Visione artificiale: nuove applicazioni per il mondo della logistica integrata e scenari futuri

Confindustria Modena

Club Digital&ICT
Confindustria Modena, 27/02/2017
Nuove tecnologie ICT al servizio della logistica integrata

Prof. Rita Cucchiara
Dipartimento di Ingegneria «Enzo Ferrari»

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Centro Interdipartimentale di Ricerca
Softech: ICT per le Imprese

Imagelab

New Master UNIMORE 2017
Opening Soon!

MuMeT 2017
visual computing and multimedia technologies

Master «Visual Computing and Multimedia technology in the Deep Learning»

Logistics and industry 4.0

The seven Rs

- "Logistics is about getting the right product, to the right customer, in the right quantity, in the right condition, at the right place, at the right time, and at the right cost “

In industry 4.0:

- Everything is digital → ICT
- Everything is Smart →AI
- Everything is distributed and autonomous→IoT
- Everything is cooperative → HMI, MMI

In «Industry 4.0»

Artificial Intelligence

Sensors and IoT

Cognitive Systems

- Computer Vision
- Natural Language processing
- Pattern Recognition
- Machine Learning and Deep Learning
- …

…And many other Information technologies
- Mobile and distributed systems
- Software Engineering paradigms
- Real-time high performance systems
- CyberSecurity

….

Artificial intelligence: The scientific field which studies how to create computers and computer software that are **capable** of **intelligent** behavior, using Sensing, Perception, Knowledge, Reasoning and Learning.

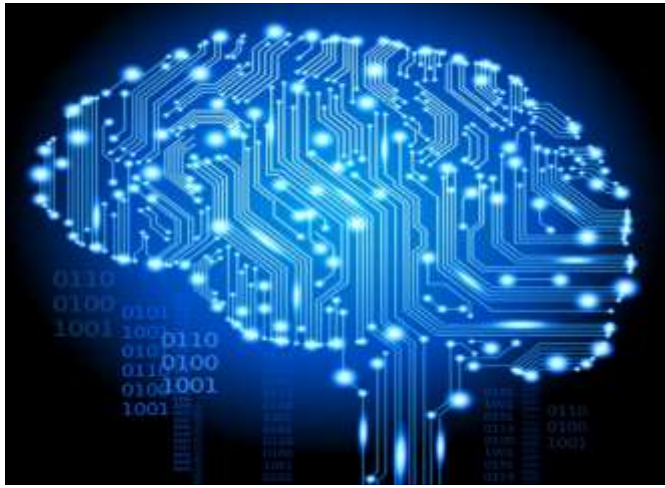
Machine Learning: The scientific discipline studying **how to constructs algorithm that can learn from and make predictions on data**, for getting computers to act without being explicitly programmed.

Deep Learning: A branch of Machine Learning for modeling and implementing deep neural network architectures and algorithms.

Pattern Recognition: The scientific discipline studying how to classify or recognize patterns and observed data using a priori knowledge, statistical information and learning

Computer Vision: the scientific discipline studying **how to perceive and understand the world through visual data by computers.**

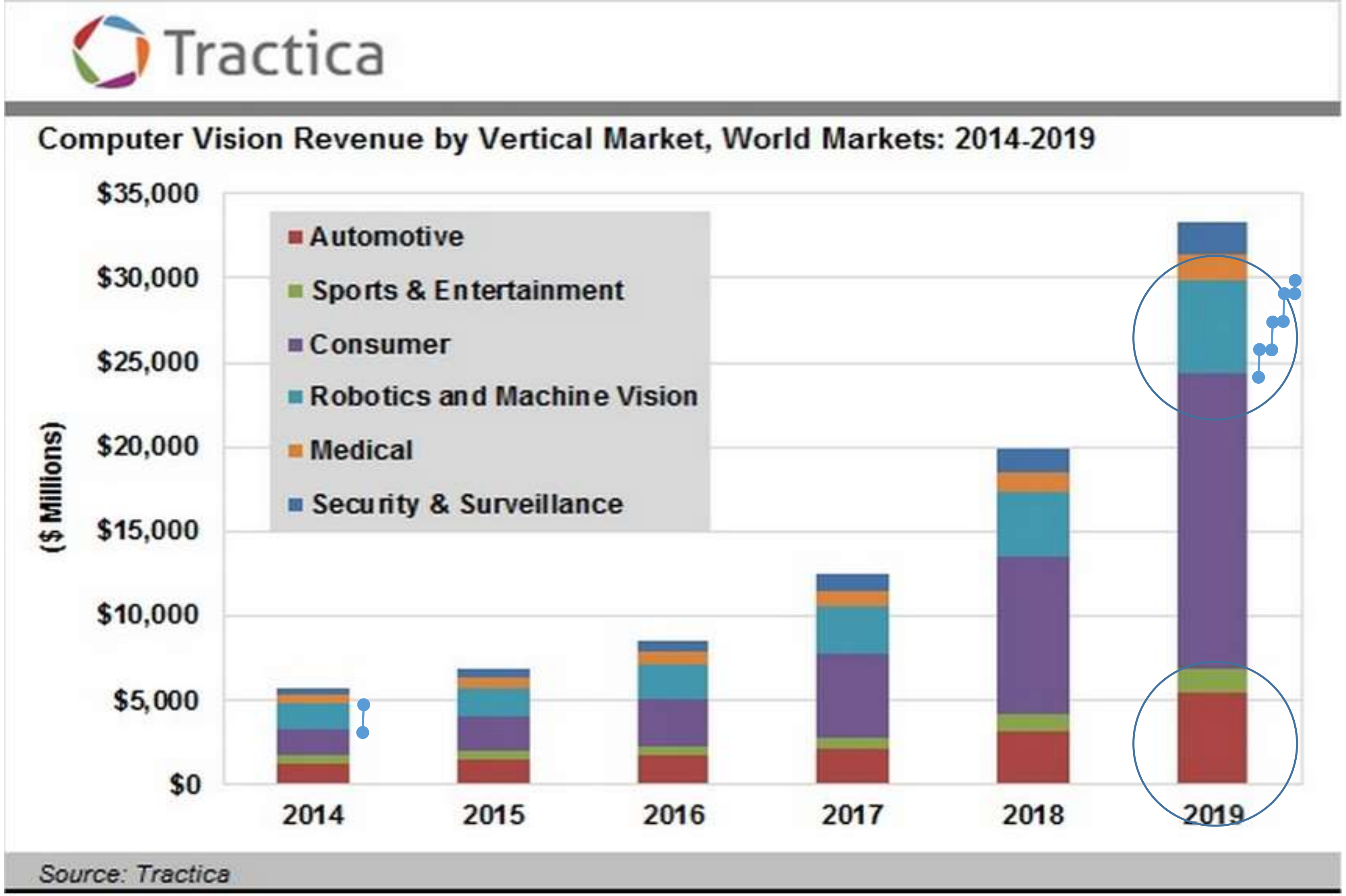
Machine Vision: the engineering field studying how to build **computer vision-based** systems, services and solutions, typically for industrial environment.



And related applications:
e.g. Video-surveillance, Medical Imaging, Machine Vision, Automotive, Biometrics, Building Automation, Smart Cities, Industry 4.0, Digital humanity, Big data analytics, Remote Sensing…

Why it is so important? The Market

The market for **computer vision** technologies will grow from $5.7 billion in 2014 to $33.3 billion by 2019, representing CAGR of 42% (CAGR compound annual growth rate)



RnRMarketResearch.com

*The **machine vision** market size is estimated to grow from USD 8.08 billion in 2015 to USD 12.49 billion by 2020, at an estimated CAGR of 9.1% from 2015 to 2020.*

***3D Machine Vision Market Global Forecast to 2020** says, the market is expected to grow at a CAGR of 10.53% during the forecast period between 2015 and 2020 driven by 3D machine vision technology is due to its growing applications in the automotive and electronics industries.*

In "Automated Guided Vehicle Market", the total market is expected to reach USD 2.81 Billion by 2022, at a CAGR of 10.2%

Computer Vision dimensions in industries

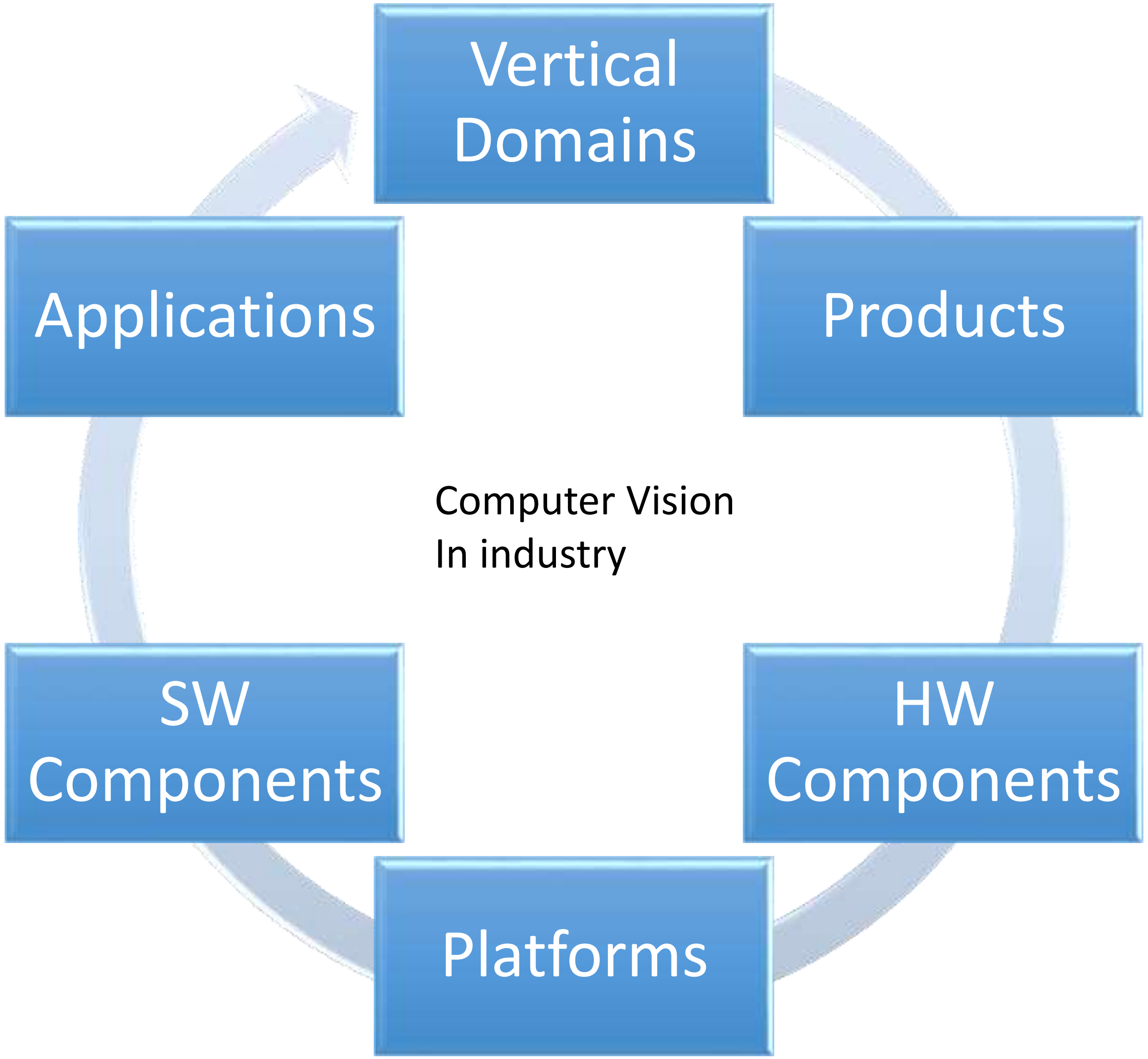
An holistic view
of Computer Vision
for industry

Web Sources
ec.europa.eu Fraunhofer 2015 (logistics in industry 4.0)

www.**ukiva.org** (UK Industrial Vision Association)

www.lapr.org (Int. Association for Pattern Recognition)
www.cvf.org (Computer Vision Foundation)
www.embedded-vision.com (Embedded Vision Alliance)

www.**vision**online.org commercial site
http://www.vision-systems.com/ commercial site



Products

- Computer /Machine Vision products
- General-purpose products (Services)
- Customized/embedded products (Systems)

Products:

- **Embedded custom systems**
- **Smart cameras-based solutions**
- **PC-based Machine vision systems**
- **Vision As A Service: services on cloud**

SoC
FPGA based (XILINX)
GPU Based platforms (NVIDIA)
DSP and microcontrollers

Smart cameras with own sw
Cognex, Datalogic, Matrox, NI, Vision Components

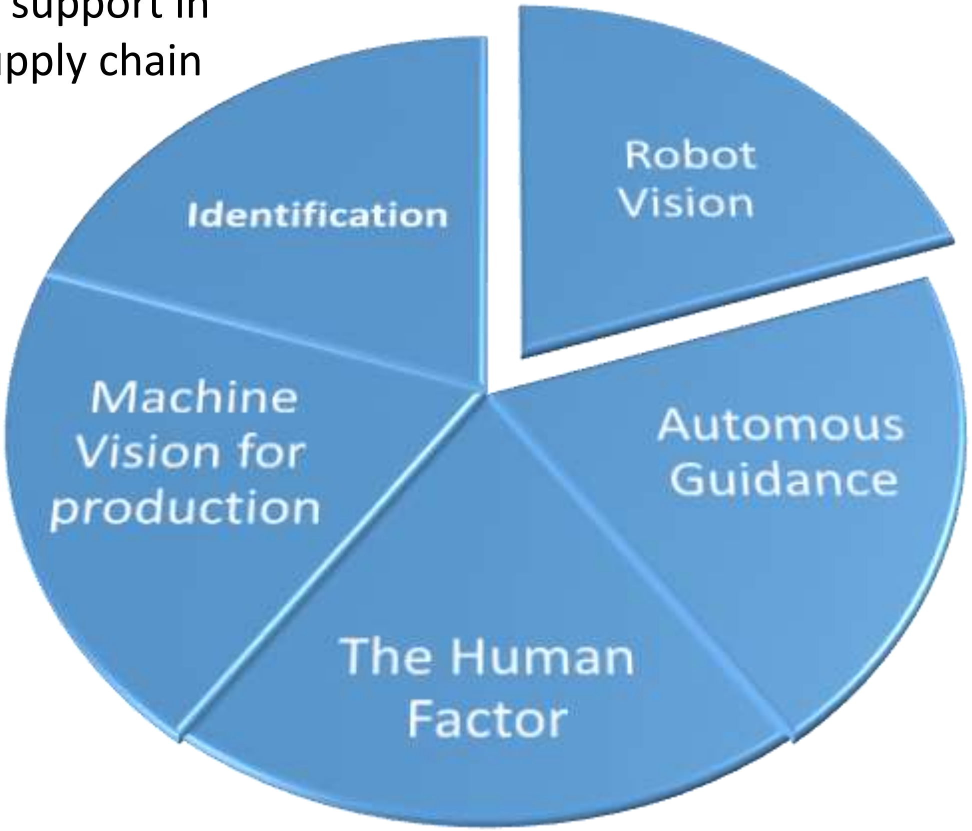
Smart cameras with third-party sw
Matrix Vision with MVTec HALCON, Adlink Tech with HALCON, Adaptive Vision..

- New solutions and new business model for software and component suppliers
- The effort is more and more in software

Applications

Visual support in the supply chain

2D/3D manipulation (picking and stowing)



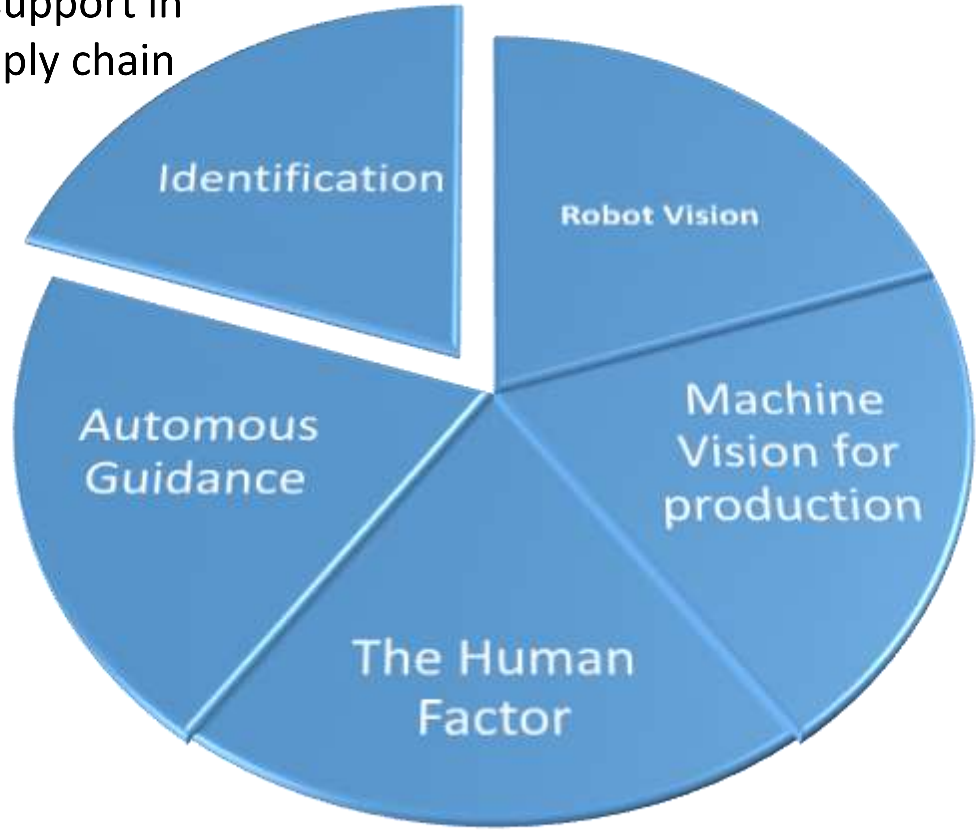
AVI Automatic Visual Inspection

Navigation and Situation Awareness

HMI Human Machine Interaction

Applications

Visual support in the supply chain



Identification- Label reading

A Code 39: non-retail environment ; standard DoD and E Health industry contains 43 chars

B Atzec code: 2D code more compact, used in transportation industries and in Airlines

C Maxicode: defined in 1992, stores 93 chars, used for fast packages in automated conveling systems

D Code 128: high density 128 chars ASCII, used in United States Postal Services

E PDF417 : a stacked linear barcode symbol used primarily transportation, identification cards, and inventory management.

