

異常行動検知のための画像処理手法の開発

鵜飼 正人* 長峯 望*

Development of Image Processing Method for Detection of Suspicious Acts

Masato UKAI Nozomi NAGAMINE

We are developing an intelligent video surveillance system to maintain security at railway premises. In this paper, we propose a method to detect extraordinary people behaviors such as fighting or vandalism occurring at a station or aboard a train. This method comprises three processes: (i) Object Detection and Tracking, (ii) Video Scenario Recognition, (iii) Raising an Alarm over Critical Events. From acquired data by the tracking, we have analyzed actual image sequence by a combination of several properties attached to a target. Furthermore, we apply a principal component analysis (PCA) to distinguish suspicious acts from ordinary behavior. Some experiential results on actual paths have indicated the effectiveness and robustness of our proposed method.

キーワード：セキュリティ，監視カメラ，画像処理，画像認識

1. はじめに

交通機関や重要施設等に対する昨今のテロ不安に鑑み、鉄道においてもテロ対策の強化は大きな課題となっている。不特定多数の利用者が行き交う駅構内や、線路内等においてテロが発生した場合、社会的影響が甚大である。

海外の鉄道テロ事件などで、監視カメラ映像が容疑者特定の重要な手がかりとなったことを踏まえて、わが国の鉄道駅でも監視カメラの設置が進んでいる(図1)。ただし全てのカメラ映像を常時注視しているわけではないので、リアルタイムで異常を発見するのは難しい。むしろ証拠目的に録画しているというのが実状である。そこで鉄道総研では、主に駅及び車内監視を対象として、映像から異常を自動的に検知し、アラームと共にリアルタイムで人間に判定画像を提示する、知的なビデオ監視システムの開発を進めている。

まず最初に、異常検知システムの要素技術である人物の追跡アルゴリズムを検討した。次に、物の置き去り、転



図1 駅構内に設置された監視カメラ

倒、暴力等のシーンを解析し、不自然な人の姿勢や手足の動きなどを捉えることにより、これらの異常事象を検知する画像処理手法を開発した。さらに雑踏下の動画像から異常を検出するため、歩行軌跡に対して主成分分析を行い、固有空間への投影により異常動作と通常動作を識別するアルゴリズムを検討し、駅等の標準的な監視カメラ設置条件のもとで検証実験を行い有効性を示した。本稿ではこれらの結果について報告する。

2. 異常検知のための検証用映像の撮影

知的なビデオ監視システム^{1), 2)}を実現するための最初のステップは、映像から人間や荷物などを確実に検知し、その位置や動線を正確に把握することである。これらの画像処理に使用するテスト映像として、スタジオ及び駅構内にて、30人のエキストラにより通常の通行シーンと、図2に示す異常事象を撮影した。駅等にある一般的な監視カメラは大体、高さは3~4m、俯角は30~40度で設置されている(図3)。そこで我々も、カメラの高さと俯角が同じ条件になるよう調整し、画面内を人が上下左右、ランダムな方向に動くよう、歩く方向をいろいろ変えて撮影した。物体の検知率を算出するためにはある程度長時間のデータが必要なので、一つのシーンを20分程連続して撮影した。これらの撮影の様子を図4、図5に、暴力等の異常シーンの撮影の様子を図6に示す。

3. 対象物体の検知と追跡アルゴリズム

基本的な人物検知の手法には、フレーム間差分法や背

* 信号通信技術研究部(信号)

