

背景モデル構築のための 画像類似性と時間的スパース性を考慮した画像選択法

清水 渚佐^{1,a)} 川西 康友² 棕木 雅之² 美濃 導彦²

概要：直近の観測画像を用いて背景モデルを構築する手法が、従来から多数提案されている。ここで、モデル構築に用いる観測画像の期間が短い場合、背景の変動は小さいため前景と背景を区別しやすいが、前景が連続して観測されると背景モデルに含まれてしまう。一方、期間が長い場合、前景の影響は小さいが、背景の変動が大きくなるため前景と背景の区別が難しい。本手法は、類似する画像を選択することにより背景の変動を抑え、かつ、互いに時間的に離れている画像を選択することにより前景の影響を抑えた背景モデルの構築を可能にする。実験では、直近の短期間/長期間の観測画像をそのまま用いた場合と提案手法により選択した観測画像を用いた場合について、背景モデルを構築し、背景差分の結果を比較することにより有効性を示す。

キーワード：背景差分，前景抽出，大域的变化，時間的スパース性，背景モデル

1. はじめに

観測画像中から、人物や車など前景領域を抽出する処理は、防犯カメラにおける前景物体の同定や認識のような高度な処理の前処理として重要である。背景差分は、前景の知識を必要としない前景領域抽出の手法であり、あらかじめ用意した背景画像と現在の観測画像の差分をとることで容易に前景領域を抽出することができる。しかし、屋外においては、1枚の背景画像を用意しただけでは、木々の揺らぎなどの局所的な変化や照明条件の変化などの大域的变化による影響を受けるため、様々なノイズが前景として抽出されてしまう。このような背景の変化に対応するため、複数枚の過去の観測画像から背景の変化をモデル化した背景モデルを構築し、背景差分を行う手法が多数提案されている [1]。また、時々刻々と起こる大域的变化に対応するため、背景モデルはオンラインで更新する必要がある。そのため、あらかじめ用意した背景モデルではなく、直近の観測画像を用いて背景モデルの構築が行われる [2]。

しかし、短期間の観測画像を用いた場合、直近の時間的に連続した観測画像が前景を含んでいると、前景も背景としてモデル化されてしまうため、検出漏れや誤検出が増え

る。また、照明変化のような大域的な変化が急に起こると、1画素単位では画素値の変化が大域的な変化によるものか前景によるものか判断できないため、画素値が大きく変化した画素が誤って前景として抽出される恐れがある。

急激な変化や前景の影響に対処するため、長期間の観測画像をモデルの構築に利用することが考えられる。長期間の観測画像の中には、様々な変化が含まれており、また、同じ場所に同じ前景物体が含まれた画像は全観測画像に対して十分に少ない。しかし、長期間の観測画像を用いて構築されたモデルでは、大域的に類似していない観測画像が混ざった状態で分布が構築され、背景と前景の区別が難しい。

そこで、本研究では、背景モデルの構築のために、観測画像ごとに、長期間の観測画像中から現在の観測画像と大域的に類似し、かつ、互いに時間的に離れている観測画像を選択する手法を提案する。大域的に類似した観測画像を選択することにより、現在の観測画像に応じたモデルを構築することができる。また、互いに時間的に離れている観測画像を選択することにより、同じ場所に同じ前景が連続して観測され、前景が背景モデルに含まれる問題を防ぐことができる。

2. 関連研究と問題点

従来、背景の大域的・局所的な変化に対処するため、複数枚の過去の観測画像を利用した前景領域抽出の手法が提案されている。

¹ 京都大学大学院 情報学研究科
Graduate School of Informatics, Kyoto University

² 京都大学 学術情報センター
ACCMS, Kyoto University

a) shimizu@mm.media.kyoto-u.ac.jp

まず、大域的な変化を扱うために、画素間の共起性に着目した手法がある。観測画像1枚を1つのベクトルとみなし、主成分分析によって背景の大域的な変化の部分空間を求め、その部分空間へ現在の観測画像を変換・逆変換することによって前景物体が除去された現在の背景画像を推定する Eigen Background 法が提案されている [3]。川端らは、部分空間を使った背景画像の推定と未知物体領域の推定を繰り返すことにより、実時間で未知物体領域を抽出する手法を提案している [4]。しかし、これらの手法では、局所的な変化に上手く対処することができない。

