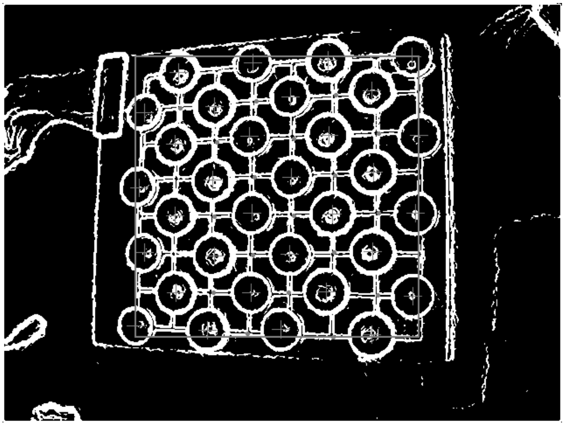
Artificial Intelligence and Image Processing



Dr.techn. Alexander K. Seewald

Outline

Biological Background: The Human Visual System

Intelligent Systems: DARPA Challenge

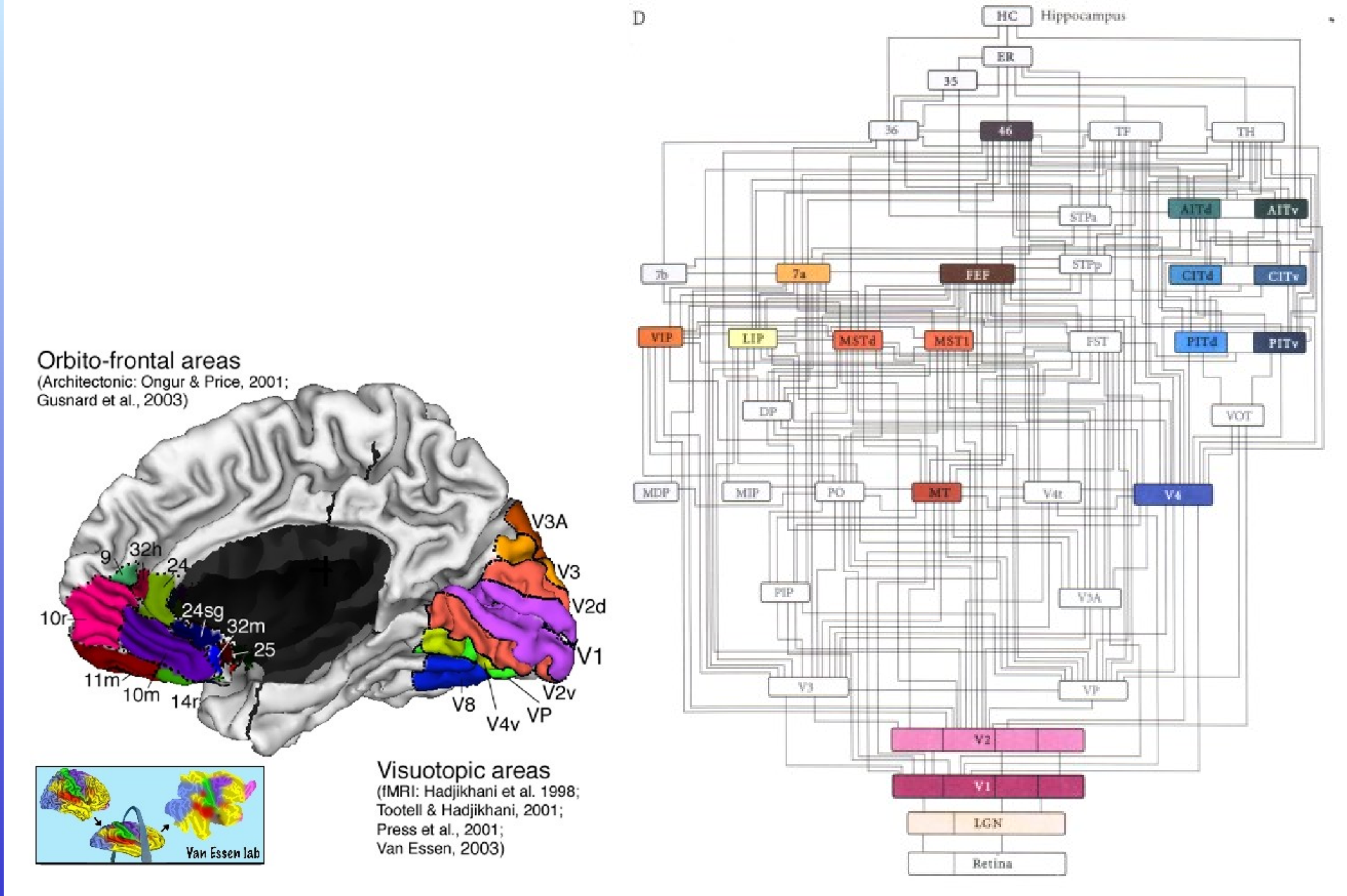
Algorithms: Scale-Invariant Feature Transform (SIFT)

