

# パンチルトカメラを用いたオンライン放置物検出

○ 小川 拓真\*, 平岡 大樹†, 伊藤 伸一‡, 伊藤 桃代§, 福見 稔¶

○ Takuma Ogawa\*, Daiki Hiraoka†, Shin-ichi Ito‡,

Momoyo Ito§ and Minoru Fukumi¶

\*: 徳島大学先端技術教育科学部, c501537046@tokushima-u.ac.jp

†: 徳島大学先端技術教育科学部, c501537004@tokushima-u.ac.jp

‡: 徳島大学ソシオテクノサイエンス研究部, s.ito@tokushima-u.ac.jp

§: 徳島大学ソシオテクノサイエンス研究部, momoito@is.tokushima-u.ac.jp

¶: 徳島大学ソシオテクノサイエンス研究部, fukumi@is.tokushima-u.ac.jp

＜要約＞ 本稿ではパンチルトカメラを用いたオンライン放置物検出手法を提案する．提案手法として，まずパンチルトカメラで動画像の取得を行う．その動画像に対して放置物があると考えられる対象領域の抽出を ST-Patch 特徴量を用いて行う．そして抽出された対象領域に対して放置物の存在を確かめる物体識別を行い，放置物の検出を行う．提案手法の有効性を検証するために複数の場면을想定した実験を行う．

＜キーワード＞ パンチルトカメラ, ST-Patch, Deep Learning, 人検出, 放置物検出

## 1. 背景・目的

近年，広大な敷地を有し大勢の人が集まる場所(空港, 駅, 病院など)に，高い割合で監視カメラが設置されている．監視カメラの設置台数は年々増加傾向にあり[1], その主な理由としては防犯の強化である．しかしながら，監視カメラから得られる情報は手動で確認を行わなければならない．特に最初に述べたような場所では，広範囲を監視する多数のカメラが必要である．そのため監視対象が多岐に渡り，すべてを手動で管理することは難しい．そこで監視カメラから得られる情報を自動で管理するシステムが必要であると考えられる．

以前，放置物検出システムを提案した[2]．放置物は人が近くに存在しない物体と定義する．しかしながら，このシステムには問題点が存在した．本稿では提案手法の変更を行い従来手法における問題点の解決を図る．

## 2. 従来手法

### 2.1 動画像入力部

パンチルトカメラにより動画像の入力を行う．パンチルトカメラを使用することにより，定点カメラよりも広範囲を監視することが可能となる．それに加えズーム機能を使うことによって，細部まで鮮明に観測することが可能となる．

### 2.2 対象領域抽出部

パンチルトカメラにより入力された動画像に対して対象領域の抽出を行う．ST-Patch 特徴量の表れた部分を対象領域としている．ST-Patch 特徴量は物体のアピアランスやモーションに着目した特徴量であり，対象物の大まかな形状と動きを表す．

### 2.3 物体識別部

人検出を行い対象領域における人の存在を判断する．人検出には，HOG 特徴量を用いた Real AdaBoost を使用する．人が存在しない場合，対象領域にある物体は放置物であると判断する．人が存在する場合，対象領域には放置物がないと判断する．

## 2.4 放置物検出部

物体識別部において放置物が存在した場合のみ、物体の画像を保存する。

## 2.5 問題点

対象領域抽出部において ST-Patch 特徴量の検出精度が低かった。また、物体識別部において人検出精度も低かった。

## 3. 問題点と対処

先に述べた従来手法にはいくつかの問題点が存在した。この節ではその問題点とそれに対する対処を述べる。

### 3.1 対象領域抽出部

まず対象領域抽出部においての問題点である。それは、ST-Patch 特徴量の検出精度が低いことである。これは輝度の変化に弱いことが原因であると考えた。この対処として私たちは拡張 ST-Patch[3]を提案し、対象領域抽出部に導入することで特徴量の検出精度を向上させる。

### 3.2 拡張 ST-Patch 特徴量

拡張 ST-Patch 特徴量とは ST-Patch 特徴量に色情報を加え、輝度の変化に強くした特徴である。ST-Patch 特徴量には 1:x 軸方向のソーベルフィルタ 2:y 軸方向のソーベルフィルタ 3:輝度差の 3 つのデータが使用されている。そこで、輝度の変化に強くするために拡張 ST-Patch 特徴量には、輝度差の代わりにフレーム間の RGB を用いた色情報の差を使用する。以下が拡張 ST-Patch 特徴量の計算である(式(1)).  $P_{ti}$  が拡張 ST-Patch 特徴量の成分である。

$$\begin{aligned} P_{t1} &= (|R_1(1,1) - R_0(1,1)| \\ &+ |G_1(1,1) - G_0(1,1)| \\ &+ |B_1(1,1) - B_0(1,1)|)/3 \\ P_{t2} &= (|R_2(1,2) - R_0(1,2)| \\ &+ |G_2(1,2) - G_0(1,2)| \\ &+ |B_2(1,2) - B_0(1,2)|)/3 \\ &\vdots \\ P_{ti} &= (|R_i(x,y) - R_0(x,y)| \\ &+ |G_i(x,y) - G_0(x,y)| \\ &+ |B_i(x,y) - B_0(x,y)|)/3 \end{aligned} \quad (1)$$

