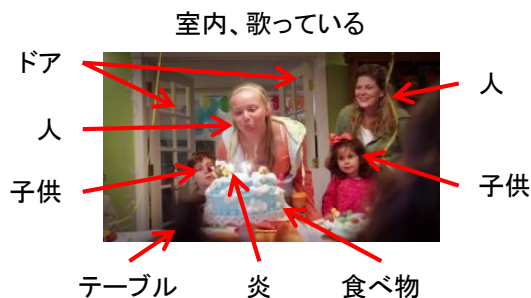


研究概要

機械学習やデータマイニング技術を用いた、マルチメディアデータからの意味情報抽出

映像検索

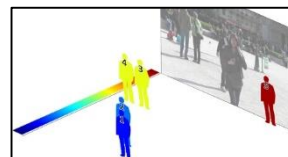


行動認識



精度改善

コンピュータビジョン



意味

セマンティック
ギャップ

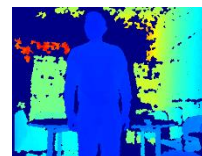
生データ

機械学習、データマイニング

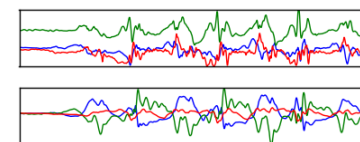
00110101001
11101000111



(ビデオ)



(深度センサ)



(ウェアラブル端末)

概念に基づく映像検索

テキスト形式で与えられたクエリ(検索要求)に適合する映像を検索する
(手動で付与されたタグは一切使用しない)

クエリ:
誕生日パーティ



関連概念:
Person
Food
Fire
Indoors
Singing etc.



物体	Person: 0.9
	Car: 0.0
	Building: 0.1
	Sky: 0.0
シーン	Food: 1.0
	Fire: 0.8
	Outdoors: 0.3
	Indoors: 0.9
動作	Beach: 0.0
	Walking: 0.2
	Throwing: 0.1
	Singing: 0.7
	⋮

