
# Enabling Ego Vision Applications on Smart Eyewear Devices

Francesca Palermo

Research Principal Investigator

EssilorLuxottica

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

- The content of this presentation is proprietary and confidential information of EssilorLuxottica Group and shall not be reproduced or distributed to any third party except for the recipient.

- Please be aware that the information included herein is for informative purposes only, protected by copyright and/or other third party's intellectual property and it is not intended to be copied or reproduced in whole or in part.

# Outline

1. Introduction to Ego Vision for Smart Eyewear
   - What is Ego Vision
   - Challenges of Embedded Vision
   - Benefits of Ego Vision in Smart Eyewear
   - Key Applications of Ego Vision
2. Ego Vision Functionalities: Ego Action, Human Pose Estimation, SLAM
3. Conclusions and Future Work

# Introduction to Ego Vision – What is Ego Vision?

# Introduction to Ego Vision – Benefits of Ego Vision in Smart Eyewear

**Hands-free functionality enhances usability**

- Seamless user interaction
- Improved mobility
- Natural experience

**Real-time context-aware insights for decision-making**

- Situational awareness
- Personalized information
- Actionable alerts

**Enhances accessibility and inclusivity**

- Empowering the visually impaired
- Support for elderly users
- Breaking language barriers

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Introduction to Ego Vision – Challenges of Embedded Vision

**Hardware constraints**

- Limited power and storage
- Small form factor
- Battery life constraints

**Real-time processing**

- Real-time performance
- Balancing latency and computation

**Privacy and security**

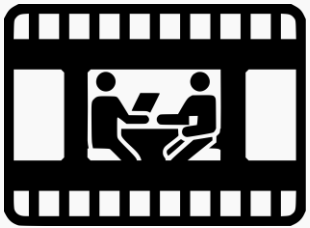
- Sensitive data
- Ensuring privacy
- Secure data transmission

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Introduction to Ego Vision – Key Applications

**Scene understanding**

- Interpret complex scenes
- Detect user context
- Generate audio-visual summaries



**Localization and mapping**

- Real-time map of surroundings
- Outdoor and indoor navigation

**Activity recognition**

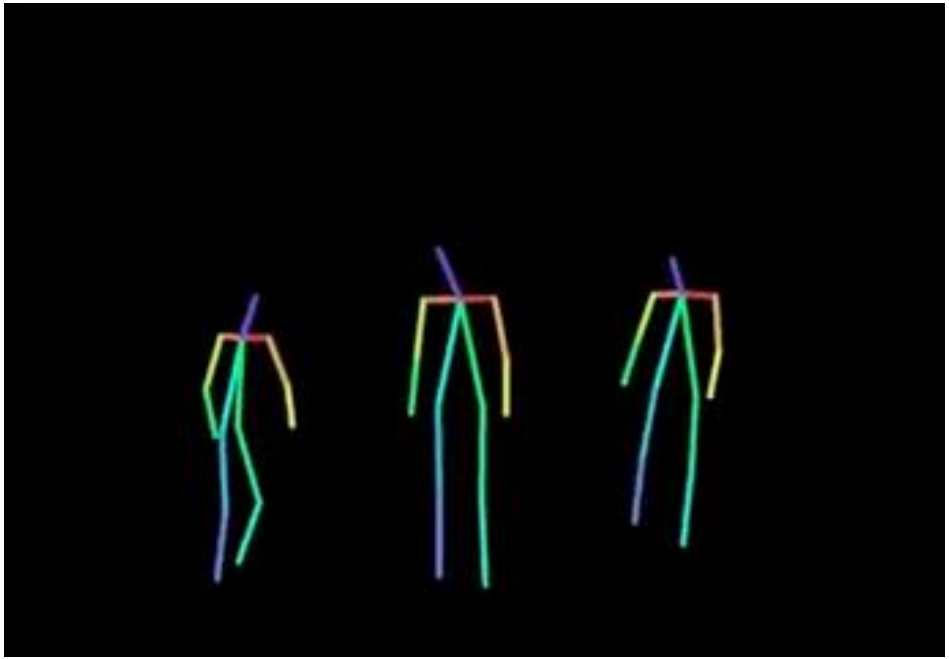
- Understand user's actions
- Recognize hand gestures for interface control

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Ego Vision Functionalities – Human Pose Estimation

Detecting people's body movements to enhance user interaction and contextual awareness