Localization via Vision: FIRA RoboCup 2005/2006

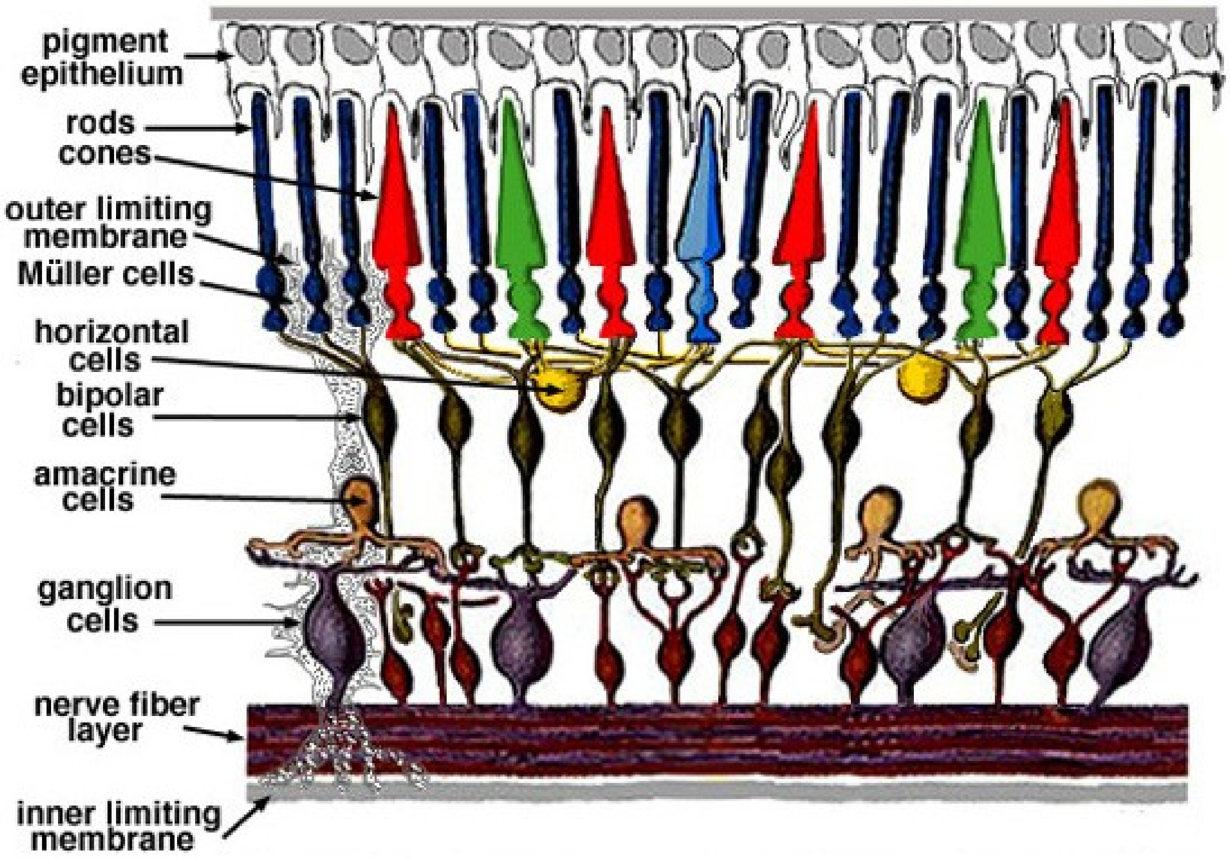
Image Classification: Handwritten Digit Recognition

