Fast Learning and Detection of Edge Shapes

Walterio Mayol-Cuevas

Computer Science Department, University of Bristol

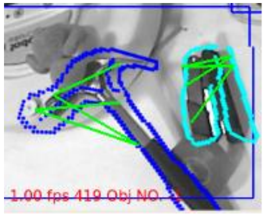
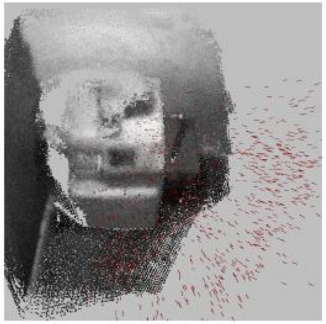
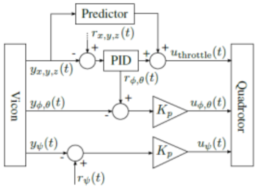
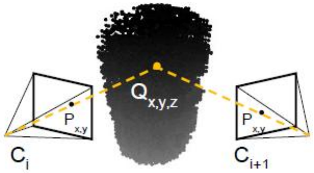
The 36th Pattern Recognition and Computer Vision Colloquium

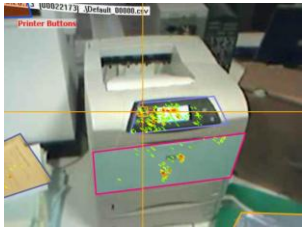
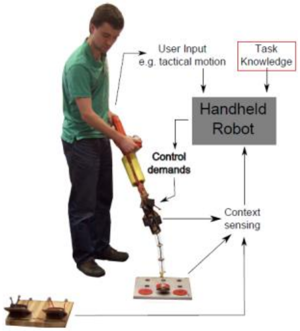
9th April 2015

At Bristol






1.00 fps 419 Obj NO.


Predictor
PID
Vicon
Quadrotor
$r_{x,y,z}(t)$
$y_{x,y,z}(t)$
$u_{throttle}(t)$
$r_{\phi,\theta}(t)$
$y_{\phi,\theta}(t)$
$u_{\phi,\theta}(t)$
K_p
$y_{\psi}(t)$
$u_{\psi}(t)$
$r_{\psi}(t)$

$Q_{x,y,z}$
$P_{x,y}$
C_i
C_{i+1}




User Input
e.g. tactical motion
Task
Knowledge
Handheld
Robot
Control
demands
Context
sensing

Multiple-object, real-time, texture-less detection and learning

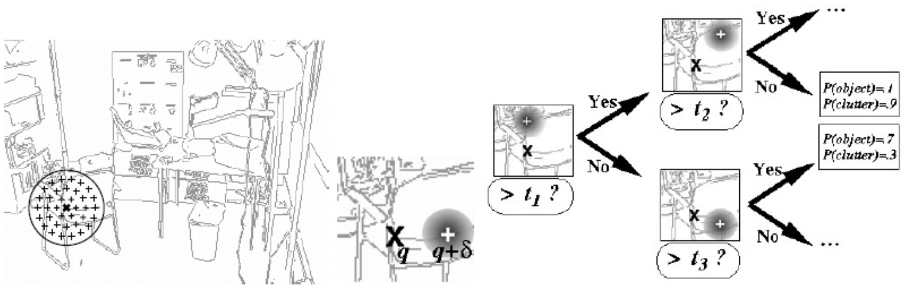
Work from these papers and with these colleagues:

Dima Damen, Teesid Leelasawassuk, Osian Haines, Andrew Calway, Walterio Mayol-Cuevas, You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video. British Machine Vision Conference (BMVC). September 2014.

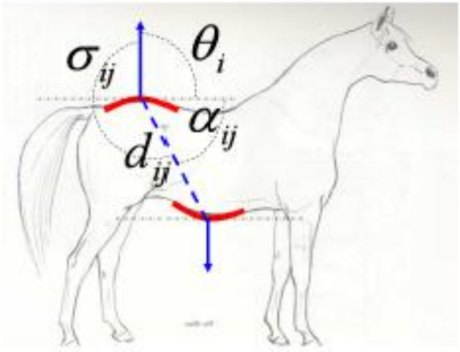
Pished Bunnun, Dima Damen, Andrew Calway, Walterio Mayol-Cuevas, Integrating 3D Object Detection, Modelling and Tracking on a Mobile Phone. International Symposium on Mixed and Augmented Reality (ISMAR). November 2012.

Dima Damen, Pished Bunnun, Andrew Calway, Walterio Mayol-Cuevas, Real-time Learning and Detection of 3D Texture-less Objects: A Scalable Approach. British Machine Vision Conference. September 2012.

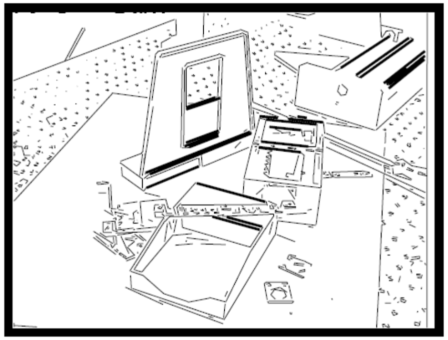
Texture-less object detection



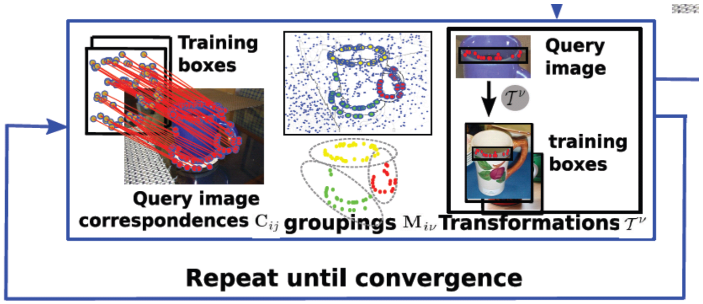
O Carmichael and M Hebert. *BMVC2002.*



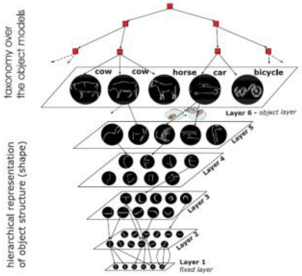
M Leordeanu etal CVPR2007



Beis & Lowe, 1990s



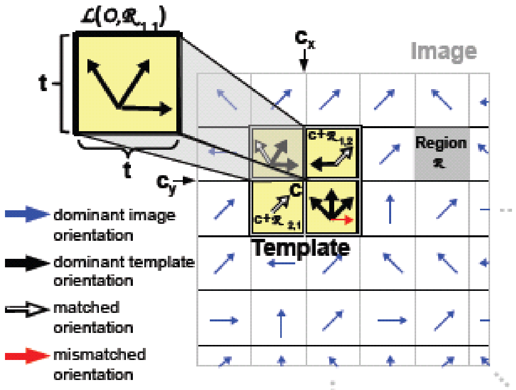
P Yarlagaddaet al *ECCV 2010*



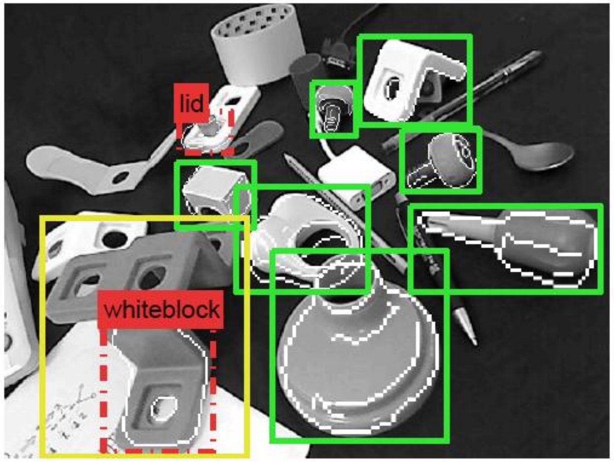
S Fidler et al *ECCV* 2010

Texture-less object detection

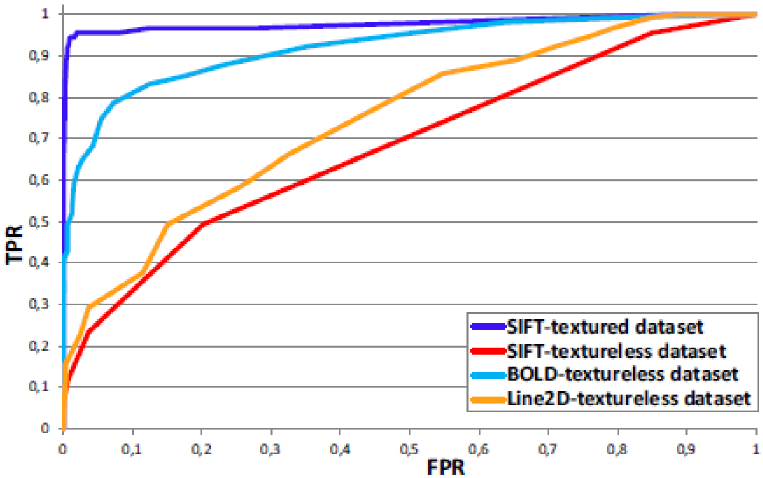
Dominant Orientation Templates:



S Hinterstoisser, et al. *CVPR2010.*



Cai, Werner and Matas, ICVS 2013



Tombari, etal ICCV 2013

Our interest and motivation

- There are less methods out there able to handle well texture-less objects (*Is SIFT helping us to do the indexing rather than to do visual description?*).

- In-situ, generative, “teach-and-use” is more difficult to get right (vs those with luxury of off-line optimization).

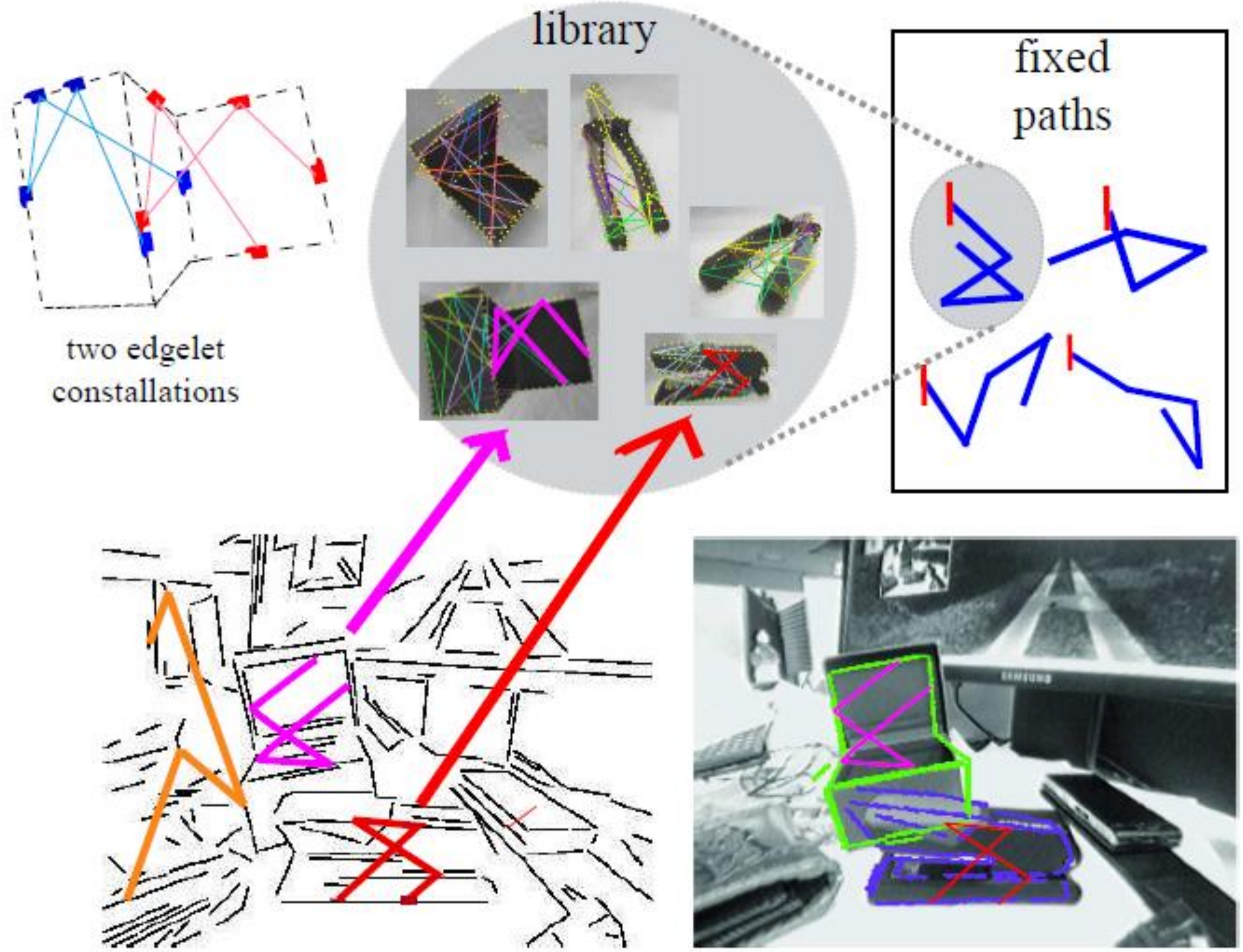
- On a non-networked, self-contained hardware is more difficult (vs with “cloud” servers).

