

条件付き確率場を用いた 同一物体検出による 三次元位置推定

松田脩佑 テチャサンティクーンナタオン 義久智樹 下西英之

大阪大学 大学院情報科学研究科

大阪大学 サイバーメディアセンター



背景

三次元空間のデジタルツイン構築の需要

例) スマート倉庫

人の安全に配慮した高効率なロボット制御の実現

- 人手不足解決のため、搬送ロボットが実践投入
- 無人-高効率/ 有人-非効率の二極化
- 有人状況下におけるロボットの作業効率向上

三次元位置推定方法

- LiDAR: 高コスト
 - 深度センサ: ノイズが大きい, 範囲が狭い
- **複数RGBカメラによる三次元位置推定**

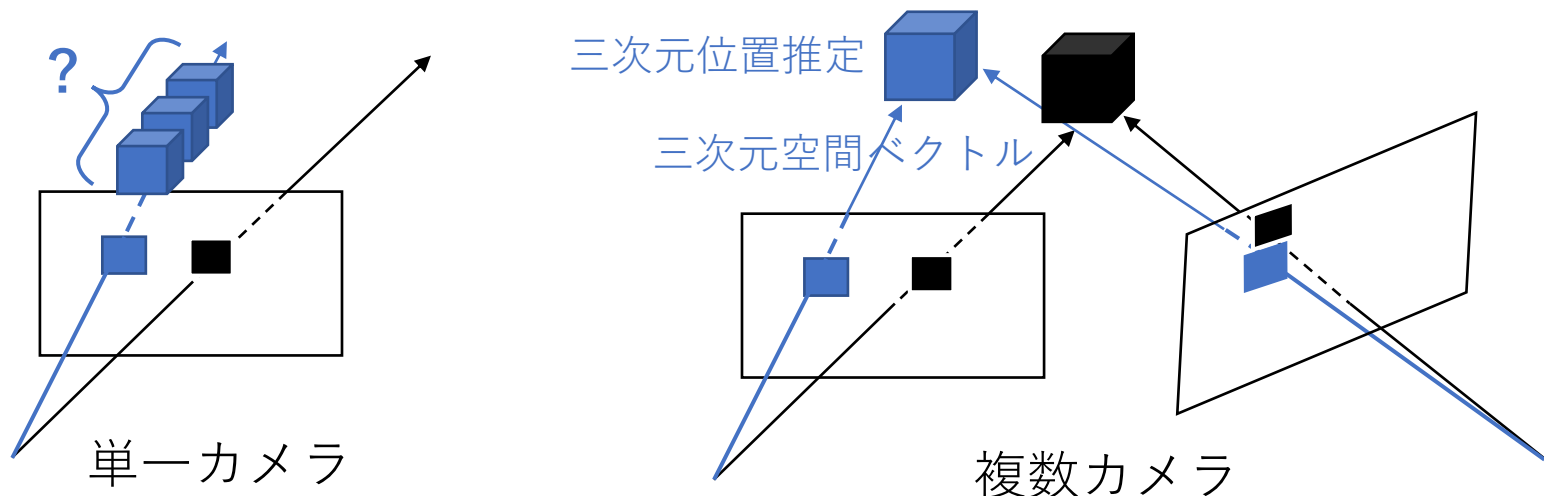
三次元位置推定

単一カメラ

- 画像平面内の二次元位置を高精度に計測
- 奥行方向の情報が欠落

複数カメラ

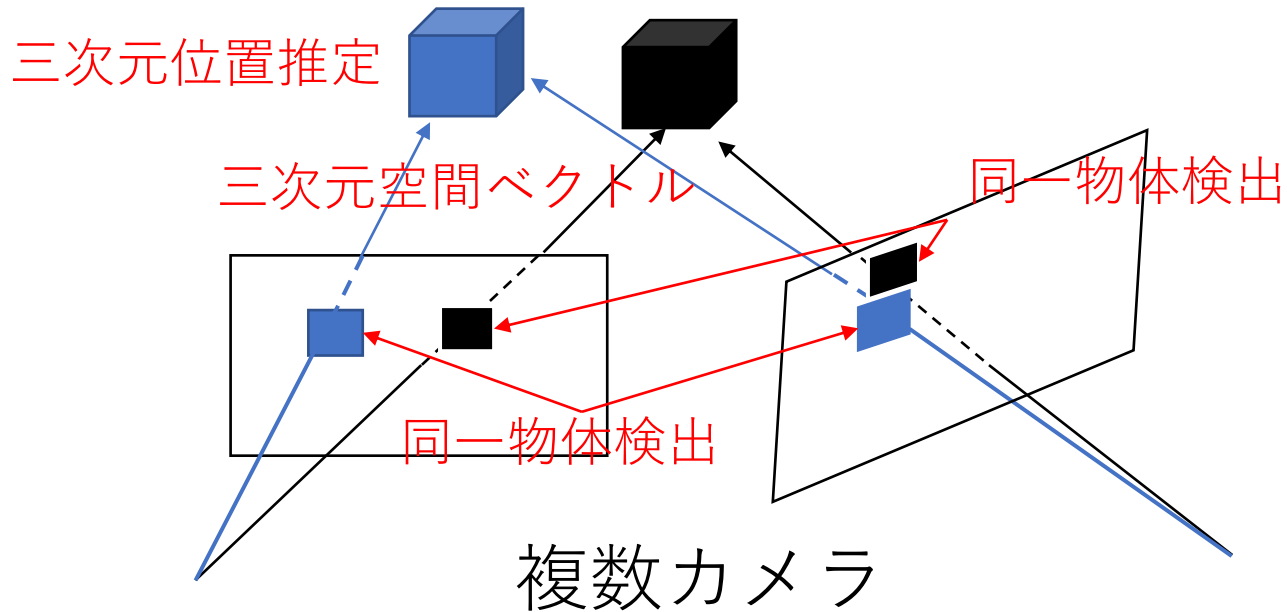
- 画角の異なる複数のカメラで同一空間をカバー
- 複数の三次元空間ベクトル(カメラから物体に向けて引いた直線)の近接点から三次元位置を推定