R,G,B はそれぞれ RED, GREEN, BLUE の頭文字であり、各画素の各色成分を表している。行列計算は ST-Patch 特徴量と同様である。

## 3.3 物体識別部

物体識別部で用いている人検出の精度が低いことが問題であった。これは AdaBoost の学習の段階に問題があると考え、学習データの見直しなど行ったが向上しなかった。そこで、人検出に使用する手法の変更を行った。本稿では人検出に畳み込みニューラルネットワーク[4]と呼ばれる Deep Learning を使用して検出を行っている。畳み込みニューラルネットワークは画像処理に特化した構造を有している人検出に有効であると考えた。

## 3.4 畳み込みニューラルネットワーク

畳み込みニューラルネットワークとは生体の脳の、特に視覚系をモデル化したニューラルネットワークである。単純細胞に相当する畳み込み層と複雑細胞に相当するプーリング層とを構成要素とする。認識対象の移動や変形に対して頑健である。

## 4. 提案手法

### 4.1 提案手法の流れ

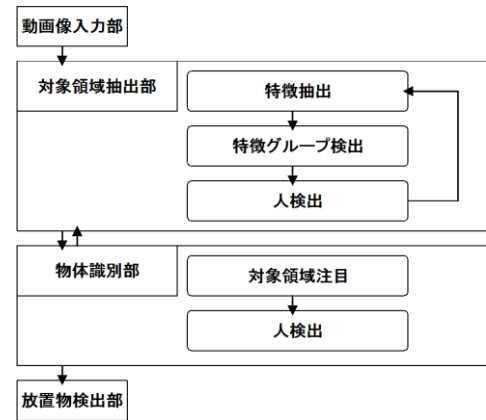


図 1 提案手法

提案手法は従来手法と同じく、大きく分けて動画画像入力部、対象領域抽出部、物体識別部、放置物検出部で構成されている(図 1)。従来手法の問題をより改善するために前述の手法変更に加え対象領域抽出部と物体識別部の処理を変更した。

動画画像入力部では従来通りパンチルトカメラを用いる。

### 4.2 対象領域抽出部

対象領域抽出部は特徴抽出・特徴グループ検出・人検出から成る。まず、特徴抽出において拡張 ST-

Patch 特徴量から時間変化特徴を抽出する (図 2).



図 2 拡張 ST-Patch 特徴量

次に, 時間変化特徴の現れた位置が近い特徴 (前景) をグループとしてまとめる. 最後に, 各グループに対して Deep Learning を用いた人検出を行う. 人が検出されれば拡張 ST-Patch 特徴量が出ていてもその前景を対象領域には含まない.

#### 4.3 物体識別部

物体識別部は対象領域注目・人検出からなる. 対象領域注目では, 対象領域となった部分に対してパンチルト機能を使用しズームする. その後, 物体識別部と同様の人検出を行う. ズームすることで, より正確な人検出が行える. 人が存在しなければ, 放置物があると判断する.

#### 4.4 放置物検出部

物体識別部において放置物と判断した場合, 物体の画像を保存する.

### 5. 予備実験

この章では, 畳み込みニューラルネットワークのプーリング法を検証する予備実験を行う.

#### 5.1 実験目的

MAXプーリングと平均プーリングの比較を行う. 最適なプーリング法は状況によって異なる. 人検出における最適なプーリング法を確認するために実験を行う. 実験後に優れているプーリング法で人検出を行う.

#### 5.2 データセット

人とそうでない画像それぞれ 900 枚準備し, 学習データに用いる. それと別に, それぞれ 700 枚の画像を用意し, テストデータに用いている. すべての画像のサイズは  $64 \times 128$  ピクセルとなっている. 学習用の人の画像を[5]から 900 枚取得した (図 3). テスト用の人の画像は研究室内で撮影した. 7 人の

人からそれぞれ 10 枚ずつ画像を取得した (図 4). 人でない学習画像とテスト画像は研究室内の画像をランダムに切り取り作成した (図 5).



