Computer Vision
22.10.2008

Bastian Leibe
RWTH Aachen
http://www.umic.rwth-aachen.de/multimedia
leibe@umic.rwth-aachen.de

Organization
- Lecturer
  - Prof. Bastian Leibe (leibe@umic.rwth-aachen.de)
- Language
  - English or German?
- Goals of the lecture
  - Develop methods that enable a machine to “understand” images and videos
  - Teach fundamental computer vision techniques
  - Show current research developments and real-world applications

Course Webpage & Mailing List
- Course webpage
  - http://www.umic.rwth-aachen.de/multimedia
    - Teaching → Computer Vision
    - Slides will be made available on the webpage
- Please subscribe to the lecture on the Campus system!
  - Important to get email announcements!
  - May also be needed to get a larger room.

Exercises and Supplementary Material
- Exercises
  - Typically 1 exercise sheet every 2 weeks
  - Matlab based exercises
  - Hands-on experience with the algorithms from the lecture.
  - Send me your solutions the night before the exercise class.
- Supplementary material
  - Research papers and book chapters
  - Will be provided on the webpage
  - Optional reading if you take the regular 3V+1Ü exam.
  - Mandatory reading if you need to take a V4 exam (Vertiefungslinie Diplom).

Textbooks
- No single textbook for the class.
- Basic material is covered in the following two books.
  - *Computer Vision – A Modern Approach* by D. Forsyth, J. Ponce
    - Prentice Hall, 2002
  - *Multiple View Geometry in Computer Vision* by R. Hartley, A. Zisserman
    - 2nd Ed., Cambridge Univ. Press, 2004
- Research papers will be given out for newer topics.
  - Suggested reading, not mandatory (exception: V4 exam).
How to Find Me

- Office 1: Informatik 8
  - Computer Science building, room 6302

- Office 2: UMIC building
  - UMIC Research Centre
  - Otto-Blumenthal-Strasse, room 112
  - (Address will change in December)

- Open door policy (until further notice)
  - If you have questions to the lecture, come to me.
  - Send me an email before to confirm a time slot.

Questions are welcome!

Topics of Today’s Lecture

- What is computer vision?
- What does it mean to see and how do we do it?
- How can we make this computational?

- First Topic: Image Formation
  - Details in Forsyth & Ponce, chapter 1.

Images and video are everywhere...

Cameras are all around us...

Why Computer Vision?

- Goal of Computer Vision
  - Enable a machine to “understand” images and videos

- Automatic understanding
  - Computing properties of the 3D world from visual data (measurement)
  - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (perception and interpretation)

Vision for Measurement

- Real-time stereo
- Structure from motion
- Multi-view stereo for community photo collections

Slide credits: Kristen Grauman
Vision for Perception, Interpretation

- Sky
- Amusement park
- Ferris wheel
- Lake Erie
- Tree
- Carousel
- Deck
- People waiting in line
- Ride
- Umbrellas
- Pedestrians
- Maxair
- Bench
- Tree
- Lake Erie
- People sitting on ride

Objects
- Activities
- Scenes
- Locations
- Text/writing
- Faces
- Gestures
- Motions
- Emotions...

Related Disciplines
- Artificial intelligence
- Graphics
- Machine learning
- Cognitive science
- Image processing
- Algorithms
- Algorithms
- Cognitive science
- Computer vision
- Artificial intelligence
- Graphics
- Machine learning
- Cognitive science
- Image processing
- Algorithms

Directions to Computer Vision
- Science
  - Foundations of perception. How do WE see?
- Engineering
  - How do we build systems that perceive the world?
- Many applications
  - Medical imaging, surveillance, entertainment, graphics, ...

Applications: Faces and Digital Cameras
- Setting camera focus via face detection
- Camera waits for everyone to smile to take a photo [Canon]
- Automatic lighting correction based on face detection

Applications: Vision for Mobile Phones
- Situated search Yeh et al., MIT
- MSR Lincoln
- kooaba
- ~30'000 movie posters indexed
- Query-by-image from mobile phone available in Germany and Switzerland

Demo: Movie Poster Recognition
- Commercial services coming out...
- 30'000 movie posters indexed
- Available in Germany and Switzerland

Source: http://www.kooaba.com
Applications: Vision-based Interfaces

- Games (Sony EyeToy)
- Assistive technology systems
  - Camera Mouse
  - Boston College

Applications: Medical & Neuroimaging

- Image guided surgery
  - MIT AI Vision Group

Applications: Medical & Neuroimaging

- FMRI data
  - Golland et al.

Applications: Visual Special Effects

- The Matrix
- What Dreams May Come

- MoCap for Pirates of the Caribbean, Industrial Light and Magic
  - Source: S. Seitz

Applications: Safety & Security

- Autonomous robots
- Driver assistance

- Monitoring pools
  - Poseidon

- Surveillance

Ok, Let’s Do It - Any Obstacles?

- 1966: Marvin Minsky directs an undergraduate student to solve "the problem of computer vision" as a summer project.

- Obviously, computer vision was too difficult for that...

Challenges: Many Nuisance Parameters

- Illumination
- Object pose
- Clutter

- Occlusions
- Intra-class appearance
- Viewpoint
Challenges: Intra-Category Variation

- Intra-Category Variation

Challenges: Complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991].

So, Should We Give Up?

