Computer Vision 2: Advanced Topics in Computer Vision

Theo Gevers, Jan-Mark Geusebroek, Sezer Karaoglu, Andrew Brown, Minh Ngo, Jian Han, Wei Zeng, Partha Das, Yunlu Chen, Shuo Chen, Jiaojiao Zhao

Course Organization

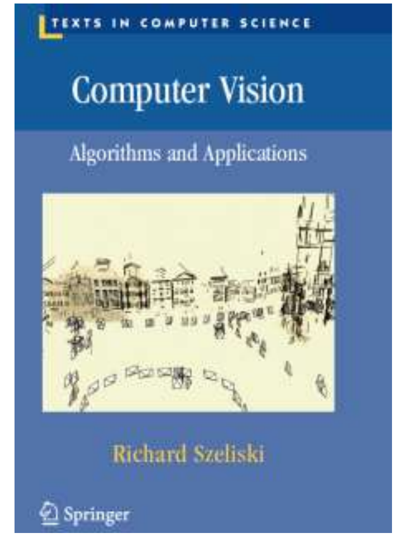
- Lectures by invited speakers (30%)
  - Lectures (3D reconstruction theory) and invited speakers
  - Read the book
- Paper presentation students (30%)
  - Read the papers
  - Submit 3 questions
  - One Presentation (20 min. + 5 min. Q&A)
- Practice (40%)
  - 3D reconstruction project

Schedule Lectures and Speakers

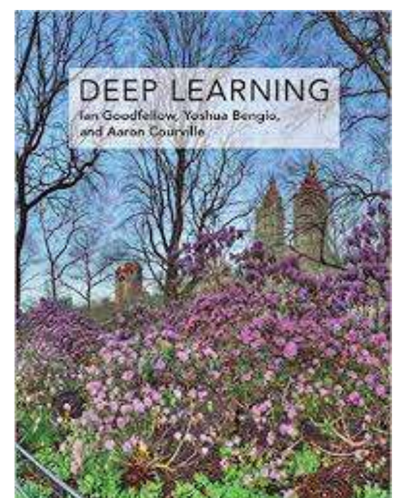
- **This week:** Introduction / paper-student assignment.
- **Week 2 (April 8):** Dr. Jan van Gemert, multi-view stereo & SFM (**required** for exam).
- **Week 3 (April 15):** Dr. Zeynep Akata (**not required** for exam).
- **Week 4 (April 22):** Easter.
- **Week 5 (April 29):** Dr. Pascal Mettes (**not required** for exam).
- **Week 6 (May 6):** Dr. Albert Ali Salah (**not required** for exam).
- **Week 7 (May 13):** TBA.

Textbooks

- CV2 is based on "*Computer Vision: Algorithms and Applications*" by Richard Szeliski
  - Freely available for download from http://szeliski.org/Book/



- and "*Deep Learning*" by Ian Goodfellow, Yoshua Bengio and Aaron Courville
  - Freely available for download from http://www.deeplearningbook.org/

