Foundations of Computer Vision

Introductory meeting

Nassir Navab, Federico Tombari,
Fabian Manhardt, Johanna Wald

Seminar contents

- this seminar includes a selection of the most relevant and impactful papers in the field of computer vision
- papers have been selected to cover different aspects of the topic:
  - Keypoints and Tracking
  - Object Detection
  - Filtering and Segmentation
  - Human Body/Face Detection
  - Simultaneous Localization and Mapping
  - Deep Learning for Computer Vision

Goals

- You are going to learn:
  - about relevant works in the field of Computer Vision
  - how to read and understand a scientific article
  - how to write a scientific report
  - how to give a talk to an audience, and deal with related questions afterwards

Seminar Schedule

- 6 sessions, 1 every Thursday, 2pm-3.30pm
- Two presentations per session
- Seminarraum 03.13.010

- Paper assignments:
  - selected students can express up to 3 preferences
  - We will then match them to a paper and tutor trying to maximize global happiness

Presentation

- Each presentation is 20 minutes + 10 minutes for Q&A
- Slides templates (Powerpoints, Latex, ..) provided on website

- The presentation should cover all relevant aspects of the paper
  - Introduction and state of the art
  - Main contribution(s)
  - Experimental results
  - Discussion, summary and future work
- The presentation should be self-contained
- All students are expected to attend all presentations and interact during Q&A (this will influence your final mark)

Report

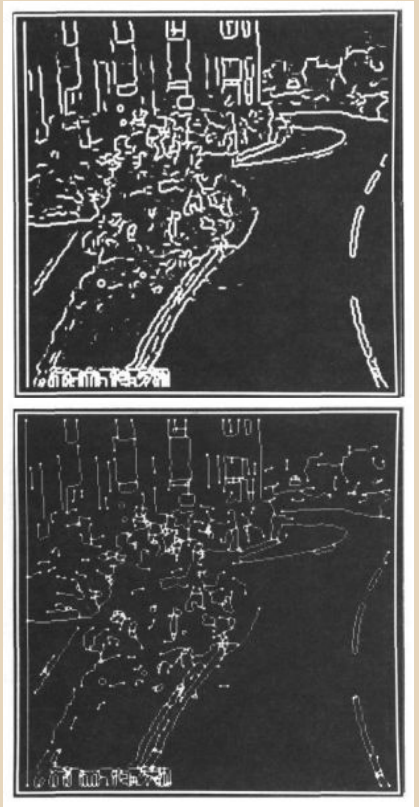
- The report should summarize the paper in the way it has been presented during the talk, and provide the student's opinion concerning the main contributions and impact
- Language: English
- Max 8 pages
- Template on course website
- Once ready, send the report to supervisor, within **two weeks** from the day of the presentation

Evaluation criteria

- Quality of presentation (both regarding slides and speech)
- Quality of the report
- Comprehension of the scientific contents of the presented work
- Interaction and participation during the other talks

C. Harris, M. Stephens, “A Combined Corner And Edge Detector”, Alvey vision conference 1988

- Goal is to find points / regions of interest in an image
- Idea
  - Find regions of high intensity values within a small window by using a response function based on eigenvalue analysis
- Application
  - Feature Points Detection



B.D. Lucas, T. Kanade, “An iterative image registration technique with an application to stereo vision”,
Proceedings of Imaging Understanding Workshop, pages 121-130, 1981

- Goal is to find the affine registration between two frames

- Idea
  - Use image gradients
  - Use multi-resolution pyramid to converge in a stable fashion
- Application
  - Optical Flow
  - Stereo Image matching
  - Depth map generation
- Implemented in OpenCV

