Zurich University
of Applied Sciences

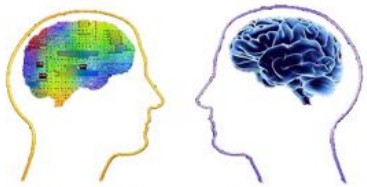
InIT Institute of Applied Information Technology
Human Information Interaction
ICT Accessibility Lab

PROJECT: AUTOMATIC TRANSLATION FROM SIGN LANGUAGE

Andri Reichenbacher

Conference Barrier-free Communication: Methods and Products

ZHAW Event, 14th September 2017



Overview

Zurich University of Applied Sciences

zh aw

- **Introduction** Motivation, Problem Definition
- **State of Art** Research, Practicality
- **Analysis** Demo, Results
- **Conclusion** Application Area, Technical Progress

Motivation

Zurich University of Applied Sciences

- **Project Idea**
  - Relief for deaf and hard of hearing by automatic translation from sign language to text or audio
  - Evaluation project in 2013 with helping and usage of the available and different sensors for recognition of gestures
- **Proposals**
  - Useful as complement to barrier-free communication in sign language
  - Usage of modern technologies and research methods for the development of communication platform for people with hearing impairment
  - Innovative approaches for image processing, machine learning and deep learning

Problem Definition

Zurich University of Applied Sciences

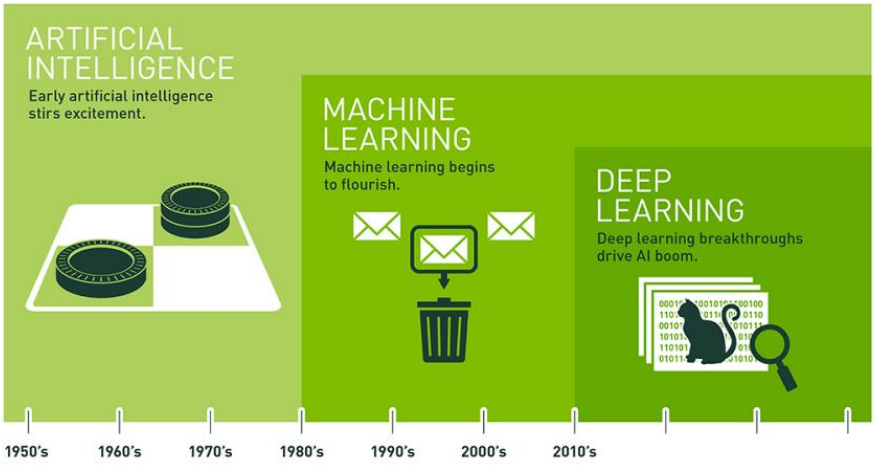
zh aw

What is Sign Language?

- Body position = Association of Hands + Arms + Torso + Mimic
- Gestures mean language expression for thinking in flow images as movie or flip book, not primitive single image
- Visually perception by associative production of body position in three-dimensional space, that is some composited informations can reflect at once
- In contrast spoken language is a linear, sequential production of words by voice, hence notion for one-dimensional space
- Therefore this own grammatical structure in 3D and totally different to spoken language
- Sight contact always essential, otherwise break in communication

Research

What is State of the Art?

