

# 映像監視システムにおける

## 群相互作用の検出及び解析による異常の自動検出

研究代表者 陳 凡 (CHEN Fan) 北陸先端科学技術大学院大学・情報科学研究科 助教

### 1. はじめに

異常事件の自動検出が知的映像監視システムの実現に欠かせない基盤技術の一つである。特に大規模な映像監視システムが普及している現在、数多くのカメラから撮影されたデータの量が爆発的に増大した。異常の自動検出が操作員の作業負担を低減し、システムの運営費用を抑え、予測できない危険な事件を速やかに対応するなど多くの面で役に立っている。

一般的な分類問題に比べ、異常検出には幾つかの難点が含まれている：

a) 異常事件の予測性が低いため、検出すべき異常事件の事前定義が容易でない。異常事件の事前定義を回避するためには、教師なし学習による異常検出が考えられる。代表的な手法として、クラスタリング、密度推定、単クラスサポートベクトルマシンなどが挙げられる；

b) 目標異常事件を定義できても、発生頻度が低い。その結果、十分な学習サンプルを集まって教師あり分類器を訓練する事が困難である。学習サンプルを十分に集まるためには、事件の発展と共にオンラインで異常事件に関する知識を累積することが重要である。それを実現するためには、検出手法の調整上の自由度及び計算効率上の実時間性が要求される；

c) 正常サンプルが異常サンプルより集まりやすいため、クラス間の非平衡が起きやすく、検出手法と検出結果の両方に大きな影響を与える。データの再サンプリングや重み付けなどの手法により、クラス間の非平衡を解消することが考えられる。本研究には、特徴抽出の段階において軌跡の抽出や背景の除去を行い、容易に判別できる正常サンプルの大部を事前に除去することにより、クラス間の非平衡問題を低減しようとする。

異常検出の一般的流れには、視覚特徴の抽出、相互関係の表現、及び異常度の評価が含まれる：

a) 前景マスク、テクスチャー[1]、運動ベクトル[2]など低レベルの視覚特徴から軌跡[3]、物体輪郭など高レベルの視覚特徴まで、異常検出には様々な特徴が使われる。異常検出の性能を大きく左右する重要な要素である一方、特徴の抽出及び選択が実用環境に強く依存するため、応用環境を分析して決められる必要がある。

b) 孤立した特徴点に加え、特徴点間の群体的相互作用が共存 (Co-occurrence) [4], 時空間依存性[5], 特徴集団[6]などにより記述され、物体間の空間上、時間上及び動力学上の相互関係を表している。相互関係を特徴に取り入れることにより、語義上でより複雑な事件を記述することができる。特に厳密に規制されている実用環境 (例、交通流管理[7]) において、相互関係により複雑な運動パターンを抽出できる。相互関係の記述には文脈上でより広い近傍範囲が必要とされるため、検出結果の分解能が低下されることがある。

c) 異常度を評価する際に、様々な基準が考えられる。異常の発生頻度が低いことにより、確率密度に基づき事件の異常度を定義できる。特徴集団の異常度が集団内特徴ペア間の相互関係を累積することにより算出できる。クラスタリングの場合には、クラスタの中心までの距離により異常度を算出できる。特徴空間において、異常サンプルの周辺にある空白領域がより広いため、各特徴点を切り出す容易度により異常度を定義する手法も挙げられる[8]。我々も主に異常事件の確率密度に基づき、異常度を定義する。

本研究には、軌跡情報を使用しない手法と軌跡情報を用いた手法を検討した。小範囲の監視システムには、運動物体を高精度で追跡できる。追跡された物体の運動軌跡情報に含まれた群相互作用の子集団を解析するこ

とにより、異常事件を検出する手法が提案された[9], [10]。広範囲で混雑した監視環境には物体を精度良く追跡することが困難である。テクスチャーや運動ベクトルなど低レベルの視覚情報に基づき、密度推定などにより異常度を判断する手法が提案された[11]。本文には、教師なし学習及びオンライン更新可能な検出手法を中心に、我々が提案した Random Projection Forest を用いた異常検出手法を主に紹介する。最後に、知的映像監視システムを応用例とし、検出された異常情報を用いた映像内容の自動要約も簡単に紹介したい。

## 2. Random Projection Forest を用いた異常検出手法

大範囲で混雑する環境における異常検出には安定した軌跡検出が困難であるため、テクスチャーや運動特徴を用いて異常検出を検討する。オンライン更新可能な教師なし検出アルゴリズムの提案に集中、テスト用の公開データセットに最適な視覚特徴及び相互関係をヒューリスティック的に決める。事前に明示的に定義されている目標異常事件のリストを使わず、事件の希少性により異常度を定義することが自然である。確率密度が主に正常値に基づき算出されるため、正常サンプルのみを用いた学習枠組みに適しており、クラス間の非平衡もより容易に解消できる。それゆえ、確率密度推定に基づく検出手法を中心に検討する。

ガウス混合モデルなどモデルベースの確率密度推定手法には最適なモデルの選択が大切である。モデルがデータより低い自由度を持つ場合には、データの分布を示す十分な分解能が得られず、異常と正常サンプルを精度よく区別できない可能性がある。データよりモデルがはるかに高い自由度を持つ場合には、過剰適合が起きてしまう。異常値が正しくフィットされ、誤判別されやすい。

カーネル密度推定にはモデルの選択が要らないが、新たなサンプルの密度を推定するのに学習済みの訓練サンプルが全て必要とされる。それゆえ、密度推定に必要な計算負担が大きい。

Random Forest は教師あり分類問題によく使われる手法である。従来研究に提案されている Random Forest に基づく密度推定手法には、分配関数を容易に計算するため、座標軸に平行する投影方向を用いた[12]。ここではランダムで選ばれた任意な投影方向を用いた Random Projection Forest[13]に基づき、密度推定を考える。異常検出に Random Projection Forest を適用することに幾つかの理由がある：

- 1) 階層的クラスタリングによりデータを木構造に組織化できる。木構造の特性により、Random Forest が特徴空間において入力サンプルの近傍範囲を効率的に特定することができ、計算効率を著しく改善できる。さらに、複数の決定木がそれぞれ独立して計算されるので、処理の並列化を実現しやすい；
- 2) 決定木ごとで過度なクラスタを行い、各投影方向で入力サンプルの顕著度を十分に評価する。投影方向選択上のランダム性と複数決定木の結果統合により、過度なクラスタによる過学習を抑える。結果として、モデルの選択問題を回避し、安定した密度推定が得られる；
- 3) Random Projection Forest が探索的なデータ解析を行い、様々な投影方向でのデータの異常度を繰り返し返してチェックする。異常性を示した投影方向を強調することで、異常検出の精度を改善できる；
- 4) 決定木に基づく探索的なデータ解析には、あるサンプルが異常と判別される原因を解釈しやすい；
- 5) Random Projection Forest には、節点上のパラメータ値の更新と木構造の修正などを含む自由度の高いモデル調整が可能である。知識のオンライン更新を実現しやすく、変化し続ける実環境に対応できる。

