Machine Vision
In the Deep Learning Era

CRIT, 1/02/201

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»

From AI to Computer Vision

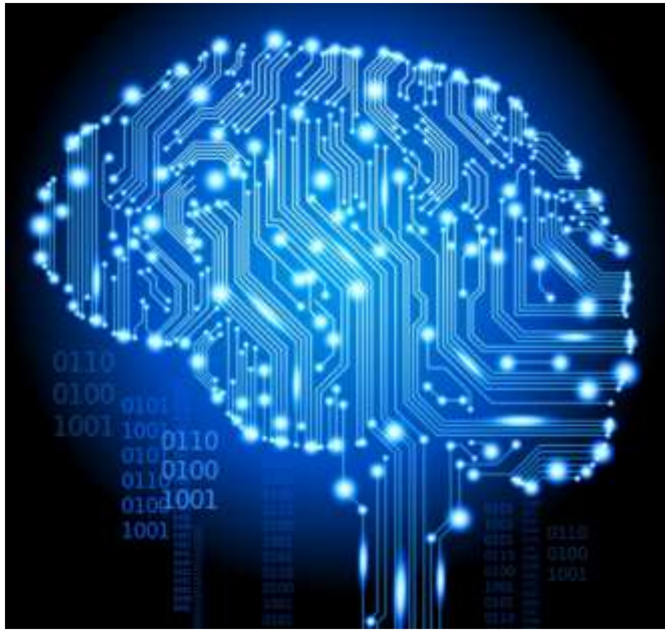
Artificial intelligence: The scientific field which studies how to create computers and computer software that are **capable** of **intelligent** behavior, using Sensing, Perception, Knowledge, Reasoning and Learning.

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.**



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…

From Computer Vision to Machine Vision

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

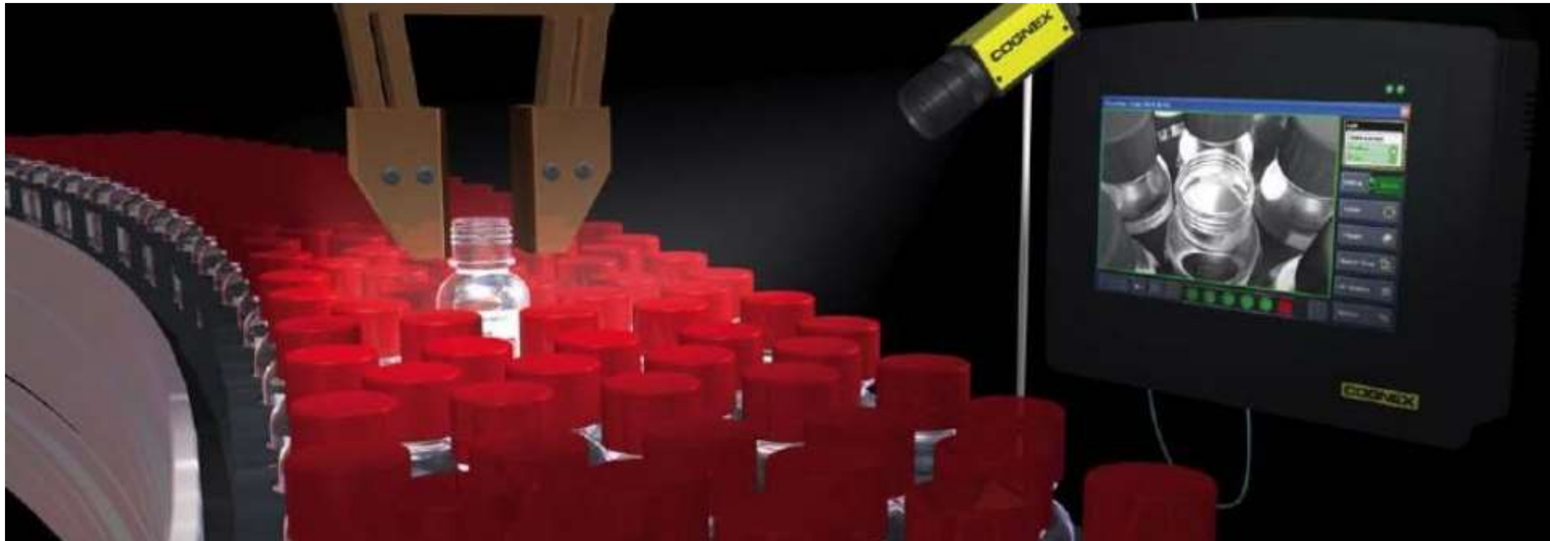
Machine Vision: the engineering field studying how to build **computer vision-based** systems, services and solutions, typically for industrial environment.

A (theoretical) Computer Science and Engineering discipline

Machine vision fundamentals:

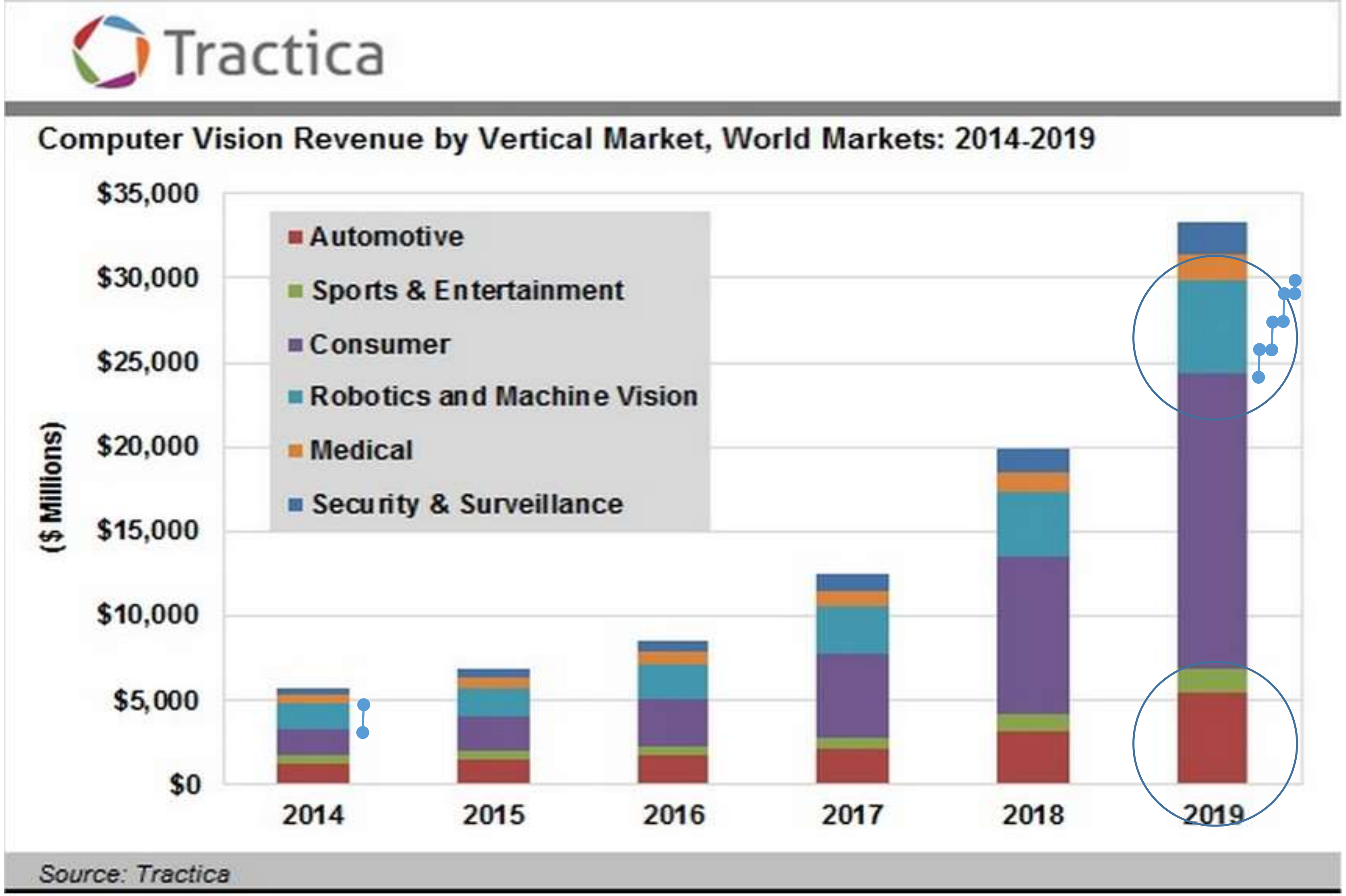
- Constrained environment
- Speed-based and Real-time solutions (w.r.t. actions to do)
- Defined precision/recall and performance
- Low Cost design and production
- Standardization

Courtesy of ABCON UK



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%



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.*

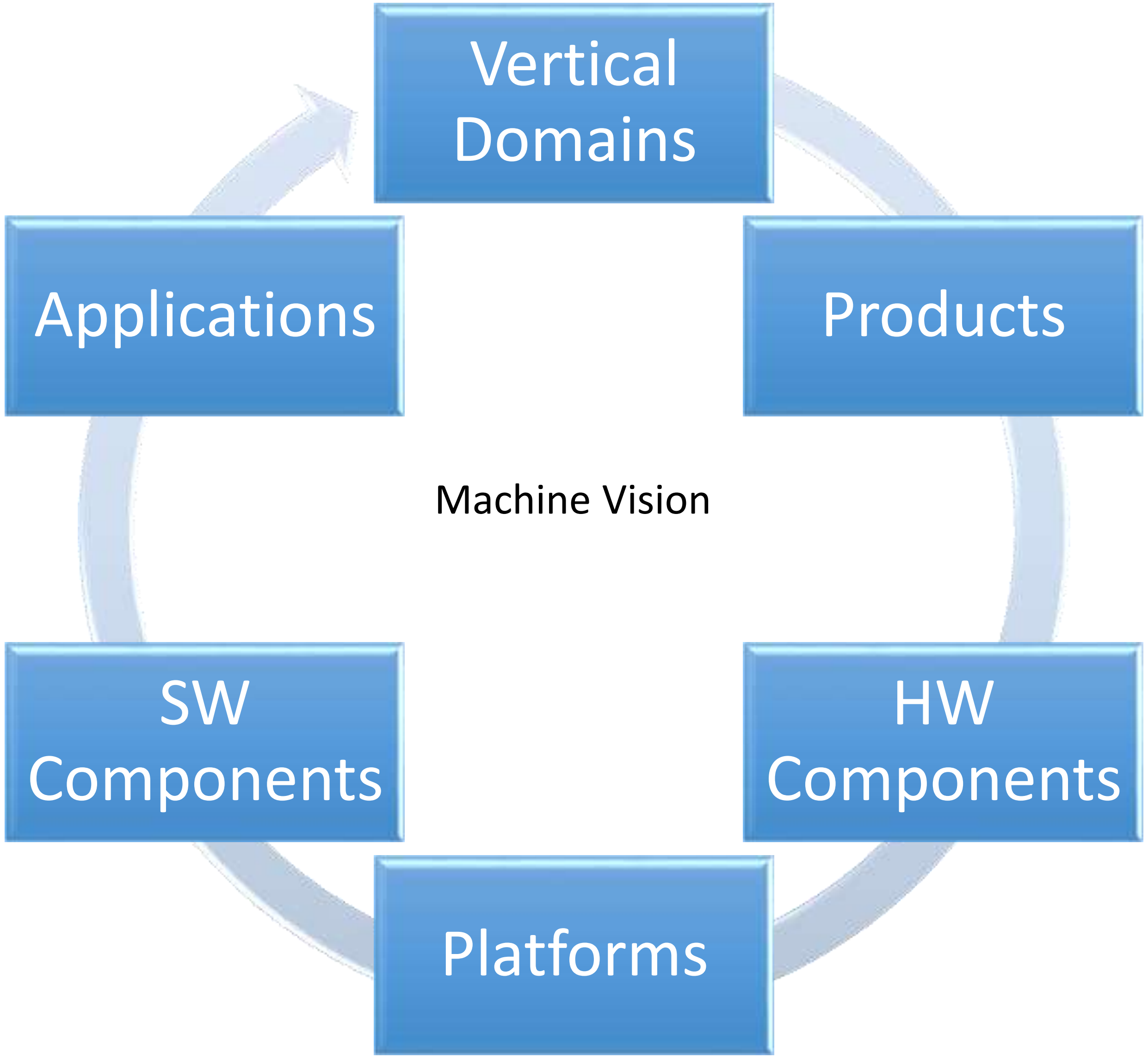
3D Machine Vision Market Global Forecast to 2020 *says, the market is expected to grow at a CAGR of 10.53% during the forecast period between 2015 and 2020 driven by 3D machine vision technology is due to its growing applications in the automotive and electronics industries.*

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

An Overview of Machine Vision dimensions

An holistic view
of Machine Vision
for industry

Web Sources
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.visiononline.org commercial site
http://www.vision-systems.com/ commercial site



Vertical Domains

Vertical Domains

Industrial

- Automotive
- Consumer Electronics
- Electronics & Semiconductor
- Printing
- Metals
- Wood & Paper
- Food
- Packaging
- Robotics
- Manufacturing
- Rubber & Plastics
- Pharmaceutical
- Glass
- Machinery
- Solar Panel Manufacturing
- ...

Non-industrial

- Healthcare & Medical Imaging
- Postal & Logistics
- Intelligent Transportation System (ITS)
- Security & Surveillance
- Agriculture
- Consumer Electronics
- Autonomous Cars
- Smart cities
- Cultural Heritage
- Education
- …

Enabling Embedded Vision

ADAS, Machine Vision, Surveillance, Drones, Medical, ProAV ...

Courtesy of APEX - http://apexcontrols.cc/

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**

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

HW Components

HW Components

- **Cameras**
- Sensor types
- LED Lightings
- Frame Grabbers
- Selections of characteristics
  - **Interface standards**
  - Frame Rate (Area Scan and Line Scan)
  - Format (25–125 fps, >125 fps, <25 fps)

- 3D Cameras
- Multispectral/Hyperspectral Cameras
- Smart Cameras
- CCD – CMOS Cameras
- High-Speed Video Cameras
- Line-Scan Cameras
- IR Cameras & Detectors
- Barcode Scanners
- Low-Light/Intensified Cameras
- X-Ray Cameras
- Analog Cameras
- Standard cameras

Camera Link
CoaXPress
FireWire
GigE Vision (1.2 Gbps)
USB 3.0 (3 Gbps)
…

Interface	Cable length in m	Bandwidth max. in MB/s.	Multi-camera	Cable costs	"Real-time"	"Plug & Play"
USB 2.0	5	40	🟧	🟧	🟥	🟩
FireWire	4.5	64	🟧	🟧	🟩	🟩
GigE Vision	100	100	🟩	🟩	🟧	🟧
USB3 Vision	8	350	🟧	🟧	🟩	🟩
Camera Link	10	850	🟥	🟥	🟩	🟥

	USB 2.0	USB3 Vision	FireWire A	FireWire B	GigE Vision	Camera Link
Bandwidth	50 MB/s	400 MB/s	50 MB/s	100 MB/s	125 MB/s	850 MB/s
Cable Length	3 m	3 m	4, 5 m	10 m	100 m	10 m
Camera Standard	N/A	USB3 Vision	IIDC (DCAM)	IIDC (DCAM)	GigE Vision	Camera Link
CPU Usage	High	Low	Medium	Medium	Medium	Low
Cost	Low	Low/ Medium	Medium	Medium	Medium	High
Power Over Cable	Yes	Yes	Yes	Yes	Yes	Yes

Computing Platforms

Platforms

1. Acquisition devices
2. Cable and Camera System Interfaces
3. Processing units:
   - FPGAs,
   - DSPs, Microcontrollers,
   - **ARM-based or embedded** processors   (Cortex A9 … Snapdragon Neural processor Qualcomm)
   - **General Purpose** Multi-cores, many-cores (PC-based solutions)
   - **GPUs** (Nvidia)
   - Cloud based and HPC- to Exascale computing

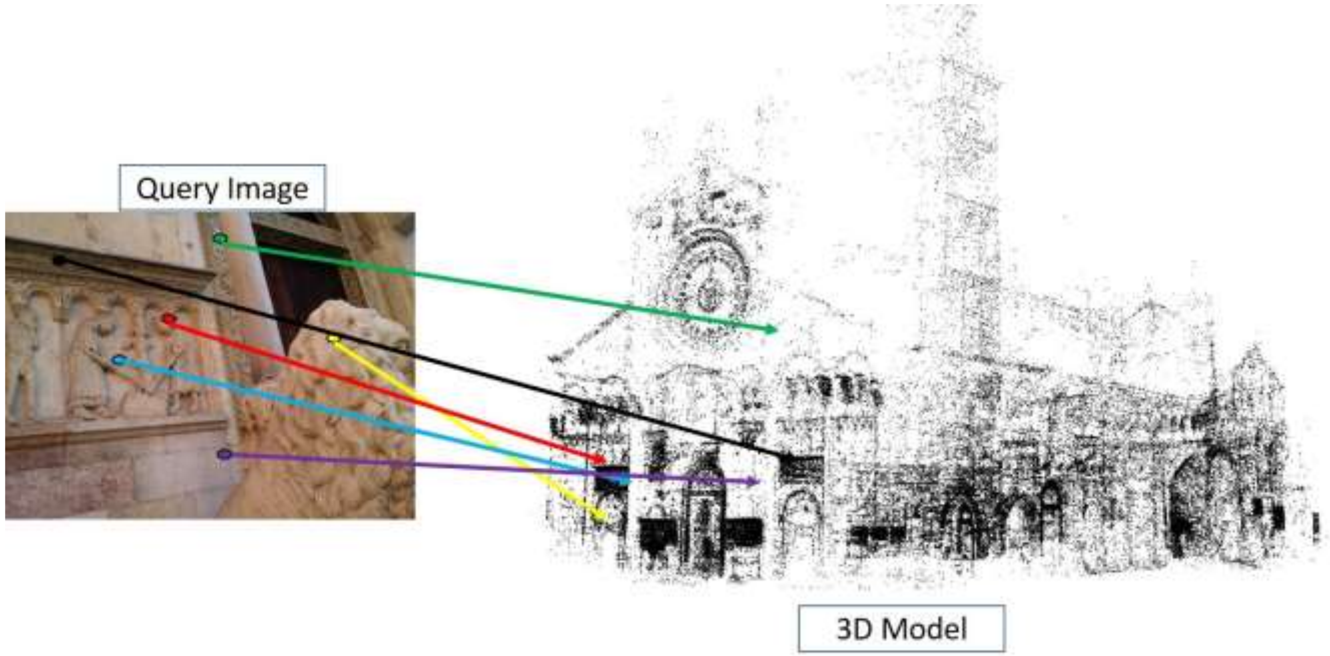
"Processors for Embedded Vision: Technology and Market Trends” 2015 and 2016

An exampleof 3D Reconstruction : Very different computational time.

- 2D to 3D feature mapping O(N x M x D)
  - N : query keypoints, (c.a. 2000/image)
  - M : dataset keypoints (in the order of 10^6)
  - D : keypoints vector size (128 for SIFT)

256 x 10^9 computations of MSR

CUDA	OPENCL
Proprietary	Open source
No CPU support	Supports OpenCL CPU devices
Works only on Nvidia GPU	Cross Platform
More mature tools, including a debugger and a profiler	Old C99 style programming, moving now to C++ 1 1 and C++14 standards. Native / low-level programming
Easier to use, high level programming	
Data must be allocated and copied to GPU	On Intel and AMD architecture we can use zero-copy buffers



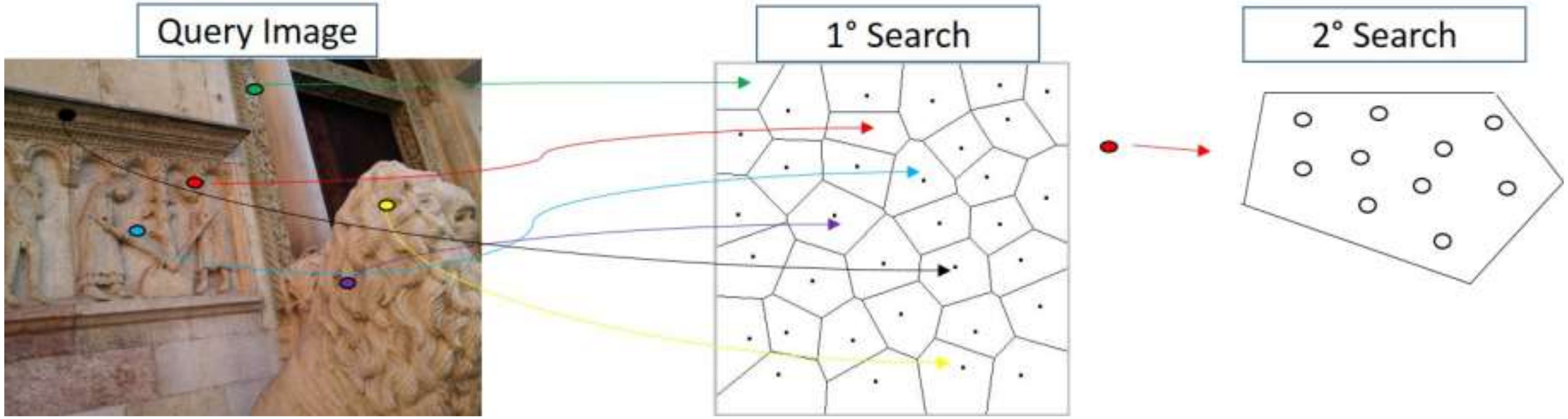
An Example: Time comparison

CPU - time	CUDA - time	OPENCL - time	OPENCL (Zero Copy) - time
30 min	104 sec	147 sec	**11 sec**
1x	Up to 17x	Up to 12x	**Up to 166x**

Standard approach

CPU - time	CPU BoW - time
30 min	0.511 sec
1x	Up to 3600x

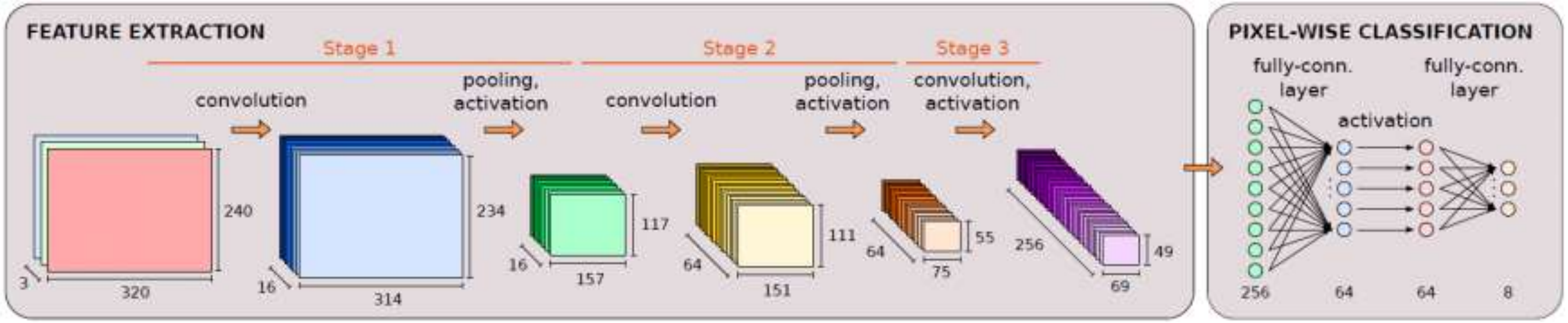
BoW approach



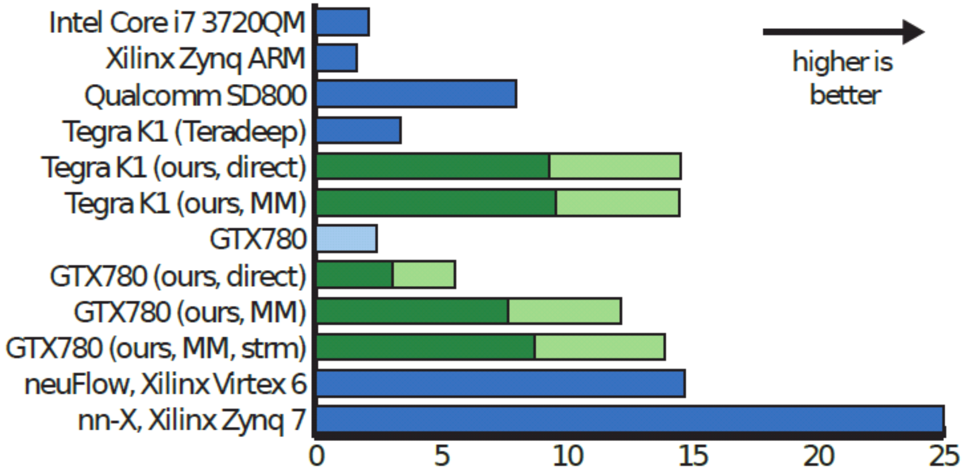
«Low» computation
« High» Memory

Optimizing Image Registration for Interactive Applications. AVR (2016)
Gasparini, R., Alletto, S., Serra, G., Cucchiara, R. AVR (2016)

