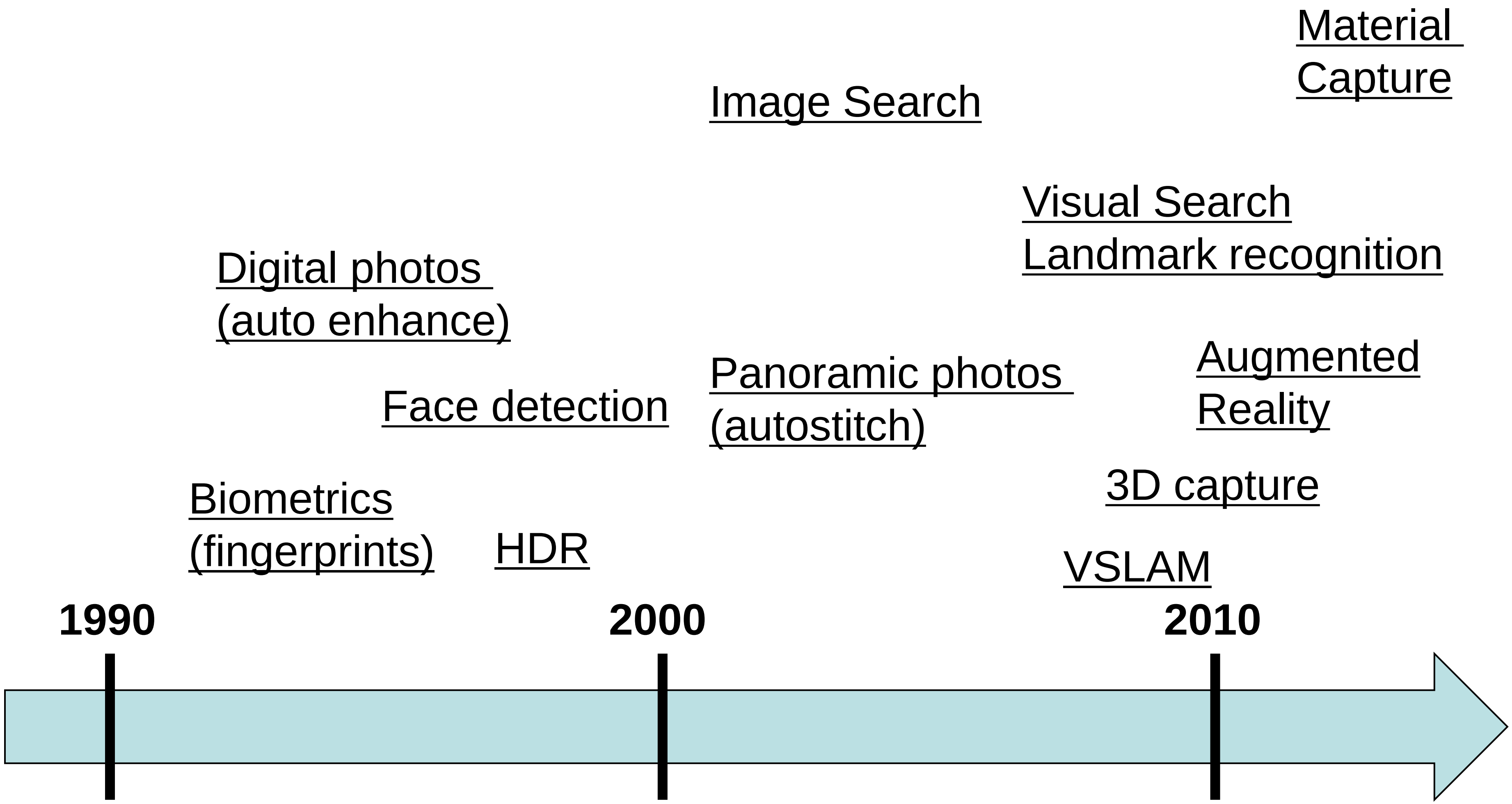
Visual Computing Group

Part 4

Mobile metric capture and reconstruction

Computer vision and mobile applications

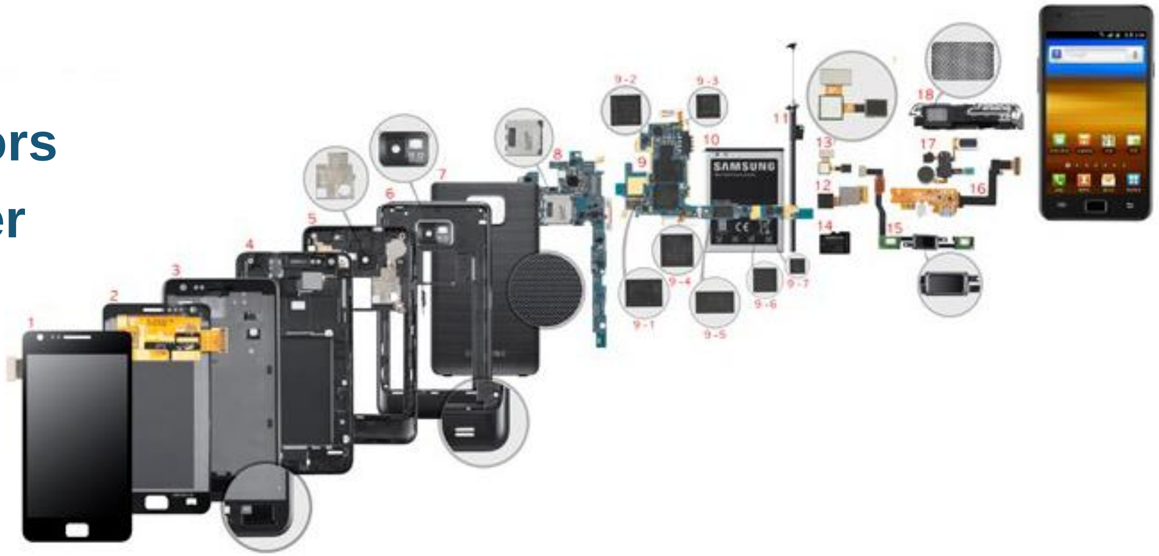


Computer vision and mobile applications

- **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
  - …

Applications made possible by specific features of mobile devices!

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



Features (1/7): Mobility

- **Consumer**
  - Smartphones
  - Tablets


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


- **Specific**
  - Drones
  - Robots


Features (1/7): Mobility

- **Consumer**
  - Smartphones
  - Tablets

- **Embe**
  - Auto
  - Assi

On-site applications / Personal applications / Motion and/or location taken into account / Embedded solutions

- **Specific**
  - Drones
  - Robots

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

- **Common features**
  - High resolution and good color range (>12 MP, HDR)
  - Small sensors (similar to point and shoot cameras – approx. 1/3”)
  - 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…

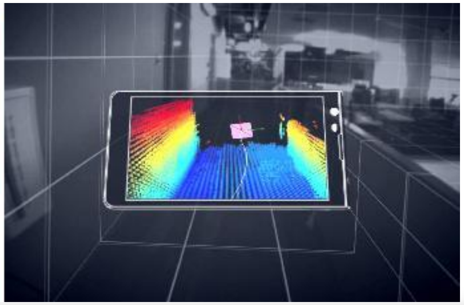
Features (2/7): High-res/flexible camera

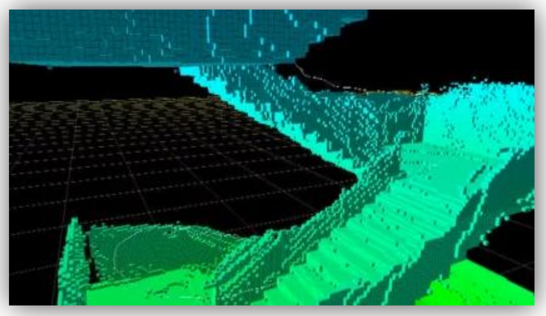
- **Common features**
  - High resolution and good color range (>12 MP, HDR)
  - Small sensors (similar to point and shoot cameras – approx. 1/3”)
  - High
    (4K
- **Wide v**
  - standard, fisheye, spherical
- **Specialized embedded cameras…**
  - Better lenses and sensors…

Visual channel is the primary one
Computational photography
Apps analyze/use snapshots or videos

