

# 実環境における移動体の検出・追跡技術に関する研究

堀 武司, 波 通隆, 飯島 俊匡

## A Technique of Moving Object Detection and Tracking in Real Environments

Takeshi HORI, Michitaka NAMI, Toshimasa IJIMA

### 抄 録

近年、通行量調査や防犯などの目的のために、街頭に設置される監視カメラが増加している。これらのカメラから得られる大量の動画データを有効に活用するために、コンピュータを用いた自動処理技術のニーズが高まっている。こうしたシステムを構築するためには、動画から人や車両などの移動体検出が共通の処理として必要となる。しかし、移動体検出のために一般的に用いられる背景差分法などの手法は周囲の環境変動などに対するロバスト性が低く、屋外などの実環境下での使用は困難であった。本研究では、ロバスト統計手法の応用により外乱を含む入力画像列から背景モデルの自動生成と逐次更新を行う手法を開発した。また、得られた背景モデルを用いて移動体の検出と追跡を行うシステムの実装を行った。車両などの通行量調査を模した環境で試作システムの試験を行い、良好な結果を得た。

キーワード：動画解析、背景差分法、ロバスト統計、LMedS推定

### Abstract

In recent years, video surveillance cameras installed in streets are increasing, for the purpose of traffic census, crime-prevention and so on. To utilize the video data obtained from these cameras effectively, automatic video image analysis technology is strongly desired. Detection and tracking of moving objects from video images is one of the most fundamental processes for the video analysis systems. However, generally used methods such as background subtraction, cannot apply to the application used in real environment for lack of robustness against the disturbance such as change of lighting condition. In this study, we developed a method for estimating the background model from the input image sequence including disturbance. Moreover, we implemented a system for moving object detection and tracking with the proposed background estimation method. To confirm the effectiveness of the system, results of the experiments in the real environment are illustrated.

KEY-WORDS : video image analysis, background subtraction, robust statistics, LMedS estimation

### 1. はじめに

近年、様々な目的で街頭などへの監視・モニタ用ビデオカメラの設置が増加しており、これらのカメラから得られる大量の動画データ処理を自動化する技術へのニーズが高まっている。

店舗の新規出店計画のためのマーケティング調査や、商店街の活性化のための基礎資料などとして、市街地での通行量調査が盛んに行われている。特に最近では、多様化する消費者ニーズに合わせたきめ細かなサービスを提供するために、歩行者の数や性別だけでなく、世代層などの情報を含む、より詳細な調査を行いたいとの要望が増えている。現状では、これらの調査は調査員を現場に配置し目視により計数を行う場合が多いが、人件費などのコストが高いため調査場所や期間が制限されるという問題がある。そこで、ビデオカメラや

事業名：一般試験研究

課題名：実環境における移動体の検出追跡技術に関する研究

各種センサを用いた自動計測技術を導入し、より詳細かつ継続的に調査を行う事が望まれている。

また、最近の子供などを狙った凶悪犯罪の増加により安全に対する市民の意識が高まっており、犯罪を抑止するために学校周辺やマンションなどに監視カメラを設置する動きが全国に広がっている。また、犯罪発生率の高い都市の繁華街などでは、自治体や警察の主導により監視カメラ設置が進められている。公共空間における監視カメラの運用に関してはプライバシー面での問題を指摘する声もあるが、地域団体や自治体による監視カメラ設置は、今後も増加するものと予想される。防犯用途の場合、対象地区の全体を死角なくカバーするためにカメラ設置台数が増える傾向がある(例えば、新宿区の事例では、50台以上のカメラが設置されている)。このような多数の映像を人間が常時監視するのは現実的ではなく、カメラ情報の有効活用のためには自動監視技術の導入は必須となる。

監視カメラ用の自動化システムに要求される機能は、通行量調査、防犯監視などの用途により様々であるが、動画の中の人や車両などの移動物体を検出・追跡する処理は、ほぼ全ての動画処理システムにおける基本的機能として必要であり、この部分の安定した動作はシステム全体の有用性に大きく影響する。しかし、現在移動体検出のために一般的に用いられている背景差分法などの手法は、日照条件の変化など外乱要因に対するロバスト性が十分ではなく、街頭や店舗内などの実環境で使用されるシステムにそのまま適用する事は困難である。