An Example: Scenel Labeling on GPUs



80.6% accuracy



11 frame/s (320x240) @ 11W On NVIDIA Tegra K1

Thanks to L. Cavigelli, M. Magno, L. Benini, «Accelerating Real-Time Embedded Scene Labeling with Convolutional Networks», DAC 2015

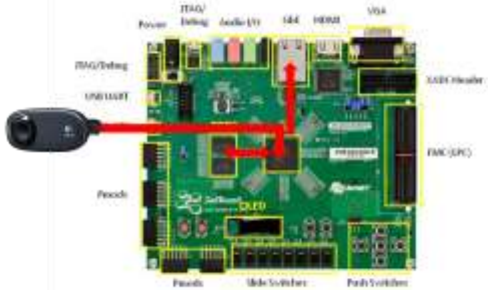
An Examples with Xilinx

Camera Trap equipped with Cellular ~400$

CNN: ResNet-18

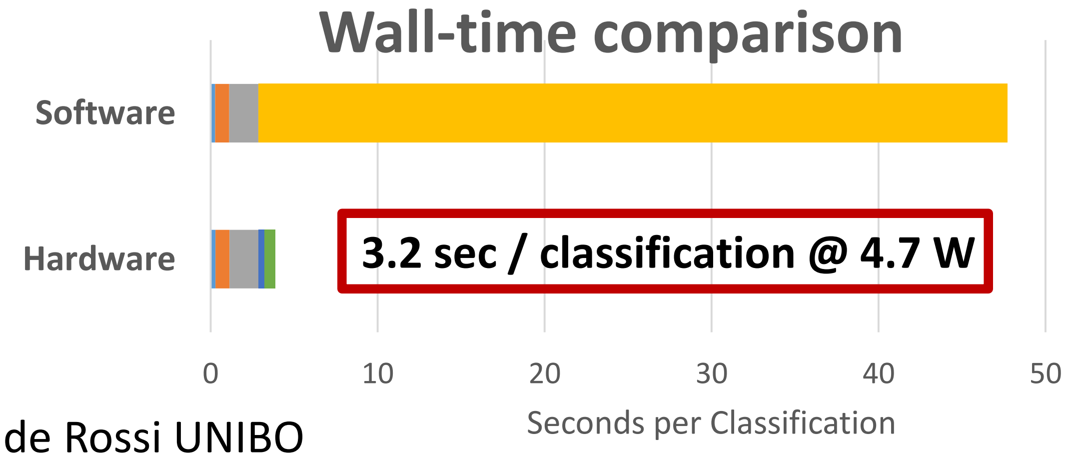
Computing System: Xilinx Zedboard

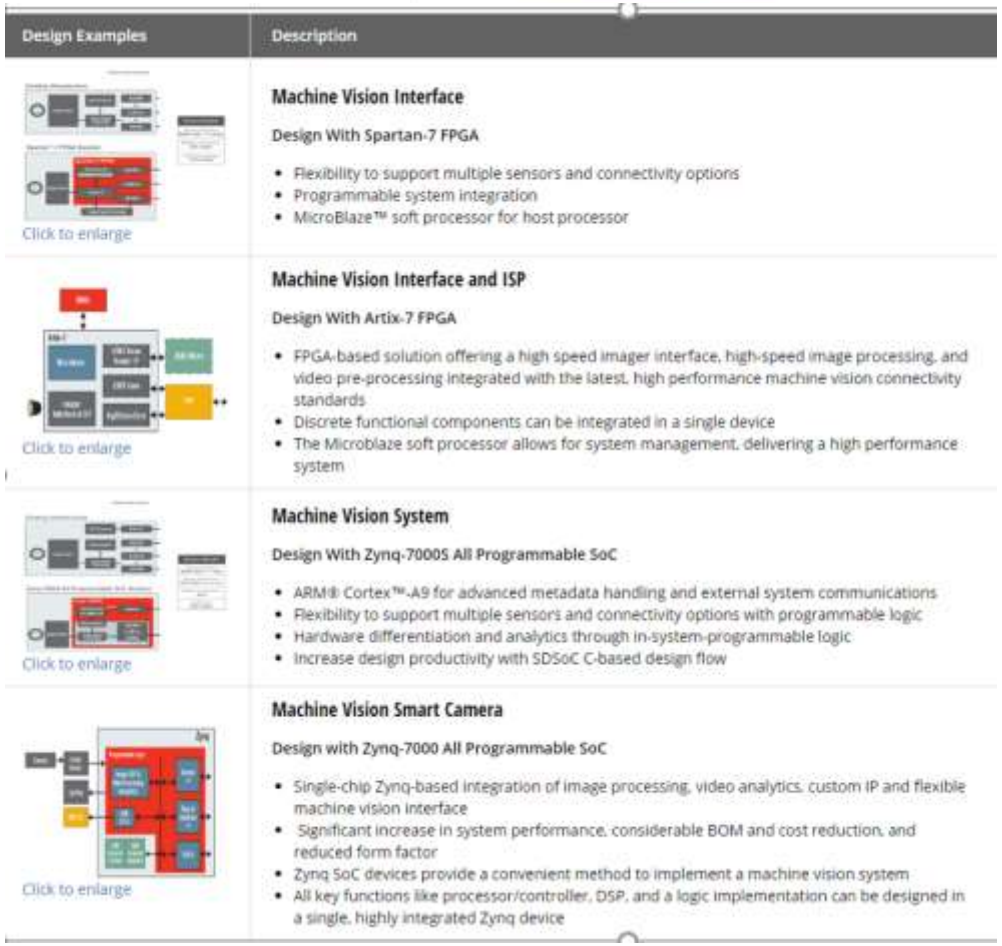
Running Linux!



Other Applications of Embedded Classifiers:

- Robotics and Drones
- Autonomous Cars
- CCTV





Design Examples	Description
Click to enlarge	**Machine Vision Interface** Design With Spartan-7 FPGA • Flexibility to support multiple sensors and connectivity options • Programmable system integration • MicroBlaze™ soft processor for host processor
Click to enlarge	**Machine Vision Interface and ISP** Design With Artix-7 FPGA • FPGA-based solution offering a high speed imager interface, high-speed image processing, and video pre-processing integrated with the latest, high performance machine vision connectivity standards • Discrete functional components can be integrated in a single device • The Microblaze soft processor allows for system management, delivering a high performance system
Click to enlarge	**Machine Vision System** Design With Zynq-7000S All Programmable SoC • ARM® Cortex™-A9 for advanced metadata handling and external system communications • Flexibility to support multiple sensors and connectivity options with programmable logic • Hardware differentiation and analytics through in-system-programmable logic • Increase design productivity with SDSoC C-based design flow
Click to enlarge	**Machine Vision Smart Camera** Design with Zynq-7000 All Programmable SoC • Single-chip Zynq-based integration of image processing, video analytics, custom IP and flexible machine vision interface • Significant increase in system performance, considerable BOM and cost reduction, and reduced form factor • Zynq SoC devices provide a convenient method to implement a machine vision system • All key functions like processor/controller, DSP, and a logic implementation can be designed in a single, highly integrated Zynq device

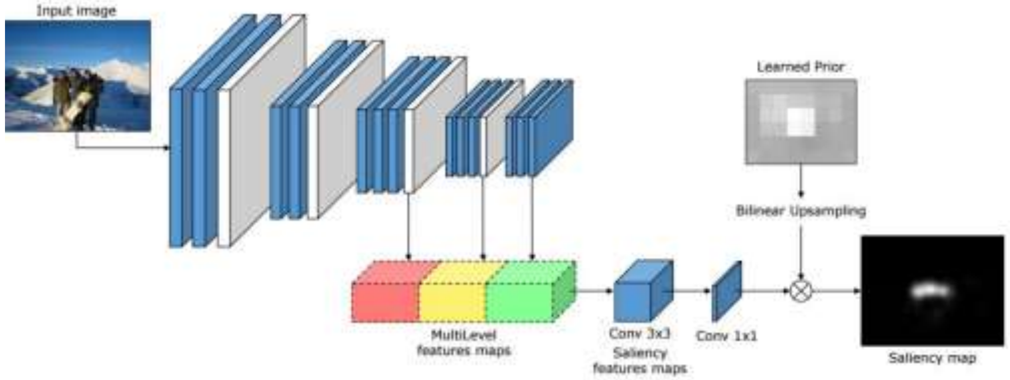
Thanks to Davide Rossi UNIBO

Software components

SW Components

- Data acquistion/compression tools
- Image Processing/Analysis
- 3D Computer Vision
- Motion-based Computer Vision
- Pattern Recognition and Statistical-based Learning
- Deep Learning
- General AI solutions

- Algorithms
- Libraries
- Tools
- Datasets



Software components- Libraries

- **MVTec HALCON :** 2D and 3D vision, GPU , mPc-based and embedded support , expensive
- **MvT Merlic :** simplified interfaces also for Mobile
- **Cognex VisionPro : oriented to robotic, easy interface Quickbuileder , bar code sw , limited in 3D**
- **Matrox** Imaging
- **National Instrument**
- **VisionServer** Acceleretor for Allen-Bradley PLCs

OpenSource

- **Open CV** Open Source Computer Vision Library , (http://opencv.org).
- **Simple CV** (http://simplecv.org) Sight Machine
- **Darwin**, an open-source platform-independent C++ framework for machine learning and computer vision Australian National University (Canberra, Australia; www.anu.edu.au).
- **Open Vision Control**, object motion detection (http://openvisionc.sourceforge.net).

- … and all libs for machine learning
- Standard machine vision libaries are often not enough after the DL Revolution

http://www.jmakautomation.com/halcon-vs-cognex-visionpro

Applications

Applications

New Deep-Learned based solutions

AVI Automatic Visual Inspection

2D/3D Robot manipulation

(picking and stowing)

The Human Factor and HMI

Automous Guidance and Situation Awareness

Amazon Picking challenge
Winner 2016 Delt Univeristy
No CAD MODEL **Deep Learning recognition**

[Past experience UNIMORE – Marchesini Patent 2009]

Autonomous Guidance in cars

UNIMORE – Dr(eye)ve
Modena Smart Area Maserati

Vision for Human-Machine Interaction with Vehicles and Robots

UNIMORE POSEeidon Depth based deep learning

HAxIA Regional project for egocentric Vision in industry

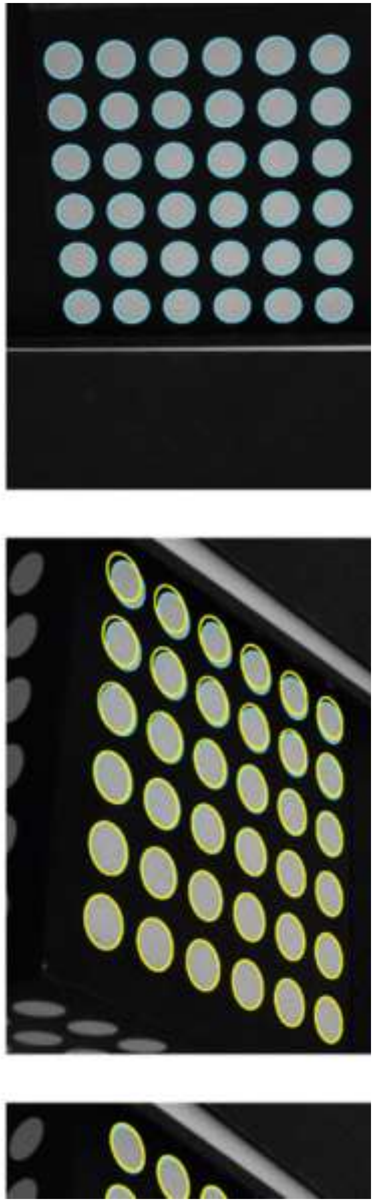
Industrial AGV (automatic Guided Vehicles) situation awareness

[UNIMORE – Infolog 2016]

Quality Control and Inspection
Measurement

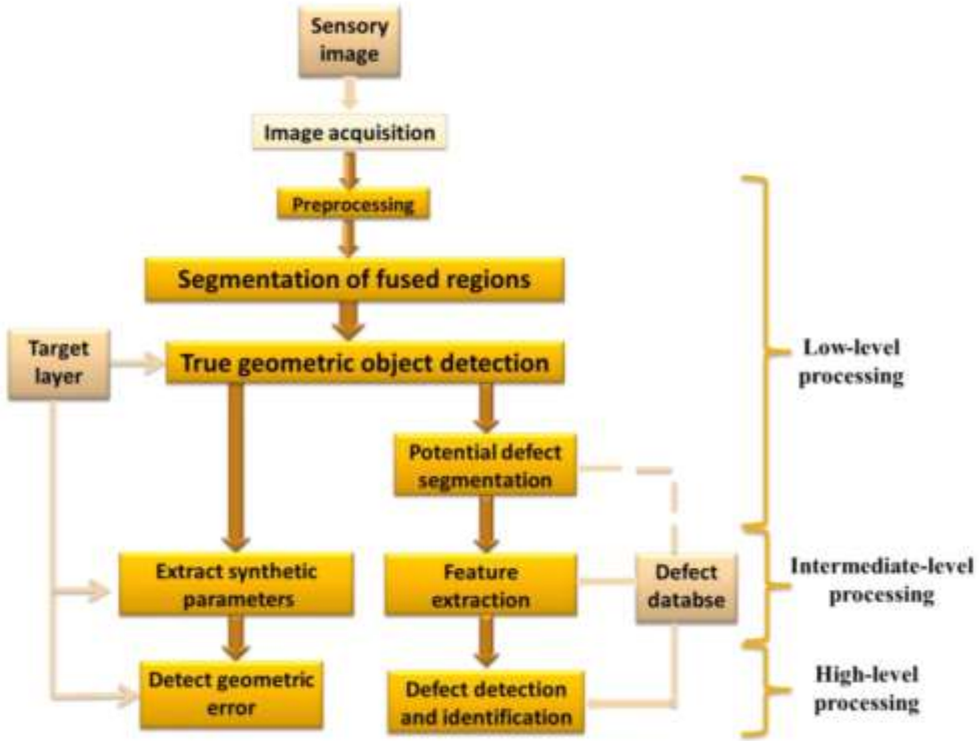
Seven challenges for Machine Vision

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.



Omography for inspection

- Automated Visual Inspection
- Classical machine vision approaches



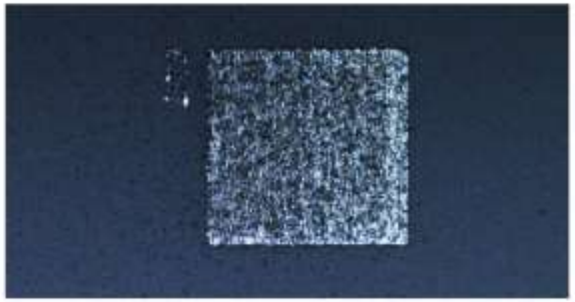
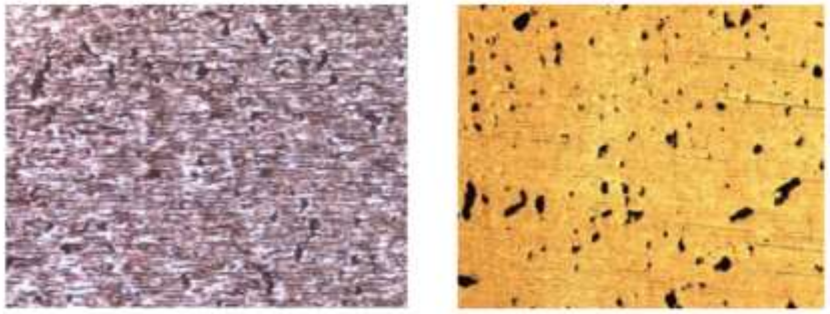
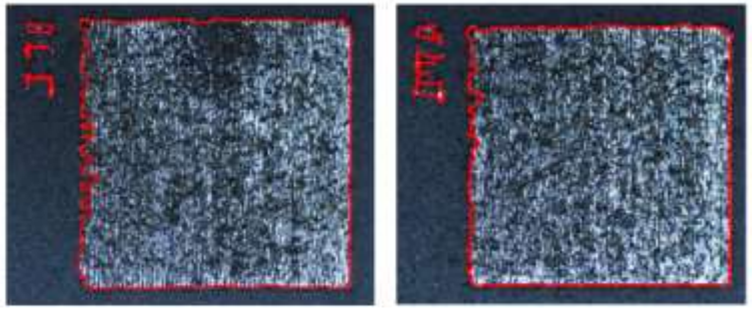


Figure 44- Example of image with the setup using 8 megapixel camera and adjusted level of intensity for square LED mounted at an angle from the build.





Thanks to M. Aminzadeh, Phd Dissertation in Mechanical Engineering Georgia Tech 2015

Quality control, inspection , measurement

The standard pipeline

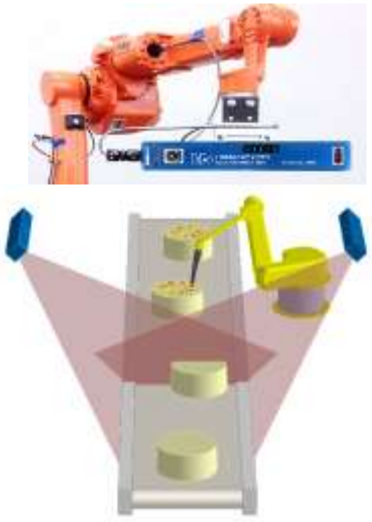
Model

Data

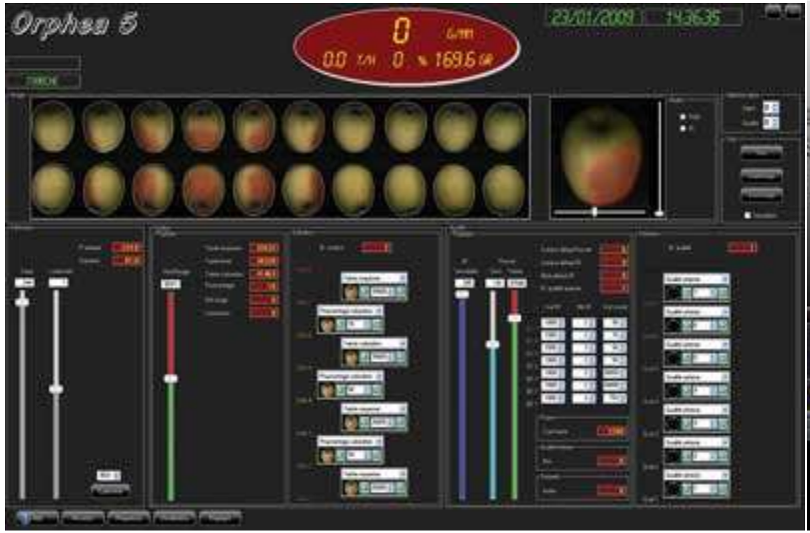
Image processing → [Segmentation and detection] → Feature extraction → Classification, identification..

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

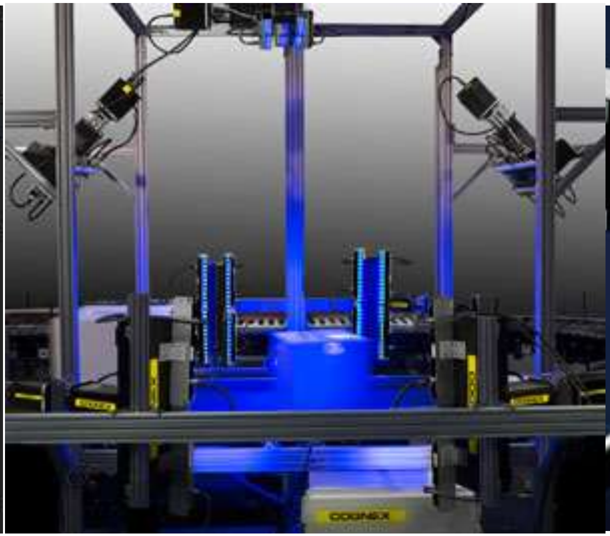
3D Hermary Light Scanner



Matrox Imaging based apple measurement



Cognex Dataman 503 UV-based scanner

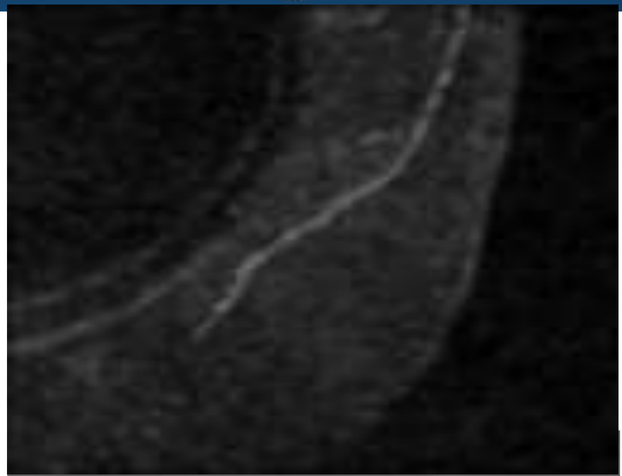


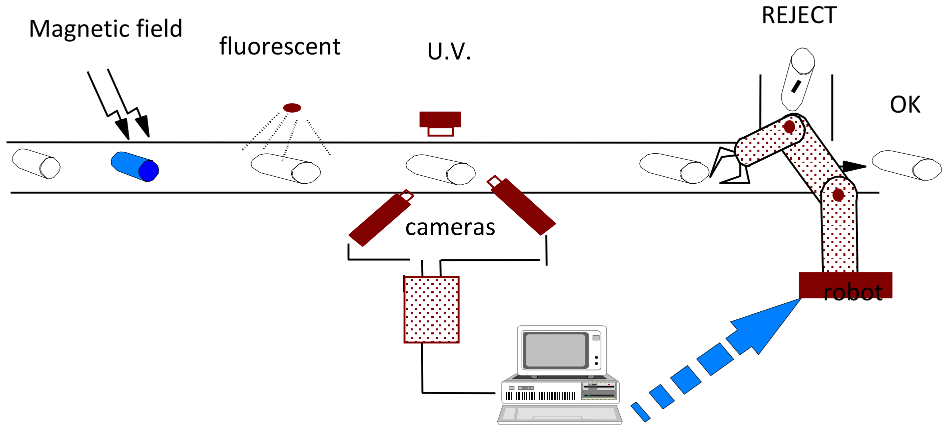
AIS http://aisukltd.com/machine-vision-systems/