密度推定のほかには、Random Forest を用いた異常検出手法として、Isolation Forest が挙げられる[8]。特徴空間を区切りする際に、異常値周辺の空白がより広いである。Isolation Forest には、異常値が最も分離しやすいサンプルとして検出される。分離の容易さが根節点から葉節点まで訪れた経路の平均的長さにより評価される。正常値の代わりに、異常値に着目してモデルを構築することが Isolation Forest の特色である。異常検出には、訓練サンプルに正常サンプルしか含まれない学習枠組みを取り扱うことが多い。データの大部を占める正常サンプルの特性を直接的モデリングできる手法を望み、密度推定手法を考える。

上述の考えに加え、本研究の提案手法には主に以下の貢献が挙げられる：a) Random Projection Forest を用いた異常検出手法の基本枠組みを提案した；異常値周辺の空白領域の最大化により、異常検出に有効な節点分割基準を定義した；射影追跡法の考えを用いて、低い密度を示した特定な投影方向を強調できる密度推定式を定義した；異常値の判別には、低い確率密度を得た決定木を強調できる幾何平均による結果統合を用いた。

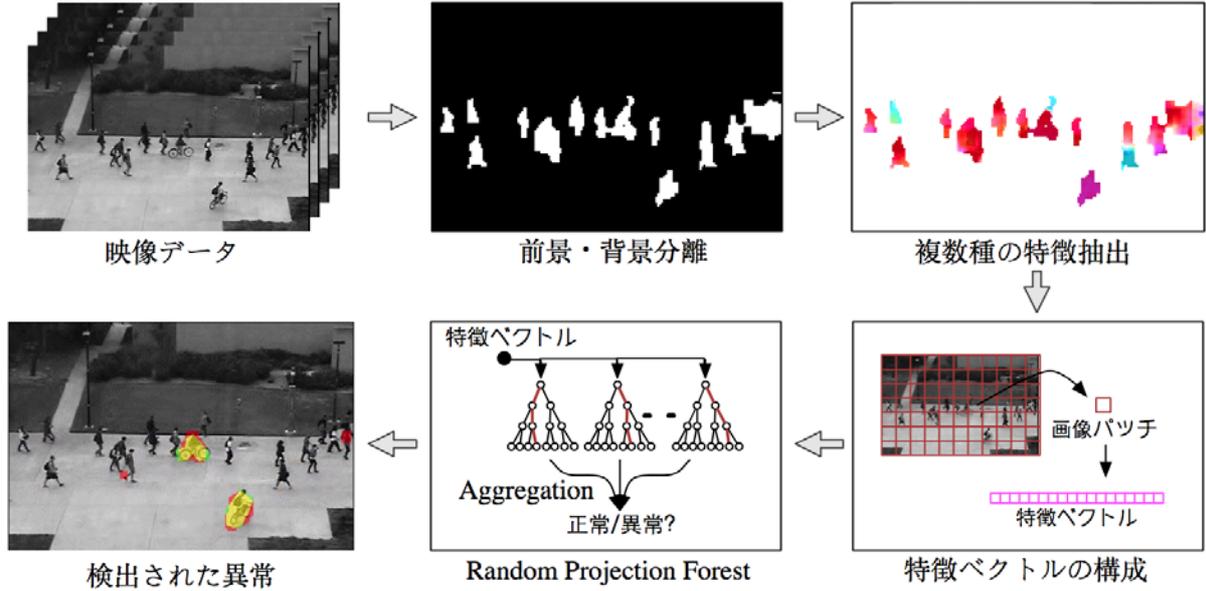


図 1 Random Projection Forest を用いた異常検出処理の流れ

映像監視システムの一般状況として、異常事件のほとんどがカメラの画面内で変化のある部分に関わっている。それゆえ、変化のない背景情報を事前に排除することが安全である。さらに、明らかに正常であるサンプルの大部を除去することにより、クラス間の非平衡問題を低減できる。カメラ画面の上に二次元の格子を置き、密なサンプリングを行う。各サンプリング点において、当該パッチの視覚特徴を表す一つのベクトルが計算される。算出された特徴ベクトルを Random Projection Forest に入力し、異常度を評価する。異常度の高い特徴点が異常点として分類され、出力される。図 1 には異常検出処理の全体的な流れが示されている。

## 2.1 Random Projection Tree を用いた異常検出のための節点分割基準

決定木には階層的なクラスタリングが行われる。サンプルがまず決定木の根節点に入力される。途中に到達した各非葉節点において、ある分割基準によりデータを二つのクラスタに分類する。分割後のデータがそれぞれ左右子節点に入る。葉節点に到着するまで前述の分制作業が繰り返される。最終的に、複雑な確率分布を持つデータが幾つかのクラスタに分割され、各クラスタがより簡単な確率分布で表せる。

Random Projection Tree には、投影方向  $\mathbf{w}$  がランダムに選ばれる。データセット  $\mathbf{S}$  にあるサンプル  $\mathbf{x}_i$  が投影方向  $\mathbf{w}$  に投影され、閾値  $\theta$  と比較する。比較結果によりデータが二つの子集合の一つに分類される。二つの子集合がそれぞれ左右子節点に対応し、 $\mathbf{S}^L$  と  $\mathbf{S}^R$  で表す。

決定木によるクラスタリングの目的は、モデリングされた確率分布  $q(\mathbf{w}^T \mathbf{x}_i)$  を投影データの経験分布関数  $p(\mathbf{w}^T \mathbf{x}_i)$  にできるだけ近づくことである。ここでは、カルバック・ライブラー距離により、モデル  $q(\mathbf{w}^T \mathbf{x}_i)$  で計算された確率分布と経験分布関数  $p(\mathbf{w}^T \mathbf{x}_i)$  との差異を評価する。

$$\mathcal{KL}(p, q|\mathbf{S}, \mathbf{w}) = \sum_{\mathbf{x}_i \in \mathbf{S}} p(\mathbf{w}^T \mathbf{x}_i) \log \frac{p(\mathbf{w}^T \mathbf{x}_i)}{q(\mathbf{w}^T \mathbf{x}_i)}. \quad (1)$$