研究目的

同一物体検出による三次元位置推定

- Step. 1 : 各カメラで捉えた物体から同一物体を検出
- Step. 2 : 三次元空間ベクトルの近接点から三次元位置を推定



同一物体検出に関する関連研究

画像による同一物体検出[1]

- 背景や服装による精度低下
- ロボットや貨物等の分類が困難

位置による同一物体検出[2,3]

- 人物のみを対象
- 高さ情報の欠落
- 特定の平面に接していることが前提

- [1] Han, R., Wang, Y., Yan, H., Feng, W., & Wang, S. (2022). Multi-View Multi-Human Association With Deep Assignment Network. *IEEE Transactions on Image Processing*, 31, 1830-1840.
- [2] Lima, J. P., Roberto, R., Figueiredo, L., Simoes, F., & Teichrieb, V. (2021). Generalizable multi-camera 3d pedestrian detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.
- [3] López-Cifuentes, A., Escudero-Viñolo, M., Bescós, J., & Carballeira, P. (2022). Semantic-driven multi-camera pedestrian detection. *Knowledge and Information Systems*, 64(5), 1211-1237.

同一物体検出手法の提案

マルチモーダルな同一物体検出手法

1+2によるマルチモーダルな処理を行うことで
三次元情報のまま検出精度を向上

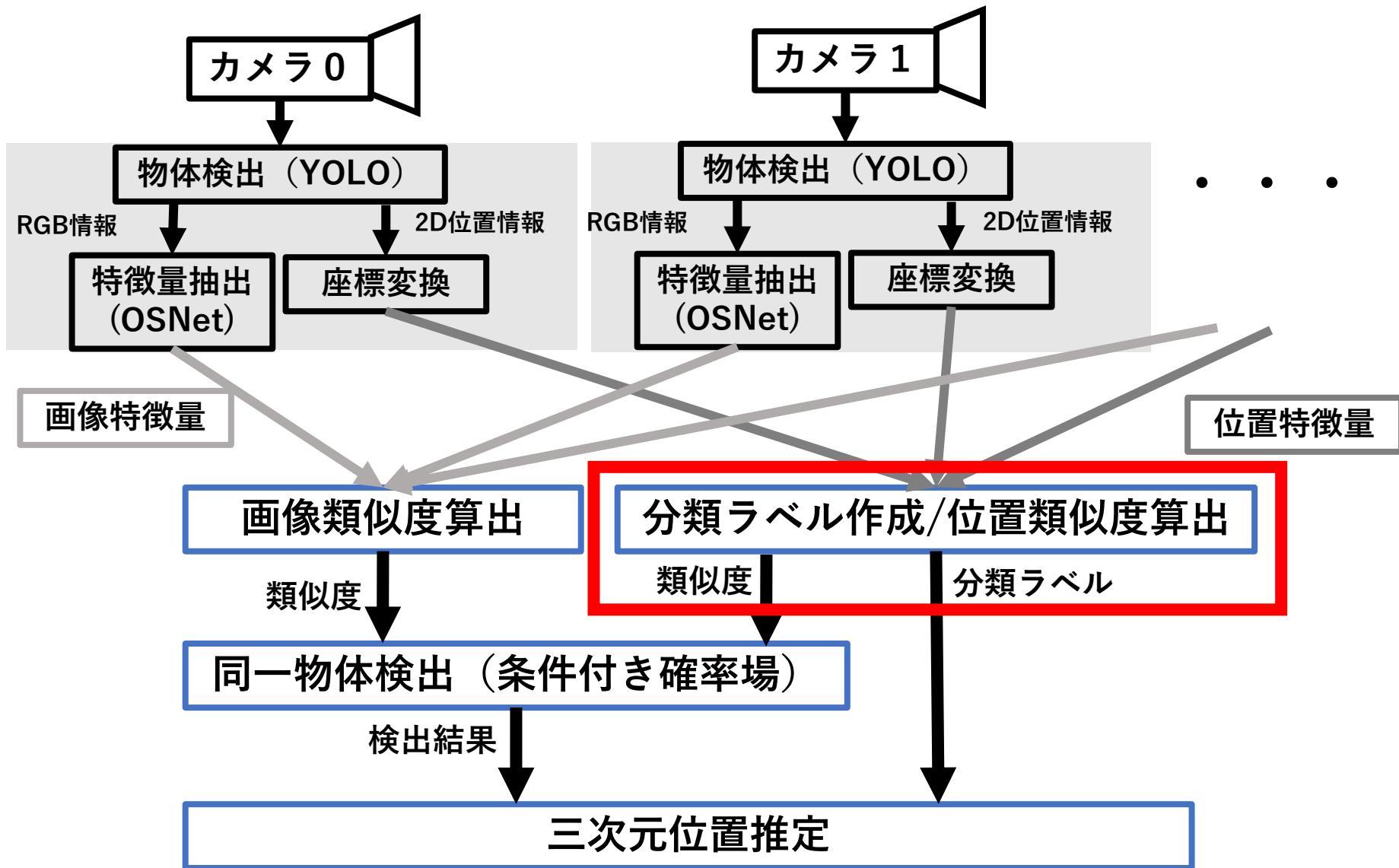
1. 位置による類似度

- 三次元空間上の二次元ベクトルの近さによって類似度を表現(地面に投影せず三次元情報のまま扱う)

2. 画像による類似度

- 画像特徴量の近さによって類似度を表現(既存手法と同様)

