Computer Vision-Based Assistive Technology for the Visually Impaired

James M. Coughlan, Ph.D.

THE SMITH-KETTLEWELL EYE RESEARCH INSTITUTE

Commonly used assistive technologies (ATs) for visually impaired persons

The most widely used ATs are *simple* and *reliable:*

White cane

Lenses, screen magnifiers

Accessible Pedestrian Signals (crosswalks)

Detectable warning surfaces (e.g., bumps on curb ramps)

Commonly used assistive technologies (ATs) for visually impaired persons

Smartphone-based ATs:

- text-to-speech display
- GPS apps → useful for outdoor navigation, transit, etc.

Much more assistance is needed!

Wayfinding (getting from one point to another) assistance needed indoors and outdoors – but GPS unavailable indoors

Orientation and mobility require high-resolution self-localization and environmental sensing: where exactly do I walk to board the bus, while avoiding obstacles, etc.?

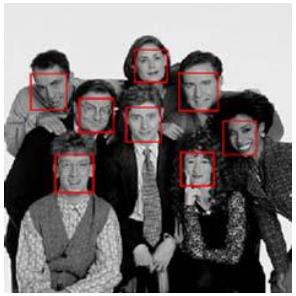
Need to read printed documents and signs

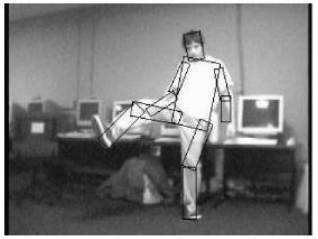
Many devices and appliances require use of visual displays (e.g., LED/LCD)

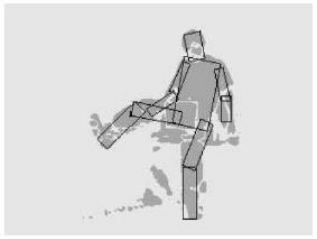
…

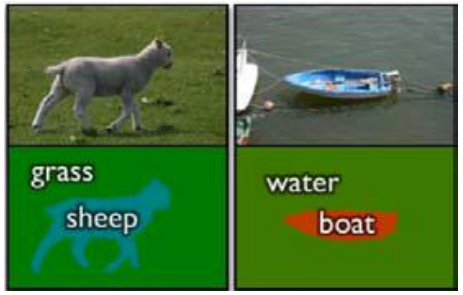
Solution: why not use computer vision?

Q: “Computer vision for the blind – what could be more obvious? Why aren’t more people doing it?”








grass
sheep
water
boat

+

= ?

Why not use computer vision?

Q: “Computer vision for the blind – what could be more obvious? Why aren’t more people doing it?”

A1: It’s extremely difficult to make assistive technology that is useful, reliable and convenient.

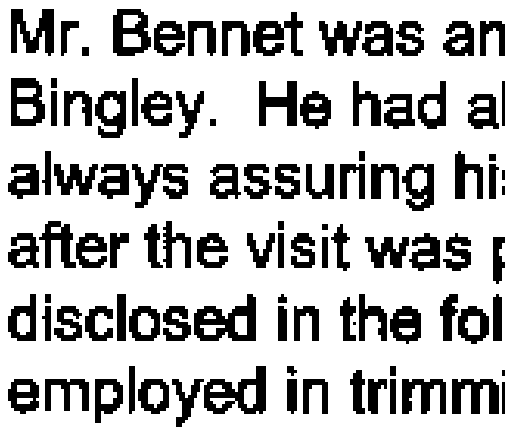
Why not use computer vision?

Q: “Computer vision for the blind – what could be more obvious? Why aren’t more people doing it?”

A2: In particular, it’s even harder to make computer vision work reliably under the kinds of *unconstrained conditions* typically demanded by visually impaired users.