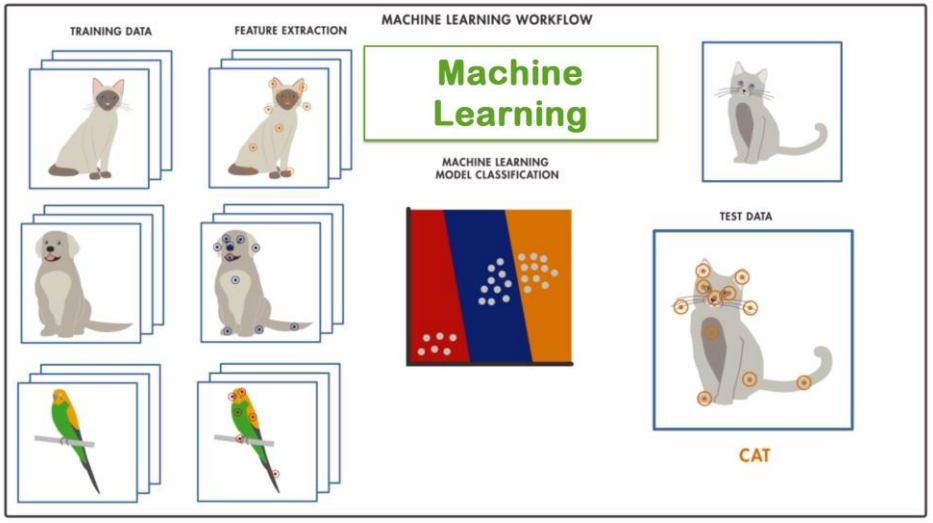


- **Artificial Intelligence**
  - IBM Deep Blue defeated world chess champion
  - Was an important fundament for development
- **Machine Learning**
  - Is always still current used and well established
  - Is considered as transition to Deep Learning
  - Uses different efficient algorithms for example as Adaptive Boosting and Random Forests
- **Deep Learning**
  - Is at present and has trend towards Big Data, Smart Data, Data Science etc.
  - Is a specialized form of Machine Learning
  - Uses different kind of neural networks such as Convolutional Neural Network

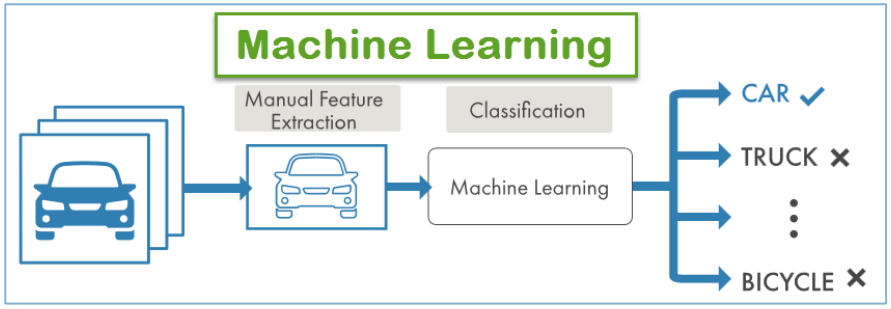
Research

Machine Learning (ML)

- Look at a cat, a dog or a parrot
- Learning for many images of animals as object to identify them over time
- Three objects are divided into classes as error free as possible
- Each object has relevant features of image as edges, corners, pointed ears, tails etc.



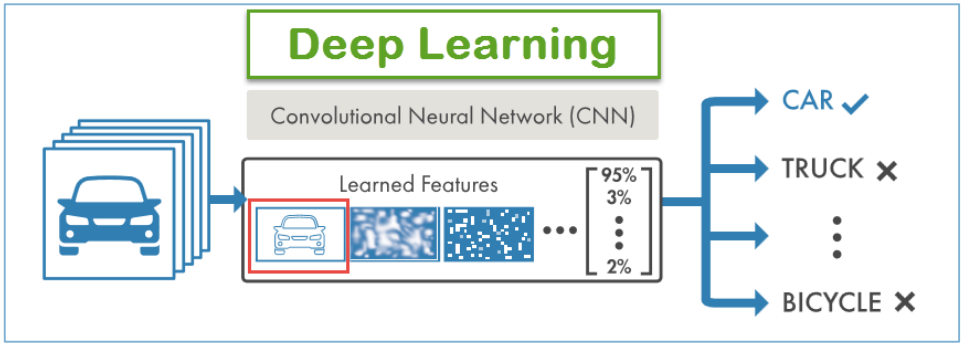
- ML requires manual feature extraction from images
- Features are used to create a model that categorizes the objects in the image



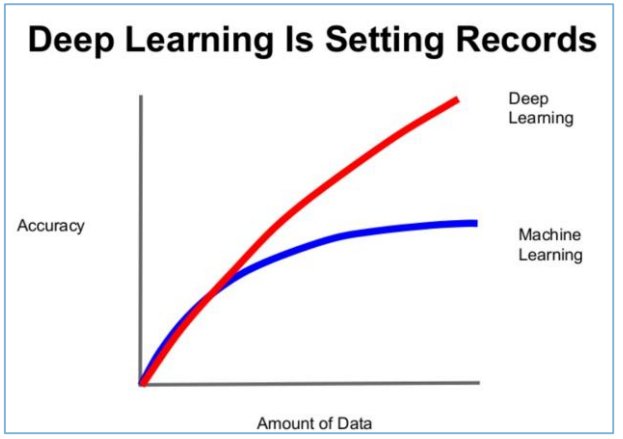
Research

Deep Learning (DL)