特集：信号通信技術

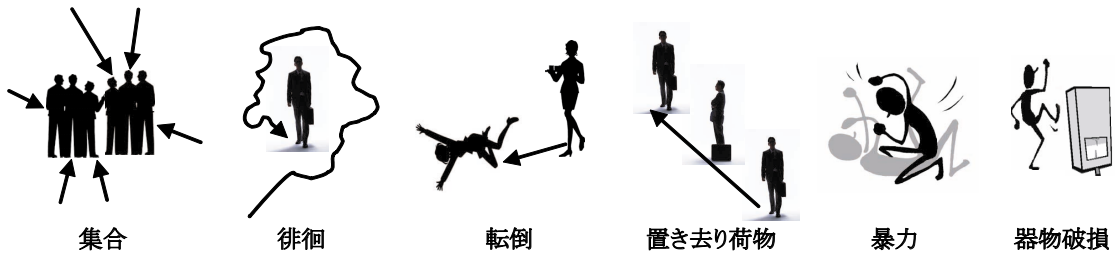


図2 検知目標とした異常事象（シーン）

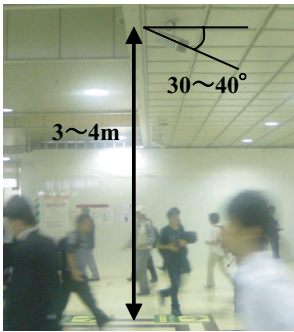


図3 一般的な駅構内の監視カメラ



図4 スタジオでの撮影風景と監視カメラ



図5 駅での撮影風景と監視カメラ

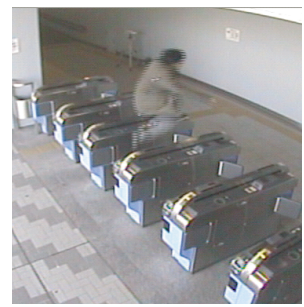
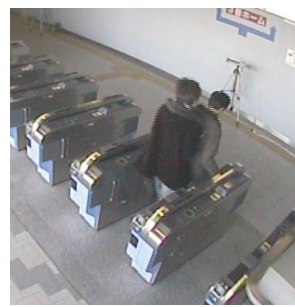


図6 駅での異常シーン（集合、暴力、改札共連れ、改札ジャンプ）の撮影風景

景差分法があるが、我々はこれらの手法をベースに、色による識別やカルマンフィルタによる動き推定アルゴリズムを適用した物体検知と追跡手法を考案した。カルマンフィルタは、前の時刻の物体の位置と移動速度の観測値から、現時点の推定値を計算する。そして今の時刻の観測値を用いて、推定値を補正してより正確な状態を推定する。さらに次の時刻の予測位置を推定するという、予測と更新の二つの手続きを再帰的に繰り返す。これにより、柱等による隠れや重なりで頑強な追跡を実現している（図7）。まず画面上での人の大きさや通常の歩行速度等、人を対象とする情報を登録し、最後にカルマンフィルタの詳細な感度調整を行い、リアルタイムで物体

の検知と追跡を行った（図8）。楕円で囲まれた検知人物の傍に表示している数字は、個々人を識別するIDである。実験環境における追跡性能を評価するために、図8のシーンにおいて、改札に向かう人と改札から出てくる人をそれぞれ計測し、人数をカウントした。変数 Count は実際の測定値である。本来いるのに検知できなかったカウント漏れと、重なりにより2人を1人として数えてしまったカウント不良を考慮した計測精度 *global accuracy* 及び *real accuracy* は表1の通りである。実験結果より、一般的な監視カメラの設置条件下では、10人程度の人物を85%程度の精度で同時に追跡できることがわかった。



図7 カルマンフィルタによる隠れに頑強な追跡アルゴリズム

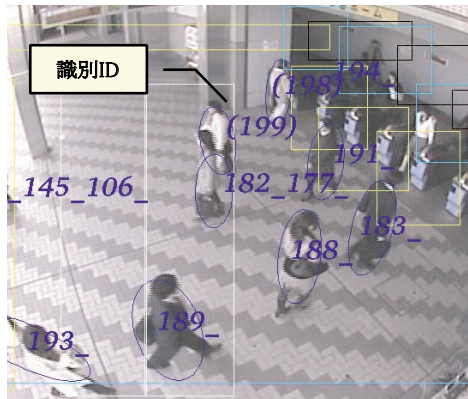


図8 雑踏下での複数人の追跡結果

4. シーン理解のための特徴解析

物体の追跡結果から得られる基本的な情報には、ターゲットの位置（座標値）、大きさ、輝度がある。一例として、画面上（スクリーン座標系）の物体の移動軌跡と、実空間（世界座標系）へ座標変換した軌跡をそれぞれプロットした（図9）。これらの情報を時系列に取得することで表2に示すような10数種類の特徴量を計測し、これらの組み合わせで特定の行動を検知することを考えた（表3）。

