

プライバシー保護画像を用いた異常行動検出システムの開発

Abnormal behavior detection using privacy protected videos

高木修平 (九州大) 岩下友美 (九州大) 長原 一 (九州大)
 諸岡健一 (九州大) 正 倉爪 亮 (九州大)

Shuhei TAKAKI[†], Yumi IWASHITA[†], yumi@ieee.org
 Hajime NAGAHARA[†], Kenichi MOROOKA[†], Ryo KURAZUME[†]
[†] Kyushu University

Visual surveillance, which relies on human motion recognition and people recognition, has received a lot of attention for its use in effective monitoring of public places. However, there is a concern of loss of privacy due to distinguishable facial information. To deal with this issue, we developed a camera system which does not capture any facial information. In this paper we propose an abnormal-behavior detection method using privacy-protected videos taken by the proposed camera system. In the proposed method, we extract both motion-based and appearance-based features, and we combine these two methods by taking advantages of each of them. We build a database including normal and abnormal behaviors, and we show the effectiveness of the proposed method on cases from the database.

Key Words: Privacy-protected videos, Activity recognition, Abnormal behavior

1 はじめに

現代社会では、分散カメラを用いた監視システムは、安心安全な社会の実現に欠くことのできない基本的な社会インフラとして定着しつつある。一方、カメラを用いた監視システムによる個人特定や人物追跡の精度が向上するほど、常に個人のプライバシーの侵害との兼ね合いが問題となる。例えば、Google street viewなどで用いられているプライバシー保護手法では、通常のカメラで撮影した画像に対して、人物の顔部分などプライバシー性の高い箇所を自動（もしくは手動）で検出し、ぼかし処理などにより隠蔽している。また、馬場口らは、監視カメラに映る人物に対して隠蔽・透明化・抽象化・アノテーション処理を施す手法を提案している [1]。しかしこれらの手法では、プライバシー保護の対象も一旦は完全に画像化されており、心理的抵抗や情報流出の危険性など、完全なプライバシー保護とはなっていない。

これに対し我々は現在、顔のような特定対象のみを光学的に排除することでプライバシー問題を抜本的に解決する「匿名カメラ」を開発している [2]。匿名カメラは、まず赤外線カメラより撮影された人物の温度分布を基にピクセル毎に人物の顔や手の領域の推定を行い、次に推定された領域に対してマスク処理を行うことで、顔や手の輪郭を保持したまま人物のプライバシー情報を能動的に排除することができる。また、匿名カメラは顔や手などのマスクの位置を同時に取得することができる。以上のことから、匿名カメラはプライバシー保護対象を一度も光学的に画像化しないことから完全なプライバシー保護システムである。この匿名カメラを公共の場に設置することで、一般市民の監視システムに対する不安感を払拭でき、本来対立すべき監視とプライバシーの両立が可能となる。特に、匿名カメラを用いた監視システムは、患者の転倒などの異常動作が多発するにも関わらず、極めて高い秘匿性が求められる病院内環境において、非常に有効であると考えられる。そこで本論文では、匿名カメラを用いた病院内における監視システム (図 1) を提案する。

本論文では、まずマスクの位置の時系列情報から移動ベクトルの大きさと方向の変化を表現可能な Histogram of oriented Changes (HoC) [3] と、回転やスケールの変化に対して頑健に人物の形状情報を記述できる 2 次元アフィンモーメント不変量 [4] をそれぞれ用いた 2 つの動作検出手法を提案する。さらに、前述した 2 つの手法を組み合わせ、立ち止まりなどの動きを伴わない動作や複数人物の交差によるオクルージョンに対しても頑強な動作検出手法を提案する。最後に、病院内で撮影した動作データ