確率モデルがデータをより正確にフィットできるように、節点上の二分割を繰り返して行う。分割前後の差異を最大化することにより、各節点で最適な投影方向と分割閾値を選出する。

$$(\mathbf{w}^*, \theta^*) = \arg \max_{\mathbf{w}, \theta} \mathcal{D}, \quad (2)$$

$$\mathcal{D} = \mathcal{KL}(p, q|\mathbf{S}, \mathbf{w}) - \mathcal{KL}(p, q_\theta^{LR}|\mathbf{S}, \mathbf{w}), \quad (3)$$

$q_\theta^{LR}(\mathbf{w}^T \mathbf{x}_i)$  が分割後の確率モデルを表す。その値が所属の子節点により定義される。

$$q_{\theta}^{LR}(\mathbf{w}^T \mathbf{x}_i) = \begin{cases} q_{\theta}^L(\mathbf{w}^T \mathbf{x}_i), & \mathbf{x}_i \in \mathbf{S}^L \\ q_{\theta}^R(\mathbf{w}^T \mathbf{x}_i), & \mathbf{x}_i \in \mathbf{S}^R \end{cases} \quad (4)$$

投影データの分割方式が確率分布の特性により決められる（図 2）。

a) 投影データの確率分布に主要なクラスタ成分が二つ以上ある場合には、分割後の投影データをガウス成分  $q_{\theta}^L(\mathbf{w}^T \mathbf{x}_i) \sim \mathcal{N}(\mathbf{w}^T \mathbf{x}_i; u^L, \sigma^L)$  と  $q_{\theta}^R(\mathbf{w}^T \mathbf{x}_i) \sim \mathcal{N}(\mathbf{w}^T \mathbf{x}_i; u^R, \sigma^R)$  のいずれにより記述できるとする。

$$\mathbf{x}_i \in \begin{cases} \mathbf{S}^L, & \mathbf{w}^T \mathbf{x}_i \leq \theta \\ \mathbf{S}^R, & \mathbf{w}^T \mathbf{x}_i > \theta \end{cases} \quad (5)$$

$u, \sigma, u^L, u^R, \sigma^L, \sigma^R$  がモデルのハイパーパラメータであり、最大尤度推定により簡単に求められる。

b) 投影データの確率分布に主要なクラスタ成分が一つしか存在しない場合には、主要な成分を切り出した。下式のように、全体の期待値までの距離を計算し、投影データを二つのクラスタに分割する。

$$\mathbf{x}_i \in \begin{cases} \mathbf{S}^L, & |\mathbf{w}^T \mathbf{x}_i - u| \leq \theta \\ \mathbf{S}^R, & |\mathbf{w}^T \mathbf{x}_i - u| > \theta \end{cases} \quad (6)$$

主要なクラスタ成分に分類されるデータがガウス成分  $q_{\theta}^L(\mathbf{w}^T \mathbf{x}_i) \sim \mathcal{N}(\mathbf{w}^T \mathbf{x}_i; u^L, \sigma^L)$  により記述される。それ以外のデータがガウス成分  $q_{\theta}^R(\mathbf{w}^T \mathbf{x}_i) \sim \mathcal{N}(|\mathbf{w}^T \mathbf{x}_i - u|; u^R, \sigma^R)$  により表れる。

式 5 と式 6 を式 1 に代入し、式 3 に定義された分割基準が下記のように求められる。

$$\begin{aligned} D &= \mathcal{KL}(p, q|\mathbf{S}, \mathbf{w}) - \mathcal{KL}(p, q_{\theta}^{LR}|\mathbf{S}, \mathbf{w}) \\ &= \log \sigma - \frac{|\mathbf{S}^L|}{|\mathbf{S}|} \log \sigma^L - \frac{|\mathbf{S}^R|}{|\mathbf{S}|} \log \sigma^R + \text{const}. \end{aligned} \quad (7)$$

$|\mathbf{S}|$  が集合  $\mathbf{S}$  に所属するサンプルの数を計算する。標準偏差が期待値の周囲に確率変数のばらつきを表している。式 7 には、分割基準  $D$  の最大化によりデータを最も圧縮できる投影方向と分割閾値が求められることが分かる。クラスタ内の距離を最小化することにより、クラスタ成分間の境界区域を最大化できる。境界区域に散布される異常値がより検出されやすい。これが Random Projection Tree を異常検出に適用する主要な動機である。

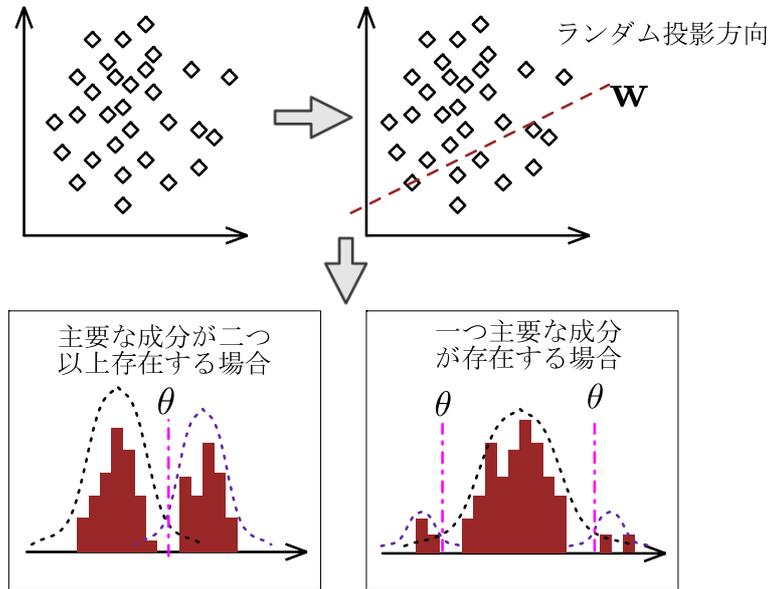


図 2 Random Projection Forest における節点の分割基準

## 2.2 Random Projection Forest を用いた確率密度評価

各葉節点における局所的確率密度がクラスタを表すガウス成分により記述される。サンプル全体における確率密度を計算するため、クラスタごとでの局所密度を全特徴空間で正規化する分配関数を計算すべき。Random forest による密度推定問題には、葉節点ごとでの分配関数の計算が大きな難点である。従来手法 [9]

には、ランダムな投影方向が座標軸に平行する方向に限定され、積分により分配関数が計算された。それに対し、任意な投影方向を考える場合には密度関数の積分が大変複雑であるため、分配関数の計算が容易でない。

繰り返してサンプルの近傍を狭くことで、Random Projection Tree が様々な投影方向でデータの確率分布を調べ、探索的にデータのモデリングを精緻化する。直接的な積分の代わりに、我々は探索的射影追跡による確率密度推定の考え[14]を用いて、入力サンプルの確率密度を計算する。

