Assisting the Visually Impaired Using Depth Inference on Mobile Devices via Stereo Matching

Benjamin Chidester, Minh N. Do

7/15/2013

Desired Elements of a Navigation Assistance Tool

Qualities:

- Light, comfortable, convenient, non-intrusive, avoids negative social side-effects, inexpensive

Functionality:

- Obstacle detection and avoidance
- Environment enrichment features
  - Beacons or waypoints
  - Object recognition and scene description

Desired Elements of a Navigation Assistance Tool

Qualities:

- Light, comfortable, convenient, non-intrusive, avoids negative social side-effects, inexpensive

Functionality:

- Obstacle detection and avoidance
- Environment enrichment features
  - Beacons or waypoints
  - Object recognition and scene description

We focus on this essential element

Desired Elements of a Navigation Assistance Tool

Qualities:

- Light, comfortable, convenient, non-intrusive, avoids negative social side-effects, inexpensive

Functionality:

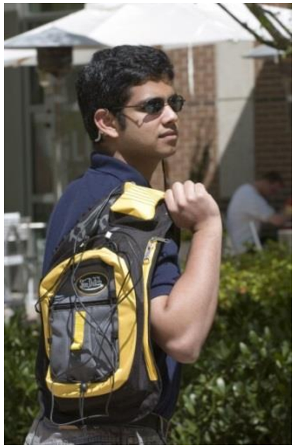
- Obstacle detection and avoidance
- Environment enrichment features
  - Beacons or waypoints
  - Object recognition and scene description

We focus on this essential element

Requires depth information

Related Tools for Navigation Assistance

- SWAN: System for Wearable Audio Navigation
  - Beacons and waypoints guide user toward destination
  - Voice recordings and GPS allow user to save notes about a particular location on the route
  - Object recognition describes elements of scene to user
- The vOICe
  - Captured image is described to user through sound



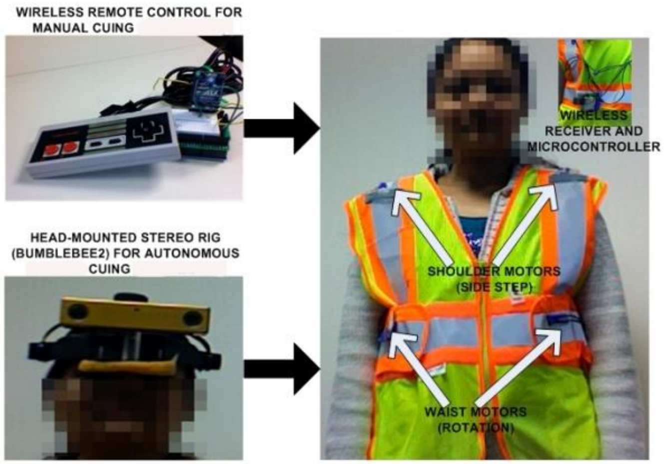
http://sonify.psych.gatech.edu/research/swan/index.html

http://www.seeingwithsound.com

- Listen2dRoom
  - Elements of room are identified and spoken to user

Related Tools for Navigation Assistance

- SLAM - Univ. of Southern California
  - Depth is detected from stereo cameras
  - Wearable motors provide cues to user for directions
- BrainPort
  - Images are described to user using a touch device on the tongue



V. Pradeep, G. Medioni, and J. Weiland, "Robot vision for the visually impaired,"

http://www.scientificamerican.com/article.cfm?id=device-lets-blind-see-with-tongues

Mobile Devices as a Platform for Depth Inference

Benefits:

- Convenient – user may already own a mobile device
- Non-intrusive, light, comfortable – no additional hardware required, avoids negative social side-effects
- Computational power – enough computational power housed within the device to perform computer vision tasks

Trend in Mobile Imaging:

- Camera arrays
  - Thinner devices
  - Computational photography applications
  - 3D video

http://www.pelicanimaging.com/

Mobile Devices as a Platform for Depth Inference

Benefits:

- Convenient – user may already own a mobile device
- Non-intrusive, light, comfortable – no additional hardware required, avoids negative social side-effects
- Computational power – enough computational power housed within the device to perform computer vision tasks

