AudioVision

Audio Depth Map as a Replacement
for Traditional Vision

JP Verkamp

Basic Overview

- Why?
- Required components
  - Stereo camera setup
  - Stereo vision algorithm
  - Depth map conversion
  - Audio representation

Why?

- Senses have been substituted for each other before
  - Braille uses touch to replace sight (for reading)
  - Sign language uses vision to replace hearing
  - A similar project uses the tongue's sense of touch to replace vision (in general)
- Related to Computer Vision / Image Processing
- Related to Human Computer Interfaces

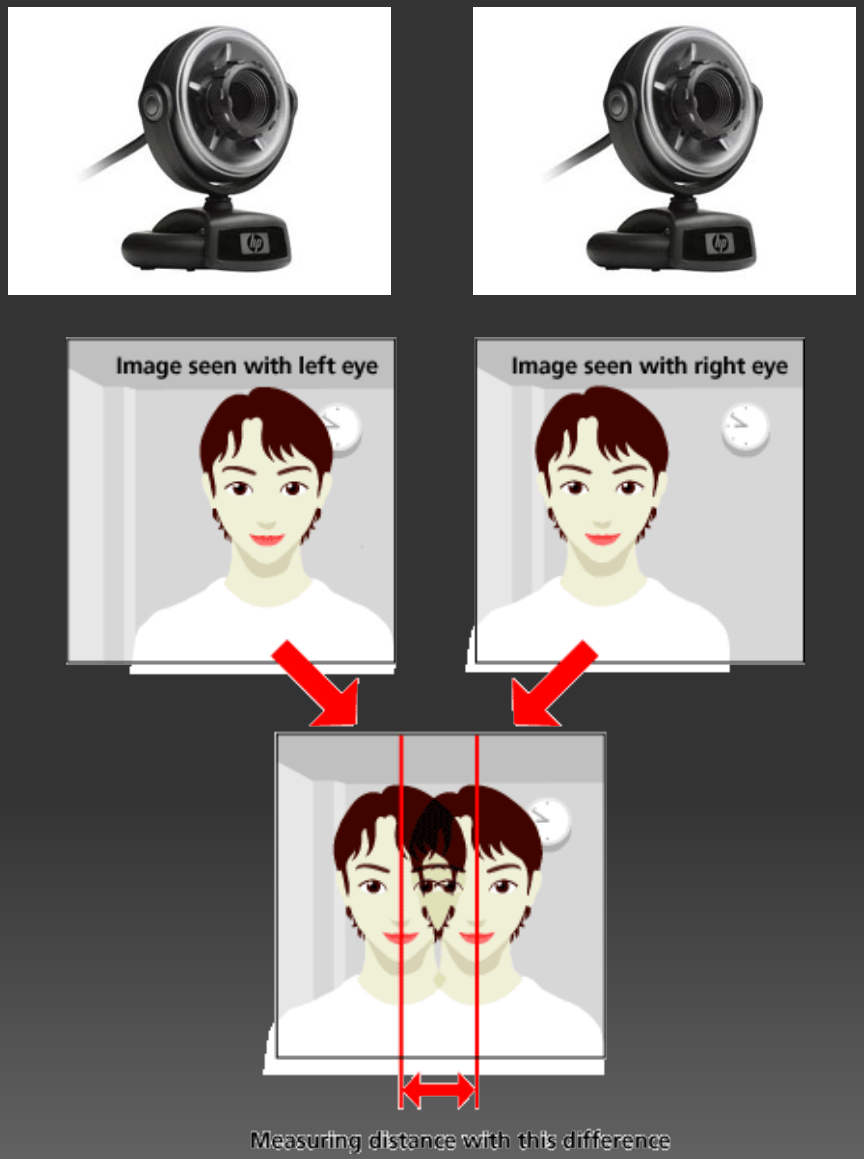
Required Components

- Stereo camera setup