シーンの理解に有効と思われる13種類の特徴量について具体的に説明する。

- ▲ Age [sec]：物体を検知してから現在までの時間
- ▲ Size [pixel]：検知物体のピクセル数
- ▲ Speed [pixel/sec]：検知物体の単位秒あたりの移動pixel数

表1 改札通行シーンにおける検知結果

解像度 352 x 288 俯角 40°	左→右	右→左
カウント値（観測値） <i>Rcount</i>	92	96
カウント漏れ（未検知等） <i>MissCount</i>	23	19
カウント不良（人の重なり等） <i>FalseCount</i>	1	1
<i>global accuracy</i>	80.7%	84.2%
<i>real accuracy</i>	79.8%	83.3%

解像度 720 x 780 俯角 40°	左→右	右→左
カウント値（観測値） <i>Rcount</i>	92	96
カウント漏れ（未検知等） <i>MissCount</i>	18	15
カウント不良（人の重なり等） <i>FalseCount</i>	2	3
<i>global accuracy</i>	85.1%	88.8%
<i>real accuracy</i>	83.43	86.3%

$$Rcount = Count - FalseCount + MissCount$$

$$global\ accuracy = Count \div Rcount \times 100$$

$$real\ accuracy = (Count - FalseCount) \div Rcount \times 100$$

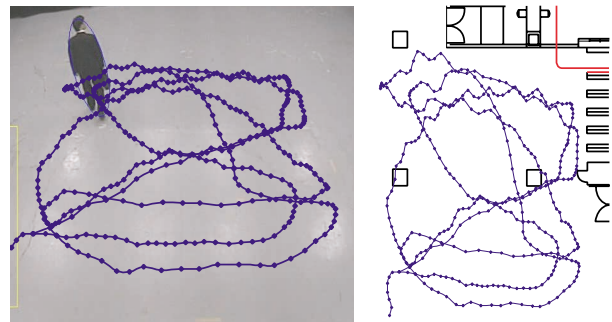


図9 物体の追跡結果（移動軌跡の表示）

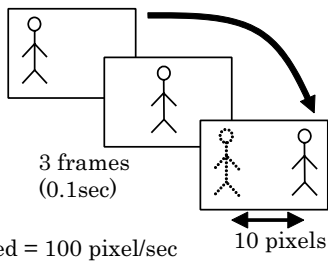
表2 追跡情報を基に計測される特徴量の例

特徴量	内容	特徴量	内容
Age	ターゲットが存在している時間	Weight	オブジェクト同士が融合する頻度。融合で+1，分離で-1
Size	ターゲットの面積	Occupancy	観察領域のうち、ターゲットの占有面積の比率
Speed	ターゲットの移動速度	Fsm	Frequency split & merge：融合／分離した回数を表す
Jitter	ターゲットの動きの変化（変化した物体のサイズ）を表す	CovRatio	ターゲットの縦と横のピクセル比
Mobility	単位時間（1秒）あたりの重心からの平均距離	Direction	ターゲットが移動している方向

表3 シーンの分析と検知に有効と思われる特徴量

シーン	定義	検知に有効と思われる特徴量
集合	3人以上が1箇所に集まってくる。また、集合状態から離れる。	Size, Speed, Weight, Age 等
徘徊	特定の人物が、ある範囲内を行ったり来たりうろろしている	Speed, Mobility, Age 等
転倒	一人の人物が歩いている、または走っている状態から倒れる	CovRatio, Mobility, Speed, Age 等
置き去り	荷物を持っていた人物が、荷物を置いて去っていく	Speed, Mobility, Age, Size 等
けんか	2人もしくは3人で、組み合いながら激しく手足を動かしている	Jitter, Mobility, Weight, Fsm, Size, Speed 等
暴力	1人が相手に対して、激しく手足を動かしている	Jitter, Mobility, Fsm, Size, Speed 等
破壊	1人が物の傍らで激しく手足を動かしている	Jitter, Mobility, Fsm, Size, Speed 等