概念: 人間が映像から知覚可能な意味内容の総称

1. 概念認識: 各概念の出現可能性を表す認識スコアを算出する

2. 検索: クエリに関連する概念に対する認識スコアが高い映像を検索

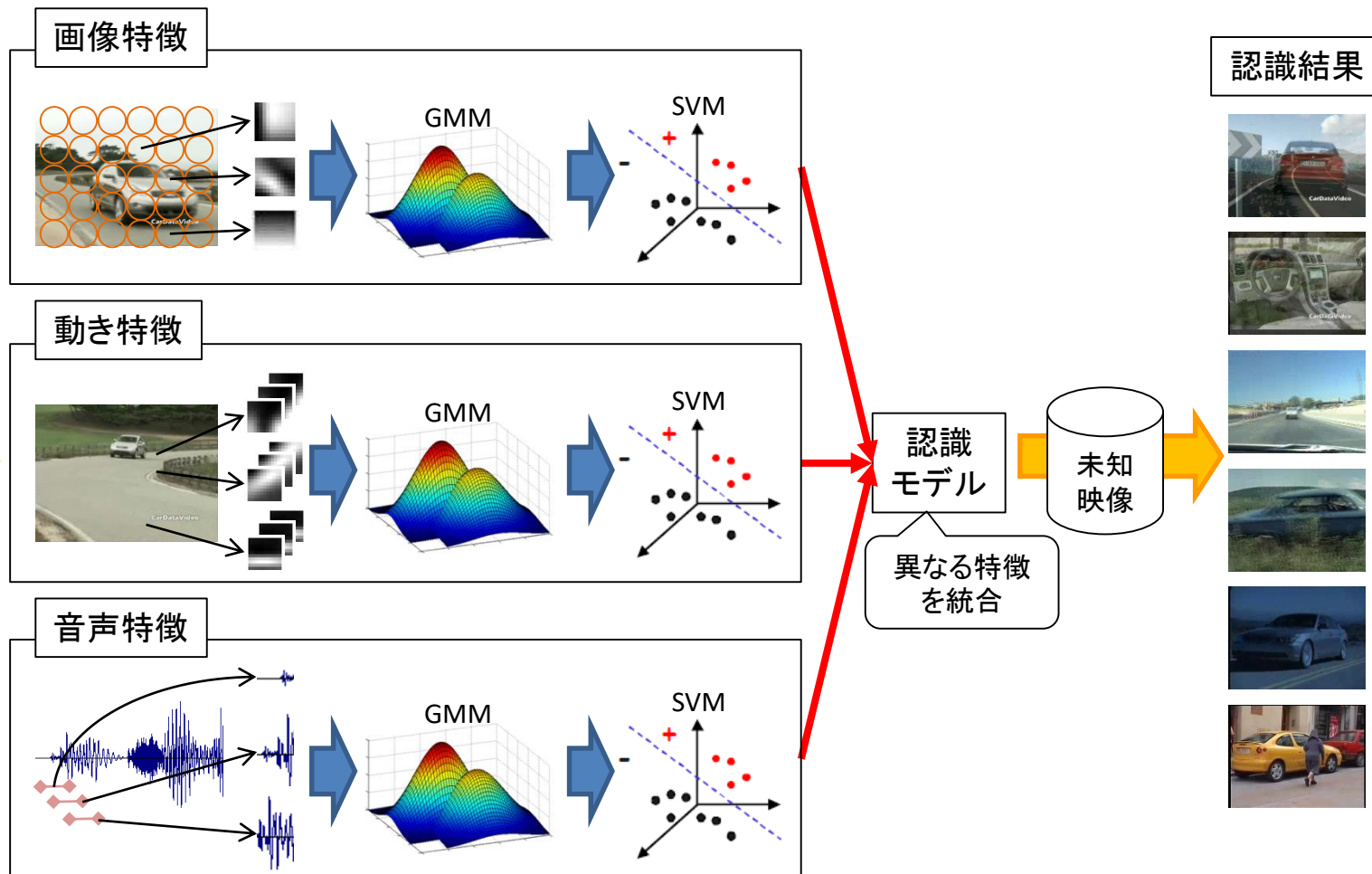
多くの概念を認識しておけば、あらゆるクエリに対応可能!

