

複数物体の重なりを理解するレイヤー型検出法

Layered Detection for Multiple Overlapping Objects

藤吉 弘巨

Hironobu Fujiyoshi

中部大学 工学部情報工学科

College of Engineering, Chubu University

金出 武雄

Takeo Kanade

カーネギーメロン大学ロボティクス工学研究所

The Robotics Institute, Carnegie Mellon University

EMAIL: hf@cs.chubu.ac.jp

Abstract 本論文では、複数物体の重なりを理解するレイヤー型検出法を提案する。本手法は、ピクセル分析とリージョン分析の二つの処理からなる。ピクセル分析では、各ピクセルの輝度値の時間変化を観測し、その変化軌跡によりピクセルの状態を静もしくは動と判定する。判定にある時間幅における輝度変化を用いることで、太陽光等の環境変化の影響を受けにくくなる。リージョン分析では、動とラベル化されたピクセル領域を移動物体と判定する。静とラベル化された領域は静止物体と判定し、背景上にレイヤーとして記憶する。これにより、レイヤー上を通過する移動物体を区別して検出することが可能となる。実環境下における評価実験を行い、本検出手法の有効性を確認した。

1 はじめに

近年、犯罪発生率の急増に伴い、画像理解技術を用いた監視システムの研究が盛んになってきている。米国ではDARPA(Defence Advanced Research Projects Agency)の基、複数のカメラが協調して動作するビデオ監視システムの研究プロジェクトVSAM(Video Surveillance and Monitoring)が行われている[1, 2]。このような新しいビデオ監視システムにおいて、屋外の画像を対象とする場合、天候等による環境変動の影響を受けずに侵入物体を検出することが重要となる。移動物体の検出には、以下の三つの手法が多く用いられている。

(1) フレーム間差分法 [3]

フレーム間差分法は、現在の入力画像と前回の画像との差分を計算し、差分値の大きい領域を移動物体として検出する。動的な環境変化に適応的であるが、一般に移動物体すべての領域を抽出することは不可能である。

(2) 背景差分法 [4, 5, 7]

背景差分法は、検出すべき物体が存在しない背景画像を予め用意しておき、入力画像と背景画像の差分を計算する手法である。ほぼ完全な移動物体領域を検出することができるが、天候等

による環境変化に対応して背景画像を更新する必要がある。

(3) オプティカルフロー法 [6]

オプティカルフロー法は、カメラが動いている状態でも物体を検出することができるが、計算量が多いため特別なハードウェアを使用しない場合、リアルタイムでの演算は困難である。

“人が自動車から降りた”等の人と自動車間の行動(アクティビティ)を認識するには、人が自動車から降りた瞬間を観測する必要がある。その際には、人と自動車画像上で重なっている場合が多く、従来の物体検出と領域クラスタリングを組み合わせた手法では、それぞれの領域を分けて検出することが難しい。

本論文では、複数物体の重なりを理解する物体検出法として、ピクセル分析とリージョン分析の二つの処理からなるレイヤー型検出法を提案する。本手法は、一度静止した物体を背景上のレイヤーとして登録するため、その後、そのレイヤー上を通過する移動物体を区別して検出することが可能となる。

まず、2では検出アルゴリズムについて述べ、3では実環境下における提案手法と背景差分法の実験結果を示す。

2 レイヤー型検出アルゴリズム

レイヤー型検出法は、ピクセル分析とリージョン分析の二つの過程からなる。ピクセル分析の目的は、各ピクセルの時間的変化を観測することにより、そのピクセルの状態を静 (stationary) もしくは動 (transient) と判定することである。リージョン分析の目的は、ピクセルの集まりを移動物体 (moving object) もしくは静止物体 (stopped object) と判定することである。図 1 はピクセル分析とリージョン分析の概念を示す。あるピクセルにおいて、その輝度値の状態は初期値である背景から始まり、物体が通過するとき transient の状態へ遷移する。その物体が停止したときは、stationary の状態へ遷移する。