Trend in Mobile Imaging:

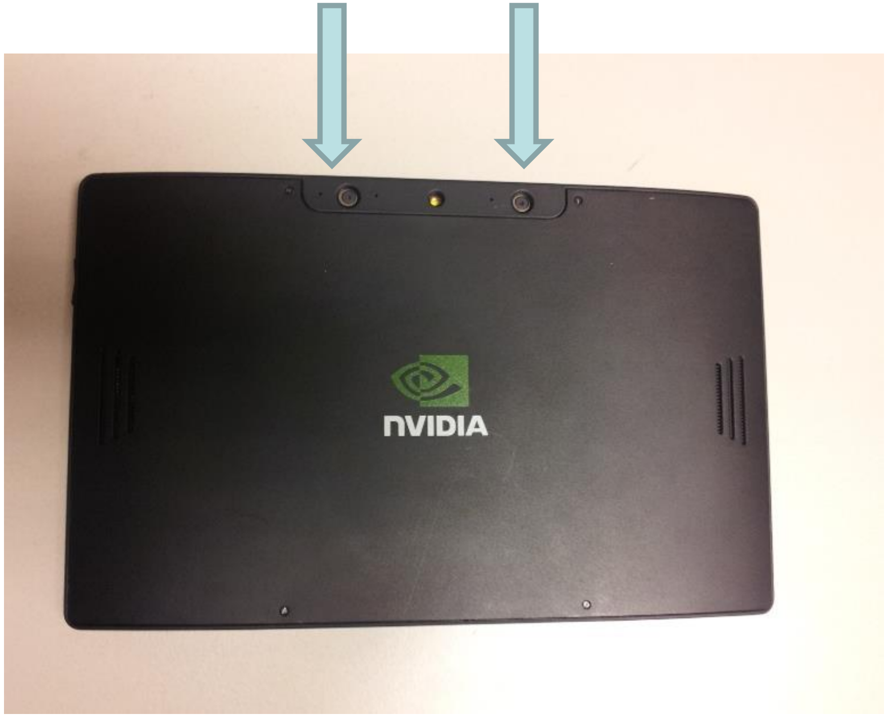
- Camera arrays
  - Thinner devices
  - Computational photography applications
  - 3D video

Camera arrays provide depth information

http://www.pelicanimaging.com/

NVIDIA Tegra 3 Developer Tablet

Stereo Cameras



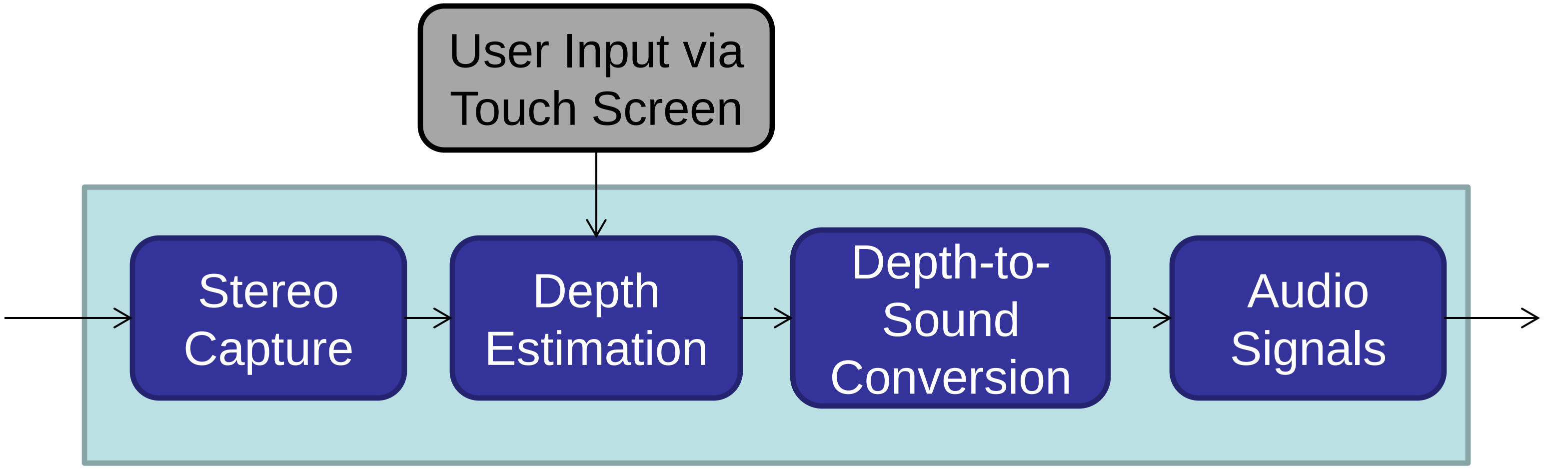
FCam API provides access to camera parameters