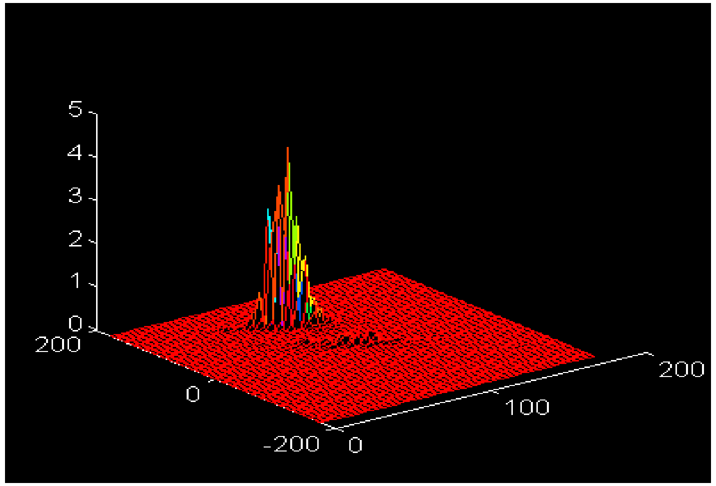
Vision and learning in quality inspection

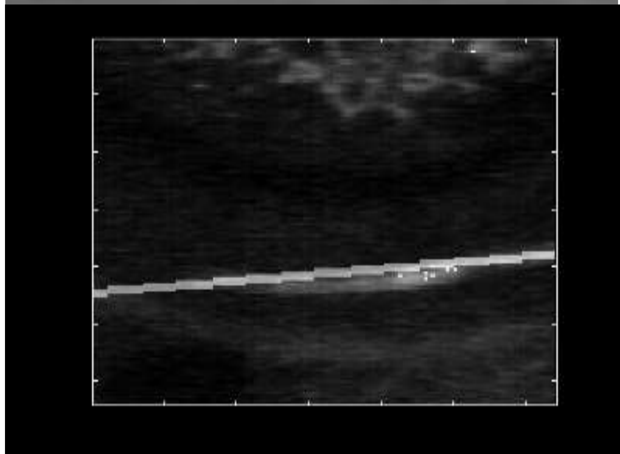
- when the CAD model is not necessary
- The defect model is needed

- 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,

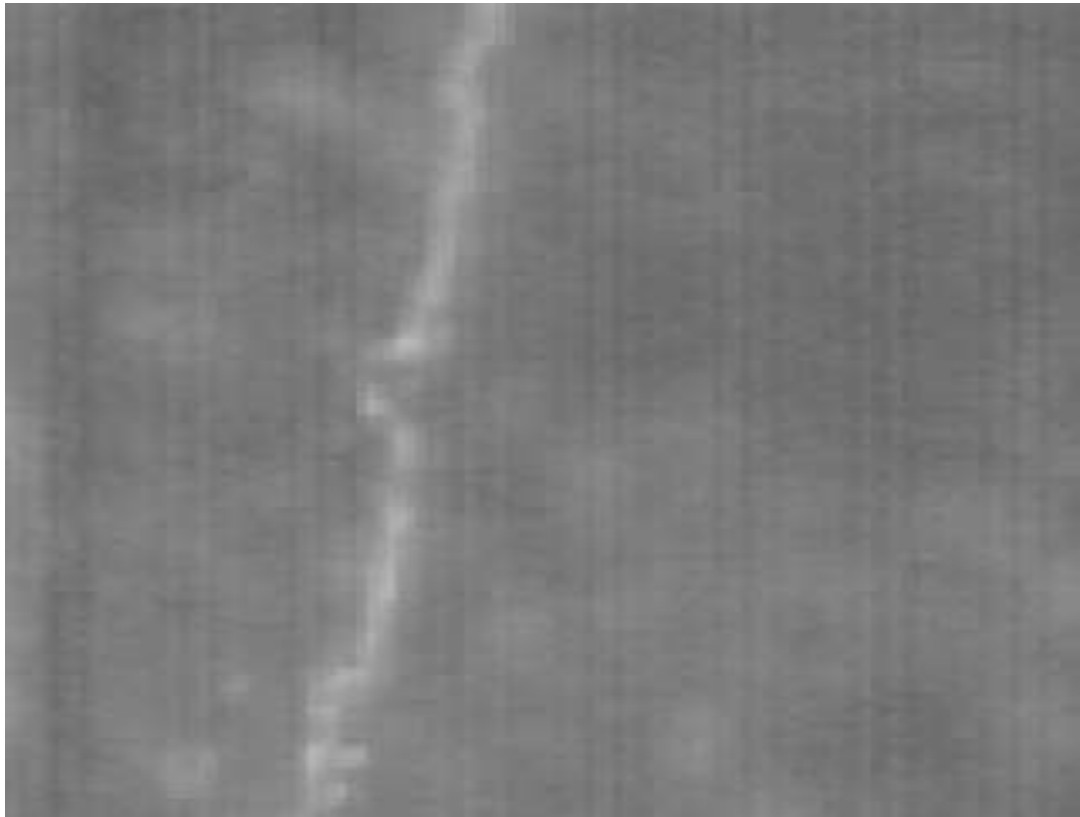
✓ Pattern Recognition: a k-nearest neighbour algorithm,

✓ 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

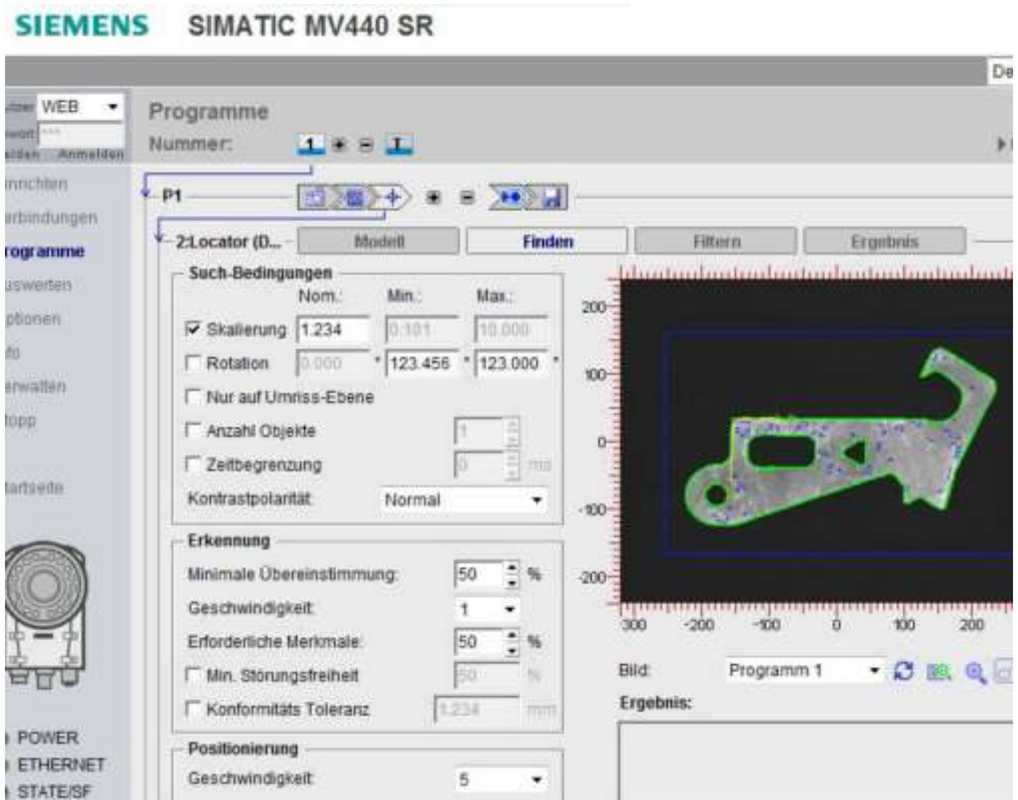
	Discrim	Logdisc	Quadisc	NN	Predict	c4.5 tree	c4.5 rules
CH	0.853	0.857	0.853	0.885	0.873	0.959	0.959
H1 H2	0.855	0.928	0.316	0.845	0.864	0.933	0.933



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

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

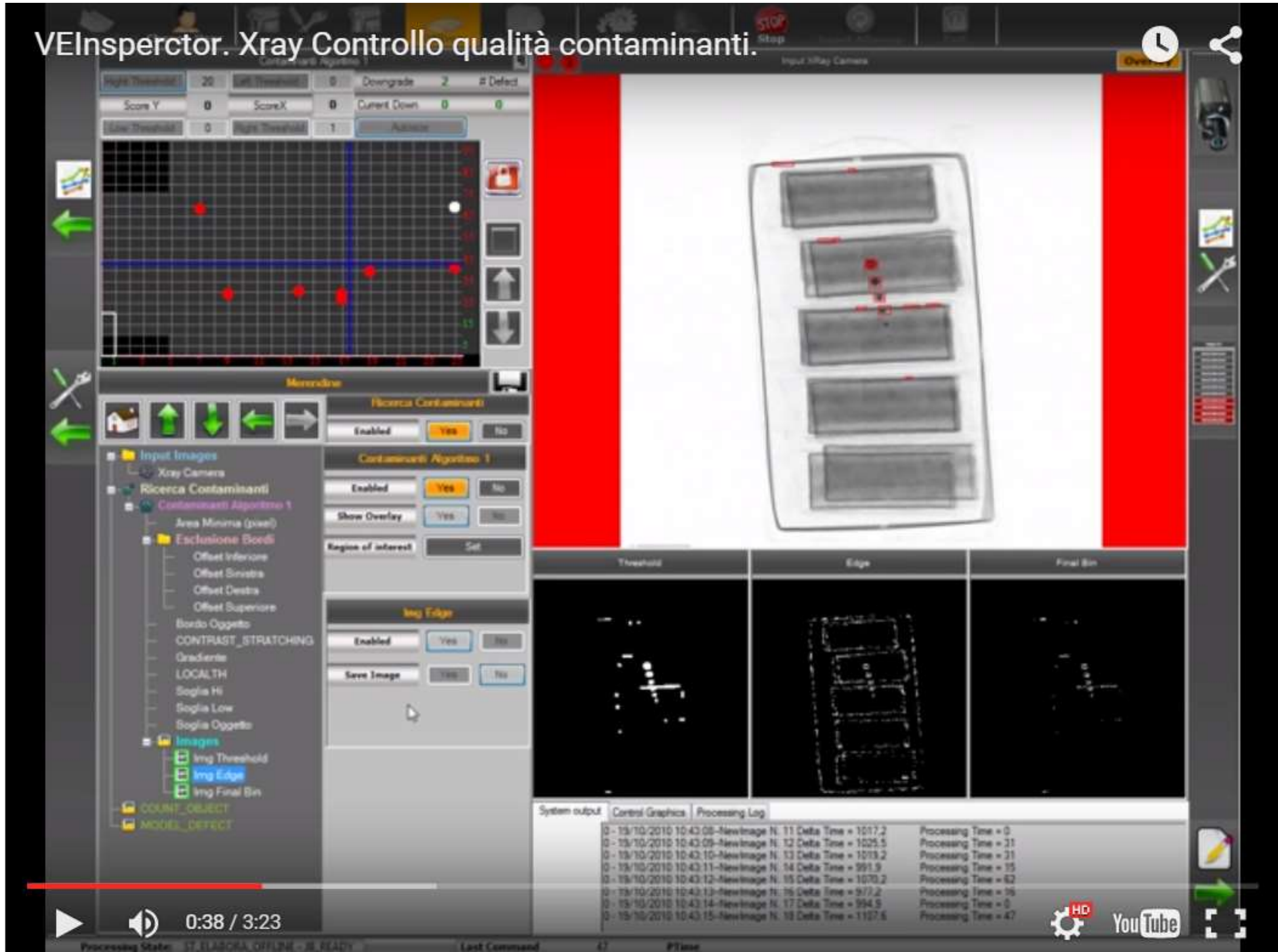
- Eg. SIEMENS SIMATIC MV 440
- Pat-Genius" object recognition license, SIMATIC MV440 for object recognition position detection, counting etc., reading 1D bar codes and 2D matrix codes, text recognition, to check the position of a label and check the inscription (reading and comparing) of plain text in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



Nowadays ,they are commercial tools

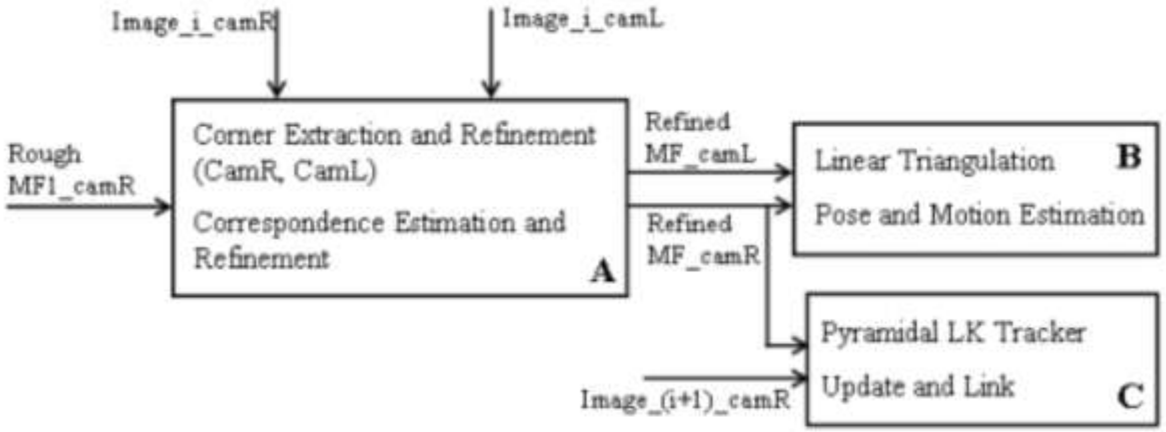
- ✓ Machine vision in constrained scenario
- ✓ Structured light,
- ✓ Mainly model-based
- ✓ Image processing and measurement
- ✓ 2D and 3D geometry

- NOW commercial tools.

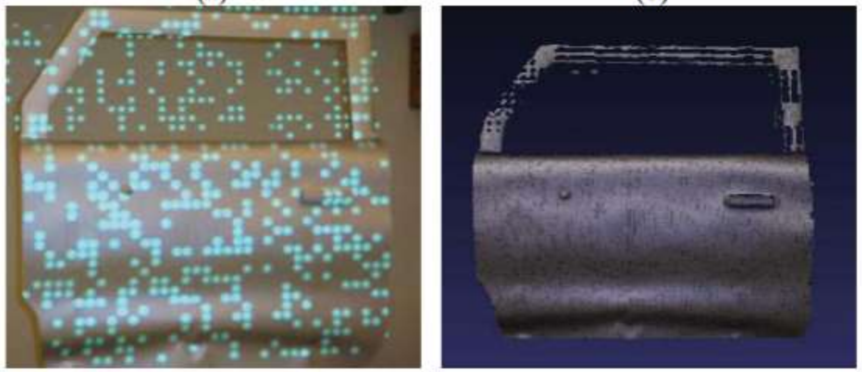


- Thanks to VISION-E srl
- UNIMORE spin-off

Now improved in 3D (2010)

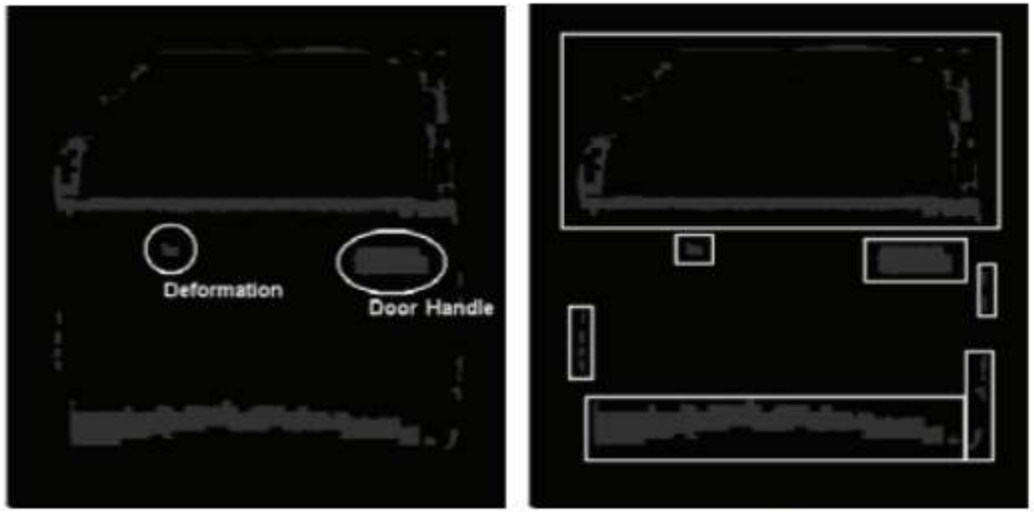


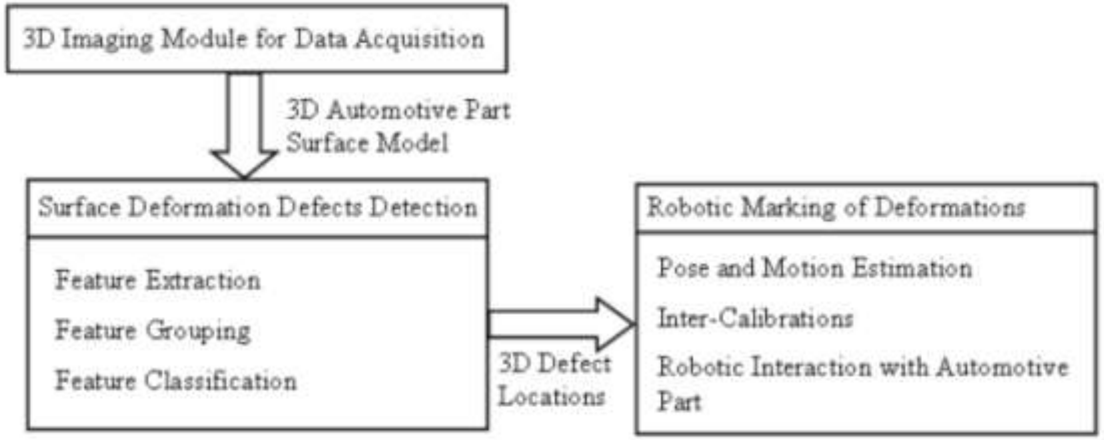
(a) (b)



Automated Surface Deformations Detection and Marking on Automotive Body Panels

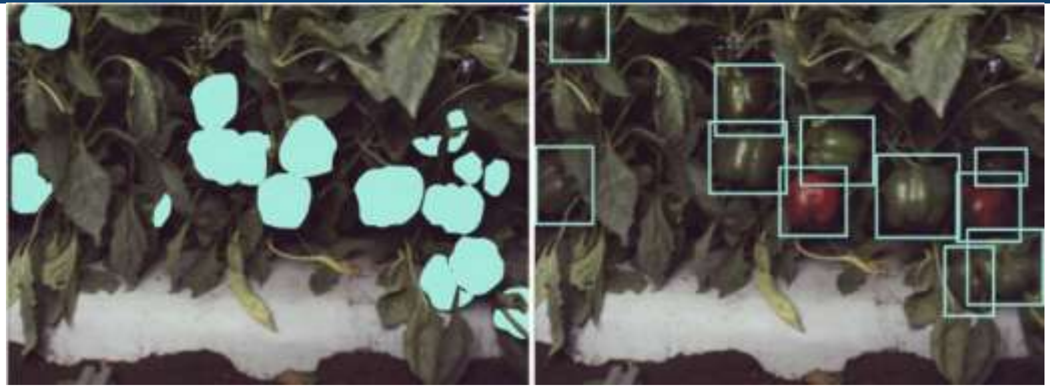
Valentin Borsu, Arjun Yogeswaran, and Pierre Payeur



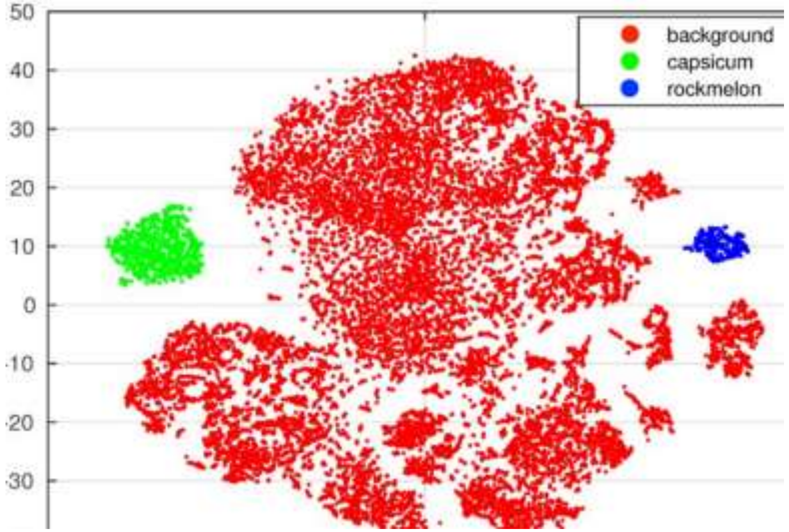


Towards Deep Learning based Pattern recogntion approaches

- Why?
- Do you really need to have a model?
- Can you learn by the appearance what is good or not?
- What can you do when the target example are few?
- Can you find a general-purpose approaches?

