Visualizing Body-Interaction Importance for Conversation Activity Recognition in Pedestrian Groups

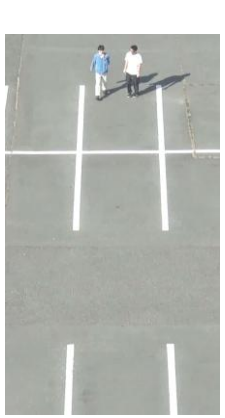
Wataru Ganaha∗, Takumi Ozaki∗, Masashi Nishiyama∗

∗Graduate School of Engineering, Tottori University, Japan

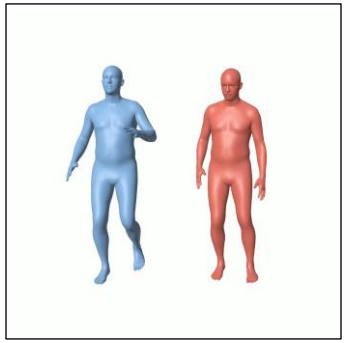
鳥取大学
Tottori University

Introduction

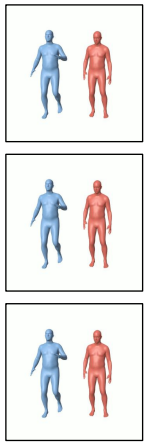
- Video analysis supports a deeper understanding of human behavior.
  We focus on conversation activity recognition in pedestrian groups walking outdoors.
- Conversation activity indicates whether conversation occurs.
  If conversation occurs, it indicates whether it is active or inactive.
- Our previous method recognizes conversation activity using physical body interaction.
  This term refers to temporal changes in gestures and mutual body orientation.



Pedestrian group video sequence



Interaction video sequence in a virtual space



Short video sequence

Classification
(C3D model)



Active



Inactive

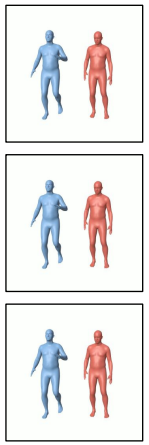


None

Conversation activity

Applying IG to visualize importance

- Previous evaluation showed high recognition accuracy. Pixel-level importance over time and body states in high-importance regions remained unexplored.
- To visualize body-interaction importance,
  we apply **Integrated Gradients (IG)** to each short video sequence.
  [M. Sundararajan+, 2017]
- IG attributes the model output for the predicted class to input features through gradients.



Short video sequence

Visualization of importance
(IG applied to the C3D model)



Importance map for each frame

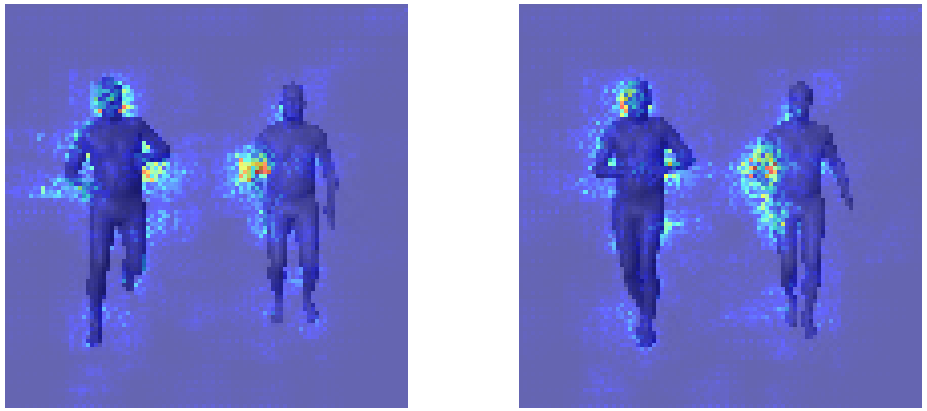
Body-interaction importance is the attribution magnitude at each pixel.
It does not distinguish positive from negative effects on the model output.

Limitations of direct IG visualization

Direct application of IG has **spatial** and **temporal** limitations.