概念認識手法

- 概念の見た目の多様性 → 大量の学習用映像
- 概念の出現時間・位置の多様性 → 局所特徴の時空間的に密なサンプリング

Carの認識

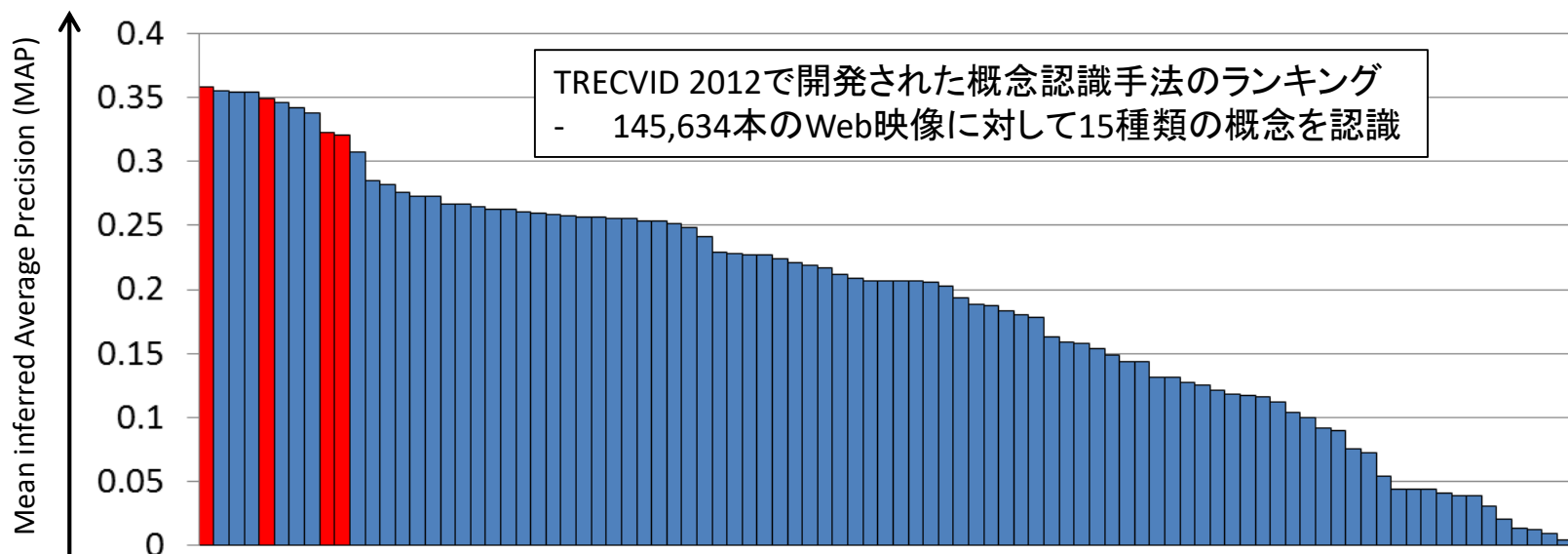
学習用映像



概念認識精度

TREC Video Retrieval Evaluation (TRECVID):

