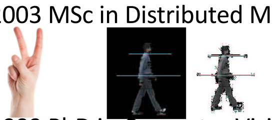
Egocentric Vision

Dr Dima Damen
Department of Computer Science

Short Bio

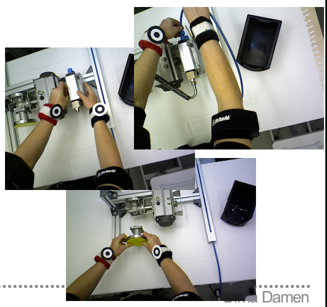
- 1998-2002 BSC in Computer Science
- 2002-2003 MSc in Distributed Multimedia Sys.



- 2006-2009 PhD in Computer Vision

Short Bio

- 2010-2012 Postdoc on EU-FP7 project



Short Bio

- 2013-2017 Assistant Prof in Computer Vision
- 2017- Associate Prof in Computer Vision

Egocentric Vision?

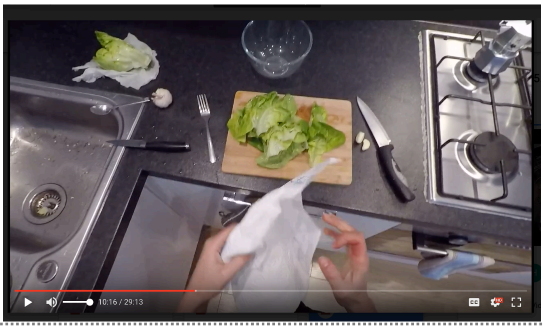
- Research interests: action and activity recognition
- Particularly centred around the perception of object interactions

Ego…

Ego… a person's sense of self-esteem or self-importance

Egocentric vision… the wearer serves as the central reference point in the study of interesting entities: objects, actions, interactions and intentions

Ego…



Visual Sensing – the landscape

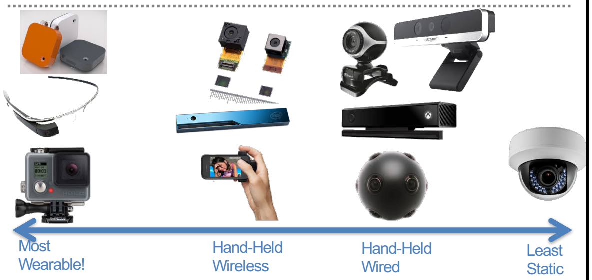
Visual Sensing – the landscape

Expensive

Visual Sensing – the landscape

Moveable

Visual Sensing – the landscape



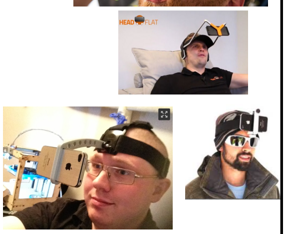
Wearable

Wearable?

Wearable?

Wearable?

Wearable?

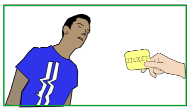
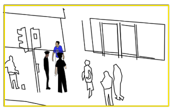


Wearable?

- Hat-Mounted
- Head-Mounted
- Glass-Mounted
- Shoulder-Mounted
- Chest-Mounted
- Wrist-Mounted
- Belt-Mounted
- Ankle-Mounted

But why do we care about… hardware???

- OPV (Ordinal-Person Views)
  - FPV (First-Person View)
  - SPV (Second-Person View)
  - TPV (Third-Person View)

See for yourself!

- Videos…

Top-Left: Fathi – Disney dataset for social interactions 2012
Top-Right Damen EPIC-Kitchen 2017
Bottom-Left Damen BEOID 2014
Bottom-Right Kitani (Youtube) Sports 2011

Conclusions?

- Just another camera?
- Just a shaking camera?

Egocentric Vision

- The Unique Problems
  1. Camera Motion
  2. Mapping and Localisation
  3. Attention and Task-Relevance
  4. Object Interactions
  5. Multi-view Solutions
- The Unique Applications
  1. Video Summarisation
  2. Skill Determination
  3. Real-time solutions

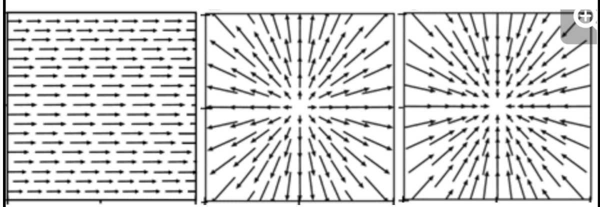
The Unique Problems

1. Camera Motion

1. Camera Motion

- Two types of motion
  - Egomotion
  - Foreground motion

Ego-motion



rotating approaching receding

Ego-motion

- Detect to:
  - Use?
  - Remove?

Hyperlapse

- https://youtu.be/sA4Za3Hv6ng

The Unique Problems

2. Mapping and Localisation

Mapping and Localisation

- https://youtu.be/ufBLu1VUQ-E

The Unique Problems

3. Attention and Task Relevance

Attention and Task Relevance

- What is attention?
  - Non-Egocentric Attention Models (→ Saliency)

DUT-OMRON dataset
http://saliencydetection.net/dut-omron

Attention and Task Relevance

- What is attention?
  - Non-Egocentric Attention Models (→ Saliency)

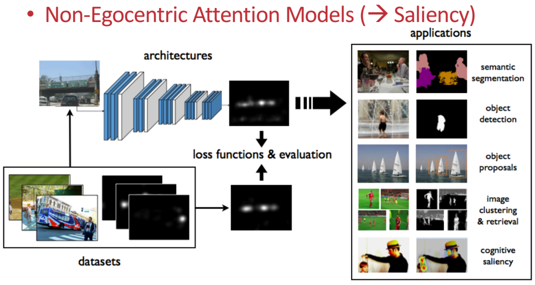
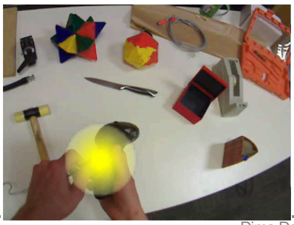


Figure: Zoya Bylinskii & Tilke Judd, ECCV 2016 Tutorial
http://saliency.mit.edu/ECCVTutorial/newDirectionsInSaliency.pdf

Attention and Task Relevance

Attention and Task Relevance

- Attention in egocentric vision
  - Foreground segmentation
  - Hand-region segmentation
  - Gaze tracking



