

## MRF モデルを用いた動画からの移動物体検出

保坂 忠明<sup>†</sup> 小林 匠<sup>†,††</sup> 大津 展之<sup>†,††</sup>

<sup>†</sup> 産業技術総合研究所 〒305-8568 茨城県つくば市梅園 1-1-1

<sup>††</sup> 筑波大学 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: †{hosaka-t,takumi.kobayashi,otsu.n}@aist.go.jp

**あらまし** マルコフ確率場 (MRF) に基づく移動物体検出手法を提案する。本手法の特徴は、複数枚の背景画像に対する最近傍法を用いたピクセル毎の背景差分処理により少ない計算量で効果的に背景変動に対応し、更にフレーム間のブロックマッチングを用いて対象の移動情報を取り入れたことである。MRF モデルの立場から、それらの特徴を踏まえたエネルギー関数を定義し、その大域的最適解をグラフカットにより求める。評価実験により、従来法と比較して性能・処理速度の面において提案手法の有効性を示すことができた。

**キーワード** 移動物体検出, 背景差分, マルコフ確率場, グラフカット。

## Moving Object Detection by Using Markov Random Field Model

Tadaaki HOSAKA<sup>†</sup>, Takumi KOBAYASHI<sup>†,††</sup>, and Nobuyuki OTSU<sup>†,††</sup>

<sup>†</sup> National Institute of Advanced Industrial Science and Technology  
1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 Japan

<sup>††</sup> University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573 Japan  
E-mail: †{hosaka-t,takumi.kobayashi,otsu.n}@aist.go.jp

**Abstract** A method of moving object detection based on Markov random field (MRF) is proposed. The characteristics of our method are less computational background subtraction using the nearest neighbor method for multiple background images to cope with dynamic scenes, and interframe block matching to incorporate object movements. From the viewpoint of MRF, we define the energy function by considering those characteristics, and optimize the function by using graph cut algorithm. Experimental results showed a favorable detection performance and less computational cost in many cases.

**Key words** Moving object detection, background subtraction, Markov random field, graph cut.

### 1. はじめに

動画から移動物体を適切に検出する技術は、画像認識やコンピュータビジョンの分野における基本的な問題の一つであり、モニターによる環境監視、人物のトラッキング、動作解析、映画撮影やテレビ会議における合成画像作成など幅広く適用可能な技術である。一般に利用されている移動体検出手法として、背景差分を用いる方法がある。本研究の目的は、背景差分を利用して性能面と処理速度の面から優れた移動物体検出手法を提案することである。

最も簡単な方法は、あらかじめ取得または推定した背

景画像と現在のフレームとの差分を計算し、その値と定義した閾値との大小関係を比較することで物体が背景かを推定する。この手法は、単純背景差分と呼ばれ、実装が簡単であり計算量も少ないため、多くのアプリケーションで利用されることも多いが、木の葉の揺れや明度変化、カメラの揺れなどが存在する動的な背景に対応できないという欠点がある。動的な背景に対応するために、これまで数多くの研究がなされているが、ここでは代表的な手法や本研究と関連の深い手法のいくつかを紹介する。

Stauffer らは、ピクセル毎に混合ガウス分布により背景をモデル化し、K-means 法に基づいた簡便な更新則を与えた [1]。そして、混合率の小さいコンポーネントに所属

するか、またはいずれのコンポーネントからも離れた色を移動物体とすることで推定がなされる。Pfinder [2] では、単一のガウス分布を用いているが、背景だけでなく抽出対象となる人についてもモデル化を行い、それらはフレーム毎に更新される。 $W^4$  [3] では、複数の背景画像に対し、ピクセル毎にピクセル値の最大値と最小値を求め、適応的に閾値を変化させている。それにより、おおまかに対象物体の位置を掴み、過去のフレームにおける結果との重複領域に基づきトラッキングを実現している。また、より複雑な背景変化に対応すべく、特定の分布形を仮定するのではなく、ノンパラメトリックな確率分布を用いる研究もされており、ElgammalらはParzen推定を用いて背景モデルを構築した[4]。これらの研究は、いずれもピクセル毎の推定を行うため、ちらつきなどのノイズには一般に弱く、後処理として膨張と収縮などを採用することもしばしば行われる。