提案手法の概要①



分類ラベル作成

各カメラで捉えた物体の位置から三次元位置の候補(分類ラベル)を作成

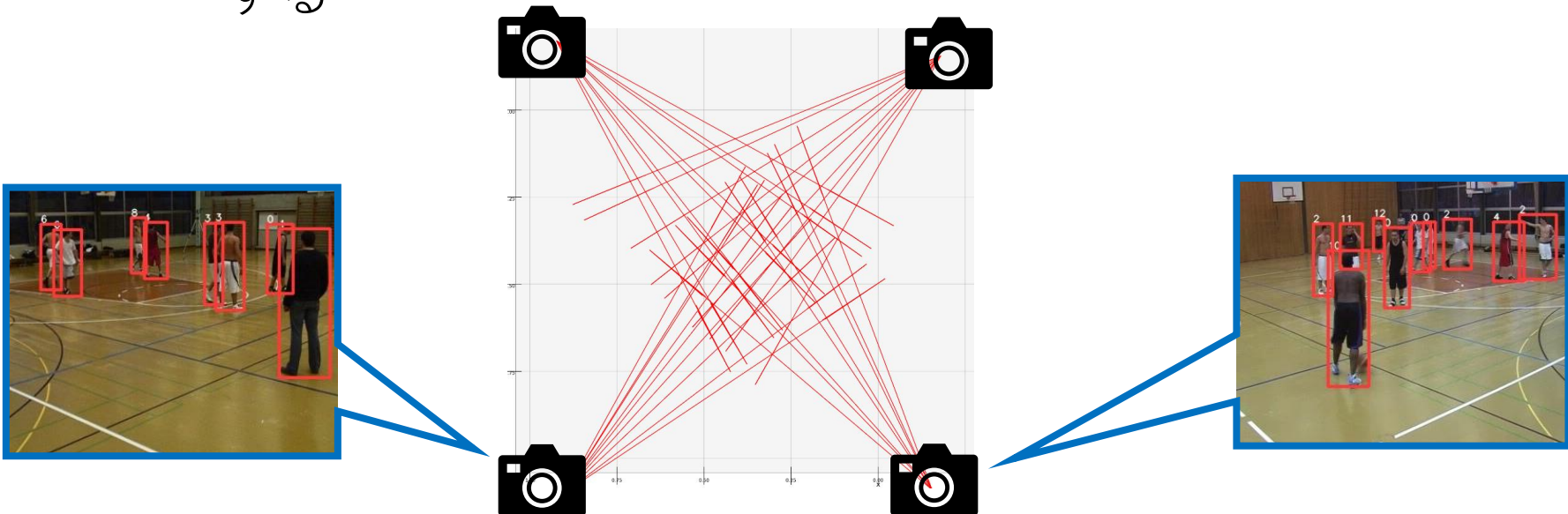
1. 各カメラの設置位置や角度，焦点距離の導出
2. 検出物体の中心点を三次元の直線に変換
3. 全直線間の共通垂線の中点座標を算出
4. クラスタを分類ラベルに，
クラスタと3.の座標の距離をラベルとの類似度とする



分類ラベル作成

各カメラで捉えた物体の位置から三次元位置の候補(分類ラベル)を作成

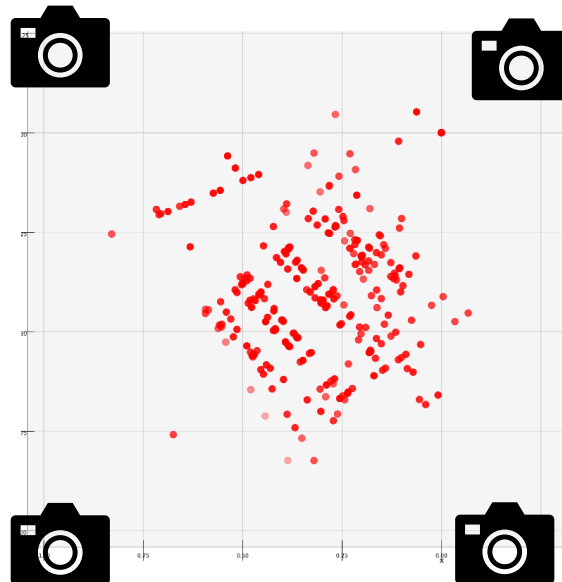
1. 各カメラの設置位置や角度, 焦点距離の導出
2. 検出物体の中心点を三次元の直線に変換
3. 全直線間の共通垂線の中点座標を算出
4. クラスタを分類ラベルに,
クラスタと3.の座標の距離をラベルとの類似度とする



分類ラベル作成

各カメラで捉えた物体の位置から三次元位置の候補(分類ラベル)を作成

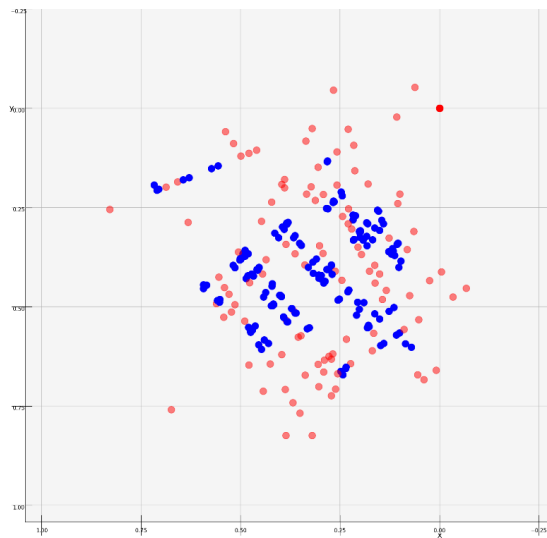
1. 各カメラの設置位置や角度，焦点距離の導出
2. 検出物体の中心点を三次元の直線に変換
3. 全直線間の共通垂線の中点座標を算出
4. クラスタを分類ラベルに，
クラスタと3.の座標の距離をラベルとの類似度とする



