

複雑な背景における物体検出法の検討

伊谷 裕介[†] 渡辺 裕[†]

[†] 早稲田大学大学院 国際情報通信研究科

〒 367-0035 埼玉県本庄市西富田大久保山 1011

E-mail: †y-itani@tom.comm.waseda.ac.jp, ††hiroshi.watanabe@waseda.jp

あらまし 近年、セキュリティへの不安から監視カメラを用いた映像監視システムに対する需要が高まっている。現状では、集中監視センターなどに伝送されて監視員が表示映像を確認する例が多い。人間に代わって異常を発見するための様々な画像認識技術が開発されており、機械による自動監視に対する需要は極めて高い。自動監視システムの要素として物体検出があるが、特に屋外環境において木のゆれなど背景変動のある複雑な背景に対し検出精度が落ちることが問題として挙げられる。本稿では背景変動がある環境における検出精度の向上をめざす。提案手法では、まず、色情報と動きベクトルの向きの類似度により領域に分ける。そして、領域ごとにフレーム間の輝度値の差分和をとり、しきい値処理により動物体を検出する。最後に実験により提案手法の有効性を示す。

キーワード 物体検出, 監視システム

A Study on Object Detection in Complex Background

Yusuke ITANI[†] and Hiroshi WATANABE[†]

[†] Graduate School of Global Information and Telecommunication Studies, Waseda Univ.,

1011 Okuboyama Nishi-Tomida Honjo-shi Saitama 367-0035 Japan

E-mail: †y-itani@tom.comm.waseda.ac.jp, ††hiroshi.watanabe@waseda.jp

Abstract Recently, a demand for video surveillance system using surveillance cameras has been increasing because of the uneasiness to secure the public peace. At present, It is often the case that people look video transmitted to a monitoring center. A variety of video recognition technologies to find abnormality are developed, and demand for a smart surveillance by the machine is increasing. Though object detection is an important element of smart surveillance system, the detection accuracy is often degraded in a complex background. This kind of situation can be seen in the environmental change like shakes of the tree. In this paper, we propose a method of object detection in such a complex environment. First, regions in the image is divided according to a similar level of the color information and the direction of the motion vector. Then, the difference of the luminance value in each area between frames is calculated. Finally, foreground regions are detected using a threshold. Experimental results show that the proposal method can detect a moving object effectively.

Key words object detection, surveillance

1. ま え が き

近年、安全・安心を目的として、監視カメラを用いた映像監視が注目されている。[1]～[3] 監視カメラにより得られた画像は集中監視センターなどに伝送されて監視員が表示映像を確認する例が多い。しかし、光ファイバの敷設などにより多くの画像が伝送可能となり、一人が担当するモニタ数が多くなったため、監視作業の負担が大きい[4]。人間に代わり自動的に異常を発見するための様々な画像認識技術が開発されており、機械による

自動監視に対するニーズは極めて高い。

監視カメラによる映像監視の流れを図1に示す。映像監視ではまず、映像から動物体を検出し、追跡する。そして、その物体が何であるか分類し、物体の構造を把握する。この流れ中で、物体検出は第一段階である。物体検出は木の揺れなどの背景変動や急激な照明変化などの環境の変化など様々な要因により、識別に影響が出る。[6] そのため、実際の観察現場における様々な状況変化の下で安定して観察対象を発見・追尾・確認できるような手法が求められる。

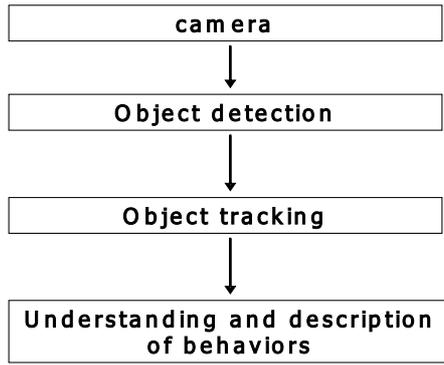


図 1: 監視カメラの処理の流れ

特に重要施設の映像監視セキュリティにおいて、見逃し（検出漏れ）が許されないと同時に日照条件や木の揺れ、光の乱反射など環境の変動に起因する誤検出を低減することが求められる。