- NO! Very active research area with exciting progress!

Course Outline

- Image Processing Basics
  - Segmentation
  - Local Features & Matching
  - Object Recognition and Categorization
  - 3D Reconstruction
  - Motion and Tracking

Course Outline

- Image Processing Basics
  - Segmentation
  - Local Features & Matching
  - Object Recognition and Categorization
  - 3D Reconstruction
  - Motion and Tracking
Course Outline
- Image Processing Basics
- Segmentation
- Local Features & Matching
  - Object Recognition and Categorization
  - 3D Reconstruction
  - Motion and Tracking

Course Outline
- Image Processing Basics
- Segmentation
- Local Features & Matching
  - Object Recognition and Categorization
  - 3D Reconstruction
  - Motion and Tracking

Topics of Today's Lecture
- What is computer vision?
- What does it mean to see and how do we do it?
- How can we make this computational?

First Topic: Image Formation
- Details in Forsyth & Ponce, chapter 1.

Camera Obscura
- Around 1519, Leonardo da Vinci (1452 - 1519)

“...When images of illuminated objects ... penetrate through a small hole into a very dark room ... you will see [on the opposite wall] these objects in their proper form and color, reduced in size ... in a reversed position owing to the intersection of the rays”
Camera Obscura

- Used by artists (e.g. Vermeer 17th century) and scientists

Jetty at Margate England, 1898.
An attraction in the late 19th century

Camera Obscura

Pinhole Camera

- (Simple) standard and abstract model today
  - Box with a small hole in it
  - Works in practice

Pinhole Size / Aperture

- Pinhole too big - many directions are averaged, blurring the image
- Pinhole too small - diffraction effects blur the image
- Generally, pinhole cameras are dark, because a very small set of rays from a particular point hits the screen.

The Reason for Lenses

- Keep the image in sharp focus while gathering light from a large area

The Thin Lens

\[ \frac{1}{z'} - \frac{1}{z} = \frac{1}{f} \]
**Focus and Depth of Field**

- Depth of field: distance between image planes where blur is tolerable

**Thin lens:** scene points at distinct depths come in focus at different image planes. (Real camera lens systems have greater depth of field.)

- How does the aperture affect the depth of field?

**A smaller aperture increases the range in which the object is approximately in focus**


Slide from S. Seitz

**Application: Depth from (De-)Focus**

- Images from same point of view, different camera parameters

3D Shape / depth estimates

*(figs from H. Jin and P. Favaro, 2002)*

Slide credit: Kristen Grauman

**Field of View**

- Angular measure of the portion of 3D space seen by the camera

**Field of View Depends on Focal Length**

- As $f$ gets smaller, image becomes more wide angle
  - More world points project onto the finite image plane

- As $f$ gets larger, image becomes more telescopic
  - Smaller part of the world projects onto the finite image plane

**Digital Images**

- Film is replaced by a sensor array
- Current technology: arrays of charge coupled devices (CCD)
- *Discretize* the image into pixels
- *Quantize* light intensities into pixel values.

Image source: Michael Black

Slide credit: Kristen Grauman
### Resolution

- **Sensor**: size of real world scene element that images to a single pixel.
- **Image**: number of pixels.
- Influences what analysis is feasible, affects best representation choice.

---

### Color Sensing in Digital Cameras

- **Bayer grid**: Estimate missing components from neighboring values (demosaicing).

---

### Grayscale Image

**Problem of Computer Vision**

- How can we recognize fruits from an array of (gray-scale) numbers?
- How can we perceive depth from an array of (gray-scale) numbers?
- How do we humans do it? How can we make a computer do it?

---

### Looking At What Artists Do

- **Visual Cues: Color & Lighting**

---

### Visual Cues: Perspective & Shading

- **Visual Cues: Perspective & Shading**

---
Computer Vision

- One way of looking at it: Take all the cues of artists and ‘turn them around’
- Exploit these cues to infer the structure of the world
- Need mathematical and computational models of these cues
- Sometimes called ‘inverse graphics’

One way of looking at it:

Take all the cues of artists and ‘turn them around’

Exploit these cues to infer the structure of the world

Need mathematical and computational models of these cues

Sometimes called ‘inverse graphics’

Is it more than inverse graphics?

- How do you recognize
  - The banana?
  - The glass?
  - The towel?

- Ill-posed problem: real world much more complex than what we can measure in images
  - 3D → 2D
- Impossible to literally “invert” image formation process

Scene Structure: Occlusion & Interposition

Magritte, 1957

Scene Structure: Relative Size

Magritte: The Listening Room

Scene Structure: Relative Position

Magritte: Personal Values
Many of Those Ideas Are Very Old

- Why didn’t computer vision work back in the 70’s?
- People didn’t lack the general ideas…
  - but they lacked
    - the algorithms
    - the data
    - and some reliable bottom-up signal
- Now we have
  - Better low level vision (segmentation, motion estimation, stereo, …)
  - Better learning algorithms
  - A lot more data and hardware and techniques to deal with it

Next Lectures

- First few lectures: low-level vision
  - Binary image processing
  - Filtering operations
  - Edge and structure extraction
  - Color
  - Segmentation and grouping
- Tomorrow: Binary image processing
- Wednesday 29.10.: Exercise 1
  - Intro Matlab, basic image operations

Questions?