一方で、画像修復などによく用いられるマルコフ確率場(MRF)に基づいた手法も近年提案されており、それらの研究では、隣接ピクセル同士は同じラベルをとりやすいというスムージングの効果が自然な形で取り込まれ比較的ちらつきなどのノイズには頑健となる。Howeらは、文献[5]において、単純背景差分を基本としたエネルギー関数を定義し、その最小化をグラフカット[6]を使って行った。ただし、単一の背景画像を用いているため、動的な背景には必ずしも対応できない。Migdalらは、文献[1]の混合ガウス分布を用いた方法により背景モデルを生成し、さらに、MRFの観点から隣接ピクセル間、および時間方向にもスムージング効果を取り入れてエネルギー関数を定義した[7]。この手法は、動的な背景にも対応できるものの、解を求めるためにマルコフ連鎖モンテカルロ法(Gibbs サンプラー)を用いており計算時間は必ずしも実用的ではない。Sheikhらは、ピクセル単位ではなく画像全体で一つの背景分布および対象物体の分布を生成し、それをフレーム毎に逐次更新していきながら、Parzen推定により差分値を評価した[8]。それに、スムージングの効果をとり入れることで、エネルギー関数を定義してグラフカットにより最適解を求めている。背景差分技術としては、精度の高いものであるが、Parzen推定の実行に多くの計算量が必要なことと、対象物体と背景に類似の色が含まれる場合に適切な推定が困難になるなどの解決すべき問題も残されている。

本研究では、MRFモデルを用いた移動体検出手法を提案する。本手法は、複数枚の背景画像を用いた最近傍法により動的な背景に対応しつつ、従来法よりも処理速度が向上した。また対象物体に対するフレーム間ブロックマッチングを取り入れることで、物体と背景に類似の色が含まれる場合にも対応することができる。

## 2. エネルギー関数

以下で、 $c_t^i$  は与えられた動画の  $t$  番目のフレームにおけるピクセル  $i$  の RGB 値を表す 3 次元ベクトルとし、また  $c^t = \{c_1^t, c_2^t, \dots, c_N^t\}$  ( $N$  は、ピクセルの総数) とする。移動体検出のために各ピクセルにラベル  $l_i^t \in \{0, 1\}$  を割り当てるのが最終的な目的であり、ピクセル  $i$  が移動物体であれば  $l_i^t = 1$ 、背景であれば  $l_i^t = 0$  となる。なお、誤解の恐れがない場合にはフレーム番号を表す添字  $t$  を省略することができるので注意されたい。

### 2.1 MRF の MAP 推定とエネルギー最小化

画像  $c^t$  が与えられた場合にラベル  $l^t$  の事後確率を

$$P(l^t | c^t) = \frac{1}{Z(c^t)} \exp\{-U^t(l^t; c^t)\} \quad (1)$$

で定義すると、検出結果はこの事後確率を最大化する (maximum a posteriori, MAP) 解として得られることになる。ここで、 $U^t(l^t; c^t)$  はエネルギー関数であり、 $Z(c^t)$  は規格化定数である。

ラベル  $l$  の同時確率 (1) から得られる各ピクセルに対する条件付確率  $p_i$  について

$$p_i(l_i | l \setminus l_i; c) = f_i(l_i | \{l_j\}; c) \quad (j \in \mathcal{M}(i)) \quad (2)$$

が満たされる場合に、このシステムをマルコフ確率場 (Markov random field, MRF) と呼ぶ。上式で  $\mathcal{M}(i)$  は、ピクセル  $i$  によって決まるピクセルの集合であり、 $f_i$  は集合  $\mathcal{M}(i)$  に属すピクセルのラベルを含む関数である。画像では、近傍ピクセル間の相関は一般に非常に強いため、マルコフ確率場を画像処理に応用する際は、 $\mathcal{M}(i)$  としてピクセル  $i$  周りの最近傍ピクセル集合 (4 近傍または、8 近傍) を用いる場合が多い。

