Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

多数のカメラを活用した空間認識処理に関する研究

神村涼[1]　池上努[2]　工藤知宏[1]

概要：同一領域を複数のカメラで同時撮影し，それらのデータを統合することで空間の三次元的な構成要素を特定する手法について報告する．エッジデバイス自身で一次解析処理を行うスマートカメラを用いることで，撮影画像の集約を避け，通信量の削減とプライバシー保護の両面を実現する．撮影した画像を個々のカメラ上で CNN を用いた物体検出手法である YOLO により処理し，得られた画像解析情報をサーバに送信する．サーバ側ではこれらの情報とカメラの三次元位置情報から，領域に存在するオブジェクトの種類と位置を再構成する．3 台のカメラの情報を用いることで，2 台の情報では確定できない複雑なシーンであっても一意的な再構成が可能であることを示した．

Spatial Recognition by using Surrounding Observation Cameras

RYO KAMIMURA[†1]　TSUTOMU IKEGAMI[†2]
TOMOHIRO KUDOH[†1]

1. はじめに

昨今，監視や見守りの自動化を目的として，機械学習のビデオモニタリングへの応用が進められている．これらの研究の多くは，撮影した生データをクラウドに集約し分析する手法を用いている[1]．このような手法では，通信量が膨大になるだけでなく，プライバシーなどの問題が生じる可能性がある[2]．また，ビデオモニタリングではカメラと対象物の位置関係次第では，形状や数量などについて正しい情報が得られない恐れがある．後者の問題は複数カメラの情報を統合して解析・活用することで解決することができるが，この場合，通信量やプライバシーなどの問題はより深刻になる．

このジレンマを解消する上で，エッジデバイス上での画像データ処理は有効な手段である．カメラに付随したエッジデバイス上で撮影画像に対して物体検出などの画像処理を施し，抽象化されたデータのみをクラウド上に送信する．これにより，通信量の削減だけでなく，プライバシーの保護を実現できる．また，同一シーンを多数のカメラで同時撮影しても通信量は小さく，死角の問題を容易に解決できる．これを実現するためには，高度なデータ処理機能を備えた監視カメラを多数配置する必要があるが，近年，この目的に則した安価なカメラ付きデバイスが製品化されるようになってきた．その一つ，HUSKYLENS は，顔認識やオブジェクト追跡などの機械学習アルゴリズムが組み込まれたシングルボードコンピュータで，検出結果をシリアルインターフェース経由で外部に送信する．HUSKYLENS の MPU，Kendryte K210 は機械学習専用のモジュールを備えた System On Chip (SoC)で，物体検出や音声処理を低消費電力で実現する．同じ MPU を用いたテストボード，Sipeed MAIX Bit の動作例を Figure 1 に示す．ここでは Micro Python を介して Tiny YOLOv2 を動かし，カメラで撮影したデータをリアルタイムで物体検出処理をしている．LCD 上に物体検出した枠とそのカテゴリ（Potted Plant と Car）が表示されていることがわかる．これら一連の処理にかかる電力は LCD（バックライト付）・カメラ込みでも 1 W 未満(5 V で 200 mA 未満)であり，バッテリ動作も可能な低消費電力を実現している．

Figure 1　Example of object detection by edge device.

このようなインテリジェントなデバイスを多数配置し，同一シーンを多方向から撮影・画像処理することで，シーンに関する抽象化後のデータを収集することができる．本稿では，このようにカメラ上での解析で得られた多面的な情報を統合し，シーンに関する三次元的な情報を再構成する手法について論じる．すなわち，各カメラの絶対位置と撮影範囲を所与とし，カメラから出力される物体検出結果（物体のカテゴリおよび撮影画像内における位置とサイズ）から物体の絶対位置を推定する．

1 東京大学
The University of Tokyo
2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

2. 関連研究

2.1 複数カメラの活用

近年，スポーツ映像のマルチカメラ解析への関心が高まっている．例えば，多視点映像の同期処理によるスポーツ選手のトラッキングや，自由視点映像の生成，プレーの分類や戦術分析などに関する研究が行われている[5]．スポーツ映像ではオクルージョンが多く動きが速いため，選手などの特定が困難な傾向がある．そのため，既存研究のほとんどの手法は，異なる方向から撮影された重なりを持つ複数の映像の情報を統合している[6-8]．このように複数の映像の情報を統合するためには，画像上の全ての画素を対応する三次元空間座標と結びつけるために正確なカメラキャリブレーションが必要である[8]．上述の既存研究はいずれも，カメラで撮影した映像データを集約しており，前節で述べたように通信量やプライバシーなどについて問題となる可能性がある．

また，正確な三次元情報を取得するために点群を用いる研究[9]も報告されている．点群に基づく手法は，高いコストを要することやリアルタイム性に欠けることなどが課題となる．