特集：信号通信技術



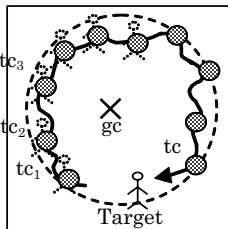
- ▲ Jitter [pixel/sec]: 単位秒あたりに変化した検知物体のサイズ。激しい動きほど大きくなる。

$$\frac{1}{30} \sum |size_i - size_{i-1}| \dots \dots \dots (1)$$

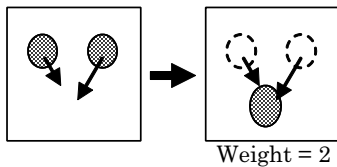
- ▲ Mobility [pixel]: 単位時間当たりの物体の軌跡の重心と物体の中心の距離の平均

$$\frac{1}{30} \sum |gc - tc| \dots \dots \dots (2)$$

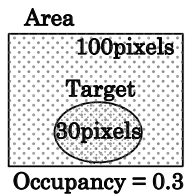
(gc: 軌跡の重心 tc: ターゲットの中心)



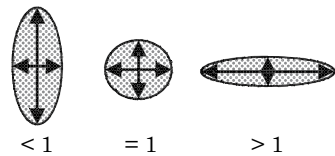
- ▲ Weight [none]: マージしている検知物体の数



- ▲ Occupancy [none]: エリアにおける物体の占有面積率



- ▲ FreqSplitMerge [none]: 単位時間当たりの分離/融合回数
- ▲ CovRatio: ターゲットの縦と横のピクセル比 width/height
- ▲ Direction [degree]: 物体が移動している方向



- ▲ AveSpeed [pixel/sec]: 物体が検知されてから現在までの平均速度

$$\frac{1}{age} \sum speed \dots \dots \dots (3)$$

- ▲ TimeToStopped [sec]: 物体が停止してから現在までの時間
- ▲ Stops [times]: 物体が停止した回数

5. 特徴解析手法に基づく異常行動検知結果

計測した時々刻々の特徴量を監視できるツールを作成した。例えば図の暴力シーンにおいて、③のタイミングでアラームを出すようにしたい場合でも、モニタリング機能により適切に閾値を決定することができる(図10)。

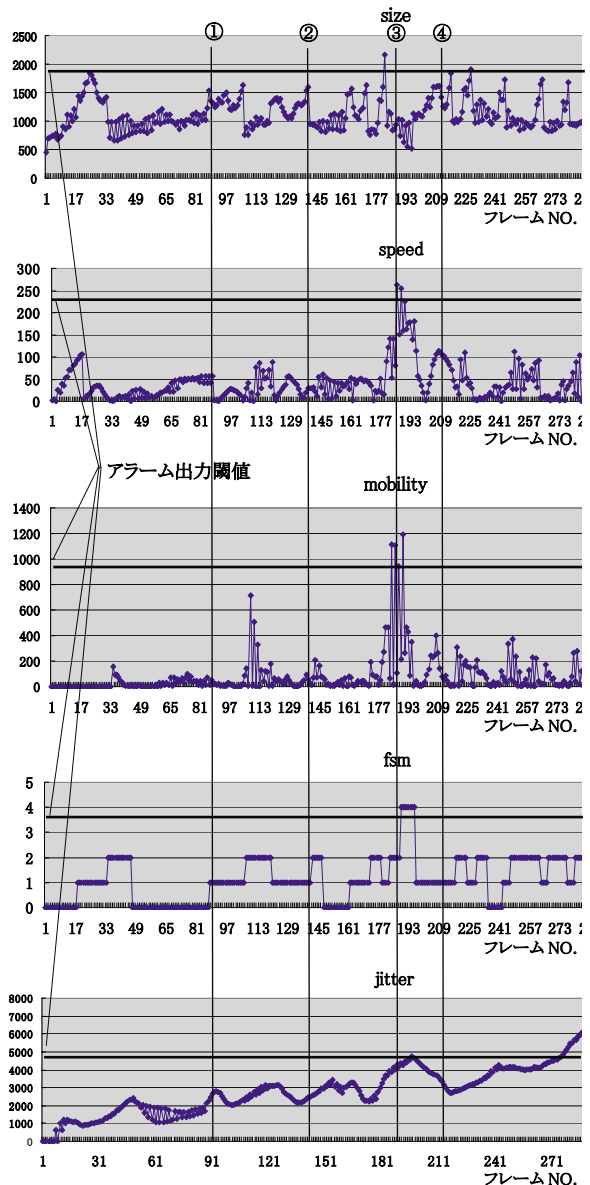
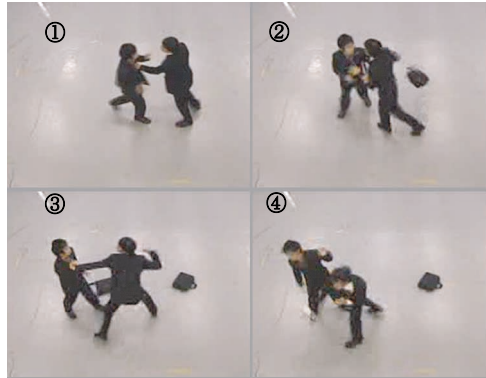