一方、局所的な変化を扱うために、単一の画素に着目した手法がある。最も単純なものとして、各画素に対して時間方向の画素値の平均やメジアンを背景の推定値として利用する手法がある [5][6]。これは、ほとんどの時間で観測されているのは背景であり、前景物体の観測は全体の中では十分に少ないため、平均やメジアンをとることで前景の影響を無視できるという考えに基づいている。また、観測画像から画素値の観測確率を混合正規分布やノンパラメトリックにモデル化し、背景差分を行う手法が提案されている [7][8]。これは、背景の変化は繰り返し何度も観測される局所的な変化であり、一方、前景は局所的に見ると一時的にしか観測されないことを利用している。現在の画素値の観測確率が閾値以上の画素は背景、そうでなければ前景と判定する。この手法は、木々の揺れや水面の揺らぎのような繰り返し観測される局所的な変化を上手くモデル化することができる。

これらの手法では、過去の観測画像を用いて背景モデルを構築しているが、直近の過去の観測画像をそのまま用いる場合には次のような問題点がある。図1、図2は、背景モデルの構築に短期間及び長期間の観測画像を用いる場合の問題点を示した図である。横軸 x はある画素の画素値、縦軸 $P(x)$ は画素値 x の観測確率を表している。

まず、短期間の観測画像を用いる場合、同じ前景が同じような位置に連続して含まれ、また、時間方向に関して前景を含む割合が高くなるため、前景も背景モデルに含まれるという問題がある。図1(a)のように、前景が連続して観測されると、本来は背景の画素値によって破線の分布のような背景モデルが推定されるはずが、実際は前景の画素値によって実線の分布のような間違った背景モデルが推定されてしまい、前景を背景と判定してしまう。また、急な大域的变化が起こったとき、背景の画素値が背景モデル構築時と現在で大きく異なり、背景モデルの更新が間に合わないために誤検出されるという問題がある。図1(b)のように、曇りから晴れへと急激な照明変化が起こると、同じ背景の画素でも曇りと晴れでは画素値は大きく異なるため、背景を間違っ前景と判定してしまう。

一方、長期間の観測画像を用いる場合、画素ごとに見ると前景が観測される確率は十分小さくなるため、前景が背

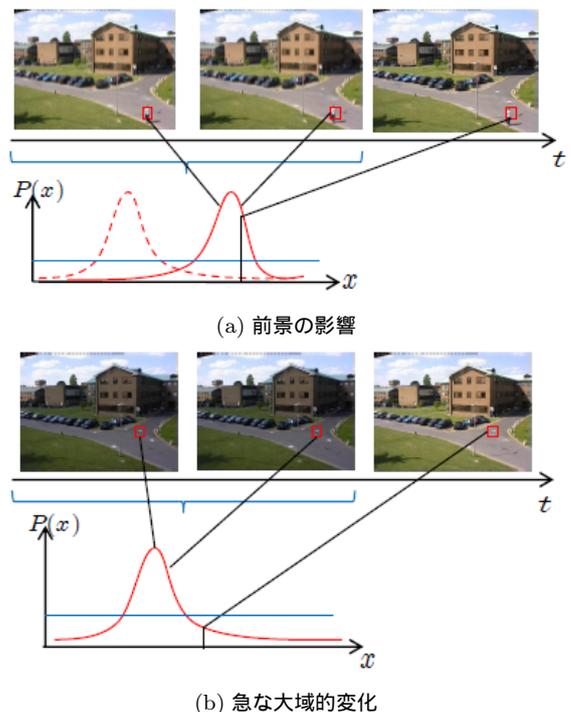


図1 短期間の観測画像を用いた場合の問題点

Fig. 1 The problems with images during a short period of time.