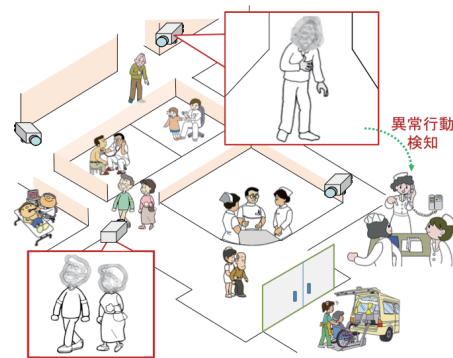


Fig.1 An example scenario of an abnormal behavior detection using privacy-protected videos

ベースを用いて実験を行い、提案手法の有効性を示す。

2 匿名カメラシステム

まず、本研究で用いる匿名カメラシステムの概要を図 2 に、匿名カメラの概要を図 3 に示す。図 2 および図 3-a,b に示す様に、匿名カメラは赤外線カメラと LCoS カメラ、コールドミラーにより構成される。ゲルマニウム製のコールドミラーは赤外線を透過し可視光を反射する。これにより匿名カメラでは、赤外線カメラと LCoS カメラによって、同一視野の赤外線画像と可視光画像を同時に撮影できる。プロトタイプカメラでは、赤外線カメラとして ARTCAM-320-THERMO QVGA カメラ (320×240 pixels at 60fps) を用いた。LCoS カメラは、図 3-c,d に示す様に、対物レンズやリレーレンズ、ビームスプリッタ、LCoS (800×600 pixels at 60fps, 島津製 HMD Data Glass 3/A から取り出した)、CCD カメラ (Point Grey, Flea3, 1280×1024 pixels at 60fps) によって構成される。シーンより LCoS カメラに入射してきた光は対物レンズおよびリレーレンズを通ったあと、ビームスプリッタにより LCoS の方向に反射され、リレーレンズにより LCoS 上に結像する。LCoS は電気信号により、反射輝度を制御することができる。その後、LCoS により反射された光が、リレーレンズとビームスプリッタを通過し、再度 CCD カメラのセンサ上に再結像する。つまり、LCoS により一部が光学的にマスクされ、反射された光のみが CCD に到達し電子データとして記録される。