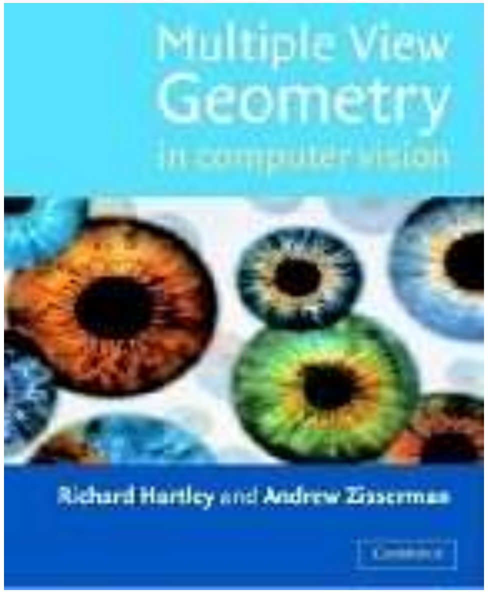


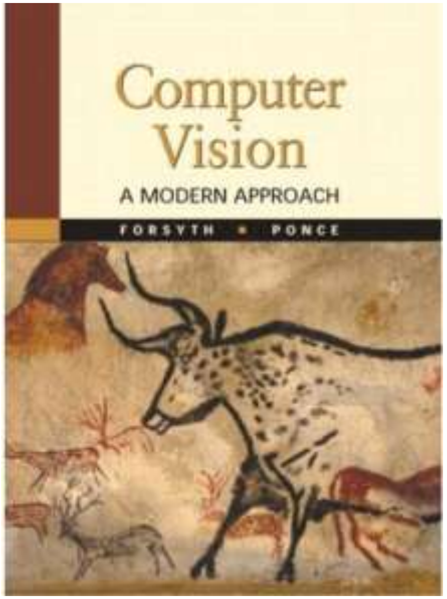
Background

- Two other useful books

  Forsyth, David A., and Ponce, J. Computer Vision: A Modern Approach, Prentice Hall, 2003.

  Hartley, R. and Zisserman, A. Multiple View Geometry in Computer Vision, Academic Press, 2002.





Exam (30%)

- **50%** of the questions will cover the **basic theory on 3D reconstruction**. The slides, lab assignment and book will form the basis for the (closed book) exam.
- **Focus on chapters**: 4, 6.1, 7-7.7, 11.1-11.4, 12.1, 12.2 of the book (http://szeliski.org/Book/)
- The other **50%** of the questions will be **related to the papers** presented during the course.

Course Organization

- Lectures with invited speakers: Monday's
  - Lectures (3D reconstruction theory) and invited speakers
  - Read the book
- Paper presentation students:
  - Read the papers
  - Submit 3 questions
  - One Presentation
- Practice:
  - 3D reconstruction project

Presentation + Questions (30%)

Each person has to **1)** Present one paper in-depth and **2)** Think of 3 questions for each paper.

Presentation

Present the selected paper for 30 minutes (#students 2-3)

Put the emphasis on the Computer Vision parts, make sure you include:

- the main contribution of the paper,
- the research that this paper builds upon,
- the research that builds upon this paper and competing methods (use Google scholar)
- strengths, weaknesses, improvements and discussion points

Create questions

I do not want to force a (time consuming) full paper review on you, so, I ask you for some good questions to show that you (at least at a high-level) have read the paper. The questions have to be submitted 1 hour before the presentation. We will grade them Pass or Fail. If you do not submit or Fail, I'll ask you to do a review in compensation.

Paper Themes

Perception

Imaging

Features, Segmentation and Detection

Tracking and Flow

Recognition and Detection

3D Reconstruction and Recognition

Faces, People and Crowds

Deep Learning

Papers per Theme

Perception

The Plenoptic Function and the Elements of Early Vision, Adelson and Bergen, 1991:

http://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/adelson-elements-91.pdf

On Seeing Stuff: The Perception of Materials by Humans and Machines, Adelson, 2001:

http://web.mit.edu/persci/people/adelson/pub_pdfs/adelson_spie_01.pdf

Throwing Down the Visual Intelligence Gauntlet in Machine Learning for Computer Vision, Tan, 2012:

C., J.Z. Leibo, and T. Poggio, 2012.

http://cbcl.mit.edu/publications/ps/MIT-CSAIL-TR-2012-016.pdf

Papers per Theme

Imaging

- Fast End-to-End Trainable Guided Filter, CVPR 2018: http://wuhuikai.me/DeepGuidedFilterProject/deep_guided_filter.pdf
- CNN based Learning using Reflection and Retinex Models for Intrinsic Image Decomposition by A. S. Baslamisli, H. A. Le, T. Gevers, 2018: https://arxiv.org/pdf/1712.01056.pdf
- Learning Non-Lambertian Object Intrinsics across ShapeNet Categories by J. Shi, Y. Dong, H. Su, S. X. Yu, 2017: https://arxiv.org/pdf/1612.08510.pdf
- Revisiting Deep Intrinsic Image Decompositions by Q. Fan, D. Wipf, J. Yang, G. Hua, B. Chen, 2017: https://arxiv.org/pdf/1701.02965.pdf
- Accidental Pinhole and Pinspeck Cameras: Revealing the Scene Outside the Picture, Torralba, 2012: http://people.csail.mit.edu/torralba/publications/shadows.pdf
- Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance, Gehler, 2011: http://people.tuebingen.mpg.de/mkiefel/projects/intrinsic/nips11intrinsic.pdf