2.2 スマートカメラの活用

AI，ML，IoT などの進歩により，人手を介さずに映像の中からセキュリティ上の脅威を自動的に検出する，インテリジェントなモニタリングシステムが提案されている．[1]では，画像解析をクラウド上の集中管理されたサーバで実行される部分と，スマートカメラ上でローカルに実行される部分に分けている．AI を搭載したモニタリングアプリケーションをエッジに展開することで，撮影された映像の初期分析をカメラを搭載したデバイス上で行い，通信のオーバーヘッドの削減と，低遅延での脅威検出を実現している．ただし，カメラで撮影した画像は二次元情報しか持たないため，[1]のような単視点カメラによるモニタリングでは奥行き情報が定まらず三次元再構成が困難である．LiDAR (Light Detection and Ranging) センサなどと組み合わせることで奥行き情報が得られるが，一般に高いコストを要する．

3. 提案手法

3.1 概要

本研究では，前節で述べた複数カメラを用いるソリューションとスマートカメラによるソリューションを統合している．本研究の全体像を Figure 2 に示す．カメラ付きのエッジデバイスを複数用意し，同じシーンを多方向から撮影するよう配置する．カメラの位置・画角情報はあらかじめ取得しておく．各カメラは撮影した画像の一次解析処理を行い，得られた画像解析データをクラウド上のサーバに集約する．サーバ上では複数カメラの情報を統合し，撮影シーンの三次元的な状況を再構成する．カメラ上で画像処理することで，サーバとの通信量やプライバシーなどの課題を解消できる．

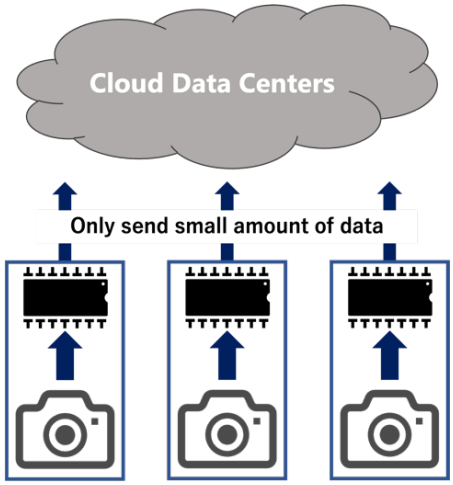

Figure 2　Research overview.

本研究では，画像の一次解析処理に You Only Look Once (YOLO) [10-12]を用いた．YOLO は機械学習による物体検出のアルゴリズムで, 画像内の物体のバウンディングボックスを検出すると同時に, その物体のカテゴリを判定する. YOLO では検出と判定を分離せずに処理するため, 高速な物体検出が可能となっている. 以下に述べる提案手法の検証では, あらかじめ撮影した高解像度の画像をパソコン上の YOLOv7 で処理し，検出した物体のカテゴリとバウンディングボックス（中心位置とサイズ）のリストを一次解析結果として用いた. 物体検出の例を Figure 3 および Table 1 に示す.

Figure 3　Object detection by YOLOv7.

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

Table 1　Object detection result.

class	*x*	*y*	*width*	*height*
2 (Car)	0.192448	0.608681	0.170312	0.163194
75 (Vase)	0.590885	0.678819	0.167187	0.302083

YOLO は, 画像左上隅を原点 $(0.0, 0.0)$ とし, 右下隅が $(1.0, 1.0)$ になるよう規格化した座標系について，検出した物体のカテゴリと中心点, およびバウンディングボックスの幅・高さを出力する．この例では，容量 2.8 MB の JPEG 画像を一次解析処理することで，77 B のテキストデータに縮約している．なお，検証に用いたセットアップでは，カテゴリのクラス数は 80 である．

3.2　エピポーラ線

Table 1 で見たように，各カメラからは二次元の画像解析結果が送信される．カメラの位置と方角および画角情報を既知として，解析結果から三次元空間内のエピポーラ線を求める．

エピポーラ線とはカメラと物体を結ぶ直線のことである．Figure 4 に幾何学的な配置図を示した．赤で示した線がここで求めるエピポーラ線である．いま，三次元空間内のカメラの位置を $\boldsymbol{o}$，撮影方向の単位ベクトルを $\boldsymbol{n}$ とし，カメラ上方の単位ベクトルを $\boldsymbol{p}$ とする．$\boldsymbol{p}$ は $\boldsymbol{n}$ 軸周りのカメラの傾き（ロール角）を表すもので，$\boldsymbol{n}$ と直交する．カメラの仮想的な投影面を距離 1 の位置に設定すると，そのサイズ (U, V) は画角情報から求めることができる. 投影面中央から右端までのベクトルを $\boldsymbol{u}$，中央から上端までのベクトルを $\boldsymbol{v}$ とすると，

$$\boldsymbol{u} = \frac{U}{2}(\boldsymbol{n} \times \boldsymbol{p})$$

$$\boldsymbol{v} = \frac{V}{2}\boldsymbol{p}$$

となる．典型的なケースとしてカメラのロール角をゼロとして配置すると，鉛直上方の単位ベクトルを $\boldsymbol{z}$ として

$$\boldsymbol{p} = \frac{\boldsymbol{z} - (\boldsymbol{z}\cdot\boldsymbol{n})\boldsymbol{n}}{|\boldsymbol{z} - (\boldsymbol{z}\cdot\boldsymbol{n})\boldsymbol{n}|}$$