事後確率 (1) に対する MAP 推定により得られる解は、エネルギー関数  $U^t(l^t; c^t)$  を最小にする解

$$l^{t*} = \underset{l^t}{\operatorname{argmin}} U^t(l^t; c^t) \quad (3)$$

と等しく、本研究において実際に最適化問題を解く際は、陽には確率を意識せずに 3 節で述べるグラフカットを用いて解を求めることができる。

本研究で用いるエネルギー関数は、2 つの項からなり

$$U^t(l^t; c^t) = \lambda \sum_{i \in \mathcal{P}} U_L^t(l_i^t; c_i^t) + \sum_{(ij) \in \mathcal{N}} U_S^t(l_i^t, l_j^t; c_i^t, c_j^t) \quad (4)$$

と表される。上式で、 $U_L^t, U_S^t$  は以下で詳しく述べるが、それぞれ物体/背景らしさに関する項とラベルに対するスムージング項である。記号  $\mathcal{P}, \mathcal{N}$  はそれぞれ全ピクセルの集合および最近傍 (4 近傍) ピクセル対の集合を表している。パラメータ  $\lambda (> 0)$  によってこれら 2 つの項の重みを調整する。

## 2.2 物体/背景らしさに関する項 $U_L^t$

### 2.2.1 複数背景画像を用いた最近傍法

背景差分における背景モデルとしては、木の葉や水面の揺れ、照明条件の変化、ファンや信号機のランプのような周期的な動きなどを含む動的な背景に十分に対応できるものが望ましい。1節で述べた様に、その目的のために混合ガウス分布や複数背景画像を用いた Parzen 推定が一般に使われている。しかし、精度のよい推定を行うためには、混合ガウス分布では適切な学習則、Parzen 推定では背景サンプル数に比例する和の計算が必要となり、計算時間の面では必ずしも望ましいとは言えない。

本研究では、 $t$  番目のフレームの処理時において、 $L$  (本稿記載の実験では、 $L = 50$ ) 枚の過去に観測された画像  $c^{t\alpha}$  ( $t_\alpha < t; \alpha = 1, 2, \dots, L$ ) を背景画像サンプルとして保持する。これらは、必ずしも連続した  $L$  個のフレームでなくてもよい。ここで、それら背景画像を  $b^\alpha = \{b_1^\alpha, b_2^\alpha, \dots, b_N^\alpha\}$  ( $\alpha = 1, 2, \dots, L$ ) と別の変数で書き直しておく。併せて、対応するフレーム  $t_\alpha$  での推定結果  $l^{t_\alpha}$  も保持しておく。この推定結果も同様に  $m^\alpha = \{m_1^\alpha, m_2^\alpha, \dots, m_N^\alpha\}$  と表す。

$t$  番目のフレームにおけるピクセル  $i$  が物体であるか背景であるかを判定する基準となる差分値  $d_i^t$  を次式で定義する。

$$d_i^t = \min_{\alpha \text{ s.t. } m_i^\alpha = 0} \|c_i^t - b_i^\alpha\|^2. \quad (5)$$

さらにこの値を閾値  $T_A$  と比較することで、 $U_L^t(l_i^t; c_i^t)$  を

$$\begin{aligned} U_L^t(l_i^t = 0; c_i^t) &= \Theta(d_i^t - T_A), \\ U_L^t(l_i^t = 1; c_i^t) &= 1 - U_L^t(l_i^t = 0; c_i^t) \end{aligned} \quad (6)$$

