Course Introduction

Lecture 01

COMPSCI 775
Lecturer: Professor Reinhard Klette
Tutors: Dongwei Liu and Haokun Geng

Semester 2, 2014, Tamaki campus

1See last slide for copyright information.
Contact Addresses and Data Recording

Reinhard KLETTE (r.klette@auckland.ac.nz), 723.319
Dongwei LIU (dliu697@aucklanduni.ac.nz), 723.321
Haokun GENG (hgen001@aucklanduni.ac.nz), 723.321

In the second half of the semester you record – under the supervision of one or two of the tutors – your own stereo sequences with trinocular cameras installed in HAKA1, the test vehicle in the enpeda.. project.

You do not need a driver’s licence for taking this course.
Agenda

1. Course Outline
2. Computer Vision
3. Computational Photography
4. Semantic Image Segmentation
5. Stereo Vision and Visual Odometry
6. Summary and Computer Science, Tamaki Campus
Course Abstract

COMPSCI 775 is an advanced course on multimedia imaging, presenting subjects in computer vision (image segmentation, camera calibration, binocular stereo, motion analysis, image features, tracking, object detection, and so forth). This 2014 course has a particular focus on

1. *computational photography*,
2. *semantic segmentation*, and
3. *stereo vision and visual odometry*.

The online lecture notes cover the material taught in the lectures.

There are 36 meetings (mostly lectures, but also a few seminars), and also a few tutorials (e.g. for introducing into OpenCV and the used test vehicle, called HAKA1)
Web Site, Course Work, and Marking

The 775 course web site is

www.cs.auckland.ac.nz/~rklette/TeachAuckland.html/775/

with possible updates from time to time.

The three assignments are on computational photography, semantic segmentation, and stereo vision and visual odometry.

You have to submit *three* assignments (individual work) which count 30% towards the final mark. Assignments are in C/C++ using OpenCV in Visual Studio. For each assignment, submit a report (pdf format, optional: Latex template provided) and your sources.

The exam (not multiple-choice; only about subjects as taught in the 775 lectures; closed book) counts 70% towards the final mark.
Available Sources and Team Work

You can use any available material or sources (in publications, on the net, etc.), including advice by your mates, **but** with proper references or acknowledgments, for not risking any plagiarism accusation.

Using sources by others requires proper understanding and testing. OpenCV does not come with a warranty.

Team work is encouraged, but your assignments have to be done just by you. You may contact .enpeda.. staff or students if you wish to discuss your course work (with prior confirmation of appointments via email).
General Hints for Assignments

Working with real-world images or videos, recorded under various lighting conditions, defines in general a high level of complexity. Computer vision is a difficult subject. Don’t expect to have excellent results easily.

Comparative performance evaluation is a good way to understand drawbacks or positive features of your implemented technique.

A solution is “COMPSCI 775-real time” if it allows you to test it on various input data without losing patience.
Agenda

1. Course Outline
2. Computer Vision
3. Computational Photography
4. Semantic Image Segmentation
5. Stereo Vision and Visual Odometry
6. Summary and Computer Science, Tamaki Campus
Computer Vision

*Computer vision* aims at understanding or modelling static or dynamic 3D objects or scenes based on captured image data.

The task is related to projective geometry (how 3D objects are mapped into one or several 2D images, possibly from different perspectives). Computer vision combines (static) image analysis with image sequence (i.e. video) analysis.

Image sequences are captured by normal video cameras or more advanced specialised cameras and processed in a computer. Computer vision studies methodological and algorithmic problems as well as topics related to the implementation of designed solution.
Origins in Early 1960s

First Textbook

First Journal
Founded by A. Rosenfeld, after 1982 continued as *Computer Vision, Graphics, and Image Processing*
From Image Analysis to Computer Vision

Lawrence (Larry) Roberts finished in 1962 a PhD thesis at the Massachusetts Institute of Technology (MIT) about ”blocks world” analysis based on images. His contribution was a first systematic study of the problem “How to understand 3D geometry based on perspective views?” (such as recorded by a camera), and basically the start of computer vision. See packet.cc/files/mach-per-3D-solids.html for a publication which resulted from his PhD.

