Behaviour modelling and life logging

Research topic no. 10:

Aachen, Germany
05.05.2022

Wiktor Mucha

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091".

Universitat d'Alacant
Universidad de Alicante
Project Coordinator

RWTH AACHEN UNIVERSITY

Stockholm University

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

TU WIEN

Presentation agenda

- Introduction to the topic
- Motivation
- Available work
- Open problems
- Methodology
- Publications
- Presentation summary

Introduction to the topic

- ✓ Life logging - a technology that uses wearable sensors to gather, store, and process data from the daily lives of an individual.
- ✓ Using a wearable camera can illustrate in detail which activities the person wearing the camera has done.
- ✓ This thesis aims to create computer vision lifelogging systems for health improvement.

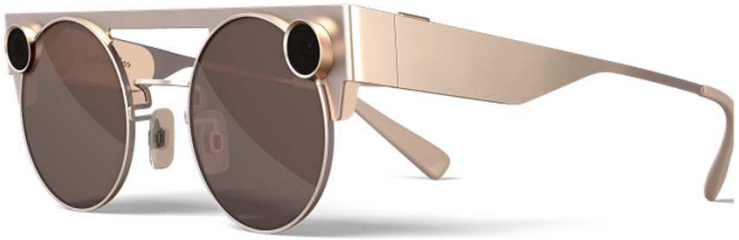
User wearing a lifelogging device[1]

[1] https://newatlas.com/narrative-clip-2/35422/

Devices for egocentric video life logging

Narrative Clip[2]

RayBan Stories[3]



Spectacles 3[4]

GoPro Hero[5]

[2] http://getnarrative.com
[3] https://www.ray-ban.com/canada/en/electronics/ray-ban%20stories%20%7C%20round-shiny%20black/8056597705035
[4] https://www.spectacles.com/at/shop/spectacles-3/
[5] https://gopro.com/de/at/shop/cameras/hero9-black/CHDHX-901-master.html

Motivation behind the project

- ✓ The aging population demands new technological solutions to reduce the involved medical personnel.
- ✓ The technological progress and miniaturization of electronic devices and sensors provide new systems for life logging, including cameras.
- ✓ Available work is focused in general on processing egocentric image streams or videos.
- ✓ This work will study the possibilities of egocentric visual data processing for health and lifestyle improvement.

Fields of egocentric images applications

- ✓ Temporal segmentation
- ✓ Action Recognition
- ✓ Routine discovery
- ✓ Food scenes classification
- ✓ Social interaction characterization

Temporal segmentation

Explanation

- ✓ Egocentric video and photo streams are large unstructured data.
- ✓ Dividing photo stream into moments/scenes.
- ✓ Information about Activities of Daily Living(ADL).

Examples:

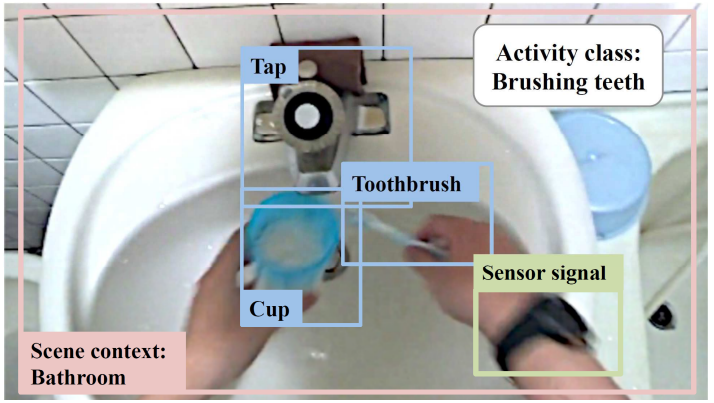
- ✓ Classifiers based on low-level features image representation[7].
- ✓ A combination of CNN and color histogram features to predict what a camera wearer does[6]
- ✓ Unsupervised clustering method based on agglomerative clustering[8]

[6] D. Castro, S. Hickson, V. Bettadapura, E. Thomaz, G. Abowd, H. Christensen, I. Essa, Predicting daily activities from egocentric images using deep learning, proceedings of the 2015 ACM International symposium 700 on Wearable Computers (2015)
[7] A. R. Doherty, A. F. Smeaton, Automatically segmenting lifelog data into events, Proceedings of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services (2008) 20–23.
[8] E. Talavera, M. Dimiccoli, M. Bola˜nos, M. Aghaei, P. Radeva, R-clustering for egocentric video segmentation, Lecture Notes in Computer Science 9117 (2015) 327–336.