- Is generally more complex
  to get reliable results
- Eliminate manual feature extraction
- Can automatically and directly learn
  relevant features in data
- Performs «end-to-end-learning» in principle



- Key advantage of DL
  - Continue often to improve the accuracy
    as the amount of data increases

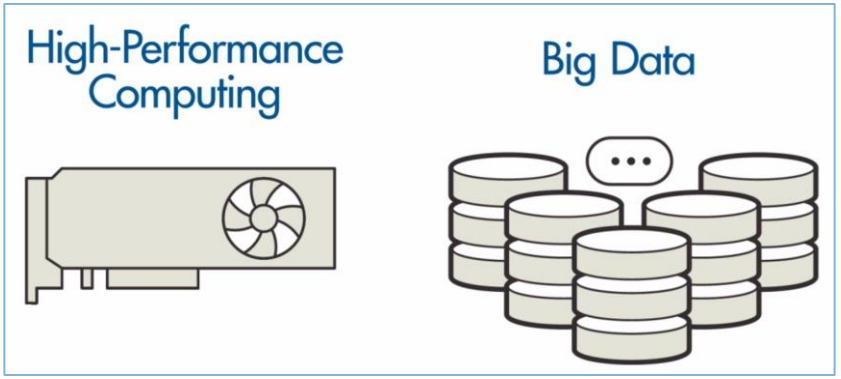


Practicality

Zurich University of Applied Sciences

Comparison of Machine and Deep Learnings

- Conditions for decision between ML and DL
- **Pro ML**
  - Is suitable especially for a small amount of data to train
  - Can achieve a short training time
  - Is enough to use an efficient CPU
  - Is possible to define own features
- **Pro DL**
  - Requires a very large amount of data (thousands of images) to train
  - Needs a long training time
  - Needs less time to analyze all images
  - Requires a high-performance GPU to rapidly process image data



	Machine Learning	Deep Learning
Training dataset	Small	Large
Choose your own features	Yes	No
# of classifiers available	Many	Few
Training time	Short	Long

Analysis

Zurich University of Applied Sciences

Which Methods?

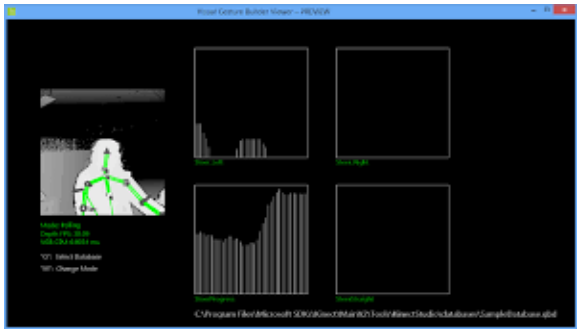
- Current usage of Kinect cam with 3D depth-sensor from Microsoft
- Current usage of tool Visual Gesture Builder (VGB) from Microsoft
  - Integral part of algorithms such AdaBoost and Random Forest
- Reasons for image processing algorithms with VGB
  - Minimum effort for record clips, tagged clips, without programming, non-engineering task etc.

Which Gestures for training data?

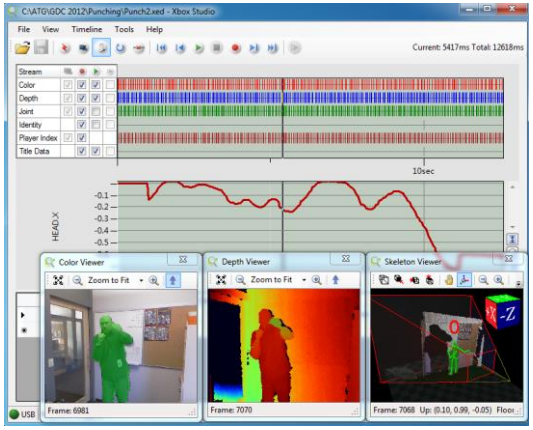
- Base idea for three different meaning of gestures
- But these motions relatively similar at first sight for example as Car, Thursday and Milk