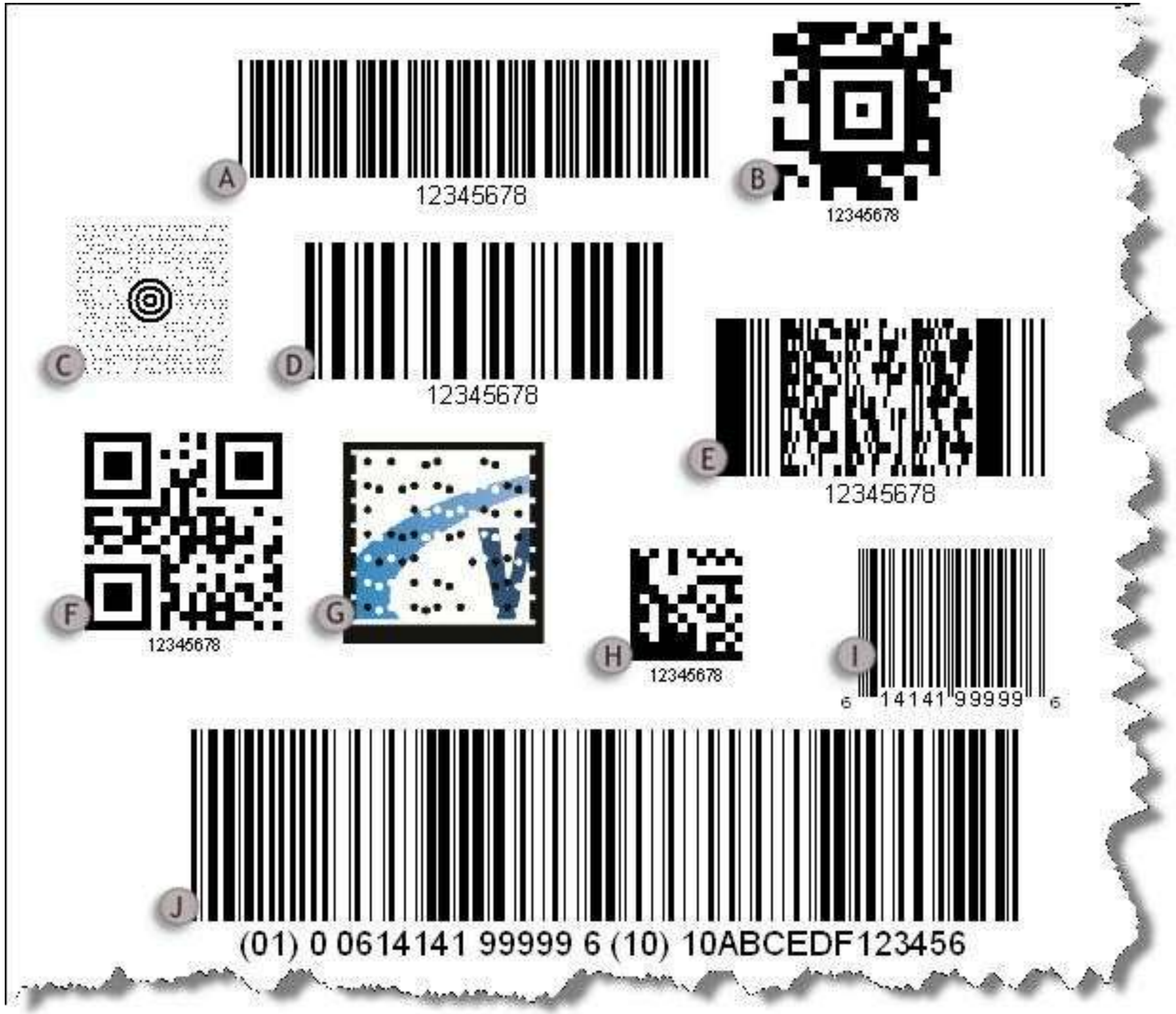
F QR Code: Quick response Code, initially in automotive in Japan, common in cunsumer adverstisment to be converted in a URL

G HCCB High capacity color barcode, is a Microsfot Tag with a palet of 4-8 colors , used expecially in mobile

H Data matrix: rectangular patterns also for 2.3 mm and readbable with a 20% of contrast, scalable in very small (600 micro) to large(1 mq) Used in USA's Electronic industry alliance

I UPC Code: Universal product Code for tarcking objects in stored 12 digits

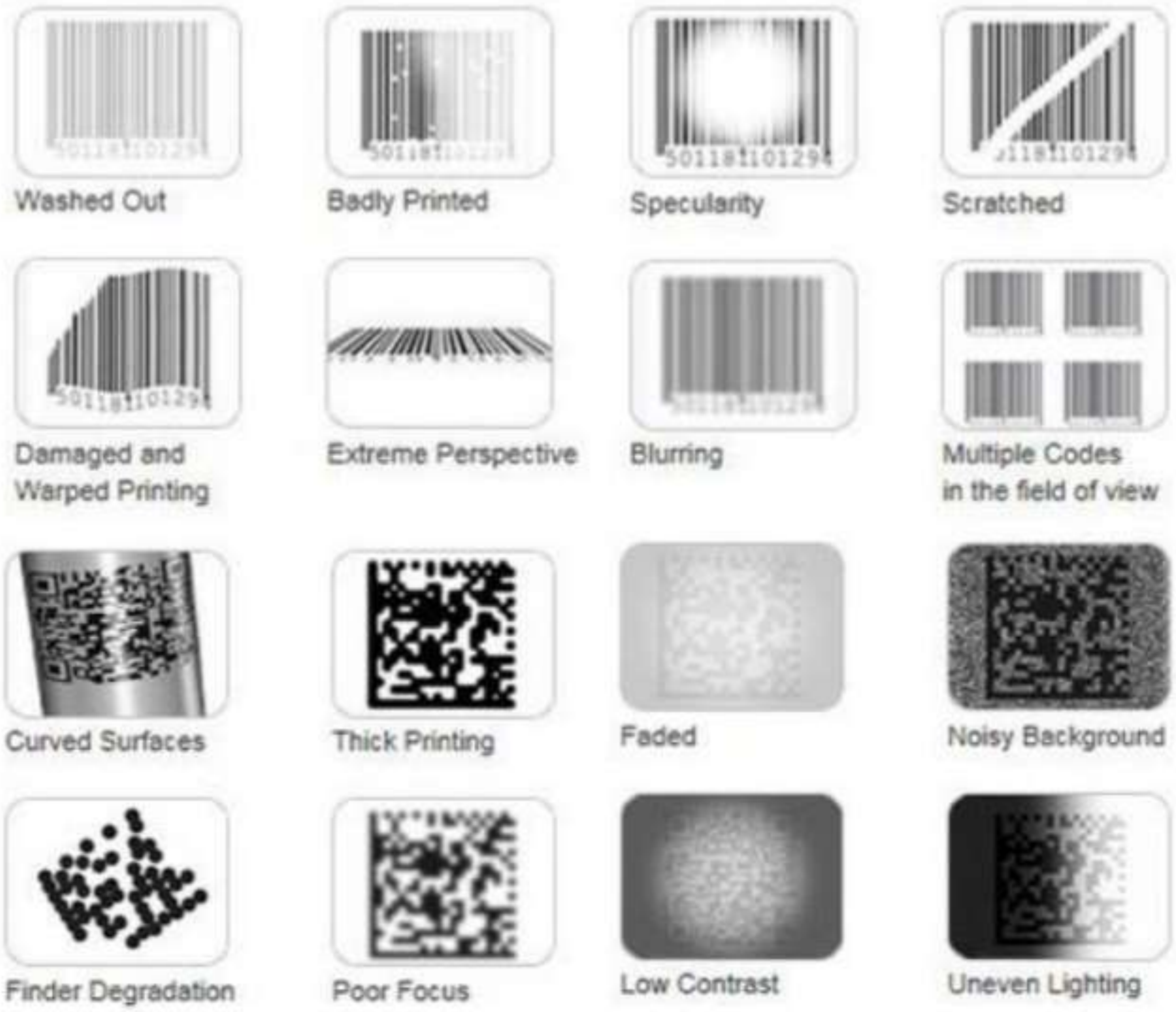
J Gs1-128 subset of Code 128 used for containers and pallets for additional data as quantities or batch numbers

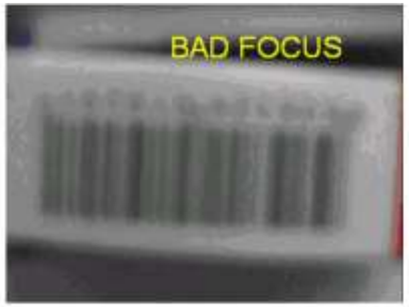


Identification- Label reading

- Laser Scanners
- Camera-based Reading
- Mobile-based reading

- Now image processing tools can correct errors and ambiguities





GOOD (after Manual Focus ring adjustments):

- Next step: direct OCR **optical character recognition**
- **A**dditional identification of uncoded text
  - Correct placement of labels (orientation detection)
  - Verification of the presence of the logo (template matching)
  - Quality inspection (color, shapes, contrasts etc)

Classical Methods
- A) template matching
- B) feature extraction and k-means clustering

Available software
- Cognitive OpenOCR
- Tesseract Google
- Now DL

Siemens Postal, Parcel & Airport Logistics GmbH

- Tesseract Google OCR

- 800 Chars needed for Training
- Avg Trainig Time 10 minutes
- Core i7 PC NO GPU

- Deep Neural Network

- 5000 chars needed for Training
- Avg Training time 30 minutes
- Core i7 PC + NVIDIA GPU CARD

Technology	Accuracy
Tesseract eng language	40%
Tesseract trained language	60%
DEEP neural network(NN)	**98%**

DEMO code @
http://christopher5106.github.io/computer/vision/2015/09/14/comparing-tesseract-and-deep-learning-for-ocr-optical-character-recognition.html

Identificiation and OBJECT AND SHAPE RECOGNITION

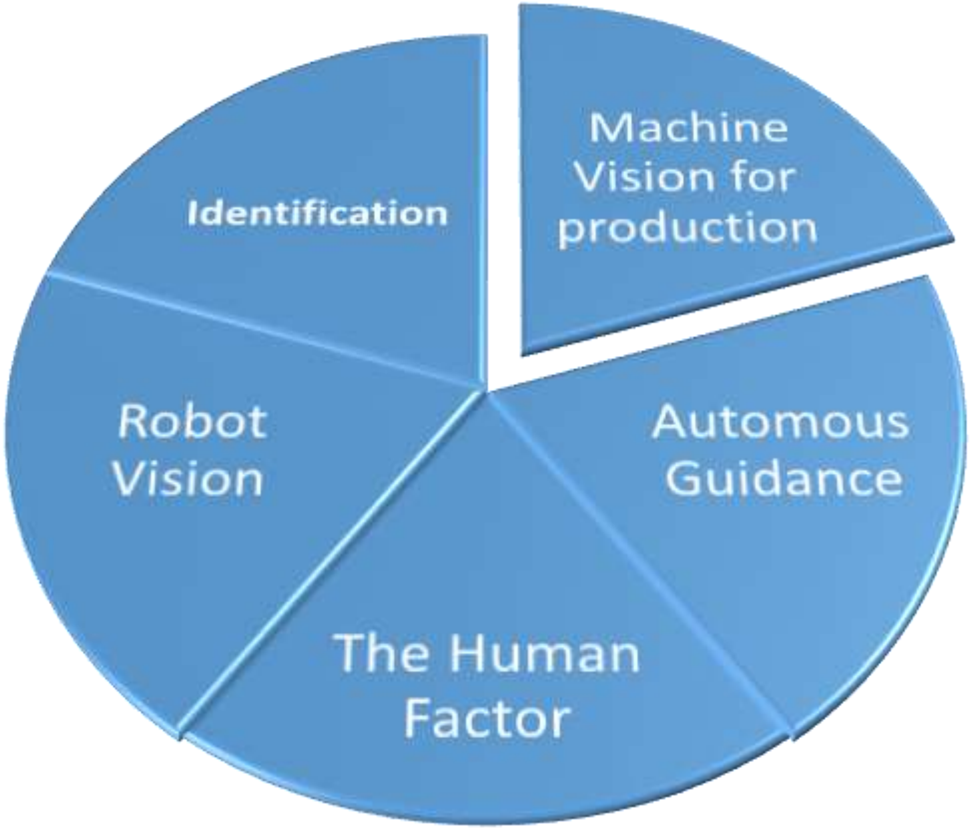
Computer vision now can:

- Read from printed labels and provide OCR
- Recognize logos
- Recognize 2D and 3D shapes also with affine/perspective transformation
- Recognize under deformation with Grab cut and Cnns…. New frontiers!

CNN

Trained with Imagenet

Applications



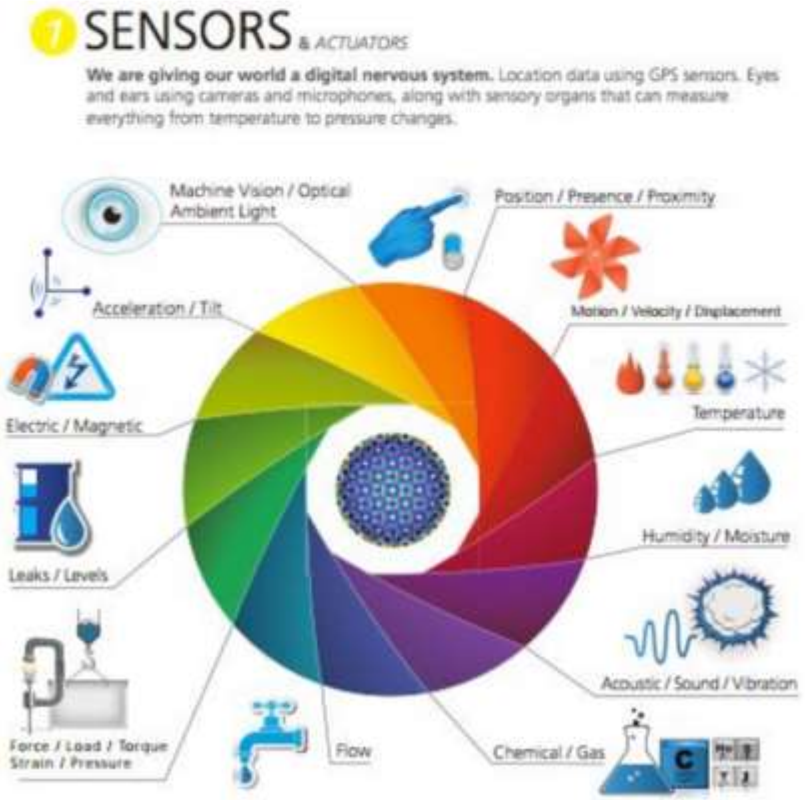
AVI
Automatic
Visual
Inspection

Automated Visual Inpection

AVI (Automated Visual Inspection) one of the first use of Vision in logistics and production

The seven AVI constraints:

1. Real-time processing
2. Illumination
3. Acquistion Issues
4. Selection of features ..
   - color,
   - shapes (Template matching, contour filling.. Convex hole)
   - texture, frequency-based (Gabor, Wavelet, Furier..),
   - keypoints, (Sift, Surf …)
   - 3D building boxes
   - Convolutional NN Features
5. Selection of suitable classifiers and computer vision tasks
   - Bayesian, SVMs, KNNs.., DL architectures
6. The lack of significant examples (eg defective ws non defective targets)
7. The need of find a new solution for each new problem.



Quality control, inspection , measurement

The standard pipeline

Model

Data

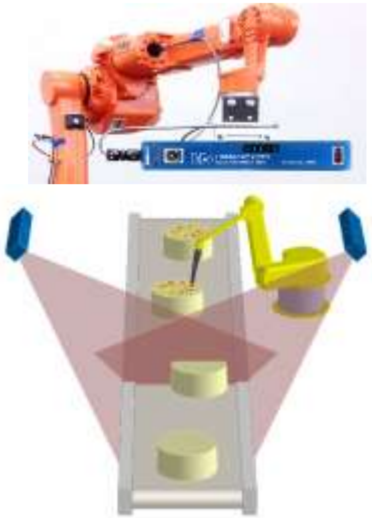
Image processing → [Segmentation and detection] → Feature extraction → Classification, identification..

DEEP LEARNING

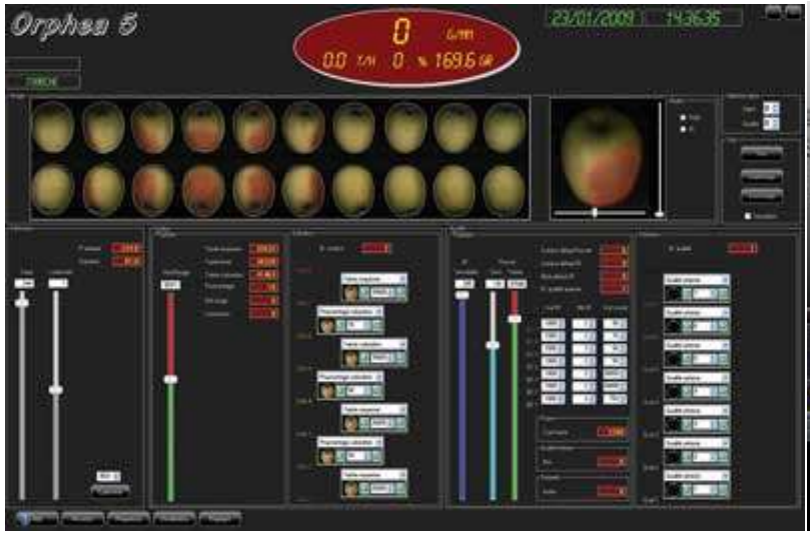
Products are designed for a class-specific application
and optimized for the specific context

often with a **precise product CAD model**

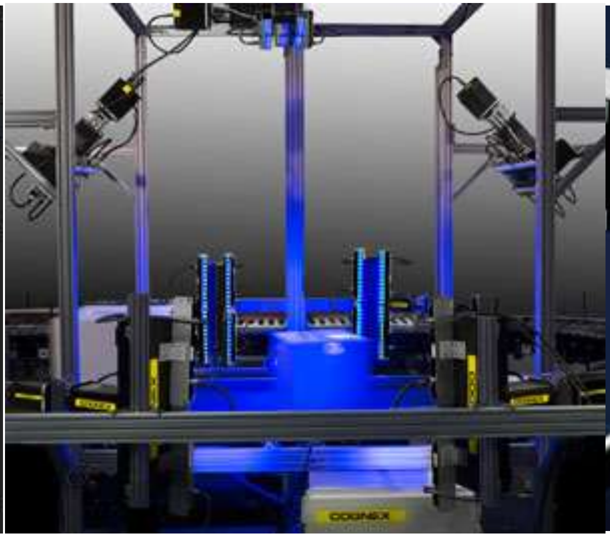
3D Hermary Light Scanner



Matrox Imaging based apple measurement



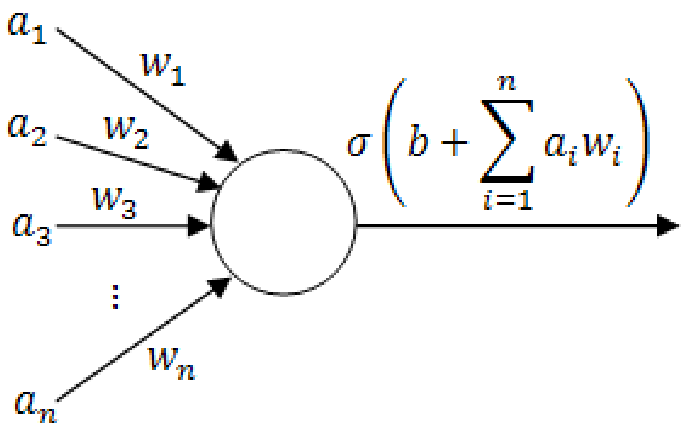
Cognex Dataman 503 UV-based scanner

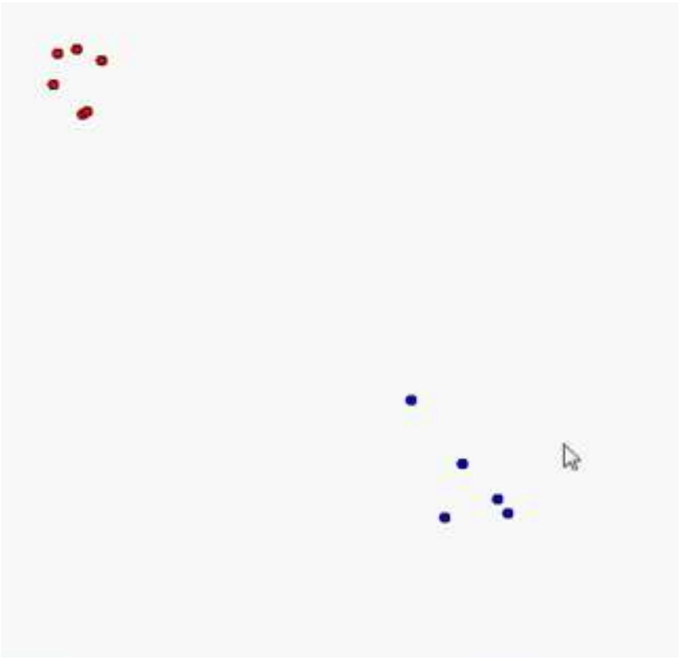


AIS http://aisukltd.com/machine-vision-systems/

- Perceptron is the analogous of a neuron
- Computational model -> perform linear classification