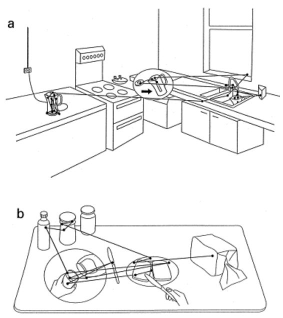
Quick introduction to human gaze

- Humans iterate between "fixations" and "saccades"
  - Fixation: short stops
  - Saccade: quick movements between fixations
- https://youtu.be/pknohrsz4Qs

Quick introduction to human gaze

Quick introduction to human gaze

Quick introduction to human gaze



Land and Hayhoe (2001) In what ways do eye movements contribute to everyday activities? Vision Research

Quick introduction to human gaze

- The notion of fixation/saccade has recently inspired attention models in vision

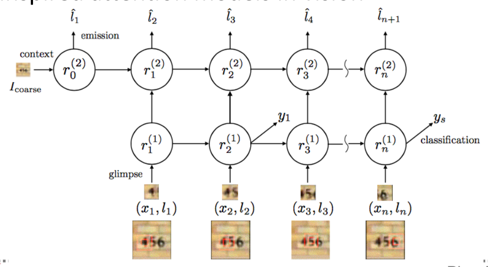


Figure: Lei Ba et al (2015). Multiple Object Recognition with Visual Attention. ICLR

The Unique Problems

3. Attention and Task Relevance

Case Study: You-Do, I-Learn

You-Do, I-Learn

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

- First-person view
- Offers a unique insight into 'used' or 'attended-to' objects
- How these objects have been used

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Try it yourself

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You-Do, I-Learn

- Q. How to 'ground-truth' objects that have been used?
- Q. How to 'ground-truth' how these objects have been used?

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

BEOID

- Ground-truth by written narration
- Released with dataset

pick the charger and plug it into the socket. Check that the screwdriver is powered by looking at the button. Pick the tape and place it in the box. Walk to the printer. Open the drawer to check the paper, and press keys on the printer pad. Use the card to unlock the door

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You Do, I Learn

- 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

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Definition

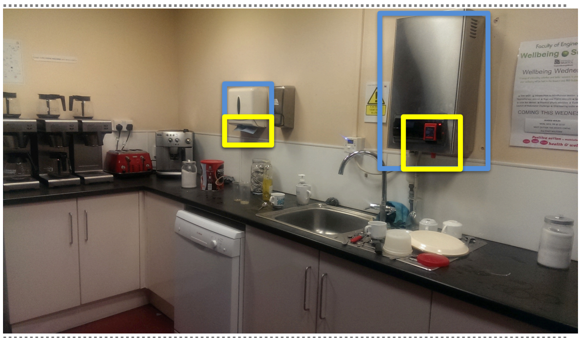
Task-Relevant Object (TRO)

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

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Which Objects?



Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects

- *Suggested* Problem Formulation…
  - Given a sequence of egocentric images $\{I_1, .., I_T\}$
  - Collected from multiple operators around a common environment
  - Automatically discover all task-relevant objects

$$\{O_k; 1 \leq k \leq K\}$$

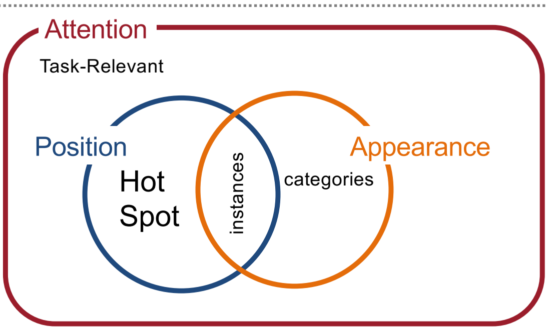
$$O_k = \{\Omega(I_t); 1 \leq t \leq T\}$$

  - *Assumption:* at most one task-relevant image part is present within each image

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

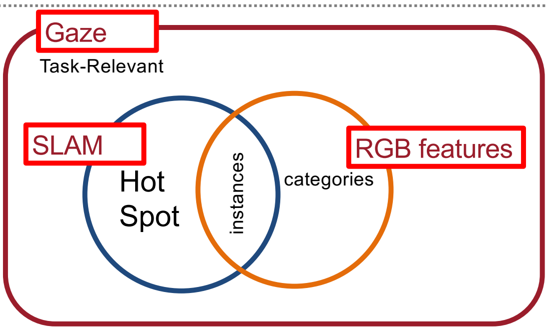
Discovering Task-Relevant Objects



Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects



University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 49

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering TROs

Discovering becomes a clustering task…

- Considers attention, position and appearance
- Unknown number of objects

University of BRISTOL

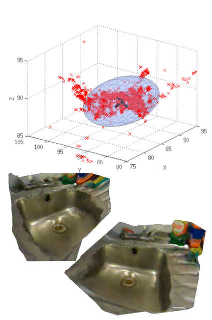
D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