と定義する。ここで、 $\Theta(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$  となる階段関数である。以上の処理は、ピクセル毎に現在フレームとの差分値がもっとも小さくなる背景画像を  $L$  個のサンプルの中から探すことになり、最近傍法を行っていることになる。しかし、式 (6) を考慮すると、式 (5) の実行に常に  $L$  回の計算が必要となるわけではなく、閾値  $T_A$  を下回る値が見つければその段階で残りの計算を省略してよいことになり、このことが、計算量の節約に大きく寄与する。Parzen 推定では、ピクセル  $i$  が背景である確率  $p_i^B(c_i^t)$  を窓関数 (カーネル密度関数)  $K$  を使って

$$p_i^B(c_i^t) = \frac{1}{L} \sum_{\alpha=1}^L K(c_i^t - b_i^\alpha) \quad (7)$$

と推定することになり、常に  $L$  回の和が必要となるのに対し、本手法は  $U_L^t(l_i^t = 0; c_i^t) = 1$  となる場合のみが  $L$  回の計算を必要とし、移動体の占める領域が画像全体に対して比較的小さければ (現実の多くの画像が該当する)、計

算量の大幅な削減が期待できる。

なお、本研究では、動画像の最初の約  $2L$  枚を取り出し、ピクセル毎にそれら  $2L$  個のピクセル値をソートし、中央付近に位置する  $L$  個の値を初期背景サンプルとして用いた。以降は、数フレーム毎に、最も古い背景サンプルを捨てて現在フレームを背景サンプルとして取り入れた。

### 2.2.2 フレーム間ブロックマッチングに基づく移動情報の利用

多くの先行研究においては、移動物体が物体と類似色の背景領域に入ると検出精度が低下するという欠点がある。そのような状況においても移動物体の動き情報が推定できれば、それを手掛かりに望ましい検出が期待できる。そこで、本研究ではオプティカルフローの考え方から  $t$  番目のフレーム処理時には、 $t-1$  番目のフレームとの局所的なブロックマッチングを以下に述べるように行い、式 (6) で求めた  $U_L^t$  の値に動き情報の影響を付加する。

$t$  番目のフレームの推定を行う際には、1 フレーム前で  $l_i^{t-1} = 1$  と推定されたピクセル  $i$  に対してのみ次の処理を行う。

- 1)  $t-1$  番目の画像に対し、ピクセル  $i$  およびその上下左右の最近傍ピクセルを含めた 5 つのピクセルの RGB 値を並べた 15 次元ベクトル  $g_i^{t-1}$  を定義する。
- 2)  $t$  番目の画像に対し、ピクセル  $i$  を中心とする 1 辺  $w$  (本稿記載の実験では  $w = 11$ ) ピクセルの矩形領域  $\mathcal{W}(i)$  に存在するピクセルに対し、同様に 15 次元ベクトル  $g_j^t$  ( $j \in \mathcal{W}(i)$ ) を定義する。
- 3) 領域  $\mathcal{W}(i)$  の中から  $g_i^{t-1}$  に最も近い 15 次元ベクトルを持つピクセル  $j^*$  を

$$j^* = \operatorname{argmin}_{j \in \mathcal{W}(i)} \|g_j^t - g_i^{t-1}\|^2 \quad (8)$$

によって求める。

- 4)  $i \neq j^*$  かつ  $\|g_{j^*}^t - g_i^{t-1}\|^2 < T_B$  の場合に、

$$U_L^t(l_{j^*}^t = 0; c_{j^*}^t) \leftarrow U_L^t(l_{j^*}^t = 0; c_{j^*}^t) + \beta \quad (9)$$

のように、ピクセル  $j^*$  が対象物体として検出されやすいようにエネルギーを修正する ( $U_L^t(l_{j^*}^t = 1; c_{j^*}^t)$  には修正を加えない)。ここで、 $\beta (> 0)$  はあらかじめ決めた定数である。1 フレーム前と同じ位置のピクセルがマッチした場合には修正対象とはしないが、これは直前のフレームで誤って背景が物体と認識された場合にも、次のフレームではその誤りが周囲に波及することを防ぐ効果がある。

上で述べた最近傍ピクセルの RGB 値を含めた 15 次元ベクトルは、静止画のセグメンテーションにおいてもその有効性が報告されており、これによりある程度のテクスチャー情報を取り入れていると考えられる [9]。

