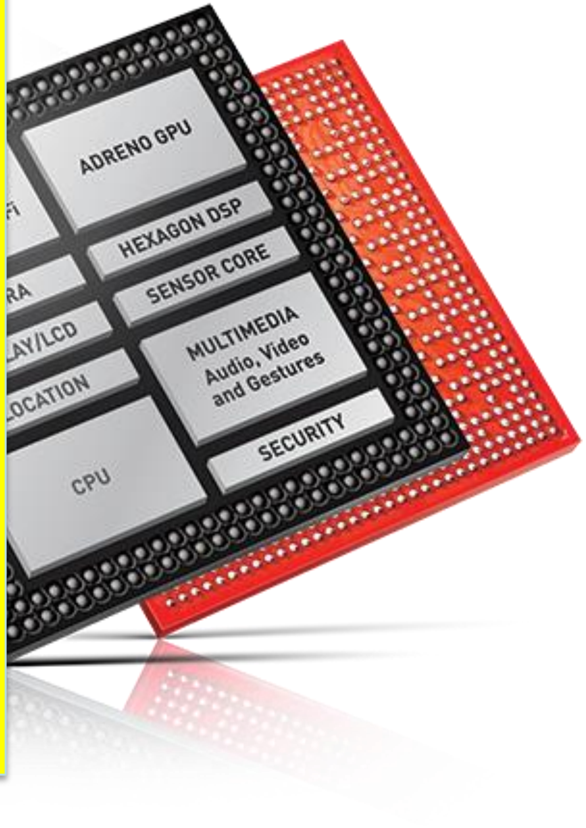
Features (4/7): Processing power

- **Growing perforr**
  CPU+GPU
  - *(see previous sec*
- **Capable to exec**
  pipeline on mob
  - i.e. *OpenCV* for *A*
- **Main limitation c**
  consumption

On-board pre-processing or even full processing

Ex. Tanskanen et al. **Live Metric 3D Reconstruction on Mobile Phones**. ICCV2013

.



Features (5/7): Connectivity

- **Many connectivity options**
  - **Local area**: NFC, Bluetooth, Bluetooth Low Energy, Wi-Fi 802.11x
  - **Wide area**: Cellular wireless networks: 3G/4G/5G

- **Mobile devices can connect at local or wide area at reasonable speed**
  - Typical LTE/4G: 18 Mbps down, 9.0 Mbps up
  - Typical Wi-Fi: 54Mbps (g), 300Mbps (n), 1Gbps (ac).

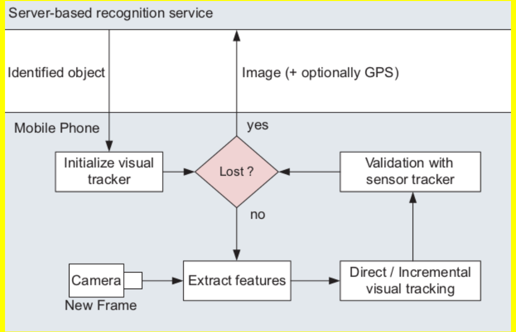
- **Lo-cost -> No-Costs**

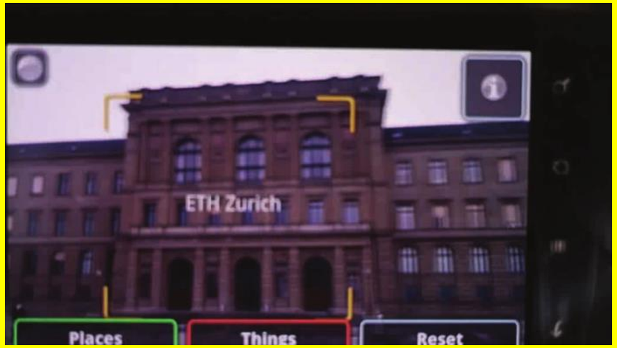
Features (5/7): Connectivity

- **Many connectiv**
  - **Local area**: NFC, Blu 802.11x
  - **Wide area**: Cellular w
- **Mobile devices c wide area at rea**
  - Typical LTE/4G: 18 M
  - Typical Wi-Fi: 54Mbp
- **Lo-cost -> No-Co**

Load balancing (client / server)
Access to large databases (e.g., search)
Communication

Ex. Gammeter et al. **Server-side object recognition and client-side object tracking for mobile augmented reality**. CVPRW 2010.



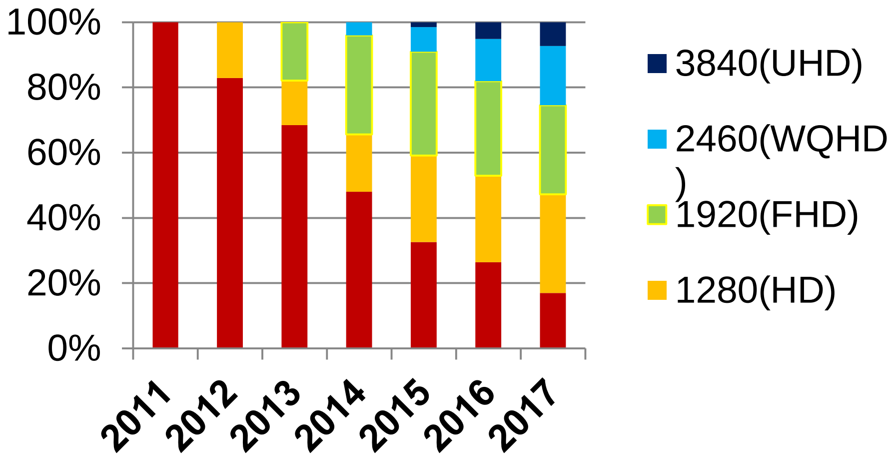


The internet on-line repositories

Computing nodes (back end) (cloud, server)

Features (6/7): Display!

- **Increasing display density**
  - Improved data presentation
  - Better touch-screen
- **Co-located with camera + other sensors**
  - Interactive capture
  - Interactive navigation



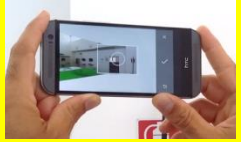
Data source: NPD DisplaySearch

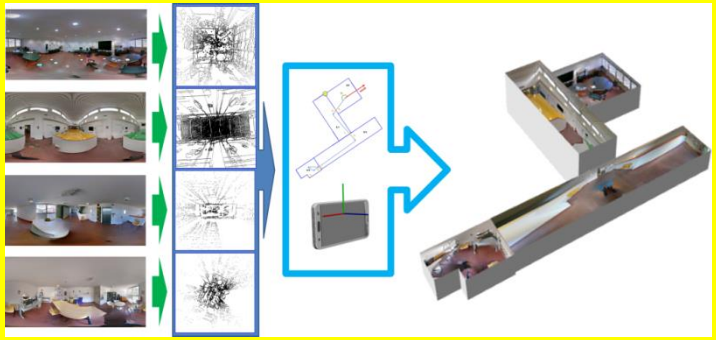
Features (6/7): Display!