The human visual system is “too good” for some tasks (e.g., object recognition); accordingly, the complexity of the vision process is often underestimated. Many researchers were very optimistic in the 1960s about creating a “seeing machine” (as one of the components of the new discipline of Artificial Intelligence); shortly after, difficulties became obvious. Even simple tasks such as defining and detecting an “edge” in an image triggered long-lasting research projects.
Today

Since the late 1990s, we experience a massive growth of computer vision applications, wherever cameras appear to be applicable.

Examples are industrial surface inspection, surveillance systems, face recognition (e.g., introduced at EU border crossings in 2008, and at the Auckland airport in 2010), 3D face modelling (e.g., for creating 3D face models for fun in crystal cubes, since about 2000), cameras in children toys with scene analysis abilities, or modelling 3D spaces or buildings using aerial views and stereo vision.

See also solutions for automatically inspecting the “blind spot” of a side mirror of a car (e.g., by Volvo in 2004), traffic sign recognition, and so forth, up to autonomous driving at low speed on public roads (e.g., Mercedes in April 2013), or fully autonomous driving on selected roads (e.g. Mercedes in September 2013).
First Issues of Computer Vision Conferences

**Int. Conf. Pattern Recognition (ICPR)**

**IEEE Conf. Computer Vision Pattern Recognition (CVPR)**
Arlington 1983, chairs: T. Kanade and D. Ballard

**Int. Conf. Computer Analysis Images Patterns (CAIP)**
Berlin 1985, chair: R. Klette

**Int. Conf. Computer Vision (ICCV)**
London 1987, chairs: J. M. Brady and A. Rosenfeld

**European Conf. Computer Vision (ECCV)**
Antibes 1990, chair: O. Faugeras

**Asian Conf. Computer Vision (ACCV)**
Osaka 1993, chair: K. Sugihara

**Pacific-Rim Symp. Image Video Technology (PSIVT)**
Hsinchu 2006, chairs: L.-W. Chang and R. Klette
Hundreds of Books on Computer Vision Subjects

Reminder: “A book is a set of written, printed, illustrated, or blank sheets, made of ink, paper, parchment, or other materials, usually fastened together to hinge at one side.” [Wikipedia, 11 July 2014]

You can find books in the library of our university, or you may purchase some, e.g. online:
Example: Computer Vision for Driver Assistance 2011

Driver monitoring, shown by PhD student Mahdi Rezaei, to S&T Minister of China, Dr. WAN Gang, on 28 March 2011

“Every truck and long-distance bus should be equipped with this technology.”
Correlate driver’s attention to road and traffic conditions by analyzing both simultaneously (in real-time, monocular recording only)

Use of a fuzzy-logic inference system for detecting configurations which require a change in the driver’s attention

[M. Rezaei and R. Klette: Look at the Driver, Look at the Road: No Distraction! No Accident! In Proc. CVPR, 2014 (full text also available on ResearchGate)]
Agenda

1. Course Outline
2. Computer Vision
3. Computational Photography
4. Semantic Image Segmentation
5. Stereo Vision and Visual Odometry
6. Summary and Computer Science, Tamaki Campus
Photo-Fun in the late 19th Century
Face Morphing and an Average Face, around 1900

[F. Naumann. Im Reiche der Kamera. Ed. Liesegangs Verlag, Leipzig 1912]
Computational Photography

“Computational photography refers to computational image capture, processing, and manipulation techniques that enhance or extend the capabilities of digital photography. Good examples of image capture and manipulation include panoramas, and high-dynamic-range imaging which is the use of differently exposed pictures of the same scene to extend dynamic range beyond even that of analog film-based media.” [Wikipedia, 11 July 2014]

Computational photography may be used for

1. removing noise,
2. inpainting (at missing pixels),
3. creating particular feelings,
4. enhancing image features,

and so forth.
Artistic Filters

**Example:** Emulating the painting style of Vincent van Gogh

Image on the right shows only one of three layers (each layer is an intermediate result only, for dots of a particular size), to be merged into a final result.

