







Spin-off of the University of Catania



Procedure Understanding from Egocentric Videos

Giovanni Maria Farinella

FPV@Image Processing Laboratory - http://iplab.dmi.unict.it/fpv

Next Vision - http://www.nextvisionlab.it/

Department of Mathematics and Computer Science - University of Catania

giovanni.farinella@unict.it - www.dmi.unict.it/farinella

After personal computers and smartphones, wearable computing is the **third wave of computing – Always on AI Personal Assistant**



Personal Computers:

computing for the mass, but not mobile and not context aware dedicated access to computing





Smartphones: mobile computing is always accessible, but forces to switch between the real and virtual world

Eyeworn Devices: computing everywere with minimal switch between real and virtual world

Università di Catania Egocentric (First Person) Vision



Egocentric vision or first-person vision is a sub-field of computer vision that entails analyzing images, videos captured by a wearable device, which is typically worn on the head or on the chest and naturally approximates the visual field of the camera wearer.

Consequently, visual data capture the part of the scene on which the user focuses to carry out the task at hand and offer a valuable perspective to understand the user's intents and activities and their context in a naturalistic setting.

This research area is sometimes referred with the name "Wearable Vision".



Università di Catania What will the future AI be? We wonder!





Imagine to wear an AI-powered device able to observe the scene from your point of view to understand what you are doing and to support you in every day life where you live and work

Università di Catania What do we need? What should be the wearable device?

An AI-Powered Wearable Virtual Assistant



A wearable device which perceives the world from our "egocentric" point of view is ideal for implementing an AI-Powered Virtual Assistant Università di Catania Success Cases - Three Examples on the Market

First Person (Egocentric) Vision Devices







OrCam MyEye Available since 2015 Health, Assistive Technologies

Microsoft HoloLens Available since 2016 Industry, Mixed Reality

Apple Vision Pro Pre-order since 2024 Health, Mixed Reality

Università What's Relevant in Egovision?

An Outlook into the Future of Egocentric Vision Sam is finally home Chiara Plizzari^{*} · Gabriele Goletto^{*} · Antonino Furnari^{*} · after a long day. EgoAI kept track of Siddhant Bansal^{*} · Francesco Ragusa^{*} · Giovanni Maria Farinella[†] · m's food intake Dima Damen^{\dagger} · Tatiana Tommasi^{\dagger} mato sou Politecnico di Torino 💈 Università di Catania EgoAI suggest half-day visit t Egyptian mus "aire feels orted Abstract What will the future be? We wonder! A 3D projection 0 2.2In this survey, we explore the gap between current reof Remy helps Audible with cooking 3D projection search in egocentric vision and the ever-anticipated fu-ture, where wearable computing, with outward facing 2.5 EGO-Designer cameras and digital overlays, is expected to be integrated in our every day lives. To understand this gap, the article starts by envisaging the future through character-based stories, showcasing through examples the limitations of current technology. We then provide a mapping between this future and previously defined research tasks. For 4.5 Gaze Understanding and Prediction 23 Cleopatra discovers : lunch while also enlig each task, we survey its seminal works, current state-history behind varie EgoAI has reserved an of-the-art methodologies and available datasets, then hitting the she 4.8 Hand and Hand-Object Interactions 30 afternoon at the therlook and sound natural mal baths. The next reflect on shortcomings that limit its applicability to bus is scheduled to future research. Note that this survey focuses on soft-arrive in 20 minutes ware models for egocentric vision, independent of any After dinner, Sam enjoys a group card specific hardware. The paper concludes with recommengame with his friends, who are connected dations for areas of immediate explorations so as to through their own Conclusion 49 EgoAI unlock our path to the future always-on, personalised EgoAI suggests Claire and life-enhancing egocentric vision. A NO a proper Italian coffee Keywords Egocentric Vision, Future, Survey,

Keywords Egocentric Vision, Future, Survey, Localisation, Scene Understanding, Recognition,