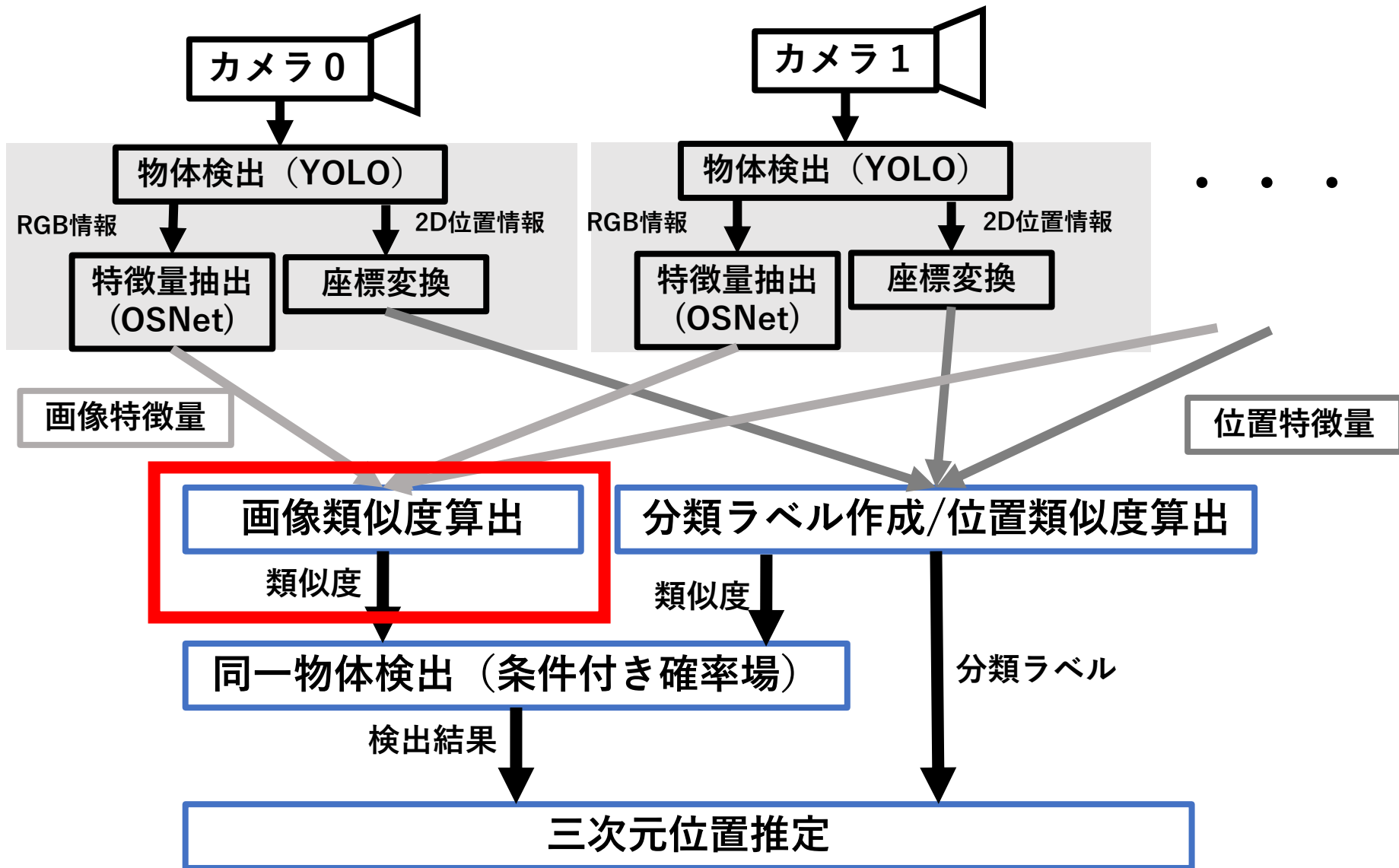
分類ラベル作成

各カメラで捉えた物体の位置から三次元位置の候補(分類ラベル)を作成

1. 各カメラの設置位置や角度, 焦点距離の導出
2. 検出物体の中心点を三次元の直線に変換
3. 全直線間の共通垂線の中点座標を算出
4. クラスタを分類ラベルに,
クラスタと3.の座標の距離をラベルとの類似度とする