**High-Precision Algorithms:**
Multi-person pose detection optimized for mobile deployment [1, 2]
**Benchmarks Achieved:** 49% mAP on mobile, 60% mAP on edge
**Cons:** Missing real-time (30 FPS) on edge device (49 FPS on mobile, 11 FPS on edge)



https://rf-action.csail.mit.edu/

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Ego Vision Functionalities – Human Pose Estimation

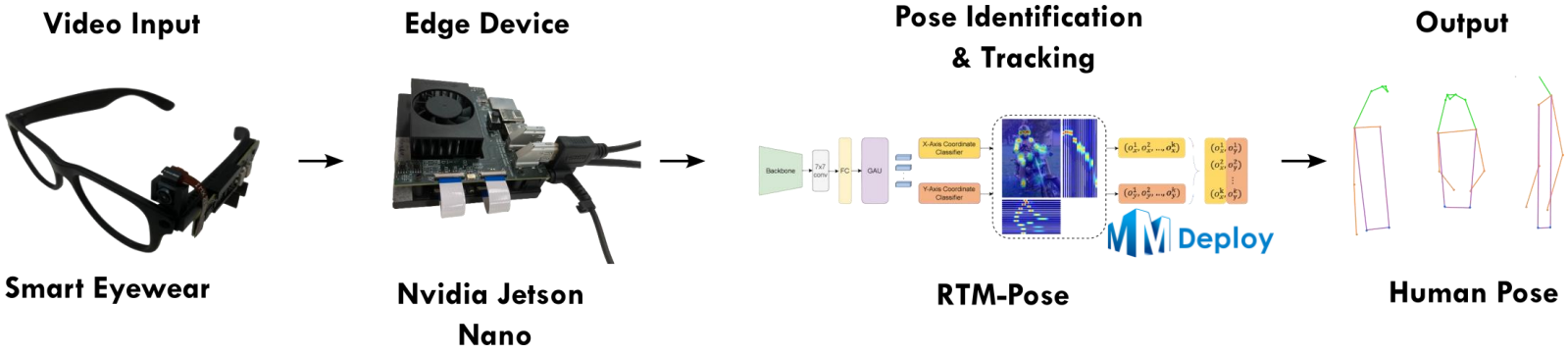
**Limitations**

RTMPose on Nvidia Jetson Nano reaches 10 fps, insufficient for real-time processing

**Improvements**

Converted RTMPose PyTorch model into C language with MMDeploy framework:

- Achieved over 30 fps on the Jetson Nano
- Process up to ten people simultaneously




H. Quan et al., Evaluating Human Pose Estimation Algorithms for Resource-Constrained Smart Eyewear Device, ECCV WS (2024)

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Ego Vision Functionalities – Human Pose Estimation

| Model | Yolo Pose 8 (n) | RTM Pose | RTM Pose Opt C |
|---|---|---|---|
| mAP % | 50.5 | 64.4 | **64.4** |
| Hz | 11 | 20 | **30** |

Developed model for Human Pose Estimation on smart eyewear achieving real-time speed (30 FPS)

**Future work:**
Human Action Recognition

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Ego Vision Functionalities – Ego Action Recognition

Understanding and interpreting the user's actions from a first-person perspective

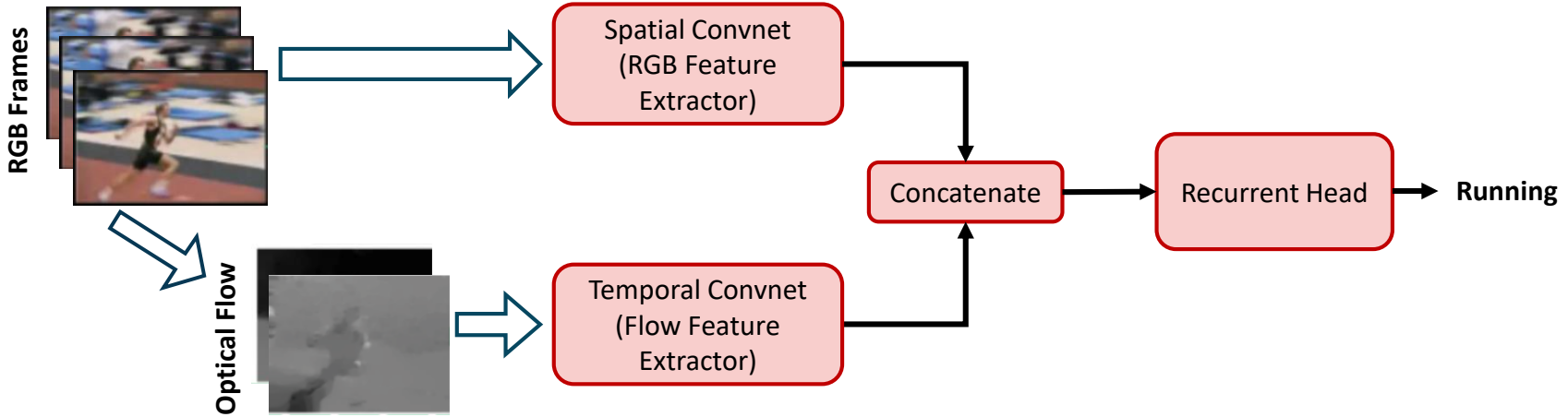
**Context-Aware Models:**
First-person vision with multi-modal sensors [1, 2]
**Benchmarks Achieved:** >90% accuracy on large-scale ego-action datasets
**Cons**: Developed models not embeddable on edge


