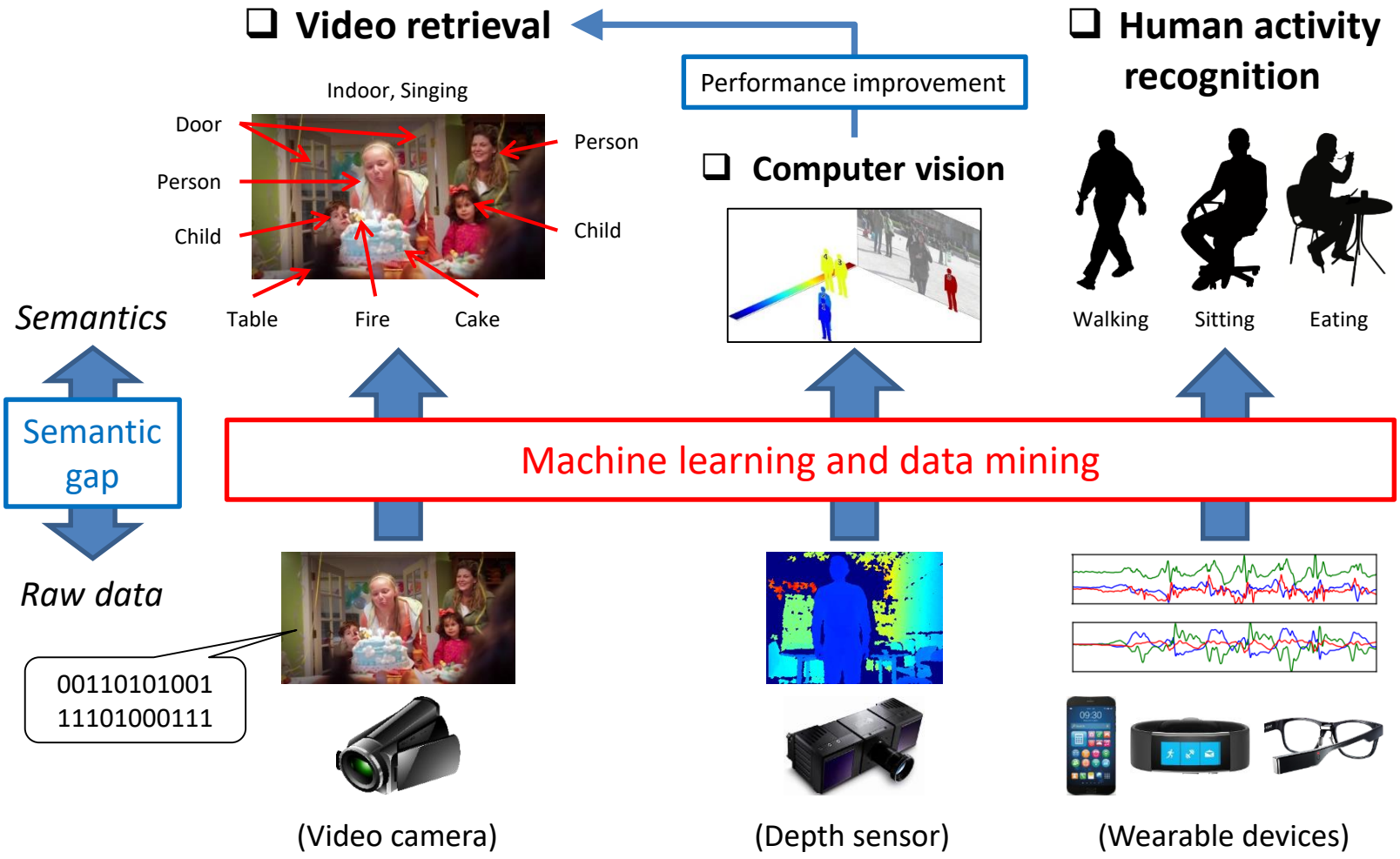
Overview of My Research

Extract semantic information from multimedia data using machine learning and data mining techniques



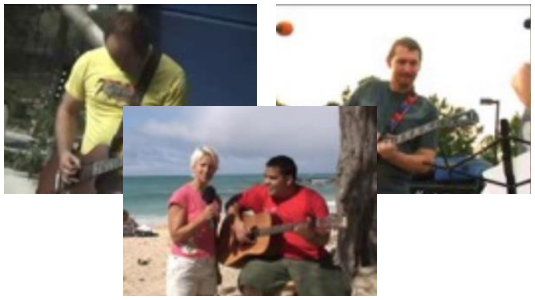
Concept-based Video Retrieval

Given a text query, return videos that are relevant to the query
(No manually annotated tag is used)