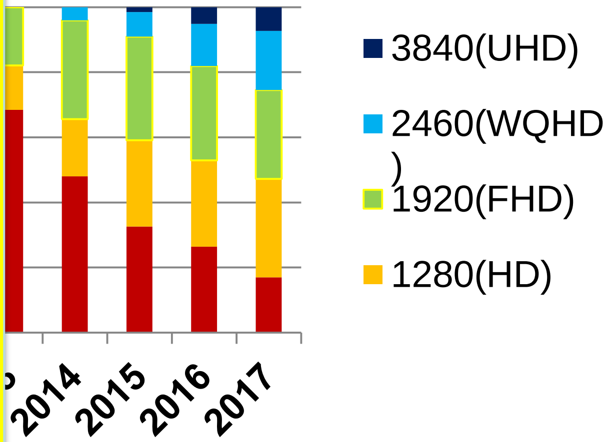
- **Increasing dis**
  - Improved data
  - Better touch-s
- **Co-located wi**
  - Interactive cap
  - Interactive nav

Data/result presentation
Guided capture / Augmentation

Ex. Pintore et al. **Mobile Mapping and Visualization of Indoor Structures to Simplify Scene Understanding and Location Awareness**. ECCV ACVR 2016



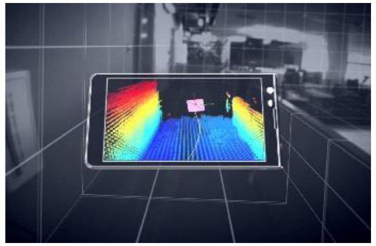


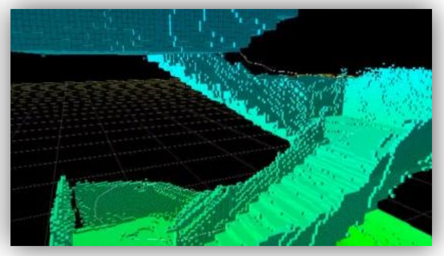


Data source: NPD DisplaySearch

Features (7/7): Active lighting

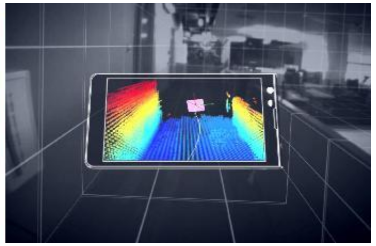
- **All smartphones have a flashlight**
  - LED source at fixed distance from camera
- **Can emulate custom (mobile) devices which have integrated emitters**
  - Google TANGO / Microsoft Kinect
    - Integrated depth sensor
- **Enables specialized capture procedures**

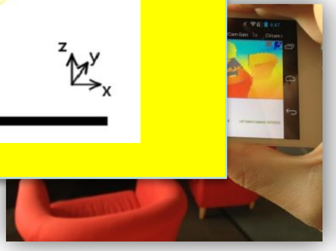


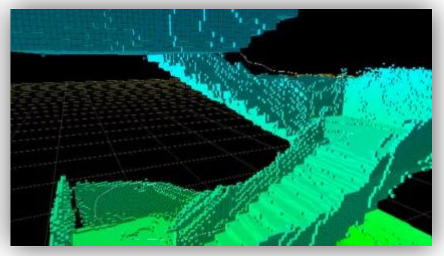


Features (7/7): Active lighting

- **All smartphones have a flashlight**
  - LED source at fix
- **Can emulate cus integrated emitt**
  - Google TANGO /
    - Integrated dept
- **Enables special**

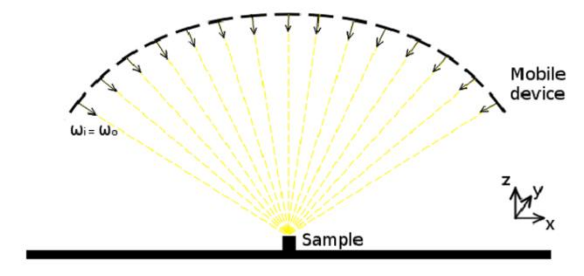






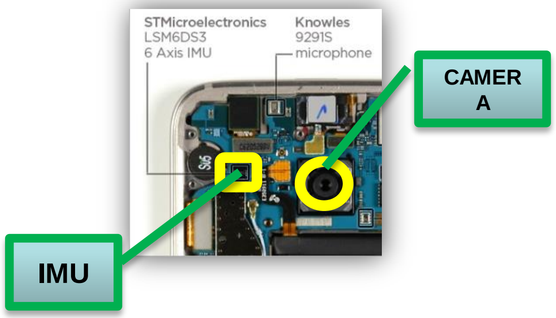
Material capture exploiting synchronization of illumination and visual sensing

Ex. Riviere et al. **Mobile surface reflectometry**. *Computer Graphics Forum*. 2015.



Wrap-up:
modern mobile features enable new applications

- **Features**
  1. **Mobility**
  2. **Camera**
  3. **Non-visual sensors**
  4. **Processing power**
  5. **Connectivity**
  6. **Display**
  7. **Active light**



- **Next: specific case studies exploiting modern mobile features**

Part 5.2

Mobile Metric Capture & Reconstruction: Case studies

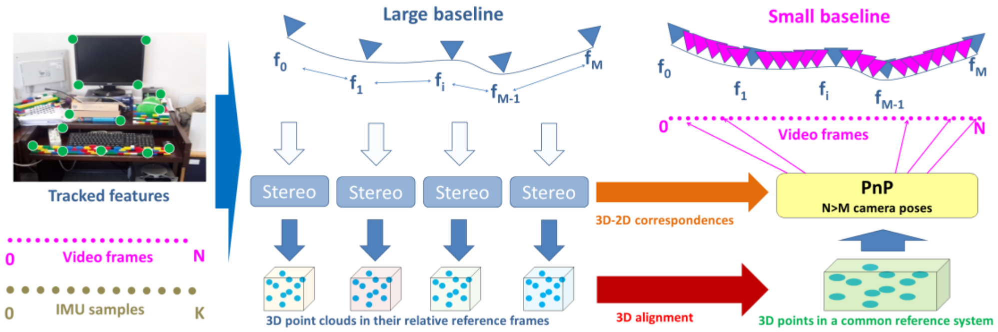
Giovanni Pintore, CRS4

Example 1

METRIC CAPTURE

Metric acquisition with a commodity mobile phone

- **Goal**
  – Capture 3D models with real-world measures
- **Mobile solution: data fusion**
  – Exploit synchronization of visual sensor & inertial sensors



