

移動ビデオカメラ画像からの運動物体の抽出

譲田 賢治 坪内 貴之 菅谷 保之 金谷 健一

岡山大学工学部情報工学科

カメラを移動しながら撮影したビデオ画像から背景とは独立に運動する物体の領域を検出する新しい方法を示す。これはビデオ画像中に特徴点を抽出して追跡し、アフィンカメラモデルに基づく拘束条件を用いて背景点の軌跡を選別し、背景のみのパノラマ画像を生成して各フレームと比較するものである。本論文では、これを実現するための画像処理手法や2値化手法を述べ、実ビデオ画像を用いてその有効性を確認する。

Extracting Moving Objects from a Moving Camera Video Sequence

Kenji Jota Takayuki Tsubouchi Yasuyuki Sugaya Kenichi Kanatani

Department of Information Technology, Okayama University, Okayama 700-8530 Japan

We present a new method for extracting objects moving independently of the background from a video sequence taken by a moving camera. We extract and track feature points in the sequence and select the trajectories of background points by exploiting the constraints based on the affine camera model. Then, we generate a panoramic image of the background and compare it with the individual frames. We describe our image processing and thresholding techniques for achieving this and confirm the effectiveness of our method using real video sequences.

1. まえがき

ビデオ画像中から移動物体を抽出することは交通監視やセキュリティなどのさまざまな応用の基本となる処理であり、多くの研究が行なわれている。しかし、ほとんどはシーン中にカメラを固定し、移動物体を静止した背景中から分離することを目的としている。首振りカメラの場合も通常、投影中心を固定したパン・チルトカメラを仮定し、その制御信号から得られるカメラ方向を用いて静止カメラの場合に帰着させている [3]。そして、背景の照明変化による変動と侵入物体とを区別するさまざまな方法が開発されている [1, 10]。

一方、シーンを移動車両中から撮影したり、手動で撮影する場合はカメラ自体が移動するので、静止したシーンも画像中では未知の運動をする。この場合には未知の背景の運動に相対的に運動する物体を検出しなければならず、非常に困難な問題である。

これに対して我々はビデオ画像中に特徴点を抽出し、それを全フレーム間に渡って追跡した軌跡を背景点と物体点に分離する研究を行ってきた [7, 8, 9, 14, 15]。そして、幾何学的拘束条件を統計的検定や最適化学習と組合せ、誤追跡が含まれたり物体や背景の運動

が退化した場合にも高精度で分離できる手法を開発した [16]。

しかし、これはあくまでも移動物体の特徴点を識別しただけであり、移動物体領域を識別したわけではない。例えば移動物体部分に特徴点が抽出されなければ移動物体の存在自体が検知されないし、特徴点が移動物体の一部分に偏在すると、特徴点の存在しない部分は背景と区別できない。

本論文ではこれまでに得られた手法を組み合わせ、各画像フレームから移動物体の占める領域そのものを抽出する手法を提案する。そして、これを実現するための画像処理手法や2値化手法を述べ、実ビデオ画像を用いてその有効性を確認する。

2. 運動物体の抽出の原理

画像中を任意に運動する任意形状の3次元シーンからそれに相対的に任意の運動をする任意の3次元物体をあらゆる場合に識別することは困難である。そこで次のような現実的な制約を課す。シーンはカメラから十分遠方にあり(例えば10メートル以上)、それに比べてカメラの移動距離は小さく(例えば1メートル以内)、カメラの向きの変化も比較的小さい(例えば90°以内)とする。実際問題への応用はほとんどこのような場合であると思われる。

このような場合、背景の異なるフレーム間の対応

†700-8530 岡山市津島中 3-1-1, 岡山大学工学部情報工学科
{k-jota, tubo, sugaya, kanatani}@suri.it.okayama-u.ac.jp

は射影変換とみなせるので、これを利用してモザイク生成が可能である [5, 6]。提案手法はそのように生成したパノラマ画像を利用して移動物体領域を抽出するものである。具体的には次の段階から成る。

1. ビデオ画像中に特徴点を抽出し、全フレームに渡って特徴点を追跡する。追跡が途切れたら新たな特徴点を検出して追跡を再開する。
2. 全フレームに渡って追跡できた特徴点軌跡から、背景点の軌跡のみを選別する。
3. それらを用いて各フレームを基準フレームに射影変換して、パノラマ画像を生成する。
4. パノラマ画像の各画素を背景部分の画素に置き換え、移動物体部分を除去する。
5. 背景のみのパノラマ画像の各画像を元のフレームに逆射影変換する。
6. その画像と原画像との差分をとり、2値化して運動物体領域を抽出する。

基本的な原理は“物体”ではなく“背景”を選別することであり、従来方法とは逆に、運動物体に特徴点が抽出されなかったり抽出される特徴点が少ないほど良好な結果が得られる。