Conclusion

The Human Visual System



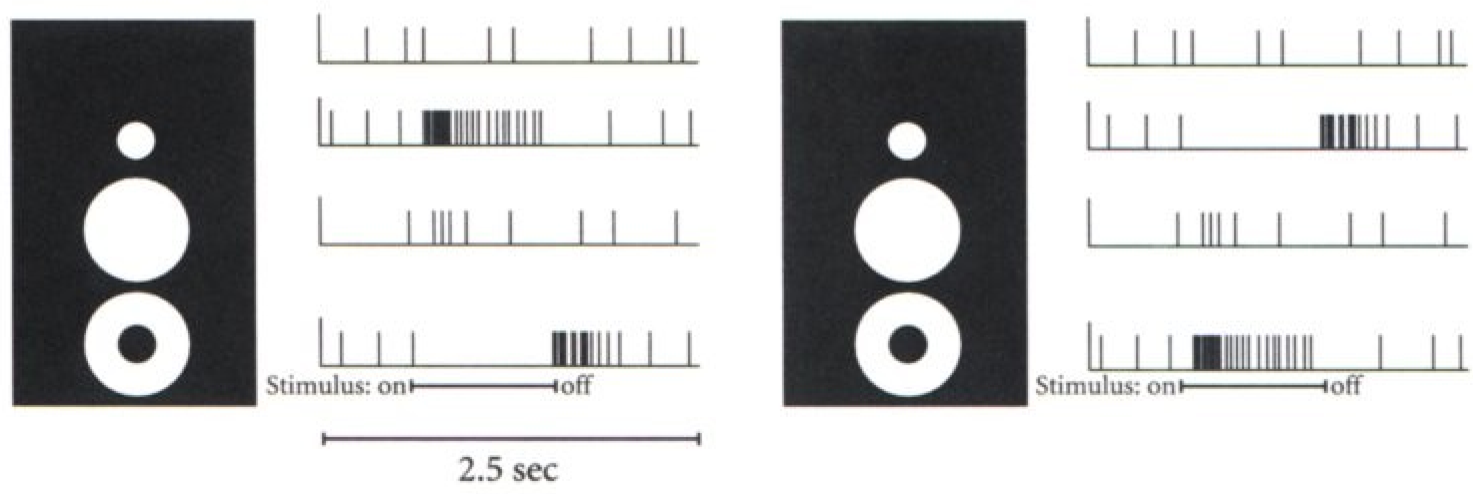
The Retina



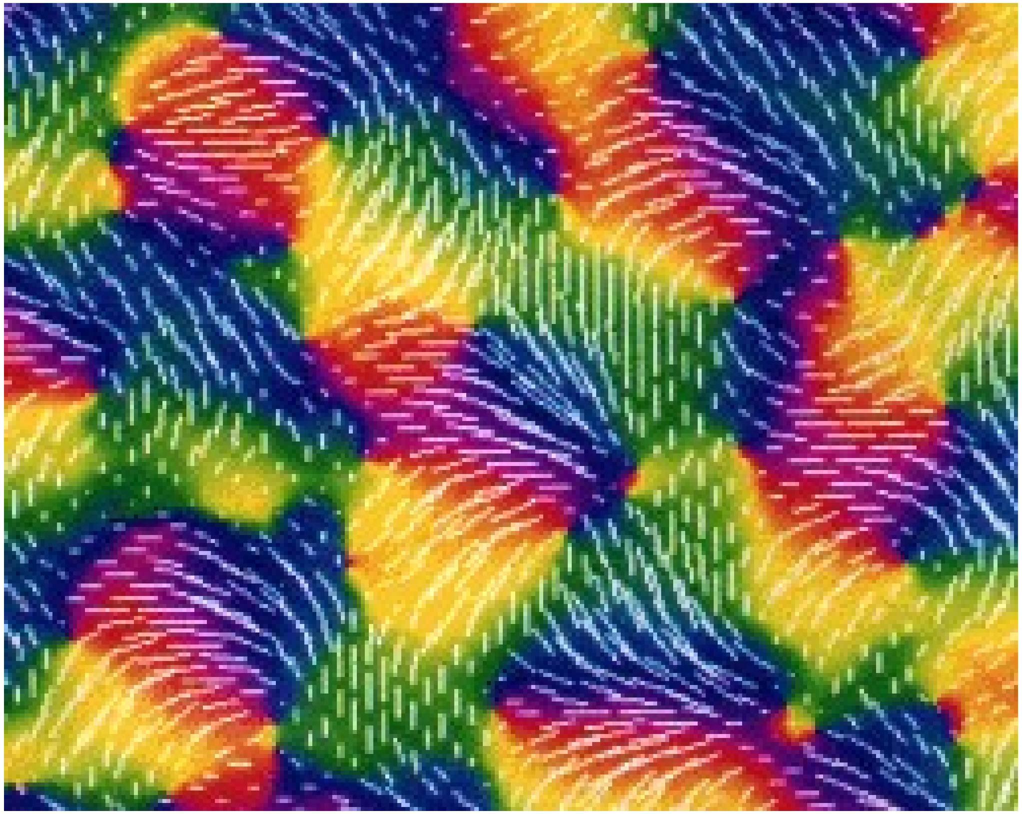
Alexander K. Seewald
alex@seewald.at / alex.seewald.at

Ganglion cells are sensitive to brightness differences

Left: on-center cell, right: off-center cell

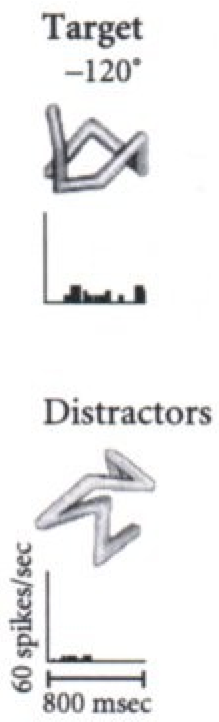


Orientation columns in the primary visual cortex are sensitive to image gradients in specific directions.