Garro et al. **Fast Metric Acquisition with Mobile Devices.** VMV 2016

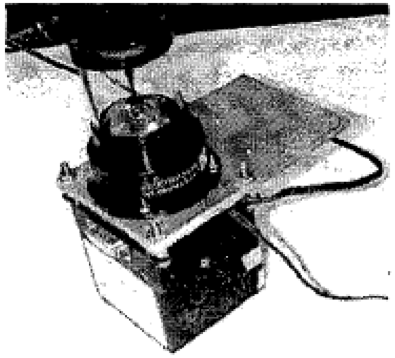
Visual sensor enables structure from motion methods

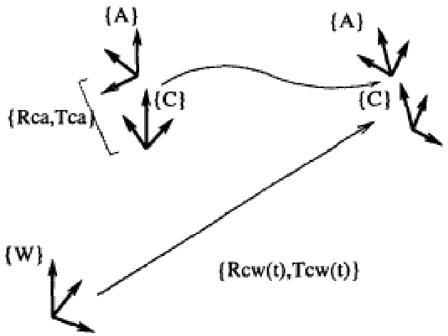
- **SfM reconstructs a point cloud from a series of images**
  - 3D positions of (sparse) matched features
  - Camera positions and orientations

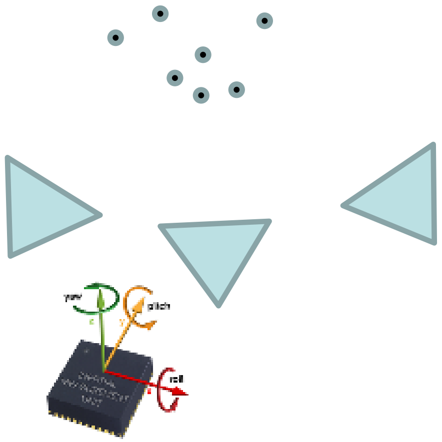
- **SCALE AMBIGUITY PROBLEM!**

Data fusion solution

- **Baseline idea**
  - Camera bundled with an IMU (inertial measurement unit)
  - Compare the camera trajectory recovered from **SfM** and the device motion detected by **inertial sensors**
  - Original **robotics** approach: *assumes **IMU more accurate than SfM***





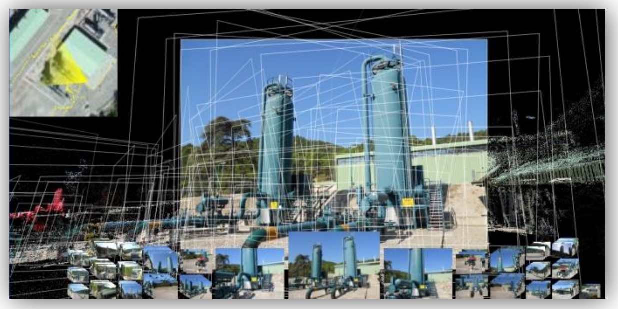


*Jung and Taylor. **Camera Trajectory Estimation using Inertial Sensor Measurements and Structure fom Motion Results.** CVPR 2001*

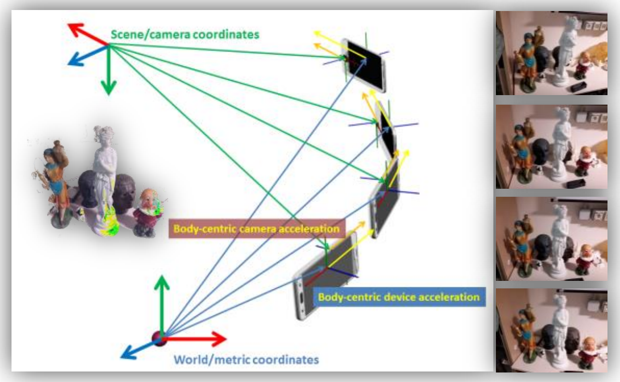
Mobile for metric acquisition

- **Outdoor**
  - **Visual+GPS (absolute reference)**

    Ex. ***Pintore et al.*** *3DNSITE: A networked interactive 3D visualization system to simplify location awareness in crisis management.* ***2012***



- **Indoor**
  - **Visual+IMU (relative reference)**
  - IMU returns relative linear accelerations in metric units



- **Mobile sensors generally less reliable than SfM information!**

First solution: comparing trajectories (1/2)

- **Straightforward solution:** to integrate the device trajectory from acceleration

$$\boldsymbol{x}(\boldsymbol{T1},\boldsymbol{T2}) = \left\| \int_{\boldsymbol{T1}}^{\boldsymbol{T2}} \left(\boldsymbol{v}(\boldsymbol{T1}) + \int_{\boldsymbol{T1}}^{\boldsymbol{t'}} \boldsymbol{a}(\boldsymbol{t})\, \boldsymbol{dt} \right) \boldsymbol{dt'} \right\|$$

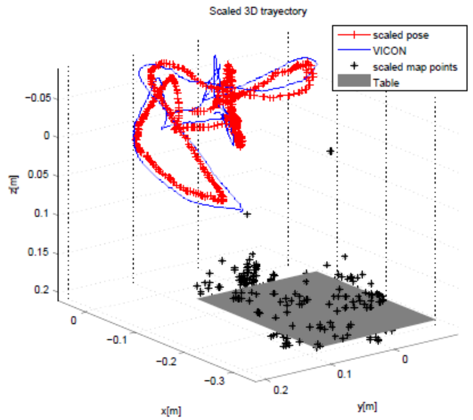
- Not so easy: onboard IMU sensors are noisy and SfM camera positions are sparse

First solution: comparing trajectories (2/2)

- *Example: Verlet* integration combined with a *Kalman* filter (Tanskanen et al.)
- Real-time comparison of **visual position** $\vec{x}_i$ and **integrated physical position** $\vec{y}_i$ to estimate the scale $\boldsymbol{\lambda}$

$$argmin = \sum_{i \in I} \|\vec{x}_i - \lambda\, \vec{y}_i\|^2$$

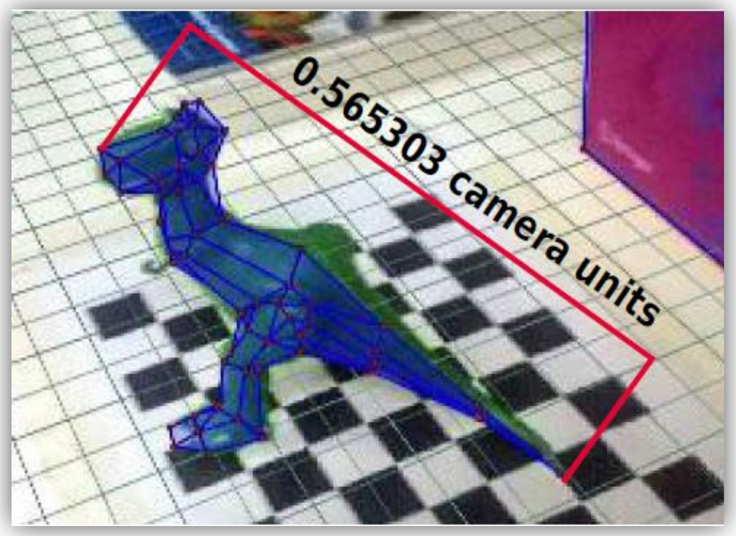
- Integration leads to a significant scale error: at its best 10% to 15%!