- *But addressing all three issues together helps to develop a type of CV that is ultimately targeted to be practical and useful.*

Out method is:

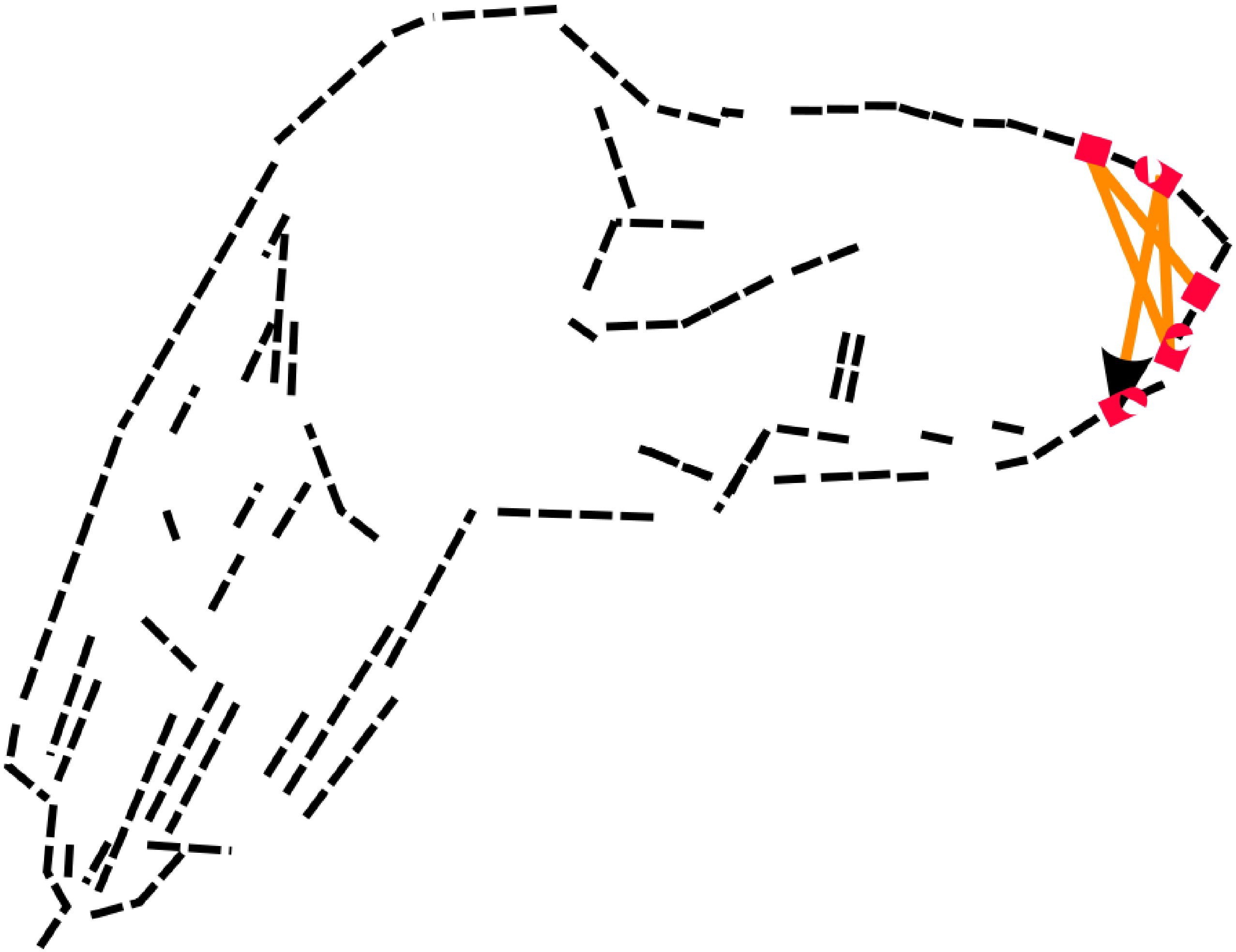
- For *multiple* known 2D/3D objects.
- For texture-minimal / texture-less objects.
- Working at multiple frames per second.
- Scalable.
- Appearance invariance built-in.
- Recovers scale, orientation and or a H.
- *Allows online training:* in-situ and "anywhere" operation.

Detecting Texture-Minimal Objects



Video at https://www.youtube.com/watch?v=4rPjN1mcKGc

Constellation of Edgelets



Constellation of Edgelets

Relative edgelet orientations

$$\phi_i = \widehat{e_i, e_{i+1}}$$

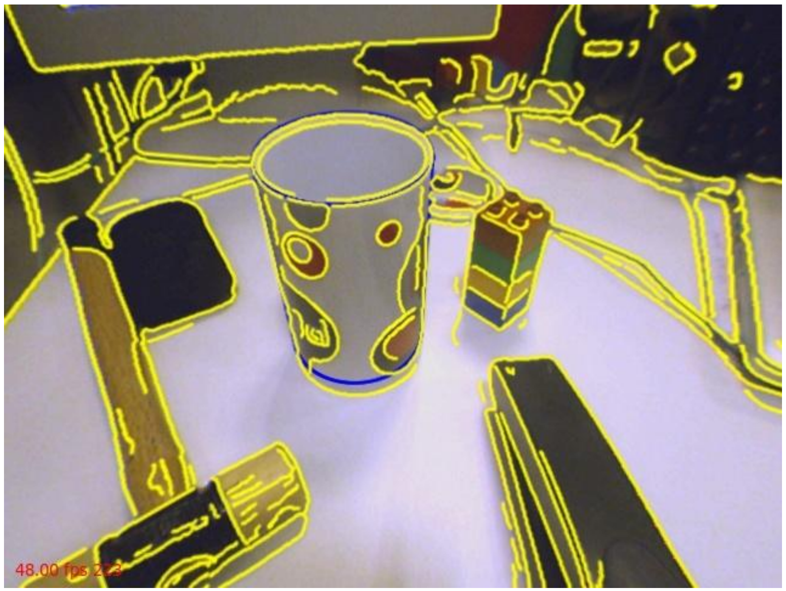
$$\delta_i = |v_{i+1}| / |v_i|$$

Relative distances between edgelets (wrt first distance)

$$f(c_i) = (\phi_1, \ldots, \phi_{n-1}, \delta_1, \ldots, \delta_{n-2})$$

Descriptor is very compact 2*n*-3 and invariant to translation, scale & rotation

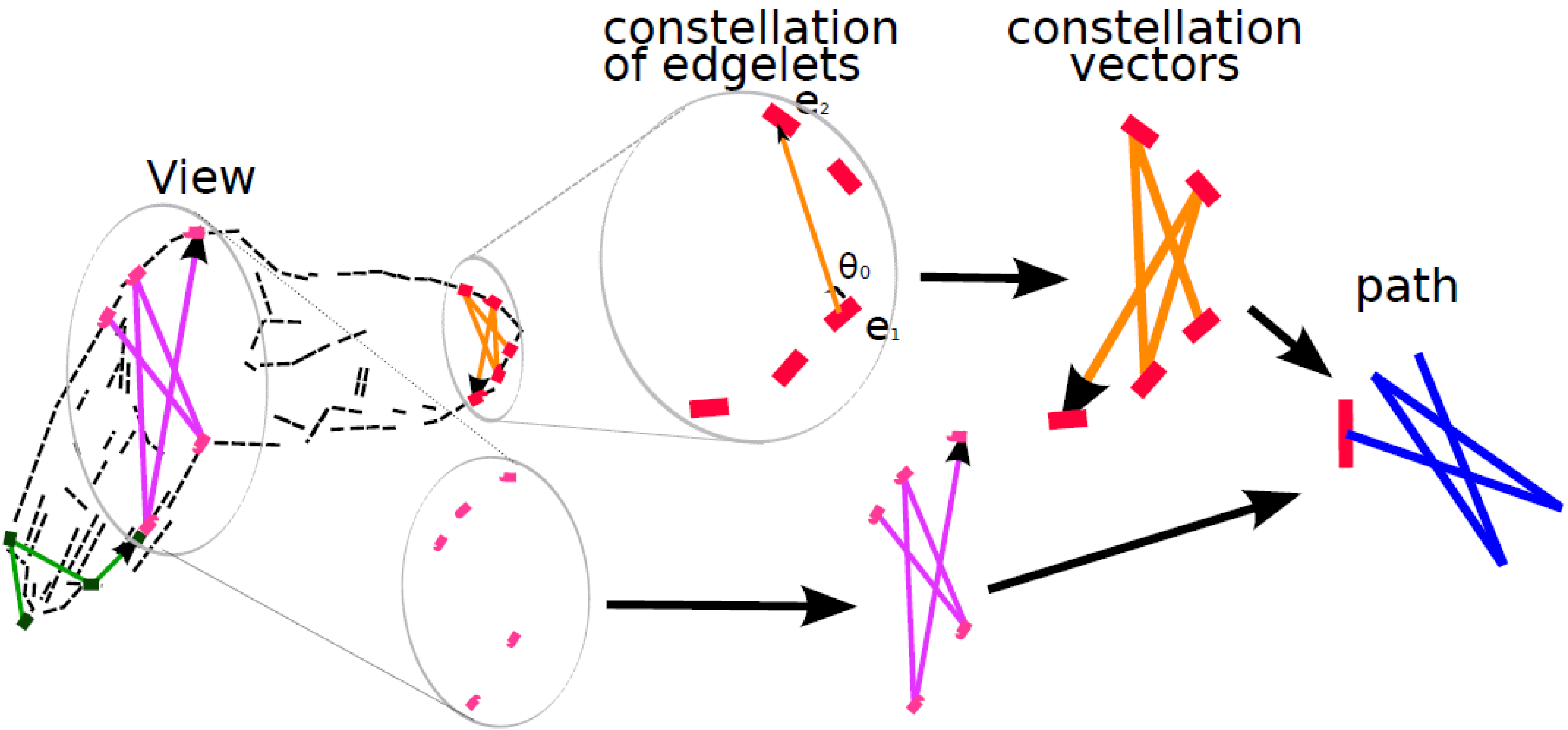
Tackling the tractability problem



On a “Simple” image like this, the maximum number of edge configurations can be of the order of 10s of thousands of millions of possibilities. For a 5-edgelet chain on an image with n edgelets this is: $\dfrac{n!}{(n-5)!}$

Key idea: use **fixed paths**. Instead of searching and training for the object in any possible way, fixing paths does this in a pre-determined manner. For this image it means 8 orders of magnitude less options instead ~1.5K only.

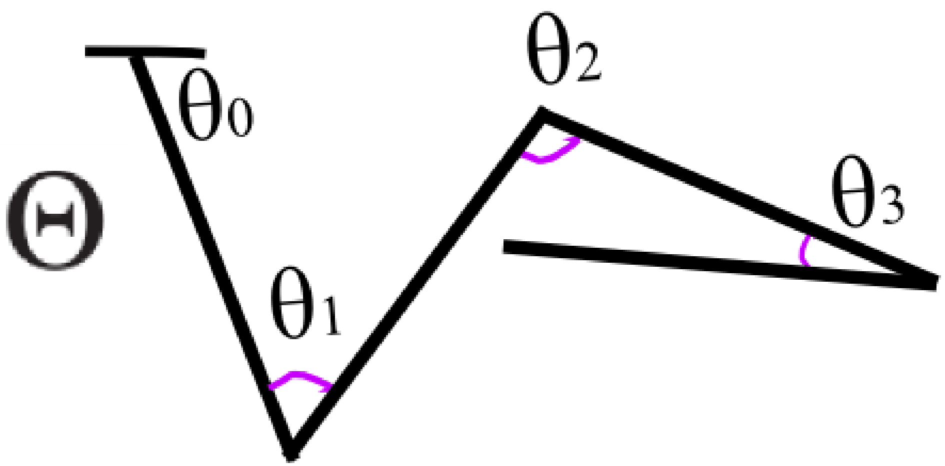
Fixed Paths

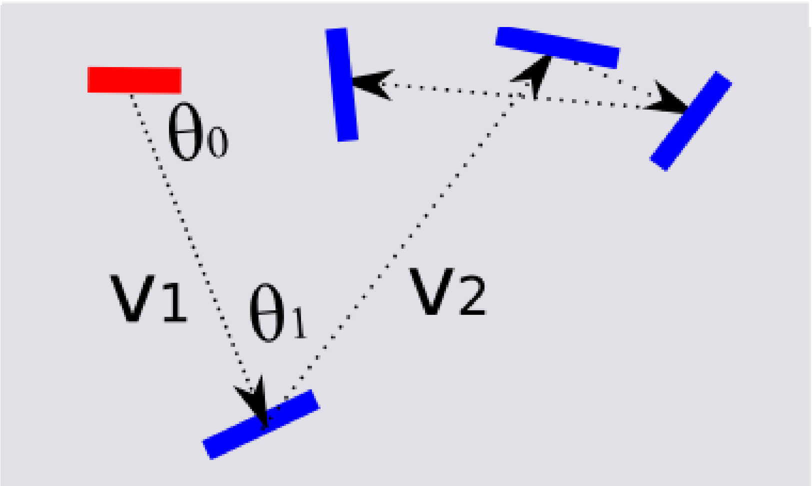


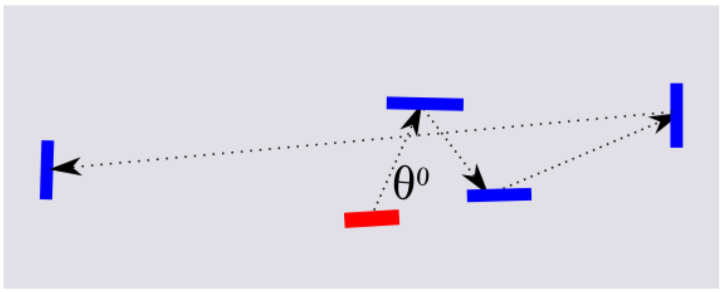
A constellation, when detected tries to verify shape via the rest of edgelets and an iterative alignment via a Homography

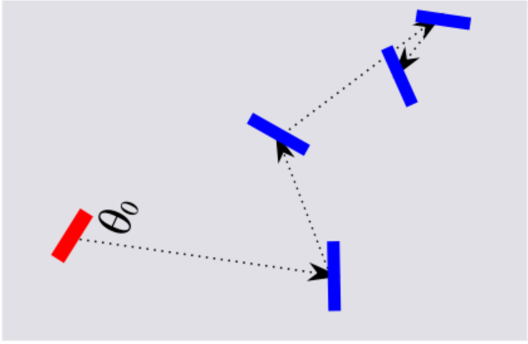
What is a *fixed* path?

