

人物領域抽出のためのステレオカメラによる背景差分法の一検討

A study on background subtraction method using stereo camera for human extraction

上野 智史
Satoshi Ueno

三功 浩嗣
Hiroshi Sankoh

内藤 整
Sei Naito

1. はじめに

我々はこれまでに、複数カメラを利用し、主に人物映像を対象として屋内スタジオ環境での自由視点映像の合成技術に取り組んできた[1]。これまでの検討により、多数のカメラを用いた撮影、もしくはクロマキーのような特殊な背景の環境においては、実用的な精度で人物領域の抽出を実現した。しかしながら、一般的な背景の環境で自由視点映像を合成するためには、各カメラ映像のフレーム毎に人物領域を高精度に抽出する必要がある。

固定カメラ環境においては、特に物体と背景の境界を高精度に検出するという観点で、背景差分法が効果的である。また、カメラからの点との 3 次元距離である、奥行き値に基づく物体抽出法は、特に物体中心部を欠損なく抽出できるという観点で優れる。そこで本論文では、固定ステレオカメラを利用して画素値と奥行き値を利用することで、より高精度に人物領域を抽出することを目的とする。

2. 画素値による背景差分法とその課題

カメラ映像のピクセル毎に、画素値をフレーム間の混合ガウス分布で分類し、背景モデルを構築する背景差分法が、多少の背景の変動や照明変動などに対してロバストに前景領域を抽出する手法として有効である[2,3]。以下、背景モデルの構築手法を述べる。背景モデルはピクセルごとに独立してフレーム間で算出するため、以降はあるピクセル (x, y) についての処理を述べる。時刻 t での画素値を X_t とするとき、時刻 t までに得られた画素値の組 $\{X_1, \dots, X_t\}$ は K 個のガウス分布を用いてモデル化できる。このとき、現在の画素 X_t がもつ背景モデルへの適合確率は、

$$P(X_t) = \sum_{k=1}^K \omega_{k,t} \eta(X_t | \mu_{k,t}, \Sigma_{k,t}) \dots (1)$$

となる。ここで、 $\omega_{k,t}$ は時刻 t での i 番目のガウス分布の重み係数、 $\mu_{i,t}$ は時刻 t での i 番目のガウス分布の平均値、 $\Sigma_{i,t}$ は時刻 t での i 番目のガウス分布の分散共分散行列である。計算コストを削減するために、各色成分が独立であるとみなして共分散部を 0 とし、 $\Sigma_{i,t} = \sigma_{k,t}^2 I$ と表現する。

X_t に対する前景または背景の推定は、 X_t が属する確率が最も高いガウス分布が前景または背景のいずれに属するかに基づき行われる。 X_t が既存のいずれのガウス分布にも属しないと判定された場合は、これまでに最も利用されていないガウス分布を破棄し、新規にガウス分布を構築する。推定後、 X_t を利用してモデルの更新を行い、各ガウス分布の重みの正規化、およびそのガウス分布の照合度 ω/σ^2 の大きい順にソートを行う。ソート後、再構成された混合ガウス分布のうち、閾値 T_B を元に次式を満足する B 個の分布を新しく背景モデルと設定する。

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b \omega_{k,t} > T_B \right) \dots (2)$$

式(2)は、これまでに利用された照合度が高いガウス分布に対して、順に背景のラベルを割り当てることを示す。

影領域に対しては、影の輝度特性を考慮して影の混合ガウス分布によるモデルを構築し、影に属するか判定を行う。

画素値による背景差分法は、監視用途などで背景が大部分を占める中で特定の動物体の有無を検出する場合には有効性が示されている。しかしながら、自由視点映像の合成のためには、前景を欠損なく抽出する必要があるという観点で下記 2 つの課題がある。

1. 一時静止物体の対応
前景の一部が一定時間停止すると、前景が背景モデルに溶け込むケースが存在し、以降同手法では前景の人物抽出は困難になる。
2. 影の対応
前フレームまでの背景モデルと比較して、輝度値が低い新規モデルは影と判断されるケースが存在する。