Tanskanen et al.
Live metric 3D reconstruction on Mobile Phones
ICCV2013

Second solution: comparing accelerations (1/2)

- **IMU acceleration compared to the instant camera acceleration**
  - Off-line approach
- **Camera acceleration recovered from the double derivative of the camera position**
- **Derivative operator leads to better accuracy than integration**



Ham et al. Hand-waving away scale. ECCV2014

Second solution: comparing accelerations (2/2)

- **Such SfM pipeline needs a large baseline**
  - **Downsample** (**D**) IMU samples at SfM frame rate
  - External pre-calibration needed $\mathbf{b}^\intercal$ **: position between camera and IMU**

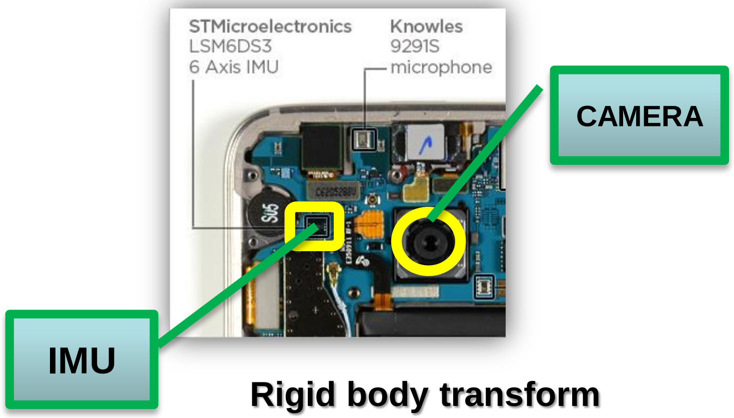
$$\underset{s,\mathbf{b}}{\arg\min}\, \eta\{s \cdot \hat{\mathbf{A}}_V + \mathbf{1} \otimes \mathbf{b}^\intercal - \mathbf{D}\mathbf{A}_I \mathbf{R}_I\}$$

- Requires very long acquisition times and pre-processing
- **Hard to be implemented on mobile systems**

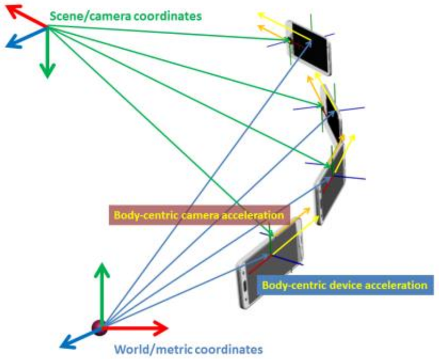
Proposed mobile solution (1/2)

- **Using robust fitting**

$$\underset{s,R}{\operatorname{argmin}}\{||A_c - sRA_s||\}$$

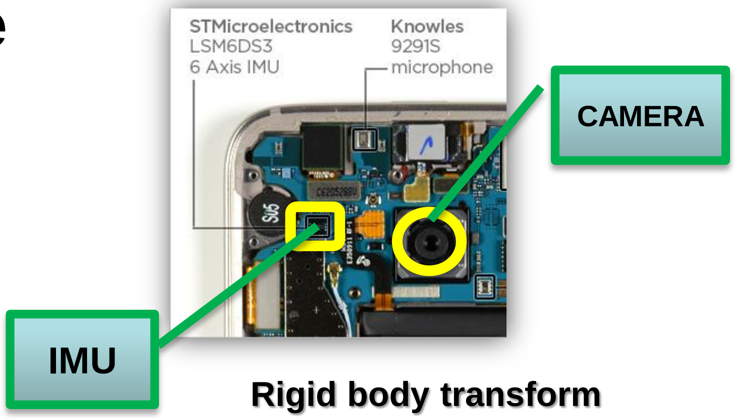


Rigid body transform

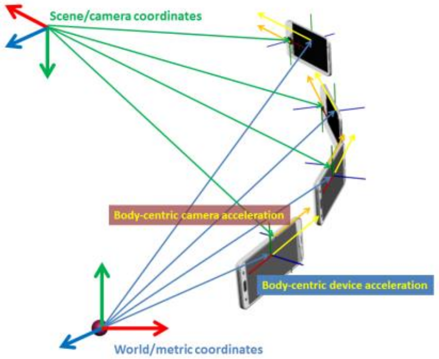


Proposed solution (2/2)

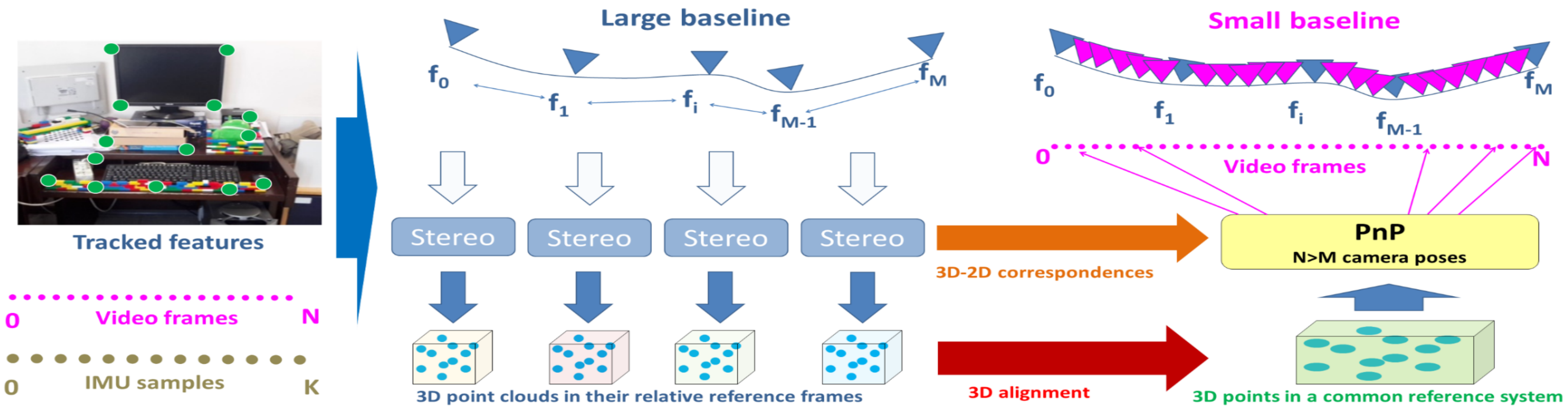
- **Match the acceleration samples at the IMU sample-rate**
  - Exploit the high and regular IMU sample-rate
- **Small SfM baseline required**
  - Video frames involved
  - **Need for a specific vision mobile pipeline**