提案手法の概要②

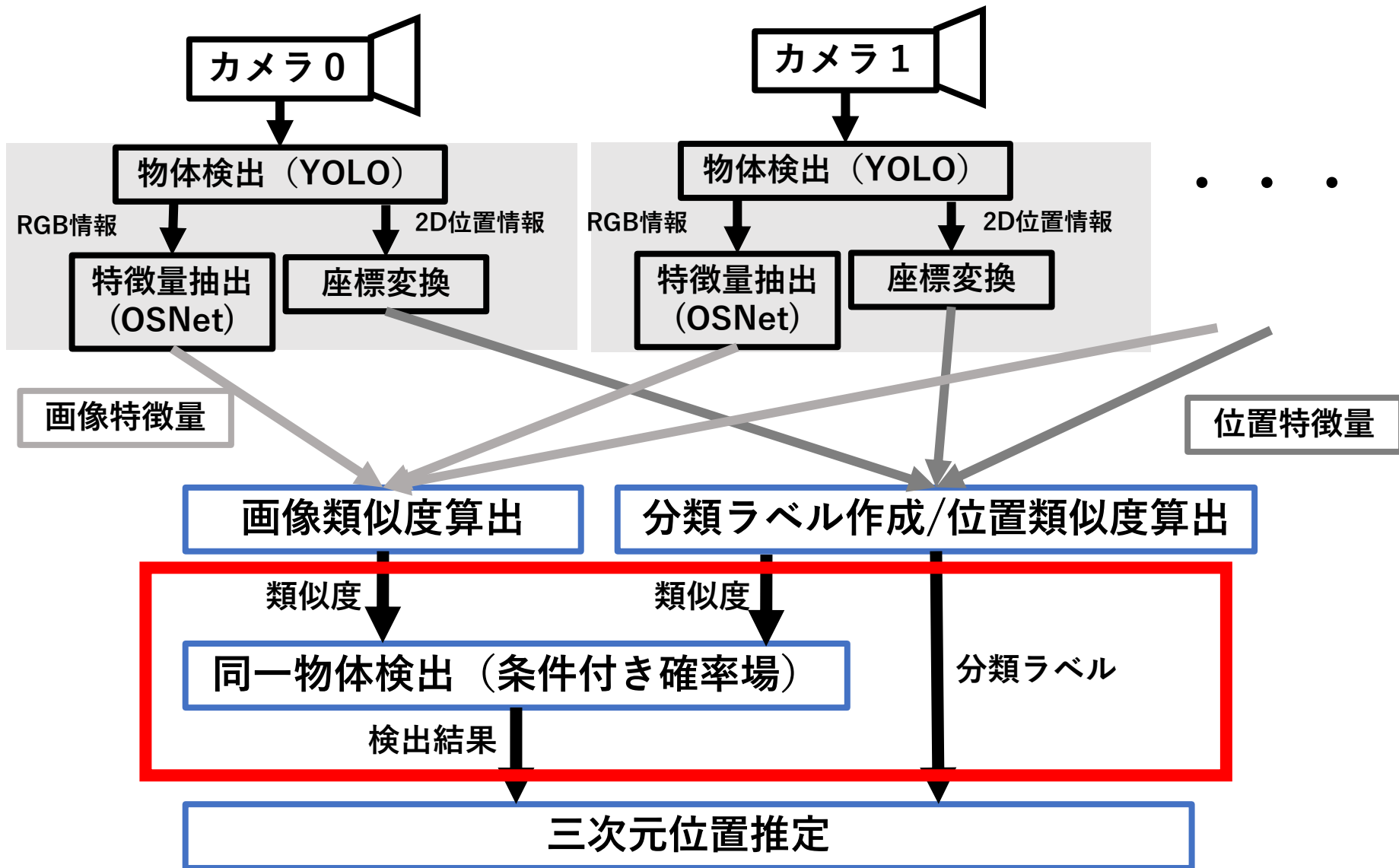


類似度算出

検出物体のRGB情報から特徴量を抽出し、
検出物体間の類似度を算出

- 特徴量抽出には既存のCNNモデルを利用
- 抽出された512次元の特徴量ベクトルから
コサイン類似度により類似度を算出

提案手法の概要③



条件付き確率場

グラフィカルなラベリング手法

- 推論対象同士の依存関係を考慮
- 更新を繰り返し高類似度な対象とのラベル付けに影響
- セグメンテーション/ 品詞ラベリングに利用

単項とペアワイズ項によるラベル更新

ペアワイズ項により高類似度物体を同ラベルに近づけつつ，単項も意識してラベルを更新

- 単項 : 分類ラベルとの類似度
- ペアワイズ項 : 物体間の類似度

条件付き確率場による同一物体検出

単項：分類ラベルとの類似度

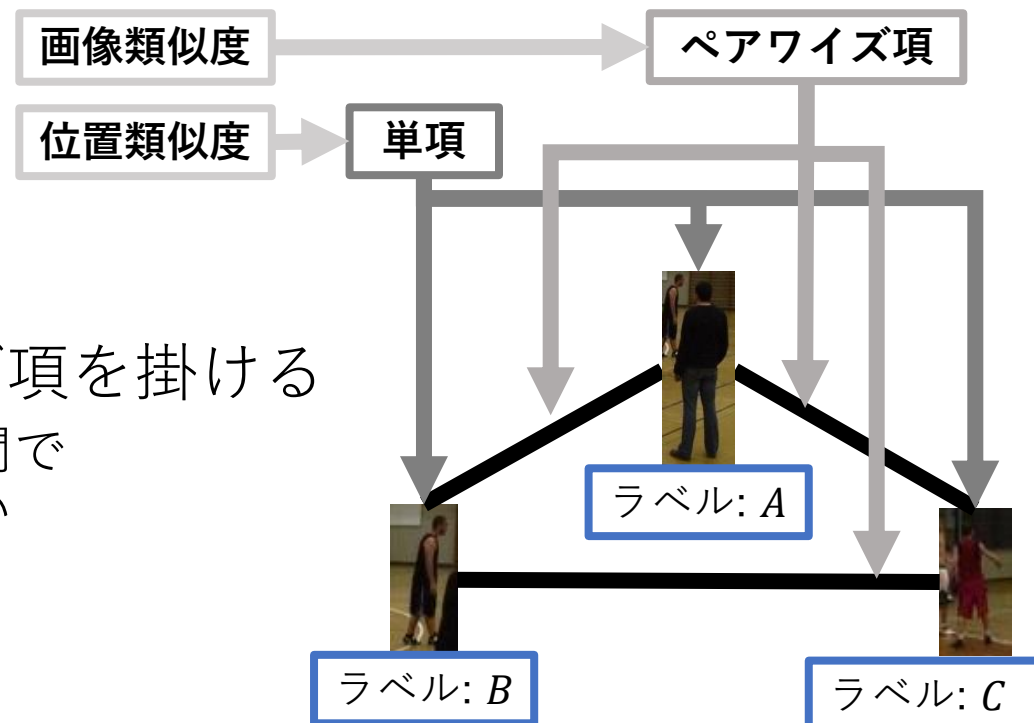