Damen et al., 2022

# Ego Vision Functionalities – Ego Action Recognition

**Goal: Understanding and interpreting the user's actions from a first-person perspective**

**MiniROAD** uses a two-stream Temporal Segment Network (TSN) to extract features from RGB and Optical Flow frames, followed by a recurrent head for action classification



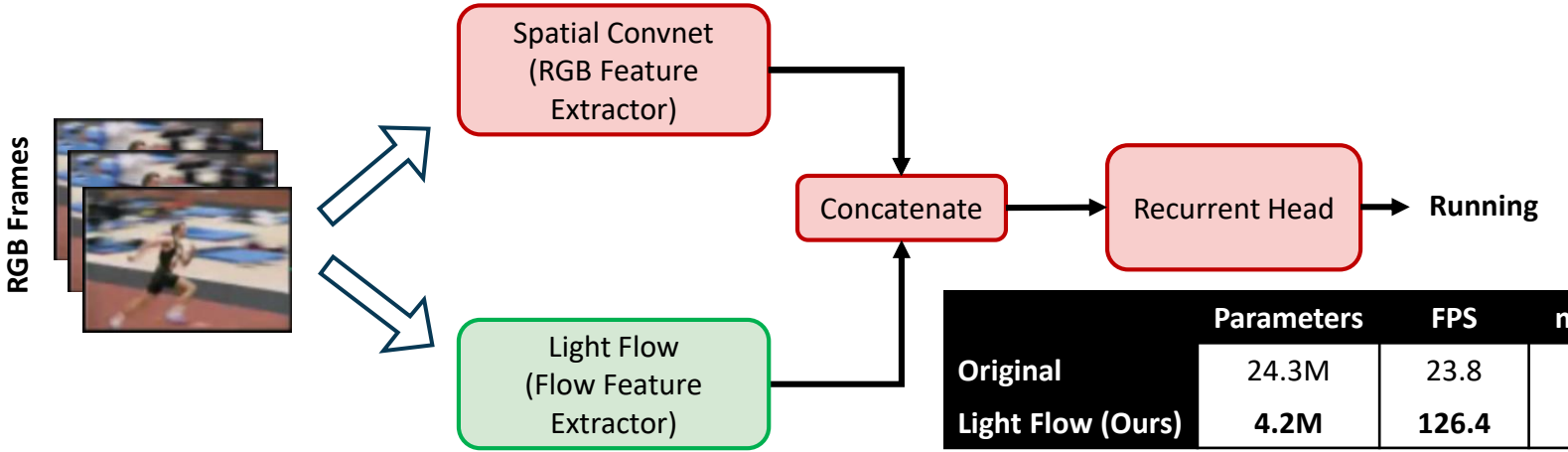

R. Santambrogio et al., Towards Real-Time Online Egocentric Action Recognition on Smart Eyewear, ECCV WS (2024)

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Ego Vision Functionalities – Ego Action Recognition

**Goal: Understanding and interpreting the user's actions from a first-person perspective**

Optical Flow computation is costly, we replaced the temporal branch with a modified encoder from **Light Flow**, trained to extract optical flow features from RGB




| | Parameters | FPS | mAP (%) |
|---|---|---|---|
| **Original** | 24.3M | 23.8 | **71.8** |
| **Light Flow (Ours)** | **4.2M** | **126.4** | 62.7 |