Features (3/7): Active lighting

- **All smartphones have a flashlight**
  - LED source at fixed distance from camera
- **Custom devices have integrated emitters**
  - Google TANGO / Microsoft Kinect
    - Integrated depth sensor
- **Leads to specialized capture procedures**



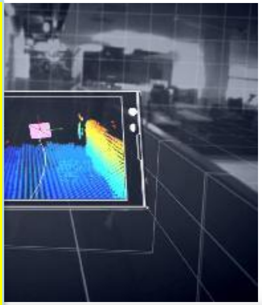


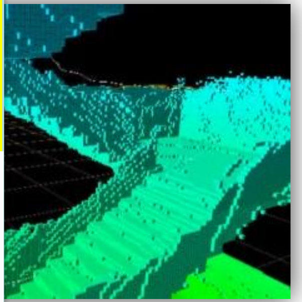
Features (3/7): Active lighting

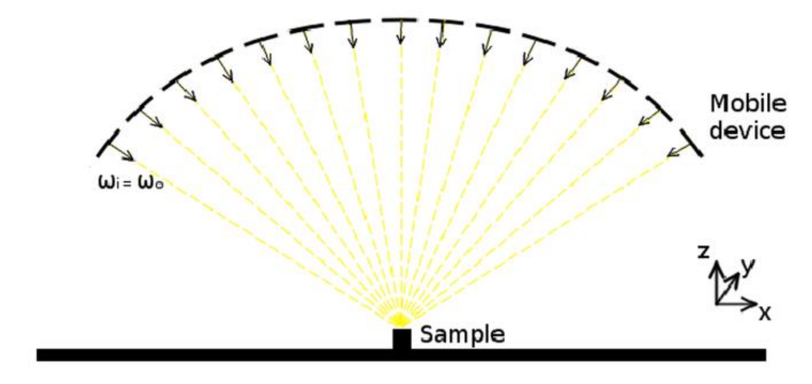
- **All sm**
  flashli
  - LED
    from
- **Custo**
  integr
  - Goo
    Kine
    - I
- **Leads**
  capture procedures

Specialized capture procedures exploiting synchronization of illumination and visual sensing

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

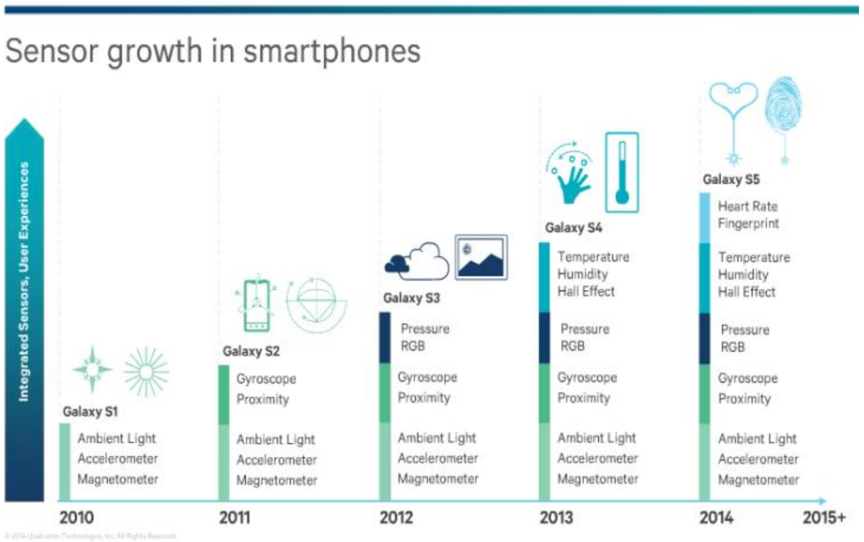


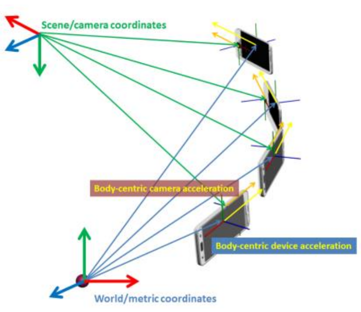




Features (4/7): Non-visual sensors

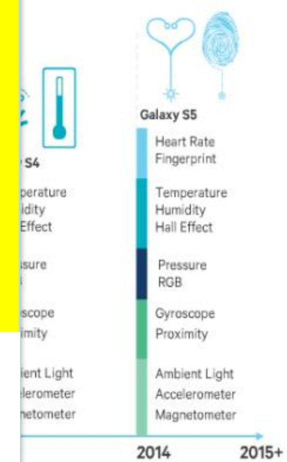
- **Absolute reference**
  - **GPS / A-GPS**
    - Mainly for outdoor applications
  - **Magnetometer**
    - Enable compass implementation
    - Often inaccurate for indoor
- **Relative reference**
  - **Accelerometer**
    - Variable accuracy (sensitive to temperature)
    - Good metric information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





Features (4/7): Non-visual sensors

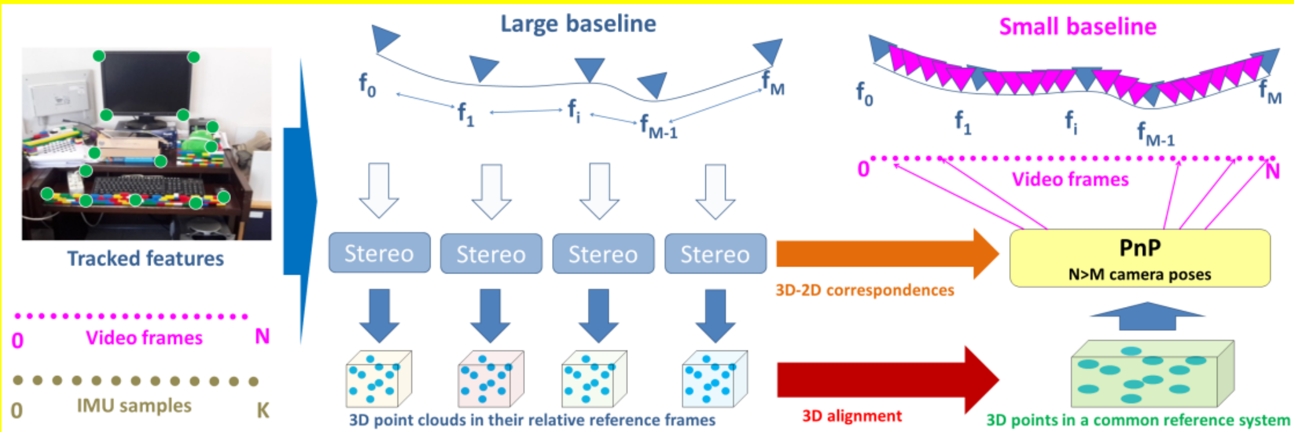
- **Absol**
  - **GPS**
  - **Mag**
- **Relati**
  - **Acc**
  - **Gyr**
- **Synced with camera!**

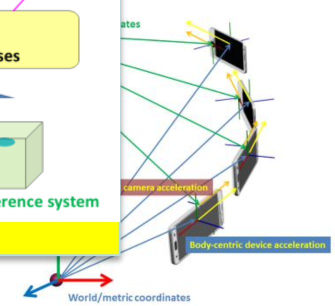


Data fusion!