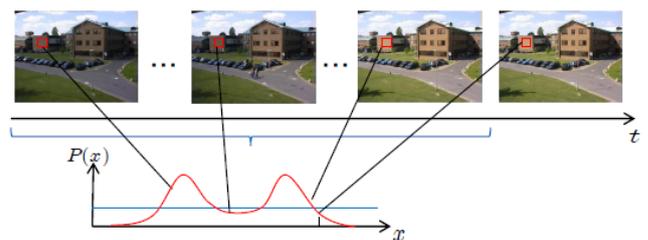


図2 長期間の観測画像を用いた場合の問題点

Fig. 2 The problem with images during a long period of time.

景モデルに与える影響を抑えることができる。しかし、背景の変化が大きくなるため、前景と背景の区別が難しくなるという問題がある。図2のように、曇りから晴れまで様々な照明条件が観測されていると、同じ背景の画素でも画素値の変動が大きく、判定を誤りやすくなる。

3. 背景モデル構築のための画像選択法

3.1 概要

従来手法では、過去の直近の観測画像をそのまま背景モデルの構築に利用していた。しかし、2章で述べたように、短期間の観測画像を用いる場合には前景の影響や急な大域的变化という問題があり、長期間の観測画像を用いる場合には大域的变化が大きいという問題がある。

本研究では、これらの問題に対し、現在の観測画像に類似していて、かつ、互いに時間的に離れた画像を選択して背景モデル構築を行うことにより、前述の問題を解決する(図3)。大域的な変化には、現在の観測画像と大域的に類



図 3 画像類似性と時間的スパース性を考慮した画像選択

Fig. 3 Select images based on image similarity and temporal sparsity.

似した画像を選択し、現在の観測画像に特化した背景モデルを構築することで対処する。前景の影響には、観測時刻が散らばるように観測画像を選択することで、同じ前景が同じような位置に連続して含まれないようにすることで対処する。

3.2 画像の類似性による選択

Kawanishi ら [9] は、複数日の同時刻に観測した画像は大域的な変化の状態が同じであると仮定しているが、規則正しい大域的な変化しか対応できず、制限が多い。

前景や局所的な変化は画像全体に対して十分小さく、画像を 1 つのベクトルとみなしたときのベクトル間類似度は、大域的な類似度を表していると言える (図 4)。しかし、次元数が非常に大きいため、計算コストが高い。

本研究では、主成分分析によって求めた低次元空間での近傍をとることにより、大域的に類似した画像の選択を行う。過去の観測画像に主成分分析を適用して低次元空間を求め、長期間の観測画像及び現在の観測画像を低次元空間に変換し、現在の観測画像との距離が近い観測画像 k_1 枚を選択する。

3.3 時間的スパース性による選択

観測時刻が互いに近い画像は大域的に類似した画像が多いため、画像の類似性のみに基づいて画像を選択すると、観測時刻が互いに近い画像がまとまって選択されやすい。時間的に連続した観測画像には、同じ前景が同じような位置に含まれている可能性が高く、構築された背景モデルに対する前景の影響が大きくなってしまう (図 5(a))。

本研究では、画像の類似性に基づいて選択された観測画像 (候補画像) の中から、最終的な選択結果が時間的にスパースとなるように絞り込みを行う (図 5(b))。候補画像から現在の観測画像と類似性が高い順に画像を取り出し、最終的な選択結果に含めるかどうかを判定する。取り出した画像の観測時刻とすでに最終的な選択結果に含まれることが決定したすべての画像の観測時刻が閾値以上離れていれば、その画像を最終的な選択結果に含める。

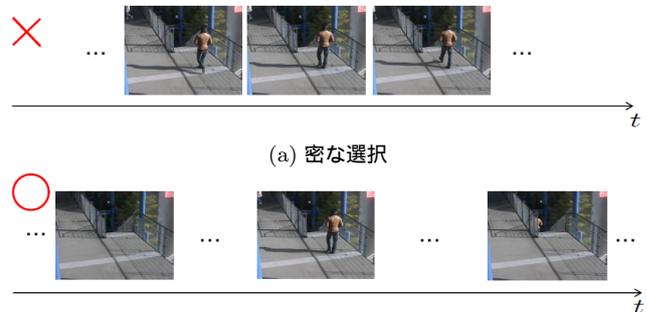
4. 提案する画像選択を用いた背景差分法

時刻 t における観測画像を 1 つのベクトルとみなした i_t



図 4 画像の類似性

Fig. 4 Image similarity.



