Computer Vision – Lecture 1

Introduction

09.04.2019

Bastian Leibe
Visual Computing Institute
RWTH Aachen University
http://www.vision.rwth-aachen.de/
leibe@vision.rwth-aachen.de

Organization

• Lecturer
  Prof. Bastian Leibe (leibe@vision.rwth-aachen.de)

• Teaching Assistants
  Istvan Sarandi (sarandi@vision.rwth-aachen.de)
  Dan Jia (jia@vision.rwth-Aachen.de)

• Course webpage
  http://www.vision.rwth-aachen.de/courses/
  → Computer Vision
  Slides will be made available on the webpage
  There is also a moodle electronic repository

  • Please subscribe to the lecture on RWTH Online!
    Important to get email announcements and moodle access!

Language

• Official course language will be English
  → If at least one English-speaking student is present.
  → If not… you can choose.

  However...
  → Please tell me when I’m talking too fast or when I should repeat
    something in German for better understanding!
  → You may at any time ask questions in German!
  → You may turn in your exercises in German.
  → You may answer exam questions in German.

Exercises and Demos

• Exercises
  → Typically 1 exercise sheet every 2 weeks (numpy/TensorFlow)
  → Hands-on experience with the algorithms from the lecture.
  → Send in your solutions the night before the exercise class.
  → No admission requirement to qualify for the exam this year!

  • Teams are encouraged!
    → You can form teams of up to 3 people for the exercises.
    → Each team should only turn in one solution.
    → But list the names of all team members in the submission.

Course Webpage

**Course Schedule**

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Content</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon, 2019-04-01</td>
<td>–</td>
<td>No class (RWTH Dies)</td>
<td></td>
</tr>
<tr>
<td>Tue, 2019-04-02</td>
<td>–</td>
<td>No class (RWTH Dies)</td>
<td></td>
</tr>
<tr>
<td>Thu, 2019-04-04</td>
<td>Introduction</td>
<td>Why video? Applications, Challenges, Image Formation</td>
<td></td>
</tr>
<tr>
<td>Mon, 2019-04-15</td>
<td>Image Processing</td>
<td>Linear filters, Gaussian smoothing, Multi-scale decompositions</td>
<td></td>
</tr>
<tr>
<td>Tue, 2019-04-16</td>
<td>–</td>
<td>TBD</td>
<td></td>
</tr>
<tr>
<td>Wed, 2019-04-17</td>
<td>–</td>
<td>TBD</td>
<td></td>
</tr>
<tr>
<td>Thu, 2019-04-18</td>
<td>Image Processing III</td>
<td>Image Derivatives, Edge detection, Contours</td>
<td></td>
</tr>
<tr>
<td>Mon, 2019-04-22</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Wed, 2019-04-24</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Thu, 2019-04-25</td>
<td>Structure</td>
<td>Linear Filtering, High/low Transform, Gen, High Transform</td>
<td></td>
</tr>
<tr>
<td>Mon, 2019-04-29</td>
<td>Segmentation II</td>
<td>Segmentation as Clustering, k-means, EM, Mean-shift segmentation</td>
<td></td>
</tr>
<tr>
<td>Thu, 2019-05-02</td>
<td>Exercise 1</td>
<td>Derivation, Edge, High Transform</td>
<td></td>
</tr>
<tr>
<td>Thu, 2019-05-07</td>
<td>Segmentation</td>
<td>Segmentation as Energy Minimization, Markov Random Fields, Graph Cuts</td>
<td></td>
</tr>
<tr>
<td>Mon, 2019-05-13</td>
<td>Classification</td>
<td>Editing Window-based Object Detection</td>
<td></td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td></td>
</tr>
</tbody>
</table>

http://www.vision.rwth-aachen.de/courses/
Textbooks

- No single textbook for the class.
- Basic material is covered in the following two books.
  - R. Hartley, A. Zisserman. Multiple View Geometry in Computer Vision. 2nd Ed., Cambridge Univ. Press, 2004
  (available in the library’s “Handapparat”)
- Additional material will be given out for some topics.
  - Tutorials and deeper introductions.
  - Application papers

How to Find Us

- Office:
  - UMIC Research Centre
  - Mies-van-der-Rohe-Strasse 15, room 124
- Office hours
  - If you have questions to the lecture, come to us.
  - Send us 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.

Why Computer Vision?

Cameras are all around us...

Images and video are everywhere...

- Personal photo albums
- Movies, news, sports
- Internet services
- Surveillance and security
- Mobile and consumer applications
- Medical and scientific images

What is 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

Vision for Perception, Interpretation

Objects

Activities

Scenes

Locations

Text / writing

Faces

Gestures

Motions

Emotions...

Related Disciplines

Artificial intelligence

Machine learning

Cognitive science

Computer vision

Graphics

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

Segmentation

• Automatic background removal from images
  > Functionality is included in Microsoft Office 2010…
Matching

- Stitch your photos together to create panoramas

Applications: Vision for Mobile Phones

- Take photos of objects as queries for visual search

Applications: Vision-based Interfaces

Games (Microsoft Kinect)

Assistive technology systems
Camera Mouse
Boston College

Applications: Medical & Neuroimaging

Image guided surgery
MIT AI Vision Group

Applications: Visual Special Effects

The Matrix

MoCap for Pirates of the Caribbean, Industrial Light and Magic (source: S. Seitz)

Applications: Safety & Security

Autonomous robots
Driver assistance
Monitoring pools (Poseidon)

Pedestrian detection [MERL, Viola et al.]

Surveillance
Ok, Let's Do It – Any Obstacles?

- 1966: Seymour Papert 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

- 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!

Things Are Starting to Work...

Computer Vision in realistic scenarios is becoming feasible!
Course Outline

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

And you might have heard of...

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

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.

Slide credit: Bernt Schiele
Source: http://www.acmi.net.au/AIC/CAMERA_OBSCURA.html

Jetty at Margate England, 1898.
An attraction in the late 19th century
Adapted from R. Duraiswami
http://brightbytes.com/cosite/collection2.html
The Reason for Lenses

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

Source: Forsyth & Ponce

The Thin Lens

\[ \frac{1}{z'} - \frac{1}{z} = \frac{1}{f} \]

Source: Forsyth & Ponce

Focus and Depth of Field

Thin lens: scene points at distinct depths come in focus at different image planes.

(Real camera lens systems have greater depth of field.)

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

Source: Shapiro & Stockman

Focus and 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


Field of View

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


Application: Depth from (De-)Focus

Images from same point of view, different camera parameters

3D Shape / depth estimates

Slide credit: Kristen Grauman

Slide credit: Kristen Grauman
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.

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

- 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?

Next Lectures

- First few lectures: low-level vision
  - Filtering operations
  - Edge and structure extraction
  - Segmentation and grouping
- Next week: Linear Filters
Questions?