具体的には、Random Forest を構築する際に、成員となる  $L$ 本の決定木をすべて  $M$ 階層まで完全に成長させる。サンプル  $\mathbf{x}_i$  が与えられた時、 $l$ 本目の決定木の根節点に入力される。葉節点まで階層的にクラスタリングされ、途中で訪問済みのノードの集合を  $\mathbf{v}_{il} = \{v_{ilm}\}$  にまとめる。下式のように、 $l$ 本目の決定木により評価されたサンプルの確率密度  $p_l(\mathbf{x}_i)$  が近似的に計算される。

$$p_l(\mathbf{x}_i) \propto \mathcal{N}(\mathbf{x}_i; u_{v_{ilM}}, \Sigma_{v_{ilM}}) \prod_{m=1}^{M-1} p_{v_{ilm}}(\mathbf{w}^T \mathbf{x}_i) \quad (8)$$

ここでは、同一葉節点に到達するサンプルが構成した局所クラスタをガウス分布  $\mathcal{N}(\mathbf{x}_i; u_{v_{ilM}}, \Sigma_{v_{ilM}})$  により表れる。分割前での広い近傍におけるサンプル点の重要度が  $p_{v_{ilm}}(\mathbf{w}^T \mathbf{x}_i)$  により評価される(図 3)。各決定木において、異なる投影方向上の密度を累積し、全特徴空間での密度を独立して評価する。

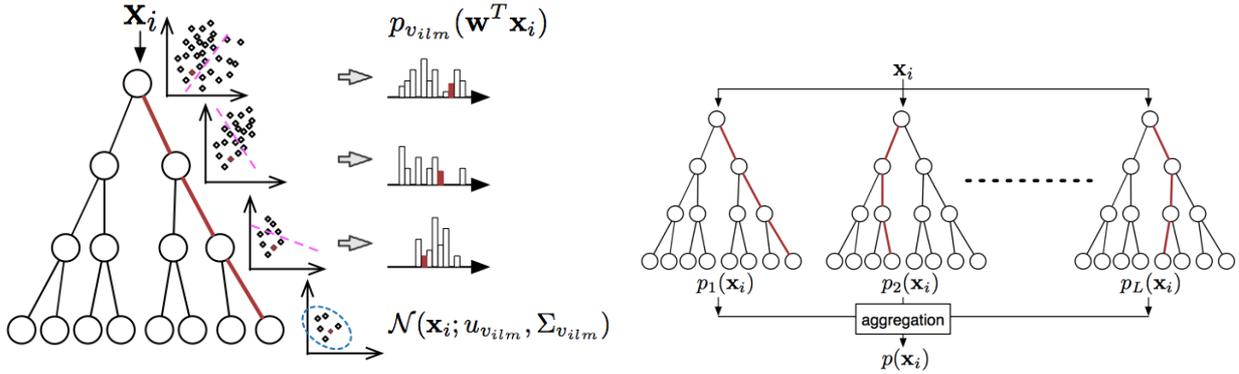


図 3 Random Projection Tree における確率密度の計算

一つの決定木が一つのクラスタリング方法を与えた。データ適合の精度を保証するためには、実際の確率分布に含まれるクラスタ数よりも多くの葉節点を用意することが一般的である。単決定木で起きる過剰適合を抑えるため、Random forest には複数の決定木の評価結果を統合し、全特徴空間においてサンプル点の密度を算出する。通常の密度推定問題には、外れ値に対する頑健性が重視され、複数の決定木による平均効果を求めている。異常値がすべての投影方向に外れ値として現れる必要がないため、異常検出のための密度推定には異常値の抑制する過度な平均を避けるべきである。結果統合する際に、低い密度を出力した決定木をより強調するため、我々は算術平均の代わりに幾何平均を使用する。

$$p(\mathbf{x}_i) = \sqrt[L]{\prod_{l=1}^L p_l(\mathbf{x}_i)} \quad (9)$$

合成されたデータセットを用いて、提案された Random Forest による密度推定を検証した。その結果が図 4 に示されている。正弦曲線型の分布をしているデータの推定結果が図 4 (上部) に示されており、二つのガウス成分を持つガウス混合モデルに従うデータの推定結果はその下部に示された。Random Forest が複雑な分布に従うデータの密度も正確に推定できることが分かった。モデルの自動選択が密度推定に重要な問題である。データよりモデルの自由度が低くなると十分な精度が得られず、自由度が高すぎると過学習が起きやすい。決定木ごとに十分な階層数を取ることで精度を保証し、複数の決定木の結果統合により過学習を抑えた。結果と

して、明示的なモデル選択をせずに、高精度な密度推定を実現した。従来のカーネル密度推定には、密度を評価する際にすべての訓練データが使われます。それに比べ、Random Forest には木構造によるデータの階層的に再組織化を行い、近傍の訓練サンプルを効率よく検索できたため、全体の計算量を著しく減らした。

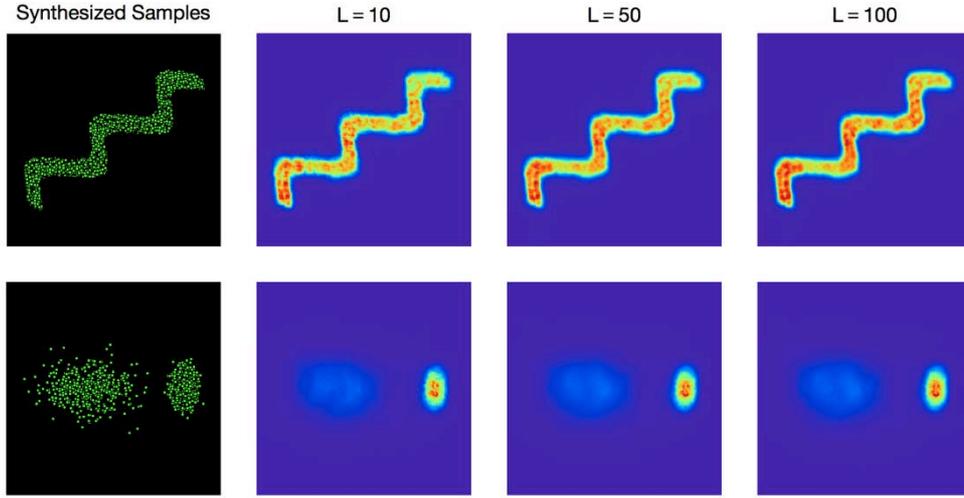


图 4 Random Projection Forest による確率密度計算の実例

### 2.3 特徴抽出

異常事件の予測性が低いであるため、正常な事件に対応する学習データのみを用いて検出器を訓練する場が多い。しかしながら、教師なし異常検出には、学習により特徴の自動選択が困難である。Random Projection Forest による異常検出の振る舞い及び性能の検証に集中するため、ここではテストされた公開データセットにとって最適な特徴を実験によって選択して決定する。明示的に、本研究には四種類の特徴が含まれている：