そこで本研究では、外乱を含む実環境下で使用する各種自動監視システムへの適用を想定した、汎用的でロバストな移動体の検出と追跡技術の開発を行った。

以下、2章では、従来用いられてきた移動体検出手法とその問題点について述べる。3章では、環境変動に対するロバスト性を確保するために、背景差分法の背景モデルの構築と更新を、ロバスト推定手法の応用により逐次的に行う手法の提案と評価を行う。4章では、得られた背景モデルを用いた移動体検出、および追跡処理の実装の詳細について説明し、5章では通行量調査を想定したテスト映像データを用い、試作システムの評価を行った結果について述べる。

## 2. 従来手法

移動物体領域の検出は、動画計測における基本処理の一つであり、様々な手法が提案されている。中でも代表的な手法としては、背景差分法とフレーム間差分法が挙げられる。

背景差分法は、移動物体を含まない静止した背景画像と、時刻  $t$  における画像との差分から移動物体を検出する手法である。背景差分法は原理が単純であり、背景データさえ取得出来れば検出対象となる移動体に関する制約も少ないため、

利用しやすい手法である。

しかし、実システムの運用場面ではシーン中に人や車両が常時存在する事も多く、移動体を含まない背景単独の画像データを確実に取得するのは困難である場合が多い。また、一度背景を取得出来たととしても、背景自体も中長期的には日照条件などにより変動するため、それに追従するために背景データを逐次更新しなければならない。また、風になびく樹木などのような、周期的な揺らぎを含むシーンでは誤検出を生じるため、揺らぎを含めた背景のモデル化が必要となる。

フレーム間差分は、時刻  $t$  の画像と時刻  $(t-n)$  の画像との差分を用いる手法である。一般的には  $n=1$  として隣接フレーム間で処理を行う。フレーム間差分法では近傍時刻の画像を用いて演算するため、背景差分法と比較して周囲の環境変動に対して頑健である。しかし、対象が画面内で静止してしまうと検出出来なくなるという欠点がある。また、移動体領域内部にテクスチャが少ない場合、移動体のエッジ部分のみが検出される、移動体の移動速度や動画のフレームレートにより検出画像の現れ方が変化する、といった性質があり、やや使いにくい面がある。

本研究では、移動体検出の基本原則として背景差分法を用い、背景の変動などの問題に対応するために、過去に取得された画像列を用いて背景モデルの自動生成、逐次更新を行う手法について検討した。

## 3. ロバスト推定手法を用いた背景モデル更新手法

背景モデルの自動生成を行うための単純な方法としては、過去に撮影された複数枚の画像の移動平均を用いる方法などが考えられる。しかし、画像には背景の他に検出対象である移動体も含まれており、これが例外値として推定結果に大きく影響を及ぼす。そこで、ロバスト推定と呼ばれる手法を導入することで、画像列に含まれる移動体の影響を取り除き、背景のみの抽出を実現する。

最小二乗法などを用いた場合、データ中に真の値から大きく外れた例外値が少数でも含まれていると、全体の推定結果

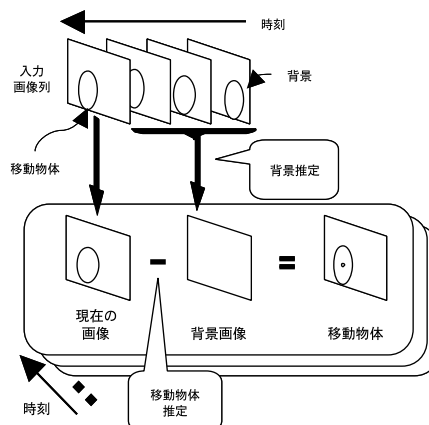


図1 背景モデル推定処理のフロー

に大きく影響を受ける。ロバスト推定とは、データにある程度まで例外値が含まれていても、その影響を最小限に抑えて安定した推定が可能な手法の総称である。本手法では、画像列の中に含まれる背景の値を推定値、時々現われる移動体を例外値と見なし、例外値をロバスト推定によって排除することで、背景データのみの抽出を実現する。

ロバスト推定の代表的な手法としては、M-推定<sup>1)</sup>、LMedS推定<sup>2)</sup>などがあるが、今回の実装ではLMedS推定による手法を用いた。