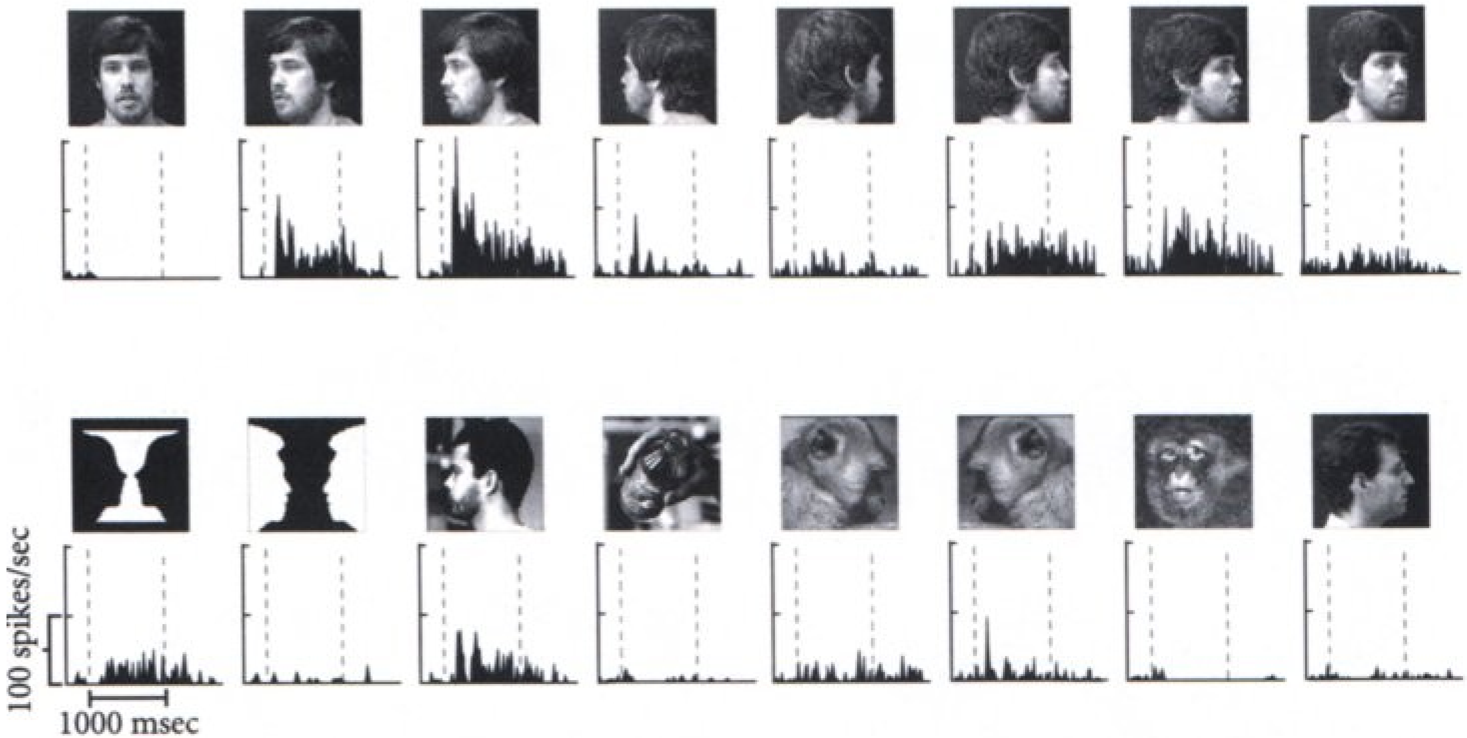
Higher vision areas have more specific feature detectors.

This is a paper-clip recognition neuron...



Alexander K. Seewald
alex@seewald.at / alex.seewald.at

... and this is a bearded-faces-in-profile neuron.



DARPA Challenge

Autonomous robot vehicle which won the DARPA Challenge 2005. Built at Stanford University in about 15 months by a team of around 35 people. Uses Machine-Learned Laser Perception and Speed Strategy.

Stanley

Stanley

Sensors

- Sight: Five laser rangefinders; monocular video camera; radar for long range sight
- Position: GPS sensor with 20cm resolution for pose estimation; measurements of wheel speed for pose estimation
- Balance: a 6 DOF inertial measurement unit; GPS compass generates 2 DOF balance information from two separate GPS antennas

Brains

- Six Pentium M motherboards in a rugged rack mount unit
- Battery-backed, electronically-controlled power system
- Custom software modules for: Planning and Optimization; Control; LIDAR - Light Detection and Ranging; *Computer Vision*; Inertial Navigation; Reliability
- Data sampling from instruments at rates varying from 10Hz to 100Hz.

Adaptive Vision

- Use laser range finder to locate a smooth patch of ground ahead
- Sample color and texture of this patch from monocular video image
- Scan for same color and texture in the whole image

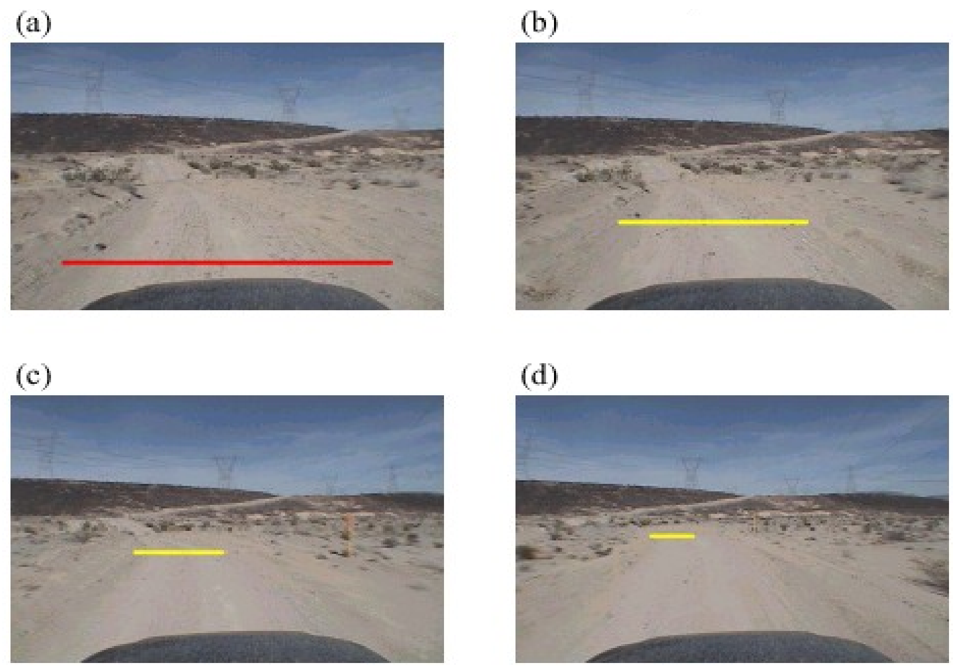
$\Rightarrow$ *Road Segmentation*

Alternative Vision System

"Adaptive Road Following using Self-Supervsied Learning and Reverse Optical Flow" (Lieb, Lookingbill & Thrun, 2005)

- Assume: region directly in front of the vehicle is drivable road.
- Sample region at various distances from past images, using reverse optical flow to determine its previous position.
- Match each sampled region at appropriate vertical pos.in the current image
- Integrate via Dyn.Prog.