3. 幾何学的拘束条件

まず背景の選別のための幾何学的拘束条件について述べる。これは文献 [7, 8, 9, 14, 15, 16, 18] で用いられたものの特殊化である。

3.1 軌跡ベクトル

N 個の特徴点 $\{p_\alpha\}$ を M 枚の画像に渡って追跡し、第 κ 画像における α 番目の特徴点 p_α の画像座標を $(x_{\kappa\alpha}, y_{\kappa\alpha})$, $\kappa = 1, \dots, M, \alpha = 1, \dots, N$ とする。そしてその運動履歴を次の $2M$ 次元ベクトルで表し、軌跡ベクトルと呼ぶ。

$$p_\alpha = (x_{1\alpha} \ y_{1\alpha} \ x_{2\alpha} \ y_{2\alpha} \ \cdots \ x_{M\alpha} \ y_{M\alpha})^\top \quad (1)$$

これによって各特徴点の軌跡を $2M$ 次元空間の 1 点と同一視できる。便宜上、フレーム番号 κ を「時刻 κ 」と呼ぶ。

3.2 カメラ座標系とシーン座標系

本論文ではカメラの光軸を Z 軸とするカメラ XYZ 座標系を基準にとり、カメラの移動を「静止したカメラに相対的にシーンが運動する」と解釈する。

シーン中のカメラから十分離れた背景のある点 T に正規直交系 $\{i, j, k\}$ を固定し、シーン座標系の基

底とする。ただし基底ベクトル k をカメラの視点と点 T を結ぶ視線の方向にとると約束する（符号は任意）。カメラがシーン中を移動しても、仮定によりカメラの移動距離は小さく、点 T は十分遠方にあるので、 k の向きはシーンの固定されているとみなせる。

カメラが移動すると、シーン座標系の位置と向きはカメラ座標系に対しては時刻 κ ごとに变化する。点 T と基底ベクトルを時刻 κ におけるカメラ座標系で表したものをそれぞれ $t_\kappa, \{i_\kappa, j_\kappa, k_\kappa\}$ とする。シーン座標系に関して 3 次元座標 $(a_\alpha, b_\alpha, c_\alpha)$ を持つ点 p_α の時刻 κ における 3 次元位置 $r_{\kappa\alpha}$ はカメラ座標系では次のように表せる。

$$r_{\kappa\alpha} = t_\kappa + a_\alpha i_\kappa + b_\alpha j_\kappa + c_\alpha k_\kappa \quad (2)$$

3.3 アフィンカメラモデル

平行投影や弱透視投影や疑似透視投影を抽象化したアフィンカメラ [4] は、3 次元点 $r_{\kappa\alpha}$ が次のように画像上に投影されると仮定するものである。

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = A_\kappa r_{\kappa\alpha} + b_\kappa \quad (3)$$

ここに A_κ, b_κ はそれぞれ時刻 κ でのカメラの位置や内部パラメータによって定まる 2×3 行列および 2 次元ベクトルである [9]。実際のカメラの撮像は透視投影であるが、シーンが十分遠方にある場合は非常によい近似を与えることが知られている。

式 (2) を代入すると、式 (3) は次のように書ける。

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \tilde{m}_{0\kappa} + a_\alpha \tilde{m}_{1\kappa} + b_\alpha \tilde{m}_{2\kappa} \quad (4)$$

ここに $\tilde{m}_{0\kappa}, \tilde{m}_{1\kappa}, \tilde{m}_{2\kappa}$ は時刻 κ でのカメラの位置や内部パラメータで決まる 2 次元ベクトルである。 c_α の項がないのは、仮定によりシーン座標系の原点 T は十分遠方にあり、 k_κ はカメラから見た視線方向あって奥行き c_α が画像上に現れないからである。

上式を時刻 $\kappa = 1, \dots, M$ に渡って式 (1) のように縦に並べると、式 (1) の軌跡ベクトル p_α は次のように書ける。

$$p_\alpha = m_0 + a_\alpha m_1 + b_\alpha m_2 \quad (5)$$

ただし、 $m_i, i = 0, 1, 2$ は $\tilde{m}_{i\kappa}$ を時刻 $\kappa = 1, \dots, M$ に渡って縦に並べた $2M$ 次元ベクトルである。

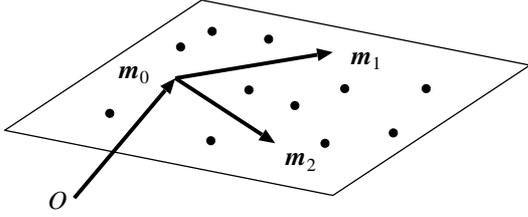


図 1: 軌跡ベクトルは 2 次元アフィン空間に含まれる .