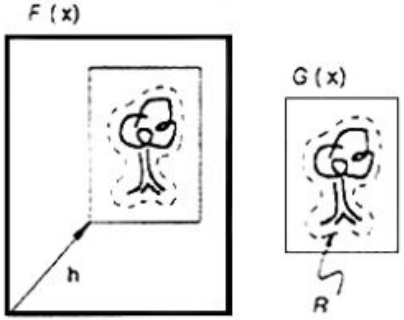


Figure 1: The image registration problem

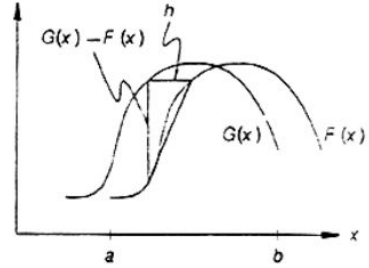


Figure 2: Two curves to be matched

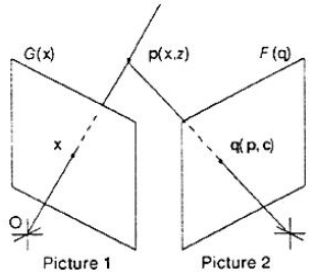
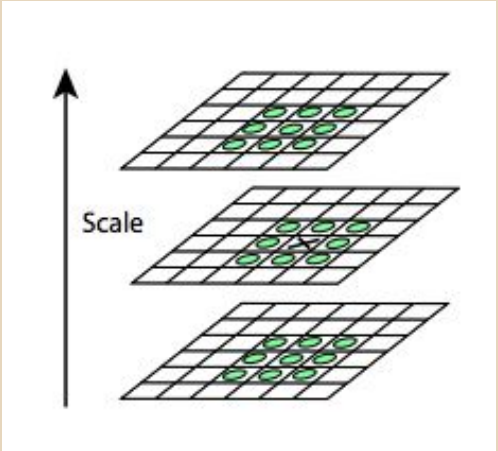
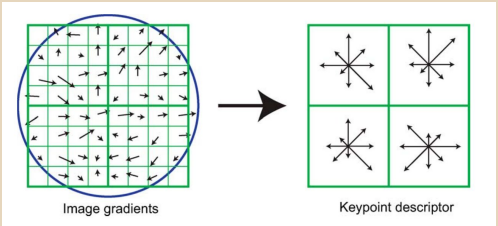


Figure 3: Stereo vision

D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, IJCV 2004

- SIFT: Local Image Descriptor
  - Highly robust against viewpoint changes
- Algorithmic pipeline
  - Find Interest points
  - Do statistics on local gradient directions
  - Accumulate information
- Application areas
  - Multi-view matching
  - Object recognition





J. Sivic, A. Zisserman, “Video Google: A Text Retrieval Approach to Object Matching in Videos”, IEEE International Conference on Computer Vision 2003

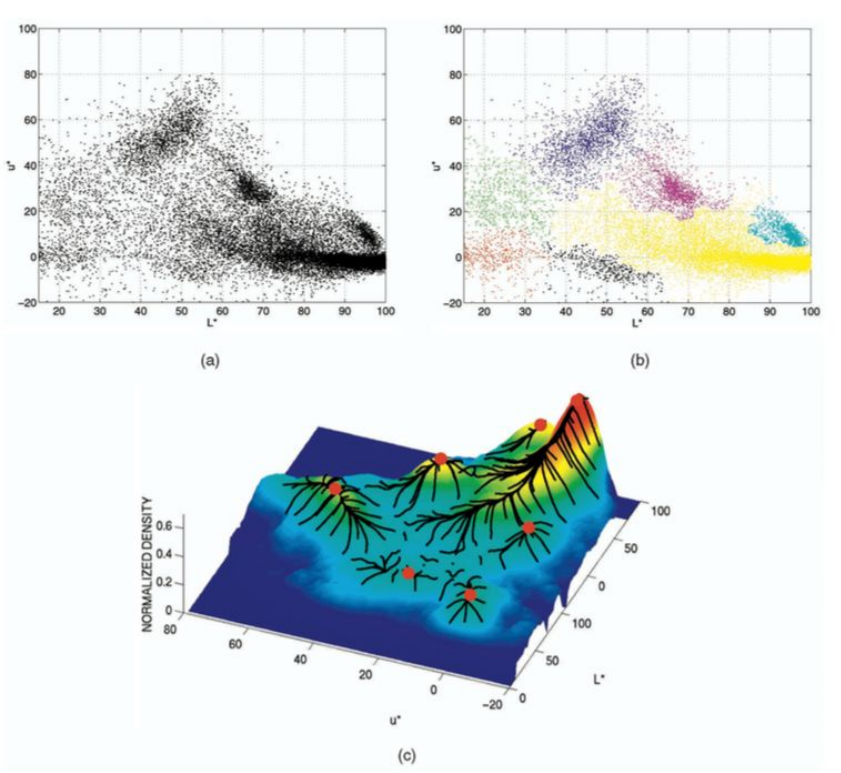
- Search and localize all occurrences of an user outlined object in a video