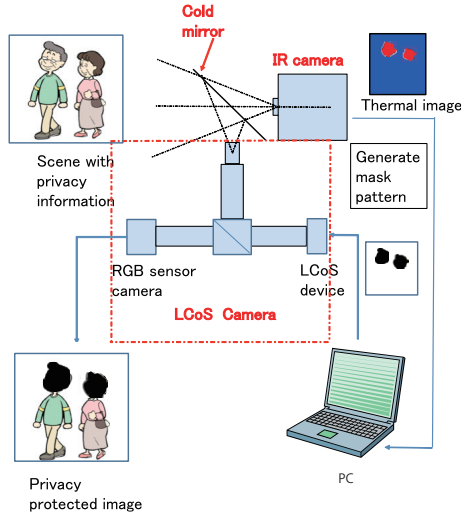


Fig.2 An overview of the camera system for privacy-protected videos

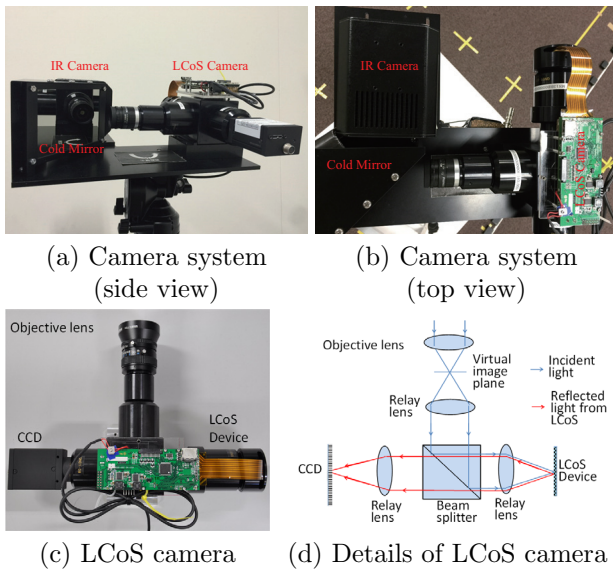


Fig.3 The camera system for privacy-protected videos

図2に示す様に、赤外線カメラにより撮影された赤外画像から体温を閾値としてPC上で二値化画像を生成することにより、顔領域のマスク画像を作成する。一方で、シーンからの可視光線は、コールドミラーで反射され、LCoSカメラに入射される。PCで作成された顔領域マスクは、LCoSに入力されシーン中の光線の顔領域を光学的にブロックする。図2、図4に示す様に、最終的にRGBカメラにより撮影される画像には顔からの光線が届かないため、顔のピアランス情報を欠落させたプライバシー保護撮影を行うことができる。

3 匿名カメラを用いた異常動作検出

本章では、提案する3つの動作検出手法について述べる。まず、動きの特徴であるHoCを用いた動作検出手法について述べ、次に見えの特徴である2次元アフィンモーメント不変量を用いた動作検出手法について述べる。最後に、動きと見えの特徴を統合することで動作を検出する手法について述べる。

3.1 動きの特徴を用いた動作検出

本節では、まずマスクの位置の時系列情報から移動ベクトルの大きさと方向の変化を表現可能なHoCの抽出方法について述べ、次に1クラスSupport Vector Machine(SVM)[5]と投票を用い



(a) An example of (b) A masked image captured images

Fig.4 Examples of captured images and masked images

た動作検出手法について述べる。

3.1.1 Histogram of oriented Changes

まず、図5にHoCの抽出方法を示す。HoCの抽出では、まずマスク画像列から取得したマスクの位置情報に対して最近傍探索を行い、重心座標に関する時系列データを取得する。次に、時系列データを基に各フレームにおける大きさの比 $\varsigma(x_{\Delta f}(n))$ と変位の方向 $\theta_{\Delta f}$ を計算をする。最後に、各フレームで計算した大きさの比と変位の方向に基づいてヒストグラムを計算する。

HoCの詳細な計算方法について述べる。まず、動きベクトルの大きさの比 $\varsigma(x_{\Delta f}(n))$ の計算方法について述べる。ここで、重心座標に関する時系列データは取得済みとする。フレーム n に対する重心座標を $p_{\Delta f}(n)$ とする。ここで、 Δf はフレーム間隔である。計算するフレーム間隔が Δf であるとき、フレーム n 、フレーム $n-1$ における重心座標 $p_{\Delta f}(n)$ 、 $p_{\Delta f}(n-1)$ から得られる動きベクトルの大きさを $L_{\Delta f}(n)$ とする。このとき、フレーム n 、フレーム $n-1$ から得られる動きベクトルの大きさは $L_{\Delta f}(n)$ 、 $L_{\Delta f}(n-1)$ であり、フレーム n に対する動きベクトルの大きさの比 $x_{\Delta f}(n)$ は次式で表される。

$$x_{\Delta f}(n) = \left| \frac{L_{\Delta f}(n)}{L_{\Delta f}(n-1)} \right| \quad (1)$$

ここで、式1は分母と分子の $L_{\Delta f}$ の値によって0から無限大の値をとり得る。そのため、次式によって正規化を行う。

$$\varsigma(x_{\Delta f}(n)) = \frac{1}{1 + e^{-4(x_{\Delta f}(n)-1)}} \quad (2)$$

$$\text{where } 0 \leq \varsigma(x_{\Delta f}(n)) < 1$$

ここで、式2はシグモイド関数に基づく式である。また、式2では入力値 $x_{\Delta f}(n)$ に対して関数の出力値 $\varsigma(x_{\Delta f}(n))$ は0から1の値をとる。