そこで、本稿では木の揺れなどの背景変動に着目し、背景変動に対してロバストに物体を検出する手法について検討する。提案手法は、まず色情報と動きベクトルを用いて領域に分割し、領域ごとの輝度の差分和を用いて物体を検出する。実験により、提案手法の有効性を示す。

2. 従来手法と問題点

本節では、物体検出に関する手法の従来手法と問題点について述べる。自動監視のための物体検出として、代表的な手法に背景差分法 [7] やモデルを用いた手法 [5] が挙げられる。背景差分法は、あらかじめ移動物体の存在しない背景画像を取得し、入力画像と背景画像との差分処理結果に応じて移動物体検出する手法である。この手法は木のゆれや窓や水による光の乱反射など環境変化や複数人物がいる状況など複雑な環境に対するロバスト性が十分でない [3]。また、モデルを用いる手法は、人の全身や頭部などのモデルを用意し、画像中で人の候補領域が存在する確率により移動する人を検出する手法である。この手法は体の一部が遮蔽されている場合に検出精度が低下する。

複雑背景に対する手法として、画像中に現れる局所的な動きに着目し、動物体の動きとその他背景変動を分離する手法が提案されている [7]。図 2 に示すように、まず背景差分により検出した領域 R_j の中に割り当てた各ブロックごとに輝度差分の絶対値総和 (SAD) を計算する。この値を用い領域ごとに代表速度を求め時間軸をさかのぼり追跡する。このとき、

- SAD の値、
 - 全 L 回繰り返した際の SAD の総和の最小値、
 - 各追跡過程で得られる変化領域の代表速度が 0 でない回数、
- を特徴量として定義し用いる。

この手法では代表速度の検出精度が悪い場合、違った箇所を時間軸をさかのぼりながら追跡することになり誤検出の原因となる。また、フレーム間の追跡で得られた代表速度が一定でなくとも、時間平均の特徴量が小さくなり誤検出の原因となる。

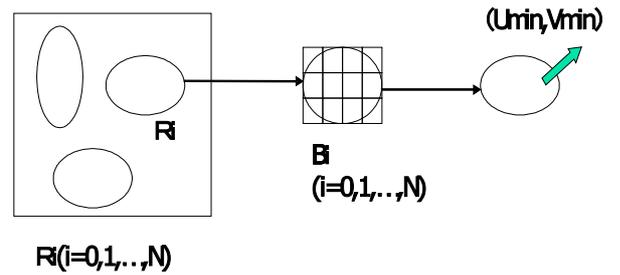


図 2: 従来手法

3. 予備実験

3.1 特徴量の定義

本節では、背景変動と動物体を分類するために、必要となる特徴について述べる。図 3 のように、木が左から右側に動いた場合を考える。この場合、左上のブロックは木のない領域に対し木が入る変化が起きる。そのため、輝度値が下がる変化である。一方で、右下のブロックは木がなくなる変化が起きる。そのため、輝度値が増える変化である。つまり、左上のブロックと右上のブロックとは全く逆の変化が起きていることがわかる。これより、時間 t のフレームの (x, y) 座標における輝度を $f(x, y, t)$ とすると、木などの揺れの場合、注目画素の輝度値 $f(x, y, t)$ の時間方向の差分 $f(x+i, y+j, t) - f(x+i, y+j, t-1)$ をその周囲で和をとると 0 に近づくことが予想される。

$$\sum_{i,j} f(x+i, y+j, t) - f(x+i, y+j, t-1) > TH \quad (1)$$

$$\sum_{i,j} f(x+i, y+j, t) - f(x+i, y+j, t-1) < TH \quad (2)$$

つまり、式 (1) のときは動物体、式 (2) のときは背景であると考えられる。また、木の揺れなどの周期性を持つ変化の場合、時間方向の変化を考えたとき、その領域の輝度の差分和の変化はほとんどないと考えられる。