- **Perceptron is a linear Classifier**

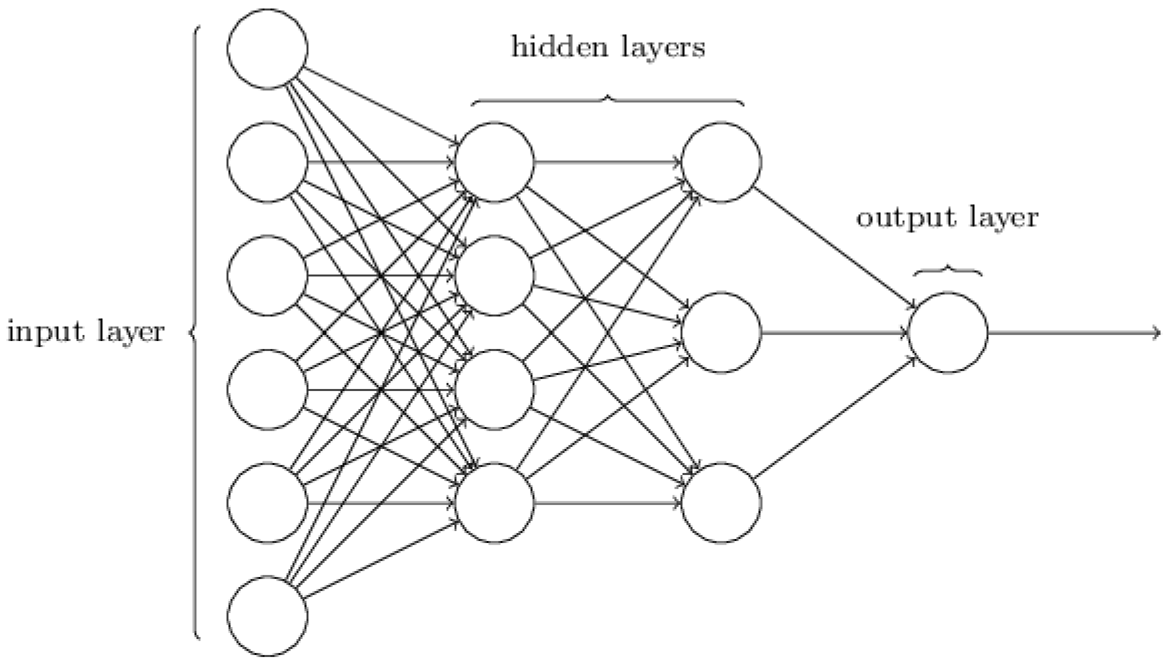


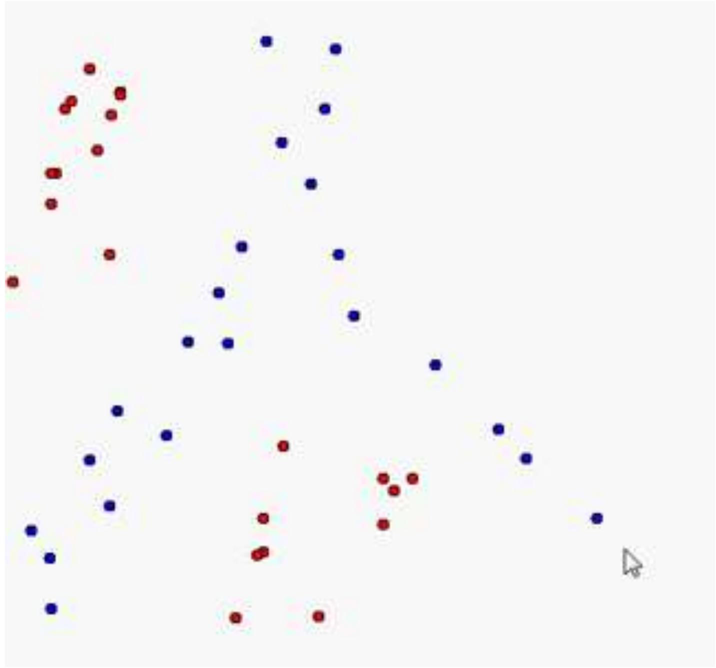


Multilayered Neural Network

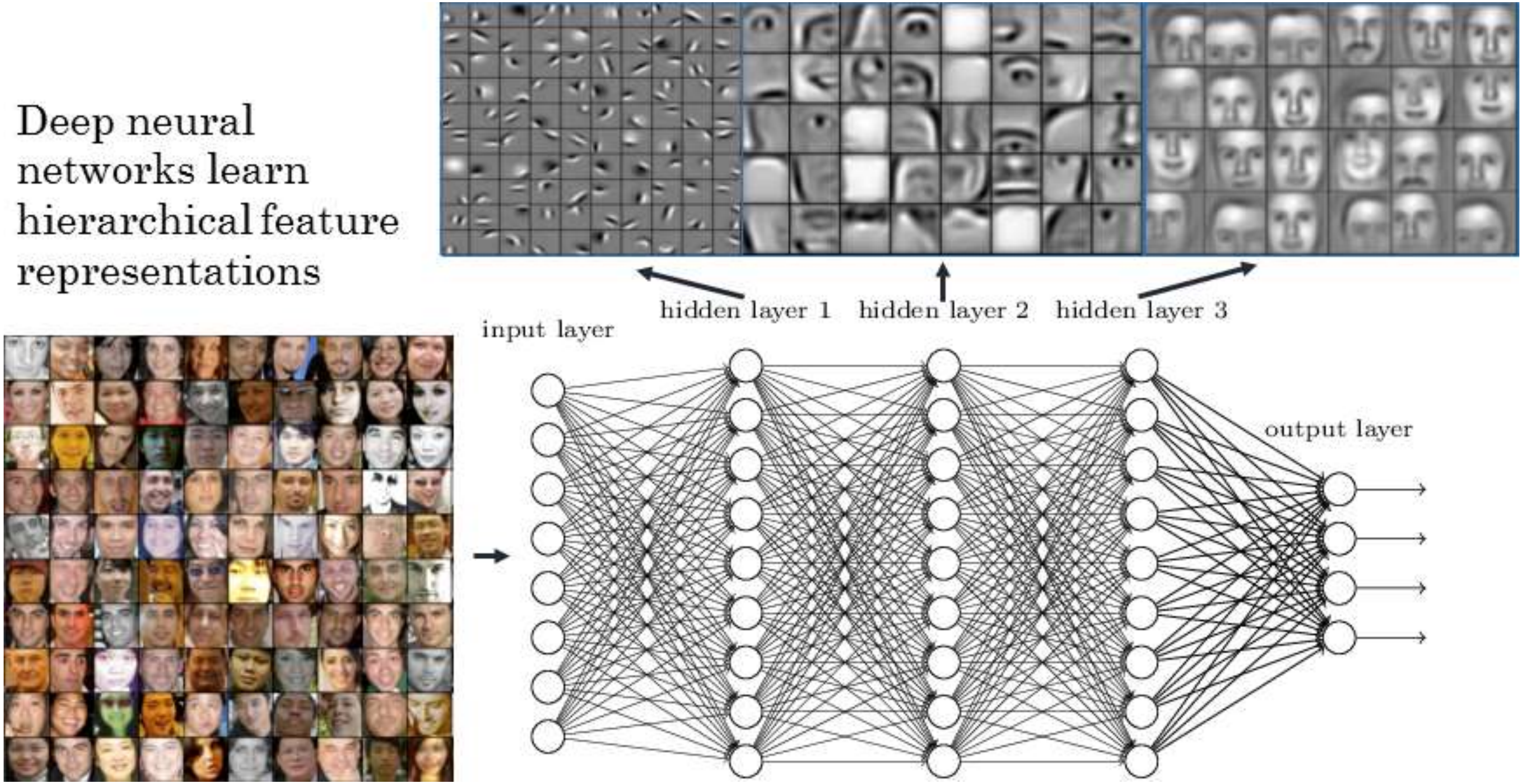
- Stacking perceptrons **vertically** we obtain a layer
- Stacking layers **horizontally** we obtain a network

- **Network With 3 layers is a non-linear classifer**





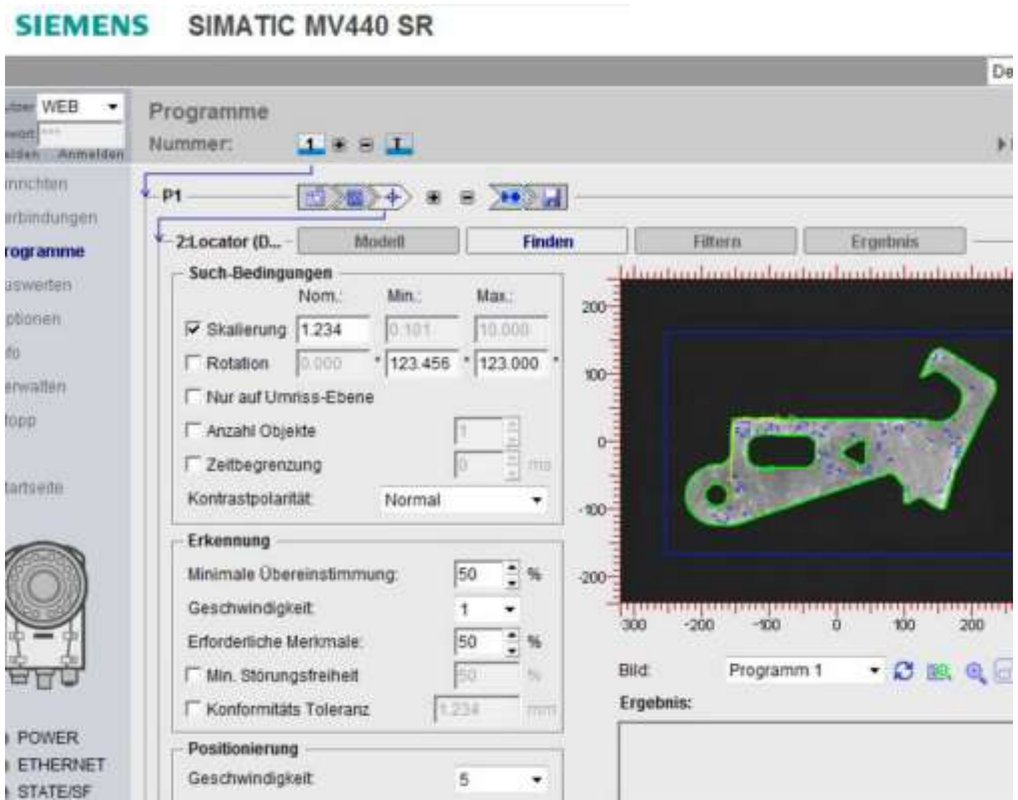
Deep neural networks learn hierarchical feature representations



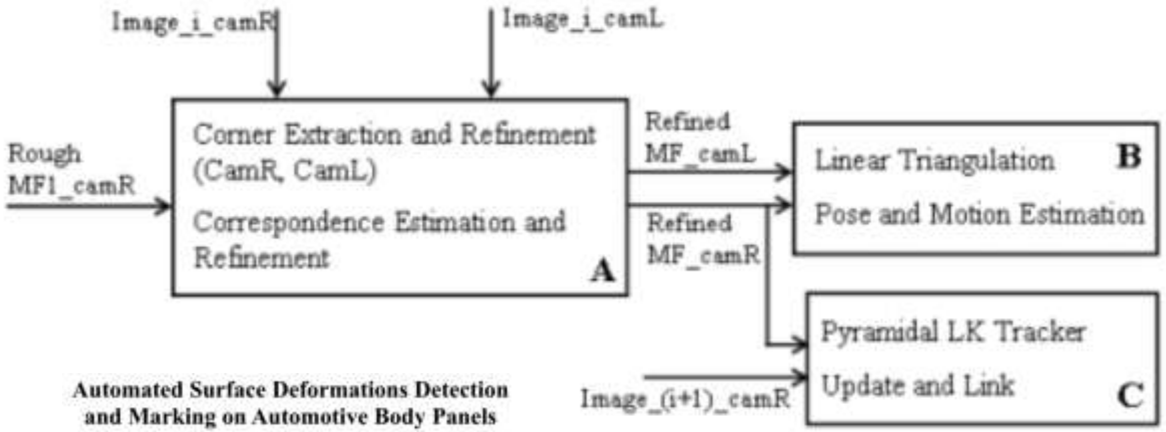
Model-based vision: products and tools are on-the-shelf

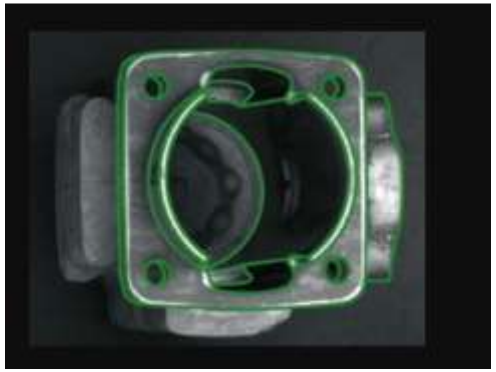
- Eg. SIEMENS SIMATIC MV 440
- Pat-Genius" object recognition license, SIMATIC MV440 for object recognition position detection, counting etc., reading 1D bar codes and 2D matrix codes, text recognition, to check the position of a label and check the inscription (reading and comparing) of plain text in an image field.
- 2500 checks/min
- the object CAD model is required

- Many products and companies use it.
- e.g. SACMI



Now improved inspection in 3D , with Stereo and ToF

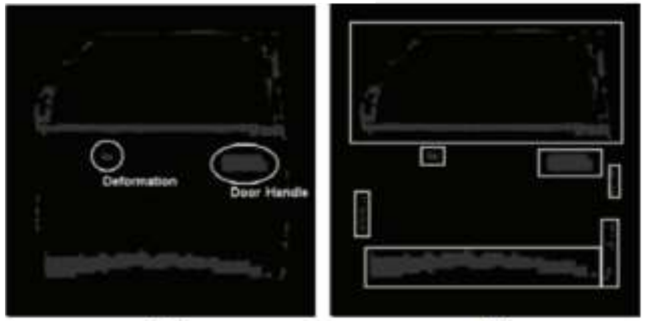




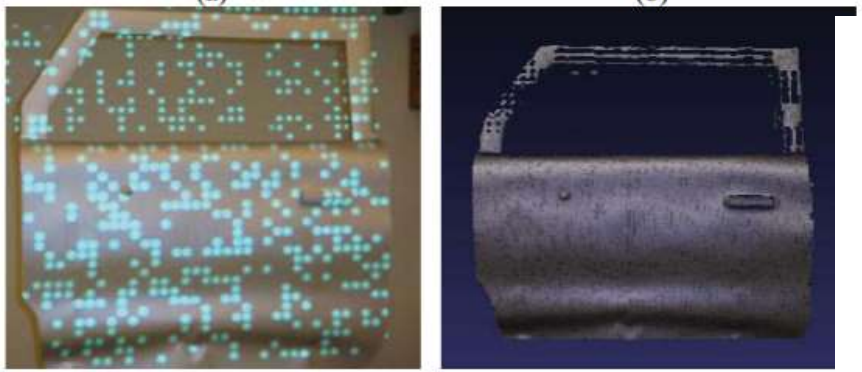
Automated Surface Deformations Detection and Marking on Automotive Body Panels

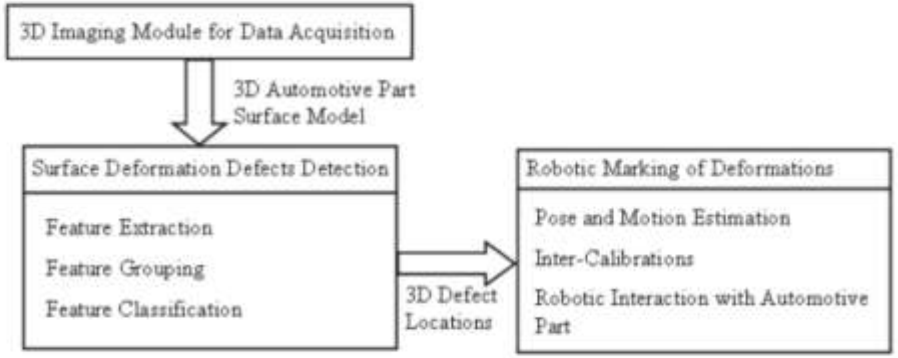
Valentin Borsu, Arjun Yogeswaran, and Pierre Payeur

(a) (b)



(a) (b)

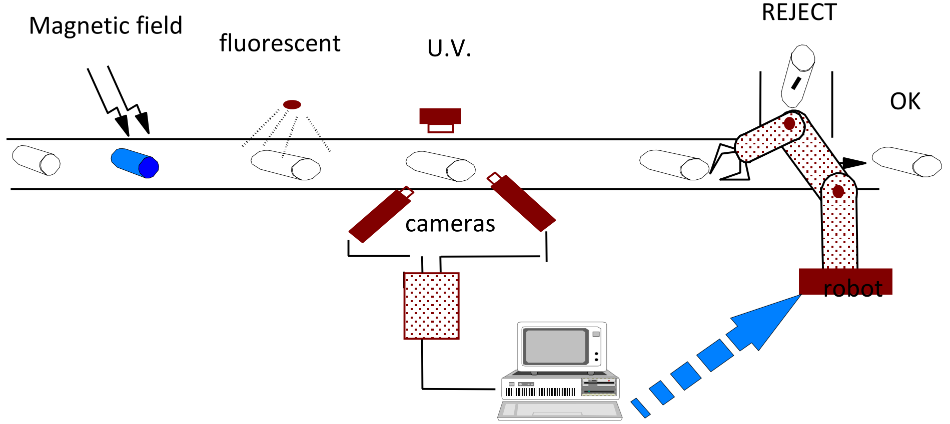


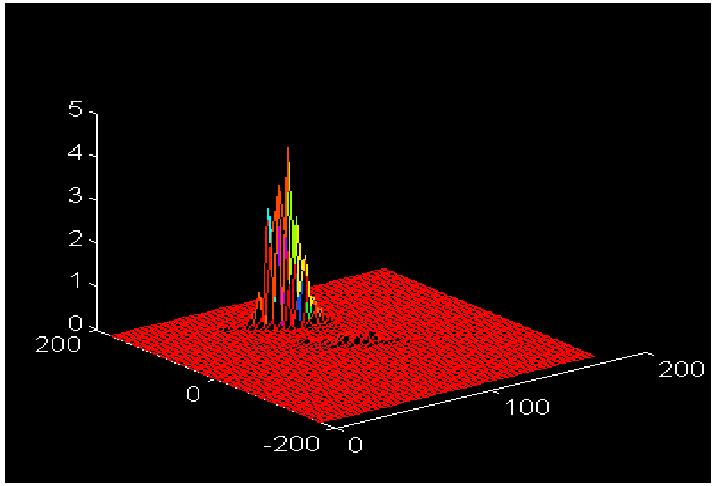


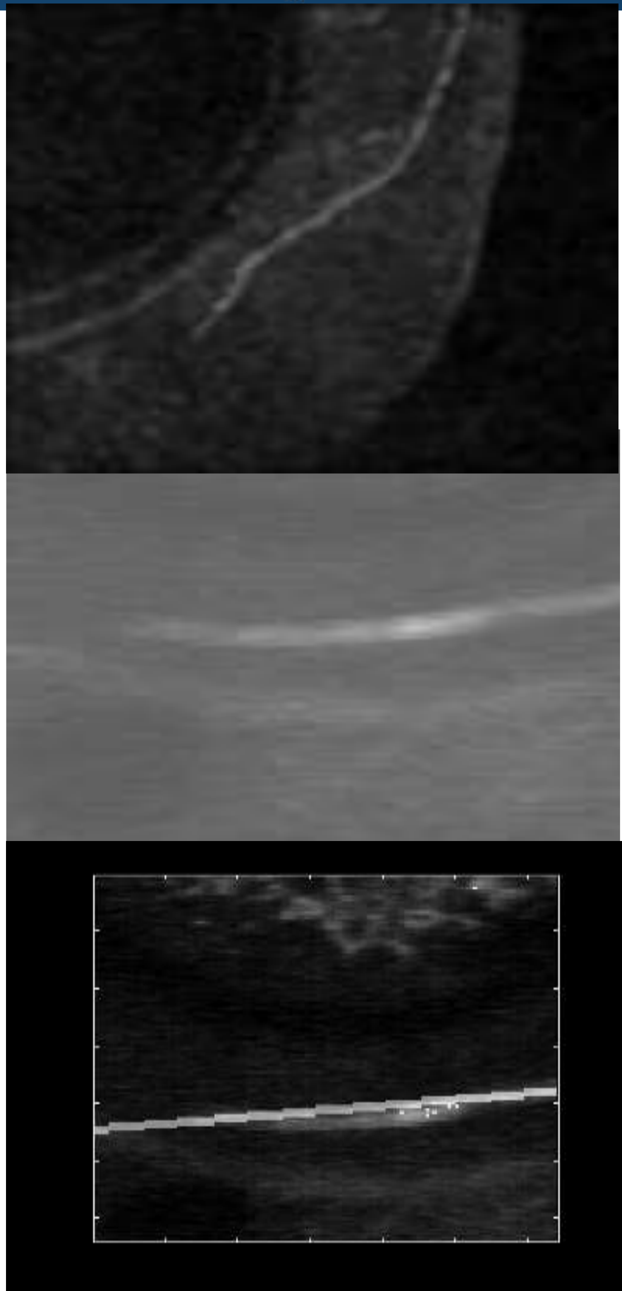
Vision and learning in quality inspection

- When the CAD model is not necessary
- The defect model is needed

- a VERY OLD story. 1995 BERCO,
- Thin and straigth crack detection







*R. Cucchiara, F. Filicori, R. Andreetta,"Detecting micro cracks in ferromagnetic material with automatic visual inspection"in Proceedings of the Intern Conf. Quality Control by Artificial Vision QCAV' France 1995,
R. Cucchiara, F. Filicori,"The Vector-Gradient Hough Transform« IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, n. 7, pp. 746-751, 1998

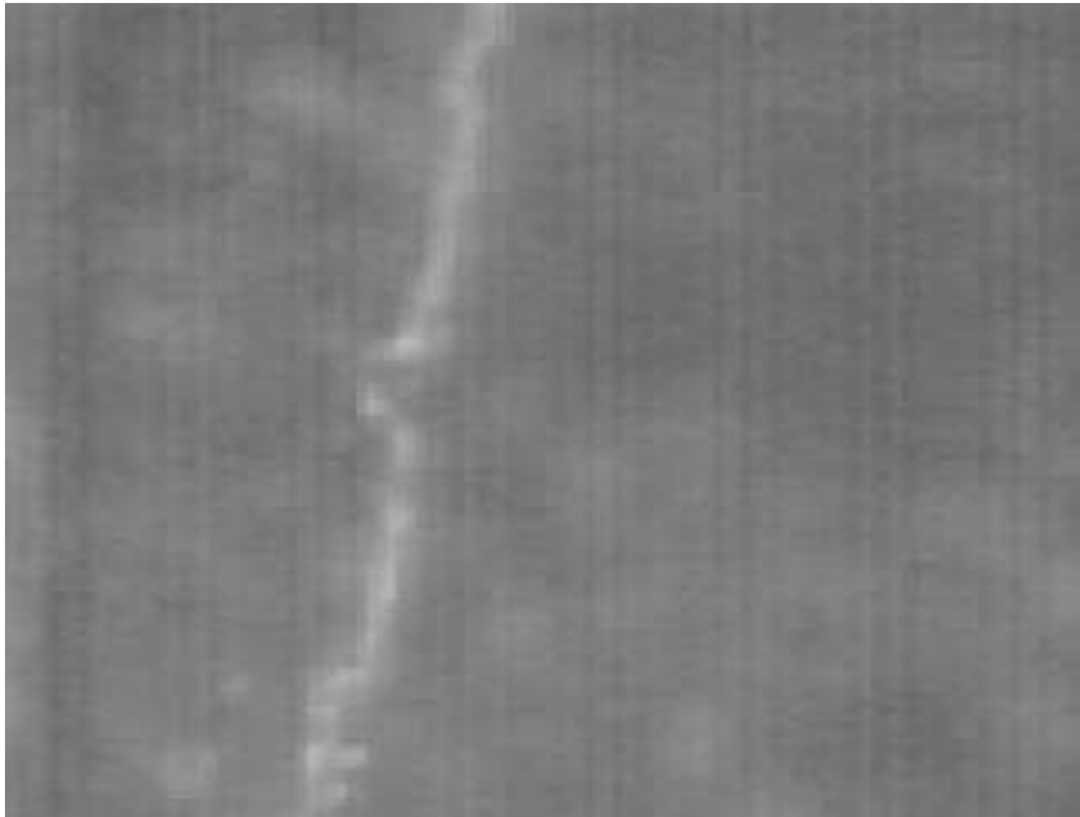
- Defective and non-defective industrial workpieces
- Six different learning algorithms: (in ’98*)

✓ **Artificial Intelligence**: an attribute-value learner, C4.5,

✓ Pattern Recognition: a **backpropagation neural network**, NeuralWorks Predict,

✓ Pattern Recognition: a k-nearest neighbour algorithm,

✓ Statistical analysis: 3 techniques, linear, logistic and quadratic discriminant.