図10 時系列処理映像における特徴量のモニタリング

車内で発生するシーンを想定して、魚眼映像を透視変換した映像に対して、窓ガラスへのいたずら、車内暴力を検知した様子を図11に示す。また異常アラームが発生した際の詳しい状況は、例えばPTZカメラと連動させることにより、効率的に記録することができる(図12)。

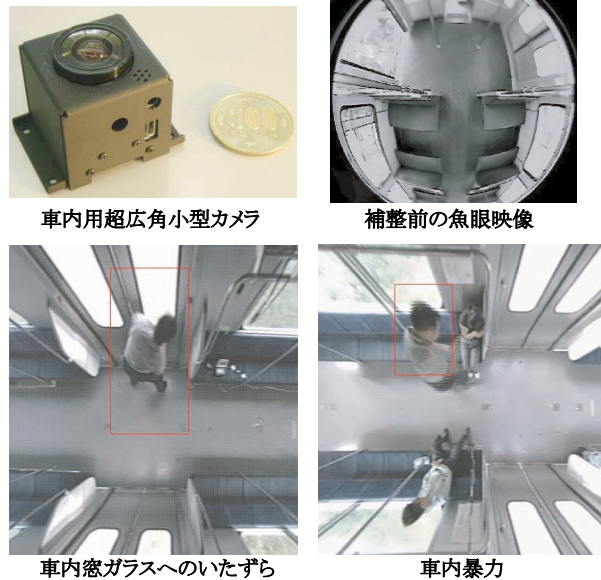


図11 超広角小型カメラによる車内異常シーン検出



図12 異常アラーム出力とPTZカメラとの連動例

6. 主成分分析と固有空間を用いた行動認識

前節までは予め想定したシーンを異常な事象として、これを直接検知する手法を検討してきたが、一般的に“異常”そのものを定義することはできない。しかし、異常は「普通と違うこと、いつもと違うこと」(広辞苑)とあるように、通常動作を分析、認識できれば、それに該当しない動作として検知することができる。通常動作は、統計的な分布を考えた場合に分布の集中する動作と考え、その分布から大きく逸脱する動作を異常な動作とする³⁾。そこで歩行軌跡に対して主成分分析を行い、固有空間への投影により通常の動きと異常な動きを

識別する手法を検討した。ある人物*i*の二次元座標の時系列(移動軌跡) Y_i を構成する M_i 個の二次元座標は、 t 番目の座標を $P_t = (p_{xt}, p_{yt})^T$ とすると、次のように長さ $2M_i$ のベクトル y_i で表される。

$$y_i = (P_1^T, P_2^T, \dots, P_{M_i}^T)^T \dots \dots \dots (4)$$

移動軌跡はオプティカルフローなどを用いた追跡手法により取得する。人によって軌跡は異なるため、それらを正規化する必要がある。例えば歩行速度の遅い人の場合、同じ経路を歩く足の早い人に比べて座標の数が多くなるし、軌跡の開始と終了場所は人によっても異なる。これらの差異を吸収するために線形補間による間引きなどで座標の数を同数にそろえた。図13に歩行軌跡の学習に適用した固有空間法の処理フローを示す。全サンプル画像ベクトル $y_i (i=1, \dots, M)$ の平均 C