Papers per Theme

Features, Segmentation and Detection (1)

- DeepLab.v3: Rethinking Atrous Convolution for Semantic Image Segmentation, 2017: https://arxiv.org/pdf/1706.05587.pdf
- Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation, 2018: https://arxiv.org/pdf/1801.04381.pdf
- SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation by V. Badrinarayanan, A. Kendall, R. Cipolla, 2017 https://arxiv.org/pdf/1511.00561.pdf
- Fully Convolutional Networks for Semantic Segmentation by J. Long, E. Shelhamer, T. Darrell, 2015: https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf
- DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs by L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, 2016: https://arxiv.org/pdf/1606.00915.pdf
- Constrained Convolutional Neural Networks for Weakly Supervised Segmentation by Pathak, Krahenbuhl, Darrell, 2015: http://www.robots.ox.ac.uk/~vgg/rg/papers/ccnn.pdf

Papers per Theme

Features, Segmentation and Detection (2)

- CPMC: Automatic Object Segmentation Using Constrained Parametric MinCuts, Carreira and Sminchisescu, 2010: http://www.maths.lth.se/matematiklth/personal/sminchis/papers/cs-cvpr10.pdf
- HOGgles: Visualizing Object Detection Features, Vondrick et. al., 2013: http://people.csail.mit.edu/khosla/papers/iccv2013_ihog.pdf
- Diagnosing Error in Object Detectors, Hoiem, 2012: http://dhoiem.cs.illinois.edu/publications/eccv2012_detanalysis_derek.pdf
- Regionlets for Generic Object Detection, Wang et. al., 2013: http://www.ece.northwestern.edu/~mya671/mypapers/ICCV13_Wang_Yang_Zhu_Lin.pdf
- Scale and Affine Invariant Interest Point Detectors. Mikolajczyk and Schmid, 2004: http://www.robots.ox.ac.uk/~vgg/research/affine/det_eval_files/mikolajczyk_ijcv2004.pdf
- On SIFTs and Their Scales, Hassner, 2012: http://www.openu.ac.il/home/hassner/projects/siftscales/OnSiftsAndTheirScales-CVPR12.pdf
- Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics, Kendall et al., 2018: https://arxiv.org/pdf/1705.07115v3.pdf

Papers per Theme

Tracking and Flow (1)

- DensePose: Dense Human Pose Estimation in the Wild, 2018: https://arxiv.org/pdf/1802.00434.pdf
- FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks by E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, 2017: https://lmb.informatik.uni-freiburg.de/Publications/2017/IMSKDB17/paper-FlowNet_2_0__CVPR.pdf
- Tracking by Natural Language Specification, Zhenyang Li, Ran Tao, Efstratios Gavves, Cees G. M. Snoek, Arnold W.M. Smeulders, 2017: http://openaccess.thecvf.com/content_cvpr_2017/papers/Li_Tracking_by_Natural_CVPR_2017_paper.pdf
- Detect to Track and Track to Detect, Christoph Feichtenhofer, Axel Pinz, Andrew Zisserman, 2017: http://openaccess.thecvf.com/content_ICCV_2017/papers/Feichtenhofer_Detect_to_Track_ICCV_2017_paper.pdf
- Scene Chronology, Matzen and Snavely, 2014: http://www.cs.cornell.edu/projects/chronology/
- Tracking-Learning-Detection, Kalal, Mikolajczyk, and Matas, 2010: http://epubs.surrey.ac.uk/713800/1/Kalal-PAMI-2011(1).pdf

Papers per Theme

Tracking and Flow (2)