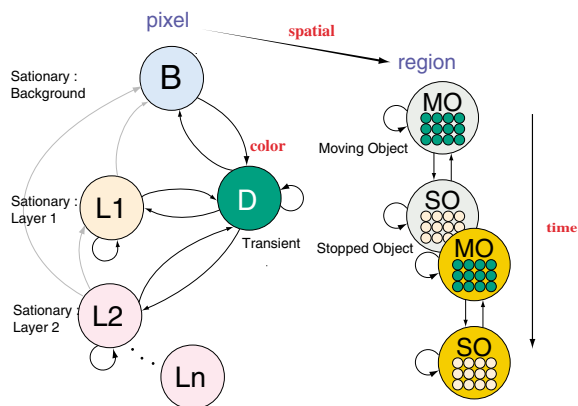


図 1: ピクセルとリージョン分析の概念

図 2 は、検出過程を示す。本検出アルゴリズムは、各ピクセルが transient であるかどうかを決定するために、ある時間幅における輝度値の振舞を観測する必要がある。ピクセル分析により transient, stationary と判定されたピクセル群は、クラスタリングにより各領域にまとめられる。リージョン分析は、領域が transient ピクセルの場合移動物体、stationary ピクセルの場合静止物体と判定する。レイヤーマネージメントは、静止物体と判定された領域を新規レイヤーとして登録する。再び静止物体が動き始めるとレイヤーの削除を行う。

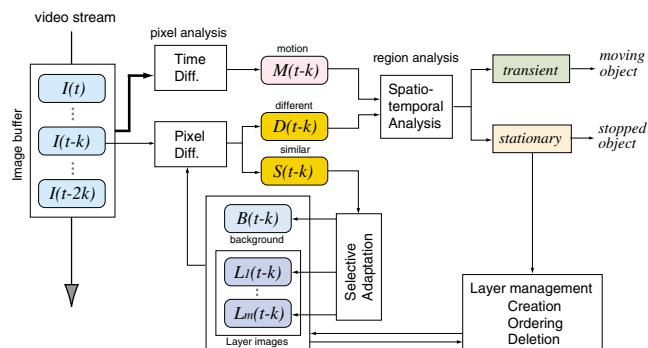


図 2: レイヤー型検出の過程

2.1 ピクセル分析

屋外カメラで撮影した画像上のピクセルの輝度変化は、状況に応じて図 3 に示すように三つの特徴に大別できる。

- 物体がピクセル上を通過するとき、そのピクセルの輝度値は急激な変化を伴う。その後、一時的に不安定な状態が続き、再度急激な変化の後、背景である元の輝度値に戻る (図 3(a) 参照)。
- 物体がピクセル上で停止したとき、そのピクセルの輝度値は急激な変化の後、一時的に不安定な状態が続き、最終的には物体の輝度値に安定する (図 3(b) 参照)。
- 太陽が雲に隠れた等の環境変化が生じたとき、輝度値は図 3(c) に示すように緩やかに変化する。

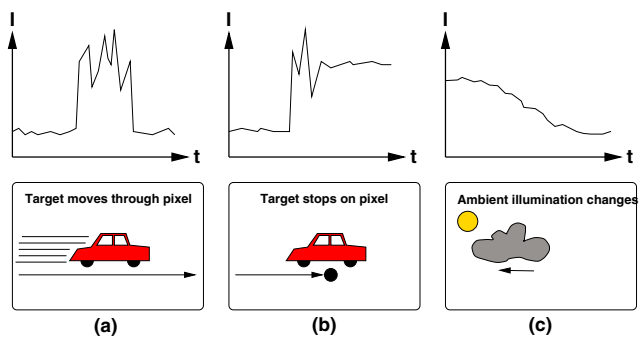


図 3: ピクセルの輝度変化

これらの特徴を捉えるためには、以下の二つの変化に着目する必要がある。

1. 輝度値の急激な変化
2. 不安定な状態から安定した状態への遷移

まず最初に、急激な輝度変化を捉えるモーショントリガーを考える。ここで、 I_t を現在のフレームから k フレーム前の輝度値とする。輝度値の変化量 T を求めるには、 t よりも前（過去）のフレームに着目する。変化量 T は次式により求められる。

$$T = \max \left\{ |I_t - I_{(t-j)}|, \forall j \in [1, 5] \right\}$$

急激な輝度値の変化がピクセル上に生じたとき、 T の値は大きくなる。

次に、ピクセルの安定度を表す S について考える。安定状態の検出には t より後（未来）のフレームに着目する。安定度 S は次式により求められる。

$$S = \frac{k \sum_{j=0}^k I_{(t+j)}^2 - \left(\sum_{j=0}^k I_{(t+j)} \right)^2}{k(k-1)}$$

