

Introduction to the special issue: Egocentric Vision and Lifelogging[☆]

Mariella Dimiccoli⁵

Gran via de les Corts Catalanes 585, 08007, Barcelona, Spain

Cathal Gurrin²

Glasnevin, Dublin 9, Ireland

David Crandall³

611 N Park Ave, USA

Xavier Giró-i-Nieto⁴

Jordi Girona 1-3, edifici D5 Campus Nord UPC 08034 Barcelona, Spain

Petia Radeva⁵

Gran via de les Corts Catalanes 585, 08007, Barcelona, Spain

Introduction

Advances in wearable technologies have driven growing interest in recording (or 'lifelogging') people's everyday activities and interactions, and in the challenges of gathering, organizing, analyzing, searching, and visualizing this rich multimedia and sensory content. In this context, wearable cameras that capture the wearer's everyday life from a first-person or "egocentric" perspective have created new research directions for computer vision and multimedia analysis, aiming to extract valuable semantic information from huge volumes of imagery

[☆]Fully documented templates are available in the elsarticle package on CTAN.

¹University of Barcelona and Computer Vision Center

²School of Computing, Dublin City University

³School of Informatics and Computing, Indiana University

⁴Technical University of Catalonia

⁵University of Barcelona and Computer Vision Center

and other first-person sensor data. Furthermore, egocentric vision and lifelog-
10 ging have inspired new research opportunities in other fields that require data
about human activities, including healthcare and wellness, safety and security,
psychology, etc. Enabling such applications requires egocentric and lifelogging
vision techniques to be robust and flexible to operate in the diverse, real-world
scenarios of everyday life.

15 The goal of this special issue is to present recent developments and appli-
cations of egocentric vision and lifelogging in general. The submitted papers
were rigorously peer-reviewed according to the guidelines and standards of JV-
CIR, and eight papers were accepted for publication. These papers cover ar-
eas related to image quality assessment, video summarization and description,
20 object-interaction prediction, multi-modal activity recognition, and temporal
segmentation based on personal location.

In their paper, “Image Quality Assessment in First-Person Videos,” Bai
and Reibman propose a new quality assessment measure specially conceived for
egocentric images. The measure is called Local Visual Information (LVI) and
25 primarily measures the relative blur between images, allowing them to select
the image with the best quality from a set of images. The key contribution is a
framework of mutual reference frame quality assessment for first-person vision
which measures the LVI score of each frame.

In “Next-Active-Object Prediction from Egocentric Videos,” Furnari et al.
30 address the problem of anticipating user-object interactions — i.e., predicting
the next object with which the user will interact. In particular, the authors train
a classifier to discriminate trajectories leading to an object interaction from
all others, and forecast next-active-objects by analyzing fixed-length trajectory
segments within a temporal sliding window.

35 Paper “Egocentric Video Description based on Temporally-Linked Sequences,”
by Bolaños et al., address the problem of egocentric day sequence captioning.
Specifically, the proposed model treats image sequences from different events as
temporally-linked units and exploits information from temporally neighboring
events for generating descriptions of the current event. A different approach for

40 sequence captioning is proposed in “DeepDiary: Lifelogging Image Captioning
and Summarization,” by Fan et al. The authors propose a MRF-based formula-
tion on top of more classical deep learning-based image captioning approaches.
Additionally, they demonstrate the utility of the proposed approach in practi-
cal user applications such as daily activity summarization, photo browsing, and
45 removal of photos with sensitive content.

A novel approach for video summarization is proposed in “Edited Near-
est Neighbour for Selecting Keyframe Summaries of Egocentric Videos,” by
Kuncheva et al. The authors propose to summarize egocentric videos with a set
of representative key-frames. Under the assumption that the video is already
50 segmented into events of interest (classes), the problem of selecting a keyframe
summary is cast as prototype (instance) selection for a nearest neighbor classifier
(1-nn).

In “Making a long story short: A Multi-Importance fast-forwarding ego-
centric videos with the emphasis on relevant objects,” Silva et al. propose a
55 fast-forward method for egocentric video that emphasizes segments in propor-
tion to their estimated relevance. They further propose a stabilization method
for fast-forward videos that ultimately provides a trade-off between smoothness
in visual flow and emphasis on the most relevant parts.

The problem of egocentric activity monitoring is addressed in “Multi-modal
60 Activity Recognition from Egocentric Vision, Semantic Enrichment and Lifelog-
ging Applications for the Care of Dementia,” by Meditskos et al. In particular,
the authors propose a multi-modal approach that combines visual and iner-
tial data acquired by wearable sensors through a fusion process. In addition,
they leverage Semantic Web technologies to build inter-operable activity graphs
65 for semantic activity representation and interpretation in a clinical intervention
setting.

Finally, in “Personal-Location-Based Temporal Segmentation of Egocentric
Video for Lifelogging Applications,” Furnari et al. propose to segment egocentric
videos according to the personal locations visited by the user. For this purpose,
70 they first classify each frame as one of the positive locations, then reject non-

personal locations and discourage spurious changes by using a Hidden Markov Model.

The papers of this Special Issue provide a snapshot of the current state of the art in egocentric vision and lifelogging. Still, there are plenty of new and
75 challenging problems to explore in this emerging field, such as egocentric vision for personal heritage and sport analysis, retrieval of personal memories, and first-person activity forecasting, as well as experiences from various application domains. We hope that the problems addressed and the techniques introduced in this Special Issue inspire readers with innovative and interesting ideas to
80 further advance the field.

ACKNOWLEDGEMENT

We express our sincere thanks to all of the people who have contributed their time and effort to the Special Issue. We thank all the authors who submitted and contributed their work to this special issue and worked under a tight sched-
85 ule. We also thank the reviewers of these manuscripts, who made many helpful suggestions for improving accepted papers. We also thank the editorial staff of JVCIR for their support. We hope that the papers selected in this special section will become useful resources for researchers and practitioners in egocentric vision, lifelogging and beyond.