3.4 アフィン空間拘束条件

式 (5) は, 背景点 p_α の軌跡が $2M$ 次元空間中の点 m_0 を通り, $\{m_1, m_2\}$ の張る「2 次元アフィン空間」に含まれることを意味する (図 1). 物体が任意の剛体変換をする場合は c_α の項も現れ, その軌跡ベクトルは「3 次元アフィン空間」に含まれる. これはアフィン空間拘束条件と呼ばれている [9]. 式 (5) はそれが 2 次元アフィン空間に退化したものである.

このことから, 背景点の軌跡の作る 2 次元アフィン空間から大きく離れた軌跡は背景点の軌跡ではないとみなせる. これが背景点の軌跡を選別する原理である.

本論文ではアフィンカメラモデルを用いるが, 厳密な透視投影を仮定すると Shashua ら [12, 19] の「H テンソル」の理論となる. しかし, 透視投影は非線形な関係式であり, 解析が非常に複雑になる, 上述のアフィン空間拘束条件はそれを線形近似したものに相当している.

4. 背景点の軌跡の選別

特徴点を追跡した軌跡から背景点の軌跡を選別する方法を述べる. これは菅谷・金谷のアウトライア除去の手法 [14] を式 (5) の退化したアフィン空間拘束条件に適用するものである.

N 個の特徴点が全フレームに渡って追跡できたとし, それらの軌跡ベクトルを $\{p_\alpha\}$, $\alpha = 1, \dots, N$ とする. 以下, 軌跡ベクトルの次元を $n (= 2M)$ と置く. 2 次元アフィン空間拘束条件より, この中から背景点の軌跡とみなせないもの (“アウトライア”) を RANSAC [2] によって次のように除去する.

1. 軌跡ベクトル $\{p_\alpha\}$ からランダムに 3 個を取り出し, q_1, q_2, q_3 とする.
2. 取り出したベクトルの重心を q_C とし, その周りの $n \times n$ モーメント行列 M_2 を次のように計算する.

$$M_2 = \sum_{i=1}^3 (q_i - q_C)(q_i - q_C)^\top \quad (6)$$

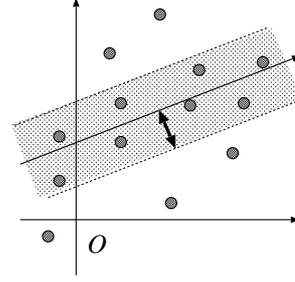


図 2: アフィン空間の当てはめによる背景点の軌跡の選別 .

3. M_2 の大きい 2 個の固有値 $\lambda_1 \geq \lambda_2$ と対応する単位固有ベクトル u_1, u_2 を計算する.
4. $n \times n$ 射影行列 P_{n-2} を次のように計算する.

$$P_{n-2} = I - \sum_{i=1}^2 u_i u_i^\top \quad (7)$$

5. 軌跡ベクトル $\{p_\alpha\}$ のうち

$$\|P_{n-2}(p_\alpha - q_C)\|^2 < (n-2)\sigma^2 \quad (8)$$

となるものの個数を S とする.

6. 以上の処理を反復し, S を最大とする射影行列 P_{n-2} を求める¹.
7. 次式を満たすベクトル p_α を除去する.

$$\|P_{n-2}(p_\alpha - q_C)\|^2 \geq \sigma^2 \chi_{n-2;99}^2 \quad (9)$$

ただし $\chi_{r;a}^2$ は自由度 r の χ^2 分布の $a\%$ 点である.

式 (8) 中の $\|P_{n-2}(p_\alpha - q_C)\|^2$ は 3 点 q_1, q_2, q_3 を通る 2 次元アフィン空間から点 p_α までの距離の 2 乗である. p_α が背景点の軌跡ベクトルであり, 各座標に独立に期待値 0, 標準偏差 σ の正規分布に従う誤差が入るとすれば, これは 2 次元アフィン空間に直交する $n-2$ 個の誤差成分の 2 乗和であり, これを σ^2 で割ったものは自由度 $n-2$ の χ^2 分布に従う. したがって残差の期待値は $(n-2)\sigma^2$ である. 上の手順ではそれ以下の軌跡ベクトルの個数が最大になるように 2 次元アフィン空間を選び, 有意水準 1% でインライアと見なされないものを除去している (図 2). 実験によれば, 通常の動画像では $\sigma = 0.5$ 程度が適当であることが確認されている [14].

5. 運動物体の抽出の手順

次に運動物体を抽出する手順を述べる.

¹実験では 200 回連続して更新がないことを収束条件とした.