Action recognition

Explanation

- ✓ Task of discovering action happening in the image.
- ✓ Object-based, motion-based, hybrid approaches.
- ✓ Detection of actions during the day allows describing the user's activity.



Example of a detected activitiy basing on objects[9]

Examples:

- ✓ Action recognition performance improvement by modeling the visual tempo of actions that characterize the dynamics and the temporal scale of action[10].

[9] P.-J. Hsieh, Y.-L. Lin, Y.-H. Chen, and W. Hsu, "Egocentric activity recognition by leveraging multiple mid-level representations," in 2016 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6, IEEE, 2016.
[10] C. Yang, Y. Xu, J. Shi, B. Dai, and B. Zhou, "Temporal pyramid network for action recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 591–600, 2020.

Routine discovery

Explanation

- ✓ Automated solution for routine discoveries of a camera wearer.
- ✓ Routines are sequences of actions that are happening regularly or at specific intervals.
- ✓ Several studies based on GPS modules with satisfactory results

Examples:

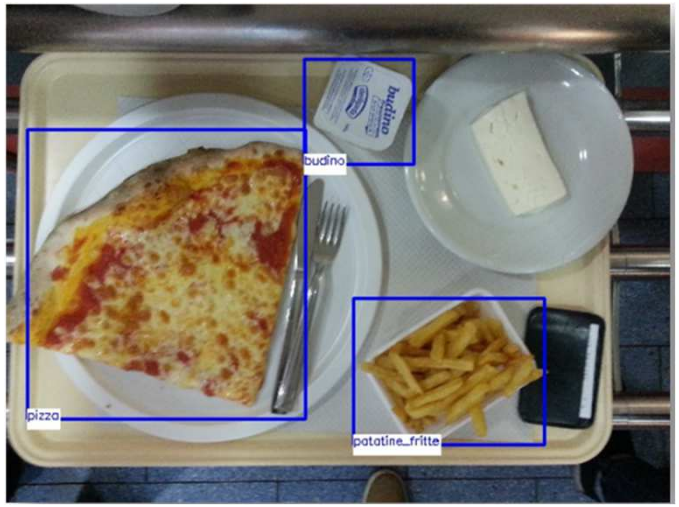
- ✓ Egocentric visual routines discovery into routine vs non-routine days[11].
- ✓ Employing GPS data sequences to cluster days with similar routines[12].

[11] E. Talavera, N. Petkov, P. Radeva, Unsupervised routine discovery in egocentric photo-streams, 18th Conference on Computer Analysis of Images and Patterns (2019).
[12] J. Biagioni, J. Krumm, Days of our lives: assessing day similarity from location traces, in: International Conference on User Modeling, Adaptation, and Personalization, Springer, 2013, pp. 89–101.

Food scene monitoring

Explanation:

- ✓ Detection and processing of scenes containing meals and food.
- ✓ The part of a healthy diet is food that is being eaten, ethics regarding the time on consumption and schedule, and places of food intake.
- ✓ By analyzing the environment where the camera wearers spend time, their behavior, habits, and even health can be described, studied, and improved.