で与えられる．ここで，・は内積を，× は外積を表す．ここに定義した各ベクトルは，カメラごとに定数として求められる．

YOLO により物体が画像内 (x, y) の位置に検出されたとき，カメラとその物体を結ぶエピポーラ線は，t をパラメータとして次のように定められる．

$$\boldsymbol{e} = \boldsymbol{o} + t\boldsymbol{d}$$

$$\boldsymbol{d} = \frac{\boldsymbol{n} + (x - 0.5)\boldsymbol{u} + (0.5 - y)\boldsymbol{v}}{|\boldsymbol{n} + (x - 0.5)\boldsymbol{u} + (0.5 - y)\boldsymbol{v}|}$$

$\boldsymbol{e}$ はエピポーラ線上の点である. カメラと物体の距離を t に代入すると，物体の位置を定めることができる．

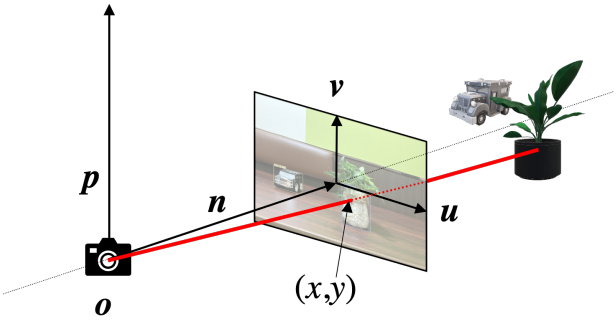

Figure 4　Epipolar geometry.

3.3　三次元再構成

カメラが 1 台しかない場合，検出した物体に対してエピポーラ線が定まるが，線上のどこに位置するか特定することができない．そこで配置の異なる 2 台以上のカメラを用いて同一物体を検出し，複数のエピポーラ線の交点として物体の位置を決定する．一般にカメラの位置や検出座標は誤差を含むため，エピポーラ線は厳密には交差しない．ここではエピポーラ線間の最近接点より物体座標を推定する．

2 台のカメラから同一物体に向けたエピポーラ線をそれぞれ

$$\boldsymbol{e_1} = \boldsymbol{o_1} + t_1\boldsymbol{d_1}$$

$$\boldsymbol{e_2} = \boldsymbol{o_2} + t_2\boldsymbol{d_2}$$

とする．方向ベクトル $\boldsymbol{d}_1$，$\boldsymbol{d}_2$ が非平行であれば $\boldsymbol{e}_1$，$\boldsymbol{e}_2$ の最近接点は $\boldsymbol{e_2} - \boldsymbol{e_1}$ と $\boldsymbol{d}_1$，$\boldsymbol{d}_2$ の直交条件より一意に定まり，

$$t_1 = \frac{(\boldsymbol{d_1} - a\boldsymbol{d_2})\cdot\boldsymbol{b}}{1 - a^2}$$

$$t_2 = \frac{(-\boldsymbol{d_2} - a\boldsymbol{d_1})\cdot\boldsymbol{b}}{1 - a^2}$$

で与えられる．ここで $a = \boldsymbol{d_1}\cdot\boldsymbol{d_2}$，$\boldsymbol{b} = \boldsymbol{o_2} - \boldsymbol{o_1}$ である．物体の推定位置は，この最近接点の中点にとる．

3 台以上のカメラが同一物体を検出する場合，エピポーラ線の全組み合わせについて最近接点が求められる．この場合，最近接点分布の中心を推定位置とし，分布の半径を推定誤差とする．

3.4　計算例

カメラ 2 台で同一シーンを撮影したケースについて，計算例を示す. 物体（Car と Vase）およびカメラの配置は Figure 5 に示す通りである．カメラ 1，2 の撮影画像と検出結果をそれぞれ Figure 3，Table 1 および Figure 6，Table 2 に示す．これらの情報をもとに計算したエピポーラ線と位置推定の結果を Figure 7 に図示した．

