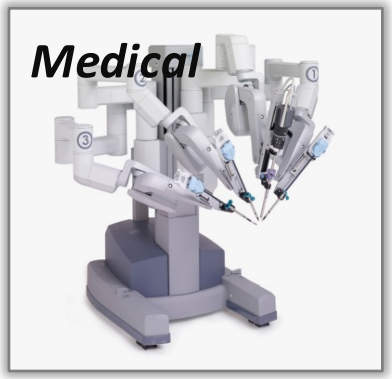
Extracting and Quantifying Visual Information from Mobile Devices

Minh N. Do

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Mobile Devices with Visual Perception

- Automobile
- Service
- Consumer
- Medical
- Entertainment
- Education
- Domestic
- Manufacturing
- Military
- Augmented Reality



Required Perception of Mobile Devices

Mapping

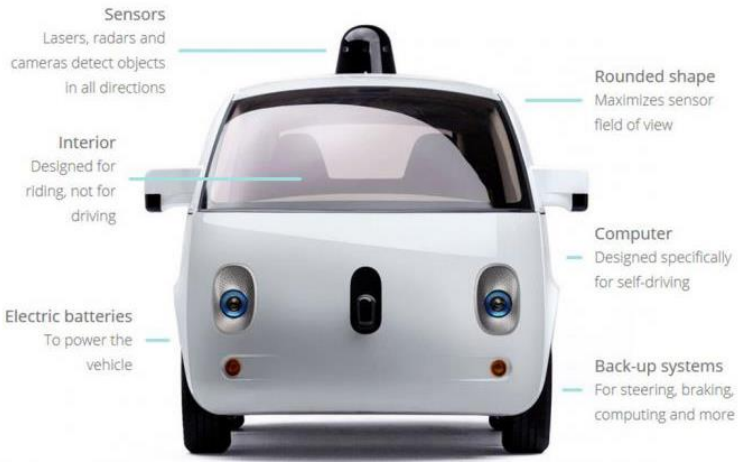
- Location
- Geometry
- Semantics
- Updates