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

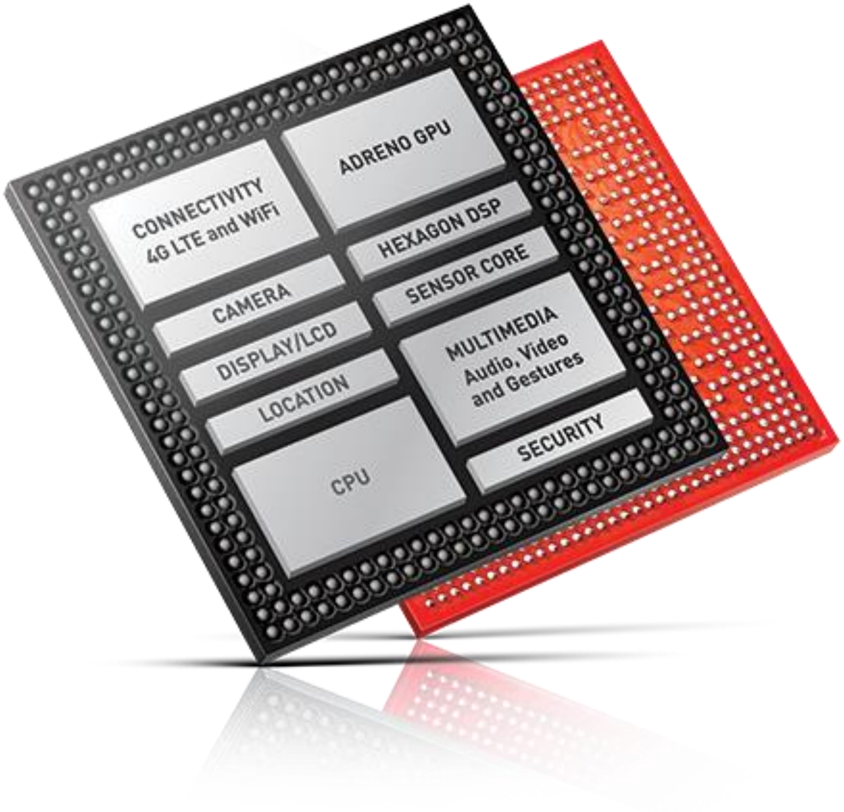
.





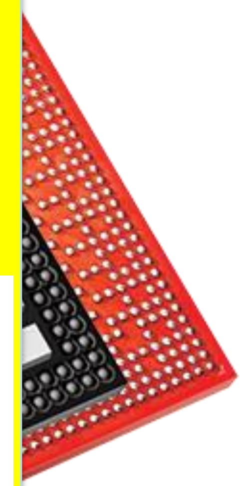
Features (5/7): Processing power

- **Growing performance of mobile CPU+GPU**
  - *(see previous sections)*
- **Capable to execture computer vision pipeline on mobile device**
  - i.e. *OpenCV* for Android
- **Some limitations due to power consumption**



Features (5/7): Processing power

- **Growi**
  mobil
  - *(see*
- **Capab**
  comp
  on mo
  - i.e. (
- **Some**
  powe



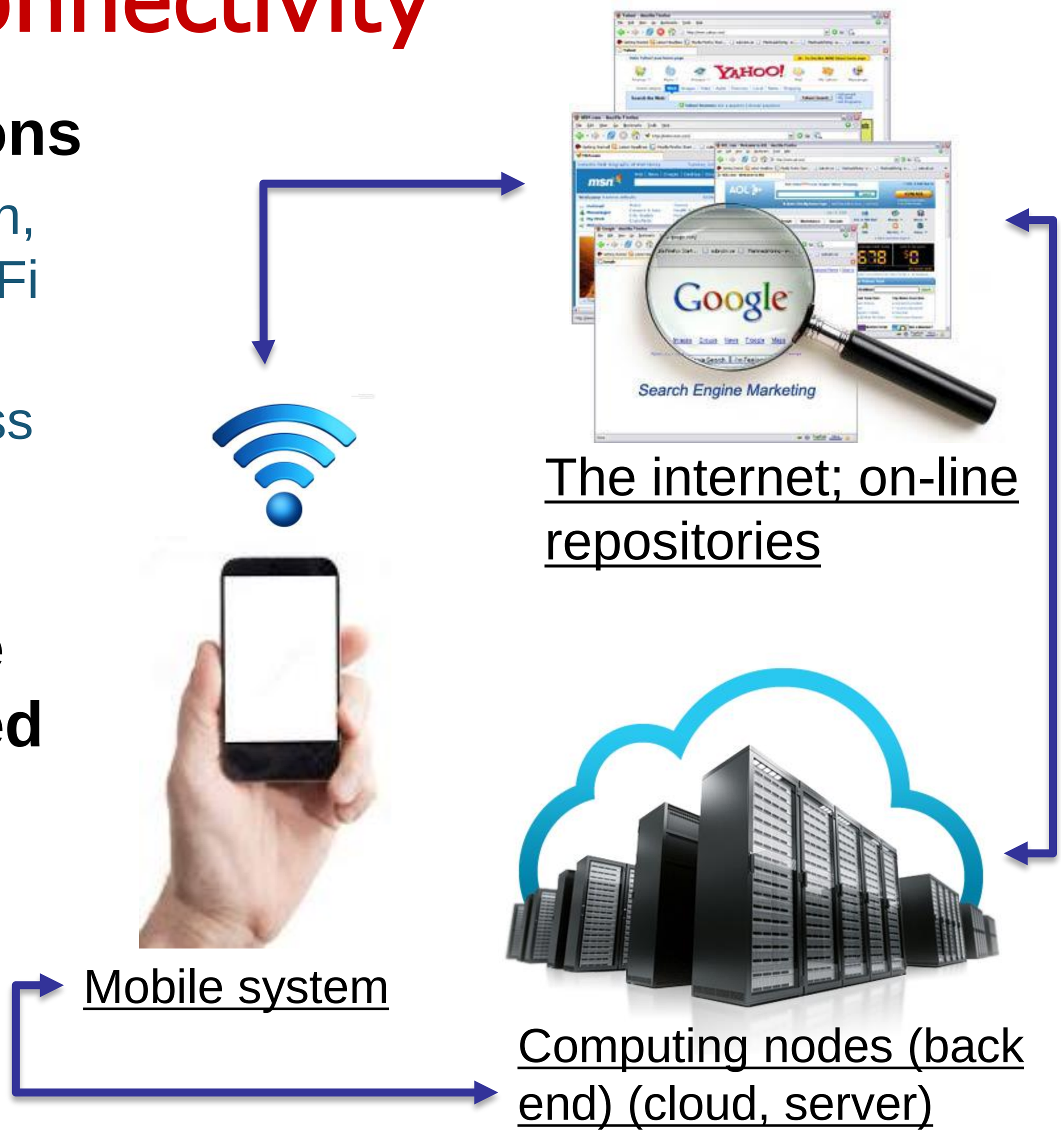
On-board pre-processing or even full processing

<u>Ex.</u> Tanskanen et al. **Live Metric 3D Reconstruction on Mobile Phones**. ICCV2013

.

Features (6/7): 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**

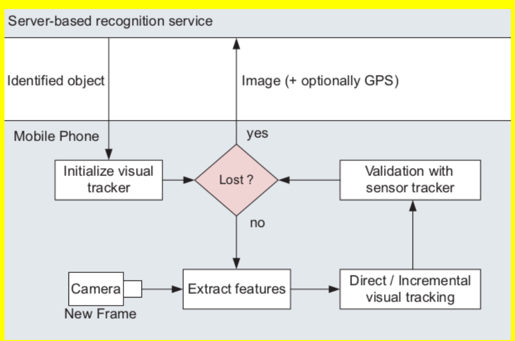


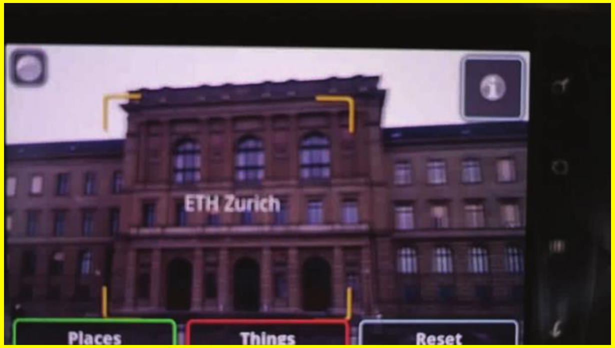
Features (6/7): Connectivity

- **Many**
  - **Loc**
    Blue
    802
  - **Wid**
    netw
- **Mobil**
  conne
  area a
  - Typ
    dow
  - Typ
    300Mbps (n), 1Gbps (ac).
- **Lo-cost -> No-Costs**

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.