### 3.1 LMedS推定

Least Median of Squares(LMedS)推定は、最小二乗(LMS)基準

$$LMS = \min \sum_i \varepsilon_i^2 \quad (1)$$

の代わりに、LMedS基準

$$LMedS = \min \text{med} \varepsilon^2 \quad (2)$$

を最小化する事で、パラメータの推定を行う手法である。ここで、medは中央値を表す。ロバスト性を評価するための尺度として、推定結果に大きな影響を与えずにどれだけ割合の例外値を許容出来るかを示すbreakdown pointと呼ばれる指標があるが、LMedS推定のbreakdown pointは0.5であり非常に高いロバスト性を有している。

LMedS推定、すなわち式(2)の最小化は非線形の最適化問題であるため、特に探索空間が高次元である場合は、厳密解の計算は困難である。そこで、一般的には準最適解を高速に求める事が可能なランダムサンプリング法<sup>3)</sup>が用いられている。

ランダムサンプリング法では、以下の手順で処理を行う。

1. 全データから、 $F$ 個のサンプルをランダムに選択する( $F$ はモデルのパラメータ数、すなわち探索空間の次元)。
2. 1. で選択されたサンプルを用い、モデルを構築する。
3. 2. で得られたモデルを用いて、全データに対して式(2)のLMedS基準を計算する。
4. 以上の処理を  $N$  回反復して行い、最良のパラメータを採用する。

反復回数  $N$  は、要求される解の信頼性から決定される。例外値に影響されないためには、少なくとも一回以上の反復において、 $F$ 個のサンプルの中に例外値が含まれなければよい。そこで、データに含まれる例外値の割合を  $\varepsilon$  とすると、その確率  $P$  は

$$P = 1 - \{1 - (1 - \varepsilon)^F\}^N \quad (3)$$

となる。例えば、 $P = 0.99$ 、 $\varepsilon = 0.5$ 、 $F = 2$  とすると、反復回数  $N$  は18回以上となる。

### 3.2 背景モデル推定への応用

背景モデルとして単純な輝度値を用いた場合、時刻  $t$  の背景画像  $I(t)$  の位置  $(x, y)$  における輝度値  $I_{xy}(t)$  を一次元のデータ列と見なす事が出来る。これにLMedS推定を適用する。この場合、パラメータ数  $F = 1$  となる。ランダムサンプリング法を用い、例外値の比率  $\varepsilon$  をLMedS推定の上限値0.5、 $P = 0.99$  とすると、式(3)から必要な反復回数は  $N = 7$  となる。

移動体検出処理の精度向上のため、背景モデルとして画素の色情報や、エッジ、テクスチャ特徴など、より複雑な表現を用いた場合でも、ほぼ同様の手法で扱うことが可能である。

### 3.3 性能評価

背景モデルとして各画素の色情報(Y, Cr, Cb)の3チャンネルを用い、ランダムサンプリング法を用いて背景推定を行うプログラムを実装し、性能評価を行った。

背景推定に使用する過去画像のフレーム数と画像解像度を変化させ、背景更新処理の速度を計測した結果を表1に示す。画像解像度と処理速度は、ほぼ比例関係にあることがわかる。320×240画素では一回の更新に14秒～25秒を要しているが、160×120画素や80×60画素では十分に実用的な速度が得られている。一般的な動画像解析アプリケーションにおいては、前処理としての移動体検出処理では物体の存在や動きを大まかに捉える事が出来れば十分であるため、ある程度まで解像度を落とした画像を用いても問題は少ないと考えられる。

また、LMedS推定の計算を行うためには、過去の画像データを保持しておく必要がある。高解像度で処理を行った場合は、10～30フレーム分の画像データを主記憶上に常時保持する事となり、メモリ消費量の大幅な増大を招く。この点からも、低解像度での処理が現実的といえる。

実際の背景推定処理の過程を図2に示す。これは、室内の蛍光灯の消灯により、照明条件が大きく変化した場合の例であり、上段がLMedS推定により得られた背景モデル画像、下段が背景差分画像である。照明が変化した直後の状態(左端画像)では、不正確な背景モデルを用いて背景差分処理を行った結果、画面全体のかなりの部分を移動体領域として誤検出している。しかし、時間経過と共に背景モデルが徐々に正しい値に更新され、背景差分処理における誤検出も解消されている。