⇒ *Road Segmentation*

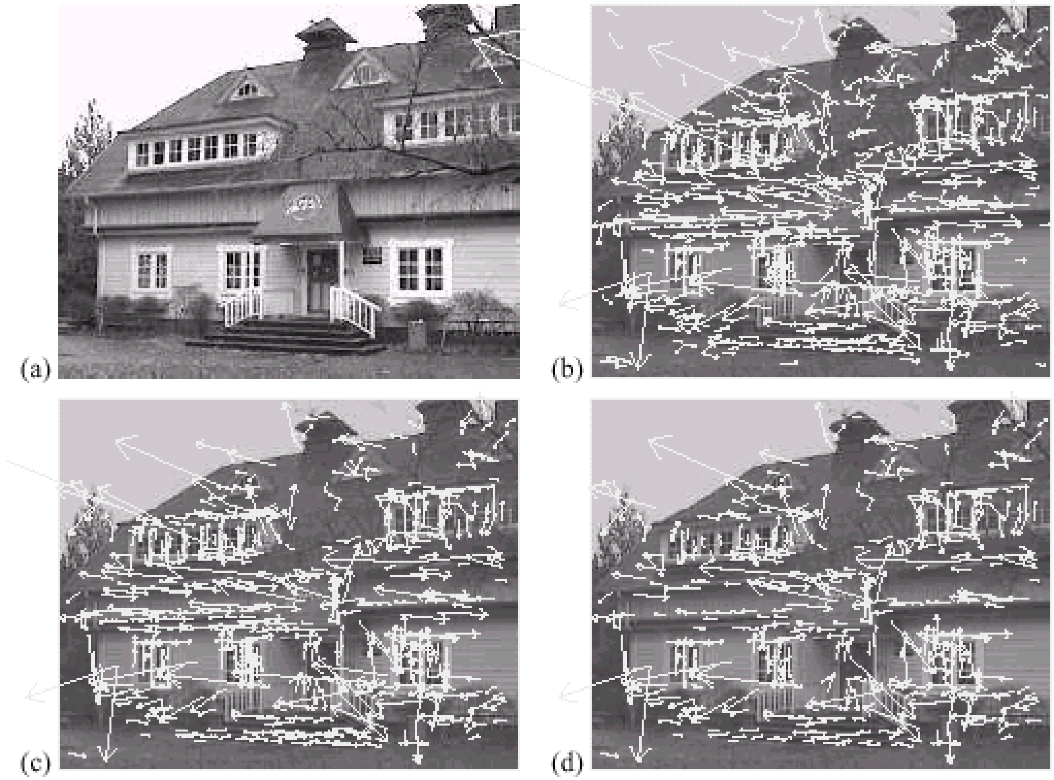


Alexander K. Seewald
alex@seewald.at / alex.seewald.at

Scale Invariant Feature Transform (SIFT)

[Lowe, 1999] & [Lowe, 2004]

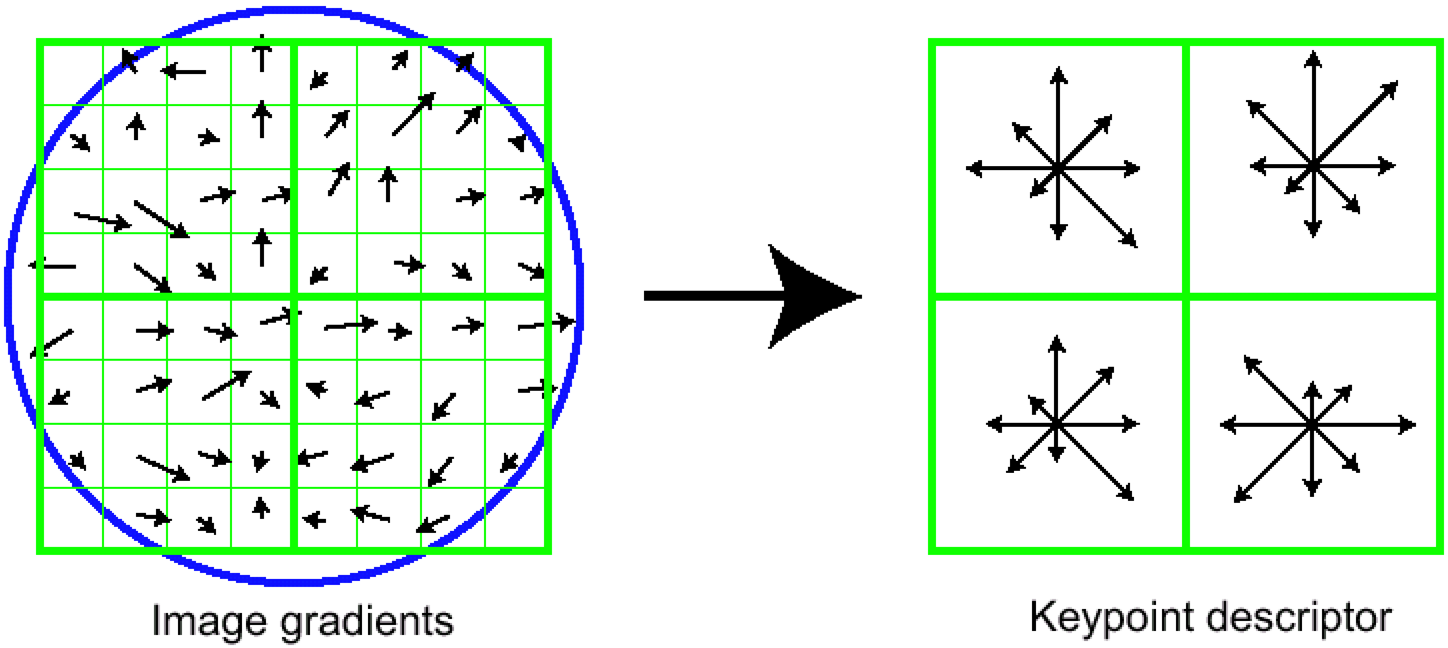
2. Scale-space extrema detection via difference-of-gaussians
3. Keypoint localization by local quadratic interpolation