- 1) 画像パッチ内の前景画素の割合。背景・前景画像の分離にはガウス混合モデルに基づく手法を用いた[15]。この特徴は主に物体の大きさを表すため導入されたものである；
- 2) 画像パッチ内運動物体の平均的運動ベクトル。運動ベクトルの計算にはオプティカル・フロー推定のための多階層 Lucas-Kanade アルゴリズムが使われた[16]。これは物体の運動特徴を表している；
- 3) テクスチャー特徴が画像パッチ内異なる方向を持つ複数のガボールフィルタの応答により抽出される；
- 4) 異常事件の場所依存性を表すため、画像パッチの中心画素の XY 座標値も特徴ベクトルに取り込まれた。座標値が  $[0, 1]$  範囲内に正規化される。

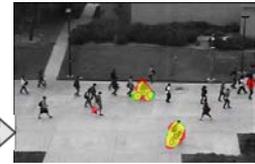
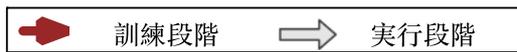
カメラ画面に対しグリッドサンプリングを行う。分割セルの大きさが  $N \times N$  画素とする。  $t$  番目のフレームにある  $i$  番目のセルにおいて、9次元の特徴ベクトル  $\mathbf{x}_{ti} = [p_{ti}^X, p_{ti}^Y, r_{ti}^{FG}, m_{ti}^X, m_{ti}^Y, g_{ti}^{0^\circ}, g_{ti}^{45^\circ}, g_{ti}^{90^\circ}, g_{ti}^{135^\circ}]$  が算出される。  $p_{ti}^X$  と  $p_{ti}^Y$  が画素の座標値であり、  $r_{ti}^{FG}$  が前景画素の割合を表す。  $m_{ti}^X$  と  $m_{ti}^Y$  がパッチ内の平均運動ベクトルであり、  $g_{ti}^{0^\circ}, g_{ti}^{45^\circ}, g_{ti}^{90^\circ}, g_{ti}^{135^\circ}$  が四方向のガボールフィルタの平均応答を表している。

### 2.4 異常検出

提案手法が訓練段階と実行段階との二段階に分けられる (図 5)：訓練段階には、各時間  $t$  において、決定木の部分集合をランダムで選出する。現フレームの各サンプリング位置から特徴ベクトルを算出し、節点の成長及びパラメーターの調整により選ばれた決定木を学習させる；実行段階には、抽出された特徴ベクトルを各決定木に入力する。すべての決定木の出力を統合した密度値を閾値  $\phi$  と比較し、異常判別を行う。

$$a(\mathbf{x}_{ti}) = \begin{cases} 1, & p(\mathbf{x}_{ti}) \leq \phi \\ 0, & p(\mathbf{x}_{ti}) > \phi \end{cases} \quad (10)$$

最後に、得られた異常値マップ  $a(\mathbf{x}_{ti})$  に簡単な平滑化処理をかけ、孤立した異常点を除去する。



別 検出結果

図 5 Random Projection Forest による異常検出の枠組み

## 2.5 実験結果

本研究には、混雑した環境で歩行者を扱った公開データセット UCSD Pedestrian dataset[17]を用いて、提案手法の性能を検証した。UCSD Pedestrian Dataset にはさらに二つのデータセットが含まれている：UCSD Ped 1 には、画面サイズが  $238 \times 158$ 画素となり、正常事件のみが含まれる学習用のビデオクリップ 34 点と正常・異常事件が混在したテスト用のビデオクリップ 36 点が提供されている；UCSD Ped 2 には、画面サイズが  $360 \times 240$ 画素となり、正常事件のみが含まれる学習用のビデオクリップ 16 点と正常・異常事件が混在したテスト用のビデオクリップ 12 点が提供されている。実験中には、 $L = 96$ 本の決定木を持つ Random Projection Forest を使い、最大階層数が  $M = 8$ とする。

従来研究[1][2]のように、異常の検出と異常の位置特定の両方から性能を評価するため、我々がフレームレベルと画素レベルの両方で検出精度を定義した：

- ・フレームレベルの評価基準には、異常画素のあるフレームを ポジティブ 検出とする。それを手標記で得られた正解値と比較し、正検出と誤検出のフレームの割合を算出する。

- ・画素レベルの評価基準には、異常画素の正解値を用いて、正検出と誤検出のフレームの割合を計算する。ここでは、正検出のフレームとして認定されるためには、二つの条件を満たすべき：正解値として当該フレームが異常であること；40%異常の正解異常画素が検出されること。それに対し、誤検出のフレームは、正解値として当該フレームが正常であり、検出結果には異常画素が含まれることから定義される。

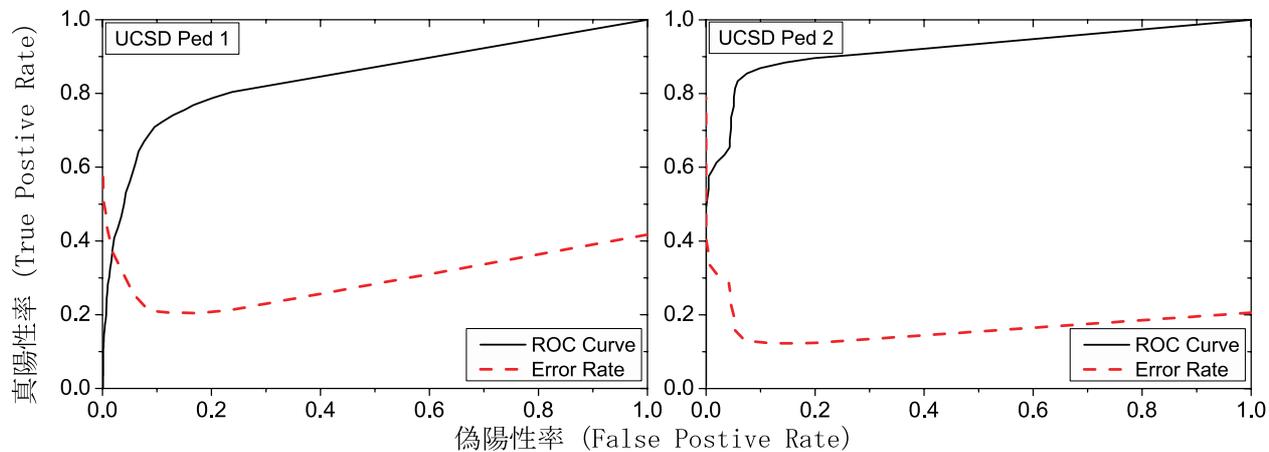
真陽性率 (TPR) と偽陽性率 (FPR) に加え、エラーレット (ER) が誤分類フレームの割合として定義され、性能評価に使われた。FPR に対して、フレームレベルと画素レベルの TPR 及び ER の結果が図 6 に示されている。表 1 には、フレームレベルの評価に平等エラーレット (EER,  $FPR = 1 - TPR$ の時の ER 値として定義される) を、画素レベルの評価に検出率 (RD) ( $= 1 - EER$ ) が示された。同データセットにおいて、従来手法に比べ、提案手法がより正確に異常を検出できることが分かった。