- Stereo vision algorithm

- Depth map conversion

- Audio representation

Stereo Camera Setup



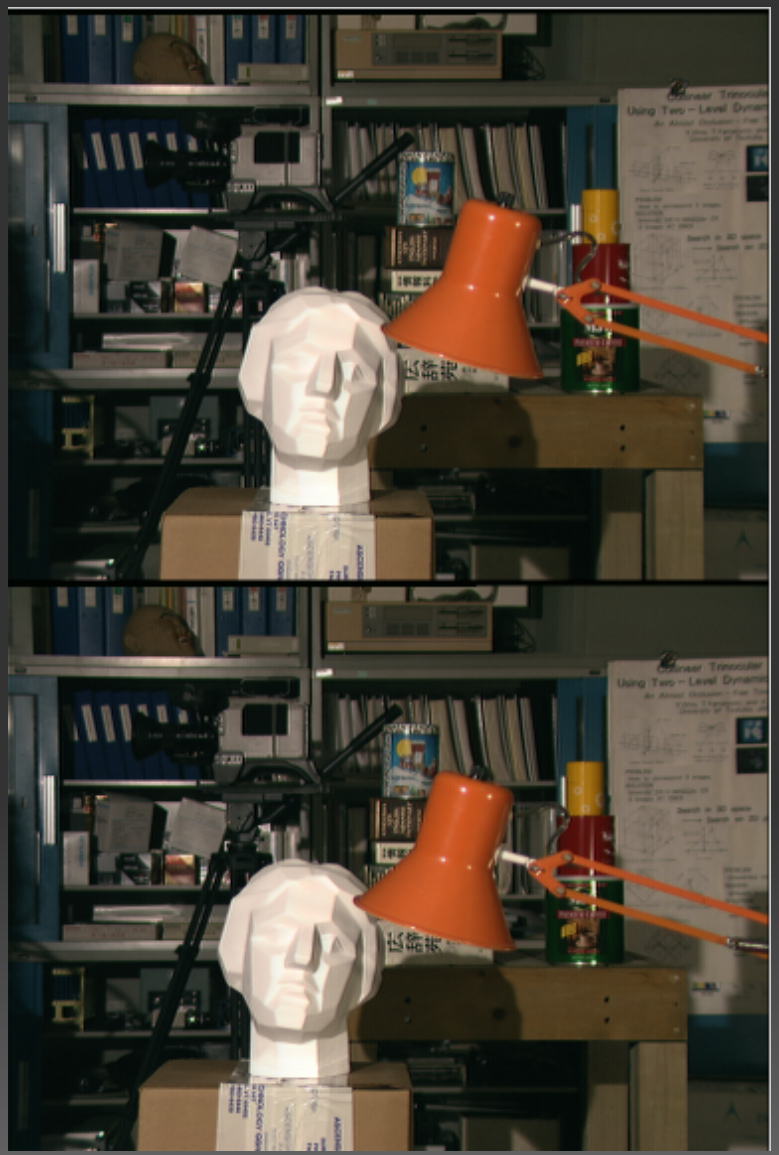
- Two webcams
- With two cameras, depth can be calculated
- Cheap webcams are perfectly acceptable for this resolution

http://www.nec.co.jp/products/robot/english/tech/img/im04_06.gif

Stereo Vision Algorithm

- Start with two cameras with a known displacement
- Match regions in the two images/video streams
- Calculate the disparity (distance) between those regions using trigonometry
- Smooth the disparity map for best results
- Result is a decently accurate map of relative distance to each point in the image