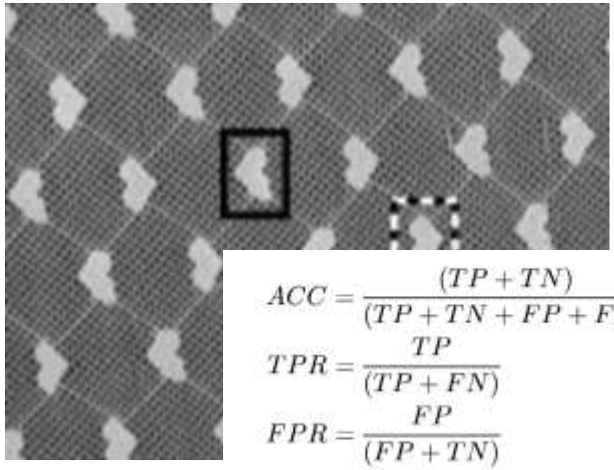


DEEPFruit 2016



96 images, Fasr R-CNN VGG,
RGB + NIF images
t-Distributed Stochastic Neighbour Embedding space

- Deformable defect detection in warp knitted fabric



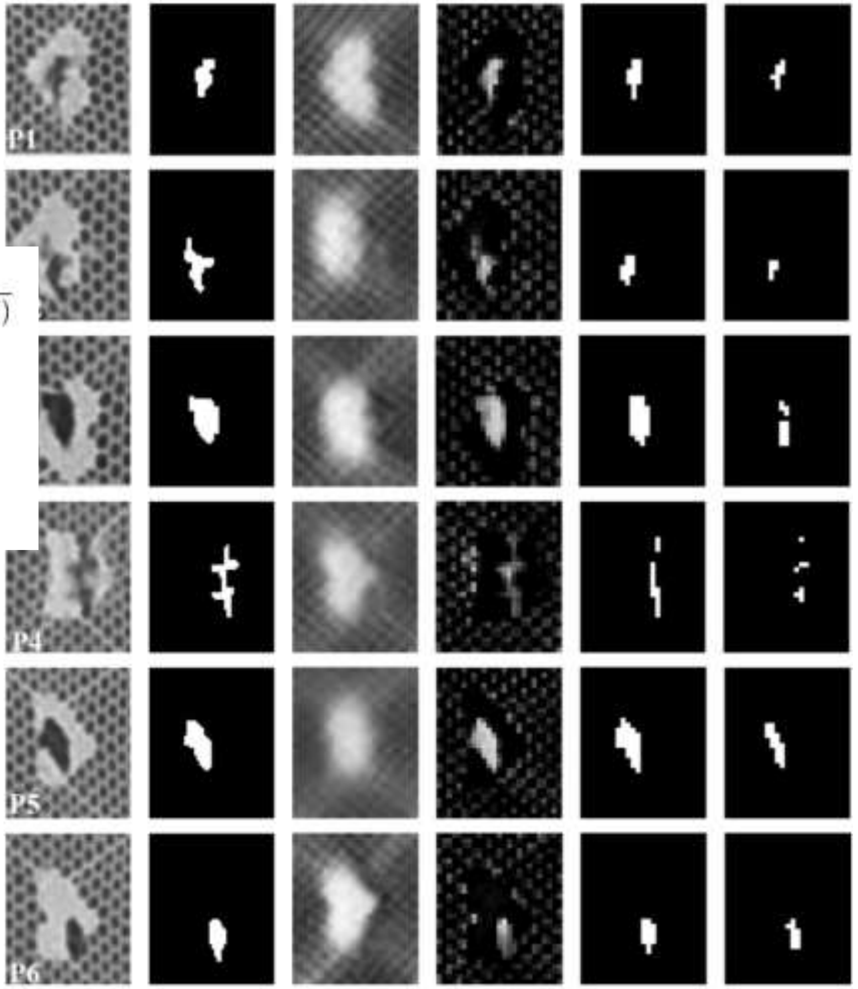
$$ACC = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

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

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

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

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



(a) (b) (c) (d) (e) (f)

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

Yundong Li, Weigang Zhao, and Jiahao Pan

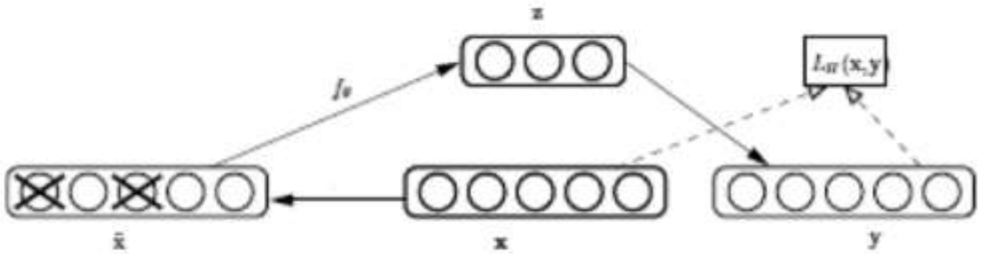
LOCATING ACCURACY COMPARISON OF EXPERIMENT 1

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**

Denoising Autoencoder (DA)

- Data $x \in [0,1]^d$ be the input vector and $y \in [0,1]^d$

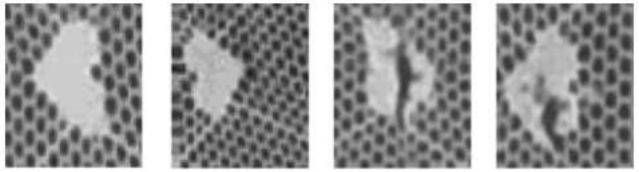
- Autoencoder
$$z = f_\theta(\tilde{x}) = s(W\tilde{x} + b)$$
$$y = g_{\theta'}(z) = s(W'z + b').$$



- Loss
$$L(x, y) = \|x - y\|^2.$$

- Denoising Autoencoder
$$z = f_\theta(x) = s(Wx + b)$$
$$y = g_{\theta'}(z) = s(W'z + b')$$

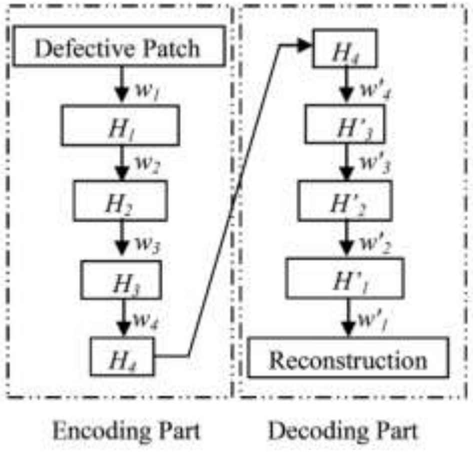
- Stacked DA with Fisher Criterion
- (ratio between intra class distance and interclass distance)

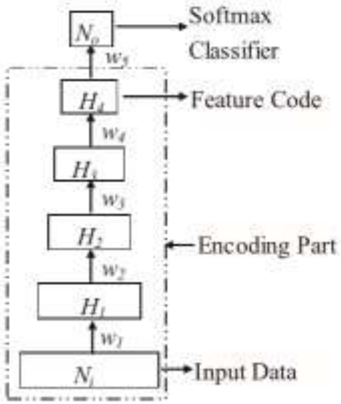


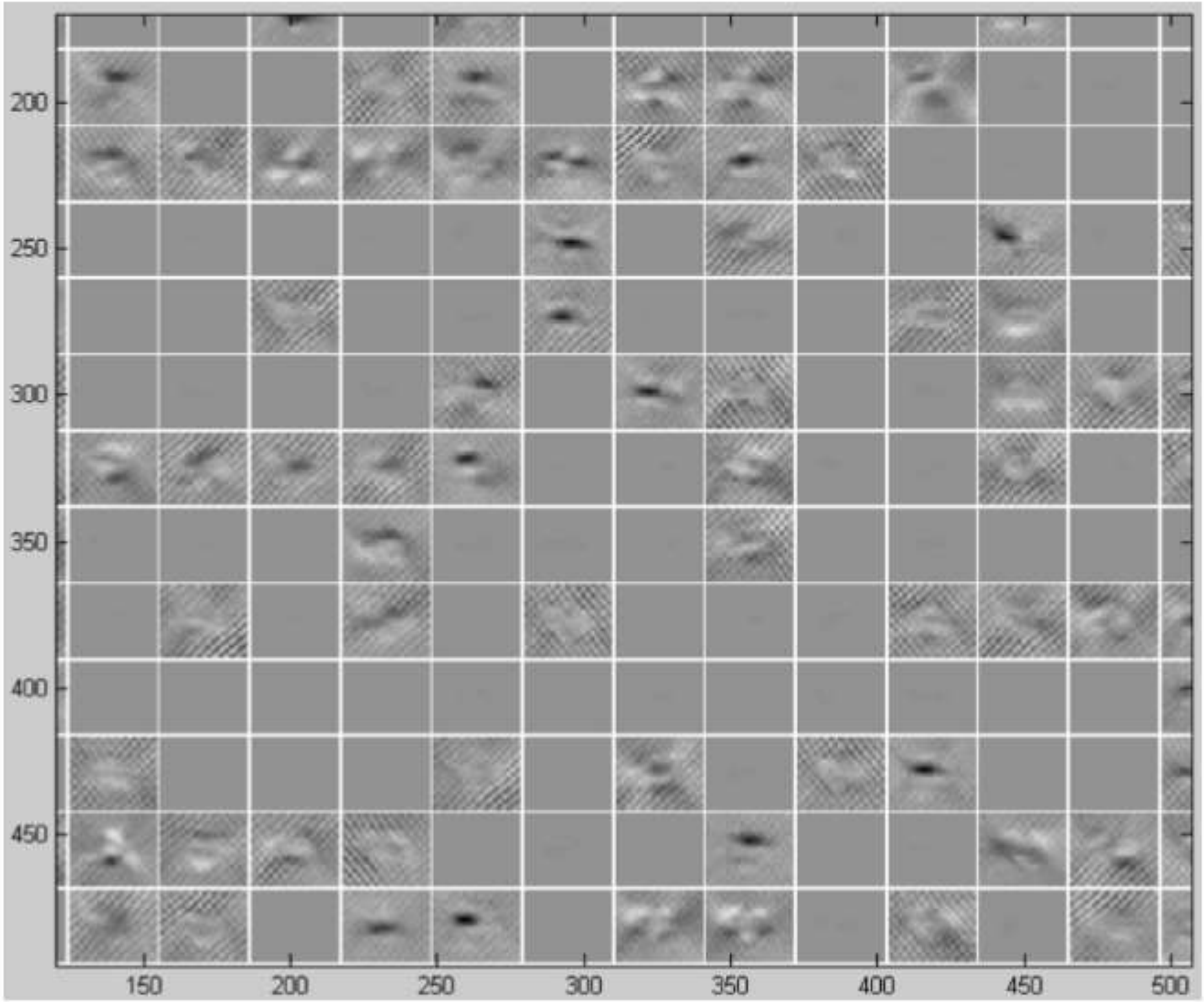
$$J_{(W,b)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\left\|h_{w,b}(x^{(i)}) - y^{(i)}\right\|^2\right) + \lambda \frac{J_{\text{int } ra}}{J_{\text{int } er}}$$

a detail of the autoencoder

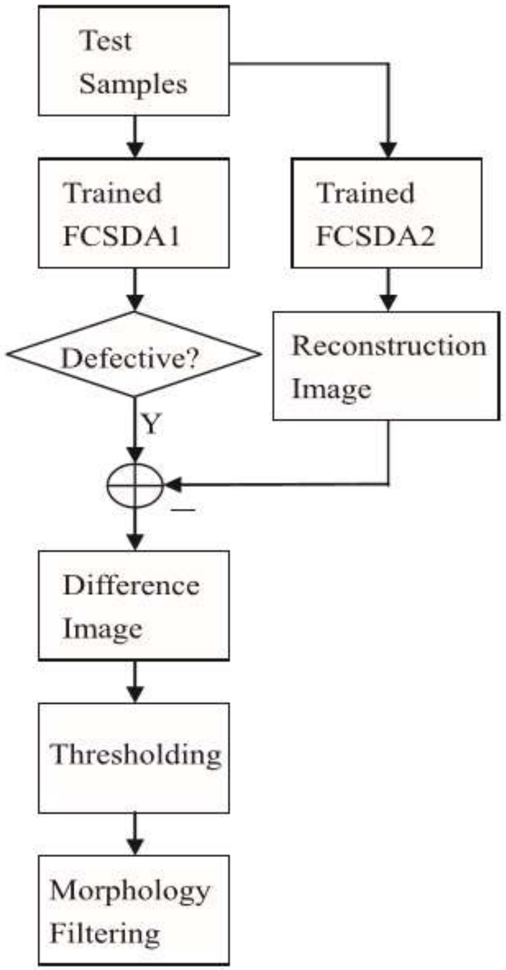
- The encoder (and decoder for reconstruction)
- 30x25 patches = 750 vector
- 750,600,400,200,100,3
- 2000 positive, 600 defect







• 0,21ms detection on a standard Corei5



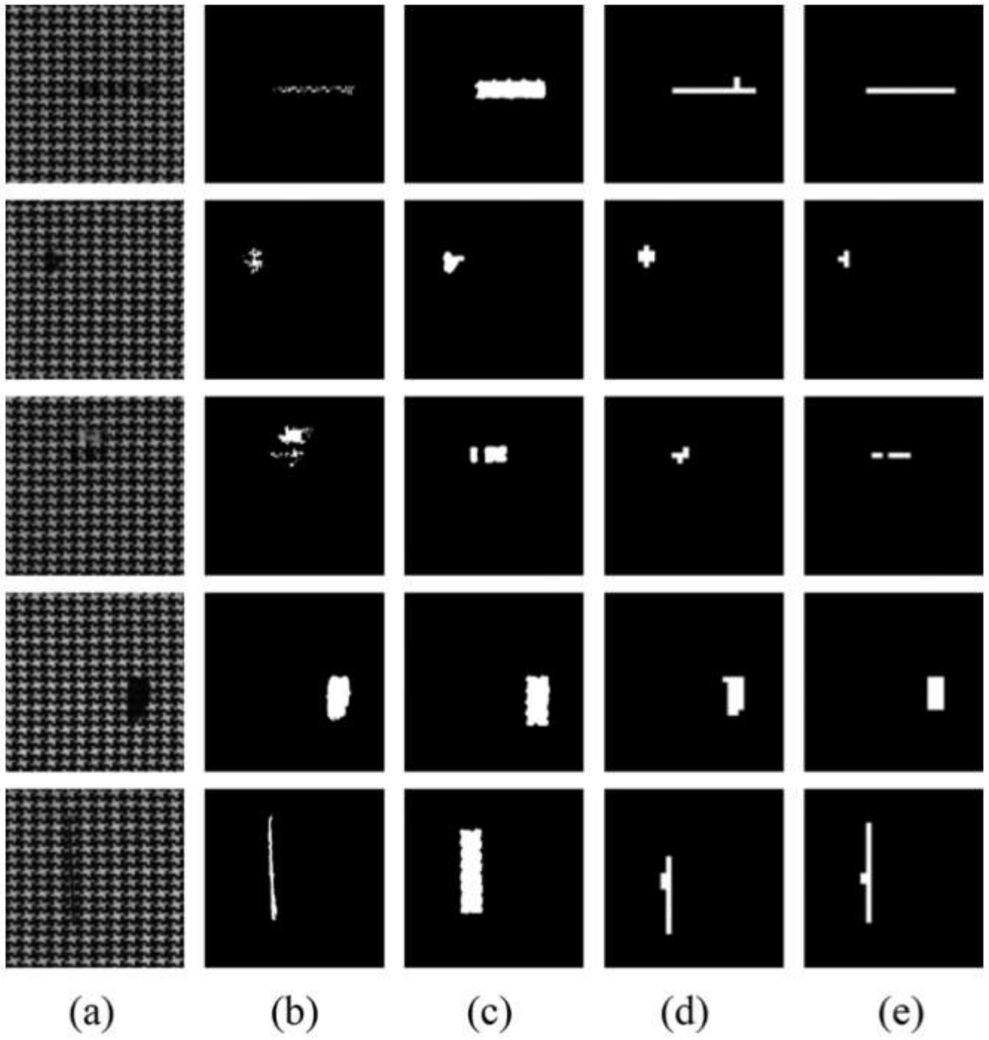


Fig. 10. Defect locating results comparison. (a) Defective images. (b) The ground-truths labeled by manual. (c) Results of ID method. (d) Results of SDA method. (e) Results of FCSDA method.

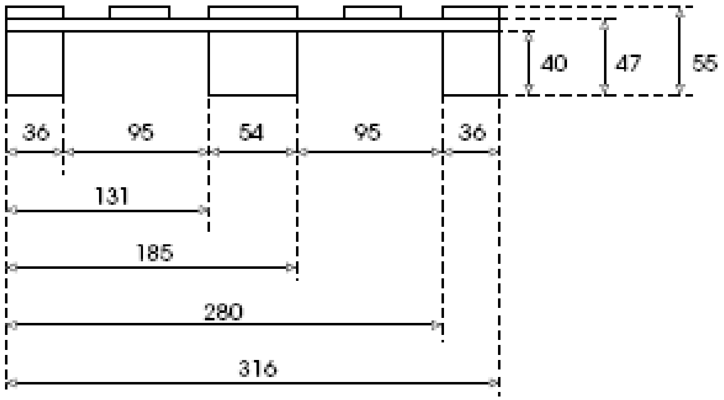
Identification and Localization
3D localization for Grasping

Localization in 2D images- classic methods

- Image segmentation
- Object localization in 2D images

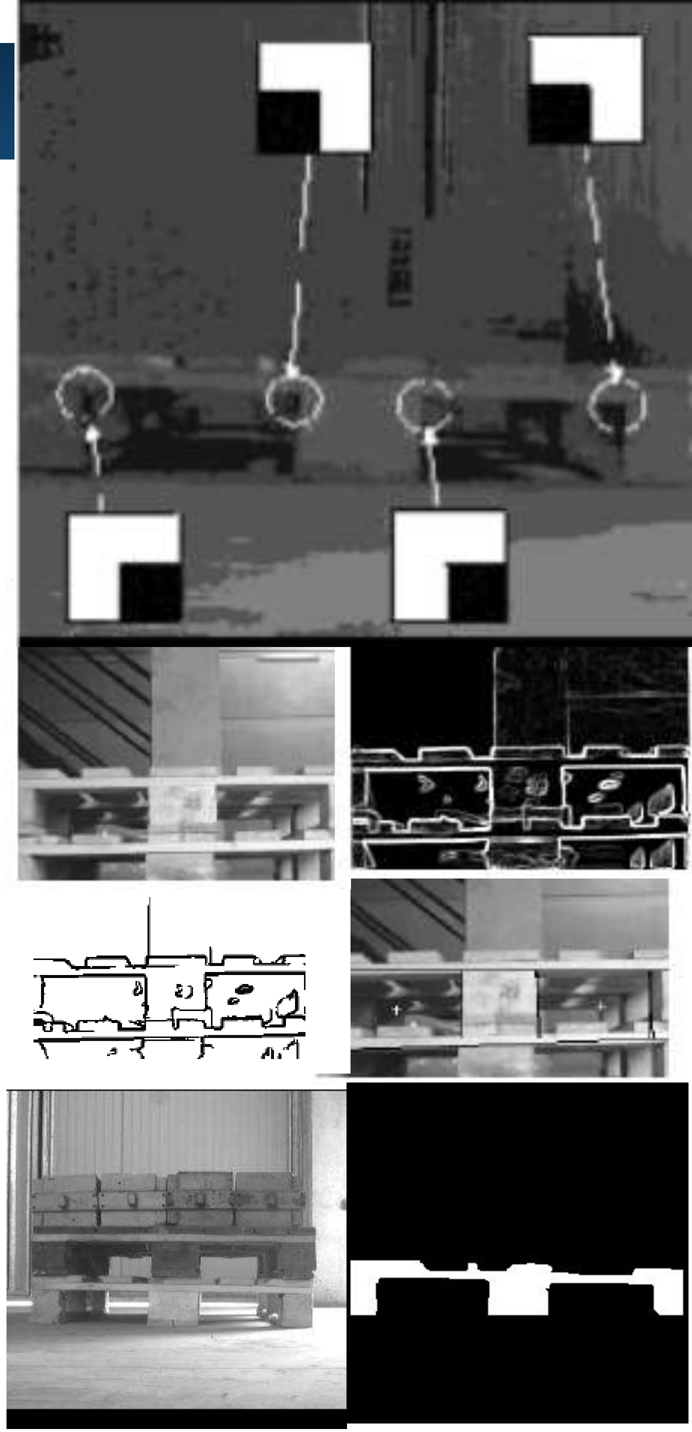
  Traditionally

  - 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

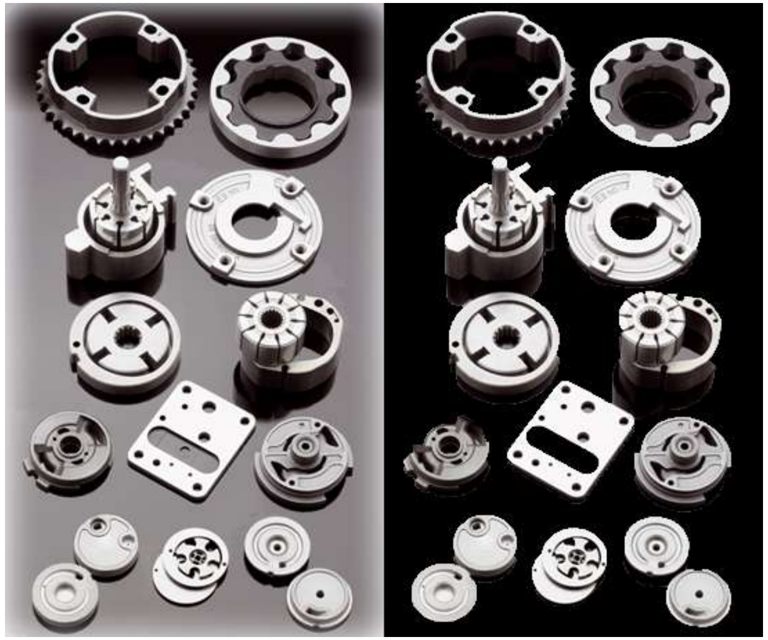
Localization in 2D images Grab cut (and Deep contour)

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

"GrabCut" — Interactive Foreground Extraction using Iterated Graph Cuts

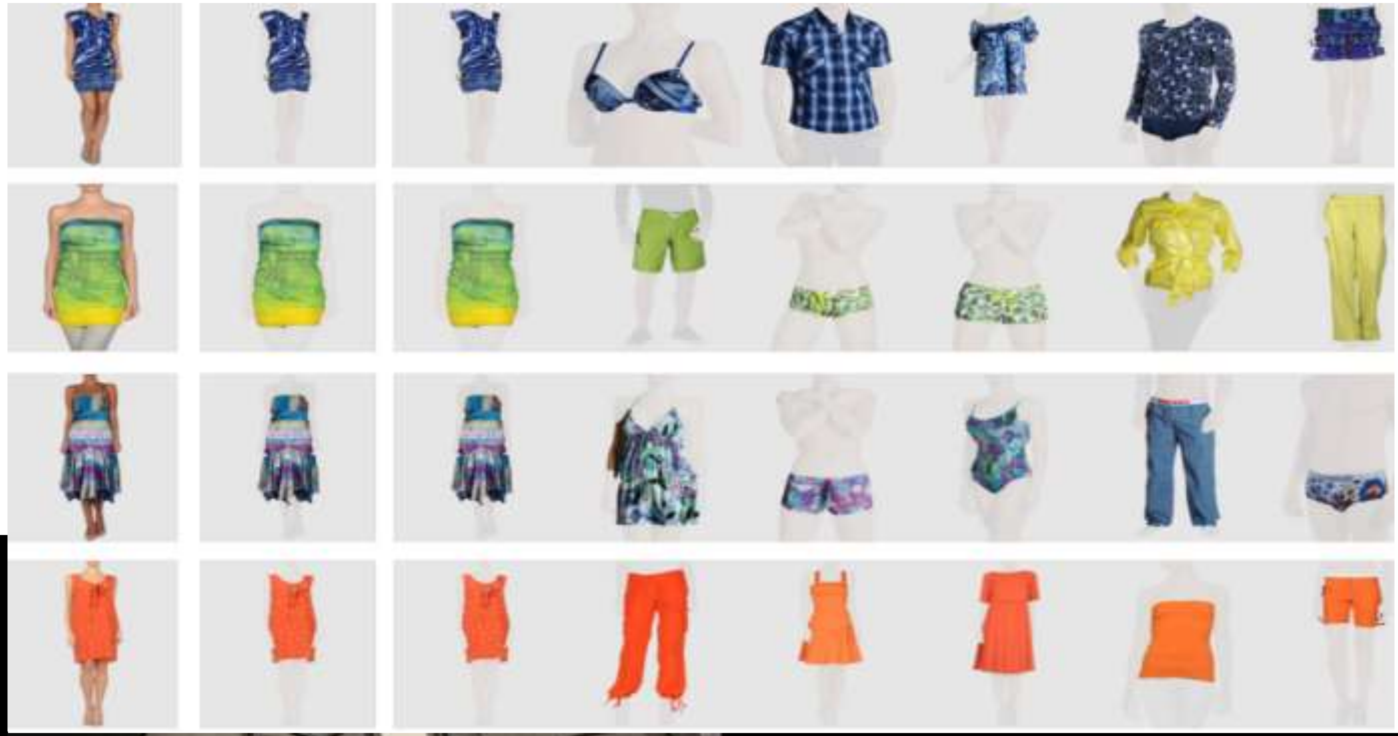
Carsten Rother* Vladimir Kolmogorov† Andrew Blake‡
Microsoft Research Cambridge, UK

Figure 1: **Three examples of GrabCut.** The user drags a rectangle loosely around an object. The object is then extracted automatically.