$$C = \frac{1}{M} \times \sum_{i=1}^M y_i \dots \dots \dots (5)$$

を計算し、次に各画像ベクトルから平均 C を差し引き、行列 Y_i を求める。

$$Y_i \equiv [y_1 - C, y_2 - C, \dots, y_{2M} - C] \dots \dots \dots (6)$$

この時、画像集合の共分散行列 Q は式(7)で表される。

$$Q_i \equiv Y_i Y_i^T \dots \dots \dots (7)$$

例えば k 次元の固有空間は、次の固有方程式

$$\lambda_k e_k = Q e_k \dots \dots \dots (8)$$

を解き、固有値の大きい順の k 個に対応する固有ベクトル (e_1, \dots, e_k) を基底ベクトルにすることにより得られる。固有空間の次元は、累積寄与率が90%を超える次元数とした。歩行軌跡の相関が大きければ、散布図にプロットされた主成分得点の点間距離が小さくなるので、これより通常歩行と異常行動の識別を試みた。

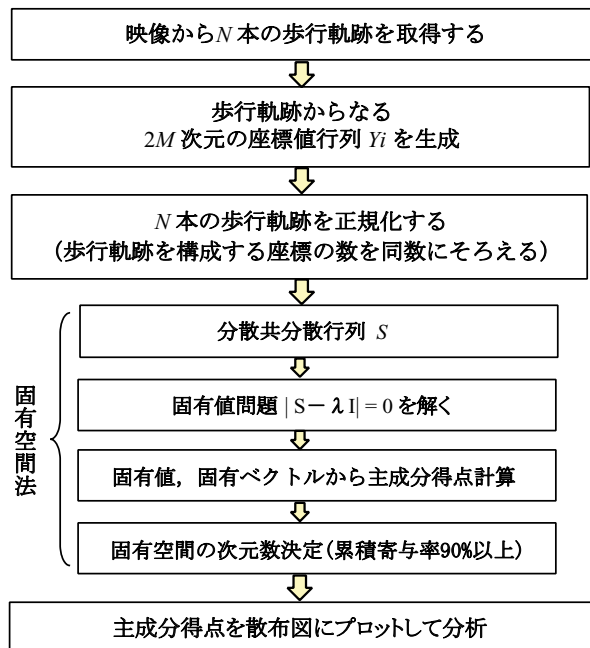


図13 歩行軌跡の学習と固有空間法の処理フロー

特集：信号通信技術

実験には、改札出口付近を俯角40度で捉えた映像を用いた。図14に示すように、通常のシーンとして改札へ向かう人及び改札から出てくる人を、普通でないシーンとして「徘徊」を撮影した。65人分の歩行軌跡をプロットした結果を図15に示す。ほぼ一定の速さで直線的な軌跡を描く大部分の通行人の中で、徘徊者の複雑な軌跡が観測されている。固有空間法による分析を行い、65人分の主成分得点と累積寄与率を計算した(表4)。累積寄与率が約90%となる第2主成分までで主成分得点をプロットした結果を図16に、第3主成分までをプロットした結果を図17にそれぞれ示す。改札へ向かう画面奥と手前の歩行者、さらに改札から出てくる歩行者の大体3つのグループに分類されているのがわかる。徘徊者のプロット点とは明らかに距離が離れていることから、識別の可能性が示された。