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

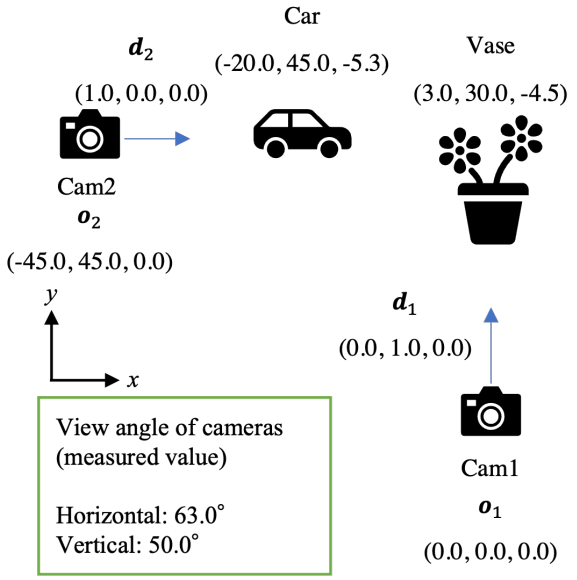

Figure 5　Position of the Car, the Vase and Cameras in cm unit.

Figure 6　Object detection by YOLOv7 (Cam2).

Table 2　Object detection result (Cam2).

class	*x*	*y*	*width*	*height*
2 (Car)	0.533594	0.722917	0.258854	0.233333
75 (Vase)	0.766667	0.637847	0.095833	0.161806

推定位置の実測値からのずれは，Car で 3.6 cm，Vase で 0.4 cm であった．一方，最近接点間の距離は Car で 1.4 cm，Vase で 0.6 cm であり，最近接点のばらつきが推定位置の誤差をよく反映していることがわかる．Car において誤差が大きいのは，検出した物体の中心点がカメラ間で一致していないことが原因と考えられる．すなわち，物体検出で得られるバウンディングボックスの中心は，必ずしも物体の重心に一致しない．改善策として，セグメンテーション[13]の情報を併用することで，物体の中心点を精密化することができる．

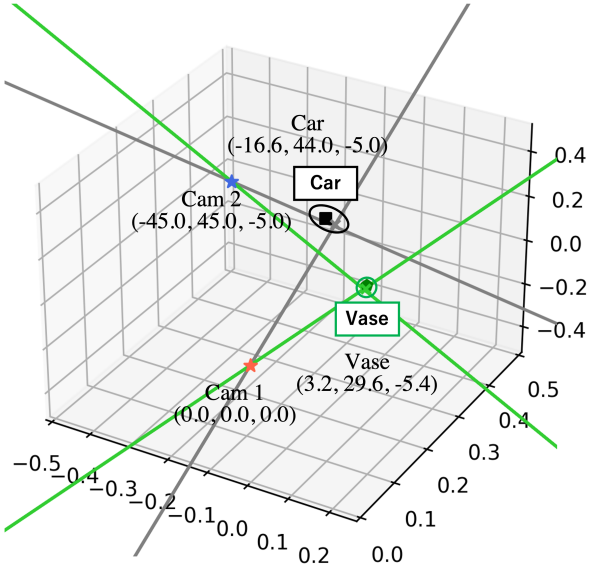

Figure 7　3D coordinate estimation.

4. オブジェクトが複数ある場合の処理

前節で述べた計算例では，物体検出結果に含まれるカテゴリを手がかりにエピポーラ線を分類することができた．しかし，群衆シーンのように同一カテゴリの物体が複数存在する場合，エピポーラ線の帰属に関して問題が生じる．群衆シーンを模擬するものとして，Figure 8 に示すシーンの解析を行なった．9 本の Bottle と 3 台のカメラの配置図を Figure 9 に示す．以下，Bottle A と B の位置決定を題材に，問題点を提示するとともに解決策を提案する．

Figure 8　Placement of Bottles.

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

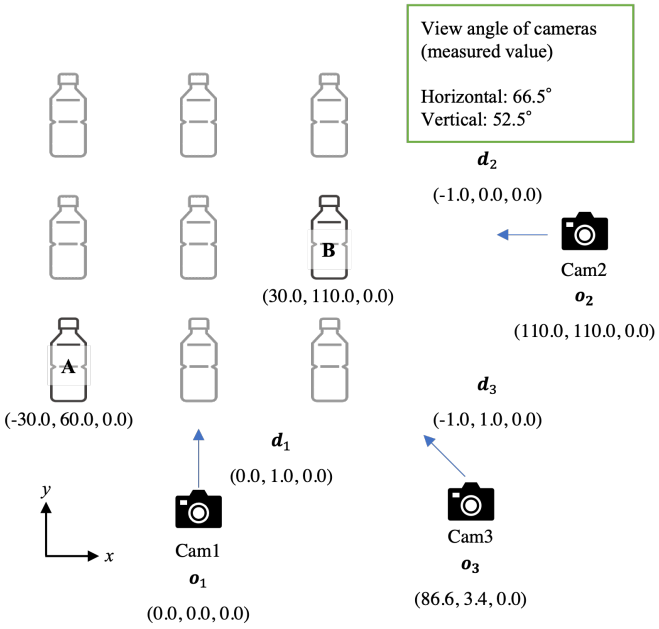

Figure 9　Position of the Bottles and Cameras in cm unit.