(b) 疎な選択

図 5 時間的スパース性

Fig. 5 Temporal sparsity

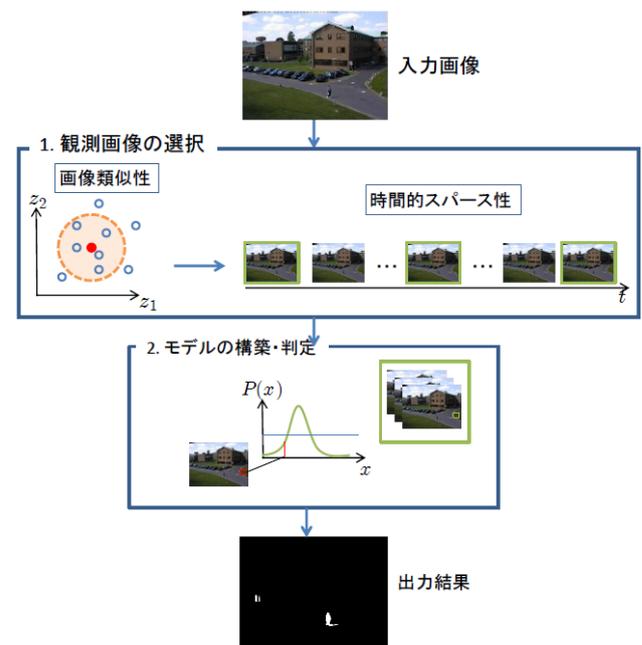


図 6 提案手法の流れ

Fig. 6 The flowchart of the proposal method.

は、背景の大域的な変化の差分ベクトル b_t^G 、背景の局所的な変化の差分ベクトル b_t^L 、前景による変化の差分ベクトル f_t を使って、次のように定式化できる。

$$i_t = b_t^G + b_t^L + f_t \quad (1)$$

提案手法は、大きく 2 つの処理に分けることができる (図 6)。

1. モデルの構築に使用する観測画像の選択
2. モデルの構築と背景・前景の判定

画像の選択及びモデルの構築は、新しく観測画像が入力されるたびに行う。また、画像選択の母集合となる長期間の観測画像は、直近の n 枚となるように随時更新する。

4.1 節, 4.2 節において、それぞれの処理について詳しく説明する。

4.1 観測画像の選択

観測画像の選択は、次の手順で行う。

Step1: 画像の類似性に基づいて候補画像を選択

Step2: 時間的なスパース性に基づいて画像を絞り込み

画像の類似性と時間的なスパース性に基づく観測画像の選択アルゴリズムを以下に示す。

Step0: 事前処理

- (1) 長期間の観測画像集合 S_0 に主成分分析を適用し、変換行列 U を求める。
- (2) S_0 を低次元空間に変換し、集合 Z とする。

$$Z = \{z_r \mid z_r = U^T i_r, i_r \in S_0\} \quad (2)$$

Step1: 画像の類似性に基づく選択

新しく観測画像 i_t が入力されるたびに、以下の処理を行う。

- (1) 観測画像を低次元空間に変換する。

$$z_t = U^T i_t \quad (3)$$

- (2) z_t の近傍 k_1 点 $\{z_s^{(t)}\}_{s=1}^{k_1}$ を Z から選択する。
- (3) 各点に対応する観測画像 $\{i_s^{(t)}\}_{s=1}^{k_1}$ をユークリッド距離の昇順に並べ、候補画像列 S_1 とする。
- (4) S_0, Z を更新する。