環境変化に対して背景モデルが追従するまで時間は、過去フレーム中における環境変化後のデータ数がLMedS推定のbreakdown point(0.5)を超えるまでの時間とほぼ等しい。実時間における背景更新の追従速度は、過去フレームの更新間隔により調整する事が出来る。間隔を狭くする(すなわち、

現在時刻に近い範囲のデータで推定を行う)と、外乱に対する追従性は向上する。一方、シーン中で移動体が静止した場合、背景と同化してしまい正常に検出できなくなってしまうまでの時間が短くなる。そのため、対象とするシーンの特性に応じて、フレーム更新間隔を適切な値に設定する必要がある。

表 1 背景更新の処理速度

解像度	フレーム数		
	10	20	30
320x240	14.1	19.4	25
160x120	3.6	5.0	6.5
80x60	0.9	1.2	1.6

(単位：秒) 実行環境：IBM-PC 互換機, Pentium4 2.6GHz, Windows-XP SP2

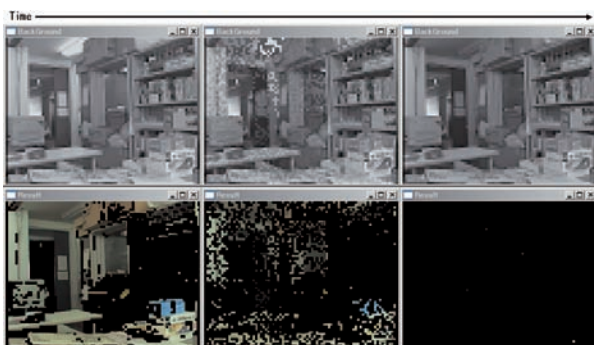


図 2 背景モデルの更新過程  
(上段：背景モデル画像, 下段：背景差分処理結果)

#### 4. 移動体の検出と追跡

3章で述べた背景推定手法を用い、実際に移動体の検出と追跡を行う試験システムの実装を行った(図3)。以下、具体的な処理内容に関して順に説明する。

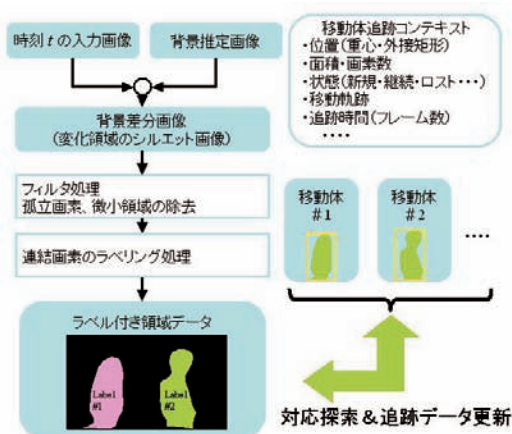


図 3 移動体の検出・追跡処理

#### 4.1 背景差分処理

LMedS推定で求められた背景モデルと、現在時刻の画像とを用いた背景差分法により、移動体領域の抽出を行う。

背景モデルとしては、単純な輝度値、Y/Cr/Cbの色情報、およびエッジ抽出データを用いた試験を行ったが、検出精度と処理速度のバランスから、本システムの実装では色情報による処理を採用した。

差分処理によって得られた移動体領域のシルエット画像には、外乱やカメラのノイズによる誤検出が含まれる。また、移動体領域の値が背景と偶然一致してしまった場合にも検出漏れが生ずる。そこで、これらを修正するためにシルエット画像に対するフィルタ処理を行った。

検出対象である歩行者や車両は極端な凹形状や孔を含まず、かつ低解像度化のために詳細形状は失われているという前提に基づき、抽出されたシルエット画像に対して図4に示す3×3フィルタを複数回反復して適用し、孤立画素の除去やシルエットの欠けの補正などを行った。