4.1　エピポーラ線の帰属問題

Figure 9 の空間配置において，Cam1 と Cam2 の 2 台のカメラのみを用いて物体座標推定を行う場合について，得られるエピポーラ線を Figure 10 に図示した．実際には A，B 以外のボトルも検出されるが，これらはあらかじめ人為的に省いた．各カメラから 2 本ずつ，合計 4 本のエピポーラ線が引かれる結果，合計 4 個の交点（最近接点の組）が算出される．ここで，Cam1 と Bottle A を結ぶエピポーラ線と，Cam2 と Bottle B を結ぶそれとの交点を，1A-2B のように記述する．前節の手順に従って位置推定すると，エピポーラ線の対応関係は自明ではないため，2 本の Bottle の三次元座標として，1A-2A および 1B-2B の組（Pattern 1）と，1A-2B および 1B-2A の組（Pattern 2）の 2 通りが考えられる．ここから正しい組み合わせ（Pattern 1）を選び取る問題が，エピポーラ線の帰属問題である．

第一に，推定誤差（最近接点間の距離）のより小さい組を選択する手法が考えられるが，カメラと物体の配置次第ではこれはよい指標とはならない．実際，Figure 9 の配置では Pattern 1 で 1.7 cm と 1.0 cm，Pattern 2 で 1.2 cm と 0.1 cm なり，前節の結果から予測される推定誤差範囲内に収まってしまう．以下，サイズ情報を併用する方法と，3 台以上のカメラを活用する手法について述べる．

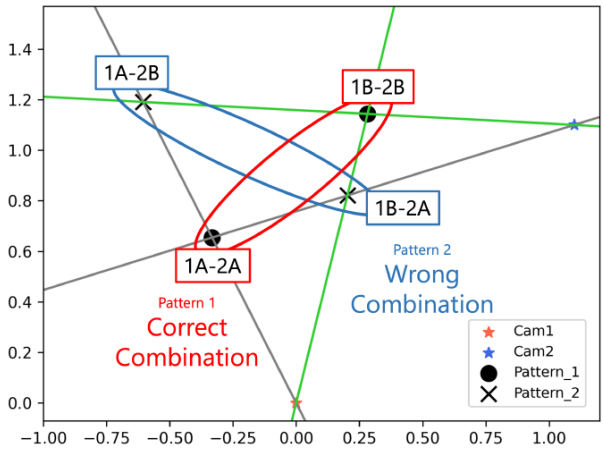

Figure 10　Object coordinate estimation (2 Cameras).

4.2　サイズ情報を用いた帰属

YOLO は検出した物体のバウンディングボックスのサイズ(w, h)を出力する．カメラと物体との距離 t が定まると，物体の実際のサイズを

$$W = U \cdot w \cdot t$$
$$H = V \cdot h \cdot t$$

のように計算できる．誤った組み合わせ（Pattern 2）では遠近法が狂っているので，各カメラから算出される物体のサイズの整合性から正しい組み合わせの判定が可能である．サイズ推定の結果を Table 3 に示す．サイズ指標として対角線の長さを用いた．

Table　3 Bottle size estimation.

	Pattern 1		Pattern 2	
	1A-2A	1B-2B	1A-2B	1B-2A
Cam1 [cm]	252.7	233.9	458.4	140.8
Cam2 [cm]	224.7	241.4	502	168.2
Absolute Error [cm]	28.0	7.5	43.6	27.4

Cam1 と Cam2 の絶対誤差が最大となるのは 1A-2B の場合である．よって，Pattern 2 が誤った組み合わせであり，Pattern 1 が正しい組み合わせであると判定することができる．ただし，正しい組み合わせである 1A-2A の場合においても，絶対誤差が比較的大きな値となっている．YOLO による物体検出では，撮影する方向に依存してサイズが変化することもあり，サイズ情報を用いた帰属の有用性はカテゴリに依存して異なる．また，物体が重なって写るケースなどバウンディングボックスを誤って検出する傾向があるため，複雑なシーンの解析では誤判定する可能性がある．そこで，4.3 ではカメラを 1 台追加することにより，サイズ情報を用いることなく座標推定を行なった．

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

4.3 3 台目のカメラを用いた帰属

3 台以上のカメラを用いることで，推定誤差に基づく判定を精緻化することができる．Figure 11 に，Cam1，Cam2 に加えて Cam3 から Bottle A, B へのエピポーラ線を図示する．図中，Bottle A，B の実際の位置を黒点で示している．