TEGRA 3 SPECIFICATIONS

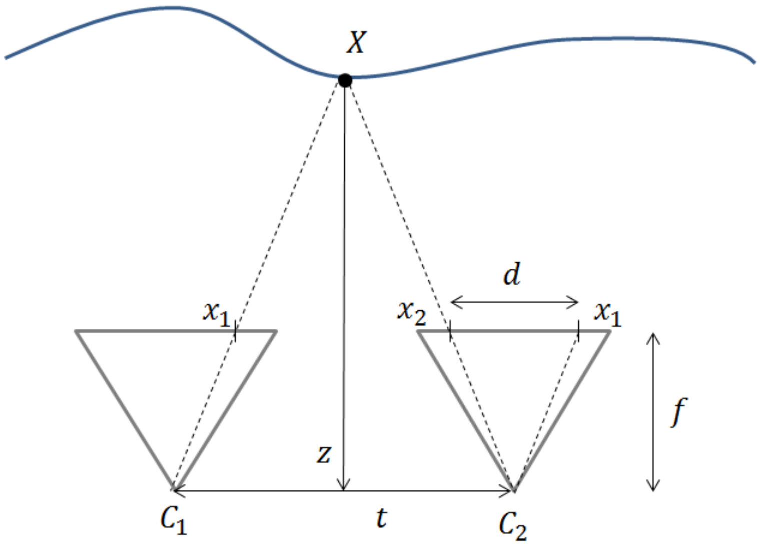
	Tegra 3 on Android
Processor	
CPU	Quad-core, with 5th battery-saver core
Max Frequency	Up to 1.7 GHz single core /1.6 GHz quad-core
L2 Cache	1 MB
L1 Cache (I/D)	(32KB / 32KB) per core
Memory	
Frequency	DDR3-L 1500 LPDDR2-1066
Memory Size	Up to 2 GB
GPU	
Architecture	ULP GeForce

http://www.nvidia.com/object/tegra-3-processor.html

Proposed Mobile System



Depth Resolution from Stereo



Depth: $z = \frac{tf}{d}$

t: baseline (m)

f: focal length (in pixels)

z: depth (m)

d: disparity (in pixels)

Conversion of Disparity to Depth for NVIDIA Tablet

Disparity (pixels)	Depth (meters)
1	38.15
2	19.08
5	7.63
10	3.82
30	1.27
50	0.76

Can infer depth of several meters, which is appropriate for navigation

Depth Inference via Stereo Matching

Local Methods

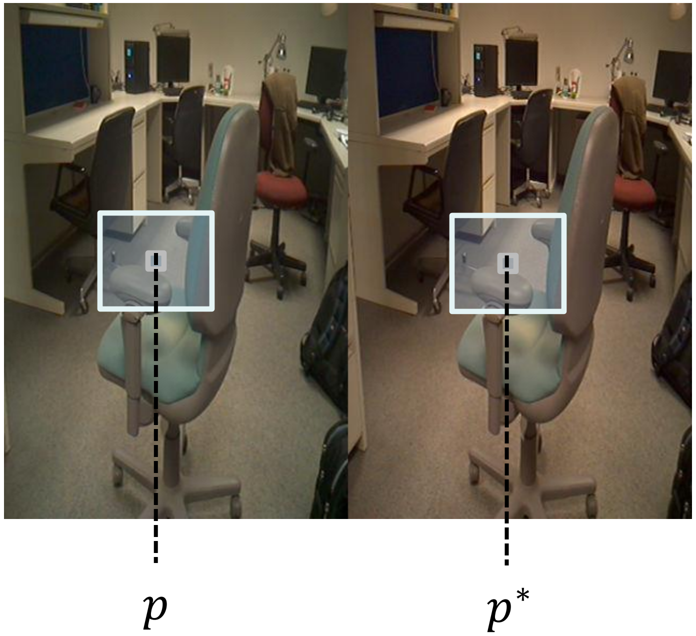
- window-based correspondence search on an individual pixel basis
- least computationally demanding approaches
- less robust