ここで、式 (1) (2) はある領域を規定したときに、その領域の差分和を考えている。そのため、領域ごとに輝度を平均した画像を考え、その差分を考えることと同じである。

$$\frac{1}{N} \sum_{i,j} f(x+i, y+j, t) - f(x+i, y+j, t-1) \quad (3)$$

$$= \sum_{i,j} \frac{f(x+i, y+j, t) - f(x+i, y+j, t-1)}{N} \quad (4)$$

$$= \sum_{i,j} \frac{f(x+i, y+j, t)}{N} - \sum_{i,j} \frac{f(x+i, y+j, t-1)}{N} \quad (5)$$

フレーム間差分と、領域ごとに平均化したフレーム間差分の性質の違いを見るために予備実験を行った。実験では、フレームを 8×8 のブロックに分け、ブロックごとに輝度値を平均化した。実験に用いた原画像を図 4 に示す。

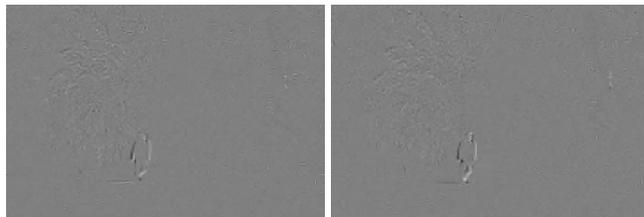


図 3: アプローチ



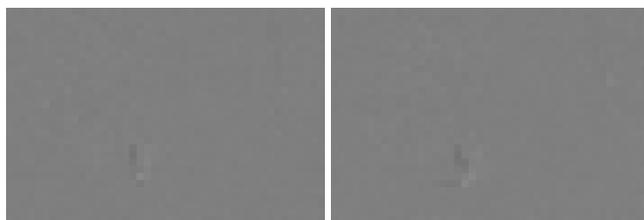
(a) 一枚目 (b) 二枚目

図 4: 原画像



(a) 一枚目 (b) 二枚目

図 5: フレーム間差分



(a) 一枚目 (b) 二枚目

図 6: 平均画像のフレーム間差分

3.2 予備実験結果

図 5, 6 に単純なフレーム間差分の結果と平均画像を用いたフレーム間差分の画像の結果を示す。図 5, 6 より、フレーム間差分の画像は木の揺れによる変化が目立つのに対し、平均画像を用いた場合のフレーム間差分はそれが目立たない。そして人の領域は平均画像の差分であっても目立つことが確認できる。

また、その際の木の人領域、人が動いた領域、木などが背景におけるフレーム間の輝度の差分値のグラフを図 7, 8 に示す。図 7, 8 より、人の動きの変化に比べ、木の揺れや背景のほうが輝度の変化が小さいことが確認される。

また、平均画像の輝度差分値を時間方向に 3 フレーム分足したもののグラフを図 9 に示す。これより、時間方向の和も考慮することで、より精度良く分類が可能になることがわかる。