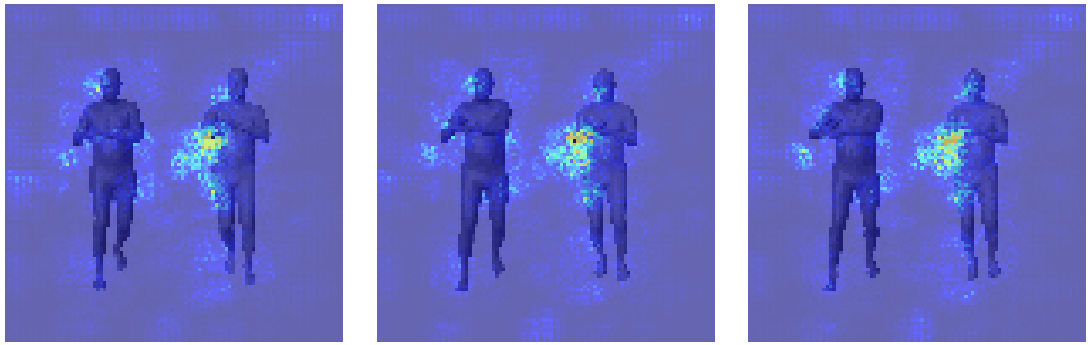
Spatial scattering

✗ Low-importance pixels scatter across body parts and mix with high-importance pixels. High-importance regions are difficult to identify.



Temporal discontinuity

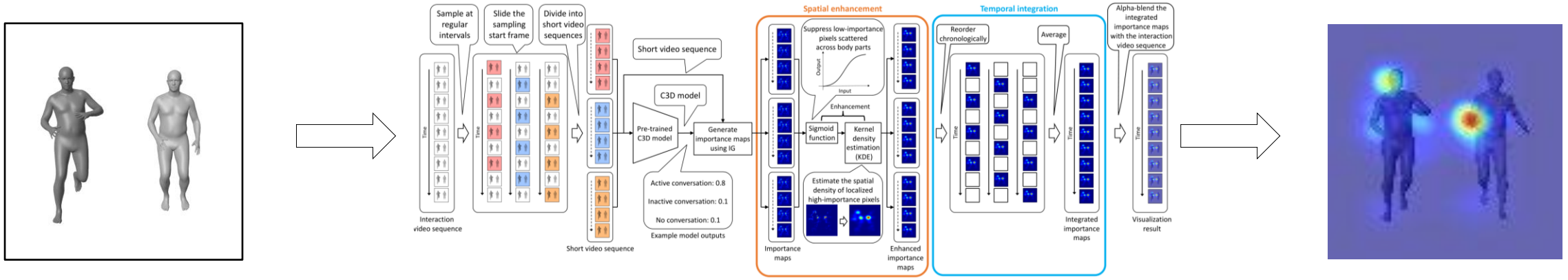
✗ Each short video sequence produces independent importance maps. Their temporal correspondence is unclear.



Purpose

We propose an IG-based method to visualize pixel-level spatiotemporal body-interaction importance.

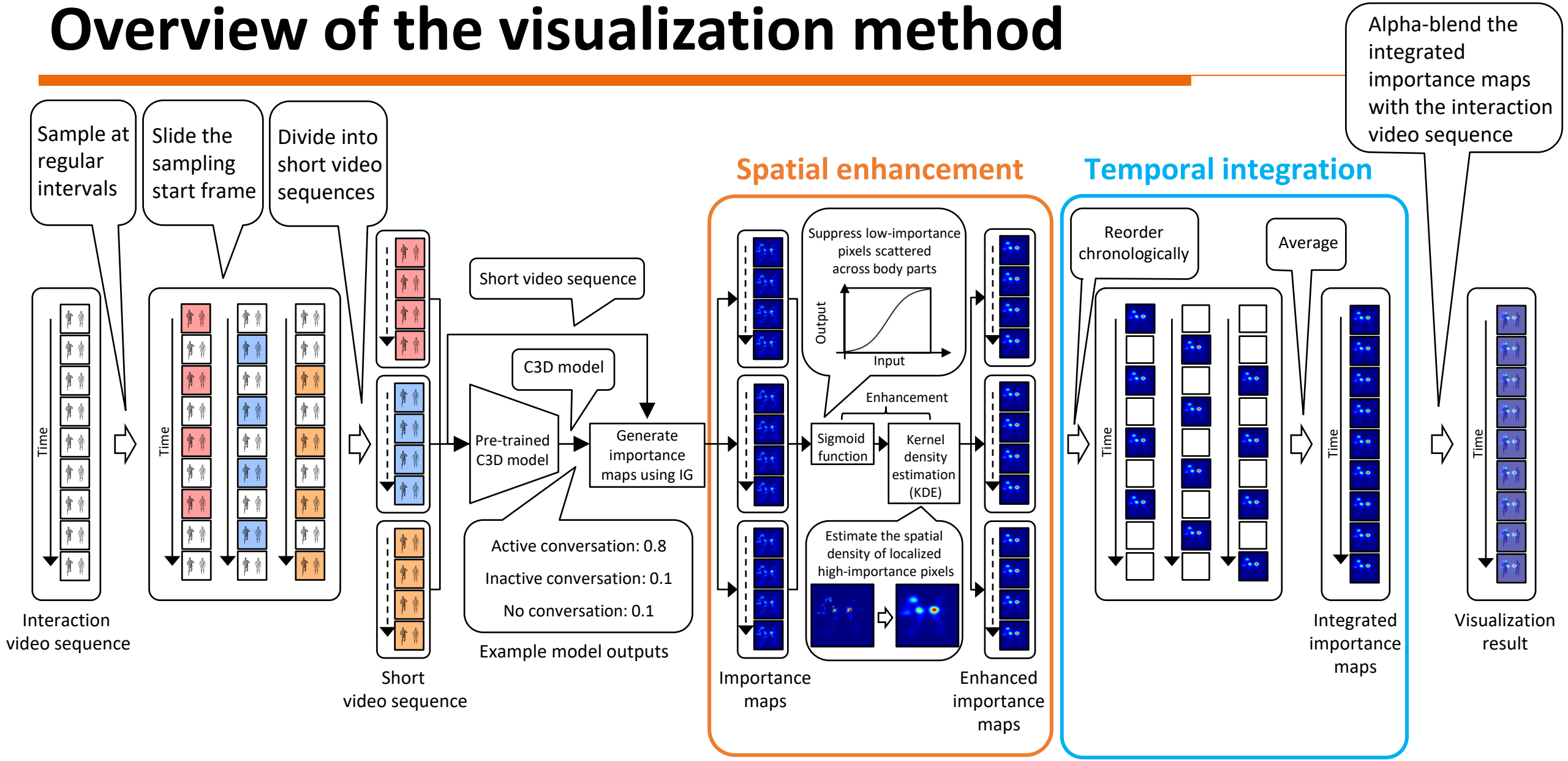
- Our method combines **spatial enhancement** and **temporal integration**.
  - **Spatial enhancement** suppresses scattered low-importance pixels.
  - **Temporal integration** reveals continuous changes in importance.