これはフレーム t から $t+k$ までの分散であり、安定した状態のとき S は小さくなる。

ここで、各ピクセルの状態を表す M は、背景のとき bg 、transient のとき tr 、stationary のとき st をとるように、以下に示すアルゴリズムにより決定される。 I は、時刻 t おける輝度値とする。 Th_t は急激な変化を判定する閾値、 Th_s は安定性を判定する閾値である。background intensity は背景画像として、予め用意しておく。

```

if ((M = st OR M = bg) AND (T > Th_t))
  M = tr
if ((M = tr) AND (S < Th_s)) {
  if (I = background intensity)
    M = bg
  else
    M = st
}

```

この操作を全ピクセルについて行う。背景と判定されたピクセルは、次式に示す IIR フィルタ

により環境変化に追従するよう適応的に更新する [7]。

$$B(t) = \alpha I(t) + (1 - \alpha)B(t-1)$$

α は、背景と判定されたピクセルを背景画像にどれだけ反映させて更新するかを決定する定数である。

2.2 リージョン分析

リージョン分析では、ピクセル分析によりラベル化された背景以外のピクセル領域について検討する。まず最初に、背景以外のピクセル群に対して、領域クラスタリングを行う。クラスタリングには、近隣領域の連結性を用いた手法が多いが、本研究では、ある半径 r_c 以内の距離に存在するピクセル同士を同一クラス $R(i)$ に属することとした。 i はクラス番号である。このとき、各領域 $R(i)$ は以下の三状態に分けられる。

1. 領域が全て transient ピクセルから構成されている場合、その領域は moving object と判定できる。
2. 領域が全て stationary ピクセルから構成されている場合、その領域は sttoped object と判定できる。その領域が登録されていない場合は新規レイヤーとして登録する。
3. 領域が transient と stationary ピクセルの混合として構成されている場合、その領域内に sttoped object と moving object が含まれていると考えられる。そこで、レイヤー操作によりレイヤーとして登録されていない領域を抽出し、再度領域クラスタリングを行う。クラスタリング後の領域が stationary ピクセルのみの場合は、sttoped object と判定でき、新規レイヤーとして登録する。クラスタリング後の領域が transient ピクセルの場合は moving object と判定できる。

すべての領域 $R(i)$ は、以下に示すアルゴリズムにより、sttoped object, moving object と

判定される．ここで， $L(j)$ はレイヤー画像， j は既に登録されているレイヤーの数とする．

```
if (R = tr) {
    R -> moving object
}
elseif (R = st) {
    %remove all pixels already assigned
    %to any layer
    R = R - (L(0) + L(1) + .. + L(j))
    %if anything is left, make a new
    %layer out of it
    if (R != 0) {
        make new layer L(j+1) = R
        R -> stopped object
    }
}
else {
    %R contains a mixture of t and s
    R = R - (L(0) + L(1) + .. + L(j))
    SR(i) = spatial_clustering(R)
    for (each region SR(i)) {
        if (SR(i) = tr) {
            SR(i) -> moving object
        }
        if (SR(i) = st) {
            make new layer L(j+1) = SR
            SR(i) -> stopped object
        }
    }
}
```

レイヤーマネージメントは，領域が全て stationary ピクセルにより構成されているとき，その領域を新規レイヤーとして背景上に登録する．再び，そのレイヤー領域である物体が動き出したときはそのレイヤーを削除する．

3 検出実験結果

3.1 分析例

図 4 に，監視カメラ映像のある一点の輝度変化とそのピクセル分析例を示す．この映像シー

ケンスには，次に示す人と自動車によるアクティビティが含まれている．

1. 自動車 A が監視領域に侵入し，ピクセル上で停止．
2. 別の自動車 B が監視領域に侵入し，自動車 A のカメラからみて手前側に停止．
3. 自動車 B から人が手前側に降車して，ピクセル上を通過．
4. 同じ人が自動車に戻るため，再度ピクセル上を通過．
5. 自動車 B が発進してピクセル上を通過．
6. 自動車 A が発進してピクセル上を通過．

これらの各事象に対して，図 3(a),(b) に示した変動特徴と同じ軌跡を観測することができる．また，ピクセル分析が正確にピクセルの状態を transient と stationary に判定しているのが分かる．

図 5 は，上記の 4. の状態におけるレイヤー型検出によるリージョン分析例を示す．これは，三つの物体が重なっているにも関わらず，二つの stopped object と一つの moving object をそれぞれ検出した例である．静止物体は一時的に背景上のレイヤーとして登録されているため，その上を通過する移動物体である人の領域を区別して検出することが可能である．また，レイヤー操作により他の物体が重なっているために隠れているオクルージョン領域 (青色で示した領域) を知ることができる．