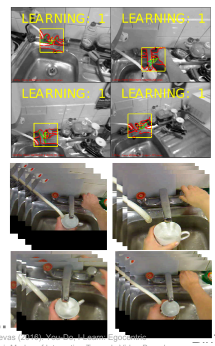
Dima Damen
5 July 2018 50

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects

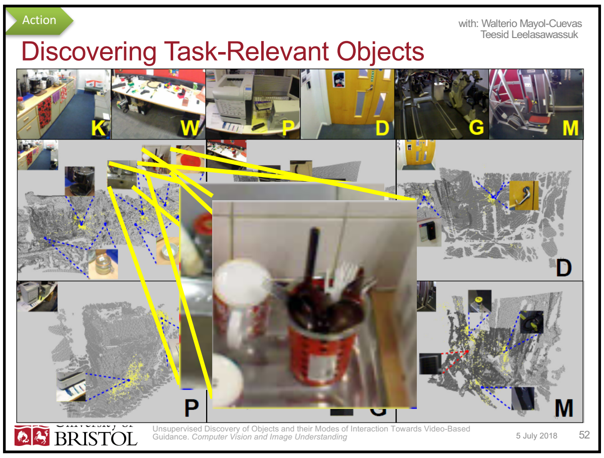


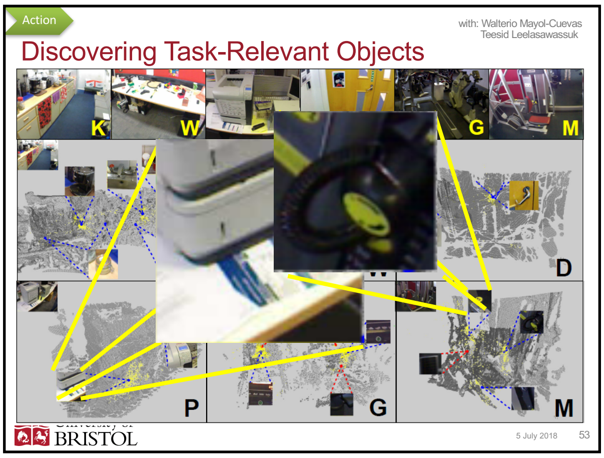


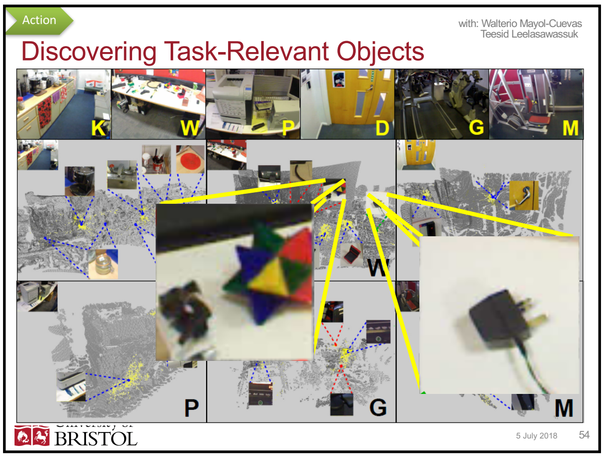
University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 51


Action
with: Walterio Mayol-Cuevas
Teesid Leelasawassuk
Discovering Task-Relevant Objects
K
W
P
D
G
M
D
P
M
BRISTOL
Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. Computer Vision and Image Understanding
5 July 2018
52

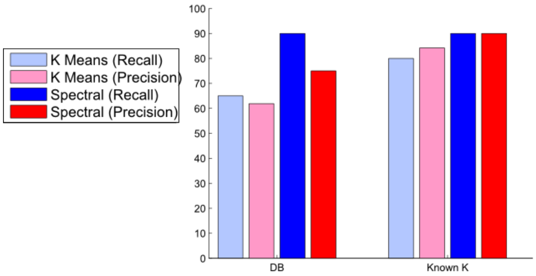

Action
with: Walterio Mayol-Cuevas
Teesid Leelasawassuk
Discovering Task-Relevant Objects
K
W
G
M
D
P
G
M
BRISTOL
5 July 2018
53


Action
with: Walterio Mayol-Cuevas
Teesid Leelasawassuk
Discovering Task-Relevant Objects
K
W
P
D
G
M
W
P
G
M
BRISTOL
5 July 2018
54

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects



University of BRISTOL

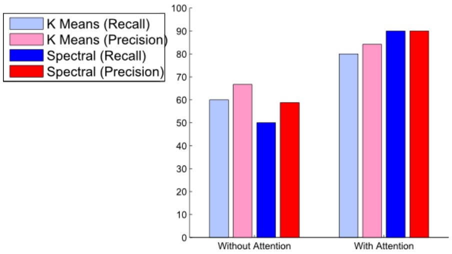
D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 55

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects



University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 56

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction

University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 57

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Definition

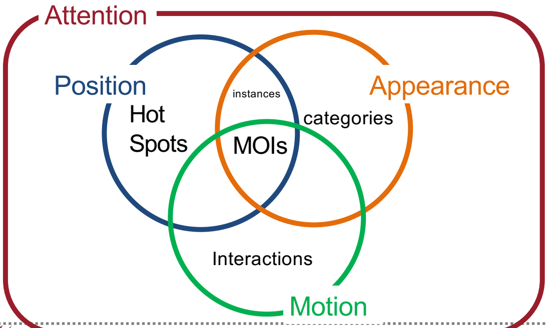
Modes of Interaction (MOI)

the different ways in which TROs are used

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction



Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

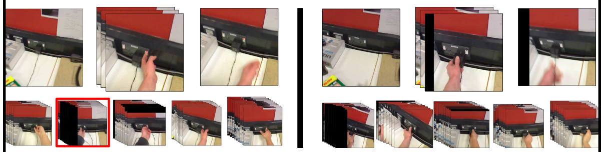
Discovering Modes of Interaction

- Motion
  - Video snippets for each discovered object
  - Descriptor per snippet
  - Clustering using DB-index

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction



University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 61

Action

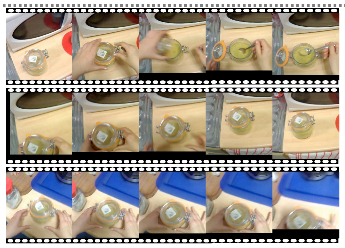
with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction

Open & get sugar

Put

Pick



Open door

University of BRISTOL

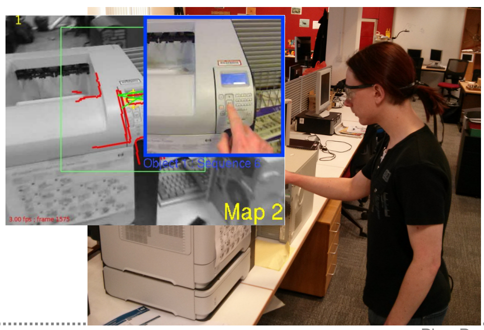
D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 62

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Back to…. the goal…



University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 63

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You Do, I Learn - Demonstration

University of BRISTOL

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Dima Damen
5 July 2018 64

More info…

Project You-Do, I-Learn

Video1 (2014), Video2 (2017)

Automated capture and delivery of assistive task guidance with an eyewear computer: The GlaciAR system. T Leelasawassuk, D Damen, W Mayol-Cuevas. Augmented Human, Mar 2017 pdf

You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video. D Damen, T Leelasawassuk, O Haines, A Calway, W Mayol-Cuevas. British Machine Vision Conference (BMVC), Sep 2014. PDF | Abstract | Dataset

Multi-user egocentric Online System for Unsupervised Assistance on Object Usage. D Damen, O Haines, T Leelasawassuk, A Calway, W Mayol-Cuevas. ECCV Workshop on Assistive Computer Vision and Robotics (ACVR), Sep 2014. PDF Preprint

Estimating Visual Attention from a Head Mounted IMU. T Leelasawassuk, D Damen, W Mayol-Cuevas. International Symposium on Wearable Computers (ISWC), Sep 2015. PDF