Interaction video sequence | Our method | Visualization result

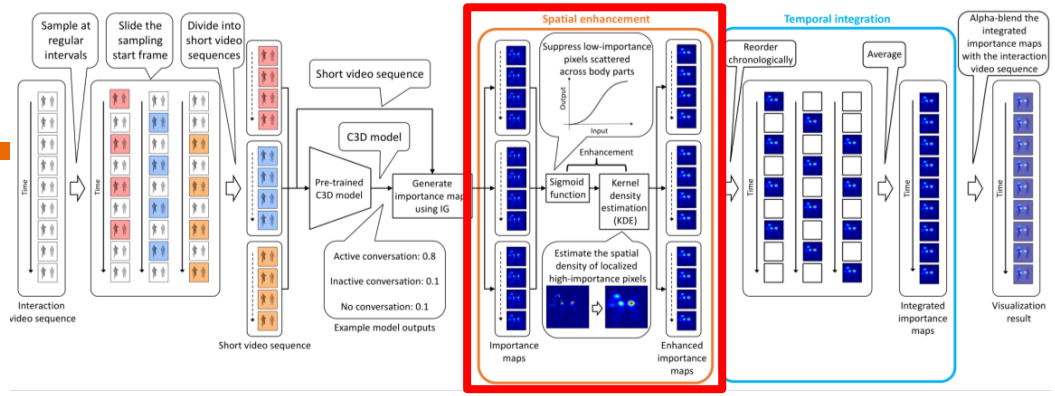
- We examine body parts and states in regions with large attribution magnitudes.