Query:
A person playing guitar outdoors

Concepts:
Person
Outdoors
playingGuitar

Video data



Person: 0.9
Car: 0.0
Building: 0.1
Road: 0.3
Bridge: 0.0
Sky: 1.0
Outdoors: 1.0
Indoors: 0.2
Beach: 0.7
Walking: 0.2
Playing_Guitar: 0.8
Throwing: 0.1
⋮

Concepts: Textual descriptions of meanings that humans can perceive from videos

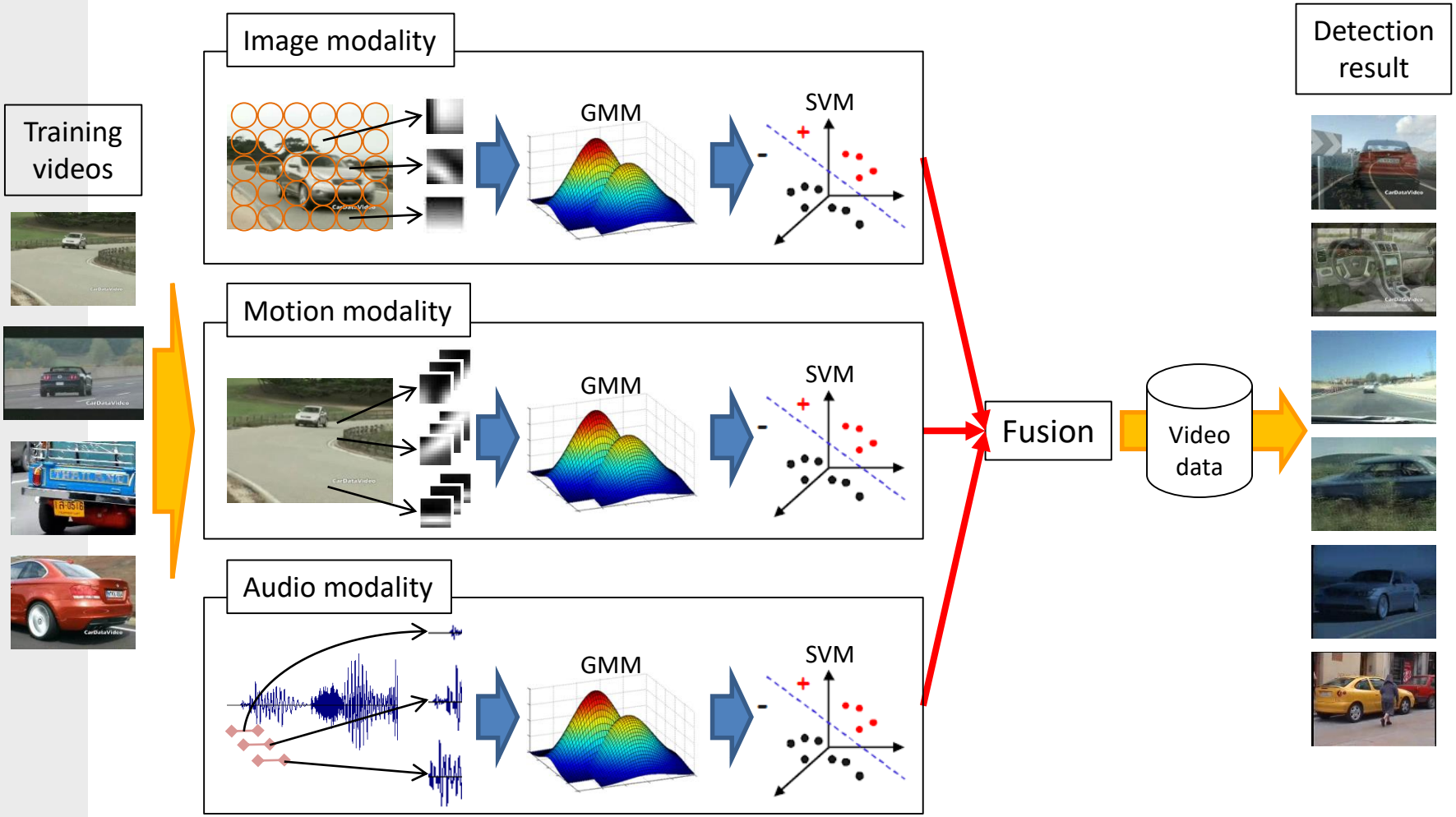
1. Concept detection: For each concept, annotate a video with a detection score
Representing the likelihood of the concept's presence