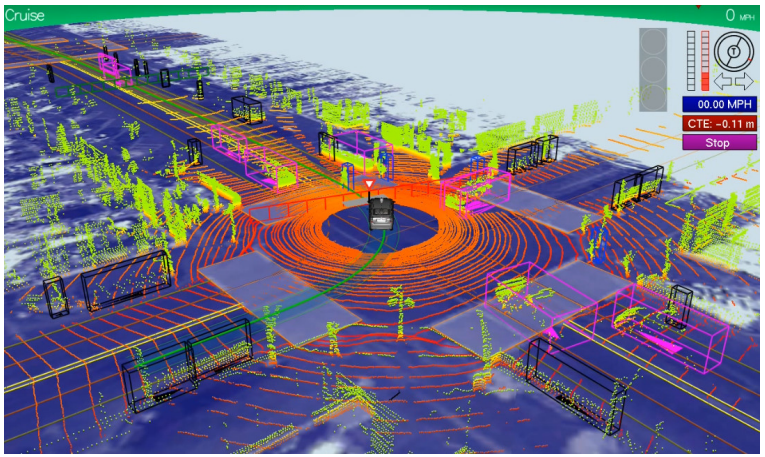
Object

- Distance
- Dimension
- Category
- Instance

Dynamics

- Motion
- Behavior
- Interaction



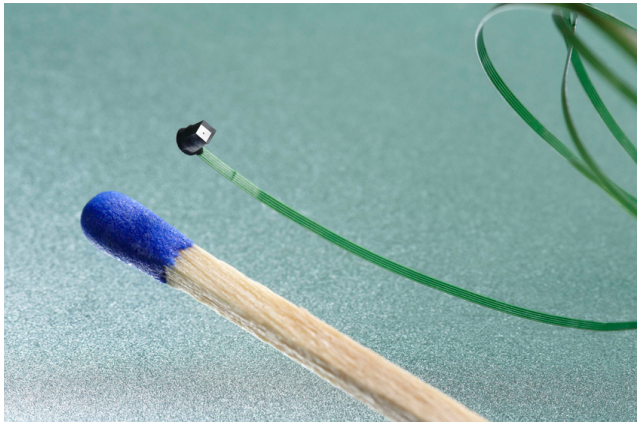


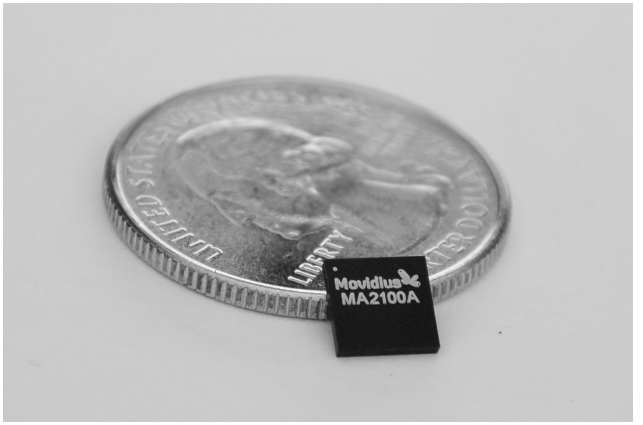
Vision as Sensing Input

- High resolution provides details about complex scenes
  - Typical camera has ~1.3 mega-pixel, running at 36FPS (~50Mbyte per second)
  - More than 70% of human body's sensors are in the eyes

- Visual information: Shape vs. Appearance
  - Most complex situations are defined by appearance (texture) more than shape:
    - e.g. road markings, traffic signs, person identity, object instance, etc.

- Cheap and versatile in size and configuration





Computer Vision

- Single image
- Static scene
- RGB only
- Limited data
- Limited computation power
- Slow algorithms

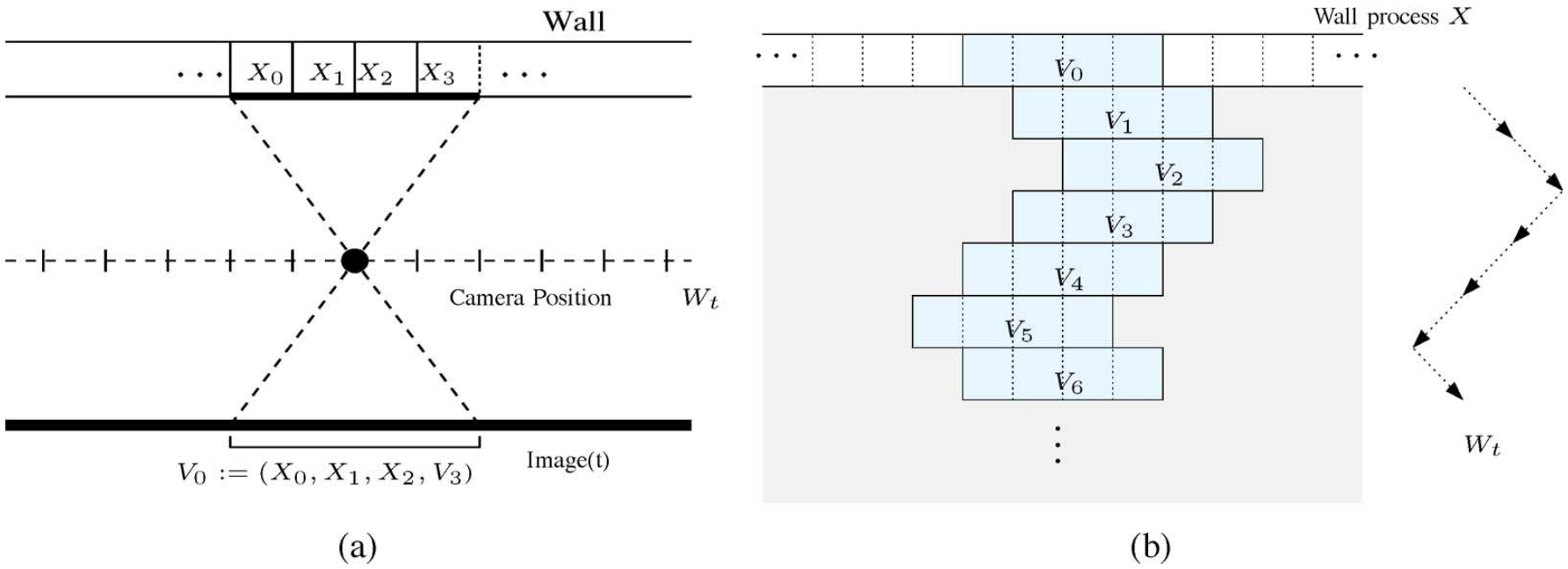
- Video
- Dynamic scene
- Depth, IMU, GPS
- Large amount of data
- Visual computing chips
- Real-time algorithms

- Key problems in mobile vision:
  - Localization and mapping
  - Object and place recognition
  - Motion and dynamics