Example of a food recognition[13]

Examples:

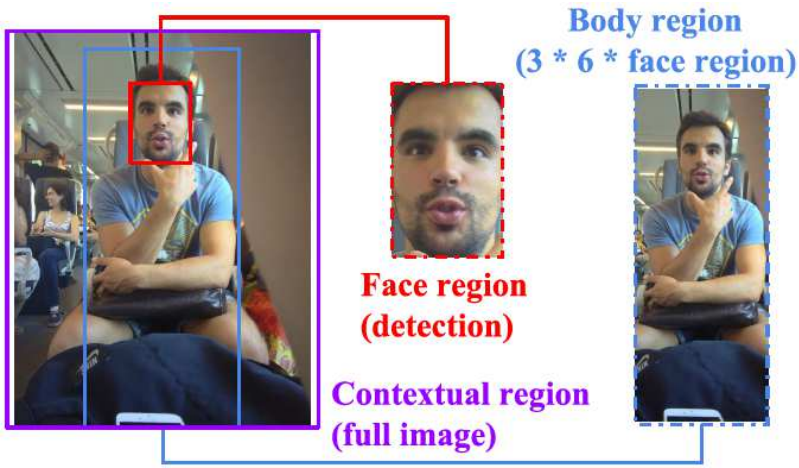
- ✓ Food Scenes classification with DL for food place classification[14].
- ✓ Hierarchical classification of food related scenes[15]
- ✓ Detecting food in egocentric images[16]

[13] E. Aguilar, B. Remeseiro, M. Bola˜nos, and P. Radeva, "Grab, pay, and eat: Semantic food detection for smart restaurants," IEEE Transactions on Multimedia, vol. 20, no. 12, pp. 3266–3275, 2018.
[14] M. K. Sarker, H. A. Rashwan, E. Talavera, S. Furruka Banu, P. Radeva, D. Puig, et al., Macnet: Multi-scale atrous convolution networks for food places classification in egocentric photo-streams, in: Proceedings of the European Conference on Computer Vision (ECCV).
[15] E. Talavera, M. Leyva-Vallina, M. Sarker, D. Puig, N. Petkov, P. Radeva, Hierarchical approach to classify food scenes in egocentric photo-streams, Journal of Biomedical and Health Informatics (2019).
[16]Wenyan Jia, Yuecheng Li, Ruowei Qu, Thomas Baranowski, Lora E Burke, Hong Zhang, Yicheng Bai, Juliet M Mancino, Guizhi Xu, Zhi-Hong Mao, et al. Automatic food detection in egocentric images using artificial intelligence technology. Public health nutrition, 22(7):1168–1179, 2019.

Social interaction characterization

Explanation:

- ✓ Automated analysis of egocentric images for the social interaction pattern descriptions.
- ✓ The faces of people from the captured scene are the primary source of information.
- ✓ Lack of social relations leads to health problems like inflammation, adolescence, hypertension, and decreasing psychological wellbeing[17].



The example of face, body and contextual regions extraction[21]

Examples:

- ✓ Multi-face tracking for first person captured photo streams[18].
- ✓ Classifications of social interactions into formal and informal type[19].
- ✓ Social characteristic of a person based on numerical statistics from interactions[20].

[17] Estefania Talavera, Nicolai Petkov, and Petia Radeva. Egocentric vision for lifestyle understanding. In Wearable Sensors, pages 415–433. Elsevier, 2021
[18] M. Aghaei, M. Dimiccoli, P. Radeva, All the people around me: face discovery in egocentric photo-streams, in: IEEE International Conference on Image Processing (ICIP), pp. 1342–1346.
[19] M. Aghaei, M. Dimiccoli, C. C. Ferrer, P. Radeva, Towards social pattern characterization in egocentric photo-streams, Computer Vision and Image Understanding 171(2018)104-117.
[20] E. Talavera, A. Cola, N. Petkov, P. Radeva, Towards egocentric person re-identification and social pattern analysis, 1st Applications of Intelligent Systems (APPIS) (2018).
[21] Emanuel Sanchez Aimar, Petia Radeva, and Mariella Dimiccoli. Social relation recognition in egocentric photostreams. In 2019 IEEE International Conference on Image Processing (ICIP), pages 3227–3231. IEEE, 2019.