次に、変位の方向 $\theta_{\Delta f}$ の計算方法について述べる。計算するフレーム間隔が Δf であるとき、フレーム n に対する前後のフレームはフレーム $n-\Delta f$ 、フレーム $n+\Delta f$ となり、フレーム n に対する変位の方向 $\theta_{\Delta f}(n)$ は次式で表される。

$$\theta_{\Delta f}(n) = \tan^{-1} \frac{(a_{\Delta f}(n+\Delta f))_y}{(a_{\Delta f}(n+\Delta f))_x} - \tan^{-1} \frac{(a_{\Delta f}(n))_y}{(a_{\Delta f}(n))_x} \quad (3)$$

$$\text{where } 0 \leq \theta_{\Delta f}(n) < 360$$

ここで、 $a_{\Delta f}(n+\Delta f)$ 、 $a_{\Delta f}(n)$ はそれぞれフレーム n における前後フレームとの差分 $a_{\Delta f}(n+\Delta f) = p_{\Delta f}(n+\Delta f) - p_{\Delta f}(n)$ 、 $a_{\Delta f}(n) = p_{\Delta f}(n) - p_{\Delta f}(n-\Delta f)$ である。

最後に、ヒストグラムの計算方法について述べる。あるフレーム n における動きベクトルの大きさの比 $\varsigma(x_{\Delta f}(n))$ と変位の方向 $\theta_{\Delta f}$ が計算されているとする。このとき、フレーム n は式4、5を同時に満たすヒストグラムのビンに格納される。

$$(l-1)\Delta a \leq \varsigma(x_{\Delta f}(n)) < l\Delta a \quad (4)$$

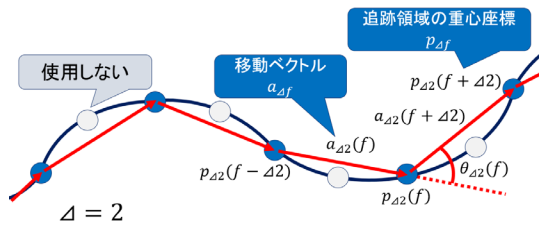


Fig.5 The idea of the HoC

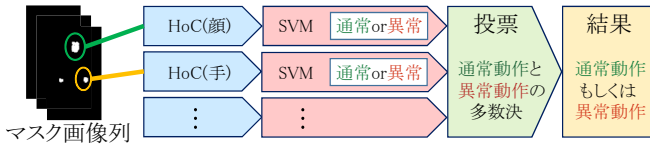


Fig.6 Abnormal behavior detection using voting-based method

ここで、 Δa は大きさの比の分割数 (式 2 のとり得る値 0~1 を何分割するか) によって決定され、大きさの分割数を L としたとき $\Delta a = \frac{1}{L}$ となる。

$$(w-1)\Delta\theta \leq \theta_{\Delta f}(n) < w\Delta\theta \quad (5)$$

ここで、 $\Delta\theta$ は角度の分解能 (360 度を何分割するか) によって決定され、角度の分解能を W としたとき $\Delta\theta = \frac{360}{W}$ となる。この計算を N 枚の全てのフレームに対して計算することで、HoC 特徴ヒストグラムを得る。

3.1.2 投票による動作検出

まず、投票による動作検出の流れを図 6 に示す。投票による動作検出手法では、まず顔や手などから抽出した全ての HoC に対して SVM による識別を行う。ここで、SVM には 1 クラス SVM を用いて歩行動作のみを学習データとし、識別対象となる HoC が通常か異常かの判定を行う。次に、全ての SVM の結果から通常と異常の多数決を行い、多数決結果を入力テストデータの識別結果として出力する。

3.2 見えの特徴を用いた動作検出

本節では、まず背景差分を用いて抽出した対象人物領域から見えの特徴である 2 次元アフィンモーメント不変量を抽出する方法について述べる。次に、2 次元アフィンモーメント不変量を用いた動作検出手法について述べる。

3.2.1 2 次元アフィンモーメント不変量

2 次元アフィンモーメント不変量の抽出では、まず撮影画像に対して背景差分を行い対象人物領域を抽出する。次に、抽出した対象領域から見えの特徴として 2 次元アフィンモーメント不変量を抽出する。