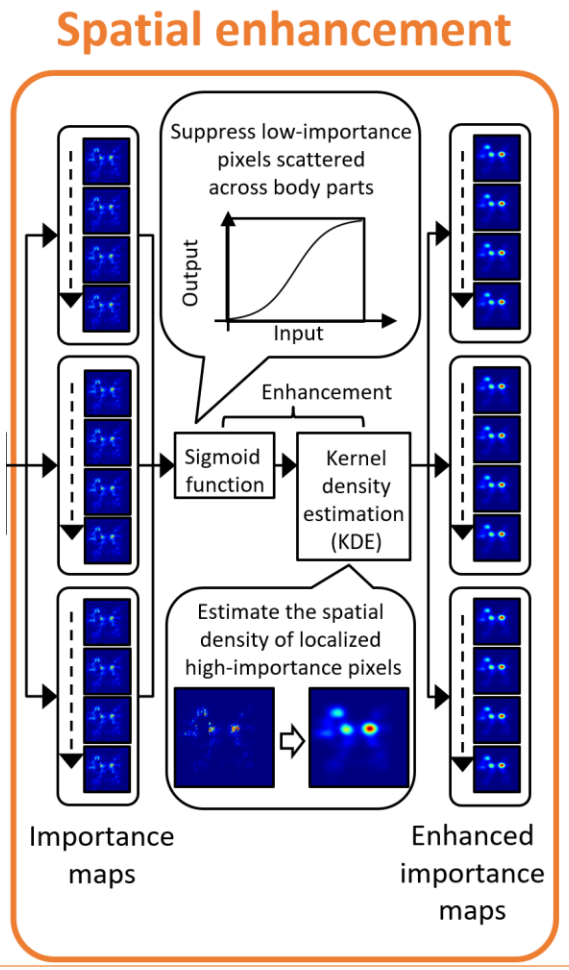
Overview of the visualization method



Spatial enhancement

Spatial enhancement clarifies regions in which high-importance pixels are concentrated.



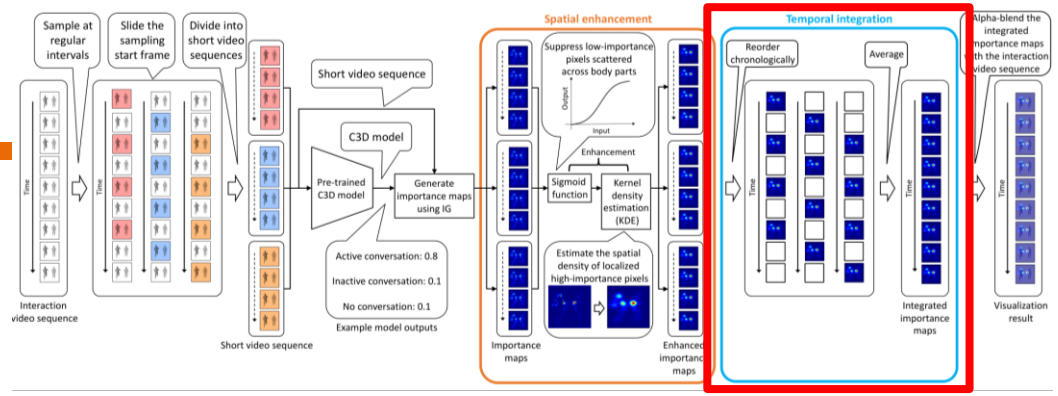


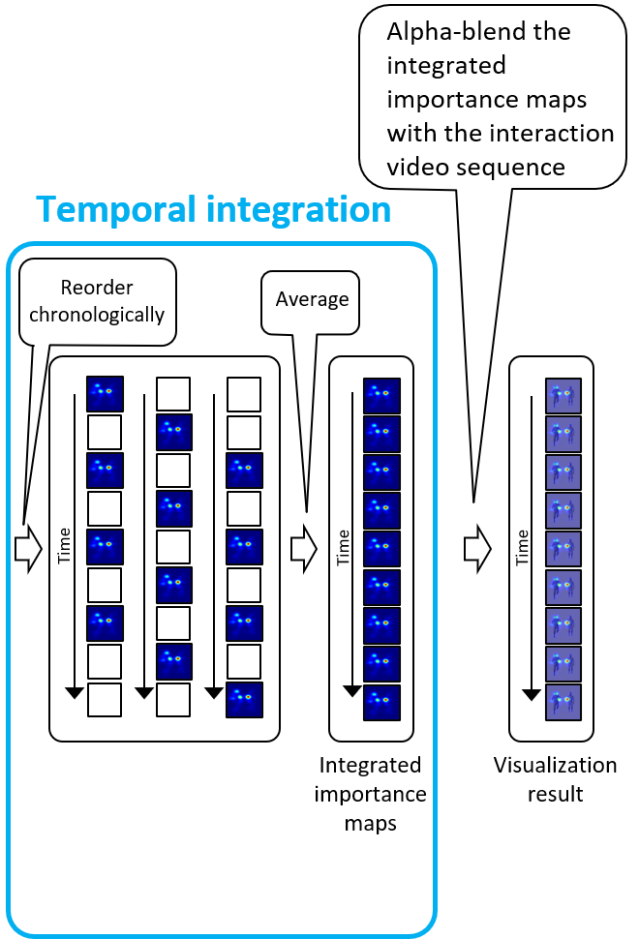
- We normalize each importance map to the range 0 to 1.
- We apply a sigmoid function.
  It suppresses low importance and enhances high importance.
- We apply Kernel Density Estimation (KDE) with a Gaussian kernel. It represents locally concentrated high-importance pixels as continuous regions.