- Track to the Future: Spatiotemporal Video Segmentation with Long-Range Motion Cues, Lezema, 2011: http://www.di.ens.fr/~josef/publications/lezama11.pdf
- On Space-Time Interest Points, Ivan Laptev, 2005: http://www.irisa.fr/vista/Papers/2005_ijcv_laptev.pdf
- Event Retrieval in Large Video Collections with Circulant Temporal Encoding, Revaud et. al., 2013: http://hal.archives-ouvertes.fr/docs/00/80/17/14/PDF/revaud_event.pdf
- DeepFlow: Large Displacement Optical Flow with Deep Matching, Weinzaepfel et. al., 2013: http://hal.archives-ouvertes.fr/docs/00/87/35/92/PDF/DeepFlow_iccv2013.pdf
- InterpoNet, A brain inspired neural network for optical flow dense interpolation, 2017, Zweig and Wolf: http://arxiv.org/pdf/1611.09803v3.pdf
- FlowNet: Learning Optical Flow with Convolutional Networks, Dosovitskiy et al., 2015: http://lmb.informatik.uni-freiburg.de/Publications/2015/DFIB15/flownet.pdf

Papers per Theme

Tracking and Flow (3)

- Optical Flow Requires Multiple Strategies (but only one network), Schuster, Wolf and Gadot, 2017: http://arxiv.org/pdf/1611.05607v3.pdf
- Optical Flow Estimation using a Spatial Pyramid Network, Ranjan and Black, 2016: http://arxiv.org/pdf/1611.00850v2.pdf
- Dense Correspondence Fields forHighly Accurate Large Displacement Optical Flow Estimation, Bailer et al., Flow Fields, 2018: https://arxiv.org/pdf/1703.02563v2.pdf
- Learning monocular visual odometry with dense 3D mapping from dense 3D flow, - Zhao et al., 2018: https://arxiv.org/pdf/1803.02286v2.pdf,

Papers per Theme

Recognition and Detection (1)

- What Makes for Effective Detection Proposals?, J. Hosang et. al., 2015: http://arxiv.org/pdf/1502.05082.pdf
- Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Girshick, Donahue, Darrell, Malik, 2014: http://arxiv.org/pdf/1311.2524v5.pdf
- Matching Local Self-Similarities across Images and Videos, Shechtman, Irani, 2007: http://www.wisdom.weizmann.ac.il/~vision/VideoAnalysis/Demos/SelfSimilarities/SelfSimilarities_ShechtmanIrani07.pdf
- What is an Object?, Alexe, Deselaers, Ferrari, 2010: ftp://ftp.vision.ee.ethz.ch/publications/proceedings/eth_biwi_00728.pdf
- Object Detectors Emerge in Deep Scene CNNs, Bolei et. al, 2015: http://www.robots.ox.ac.uk/~vgg/rg/papers/zhou_iclr15.pdf
- Cascade Object Detection with Deformable Part Models, Felzenszwalb, 2010: http://cs.brown.edu/~pff/papers/cascade.pdf

Papers per Theme

Recognition and Detection (2)

- You only look once: Unified, real-time object detection, Redmon et. al., 2016: http://pjreddie.com/media/files/papers/yolo.pdf
- The Devil is in the Details: an Evaluation of Recent Feature Encoding Methods, Chatfield, 2011: http://www.robots.ox.ac.uk/~vgg/research/encoding_eval/
- Aggregating Local Image Descriptors into Compact Codes. Jégou, 2012: http://hal.inria.fr/docs/00/63/30/13/PDF/jegou_aggregate.pdf
- BING: Binarized Normed Gradients for Objectness Estimation at 300fps, Cheng et. al., 2014: http://mmcheng.net/mftp/Papers/ObjectnessBING.pdf
- In Defense of Nearest Neighbor Based Image Classification, Boiman, 2008: http://www.wisdom.weizmann.ac.il/~irani/PAPERS/InDefenceOfNN_CVPR08.pdf
- YOLO9000: Better, Faster, Stronger, Redmon and Farhadi, 2016: http://arxiv.org/pdf/1612.08242v1.pdf

Papers per Theme

Recognition and Detection (3)