### 2.3 スムージング項 $U_S^s$

与画像において近傍ピクセル同士の RGB 値が近い場合は、それらのピクセルのラベルも同様に近くなるのが予想できる。そのため、MRF の観点から以下のようなスムージングの効果をエネルギー関数に取り入れる。

$$U_S(l_i, l_j; c_i, c_j) = \frac{1}{\ln(\|c_i - c_j\|^2 + 1 + \epsilon)} |l_i - l_j|. \quad (10)$$

ここで、 $\epsilon \ll 1$  である。これまで、文献 [5]-[8] などにおいてスムージング項は導入されたが、上式のように係数にグラディエントの大きさを反映させてはなかった。しかし、静止画に対するセグメンテーション問題などでは、ピクセル間のグラディエントの大小によってスムージング項の影響に差をつけることの有効性が知られている [9]。

### 3. グラフカットによる最適化

式 (3) の最小化を  $l = \{l_1, l_2, \dots, l_N\}$  に対して同時に行うことは組合せ最適化問題であり、一見すると困難なように思われる。しかし、グラフ理論で培われ近年、画像処理にも盛んに用いられているグラフカットを利用すると大域的最適解を実用的な時間で求めることが可能となる。ここでは、簡単な説明に留めるが、グラフカットについての詳細な解説および画像処理への応用については、文献 [10] などを参照されたい。

グラフ  $G = (V, E)$  とは、頂点の集合  $V$  とそれら頂点間に存在する有向辺の集合  $E \subset V \times V$  からなる。辺  $(u, v)$  には、辺容量 (重みやコストなどとも呼ばれる) が付与されており、それを  $q(u, v)$  と表すことにする。頂点集合  $V$  の中から 2 点を選び、それらを特別視して source  $s$  と sink  $r$  と呼ぶ。ここで、集合  $V$  を 2 つの部分集合  $S \subset V, R = V \setminus S$  に  $s \in S, r \in R$  の条件を満たすように 2 分割することを考える。この時、分割のコストを source  $s$  側から sink  $r$  側へ向かう辺容量の和

$$Q(S, R) = \sum_{u \in S, v \in R} q(u, v) \quad (11)$$

として定義する。このとき、sink  $r$  側から source  $s$  側へ向かう辺容量は和に加えないことに注意する。グラフカット問題もしくは、最小カット問題とは、分割コストが最小になる分割の仕方を求める問題である。これは、組合せ最適化問題であるが、古くから頂点、辺の数に対して線形オーダーの計算量で解けることが知られており、その後洗練されたアルゴリズムが報告されている [6]。

画像とグラフの関係は、図 1 のようになる。各ピクセルが円で描かれた頂点に対応し、隣接ピクセル間に矢印で表現されたエッジが存在する。ピクセルに対応する頂点とは別に、四角で表された source  $s$  と sink  $r$  が用意さ

れ、source は移動物体に、sink は背景に対応する。最小カットの結果、頂点  $i$  が  $i \in S$  となれば、それはピクセル  $i$  が移動物体と推定されたことを意味する。

最小カットは、各辺に付与される辺容量に対する最適な分割結果を与える。従って、本来達成したい最適化問題 (3) の解が得られるようにグラフの辺容量を定めなくてはならない。以下の様に辺容量を定義すれば、最小カット問題の解は、式 (3) の解と一致することになる [11]。

$$q(i, j) = q(j, i) = U_S(l_i, l_j; c_i, c_j) \quad ((ij) \in \mathcal{N}), \quad (12)$$