✓ **The associated body parts become easier to identify.**

Temporal integration

Temporal integration combines enhanced importance maps into one chronologically ordered visualization.





- We shift the sampling start frame by one frame.
  This covers every source frame.
- We reorder importance maps at their original frame positions.
- We average maps that correspond to the same source frame.
- We alpha-blend the integrated importance maps with the interaction video sequence.

✓ **Continuous changes in importance become easier to follow.**

Experimental setup

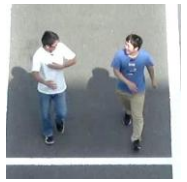
Recording conditions

We used outdoor overhead videos of 20 participants and 52 unique pairs.



Active conversation

Introduce hobbies to one another.





Inactive conversation

Discuss a topic of little interest to either participant.





No conversation

Walk without speaking.

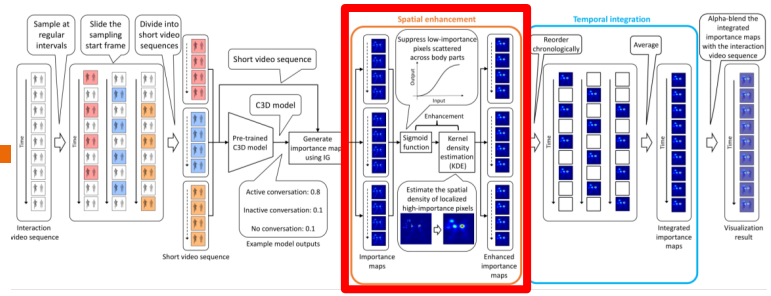


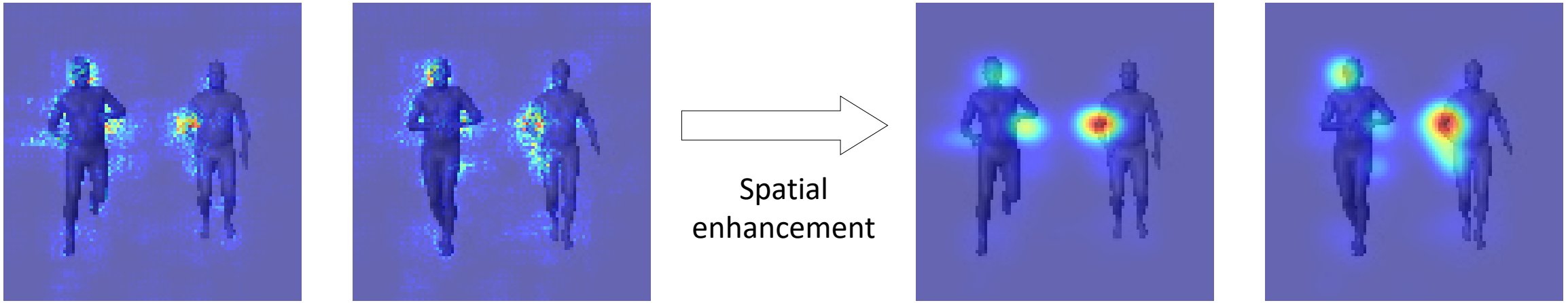
Each pair walked under all three conditions. We did not ask for predefined body movements.

Common visualization settings

- Input: 100 × 100 pixels, 16 frames, sampled at 18-frame intervals.
- IG: 50 steps with Gauss–Legendre quadrature, all-zero input baseline.
- Spatial enhancement: sigmoid a = 8, KDE with a Gaussian kernel.