Computer vision successes and limitations

#1. Optical Character Recognition (OCR)



→

Mr. Bennet was an
Bingley. He had al
always assuring his
after the visit was p
disclosed in the fol
employed in trimmi

Image

Text read by OCR

Computer vision successes and limitations

OCR works well under good conditions:

Standard fonts

Uniform background

High-quality, high-res., high-contrast image

Minimal amount of non-text clutter in image

Computer vision successes and limitations

But OCR falters when conditions are less than ideal:

OCR reading:

"MECHANICAL" "MIOHANICAL"

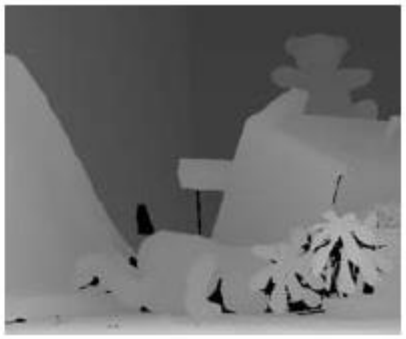
Computer vision successes and limitations

#2. Stereo depth reconstruction

Infer depth from two slightly different views of scene

Left

Right



Disparity

Computer vision successes and limitations

Complications in using stereo:

Challenging lighting conditions → many pixels under/over-exposed

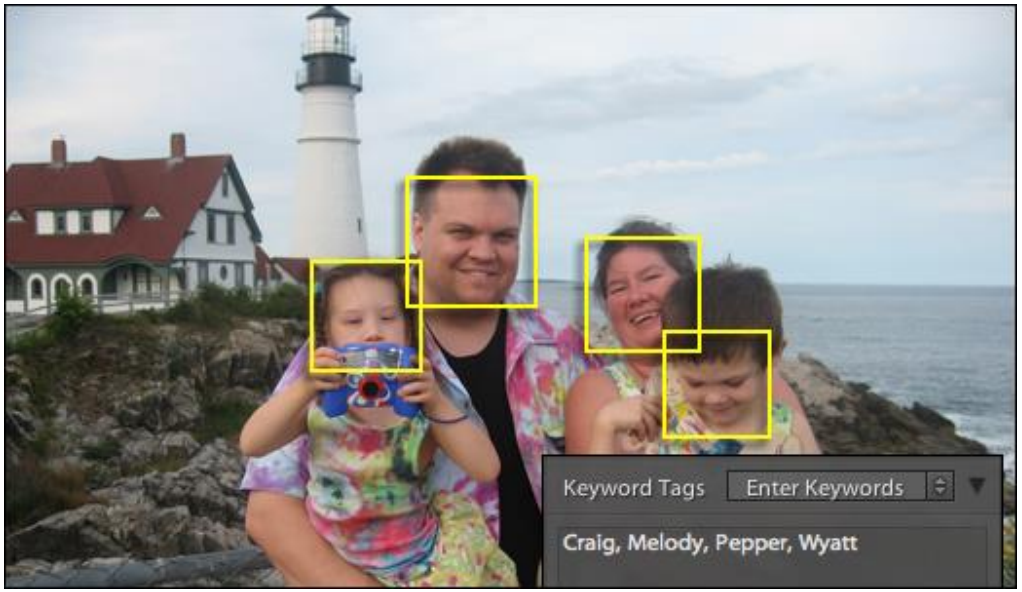
Specular surfaces

Texture needed for stereo processing may be absent or ambiguous

Computer vision successes and limitations

#3. Object detection/recognition

Works very well under constrained conditions - e.g., face recognition when there are few possible people to distinguish (consumer applications such as Picasa and iPhoto)



Computer vision successes and limitations

However, face recognition is very difficult in uncomposed images with many possible people to recognize (e.g., surveillance video in airport, cocktail party)

Computer vision successes and limitations

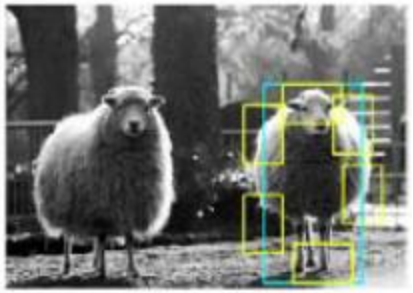
Object recognition is hard for deformable/articulated objects like people/animals…