3. 提案手法

以上の課題を踏まえ、本稿では、画素値に基づく背景差分法で算出した前景領域の更新時に、奥行き値の変化の判定を加えることで、一時静止物体及び影の影響を軽減する手法を提案する。

3.1 奥行き値の取得

入力のステレオカメラ画像に対して、平行ステレオ処理を行う。平行ステレオ画像に対して、ブロックマッチングによりピクセルごとに視差値を算出する[4]。ここで、ステレオ視による奥行き値算出は、対応する領域の視差値 d_t に基づき算出する。視差値 d_t と奥行き値 Z_t の関係は、カメラの焦点距離を f 、カメラ間の基線長を b とするとき、

$$Z_t = \frac{bf}{d_t} \dots (3)$$

と表現される。式(3)は、ステレオ環境で奥行き値を算出する際に、視差値が小さいピクセルでは、視差値の誤差の影響が奥行き値に大きく反映されることを示す。

3.2 奥行き値による背景モデルの構築

視差値から算出される奥行き値は、その視差算出における対応領域検出の精度により、時系列で大きく変動するケースが存在する。そのため、変動する結果の影響を軽減させるために、人物がいない背景領域である空舞台を複数フレーム撮影して、奥行き値による背景モデルを構築する。背景モデルは奥行き値 Z_t とその信頼度 M_t を入力値として、画素値によるモデルと同様に(1)で表現する。奥行き値の信頼度 M_t は、ピクセルごとに視差を算出する際に利用するコスト関数[4]の最小値と 2 番目の値のマージンを利用する。本値が高いほど、視差の算出結果の信頼度が高いことを意味する。

3.3 奥行き値を併用した前景および背景の判定

奥行き値の算出結果 Z_t は、安定して奥行き値が取得できているピクセルにおいてのみ利用する。すなわち、上記

表 1 実験結果(精度)

	Precision	Recall	F値
従来手法	0.92	0.89	0.90
提案手法	0.91	0.97	0.94

信頼度 M_t による背景モデルで、適合モデルの平均が高く分散が低いピクセルにおいて、奥行き値 Z_t の背景モデルの平均が低い場合に利用する。上記に適合しないピクセルに対しては、終始画素値による背景モデルの結果を利用することにする。奥行き値の信頼度が高い領域に対して、2 節で述べた 2 つの課題を解消するために、画素値に加え、奥行き値による判定を導入する。

1. 一時静止物体の対応

画素値判定により、フレーム間で前景から背景に変化するピクセルについて、奥行き値 Z_t の背景モデルとの適合度を算出する。背景モデルとの適合度の判定は、文献[2]での画素値の背景モデルで用いている判定と同様に行い、各モデルとの類似度をその分散 σ^2 を閾値にして決定する。奥行き値 Z_t が背景モデルと適合しない場合は、前景が遮蔽していると判断し前景と判定する。

2. 影の対応

画素値判定により、フレーム間で新しく影モデルと判定されたピクセルに対して、奥行き値 Z_t の背景モデルとの適合性を確認する。奥行き値 Z_t が背景モデルと適合しない場合は、輝度値が低い新しい前景が遮蔽していると判断して、前景と判定する。

4. 評価実験

4.1 実験環境

市販のステレオカメラ HDR-TD20V で撮影した映像データを利用して評価を行った。画像の解像度は 640×360 で、360 フレームのデータを利用した。うち、はじめの 185 フレームは空舞台であり、それぞれ画素値によるモデル構築および奥行き値によるモデル構築を行う。映像は、空舞台の後、一人の人物が移動してきて、一時停止の上、元に戻る構成である。画素値に基づく背景モデルによる前景抽出手法(従来手法)と、奥行き値に基づく背景モデルを併用した手法(提案手法)の算出結果に対して、細かなノイズ領域

を取り除くために、GrabCut[5]により領域の統廃合を行う。精度は、上記人物領域抽出結果と、手動でシードを選択し GrabCut により分割した人物領域との比較で評価する。評価尺度は、抽出した人物領域に占める正解人物領域の割合を Precision、および正解人物領域に占める抽出した前景領域の割合を Recall とし、そしてそれらの F 値を利用した。

