Sapienza University | a.a. 2025-2026

Frontal Gait Flow Recognition

Computer Vision Project

Lorenzo Musso (2049518) - Giulia Pietrangeli (2057291)

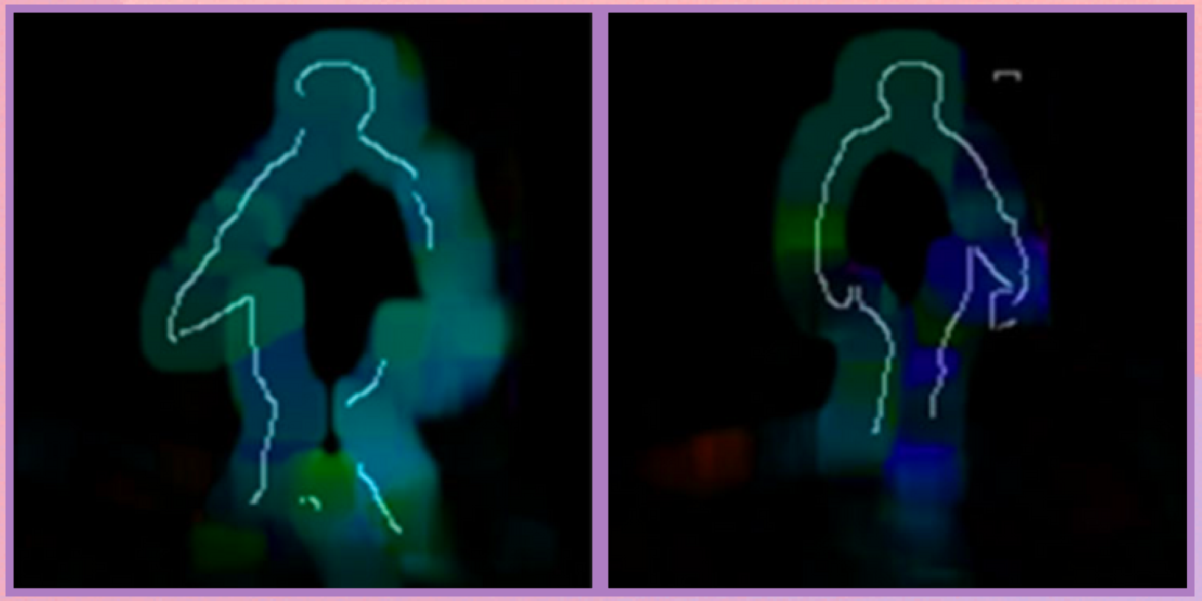
Reference Paper:

Gait Recognition Based on Gait Optical Flow Network with Inherent Feature Pyramid

Hongyi Ye, Tanfeng Sun and Ke Xu

The Inspiration: Validating Instantaneous Motion

- **Context:** Traditional methods (GEI - Gait Energy Image) focus on the silhouette.
- **Problem:** Silhouettes are sensitive to clothing and don't capture how a person moves.
- **Paper's Proposal:** GOFI (Gait Optical Flow Image)
  - Combines Optical Flow (direction/intensity of movement) + Silhouette Edges.



The GOFI Representation

Methodology:

Calculates Optical Flow (u,v) between frames.

Converts to HSV (Hue = Angle, Saturation = Magnitude).

Fuses with Canny Edge Detection to keep the body shape.

Why it matters:

It captures the dynamics of the walk, not just the static shape.

Robust against spatial noise (clothing) and temporal noise (speed).

The Paper's Architecture

Gait Optical Flow Network (GOFN)

Key Components:

- **Set Transition (ST):** Treats the gait sequence as an "unordered set" (order doesn't matter, global features do).
- **Inherent Feature Pyramid (IFP):** Extracts features at multiple scales.

Results: Achieved state-of-the-art on CASIA-B dataset, proving that Optical Flow is a superior feature for gait recognition.