Stereo Vision Algorithm Example

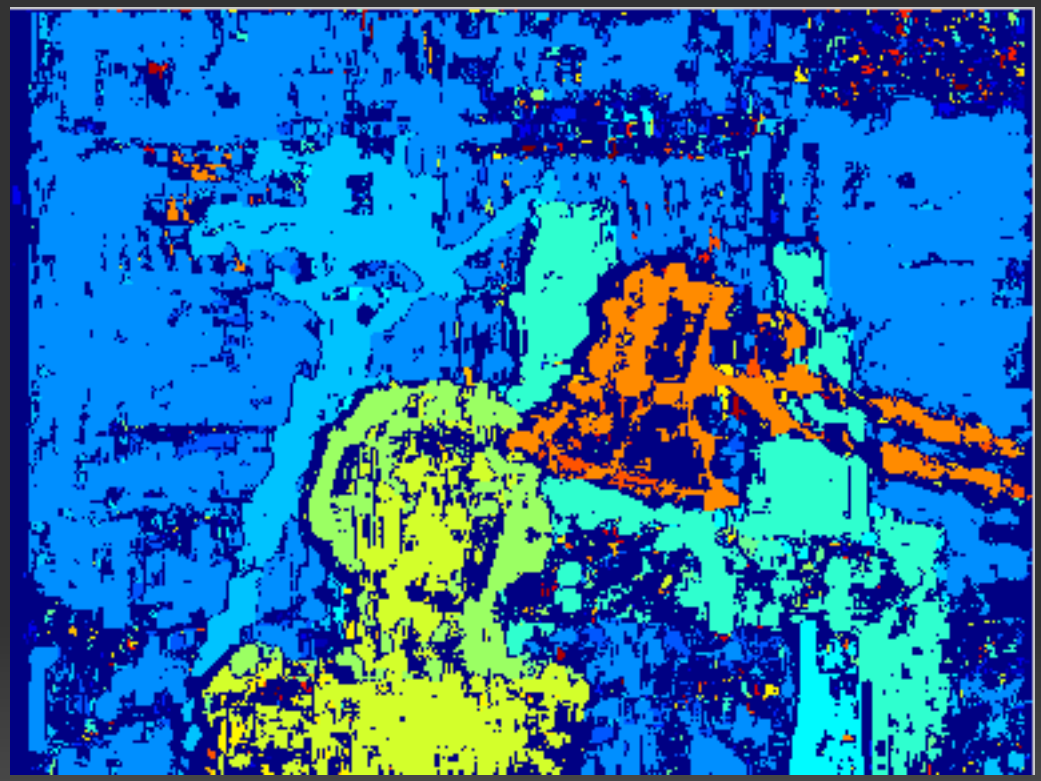


- Sample test image used to test stereo vision libraries
- Lamp is closest, followed by bust, followed by shelves
- Distances are known making testing easier

http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/

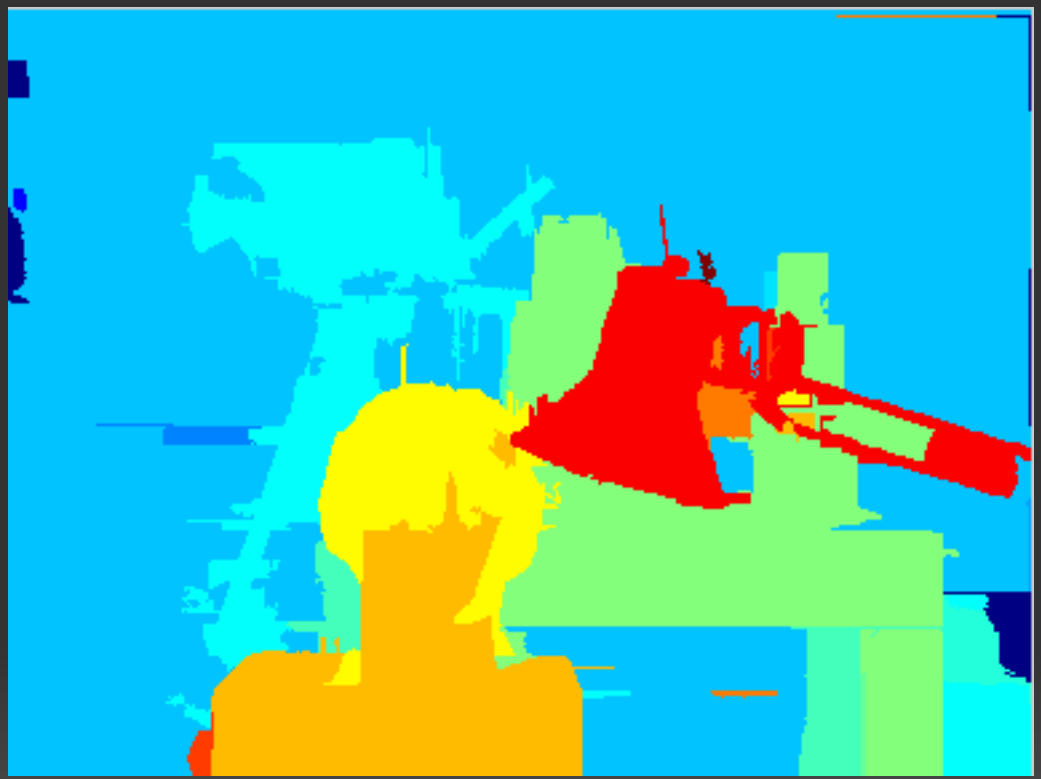
Stereo Vision Algorithm Example

- Original disparity map
- Red points are closest; blue points are furthest away
- Rough disparity is already visible
- A number of incorrect results from complex sections of the image (the book shelves)