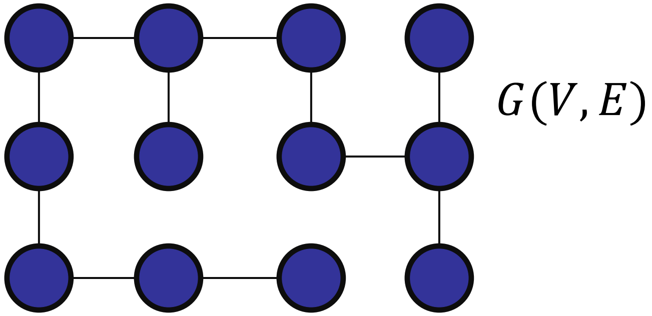
Semi-Global Methods

- optimization includes global smoothness penalty:

  $$E(\mathcal{D}) = E_{data}(\mathcal{D}) + \lambda E_{smooth}(\mathcal{D})$$

- more accurate inference
- computationally demanding





Mobile Stereo Matching Potential

Local Method

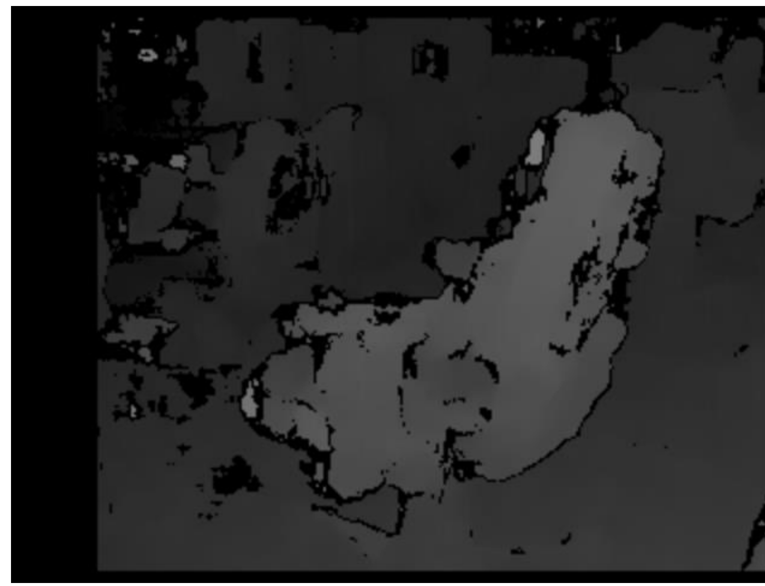
- ~5 frames per second
- depth-to-sound (or depth-to-touch) mapping will reduce dimension, so some inaccuracies can be tolerated
- timing can be reduced

Semi-Global Method

- ~1.5 frames per second
- Accuracy might not be worth the speed trade-off

320x480	Local Method	Semi-Global Method
	202 ms	672 ms

Another Real-World Example



Conveying Depth Information to the User

- Depth-to-Sound
  - interferes with sounds from surroundings
  - only required additional hardware are headphones
- Depth-to-Touch
  - limited resolution
  - does not interfere with sound
  - discomfort to user

Both can require extensive training.
Unclear which is more effective.