図14 通常の動きと普通でない(徘徊)動き

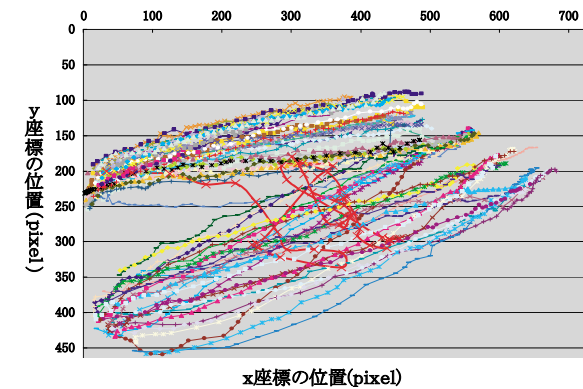


図15 歩行軌跡データのプロット図

表4 固有空間法による分析結果

	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分
主成分得点					
人物1	-5.322539	-9.479302	3.28128337	1.06921355	0.40632213
人物2	-3.2625614	-9.5494614	3.46100734	1.31183663	-0.0492905
人物3	-6.671356	1.8962467	4.71558938	-0.8191095	-0.0582618
人物4	-2.9344666	-10.62476	2.94899058	2.3005265	0.55419123
人物5	-2.3271841	-10.178311	-0.4461162	3.05376597	0.45001439
人物6	-1.8662368	-13.44378	1.26724011	3.80303377	0.11472197
人物7	-6.6937547	-0.3969362	-0.6631118	-0.5565511	-0.089031
人物8	-6.8887851	0.63402398	-0.183395	-0.767549	-0.0581471
⋮					
人物65(徘徊)	4.47278771	-2.5209729	0.78960382	1.04003522	-2.9429405
寄与率	54.6679	35.3066	6.3718	2.2733	0.5892
累積寄与率	54.6679	89.9746	96.3464	98.6198	99.2090

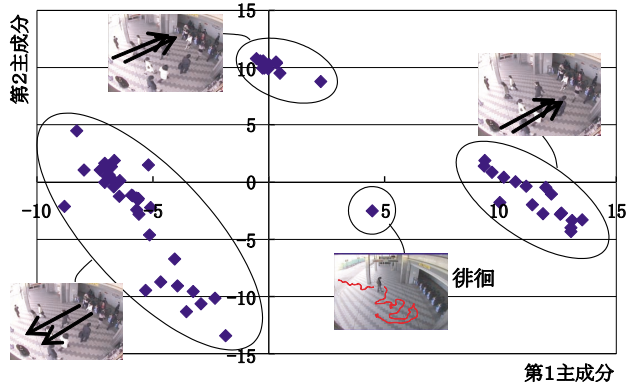


図16 主成分得点の散布図(2次元)による異常の識別

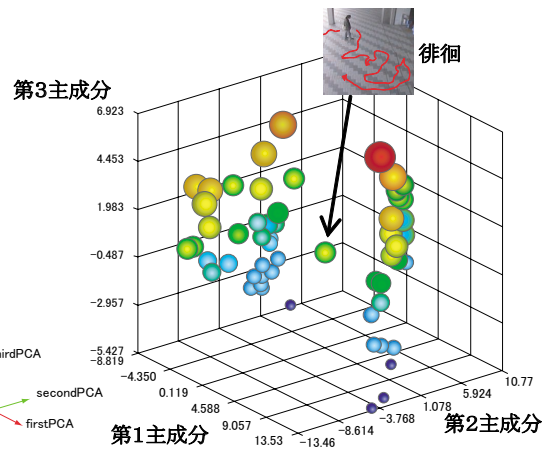


図17 主成分得点の散布図(3次元)による異常の識別

7. まとめ

画像認識による異常検知システムの要素技術として、物の隠れに強いカルマンフィルタを適用した追跡アルゴリズムを開発し、高さ4m、俯角40度程度の標準的なカメラ設置条件の下で、同時に10人程度を約85%の精度で識別できることを確認した。さらに物の置き去り、転倒、暴力等の挙動を解析し、不自然な人の姿勢や手足の動きなどを捉えることにより、これらの異常事象を検知する画像処理手法、さらに歩行軌跡に対して主成分分析を行い、固有空間への投影により異常な挙動を識別するアルゴリズムを検討し、検証実験により有効性を示した。なお、本研究は国土交通省の補助金を受けて実施した。

文献

- 1) Y.A. Ivanov.: "Recognition of visual activities and interactions by stochastic parsing", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):852-872, August 2000.
- 2) N. Oliver.: "A Bayesian Computer Vision System for Modeling Human Interactions", Proc. of ICVS'99, Gran Canaria, Spain, Jan 1999.
- 3) 篠村祐司他: 固有空間による人物の歩行軌跡の予測と零空間を用いた予測修正, 電子情報通信学会技法, pp.25-30, 2007