H-028

カメラパラメータによる射影変換を用いた深層学習型物体検出の改善手法
Improvement of Deep Learning-based Object Detection Method by Projective Transformation using Camera Parameter

段　清柱† 福田　竣† 村上　智一†
Duan Qingzhu Fukuda Shun Murakami Tomokazu

1. はじめに

主に防犯を目的として、道路や駅などの公共スペースに多くのビデオカメラが設置されている。また近年、深層学習などの技術進化により、従来の事後調査目的の録画に加え、リアルタイム認識による人物検出や不正挙動検知のニーズが高まっている。人物検出技術の代表例として SSD[1]などの物体検知手法は速い処理速度と高精度で実応用に適しているが、当該技術では検知対象の画面上サイズが小さい場合、検知が困難であることが知られている。監視領域が奥行き方向に細長い場合、奥側の検知対象が非常に小さくなり検知率が低下する。この課題に対し、入力画像の前処理を追加することにより奥側の対象の検出精度を改善する手法を検討したので報告する。

2. 提案手法

本章では課題の分析と、解決方法及び全体システム構成について説明する。

2.1 課題分析

画面上サイズが小さい物体の検知性能が低い理由は、学習用正解画像の検知対象サイズのばらつきや、連続複数回の畳み込み層やプーリング層の処理後に入力画像の情報が劣化していることなどが考えられるが、本稿では後者に関して検討を行う。例えば SSD300 の場合、300x300 の画像を約 1/8(38x38)までに圧縮してから、物体検知を行うことになる。画面上サイズが小さい場合、物体関連の情報が非常に少なくなると考えられる。

入力画像を大きくすると、検知対象の画面上サイズが大きくなって検知率を改善できる。文献[1]では PASCAL VOC 07+ PASCAL VOC 12+COCO Dataset を学習画像とし、PASCAL VOC2012 を評価した結果、SSD512 において SSD300 と比較して mAP を 2.5 ポイント改善することができた。しかし、計算時間は 2.42 倍長くなり、リアルタイム性及び計算コストについて課題がある。

防犯用映像監視の場合、監視領域が固定となり、監視対象が地面にあるなどの特徴がある。例えば、人物検知の場合、基本的に人物が地面に立っているはずである。また道路や駅構内通路など監視領域が変化せず、固定的であることが多い。本稿ではその特性を生かして、検知対象の画面上サイズが小さい領域を拡大し、検知対象の画面上サイズが大きい領域を縮小することで物体検出の対象画像に初期処理を行い、検知精度の改善を提案する。

人物画像が場所によって画面上サイズが異なる原因はカメラの視点や設置姿勢によるものである。文献[2]では、建物の認識において、建物を撮影するカメラ視点の差異を吸収するため、建物の真正面の仮想画像を生成することを提案している。仮想画像の生成はカメラの姿勢情報を用いて、射影変換の手法を採用した。射影変換操作によって検知対象の画面上サイズを調整できる効果がある。例えば、平面上にある矩形の道路が画面上では台形となるが、平面の真上の画像を生成する場合、矩形に戻し、台形上底の周辺を拡大することができる。しかし、人物などの 3D 空間の物体では平面の画像と異なるため、ビデオカメラの画像を真上から見る場合の仮想画像に変換する場合、人物の形に歪みが生じ、検知率が逆に下がる恐れ可能性がある。このため適切な射影変換のパラメータを検討する必要がある。

2.2 提案するシステム構成

図 1 に提案するシステム構成を示す。

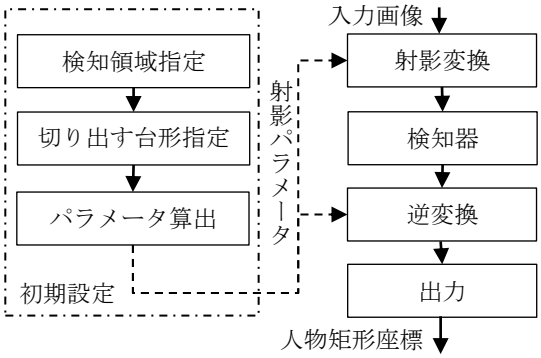

図 1. 提案システム構成

図 1 の左半分が初期設定のフローを示す。入力画像から人物検知領域を指定し、その領域をカバーする台形を指定する。台形領域を射影変換により矩形に変換する。台形から矩形への変換行列を射影変換パラメータとする。逆に、矩形から台形の座標への変換行列を逆射影変換パラメータとする。台形の上底の周辺が下底の周辺と比べて拡大されることになる。以下、台形の上底 l_1、下底 l_2 の比を人物の最大拡大率 θ とする。

$$\theta = l_2 / l_1 \quad (1)$$

図 1 の右半分が人物検知のフローを示す。入力画像に射影変換を行い、検知器の入力とする。検知器は SSD など深層学習を用いる人物検出器となる。本稿では文献[1]の学習済み SSD300 のモデルを採用する。検知器の検出結果が射影変換後の画像座標になるため、逆射影変換を用いて元入力画像の座標系の位置を算出し、検知結果として出力する。