表 1 平等エラーレット (EER) と検出率 (RD) による結果比較

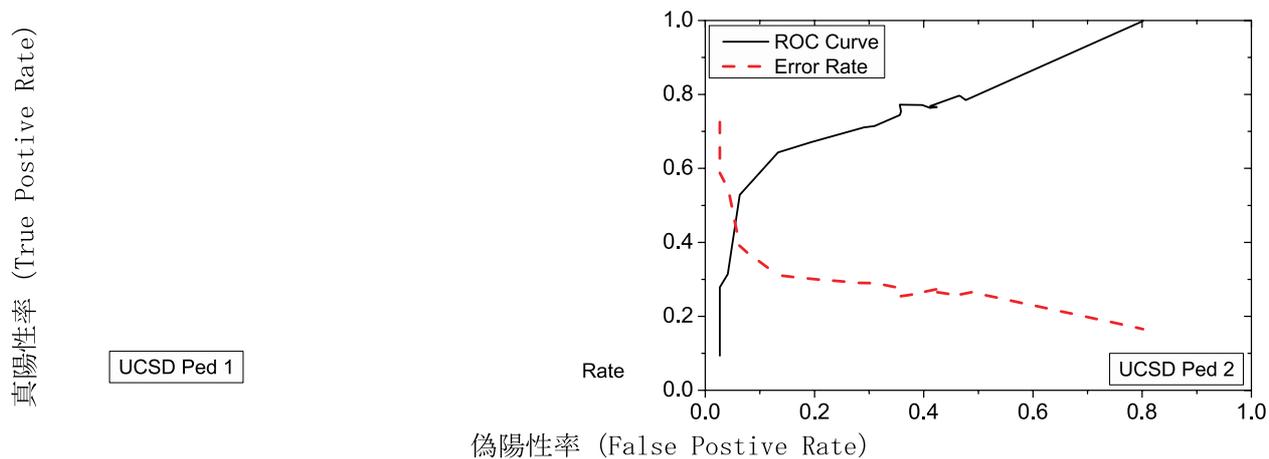
		%	508%	634%
MPPCA †	35.6%	35.8%	23.2%	22.4%
force flow †	36.5%	35.0%	40.9%	27.6%
LMH †	38.9%	45.8%	32.6%	22.4%

正解値画素の40%以上を検出結果に含まれば、誤検出された画素がいくらあっても、入力動画フレームが正検出と認定されることになる。これが UCSD Ped2 には偽陽性率の増加と共に、画素レベルのエラーレットが減少し続ける理由と考えられる。それゆえ、画素レベルの検出精度を評価するのに、従来手法によく使われた前述の評価基準が最適な基準でないことが分かった。より正確に画素レベルの検出精度を評価するため、

我々が再現率 (Recalling Rate) と選別精度 (Precision) を定義した。全正解値画素のうち正しく異常画素と検出された画素の割合が再現率と呼ばれ、異常と判定された画素のうち正解値と一致する画素の割合を選別精度と呼ばれる。再現率 (Recalling Rate) と選別精度 (Precision) の結果が図 7 に示されている。UCSD Ped1 データセットでは、74%の再現率を得た時、約半分の誤検出 (選別精度が48%) が含まれている。UCSD Ped2 データセットでは、73%の再現率を得た時、3割以下の誤検出 (選別精度が74%) が含まれている。