University of BRISTOL

Dima Damen
5 July 2018 65

The Unique Problems

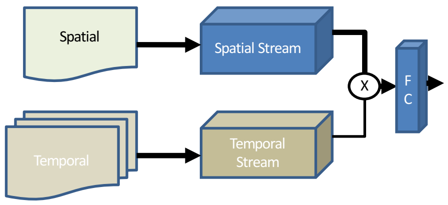
4. Object Interactions

University of BRISTOL

Dima Damen
5 July 2018 66

Action Recognition – an Introduction

- CNNs for Action Recognition
  1. Dual-Stream Neural Networks



Action Recognition – an Introduction

- CNNs for Action Recognition
  1. Dual-Stream Neural Networks

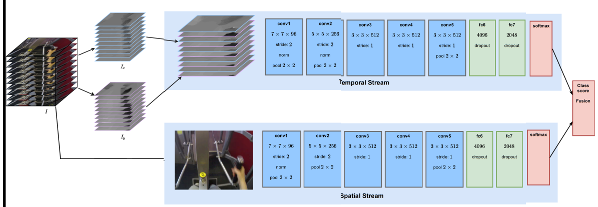


Figure by: Will Price, BSc Project, University of Bristol

Egocentric Action Recognition

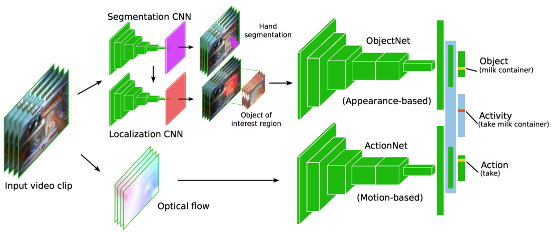
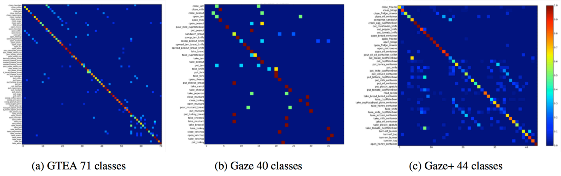
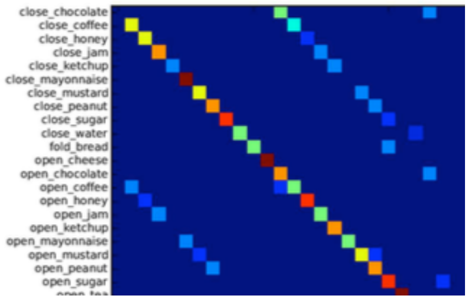


Figure from: Ma et al. Going Deeper into First-Person Activity Recognition. CVPR 2016

Egocentric Action Recognition

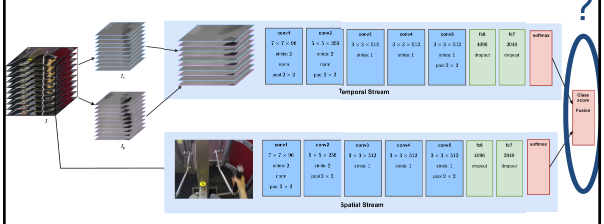


Egocentric Action Recognition



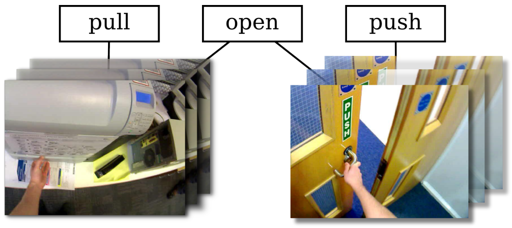
Action Recognition – an Introduction

- CNNs for Action Recognition
  1. Dual-Stream Neural Networks



with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas

Object Interactions – the Dilemma



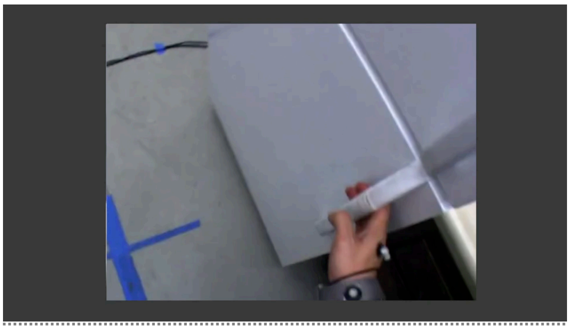
University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 73

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas

Object Interactions – the Dilemma



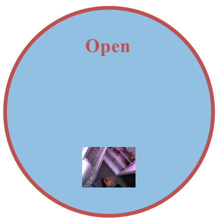
University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 74

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas

Object Interactions – the Dilemma



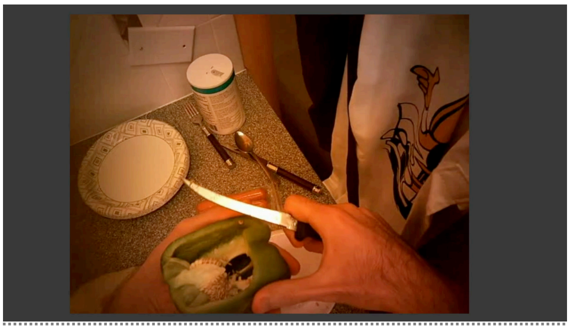
University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 75

Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



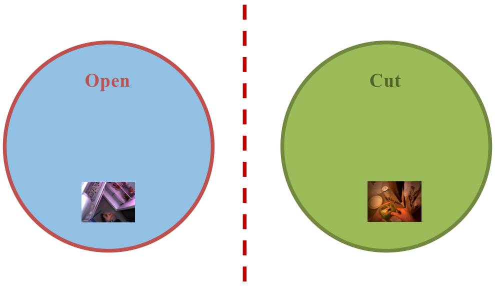
University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 76

Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



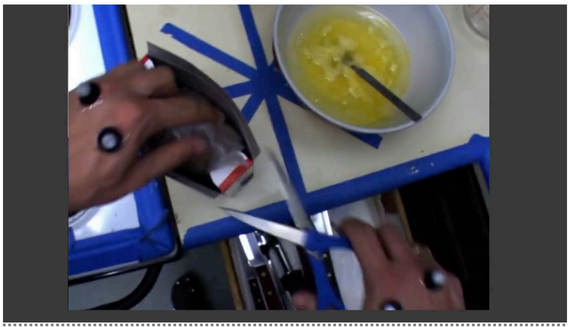
University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 77

Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



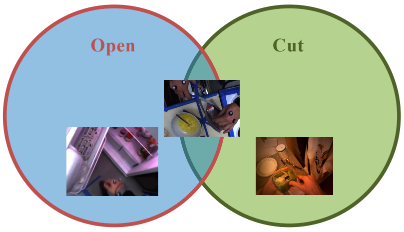
University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 78

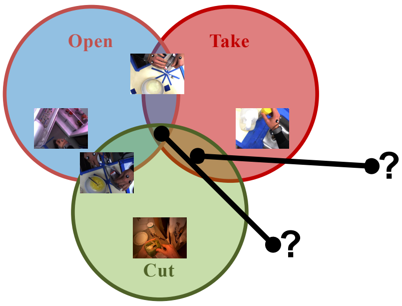
Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas

- Verbs cannot be separated into classes with hard boundaries.
- Rather the boundaries are more nuanced – what is correct in one video is incorrect for another.
- Singular classes are not enough.