Open problems

- ✓ What specific actions factoring subjects' health can be tracked throughout the day to improve their lifestyle and well-being?
  - ✓ Can we accurately model negative behaviors?
- ✓ Majority of works focuses on the food scenes. Is it possible to apply egocentric image life logs to determine the fluid intake habits of subjects?
  - ✓ Can we track the amount of taken fluids?
  - ✓ Can we track what kind of fluids is taken?
- ✓ What social signals can we detect in social events between people wearing a camera to improve their health?
  - ✓ What psychological diseases can we indicate in such data?
  - ✓ How can we employ facial and posture analysis?

Next steps

- ✓ Defining negative and positive behaviors for monitoring purposes
  - ✓ Hierarchical sorting of them in terms of difficulty level
  - ✓ Implementation plan including data planning
- ✓ Defining trackable social signals:
  - ✓ Data preparation
  - ✓ Creation of automatic system
  - ✓ Evaluation of the system
- ✓ Plan of fluid intake system:
  - ✓ Implementation plan
  - ✓ Data preparation with additional creation
  - ✓ Implementation and tests

Methodology

- ✓ Computer Vision techniques
  - ✓ Deep Learning models and AI
- ✓ Own data creation
- ✓ Employment of available datasets
- ✓ Interdisciplinary research and collaboration

Table of available datasets in egocentric vision.

Dataset name	Purpose	Type
EgoSocialStyle	Social	Images
EGO-GROUP	Social	Video
EgoRoutine	Routine	Images
EDUB-Obj	Segmentation, Obj. Detc.	Images
EGO4D	Social	Video
LBSoEV	Location	Video
OEVDLA	Action Recognition	Video
Home Action Genome	Action Recognition	Video
EDUB-Seg	Segmentation	Images
DoMSEV	Action Recognition	Video
GTEA	Action Recognition	Video

Publications

- **Depth and Thermal Images in Face Detection - A Detailed Comparison Between Image Modalities**

✓ Accepted: The 5th International Conference on Machine Vision and Applications

(ICMVA 2022), February 18-20 2022, Singapore,

- **Beyond Privacy of Depth Sensors in Active and Assisted Living Devices**

✓ Accepted**:** PETRA 2022 - Beyond Privacy of Depth Sensors in Active and Assisted Living Devices, Corfu, Greece

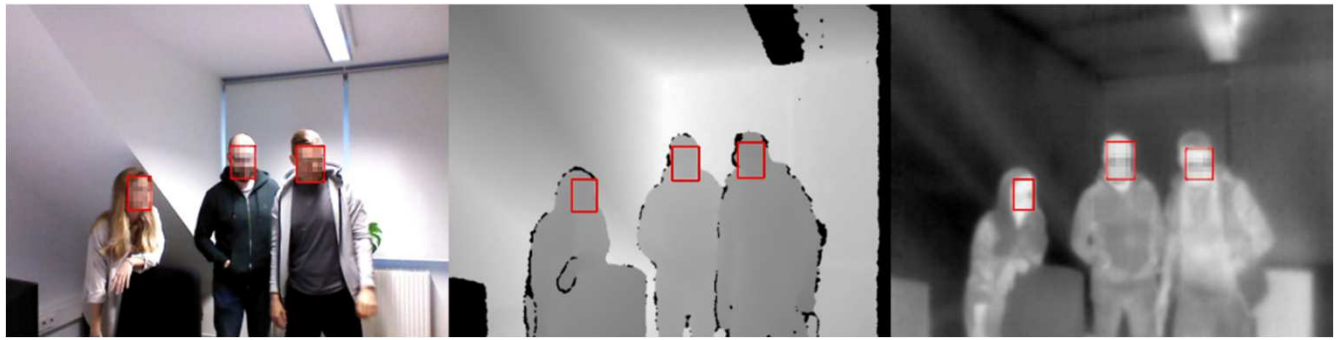
- **Addressing Privacy Concerns in Depth Sensors**

✓ Accepted**:** Joint International Conference on Digital Inclusion,

Assistive Technology & Accessibility - ICCHP-AAATE 2022, Lecco, Italy

Depth and Thermal Images in Face Detection - A Detailed Comparison Between Image Modalities

- ✓ Implementation of Face Detection (FD) model for three different image modalities (depth, thermal, RGB).
- ✓ Creating annotations for faces in existing datasets.
- ✓ A trimodal dataset for FD task - Contains perfectly aligned images of the same scene in depth, thermal and RGB.
- ✓ A detailed comparison of three image modalities in the example of FD task.