a) フレームレベルの精度



b) 画素レベルの精度

図 6. Random Projection Forest による異常検出の精度

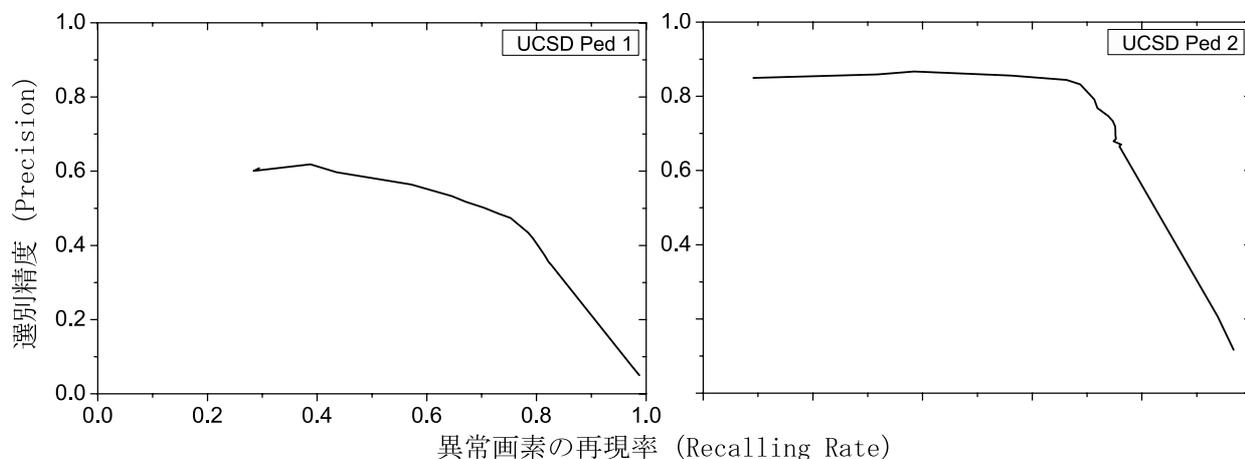


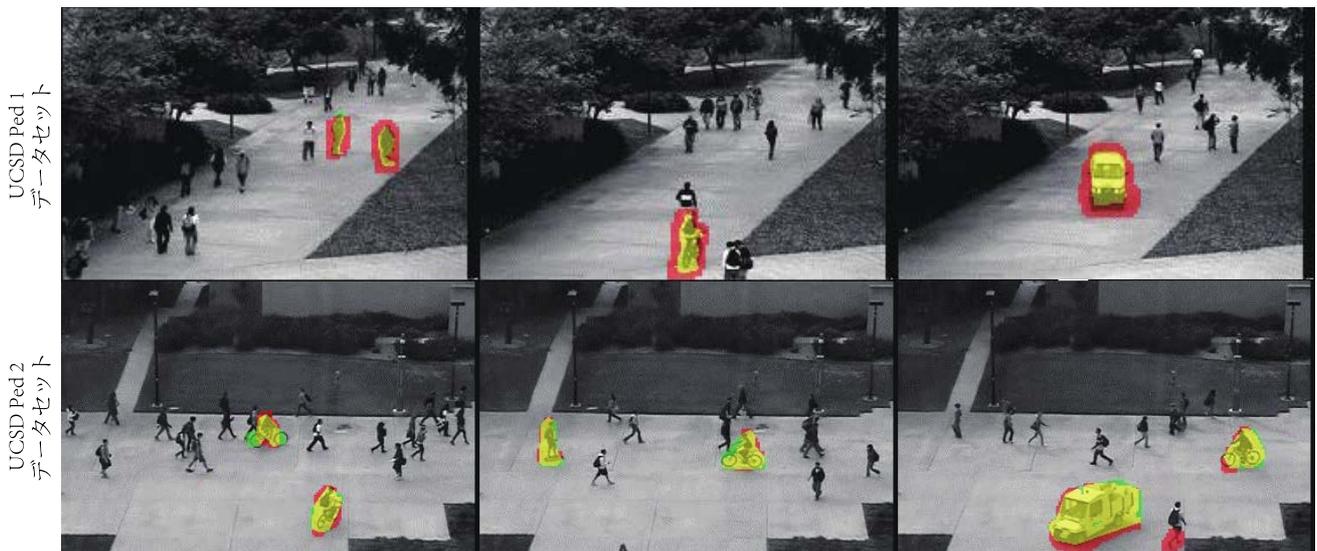
図 7. 再現率と選別精度による画素レベルの精度評価の結果

表2には計算時間が示されている。異常検出のフレームレートにより、特徴抽出を含むシステム全体の処理時間と事前に算出された特徴を用いた異常検出のみの処理時間を評価される。異常検出自体が既に実時間で計算できる。全システムを実時間化するためには、特徴抽出の部分がさらに高速化する必要がある。

表2 提案手法の計算時間 (FPS)

) 29.9 80.6 26.6 53.1

検出の結果例が図8に示されている。スケート、自転車、車など物体が正しく異常として検出された。従来手法[1]に示されている検出結果と比較すれば、画素レベルの検出精度がかなり良くなることが分かった。提案手法による検出結果のビデオが本研究の添付資料[18]に掲載されている。



\* 黄色: 正しく検出された異常画素; 赤色: 偽陽性画素; 緑色: 検出されなかった正解異常画素

図8. Random Projection Forest による異常検出の結果例

提案手法には下記の理由で性能改善が得られたと考えられる：

- 1) 様々な組み合わせから、UCSD Dataset に対して最適な特徴集合と重み付けが実験により決められた。
- 2) 前景画像の事前抽出により、正常サイクルの大部を除外し、クラス間の不均衡を著しく低減した。
- 3) 位置情報を明示的に特徴に取り入れることが場所に依存する異常事件と場所に依存しない異常事件の両方の検出に有効である。テスト用のデータセットにおいても、高速で移動する物体を検出するのに、位置依存性のない異常事件を検出すべき。立ち入れ禁止の領域に入る物体を検出するため、位置依存性のある異常事件にも対応すべき。従来手法にはサンプリングのセルごとに独立した検出器を訓練する手法が挙げられる。それに比べ、提案手法には近傍のセルからの学習データも利用でき、学習用のサンプルの数を増やしてより頑健な統計学習をできる。特に UCSD Ped2 のカメラ画面の下部には、学習として使われるサンプルがほとんどないため、位置情報を特徴に取り入れることがより重要である。
- 4) 異常検出に使われる特徴がサンプリングの各セルから独立して抽出される。UCSD のデータセットに含まれる異常事件のほとんどが歩行者、自転車、自動車など孤立した物体であり、物体間の相関が低い。実験結果として、特徴を構成する際に、時間軸上でより広い近傍範囲を取ることが有効だが、空間上で広い近傍を取ると総合的検出精度を落ちることが見られた。
- 5) Random Projection Tree の分割基準が正常サンプルのクラス間距離を最大化し、外れ値となる異常事件の検出に有効である。異常事件が一部の投影方向に外れ値として現れる場合が多い。それらの投影方向

を強調するため、階層クラスタリングのような探索的なデータ解析が有効である。異常と判断される理由も解釈しやすい。

6) 異常事件が全部の投影方向及び決定木において外れ値として現す必要がないため、算術平均による結果統合には過剰な平滑化が望ましくない。提案手法には、射影追跡のような密度推定(式8)及び幾何平均による結果統合(式9)には、低い確率密度のある投影方向及び決定木が強調され、異常をより高精度で検出できる。

上述の原因に加え、Random Projection Forest が階層的クラスタリング手法であるため、データの近傍を効率的にアクセスできる。集団学習枠組として計算が並列化処理しやすい。節点レベルと木レベルの更新など柔軟な調整方法を使用でき、大規模なデータセットに対応できる優れた拡張性を持っている。それゆえ、Random Projection Forest が大量な映像データや変動する環境における異常検出に適している。

### 3. 異常検出の監視映像要約への応用

異常検知による警報出力のほか、大規模な監視システムに設置された複数台のカメラから収集した監視映像を異常情報に基づき効率に要約することが、作業員の作業負担の低減に重要である[19]。決められた映像の長さで、検出された異常事件の経過をうまく表現するため、我々が内容の連続性と閲覧上の快適さを重視して内容要約の手法を考えた。特に、映像監視システムには24時間の映像を10分間に要約するなど高圧縮比の要約タスクが多いため、早送りの場合には再生速度が理解できる範囲を超え、ちらつきによる不快感が生じやすい。それを抑えるため、早送りと内容の飛ばしを適応的に切り替える内容要約手法を提案した[20]。