Effect of spatial enhancement





Spatial enhancement

Direct application of IG

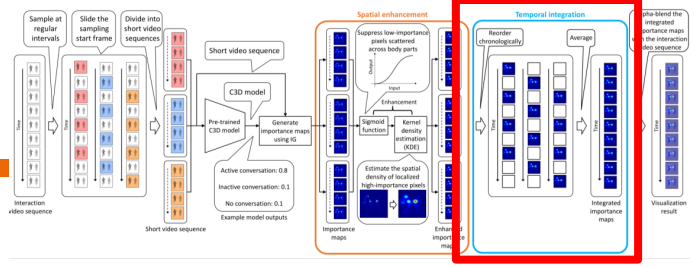
Our method

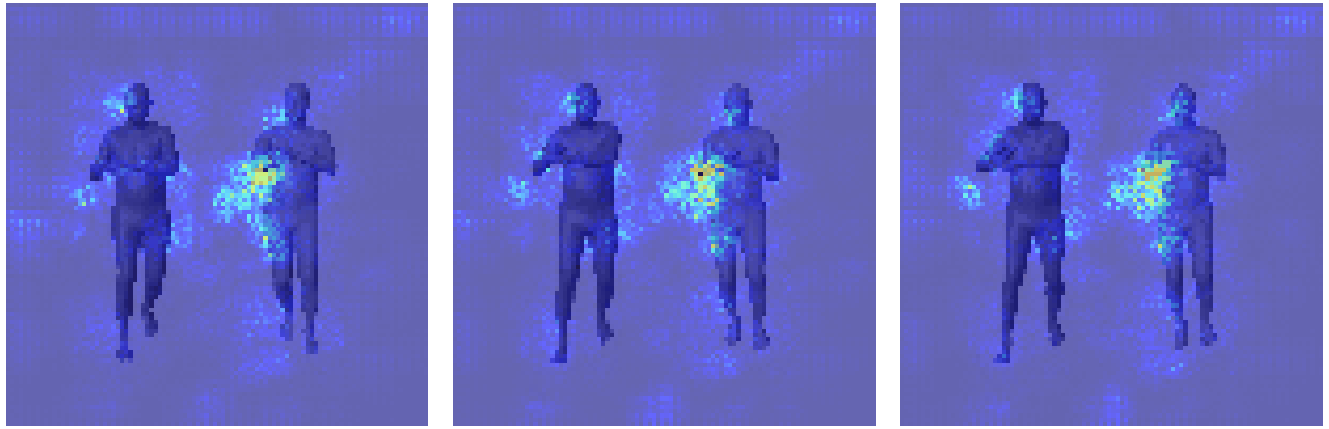
Low — High

Body-interaction importance

- We see low-importance pixels scattered across body parts.
- The sigmoid function increases contrast between high and low importance.
- KDE groups locally concentrated high-importance pixels into continuous regions.
- These operations make the associated body parts easier to identify.

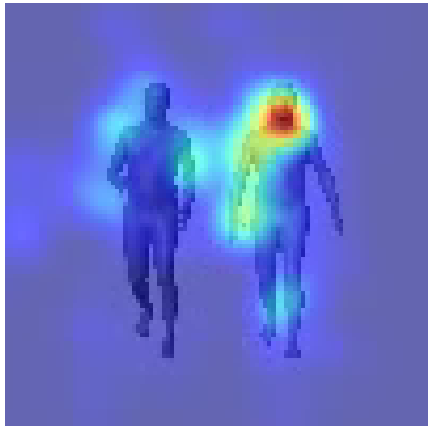
Effect of temporal integration





Direct application of IG
Individual short video sequences

Reorder
and average



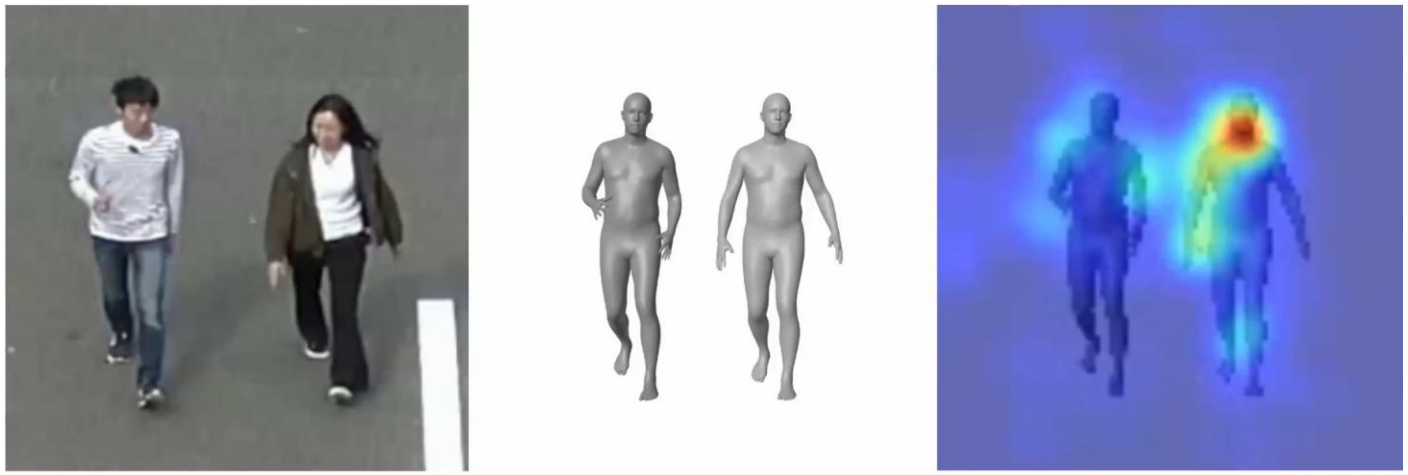
Our method
Integrated importance maps

Low — High

Body-interaction importance

- Before integration, short video sequences have unclear temporal correspondence.
- Reordering and averaging reduce discontinuities between short video sequences.
- We can follow changes in importance from beginning to end.
- Averaging reduces variation seen in only one short video sequence.