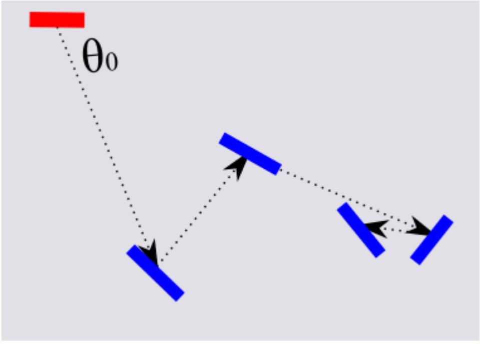
$$cos(\theta_1) = (v_1 \cdot v_2)/(|v_1||v_2|)$$





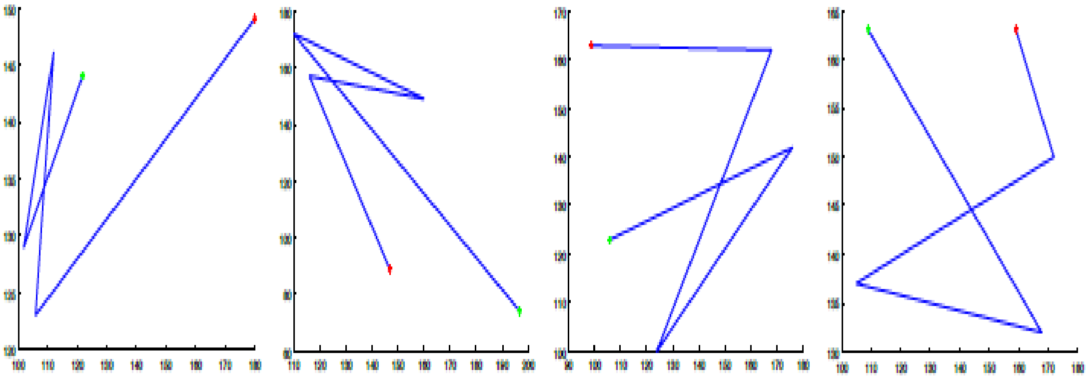




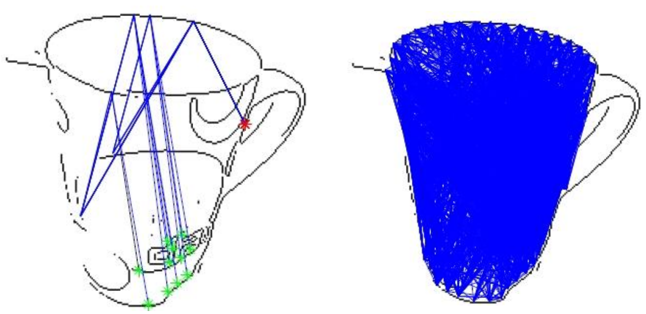


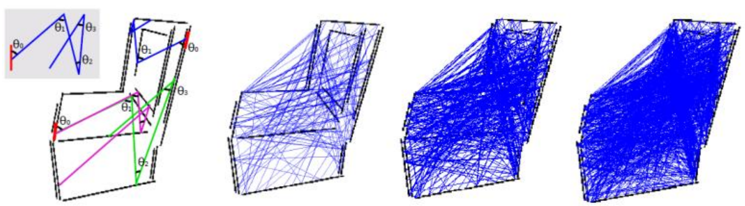
Damen, et al. BMVC 2012

A fixed path:



Extracts many different descriptors i.e. lengths and edgelet's relative orientations



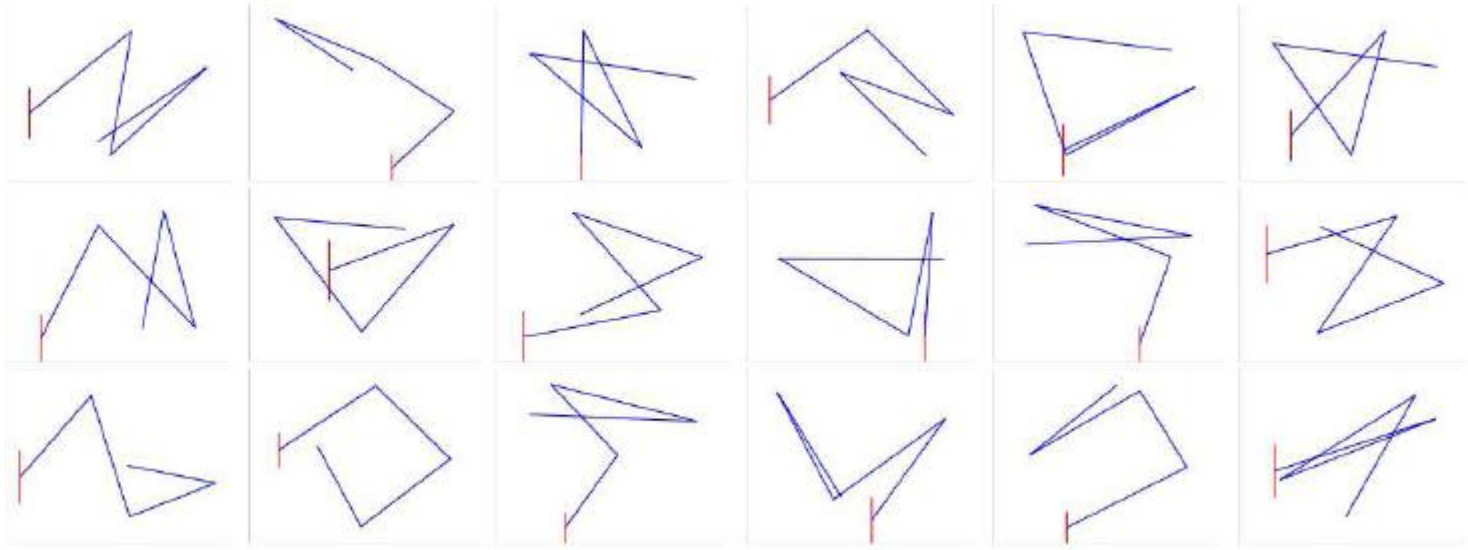


A single, fixed path still covers well different objects

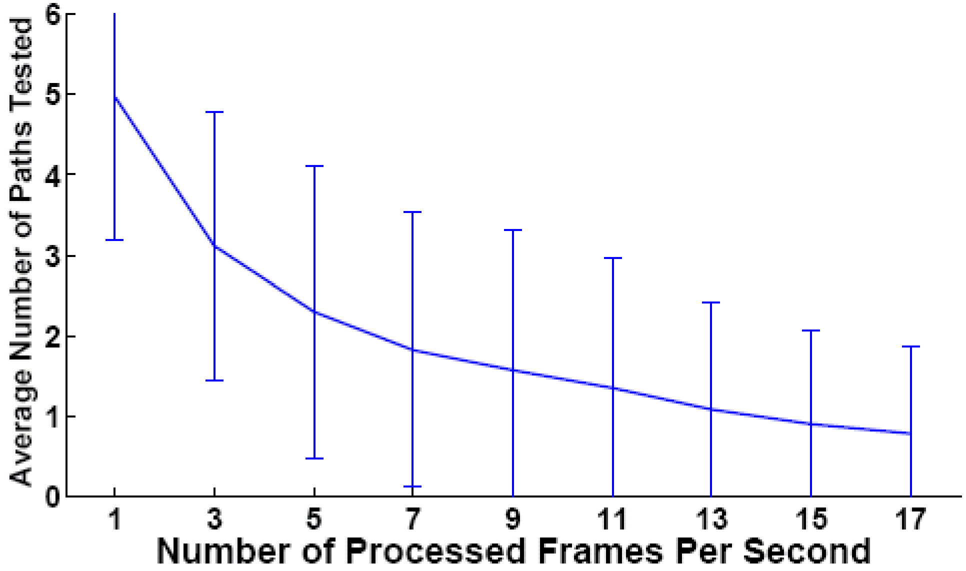
Damen, et al. BMVC 2012

How to Select the Paths??

- We performed this once:
- Randomly selected 100 angle tuples
- Test performance on an independent set of objects
- Test # of extracted constellations + ambiguity of descriptor
- Best 6 paths were selected

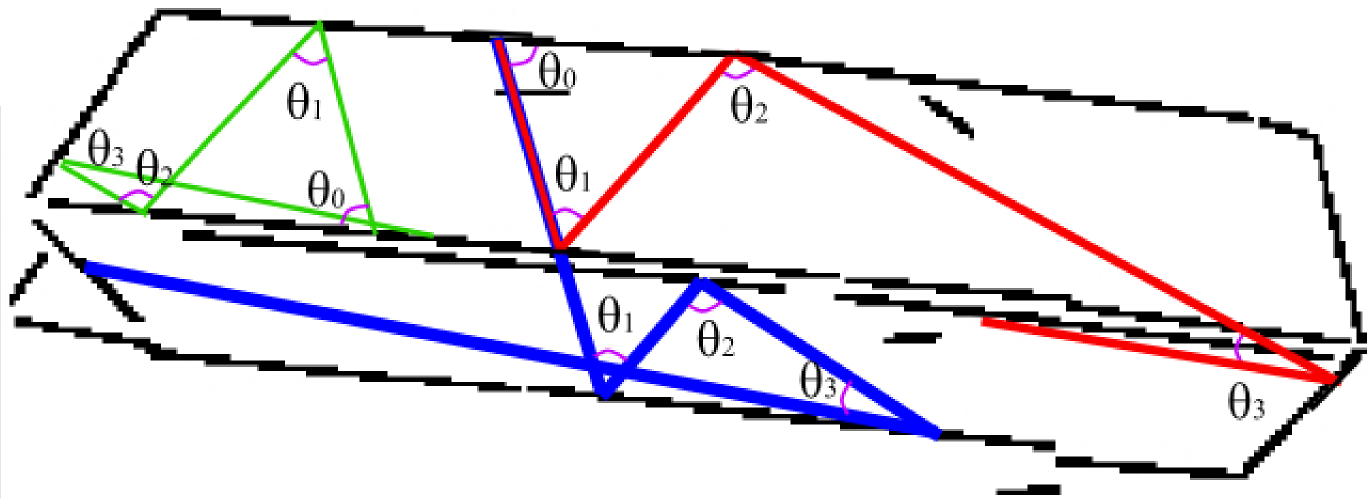


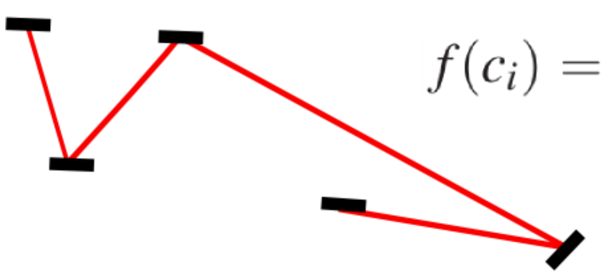
How to Select the Paths??



Order of paths	Acc. % of detections after n paths					
	1	2	3	4	5	6
(1,3,4,5,6,2)	75.61	90.42	91.04	94.13	98.45	100
(2,3,4,5,6,1)	51.84	82.61	89.3	94.65	96.99	100
(3,1,2,4,5,6)	61.07	86.26	87.28	90.33	95.67	100
(4,3,5,1,6,2)	78.12	90.89	95.45	98.19	99.41	100
(5,1,2,4,6,3)	80.79	88.90	89.50	91.6	94.9	100
(6,5,4,3,2,1)	67.91	84.70	88.06	95.90	99.24	100
Avg.	69.22	87.30	90.10	94.13	97.44	100