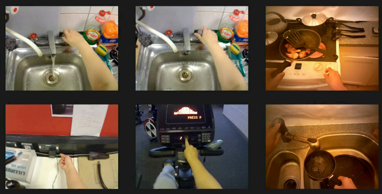
Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

Top 3 retrieved classes across all datasets.

Turn On/Off
Press
Rotate

Turn On/Off
Press
Rotate



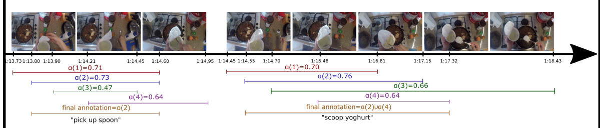
Labelling Method can differentiate turn On/Off tap by pressing and by rotating.

University of BRISTOL

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Dima Damen
5 July 2018 82

Temporal Boundaries for Object Interactions



- How robust are current state-of-the-art approaches to annotated boundaries in test segments?
- Modify test segment boundaries, maintaining significant overlap of segments IoU > 0.5
- **Correct in Green** – **Incorrect in Red**

University of BRISTOL

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

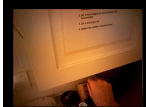
Dima Damen
5 July 2018 83

Trespassing the Boundaries

with: Davide Moltisanti
Michael Wray

GTEA Gaze+

ground truth



predicted class: take knife

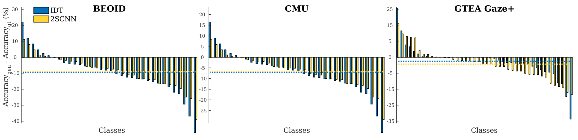
University of BRISTOL

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

Dima Damen
5 July 2018 84

with: Davide Moltisanti
Michael Wray

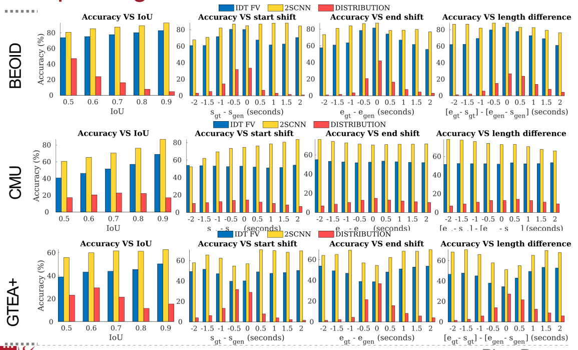
Trespassing the Boundaries



Action

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

Trespassing the Boundaries



Action

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

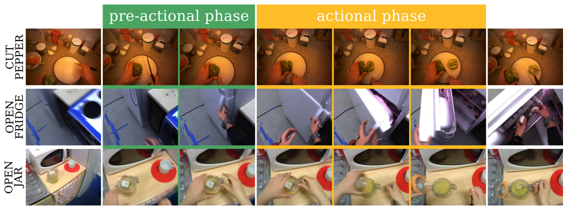
The Rubicon Boundaries

- Labelling approach proposal for temporally consistent annotations
- Decomposes an object interaction into two phases:
  - *pre-actional* phase
  - *actional* phase

Action

The Rubicon Boundaries

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas



Action

The Rubicon Boundaries

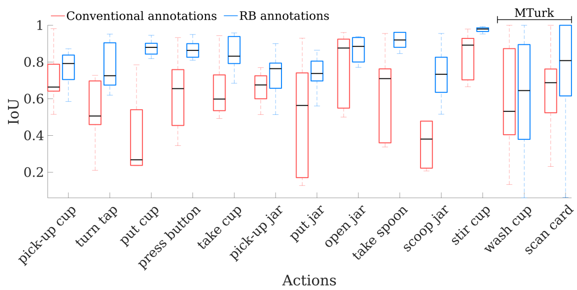
with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

Cut pepper (GTEA Gaze+)

Action

The Rubicon Boundaries

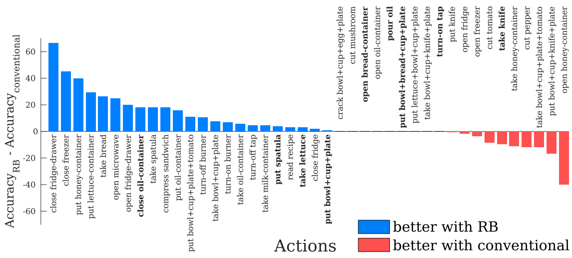
with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas



Action

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

The Rubicon Boundaries



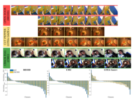
University of BRISTOL

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

Dima Damen
5 July 2018 91

More info…

Project Trespassing the Boundaries of Object Interactions



Video

Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. D Moltisanti, M Wray, W Mayol-Cuevas, D Damen. International Conference on Computer Vision (ICCV), 2017. pdf (camera ready) | arxiv

University of BRISTOL

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

Dima Damen
5 July 2018 92

with: Will Price

Visualising Learnt Models

- BEOID EBP videos:
- http://youtu.be/4cZS39c7IL0

University of BRISTOL

Dima Damen
5 July 2018 93

The Unique Problems

5. Multi-View Action Recognition

FPV with SPV

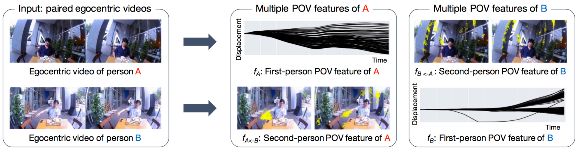


Figure from: Yonetani et al (2016). Recognizing Micro-Actions and Reactions From Paired Egocentric Videos. CVPR