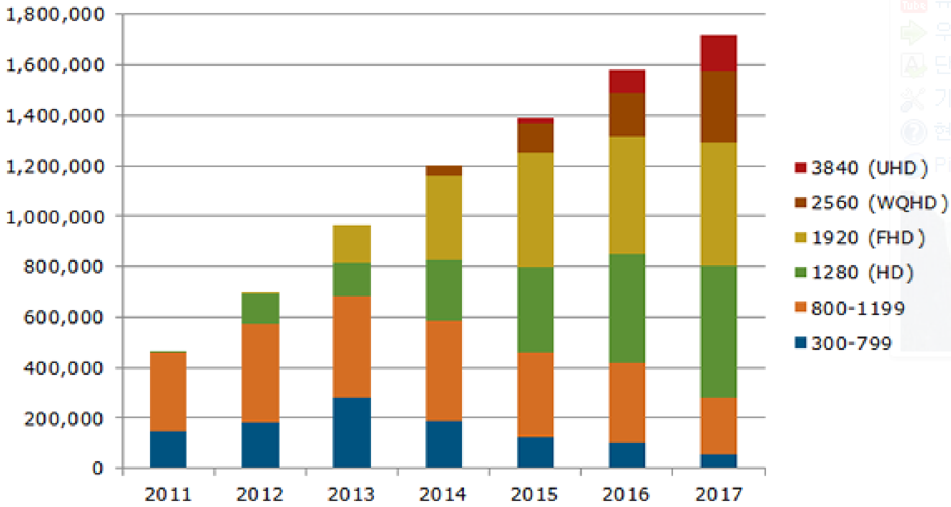
net; on-line
es

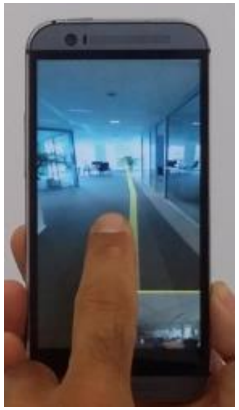
Mobile system

Computing nodes (back end) (cloud, server)

Features (7/7): Display!

- **Hi-res/hi-density display**
  - Data presentation!
- **Co-located with camera + other sensors**
  - Tracking during capture!
- **Touch screen**
  - Co-located user-interface
  - (UI also may exploits other sensors)



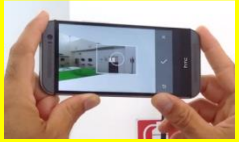


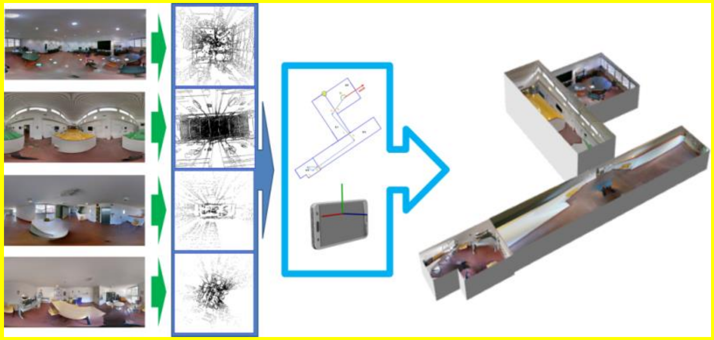
Features (7/7): Display!

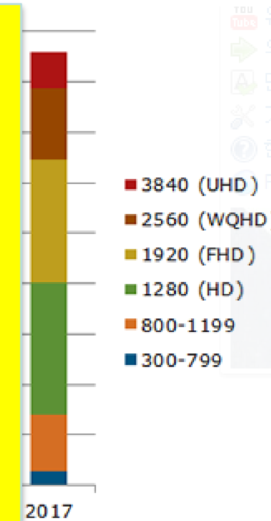
- **Hi-res**
  - Data
- **Co-loc other**
  - Trac
- **Touch**
  - Co-l
  - (UI a sens

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

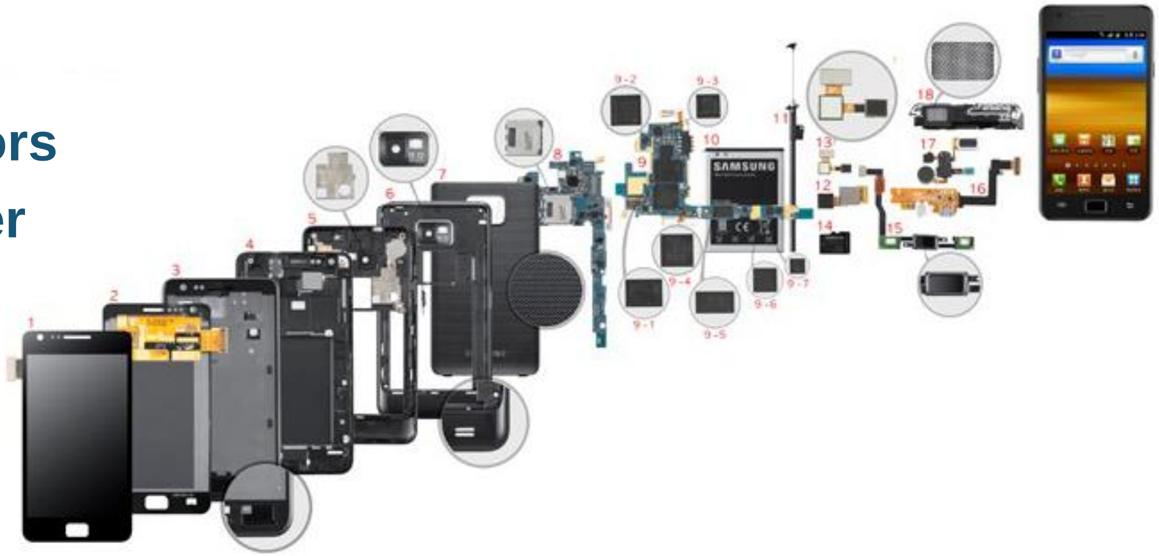






Wrap-up: mobile apps characterized by the exploitation of mobile device features

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

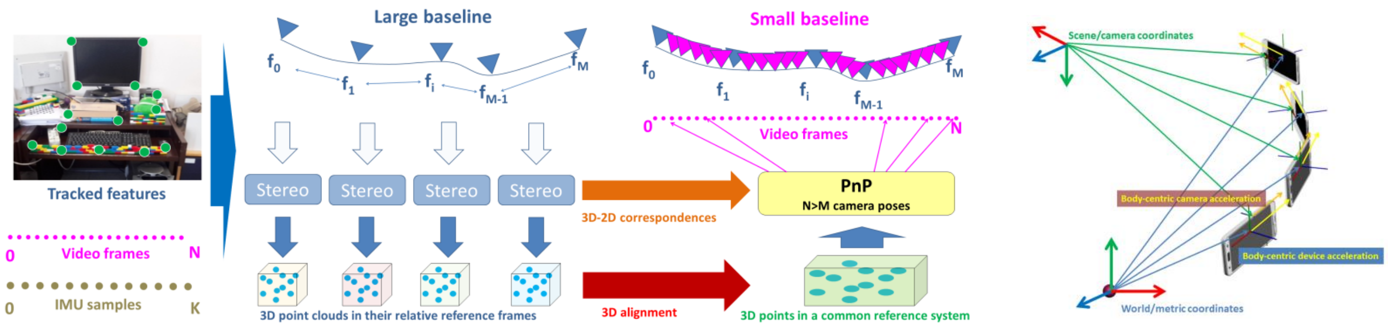


Example 1

DATA FUSION FOR METRIC CAPTURE

Metric acquisition with a commodity mobile phone

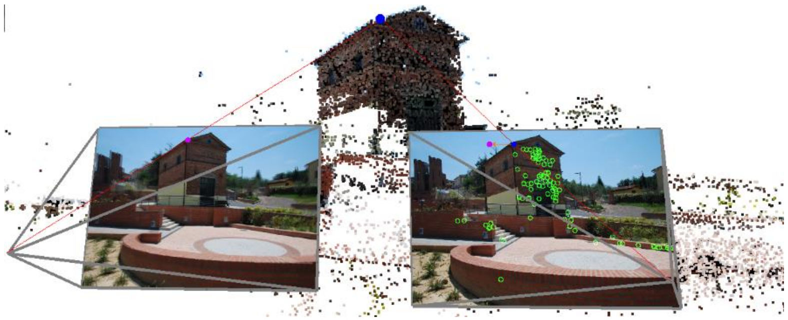
- **Goal**
  - Capture 3D models with real-world measures
- **Data fusion approach**
  - Exploit synchronization of visual sensor & IMU to capture scenes in real-world units