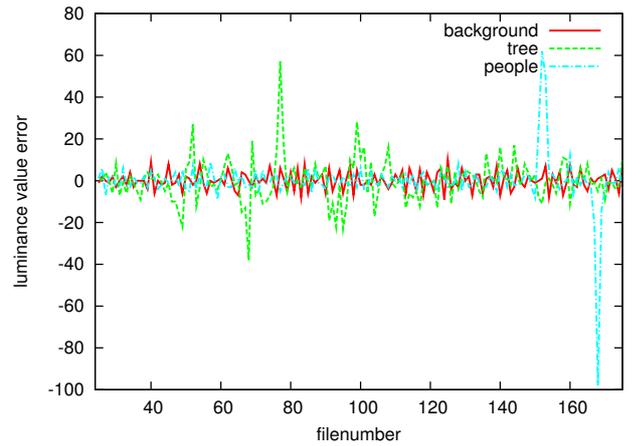


図 7: フレーム間差分の輝度変化

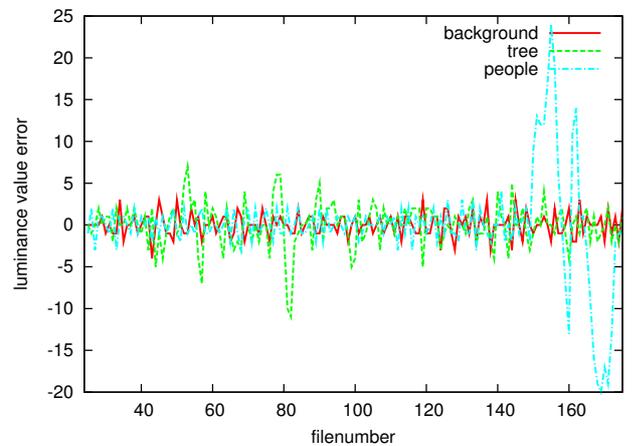


図 8: 平均画像の輝度変化

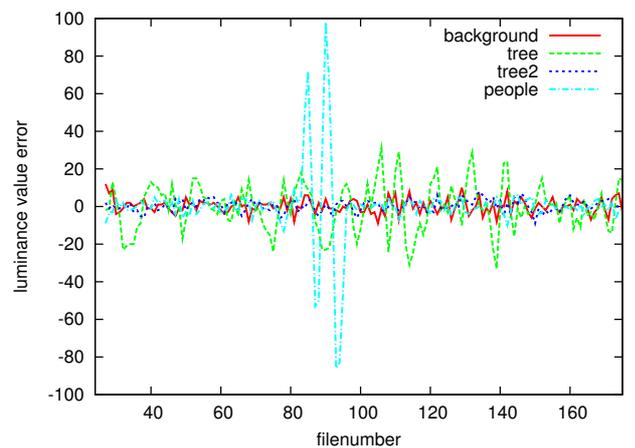


図 9: 平均画像の輝度変化

4. 提案手法

本節では、前節の予備実験の結果を踏まえ、提案手法について述べる。提案手法の流れを図 10 に示す。提案手法では、まず領域分割を行う。領域分割は以下の手順で行う。

- 8×8 のブロックにわけると。
- 隣接ブロックを比較し、濃度平均がしきい値 TH_{node} 以下の場合は統合する。

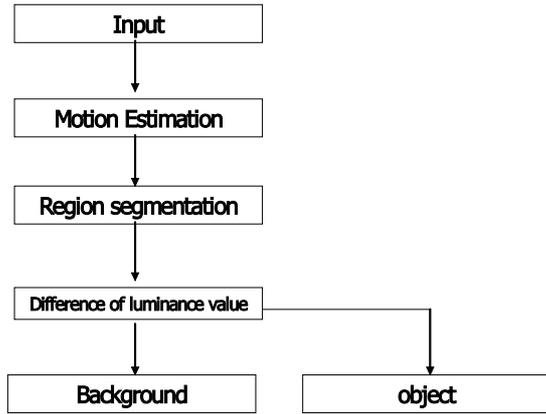


図 10: 提案手法の流れ

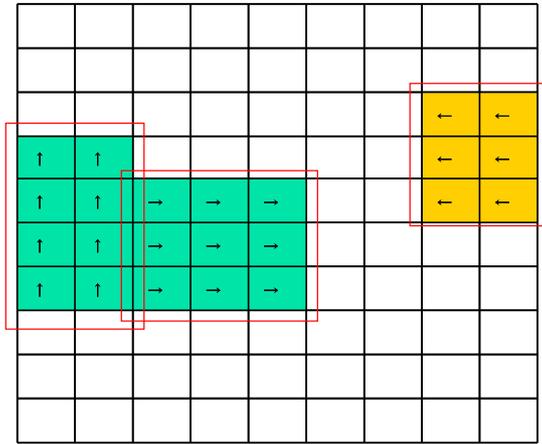


図 11: 領域分割

$$\frac{1}{N} \sum_{i,j} f(x, y, t) - f(x + i, y + j, t) > TH_{node} \quad (6)$$

- 動きベクトルの向きが類似している場合は、同じ領域として統合する。

$$\arg MV < TH_{arg} \quad (7)$$

今回、しきい値 TH_{arg} は 60 度とした。

領域分割の例を図 11 に示す。図 11 のように、色と動きの向きが似ている領域を統合する。

領域分割の後、領域ごとに輝度の差分値の和を計算する。

$$\sum_{i,j} f(x + i, y + j, t) - f(x + i, y + j, t - 1) \quad (8)$$

その差分値を 3 フレーム分の和を求める。そして領域内のブロック数 N で割り平均をとり、しきい値 TH により、動物体と背景に分類する。

$$\frac{1}{N} \sum_{i,j,k} f(x + i, y + j, t) - f(x + i, y + j, t - k) > TH \quad (9)$$