Computer vision successes and limitations

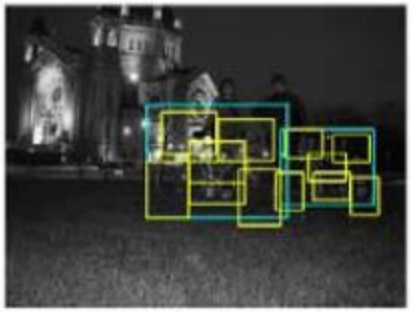
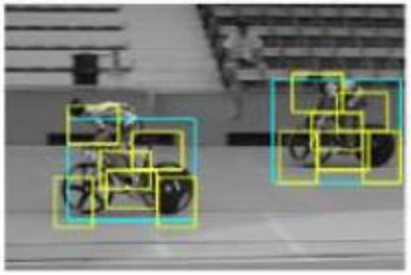
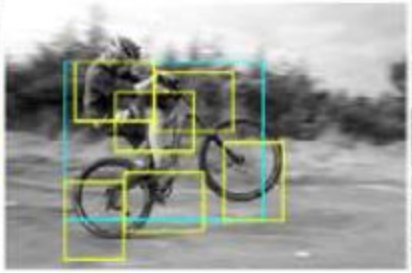
Successful detections *Failures*

People

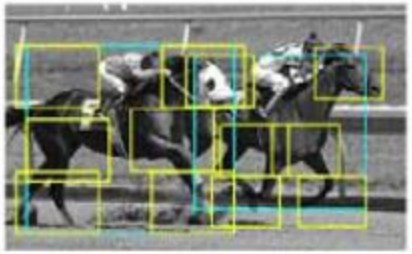


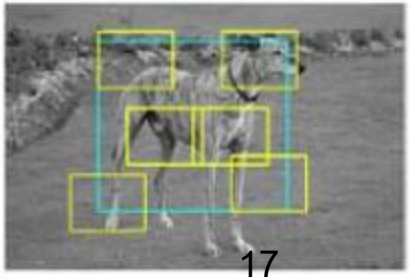



Bikes


Horses



Computer vision successes and limitations

Themes:

The state-of-the-art is always improving

But most algorithms are “brittle,” requiring certain assumptions to be met, and are often trained on “well-behaved” images

Uncomposed images, difficult lighting are always challenging

Computer vision successes and limitations

Narrowly defined tasks are most likely to be tractable (e.g., detect and read the barcode in an image)

High-level tasks such as “Is there a place for me to sit down?” or “How do I get to the conference room?” are nearly impossible to solve at this time!

Important consideration: what are the consequences of failure? (Self-driving car vs. face recognition.)

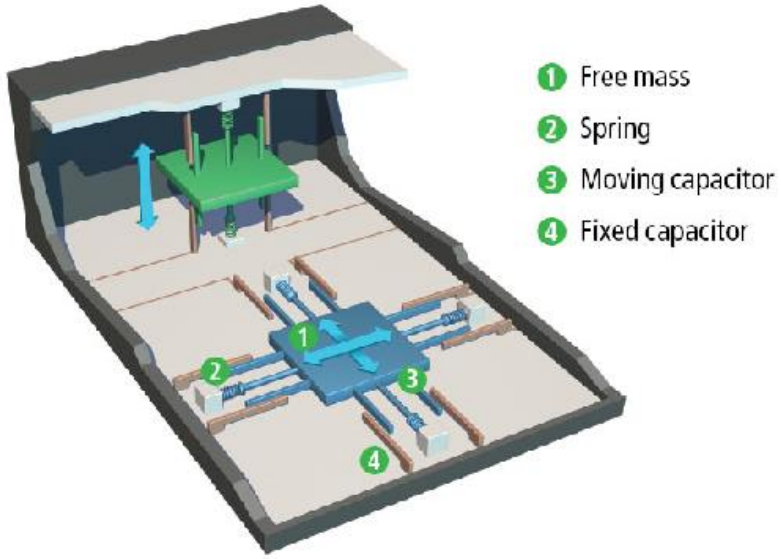
What other technologies can augment computer vision?