Rigid body transform



Vision mobile Pipeline



Fast Metric Acquisition with Mobile Devices. [Garro et al. 2016]

- Features tracked along all frames
- Only few seconds needed to obtain metric measures

- Essential Matrix estimated when baseline is large enough

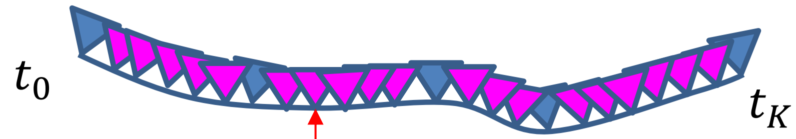
- Exploit global registration to estimate all cameras with Perspective-n-Point
- Returns densified track

Matching accelerations (1/2)

IMU accelerations

$$A_s = \begin{pmatrix} a_s^x(t_0) & a_s^y(t_0) & a_s^z(t_0) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ a_s^x(t_K) & a_s^y(t_K) & a_s^z(t_K) \end{pmatrix}$$

t_k



$$p_c''(t_k) = \frac{\Sigma_{i=0}^{8}(-1)^{(i+1)}\delta_i * p_c(t_{k+i-4})}{\Delta t^2}$$

Camera accelerations

$$A_c = \begin{pmatrix} p_c''(t_0)^T R_c(t_0) \\ \cdot \\ \cdot \\ p_c''(t_K)^T R_c(t_K) \end{pmatrix}$$

Problem to solve

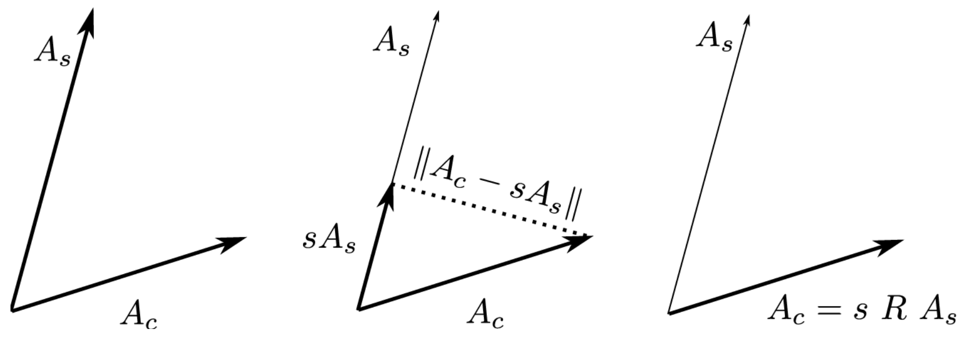
$$\underset{s}{\operatorname{argmin}}\{||A_c - sA_s||\}$$

Matching accelerations (2/2)

- **LS, gradient descent (et similia) poorly conditioned**
  - Not so many data
  - Severe outliers
- **Robust fitting using RANSAC approach**
  - Maximizes likelihood rather than just the number of inliers
- **Introduce rotation matrix R**
  - Account for orientation bias
  - Improve RANSAC performance
- **Fast, coping with large errors and noise**

$$\cancel{\underset{s}{\operatorname{argmin}}\{\|A_c - sA_s\|\}}$$

$$\underset{s,R}{\operatorname{argmin}}\{\|A_c - sRA_s\|\}$$



Results

- **Median error 4%**
  - 10-15% of other STAR solutions
- **Implementable on any mobile device**
  - IMU and video capture/stream required
  - i.e. **mobile spherical camera**!
- **Currently implemented for limited bounding volumes applications**

Scene Name		Real scale m / s.u.	Acquisition info Seconds	Poses	Samples	Our approach m / s.u.	Error	Simple scaling m / s.u.	Error
3D printer		2.094	17.0	65	883	2.01	4.0%	2.85	36.1%
Scanner setup		3.565	9.8	53	641	3.45	3.1%	3.12	12.4%
Desktop		6.520	11.3	48	596	6.24	4.2%	5.16	20.8%
Statuettes		2.602	11.5	53	607	2.49	4.5%	2.48	4.9%
Office desk		1.977	30.4	88	471	2.01	1.8%	2.01	1.8%
Office workstation		3.95	12.3	37	1307	3.94	0.3%	3.98	0.6%
Ara pacis		1.568	30.07	77	1569	1.52	2.8%	1.80	13.0%
Workstation (Fastest)		0.707	9.9	34	1305	0.73	2.7%	0.89	20.4%
Desk fast motion		6.918	14.8	74	1718	6.28	9.1%	3.88	44.0%

Example 2

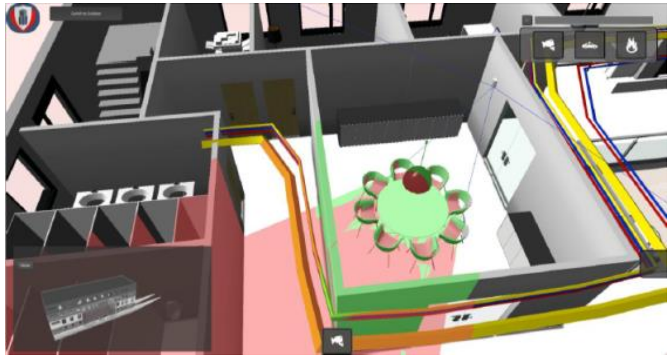
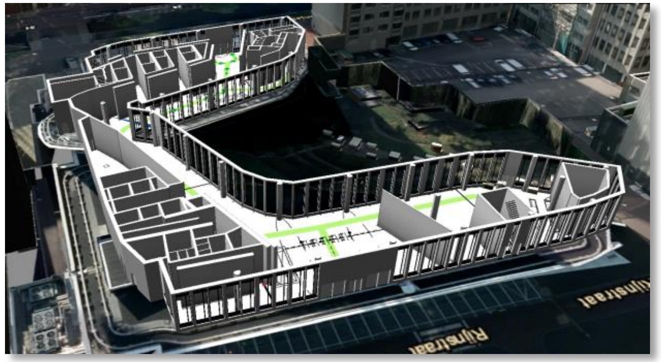
INDOOR CAPTURE AND INTERACTIVE VISUALIZATION