R. Santambrogio et al., Towards Real-Time Online Egocentric Action Recognition on Smart Eyewear, ECCV WS (2024)

# Ego Vision Functionalities – Ego Action Recognition

**Goal: Understanding and interpreting the user's actions from a first-person perspective**

- Jetson Orin Nano as computing edge device
- 3D printed eyewear frame + RGB camera

| Model | LaViLa | MiniROAD - L |
|---|---|---|
| mAP % | - | **76.2** |
| Mean Acc | 79.2 | **77.3** |
| Hz | 9.74 | **34.42** |

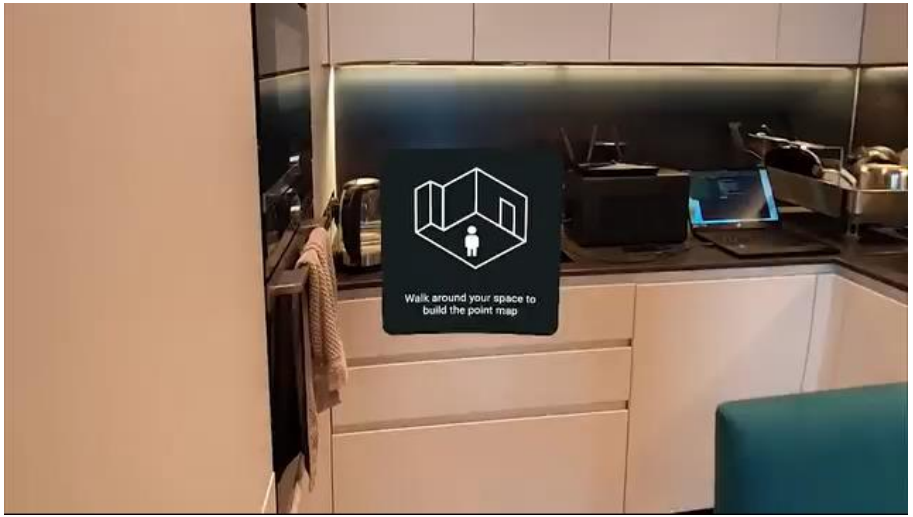

R. Santambrogio et al., Towards Real-Time Online Egocentric Action Recognition on Smart Eyewear, ECCV WS (2024)

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Ego Vision Functionalities – SLAM

**Goal: Building a real-time map of the surroundings while tracking the user's position within it**

**Enhanced Localization and Mapping:** Meter-level mapping accuracy [1, 2]
**Benchmarks Achieved:** 13 FPS on edge
**Cons:** Models are difficult to embed and missing real-time (30 FPS) on edge device



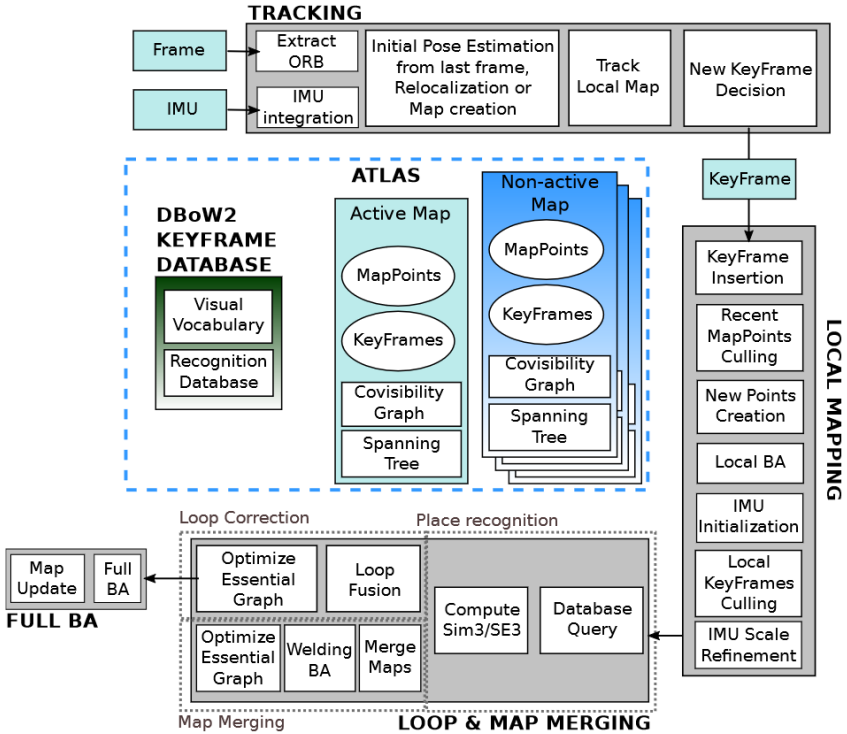

SceneScript

# Ego Vision Functionalities – SLAM

**Goal: Building a real-time map of the surroundings while tracking the user's position within it**

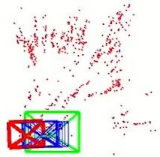
1. Able to work with different configurations
2. As fast as possible
3. Able to achieve high accuracy