2 次元アフィンモーメント不変量はモーメントベースの記述子であり、一般的なアフィン変換に対して不変である。二次元平面上における $(p+q)$ 次のモーメントは次の式で表される。

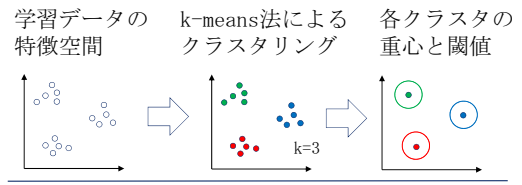
$$\mu_{pq} = \sum \sum_{(x,y) \in O} (x-x_g)^p (y-y_g)^q I(x,y) \quad (6)$$

ここで、 $I(x,y)$ は座標 (x,y) における画像の輝度値を示し、また x_g および y_g は人物領域の重心位置を示す。

また、モーメントは形状の特徴を示しており、0 次モーメントは対象物体領域の面積、2 次モーメントは対象の重心周りの分散、3 次モーメントは対象物体の歪度を示す。以下に本研究で用いる計 80 個の 2 次元アフィンモーメント不変量 $I = \{I_1, I_2, \dots, I_{80}\}$ のうちの 1 つを示す。

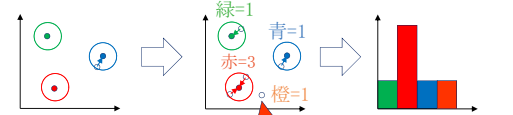
$$I_1 = \frac{1}{\mu_{00}^4} (\mu_{20}\mu_{02} - \mu_{11}^2) \quad (7)$$

学習 (通常動作のみ入力)



テストデータ入力

1 つめのテストデータ入力



外れ値

Fig.7 Bag-of-Visual Words

3.2.2 Bag-of-Visual Words (BoVW) + SVM による動作検出

本節では 1 クラス学習を行うため BoVW と SVM を組み合わせた手法について述べる。図 7 に BoVW によるベクトル量子化を示す。BoVW の学習では、まず学習データを特徴空間に投影し、k-means 法によるクラスタリングを行う。図 7 の例では、クラスタ数を 3 としており、各色がそれぞれのクラスタを示す。次に、各クラスタの重心を求め、計算したクラスタ重心を基にクラスタ毎の平均と分散を求める。最後に、クラスタ毎に計算した平均と分散を基に、クラスタ毎の閾値を決める。ここで、本研究ではクラスタ毎の閾値はクラスタ平均 + クラスタ分散と定義した。以上が BoVW の 1 クラス学習である。テストデータが入力された場合は、まずテストデータに対して最近傍となるクラスタをクラスタ重心から求める。次に、特徴空間上においてテストデータと最近傍となるクラスタ重心との距離がクラスタの閾値を越えていないかを計算する。ここで、テストデータがクラスタの閾値を越えていない場合はテストデータが最近傍クラスタに所属していると扱い、クラスタの閾値を越えていた場合は、そのテストデータを外れ値として扱う。ここまでの処理を 1 シーケンス全てのテストデータに対して行う。最後に、各計算結果を基にヒストグラムを計算し、新たな特徴ベクトルを得る。動作検出は BoVW によって得られた新たな特徴ベクトルを基に SVM による識別を行うことで実現する。

3.3 動きと見えの特徴を組み合わせた動作検出

本節では動きと見えの特徴を組み合わせた動作検出について述べる。動きと見えの特徴を組み合わせた動作検出は 2 ステップから構成される。まず、あるテストデータに対して動きの特徴を用いた動作検出と見えの特徴を用いた動作検出を行い、この 2 つの手法に対して識別結果を出力する。次に、これらの識別結果を統合し、2 つの結果の片方でも通常動作であれば最終識別結果は通常動作とする。また、2 つの識別結果が共に異常動作であれば最終識別結果は異常動作とする。