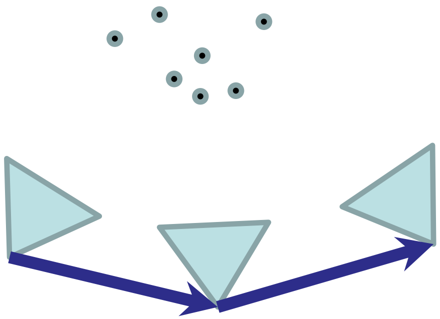


Garro et al. **Fast Metric Acquisition with Mobile Devices.** VMV 2016

Structure-from-Motion + Dense reconstruction

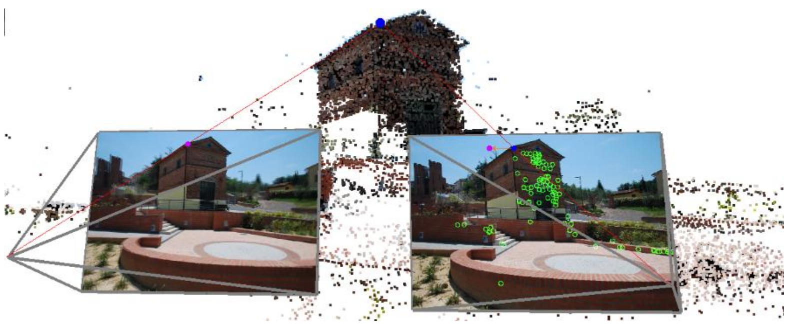
- **SfM reconstructs a point cloud from a series of images**
  - 3D positions of (sparse) matched features
  - Camera positions and orientations
- **Many approaches for densification**
  - Pipeline showed to work at interactive rates on phones (Taskanen et al 2013)
- **SCALE AMBIGUITY**

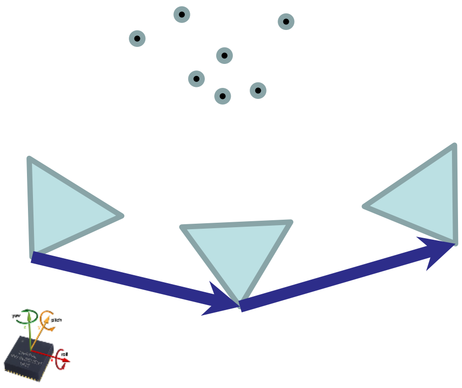




Data fusion: Visual + IMU

- **Use sensors synced with visual channel**
  - **GPS+Magnetometer** generally not applicable
  - **IMU** returns orientation and acceleration in real world units
- **Idea**
  - track camera movement with IMU during visual capture
  - use IMU data to find out the real-world distance between SfM camera positions, resolving the scale ambiguity





Data fusion: Visual + IMU

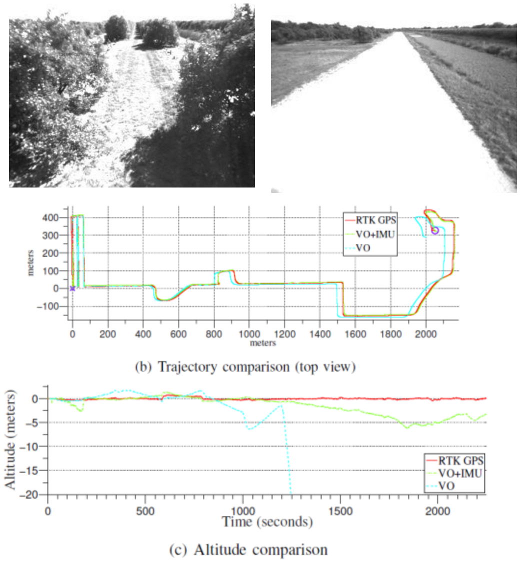
- **The accelerometer returns acceleration**
- **Therefore, we should be able to compute the displacement between two camera positions as**

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

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

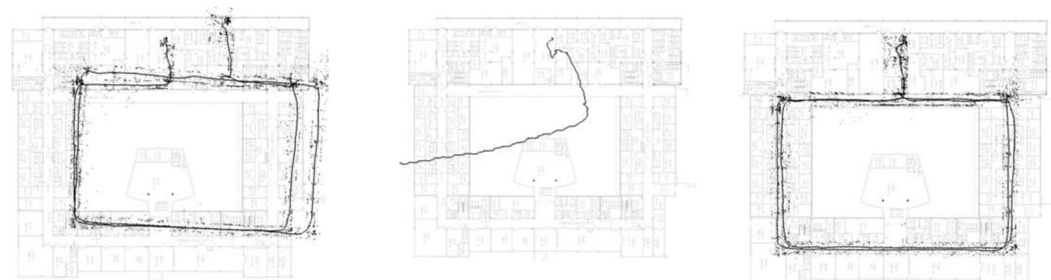
Data fusion approaches (1/5)

- **Match position from IMU integration with position from SfM, coping with noise/bias by extensive filtering**



A new approach to vision-aided inertial navigation [Tardif et al 2010]

Vision only | IMU only | Vision+IMU



Visual-Inertial Navigation, Mapping and Localization:
A Scalable Real-Time Causal Approach [Jones,Soatto 2011]

- **Requires LONG acquisition times and LONG offline processing times**

Data fusion approaches (2/5)

- **Ad-hoc online solutions taking into account IMU characteristics**
  - Segment motion in “swift movements” with large accelerations
  - Integration of IMU acceleration to derive position matched with SfM
  - Continuous process of outlier rejection and re-estimation of scale

Live metric 3D reconstruction on Mobile Phones [Tanskane et al. 2013]

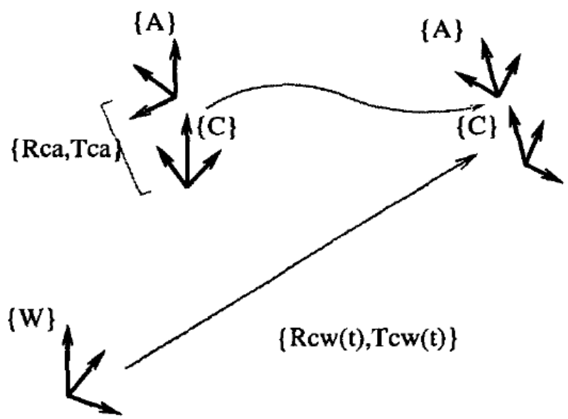
$$\arg\min_{\lambda} = \sum_{i \in I} \|\vec{x}_i - \lambda \vec{y}_i\|^2$$

One estimate of λ at the end of each swift movement
Estimation of scale λ only on inlier set *I*

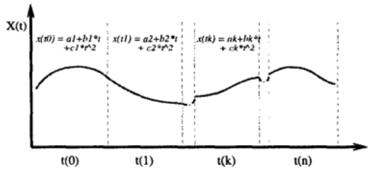
- **Working but motion-dependent and prone to accumulation error due to integration**

Data fusion approaches (3/5)

- **Match accelerations from IMU with accelerations from SfM**



$$\min \Sigma_{k=1}^{m} \|a_w(t_k) - \hat{a}_w(t_k)\|^2$$



spline parameters

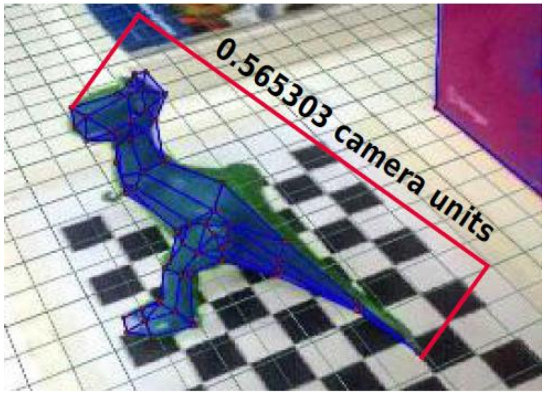
Camera trajectory estimation using inertial measurements and Structure from Motion results [JungTaylor2001]

- **Works off-line and assumes high-accuracy (robotics) IMU**

Data fusion approaches (4/5)

- **Match accelerations from IMU with accelerations from SfM at SfM frame-rate (large baseline!)**
  - **Downsample and anti-alias** IMU samples at SfM frame rate
  - Optimize scale and bias

Hand-waving away scale [Ham et al. 2014]



$$\arg\min_{s,\mathbf{b}} \eta\{s \cdot \hat{\mathbf{A}}_V + \mathbf{1} \otimes \mathbf{b}^{\mathsf{T}} - \mathbf{D}\mathbf{A}_I\mathbf{R}_I\}$$