Our Project:

Frontal Gait Flow Recognition

From Optical Flow Theory to Multimodal Security Applications

Project Scope & Dataset

Objective: Apply the GOFI concept to a Frontal View scenario using an explainable ML pipeline.

Dataset Source: VisionLab | Sapienza

Challenges:

* Heterogeneous Data: *Raw ROS Bags + Fragmented CSVs (IMU).*
* Geometry: *Frontal view is the hardest for gait (occlusion).*
* Missing Data: *"Slope" task has no video, only IMU.*

Dataset Composition

Subjects: 26 distinct individuals.

For every video sequence, we have available:

* RGB (Color Video)
* Depth (3D Distance Map)
* Infrared (IR Stream)
* IMU Data

Task	Repetitions	Available Modalities
Walk	6	Video + IMU
Stairs	3+3	Video + IMU
Slope	3+3	IMU

Data Engineering Pipeline

The Problem: Raw data was unusable for ML (binary files, unorganized folders).

Our Solution: A custom Python ETL Pipeline made of different scripts that:

* Decode RGB-D streams
* Clip depth at 15m
* Normalise to 8-bit
* Merge fragmented sessions
* Canonicalise labels

Outcome: A structured, synchronized dataset ready for training.

Quality Assurance & Data Leakage Prevention

Health Audit: Automated script to reject corrupt/short samples.

Dataset Completeness: Validated 100% mapping and confirmed zero missing subjects or runs

Anti-Leakage Protocol:

Split Logic: *Walk (Runs 5-6 Test), Others (Run 3 Test).*

Verification: *Computed MD5/SHA256 hashes of every file to detect duplicates.*

Result: ***Zero overlap** confirmed between Train and Test sets.*

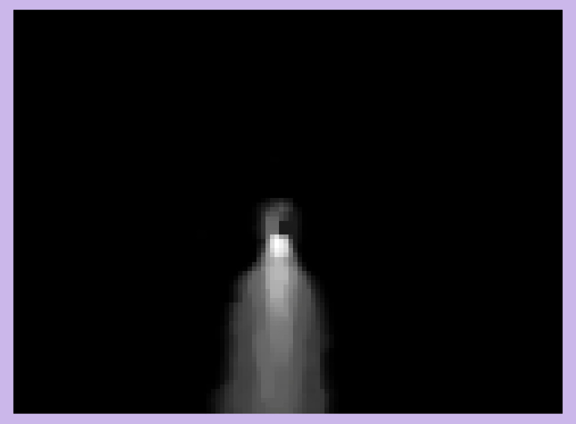
Computer Vision Pipeline

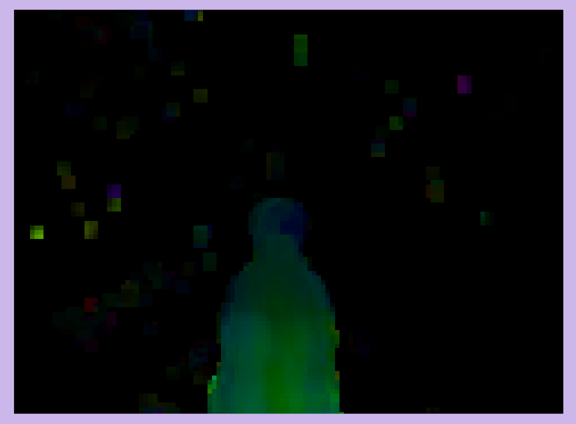
Static Background: Median filtering (N=20) to isolate the subject.

Dense Optical Flow (GOFI): Implemented the paper's concept using Farneback algorithm.

Innovation: We accumulate magnitude over the *whole video* to create a single signature image.

Sparse Optical Flow (Trace): Added Lucas-Kanade tracking on Shi-Tomasi corners to capture skeletal swing trajectories.





Multimodality & Early Fusion

The Missing Piece: Visual data fails in "Slope" tasks (no video) or heavy occlusion.