図 3 学習人データ



図 4 テスト人データ



図 5 人でない画像

#### 5.3 実験条件

この実験では学習のために誤差逆伝番法を使用している. 学習回数は 2,000 回である. ネットワーク構造を以下に示す.

##### 1. 第一層

この層は入力層である. 特徴マップのサイズは  $64 \times 128$  ピクセルである.

##### 2. 第二層

畳み込み層である. 5 つの特徴マップを持っている.  $5 \times 5$  のフィルターを 1 ピクセルずつずらし, 掛け合わせ畳み込みを行う. 特徴マップのサイズは  $60 \times 124$  である.

3. 第三層  
プーリング層である。5 つの特徴マップを持っている。ここで、 $2 \times 2$  の領域に圧縮する。この領域が重なることはない。特徴マップのサイズは  $30 \times 64$  である。
4. 第四層  
第二層と同じ畳み込み層である。特徴マップのサイズは  $26 \times 58$  である。
5. 第五層  
第三層と同じプーリング層である。特徴マップのサイズは  $13 \times 29$  である。
6. 第六層  
全結合層である。30 ユニットの所持している。
7. 第七層  
この層は出力層である。この層は 2 ユニット所持していて、前の層と完全結合しています。  
第二層は同じ画像から特徴を抽出している。第二層と第三層の接続は表 1 が表している。X が接続を示している。

表 1 接続マップ

	0	1	2	3	4	5	6	7
0	x				x	x		
1	x	x				x	x	
2		x	x			x	x	x
3			x	x			x	x
4				x	x			x

	8	9	10	11	12	13	14	15
0	x	x	x		x	x	x	x
1		x	x	x		x	x	x
2			x	x	x		x	x
3	x		x	x	x	x		x
4	x	x		x	x	x	x	x

本実験では MAX プーリングと平均プーリングの識別制度をこの条件で検証する。

## 5.4 実験結果

表 2 実験結果

	MAX	平均
識別結果	92.7%	80.2%

表 2 は識別結果を表している。識別精度は MAX プーリングが 92.7%，平均プーリングが 80.2%であった。

## 5.5 考察

今回の実験結果より平均プーリングよりも MAX プーリングの識別精度が高いことがわかった。MAX プーリングは人のエッジを検出するこ

とに向いている方法であると考え。この結果より、人検出には MAX プーリングが有効であると考え。

## 6. 実験

本稿では、提案手法が有効であるか調べるために、以下の 5 つの状況を想定した実験を従来手法と同様に研究室内（図 6）で行った。入力部から放置物検出部までの流れでそれぞれの状況において適した処理が行われていることが確認できれば実験は成功である。

1. 複数の物体を置いた場合
2. 人だけが映る場合
3. 人と物をとり合わせて配置した場合
4. 人と物を離して配置した場合
5. 複数の物体と複数の人を配置した場合



図 6 研究室内

## 7. 結果・考察

### 7.1 複数の物体を置いた場合

まず図 7 のように物体のみ研究室に配置して実験を行った。これは物体を放置して 10 秒後の画像である。



図 7 放置物の配置

図 8 から ST-Patch 特徴量がそれぞれの物体に表



れていることがわかる。

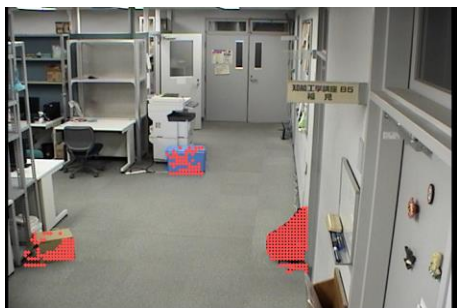


図 8 拡張 ST-Patch 特徴量

ここからグループ分けが行われる。特徴の現れている部分（図 9）に対してラベリング処理を行う。ラベリング処理とは、近くの特徴同士を一つのまとまりとして考えるための処理である。

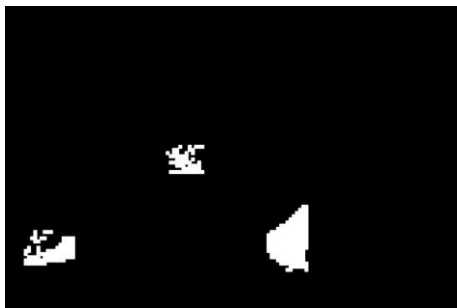


図 9 ST-Patch 特徴量

次に、ラベリングされたまとまりを少し大きめの四角で囲む（図 10）。この処理によりグループ分けが完了し、特徴グループが完成する。これを 2 段階ラベリング処理と呼ぶ。