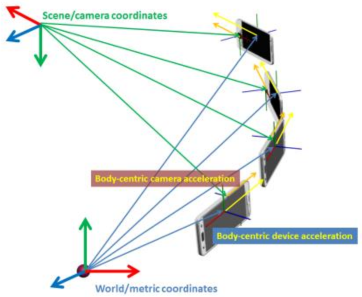
D: convolutional matrix for antialising and downsampling IMU signal

- **Requires very long acquisition times due to downsampling at SfM rate**

Data fusion approaches (5/5)

- **Match accelerations from IMU with accelerations from SfM at IMU frame-rate (small baseline!)**
  - **Upsample** SfM samples at high rate using all available visual data
  - Estimate acceleration from upsampled transforms and match them to IMU samples using robust fitting

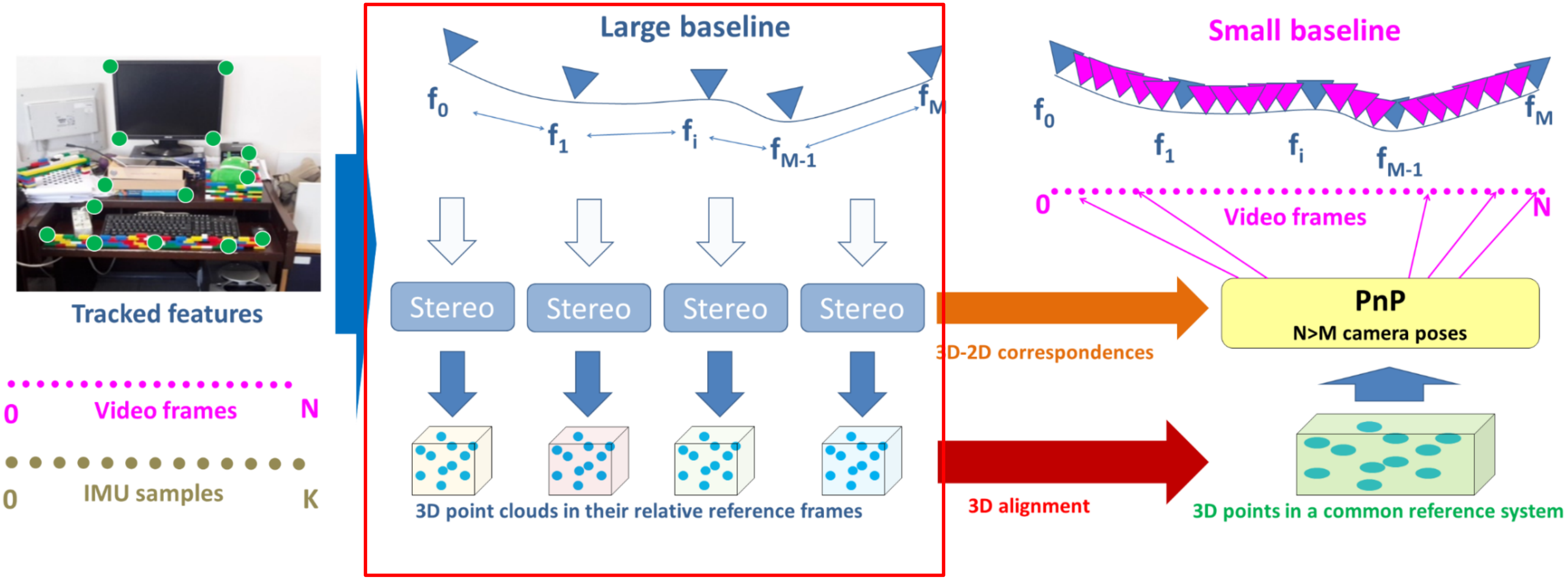
Fast Metric Acquisition with Mobile Devices. [Garro et al. 2016}



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

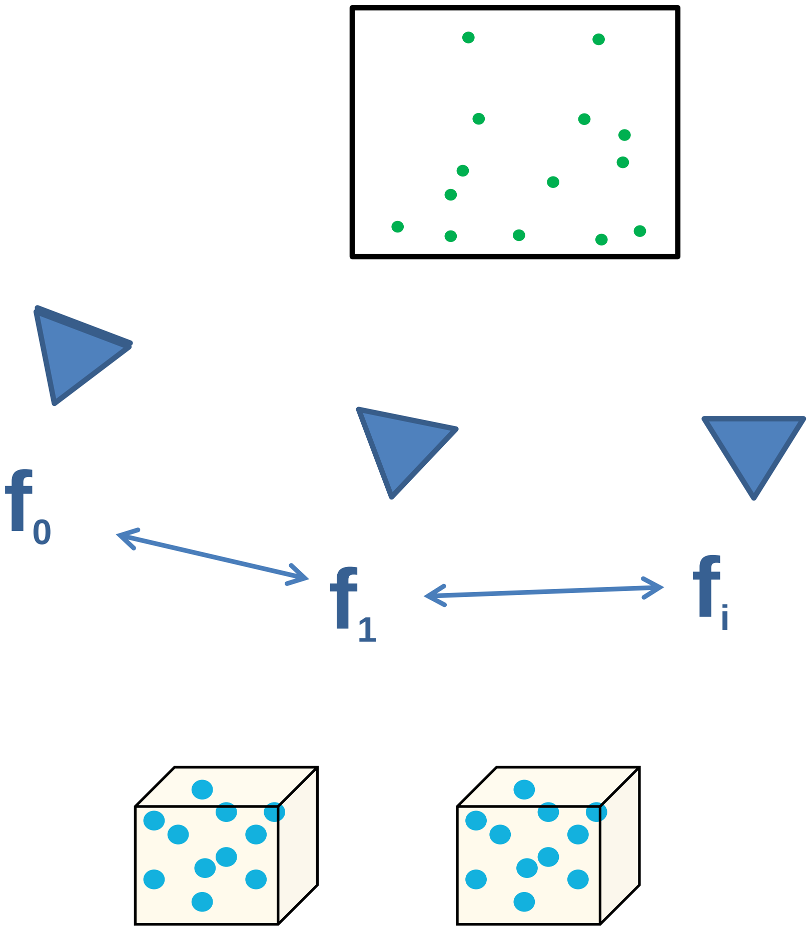
- **Fast, coping with large errors and noise**

Vision Module Pipeline



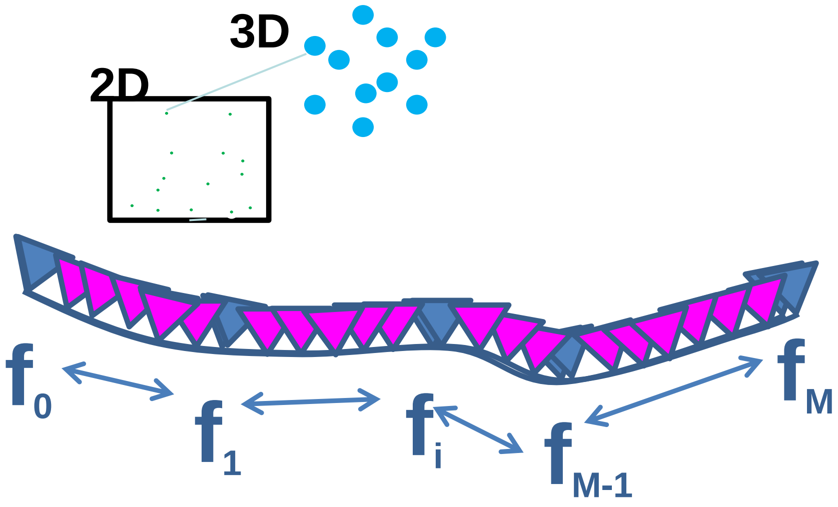
Vision Module

- **Traces Shi-Thomasi features**
- **When baseline is large enough**
  - Estimate Essential Matrix, that is, relative camera pose between f0 and fi
  - Calculate a 3D point for each feature point
- **Note: each pair of cameras has its own reference system**



Vision Module

- **Global registration**
  - M point clouds
  - A subset of features is present in each point cloud
  - Use feature correspondence to align all the point cloud in the same reference system
- **Cameras upsampling**
  - Features are tracked for **all** frames
  - Use aligned point cloud and tracking position to estimate cameras for all frames with Perspective-n-Point (**PnP)**