2. Retrieval: Select concepts related to a query, and find videos with high detection scores for the selected concepts

Detect thousands of concepts to respond to various queries

Our Concept Detection Method

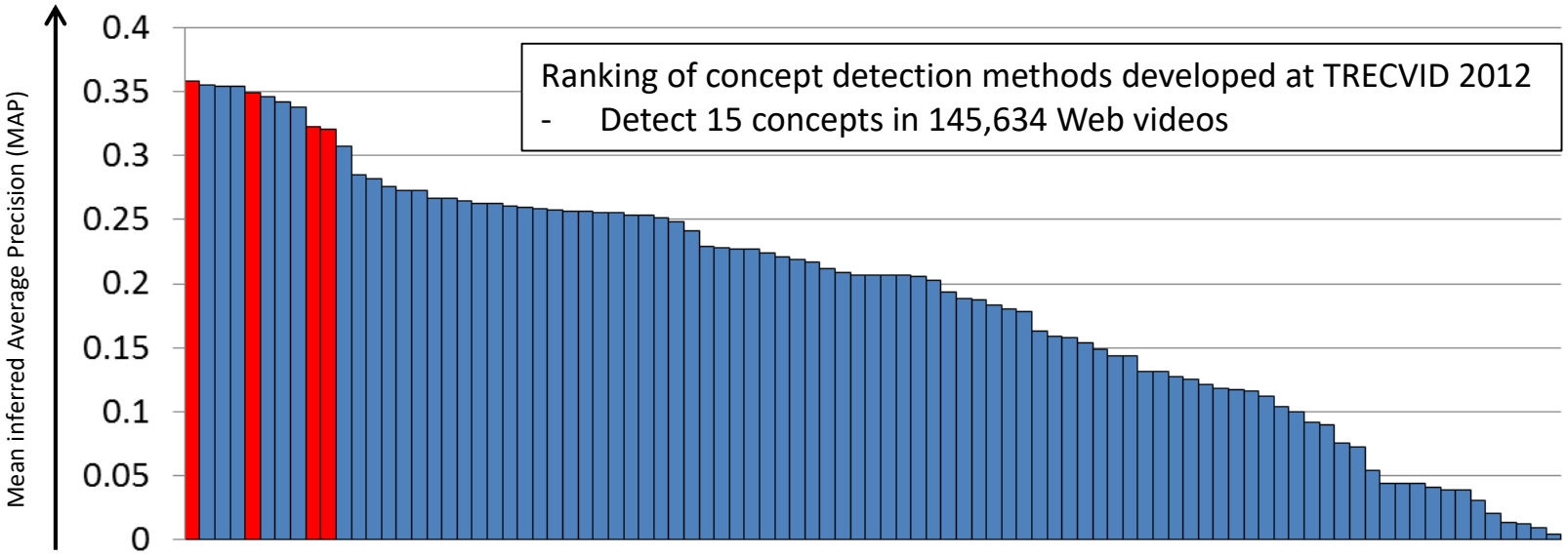
- Diversity of a concept's appearances → Large amount of training data
- Locality of a concept's appearance → Spatially and temporally dense sampling of local features



Performance of Our Concept Detection Method

TREC Video Retrieval Evaluation (TRECVID):

NIST-sponsored annual worldwide competition on video analysis and retrieval

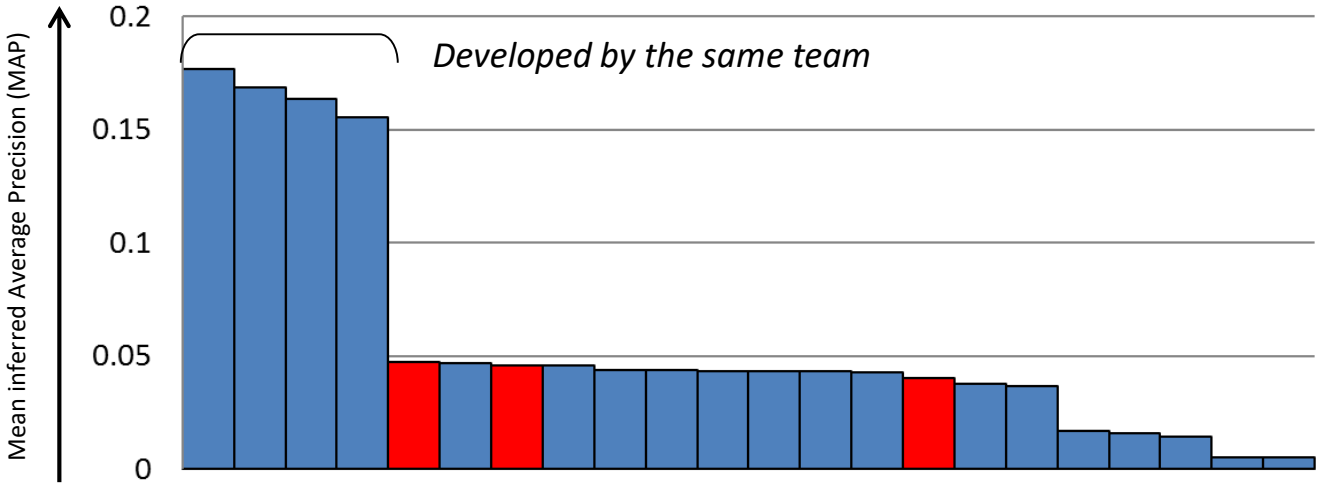


Examples of target concepts

Our method achieved the highest performance among 91 methods developed at 25 research institutes (e.g., IBM, CMU, Stanford Univ., Canon etc.)!