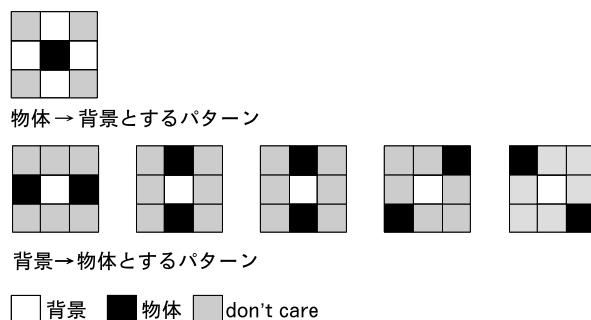


図 4 差分画像に対するフィルタ処理

#### 4.2 ラベリング処理

背景差分処理結果のシルエット画像から移動体を個別に識別するために、連結画素のラベリング処理を行う。ラベリングと同時に各ラベル領域の面積(画素数)、重心位置、バウンディングボックス(領域に外接する最小の矩形領域)なども同時に計算し、ラベル領域の属性データとして記録する。

ラベル付けされた領域のうち、一定の条件を満たさないもの(例えば面積がしきい値未満など)は、誤検出としてこの段階で除去する。

ラベリング処理の結果、画面の大部分を単一のラベル領域が占めるような場合は、背景の急激な変動などにより背景差分処理自体が失敗している可能性が高い。このような場合はエラーとして処理全体を一旦停止し、回復のための例外処理(現在の背景モデルを破棄し再生成するなど)に移行する。

#### 4.3 移動体の追跡処理

4.2節で生成されたラベル画像は、フレーム毎に独立に計算されたものである。これら各フレームのラベリング結果を時系列的に対処付ける事で、移動体の識別と追跡処理を行う。

画像中の複数の移動体を独立して追跡するため、移動体追跡処理に関するコンテキストを管理する「トラック」を単位として処理を行う。各トラックは以下の情報を保持する。

- ・ 重心位置座標
- ・ 外接矩形領域(バウンディングボックス)
- ・ 領域面積(画素数)
- ・ 状態(新規・安定追跡中・ロスト)
- ・ (重心の)移動軌跡
- ・ 追跡時間(フレーム数)

フレーム毎の追跡処理は、以下の手順で行われる。

1. 現在有効なトラックについて、対応するラベル領域を探索する。
2. 領域が見つかった場合は、当該領域を移動後の新しい物体として認識し、トラックのコンテキストを更新する。見つからなかった場合は物体が消滅したものとして追跡を中断する。
3. 既存のどのトラックにも対応付けられなかったラベル領域があれば、新規のトラックを割り当て、追跡処理を開始する。

外乱や移動体抽出処理の精度の問題により、ラベル画像データ中の移動体領域が一時的に消失する、また逆に誤検出領域が出現する場合がある。追跡処理に対するこれらのノイズの影響を抑えるため、トラックに対して図5に示す状態遷移を導入した。追跡処理中に移動体を見失った場合でも直ちに消失とはせず、一定期間内に対応領域を再発見出来た場合はそのまま追跡処理を継続する。逆に、新規検出された物体については、一定期間以上連続して検出された場合のみ正規の追跡対象と扱う。

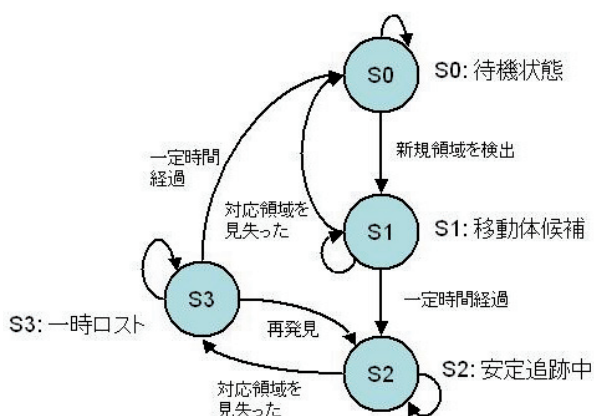


図5 トラックの状態遷移

図6は、今回実装したシステムを用いて、室内の歩行者を検出・追跡した例である。処理結果を可視化するために、背

景差分法で得られたシルエット画像に追跡処理によって認識された移動体の重心、外接矩形、トラックの状態、および移動軌跡などの情報を重畳して表示している。



図6 移動体検出・追跡処理の例

## 5. 実験と評価

実装した試験システムを用い、屋外の実環境下での移動体検出・追跡に関する評価試験を行った。

### 5.1 撮影環境と実験条件

通行量調査の自動化アプリケーションを想定して条件設定を行い、歩行者と車両が通行する一般道路を横方向からビデオカメラ(miniDV規格)で撮影した。撮影されたシーンの代表例を図7に示す。