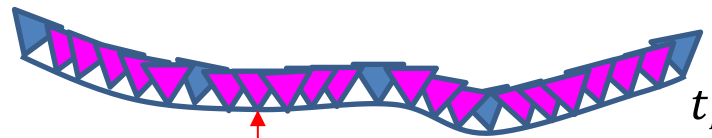
Recovering the scale factor (1/2)

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

t_0



t_K

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

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

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

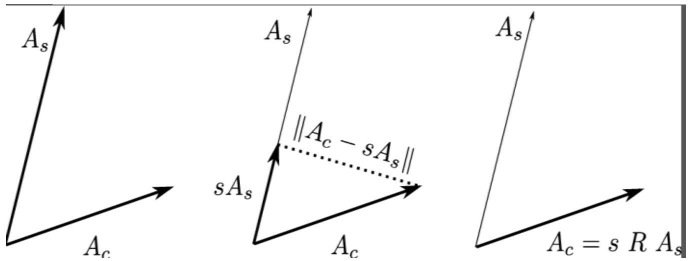
Recovering the scale factor (2/2)

- **LS, gradient descent (et similia) poorly conditioned**
  - Not so many data
  - Severe outliers

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

- **Robust fitting use RANSAC approach**
  - Use MLESAC robust estimator to maximize likelihood rather than just the number of inliers
- **Introduce rotation matrix R**
  - Account for orientation bias
  - Improve RANSAC performance

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



Results

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%

- **Median error 4% (wrt 10-15% of other STAR solutions)**

Example 2

DATA FUSION AND COMMUNICATION FOR INDOOR CAPTURE

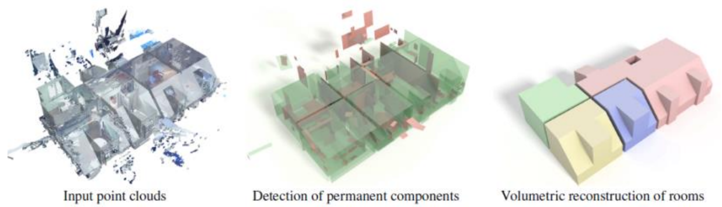
Indoor capture + presentation

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

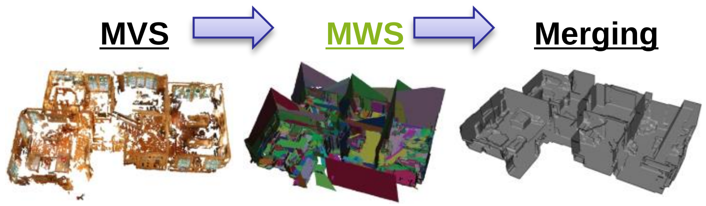
- **Much interest/applications (security, location awareness, …)**
  - Need to capture visual information together with room structure

Typical solutions

- **Indoor capture and modeling**
  - **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



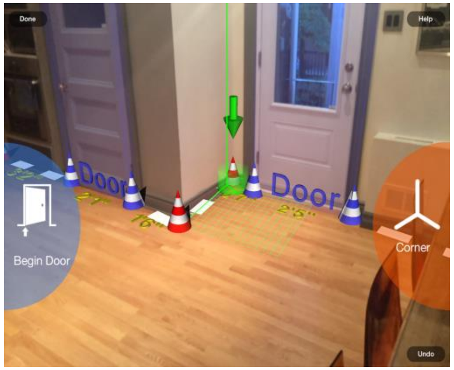
Mura et al. **Piecewise-planar Reconstruction of Multi-room Interiors with Arbitrary Wall Arrangements.** Computer Graphics Forum – Pacific Graphics 2016



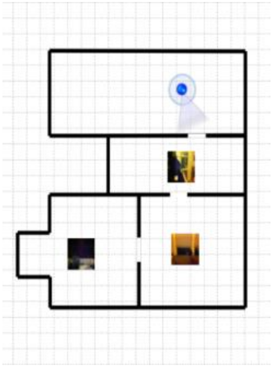
Furukawa et al. **Reconstructing Building Interiors from Images.** ICCV 2009

Examples using low-cost mobile devices

- **Interactive capture and mapping of indoor environment**
  - MagicPlan - http://www.sensopia.com
    - Floor corners marked via an augmented reality interface
    - Manual editing of the room and floor plan merging using the screen interface
  - Sankar and Seitz: Capturing indoor scenes with smartphones (UIST2012)
    - Corners marked on the screen during video playback

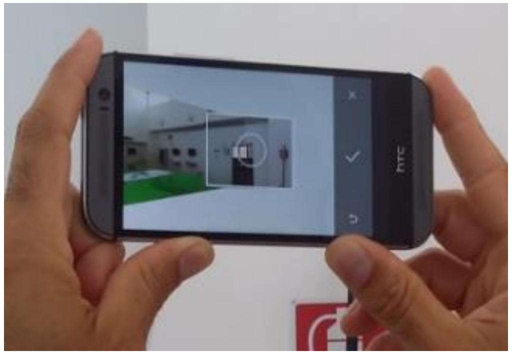






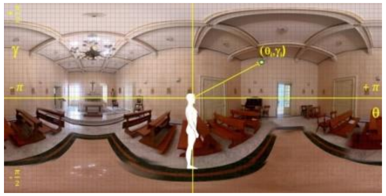
Exploiting panoramic images

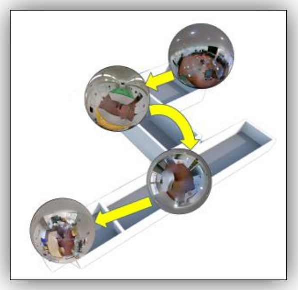
- **360 degrees images are easy to capture using common devices**
  - Interactive apps using IMU + GUI + automatic stitching
  - Dedicated cameras
- **360 degrees images are easy to navigate**
  - Spheremaps + emerging formats video+image formats
  - VR devices for immersion
- **What about analyzing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



Finding the room structure

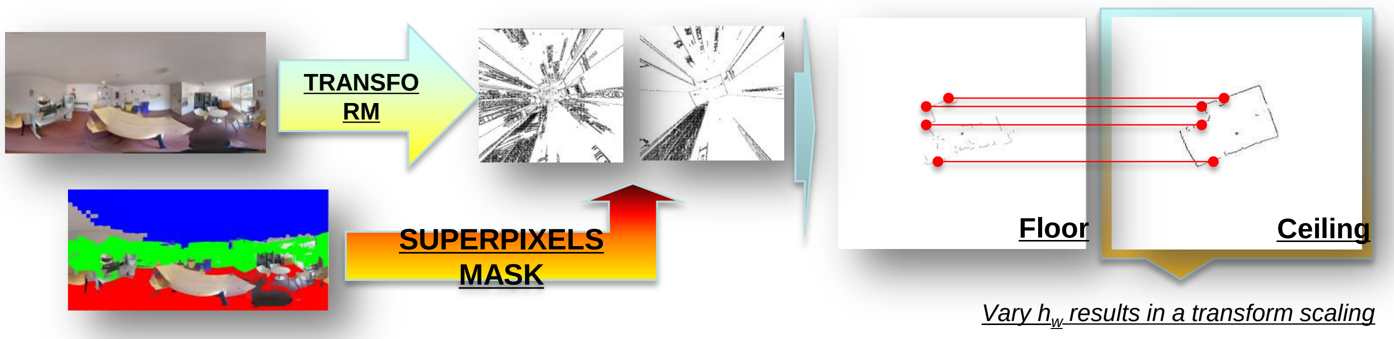
- **Take one spheremap per room**
  - Equirectangular images generated by a mobile device
    - Vertical lines aligned with the gravity vector
    - Image approx. oriented towards magnetic North
  - Eventually use IMU + Visual features for stitching
- **Track user motion to identify connections between rooms**
  - Use IMU + Visual Features for tracking
- **Solve local + global optimization to find indoor structure**
  - Multi-room environment