5.1 パノラマ背景画像の生成

背景点の軌跡からパノラマ画像が生成できる．具体的にはある基準フレームを選び，各フレームの対応する特徴点対から基準フレームとの射影変換を計算し，その変換によって各フレームを基準フレームに写像する²．射影変換の計算には「くりこみ法」[13]を用いた³．

このとき基準フレームの各画素には複数のフレームの画素が写像されるが，それから背景部分を選ぶ方法として，本論文では各 R, G, B 値のメジアンを用いた．これは各画素が移動物体によって占められる時間は全フレームの継続時間の半分以下であるという仮定に基づくものであり，静止したカメラで撮影した画像列から移動物体を除去する方法として広く用いられている．

5.2 背景差分処理

次にその背景パノラマ画像を各フレームに逆写像して各フレームの背景画像を生成する．そして原画像との差分をとり，その絶対値⁴を2値化して運動物体領域を抽出する．

しかし，そのためのしきい値の設定が困難である．固定カメラであれば，それを設置する環境で予備実験を行うことによって経験値を得ることができるが，本論文は任意の環境にカメラを運んで（例えば移動車両中から）撮影することを想定しているので，経験値を得ることが困難である．

自動的なしきい値の選択法として「大津の方法」[11]がよく知られている．しかし，これは実シーン中の物体認識を想定しているので，通常，双峰性の輝度値ヒストグラムの谷の部分が選ばれる．一方，差画像では輝度値がほぼ0の画素が圧倒的に多いので，大津の方法を適用すると不適切な判別が行なわれることが多い．特に運動物体が存在しなければ全画素が背景であり，しきい値が存在しない（理論的には ∞ となる）はずであるが，大津の方法では適当な値が選ばれ，背景が2分割される．

これを避けるために，本論文では差画像の輝度値ヒストグラムに，背景に相当する輝度値0の近傍に密集した χ^2 分布と，物体に相当する正規分布との和を当てはめ，両者が等しい値をもつ輝度値をしきい

²仮定から理論的には2画像間はアフィン変換で結ばれるが，ここでは計算の便宜から，より一般的な射影変換を用いた．

³下記のプログラムが公開されている．

<http://www.ail.cs.gunma-u.ac.jp/Labo/research.html>

⁴実験ではカラー画像を用い，R, G, B 値のそれぞれの差分の二乗和の平方根を用いた．

値とする．その手順は付録に示す．

5.3 ノイズ除去処理

差画像において背景部分の輝度値が完全に0でないのは，シーン中の同じ点でもフレームによって露光が異なり，モザイク生成の過程でメジアンに置き換えられ，必ずしも原画像の画素値とは一致しないためである．さらにモザイク生成のための射影変換が厳密ではなかったり，変換に双1次補間を用いる影響もある．このため背景部分にも差分が抽出され，2値化するとランダムなノイズパターンを生じる．そこでこれを除去するためにまずメジアンフィルタを施し⁵，次に収縮・膨張・収縮処理を行った⁶．

5.4 画像列の分割と合成

前述の方法でパノラマ画像を生成するには，4個以上の背景点が入フレームに渡って追跡されなければならない．しかし，カメラの移動によって見えていた部分がフレームの外に出ると軌跡が途絶えてしまう．また軌跡があってもその数が少ないとロバストな背景点抽出やモザイク生成ができなくなる．

そこで途絶えていない追跡の個数がある数以下になると，画像列をそこで分割し⁷，そのブロックのみからパノラマ画像を生成する．そして次のブロックに進み，以下同様にする．最後に，継ぎ目のフレームから基準フレームへの射影変換を用いて次々と各ブロックのパノラマ画像全体を張り合わせる．

6. 実験

図3の上段はカメラを移動しながらシーンを撮影した100フレーム， 310×236 画素のビデオ画像から5フレームを抜き出したものである．これにKanade-Lucas-Tomasiの方法[17]で特徴点を抽出して追跡すると，119個の特徴点が最終フレームまで追跡できた．これから4節の方法で背景点を選別すると，92個の軌跡が得られた．図3の上段の画像中には得られた背景点の位置を白の四角，除去された点の位置を黒の四角で示している．

