H-021

# HLAC 特徴を用いた暴力シーンリアルタイム検出
# Real-Time Violence Scene Detection Using HLAC Feature

千田 恭平†　　菊池 拓磨†　　伊藤 慶明†　　小嶋 和徳†
Kyohei Chida　Takuma Kikuchi　Yoshiaki Ito　Kazunori Kojima

## 1　はじめに

近年，AKB 握手会傷害事件といった監視カメラの重要性が問われる事件が多発している．暴力・傷害事件に焦点を当てた場合，警察庁の統計[1]によると，平成 25 年の暴力・傷害事件の検挙率は約 70%となっている．検挙率を年別に見てもほぼ横ばいであり，監視カメラを設置していても検挙するに至らないことが多い．その理由として，監視カメラの解像度が低いために個人の特定が難しい，事件発生後に迅速に対応できないために加害者に逃げられてしまうといったことが考えられる．監視カメラからの映像を人の目でリアルタイムに監視するには負担が大きく，事件発生後に監視カメラの映像を一から確認するにも時間がかかり，人手の負担が大きくなっていることが予想される．そこで，暴力動作を監視カメラからリアルタイムで検出し，そのシーンの解像度を一時的に上げ，別途保存しておく等することで検挙率向上の手助けに繋がるのではないかと考えられる．

本研究では，暴力事件発生後に迅速に対応できるよう監視カメラから暴力動作をリアルタイムで検出する手法を提案する．ここでは，暴力動作をパンチとキックと定義している．

## 2　暴力シーンリアルタイム検出

本研究の大まかな流れを図 1 に示す．

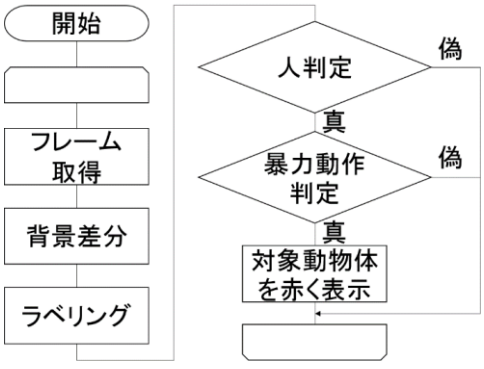


図 1　フローチャート

まず，監視カメラからフレームを取得し，背景差分法によって動物体を検出する．得られた動物体に対してラベリングを行い，その動物体が人かどうかを HOG 特徴量と Real AdaBoost により判定する．人と判定された動物体に対しては，暴力動作を行っているかどうかを HLAC 特徴量と SVM により判定する．暴力動作と判定された場合は，対象動物体を赤く表示する．

†岩手県立大学大学院ソフトウェア情報学研究科

### 2.1　動物体検出

通常の人検出では，任意の大きさの走査ウィンドウを用意し，大きさを変えながらラスタスキャンを行っている．走査ウィンドウのスケーリング率，ずらし幅の細かい設定によって高精度な検出が可能である反面，時間がかかってしまい，リアルタイムでの処理が難しい．そこで，背景差分法によって動物体のみを検出し，ラベリングを行った後，それぞれの動物体が人かどうかを判定する．動物体のみを対象としているため，何度もラスタスキャンを行う手法よりも非常に高速でリアルタイムでの検出が可能である．通常の背景差分法では，背景画像として用いる画像が不変のため，時間の経過による照明の変化等に対応できない．背景差分法をベースとした手法がいくつも提案されているが，本研究では Hofmann らによる手法[2]を用いた．この手法では，学習パラメータに基づいて背景画像を更新するため，時間の経過による照明の変化等にも頑健な動物体検出が可能である．2 値のマスク画像を取得することになるが，ここでラベリングを行い，ピクセル数が少ない動物体領域はノイズとして削除する．残った動物体領域にのみ，人判定を行う．

### 2.2　人判定

人判定では，動物体検出で得られた領域に対し，HOG 特徴量を用いて Real AdaBoost による判定を行う．HOG 特徴量とは，1 つの局所領域内におけるエッジ方向ごとのエッジ強度に着目した特徴量である．おおまかな物体の形状を捉えることが可能なため，形状や姿勢の識別に有効的な特徴量である．Dalal らの検証[3]では，人判定に最適なパラメータはセルサイズを 8×8 画素，エッジ方向(0°～180°までの範囲)を 20°ずつ 9 方向とし，ブロックは 2×2 セルで構成される 36 次元ベクトルとしている．本研究でも同様のパラメータを用いた．本来の特徴抽出ではグレースケール画像を用いるが，暴力動作判定でマスク画像を用いるため，ここでセグメンテーションが成功しているかどうかも検証する意味合いも兼ねて 2 値画像を使用した．

Real AdaBoost とは，Positive クラスと Negative クラスから抽出した特徴量を用いて学習し，識別器を生成する手法である．弱識別器の出力値の総和が閾値以上であれば，人と判定される．形状や姿勢を細かく識別するため，図 2 のような 7 つの Positive クラスと 1 つの Negative クラスを用意し，それぞれ個別に学習する．(1)と(2)は右・左向きの直立，(3)は前面・背面の直立，(4)と(5)は右・左向きのパンチ，