A rule-based (or tree based) classifier capable of reasoning as humans do

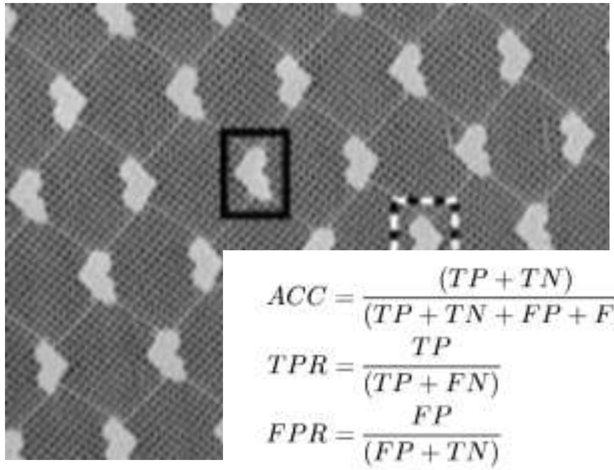
Table 1. Average accuracies

	Discrim	Logdisc	Quadisc	NN	Predict	c4.5 tree	c4.5 rules
CH	0.853	0.857	0.853	0.885	0.873	0.959	0.959
H1 H2	0.855	0.928	0.316	0.845	0.864	0.933	0.933

*R. Cucchiara, P. Mello, M. Piccardi, F. Riguzzi «An Application of Machine Learning and Statistics to Defect Detection”*Journal of Intelligent Data 1998*

Defect detection with Fisher criterion-based stacked denoising autoencoders (2016)

- Deformable defect detection in warp knitted fabric



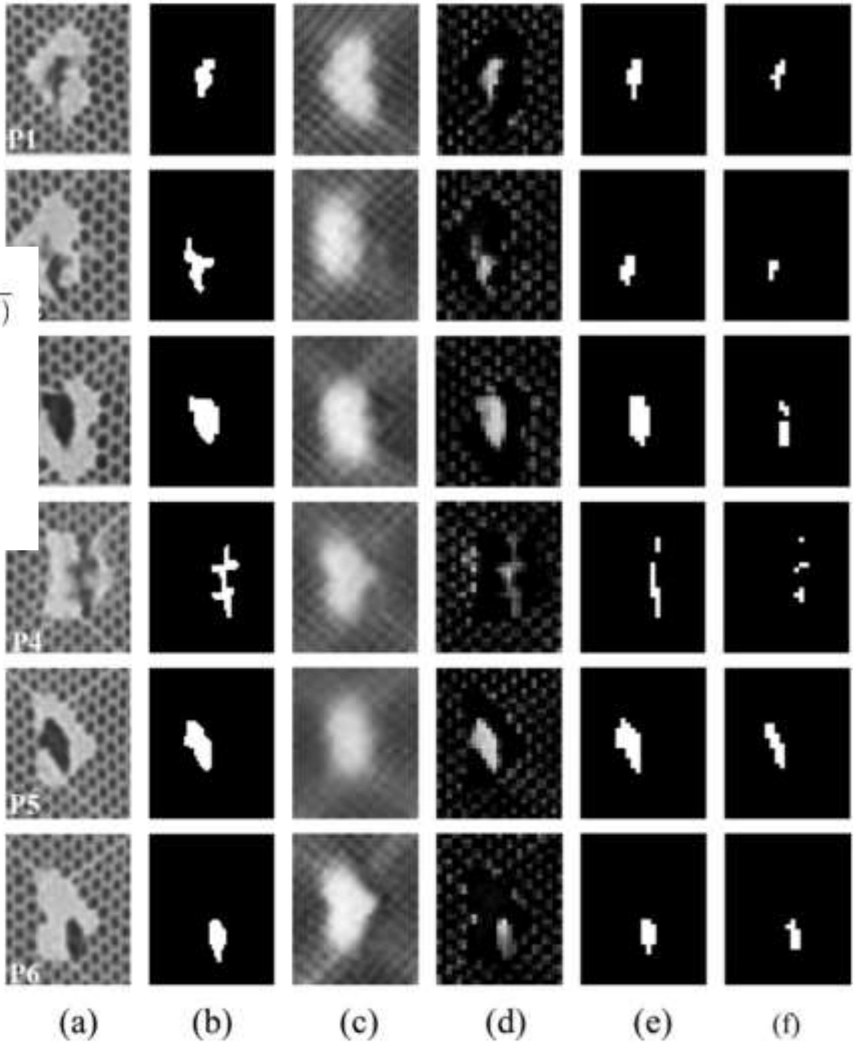
$$ACC = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$TPR = \frac{TP}{(TP+FN)}$$

$$FPR = \frac{FP}{(FP+TN)}$$

$$PPV = \frac{TP}{(TP+FP)}$$

a) examples
b) Ground truth
c) Image reconstruction
d) Difference with FCSDA
e) Detection with SDA
f) Detection with ID (texture)



Deformable Patterned Fabric Defect Detection With Fisher Criterion-Based Deep Learning

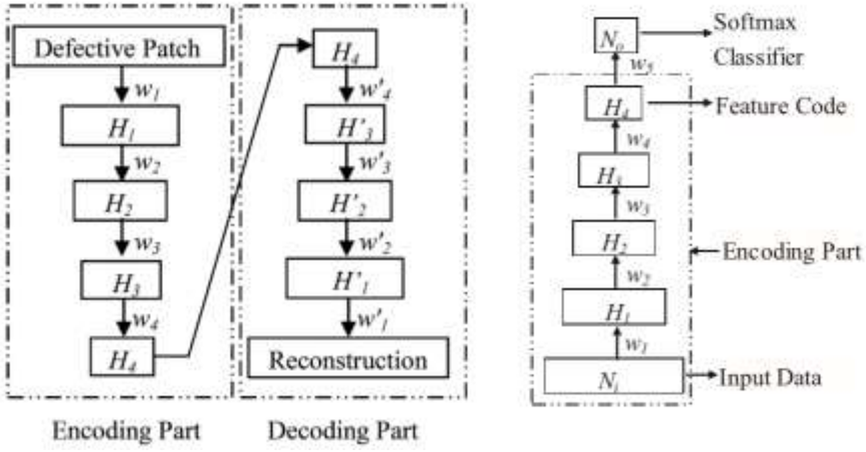
Yundong Li, Weigang Zhao, and Jiahao Pan

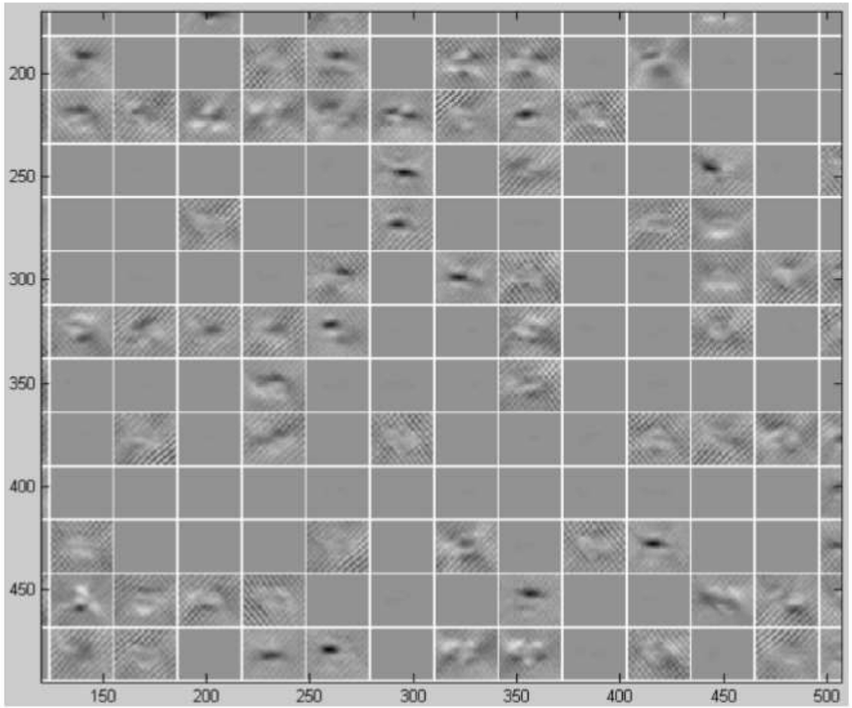
LOCATING ACCURACY COMPARISON OF EXPERIMENT 1

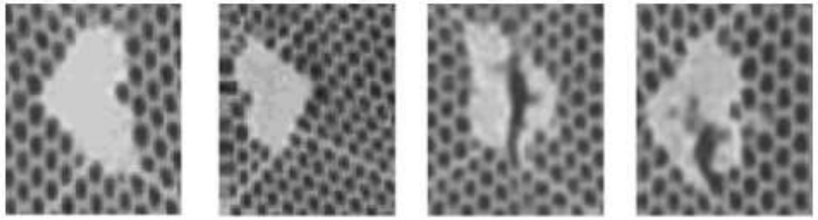
Defect Types	Methods	ACC (%)	TPR (%)	FPR (%)	PPV (%)
Broken End	ID	96.50	**81.99**	3.42	11.73
	SDA	98.58	73.96	1.28	24.18
	FCSDA	**98.66**	70.08	**1.18**	**24.71**
Hole	ID	99.14	**50.89**	0.61	**30.16**
	SDA	99.02	30.65	0.63	20.12
	FCSDA	**99.21**	10.71	**0.34**	14.06
Netting Multiple	ID	97.76	**19.31**	1.22	17.14
	SDA	98.42	8.41	0.41	21.13
	FCSDA	**98.50**	14.34	**0.41**	**31.51**
Thick Bar	ID	**99.19**	**97.99**	0.77	78.07
	SDA	99.06	70.23	0.13	93.94
	FCSDA	98.95	63.07	**0.04**	**97.83**
Thin Bar	ID	95.20	**83.29**	4.66	17.35
	SDA	97.89	30.79	1.32	21.51
	FCSDA	**98.21**	57.11	**1.31**	**33.91**

a detail of the autoencoder

- The encoder (and decoder for reconstruction)
- 30x25 patches = 750 vector
- 750,600,400,200,100,3
- 2000 positive, 600 defect







0,21ms detection on a standard Corei5 -- (300 /min)

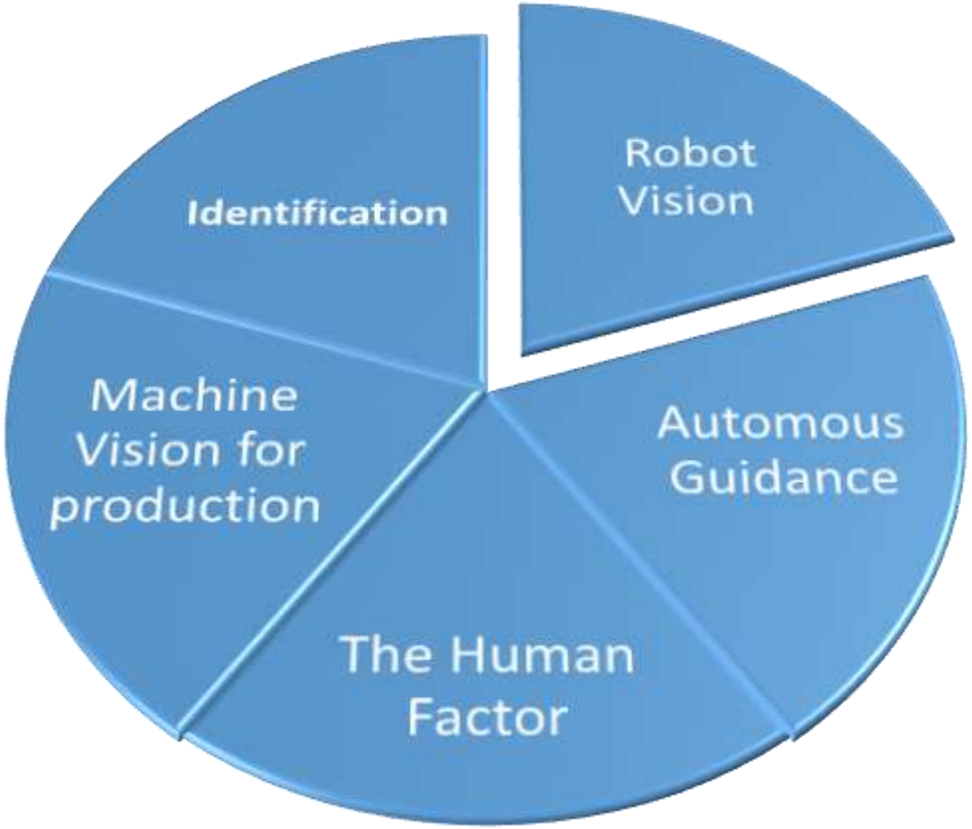
Target segmentation in images (Yoox-UNIMORE 2014)

- Object Segmentation
- Object aspect recognition
- for the supply chain

- Imaging for precise segmentation
- Learning people for fashion element Segmentation
- automatic color analysis

- Next step: learning by examples for precise categorization

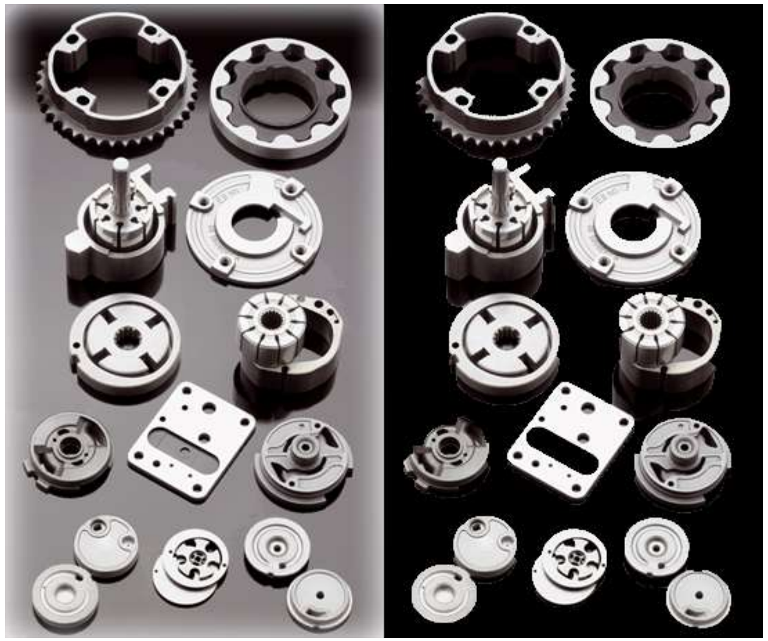
Applications



2D/3D manipulation
(picking and stowing)

Localization in 2D images Grab cut (and Deep contour)

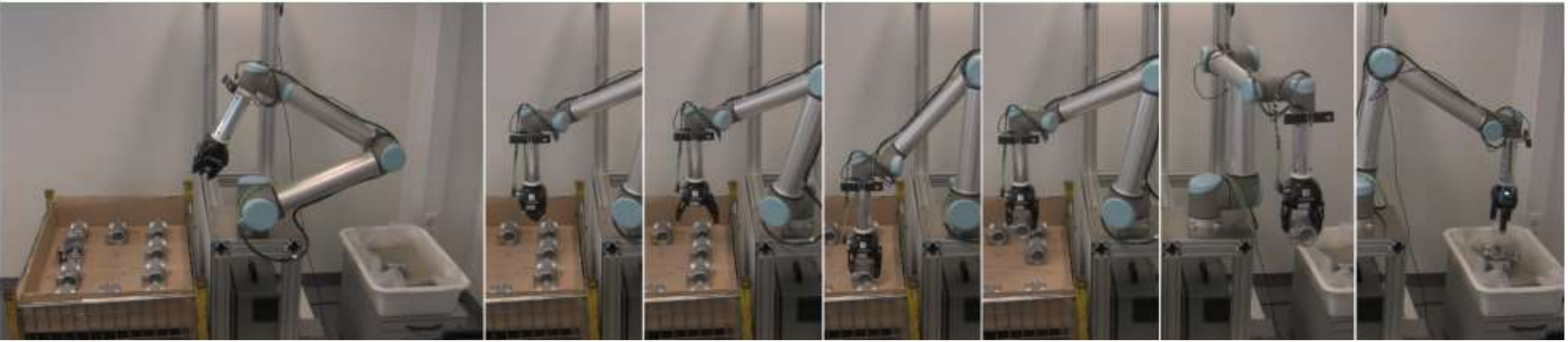
- Object segmentation for localization and grasping and supporting
- Image segmentation; enormous improvements in the last 5 years with
- Boundary detection
- Grab-cut based segmentation
- DeepContour (CVPR 2015) and DeepEdge (CVPR2015) based methods

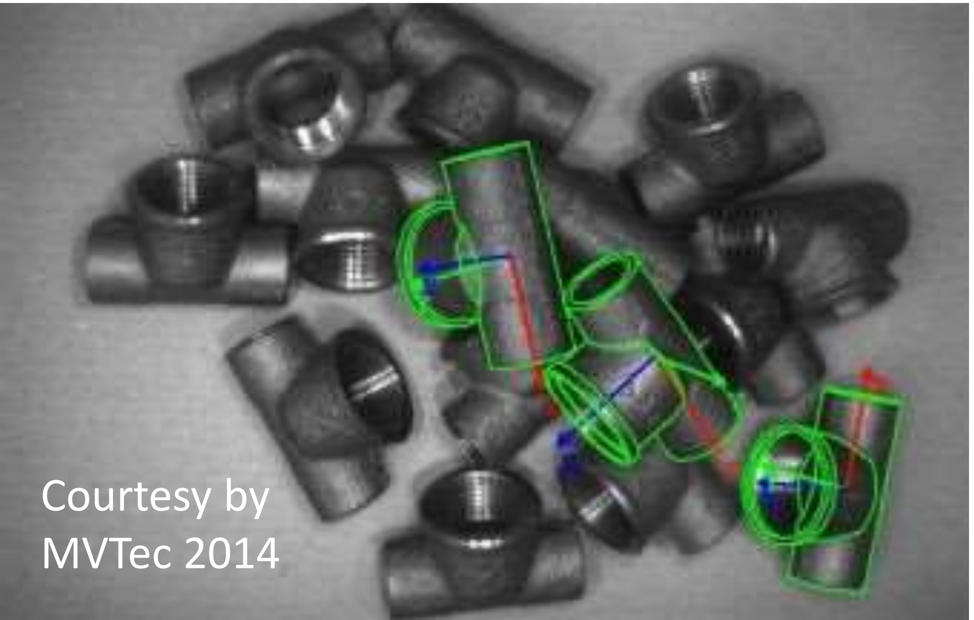


Precise object segmentation @Imagelab

2D-3D identification and localization for grasping

- Research in’90 Model based vision





- Now
- From single to multiple target
- Recognition and identification in a disordered bounch of objects
- Grasping and tiwing objects controlled by vision

Vision and learning for picking-and-place

- UNIMORE & Marchesini spa, Bologna (Patent . BO2009A 000278 2012)
- Paolo Piccinini PhD