Scale Invariant Feature Transform (SIFT)

3. Orientation assignment via direction histogram peaks
4. Keypoint descriptor by sampling & trilinear interpolation

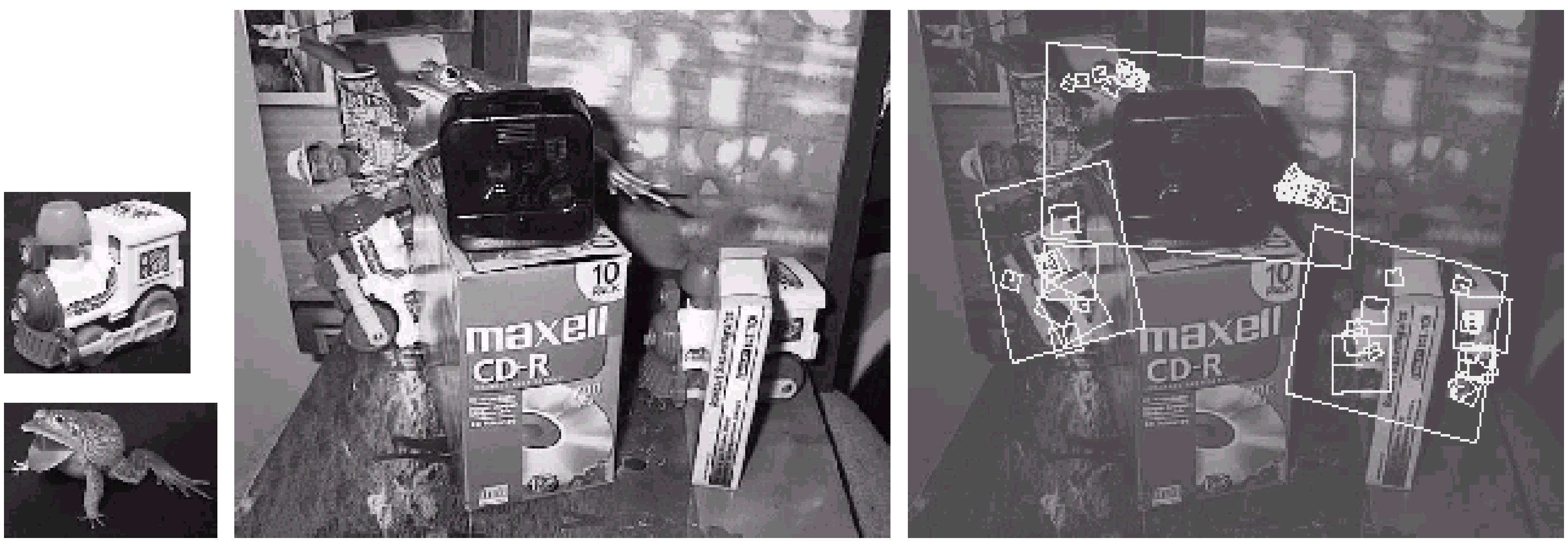
⇒ *Set of image keypoints with 128-dimensional descriptors*



Scale Invariant Feature Transform (SIFT)

Matching keypoints with an image database

- Best-Bin-First: fast approximation of nearest-neighbor
- Clustering features with Hough Transform in Pose Space
- Least-squares solution of pose to determine final affine transform of the object for accurate localization



European RoboCup (FIRA)

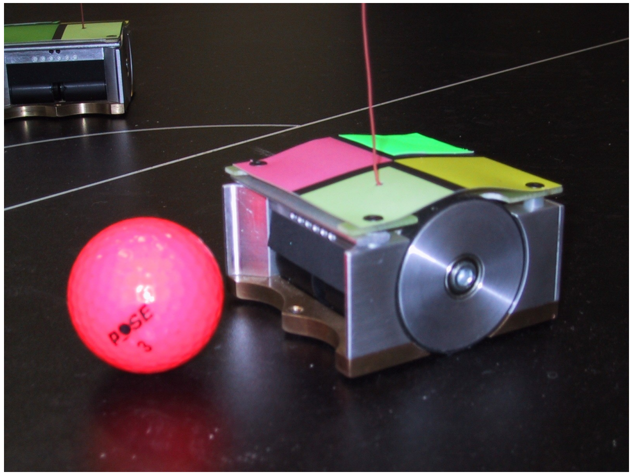
The IHRT Robot Soccer team at the Technical University of Vienna is currently one of the top teams in the European RoboCup (FIRA).