- Mask R-CNN, He et al., 2017: http://arxiv.org/pdf/1703.06870v1.pdf
- SSD: Single Shot MultiBox Detector, Liu et al, 2016: http://arxiv.org/pdf/1512.02325v5.pdf
- Speed/accuracy trade-offs for modern convolutional object detectors, Huang et al., 2016: http://arxiv.org/pdf/1611.10012v1.pdf
- R-FCN: Object Detection via Region-based Fully Convolutional Networks, Dai et al.: https://arxiv.org/pdf/1605.06409.pdf
- Learning Rich Features from RGB-D Images for Object Detection and Segmentation, Gupta et al, 2014: https://people.eecs.berkeley.edu/~sgupta/pdf/rcnn-depth.pdf
- Image-to-Image Translation with Conditional Adversarial Nets, Isola et al, 2016: https://arxiv.org/pdf/1611.07004v1.pdf https://phillipi.github.io/pix2pix/

Papers per Theme

3D Reconstruction and Recognition (1)

- 3DMV: Joint 3D-Multi-View Prediction for 3D Semantic Scene Segmentation, 2018: **https://arxiv.org/pdf/1803.10409.pdf**
- ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans, CVPR 2018 **https://arxiv.org/pdf/1702.04405.pdf**
- State of the Art on 3D Reconstruction with RGB-D Cameras, Eurographics 2018 **https://web.stanford.edu/~zollhoef/papers/EG18_RecoSTAR/paper.pdf**
- Matterport3D: Learning from RGB-D Data in Indoor Environments, 3DV 2017: **https://arxiv.org/pdf/1709.06158.pdf**
- Transformation-Grounded Image Generation Network for Novel 3D View Synthesis, Eunbyung Park, Jimei Yang, Ersin Yumer, Duygu Ceylan, Alexander C. Berg, 2017: http://openaccess.thecvf.com/content_cvpr_2017/papers/Park_Transformation-Grounded_Image_Generation_CVPR_2017_paper.pdf
- DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time, Newcombe et. al., 2015: http://grail.cs.washington.edu/projects/dynamicfusion/papers/DynamicFusion.pdf

Papers per Theme

3D Reconstruction and Recognition (2)

- Structured Indoor Modeling, Ikehata et. al., 2015: http://www.cse.wustl.edu/~sikehata/sim/iccv2015_ikehata.pdf
- BundleFusion: Real-time Globally Consistent 3D Reconstruction using Online Surface Re-integration, Dai et al, 2016: http://graphics.stanford.edu/projects/bundlefusion/ https://arxiv.org/pdf/1604.01093v1.pdf
- Sliding Shapes for 3D Object Detection in Depth Images, Shuran Song Jianxiong Xiao, 2014: http://slidingshapes.cs.princeton.edu/paper.pdf
- Real-Time Human Pose Recognition in Parts from Single Depth Images, Shotton et al. 2011: http://research.microsoft.com/pubs/145347/BodyPartRecognition.pdf
- MultiView Stereo Revisited, Goesele, 2006: http://www.gcc.tu-darmstadt.de/media/gcc/papers/Goesele-2006-MSR.pdf
- KinectFusion: Realtime 3D Reconstruction and Interaction Using a Moving Depth Camera, Izadi, 2011: http://research.microsoft.com/pubs/155416/kinectfusion-uist-comp.pdf

Papers per Theme

3D Reconstruction and Recognition (3)

- Building Rome in a Day, Agarwal, 2009: http://grail.cs.washington.edu/rome/rome_paper.pdf
- Depth estimation from image structure, Torralba, 2001: http://cvcl.mit.edu/Papers/Torralba-Oliva02.pdf
- Make3D: Learning 3D Scene Structure from a Single Still Image, Saxena, 2008: http://ai.stanford.edu/~asaxena/reconstruction3d/saxena_make3d_learning3dstructure.pdf
- Recovering Surface Layout from an Image, Hoiem, 2007: http://www.ri.cmu.edu/publication_view.html?pub_id=5818
- SegICP: Integrated Deep Semantic Segmentation and Pose Estimation, 2017: http://arxiv.org/pdf/1703.01661v1.pdf
- SemanticFusion: Dense 3D Semantic Mapping with Convolutional Neural Networks, 2016: http://arxiv.org/pdf/1609.05130v2.pdf