FPV with TPV (top-view)

Figure from: Ardeshir and Borji (2016). Egocentric Meets Top-view CVPR

FPV with TPV (top-view)

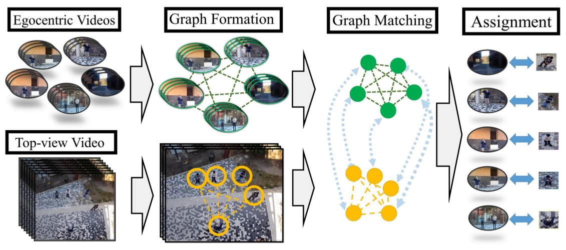


Figure from: Ardeshir and Borji (2016). Ego2Top: Matching Viewers in Egocentric and Top-view Videos. ECCV

Egocentric Vision

- The Unique Problems
  1. Camera Motion
  2. Mapping and Localisation (ref tomorrow's talk)
  3. Attention and Task-Relevance
  4. Object Interactions
  5. Multi-view Solutions
- The Unique Applications
  1. Video Summarisation
  2. Skill Determination
  3. Real-time solutions

The Unique Applications

1. Video Summarisation

Video Summarisation

- Fixations
- Highlight Detection

Egocentric Video Summarisation

- Object-Driven

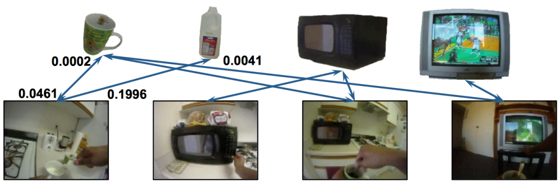


Figure from: Lu and Grauman (2013). Story-Driven Summarization for Egocentric Video. CVPR

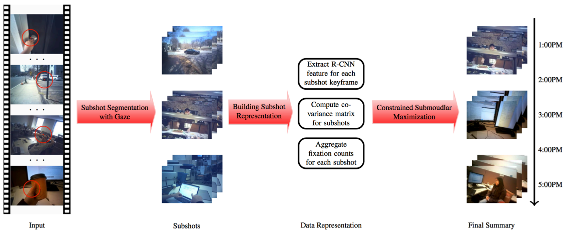
Egocentric Video Summarisation

- Object-Driven

Figure from: Lu and Grauman (2013). Story-Driven Summarization for Egocentric Video. CVPR

Egocentric Video Summarisation

- Fixation-Driven with Constraints



University of BRISTOL

Figure from: Xu et al (2015). Gaze-enabled Egocentric Video Summarization via Constrained Submodular Maximization . CVPR

Dima Damen

5 July 2018 103

Egocentric Video Summarisation

- Fixations from IMUs

University of BRISTOL

T Leelasawassuk, D Damen, W Mayol-Cuevas (2015). Estimating Visual Attention from a Head Mounted IMU. International Symposium on Wearable Computers (ISWC)

Dima Damen

5 July 2018 104

The Unique Applications

2. Skill Determination

University of BRISTOL

Dima Damen

5 July 2018 105

Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

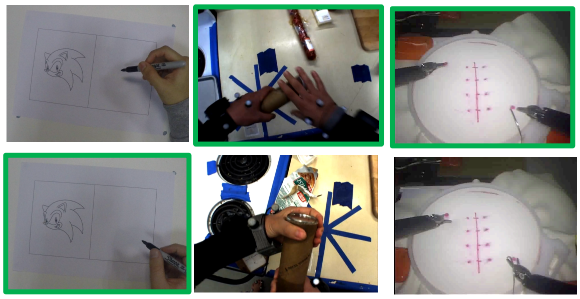
with: Hazel Doughty
Walterio Mayol-Cuevas

Assess relative skill for a collection of video sequences, applicable to a variety of tasks.

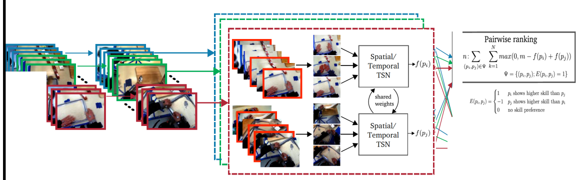
Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

Input: Pairwise annotations of videos, indicating higher skill or no skill preference



Who's Better? Who's Best? Skill Determination in Video using Deep Ranking



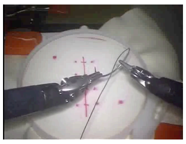
Video 1 ~ Video 2

Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

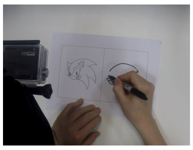
with: Hazel Doughty
Walterio Mayol-Cuevas

EPIC-SKILLS 2018

Surgery[1]



Drawing



Dough-Rolling[2]

Chopstick-Using

[1] Gao, Yixin, et al. "The JHU-ISI gesture and skill assessment dataset (JIGSAWS): A surgical activity working set for human motion modeling." Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2014.

[2] De la Torre, Fernando, et al. "Guide to the carnegie mellon university multimodal activity (CMU-MMAC) database." *Robotics Institute* (2008): 135.

University of BRISTOL

Dima Damen

5 July 2018 109

Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

$$L_{rank1} = \sum_{(p_i,p_j)\in\Psi} max(0, m - f(p_i) + f(p_j)) \quad (3)$$

$$L_{rank2} = \sum_{(p_i,p_j)\in\Psi} \sum_{k=1}^{N} max(0, m - f_k(p_i) + f_k(p_j)) \quad (5)$$

$$L_{sim} = \sum_{(p_i,p_j)\in\Phi} \sum_{k=1}^{N} max(0, |f(p_i) - f(p_j)| - m) \quad (7)$$

$$L_{rank3} = \beta L_{rank2} + (1 - \beta) L_{sim} \quad (8)$$

Method	Surgery			Dough-Rolling			Drawing			Chopstick-Using		
	S	T	TS	S	T	TS	S	T	TS	S	T	TS
Siamese TSN with margin loss	64.7	72.8	69.1	77.6	79.4	78.5	75.6	77.4	78.0	67.2	67.9	68.8
+ splits	64.4	**73.3**	69.0	79.1	**80.4**	78.5	74.9	81.8	79.1	67.2	69.9	68.8
+ similarity loss	**66.4**	72.5	**70.2**	**79.5**	79.5	**79.4**	**77.6**	**82.7**	**83.2**	**70.8**	**70.6**	**71.5**