Performance of Our Concept-based Video Retrieval Method

Ranking of methods developed at TRECVID 2016 Ad-hoc Video Search (AVS) task (manually-assisted)
- Perform retrieval for 30 queries on 335,944 web videos
- Compute the overall score of a video as the sum of detection scores for concepts related to a query



Examples of queries

A person playing guitar outdoors

A diver wearing diving suit and swimming under water

A person drinking from a cup, etc.

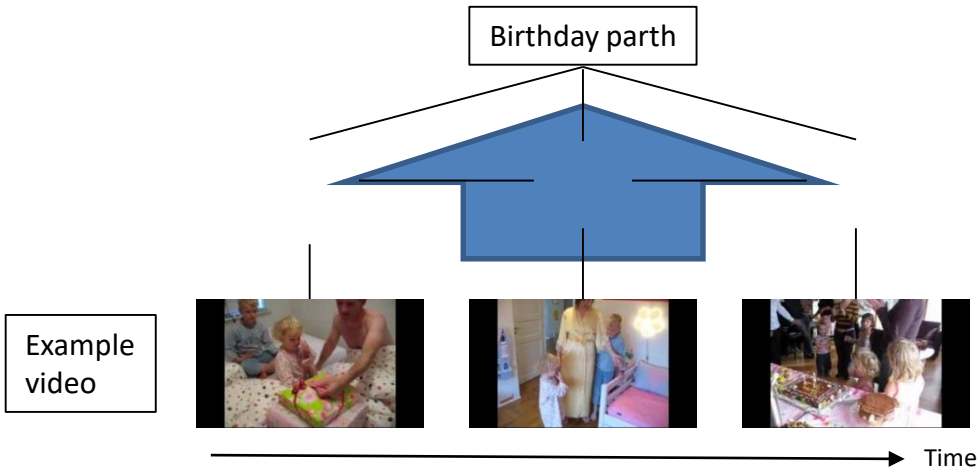
A man indoors looking at camera where a bookcase is behind him

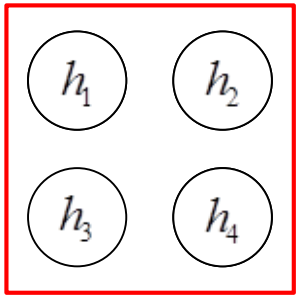
A crowd demonstrating in a city street at night

We are at the second place in terms of participating teams!
(Our best method is ranked at the fifth place among 22 methods)

Concept Selection Using Example Videos

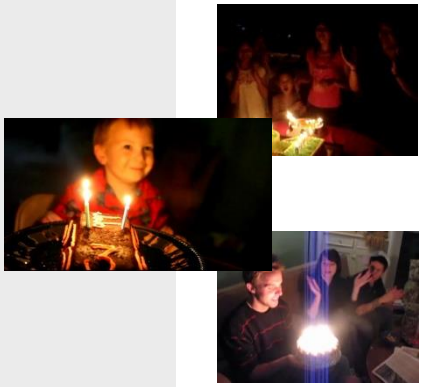
Probabilistic model with hidden states representing the relevance of each concept to a query





Optimise hidden states to accurately classify example videos

- An example hidden state for "Birthday party"