Solution: Integration of Xsens IMU sensors.

Engineering: Extracted statistical descriptors (μ,σ,RMS,min,max) for 19 sensor channels.

2 Models:

* Concatenation of Video Vector (49k) + IMU Vector (5k). Dedicated to Stairs & Walk totalling 54,082 features per sample.
* Only IMU Vector (5k) dedicated to Slope.

Machine Learning Architecture

After a Grid Search on the SVM parameters and the PCA variance we decided:

Step 1:

PCA (Dimensionality Reduction):

Retained 95% Variance.

Main Model: *Reduced 54k features → 353 components.*

Slope Model: *Reduced 4k features → 53 components.*

Step 2:

Linear SVM:

Selected parameters: *C=1 and Kernel=Linear*

Why? Efficient, effective for high-dimensional data.

Experimental Results

Main Model (Fusion)

99,04%

Slope Model

100%

Modality	Accuracy	Impact vs Fusion
Only Video	91,35%	-7,69%
Only IMU	100%	+0,96%
Fusion	99,04%	-

Fusion caused slight dilution (-0.96%) but adds robustness in real scenarios.

Only 2 missclassifications → High degree of similarity in the signature

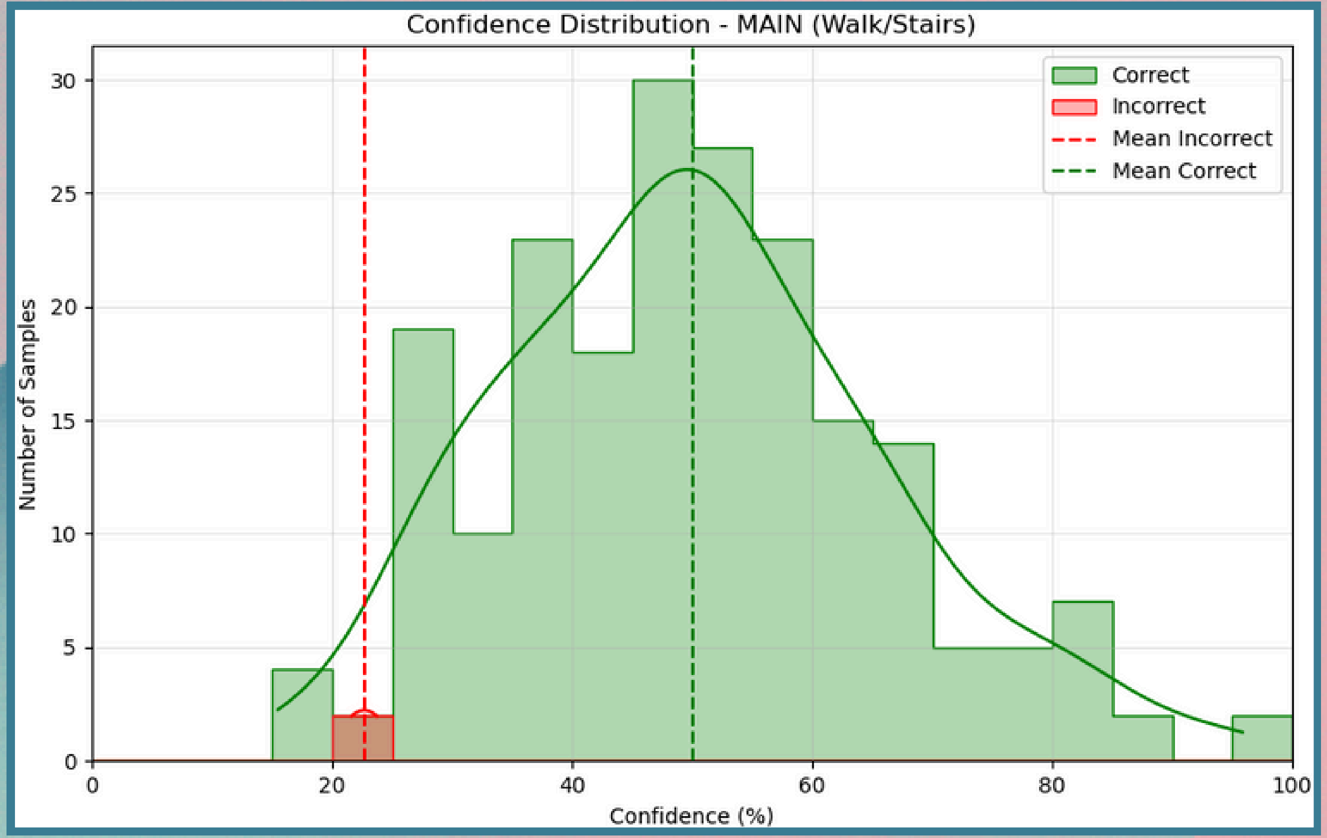
Confidence Analysis & Security Logic

Main model Analysis: We calibrated SVM probabilities.

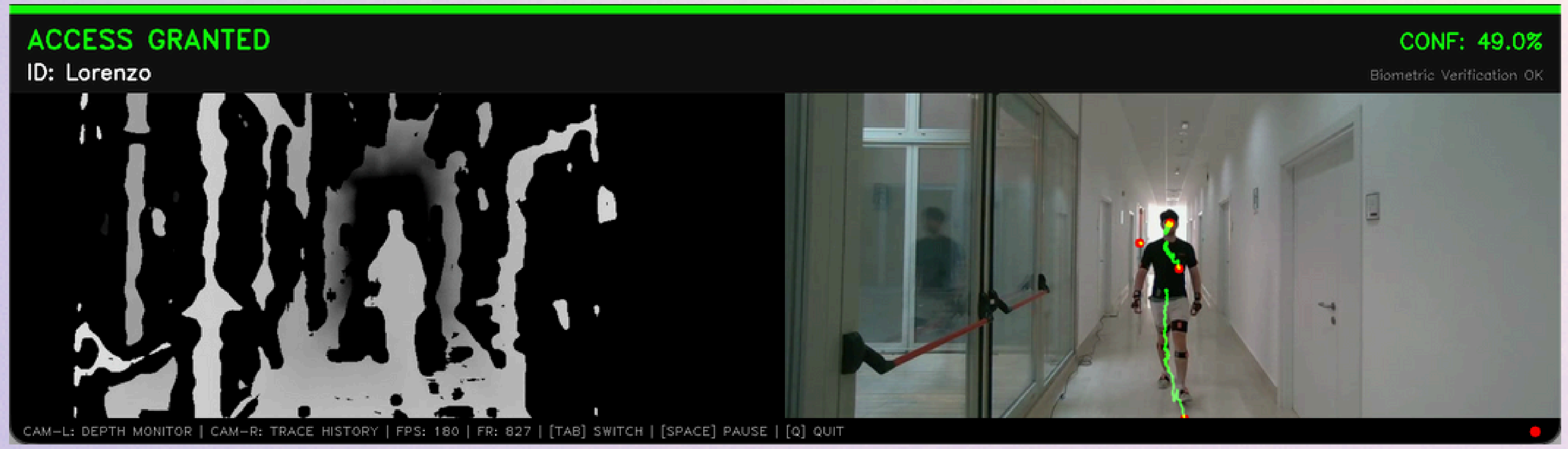
- Correct predictions have high confidence (Mean ~51%, Max 96%, Min 15%).
- Errors have low confidence (~26%).

Strategic Threshold: Established a cut-off at 26.6%.

- Below this = "Unknown Subject".
- Above this = "Identity Confirmed".