Vision from a Mobile Device

Source: Seattle Police Department

On the Information Rates of Mobile Vision

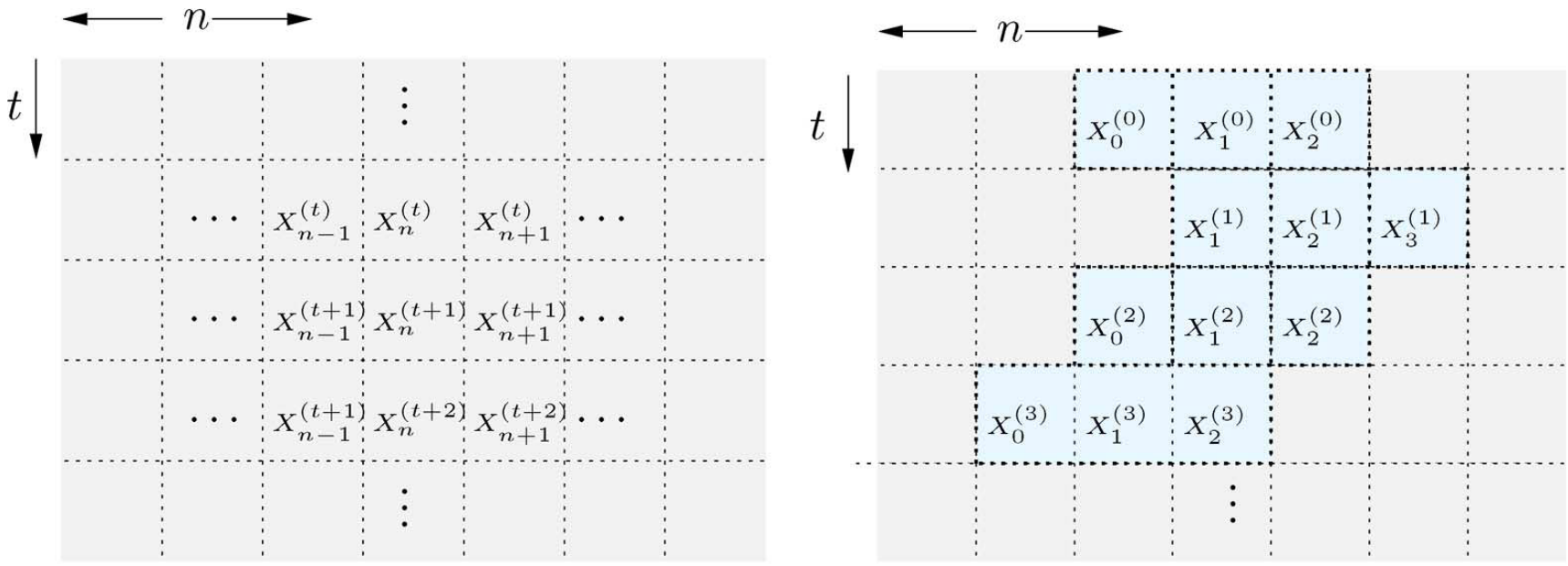


Theorem: The entropy rate and rate-distortion function of V satisfies

$$(1-2p_W)H(X) + H(p_W) - H(P_e) \leq H(V) \leq (1-2p_W)H(X) + H(p_W)$$
$$R_V(D) \leq (1-2p_W)R_D(X) + H(p_W)$$

Cunha, Do, Vetterli (2010)

Mobile Vision of Dynamic Environments



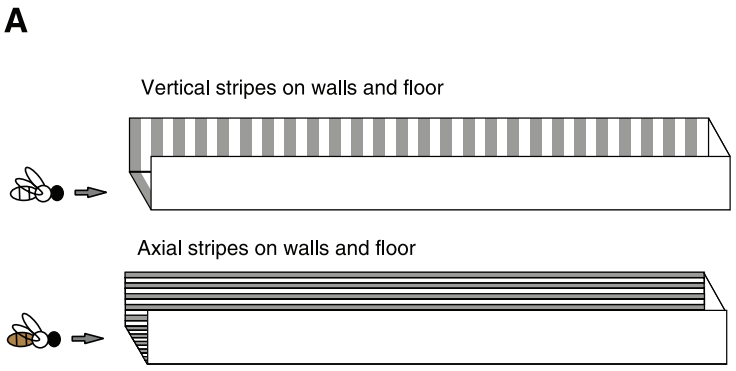
(a) (b)