Adding Fog to a Photo based on Distance Analysis, 2014

A filter designed by PhD student Dongwei Liu:

Upper row: Original photos. Lower row: Fog added using different settings, all based on distance analysis [work is submitted to a journal].
Agenda

1. Course Outline
2. Computer Vision
3. Computational Photography
4. Semantic Image Segmentation
5. Stereo Vision and Visual Odometry
6. Summary and Computer Science, Tamaki Campus
Segmentation

In *segmentation* we partition a given image into a family of pairwise disjoint segments (each is a connected set of pixels, called a *region*).

The shown original photo (left) is mapped into colour-labelled segments (right). But these segments do not yet correspond to different *categories* such as *people, lawn, trees, buildings, or concreted ground*. 
Black Connected or White Connected?

Connectedness of pixels depends on the chosen model:

The topology and geometry of sets of pixels is studied in the field of *digital geometry*. This course discusses geometric properties of segments (as a step towards image analysis or pattern recognition).

Semantic segmentation needs to achieve exactly that: segments should correspond to defined categories of visible objects or scene components.
These manually defined ground-truth examples for semantic segmentation are from Set 7 of EISATS ccv.wordpress.fos.auckland.ac.nz/eisats/.
They are originally published in [A. Barth, J. Siegemund, A. Meissner, U. Franke, and W. Förstner, in Proc. DAGM, 2010].
Agenda

1. Course Outline
2. Computer Vision
3. Computational Photography
4. Semantic Image Segmentation
5. Stereo Vision and Visual Odometry
6. Summary and Computer Science, Tamaki Campus
Stereo Analysis

How far away is a particular object?

Throw a ball with known constant speed straight into the scene; assume that it bounces straight back to you from an obstacle surface, again with the known constant speed. The time between throwing and returning, and the known speed gives you the distance to the obstacle surface.

Replace “ball” by “impulse of light”, then we have the principle of a laser range-scanner. The mathematical formula for distance calculation is trivial; the implementation is difficult.

The speed of light requires that the impulse has some measurable length. By modulating a wave onto the impulse we may understand even better the distance to the obstacle.
Computational Stereo Vision

Now we look with two cameras into the scene. A point on the obstacle surface, visible in both recorded (time-synchronised) images, allows us to understand the distance using a calculation based on triangulation.

Computational stereo matching is used for identifying *corresponding pixels* (i.e. projections of the same surface point) in both recorded images.

SGM stereo matching implemented by John Lin (PhD student at UoA) and Dr. Stefan Gehrig in 2006 at Daimler AG. Colours visualise distances.
Dense Distance Maps

Black pixels: Occluded or low confidence in results at those pixels

Implementation of modified SGM (semi-global matcher) by PhD student Simon Hermann, see [S. Hermann, S. Morales, and R. Klette, in Proc. IEEE Conf. Intelligent Vehicles (IV), Baden-Baden, 2011]
Robust Vision Challenge 2012

Robust Vision Challenge

in Association with the 2012 ECCV Workshop on Unsolved Problems in Optical Flow and Stereo Estimation

Video cameras provide information on a scene with low cost in acquisition, space and energy, and at the same time high spatio-temporal resolution. To extract depth information from a video computer vision algorithms make strong assumptions on a scene. The algorithms are thus easily distracted by phenomena that violate these assumptions, such as reflecting or transparent surfaces, lens flares, and changing illuminations. We recorded multiple real world scenes that contain instances of challenging phenomena such as they occur in everyday traffic scenarios. To be applicable in real-time, image-based depth and motion estimation needs to deal robustly and reliably with these scenes.

We pose the estimation of depth and motion on our recorded sequences as a challenge to the scientific community. Can you develop an stereo or optical flow algorithm that can deal with these sequences?

For the evaluation of participating algorithms we bring together a jury of renowned experts on the application of stereo and motion estimation. In the absence of ground information, the jury will thoroughly inspect and evaluate the submitted correspondences and their applicability in industry. A prize for the best-performing algorithm is awarded by Bosch.