**Identified Algorithm: Orb-SLAM 3**



# Ego Vision Functionalities – SLAM

We introduced a feedback loop to adaptively choose the number of points to track based on computation time



| Model | ORB SLAM3 |
|---|---|
| Absolute Traj Error (m) | 0.35 ± 0.58 |
| Hz | 34.42 |


ORB-SLAM3 algorithm for real-time indoor localization on edge devices (Jetson Orin Nano)

**Future work:** SLAM in dynamic environments

G. Affatato et al., Towards Resource-aware Visual Inertial SLAM, ECCV WS (2024)

SMART EYEWEAR LAB | EssilorLuxottica | POLITECNICO MILANO 1863

# Conclusions and Future Work

**Optimized Algorithms for Ego Action, Human Pose Estimation, and SLAM**

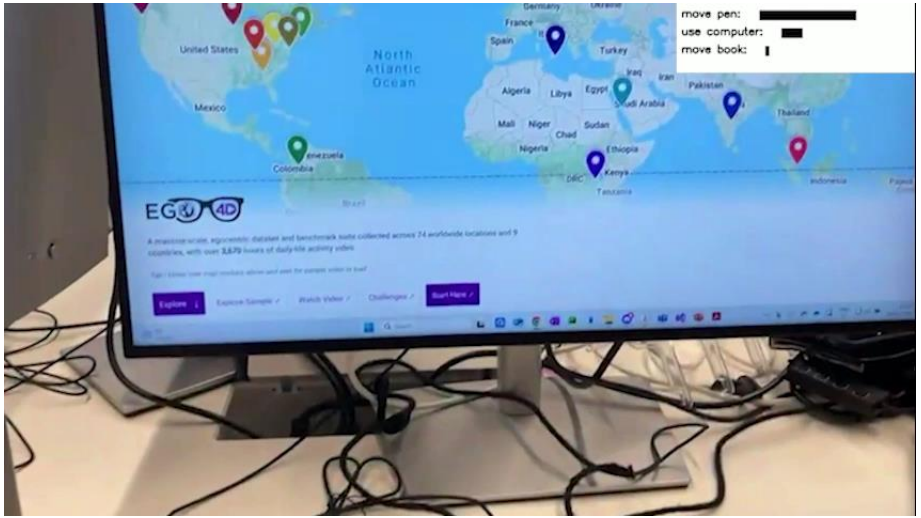
Achieved real-time performance (30 FPS) on Nvidia Jetson Orin Nano while balancing accuracy and efficiency

**Key Takeaways:**

- Prioritize accuracy over generalization
- Optimization is essential
- Trade-offs must be carefully managed

**Future Work:**

- Enhancing robustness (more scenarios)
- Expanding modalities (audio, IMU, etc.)



# Thank You for Your Attention!

Diana Trojaniello
PM,
EssilorLuxottica

Francesca Palermo
Jr. PI,
EssilorLuxottica

Matteo Matteucci
PM,
POLIMI

Marco Marcon
Sr. PI,
POLIMI

Simone Mentasti
RTDA,
POLIMI

Marco Paracchini
RTDA,
POLIMI

Hao Quan
Post-Doc,
POLIMI

Riccardo Santambrogio
PhD,
POLIMI

**Contact Details: francesca.palermo@luxottica.com**

# Resources

**Smart Eyewear Lab (Industry Research & Development)**

- Smart Eyewear Lab

**Ego Action Recognition & Human Pose Estimation**

- Ego4D: Large-Scale Dataset for Egocentric Perception
- OpenPose: Multi-Person 2D Pose Estimation
- MMPose: Comprehensive Pose Estimation Framework

**SLAM (Simultaneous Localization and Mapping)**

- ORB-SLAM3: State-of-the-Art Visual SLAM
- LIO-SAM: SLAM for LiDAR-IMU Fusion