Visualization results: active conversation



Pedestrian group video sequence | Interaction video sequence | Visualization result

Low — High

Body-interaction importance

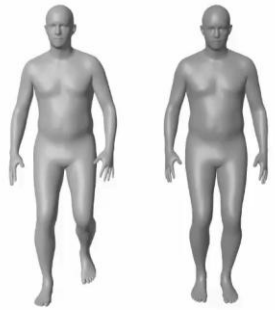
Video example of active conversation

- The pedestrians raise their hands or elbows while gesturing. They turn their heads and upper bodies toward one another.
- High-importance pixels are concentrated around raised hands and elbows, and mutually oriented heads and upper bodies.

Visualization results: inactive conversation



Pedestrian group video sequence



Interaction video sequence



Visualization result

Low — High

Body-interaction importance

Video example of inactive conversation

- The pedestrians keep their hands lowered and make no gestures. They turn their heads toward one another.
- High-importance pixels are concentrated around lowered hands and mutually oriented heads.

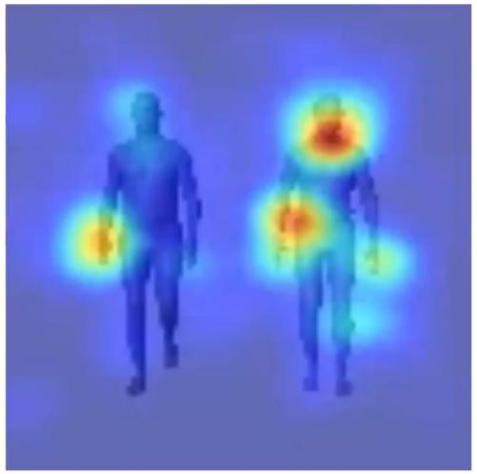
Visualization results: no conversation



Pedestrian group video sequence



Interaction video sequence



Visualization result

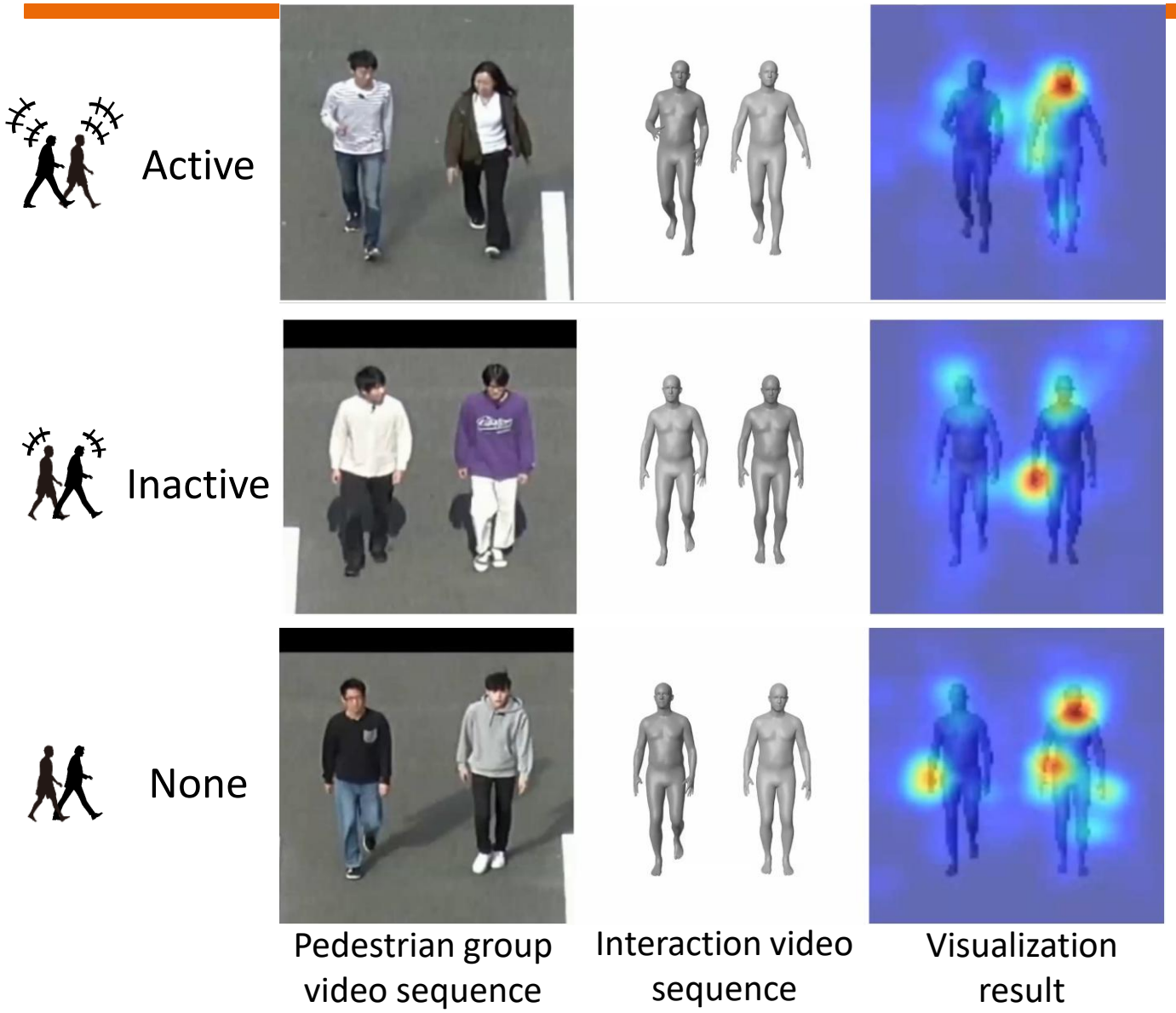
Low — High

Body-interaction importance

Video example of no conversation

- The pedestrians keep their hands lowered and make no gestures. Their heads face the walking direction.