Demo

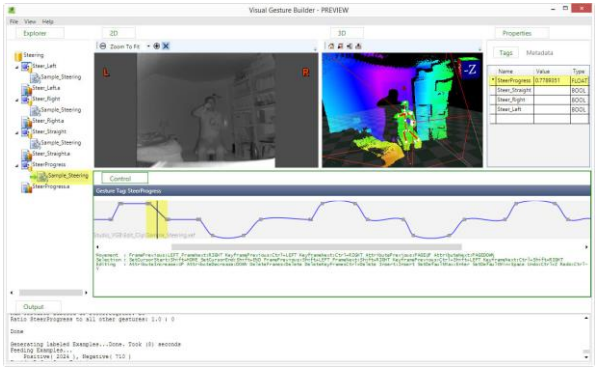
Data-driven process of creating a gesture detector using VGB



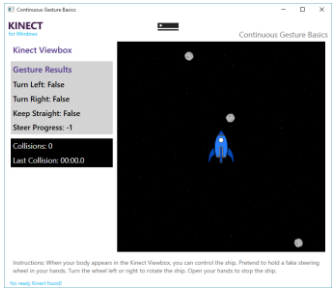
Preview gestures with VGBView



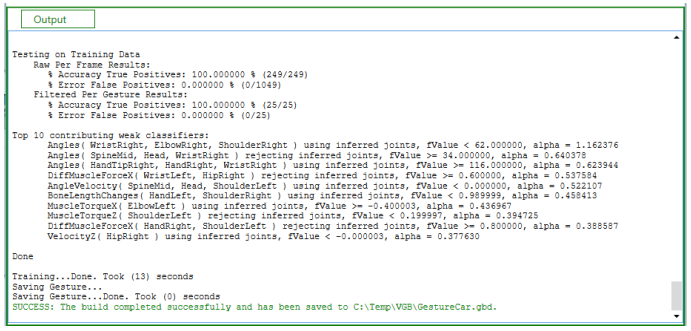
Record gestures using Kinect Studio



Tag and analyze gestures using VGB



Detect gestures in application



Build gestures using VGB to output

Demo

Zurich University of Applied Sciences

How Gesture Recognition for Testing Data?

- Needs at least programming for own application and is an engineering task
- Therefore a good direct comparison between Steering and Car as gesture

Steering	Car
Actions	Detections
3 discrete states: *SteerLeft, SteerRight, KeepStraight*	2 discrete states: *CarHandUpLeft, CarHandUpRight*
1 continuous state: *SteerProgress*	no
• Actions by discrete states for change of direction • Action by continuous state for change of angle as more or less rotation	• Detections for start state and end state • Additional number of sequence: repeated twice
Used both AdaBoost and Random Forest	Used only AdaBoost

Demo

Gesture detection in application

- Correct display of a gesture data set of test examples
- Following illustration shows gestures for Car, Thursday and Milk

Results

Zurich University of Applied Sciences

zh aw

Result of Gesture Recognition

- Each gesture data contains nearly 5 similar clips by the same person
- Evaluation for results is subjective because training data is too little at the moment
- Accuracy
  - Mostly True Positives successful deteced (without specification of perecentage)
  - Rare False Positives detected
- Latency
  - Relatively very little to none, but speed of movement of gesture should be fair

Difficulties

- Examples for gestures as Ship and Plow are nearly identical because hand detection has only simple hand position as open, close and lasso
- Problem with triggering of lower confidence value at change of discrete state from false to true in the tagging frames

Conclusion

Zurich University of Applied Sciences

Takeaways

- Use Visual Gesture Builder
  - Results speak for themselves: rapidly productivity with tagging data and non-engineering task
- Invest in quality assurance for tagging gesture data
  - Tagging plays an important role in good results and increased accuracy
- Improve Accuracy by
  - Using enough positive and negative training examples
  - Of a wide variety of different signing persons

Application Area

- Study course for Applied Linguistics in the Institute for Translation and Interpreting
- Knownledge transfer to research and teaching in the Institute for Information Technology

Technical Progress

- Possible usage of prototype with approach to Deep Learning
  - Better performance for complex grammatically structure of sign language

Questions

Zurich University
of Applied Sciences

zh
aw