図 10 2 段階ラベリング後



図 11 特徴グループに対しての人検出

特徴グループに対して人検出を行った結果が図 11 である。赤枠しかないので人として検出されている物が存在していないことがわかる。ここから物体識別部の処理に入る。人でない物体として認識された物体に対してそれぞれに注目するために、ズームを行う。ズームし、人検出を行いそれでも人が検出されなければ、そこにある物体は放置物として認識し画像として保存される。今回の実験状況では放置物として置いた 3 個の物体が正確に放置物として画像を保存できた（図 12-14）。この結果より、画面内に映っている複数の放置物を同時に検出できることがわかる。

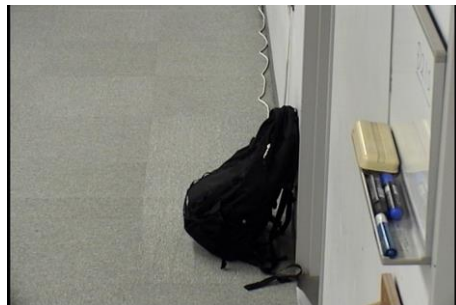


図 12 放置物 1



図 13 放置物 2



図 14 放置物 3

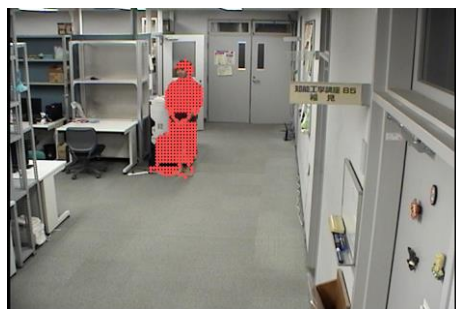


図 17 状況 3 での ST-Patch 特徴量

## 7.2 人だけが映る場合

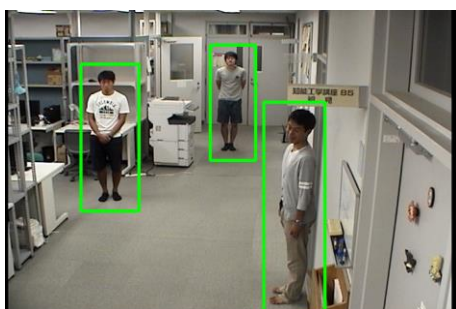


図 15 人だけが映る状況

次に人だけが映る状況である．物体は映っていないので，放置物が認識されなければ成功である．図 15 がそのときの対象領域抽出部の人検出処理後である．人のみが検出され放置物は認識されていないことがわかる．この結果より，画面内に複数の人物が存在していても正確に人検出は行われることがわかる．

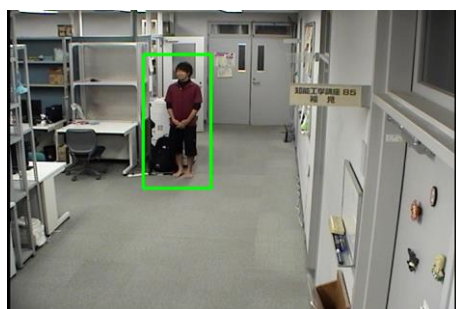


図 18 近くに物体がある場合の人検出

## 7.3 人と物をとり合わせて配置した場合



図 16 隣り合った人と物

人と物を近くに配置した場合の結果である．図 16 のような配置で実験を行った．人の近くに存在する物体は放置物としないと定義しているのでこの場合は物体が検出されなければ成功である．

## 7.4 人と物を離して配置した場合

次に，人から物を離して配置した場合である．人が近くにいない物体は放置物である可能性が高いため，人から離れた物体は放置物として検出する．画面内に人と物が同時に映っていても物体を放置物として検出できるかの実験である．配置は図 19 のようになっている．



図 19 人から離れて配置した物体

ST-Patch 特徴量は図 20 のように検出される．状況 3 のときとは違い，人と物体の特徴量が離れて検出されていることがわかる．



図 20 離れた ST-Patch 特徴量

この場合，人と物の二つの特徴グループに分けられ，それぞれに人検出を行う．図 21 より，人の特徴は緑枠で囲まれ，物体は赤枠で囲まれていることがわかる．これにより，対象領域抽出部において，人と放置物と予測される物体がそれぞれ正確に検出されることがわかる．



図 21 人と物体の検出

その後物体識別部に処理が移り，放置物の画像が保存された（図 22）．



図 22 状況 4 で検出された放置物

## 7.5 複数の物体と複数の人を配置した場合

5 つ目の状況として，画面内に複数の人と物体を配置して実験を行った．人間を 3 人と物体を 3 個配置した．物体のうち 1 個は人と重なる程度に配置し，持ち物として認識される配置にした．