Indoor capture + interactive visualization

- **Creation and sharing of indoor digital mock-ups**
  - Exploiting the capabilities of modern mobile devices

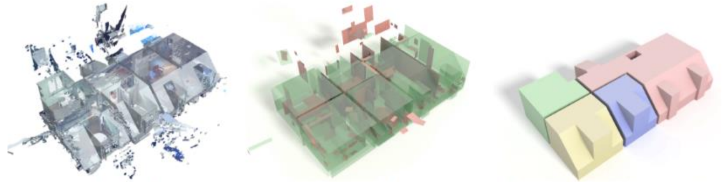
Motivations

- **Strong interest in many domains**
  - Security, smart houses design, simulations
  - …or in general when available digital models:
    - don’t represent the actual layout
    - don’t include a photorealistic representation

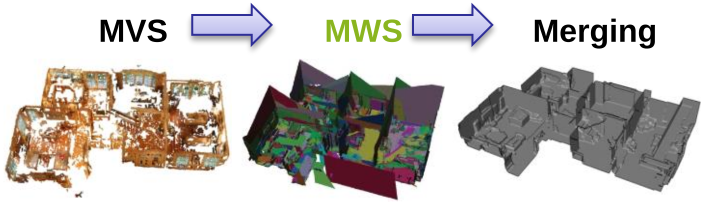
TECHNICAL CONTEXT

- **Professional solutions to create indoor models**
  - **Manual modeling**
  - **Semi-automatic methods based on high-density data**
    - **Laser scanning**
      - Professional but expensive, limited to specific applications
    - **Multi-view stereo from photographs**
      - Generally cost effective but hard to apply in the indoor environment
        - » Walls poorly textured, occlusions, clutter
        - » Furthermore: need for heavy MW constraints, computationally demanding



Input point clouds — Detection of permanent components — Volumetric reconstruction of rooms

Mura et al. **Piecewise-planar Reconstruction of Multi-room Interiors with Arbitrary Wall Arrangements.** Computer Graphics Forum – Pacific Graphics 2016



Furukawa et al. **Reconstructing Building Interiors from Images.** ICCV 2009

TECHNICAL CONTEXT

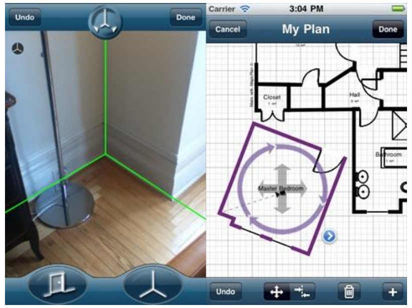
- **Common critical point of the mentioned solutions**
  - Not for anyone: require specific equipment and high professional skills
  - Considerable effort to produce structured models!

- **Growing interest in using mobile devices to simplify capture and reconstruction**
  - Wide diffusion and easiness of use
  - Increasing support (Google TANGO, Facebook 360)
  - *Example: crime scene acquisition*
    - *Usually done through laser scanner, many photographs: scene corruption!*
    - *New procedures: a preliminary and less invasive acquisition with few spherical images*

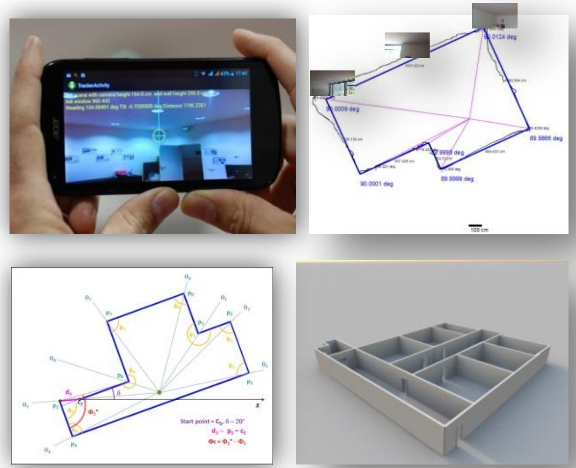
Mobile solutions: interactive indoor capture

- **MagicPlan - *http://www.sensopia.com***
  - **Floor corners marked via an augmented reality interface**
  - ***Limits:***
    - Intensive manual editing for the room and to assemble the floor plan
- **Sensors fusion methods**
  - Pintore et al. **Interactive mapping of indoor building structures through mobile devices.** In Proc. 3DV Workshop on 3D Computer Vision in the Built Environment, December Tokyo, 2014
  - Pintore et al. **Effective Mobile Mapping of Multi-room Indoor Structures.** The Visual Computer, 30(6--8): 707-716, 2014
  - **Rooms shapes recovered by merging device orientation measures and associated video frames information**

- **Both approaches focused only on the geometry**
  - **No visual representation stored!**
  - ***How to simultaneously capture the geometry and the appearance of an indoor environment?***



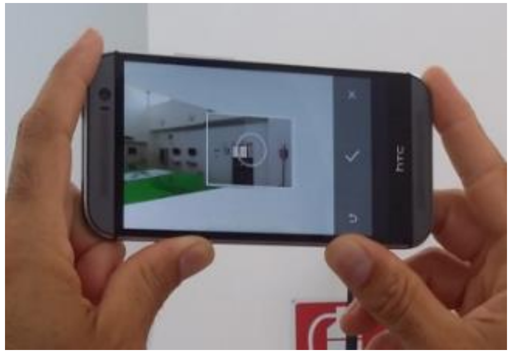
MagicPlan



Pintore et al. **Effective mobile mapping of multi-room indoor structures** The Visual Computer,2014

Solution: exploiting panoramic/360 images

- **Contain more information than perspective images**
- **360 images are easy to capture using common devices**
  - Interactive apps using IMU + GUI + automatic stitching
  - Dedicated cameras
- **360 images are easy to navigate**
  - Spheremaps + emerging image and video formats
  - VR devices for immersion
- **What about analyzing them?**