$$q(s, i) = \lambda \cdot U_L(l_i = 0; c_i), \quad (13)$$

$$q(i, r) = \lambda \cdot U_L(l_i = 1; c_i). \quad (14)$$

本研究では、文献 [6] の高速なアルゴリズムを用いて、式 (3) の解を求めている。

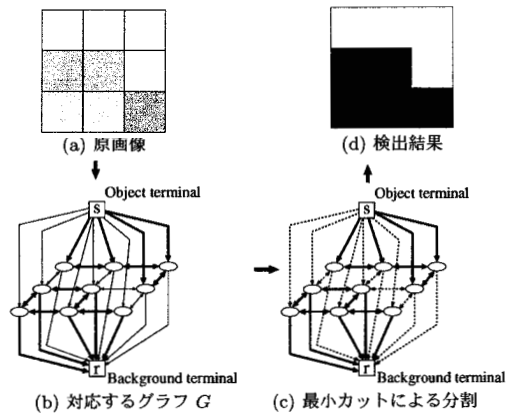


図 1 グラフカットによる移動物体検出の概略図。(a):この例では移動物体が白系、背景が黒系の色をしているとする。(b):各ピクセルは、グラフの頂点に対応し、最近傍ピクセル間には辺が配置される。線の太さは、辺容量の大きさを表しているとする。(c):破線で表される辺が、式 (11) の和に寄与し、これらの辺が取り除かれると頂点は 2 つの集合に分割される。(d):対応する推定結果。

### 4. 実験結果

本手法を用いた実験結果の例を図 2, 図 3 に示す。なお、パラメータ  $\lambda, T_A, T_B, \beta$  については、各々の実験で望ましい推定結果が得られるように手で調整しているが、フレーム間でそれらの値に変更は加えていない。また、比較対象として用いる単純背景差分では、ピクセル毎に各フレームの直前数十フレームのピクセル値に対して中央値を求め、それを背景画像として利用した。

図 2 は、2.2.2 節で述べた、フレーム間のブロックマッチングの効果を示す例である。図 2(a), (f) は、共に黒いズボンが同様に黒い背景領域に入った状況である。

図 2(b), (g) はブロックマッチングにより対象の移動情報を  $U_L$  に付加した場合の結果であり, 図 2(c), (h) はそれを行わない場合の結果である. 明らかに移動情報を考慮した場合の方が, 物体と背景の色が似ている下半身部分を適切に抽出していることが分かる. 図 2(d), (e) は, これら 2 つの場合のエネルギー関数第 1 項  $U_L$  の値を示している. 灰色で表されたピクセルが, 1 フレーム前に物体領域と推定されたピクセルとマッチしたピクセルを表しており, 図 2(e) において,  $U_L=0$  となっている腰から太股にかけての領域が図 2(d) ではブロックマッチングにより抽出されており, この移動情報を利用することで最終的に適切な推定が可能となったことが分かる. 図 2(i) に比較例として単純背景差分を用いた場合の結果を示した. 閾値を小さく取る程, 検出する物体領域は大きくなるものの, その分誤検出も多くなる. 特に, この画像では左上の部分に木の揺れがあり, 単純背景差分では, 各フレームにおいて使用する背景サンプルが 1 つであるため対応できていない. 図 2(j) は, 文献 [1] で提案された混合ガウス分布を用いて背景モデルを構築した場合の結果であり, 誤検出が少なくなるようにパラメータを設定したが, 逆に物体領域の検出もれを引き起こしている. それに対し, 本手法 (図 2(g)) では, 複数枚の背景画像を所持し, さらにグラフカットを利用することで適切にスムージングが働き誤検出がなく, 更にブロックマッチングの効果により人物領域を適切に推定している.

図 3 に示す動画は, 全体を通してカメラの揺れがあるが, 本手法では良好な結果が得られている. 比較例として, 図 2 の場合と同様に文献 [1] の方法を用いて, 今度は対象物体がなるべく多く検出できるようにパラメータを調整した場合の結果を示している. 紙面上では分かりにくい, 完全にはカメラの揺れに対応できずに, かなりの誤検出が生じている. この動画について Sheikh ら [8] と同様に正解データを利用して precision (物体と推定したピクセルに含まれる正解の割合) および recall (実際に物体であるピクセルの内, 正しく推定された割合) を求めた結果を図 4 に示す. これらは, 検出の過不足を測っており完全な推定がなされれば, 両方共が 1 となる. Sheikh らの方法と比較して, precision は優位であり, recall についても同等の結果が得られており, 本手法の有効性が示されたと言える. また, 文献 [8] の方法は Parzen 推定を用いているが, この画像 (画像サイズ:360×240) に対して毎秒 11 枚程度の処理速度であると報告されている. 一方で, 本手法では毎秒 25 枚の処理速度であり, フレーム間のブロックマッチングにも計算時間を要していることを考慮すると, 提案する最近傍法に基づく差分計算の計算量的有効性が窺える.