Papers per Theme

3D Reconstruction and Recognition (4)

- 3D Face Reconstruction by Learning from Synthetic Data, 2016: http://arxiv.org/pdf/1609.04387v2.pdf
- DSAC - Differentiable RANSAC for Camera Localization, Brachmann et al., 2017 https://arxiv.org/abs/1611.05705v4
- GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose, Yin et al., 2018: https://arxiv.org/abs/1803.02276v2
- Learning 3D Object Categories by Looking Around Them, Novotny et al., 2017: https://arxiv.org/abs/1705.03951v2
- RoomNet: End-to-End Room Layout Estimation, Lee et al., 2017: https://arxiv.org/abs/1703.06241v2
- Physically-Based Rendering for Indoor Scene Understanding Using Convolutional Neural Networks, Zhang et al., 2017: https://arxiv.org/abs/1612.07429v3
- On-the-Fly Adaptation of Regression Forests for Online Camera Relocalisation, Cavallari et al., 2017: https://arxiv.org/abs/1702.02779v2

Papers per Theme

Faces, People and Crowds (1)

- SfSNet: Learning Shape, Reflectance and Illuminance of Faces in the Wild, CVPR 2018: https://arxiv.org/pdf/1712.01261.pdf
- State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications,Eurographics 2018: https://web.stanford.edu/~zollhoef/papers/EG18_FaceSTAR/paper.pdf
- LCR-Net: Localization-Classification-Regression for Human Pose, Gregory Rogez, Philippe Weinzaepfel, Cordelia Schmid, 2017: http://openaccess.thecvf.com/content_cvpr_2017/papers/Rogez_LCR-Net_Localization-Classification-Regression_for_CVPR_2017_paper.pdf
- Learning Video Object Segmentation With Visual Memory, Pavel Tokmakov, Karteek Alahari, Cordelia Schmid, 2017: http://openaccess.thecvf.com/content_ICCV_2017/papers/Tokmakov_Learning_Video_Object_ICCV_2017_paper.pdf
- Learning without forgetting, Li and Hoiem, 2016: https://arxiv.org/pdf/1606.09282.pdf
- Faceness-Net: Face Detection through Deep Facial Part Responses, Yang et al.: https://arxiv.org/abs/1701.08393v3

Papers per Theme

Faces, People and Crowds (2)

- Scale-Aware Face Detection, Hao et al., 2017: http://arxiv.org/abs/1706.09876v1
- MoFA:Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction, Tewari et al., 2017: http://gvv.mpi-inf.mpg.de/projects/MZ/Papers/arXiv2017_FA/page.html
- Self-supervised Multi-level Face Model Learning for Monocular Reconstruction at over 250 Hz, Tewari et al., 2018: https://arxiv.org/abs/1712.02859
- Modeling Video Evolution for Action Recognition, Fernando et. al., 2015: http://www.robots.ox.ac.uk/~vgg/rg/papers/videoDarwin.pdf
- Deep Convolutional Network Cascade for Facial Point Detection, Yi Sun, Xiaogang Wang, Xiaoou Tang, 2013: http://www.ee.cuhk.edu.hk/~xgwang/papers/sunWTcvpr13.pdf
- Dense Trajectories and Motion boundary histograms for Action Recognition, Wang et al., 2013: https://hal.inria.fr/hal-00803241/document
- Recurrent Network Models for Human Dynamics, Fragkiadaki et. al., 2015, http://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Fragkiadaki_Recurrent_Network_Models_ICCV_2015_paper.pdf

Papers per Theme

Faces, People and Crowds (3)

- Cumulative Attribute Space for Age and Crowd Density Estimation, Chen, Gong, and Xiang, 2013: http://www.robots.ox.ac.uk/~vgg/rg/papers/Chen_Cumulative_Attribute_Space_CVPR13.pdf
- Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About, Sinha, 2006: http://web.mit.edu/sinhalab/Papers/19results_sinha_etal.pdf
- Histograms of Oriented Gradients for Human Detection, Dalal and Triggs, 2005: http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf
- Action Recognition with Improved Trajectories, Wang and Schmid, 2013: http://hal.archives-ouvertes.fr/docs/00/87/35/20/PDF/wang_iccv13.pdf
- 3D Face Morphable Models "In-the-Wild", 2017: http://arxiv.org/pdf/1701.05360v1.pdf
- UV-GAN: Adversarial Facial UV Map Completion for Pose-invariant Face Recognition: https://arxiv.org/abs/1712.04695
- 3D Face Morphable Models "In-the-Wild“: https://arxiv.org/abs/1701.05360