このような統合手法を採用した根拠について述べる。動きの特徴と見えの特徴には、それぞれ利点と欠点がある。例えば、動きの特徴を用いた動作検出では複数人物への拡張が可能であるが、画面内の人物に動きがない場合特徴を取得することができない。つまり、動きの特徴を用いた動作検出ではテストデータとして動きのない動作を入力した場合、正しく識別されない可能性が高い。反対に、見えの特徴を用いた動作検出では動きのない動作に対しても対応が可能であるが、人物同士が重なるなどのオクルージョンに弱い。つまり、画面内に複数人物が同時に映り込む状況下において正しく識別されない可能性が高い。そこで、提案手法では 2 つの動作検出結果を OR (片方でも動作検出結果が通常動作なら

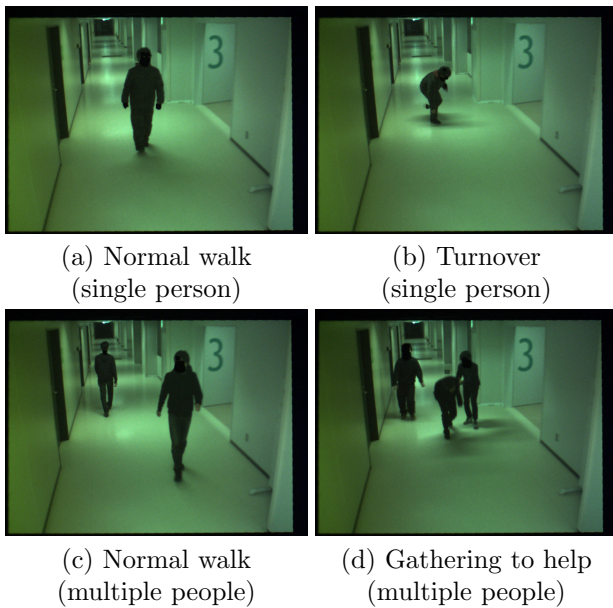


Fig.8 Database

ば通常動作, それ以外は異常動作)で統合する. OR で統合することで, 2つの動作検出手法のうちどちらかが正常に動作していれば, 最終識別結果は正常な結果となる.

4 実験

本章ではまず, 実験に用いた動作データベースの構築方法について述べ, 次に動作データベースを用いた実験について述べる.

4.1 動作データベースの構築

構築した動作データベースについて述べる. 動作データベース内の動作を大別すると, 通常動作と異常動作の2つに分けることができる. ここで, 本研究における通常動作とは, 病院内の廊下において通常とみなされる動作を意味し, 異常動作とは通常とは違う動作を意味する. また, 動作データベースは図8に示すように, 撮影時に単一人物しか映さない単一人物データベースと, 複数の人物を同時に撮影した複数人物データベースから構成される.

まず, 単一人物データベースについて述べる. 単一人物データベースは通常動作である歩行と立ち止まり, 異常動作である胸を押さえる, ぶらぶら歩く, 暴れる, 転倒の計6つの動作から構築される. 単一人物データベースの被験者は26名であり, 通常動作104($52 \times 2 = 104$)シーケンス, 異常動作104($26 \times 4 = 104$)シーケンスの計208シーケンスから構成される.

次に, 複数人物データベースについて述べる. 複数人物データベースは通常動作として歩行と立ち止まり, 異常動作として転倒した一人に対して他の人物が駆けつけるの計3つの動作から構築される. 複数人物データベースは, 単一人物データベースの被験者26名からランダムに2人もしくは3人を選び, 通常動作と異常動作についてそれぞれ撮影を行うことで構築した. また, 複数人物データベースは通常動作12($2 \times 6 = 12$)シーケンス, 異常動作12($3 \times 4 = 12$)シーケンスの計24シーケンスから構築される.

4.2 動作データベースを用いた実験

構築した動作データベースに対して3つの提案手法を適用した. ここで, すべての実験における学習データは単一人物データベースを用いて構築した. また, ある人物の動作をテストデータとする際には, その人物以外の歩行動作を用いて学習データを構築し, 識別率を求めた.