米国標準技術局 (NIST) 主催の映像解析・検索に関する国際競争型ワークショップ



認識対象
概念の例

Airplane_Flying



Boat_Ship



Instrumental_Musician



Landscape



Throwing

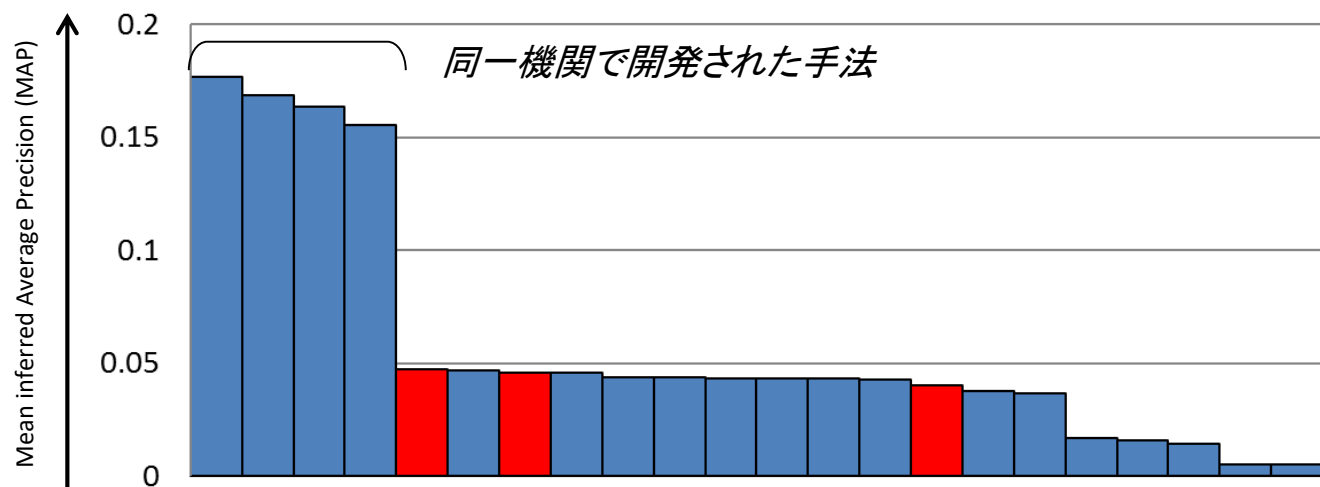


世界25機関 (IBM、スタンフォード大、CMUなど) で開発された全91手法中、
トップの概念認識精度を達成！

検索精度

TRECVID 2016 アドホック映像検索 (manually-assisted) 部門で開発された手法のランキング

- 335,944本のWeb映像から30種類のクエリに適合する映像を検索
- 各映像の検索スコアは、関連する概念に対する認識スコアの和として計算



屋外でギターを弾いている



ダイビングスーツを着て水中にいる



何かを飲んでいる



本棚を背に、カメラに向かって話している



夜間にストリートで群衆がデモを行っている

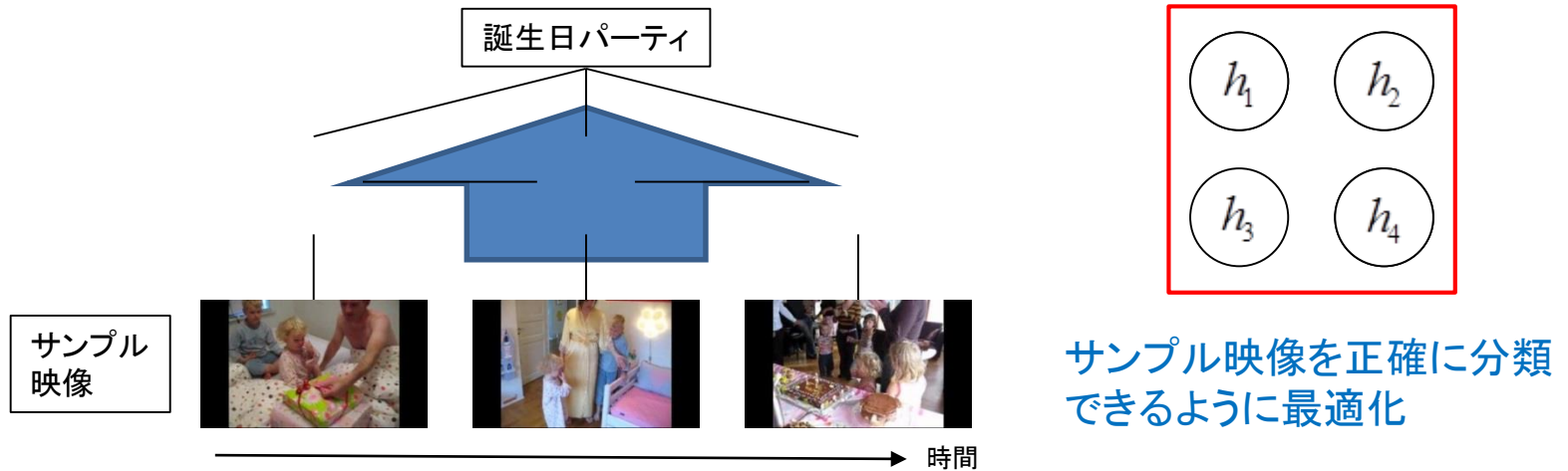


クエリ
の例

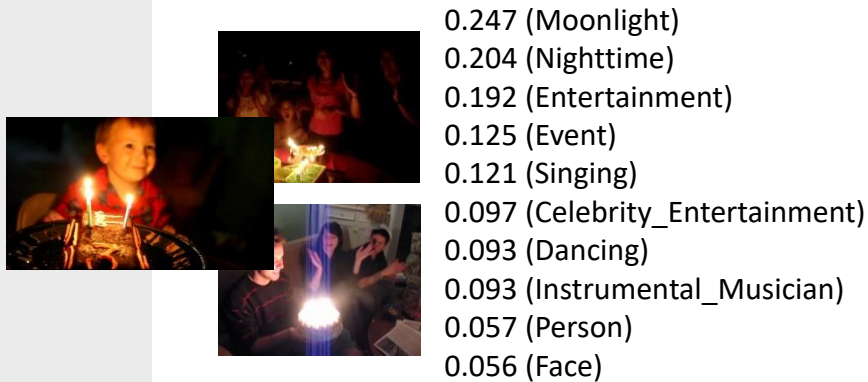
参加8機関中、第2位(全22手法中第5位)の検索精度を達成!