解析処理は、録画されたビデオ映像を処理用PCでキャプチャしながら実時間処理で行った。元画像の解像度は720×480画素、背景モデル推定と差分処理は元画像を縮小した90×60画素で行った。処理速度は、Pentium4 2.6 GHzのPCを用いた場合で約30 frame/secであり、ビデオレートでの処理を実現出来た。

### 5.2 背景更新処理

3.3節で述べたように、背景モデル更新処理では、画面内で停止する移動体が背景として誤認識されないように適切な

更新間隔を設定する必要がある。

撮影場所の右側には交差点があり、信号機が設置されているため、信号待ちの車両が最大40秒程度画面内に滞留する可能性がある。そこで、背景推定に使用する過去フレーム数を10、フレーム更新間隔を10秒とした。この場合、過去100秒間のデータから背景推定を行う事になり、移動体が同一位置に40秒間静止していたとしてもLMedS推定のbreakdown point(100秒 × 0.5 = 50秒)を超えないため、背景と誤認識される事はない。実際の試験結果でも、背景モデルの更新が正しく行われている事を確認した。

### 5.3 誤検出の防止と制約条件

シーン中には数本の街路樹が含まれており、常時風で揺らいているため、街路樹自身やその影が移動体領域として誤検出される。また、遠方には別の道路が存在しており、そこを通行する車両、歩行者も移動体領域として検出される。

これらの影響を除去し、計測対象である手前側道路を通行する歩行者、車両のみを検出・追跡するため、移動体追跡処理の段階で、4章で説明したものに加えて以下の制約条件を追加した。

- ・明らかに歩行者、車両ではない大きさ、縦横比を持つ領域は追跡対象としない。
- ・道路から外れた場所の領域(重心位置が画面上側1/3、もしくは下側1/3にある領域)は追跡対象としない。
- ・新規の追跡対象とする領域は、1) 画面左右端に接するもの、2) 既に処理中の移動体領域から分離したものに限定する。

今回の実験では、上記の制約を加える事により外乱の影響をほぼ排除し、計測対象の移動体のみを安定して追跡する事が出来た。

### 5.4 オクルージョンの扱い

現在の試作システムでは、複数の移動体の重なりやオクルージョン(隠蔽)については特別な考慮を行っていない。図7のシーンでは、画面の左右から歩行者と車両それぞれ反対方向に移動しているが、両者が画面中央部で重なった時点(図7上から三枚目)で一つの物体として誤認識されている。

通行量調査などの場合は、

- ・移動体の分布が、それほど密ではない。
- ・移動体は、原則として方向転換や不規則な移動を行わない。

といった制約があるため、すれ違いなどによって生じる一時的なオクルージョンについては、前後の履歴情報からある程度の解決が可能である。図8は、図7のシーンでの移動体検出結果(重心のx座標)を時系列順に表示したものであるが、前後の関係から移動体ID #89と#91、#88と#90がそれぞれ同一