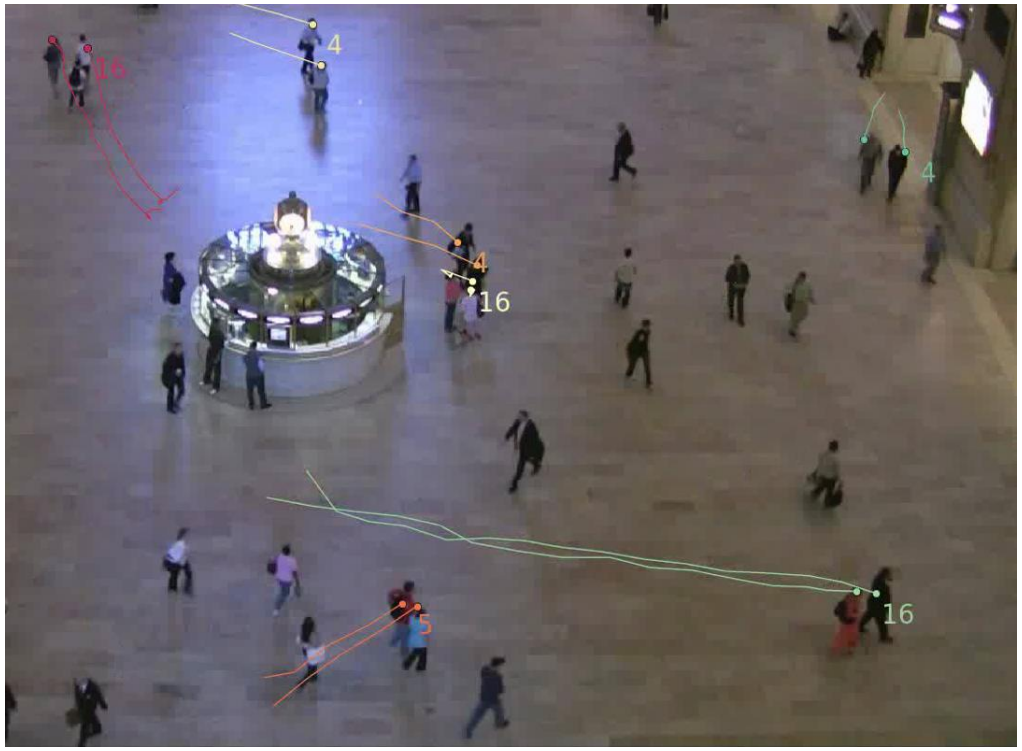
0.247 (Moonlight)
0.204 (Nighttime)
0.192 (Entertainment)
0.125 (Event)
0.121 (Singing)
0.097 (Celebrity_Entertainment)
0.093 (Dancing)
0.093 (Instrumental_Musician)
0.057 (Person)
0.056 (Face)

- An example hidden state for "Getting a vehicle unstuck"

1.665 (Text_On_Artificial_Background)
1.421 (Waterscape_Waterfront)
1.342 (Head_And_Shoulder)
1.316 (Car)
1.208 (Infants)
1.112 (Outdoor)
1.085 (Adult_Male_Human)
1.081 (Daytime_Outdoor)
1.065 (Driver)
1.051 (Human_Young_Adult)

Extract Human Groups as Convoys

Too much information in a crowd surveillance video



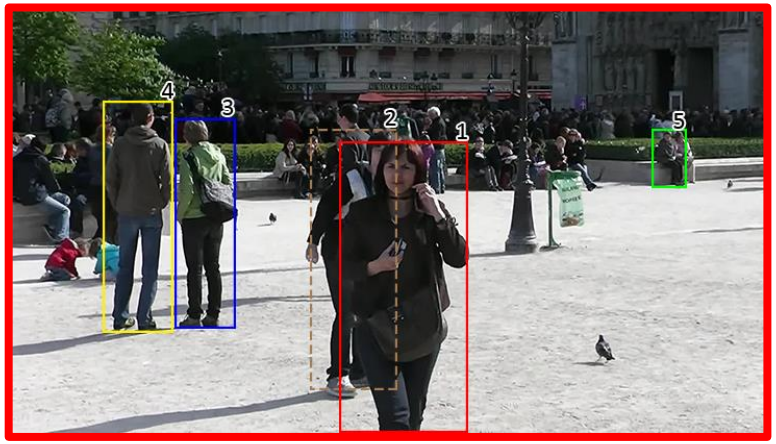
Need for automatic or assistive systems to detect suspicious activities

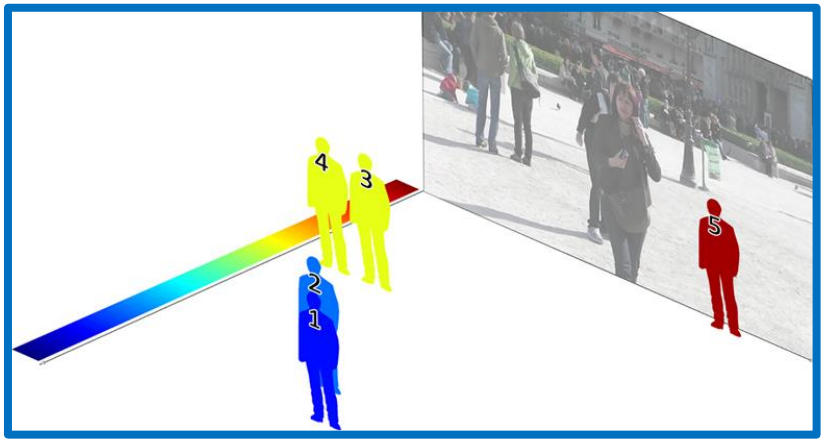
Extract groups of pedestrians moving together as convoys

1. Trajectory extraction
2. Convoy detection by trajectory analysis

Lack of 3D Information in a Video

For precise detection of convoys, we need to examine the spatial relation among people