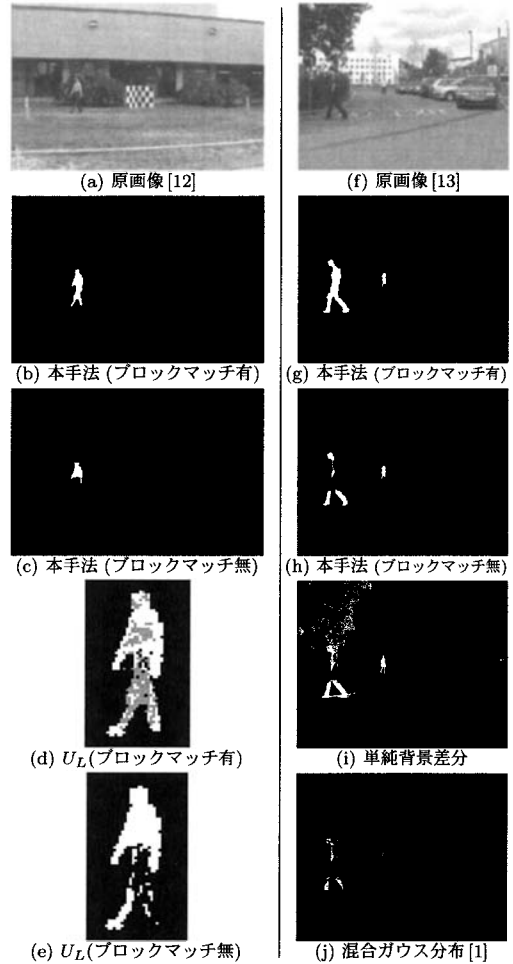


図 2 移動物体と背景が類似の色を持つ場合の例. (a, f):共に黒いズボンの人物が木の影と重なって, 境界が明瞭ではない状況である. (b, g):フレーム間のブロックマッチングに基づいて, 対象の移動情報を  $U_L$  に付加した場合の結果. (c, h): $U_L$  に対象の移動情報を付加しない場合の結果. (d, e):エネルギー関数の第 1 項  $U_L$  の値を図示しており, 灰色の箇所がブロックマッチングの結果, エネルギー関数に修正が加わったピクセルを示している. (i):単純背景差分の推定結果. 木々の揺れにより左上の領域に多くの誤検出が生じている. (j):混合ガウス分布で背景モデルを構築した場合の推定結果. 木の揺れには対応しているが逆に対象物体の検出もれがある.

## 5. おわりに

本稿では, 従来法と比較し性能面, 処理速度の面で有効な移動物体検出手法を MRF モデルに基づいて提案した. 本手法では, 複数枚の背景画像を利用した最近傍法により動的な背景に対応しつつ計算量を節約し, また対象物体に対してフレーム間ブロックマッチングを行うこ

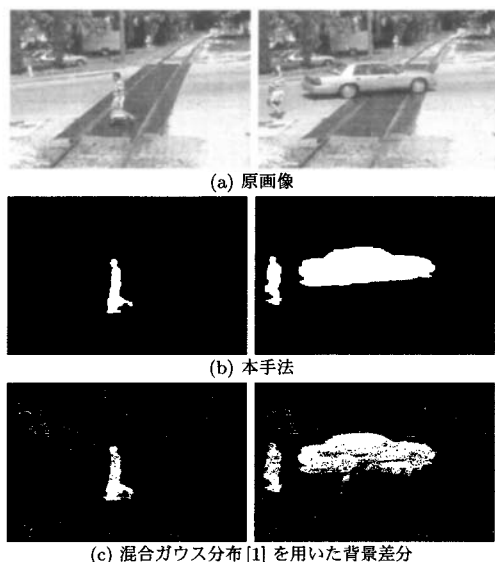


図3 カメラの揺れがある場合の例。本手法(b)は良好な結果を示している。混合ガウス分布による方法(c)は、なるべく対象物体が抽出されるようにパラメータを調整した場合の結果であり、誤検出が目立っている。