表1に単一人物データベースを用いた実験結果を, 表2に複数人物データベースを用いた実験結果を示す. 各表における(a)は動きの特徴を用いた動作検出手法を用いた場合, (b)は見えの

Table 1 Experimental results using a single-person database

| 識別率 % | 通常動作 | 異常動作 | 全体 |
|-----------|-------|-------|-------|
| 動きの特徴 (a) | 36.54 | 100 | 68.27 |
| 見えの特徴 (b) | 91.35 | 93.27 | 92.31 |
| 組み合わせ (c) | 94.23 | 90.38 | 92.31 |

Table 2 Experimental results using a multiple-people database

| 識別率 % | 通常動作 | 異常動作 | 全体 |
|-----------|-------|-------|-------|
| 動きの特徴 (a) | 50.00 | 100 | 75.00 |
| 見えの特徴 (b) | 66.67 | 91.67 | 79.17 |
| 組み合わせ (c) | 75.00 | 91.67 | 83.33 |

特徴を用いた動作検出手法を用いた場合, (c)は動きと見えの特徴を組み合わせた動作検出手法を用いた場合の実験結果である.

まず, 単一人物データベースを用いた実験結果(表1)では, 動きの特徴を用いた場合(a)の通常動作の識別率が低い. これは, 動きのない動作である立ち止まりが正しく識別できていないためである. これに対して見えの特徴を用いた場合(b)は通常動作の識別率が高いことを確認できる. また, これら2つを組み合わせた手法(c)では, 全体的に識別率が高いことを確認できる.

次に, 複数人物データベースを用いた実験結果(表2)では, 単一人物データベースを用いた実験結果よりも見えの特徴を用いた場合(b)の識別率が13%ほど低い. これは, オクルージョンにより識別率が低下したことが原因だと考えられる. 反対に, 動きの特徴を用いた場合では識別率は低下しなかった. これら2つを組み合わせた手法ではそれぞれの手法を単体で使用した場合よりも, 全体的に識別率が向上したことを確認できる.

5 まとめ

本論文では, 匿名カメラより取得できるプライバシー保護画像を用いた人物の異常動作検出手法を提案した. 匿名カメラより得られた2種類の画像列を用いて動作データベースを構築し, 提案する3つの手法を用いて異常動作検出実験を行った. 実験の結果, 動きと見えの特徴を組み合わせることで異常動作を高い識別率で検出できることを確認した. 今後は, 画面内に映るそれぞれの人物の動作を検出する手法を導入することで, より頑健な異常動作検出手法を実現する.

謝辞

本研究は, 総務省戦略的情報通信研究開発制度(SCOPE)ICTイノベーション創出型研究開発(121810005)の助成を受けた.

References

- [1] K. Chinomi, N. Nitta, Y. Ito & N. Babaguchi, "PriSurv: Privacy Protected Video Surveillance System Using Adaptive Visual Abstraction", Advances in Multimedia Modeling, Lecture Notes in Computer Science, vol.4903, pp.144-154, 2008.
- [2] Y. Zhang, Y. Lu, H. Nagahara, R. Taniguchi, "Anonymous Camera for Privacy Protection", ICPR2014, pp.4170-4175, 2014.08.
- [3] Yumi Iwahista, M. S. Ryoo, Thomas J. Fuchs, Curtis Padgett, "Recognizing Humans in Motion: Trajectory-based Aerial Video Analysis", BMVC, pp.127.1-127.11, 2013.
- [4] J. Flusser, T. Suk & B. Zitova, "Moments and Moment Invariants in Pattern Recognition", Wiley & Sons Ltd., 2009.
- [5] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola & R. C. Williamson, "Estimating the support of a high-dimensional distribution", Neural Computation, Vol. 13, No. 7, pp. 1443-1471, 2001.