検出物体 i の分類ラベル l に対する類似度: S_{il}^l

ペアワイズ項：検出物体間の類似度

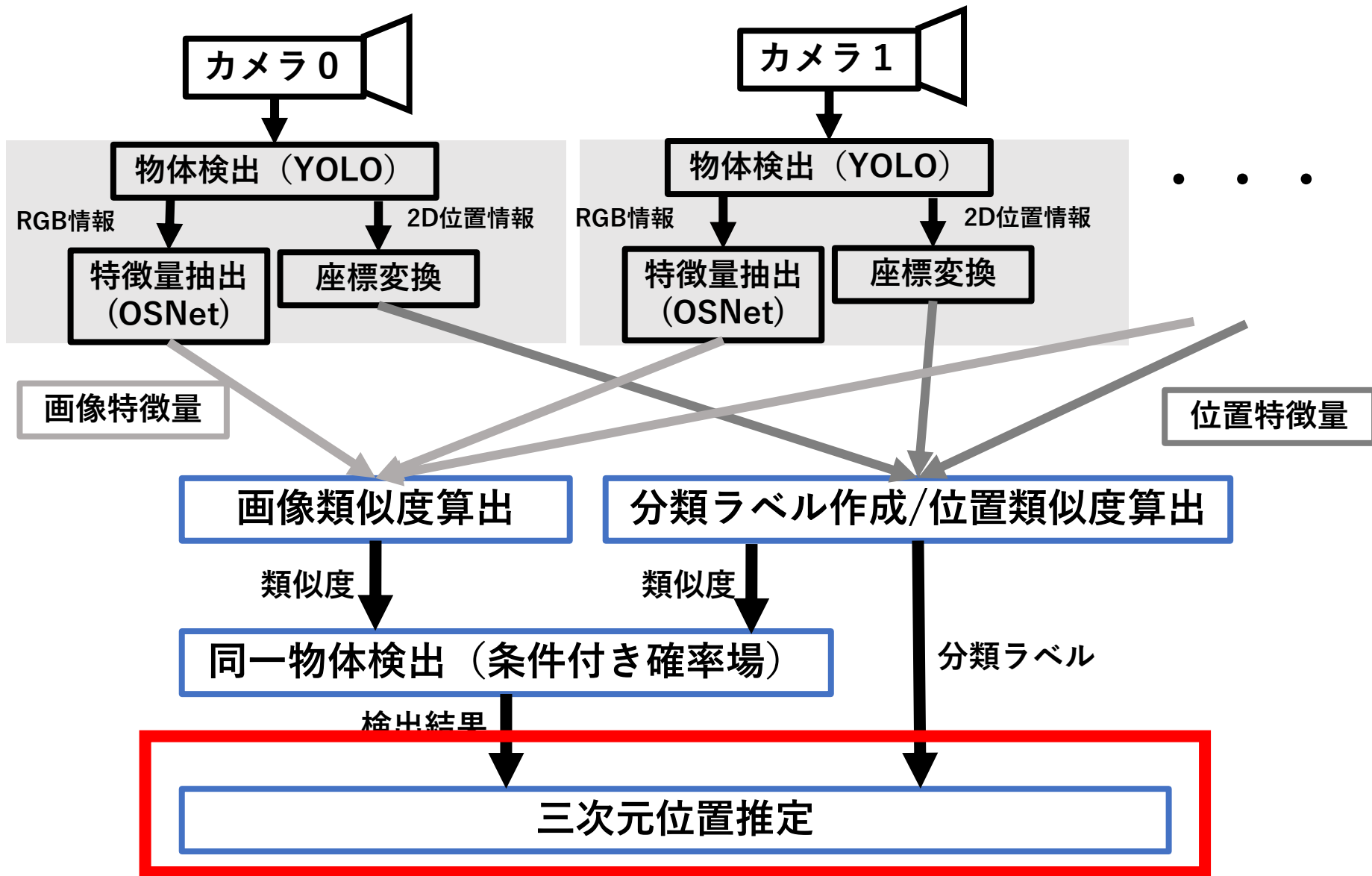
検出物体 i, j 間の画像類似度: S_{ij}^v

グラフ更新

- 初期状態：単項
- 前状態にペアワイズ項を掛ける
高類似度の検出物体間で
同ラベルになりやすい



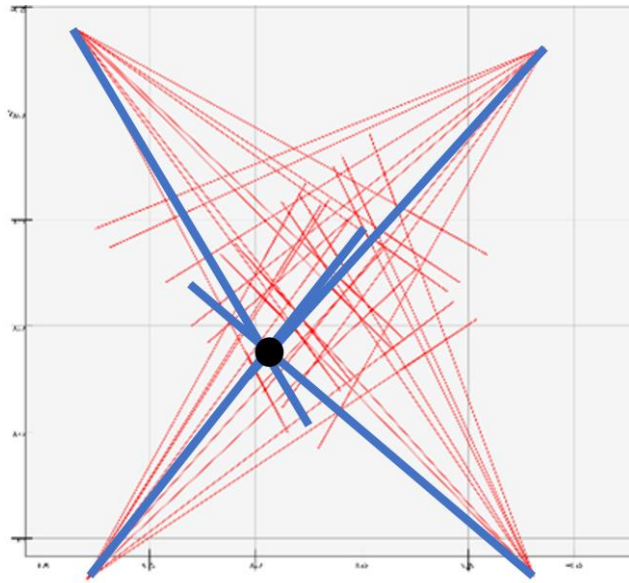
提案手法の概要④



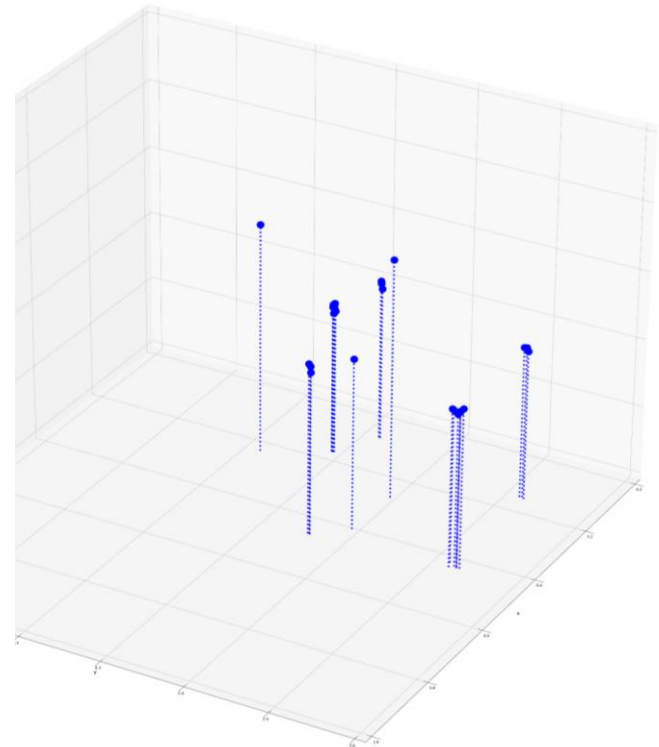
位置推定

検出した同一物体を表す直線から位置を算出

- 三次元での位置を算出可能



同一物体を表す直線の中点座標



算出された三次元位置

評価実験

評価環境

評価環境

- CPU : Intel Core i7- 9700K @ 3.60GHz
- RAM : 32GB
- GPU : NVIDIA GeForce RTX 2080 8.0GB