$$S_0 \leftarrow S_0 \cup \{i_t\} - \{i_{t-n}\} \quad (4)$$

$$Z \leftarrow Z \cup \{z_t\} - \{z_{t-n}\} \quad (5)$$

Step2: 時間的なスパース性に基づく選択

- (1) パラメータ s と選択結果の集合 S_2 を初期化する。

$$S_2 \leftarrow \emptyset \quad (6)$$

$$s \leftarrow 1 \quad (7)$$

- (2) S_1 の中から s 番目の画像 $i_s^{(t)}$ を 1 つ取り出す。
- (3) $i_s^{(t)}$ と S_2 内の全観測画像の観測時刻の差が閾値以上であれば、 $i_s^{(t)}$ を S_2 に追加する。
- (4) $S_1 = \emptyset$ もしくは $|S_2| = k_2$ (k_2 は最終的な選択数) ならば終了。そうでなければ、 $s \leftarrow s + 1$ として (b) に戻る。

このようにして得た選択結果の集合 S_2 を 4.2 節のモデル構築に利用する。

4.2 モデルの構築と背景・前景の判定

4.1 節の処理によって現在の観測画像に大域的に類似した画像が選択されているため、 b_t^G の項を考慮する必要がなく、残る $b_t^L + f_t$ についてのみ考えればよい。背景の変化は頻繁に観測されるが、時間的なスパース性により前景による変化が観測されることは少ない。そのため、 $b_t^L + f_t$ の値の大部分は b_t^L によるものである。

以下、局所的な変化についてのみ考慮すればよいため、ある一つの画素に注目した場合について述べる。画素値 x の観測確率 $P(x)$ を単純なガウス分布によってモデル化し、背景モデルとする。ここで、 μ, σ^2 はそれぞれ画素値 x の平均、分散とする。

$$P(x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_t - \mu)^2}{2\sigma^2}\right) \quad (8)$$

$$\mu = \frac{1}{k_2} \sum_{i=1}^{k_2} x_i \quad (9)$$

$$\sigma^2 = \frac{1}{k_2} \sum_{i=1}^{k_2} (x_i - \mu)^2 \quad (10)$$

提案手法では、大域的に類似した画像を選択しているため、単純なモデルで十分であると考えられる。また、単純なモデルを用いることで、計算時間も短縮することができる。

背景・前景の判定は、画素値の観測確率が閾値 θ より大きい画素は背景、それ以外の画素は前景とする。

5. 実験

5.1 実験データ

実験には、照明変化が大きい PETS2001 のデータセット*1及び商業施設の固定カメラの映像を用いた。画像サイズはどちらも 320×240 であった。

5.2 画像選択結果

まず、提案手法による画像選択の結果を調べた。観測画像の選択に関するパラメータは、低次元空間の次元数 $d = 5$ 、候補画像の数 $k_1 = 200$ 、背景モデルの構築に利用する選択結果の数 $k_2 = 20$ とした。また、低次元空間の学習には 500 枚の画像を用いた。

直近の観測画像 (短期間, short), 直近から 20 枚ずつ飛ばして選択した観測画像 (長期間, long), 画像の類似性のみに基づいて選択した観測画像 (similarity), 画像の類似性と時間的なスパース性に基づいて選択した観測画像 (similarity&sparsity) の例を図 7 に示す。

短期間の観測画像では、同じ前景が同じような位置に連続して含まれている。一方、長期間の観測画像では、同じ前景は含まれていないが、日照条件の変化による影や明るさの変化が大きい。また、画像の類似性のみに基づいて選

*1 ftp://pets.rdg.ac.uk/PETS2001/より入手可能。



図 7 背景モデルの構築に利用する画像の例
Fig. 7 The examples of images used for background modeling.