On-line Training



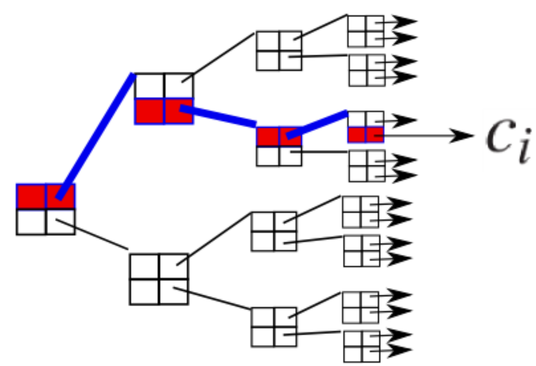


$$f(c_i) = (\phi_1, \ldots, \phi_{n-1}, \delta_1, \ldots, \delta_{n-2})$$

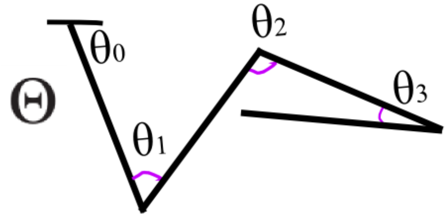
$$\phi_i = \widehat{e_i, e_{i+1}}$$

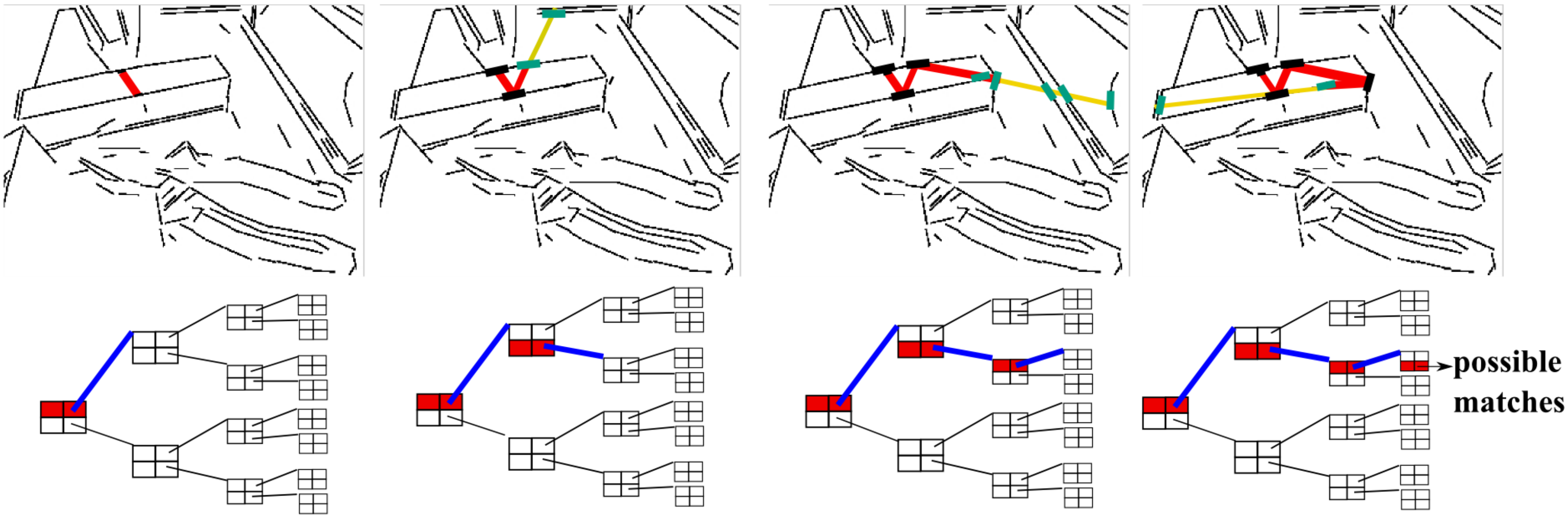
$$\delta_i = |v_{i+1}|/|v_i|$$

$$\xrightarrow{\quad}$$
$$f(c_i)$$

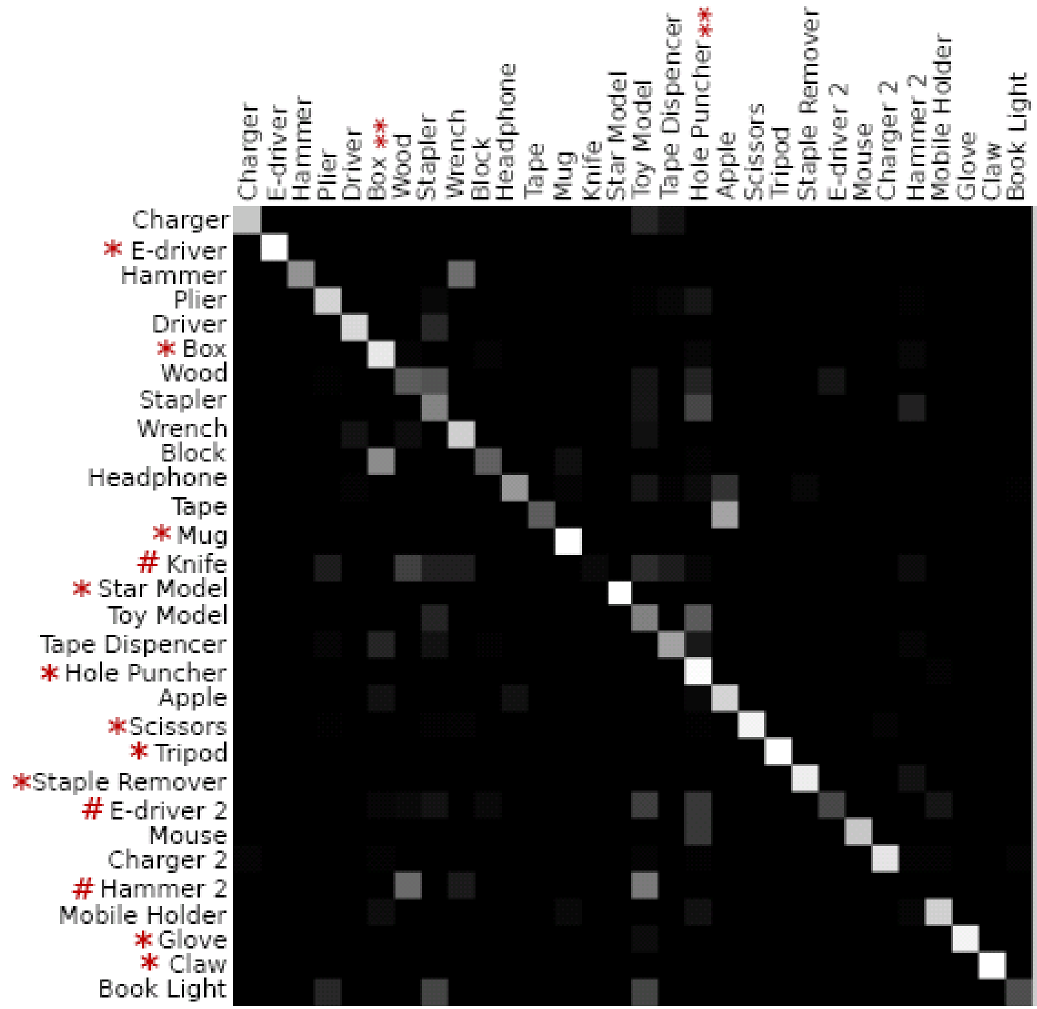


Testing



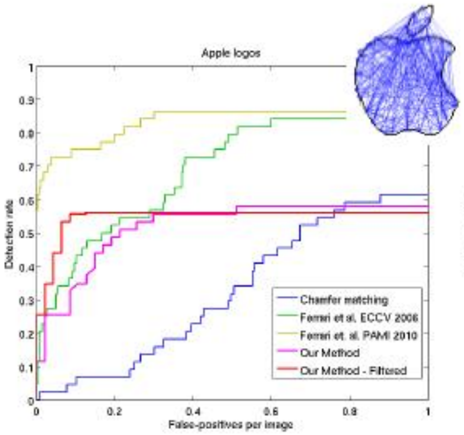


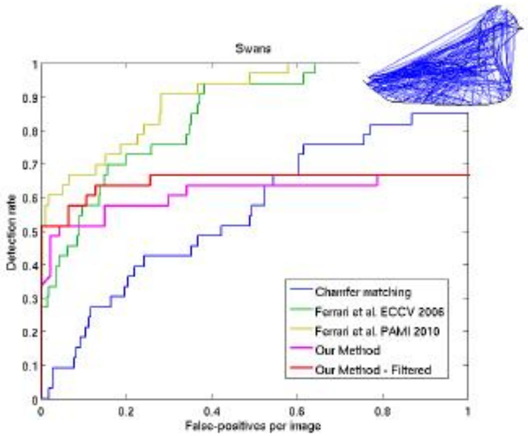
30 objects & tools



Damen, et al. BMVC 2012

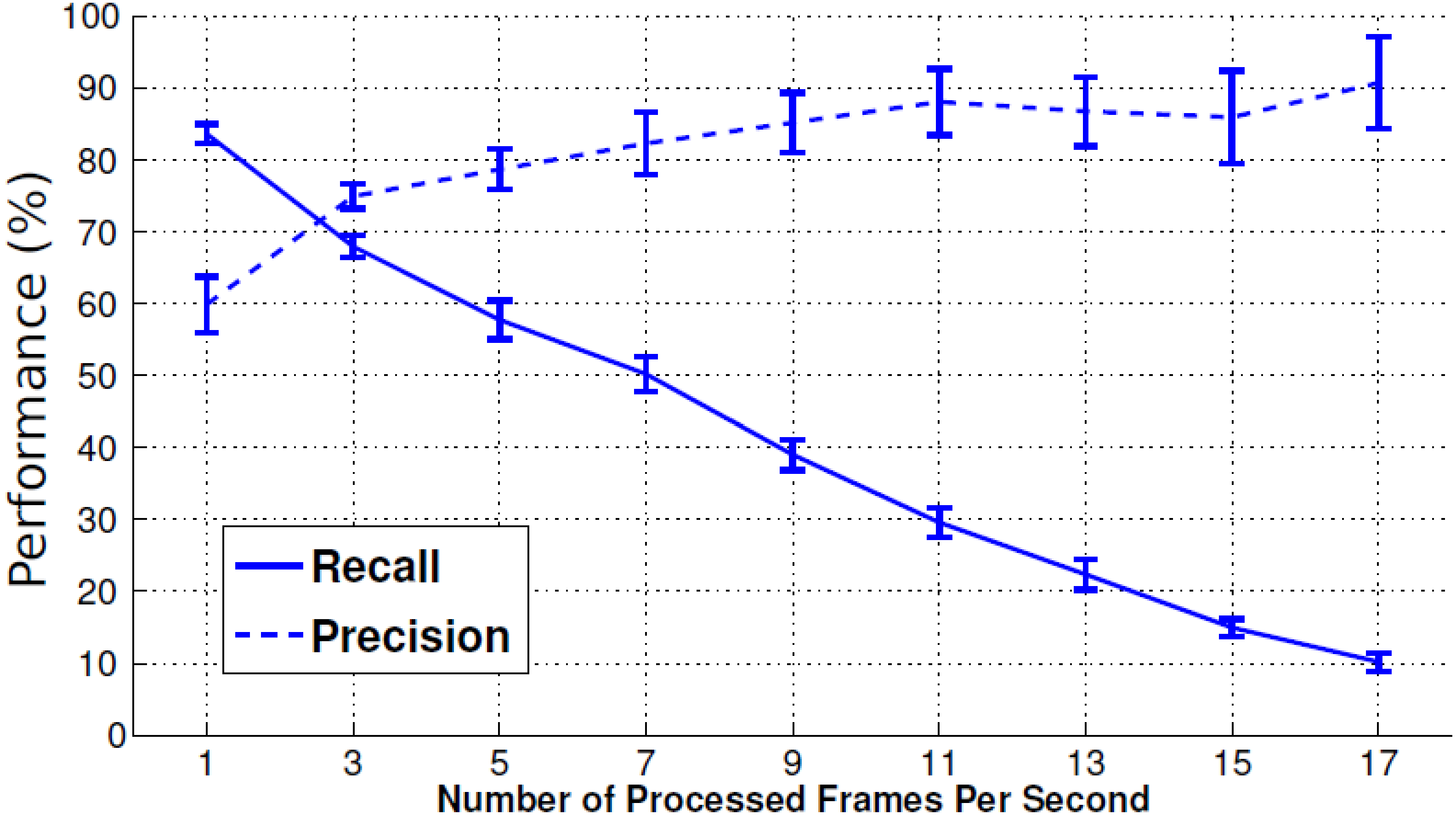
On ETHZ dataset





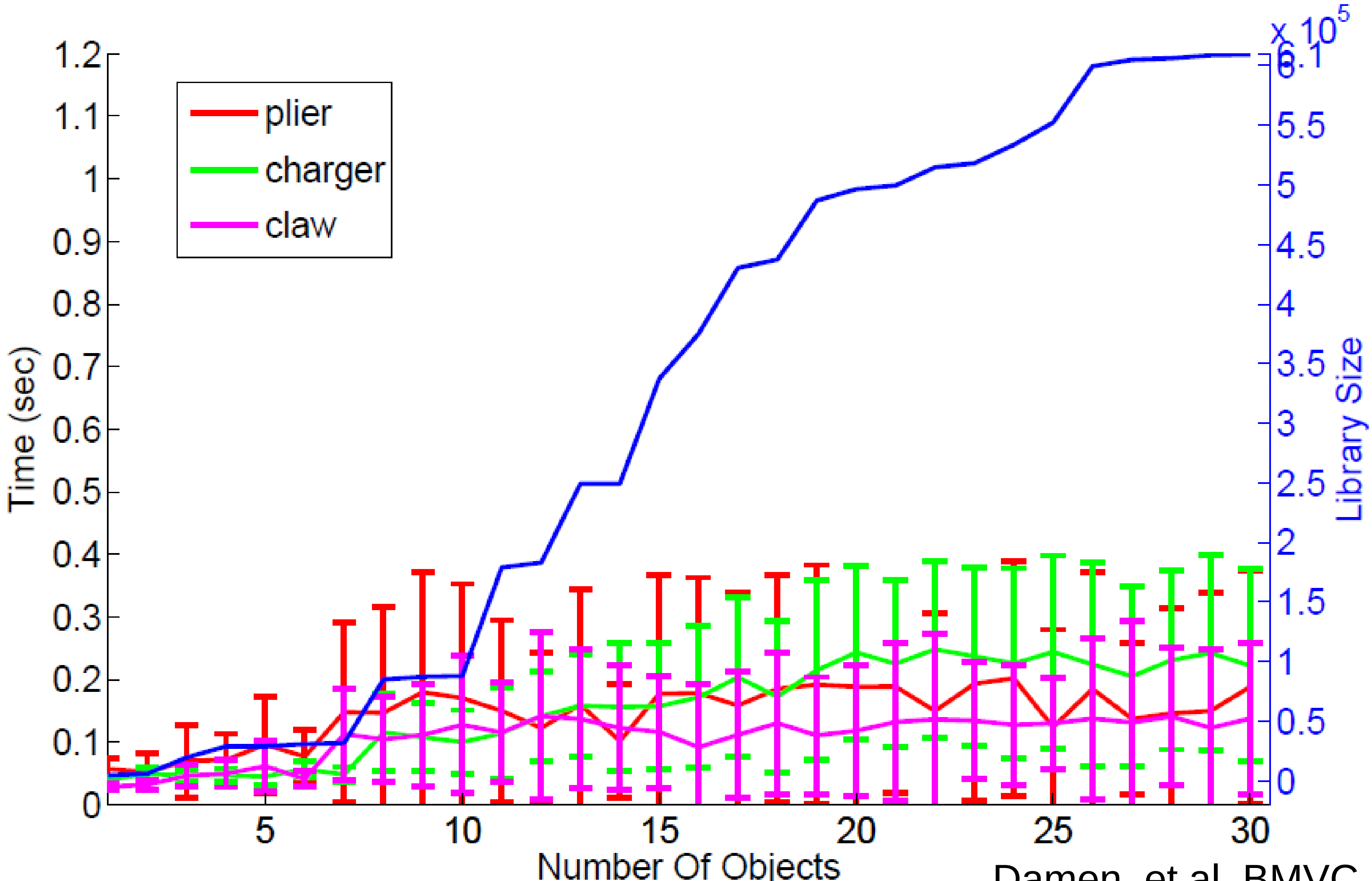
	Apple Logo	Swan	Bottle	Giraffe	Mug
[8]	83.2	75.4	83.2	58.6	83.6
[27](v)	84.0	76.7	93.1	79.5	67.0
[27](v+f)	95.8	94.1	96.3	84.1	96.4
[10]	87.3	80.0	87.6	83.5	86.1
[5]	73.0	63.5	86.9	80.3	81.6
Ours	73.2	66.1	68.97	72.4	60.9