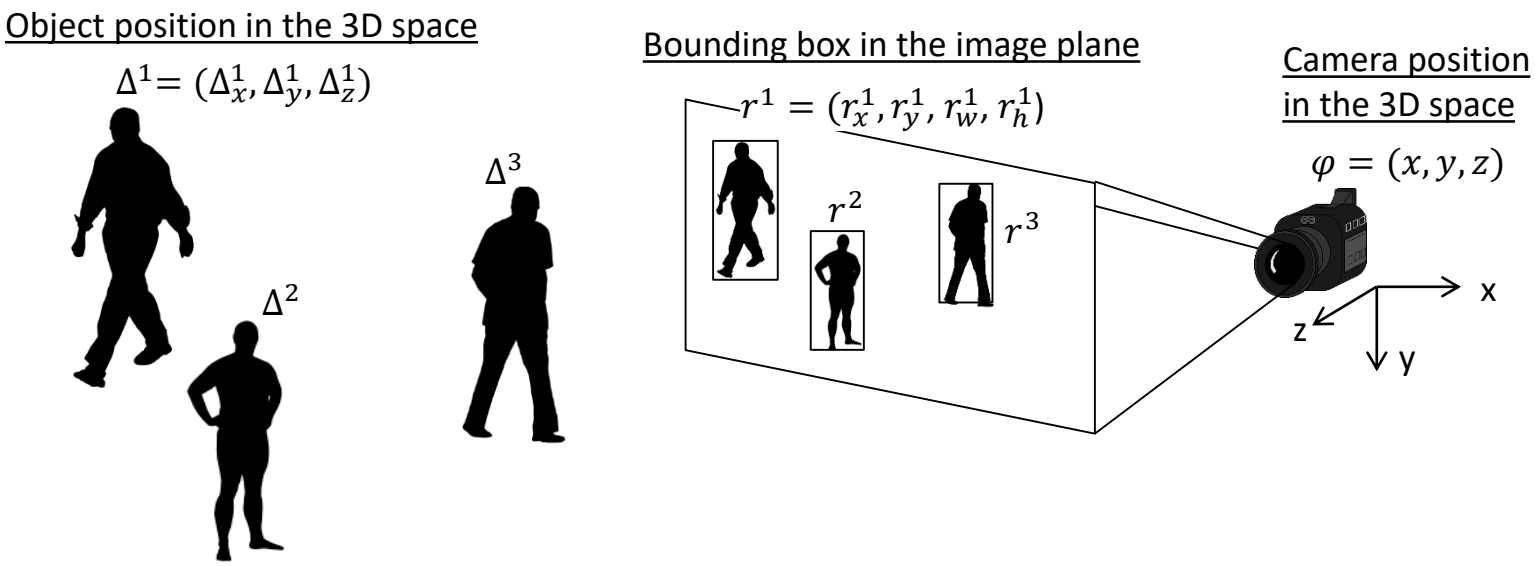


- ✓ The original 3D space is projected onto a 2D image plane
- ✓ Humans can easily recognise the 3D spatial relation from a 2D frame

From a 2D video, extract **3D trajectories** each of which represents the transition of an object's positions in the 3D space

Probabilistic 3D Trajectory Extraction

Estimate 3D positions of both of objects and the camera, so that the 2D frame can be generated with the highest probability