- Idea
  - Represent the object by a set of viewpoint invariant region descriptors
  - Vector quantizes the descriptors into clusters which will be the visual ‘words’ for text retrieval

D. Comaniciu, P. Meer, “Mean Shift: A Robust Approach towards Feature Space Analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence 2002

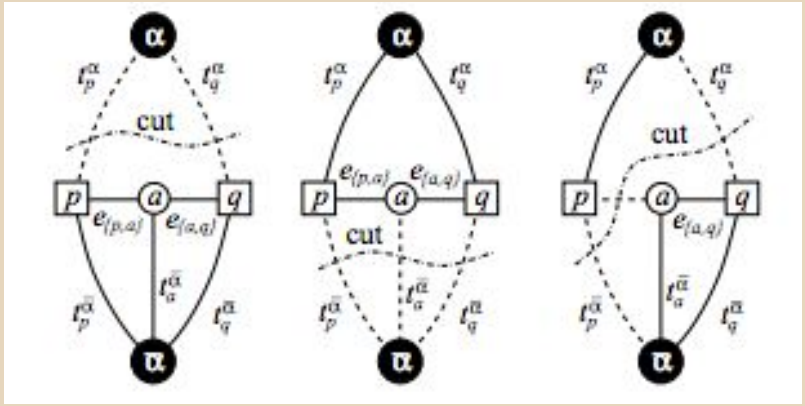
- Non-parametric technique for detecting multiple modes in density functions

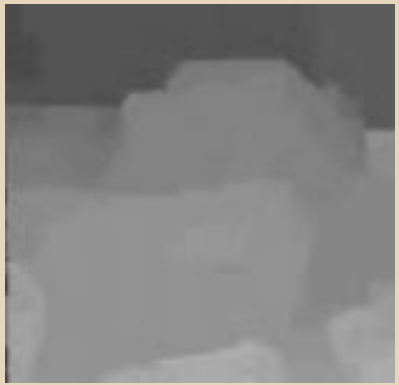
- Allows analysis of feature spaces for many tasks including
  - Segmentation
  - Smoothing
  - Clustering
  - Tracking
  - ...



Y. Boykov, O. Veksler, R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts”, IEEE Transactions on Pattern Analysis and Machine Intelligence 2001

- Assign labels (such as disparity) to the pixels of an image
- Authors propose two different approximation algorithms that are based on graph cuts to find a local minimum
- Applications: Image restoration, Segmentation

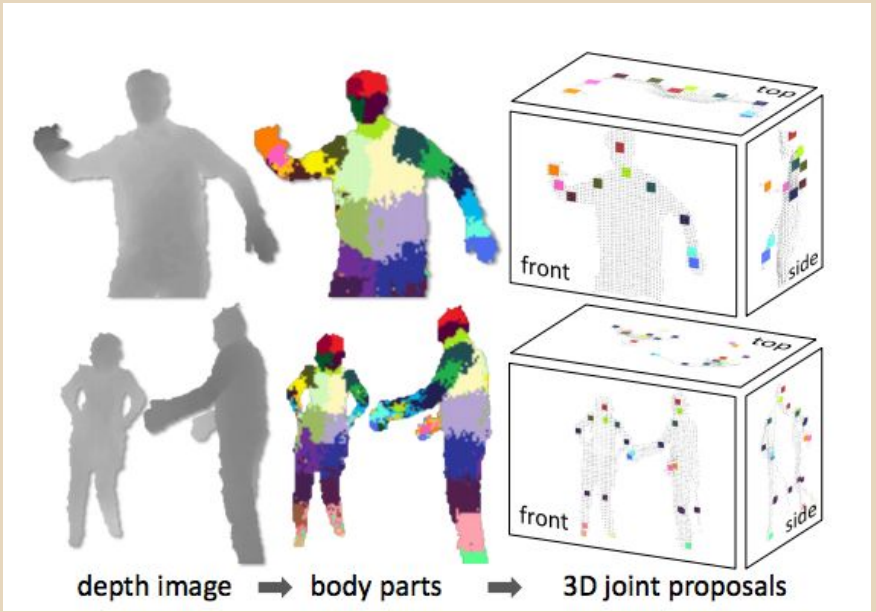




Shotton et al. “Real-Time Human Pose Recognition in Parts from Single Depth Images”, Communications of the ACM Pages, 116-124, 2011