Results – Recall vs Precision



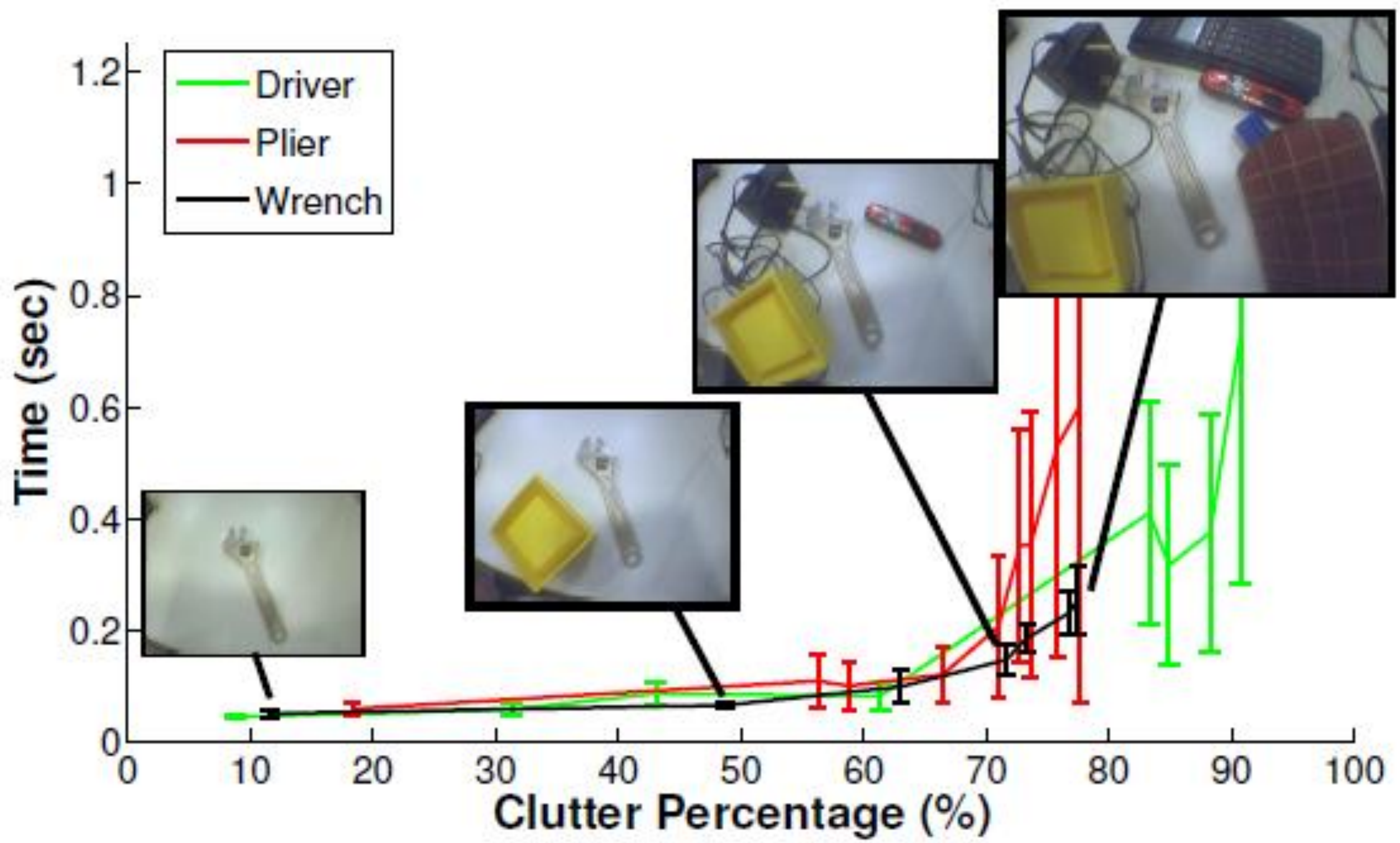
Damen, et al. BMVC 2012

Results - Scalability



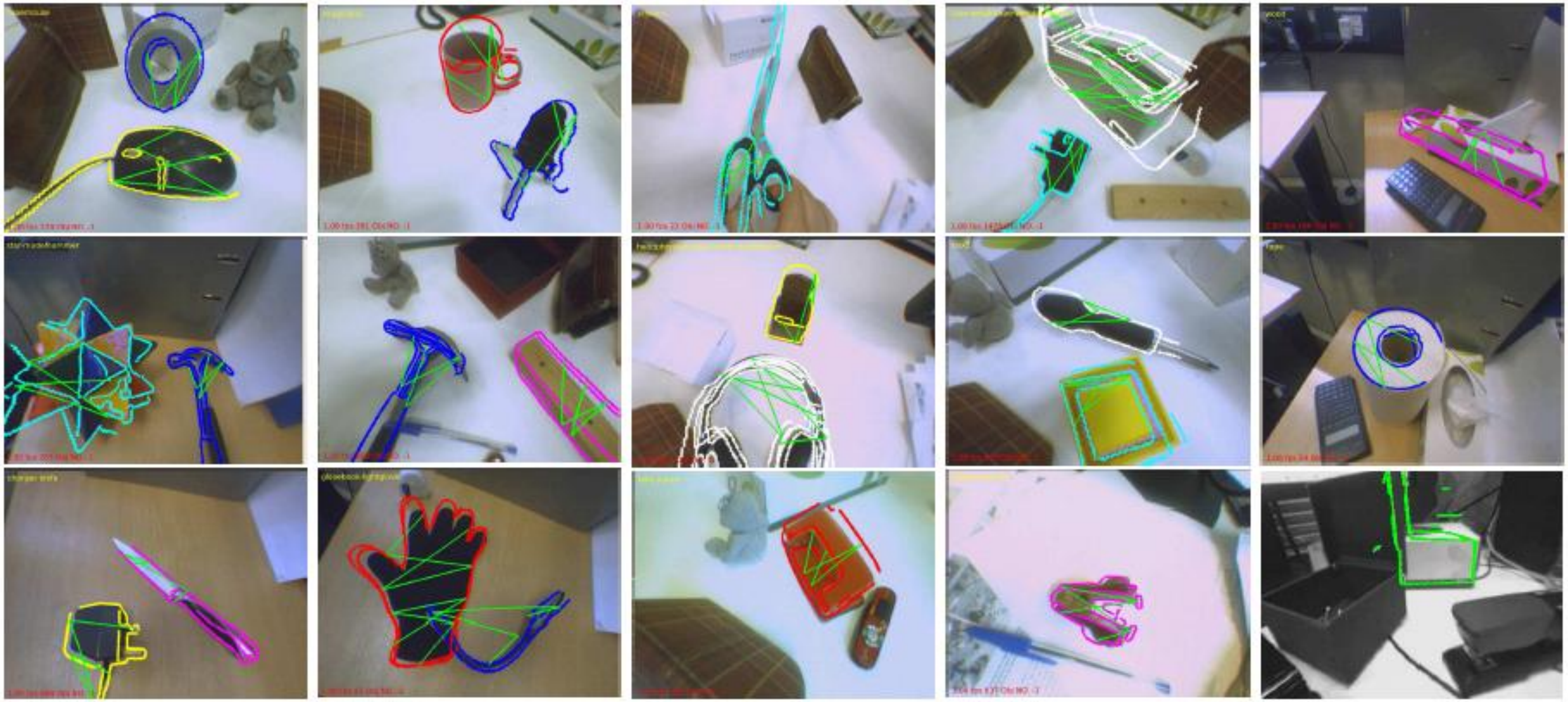
Damen, et al. BMVC 2012

Clutter handling



Damen, et al. BMVC 2012

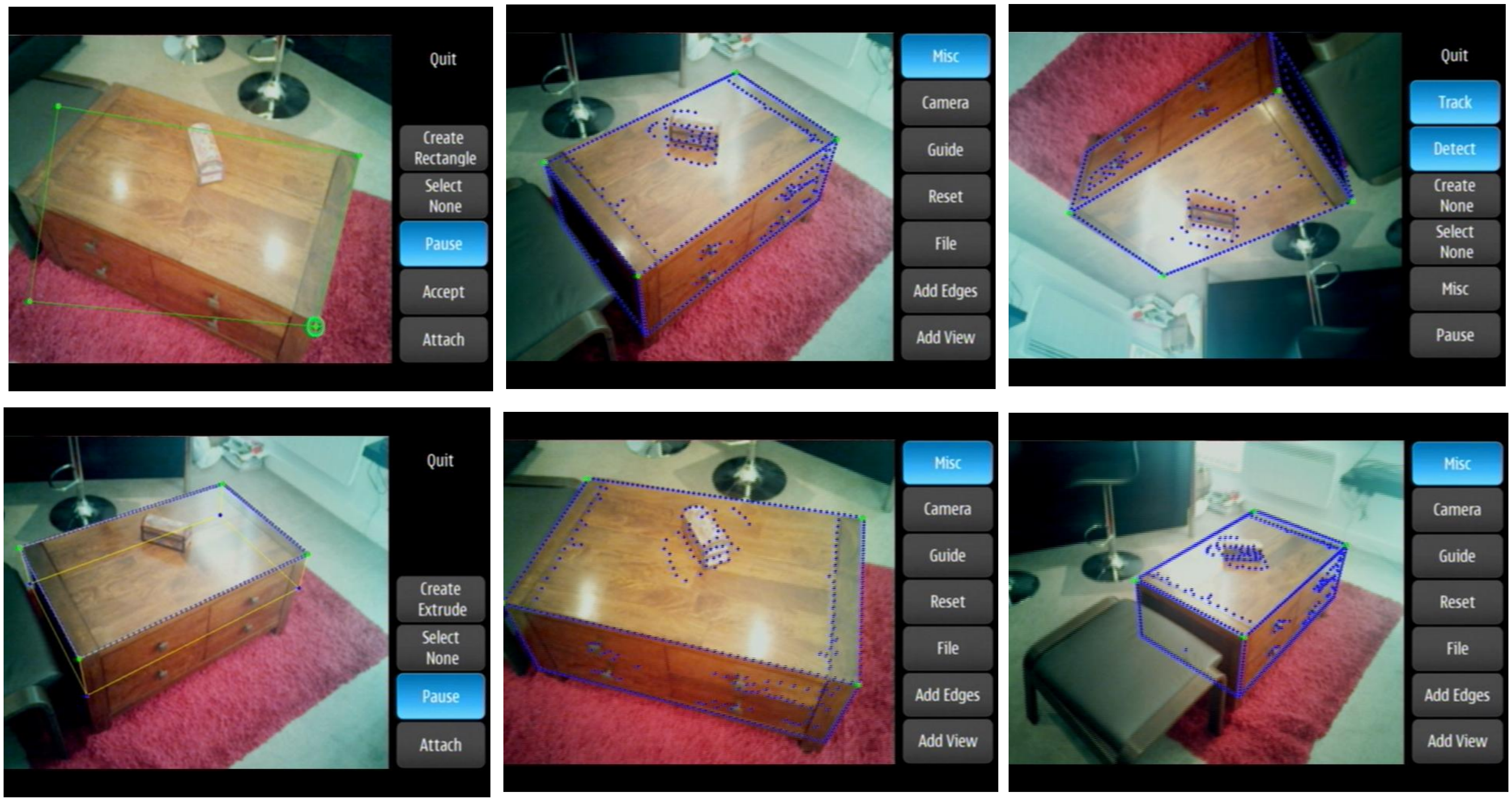
30 objects & tools



Live demo on Android

Download app from:
http://www.cs.bris.ac.uk/~damen/MultiObjDetector.htm

In-situ modelling, tracking and detection on a mobile



In-situ modelling

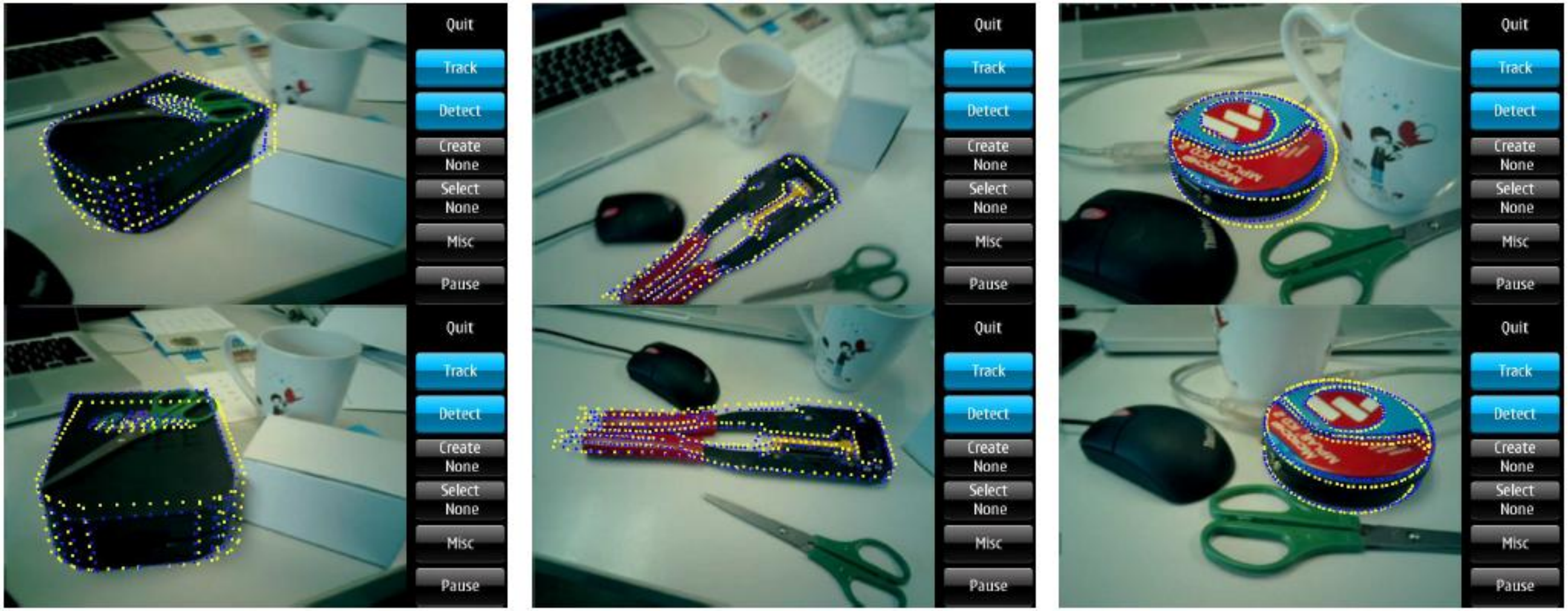
6D tracking

Detection

Bunnun, Damen, Calway, Mayol-Cuevas, ISMAR 2012

Results on mobile phone

About 0.8s per successful detection on images with about 150 edglets



Bunnun, Damen, Calway, Mayol-Cuevas, ISMAR 2012

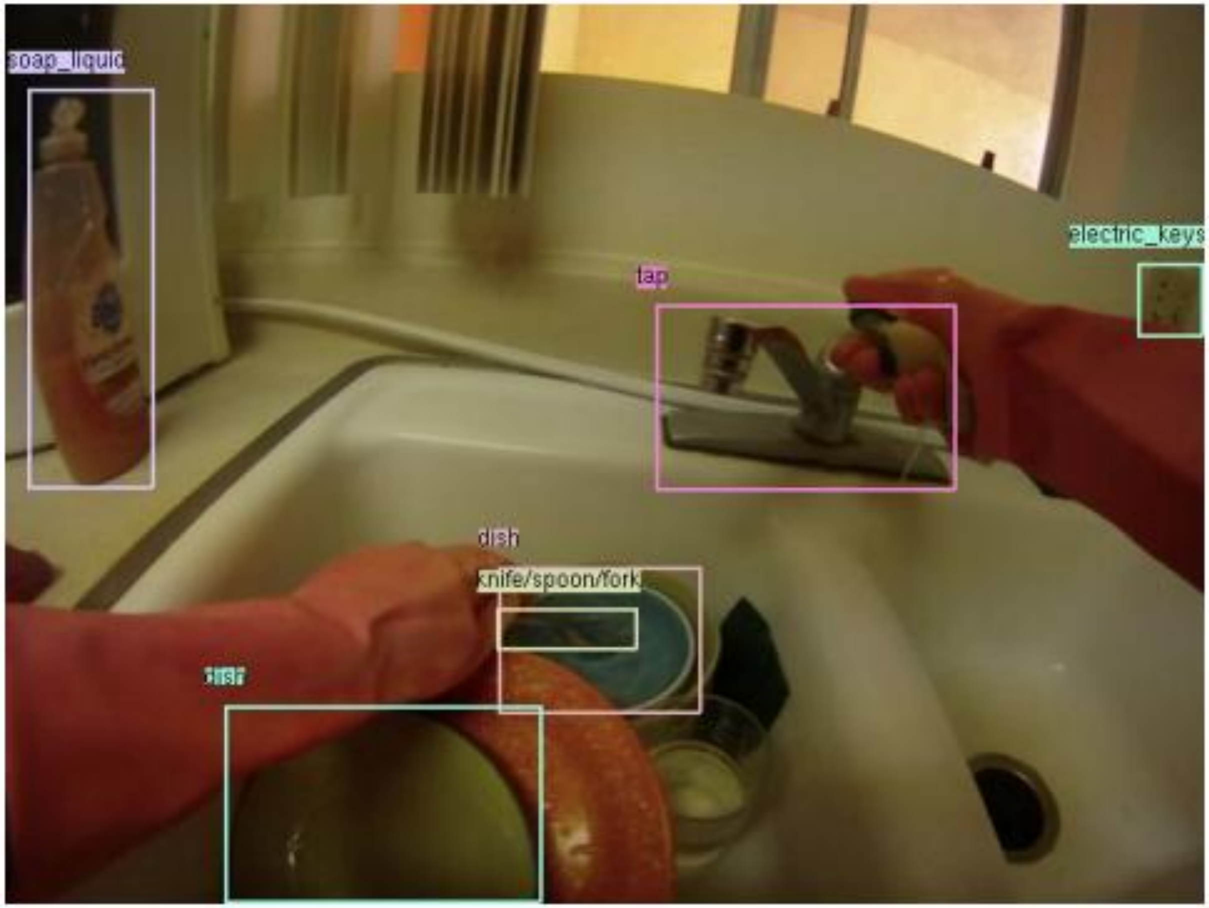
Discovering Objects of Relevance and how these are used from Wearable Vision