サンプル映像からの概念選択

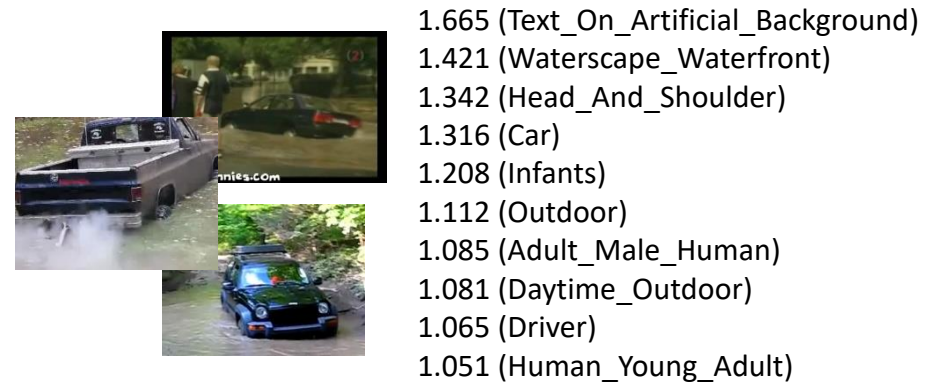
クエリに対する概念の関連性を表す隠れ状態を用いて映像を分類するモデル



➤ 「誕生日パーティ」に対する隠れ状態の例

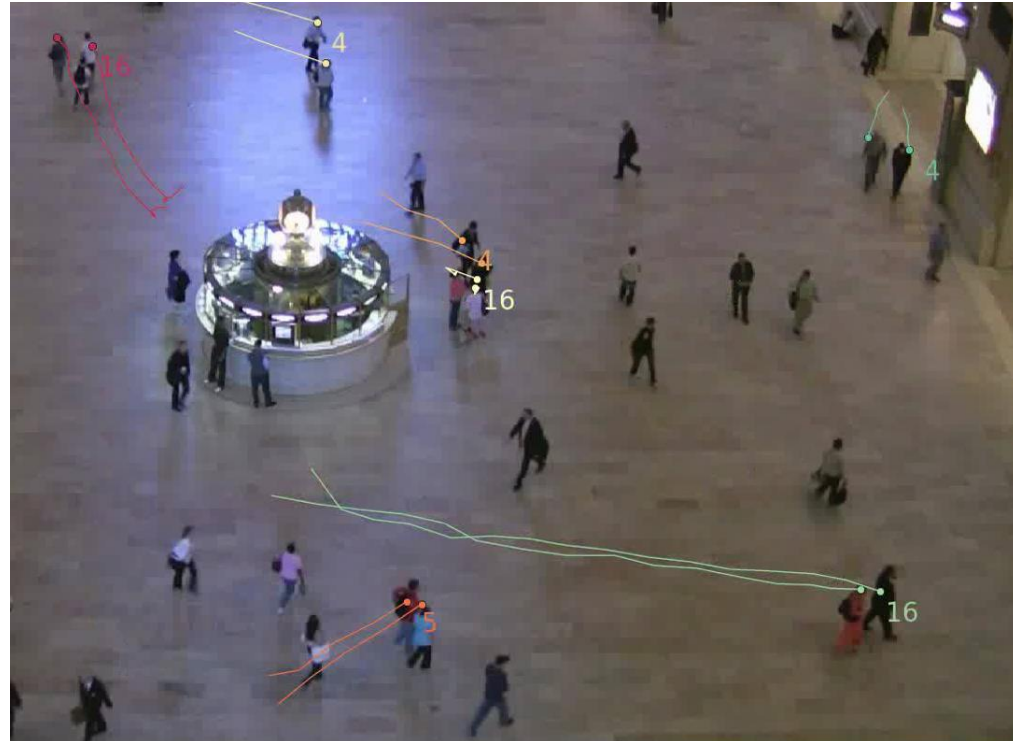


➤ 「スタックした車を脱出させる」に対する隠れ状態の例



人物のグループ(コンボイ)の抽出

多数の人物が映る監視カメラ映像における膨大な情報量



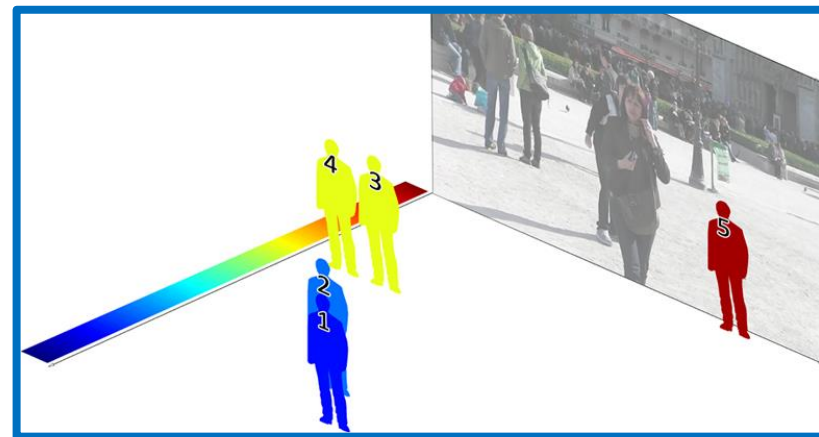
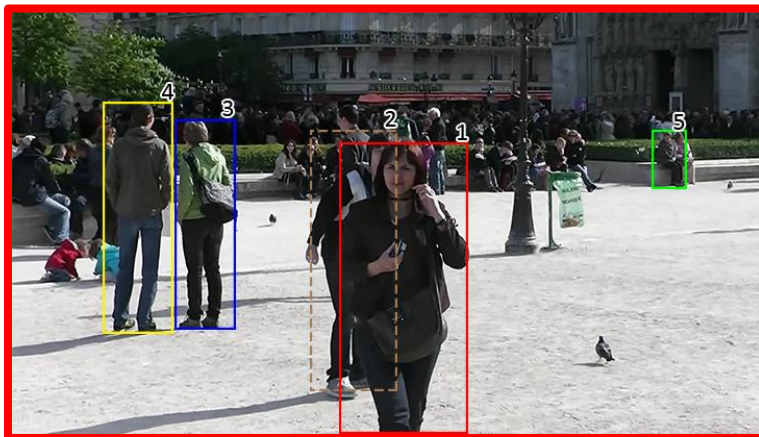