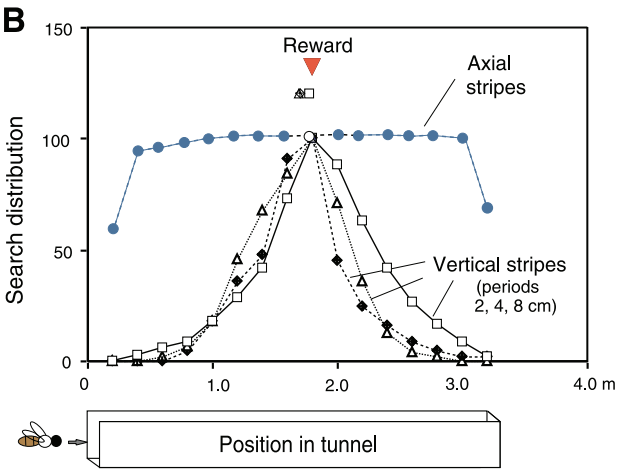
Theorem:

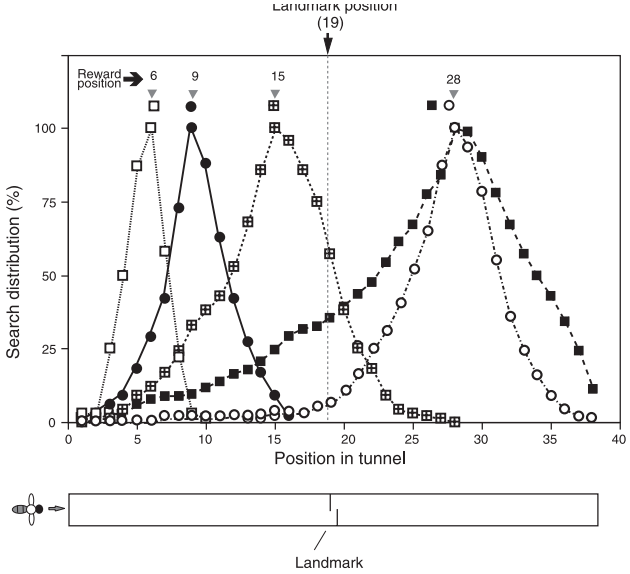
$$H(V|W) + H(p_W) - H(P_e) \leq H(V) \leq H(V|W) + H(p_W)$$

where $H(V|W) = (L-1)H(X_0^{(1)}|X_0^{(0)}) + (1-2p_W)H(X_0^{(\infty)}) + \sum_{i=0}^{\infty} H(X_0^{(i)}|X_0^{(0)})\Pr\{\mathcal{T}^i\}$

Lessons from Bees







Srinivasan (1997, 2011)

Holistic Mobile Vision

Geometric Reconstruction meets Semantic Recognition for 3D holistic vision: Real-time, robust, geometry-centric vision

***Where**: geometry, location…*

***What**: semantics, action…*

Reconstruction

- Real-time camera pose localization
- 3D environment mapping
- Depth and motion estimation
- Large-scale urban reconstruction

Recognition

- Visual place/scene recognition
- Object recognition, localization
- Human re-identification
- Action recognition, tracking

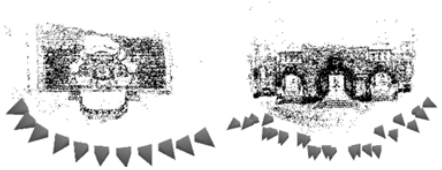
***How**: model, solve, compute*

Key techniques

- Geometry-aware filtering
- Fast randomized algorithms
- Efficient inference models
- Deep learning innovation

- Multiple sensors
- Modern vehicles
- Moving robots
- Opportunistic scan

- Mobile cameras
- Big visual data
- Rich annotations
- Powerful machine

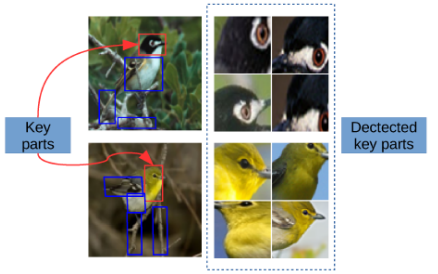


3D vision & percept.

- CODE [ECCV'14, '16] **Feature matching**
- DSE [CVPR'15] **Camera pose rec.**
- SLAM-O [TCSVT'16] **3D object prop.**

Recogn. & segment.

- **Fine-grained recog.** [TIP'16]
- PISA [CVPR'13b] **Pixel-acc. saliency**
- MHIC [TMM'16] **Recog & Co-seg.**



Scene structure

- Depth enhance.
- Slanted stereo
- Optical flow
- Live FG segment
- 3D city recons.