Papers per Theme

Faces, People and Crowds (4)

- InverseFaceNet: Deep Monocular Inverse Face Rendering: H. Kim et. al: https://arxiv.org/abs/1703.10956
- Self-supervised Multi-level Face Model Learning for Monocular Reconstruction at over 250 Hz, Ayush Tewari et. al: https://arxiv.org/abs/1712.02859
- Facelet-Bank for Fast Portrait Manipulation: Ying-Cong Chen et. al: https://arxiv.org/abs/1803.05576
- Generative Face Completion, Yijun Li , Sifei Liu, Jimei Yang, and Ming-Hsuan Yang: https://arxiv.org/pdf/1704.05838.pdf
- Finding Tiny Faces in the Wild with Generative Adversarial Network: http://openaccess.thecvf.com/content_cvpr_2018/papers/Bai_Finding_Tiny_Faces_CVPR_2018_paper.pdf
- Automatic Face Aging in Videos via Deep Reinforcement Learning: https://arxiv.org/pdf/1811.11082.pdf

Papers per Theme

GAN

- Generative Image Inpainting with Contextual Attention, Yu et al., 2018: https://arxiv.org/abs/1801.07892v2
- Fader Networks: Manipulating Images by Sliding Attributes, Lample et al., 2017: https://arxiv.org/abs/1706.00409v2
- DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks, Kupyn et al., 2018: https://arxiv.org/abs/1711.07064v3
- Interactive 3D Modeling with a Generative Adversarial Network, Liu et al., 2017: https://arxiv.org/abs/1706.05170v2
- GANimation: Anatomically-aware FacialAnimation from a Single Image, Pumarola et al., 2018: https://arxiv.org/pdf/1807.09251v2.pdf
- Deep Video Portraits, Kim et al., 2018: https://arxiv.org/pdf/1805.11714v1.pdf
- TextureGAN: Controlling Deep Image Synthesis with Texture Patches, Xian et al: openaccess.thecvf.com/content_cvpr_2018/papers/Xian_TextureGAN_Controlling_Deep_CVPR_2018_paper.pdf

Papers per Theme

Visual Reasoning, Q&A, Captions & Grounding

- Learning to Reason: End-to-End Module Networks for Visual Question Answering, Hu et al., 2017: http://openaccess.thecvf.com/content_ICCV_2017/papers/Hu_Learning_to_Reason_ICCV_2017_paper.pdf
- Show and Tell: A Neural Image Caption Generator, Vinyals et al, 2015: https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Vinyals_Show_and_Tell_2015_CVPR_paper.pdf
- Areas of Attention for Image Captioning, Perersoli, Lucas, Schmid, & Verbeek, 2017: http://openaccess.thecvf.com/content_ICCV_2017/papers/Pedersoli_Areas_of_Attention_ICCV_2017_paper.pdf
- Inferring and Executing Programs for Visual Reasoning, Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Judy Hoffman, Li Fei-Fei, C. Lawrence Zitnick, Ross Girshick, 2017: http://openaccess.thecvf.com/content_ICCV_2017/papers/Johnson_Inferring_and_Executing_ICCV_2017_paper.pdf

Papers per Theme

Self Supervision

- Unsupervised visual representation learning by context prediction, C. Doersch, A. Gupta, and A. A. Efros, 2015: https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Doersch_Unsupervised_Visual_Representation_ICCV_2015_paper.pdf

- Shuffle and learn: unsupervised learning using temporal order verification, I. Misra, C. L. Zitnick, and M. Hebert, 2016: https://arxiv.org/abs/1603.08561