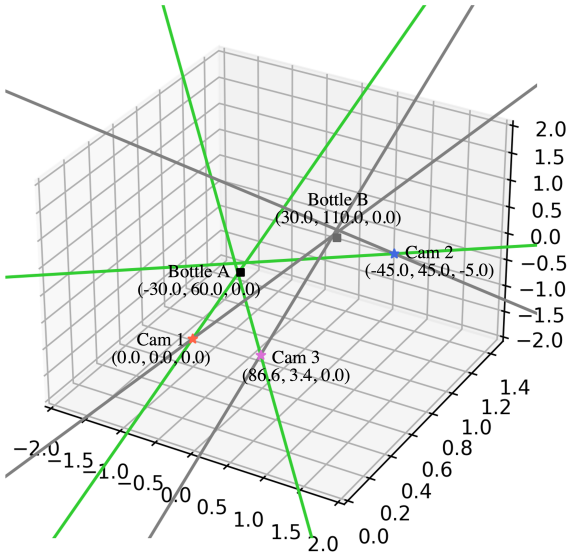

Figure 11　Object coordinate estimation (3 Cameras).

カメラが増えた結果，可能なエピポーラ線の組み合わせは 4 通りに増える．正しい組み合わせ（1A-2A-3A および 1B-2B-3B）および誤った組み合わせ（1A-2B-3A および 1B-2A-3B）の場合について，得られる最近接点を赤と青のドットでそれぞれ Figure 12 と Figure 13 に示した．正しい組み合わせでは合計 6 点の最近接点が密集する一方，誤った組み合わせではばらついていることがわかる．正誤全ての組み合わせについて，推定誤差を Table 4 にまとめた．最近接点の平均値を中心とした分布半径を推定誤差に取っている．

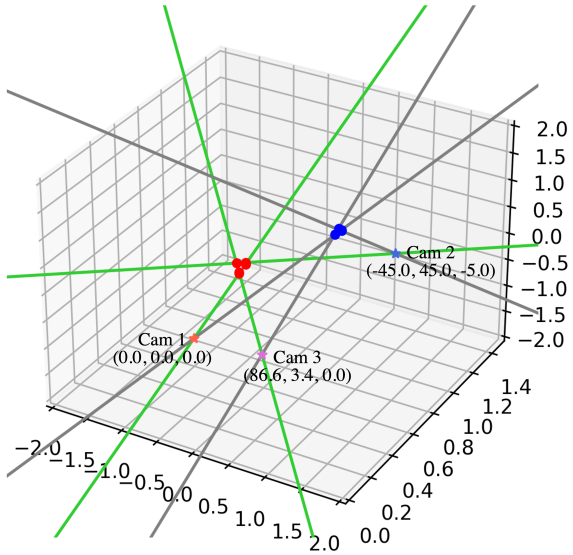

Figure 12　Correct combination (3 Cameras).

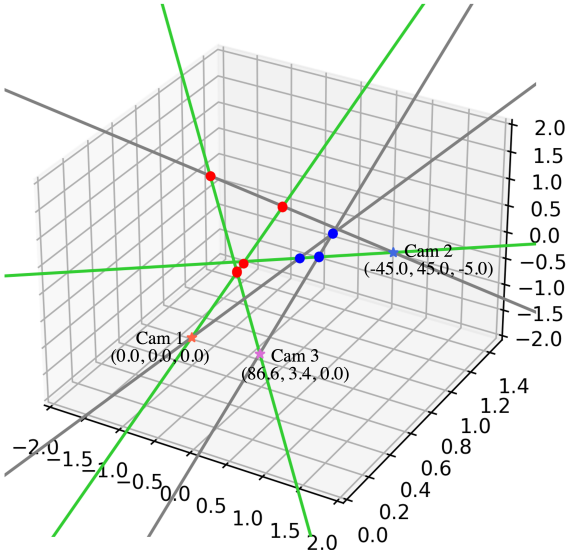

Figure 13　Wrong combination (3 Cameras).

Table 4　Distribution radius of closest points.