この背景点の対応からモザイク生成を行ない，重なる値のメジアンを表示したパノラマ画像が図4で

⁵マスクの大きさは2値化で抽出される領域の大きさに連動させた．実験ではある程度大きい領域では 5×5 ，小さい場合は 3×3 とした．

⁶実験では2画素の収縮，4画素の膨張，2画素の収縮を行った．

⁷実験では300個の特徴点を抽出して追跡し，途切れたら新たな特徴点の追跡を開始し，途切れていない軌跡が100以下になったところでブロックに分割している．

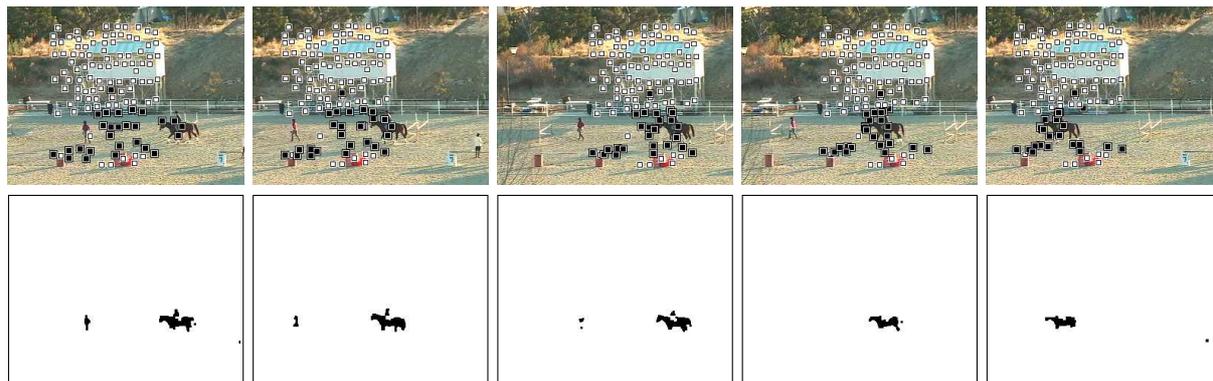


図 3: 上段：入力画像列．下段：抽出した移動物体領域．

ある．これと各フレームを比較して運動物体領域を抽出した結果が図 3 の下段である．

計算時間は特徴点の検出と追跡に 8.52 秒，背景パノラマ画像の生成に 335.47 秒，運動物体領域の抽出に 1 フレームあたり 0.25 秒かかった．ただし CPU には Pentium 4 2.6GHz，主メモリ 1GB，OS には Linux を用いた．

図 5, 6 は別の例を同様に表示したものであり，生成した背景パノラマ画像はそれぞれ図 7, 8 のようになる．これらはカメラの移動量が大きいので，図 5 では 4 ブロックに，図 6 では 5 ブロックに分割された．

7. まとめ

本論文ではカメラを移動しながら撮影したビデオ画像から背景とは独立に運動する物体の領域を検出する新しい方法を示した．これは，ビデオ画像中に特徴点を抽出して追跡し，背景点の軌跡を選別し，背景のみのパノラマ画像を生成して各フレームと比較するものである．これを実現するための画像処理手法や 2 値化手法を述べ，実ビデオ画像を用いてその有効性を確認した．

現状では微小な移動物体が見逃されたり，モザイク生成の誤差によって背景部分にランダムなノイズパターンが検出されるという問題が残る．また当然，背景と類似した輝度や色をもつ部分は検出されにくい．これらを解決するには，この手法による検出を第 1 近似としてさらなる処理を加える必要があり，今後の課題である．

一方，提案方法はカメラの移動に関する知識や計測が不要であるため，未知の環境にも容易に適用できる．全フレームを必要とするためにオフライン処理となるが，セキュリティ目的で監視結果を逐次，データベースに記憶しておくような時間遅れが許される



図 4: 図 3 から生成した背景のみのパノラマ画像．

応用では，提案方法は非常に有効であると思われる．

謝辞：本研究の一部は文部科学省科学研究費基盤研究 C (2) (No. 15500113)，テレコム先端技術研究支援センター，栢森情報科学振興財団の助成によった．

参考文献

- [1] A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis, Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, *Proc. IEEE*, **90-7** (2002-7), 1151–1163.
- [2] M. A. Fischer and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM*, **24-6** (1981-6), 381–395.
- [3] E. Haymann and J.-O. Eklundh, Statistical background subtraction for a mobile observer, *Proc. 9th Int. Conf. Comput. Vision*, Nice, France, October 2003, Vol. 1, pp. 67–74.
- [4] 金出武雄, コンラッド・ポールマン, 森田俊彦, 因子分解法による物体形状とカメラ運動の復元, 電子情報通信学会論文誌 D-II, **J74-D-II-8** (1993-8), 1497–1505.
- [5] 金澤 靖, 金谷 健一, 幾何学的 AIC による画像モザイク生成の安定化, 電子情報通信学会論文誌 A, **J83-A-6** (2000-6), 686–693.

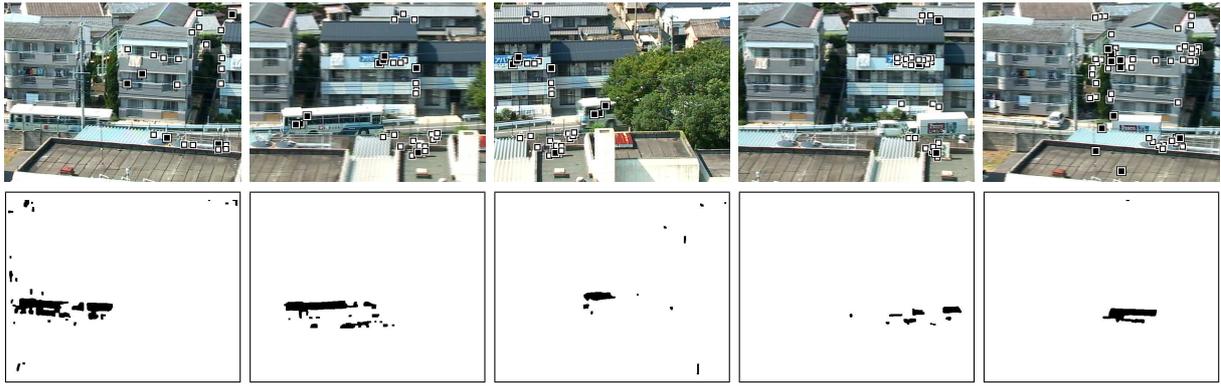


図 5: 原画像 (上段) と抽出した移動物体領域 (下段) .