Video

European RoboCup (FIRA)

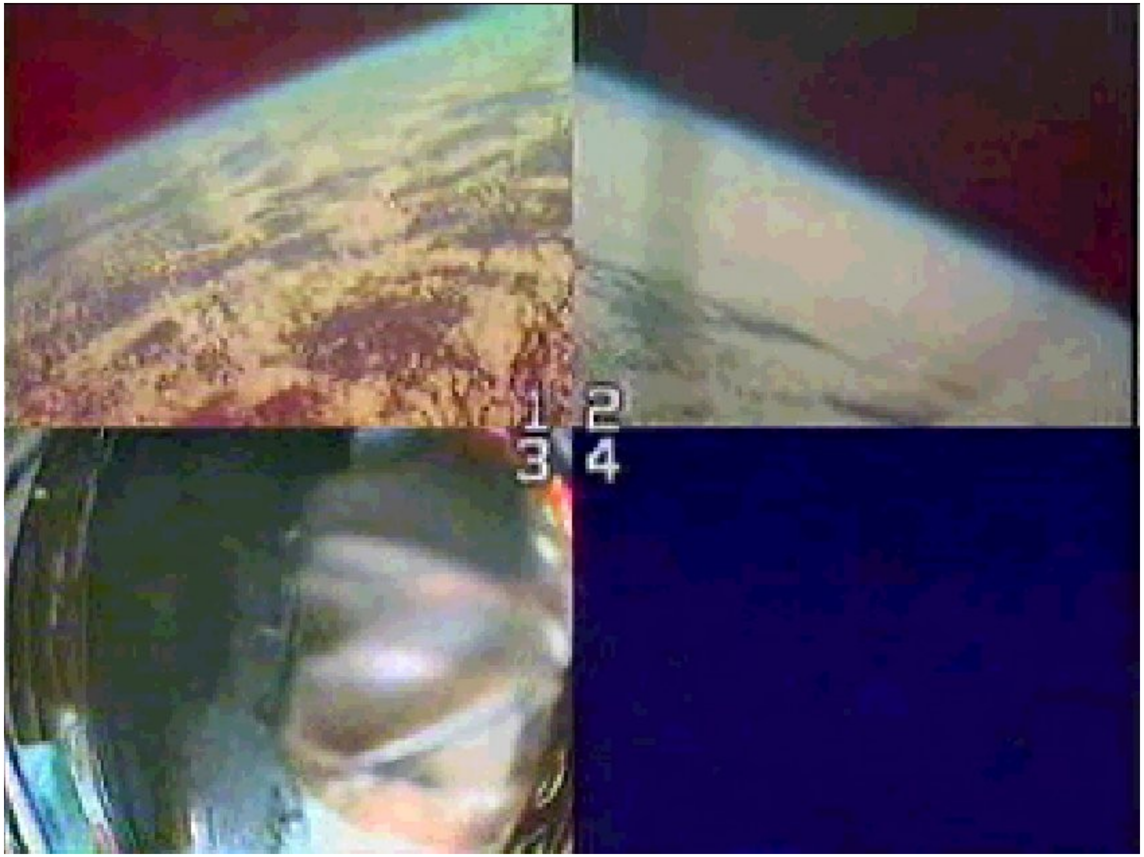
Robot position estimation

- 1-2 ceiling-mounted cameras (>=60Hz)
- Four colored fields on top of each robot in combinatorial code – at most 4-5 colors distinguished by vision.
- Calibration: manually select volumes in normalized RGB or HLS color space for each color (30-45 minutes before game)
- Robot and ball movement model
- Commands are given by radio (860Mhz)
- No collision detection and no tracking of other team robots(!)
- Around 40,000 lines of code in all; 5,000 for position estimation; 5,000 for robot behaviour controller (centralized)



Robot Soccer in Space

The same hardware platform was also used in a project that aims to build a space-based solar panel energy generator.