- Look, Listen and Learn, Arandjelovic and Zisserman, 2017: http://openaccess.thecvf.com/content_ICCV_2017/papers/Arandjelovic_Look_Listen_and_ICCV_2017_paper.pdf

Papers per Theme

Convolutional Networks (1)

- Shift: A Zero FLOP, Zero Parameter Alternative to Spatial Convolutions, B. Wu et al, 2018: https://arxiv.org/abs/1711.08141
- Squeeze-and-Excitation Networks by J. Hu, L. Shen, G. Sun, 2017 https://arxiv.org/pdf/1709.01507.pdf
- Dynamic Routing Between Capsules by S. Sabour, N. Frost, G. E. Hinton, 2017 https://papers.nips.cc/paper/6975-dynamic-routing-between-capsules.pdf
- Densely Connected Convolutional Networks, Huang et al., 2017: https://arxiv.org/abs/1608.06993v5
- Coordinating Filters for Faster Deep Neural Networks, Wen et al., 2017: http://arxiv.org/abs/1703.09746v3
- Learning Bag-of-Features Pooling for Deep Convolutional Neural Networks, Passalis et al., 2017: https://arxiv.org/abs/1707.08105v2
- SegICP: Integrated Deep Semantic Segmentation and Pose Estimation, Wong et al., 2017: https://arxiv.org/abs/1703.01661v2

Papers per Theme

Convolutional Networks (2)

- A New Method to Visualize Deep Neural Networks, 2016: http://arxiv.org/pdf/1603.02518v2.pdf
- Deep Residual Learning for Image Recognition, He, Zhang, Ren, Sun, 2015: http://www.robots.ox.ac.uk/~vgg/rg/papers/deepres.pdf
- Visualizing and Understanding Convolutional Networks, Zeiler and Fergus, 2014: http://arxiv.org/abs/1311.2901
- OverFeat:Integrated Recognition, Localization and Detection using Convolutional Networks, Sermanet et. al., 2014: http://arxiv.org/pdf/1312.6229v4.pdf
- CNN Features off the Shelf: an Astounding Baseline for Recognition, Razavian et al, 2014: http://arxiv.org/pdf/1403.6382.pdf
- Learning to Generate Chairs with Convolutional Neural Networks, Dosovitskiy et. al., 2015: http://www.robots.ox.ac.uk/~vgg/rg/papers/Dosovitskiy_Learning_to_Generate_2015_CVPR_paper.pdf

Papers per Theme

Convolutional Networks (3)

- Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Ren, He, Girshick, and Sun, 2016: http://arxiv.org/pdf/1506.01497v3.pdf
- Densely Connected Convolutional Networks, 2016: http://arxiv.org/pdf/1608.06993v3.pdf
- Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, 2016: http://arxiv.org/pdf/1602.07261v2.pdf
- Rethinking the Inception Architecture for Computer Vision (Inception-v3), 2015: http://arxiv.org/pdf/1512.00567v3.pdf
- Generative Image Inpainting with Contextual Attention: https://arxiv.org/abs/1801.07892
- Free-Form Image Inpainting with Gated Convolution: https://arxiv.org/abs/1806.03589
- Feature Pyramid Networks for Object Detection: https://arxiv.org/abs/1612.03144

Course Organization

- Lectures with invited speakers: Monday
  - Lectures (3D reconstruction theory) and invited speakers
  - Read the book
- Paper presentation students:
  - Read the papers
  - Submit 3 questions (before Wednesday 12:59am)
  - One Presentation
- Practice:
  - 3D reconstruction project

Practical (40%)

Implementing a method for 3D object reconstruction. Each week you do a part (groups of 2-3 students).

The idea is to combine computer vision modules to have a real working system in the end. In the first weeks we will do separate modules. Then, you have to combine the modules to get the final system and write a small report about. If possible, always give us your reasons why you did something and comment your code.

3D Reconstruction from RGB-D

Week 1-3: ICP (Iterative Closest Points)

Week 4-5: SFM (Structure from Motion)

Week 6-8: 3D model building and printing

Homework: Paper Assignment.

Please submit 3 paper preferences + team (2-3 students) before Thursday April 4th 11:59 pm to Wei Zeng: w.zeng@uva.nl.