$$\frac{1}{N} \sum_{i,j,k} f(x + i, y + j, t) - f(x + i, y + j, t - k) < TH \quad (10)$$

つまり式 (9) を動物体 (10) を背景とする。

5. 実験, 考察

本節では、提案手法の有効性を示すための実験とその結果について述べる。

実験条件を表 1 に示す。

表 1: 実験条件

フレームレート	15fps
解像度	CIF
カメラ	固定

実験結果の画像を図 12 に示す。木が揺れている領域に注目すると、その領域での検出はほとんど見られず、木の揺れによる誤検出を減らすことができている。また、検出すべき動物体の領域に注目すると、人がいる領域をきちんと検出している。

また、フレームごとに再現率と適合率を求めたグラフを図 13 に示す。提案手法による過検出、未検出の例を図 14,15 に示す。再現率と適合率の式を示す。

$$\text{再現率} = \frac{\text{正しく検出された動物体ブロック数}}{\text{正解マスク中のブロック数}} \quad (11)$$

$$\text{適合率} = \frac{\text{正しく検出された動物体ブロック数}}{\text{検出されたブロック数}} \quad (12)$$

正解マスクはあらかじめマニュアルで作成した。再現率についてはフレーム平均で 93% であり、検出すべき領域が検出できていることがわかる。一部再現率が低下しているフレームがあるが、これは図 15 のように、頭や手など人の一部が欠けた場合である。そのため、動物体が存在するという情報の漏れは無い。適合率はフレーム平均で 85% である。再現率に比べ低いが、これは図 14 のように動物体の周りの背景領域を検出したためである。

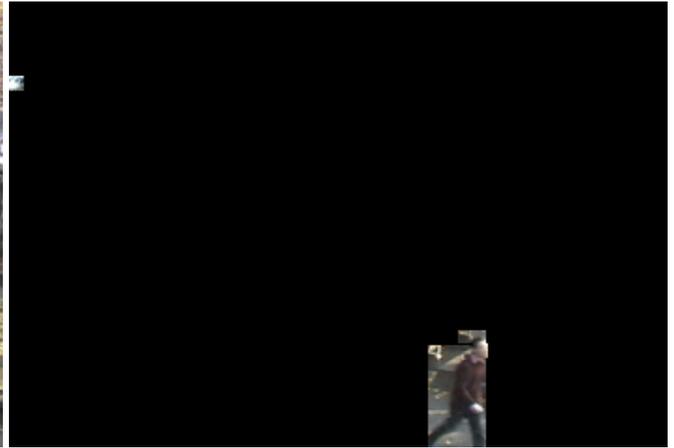
今回の目的は動物体の検出であるため、背景領域を含んでいたり、多少欠けていても問題はないと考えられる。そのため、提案手法は有効であると考えられる。

6. む す び

本稿では、背景変動のある映像での物体検出に注目した。提案手法では、まず色情報の類似性と動きベクトルの向きの一様性に注目して領域分割を行う。そして分けられた領域内の輝度を平均化した画像を考え、フレーム間差分を行うことで動物体を検出する。実験結果により、背景変動がある映像に対して精度良く動物体を検出できることを示した。