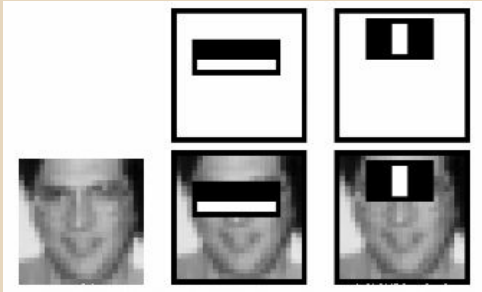
- propose a method to quickly and accurately predict human pose (3-D positions of body joints) from a single depth image, without depending on information from preceding frames

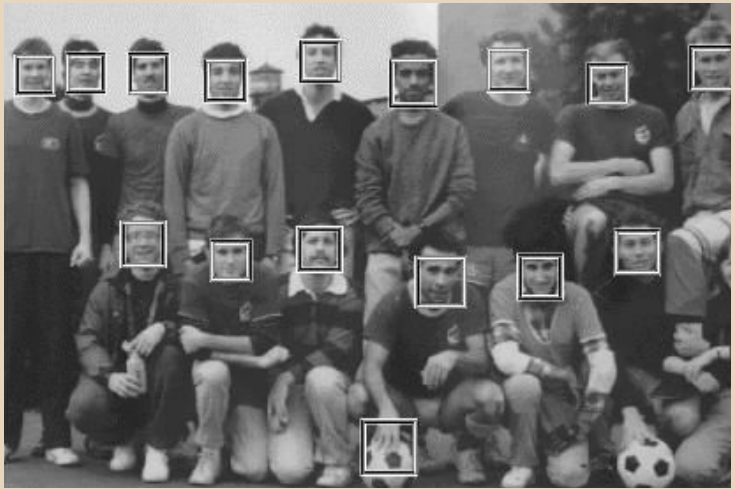
- Applications: gaming, human-computer interaction, security, telepresence, health-care



P. Viola, M. Jones, “Rapid object detection using a boosted cascade of simple features”, CVPR 2001

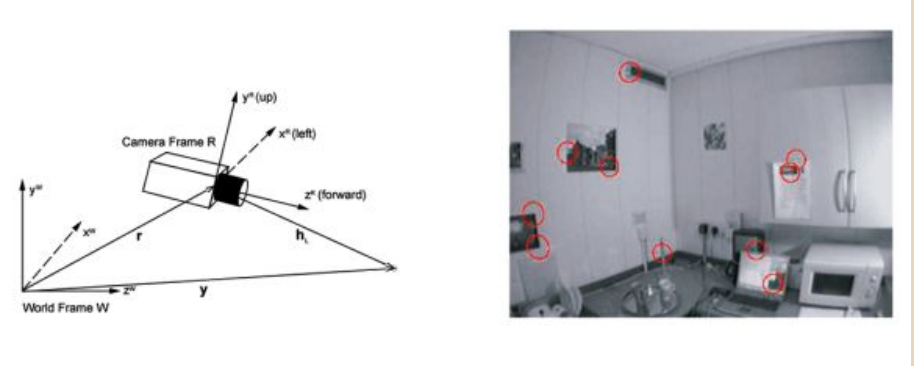
- Face detection using Haar features
- The most discriminative features for faces are automatically selected by means of AdaBoost classifier
- Efficient since computes the features via incremental schemes (Integral Images)





Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse, MonoSLAM: Real-Time Single Camera SLAM, IEEE transactions on pattern analysis and machine intelligence 2007