疑わしい行動を自動的、もしくは対話的に検出可能なシステムの必要性

共に動いている歩行者のグループをコンボイとして抽出

1. 各人物の軌跡抽出
2. 軌跡解析によるコンボイ検出

映像における3次元情報の欠如

高精度なコンボイ検出には、人物間の空間的関係性を検証する必要がある



- ✓ 元の3次元空間が、2次元フレームに写像されている
- ✓ 人間は、2次元フレームから容易に3次元の空間的関係性を認識できる



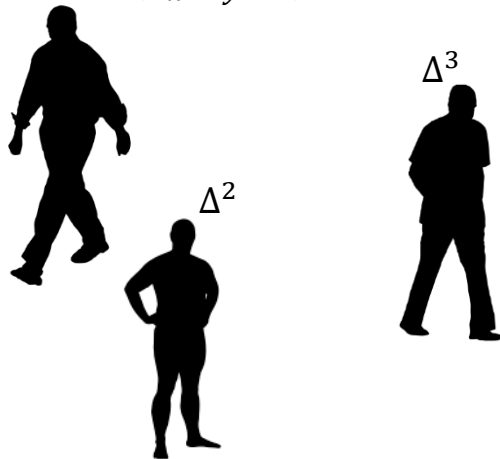
2次元フレームの系列である映像から**3次元軌跡**を抽出
(3次元空間における物体位置の推移を表す)