University of BRISTOL

H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

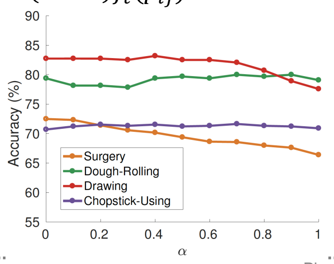
Dima Damen

110

Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

$$\frac{1}{\sigma}\sum_{j=1}^{\sigma} \alpha f_s(p_{ij}) + (1-\alpha) f_t(p_{ij})$$



University of BRISTOL

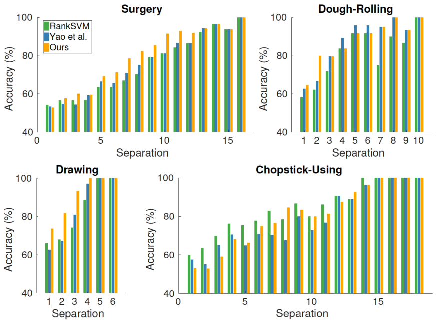
H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

Dima Damen

5 July 2018 111

Who’s Better? Who’s Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas



University of BRISTOL

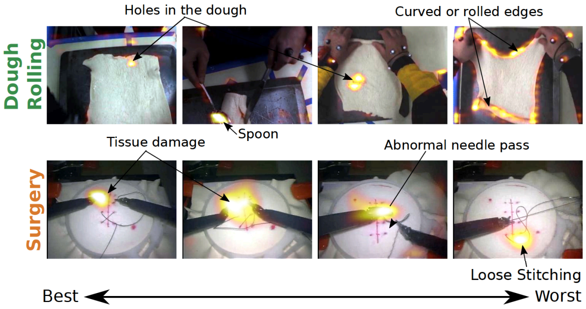
H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

Dima Damen

5 July 2018 112

Who’s Better? Who’s Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas



University of BRISTOL

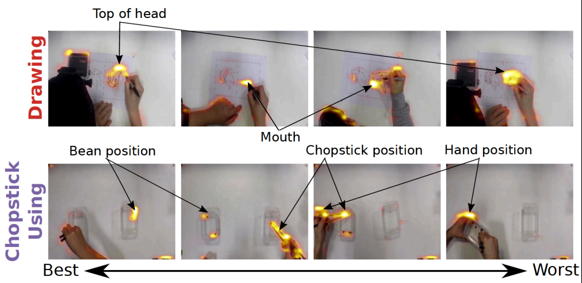
H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

Dima Damen

5 July 2018 113

Who’s Better? Who’s Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas



University of BRISTOL

H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

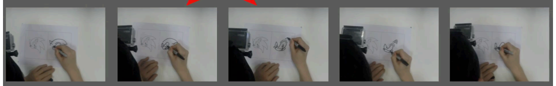
Dima Damen

5 July 2018 114

Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

Example Rankings



Sonic-Drawing task - part of new skill dataset

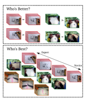
University of BRISTOL

H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

Dima Damen
5 July 2018 115

More info…

Project Who's Better, Who's Best: Skill Determination in Video



Video

Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. H Doughty, D Damen, W Mayol-Cuevas. CVPR (2018). PDF | arxiv

University of BRISTOL

Dima Damen
5 July 2018 116

The Unique Applications

3. Real-time Solutions

University of BRISTOL

Dima Damen
5 July 2018 117

Wearable (Systems)!

- On-the-cloud processing
- On-the-mobile processing
- Onboard processing!

Connecting-to-the-cloud

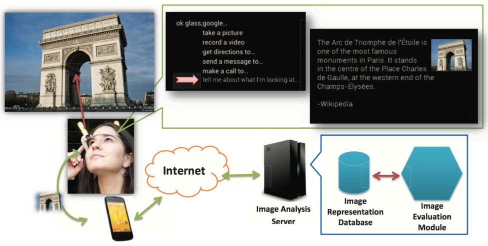


Figure 1. System overview. The user asks the device to inform her about her current view of Arc de Triomphe, and the system responds with the most relevant description in its database.

Altwaijry et al (2014). Recognizing locations with Google Glass: A case study. WACV

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You Do, I Learn – Google Glass Prototype

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

The need for large-scaled datasets…

University of BRISTOL

Dima Damen
5 July 2018 121

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

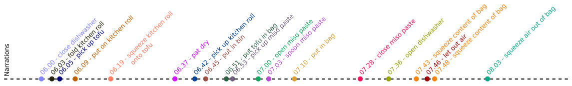
University of BRISTOL

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

Dima Damen
5 July 2018 122

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



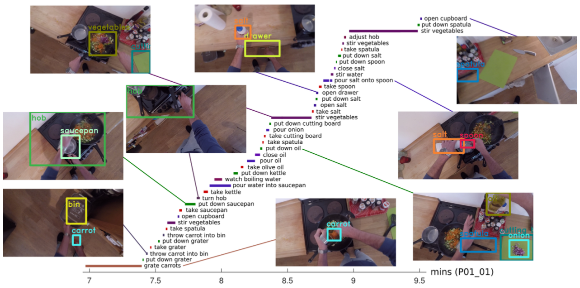
University of BRISTOL

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

Dima Damen
5 July 2018 123

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



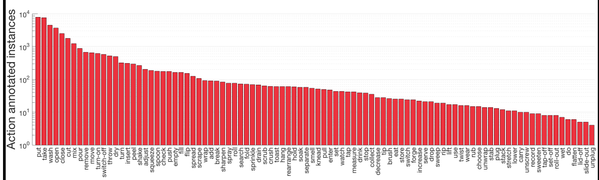
University of BRISTOL

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

Dima Damen

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



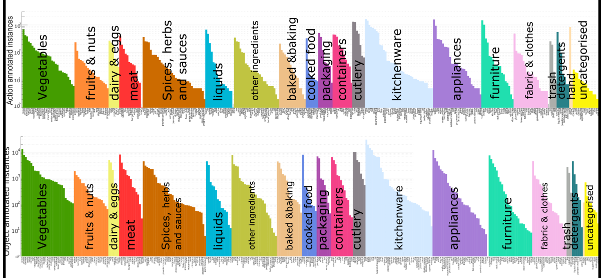
University of BRISTOL

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

Dima Damen

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



University of BRISTOL

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

Dima Damen

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

TABLE 1: Comparative overview of relevant datasets. *action classes with > 50 samples