3. 評価

3.1 評価方法

本手法の有効性を確認するための評価実験を実施した。評価指標を検知対象領域の mAP（mean Average Precision）

† 日立製作所 研究開発グループ メディア知能処理研究部

とする。本稿が想定している実環境との類似度を考慮し、評価データは文献[3]のカメラ 4 の前半のデータとする。映像は 15 フレームの間隔で画像を抽出し、評価を行った。図 2 が評価映像のスナップショットとなる。

検知領域は図 2 の点線が示す領域とし、切り出す領域は図 2 が示す台形とする。切り出す領域の画像が射影変換前処理によって SSD300 の入力サイズ 300x300 画像になる。元データの正解のタグつけが手前に限定されているため、奥側の追加タグつけを行い、評価データとして利用する。

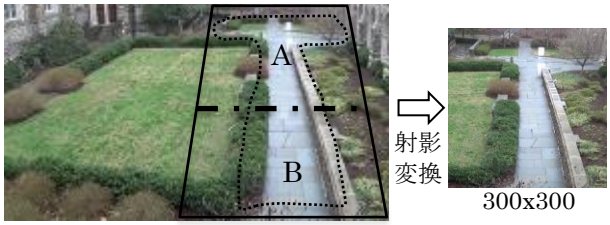

図 2. 評価対象画像

最大拡大率 θ の検知結果への影響を分析するため、最大拡大率 θ が異なる 5 パターンを評価する。また、射影変換処理による場所ごとの影響を分析するため、検知領域を図 2 に示すように、縦方向で A,B という二つの領域を分け、評価を行う。

3.2 評価結果

表 1 に 5 パターンの拡大率 θ 及び mAP を示す。

Base は θ=1 射影変換を行わず、画像の縮小に相当する。#1～#4 では順番に θ を大きくする。#4 が真正面の仮想画像を生成する場合の拡大率に相当する。

表 1．5 パターンの評価結果

Method	Base	#1	#2	#3	#4
拡大率 θ	1	1.2	1.5	2.0	3.1
mAP	76.2	84.1	90.9	85.0	67.2

表 1 が示すように、射影変換無し(Base)の mAP が 76.2 に対して、#3 の mAP が最大 90.9 を達成でき、射影変換による精度改善が確認できた。また#3､#4 の結果から、拡大率 θ が一定値を超えると検知率が悪くなることが分かる。

図 3 が検知領域全体の Precision と Recall を示す。

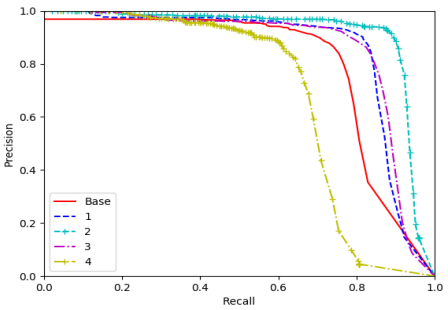

図 3. 5 パターンの Precision と Recall

図 3 が示すように、#3 の Recall が改善されていることが分る。#4 の Recall が低下していることが確認できる。検知漏れが発生していることを示す。

図 4 が領域別の mAP の変化を示す。

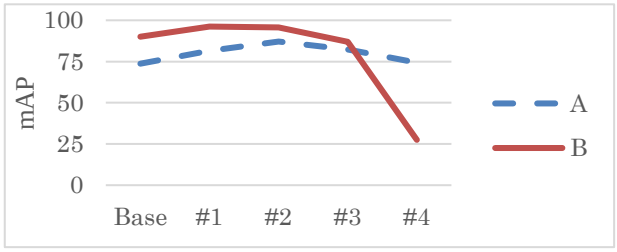

図 4. 領域別の mAP

A 領域では Base から#2 の間に mAP が改善している。#3、#4 では若干落ちているが、射影変換による検知精度が確認できると言える。

B 領域では、#1 まで若干改善できているが、#2~#4 には mAP が落ちている。特に#4 では mAP が大きく低下しているのが確認できる。その原因は手前の画像が変形され人物の外形の歪みが発生していることが考えられる。

以上の結果から、射影変換により奥側の画面上サイズが小さい人物の検知精度を改善できることが確認できた。

4. 結論

広域映像監視応用を想定し、道路や駅のホームなど監視領域が限定されかつ監視領域が細長い状況において、物体検知精度を改善する手法を提案した。入力画像に射影変換前処理を行うことによって奥側の検知対象物の画面上サイズを拡大し、検知率の改善を行った。評価の結果 mAP を 19.7 ポイント改善でき、本手法の有効性を確認した。

今後の課題としては、射影変換により人物外形の歪みが発生するため、手前の検知精度が落ちる可能性がある点が挙げられる。これに対し、学習段階において射影変換後の画像を増やして学習画像とし、歪み画像の検知率を向上させることが考えられる。

参考文献

[1] Liu W. et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham.

[2] 山口尭爾，福元和真，松下侑輝，川崎洋，小野晋太郎，池内克史:「深層学習による車載映像の都市名推定」, 生産研究， 68 巻 2 号(2016)

[3] Ristani, Ergys and Solera, Francesco and Zou, Roger and Cucchiara, Rita and Tomasi, Carlo, Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking, European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking(2016)