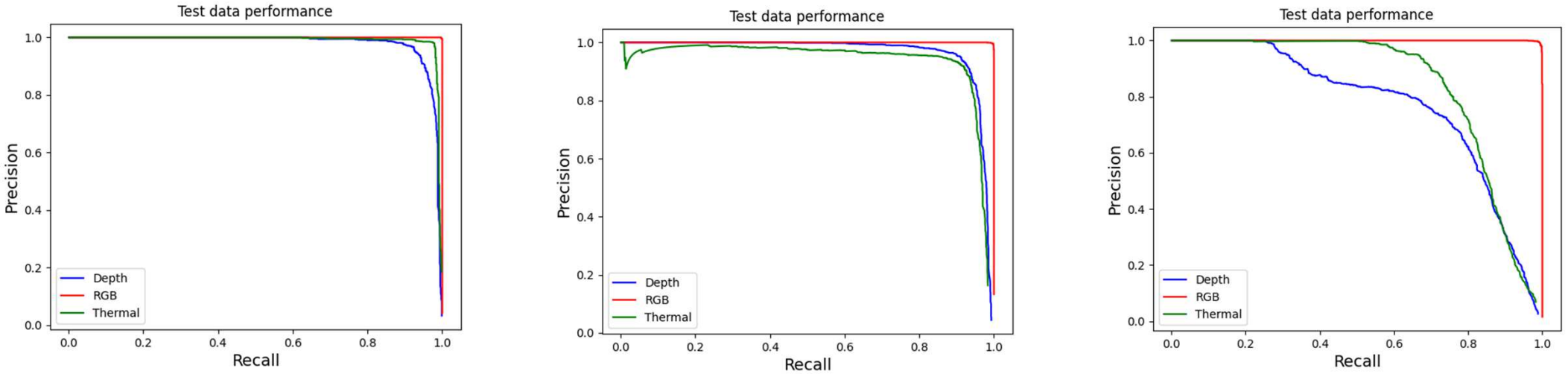


Example of a correct detection on each image modality, form left RGB, depth and thermal image

Depth and Thermal Images in Face Detection - A Detailed Comparison Between Image Modalities

Conclusions:

- ✓ FD is often a part of a pipeline in Active and Assisted Living (AAL) devices.
- ✓ Depth and thermal models are robust enough to be used in practice.
- ✓ The non-RGB models fail in the hardest conditions, where multiple faces are present in the scene, often occluded and in a distance from the camera.
- ✓ The RGB sensor is not always necessary.



Precision-recall curves for easy, medium and hard test scenarios

visuAAL

Addressing Privacy Concerns in Depth Sensors

- ✓ AAL vision-based systems raise privacy concerns due to monitoring patients
- ✓ Other types of images than RGB are used to maintain privacy. There are contradictory statements about privacy of depth cameras.
- ✓ The factors affecting privacy in depth images.
- ✓ Comparison of Face Recognition (FR) technique between RGB and depth images.

Same scene visible in deptn and RGB image.

Addressing Privacy Concerns in Depth Sensors

Conclusion:

- ✓ FR in depth is based on similar face features like in RGB image.
- ✓ A comparison favored RGB over depth with a significant accuracy difference.
- ✓ The high accuracy in depth FR was achieved in datasets with a low number of subjects.
- ✓ DL methods showed improvement in the accuracy of FR, even on small face resolutions.
- ✓ These circumstances make depth data more private than RGB, but when designing depth-based monitoring solutions, it is necessary to consider the possible disclosure of subjects' identities.

Face presented in depth and RGB modality

Presentation summary

- ✓ Life logging in computer vision and its benefits
- ✓ Fields of applying egocentric vision
- ✓ Open questiones and problems
- ✓ Future steps
- ✓ Carried out publications

Thank you!

Wiktor Mucha

TU Wien

wiktor.mucha@tuwien.ac.at

visuAAL