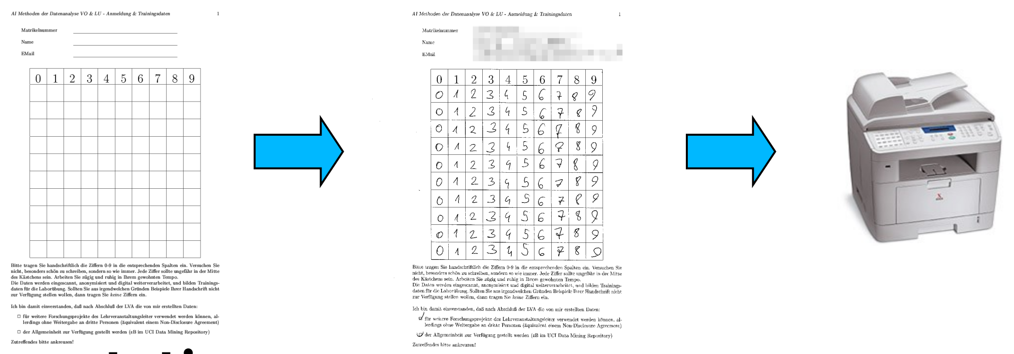
Handwritten Digit Recognition

Students of *AI Methods of Data Analysis* contributed around 4500 handwritten digit samples in 2005, and classified them with a variety of current learning algorithms.

The dataset and a technical report about the project is available at http://alex.seewald.at/digits/

Handwritten Digit Recognition

Preprocessing



```
Scan1.bmp
Scan2.bmp
Scan3.bmp
...
Scan49.bmp
Scan50.bmp
Scan51.bmp
```

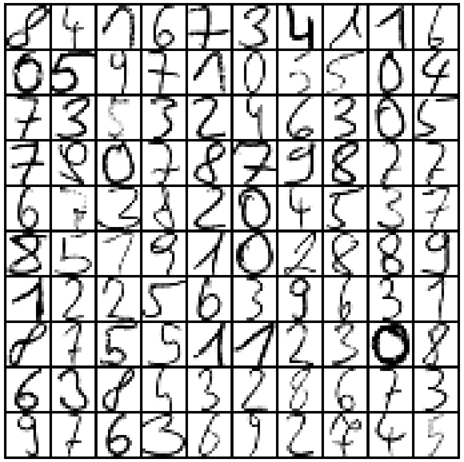
Segmentation

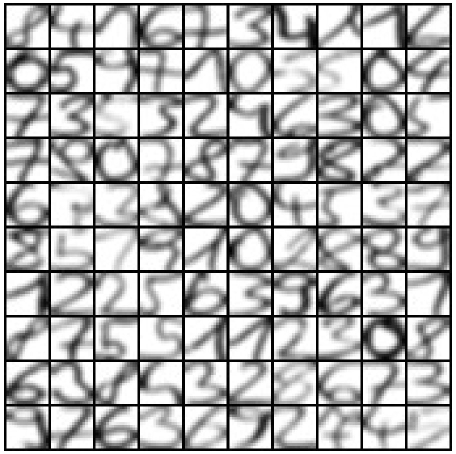
- Vertical and horizontal histogram peaks as first estimate
- Local search for the true line along each midpoint between crossings
- Linear regression on found points $\Rightarrow$ table cell positions
- Reduce table cell size uniformly until border has only white pixels
- Reduce table cell size for each direction separately, until enough black pixels are found in several consecutive steps

Manual image processing task with 8 resolution-dependent parameters

Handwritten Digit Recognition

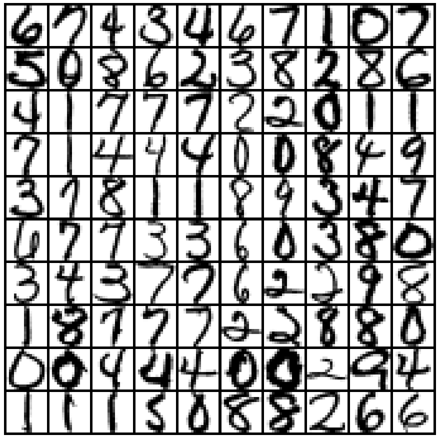
- Digits were downsampled to 16x16 pixels
- Moderate blurring and arbitrary scaling improved performance of most learners (bottom right). This data was used for image classification.



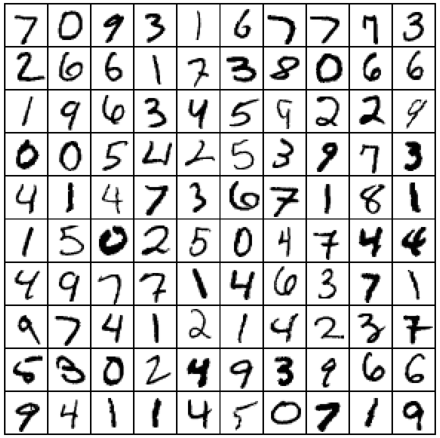


Similar to digits from other benchmark datasets:

US Postal (zip)



MNIST



Alexander K. Seewald
alex@seewald.at / alex.seewald.at

Handwritten Digit Recognition

Results of best learning algorithm on several datasets

- Our dataset (digits): 6.01% error rate (2,191 samples trained)
- US Postal (ZIP): 4.29% error rate (7,291 samples trained)
- MNIST: 1.27% error rate (60,000 samples trained)

Conclusion:

- Performance depends strongly on number of training samples
- Almost no expertise can be transferred between datasets(!)

Best bet is therefore to train on real-life data for each application

Alexander K. Seewald
alex@seewald.at / alex.seewald.at

Conclusion

- Human visual system works better, but technology is slowly gaining
- Segmentation still needs a lot of ad-hoc programming and parameter tuning. Learning for segmentation tasks has not yet been conclusively demonstrated, and may be an interesting topic for the future.
- Once segmentation is done, a variety of learning approaches can be used for image classification, and generally perform well.

Future/Current Work