- [6] 金澤 靖, 金谷 健一, 段階的マッチングによる画像モザイク生成, 電子情報通信学会論文誌 D-II, **J86-D-II-6** (2003-6), 816–824.
- [7] 金谷健一, 黒澤典義, 松永力, モデル選択によるランク推定と複数運動の分離, 情報処理学会研究報告, 2001-CVIM-126-3, 17–24, March 2001.
- [8] 黒澤 典義, 金谷 健一, 部分空間分離法とモデル選択による運動物体の分離, 情報処理学会研究報告, 2000-CVIM-124-4, 25–32, Nov. 2000.
- [9] 黒澤 典義, 金谷 健一, アフィン空間分離法による運動物体の分離, 情報処理学会研究報告, 2001-CVIM-125-3, 25–32, Jan. 2001.
- [10] A. Monnet, A. Mittal, N. Paragios and V. Ramesh, Background modeling and subtraction of dynamic scenes, *Proc. 9th Int. Conf. Comput. Vision*, Nice, France, October 2003, Vol. 2, pp. 1305–1312.
- [11] 大津展之, 判別および最小 2 乗規準に基づく自動しきい値選定法, 電子通信学会論文誌 D, **J63-D-4** (1980-4), 349–356.
- [12] A. Shashua and L. Wolf, Homography tensors: On algebraic entities that represent three views of static or moving planar points, *Proc. Euro. Conf. Comput. Vision*, Dublin, Ireland, June 2000, Vol.1, pp. 507–521.
- [13] 清水慶行, 太田直哉, 金谷健一, 信頼性評価を備えた最適な射影変換の計算プログラム, 情報処理学会研究報告, 98-CVIM-111-5 (1998-5), 33–40.
- [14] 菅谷保之, 金谷健一, 部分空間分離法による特徴点追跡のアウトラリア除去, 情報処理学会研究報告, 2002-CVIM-133-24, 177–184, May 2002.
- [15] 菅谷保之, 金谷健一, 運動物体分離のためのカメラモデルの自動選択, 情報処理学会研究報告, 2002-CVIM-134-2, 9–16, Sept. 2002.
- [16] 菅谷 保之, 金谷 健一, 複数運動の教師なし学習による多段階最適化, 情報処理学会研究報告, 2003-CVIM-138-25, 185–192, May 2003.
- [17] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*, CMU Tech. Rep. CMU-CS-91-132, Apr. 1991; <http://vision.stanford.edu/~birch/klf/>.



図 6: 図 7 から生成した背景のみのパノラマ画像 .

- [18] 坪内貴之, 菅谷保之, 金谷健一, 3次元アフィン復元のための途切れた特徴点追跡の延長, 情報処理学会研究報告, 2003-CVIM-137-17, 133–140, March 2003.
- [19] L. Wolf and A. Shashua, On projection matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$, $k = 3, \dots, 6$, and their applications in computer vision, *Int. J. Comput. Vision*, **48-1** (2002-6), 53–67.

付録：背景差分の 2 値化処理

手順．背景差分画像の 2 値化のしきい値を次のように計算した．

- 輝度値レベルを $x = 0, 1, 2, \dots, x_{\max}$ とし, それぞれの輝度値を持つ画素数 (ヒストグラム) $h(x)$ を作成する．ただし $h(0) = 0$ とする．
- 適当な初期しきい値 x_c を定める⁸．
- 輝度値 x の画素が背景である確率 $b(x)$ と侵入者である確率 $g(x)$ を表す配列の値を $x = 0, 1, \dots, x_{\max}$ に渡って次のように初期化する．

$$\begin{aligned}
 b(x) &= \begin{cases} 1 & x = 0, 1, \dots, x_c \\ 0 & x = x_c + 1, \dots, x_{\max} \end{cases} \\
 g(x) &= \begin{cases} 0 & x = 0, 1, \dots, x_c \\ 1 & x = x_c + 1, \dots, x_{\max} \end{cases} \quad (10)
 \end{aligned}$$