#1. Sensors

Many sensors on-board modern smartphones:

- Accelerometer (tilt sensor)
- Compass (magnetometer)
- GPS (only available outdoors; resolution ~10 m. in urban areas)
- Gyroscope (part of IMU = inertial measurement unit)

Etc…



Accelerometer in smartphone

Sensors: application example

Crosswatch (Coughlan, Shen, Murali, Fusco):

A smartphone-based system designed to give information and guidance to blind travelers at traffic intersections

– e.g., “Where am I? How do I align myself to the crosswalk? When is the Walk light illuminated?”

Challenge: how can a blind user aim the camera to take good images of intersection scene?

Sensors: application example

Solution: take an entire panorama of the intersection.

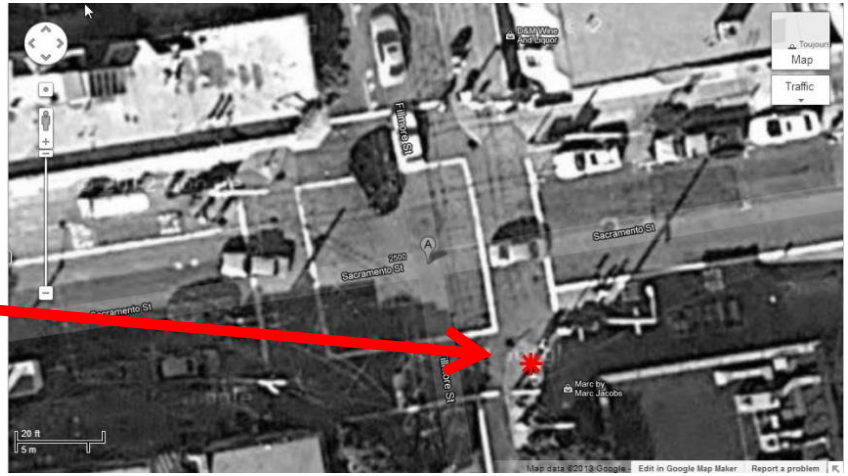
Use **accelerometer** to ensure that camera remains horizontal as it is panned in a circle. (Smartphone issues vibration warning if camera deviates too far from horizontal.)

Resulting panorama acquired by blind user

Sensors: application example

Then create aerial view, using **magnetometer** to estimate rough direction of north:

Match with template of intersection (determined using **GPS**) to estimate current location (with precision better than GPS):



What other technologies can augment computer vision?

#2. Crowdsourcing / humans-in-the-loop

Why rely solely on computer vision to analyze/interpret images?

Since smartphones have internet connection, can transmit images/video to a central server to allow sighted observers to interpret images

Crowdsourcing

Not only can a sighted operator answer *specific* questions more reliably and accurately than a computer can…

He/she can also answer open-ended questions that relate to the context of the task or goal at hand.

Crowdsourcing

Q: “Where am I?”

A: “You are outside of a bank. There is a long line of people in front of the ATM.”

Q: “Is there a place to sit down?”

A: “The chairs near you are taken, but there is a bench near the entrance.”

Computer vision vs. crowdsourcing

General principles:

- Computer vision excels at solving constrained tasks
- Often works quickly (e.g., real-time)
- Calculations can be done onboard smartphone or other portable computer → can circumvent bandwidth constraints (e.g., real-time video)

Computer vision vs. crowdsourcing (con't)

- Whereas crowdsourcing excels at less constrained problems
- Can reduce amount of computations required on smartphone
- But, quality / availability of operators may matter a lot!

Remaining challenges

Computer vision can be effectively augmented by many complementary technologies to facilitate image analysis.

However, two related challenges remain for creating effective image-based ATs:

(1) Acquiring good images

(2) Communicating spatial information about environment to user

Acquiring usable images

How to aim camera with little or no vision?
→ Appropriate feedback is essential! (And feedback needs to be real-time.)

E.g., helping blind users take photos [Vázquez & Steinfeld 2012]

Acquiring usable images

Many practical considerations compound the problem:

- camera resolution
- field of view
- speed of system
- challenging lighting conditions

(Interrelated: one consideration may often be traded off for another.)

Camera form factors

Many different possibilities are emerging, each with different properties (how camera is held/aimed, field of view, etc.)

Standard

Google Glass

Camera form factors

Wearable

Omnidirectional

Proprioception issue

It can be challenging for a user to interpret system feedback about direction (which uses camera line of sight as a reference) in terms of *body* direction.