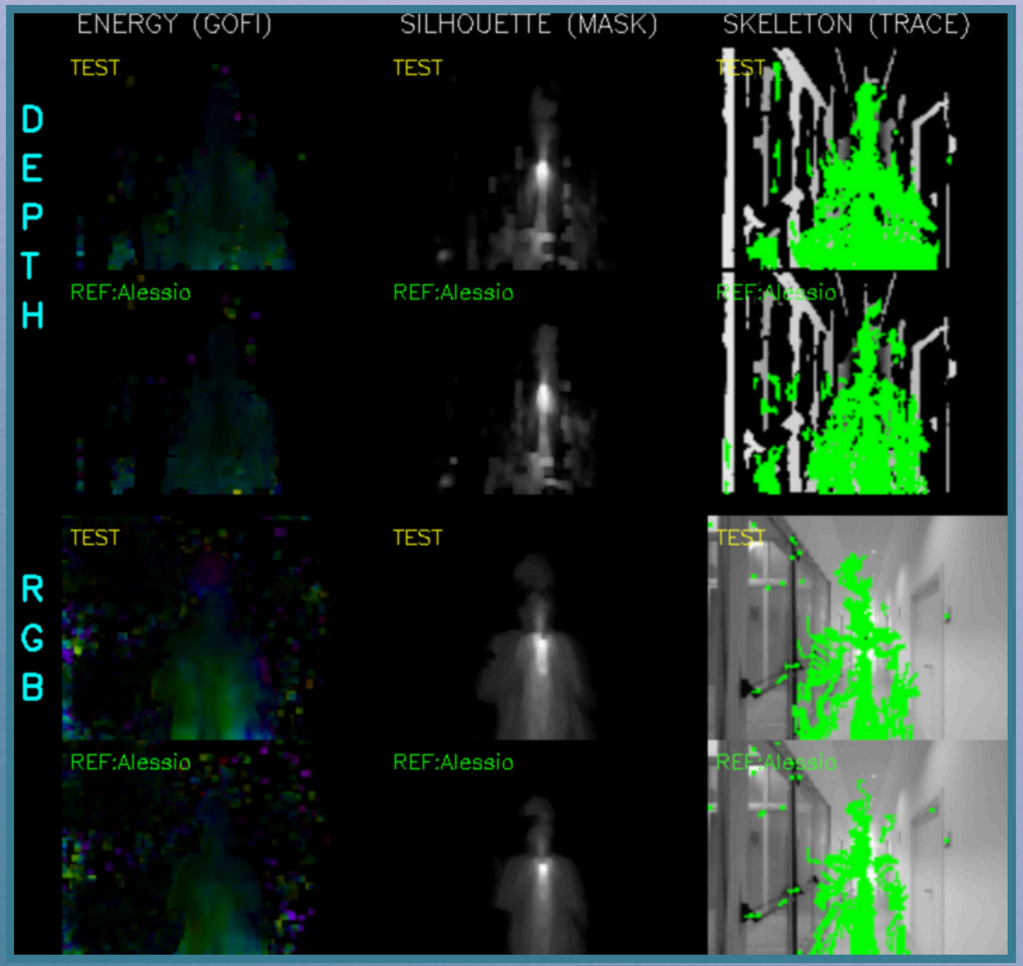


Real-Time Demonstrators



Demo 1: Gait Security Pro

- Simulates a gate / checkpoint.
- Real-time filters (if video available):



- Depth – RGB + Green Traces (Kinematics)
- Heatmap (Energy) – Silhouette
- Vibrant Depth – Cyber Edges (Shape)

Demo 2: Gait Visualizer

- Forensic tool. Compares Test Subject vs. Database Reference using a "Gait DNA Matrix".

References

H. Ye, T. Sun, and K. Xu, *Gait Recognition Based on Gait Optical Flow Network with Inherent Feature Pyramid*, Applied Sciences, vol. 13, no. 19, p. 10975, 2023.

H. Masood and H. Farooq, *Utilizing Spatio Temporal Gait Pattern and Quadratic SVM for Gait Recognition*, Electronics, vol. 11, no. 15, p. 2386, 2022.

Thank you!