物体のあると容易に推測可能である。

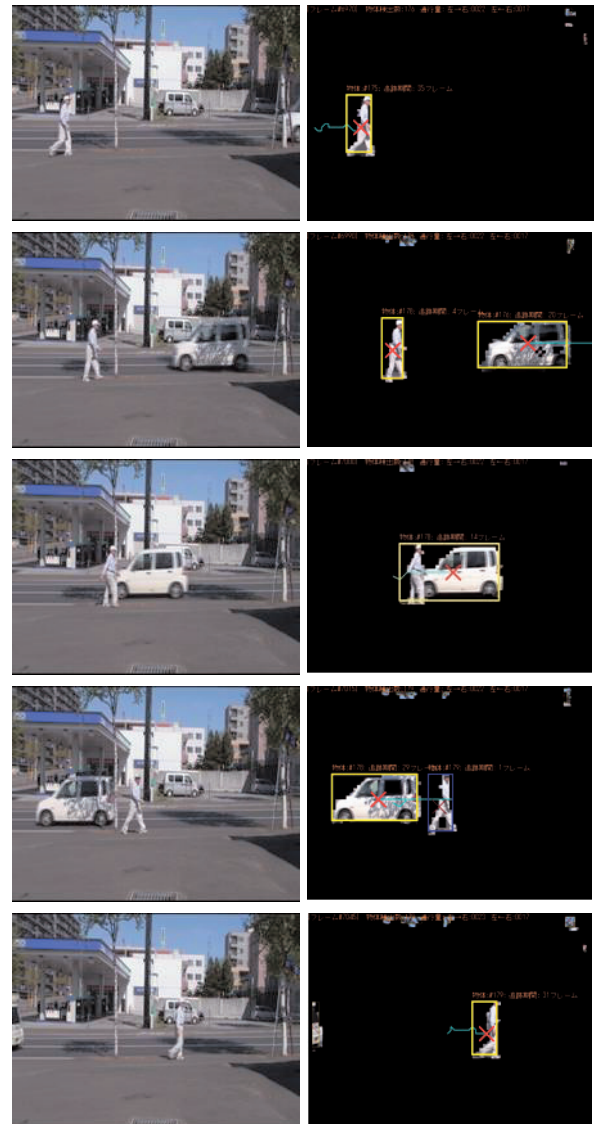


図7 屋外環境における歩行者、車両の検出

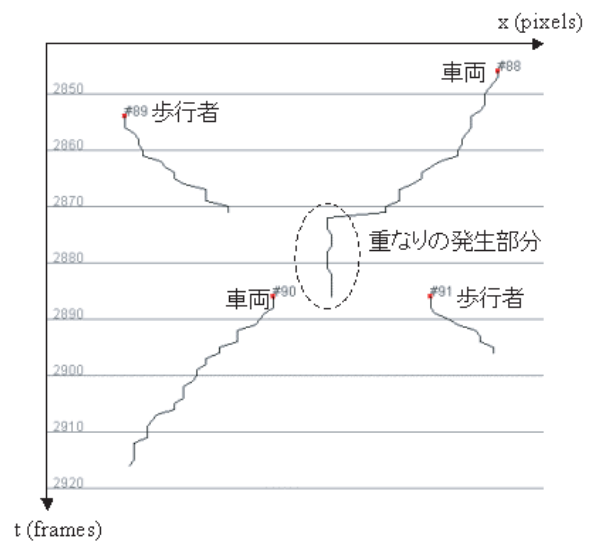


図8 図7のシーンにおける移動体追跡の履歴

より正確にオクルージョンの処理を行うには、背景差分によるシルエットの位相情報だけではなく、領域内部の画像パターンまで考慮した処理が必要である。こうした処理は個別のアプリケーションに依存する要素が大きいため、応用システム開発において本システムの上位処理として実装されるべきである。

## 6. まとめ

ロバスト推定手法の一つであるLMedS推定を用いた背景画像モデル推定手法を提案し、それを用いて汎用の移動体追跡システムの試作を行った。また、屋外の実環境において試作システムの評価を行い、安定した移動体追跡が可能であることを確認した。

今後は、通行量調査や防犯監視など具体的なニーズを持つ企業と共同で、本報告の成果に基づく各種応用システム開発を目指す予定である。

本研究では、動画像処理の初段における大まかな移動体検出に範囲を絞った技術開発を進めてきたが、これまでに当場で取り組んできた顔画像認識などのパターン認識技術との融合により、更に高度な知的センシング技術の開発を進めていきたい。

## 引用文献

- 1) 小出 治：防犯カメラの効果と利用について，予防時報，社団法人 日本損害保険協会，224号，pp.30-35，(2006)
- 2) 島井博行・栗田多喜夫・梅山伸二・田中勝・三島健稔：ロバスト統計に基づいた適応的な背景推定法，電子情報通信学会論文誌D- ， Vol.J86 No.6，pp.796-806，(2003)
- 3) 栗田多喜男：統計的画像処理手法，<http://www.neurosci.aist.go.jp/kurita/lecture/statimage/>
- 4) 竹内啓五・金子俊一・五十嵐悟・佐藤雄隆・羽根 義：ロバスト背景差分及び領域抽出に基づく歩行挙動の画像解析，画像電子学会誌，vol.31 No.2，pp.193-201，(2002)
- 5) 堀 武司・大崎恵一・波 通隆：顔画像認識システムの開発，北海道立工業試験場報告，No.301，pp.31-38，(2002)