E.g., smartphone app that signals the detection of a visual target: ask blind people to use app to walk towards a target. Easy for some people, not for others.

Proprioception issue

Which is easiest: pointing a camera that is attached to the head (Google Glass), torso (chest-worn) or hand (standard smartphone)?

How does this depend on the particular user and task?

Communicating spatial information to users

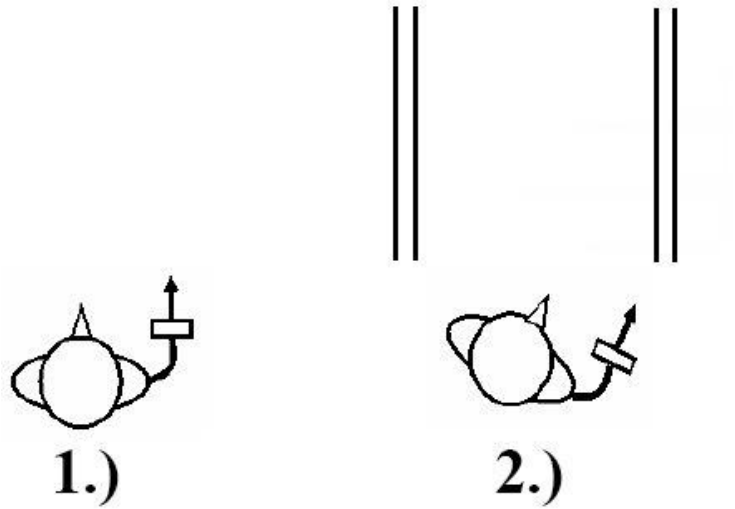
Even if system has acquired good images, what is the best way to provide spatial information about environment to user?

E.g., for wayfinding, step-by-step verbal directions may be best (“turn left at corner”) since they exploit natural reference points (such as the corner).

Communicating spatial information to users

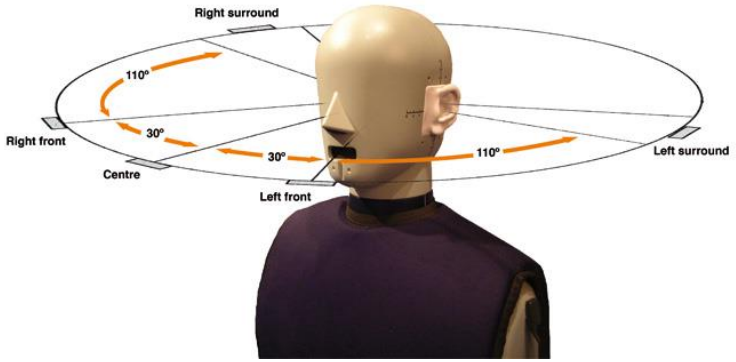
But verbal directions may be less effective for other tasks, which can require more detailed specifications of directions and displacements.

E.g., what to tell user in situation 1 or 2 to help achieve proper alignment to crosswalk?



Ingredients for possible solutions

- Spatialized sound ([Miele & Lawrence 2013]): use stereo sound to convey 3D directions



- Gesture recognition (e.g., Fifth Sense; Orcam): user can indicate 3D directions by pointing with hand/fingers

Conclusions

- Computer vision is just one of several technologies that can work together to solve problems for visually impaired persons
- It is essential to design system with appropriate user interface and hardware that facilitates the acquisition of usable images and communicates effectively with users

Final note

You won't know for sure whether a system…

- solves a real problem
- is effective
- is easy to use

until you test it with actual end users.

User testing should be done early and often!

Thanks to…

Collaborators and helpers:

Dr. John Brabyn, Tom Fowle, Dr. Giovanni Fusco, Bill Gerrey, Dr. Volodymyr Ivanchenko, Dr. Roberto Manduchi, Dr. Josh Miele, Dr. Vidya Murali, Dr. Pannag Sanketi, Dr. Huiying Shen, Dr. Ender Tekin, David Vásquez

Anonymous volunteer subjects in experiments

Funding from NIH, NIDRR and Smith-Kettlewell