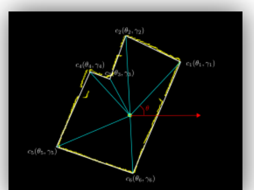
Finding the room structure

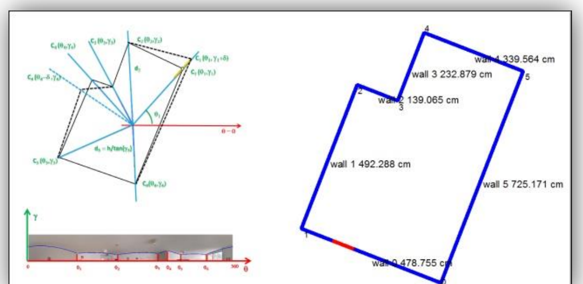
- **Analyze spheremap to extract single room structure**
  - Room model considers vertical walls
  - Extract edges and filter out regions likely far from top/bottom edges of walls
  - Find wall height
    - Voting scheme used to extract most likely wall height by maximizing pairs of matching wall-floor / wall-height edge pixels
  - Fit 2.5D room model to recovered wall edge map
- **Uses specialized transform to speed-up computation**

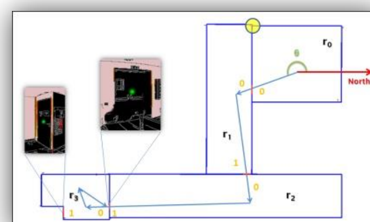


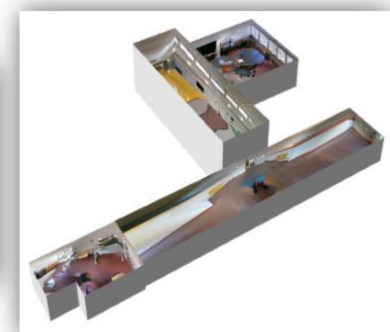
Finding the rooms structure

- **Iterated to map the entire floor-plan**
  - Mobile tracking of user's direction moving between adjacent rooms creates a connected room graph
  - Doors position identification in the image by computer vision
  - Doors matching according with graph
  - Rooms displacement
  - **Global optimization of combined model**



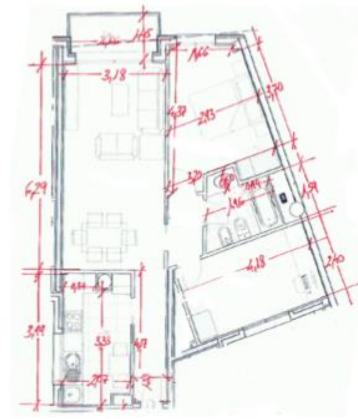






Pintore et al. **Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps.** IEEE WACV 2016

Results



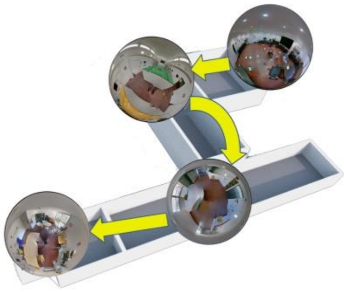
Scene	Features		Area error		Wall length error		Wall height error		Corner angle error		Editing time
Name	Area $[m^2]$	Np	MP	Ours	MP	Ours	MP	Ours	MP	Ours	MagicPlan
Office H1	720	10	2.95%	1.78%	35 cm	15 cm	2.0 cm	1.2 cm	0.8 deg	0.8 deg	26m32s
Building B2	875	25	2.50%	1.54%	30 cm	7 cm	6.0 cm	1.5 cm	1.5 deg	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

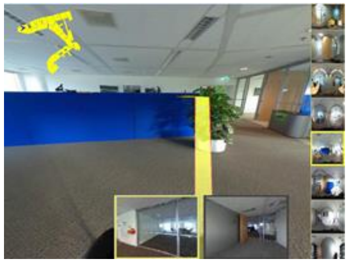
Pintore et al. **Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps.** IEEE WACV 2016

- **Reasonable, fast reconstruction with rough structure and visual features**

Sharing the indoor model

- **Indoor model**
  - Exploration graph
    - Each node is a spheremap/room
    - edges (yellow) are transitions between adjacent rooms
    - Stored on a server (standard http Apache2)
  - Panoramic images
    - Mapped according with the graph
- **Interactive exploration**
  - Room
    - **WebGL fragment shader**
    - dragging to change view orientation and pinching to zoom in/out
  - Passages
    - **Real-time rendering** of the transitions between rooms
      - Exploiting geometric model stored on the server
      - Performance improvement compared to use precomputed videos
    - Suggested paths



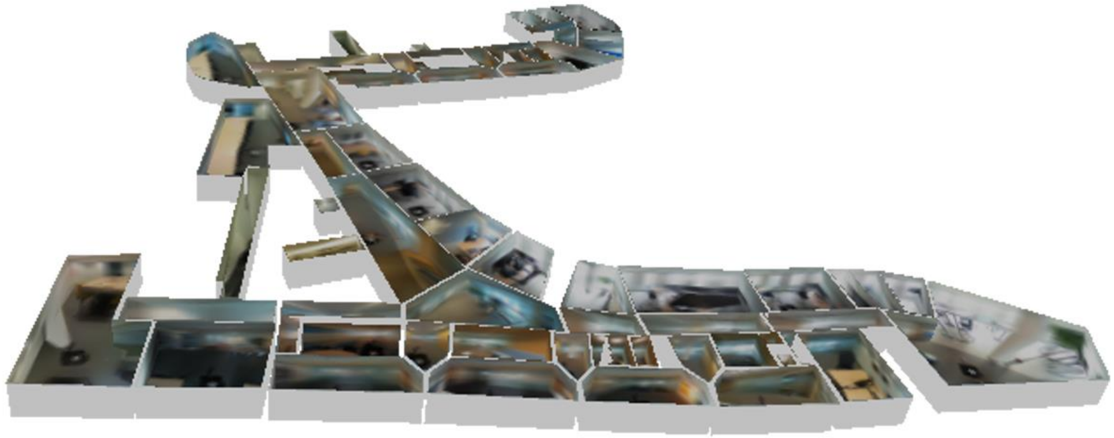


Some results

Live demo: http://vcg.isti.cnr.it/vasco/

Click on the dataset on the left column to start

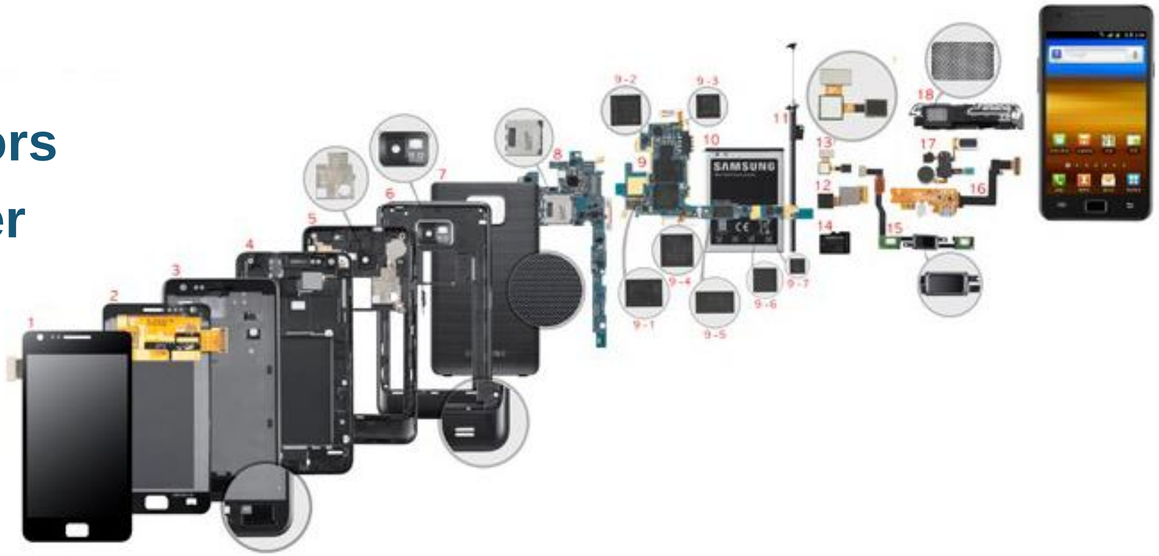
3D reconstruction of a 655 mq office with 19 rooms. This environment was acquired with a mobile phone (HTC One M8)



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

Wrap-up: mobile apps characterized by the exploitation of mobile device features

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



After the break: rendering!

BREAK!