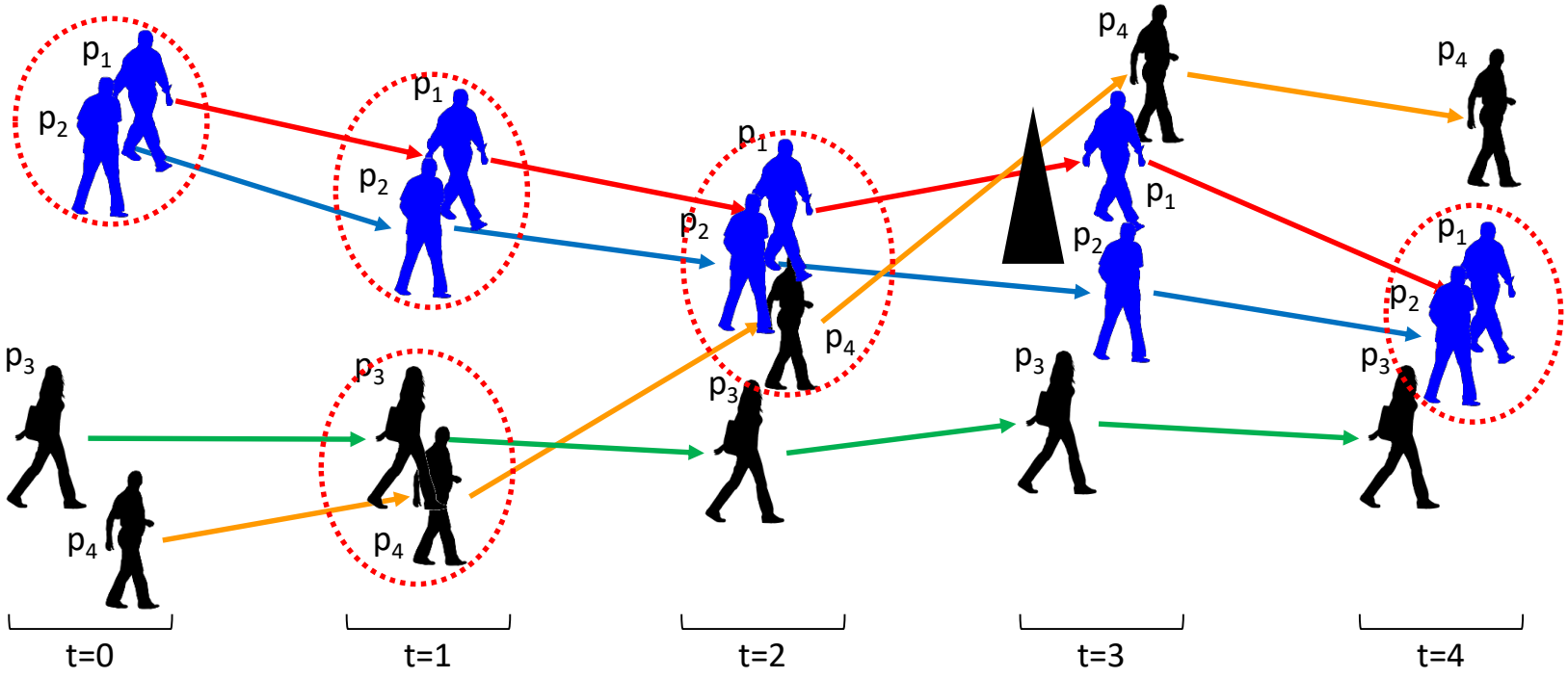


Hidden variables $\begin{pmatrix} \varphi \\ \Delta^1 \\ \Delta^2 \\ \Delta^3 \end{pmatrix}$ → **Project & Matching** → $\begin{pmatrix} r^1 \\ r^2 \\ r^3 \end{pmatrix}$ Observed variables

Demo video: https://www.youtube.com/watch?v=GgKEOTlUZxw

Convoy Detection Method

1. **Density clustering:** Find clusters of pedestrians who are close to each other
2. **Intersection:** Take intersections of clusters to extract temporally consistent ones (by relaxing the temporal continuity criterion)



Demo video: https://www.youtube.com/watch?v=p4zN39u_Waw

Cognitive Village Project

Aging population ⇔ Declining birth-rate: Lack of people who care elderly people

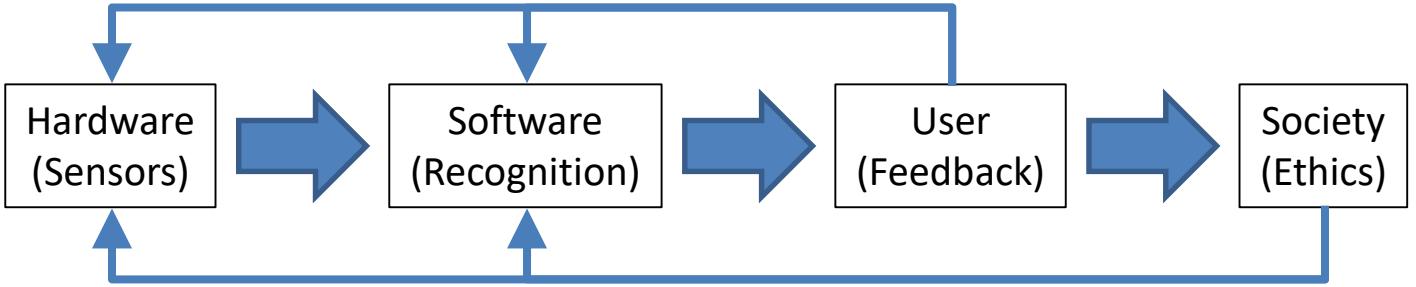
➡ Develop a system that recognises activities of an elderly person using various sensors, and support his/her independent life and healthcare

Source: Cathrin Warnke

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

"Cognitive Village: Adaptively Learning Technical Support System for Elderly" funded by German Federal Ministry of Education and Research (BMBF)

System Development through Interdisciplinary Collaboration



FoKoS ForschungsKollegSiegen

UNIVERSITÄT SIEGEN

Fraunhofer FIT

FUTURE SHAPE

Noldus Information Technology

avinotec

KSG

Website: http://www.cognitive-village.de/

Sensor-based Human Activity Recognition

Continuously record sensor data in daily life

➡ Recognise various activities of an elderly person to support his/her independent life and healthcare

JINS MEME
(JIN CO., LTD.)

