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

Mariella Dimiccoli[5]

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

Cathal Gurrin[2]

Glasnevin, Dublin 9, Ireland

David Crandall[3]

611 N Park Ave, USA

Xavier Giró-i-Nieto[4]

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

Petia Radeva[5]

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.
[1]University of Barcelona and Computer Vision Center
[2]School of Computing, Dublin City University
[3]School of Informatics and Computing, Indiana University
[4]Technical University of Catalonia
[5]University of Barcelona and Computer Vision Center

and other first-person sensor data. Furthermore, egocentric vision and lifelog-
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.

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,
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
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.
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.

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

3

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

A novel approach for video summarization is proposed in "Edited Nearest 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 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 egocentric videos with the emphasis on relevant objects," Silva et al. propose a fast-forward method for egocentric video that emphasizes segments in proportion 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 Activity Recognition from Egocentric Vision, Semantic Enrichment and Lifelogging Applications for the Care of Dementia," by Meditskos et al. In particular, the authors propose a multi-modal approach that combines visual and inertial data acquired by wearable sensors through a fusion process. In addition, they leverage Semantic Web technologies to build inter-operable activity graphs 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, 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 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 further advance the field.

ACKNOWLEDGEMENT