- High-importance pixels are concentrated around lowered hands and heads facing the walking direction.

Comparison of visualized body states



Hands and elbows

- High-importance pixels are concentrated around raised hands and elbows in active conversation, and lowered hands in the other two conditions.

Head orientation

- Inactive conversation and no conversation both show no gestures.
- Head orientations and attribution patterns around the heads differ between these conditions.

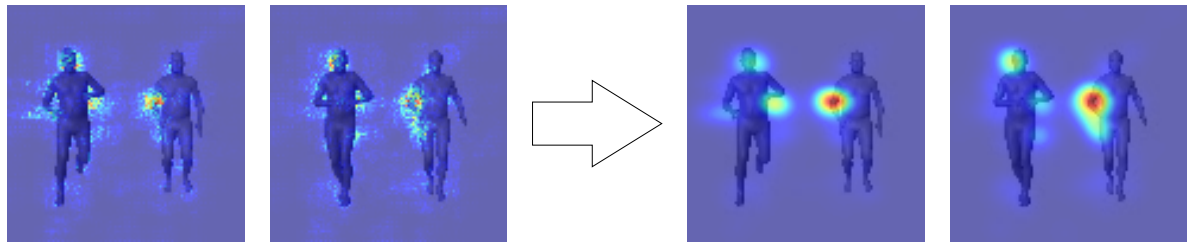
Interpretation

- These patterns suggest that the model may be sensitive to gestures and mutual body orientation.

Conclusion

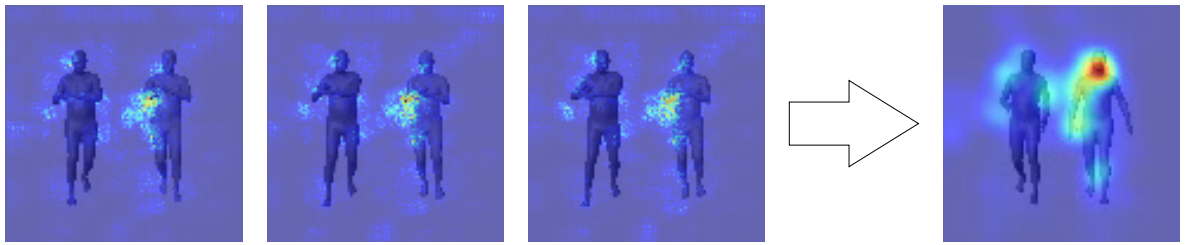
- We proposed an IG-based method to visualize pixel-level body-interaction importance.
- Spatial enhancement clarifies high-importance regions.
- Temporal integration reveals continuous changes in importance.
- The visualizations provide insights into physical states associated with C3D recognition decisions.

Spatial enhancement



✓ **The associated body parts become easier to identify.**

Temporal integration



✓ **Continuous changes in importance become easier to follow.**

- Future work will quantify body-part contributions through masking and perturbation.