Dataset	Ego?	Non-Scripted?	Native Env?	Year	Frames	Sequ-ences	Action Segments	Action Classes	Object BBs	Object Classes	Partici-pants	No. Env.s
EPIC-KITCHENS	✓	✓	✓	2018	11.5M	432	39,596	149*	454,158	323	32	32
EGTEA Gaze+ [19]	✓	×	×	2018	2.4M	86	10,325	106	0	0	32	1
BEOID [21]	✓	×	×	2014	0.1M	58	1,488	34	0	0	5	1
GTEA Gaze+ [20]	✓	×	×	2012	0.4M	35	3,371	42	0	0	13	1
ADL [23]	✓	×	✓	2012	1.0M	20	436	32	137,780	42	20	20
CMU [22]	✓	×	×	2009	0.2M	16	516	31	0	0	16	1
VLOG [15]	×	✓	✓	2017	37.2M	114K	0	0	0	0	10.7K	N/A
Charades [16]	×	×	✓	2016	7.4M	9,848	67,000	157	0	0	N/A	267
Breakfast [24]	×	✓	✓	2014	3.0M	433	3078	50	0	0	52	18
50 Salads [25]	×	×	×	2013	0.6M	50	2967	52	0	0	25	1
MPII Cooking 2 [26]	×	×	×	2012	2.9M	273	14,105	88	0	0	30	1

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

TABLE 4: Statistics of test splits: seen (S1) and unseen (S2) kitchens

	#Subjects	#Sequences	Duration (s)	%	Narrated Segments	Action Segments	Bounding Boxes
Train/Val	28	272	141731		28,588	28,561	326,298
S1 Test	28	106	39084	20%	8,069	8,064	97,865
S2 Test	4	54	13231	7%	2,939	2,939	29,995

		15 Most Frequent Object Classes															Totals		
	mAP	pan	plate	bowl	onion	tap	pot	knife	spoon	meat	food	potato	cup	pasta	cupboard	lid	few-shot	many-shot	all
S1	IoU > 0.05	74.00	72.61	71.50	60.72	84.44	69.97	44.03	40.93	29.65	58.52	62.82	53.30	78.39	51.95	62.77	9.71	49.80	38.23
	IoU > 0.5	67.60	66.21	65.98	39.96	73.80	64.71	28.80	23.89	20.75	49.85	55.48	42.99	69.75	29.20	58.48	6.98	36.50	28.06
	IoU > 0.75	21.94	44.60	39.48	3.52	25.83	19.67	3.42	2.59	5.27	15.78	13.18	8.00	24.53	4.05	26.51	0.36	8.73	6.50
S2	IoU > 0.05	75.94	87.36	72.72	47.61	78.14	75.92	55.51	41.28	71.59	38.61	N/A	44.62	80.58	53.88	58.40	6.00	51.71	40.61
	IoU > 0.5	62.88	84.86	68.61	32.18	59.75	62.86	39.60	27.52	53.54	35.47	N/A	39.19	76.27	32.54	49.36	5.32	36.27	28.57
	IoU > 0.75	14.56	62.82	38.44	2.25	4.89	14.91	3.85	1.51	9.56	8.10	N/A	7.60	43.30	5.61	25.48	0.18	9.05	7.04

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

TABLE 6: Baseline results for the action recognition challenge

		Top-1 Accuracy			Top-5 Accuracy			Avg Class Precision			Avg Class Recall		
		VERB	NOUN	ACTION	VERB	NOUN	ACTION	VERB	NOUN	ACTION	VERB	NOUN	ACTION
S1	RGB	45.25	35.78	18.91	**86.07**	**62.80**	**39.39**	54.94	40.41	07.01	**23.31**	**30.03**	**05.29**
	FLOW	43.27	17.92	09.10	79.89	39.63	21.91	**64.58**	24.51	01.52	15.35	09.72	01.28
	FUSION	**47.36**	**36.05**	**19.44**	84.27	61.05	35.45	63.12	**44.24**	**07.33**	21.95	29.25	05.22
S2	RGB	35.96	21.74	09.96	**74.70**	44.95	**24.59**	45.40	22.14	02.06	11.79	16.75	**01.91**
	FLOW	**40.56**	14.91	07.28	73.66	33.87	18.29	44.83	22.99	00.92	**14.16**	08.79	00.94
	FUSION	39.67	**22.33**	**10.84**	74.53	**45.23**	23.52	**59.60**	**23.65**	**02.09**	13.37	**16.84**	01.84

TABLE 7: Sample baseline action recognition per-class metrics (using fusion)

		15 Most Frequent Verb Classes														
		put	take	wash	open	close	cut	mix	pour	move	turn-on	remove	turn-off	throw	dry	peel
S1	RECALL	65.32	51.01	80.45	60.98	27.13	74.27	52.63	24.87	00.00	05.63	01.58	03.67	10.11	29.73	26.09
	PRECISION	35.62	41.24	63.17	72.67	72.46	69.38	69.52	66.20	-	53.33	66.67	50.00	56.25	88.00	54.55
S2	RECALL	64.16	48.03	87.76	42.06	15.10	45.69	35.85	06.06	00.00	00.00	00.81	00.00	00.00	00.00	00.00
	PRECISION	30.19	30.46	67.79	57.31	61.54	85.48	65.52	40.00	-	00.00	100.0	-	-	-	00.00

Interactive Conclusions

- Fill in the blanks:
  - Egocentric vision is ---------------------
  - Pick up an action (e.g. open door). Draw a sketch of how it looks like from FPV and TPV
  - The biggest challenge (in your opinion) in egocentric vision is ---------------------------------
  - The most interesting problem (to you) in egocentric vision is -----------------------------------

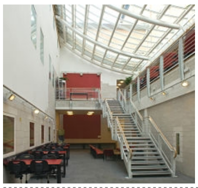
Interested in More?

- Egocentric Perception, Interaction and Computing (EPIC) Workshop Series
  - ECCV 2016 (Amsterdam)
  - ICCV 2017 (Venice)
  - ECCV 2018 (Munich)
    - Paper deadline: Tomorrow!
    - Abstract submission till 23rd of July (ongoing work)

Interested in More?

- Subscribe to the newly introduced mailing list: epic-community@bristol.ac.uk
- Instructions to subscribe:
  - send an email to: sympa@sympa.bristol.ac.uk
  - with the subject: **subscribe epic-community**
  - and blank message content

Bristol and University of Bristol



Bristol and University of Bristol

Thank you…

For further info, datasets, code, publications…

- http://www.cs.bris.ac.uk/~damen
- @dimadamen
- http://www.linkedin.com/in/dimadamen