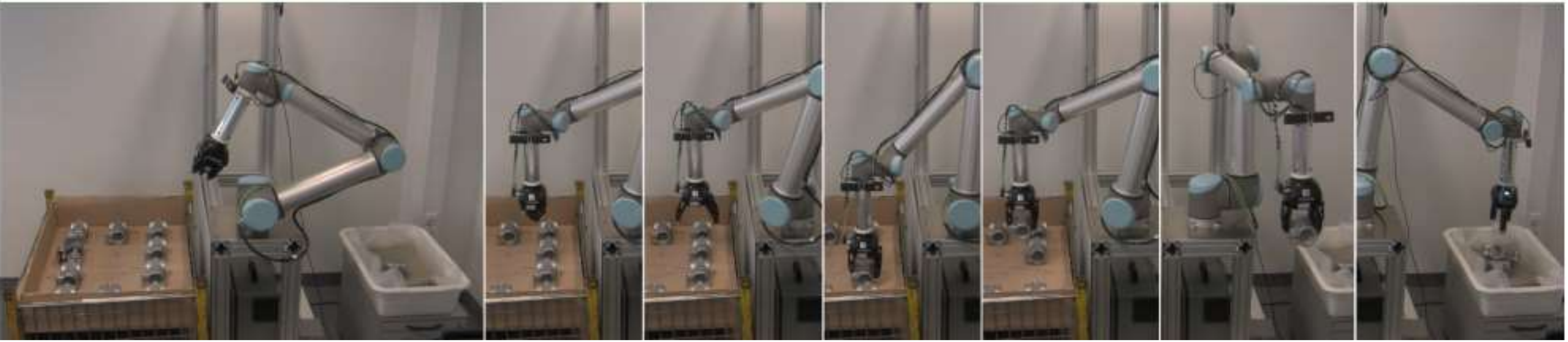
Precise object segmentation @Imagelab

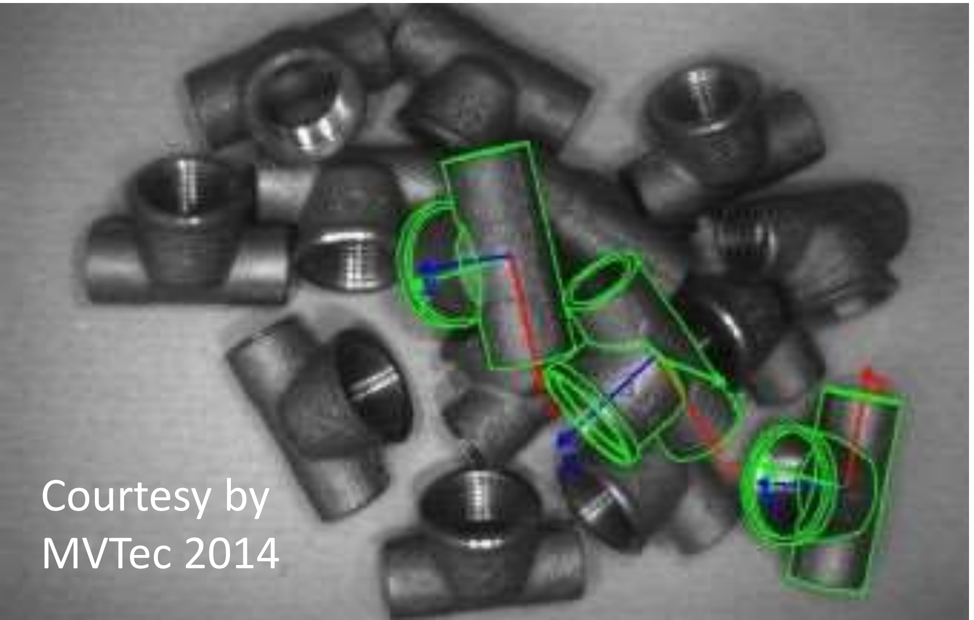
Target segmentation in images (Yoox-UNIMORE 2014)



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

➢ Random object disposal

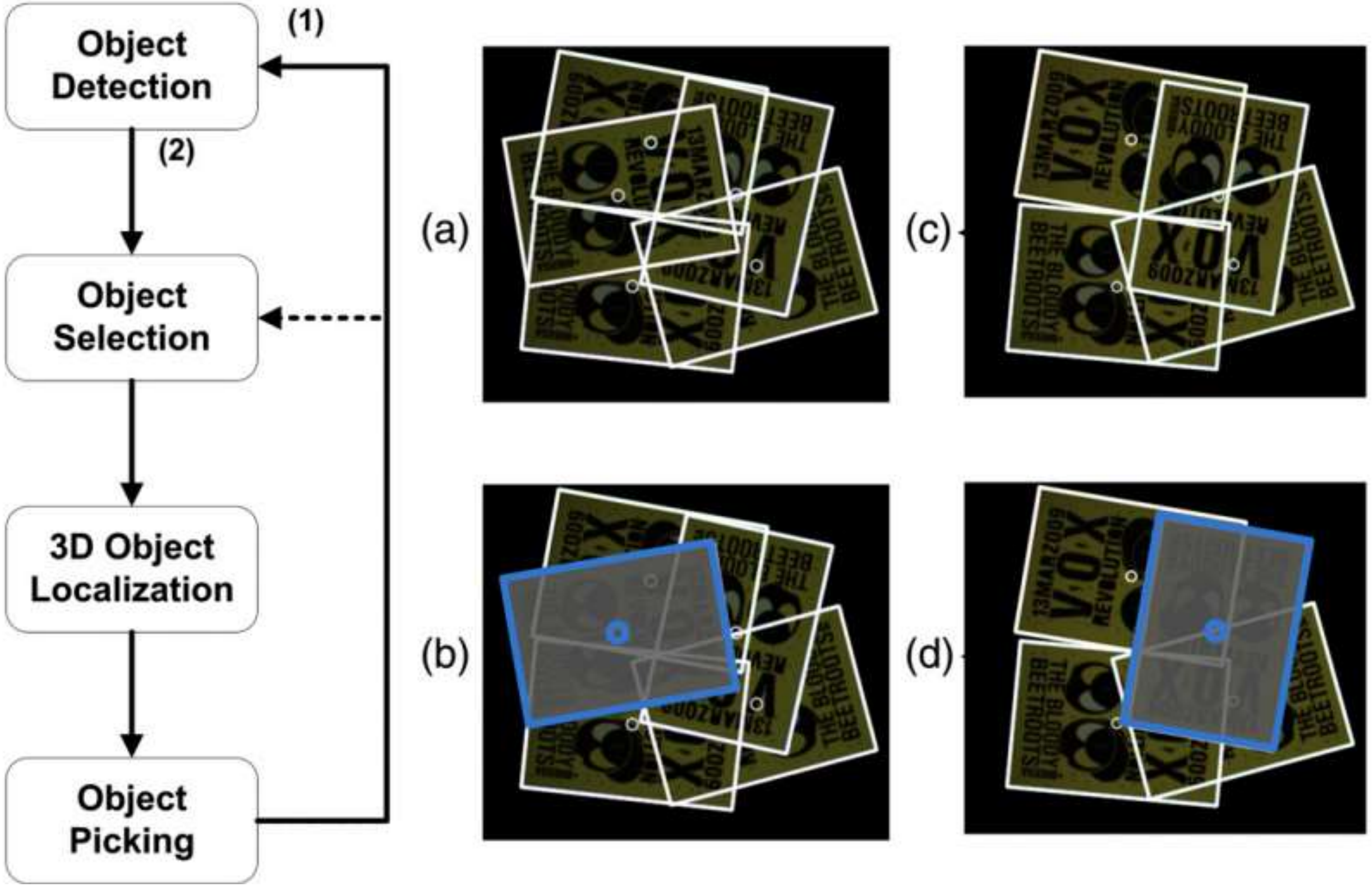
➢ Multiple instances and distractors

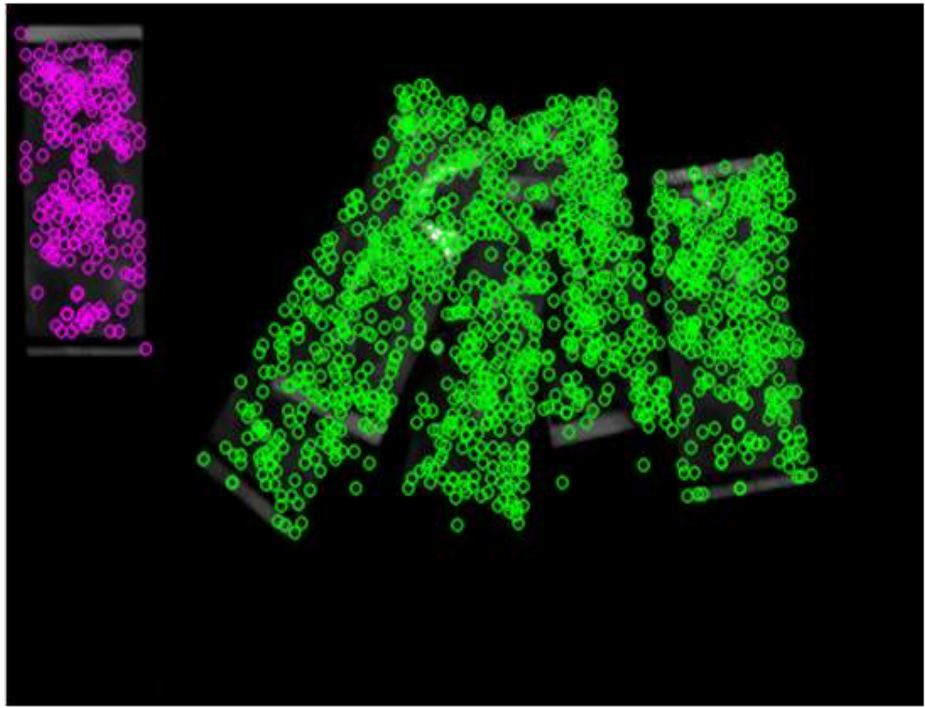
➢ Heavely occluded objects

➢ High working speed (100obj/min)

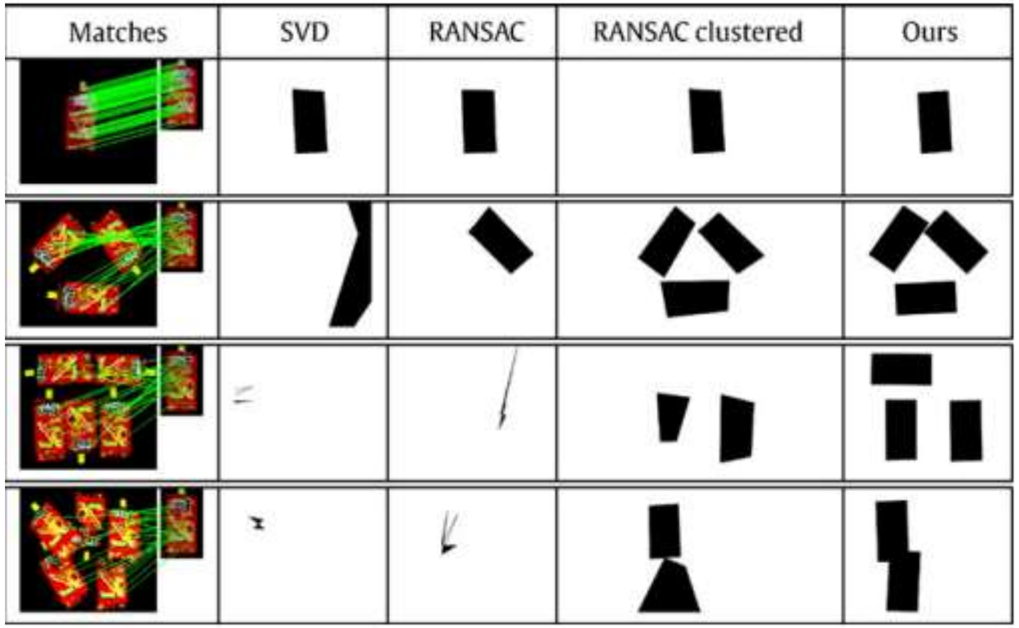
P.Piccinini, A. Prati, R.Cucchiara Real-time object detection
and localization with SIFT-based clustering
2012 Image and Video Computing 30 (2012)573-587

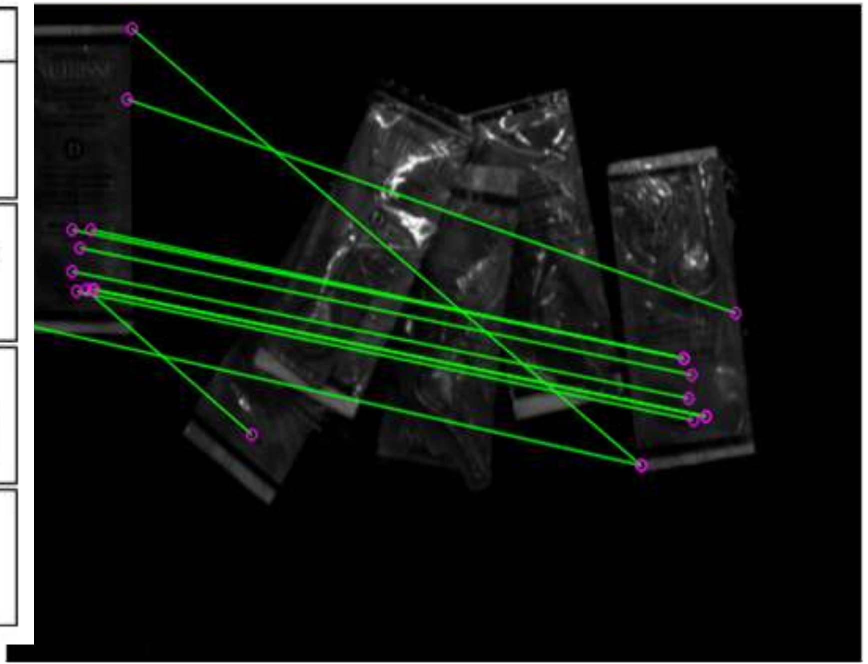
- Searching for the most visible (and easy to be picked) target
- SIFT Based (2004)



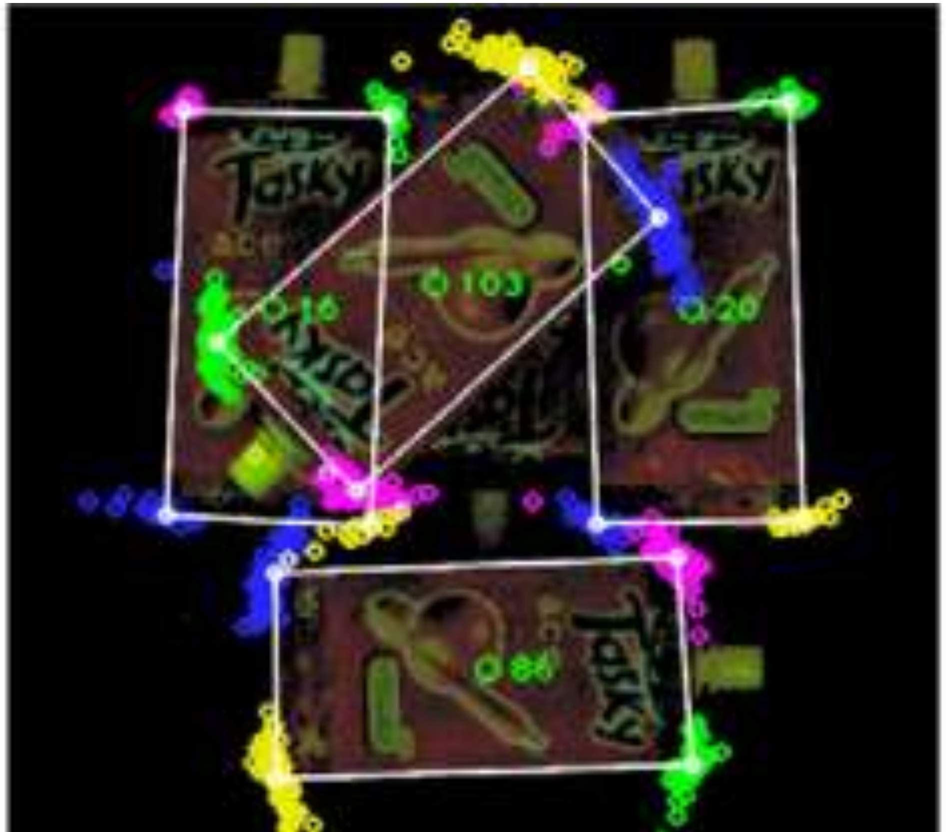


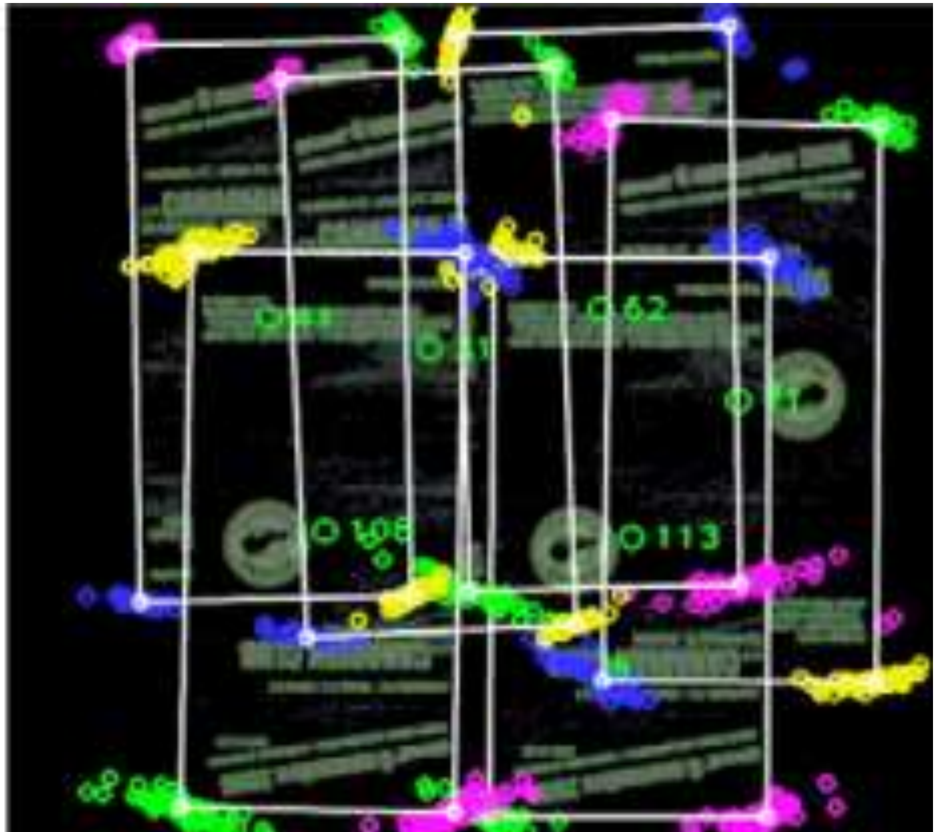
- Object localization by a registration transform
- $K = \{k_i \equiv (x_i, y_j, D_i, \vartheta_i)\}$ i=1..n features in the model M
- $KI = \{k_{Ij} \equiv (x_{Ij}, y_{Ij}, D_{Ij}, \vartheta_{Ij})\}$ j=1..r features in the image I
- Multiple matching with best-bin-first algorithm
- Registration with planar homography with 5 support points

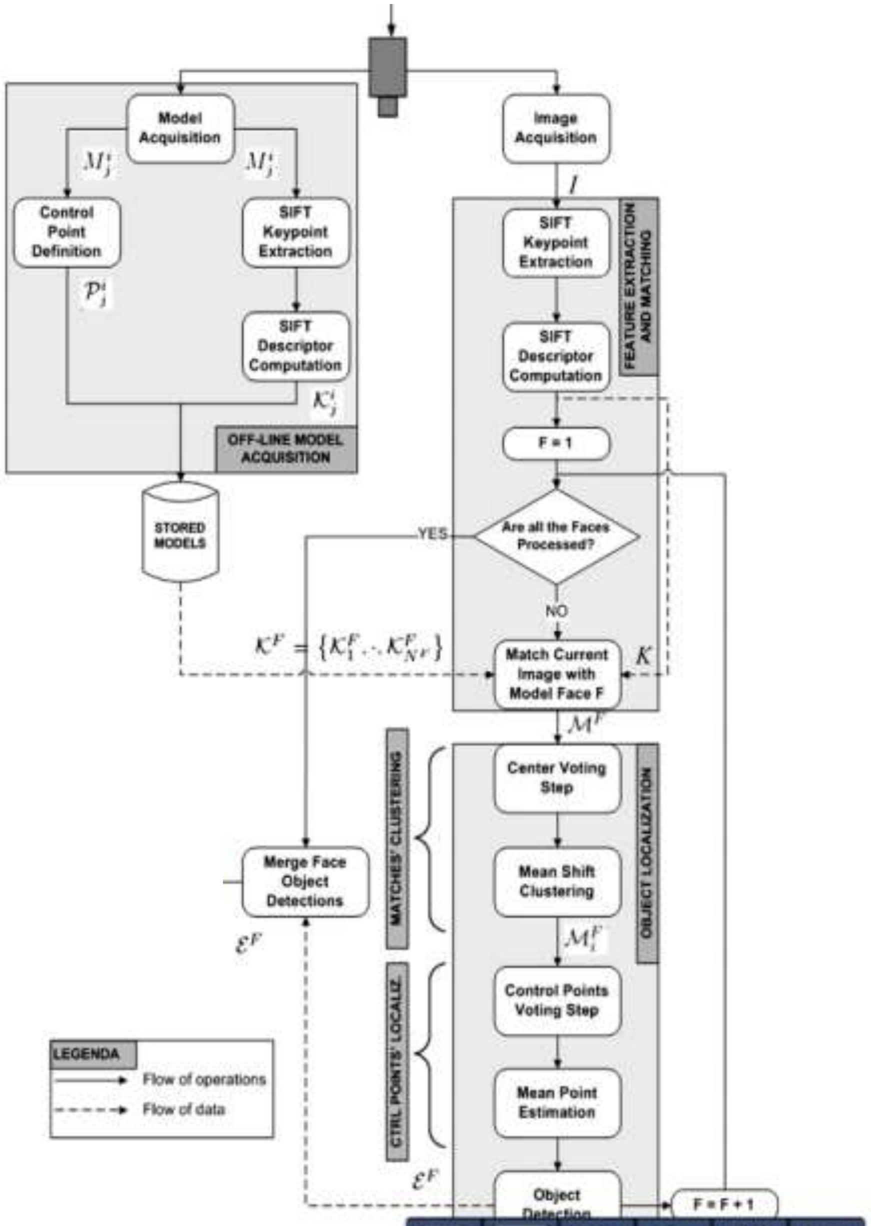


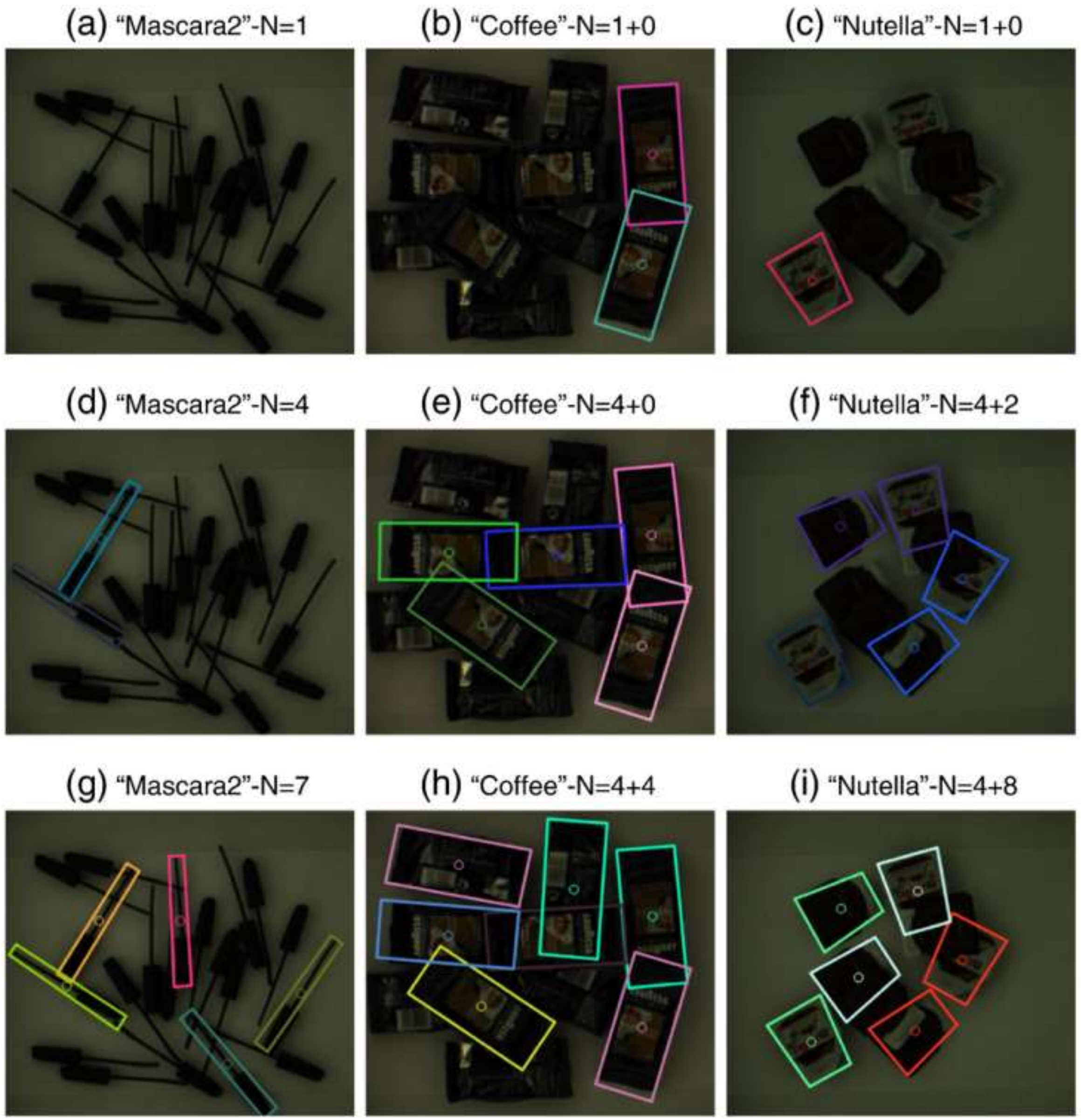


- Clustering similar features,
- Mean-shift clustering for voting to single centers:
- Estimations of center and control points by clustering