択した観測画像でも、短期間の場合と同様に、前景が連続して含まれた画像が選択されている。これら 3 つに対し、画像の類似性と時間的スパース性に基づいて選択した観測画像では、大域的に類似し、かつ、前景が同じような位置に含まれない画像が選択できている。

5.3 背景差分結果

5.2 節と同様に、4 つの観測画像からそれぞれ構築した背景モデルによる背景差分の結果を定性的・定量的に比較した。

5.3.1 定性的評価

まず、定性的な評価を行った。前景と判定された画素を白色、背景と判定された画素を黒色として表示した。結果を図 8 に示す。

上段 2 つの結果では、画像類似性と時間的スパース性に基づいて選択した観測画像を用いた結果が最も優れている。短期間の観測画像を用いた場合は、右下の人物の周辺の背景領域が前景と誤判定されている。これは、短期間の観測画像集合では、同じ人物が同じような位置に含まれているため、前景も背景モデルに誤って含まれてしまったためである。また、長期間の観測画像を用いた場合は、奥にある建物が前景と誤判定されている。これは、長期間の観測画像集合では、様々な照明条件の観測画像が含まれているため、特に照明変動が大きい部分で適切な背景モデルを構築できなかったためである。さらに、画像類似性のみに基づいて選択した観測画像を用いた場合も、前景が同じように含まれている画像がまとめて選択されてしまった

めに、前景が背景モデルに含まれてしまっている。

一方、下段の結果では、長期間の観測画像集合を用いた場合に比べて、提案手法において背景を前景と誤判定している部分が多い。これは、画像全体としての類似性に基づいて選択しており、大域的な変化が一様でない場合、画像中に占める割合が小さい部分が反映されにくいためであると考えられる。

5.3.2 定量的評価

次に、Ground Truth が公開されている^{*2}PETS2001 のデータセットを用いて、定量的な評価を行った。背景差分の性能の評価には、適合率 (Precision)、再現率 (Recall)、F 値を用いた。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

$$F = 2 / \left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}} \right) \quad (13)$$

ただし、TP (True Positive) は正しく前景と判定された画素数、FP (False Positive) は背景を前景と誤判定した画素数、FN (False Negative) は前景を背景と誤判定した画素数である。実験結果を表 1 に示す。

短期間の観測画像を用いた場合に比べて、すべての値で提案手法の方が高い値を示した。これは、提案手法の方がより適した背景モデルを構築することができたためであると考えられる。一方、長期間の観測画像を用いた場合、再現率は高いが、適合率が非常に低い。これは、モデルの構