とで性能面での向上を実現した。

今後の課題としては、更なる性能向上のために、背景モデルの改良や詳細な動き情報の利用を考察することが挙げられる。また、単純なRGB空間ではなく、規格化RGB空間や影の除去で頻りに用いられているHSV空間の利用も視野に入れている。本稿ではパラメータ $\lambda, T_A, T_B, \beta$ は最適な性能が得られるように手動で調節したが、本来はこれらのパラメータについても与画像から推定すべきであり今後の研究課題の一つである。

### 謝 辞

本研究の遂行にあたり、都市エリア産学官連携促進事業の援助を受けたことに謝意を表する。

### 文 献

- [1] C. Stauffer, and W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *IEEE PAMI*, **22**, pp.747-757 (2000).
- [2] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE PAMI*, **19**, pp.780-785 (1997).
- [3] I. Haritaoglu, D. Harwood, and L. S. Davis, "W<sup>4</sup>: Who? When? Where? What? A Real Time System for Detecting and Tracking People," *Third Face and Gesture Recognition Conference*, pp.222-227 (1998).
- [4] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance," *Proceedings of the IEEE*, **90**, pp.1151-1163 (2002).
- [5] N.R. Howe, and A. Deschamps, "Better Foreground Segmentation through Graph Cuts," *eprint*

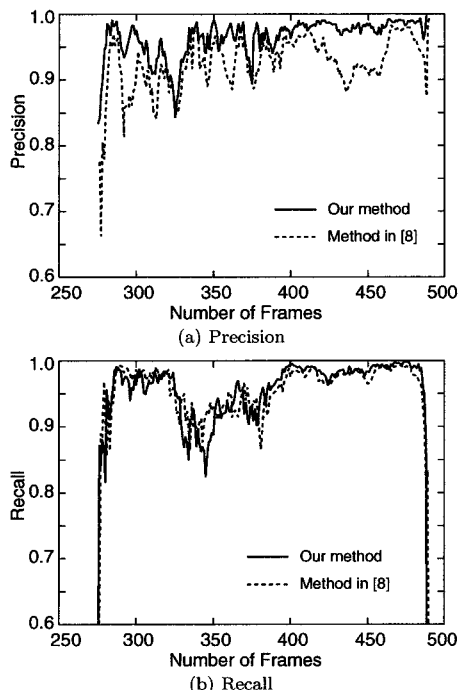


図4 定量評価。図3に示した動画像に対する precision(a) と recall(b)。270 フレームより前は、移動物体が存在しないためグラフを省略してある。

cs/0401017 (2004).

- [6] Y.Y. Boykov, and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *IEEE PAMI*, **26**, pp.1124-1137 (2004).
- [7] J. Migdal, and W.E.L. Grimson, "Background Subtraction Using Markov Thresholds," *IEEE Workshop on Motion and Video Computing*, pp.58-65 (2005).
- [8] Y. Sheikh, and M. Shah, "Bayesian Modelling of Dynamic Scenes for Object Detection," *IEEE PAMI*, **27**, pp.1778-1792 (2005).
- [9] T. Hosaka, T. Kobayashi, and N. Otsu, "Image Segmentation Using MAP-MRF Estimation and Support Vector Machine," *Interdisciplinary Information Sciences*, **13**, pp.33-42 (2007).
- [10] 石川博, 「グラフカット」, 情報処理学会 CVIM 2007 年 3 月, pp.193-204 (2007).
- [11] Y.Y. Boykov, and M.P. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images," *Proceedings of ICCV 2001*, pp.105-112 (2001).
- [12] S. Sarker, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer, "The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis," *IEEE PAMI*, **27**, pp.162-177 (2005).
- [13] J. Shutler, M. Grant, M.S. Nixon, and J.N. Carter, "On a Large Sequence-Based Human Gait Database," *Proc. Recent Advances in Soft Computing*, pp.66-71 (2002).