http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/

Stereo Vision Algorithm Example

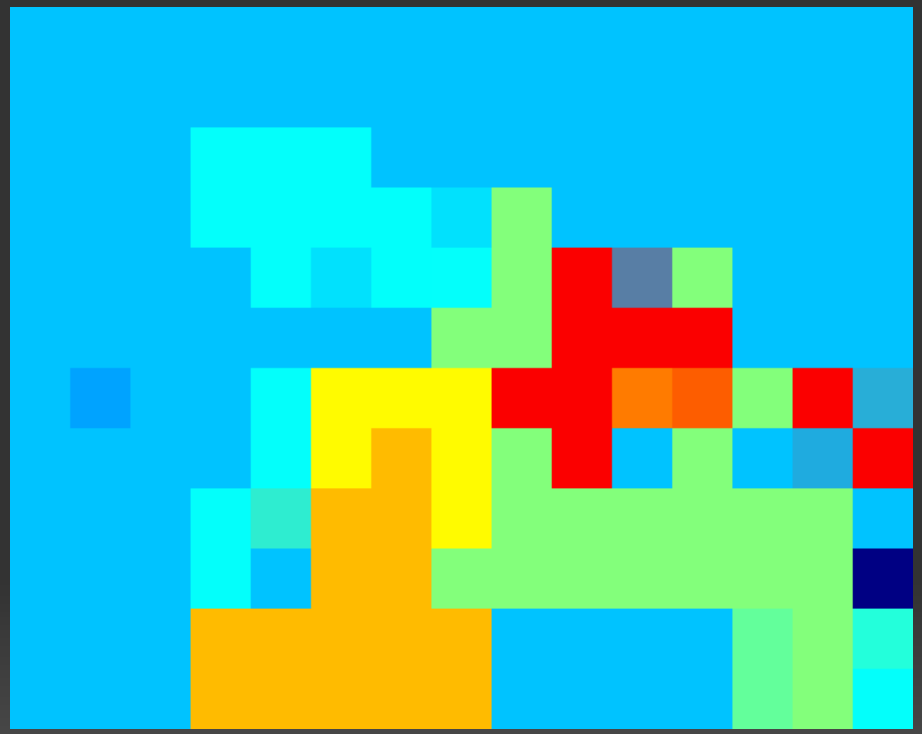


- Simple clustering algorithm used to determine regions
- Small problem regions are filtered out
- General sense of depth is achieved

http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/

Depth Map Conversion

- Need to scale down the depth map
- Audio resolution is not as high as visual
- Still possible to generally tell where objects are (in relation to one another)



Audio Representation

- Represent the world using headphones for 3D sound

- Rather difficult to demonstrate without headphones
- Directly useful only for a single dimension (horizontal)
- Need three dimensions to place a point in space:
  - Horizontal location
  - Vertical location
  - Distance

Audio Representation (Initial Idea)

- Play all points at the same time
- Horizontal location
  - Standard headphone placement
- Vertical location
  - Pitch, higher pitches along the top, lower on the bottom
- Distance
  - Volume represents distance (louder is closer)
- Drawback
  - Number of simultaneous sounds is limited by sound card

Audio Representation (Refined Idea)

- Basic representation is the same
  - Horizontal location
  - Vertical location
  - Distance
- Scan horizontally across the world
- Sounds somewhat similar to a sonar display
- Scanning speed can be adjusted based on the user's needs

Problems with AudioVision

Problems:

- Windows does not correctly recognize multiple identical webcams
- Linux video input drivers
- OpenCV's Python bindings are broken

Solutions:

- Try Linux
- Keep trying until one works
- Use C / C++

Problems with AudioVision

Problems:

- High resolution depth maps take too much time
- Sound cards limit simultaneous sources

Solutions:

- Reduce resolution before running depth map algorithm
- Scanning audio

Overview

- Why?
  - Use hearing to replace vision
- Required components
  - Stereo camera setup
  - Stereo vision algorithm
  - Depth map conversion
  - Audio representation
- Questions?