^{*2} <http://limu.ait.kyusyu-u.ac.jp/dataset/> で公開されている。

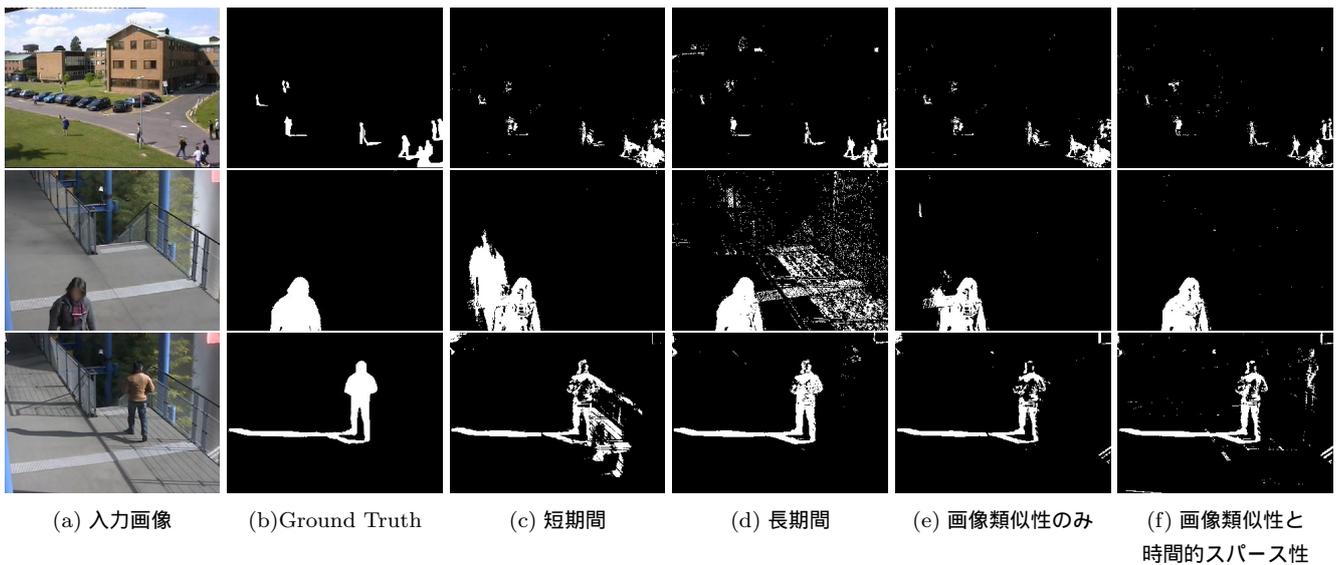


図 8 定性的評価

Fig. 8 Qualitative evaluation.

表 1 定量的評価

Table 1 Quantitative evaluation.

	short	long	similarity	similarity& sparsity
Precision	0.631	0.152	0.637	0.672
Recall	0.469	0.590	0.469	0.541
F	0.538	0.241	0.540	0.610

築に用いる観測画像の中に様々な照明条件の画像が含まれており、背景モデルの分散が大きくなってしまったためであると考えられる。

また、画像類似性のみに基づいて選択した観測画像を用いた場合は、短期間の観測画像を用いた場合とほぼ同程度の値を示した。

6. おわりに

本研究では、背景モデルの構築のために、画像の類似性と時間的なスパース性を考慮した画像選択法を提案した。背景差分の実験結果から、背景の大域的な変動と前景の影響を抑えた背景モデルが構築できることを確認した。

今後の課題として、まず、一様でない大域的な変化への対応が求められる。これは、画像をブロックに分割し、ブロックごとに画像を選択することにより解決できると考えている。また、背景モデルの構築方法や背景/前景の判定処理について検討が必要である。今回は単純なガウス分布で画素値の観測確率のモデル化を行ったが、混合ガウス分布やノンパラメトリックな手法を利用することが考えられる。背景/前景の判定処理については、今回は固定閾値処理によって判定を行ったが、画素値の分散などを考慮して画素ごとに適応的に閾値を変更するといった工夫が考えられる。

謝辞 本研究は文科省、安全・安心な社会のための犯罪・

テロ対策技術等を実用化するプログラム「環境適応型で実用的な人物照合システム」の一環として実施したものである。

参考文献

- [1] Bouwmans, T.: Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey, *Recent Patents on Computer Science*, Vol. 4, No. 3, pp. 147–176 (2011).
- [2] Stauffer, C. and Grimson, W.: Adaptive background mixture models for real-time tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 246–252 (1999).
- [3] Oliver, N., Rosario, B. and Pentland, A.: A bayesian computer vision system for modeling human interactions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 831–843 (2000).
- [4] 川端聡, 日浦慎作, 佐藤宏介: 動的背景の学習による未知物体領域の実時間抽出法, 電子情報通信学会論文誌 D, Vol. J89-D, No. 4, pp. 826–835 (2006).
- [5] Lee, B. and Hedley, M.: Background Estimation for Video Surveillance, *IVCNZ 2002*, pp. 315–320 (2002).
- [6] McFarlane, N. and Schofield, C.: Segmentation and tracking of piglets in images, *Machine Vision and Applications*, Vol. 8, No. 3, pp. 187–193 (1995).
- [7] 島田敬士, 有田大作, 谷口倫一郎: 適応的な分布数の増減法を利用した混合ガウス分布による高速な動的背景モデル, 電子情報通信学会論文誌 D, Vol. J90-D, No. 9, pp. 2606–2614 (2007).
- [8] 田中達也, 島田敬士, 有田大作, 谷口倫一郎: ノンパラメトリックな動的背景・影モデルに基づいた映像からの物体抽出, 情報処理学会研究会報告.CVIM, Vol. 159, No. 12, pp. 105–112 (2007).
- [9] Kawanishi, Y., Mitsugami, I., Mukunoki, M. and Minoh, M.: Background image generation by preserving lighting condition of outdoor scenes, *Procedia - Social and Behavioral Science*, Vol. 2, No. 1, pp. 129–136 (2010).