Anticipation, Gaze Prediction, Social Understanding,

C. Plizzari, G. Goletto and T. Tommasi, Politecnico di Torino,

Italy · A. Furnari, F. Ragusa and G. M. Farinella, University

of Catania, Italy · S. Bansal and D. Damen, University of

Bristol, UK. E-mail: Tatiana.Tommasi@polito.it

Body Pose Estimation, Hand and Hand-Object

Dialogue, Privacy

[†]: Equal Senior Author

*: Equal Contribution/First Author

Contents

Interaction, Person Identification, Summarisation,

Designing and building tools able to support human activities, improve quality of life, and enhance individuals' abilities to achieve their goals is the ever-lasting aspiration of our species. Among all inventions, digital computing has already had a revolutionary effect on human history. Of particular note is mobile technology, currently integrated in our lives through hand-held devices, i.e. mobile smart phones. These are nowadays the de facto for outdoor navigation, capturing static and moving footage of our everyday and connecting us to both familiar and novel connections and experiences.

However, humans have been dreaming about the next-version of such mobile technology — wearable computing, for a considerable amount of time. Imaginations



Plizzari, C., Goletto, G., Furnari, A., Bansal, S., Ragusa, F., Farinella, G. M., Damen., D. & Tommasi, T. (2023). An Outlook into the Future of Egocentric Vision. International Journal of Computer Vision.

Università di Catania AI for Procedure Understanding

Procedural activities are sequences of key-steps aimed at achieving specific goals.





Task: Given a video segment s_t and its previous video segment history, models have to: 1) determine previous keysteps (to be performed before s_t); infer if s_t is 2) optional or 3) a procedural mistake; 4) predict missing keysteps (should have been performed before s_t but were not); and 5) next keysteps (for which dependencies are satisfied). Università di Catania AI for Procedure Understanding – Some Needed Modules



Localization of Humans and Navigation | AI Aware of Context



(2)NAOMI Left Hand Human is interacting with X Human-Object Interaction | AI Aware of Objects 4 predicted trajector Gaze Completion Model Correct Action Anticipation and Mistake Detection | AI Aware of Risks

See UNICT publications at http://iplab.dmi.unict.it/fpv





Università Università NAOMI: Human-Object Interaction | AI Aware of Objects

Università NAOMI: Human-Object Interaction | AI Aware of Objects

QUADRO ELETTRICO

SOTTO TENSIONE

E VIETATO

 Baseguire lawoot was impained, acting tempores
Boocase gli impaneti se cost si il autorizzati
Boggiare (ciperi a la costocita di conservza prin di auso fonto lamanoni



UNIVERSIT

IMAGE PROCESSING LABORATOR

Università NAOMI: Human-Object Interaction | AI Aware of Objects





HERO is a Vision & Language Conversational Agent Wearable version in coming soon...

Understand immediately when the user makes a mistake to assist them



Università di Catania Action Anticipation | AI Aware of Risks



Anticipated Actions (in 0.50s)

WASH TABLE SPRAY LIQUID:WASHING TAKE SHEETS MOVE BOTTLE PUT LIQUID:WASHING PUT SHEETS WASH TOP OPEN TAP CLOSE CUPBOARD TAKE BAG WASH SINK MOVE BREAD

Input for $\tau_a = 0.50s$





Università Online Mistake Detection | AI Aware of Risks



Università di Catania Procedure Understanding from Egocentric Videos



Take home messages

Procedure Understanding opens **many research challenges** as well as the **new and useful applications in different domains** (e.g. Ego-Worker, Ego-Home, ...); Università di Catania AI for Procedure Understanding – Some Needed Modules



Localization of Humans and Navigation | AI Aware of Context



Human-Object Interaction | AI Aware of Objects

See UNICT publications at http://iplab.dmi.unict.it/fpv



Conversational AI Agents | AI Aware of Human Needs



Action Anticipation and Mistake Detection | AI Aware of Risks

Università di Catania Procedure Understanding from Egocentric Videos



Despite promising attempts there are <u>still a lot of</u> progress to be done to mix real and virtual worlds to support humans where they live and work.











Thank you for your attention

Procedure Understanding from Egocentric Videos

Giovanni Maria Farinella

FPV@Image Processing Laboratory - http://iplab.dmi.unict.it/fpv

Next Vision - http://www.nextvisionlab.it/

Department of Mathematics and Computer Science - University of Catania

giovanni.farinella@unict.it - www.dmi.unict.it/farinella