➢ Different objects types and distractors
  ➢ No CAD Models

➢ Learning by few examples SIFT Based

➢ Random object disposal

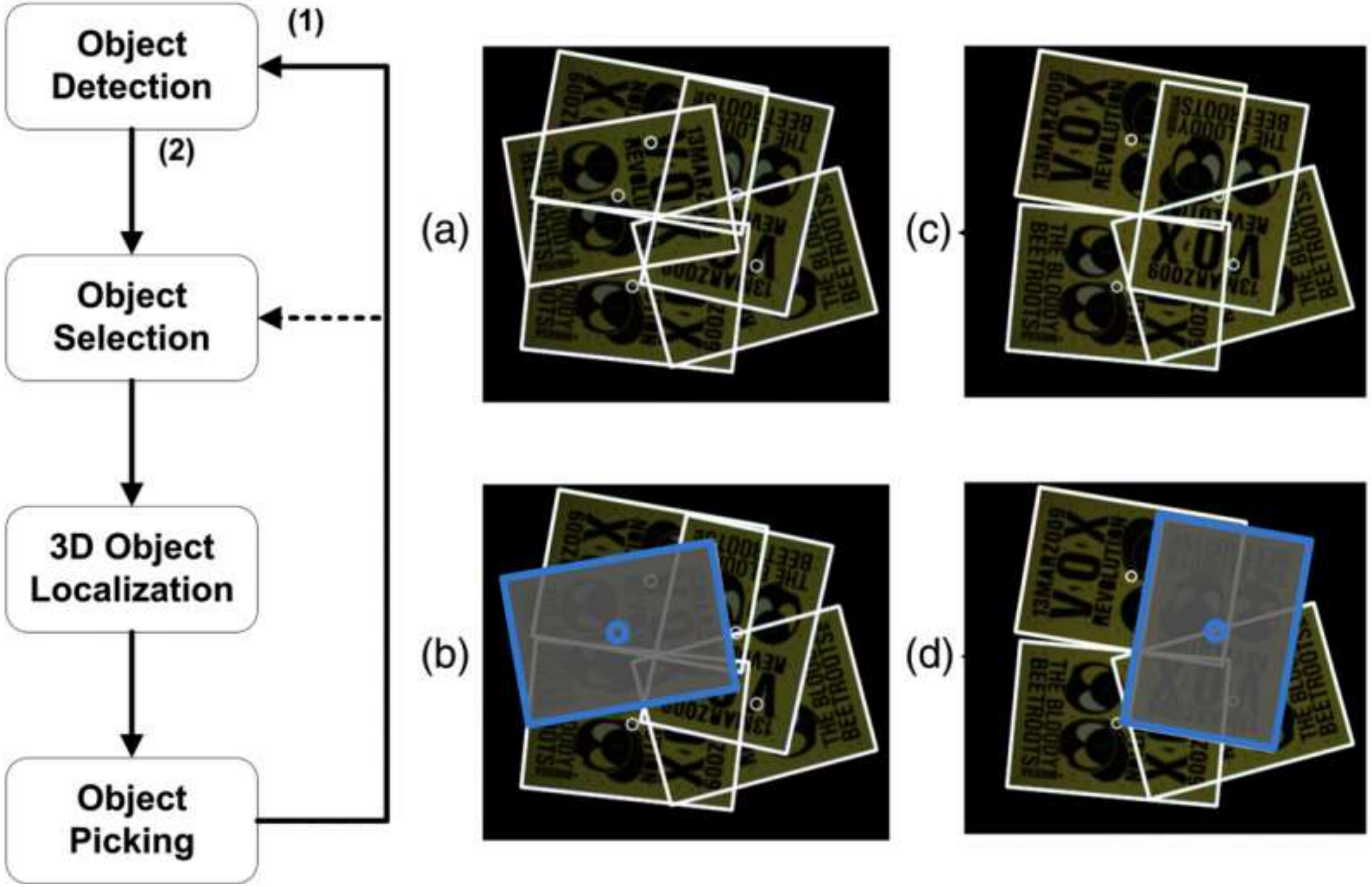
➢ Multiple instances and distractors

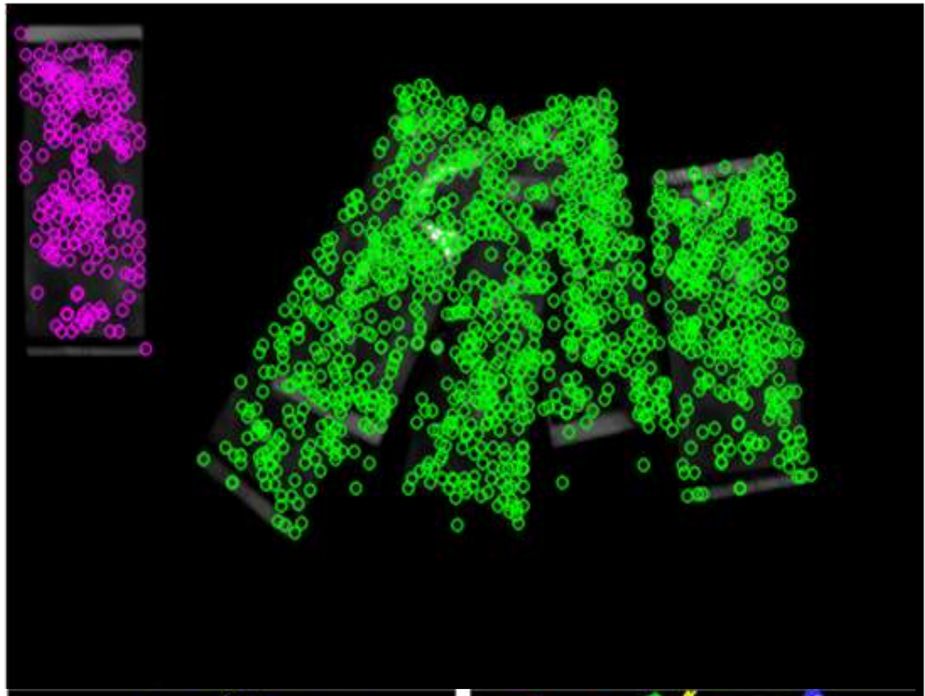
➢ Heavely occluded objects

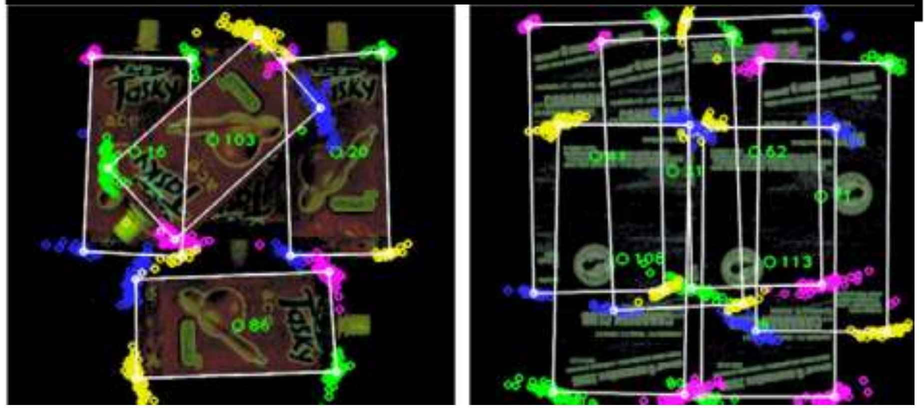
➢ High working speed (100obj/min)

P.Piccinini, A. Prati, R.Cucchiara Real-time object detection and localization with SIFT-based clustering
2012 Image and Video Computing 30 (2012)573-587

- Searching for the most visible (and easy to be picked) target
- SIFT Based (2004)

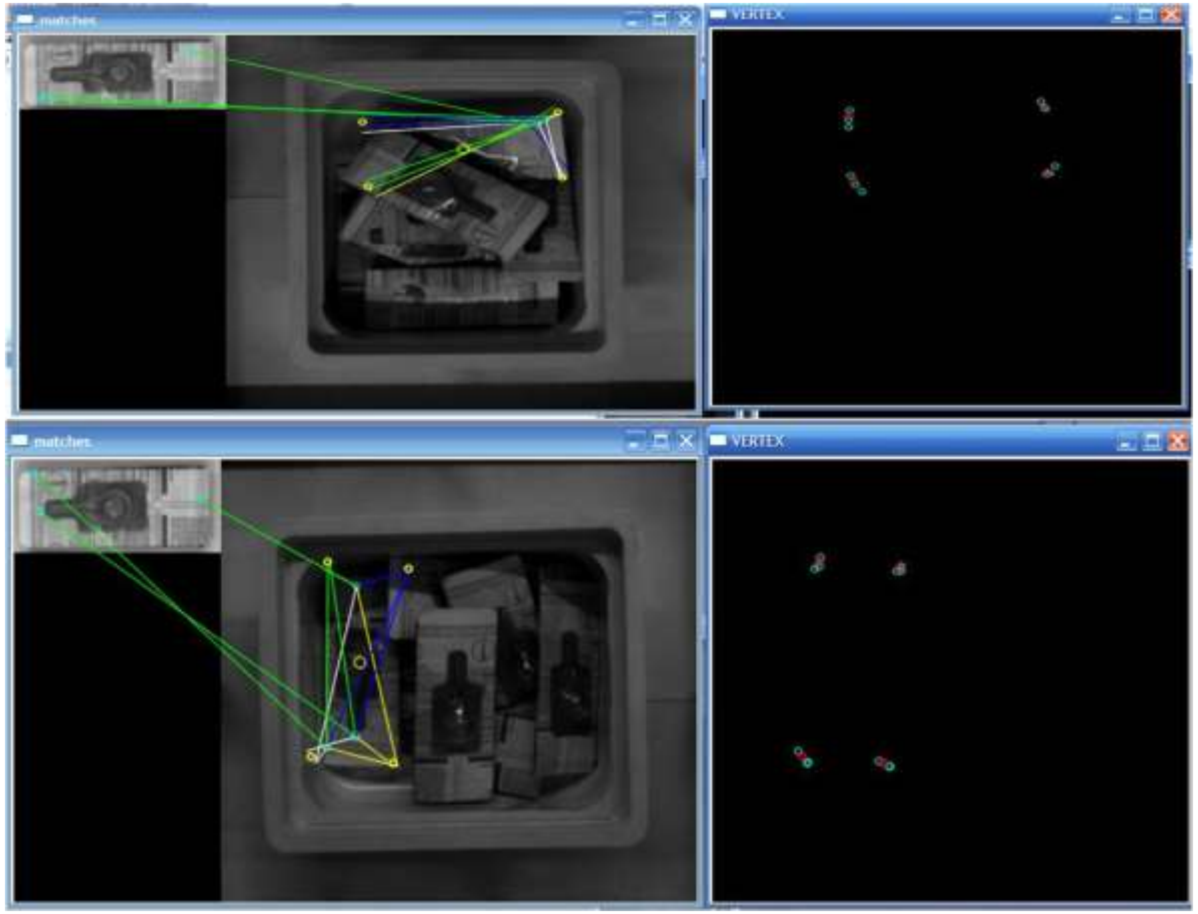


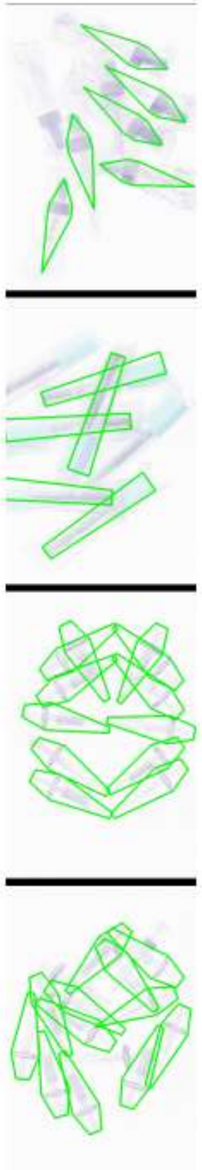


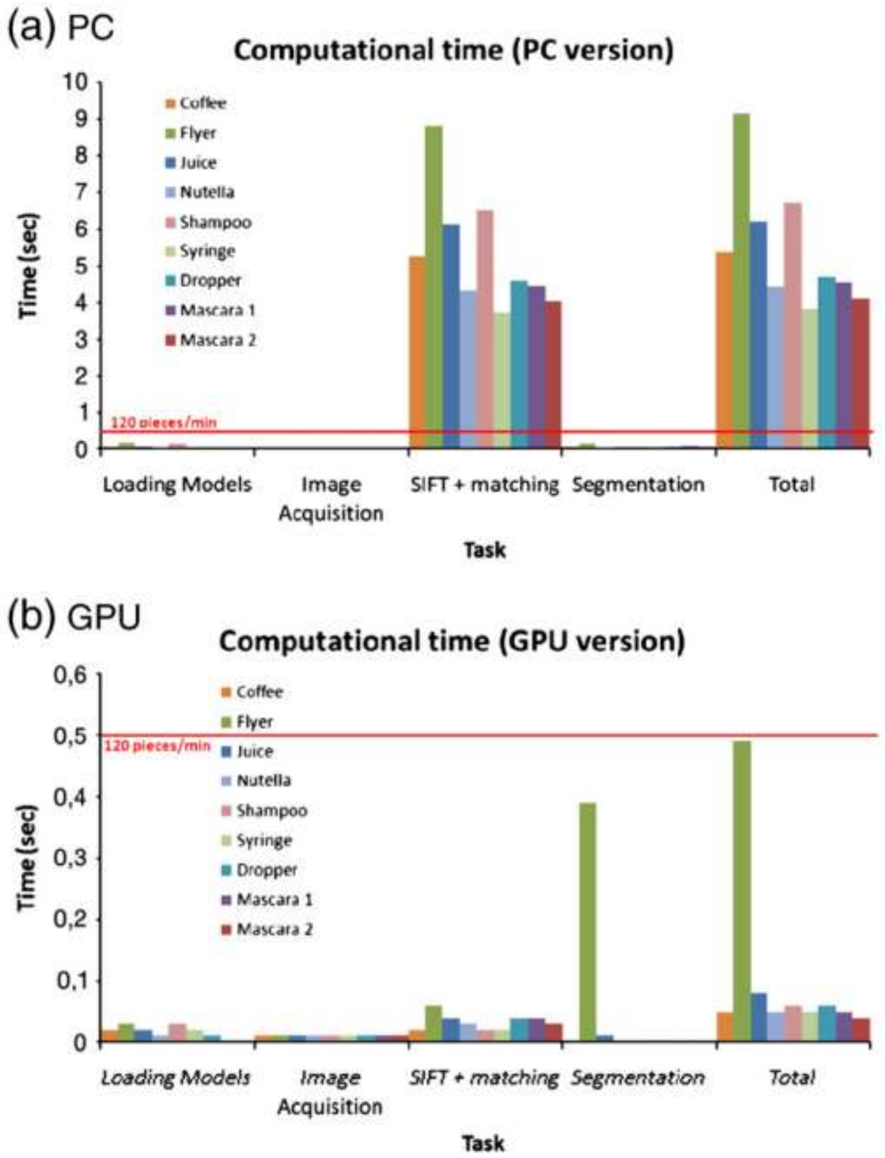


In 2012 real time with GPU only

- Real time processing
- only with GPU





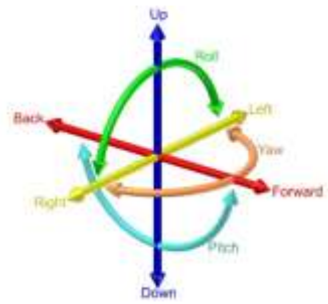


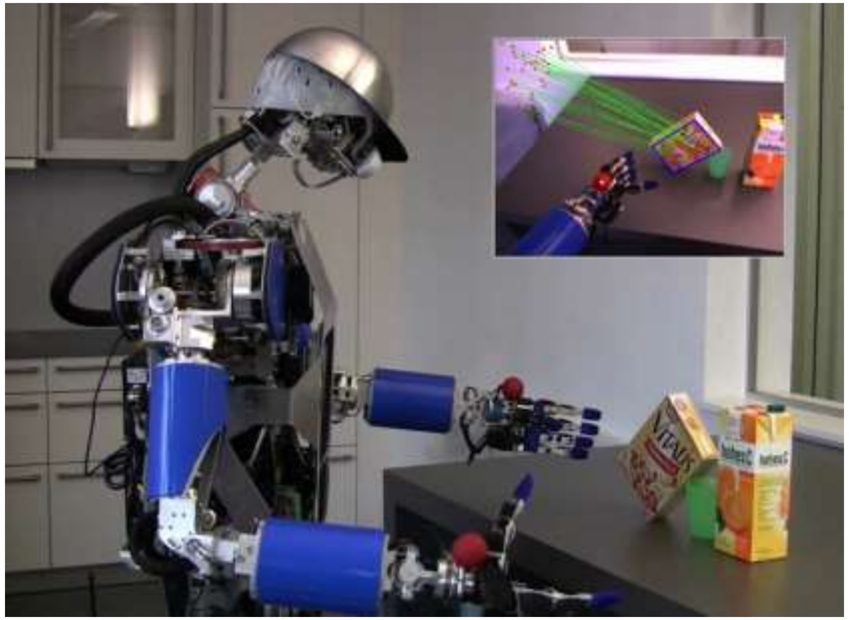
Now with DL

Automatic Bin Picking

Unconstrained picking

- Amazon Picking challenge 2015 , 2016
- two tasks:
- (1) picking: given an product ID, pick one instance out of a populated shelf and place it into a tote;
- (2) stowing: given a tote of products, pick and stow all of them into a populated shelf
- 6DoF

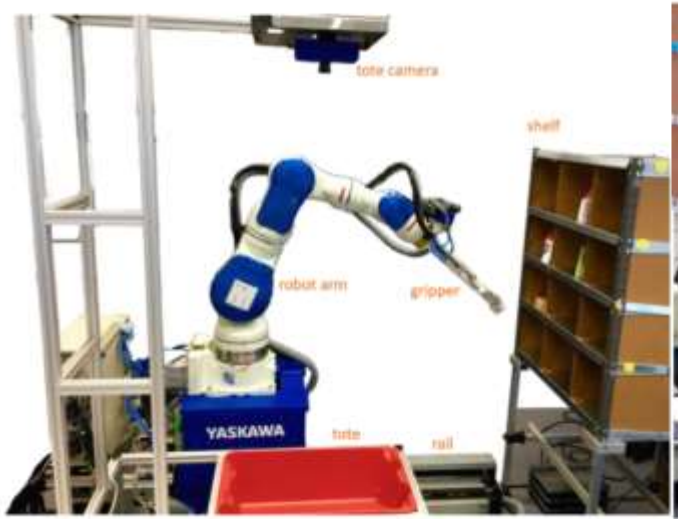




KIT Robot

Amazon Picking challenge 2016

- Winner: Delft Univ, NL29 target objects



Delft Robot

MIT Robot

a deep neural network based on Faster R-CNN classifies the objects in the RGB image and extracts their bounding boxes (trained with 500 labelled images detected in 150 ms)
Pose estimation in the 3D point cloud

- No CAD models
- No luminance setting
- No position constraints
- Real time

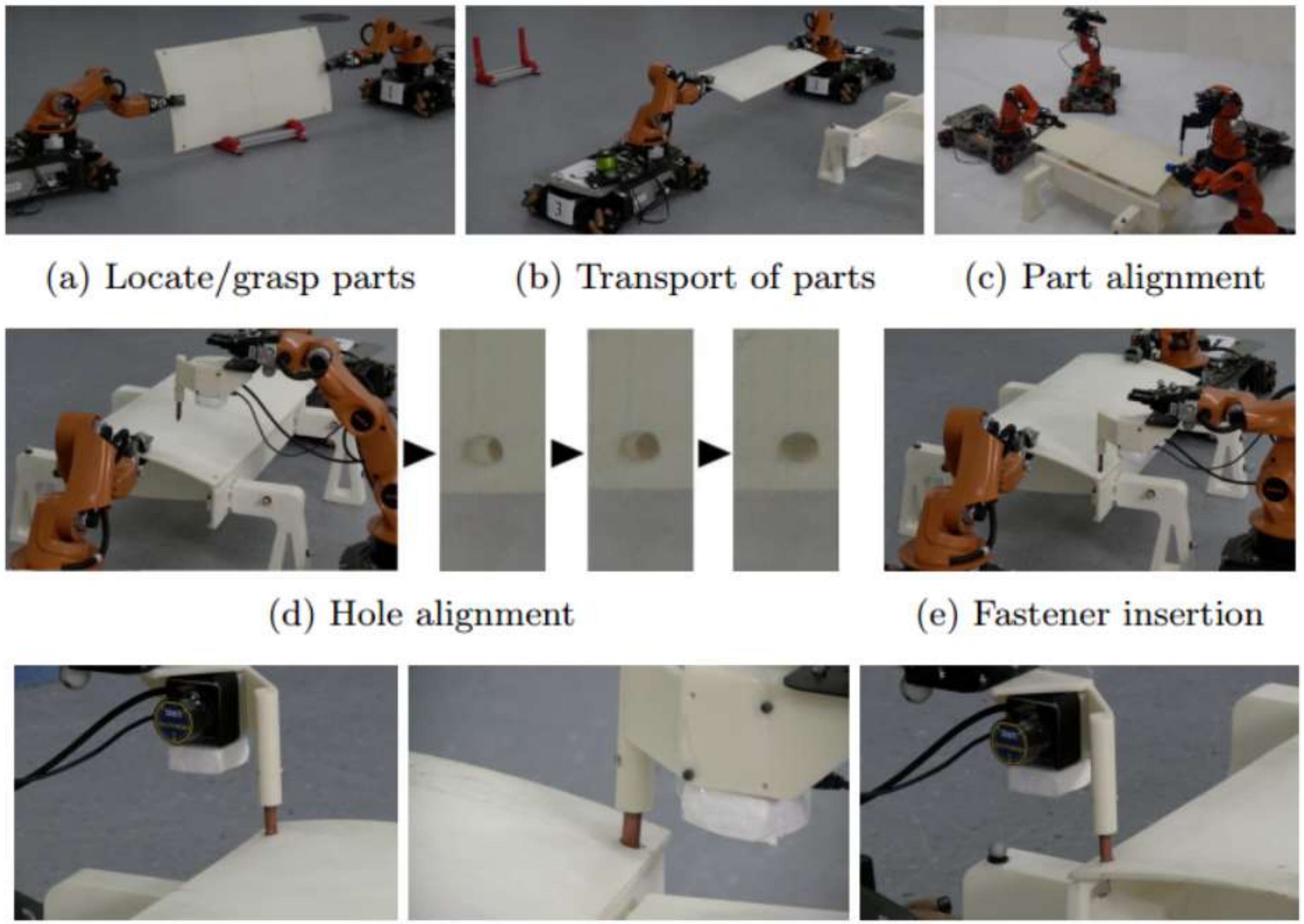
Robot coordination- collaborative working

- (MIT and Georgia Tech 2014)

Robot			
R1	R2	R3	R4
			Move to hole 1 neighborhood
Navigate to and move gripper to panel		Localize box	Find hole 1 in box
Close grippers and form fleet			Find hole 1 in box
Pick up panel			
Orient panel to horizontal			
Transport panel into neighborhood of box			
Servo panel into alignment with ladder		Localize panel	
Servo panel hole 1 into alignment with ladder hole 1			Localize panel hole 1
End fleet formation and open grippers			Insert fastener 1
Move out of the way	Align panel hole 2 to box hole 2	Move out of the way	Navigate to panel hole 2
	Move out of the way		Localize hole 2
			Insert fastener 2
			Navigate to hole 3
			Localize hole 3
			Insert fastener 3
			Navigate to hole 4
			Localize hole 4
			Insert fastener 4

Table 1: Flow of actions among four robots during attachment of a panel to a box. Time flows from top to bottom. Box colors indicate the type of localization used in each action. Blue boxes indicate fiducial based localization. Green boxes denote object-shape based tracking. Pink boxes indicate functional-feature level localization. White boxes indicate sensorless operations.