データセット

epfl basketball データセット

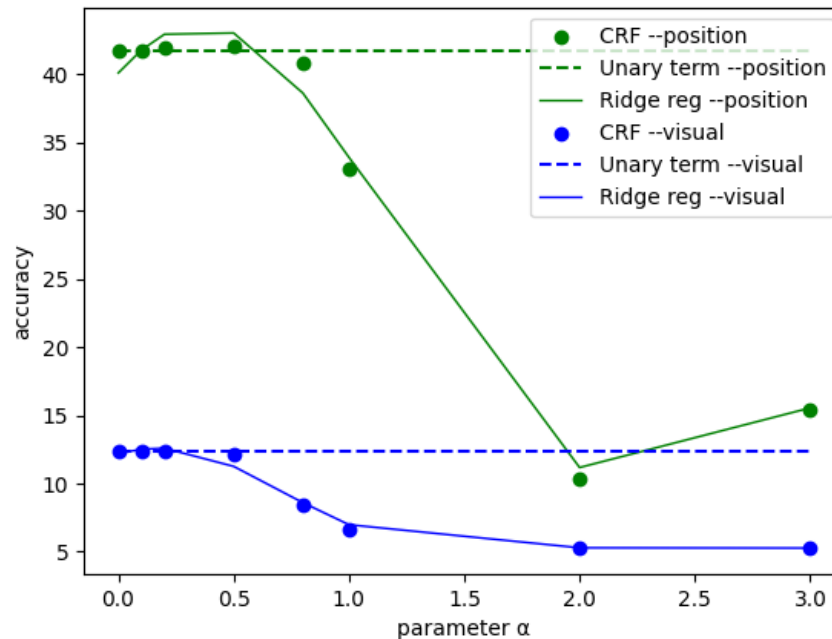
- カメラ台数 4
- 約6分, 25fps, 9367 フレームの映像データ セット



位置類似度の評価

画像類似度を単項に用いた場合と比較

- 提案手法
- 単項に画像類似度を用いた場合



- 位置類似度による精度向上
- パラメータを適切に設定することで単項のみの場合より高い精度を達成

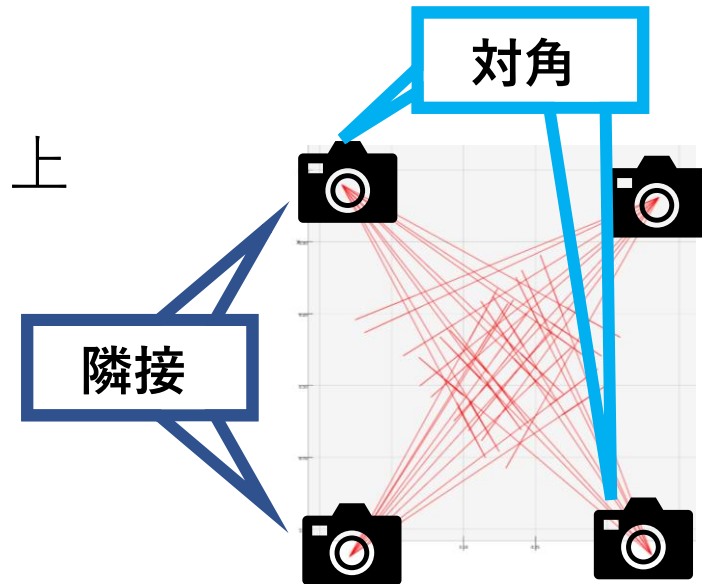
カメラ台数の評価

カメラ台数による精度の変化

- カメラ台数の増加により精度向上
- 実行時間はカメラ台数に比例

カメラ 2 台の場合

- 2 視点は対象外
2 直線の成す点をクラスタリング => 3 視点以上必要
- 対角, 隣接時で精度が大きく異なる
 - 隣接時のが密
 - クラスタが生成されるため