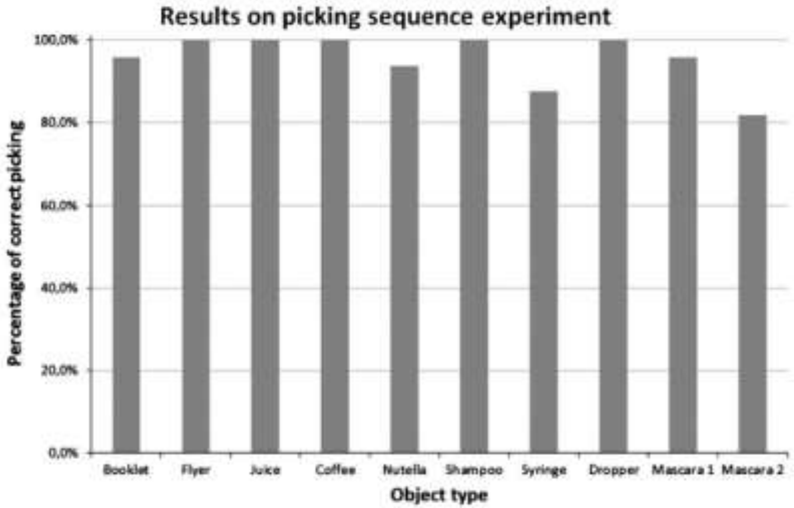


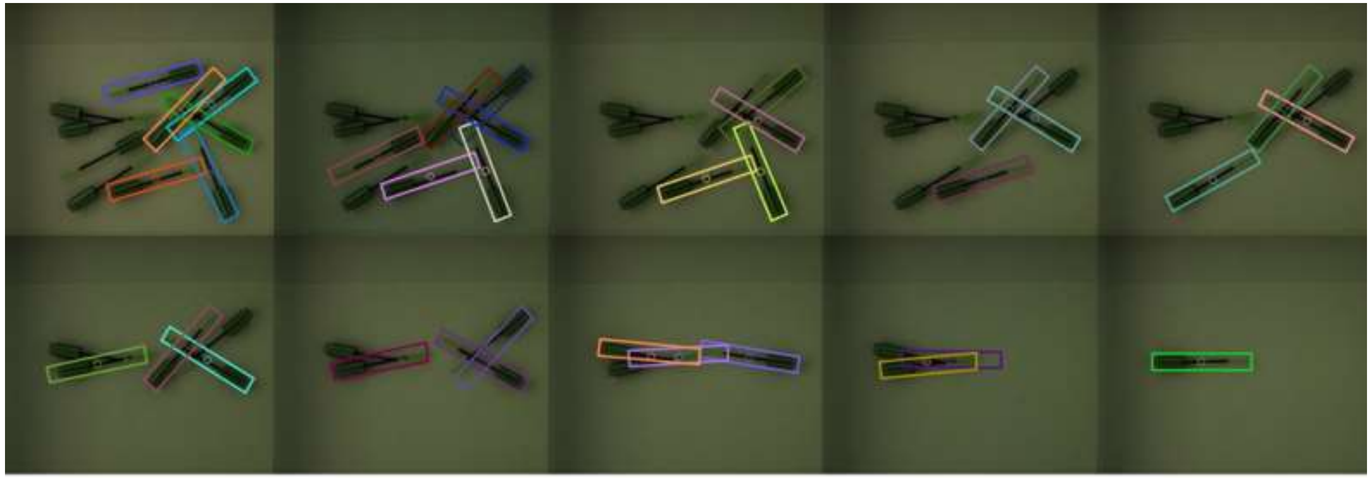

Model Acquisition
Control Point Definition
SIFT Keypoint Extraction
SIFT Descriptor Computation
OFF-LINE MODEL ACQUISITION
STORED MODELS
Image Acquisition
SIFT Keypoint Extraction
SIFT Descriptor Computation
FEATURE EXTRACTION AND MATCHING
F = 1
Are all the Faces Processed?
YES
NO
Match Current Image with Model Face F
Center Voting Step
Mean Shift Clustering
Control Points Voting Step
Mean Point Estimation
Object Detection
F = F + 1
MATCHES' CLUSTERING
CTRL POINTS' LOCALIZ.
OBJECT LOCALIZATION
Merge Face Object Detections
LEGENDA
Flow of operations
Flow of data


(a) "Mascara2"-N=1
(b) "Coffee"-N=1+0
(c) "Nutella"-N=1+0
(d) "Mascara2"-N=4
(e) "Coffee"-N=4+0
(f) "Nutella"-N=4+2
(g) "Mascara2"-N=7
(h) "Coffee"-N=4+4
(i) "Nutella"-N=4+8

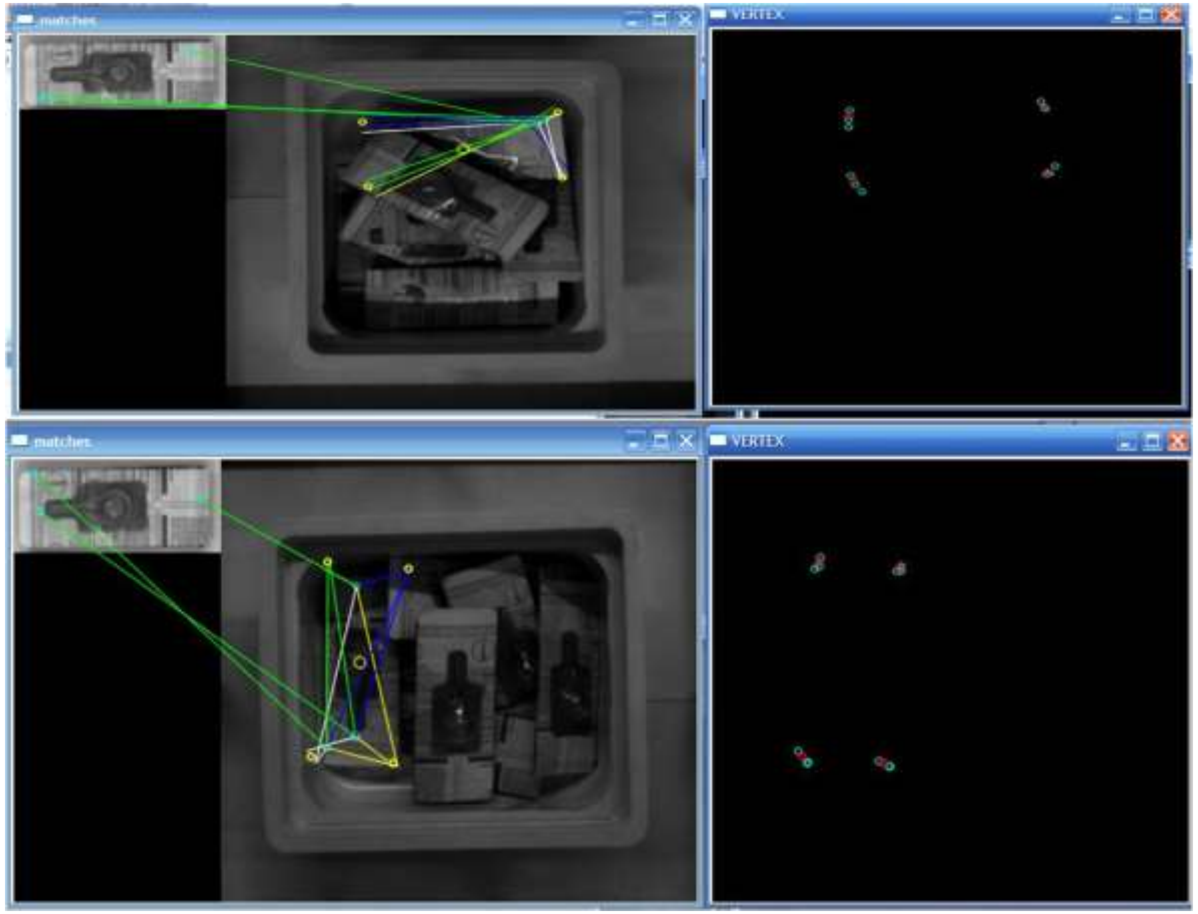
Experimental results for single model approach.

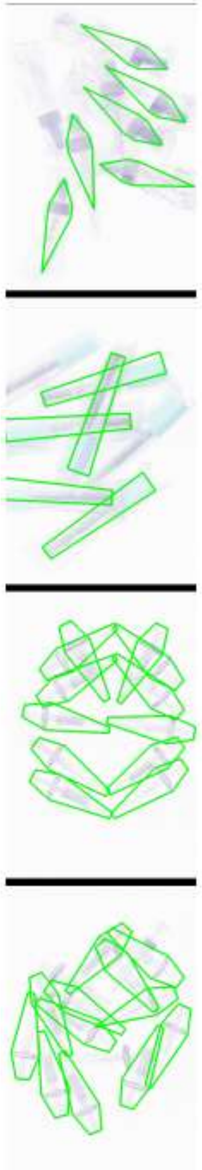
	Object-level		Pixel-level		Center dist.
	Precision	Recall	Precision	Recall	Mean (px)
Juice					
All RS	100.00%	25.00%	22.95%	23.66%	5.41
Clus RS	91.67%	82.50%	77.43%	79.93%	18.97
Ours	97.37%	92.50%	88.55%	87.64%	5.76
Nutella					
All RS	100.00%	15.38%	13.23%	14.46%	6.98
Clus RS	66.67%	33.85%	38.96%	35.82%	17.24
Ours	97.84%	86.78%	82.87%	83.13%	3.86
Flyer					
All RS	90.00%	16.36%	17.46%	15.72%	14.68
Clus RS	74.00%	64.91%	71.31%	68.46%	22.27
Ours	96.15%	90.91%	86.35%	89.39%	2.66

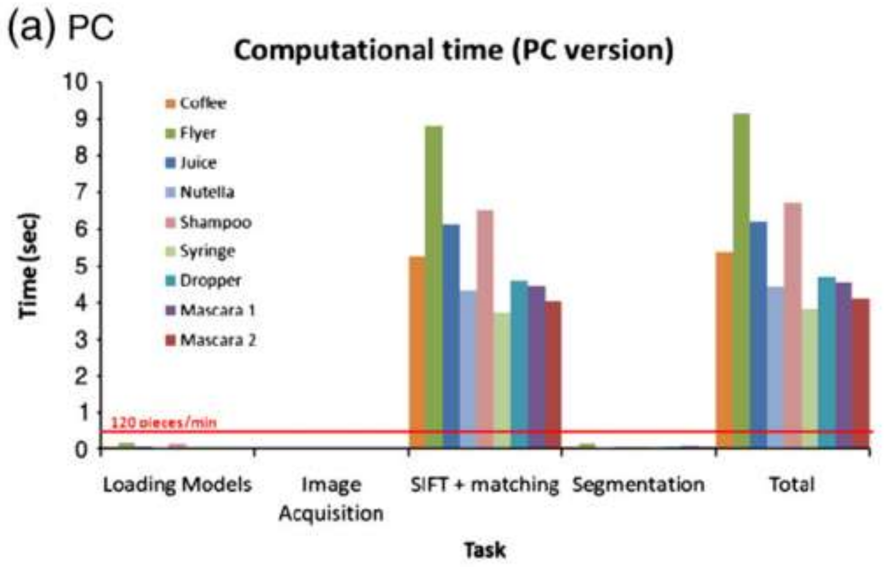


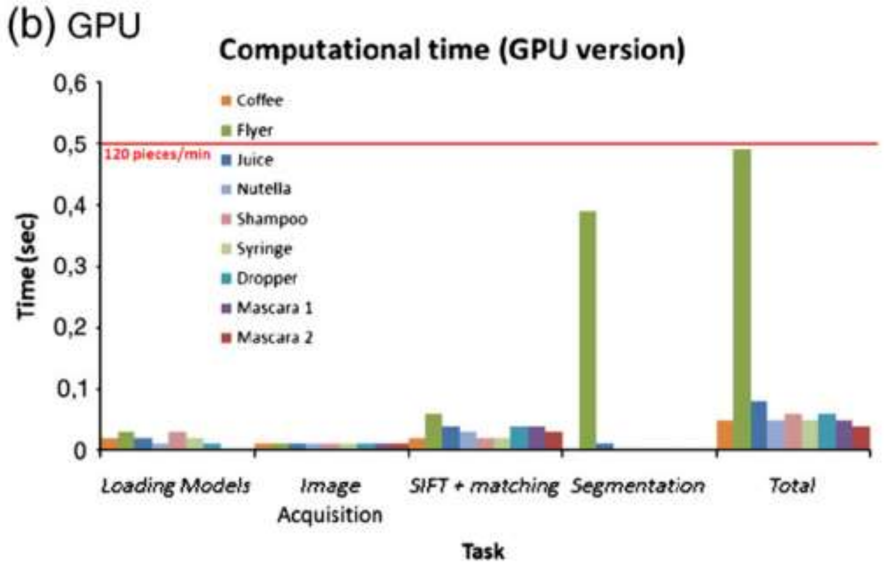


- Real time processing
- only with GPU



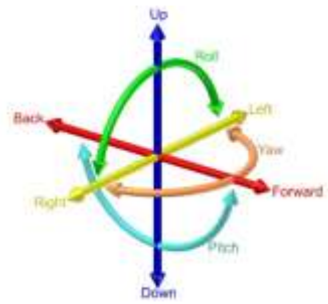


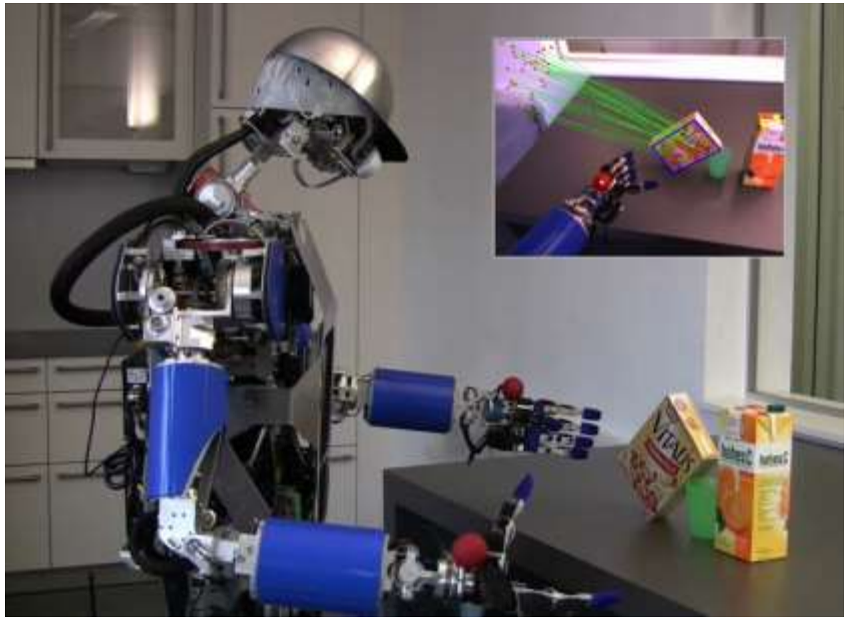




Unconstrained picking

- 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

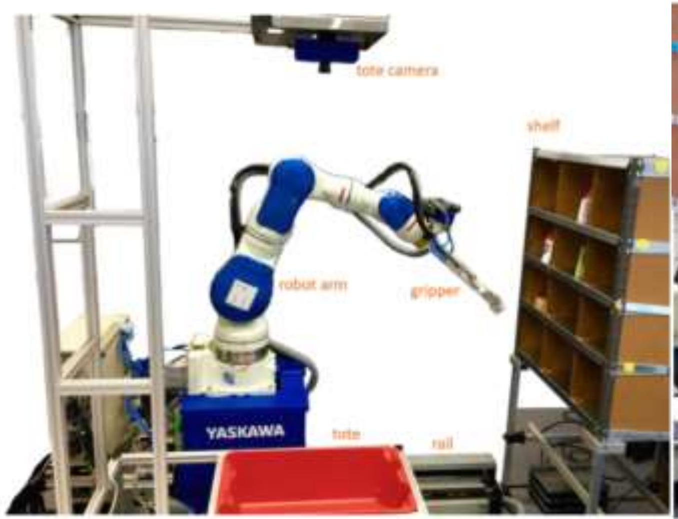




KIT Robot

Amazon Picking challenge 2016

- Winner: Delft Univ, NL29 target objects



Delft Robot

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

MIT Robot

- **Iterative Hough Forest with Histogram of Control Points for 6 DoF Object Registration from Depth Images**, (Imperial college ISOR2016)

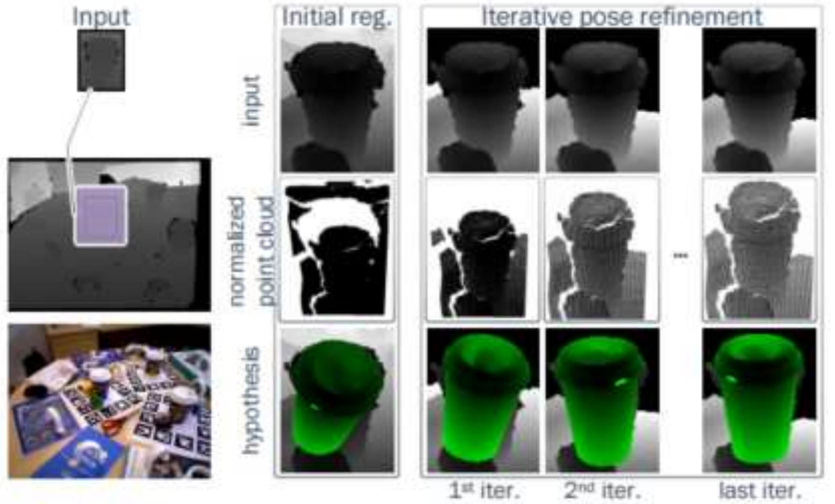
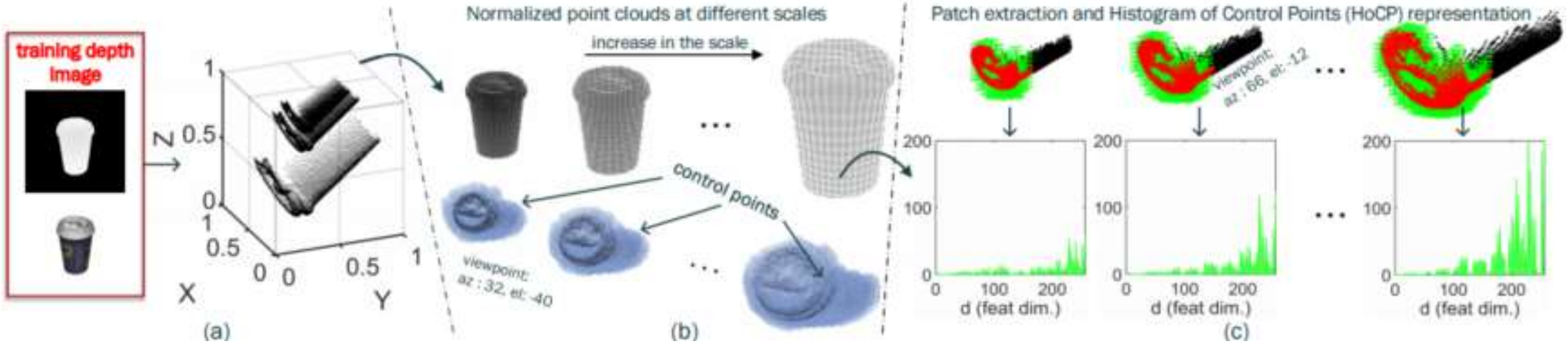


Fig. 1: Sample result of our architecture: *initial registration* roughly aligns the test object and *iterative pose refinement* further refines this alignment (The RGB image is for better visualization).