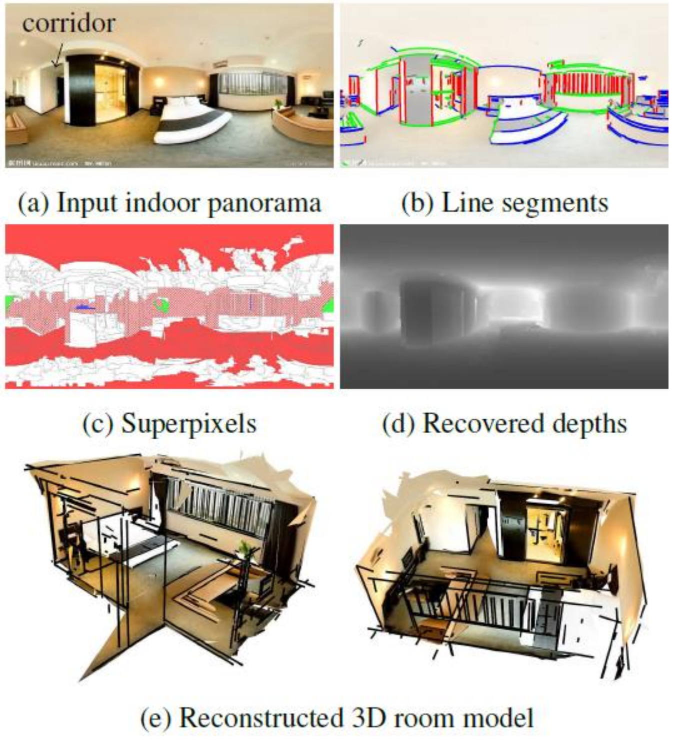


State-of-the-art approaches

- **Current SoA adopt one spherical image per room**

- **Minimize user interaction**
  - Compliant with popular navigation paradigms
  - Ready for immersive VR devices

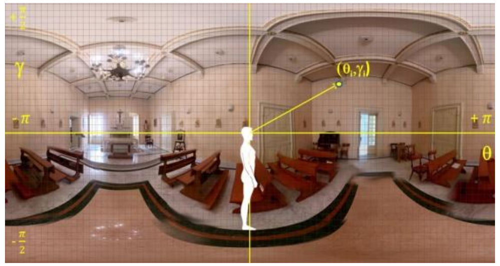
- **Example**
  - Yang et al.: indoor scene recovered from oriented super-pixel facets
    - Graph cut returning best planes
    - **Computationally demanding**
    - **Limited to single room environment**



Yang et al.
Efficient 3D Room Shape Recovery From a Single Panorama.
CVPR 2016

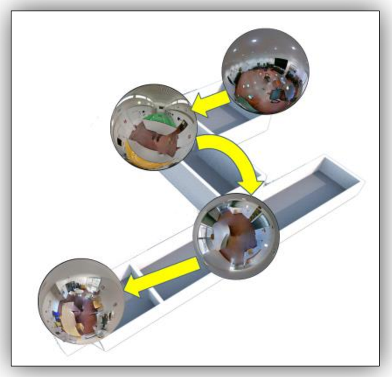
Proposed mobile solution

- **Capture setting**
  - **One equirectangular image per room generated by a mobile device**
    - Vertical lines in the image are aligned with the gravity vector
  - **Tracking of the user movement between adjacent rooms**
    - Just the movements direction during door crossing



- **Single room model**
  - **Space enclosed by vertical walls and an horizontal floor**
    - Reasonable model for almost all civil building types
      - » Enables simplified labeling: **ceiling**, **walls**, **floor**

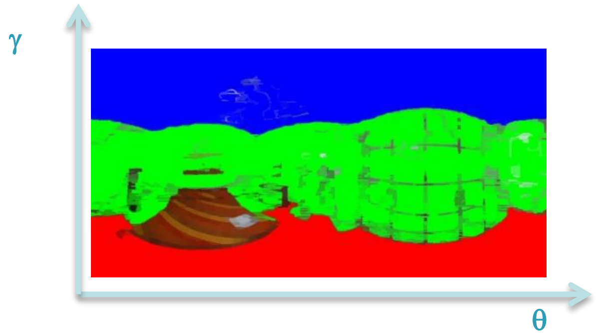
- **Multi-room model**
  - **Rooms connected by doors**



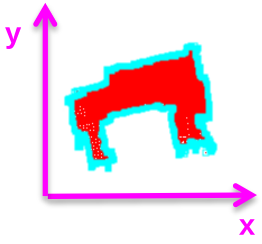
Pintore et al. ***Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps.*** *IEEE WACV 2016*

Analyzing spheremap to extract room structure (1/3)

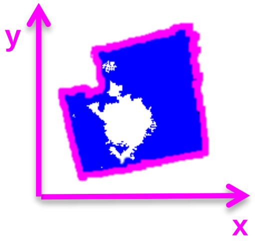
- **Super pixels labeling**
  - Wall-ceiling and wall-floor **edges**
- **Spatial transform**
  - 3D points from spherical coordinates γ and θ
  - Valid if the height ***h*** is known: i.e. : on the **edges** of the horizontal planes
- **Projected contours highlight the room shape!**
- **Actually only the ceiling edge projection defines the room shape**
  - Floor edge is often occluded by furniture, etc.



$$G_h(\theta,\gamma) = \begin{cases} x = h/\tan\gamma * \cos\theta \\ y = h/\tan\gamma * \sin\theta \\ z = h \end{cases} \qquad h = \begin{cases} -h_e & \textit{floor} \\ h_w - h_e & \textit{ceiling} \end{cases}$$



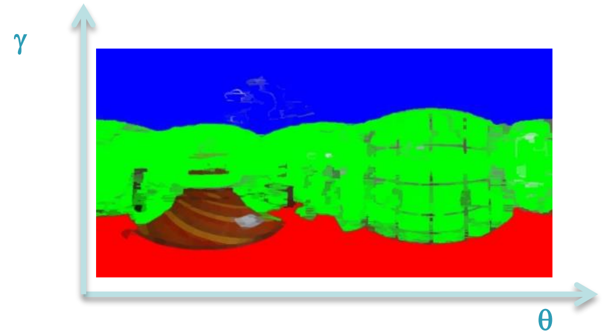
2D floor projection



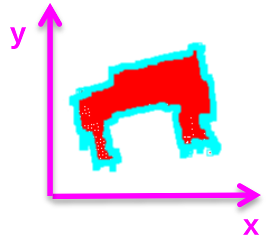
2D ceiling projection

Analyzing spheremap to extract room structure (2/3)

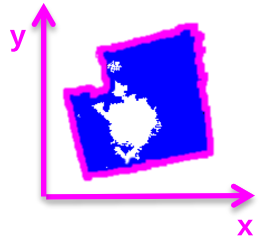
- **Height estimation**
  - According with our model ***h*** can assume only two values:
    - **$-h_e$** for the floor edge
    - **$h_w - h_e$** for the ceiling edge
  - **The height with respect to the floor is assumed fixed and known**
    - *If **h_e** is given in metric dimension, all the model results scaled in real-world dimensions*
  - **The distance from the ceiling is the only unknown (depends by h_w)**



$$G_h(\theta,\gamma) = \begin{cases} x = h/\tan\gamma * \cos\theta \\ y = h/\tan\gamma * \sin\theta \\ z = h \end{cases} \qquad h = \begin{cases} -h_e & \mathit{floor} \\ h_w - h_e & \mathit{ceiling} \end{cases}$$