確率的な3次元軌跡抽出

2次元フレームにおける物体の映り方に対して、最も適切な3次元空間における物体とカメラ位置を推定

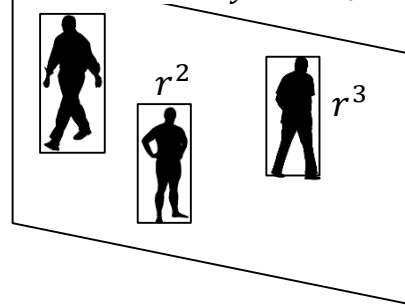
3次元空間における物体位置

$$\Delta^1 = (\Delta_x^1, \Delta_y^1, \Delta_z^1)$$



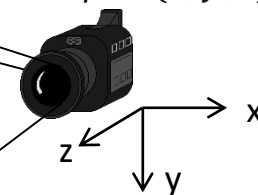
2次元フレーム中の物体領域

$$r^1 = (r_x^1, r_y^1, r_w^1, r_h^1)$$



3次元空間におけるカメラ位置

$$\varphi = (x, y, z)$$



隠れ変数

$$\begin{pmatrix} \varphi \\ \Delta^1 \\ \Delta^2 \\ \Delta^3 \end{pmatrix}$$



2次元フレームに
写像しマッチング

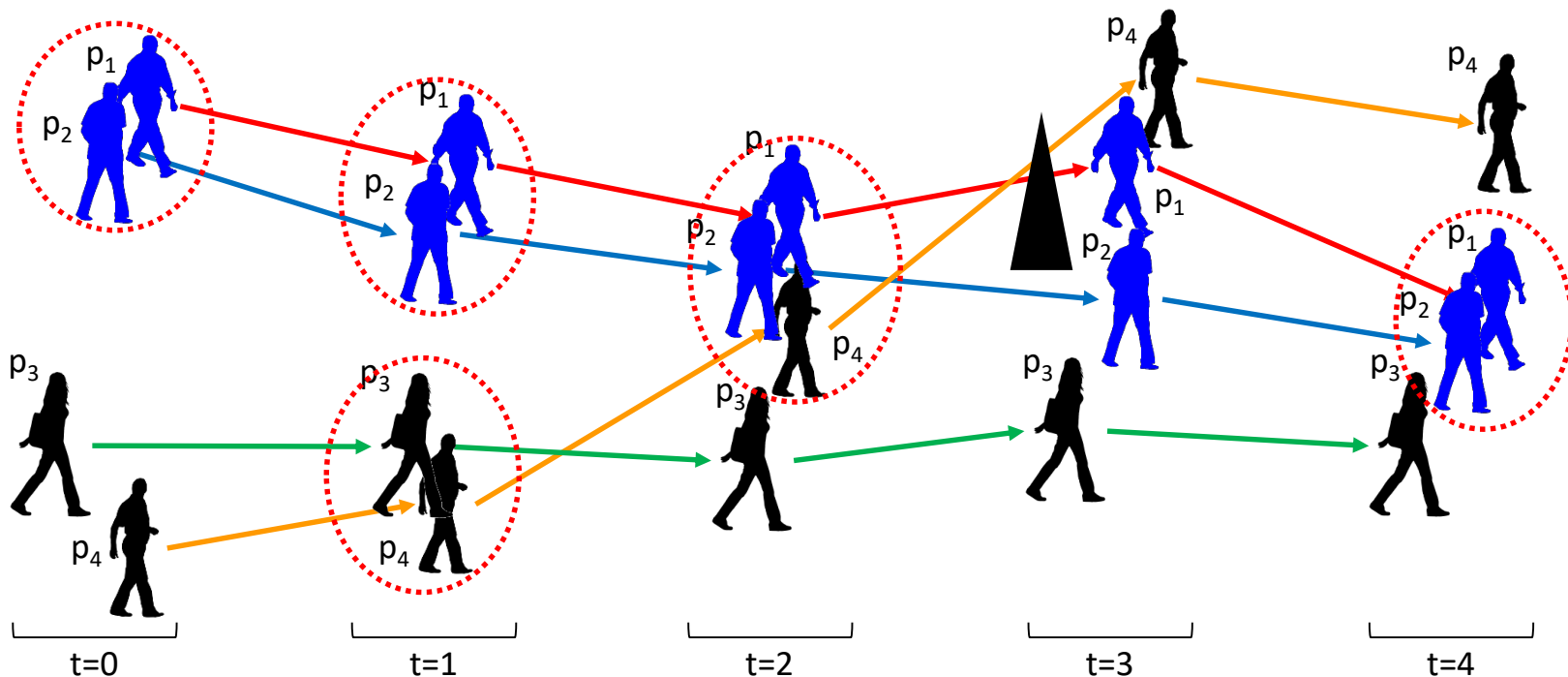
$$\begin{pmatrix} r^1 \\ r^2 \\ r^3 \end{pmatrix}$$