⁸実験では輝度値が x_c 以上の画素が全画素中の 10% になるように定めた．

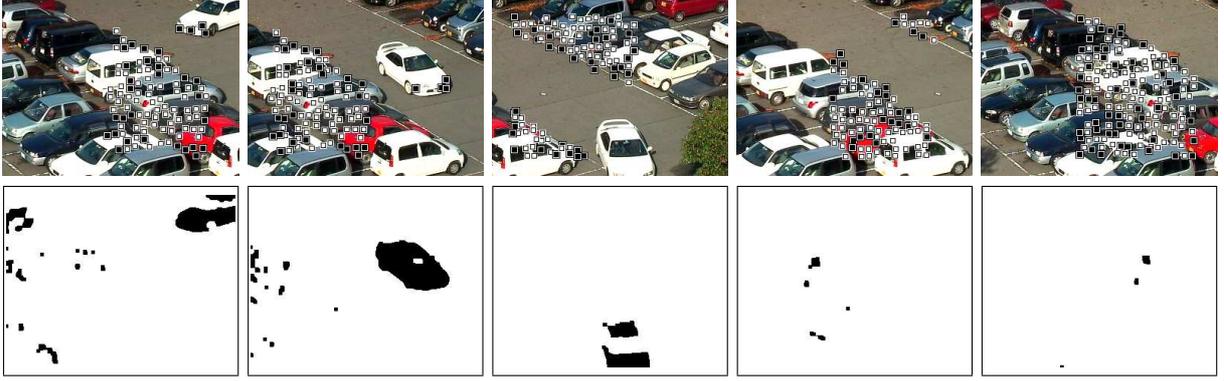


図 7: 原画像 (上段) と抽出した移動物体領域 (下段) .

- 次の計算を $N_0, N_1, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2$ が収束するまで反復する .

1. 背景画素数 N_0 とその平均輝度値 μ_0 を次のように計算する .

$$N_0 = \sum_{x=0}^{x_{\max}} b(x)h(x)$$

$$\mu_0 = 0.5 + \frac{1}{N_0} \sum_{x=0}^{x_{\max}} xb(x)h(x)$$

$$\sigma_0^2 = \frac{1}{N_0} \sum_{x=0}^{x_{\max}} x^2b(x)h(x) - (\mu_0 - 0.5)^2 \quad (11)$$

2. 物体画素数 N_1 , その平均輝度値 μ_1 , その分散 σ_1^2 を次のように計算する .

$$N_1 = \sum_{x=0}^{x_{\max}} g(x)h(x)$$

$$\mu_1 = 0.5 + \frac{1}{N_1} \sum_{x=0}^{x_{\max}} xg(x)h(x)$$

$$\sigma_1^2 = \frac{1}{N_1} \sum_{x=0}^{x_{\max}} x^2g(x)h(x) - (\mu_1 - 0.5)^2 \quad (12)$$

3. 配列 $b(x), g(x), x = 0, 1, \dots, x_{\max}$ を次のように更新する .

$$s_0 = \frac{2N_0\mu_0}{\sigma_0^2} \phi_{2\mu_0^2/\sigma_0^2} \left(\frac{2\mu_0x}{\sigma_0^2} \right)$$

$$s_1 = \frac{N_1}{\sqrt{2\pi}\sigma_1} e^{-(x-\mu_1)^2/2\sigma_1^2} \quad (13)$$

$$b(x) = \begin{cases} 0 & s_0 \approx 0 \text{ のとき} \\ \frac{s_0}{s_0 + s_1} & \text{それ以外} \end{cases}$$

$$g(x) = 1 - b(x) \quad (14)$$

- $b(0), b(1), b(2), \dots$ の値を順に調べ , 初めて $b(x)$



図 8: 図 6 から生成した背景のみのパノラマ画像 .

$\geq 0.5 \geq b(x+1)$ となる x を求め ,

$$x_c = \begin{cases} \frac{(x+1)b(x) - xb(x+1) - 0.5}{b(x) - b(x+1)} & b(x) > b(x+1) \text{ のとき} \\ x + 0.5 & b(x) = b(x+1) \text{ のとき} \end{cases} \quad (15)$$

をしきい値とする . ただし $b(x_{\max})$ に達しても $b(x_{\max}) \geq 0.5$ なら $x_c = x_{\max}$ とする .

上記の計算は輝度値ヒストグラム $h(x)$ を χ^2 分布と正規分布の和で近似し , 式 (14) のように各 x に対して χ^2 分布の画素の割合 , および正規分布の画素の割合をそれぞれ $b(x), g(x)$ としている . そして , この割合で輝度値ヒストグラム $h(x)$ を背景画素部分と物体画素部分に分け , それから式 (11), (12) によって χ^2 分布と正規分布のパラメータを計算し直している . χ^2 分布のパラメータは , 確率密度 $(1/a)\phi_r(x/a)$ の平均と分散がそれぞれ $\mu = ar, \sigma^2 = a^2r$ であることから ,

$$r = \frac{2\mu^2}{\sigma^2}, \quad a = \frac{\sigma^2}{2\mu} \quad (16)$$

と決めている . そしてステップ 1 に戻り , これを収束するまで反復している . このような計算法はクラスタリングの教師なし学習 , または EM アルゴリズムと呼ばれている [16] .