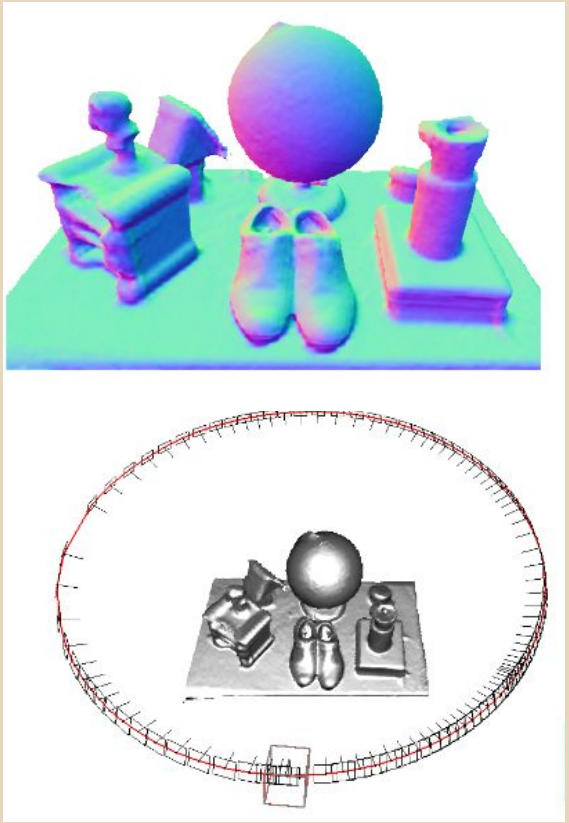
- Idea
  - real-time algorithm to recover the 3D trajectory of a monocular camera, moving rapidly through a previously unknown scene



- Applications
  - robotics, autonomous vehicles, markerless augmented reality, ...

Newcombe et al., “KinectFusion: Real-Time Dense Surface Mapping and Tracking”, ISMAR 2011

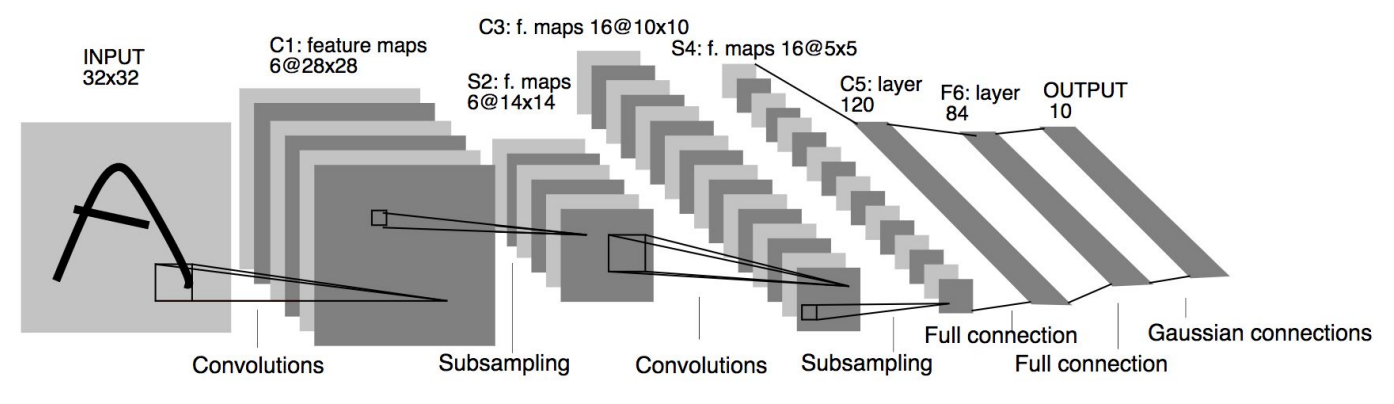
- Goal is to reconstruct the perceived environment of indoor scenes in real-time
- Idea
  - Fuse depth data into a single global implicit surface model of the observed scene
  - Simultaneously estimate the current sensor’s pose by by tracking the live depth frame relative to the global model using a coarse-to-fine iterative closest point (ICP) algorithm



LeCun et al., “Gradient-Based Learning Applied to Document Recognition”, Proceeding of the IEEE, pages 2278-2324, 1998

- Reviews various methods applied to handwritten character recognition and compares them on a standard handwritten digit recognition task

- Introduces one of the first Convolutional Network architectures which was primarily used to conduct handwritten character recognition



Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in neural information processing systems, 2012

- Large, deep convolutional neural network to classify 1.2 million images from ImageNet into 1000 different classes
- Achieved top-1 and top-5 error rates of 37.5% and 17.0%
- 60 million parameters, 650,000 neurons, five convolutional layers, max-pooling layers, three fully-connected layers with final 1000-way softmax
- Efficient GPU implementation for the convolution