(6)と(7)は右・左向きのキック，(Neg)はそれ以外を表している．個別に識別器を作成しているため，弱識別器の出力値の総和が最も大きいクラスに分類することとし，(4)～(7)に分類された動物体領域にのみ暴力動作判定を行う．


図 2　マスク画像

### 2.3　暴力動作判定

暴力動作判定では，人判定で得られた領域に対し，HLAC 特徴量を用いて SVM による判定を行う．HLAC 特徴量とは，形状を数値化した特徴量であり，図 3 のような 25 個のマスクパターンをマスク画像に当てはめ，ヒストグラムを計算する．SVM とは，与えられた教師データを用いて 2 クラスのサポートベクトルとの間のマージンが最大になるように識別境界を決定する手法である．多クラスの判定を行う場合は，クラスの組み合わせごとに SVM で 1 対 1 の判定を行い，その結果を多数決で統合する one-versus-one によって行う．人判定に用いた 7 つの Positive クラスを SVM の学習にも使用し，図 2 の(4)～(7)に判定されれば，暴力動作とした．暴力動作と判定された人領域は赤く表示する．

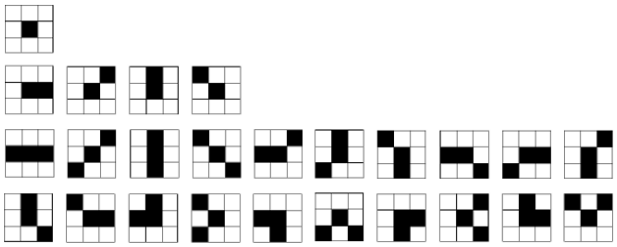

図 3　HLAC 特徴量のマスクパターン

## 3　評価実験

### 3.1　実験条件

学習では 7 つの Positive クラスと 1 つの Negative クラスを使用するため，それぞれ 3,000 枚ずつ用意し個別に学習を行った．評価動画は Web カメラを用いて撮影し，15fps で構成される 9 分程度の動画を用意した．なお，今回はパンチとキックといった動作を検出することを目的としているため，対人関係は考慮せず，フレームインできるのは 1 人のみとした．この動画には暴力動作 100 シーンが含まれており，暴力動作を行っている区間に暴力動作を検出したフレームが 5 フレーム以上含まれていれば正解とした．逆に，5 フレーム以下であれば未検出とした．暴力動作を行っていない区間に関しては，暴力動作を検出したフレームが 1 フレームでもあれば誤検出として扱うこととした．検出率と誤検出率は，以下のように定義した．また，暴力シーン区間は予め手動により決定しており，これに基づき精度評価を行なっている．

$$\text{検出率} = \frac{\text{正解数}}{\text{全暴力シーン数}} * 100$$

$$\text{誤検出率} = \frac{\text{誤検出数}}{\text{非暴力シーンの総フレーム数}} * 100$$

### 3.2　実験結果

評価実験を行った結果を表 1 に示す．暴力シーンの検出率は 89%，非暴力シーン区間の誤検出率は 3.3%という良好な結果が得られた．また，1 フレームの平均処理時間は 34.2ms となっている．

表 1　実験結果

| 検出率 | 誤検出率 | 平均処理時間 |
|---|---|---|
| 89%(89/100) | 3.3%(230/6,922) | 34.2ms |

## 4　考察

検出に失敗した区間を見てみると，動物体検出の段階で検出に失敗しているため，その後の処理に悪影響を及ぼしていると考えられる．また，誤検出が発生したフレームを見てみると，背景差分法によるセグメンテーションに失敗してしまい，正しく形状を特徴量化できなかったために誤検出が発生したのではないかと考えられる．図 4 に誤検出例を示す．しかし，全体的に見てみると，本研究は誤検出も少なく，検出率も高い手法であることを確認した．また、1 フレームの平均処理時間が 34.2ms であり，1 秒間に 29 フレームを処理できる．15fps の動画のため，リアルタイムでの検出が可能であることも確認した．


図 4　誤検出例

## 5　おわりに

背景差分法による動物体領域にのみ人判定を行う事でリアルタイムでの処理が可能であり，検出率も 89%と高く，本研究の有効性を確認した．対人関係や時系列データを含むことで，より高精度な暴力動作の検出が今後の課題となる．

## 参考文献

[1] 警察庁: “平成 25 年の犯罪情勢”, p72, 2014.
[2] M.Hofmann and P.Tiefenbacher, G.Rigoll: “Background Segmentation with Feedback: The Pixel-Based Adaptive Segmenter”, in proc of IEEE Workshop on Change Detection, 2012.
[3] N.Dalal and B.Triggs: “Histograms of Oriented Gradients for Human Detection”, Computer Vision and Pattern Recognition, Vol.1, pp886-893, 2005.