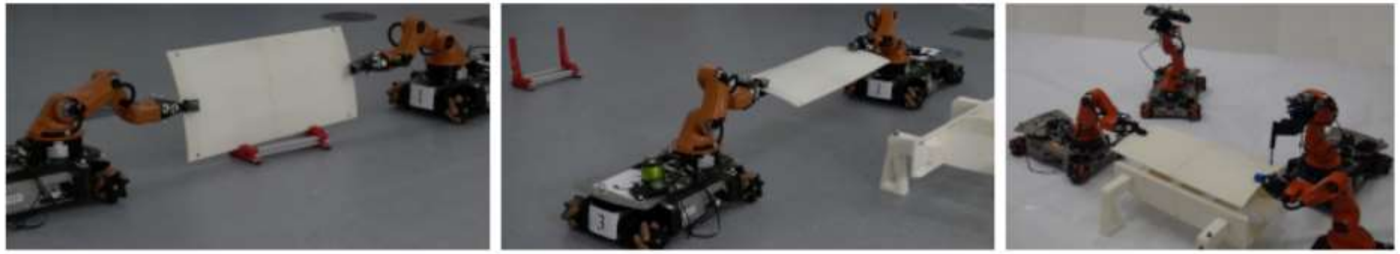
Robot coordination

- (MIT and Georgia Tech 2014)

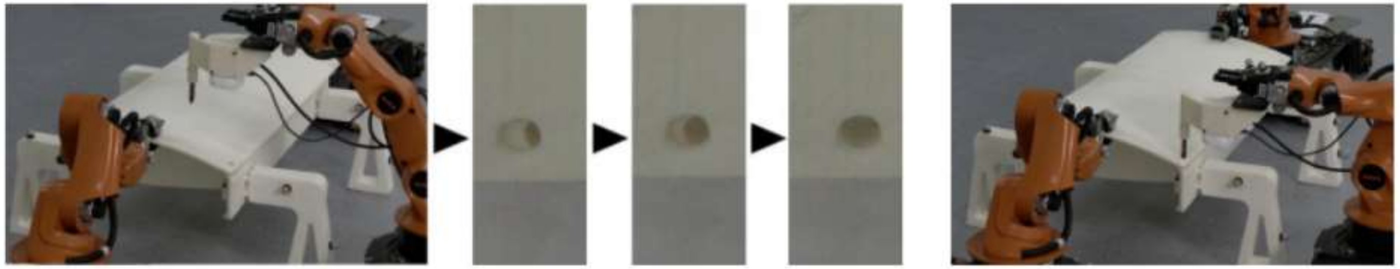
Robot			
R1	R2	R3	R4
			Move to hole 1 neighborhood
Navigate to and move gripper to panel		Localize box	Find hole 1 in box
Close grippers and form fleet			Find hole 1 in box
Pick up panel			
Orient panel to horizontal			
Transport panel into neighborhood of box			
Servo panel into alignment with ladder		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



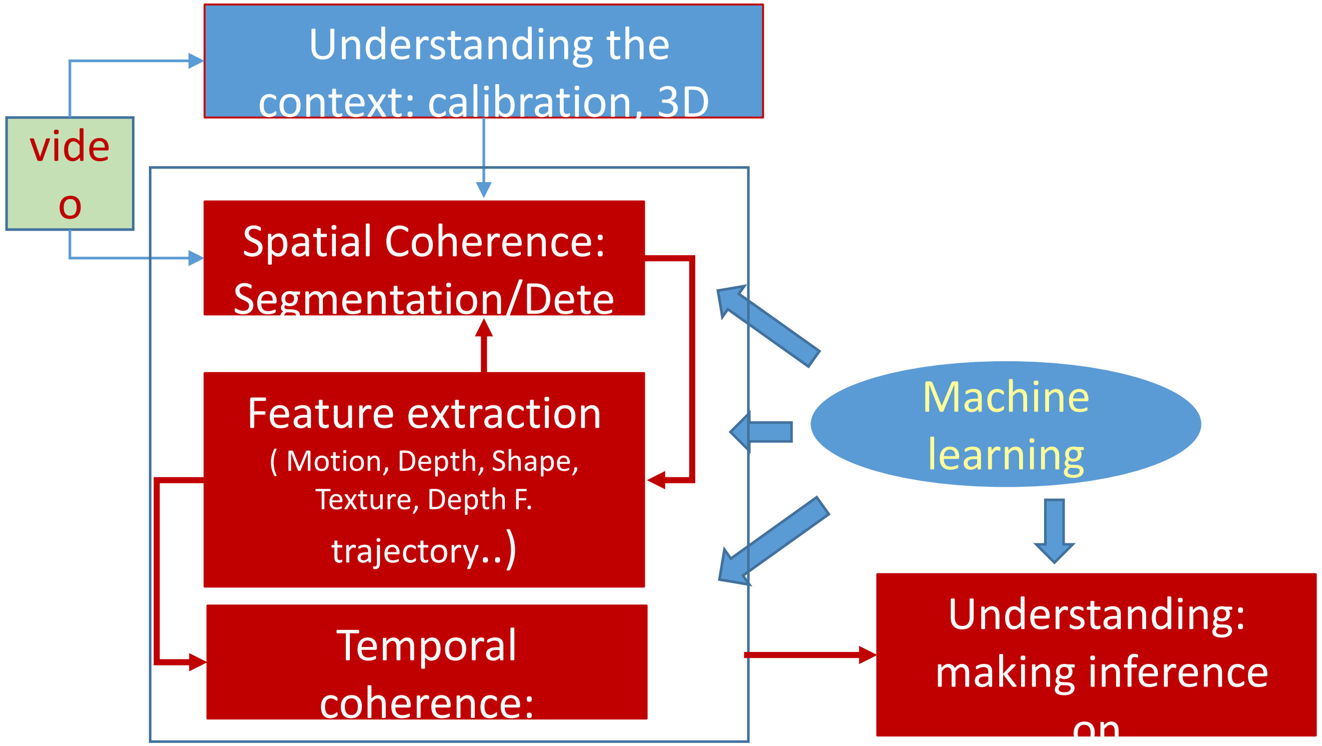
(d) Hole alignment (e) Fastener insertion

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

- 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

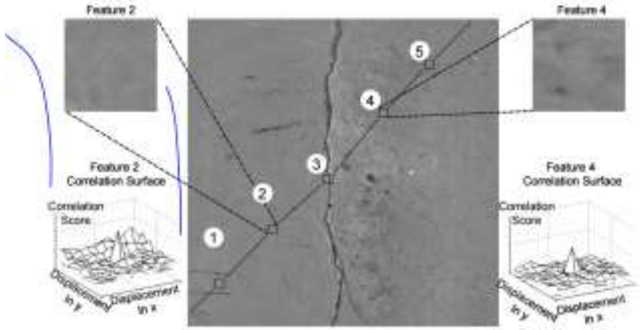


Positioning and Guidance

AGV Automatic Guided Vehicles in Industries

Autonomous guidance is more complex:

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



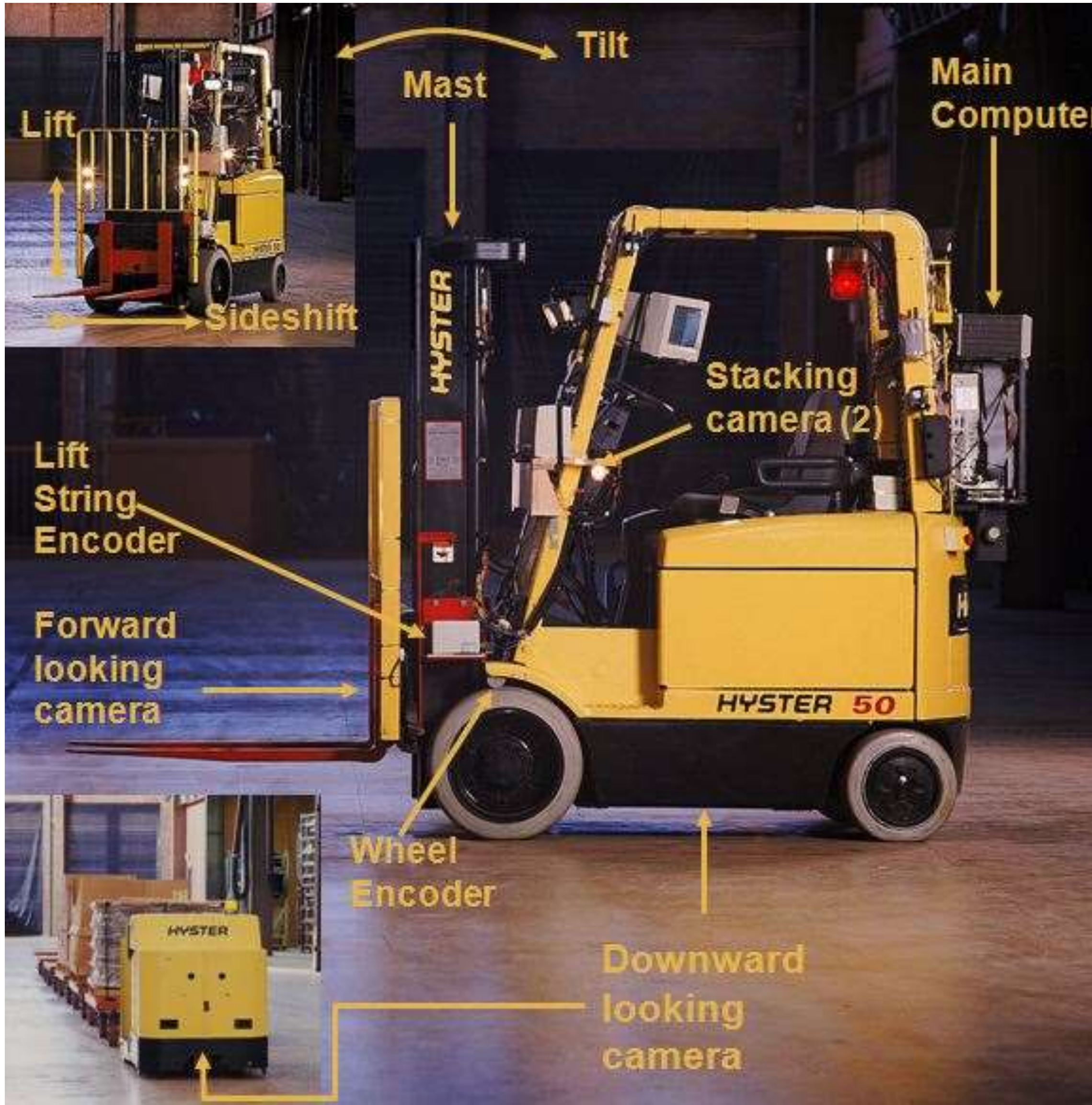
Now:

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

Markers in the environment

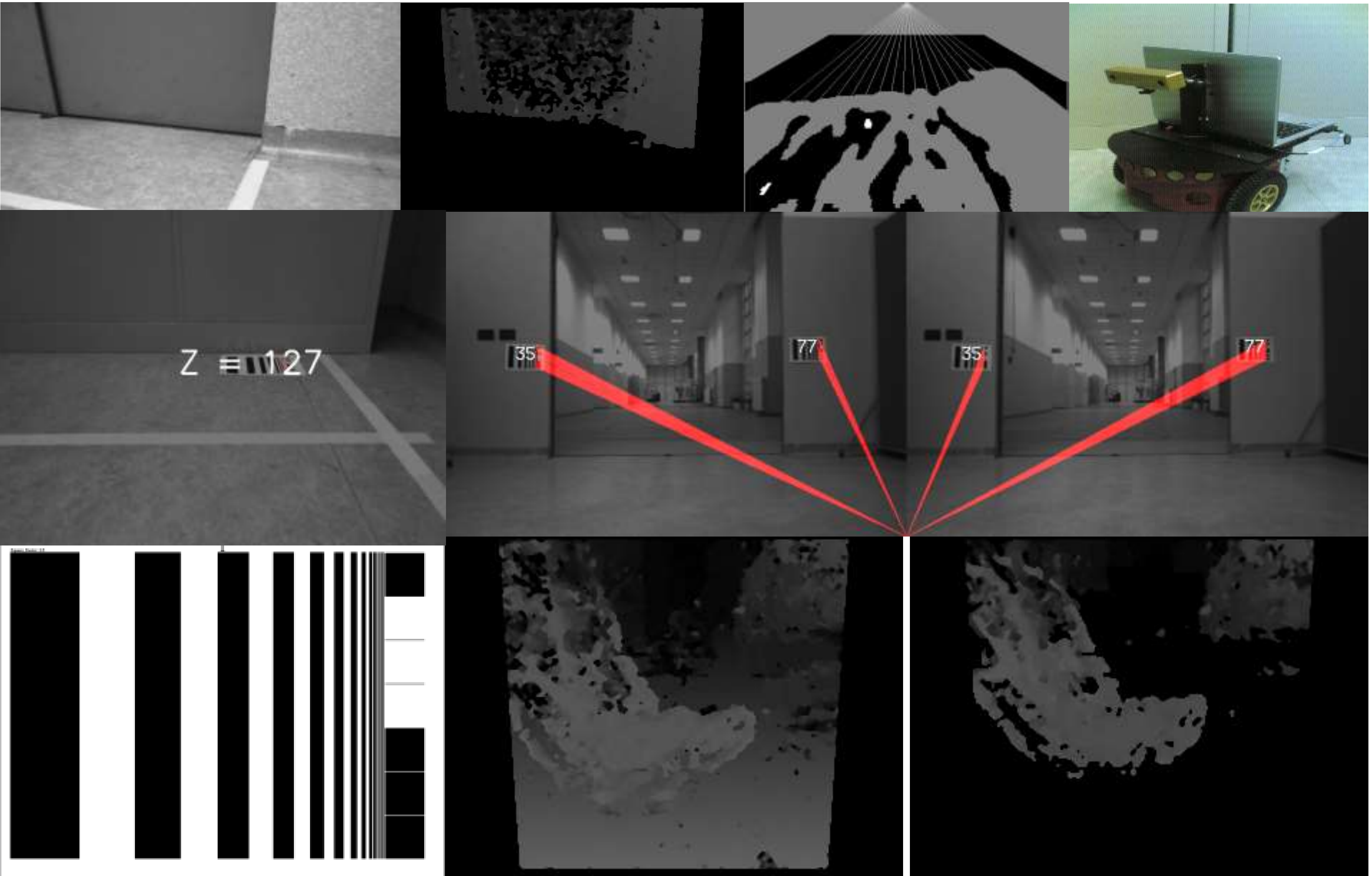
Big Data collection

A large impact of computer vision and machine learning



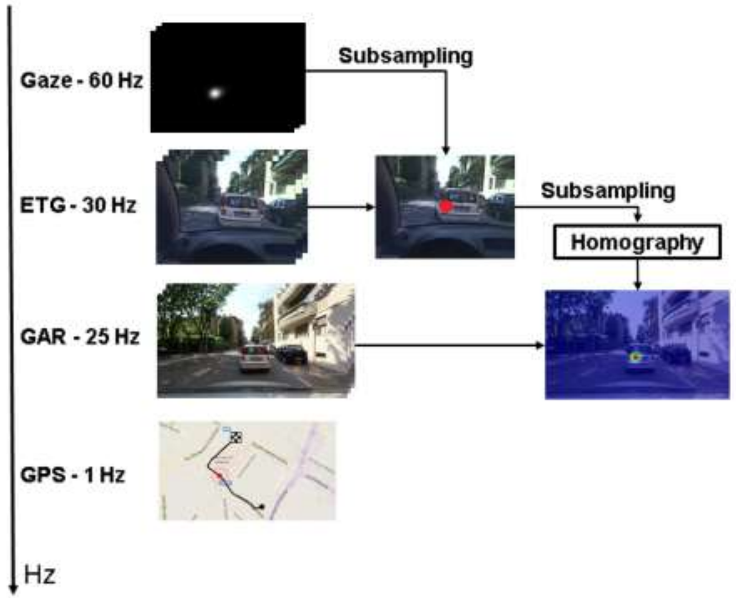
*Kelly, A.,Nagy, B. ,Stager, D., Unnikrishnan, R., “An Infrastructure-Free Automated Guided Vehicle Based on Computer Vision”, IEEE Robotics and Automation Magazine. 2007.

Stereo Vision and many sensors



2008 Imagelab
Italvision & UNIMORE
(PRRIITT)

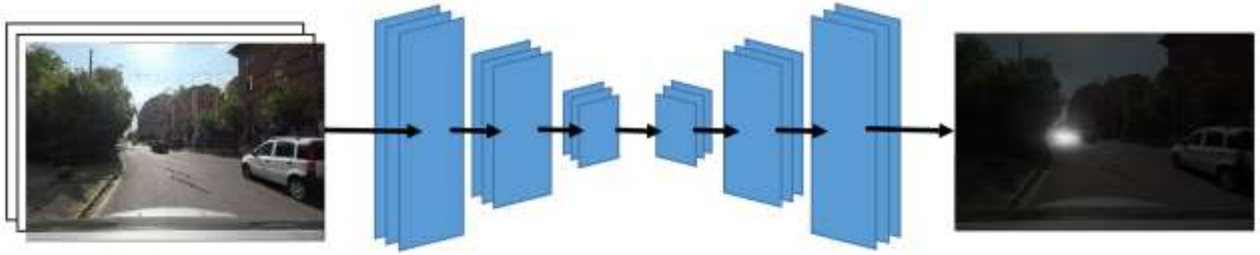
- Image registration and synchronization



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 project HBU

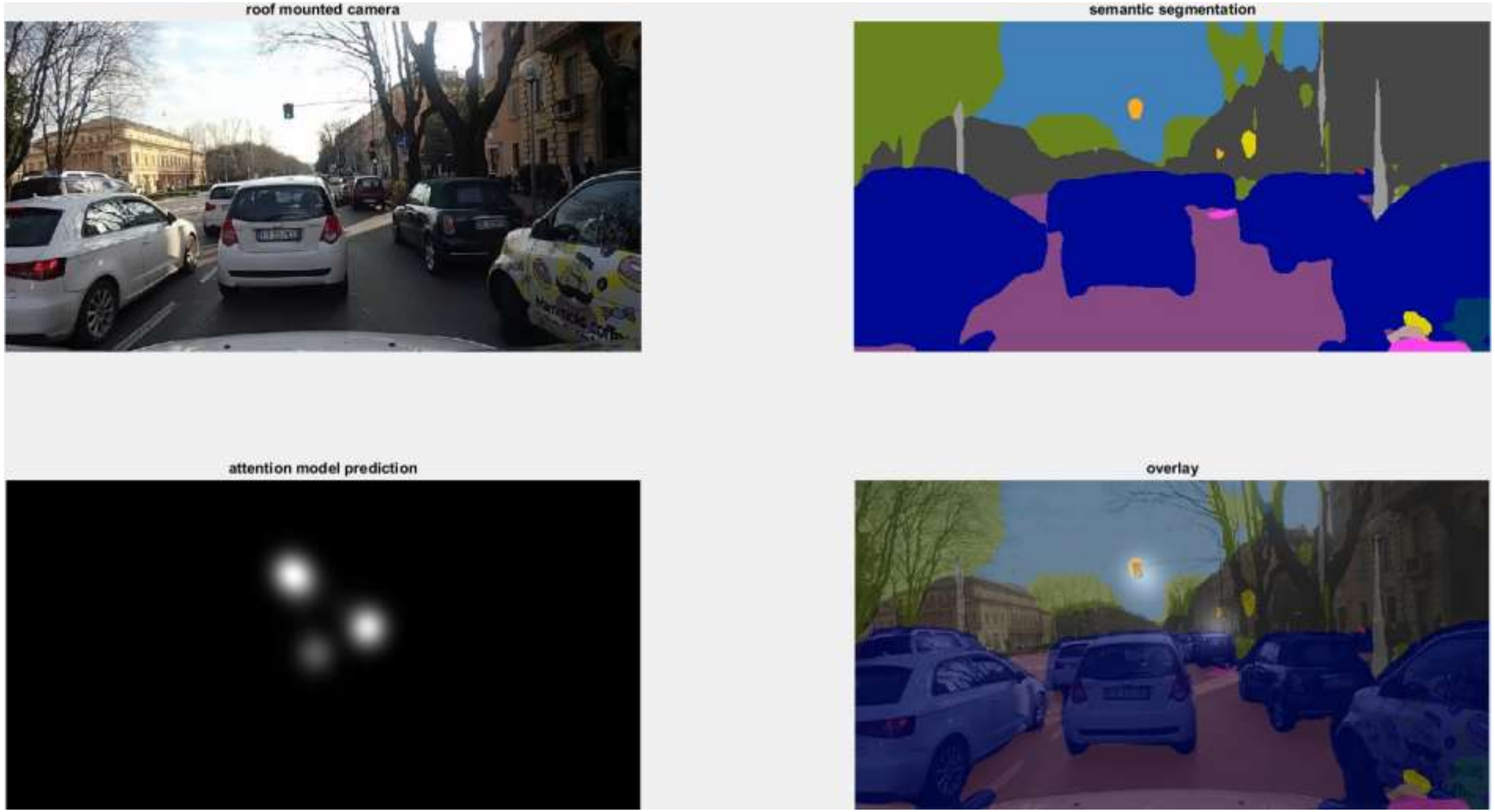


Good driving habits model: **where should we attend?**

Semantic segmentation: **what are we actually looking at?**

Look for us on http://imagelab.ing.unimore.it/dreyeve

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



Person detection
Multiple-person tracking
Face and hand detection for interaction
Gesture Recognition
Espression Recognition
Pose Estimation
..

See @Imagelab

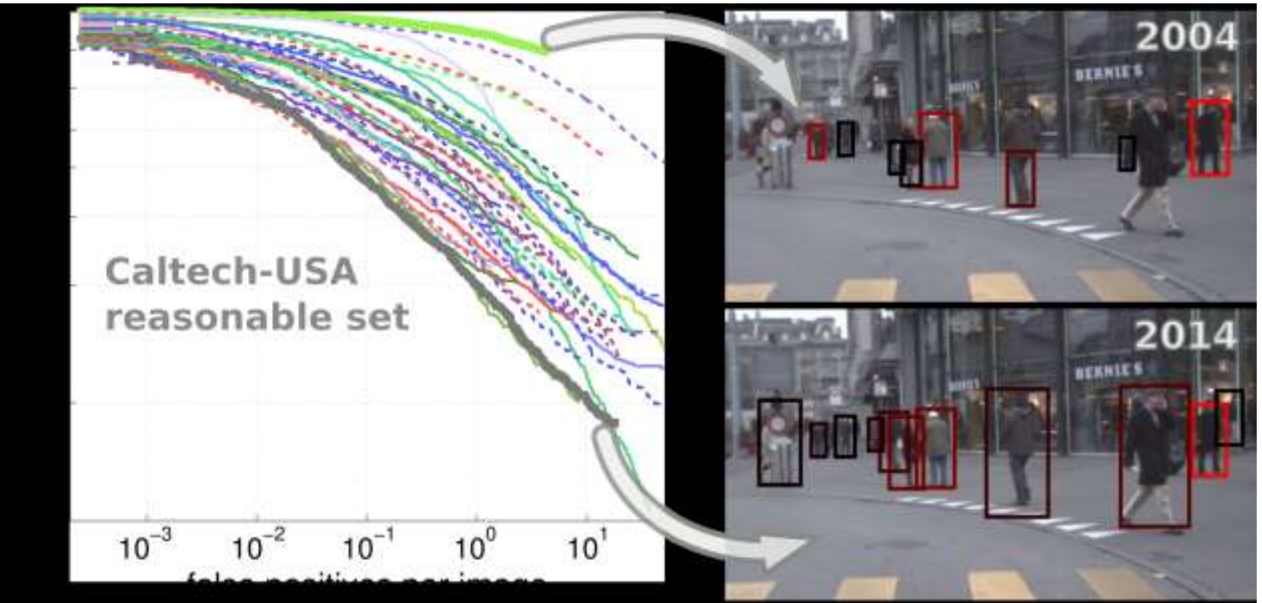
The Human Factor

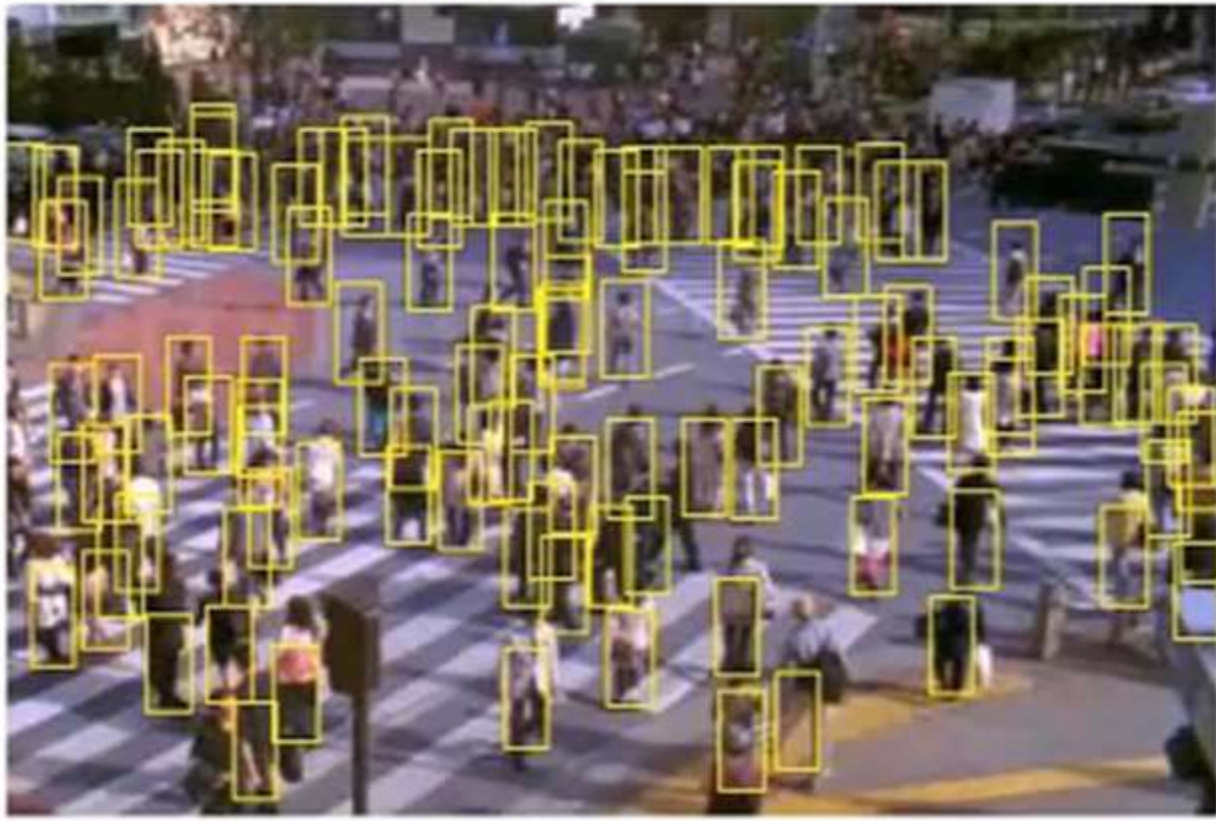
People detection...