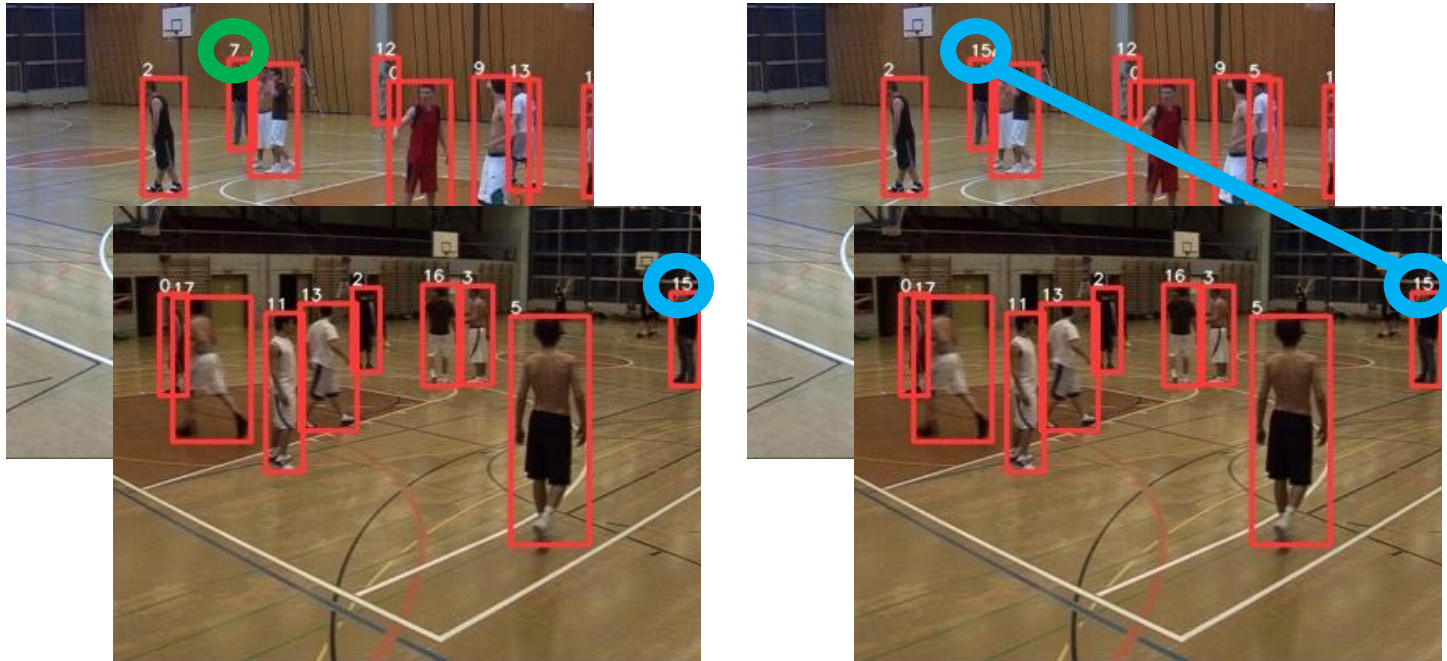
	カメラ台数			
	2 台 (対角)	2 台 (隣接)	3 台	4 台
精度 [%]	0.6	2.6	25.3	41.7
実行時間 [sec]	580	630	860	1250

条件付き確率場の評価

条件付き確率場により僅かだが精度向上

	カメラ台数			
	2台 (対角)	2台 (隣接)	3台	4台
提案手法	0.5	2.3	25.9	42.1
単項のみ	0.6	2.6	25.3	41.7

条件付き確率場により改善した例

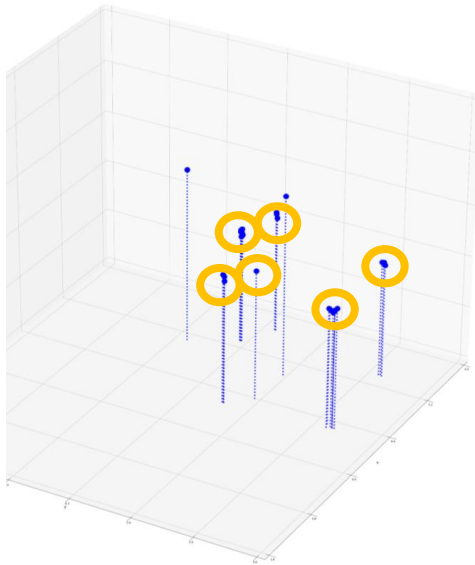


位置推定の評価

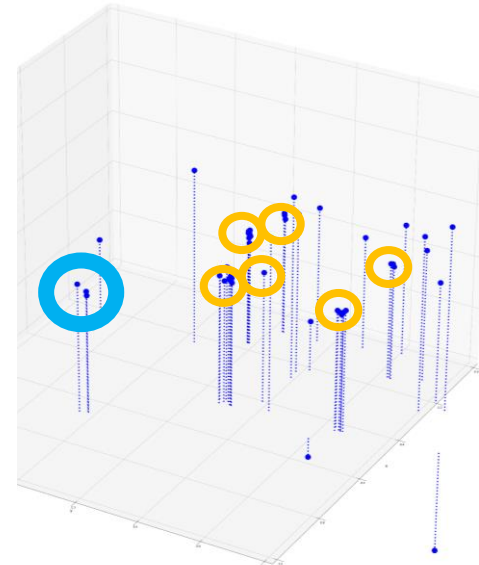
位置推定の結果

利用したデータセットに位置の正解データが無く
定量的な精度評価はできていない

単項では推定できなかった位置を推定可能



単項のみ



条件付き確率場

複数カメラによる三次元位置推定手法の提案

- 画像および位置類似度の利用
- 条件付き確率場による複数類似度を考慮した同一物体検出
- 手法内比較により精度向上を確認

課題：精度向上

- データセット (2007年公開)
 - 画像サイズが小さい(720, 576)
 - 動きが激しくブレる
- 条件付き確率場
 - セグメンテーション：同ラベル→別ラベルへの作用
類似物体のラベルを統合する作用が必要

