

# ステレオカメラによる YOLO と 3 次元点群を用いた人物検出

高橋正裕<sup>1</sup> 池勇勲<sup>1</sup> 梅田和昇<sup>1</sup>

<sup>1</sup> 中央大学理工学部精密機械工学科 〒112-8551 東京都文京区春日 1-13-27

## 概要

近年、防犯カメラやマーケティング等の分野において、画像処理を応用し、人物検出や人数カウントを行う研究が盛んに行われている。本研究では、ステレオカメラから得た画像と 3 次元点群から、オクルージョンに強い人物検出を実現する。物体検出に関する先行研究として、深層学習を用いた YOLO (you only look once) という手法が提案されているが、オクルージョンに対応できないという弱点が存在する。そこで、本研究では、ステレオカメラから得られる 3 次元情報と YOLO による人物推定領域を用いて人物領域の 3 次元点群を取得し、クラスタリングにより人物検出を行う手法を提案する。

キーワード：人物検出 ステレオカメラ 深層学習 3 次元点群 YOLO

## 1. 序論

近年、防犯カメラやマーケティング等への利用を想定し、人物検出や人数カウント等の需要が高まっている。しかし、これらを人手で行った場合、ヒューマンエラーや人件費等の問題が生じてしまう。そのため、これらの分野における自動化が求められている。

人物検出手法の代表例として、画像の輝度勾配を特徴として学習した HOG (histogram of oriented gradients) 特徴量[1]を用いた手法ある。また、近年では深層学習を用いた研究が非常に盛んであり、中でも物体認識が可能な YOLO (you only look once) [2] は高いリアルタイム性と検出率を誇っている。しかし、これらの手法はカラー画像に対してのみ適用可能であるため、距離情報を考慮することができない。そのため、人数カウントにおいて重要であるオクルージョンへの対応が難しい。

そこで本研究では、YOLO の持つオクルージョンへの弱さの解決を目的とする。具体的には、ステレオカメラを用いてカラー画像、3 次元点群を取得し、YOLO による人物推定領域内の 3 次元点群を用いて、オクルージョンへの対応を目指す。

## 2. 提案手法

提案手法のフローチャートを Fig. 1 に示す。始めに、ステレオカメラよりカラー画像と 3 次元点群を取得する。次に、取得したカラー画像に対し、YOLO による人物領域の推定を行う。その後、取得された人物領域に含まれる点群を抽出し、外れ値除去等の処理を行う。最後に、その点群に対してクラスタリングを行うことで、人物検出を行う。以下、各段階について詳細に説明する。

### 2.1 3 次元点群の取得

ステレオカメラの左右のカメラにより、カラー画像を取得する。その後、左右のカラー画像から算出される視差を用いて距離を計測する。これにより、3 次元点群の取得を行う。

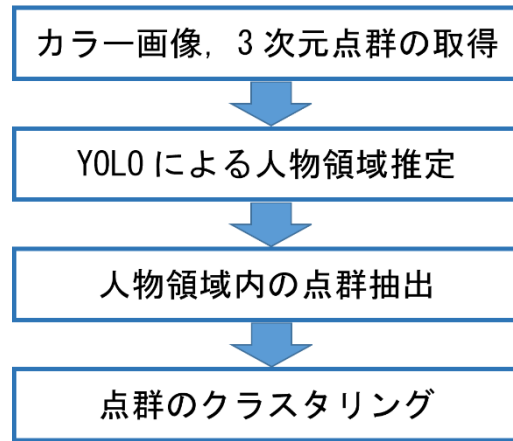


Fig. 1 提案手法のフローチャート

### 2.2 YOLO による人物領域推定

YOLO とは、ニューラルネットワークに基づく速度に特化した画像検出・認識手法である。本研究では、処理速度がより速い YOLO ver. 2 を用いる。ステレオカメラにより得られたカラー画像を入力とし、YOLO により物体認識を行った後、人物領域の推定結果を利用する。

### 2.3 人物領域内の点群抽出

YOLO による人物推定領域は、オクルージョン発生時に人物領域が結合してしまう。そこで、2.1 節で生成した 3 次元点群情報を利用し、YOLO による推定領域内の点群を抽出する。3 次元空間における点が画像中の点の対応は次式で表すことができる。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} f_x & 0 & c_u \\ 0 & f_y & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

ここで、 $(u, v)$  は画像中での座標、 $(X, Y, Z)$  はカメラ座標系における 3 次元空間内での座標を表しており、カメラの内部パラメータから成る行列によって同次座標として表現することができる。 $(f_x, f_y)$  は画像ピクセル単位の焦点距離、 $(c_u, c_v)$  は画像中心を表している。また、この段階で点群分布の統計量を用いた点群の外れ

値除去と、カメラの高さ情報を用いたしきい値による地面の除去も行う。

## 2.4 点群のクラスタリング

YOLOの推定結果を利用して人物周辺の点群のみを抽出することで、クラスタリングにより人物の点群を取得する。クラスタリングには、Kd木によるユークリッド距離における最近傍探索を用いる。

## 3 人物検出実験

本手法の有用性を検証するために、歩行者3人を対象に人物検出実験を行った。歩行者3人にステレオカメラの前をランダムに移動してもらい、オクルージョン発生時の1,018フレームについて正確な人数カウントが可能かを検証した。

オクルージョン発生時のフレームにおける、提案手法とYOLOによる検出結果の比較をTable 1に示す。提案手法によりYOLOによる検出結果が改善されていることがTable 1より分かる。具体的な例をFig. 2及びFig. 3に示す。Fig. 2はYOLOのみによる検出結果、Fig. 3は提案手法による検出結果である。YOLOでは画面左の人物と中央の人物領域が結合してしまっているが、提案手法では分割できている。それに対し、Fig. 4に失敗例を示す。この場合は複数の人物の領域が結合してしまい、1人とみなしてしまっている。これは、ステレオカメラにおけるステレオマッチングの際に距離の計測が正確に行えていない部分を補間することで、距離値の境界が曖昧になってしまうことが原因であり、より精度の良いステレオカメラを用いることで解決可能である。

また、Table 2に速度の比較結果を示す。3次元点群処理による計算量の増加により、提案手法はYOLOに比べ、処理速度が55%程度低下してしまっただが、平均15 fps程度の速さであるため、リアルタイムでの運用においては問題ないと判断可能である。

## 4 結論と今後の展望

ステレオカメラから取得した3次元点群及びYOLOを用いることで、オクルージョンに強い人物検出手法を提案し、実験により有用性を検証した。

今後の展望としては、3次元点群に含まれるノイズの除去、またYOLOの検出結果と3次元点群を入力とし、人物領域のみの点群を抽出するニューラルネットワークを用いた手法を考案する予定である。

Table 1 YOLOと提案手法の比較結果

	検出フレーム数	
	正	誤
提案手法	184	834
YOLO	117	901

Table 2 人物検出における比較結果

	画像取得時間[s]	取得フレーム数	平均速度 [fps]
提案手法	240	3,364	14.02
YOLO		7,209	30.04



Fig. 2 YOLOによる検出結果



Fig. 3 提案手法による検出結果



Fig. 4 結合した人物領域

### 文献

- [1] 奥富 正敏他, “デジタル画像処理”, CG-ARTS 協会, 2015.
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.