The Winner

The winner of the challenge where Simon Hermann and Reinhard Klette with a SGM variant. Please find details in our ECCV Workshop Winner Announcement!

The Task

Estimate robust and reliable depth or motion fields on our challenging real world videos!

Download image sequences
The sequences contain many examples for regularly occurring situations causing problems with most methods known today. We are looking for algorithms that can produce reliable depth or motion estimates for all images, including the indicated keyframes. Participation is open to all ideas that improve the state-of-the-art in automatic motion and depth estimation on the given input videos.

This could involve (but is not limited to):
- Previously unknown methods of correspondence estimation
- Correspondence estimation making use of confidence measures
- Correspondence estimation with a detection of unreliable input images or unreliable image regions
For a Video See ...

ccv.wordpress.fos.auckland.ac.nz/data/binocular-sequences/

Here is the certificate:

2012 Robust Vision Challenge

The Winner is ...

iSGM & fSGM
by
Simon Hermann
and
Reinhard Klette
Odometry

Where am I?

A human uses inertial navigation (basically the ears) and the visual system (starting with the eyes) for sensing motion, i.e. for estimating pose changes over time. *Pose* is location and direction.

The vestibular system in the inner ear is typically considered to be the primary source of input to the brain about pose changes of the body. The visual system provides further data, e.g. understanding motions with respect to visible landmarks. We need both for odometry, ears and eyes.

*Robotics* replaces the ears by an *inertial moment unit* (IMU), and the eyes by one or multiple cameras, both connected to a processor. NASA’s Mars Exploration Rover Mission (MER) is, for example, a proof that this concept works.
Visual Odometry

= use of perceived or recorded images for understanding pose changes (e.g. of a robot, a mobile phone, or a quadcopter)

Identify landmarks (i.e. image features) in subsequently recorded images, analyse for correspondence, and calculate pose changes with respect to those identified landmarks.

If possible, combine with processing IMU data. For example, many smartphones have not only one or multiple cameras, they also have a 3-axis accelerometer, 3-axis gyroscope, and a 3-axis magnetometer.

For visual odometry of people, see [J. L. Souman, I. Frissen, M. N. Sreenivasa, and M. O. Ernst: Walking straight into circles. Current Biology, 19:1538–1542, 2009]: “we tested the ability of humans to walk on a straight course through unfamiliar terrain”
Agenda

1. Course Outline
2. Computer Vision
3. Computational Photography
4. Semantic Image Segmentation
5. Stereo Vision and Visual Odometry
6. Summary and Computer Science, Tamaki Campus
Summary

In this course you will learn about

1. image and video processing, with an introduction into OpenCV,
2. concepts for image segmentation, and a general framework for pixel labelling (and error minimisation) in this context,
3. geometric analysis of segments (regions in an image),
4. setting up a camera system for stereo vision, and about three different ways how to perform stereo matching,
5. recording of trinocular stereo sequences with HAKA1, the test vehicle of the enpeda project (a 2007 A-class Mercedes),
6. motion analysis (optic flow), object detection, and object tracking, and
7. two different ways to perform visual odometry.

During the course you may also experience to present your assignment results to the class (this is not part of the course marking).
CS Tamaki Campus – ACCV 2010 and the Tamaki Group

November 2010: *enpeda..* with HAKA1 at ACCV in Queenstown

March 2011: CS Tamaki, staff, students, and a visitor from Finland
CS Tamaki is Hosting International Meetings

January 2014: *Joint Korea - New Zealand Workshop on Multimedia Imaging*, with Profs. In So Kweon and Chilwoo Lee (Korea) and Zezhung Xu (China)

Copyright Information

This slide show was prepared by Reinhard Klette with kind permission from

- Mahdi Rezaei,
- Crystal Valente,
- Dongwei Liu,
- Sandino Morales,
- John Lin and Stefan Gehrig,
- Simon Hermann, and
- Springer Science+Business Media B.V.

The slide show can be used freely for presentations. However, *all the material* is copyrighted.