観測変数

デモ映像: <https://www.youtube.com/watch?v=GgKEOTIUZxw>

コンボイ検出手法

1. **密度クラスタリング**: 互いに近い位置にいる人物のクラスターを抽出
2. **共通部分抽出**: 前後の時間に抽出されたクラスター間での共通部分をとって、時間的に一貫性のある人物のグループを特定 (共通部分の短時間の不連続性は許容)



デモ映像: https://www.youtube.com/watch?v=p4zN39u_Waw

Cognitive Villageプロジェクト



高齢化 ⇔ 少子化: 高齢者の面倒を見る人手が不足

➡ 様々なセンサを用いて、高齢者の行動を認識・監視し、自立生活や健康管理をサポートするシステムを開発



Source: Cathrin Warnke

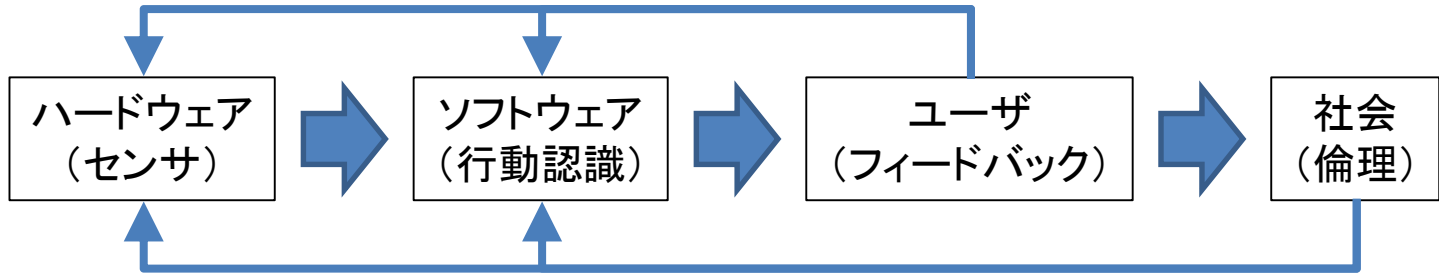
GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

“Cognitive Village: Adaptively Learning Technical Support System for Elderly”
ドイツ連邦教育研究省 (BMBF)

異分野交流に基づくシステム開発



FUTURE SHAPE



ウェブサイト: <http://www.cognitive-village.de/>

センサデータからの行動認識

日常生活下でセンサからデータを継続的に収集し、ユーザの行動を認識

➡ 高齢者の様々な行動を認識し、自立生活や健康管理をサポート



JINS MEME
(JIN CO., LTD.)

インテリジェントグラス

頭と眼の動き

- 加速度計
- 角速度計
- 眼電位 (EOG)



Microsoft Band
(Microsoft Corp.)

スマートウォッチ

手の動きと生理学データ

- 加速度計
- 角速度計
- 心拍数
- 皮膚コンダクタンス



スマートフォン

身体の動き

- 加速度計
- 角速度計
- 重力計
- 磁力計

SensFloor
(Future-Shape GmbH)
移動軌跡と歩行

プロトタイプ行動認識システム

センサデータのこういった特徴が認識に有用か分からない

➡ 特徴学習: 統計的に特徴的な部分系列(コードワード)を抽出し、コードワードの分布を表すベクトルを抽出



デモ映像(旧バージョン): https://www.youtube.com/watch?v=sIL08IE_QLE&t=115s

デモ映像(新バージョン): <https://www.youtube.com/watch?v=hr3i9I5Ga0M&t=213s>