Red		Blue	
Combination	Radius [cm]	Combination	Radius [cm]
1A-2A-3A	7.9	1B-2B-3B	3.8
1A-2B-3A	90.9	1B-2A-3B	16.1
1A-2B-3B	1041.5	1B-2A-3A	37.9
1A-2A-3B	1071.7	1B-2B-3A	133.7

正しい組み合わせの場合と比べて，誤った組み合わせの場合では，分布半径が 1 桁から 2 桁程度大きいことがわかる．したがって，分布半径の小さい組を選択することで正しい組み合わせを判定することができる．

5. 考察

カメラがn台で同一クラスカテゴリの物体がm個検出された場合，エピポーラ線の組み合わせの総数は$m!^{n-1}$通りとなる．よって，カメラの台数や物体の個数が増えるごとに，組み合わせの総数と計算量が爆発的に増加する．これは組み合わせ最適化問題に帰着され，最適解を効率的に求めるための工夫が必要となる．ここではカテゴリ分類の精緻化の可能性と，オクルージョンによる問題点について議論する．

5.1 カテゴリ分類の精緻化

(1) 色情報による分類

群衆シーンでは，色情報を特徴量として分類し組み合わせの数を削減することができると考えられる．バウンディングボックス中の色情報を手がかりとして物体を区別する．ただし，照明や外光などの周辺環境の変化によって物体色

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

も変化するため，シーン全体の色分布に基づく補正処理[14]などが必要となる．また，カメラごとの写り方の違いによって物体の色分布が異なり，誤判定する可能性がある．そのため，色情報のみではなく他の情報も組み合わせて分類を行うことが有効であると考えられる．

(2) インスタンスセグメンテーションによる分類

インスタンスセグメンテーションは，ピクセルレベルで物体のクラスカテゴリを分類する手法である．実行例を Figure 14 に示す．バウンディングボックスレベルで検出を行う物体検出と比べて，高精度な検出が可能である．オクルージョンに強いことや，正確な領域を抽出可能であることなどが特徴である．ROI (Region of Interest) を対象に，全ての物体に対してクラスカテゴリを予測し，各物体に対して一意の ID を付与する．そのため，1 枚の画像に複数の人物が写っている場合，それぞれの人物を別の物体として認識することが可能である．ただし，物体検出と比べて多くの計算量を要し，現在製品化されている安価なエッジデバイス上での実行は困難である．エッジデバイスの性能が日々進歩していることや，インスタンスセグメンテーションを軽量化する手法[15]が提案されていることなどから，将来的にはエッジデバイス上で処理可能となることが期待される．

Figure 14　Instance Segmentation.

(3) 人物姿勢推定による分類

人物姿勢推定では，画像中の人物の関節点座標の推定を行う．実行例を Figure 15 に示す．人物姿勢推定によって得られる人物の向きや姿勢などの情報から分類を行うことで組み合わせの数を削減することが可能である．また，(1)で述べた色情報による分類と併用することにより，判定の精度を高めることができると考えられる．

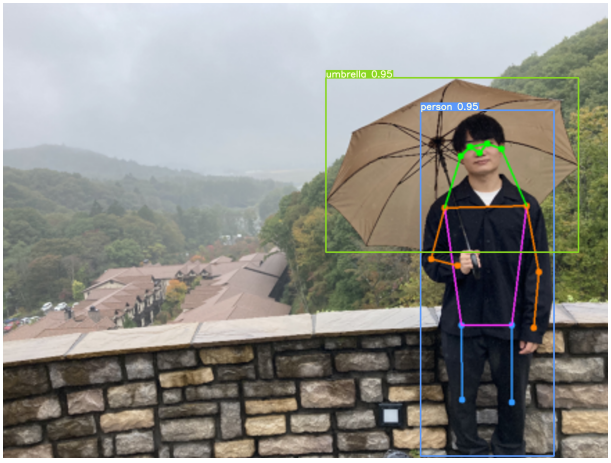

Figure 15　Pose Estimation.

5.2 遮蔽などによる問題点

カメラと物体の配置次第では，全てのカメラに全物体が検出されるとは限らない．例えば Figure 9 の配置では，Cam1-3 はそれぞれ 7 本，6 本，7 本のボトルを検出しており，数は一致しない．このような状況では正しい組み合わせの判定はより複雑となる．解決策の検討はまだ十分にできていないが，前節で用いた分布半径に基づく判定の改良などを検討している．前節では単純に最近接点の最大半径を指標としたが，例えば最近接点の分布を手がかりにエピポーラ線の取捨選択が可能ではないかと考えている．組み合わせの数が膨大である場合，サーバ側の潤沢な計算リソースを活用した並列処理も選択肢として挙げられる．

6. まとめ

本研究では，同一領域を複数のカメラで同時撮影し，機械学習による物体検出結果を統合することで空間の三次元的な構成要素を特定する手法について検討した．同一カテゴリの物体が複数存在する場合でも，3 台以上のカメラを用いることにより，領域に存在するオブジェクトの種類と位置について一意的な再構成が可能であることを示した．

カメラの台数や物体の個数が増えると，再構成にかかる計算量は爆発的に増加する．計算量の削減にはエッジデバイス側で行う一次解析処理の精緻化が必要である．現状でも色情報の付与など軽量な精緻化は可能と考えているが，将来的にはセグメンテーションや姿勢推定など，より高度な機械学習がエッジデバイス上で処理可能になると期待している．