Towards Coordinated Precision Assembly with Robot Teams



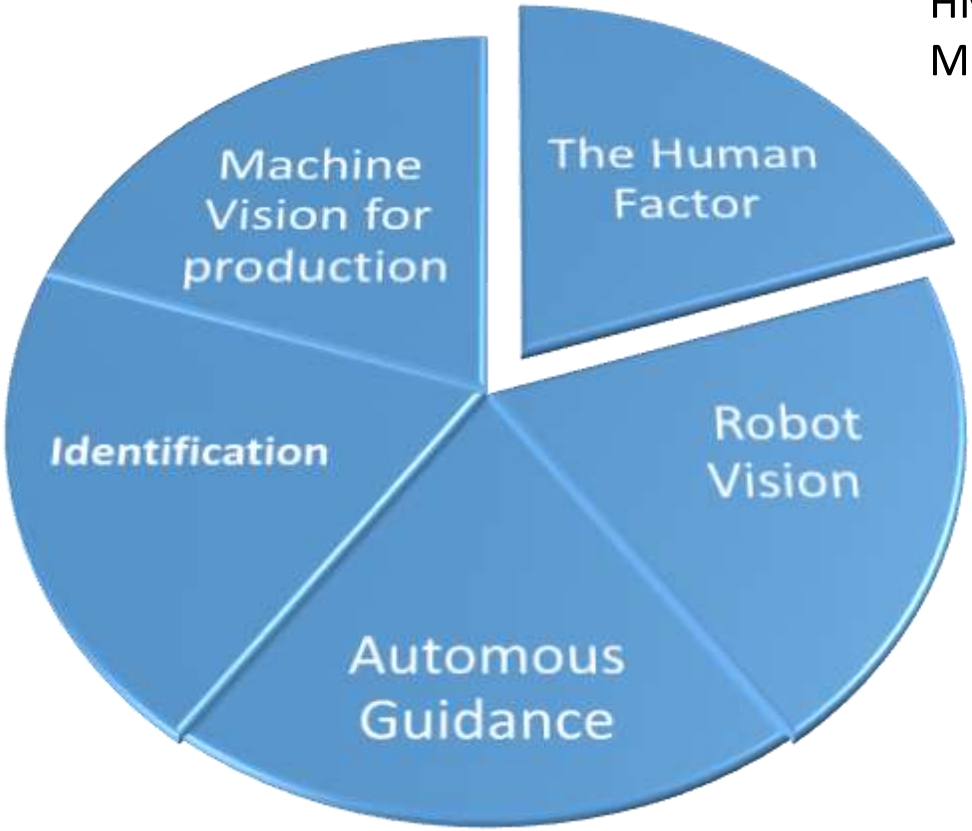
(a) Locate/grasp parts (b) Transport of parts (c) Part alignment

(d) Hole alignment (e) Fastener insertion

(f) Fastener 2 (g) Fastener 3 (h) Fastener 4

Applications

HMI Human
Machine Interacion



The Human factor

- Computer Vision useful for recognizing the Human Factor

- Human Machine Interaction (**HMI in collaborative robots**)

- Human Safety in industry (**detecting persons and autonomous machines**)

- Learning by humans
  - In automotive for autonomous driving
  - In robotics for natural grasping…

Real time processing of:

- People detection
- Semantic segmentation
- Pose estimation
- Facial gaze espression
- gesture analysis

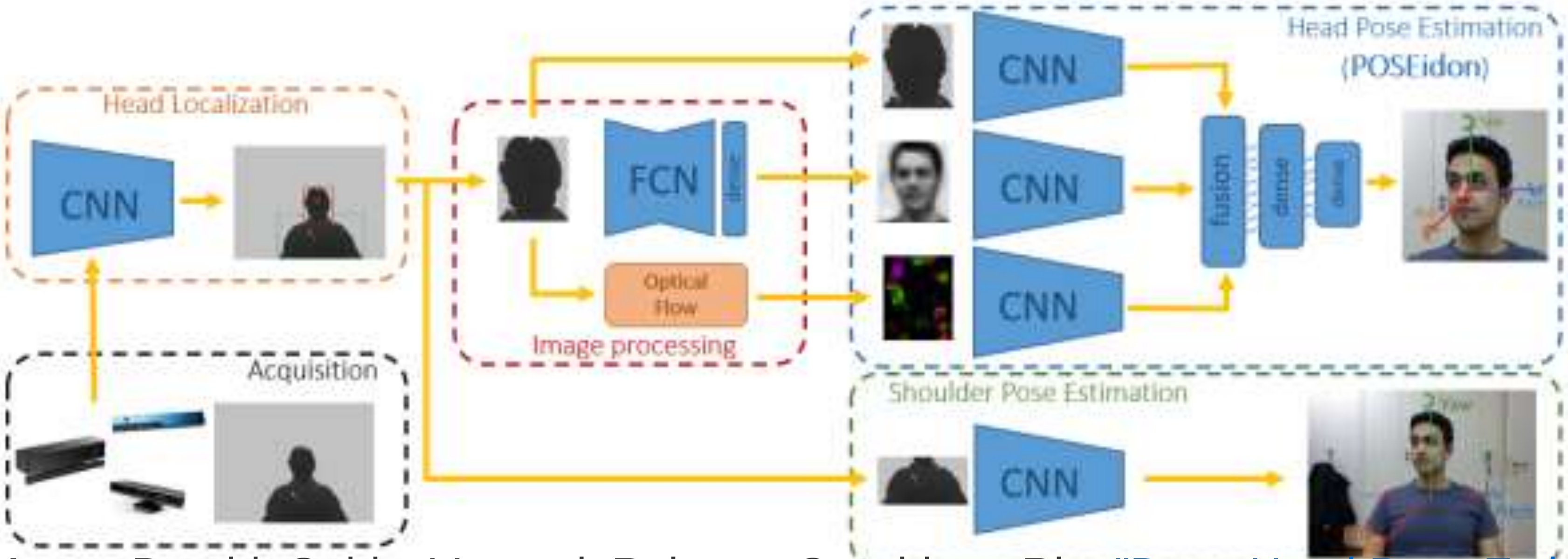
Human machine interaction with egocentric vision

- Real time object recognition for augmenting vision

- EGOCENTRIC VISION

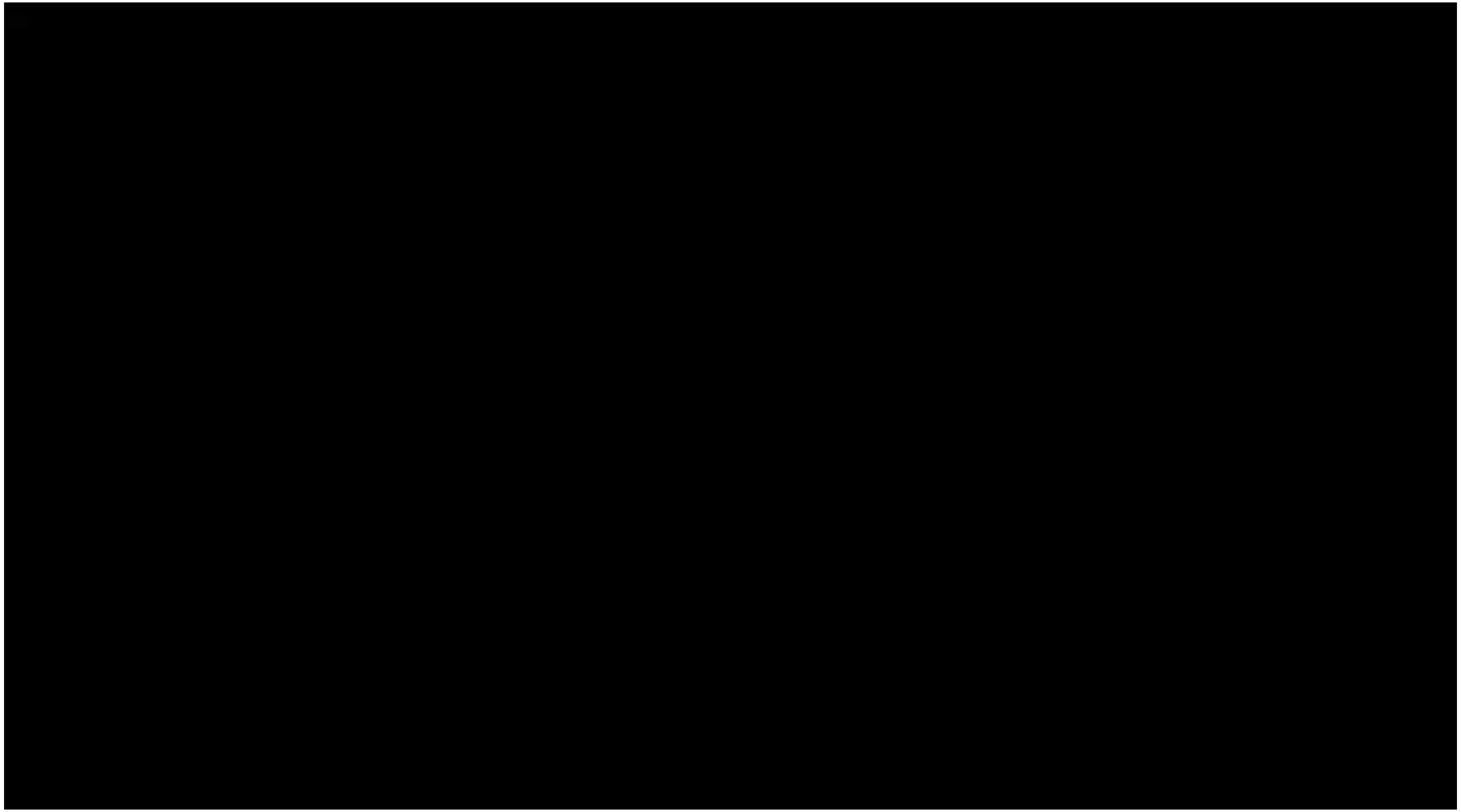
HIxIA project (Human Augmentation for Industrial Assistance) SPINAutomazione & UNIMORE

HBU IN THE CAR by depth images

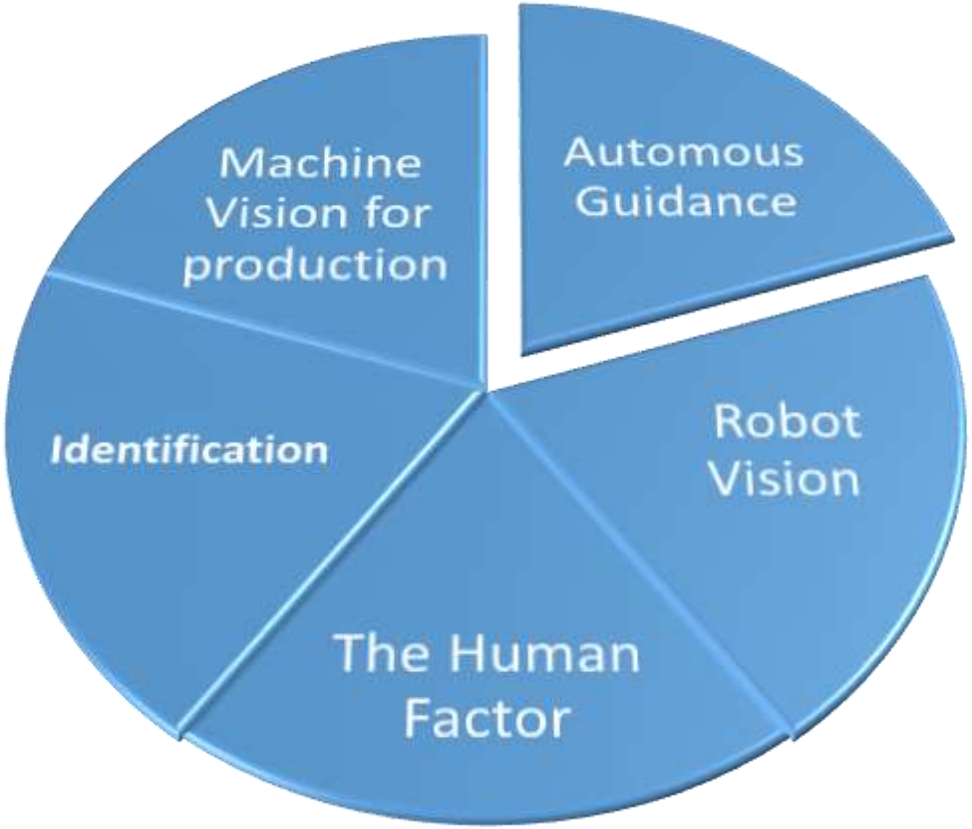


Venturelli, Marco; Borghi, Guido; Vezzani, Roberto; Cucchiara, Rita ["Deep Head Pose Estimation from Depth Data for In-car Automotive Applications"] *ProceedingsICPR 2016*

- Experiments on depth-only based pose detection



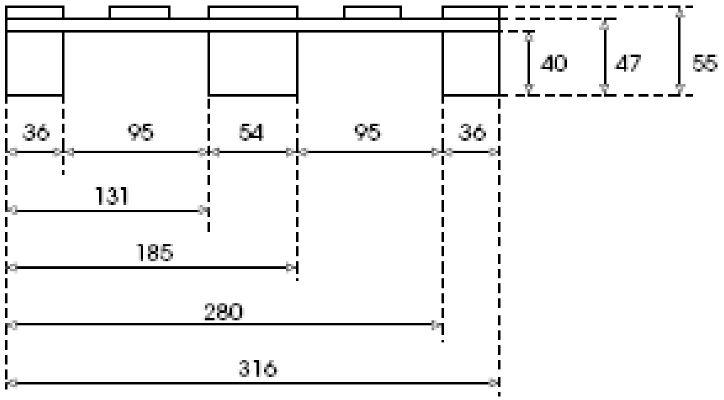
Applications

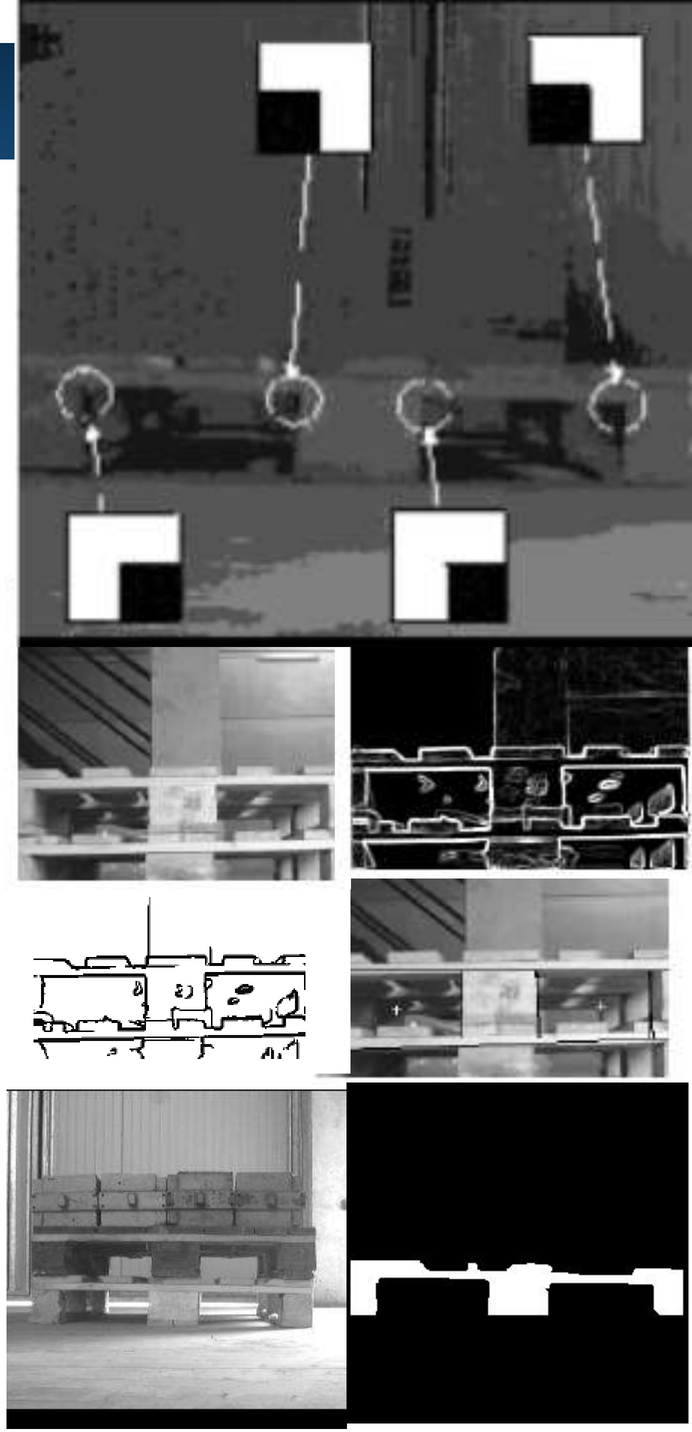


Localization in 2D images- classic methods

- Image segmentation
- Object localization in 2D images

  Traditionally
  - Template matching
  - Edge based identification
  - Model –based localization



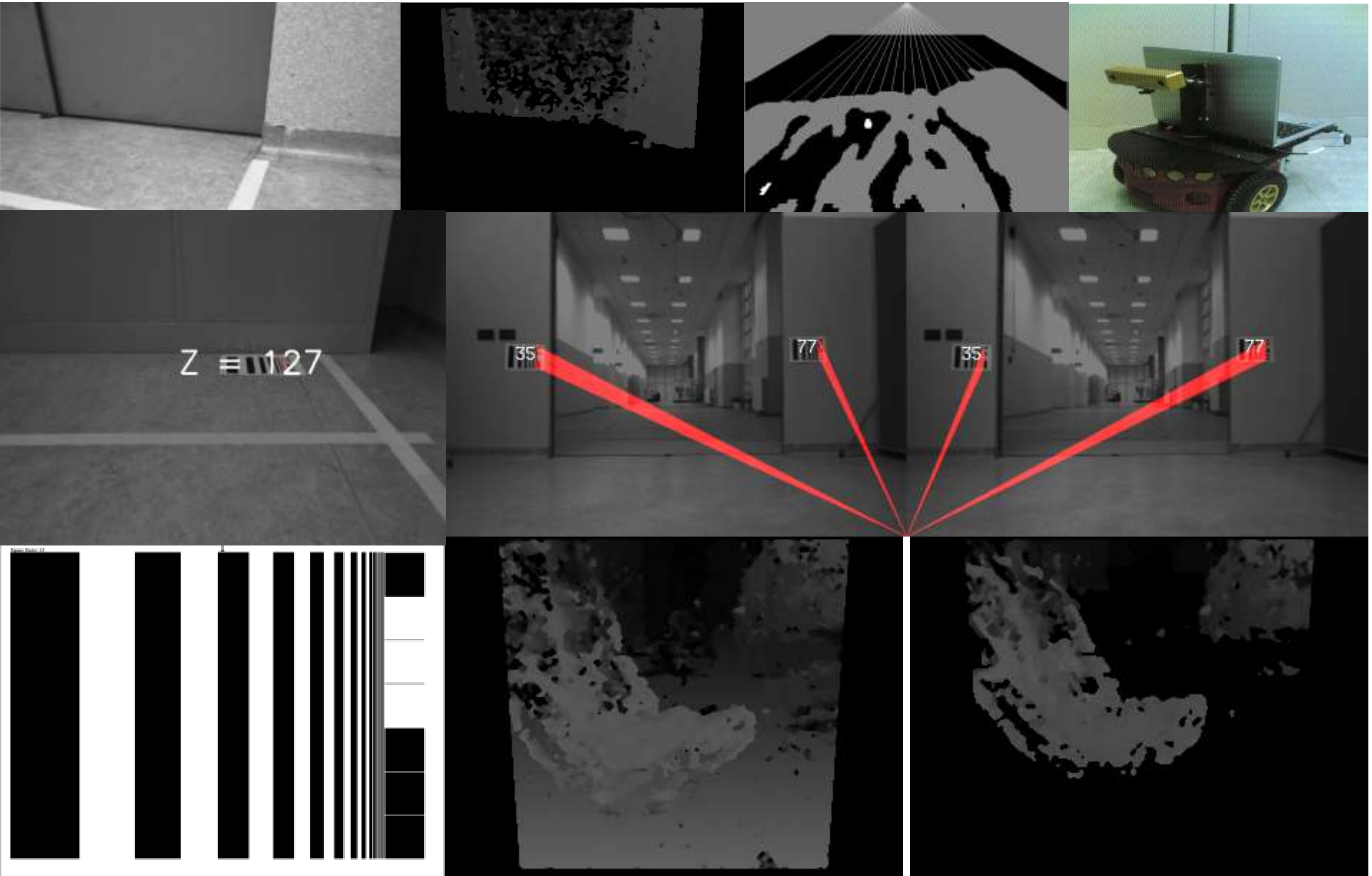


EG. Pallet recognition in un-constrained environment

- ✓ Image processing
- ✓ Hough Transform
- ✓ Harris Corner Detection
- ✓ Constraint grah analysis
- ✓ Decision Trees