図 9: 入力画像 (左) と生成した背景画像 (右) .

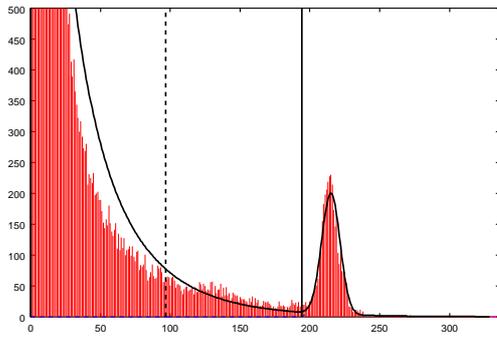


図 10: 図 9 の画像の差分の輝度値ヒストグラム．曲線は近似した分布．縦の実線は計算したしきい値．縦の破線は大津の方法によるしきい値．

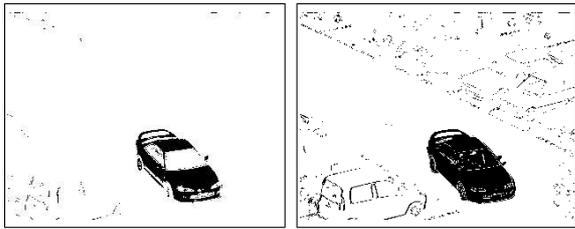


図 11: 図 9 の画像の差分の 2 値化画像．左は提案手法, 右は大津の方法 .

ヒストグラム $h(x)$ において $h(0) = 0$ としているのは次の理由による．入力画像と背景画像の輝度値が (R, G, B と) 完全に一致するのは背景画像のその部分にその入力画像そのものが使われたり, 輝度値が飽和して入力画像, 背景画像とも最高階調になったりする特殊な状況であり, しきい値の選択に用いるのは不適切と考えられるからである．

また, 式 (11), (12) の分布パラメータの計算に数値 0.5 が含まれているのは, 輝度値の区間 $[x, x + 1]$ のヒストグラムの高さを $h(x + 0.5)$ の値とみなしているためである．

実験によれば, この反復は分布のパラメータの変化率⁹が 10^{-5} 以下を収束条件とすると, ほぼ 100 ~ 200 回で収束した．これは, x_c の初期値を異常な値 (0 や

⁹ただし, 背景と移動物体の画素数 N_0, N_1 については (輝度値 0 のものを除いた) 全画素数との比とした．



図 12: 入力画像 (左) と生成した背景画像 (右) .

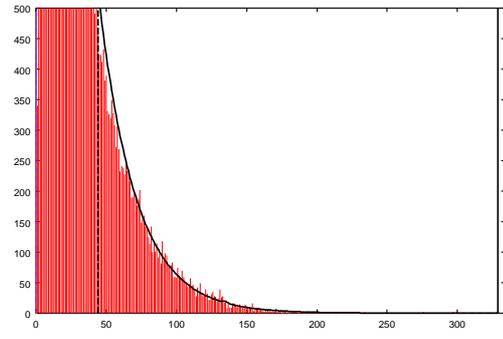


図 13: 図 12 の画像の差分の輝度値ヒストグラム．曲線は近似した分布．縦の実線は計算したしきい値．縦の破線は大津の方法によるしきい値．

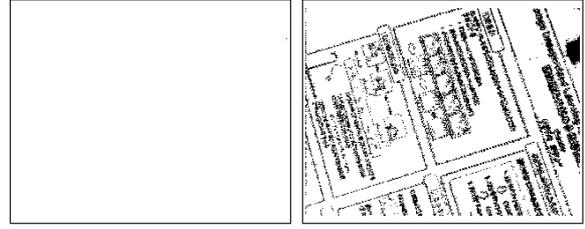


図 14: 図 12 の画像の差分の 2 値化画像．左は提案手法, 右は大津の方法 .

x_{\max} 近辺) に選ばない限り, 初期値に依存しない一意的な値に収束する．

例．図 9 は移動物体を含む画像の背景差分画像である．この画像の輝度値ヒストグラムは図 10 のようになる．これに前述の方法で分布 $f(x)$ を当てはめたものが図中の曲線であり, 縦の実線が求めたしきい値である．縦の破線は大津の方法で求めたしきい値である．そして, それぞれのしきい値で 2 値化すると図 11 のようになり, 提案手法のほうがより妥当な分割となっている．

図 12~14 は移動物体を含まない場合を同様に表示したものである．この場合でも大津の方法は画像を適当に分割してしまうのに対し, 提案手法ではしきい値が自動的に大きくなり, 移動物体が存在しないと判定されている．