4.2 実験結果

実験結果画像の例を図 1 に示す。図 1(a-1),(b-1)は入力フレームの一部であり、また(a-2),(b-2)は従来手法による前景抽出結果(白色)、および(a-3),(b-3)はその前景抽出結果をシードとして GrabCut により抽出した人物領域である。(a-2)において灰色部は影領域と判定された領域である。また同様に、(a-4),(b-4)は提案手法による前景抽出結果(白色)であり、(a-5),(b-5)はその前景抽出結果をシードとして GrabCut により抽出した人物領域である。

(a-2)で、従来手法では人物の顔領域および背中領域が影領域として背景と判定されていることが確認され、それにより最終的な人物領域抽出精度(a-3)が低下していることが確認できる。一方で、(a-4)で、提案手法により従来手法の影領域が奥行き値により前景と判定されていることが確認でき、最終的に良好に人物領域が抽出できていることが分かる(a-5)。

また、図 1 (b)は人物が 3 秒程度静止した後の移動直前のフレームであり、(b-2)で、従来手法では人物領域が背景領域として溶け込んでいることが確認でき、その結果人物抽出精度が低下していることが分かる(b-3)。一方、(b-4)で、提案手法では、一部背景モデル構築精度が低いことが原因で胴体部に欠損も存在するが、奥行き値の判定の適用により概ね胴体領域を前景として保持していることが確認され、最終的に良好に人物領域を抽出できていることが分かる(b-5)。

フレーム全体に対する、従来手法と提案手法の精度を表 1 に示す。提案手法は上記 2 つの課題に対応して検出漏れの割合を減少させることで、ほぼ Precision を保ったまま、Recall を大幅に上昇させていることが確認できる。

5. まとめと今後の課題

本論文では、固定ステレオカメラを利用して、画素値に加え奥行き値を併用することにより人物領域を安定して抽出する手法を提案した。今後は、処理の高速化および照明変動の影響を考慮した改良を行う予定である。

参考文献

- [1] 三功, 石川, 内藤, 酒澤, “確率尤度を用いた 3 次元形状モデル投影型背景分離法,” 映像情報メディア学会誌 Vol.64, No.11, pp.1685-1697(2010)
- [2] Z. Zivkovic, F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” Pattern Recognition Letters, vol. 27, no. 7, pages 773-780, 2006.
- [3] 島田, 有田, 谷口, “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構築,” 電子情報通信学会論文誌 Vol.J90-D, No.9, pp.2606-2614,2007.
- [4] Hirschmuller, H. Stereo Processing by Semiglobal Matching and Mutual Information, PAMI(30), No. 2, February 2008, pp. 328-341.
- [5] C. Rother, V. Kolmogorov and A. Blake: “GrabCut”: Interactive Foreground Extraction using Iterated Graph Cuts”, ACM Trans. Graph., 23, 3, pp. 309–314 (2004).

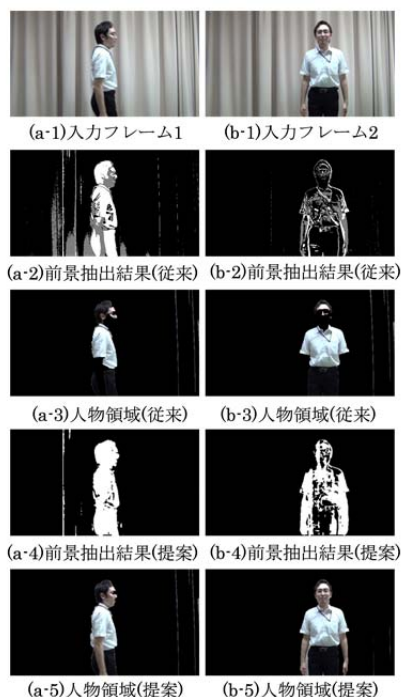


図 1 実験結果画像の例