Intelligent glasses
Head and eye movements
- Accelerometer
- Gyroscope
- Electrooculogram (EOG)

Microsoft Band
(Microsoft Corp.)

Wristband
Hand movements and physiological data
- Accelerometer
- Gyroscope
- Heart rate
- Galvanic skin conductance

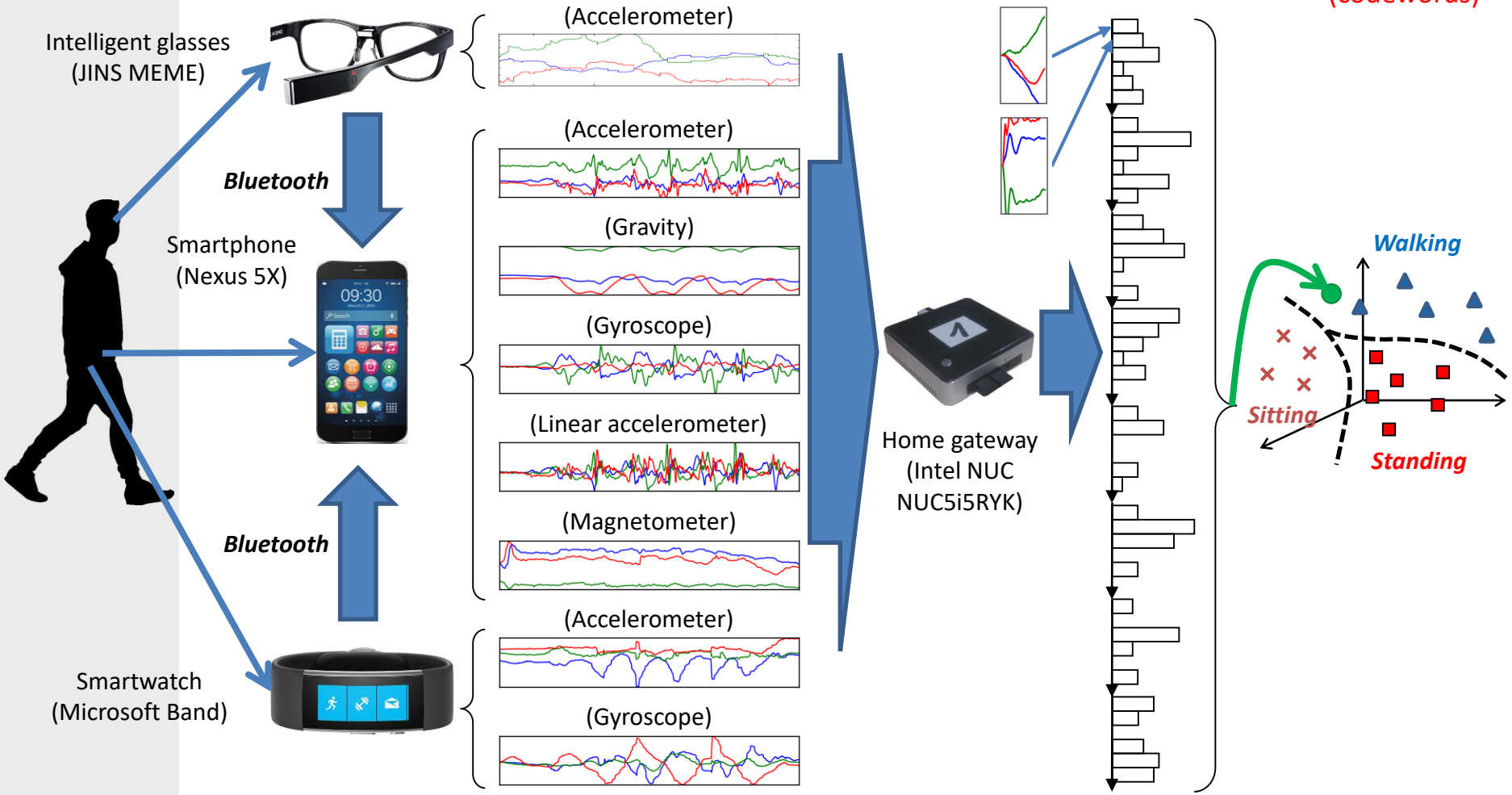
SensFloor
(Future-Shape GmbH)
Trajectories and gaits

Smartphone
Body movements
- Accelerometer
- Gyroscope
- Gravity
- Magnetometer

Prototype Activity Recognition System

It is unknown what kind of features in sensor data are useful for accurate activity recognition

➡ Feature learning: Extract a feature vector representing the distribution of statistically distinct subsequences (codewords)



Demo video (old version): https://www.youtube.com/watch?v=sIL08IE_QLE&t=115s

Demo video (new version): https://www.youtube.com/watch?v=hr3i9I5Ga0M&t=213s