- **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)

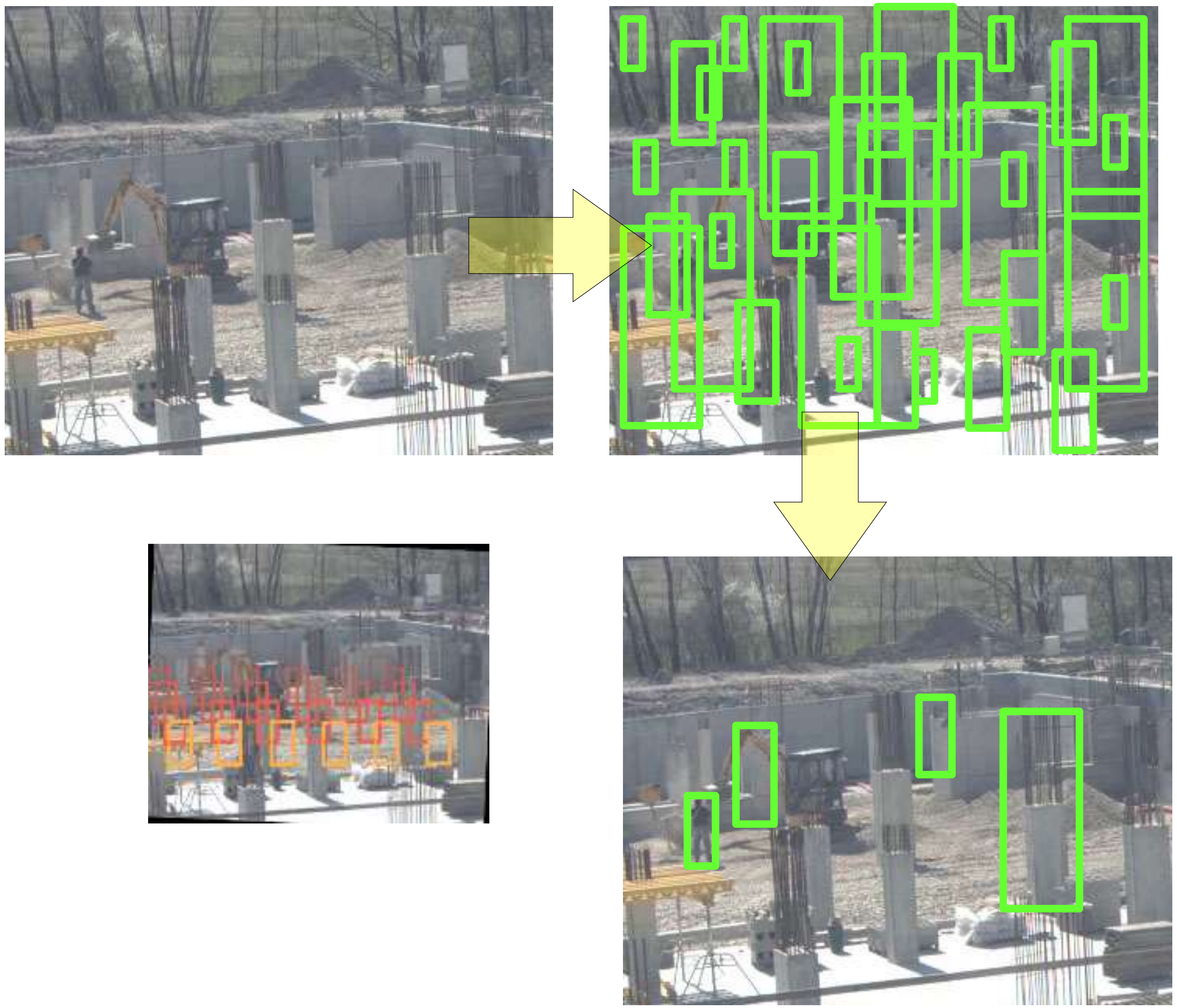
(Thanks to Shiele ECCV 2004)

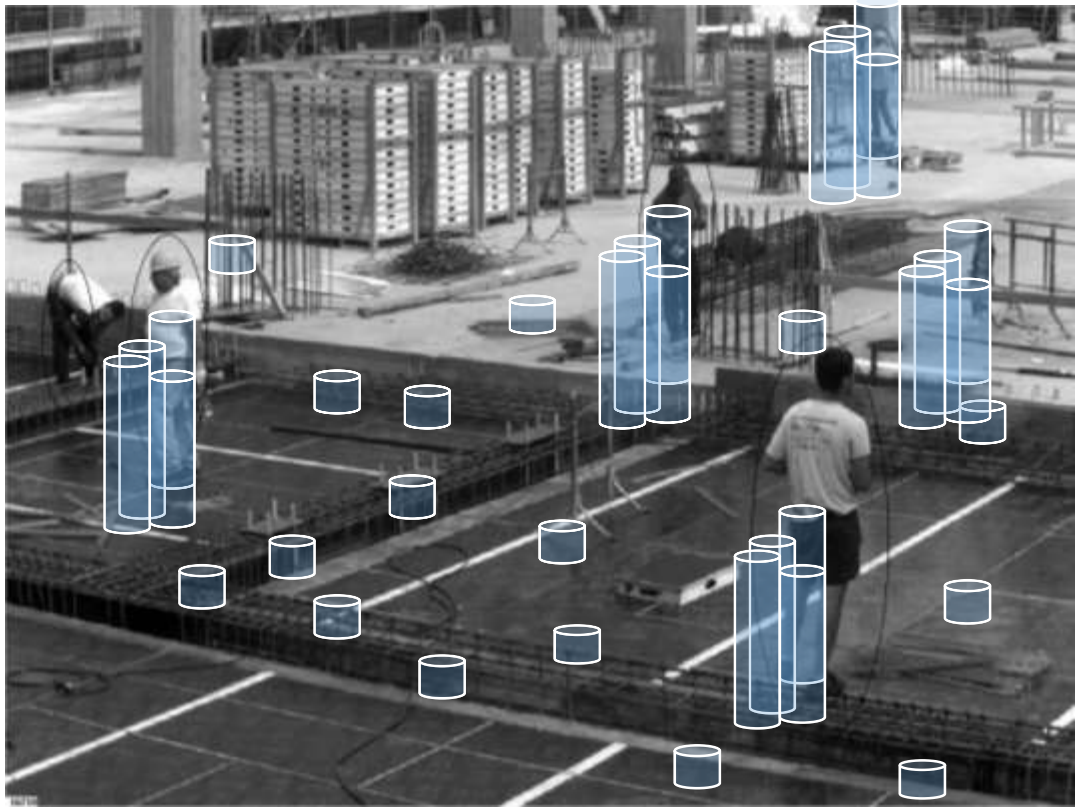


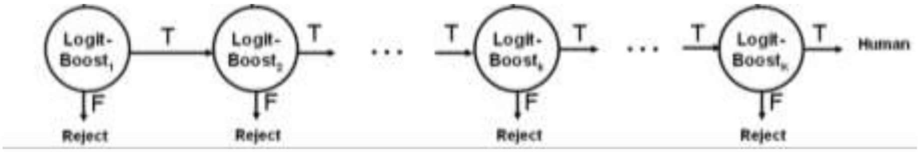


Speed and accuracy in special environments

- External construction sites



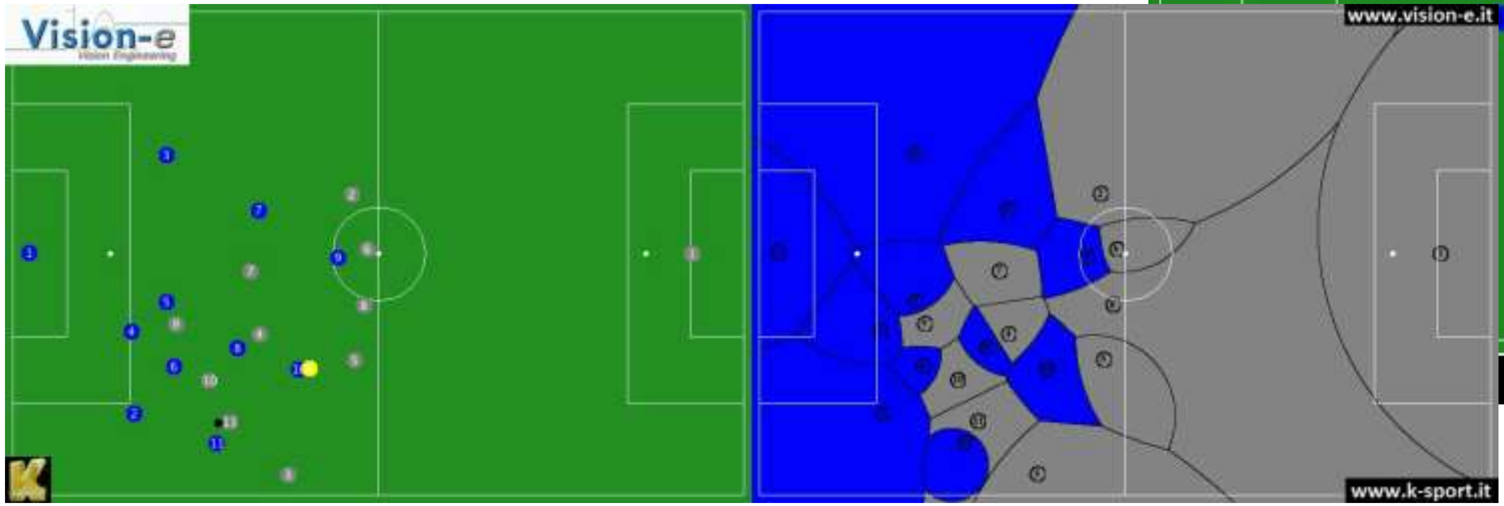




G.Gualdi, A.Prati, R.Cucchiara Multi-Stage Particle Windows for Fast and Accurate Object Detection
IEEE Transactions on PAMI Aug. 2012

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

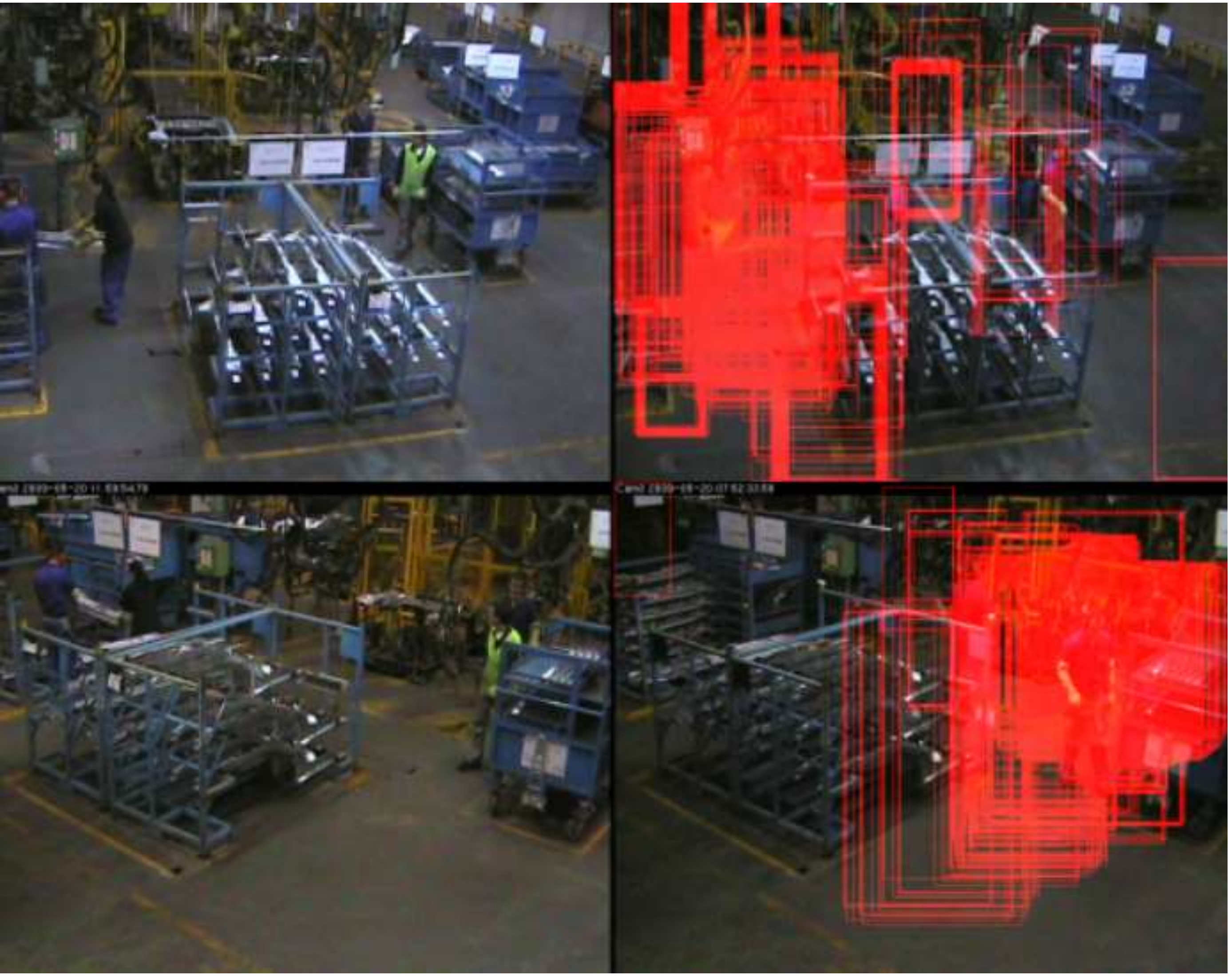
In semi-constrained world
Tracking is possible

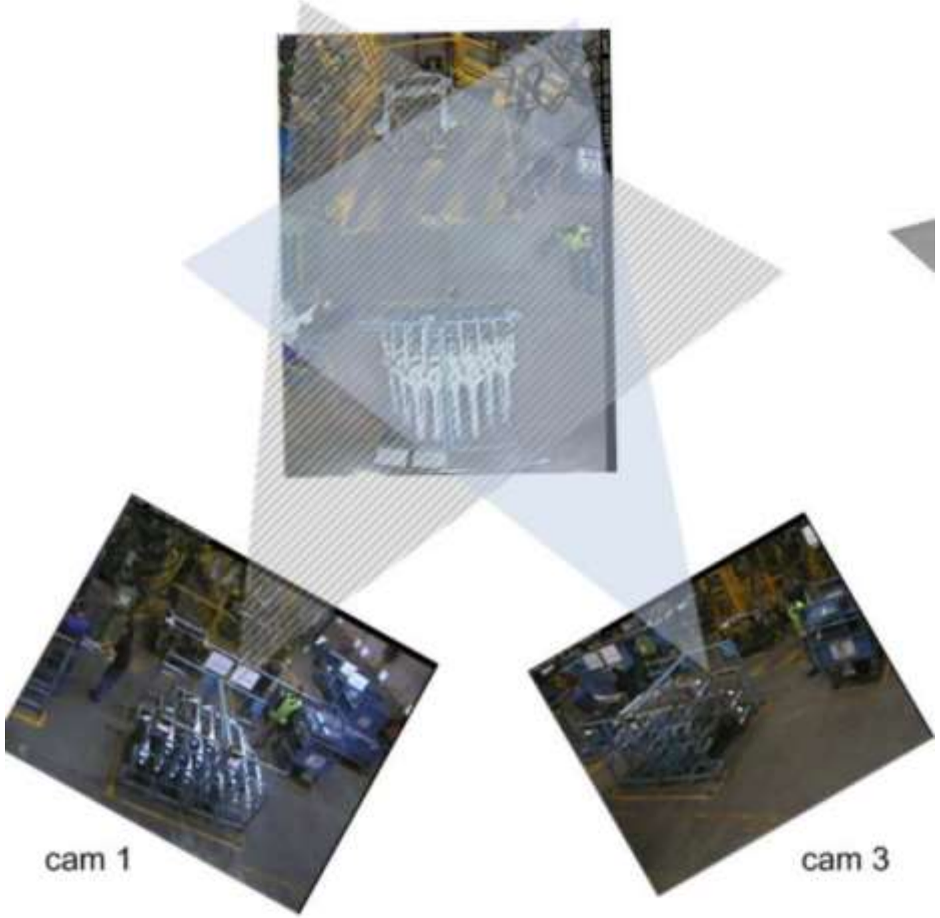


In unconstrained working area

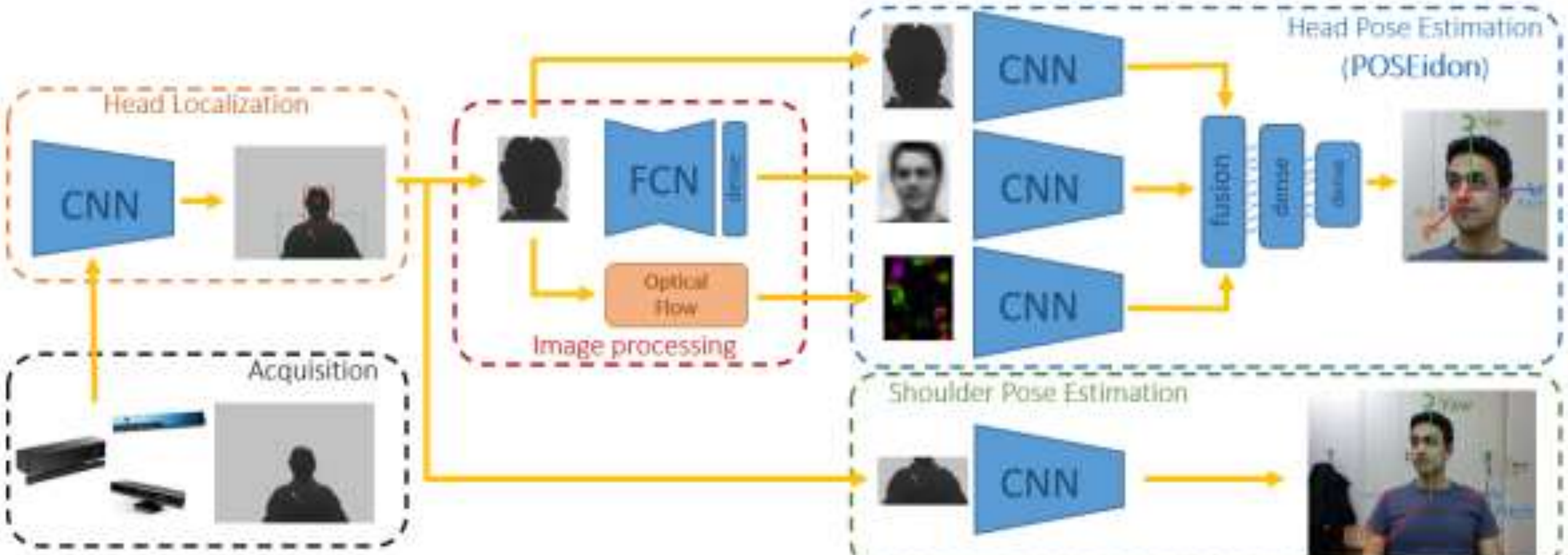
- Multiple cameras and geometric constraints

Thanks to EBVH)



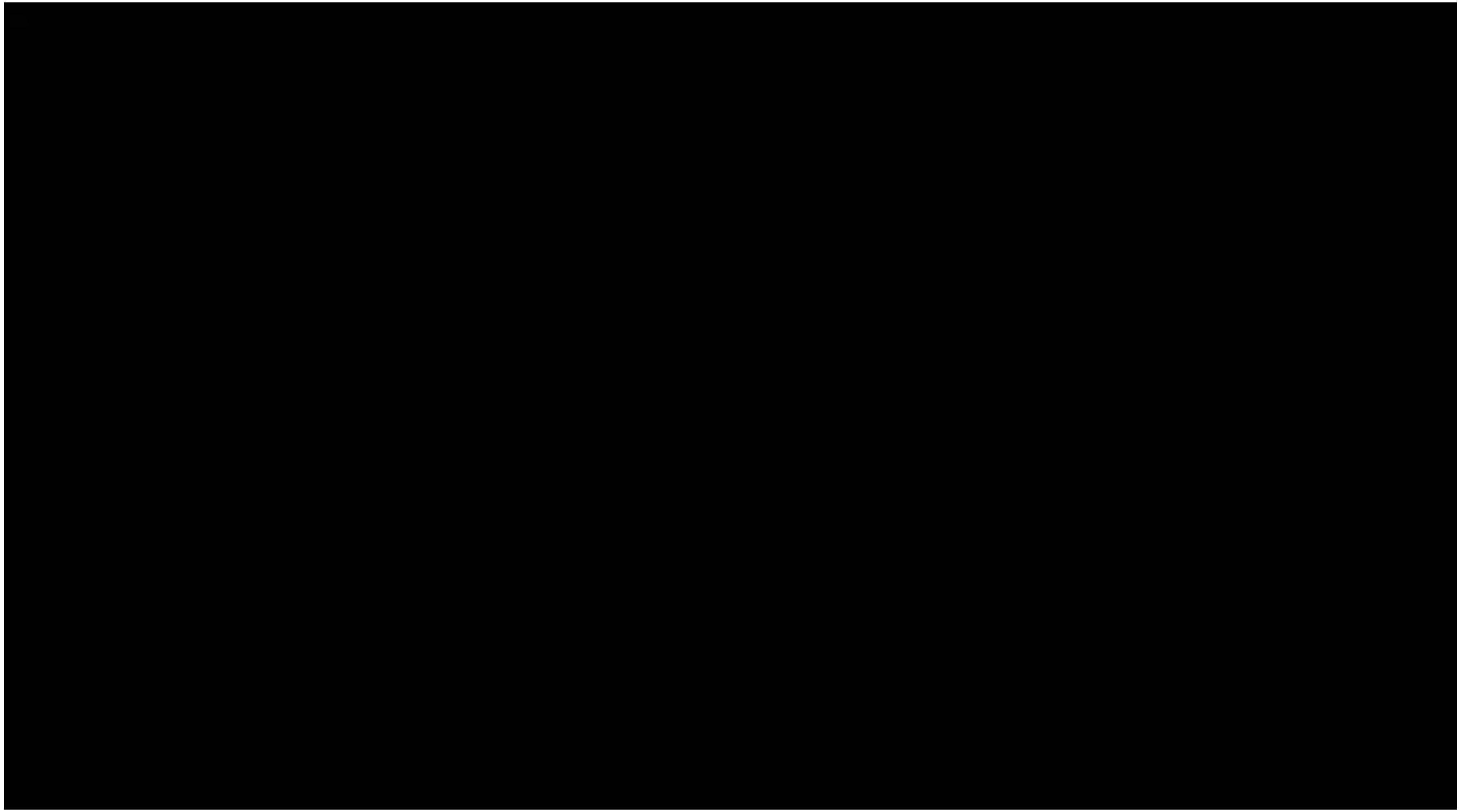


HBU IN THE CAR by depth images



Venturelli, Marco; Borghi, Guido; Vezzani, Roberto; Cucchiara, Rita ["Deep Head Pose Estimation from Depth Data for In-car Automotive Applications"]() *Proceedings We m@ICPR 2016*

- Experiments on depth-only based pose detection



Human machine interaction with egocentric vision

- Real time object recognition for augmenting vision

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

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

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

New Master   UNIMORE 2017
Opening Soon!

MuMeT 2017
visual computing and multimedia technologies