エッジ AI に関する技術は急速な進歩を続けており，新たな手法や新製品などの発表が行われている．Sense CAP[16]は 1 節で紹介したエッジデバイスと比べて性能が高く，高度な処理が実行可能となっている．本研究の提案手法と同様に，ローカルで画像を推論し抽象化された最終結果のみをクラウドに転送するため，通信量の削減や高い

Vol.2022-CG-188 No.21
Vol.2022-DCC-32 No.21
Vol.2022-CVIM-231 No.21
2022/11/19

データプライバシーを要するアプリケーションに適している. また, 抽象化されたデータの容量は非常に小さいため, LPWA(Low Power Wide Area)のように容量に強い制約がある通信方式を採用することができる．SenseCAP では，LoRaWAN モジュールを搭載することにより，超低消費電力で長距離伝送を実現している．このように，ソフトとハードの両面で技術が進歩していくことにより，本研究の応用展開が広がると考えられる．

参考文献

[1] Ahmed Abdelmoamen Ahmed and Mathias Echi.. Hawk-Eye: An AI-Powered Threat Detector for Intelligent Surveillance Cameras. IEEE Access, 2021, vol.9, p.63283-63293.
[2] Ahmed Abdel Moamen and Nadeem Jamali.. Opportunistic Sharing of Continuous Mobile Sensing Data for Energy and Power Conservation. IEEE Transaction on Services Computing, 2020, vol.13, no.3, p.503-514.
[3] “DFROBOT SKU:SEN0305”, https://wiki.dfrobot.com/HUSKYLENS_V1.0_SKU_SEN0305_SEN0336, (accessed 2022-09-21).
[4] Sipeed: “Sipeed Maix-Bit Specifications”, https://dl.sipeed.com/shareURL/MAIX/HDK/Sipeed-Maix-Bit/Specifications, (accessed 2022-09-21).
[5] 田中成典, 山本雄平, 姜文渊, 中村健二, 清尾直輝, 田中ちひろ. 複数視点からの映像を用いたスポーツ選手のトラッキングに関する研究. 日本知能情報ファジィ学会誌, 2020, vol.32, no.4, p.812-830.
[6] 姜文渊, 山本雄平, 田中成典, 中村健二, 田中ちひろ. 単視点多眼によるアメリカンフットボールプレイヤーの識別と位置特定に関する研究. 写真測量とリモートセンシング, 2018, vol.57, no.5, p.198-216.
[7] Xina CHENG, Norikazu IKOMA, Masaaki HONDA, TakeshiIKENAGA.. Multi-View 3D Ball Tracking with Abrupt Motion Adaptive System Model, Anti-Occlusion and Spatial Density Based Recovery in sports Analysis. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2017, vol.E100A, no.5, p.1215-1225.
[8] Yukun Yang, Ruiheng Zhang, Wanneng Wu, Yu Peng, Min Xu.. Multi-camera Sports Players 3D Localization with Identification Reasoning. IEEE 2020 25th International Conference on Pattern Recognition (ICPR), 2021, p.4497-4504.
[9] Yin Zhou, Oncel Tuzel.. VoxelNet: End-toEnd Learning for Point Cloud Based 3D Object Detection. arXiv: Computer Vision and Pattern Recogniton /1711.06396v1, 2017, p.1-10.
[10] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi.. You Only Look Once; Unified, Real-Time Object Detection. arXiv: Computer Vision and Pattern Recognition /1506.02640v5, 2016, p.1-10.
[11] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao.. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: Computer Vision and Pattern Recognition /2004.10934v1 2020, p.1-17.
[12] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao.. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real time object detectors. arXiv: Computer Vision and Pattern Recognition /2207.02696, 2022, p.1-17.
[13] Hao Chen, Kunyang Sun, Zhi Tian, Chunhua Shen, Yongming Haung, Youliang Yan.. BlendMask: Top-Down Meets Bottom-up for Instance Segmentation, arXiv: Computer Vision and Pattern Recognition /2001.00309, 2020, p.8573-p.8581.
[14] 辻本晃大, 土居元紀. シーンの色分布を手掛かりとして照明変動を補正した色情報による人物追跡. 日本色彩学会誌, 2017, vol.41, no.6, p.47-49.
[15] 小石原遼, 天野敏之, 渡辺義浩. 高速インスタンスセグメンテーションを用いた投影による選択的色操作の提案, 日本バーチャルリアリティ学会 複合現実感研究会, 2022, vol.25, no.1, p.1-6.
[16] “SenseCAP A1101 – LoRaWAN Vision AI Sensor, Open the Door to the TinyML world”, https://www.seeedstudio.com/SenseCAP-A1101-LoRaWAN-Vision-AI-Sensor-p-5367.html, (accessed 2022-10-21).