R. Cucchiara, M. Piccardi, A. Prati,["Focus based feature extraction for pallet recognition"]()
in *Proceedings of the 11th British Machine Vision Conference (BMVC 2000)*, Bristol, UK, pp. 695-704, 2000

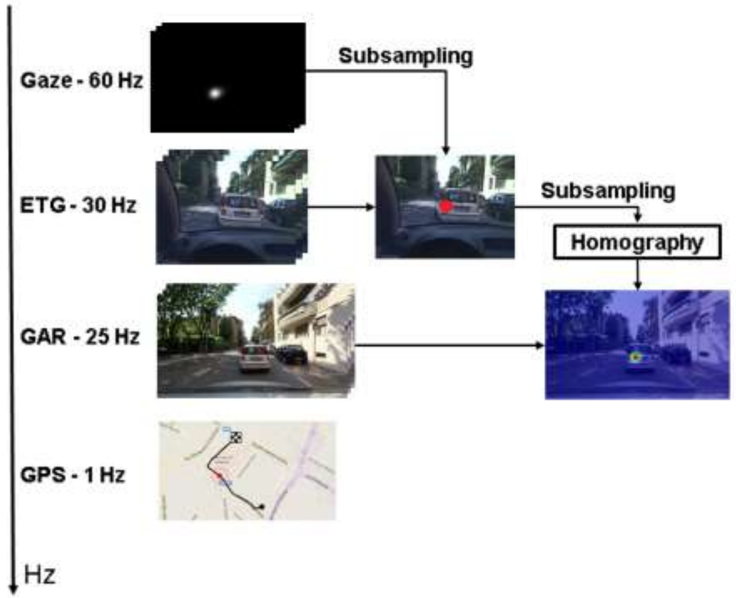
Stereo Vision and many sensors



2008 Imagelab
Italvision & UNIMORE
(PRRIITT)

Learning from a human driver

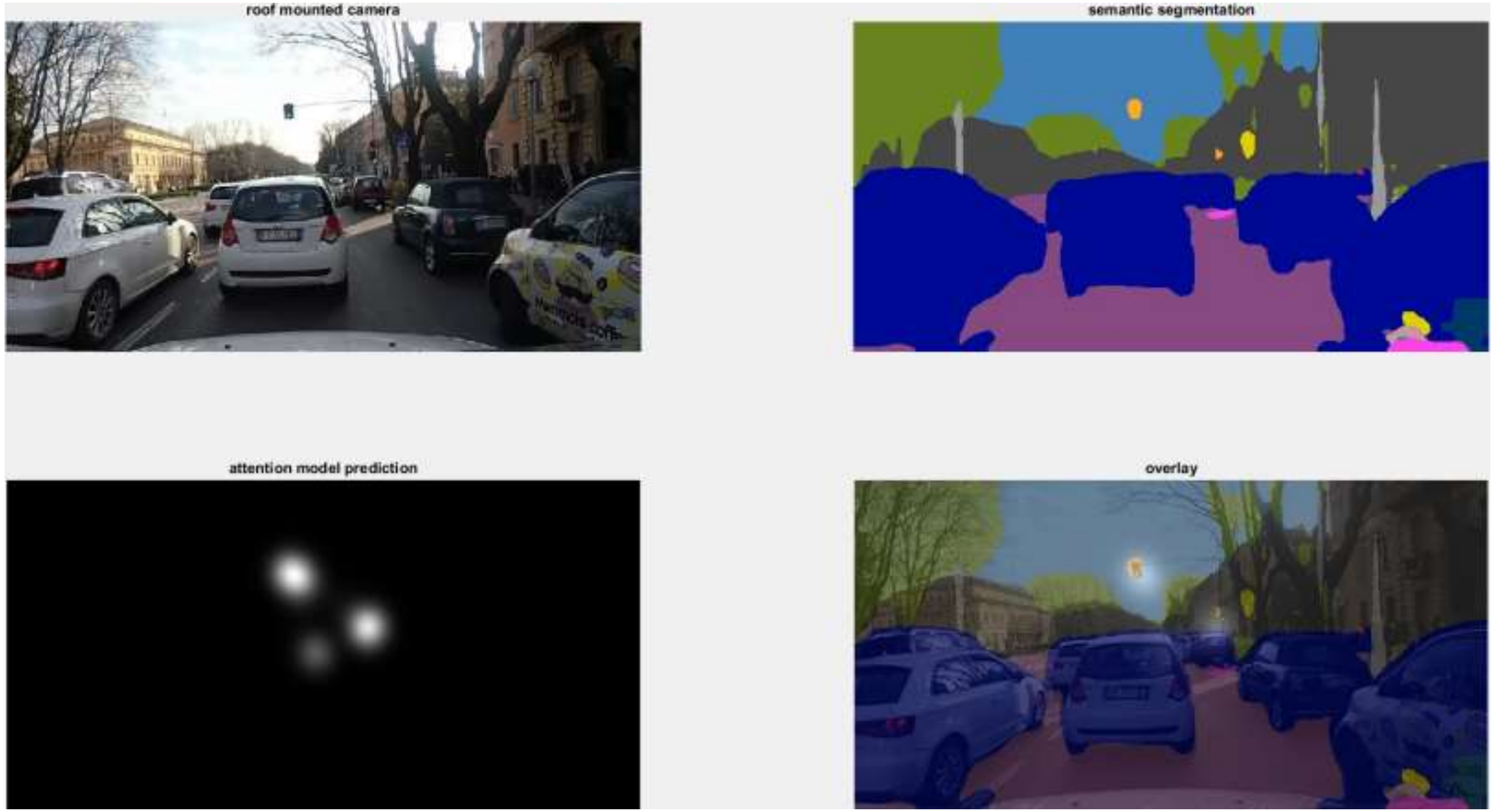
- Image registration and synchronization

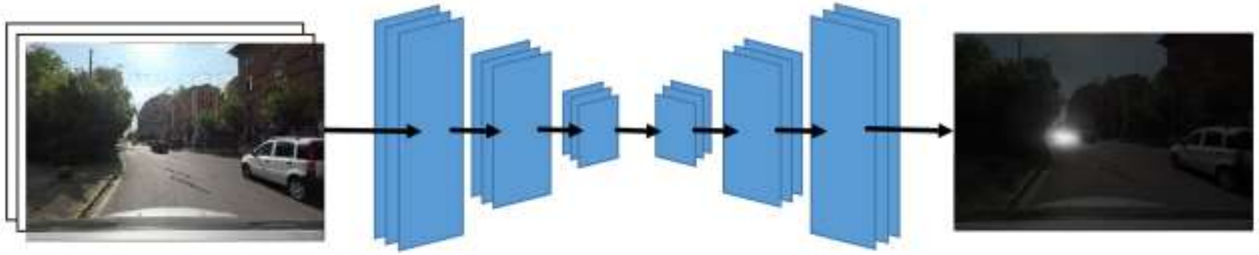


HBU –DR(eye)ve human behavior driving

Stefano Alletto∗ , Andrea Palazzi∗ , Francesco Solera∗ , Simone Calderara and Rita Cucchiara **DR(eye)VE** a Dataset for Attention-Based Tasks with Applications to Autonomous and Assisted Driving CVPRW2016

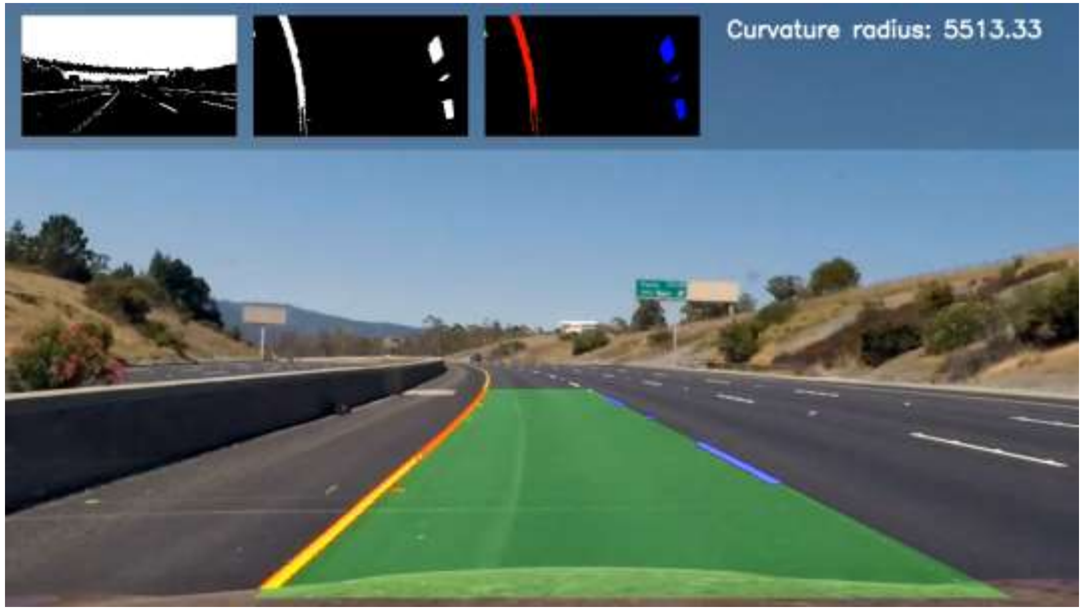
Dr(eye)ve learned where the drivers see, and what the drivers pay attention on…





Atonomous driving in street .. is «easy»

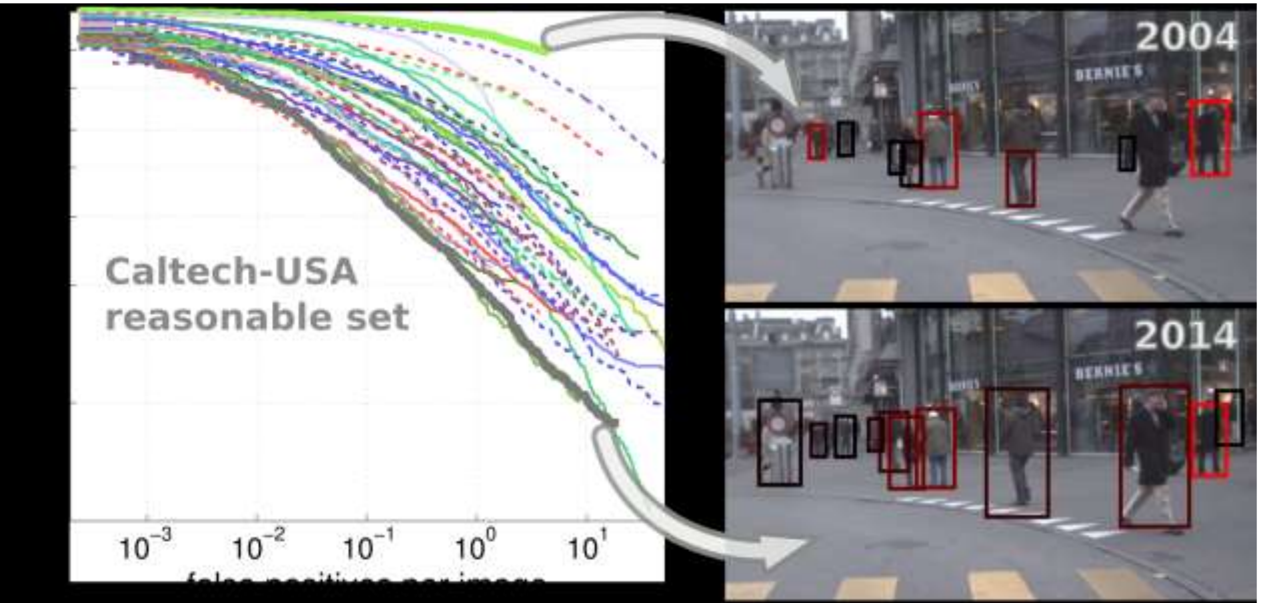
- Experiments at Modena Imagelab

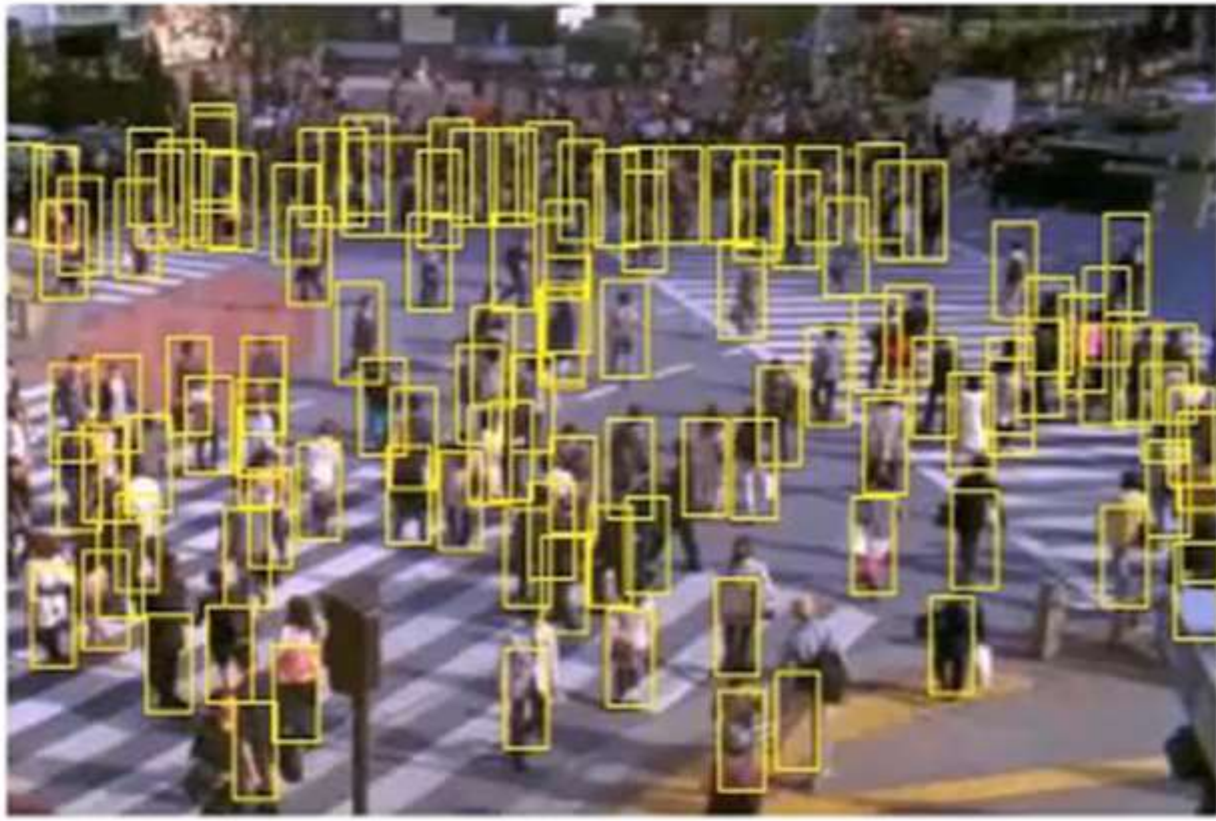


- **People detection is a robust «solved» task**
- With standard classic pattern recognition approaches
- Using motion
- With Deep Learning
- Special algorithms In case of high recall

(e.g. security in working area)

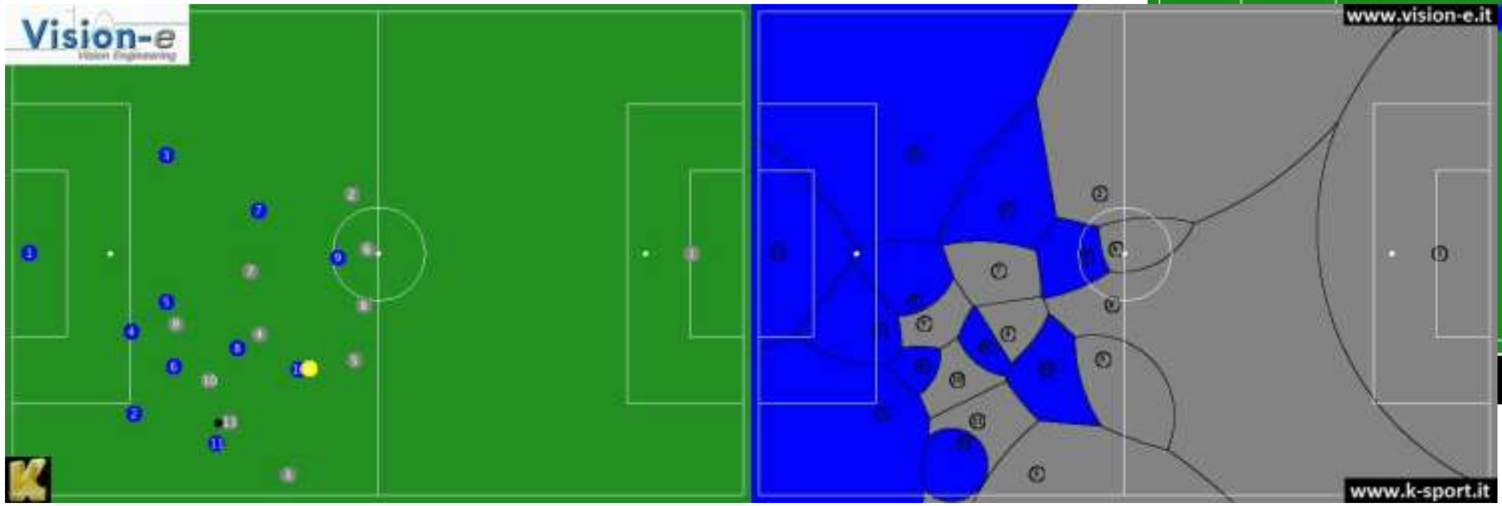
(Thanks to Shiele ECCV 2004)





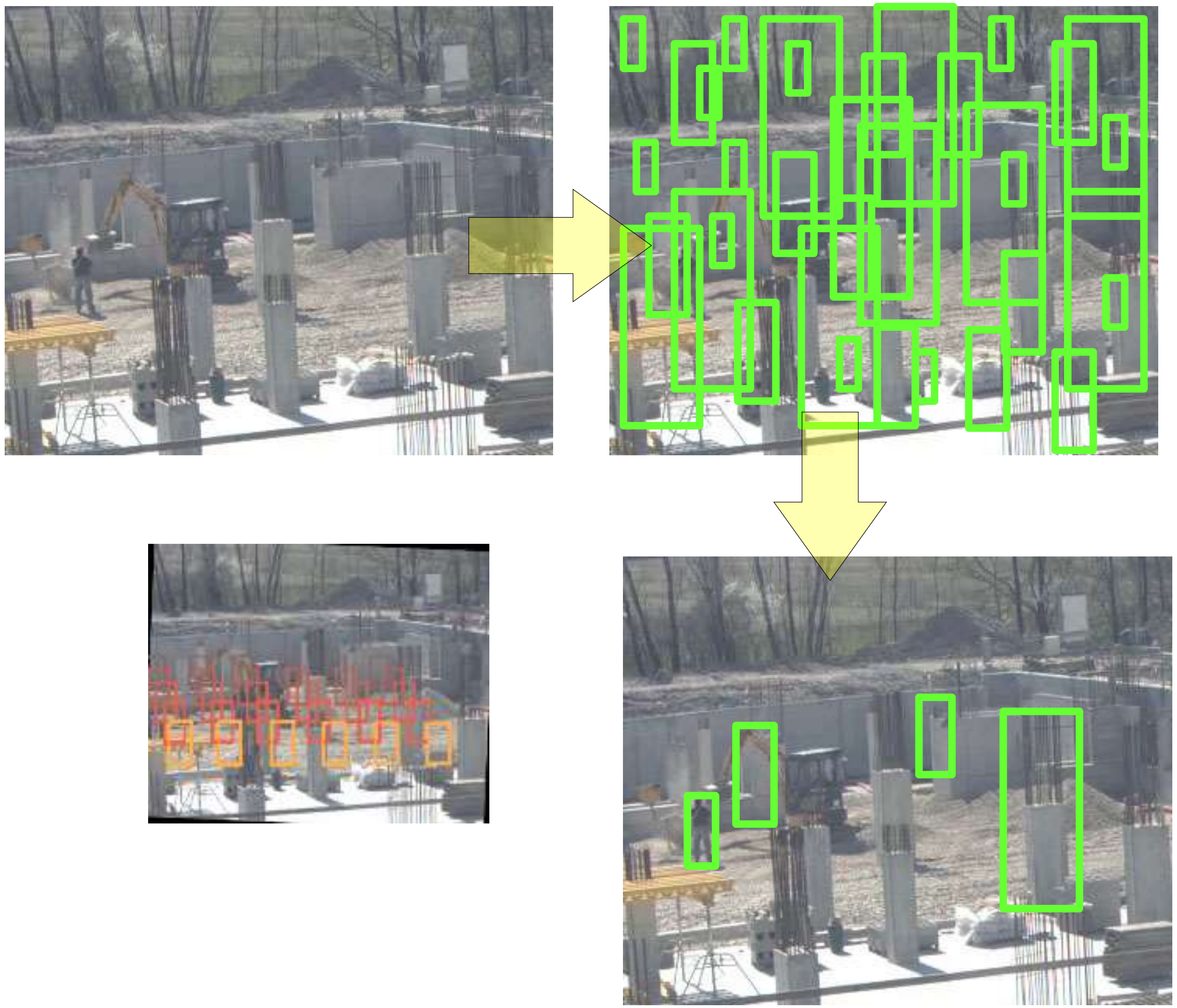
Tracking by detection: using people detection for initialize ROI-based tracking (eg particle filter)