Comp. imaging

- Photo refocus
- Rain removal
- Image stitching
- Multi-scale dec.
- Structure extra.
- Image warping
- Colorization

Edge-aware filtering

- CLMF [CVPR'12] **Local EAF**
- FGS [TIP'14] **Global EAF**
- FGI [ECCV'16] **Sparse to dense**

Efficient inference

- PMF [CVPR'13] **Local optim.**
- DFF [CVPR'14] **Generalized**
- SPM-BP [ICCV'15] **Global optim.**

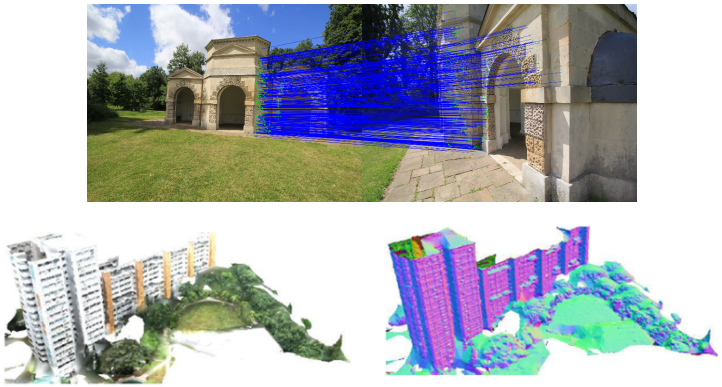
Current topics

- *Scene flow*
- *Robust SLAM*
- *Motion seg.*
- *Text detection*
- *Place recogn.*
- *Scene labeling*
- *Action recog.*
- *Deep learning*

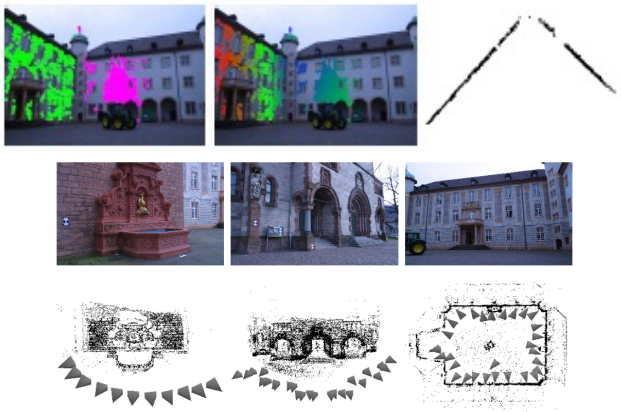
Holistic Mobile Vision

Dense depth & motion

Feature matching and mapping

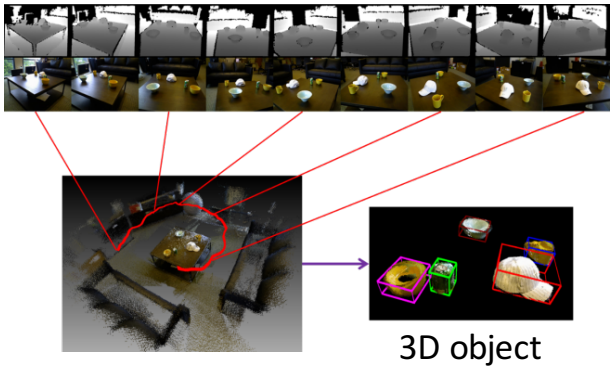


Camera pose localization

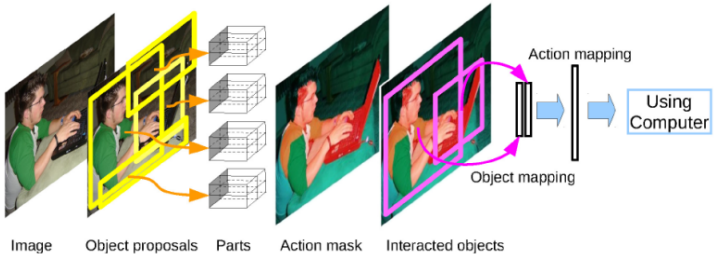


Stitching & visualization

Localizing 3D object proposals



Action recog. w/ min. labelling



Summary

- Visual perception is crucial for mobile devices and autonomous systems
  - Small
  - Cheap
  - Fast

- Key problems:
  - Localization and mapping
  - Object and place recognition
  - Motion and dynamics

- Holistic mobile vision:
  - Combine geometric reconstruction with semantic recognition
  - Supported by information rates
  - Challenges: real-time and robust methods