(a) 原画像



(b) 検出結果

図 12: 実験結果

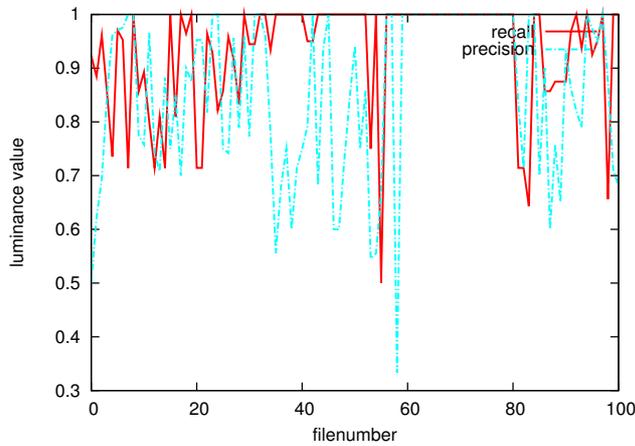


図 13: 再現率と適合率

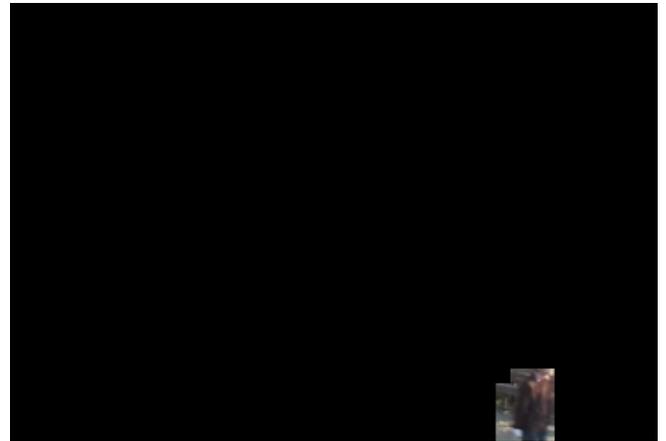


図 15: 未検出

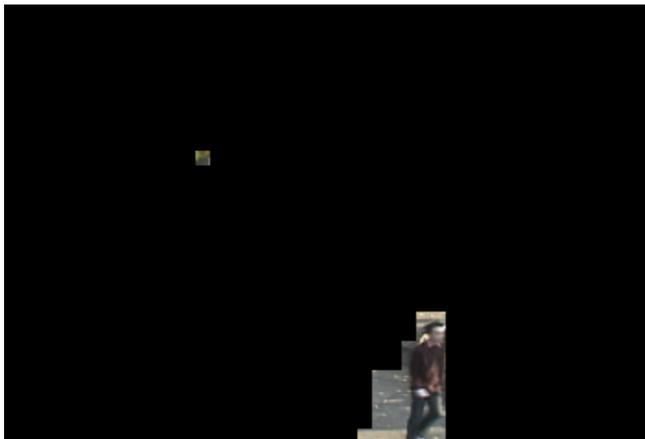


図 14: 過検出

Maybank, "A Survey on Visual Surveillance of Object Motion and Behavior," IEEE transactions, Man and Netics-Partic:Apprications and Reviews, Vol.34, No.3, Aug. 2004.

- [4] 小沢, "サーベイランスの現状と今後," 映像情報メディア学会誌, vol.57, NO.9, pp.1056-1057, Sep.2003.
- [5] Tao Zhao, and Ram Nevatia, "Tracking Multiple Humans in Complex Situations," IEEE transactions on pattern analysis and machine intelligence, vol.26, No.9, Sep. 2004.
- [6] 鷺見, 関, 波部, "物体検出-背景と検出のモデリング-, " 情報処理学会 CVIM 研究会, pp.79-98, Sep.2005.
- [7] 羽下, 鷺見, 八木, "変化領域内の動きの時空間特徴に着目した屋外情景における歩行者の検出," 信学論 D-II, Vol.J87-D-II, No.5, pp.1104-1111, May. 2004.

文 献

- [1] Arun Hampapur, Lisa Brown, Jonathan Connell, Ahmet Ekin, Norman Haas, Max Lu, Hans Merkl, Sharath Pankanti, Andrew Senior, Chiao-Fe Shu, and Ying Li Tian, "Smart video Surveillance," IEEE signal processing magazine, pp38-51, Mar. 2005.
- [2] 橋本, "映像監視における CV 実用技術の現状と課題," 情報処理学会 CVIM 研究会, Nov.2005.
- [3] Weiming Hu, Tienju Tan, Liang Wang, and Steve