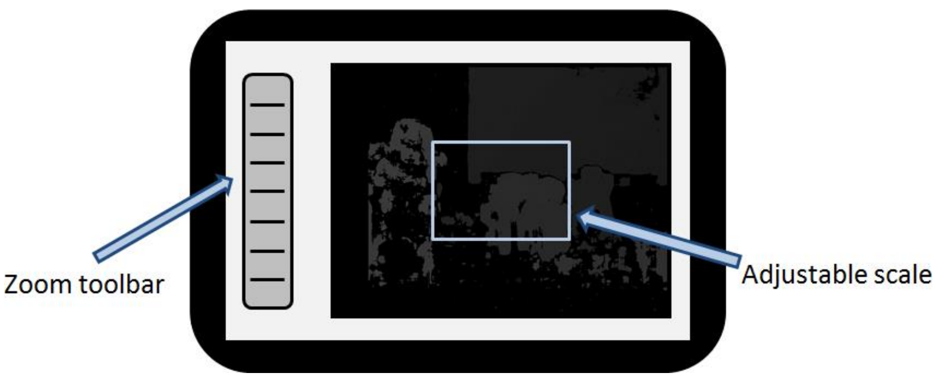
http://www.seeingwithsound.com

http://www.scientificamerican.com/article.cfm?id=device-lets-blind-see-with-tongues

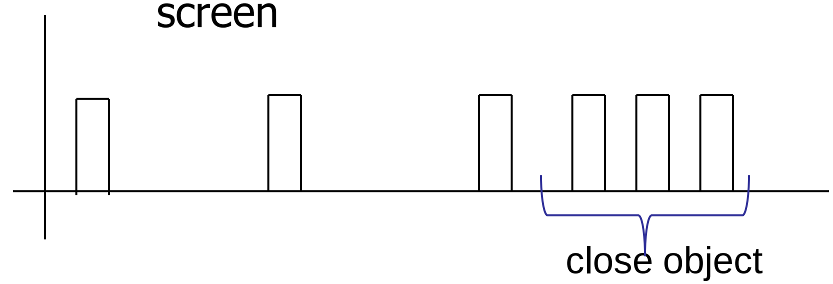
User Interface and Depth-to-Sound

Philosophy:

Do not over-interpret the data; Leave the interpretation to the user.

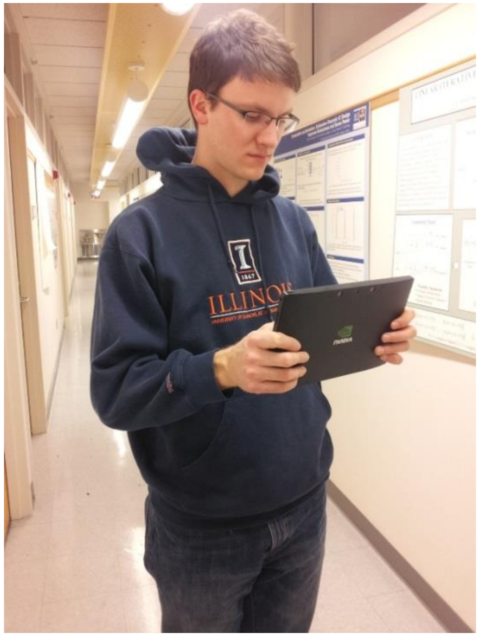


- Zoom toolbar
  - User defines the scale of the region of interest of the scene over which to aggregate depth information
- Shift-able window
  - User designates the location in the image of the windowed aggregation by touching the screen
- Depth-to-Sound
  - Average depth over window modulates pitch of output tone or frequency of beep pulse



Conclusion

- Reliable depth inference in real-time (~5 fps and greater) is achievable with stereo matching
- System demonstrates the viability of depth inference on mobile devices to assist the visually impaired
- Benefits of system:
  - Convenient, non-intrusive, no additional hardware
  - Could be easily deployed in the near future for widespread use as an app
  - A variety of depth-to-sound (or potentially depth-to-touch) mappings could be tested by owners of tablets and smart phones with camera arrays
  - GPS and 3G connectivity allow for easy integration of other possible enhancements, such as GPS waypoints, street name notifications
  - Other computer vision inference tools, such as scene understanding



Thank You