$i$ 番目の画素パッチの異常値 $a(\mathbf{x}_{ti})$ (式10)に基づき、フレーム $t$ に含む情報の重要度 $f_t$ を $f_t = 1/N \sum_i a(\mathbf{x}_{ti})$ より定義できる。 $N$ は画面内サンプリング点の総数である。閲覧しやすい要約内容を生成するのに、より多くの情報を含むのみでは足りず、画面間内容変化の速度も閲覧の快適さを大きく影響を与える。ここでは、フレーム間情報の変化 $a_t$ が $a_t = |f_t - f_{t-1}|$ と定義される。要約結果において、フレーム $t$ が占める再生時間 $s_t$ が $[0, 1]$ 範囲内の実数値を取るとする。 $s_t = 1$ のとき、フレーム $t$ を元の速度で再生する。 $s_t = 0$ のとき、フレーム $t$ が飛ばされる。 $0 < s_t < 1$ の時にはフレーム $t$ を早送りで再生する。下記の基準により最適な要約映像が求められる。

$$\mathbf{V} = \arg \max_{\mathbf{V}} \sum_t [f_t]^\alpha \left[ \frac{a_t}{C} \right]^\alpha s_t^\beta, \quad \text{s.t.} \quad \sum_t s_t = u^L \quad (11)$$

$u^L$ が目標となる要約ビデオの長さであり、 $\alpha \in (0, 1)$ ,  $\beta \in (0, 1)$ が制御用のパラメーターである。要約問題を最適化問題として解釈し、資源配分問題として解決される。文献[20]には要約手法の詳細が書かれており、生成された閲覧用監視映像の性能も一連の主観・客観実験により検証された。

### 4. おわりに

本研究には、映像監視システムにおける異常事件の自動検出が中心課題として取り扱っている。特に、異常事件の低い予測性と発生頻度により、教師なし学習手法の適用及びオンライン更新の可能性を中心に異常事件の検出手法を検討した。本文には、主に我々が提案したRandom Projection Forestによる異常検出手法を紹介した。Random Projection Forestを用いた異常事件の検出手法には、幾つかの利点が挙げられる：クラスタのパラメーターをオンラインで調整できるため、未学習の異常に柔軟に対応できる；階層的木構造を用い、異常度を効率的に計算できる；複数の決定木を統合するため、過学習を抑え、精度よく密度推定が得られる。これらの特性がオンライン異常検出システムの構築に重要である。提案手法の有効性が公開データセットにより検証され、従来研究により高い検出精度が得られた。本文の最後に、異常事件の自動検出の応用例とし、我々が提案した監視映像の自動要約手法も簡単に紹介した。

異常事件の定義には高い主観性があり、実応用環境に強く依存される。異常事件の定義上の曖昧さを解消するため、教師なし学習に加え、作業員から検出の正誤に関するフィードバック情報も検出器の訓練に入れる半教師あり学習が有効である。我々が半教師あり単クラスサポートベクターマシンによる検出手法を検討し、特徴の自動選択と判別基準の自主決定を目指している。全システムの実時間化と多視点から異常事件に関する情報の統合も今後の研究課題に含まれる。

## 文献

- [1] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 1, pp. 18-32, 2014.
- [2] D.-G. Lee, H.-I. Suk, and S.-W. Lee, "Crowd Behavior Representation Using Motion Influence Matrix for Anomaly Detection," presented at the Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on, 2013, pp. 110-114.
- [3] R. Laxhammar and G. Falkman, "Sequential Conformal Anomaly Detection in trajectories based on Hausdorff distance," *Audio, Transactions of the IRE Professional Group on*, pp. 1-8, Jun. 2011.
- [4] Y. Benezeth, P.-M. Jodoin, and V. Saligrama, "Abnormality detection using low-level co-occurring events," *Pattern Recognit Lett*, vol. 32, no. 3, pp. 423-431, Dec. 2010.
- [5] D. Xu, X. Wu, D. Song, N. Li, and Y.-L. Chen, "Hierarchical activity discovery within spatio-temporal context for video anomaly detection," presented at the Image Processing (ICIP), 2013 20th IEEE International Conference on, 2013, pp. 3597-3601.
- [6] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1436-1452, Oct. 2013.
- [7] T. M. Hospedales, J. Li, S. Gong, and T. Xiang, "Identifying Rare and Subtle Behaviors: A Weakly Supervised Joint Topic Model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 12, pp. 2451-2464, 2011.
- [8] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," presented at the Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on, 2008, pp. 413-422.
- [9] F. Chen, "Hot-spot detection by group interaction extraction from trajectories," *RO-MAN, 2013 IEEE*, pp. 406-411, 2013.
- [10] F. Chen and A. Cavallaro, "Detecting group interactions by online association of trajectory data," presented at the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 1754-1758.
- [11] F. Chen, L. Zicheng, and M.-T. Sun, "Anomaly Detection By Using Random Projection Forest," *Image Processing (ICIP), 2015 IEEE International Conference on (Submitted)*, pp. 1-5.
- [12] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning," *Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114*, vol. 5, no. 6, p. 12, 2011.
- [13] S. Dasgupta and Y. Freund, "Random projection trees and low dimensional manifolds," presented at the STOC '08: Proceedings of the fortieth annual ACM symposium on Theory of computing, 2008.
- [14] J.-N. Hwang, S.-R. Lay, and A. Lippman, "Nonparametric multivariate density estimation: a comparative study," *IEEE Trans Signal Process*, vol. 42, pp. 2795-2810, Oct. 1994.
- [15] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," presented at the Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on, 1999, vol. 2.
- [16] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision.," *the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 81, pp. 674-679, 1981.

- [17] V. Mahadevan, W. Li, and V. Bhalodia, “Anomaly Detection in Crowded Scenes ,” *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 1975-1981, Jun. 2010.
- [18] F. Chen, “Supplemental Materials of Our Anomaly Detection Method,” [Online]. Available: <http://www.jaist.ac.jp/~chen-fan/results/results-anomalydetection.htm>.
- [19] F. Chen, “監視映像の解析技術の研究—監視映像の理解から内容の表現方式まで,” *検査技術, 日本工業出版*, vol. 19, pp. 18-23, Mar. 2014.
- [20] F. Chen, C. De Vleeschouwer, and A. Cavallaro, “Resource Allocation for Personalized Video Summarization,” *Multimedia, IEEE Transactions on*, no. 99, p. 1, 2013.

〈発 表 資 料〉

題 名	掲載誌・学会名	発 表 年 月
Resource Allocation for Personalized Video Summarization	IEEE Transactions on Multimedia	2014年2月
監視映像の解析技術の研究—監視映像の理解から内容の表現方式まで	検査技術・日本工業出版	2014年3月
Automatic Extraction of Semantic Features for Real-time Action Recognition using Depth Architecture networks	IEEE International Conference on Image Processing	2014年10月
Anomaly Detection By Using Random Projection Forest	IEEE International Conference on Image Processing	2015年9月 (Submitted)