You Do, I Learn

- Egocentric view
- Multiple Operators
- Discover used objects
- Discover how objects have been used
- Extract guidance videos
- Fully unsupervised
  - No prior knowledge of objects (number, size)
  - Static and moveable objects

Related Work

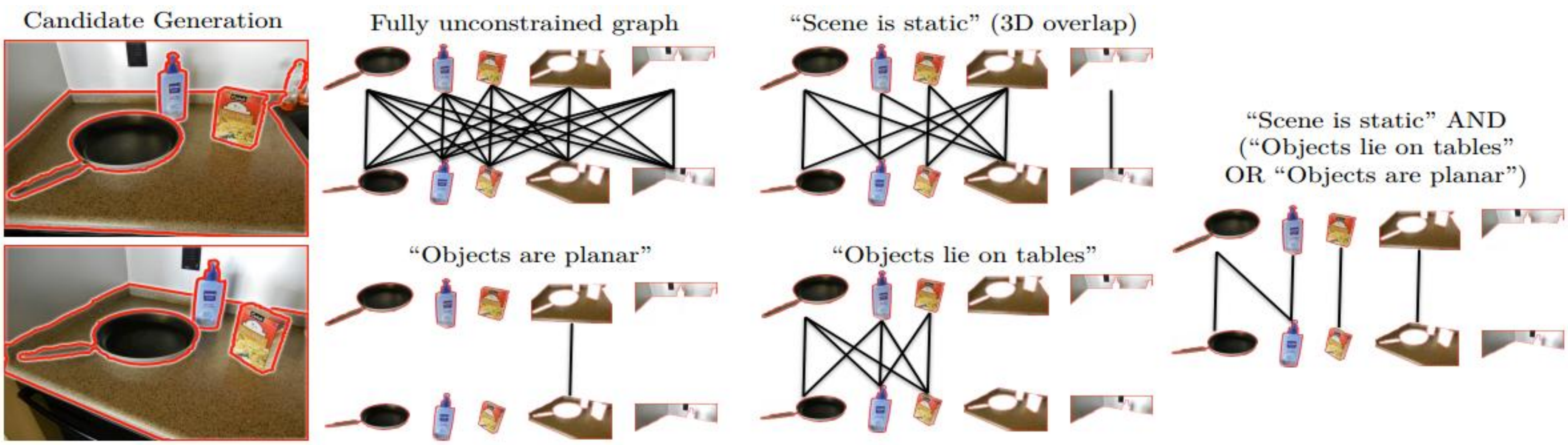
Expect objects to be known apriori…



H Pirsiavash and D Ramanan. Detecting acitivites of daily living in first-person camera views. CVPR, 2012.

Related Work

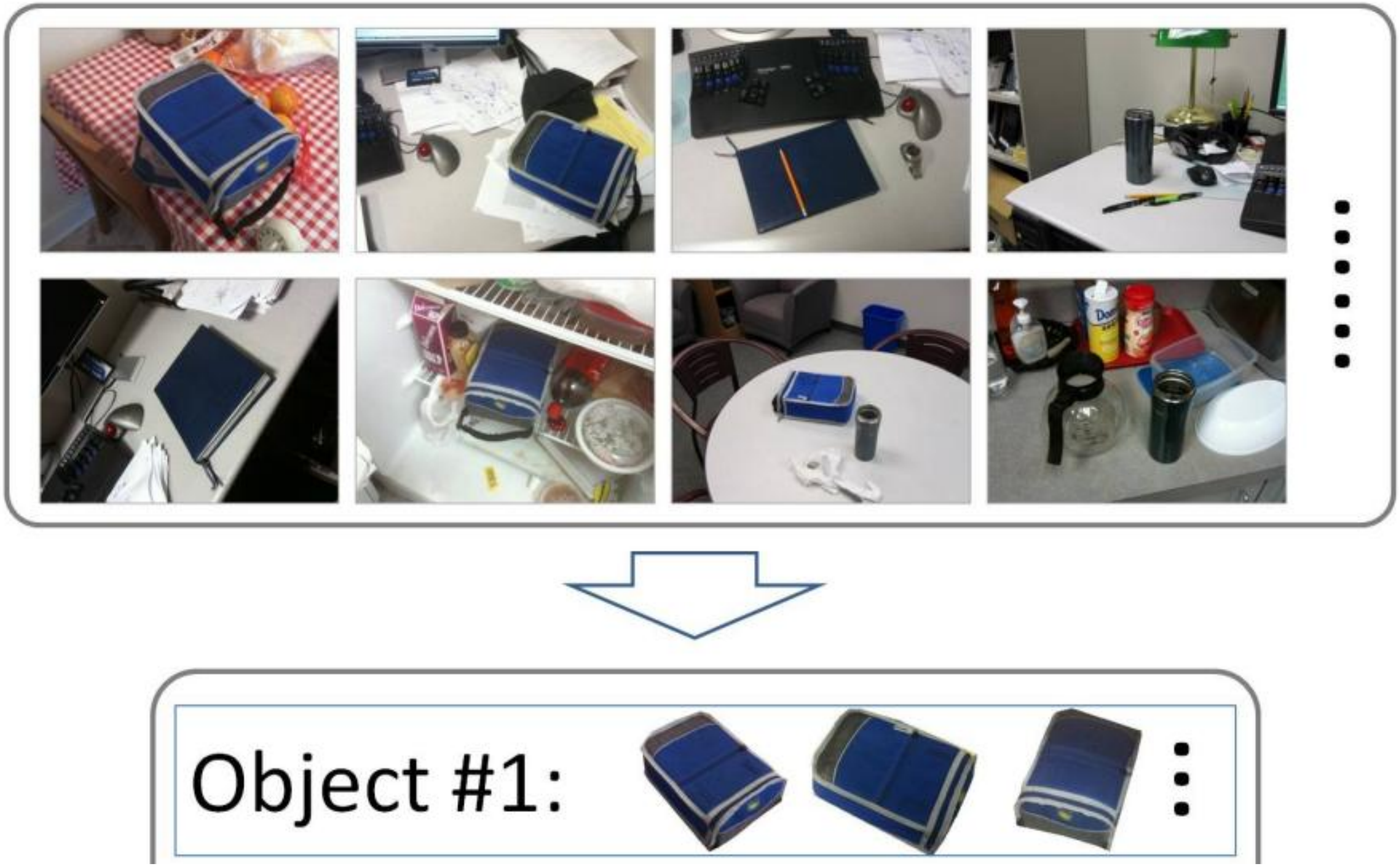
- Discover all objects in the scene
- Have assumptions on objects (planar, table-top)



A Collet, B Xiong, C Gurau, MHebert, and S Srinivasa. Exploiting domain knowledge for object discovery. ICRA, 2013.

Related Work

Expect the object to be moved



H Kang, M Hebert, and T Kanade. Discovering object instances from scenes of daily living. ICCV, 2011.