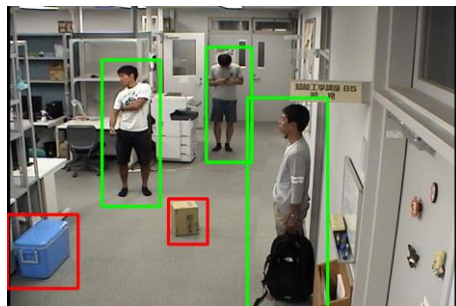


図 23 複数の人と物体の検出

図 23 がそのときの対象領域抽出処理を終えた後である．人が 3 人検出され，物体は近くに人が存在しない 2 個だけ検出され，2 個の物体のみが対象領域として検出されていることがわかる．人のすぐ横にある物体はその人の持ち物として認識され，対象領域に含まれることはなかった．この 2 個の物体を対象領域とし，物体識別に移る．物体識別部で放置物として判断され保存された画像が図 24，25 である．どちらもしっかりとズームされた画像が保存されていることがわかる．カメラに複数の人と複数の物体が写りこんでいる場合でも正確に放置物検出ができています．人の持ち物であると考えられる物体も正確に対象領域検出部で放置物から除くことに成功している．このことより，人間の人数が増え，放置された物体がいくつか映っていても提案手法により正確に放置物を検出できると考えられる．





図 24 提案手法



図 25 提案手法

## 8. まとめ

本稿ではパンチルトカメラを用いた放置物検出手法を提案した。提案手法として、従来手法にあったいくつかの問題点を改善し新たな手法を提案した。まず一つ目の問題点は、対象領域抽出部においての問題点である。時間変化による特徴を得るために従来手法では ST-Patch 特徴量を採用していた。しかし、これは輝度の変化にあまり強くなく、検出精度が低かった。これを改善するために、拡張 ST-Patch 特徴量を採用することでこの問題点を解決した。二つ目の問題点は、物体識別部においての問題点である。従来手法の物体識別部での人検出精度が高くなかった。これを改善するために、顔検出などで高い精度を出していた Deep Learning を人検出に採用した。本稿では特に、畳み込みニューラルネットワークを用いている。これにより人検出の精度を向上させることに成功した。対象領域抽出の精度向上や人検出の精度が向上したことで、手法の流れを少し変えより放置物を検出しやすい手法に変更した。対象領域抽出部においてラベリング処理を用いて対象領域を複数検出できるように改善したことで、一度に複数の放置物を検出できるようになった。そして対象領域抽出部に人検出を導入したことで、より精

度の高い放置物検出手法が完成した。

提案手法を用いて 5 つの状況における放置物検出実験を行った。人の人数や物体の数、カメラへの写り方などさまざまな状況を考えて実験を行った。その結果、いま実験している研究室内では一度に精確に放置物を検出できる結果が得られた。人の検出においても高い精度を出し、人の近くに存在する物に関しては放置物として認識することはなかった。よって提案手法は放置物検出において有効であると考えた。しかしながら、現段階では研究室内の狭い範囲でしか実験を行うことしかできていない。今後の予定として、この手法で実際に広い範囲をもった場所でも放置物検出を精確に行えるのか実験する必要があると考える。広大な敷地を持ちたくさんの人が行きかう場所で放置物検出実験を成功させたい。

## 参考文献

- [1] “防犯・監視カメラ関連業者の市場拡大加速”，株式会社帝国データバンク，  
<http://www.tdb.co.jp/report/watching/press/pdf/p150309.pdf>.
- [2] Takuma Ogawa, Momoyo Ito, Shin-ichi Ito and Minoru Fukumi: “Detection of abandoned object by pan-tilt camera”, Proceeding of 2015 International Workshop on Nonlinear Circuits, Communication and Signal Processing NCSP'15, pp.278-281, (2015).
- [3] 小川拓真, 平岡大樹, 伊藤伸一, 伊藤桃代, 福見稔: “放置物検出における ST-Patch 特徴量の改良”電気学会 電子・情報・システム部門大会, PS2-4, 1490, (2015)
- [4] LeCun, Y., Boser, B., Denker, j. S., Henderson, D., Howard, R. E., Hubbard, W. and jackel, L. D.: Backpropagation Applied to Handwritten Zip Code Recognition, Nueal Computation, Vol 1, No.4, pp.541-551,1989.
- [5] MIT cbcl (2000) CBCL Pedestrian Database[Online]  
<http://cbcl.mit.edu/software-datasets/PedestrianData.html>