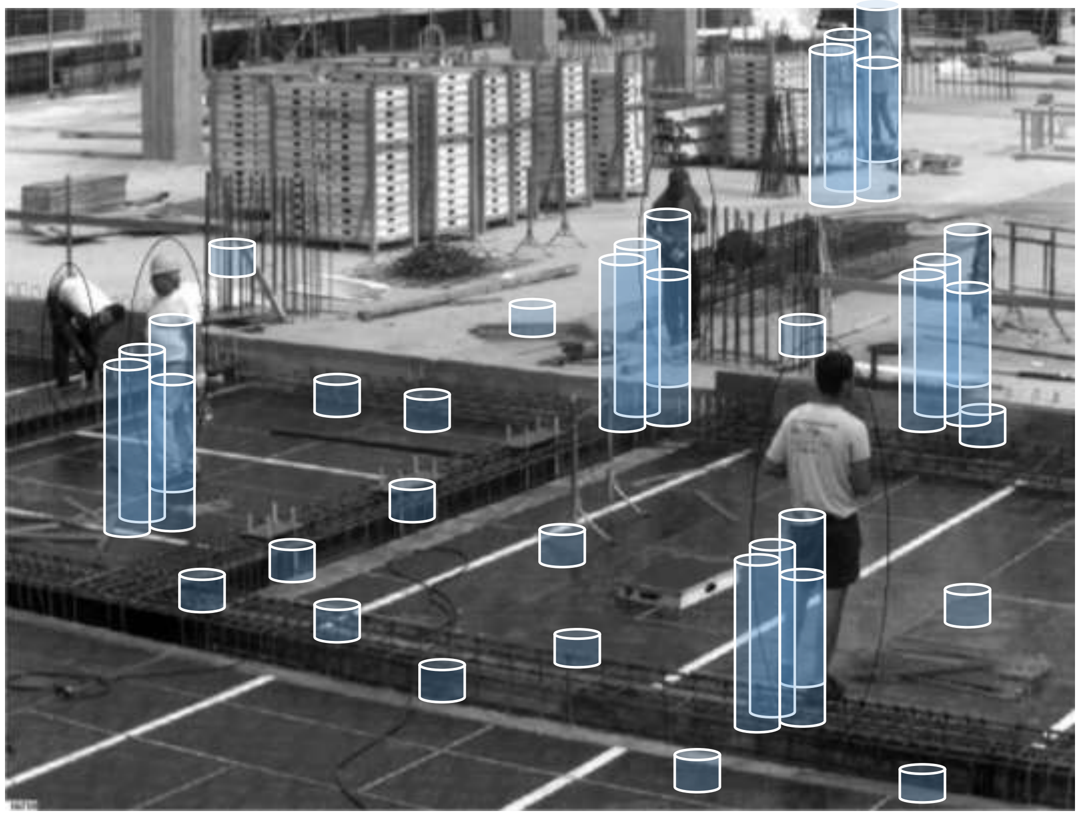
In semi-constrained world
Tracking is possible

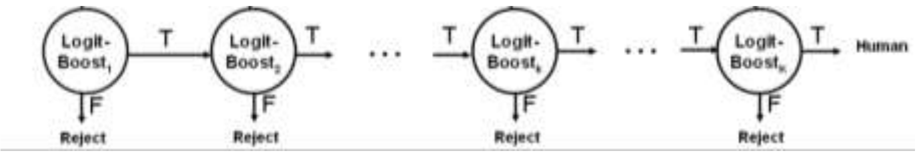


In special environments is more complex

- External construction sites



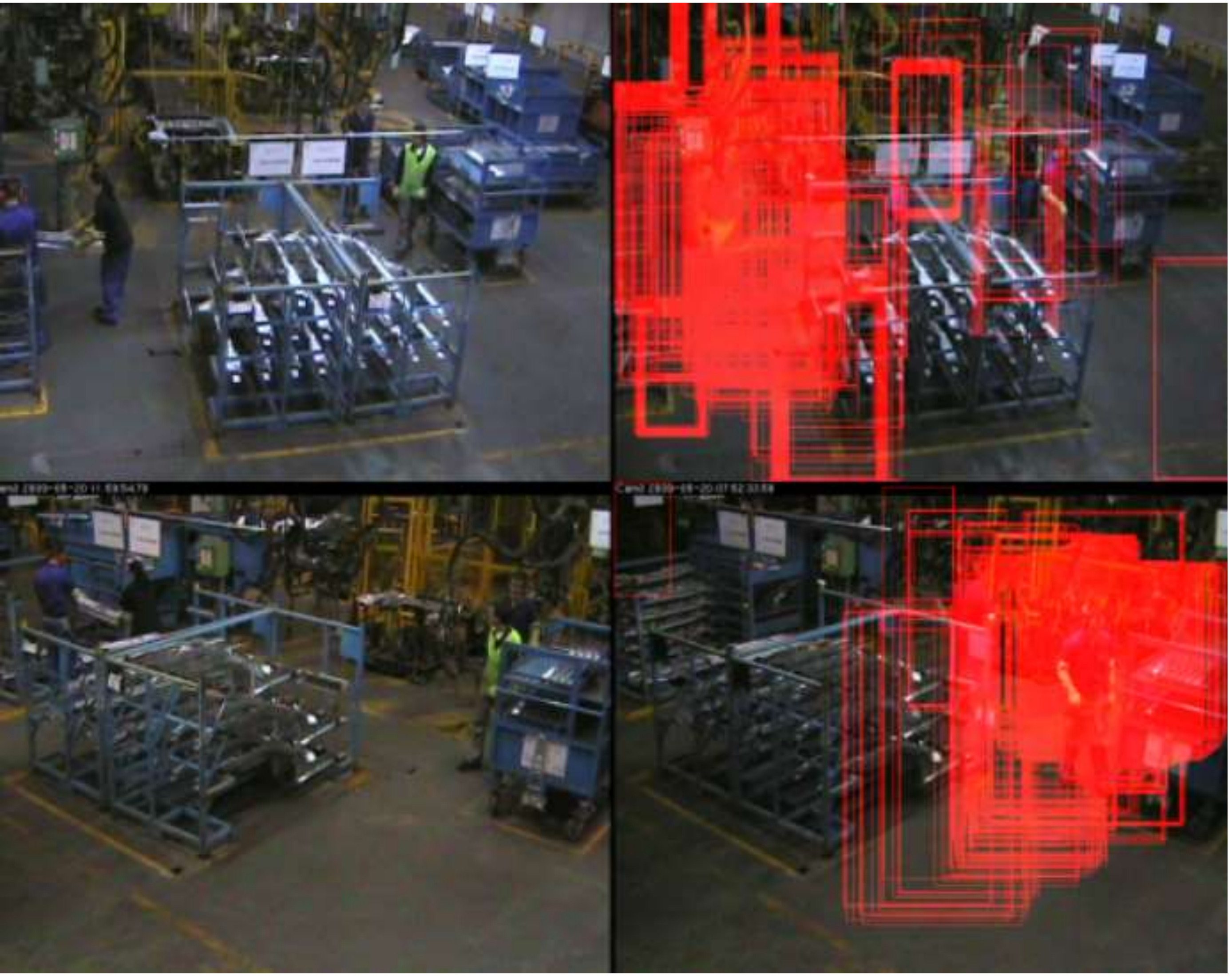




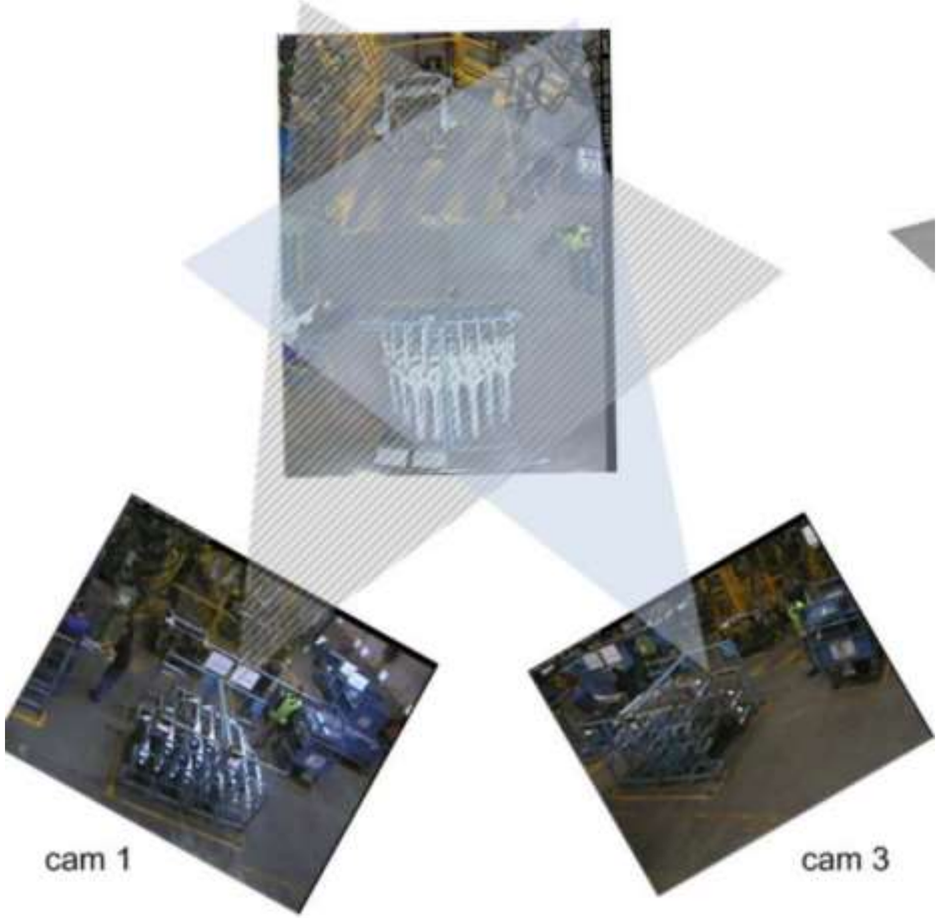
G.Gualdi, A.Prati, R.Cucchiara Multi-Stage Particle Windows for Fast and Accurate Object Detection
IEEE Transactions on PAMI Aug. 2012

In unconstrained working area is more complex!

- Multiple cameras and geometric constraints



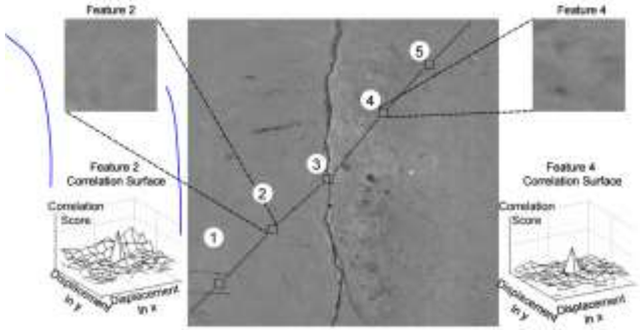
Thanks to EBVH)



AGV Automatic Guided Vehicles in Industries is the future!

Autonomous guidance is more complex:

- ✓ 3D world reconstruction
- ✓ Unconstrained scenario
- ✓ Human presence

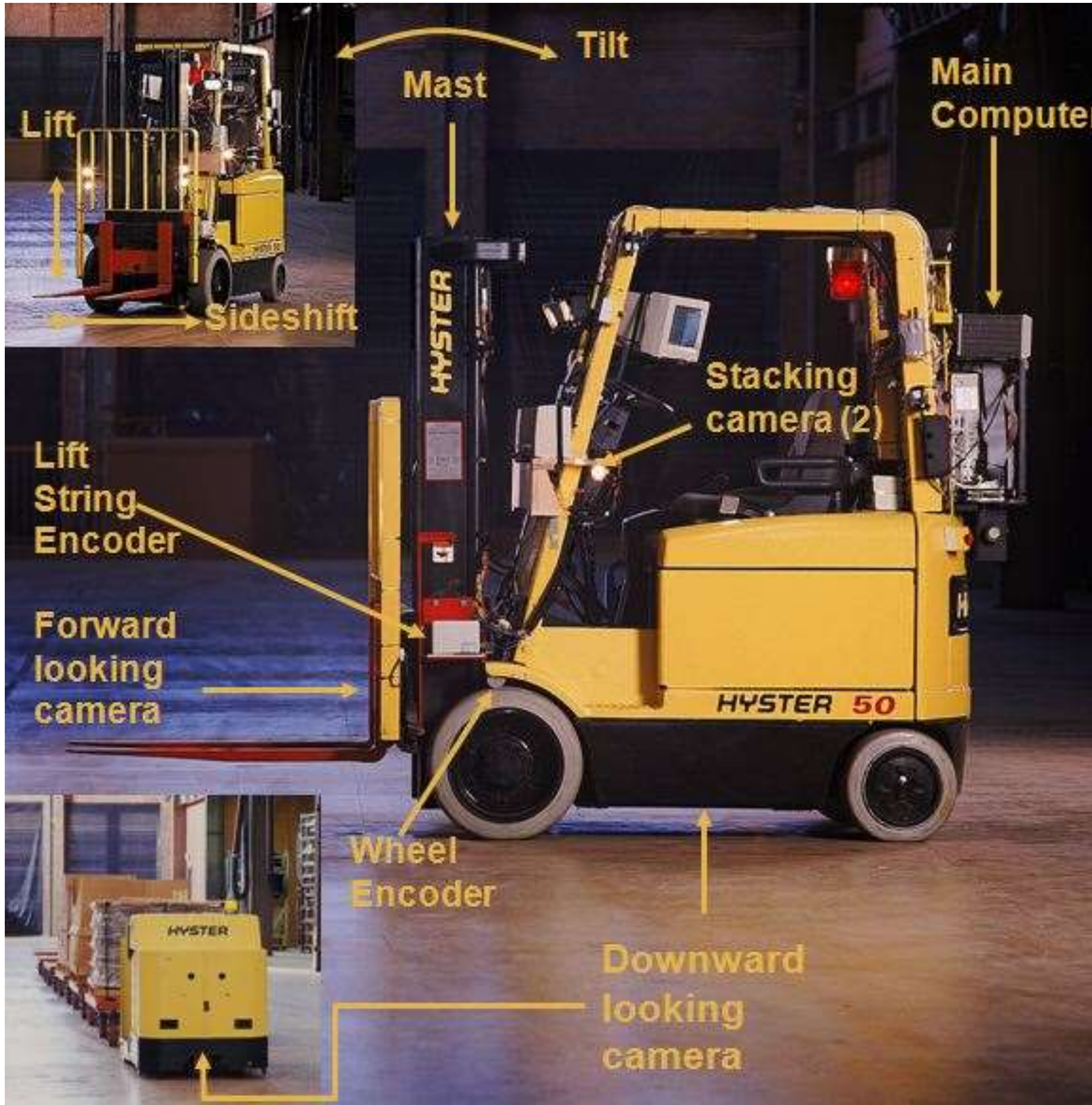


Now:

Vision and other sensors (GPS, Laser..)*

Markers in the environment

Big Data collection



A large impact of computer vision and machine learning

*Kelly, A.,Nagy, B. ,Stager, D., Unnikrishnan, R., “An Infrastructure-Free Automated Guided Vehicle Based on Computer Vision”, IEEE Robotics and Automation Magazine. 2007.

- Reading Markers
- Cognex DataMan 500 for logistics(1000 fps, up to 270 reading codes per seconds
- In Amazon Kiva robots read markers; communications for collision avoidance
- Next logistic systems, Drones

Fig. 6 Delivery of packages via drones [6]

Fig. 5 Kiva robots and their orientation in the warehouse using QR codes [5]

AGVS spreding

Positioning guided vehicles

Next future

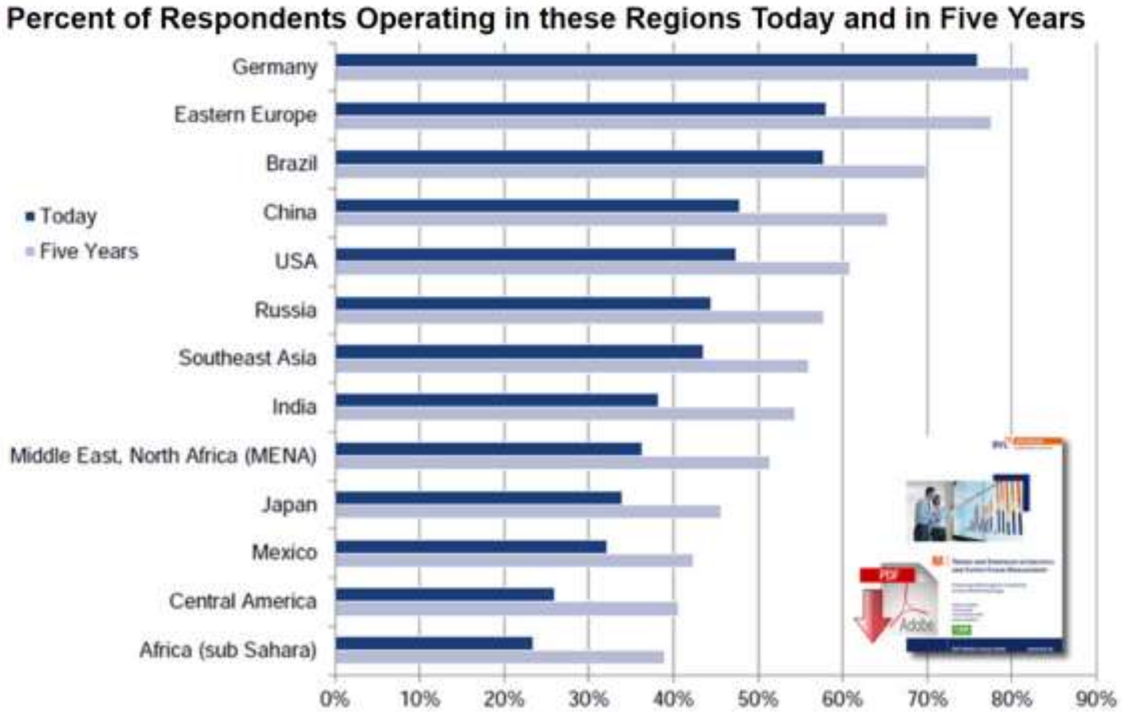
- Autonomous driving
- Safety for humans
- Autonomous decision
- HM interaction
- M-M collaboration

The TUG is an "autonomous mobile robot" (AMR). As an autonomous robot, it has considerable advantages in flexibility, compatibility and control compared to "automated guided vehicles" (AGV).

AMR	vs.	AGV
Trackless navigation		Requires "tracks"
Can go around obstacles		Obstacles stop it
Can be easily re-mapped		Difficult to re-map
No depots needed		Needs "depots"
Delivers to user location		Does not deliver to user
Travels around people		Travels in dedicated areas
Easy to expand & change		Difficult to expand

It is not only Vision (and other sensors) ---- to cope with many other issues

- Privacy low (In Italy, Europe, USA)
- Ethical issues
- Patents and Licences
- ..
- Perfomance analysis (not always predictable)
- ……..



- but is a good idea!
- (data from 2014-2019)

Conclusions.

- Computer Vision and Machine Vision are now the same discipline
- **Deep Learning approaches are fully integrated**
- Real time processing can be reached with embedded platforms
- Towards to more general-purpose approaches

- **Ready for collaborations**
- Stages
- Joined/ Funded Research projects
- Industrial Phd Programs
- Master

softech-ict

New Master   UNIMORE 2017
Opening Soon!

MuMeT 2017
visual computing and multimedia technologies

Thanks to:
Rita Cucchiara, Costantino Grana, Roberto Vezzani, Simone Calderara,[Giuseppe Serra],
Stefano Aletto, Fabrizio Balducci, Guido Borghi, Andrea Palazzi, Federico Bolelli,
[Marco Manfredi], Francesco Paci, Francesco Solera, Patrizia Varini, Lorenzo Baraldi,
Andrea Corbelli, Marcella Cornia, Augusto Pieracci, Paolo Santinelli,
Silvia Calio and Marco Venturelli