Definitions

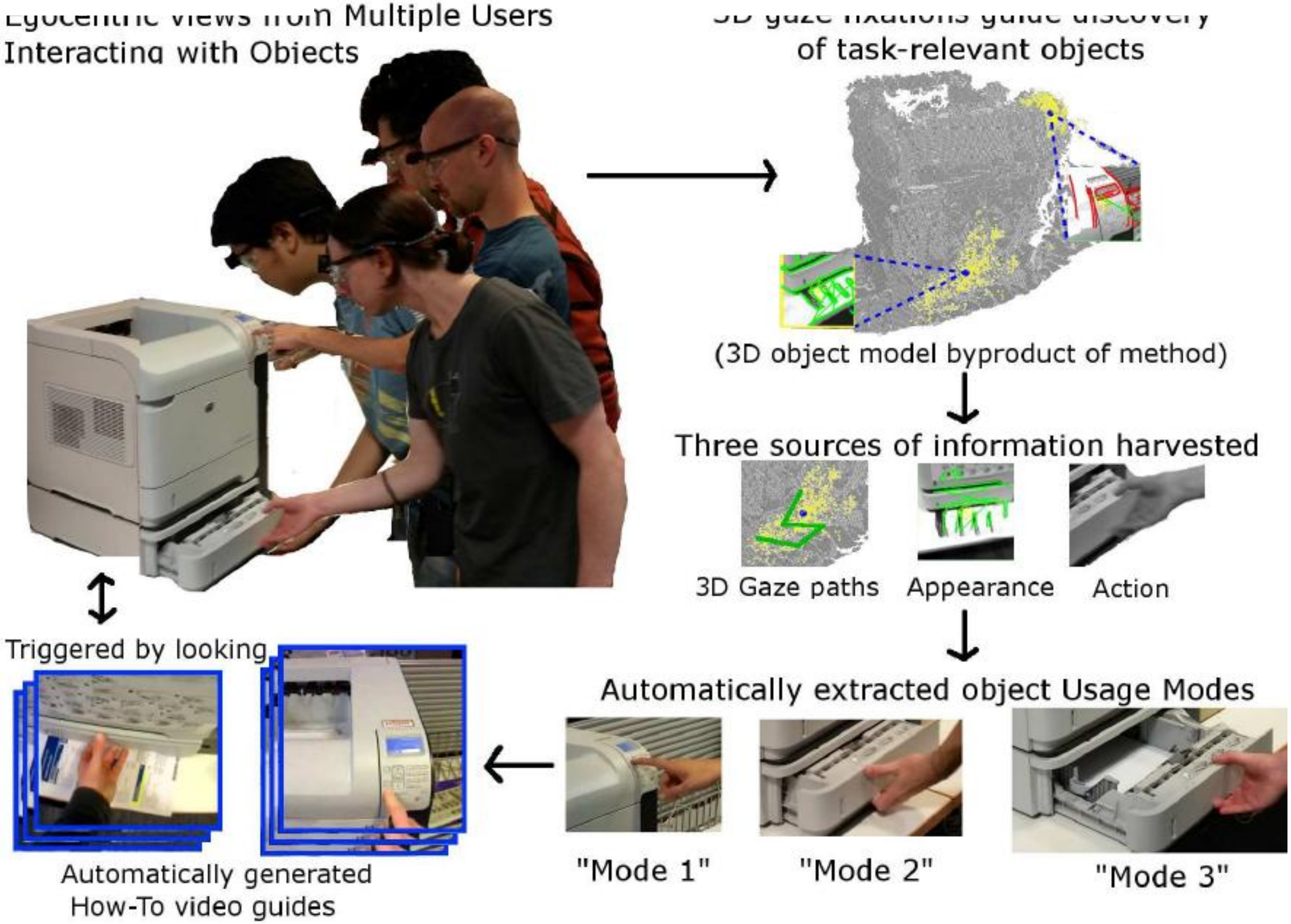
- Task-Relevant Object (TRO)

  an object, or part of an object, with which a person interacts during task performance

- Modes of Interaction (MOI)

  the different ways in which TROs are used

You Do, I Learn - Results



Video at: https://www.youtube.com/watch?v=vUeRJmwm7DA

Discovering Task Relevant Objects

- By combining attention, position and appearance

…it's a clustering task

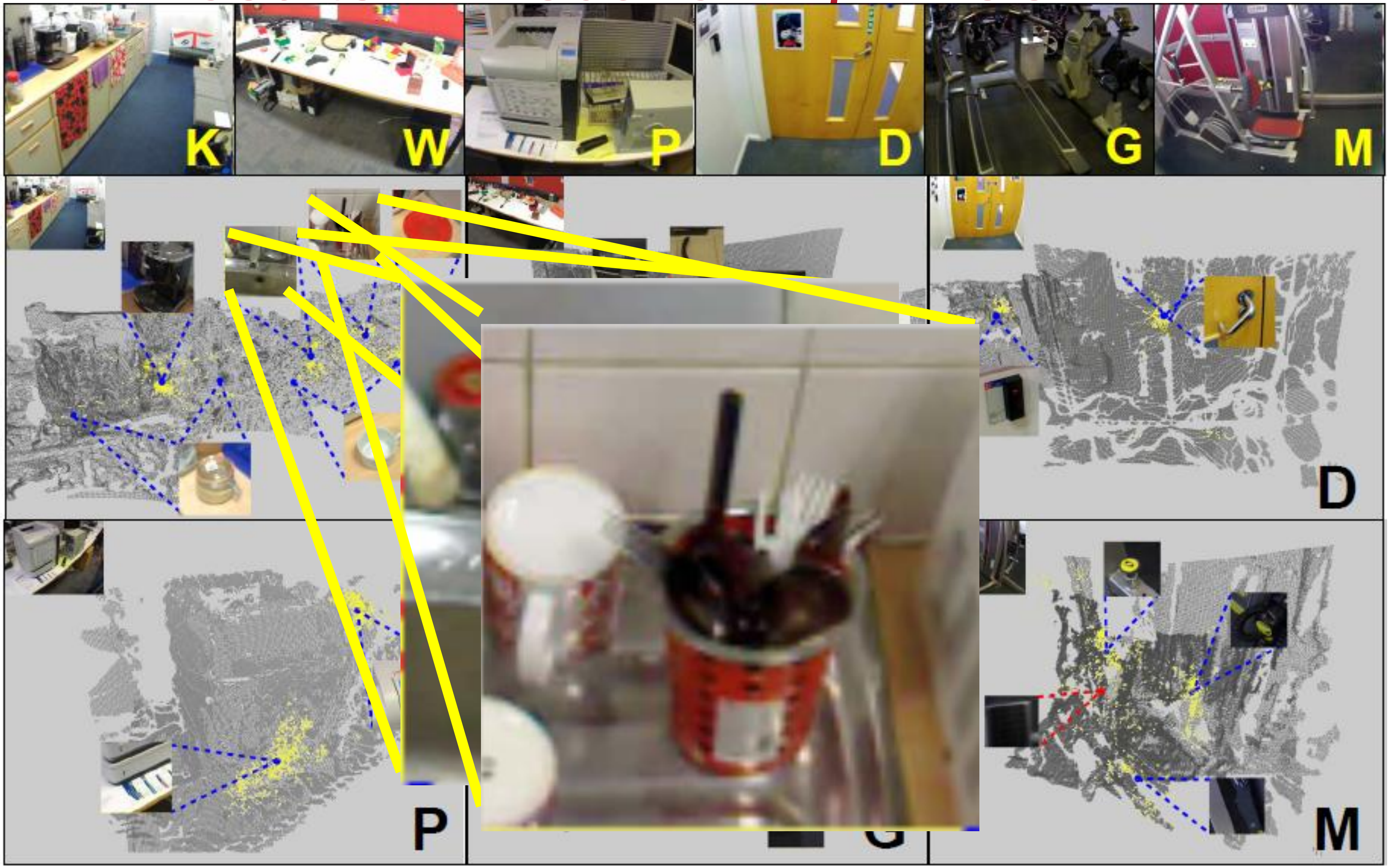
  - K-Means vs. Spectral Clustering
  - Unknown number of objects
  - Davies-Bouldin (DB) Index

Discovering Modes of Interaction

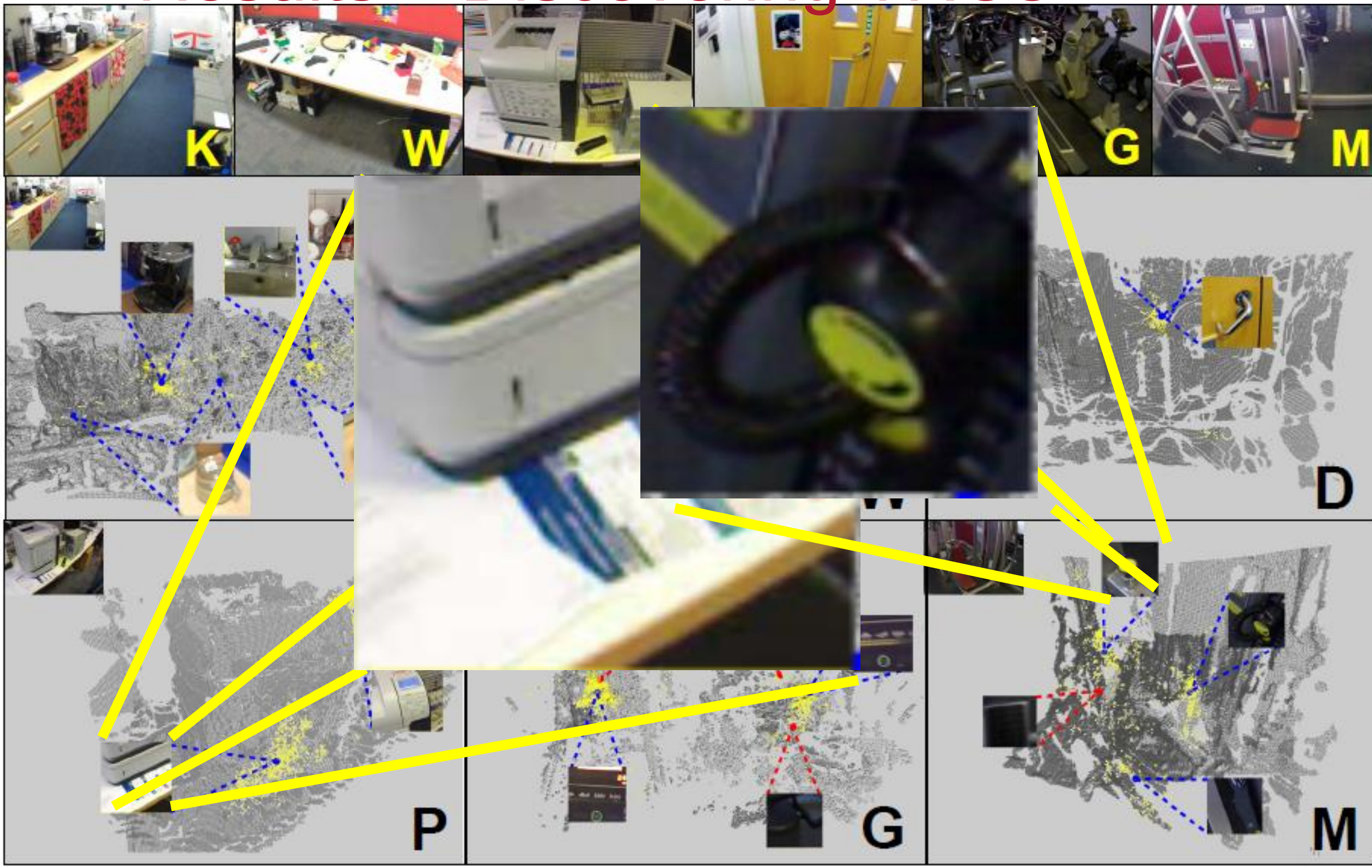
Motion

- Video snippets around each discovered object
- 3D Harris points
- Histogram of Optical Flow (HOF)
- BoW
- Temporal Pyramid