3.2 評価実験

提案した本検出アルゴリズムは，Pentium III(500MHz) の PC にインプリメントされ約 6 ~ 11[fps] で動作している．フレームレートの変動は，オブジェクトの数とレイヤーの数による．表 1 は，比較的交通量が多い大学の屋外駐車場

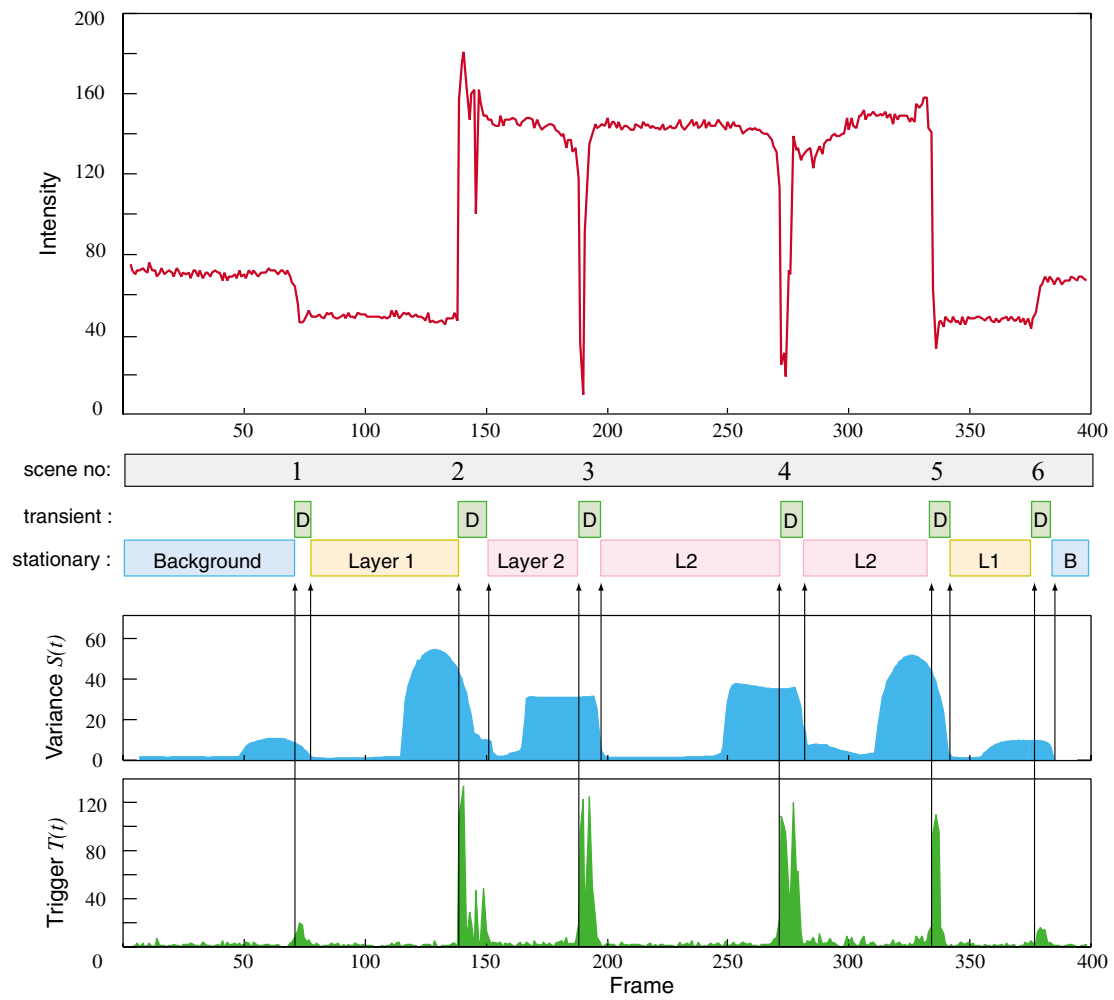


図 4: 輝度変化とその 픽セル分析例

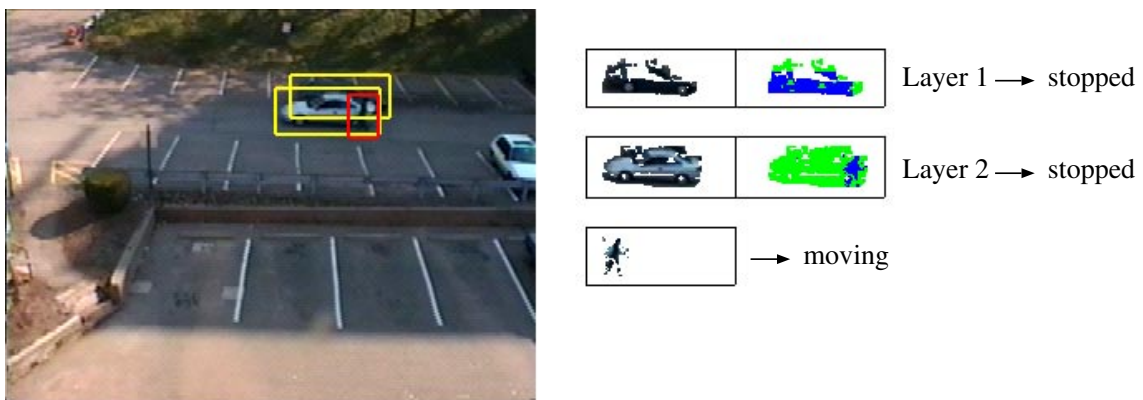


図 5: リージョン分析による検出例

の2箇所 (cam1, cam2) で撮影したビデオ映像 (約8時間) の検出結果を目視により評価したものである。MTDは、通常の背景差分手法を示す[7]。MTDの検出失敗は晴れの日が多い。これは、影の領域における背景と検出すべき物体のコントラストが低いためである。

一方、レイヤー型検出法は、ピクセルの状態を stationary もしくは transient と判定するのに、時間軸における分析窓の中心 (現在のフレームではなく k フレーム前) で決定しているため、このような低コントラストの状態でも検出が可能である。レイヤー型検出法における失敗は誤り検出が多く、その殆んどは太陽光が自動車の窓ガラスに反射したものであった。晴れの日での検出率が曇の日より低い原因はこの理由による。また、この反射による誤り検出は、太陽との位置関係によりカメラ1で多く生じた。そのため、カメラ1の検出率はカメラ2よりも低くなっている。

表 1: 検出率 [%]

	MTD		Layer method	
	cam1	cam2	cam1	cam2
sunny day	58.4	89.9	84.4	94.7
cloudy day	92.1	93.8	92.7	96.2
average	83.5		92.0	

4 まとめ

本論文では、ピクセル分析とリージョン分析の二つの処理からなるレイヤー型検出法を提案した。ピクセル分析では、各ピクセルの輝度値の時間変化を観測し、その変化軌跡によりピクセルの状態を動かと判定する。ある時間幅の変化軌跡を用いることで、低コントラストな場合でも検出が可能である。リージョン分析では、検出領域を移動物体もしくは静止物体と判定し、レイヤー操作により複数物体の重なりを区別できることを示した。屋外カメラを用いた実験を行い、従来の背景差分法より高い検出率 92[%] を得ることができた。これにより実環境下における本手法の有効性を確認した。

謝辞 本研究に助言して頂いた Dr. Robert T. Collins, Dr. Alan J. Lipton を始めとする CMU VSAM プロジェクトのメンバーに感謝致します。

参考文献

- [1] VSAM: “Section I, video surveillance and monitoring”, In *Proceedings of DARPA Image Understanding Workshop*, volume 1, pages 1–400, November 1998.
- [2] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa: “A system for video surveillance and monitoring: VSAM final report”, In *Technical report CMU-RI-TR-00-12, Robotics Institute, CMU*, May 2000.
- [3] P.L. Rosin and T. Ellis: “Image difference threshold strategies and shadow detection”, In *Proceedings of the British Machine Vision Conference*, pages 347–356, 1995.
- [4] I. Haritaoglu, D. Harwood, and L.S. Davis: “W4: Real-time surveillance of people and their activities”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830, August 2000.
- [5] C. Stauffer and W.E.L. Grimson: “Learning patterns of activity using real-time tracking”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, August 2000.
- [6] C. Cedras and M. Shah: “Motion-based recognition: A survey”, *Image and Vision Computing*, 13(2):129–155, March 1995.
- [7] A. Lipton, H. Fujiyoshi, and R.S. Patil: “Moving target detection and classification from real-time video”, In *Proceedings of the 1998 Workshop on Applications of Computer Vision*, pages 8–14, October 1998.