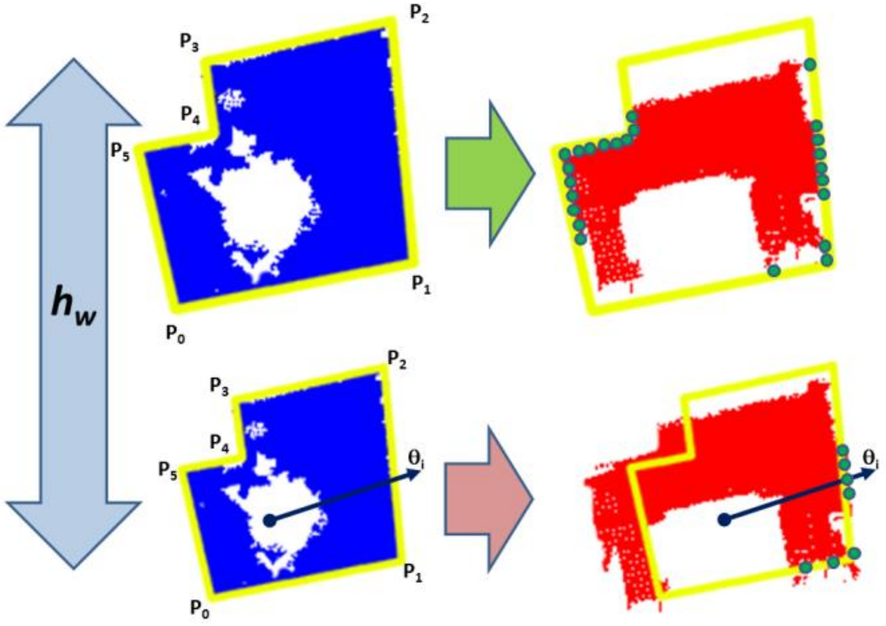
2D floor projection



2D ceiling projection

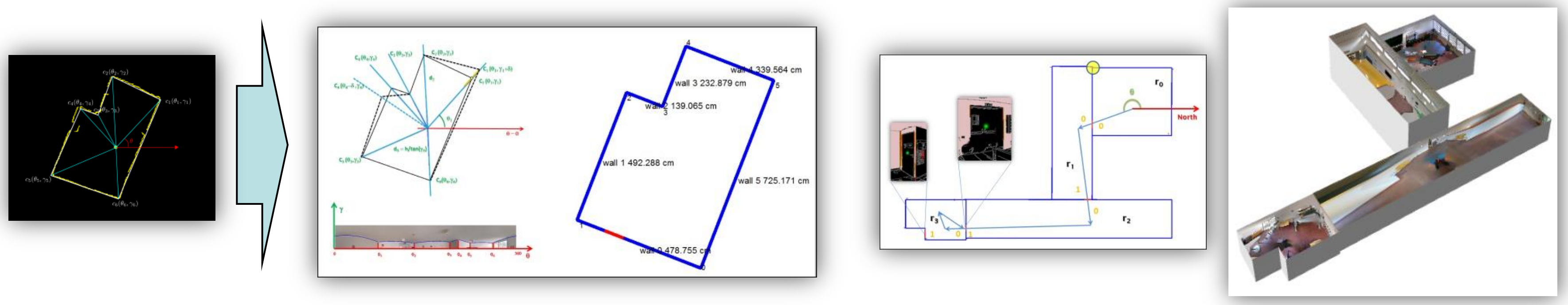
Analyzing spheremap to extract room structure (3/3)

- **h_w works as a scale factor for the ceiling 2D contour**
- **We search for the *h_w which* maximizes the ceiling-floor matches count**
  - If ***h_w*** is the real wall height the XY coordinates of the ceiling and floor edges should be the same



Finding the multi-rooms structure

- **Rooms assembly**
  - Doors position identification in the image by computer vision
  - Doors matching according with capture graph
  - Final rooms displacement



Pintore et al. **Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps.** IEEE WACV 2016

Results

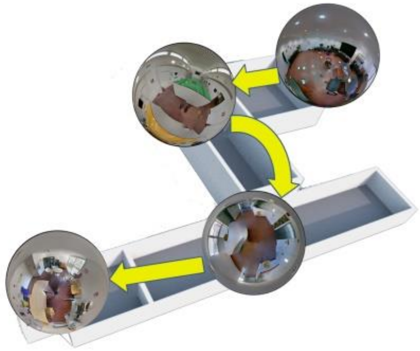
Scene Name	Features Area $[m^2]$	Features Np	Area error MP	Area error Ours	Wall length error MP	Wall length error Ours	Wall height error MP	Wall height error Ours	Corner angle error MP	Corner angle error Ours	Editing time MagicPlan
Office H1	720	10	2.95%	1.78%	35 cm	15 cm	2.0 cm	1.2 cm	0.8 deg	0.8 deg	26m32s
Building B2	875	25	2.50%	1.54%	30 cm	7 cm	6.0 cm	1.5 cm	1.5 deg	1.5 deg	42m18s
Commercial	220	6	2.30%	1.82%	25 cm	8 cm	12.0 cm	2.7 cm	1.5 deg	1.0 deg	28m05s
Palace	183	3	16.86%	0.20%	94 cm	5 cm	45.0 cm	1.3 cm	1.8 deg	0.5 deg	15m08s
House 1	55	5	21.48%	2.10%	120 cm	16 cm	15.0 cm	4.7 cm	13.7 deg	1.2 deg	25m48s
House 2	64	7	28.05%	1.67%	85 cm	8 cm	18.0 cm	3.5 cm	15.0 deg	0.5 deg	32m25s
House 3	170	8	25.10%	2.06%	115 cm	15 cm	20.0 cm	4.0 cm	18.0 deg	1.5 deg	29m12s

Pintore et al. **Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps.** IEEE WACV 2016

Reasonable, fast reconstruction with structure and visual features

Sharing and interactive exploration of the indoor model

- **Visual model stored on a server**
  - Exploration graph
    - Each node is a spheremap/room
    - edges (yellow) are transitions between adjacent rooms
- **Client-side interactive exploration**
  - Room
    - **WebGL fragment shader**
    - dragging to change view orientation
  - Passages
    - **Real-time rendering** of the transitions between rooms
    - Suggested paths
  - Exploiting network connection
    - Low bandwidth required thanks to real-time rendering



Some results

Live demo: http://vcg.isti.cnr.it/vasco/
Click on the dataset on the left column to start

3D reconstruction of a 655 mq office with 19 rooms. This environment was acquired with a mobile phone (HTC One M8)

Reconstruction of a 70 rooms floor of the NHV ministry at Den Haag, Netherlands. The whole model was acquired with a Ricoh Theta S camera

Next session:

CLOSING/Q&A