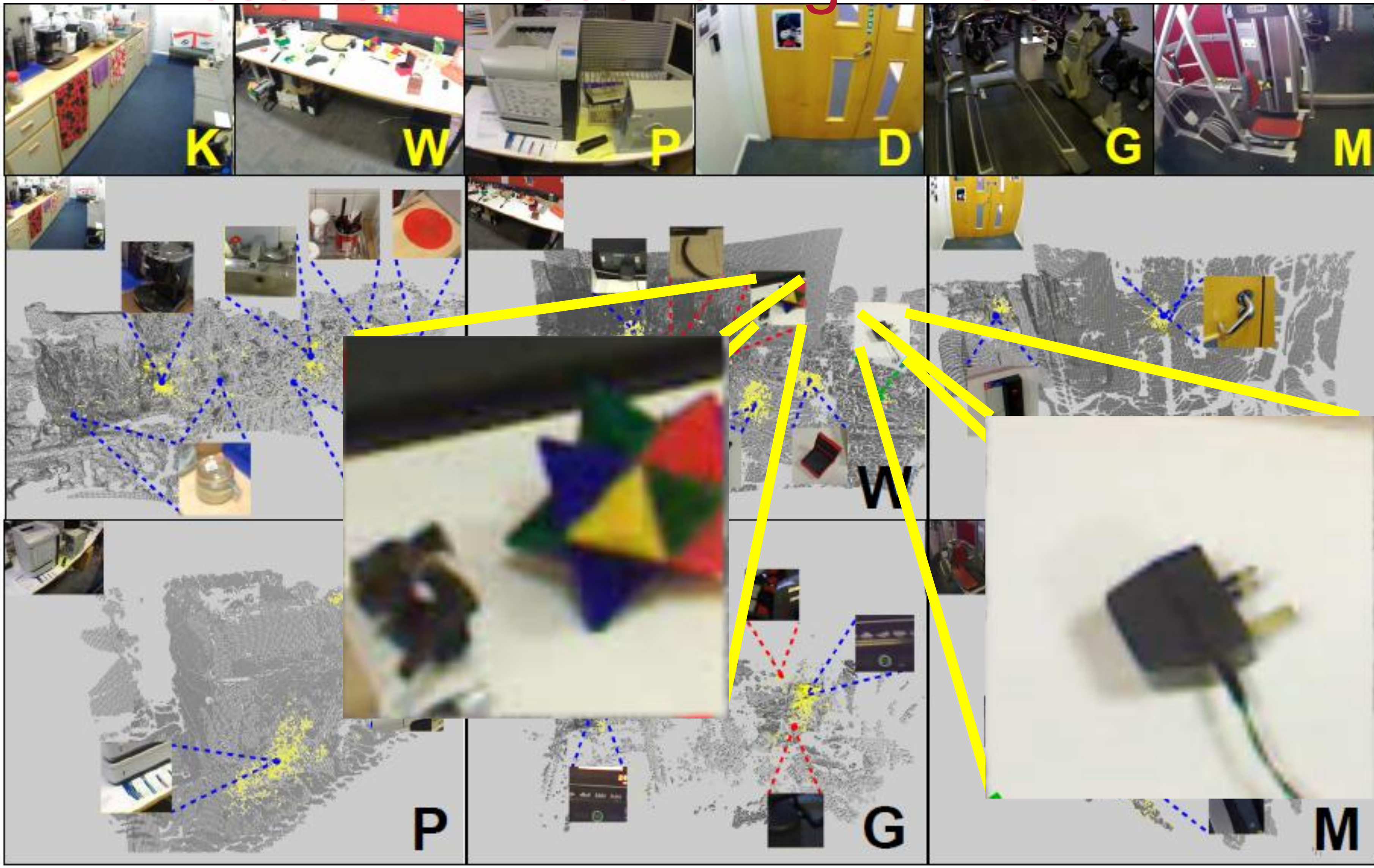
Results – Discovering TROs



Results – Discovering TROs

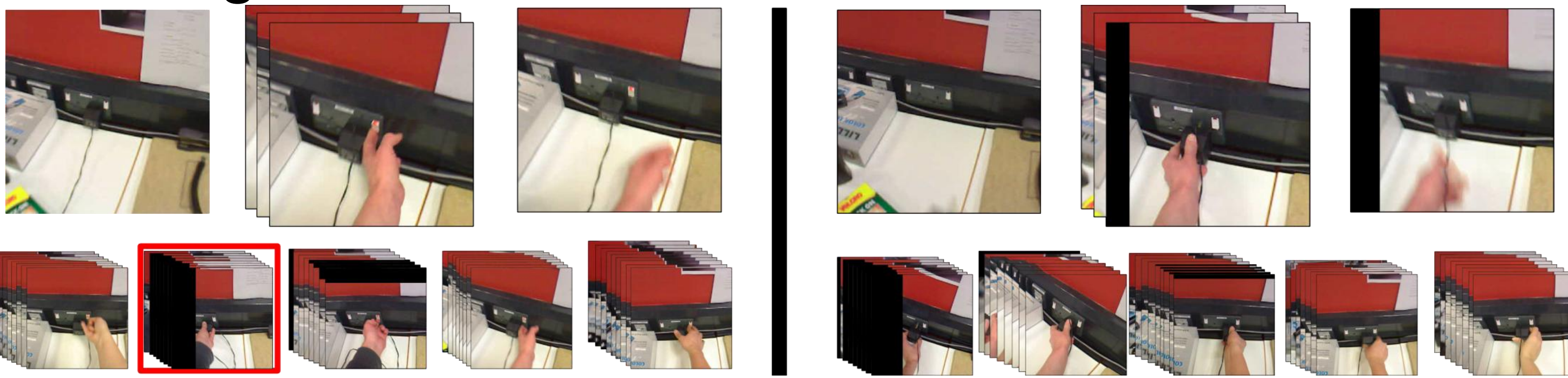


Results – Discovering TROs



Results – Discovering MOIs

- E.g. Electric Socket



Results – Discovering MOIs

Open & get sugar

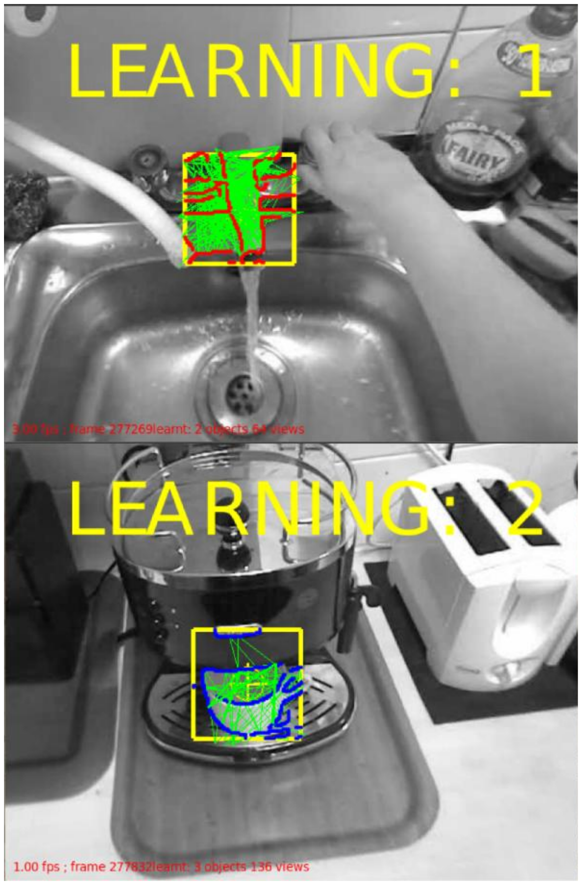
Put

Get

Open door

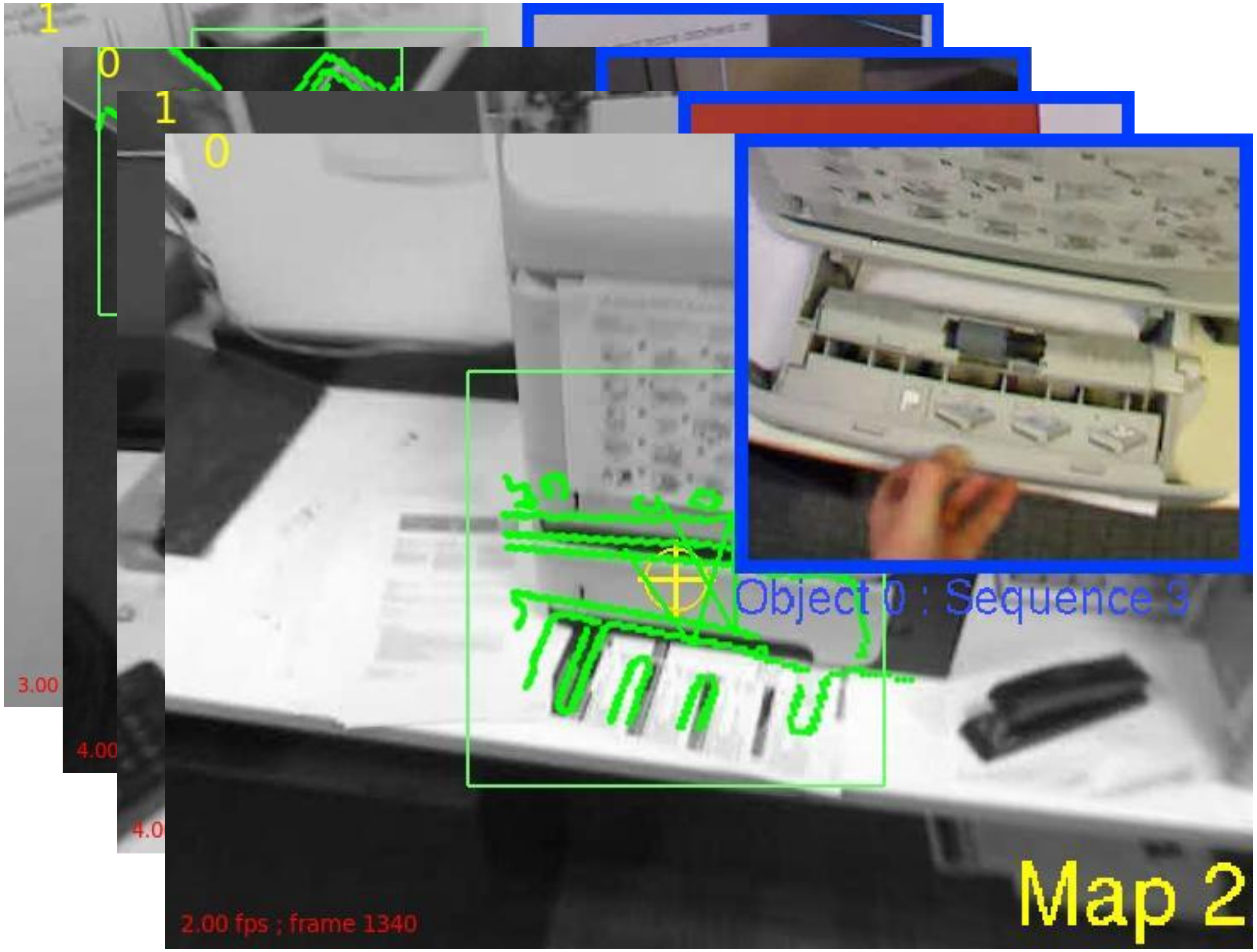
Results – Video Guides

- Real-time texture-minimal scalable detector[1]



[1] Dima Damen, Pished Bunnun, Andrew Calway and Walterio Mayol-Cuevas (2012). Real-time Learning and Detection of 3D Texture-less Objects: A Scalable Approach. British Machine Vision Conference (BMVC)

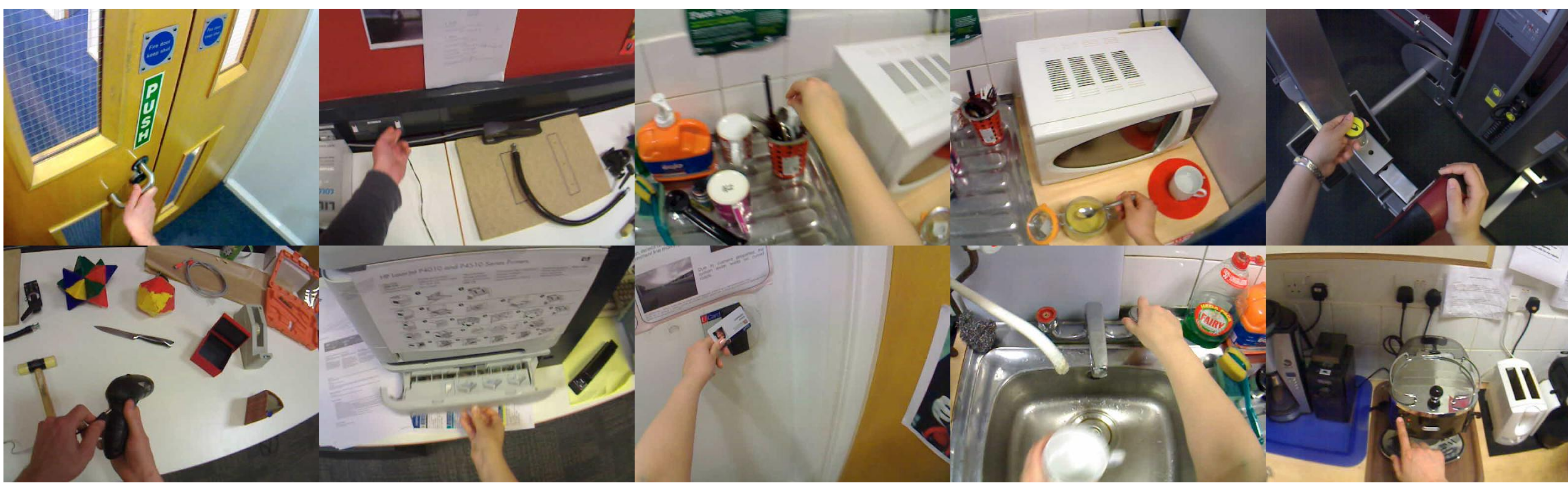
Results – Video Guides



Dataset

Bristol Egocentric Object Interactions Dataset

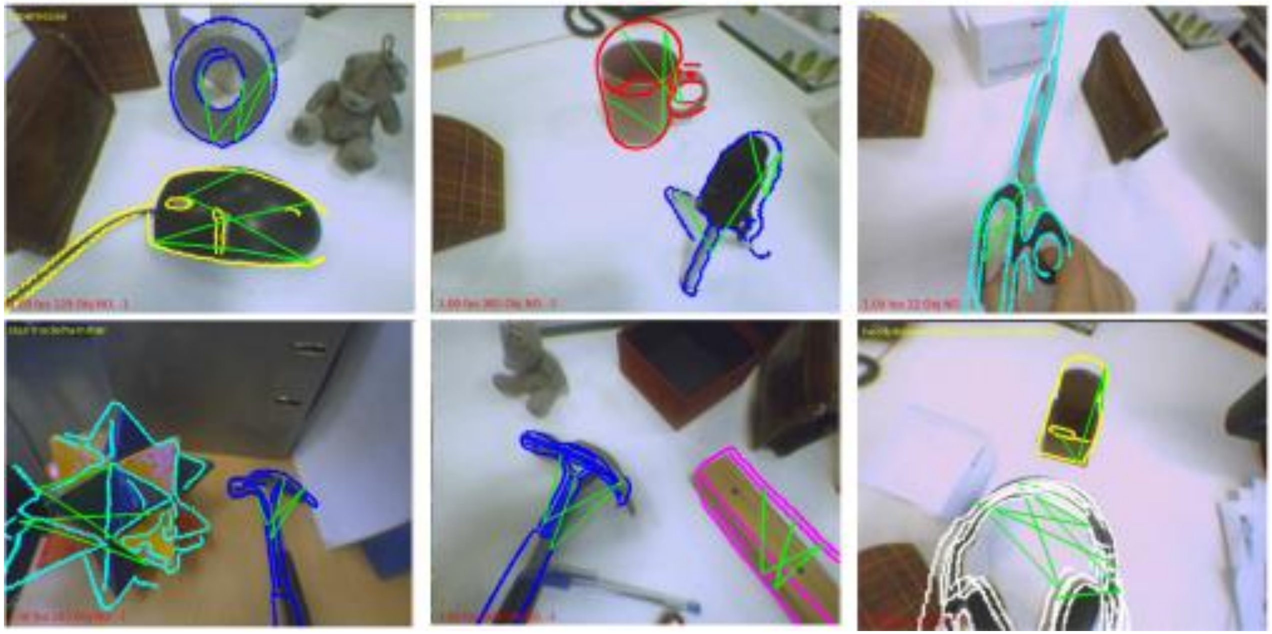
- Released (July 2014)
- wearable gaze tracker hardware (ASL Mobile Eye XG)
- 6 locations: kitchen, workspace, printer, corridor with a locked door, cardiac gym and weight-lifting machine
- 5 operators (2 sequences each) with narrations



On Glass

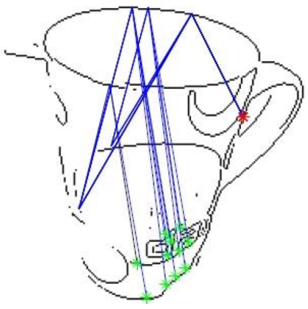
Code Released

- C++, tested on Ubuntu and works in ROS
- http://www.cs.bris.ac.uk/~damen/MultiObjDetector.htm



Summary

- Object detection via tractable edge configuration extraction. On-line training and amenable to mobile hardware.
  - Code at: http://www.cs.bris.ac.uk/~damen/MultiObjDetector.htm



- Egocentric object discovery and modes of interaction is achievable unsupervised and supported by our real-time in-situ texture-less object detector.