

ステレオマッチングによる深度情報を用いる背景差分法に基づく映像オブジェクトの抽出

Image Object Extraction with Background Subtraction using Stereo-Matching-Based Depth Information

池田 将之
Masayuki IKEDA

1 はじめに

「t-Room」^{?)}では、ディスプレイとカメラが対峙するという特殊な構造をとっている為、カメラ映像に鏡合わせのようなエコーが発生する。この問題点を解決する為、解決方法の一つとして、本来ディスプレイに表示されることが求められる映像オブジェクト、即ちディスプレイの前景となる人物や物体のみをカメラ映像から抽出する方法が模索されてきた。また、筆者らは従来使用することが想定されていなかった t-Room の中心部を使用可能にすることを目指している^{?)}。実現に向けては、t-Room 内の参加者の位置を把握しなければならず、ここで深度情報を利用することを検討している。本研究では、ステレオカメラを利用したステレオマッチングから得られる深度情報に対して背景差分を適用することで、映像オブジェクトを抽出する手法を提案し、提案手法とステレオマッチングから得られる深度情報を基にマスク処理を行う手法との比較実験を行い、提案手法によるオブジェクト抽出精度の向上を確認する。

2 ステレオマッチングによる映像オブジェクト抽出法

2.1 ステレオマッチング

2つの視点から撮影した二次元画像間には、被写体の遠近に応じてずれが生じる。このずれを視差と呼び、視差を利用して、三次元の深度を求めることができる。これをステレオマッチングと呼び、視差の大きさを画像で表現したものを視差マップと呼ぶ。視差マップでは視差の大きい部分を明るく(輝度値は高く)、視差の小さい部分を暗く(輝度値は小さく)、段階的に表示する。ずれが大きいほど、その部分が基線からより近距離にあることを示す。

2.2 固定閾値による映像オブジェクト抽出

ステレオマッチングによって出力される視差マップに対して、設定された閾値よりも輝度値の高いピクセルのみを前景と判定する手法である。この手法においては、閾値を一定に設定する。同一の被写体を撮影していても、ステレオカメラ間の基線長が変われば、視差も変わるので、基線長の変更に応じて閾値も変更しなければならない。

また、t-Room において、カメラはディスプレイの上部に空間を見下ろす角度で配置されている。故に、カメラからの直線距離上で、t-Room 内の床とオブジェクトが同距離になってしまう地点が存在する。閾値を一定の値で設定しなければならない本手法では、閾値を低く設定すると余分な抽出をしてしまう。また、閾値を高くするとオブジェクトの一部を抽出できない。このように本手法では背景と判定されるべき部分の除去精度と前景の抽出精度の間にトレードオフの関係が存在する。このために本手法のみを用いて映像オブジェクトを抽出することは

困難である。

2.3 深度情報を用いる背景差分法に基づく映像オブジェクト抽出

そこで本研究では、ステレオマッチングによる深度情報を用いる背景差分法に基づく映像オブジェクトの抽出法を提案する。本手法では事前に用意した多数の背景画像から背景の視差マップを作成し、ステレオマッチングの結果出力される視差マップと背景の視差マップとの差分をとり、値が一定以上の領域を前景と判定し、マスク画像を生成する。

本研究では、以下に述べる3手法の背景の視差マップ作成法を作成した。

平均値によるマスク作成法

元画像 150 枚の各画素について、その平均値を背景の視差マップとする。予備実験により、この手法ではノイズの影響を受けることが判明したため、新たに以下の2手法を考案した。

最大値によるマスク作成法

元画像 150 枚の各画素について、その最大値を背景の視差マップとする。この手法ではノイズには対応できるものの、本来の目的である映像オブジェクトの抽出精度が落ちてしまうので、さらに以下の手法を考案した。

分散を考慮したマスク作成法

元画像 150 枚の各画素について、その分散を計算し、その値が大きいものほど最大値に重みを、その値が小さいものほど平均値に重みをもたせて背景の視差マップを作成する。ノイズはフレームごとに離散的に出現するので、分散が大きくなるという特徴を利用している

3 評価実験

3.1 概要

本研究で提案した映像オブジェクト抽出法について評価を行った。以下でその概要について述べる。開発したステレオマッチングによる深度情報を用いる背景差分法に基づく映像オブジェクトの抽出法とステレオマッチングによる深度情報を用いる固定閾値に基づく映像オブジェクトの抽出法の2つの手法でオブジェクト抽出を行い、結果の精度を比較する。また、本実験ではウェブカメラを4台並べて設置し、基線長の異なる3対のステレオカメラで実験を行うことで、基線長の変化によるシステムの性能の変化についても評価する。ここでは、基線長の短いステレオペアから順に R1-L1, R1-L2, R1-L3 と表記する。定量的な評価を容易にするために、静止画を用いて実験を行う。各基線長に対して 10 枚のテスト画像を用意し、その元画像に対して正解画像を作成する。本研究での性

能は、特に映像オブジェクトの抽出精度に注目しており、*Precision*(適合率、性能)や*Recall*(再現率)、これらの調和平均である *F-measure*(*F* 値)等の評価指標を基に、定量的に評価する。

3.2 結果

この節でのグラフや図上の“threshold”は固定閾値によるマスクの作成法、“bs_ave”は平均値によるマスクの作成法、“bs_max”は最大値マスクの作成法、“bs_var”は分散を考慮したマスクの作成法を表す。

提案手法で作成された背景の視差マップの一例を図??に示す。

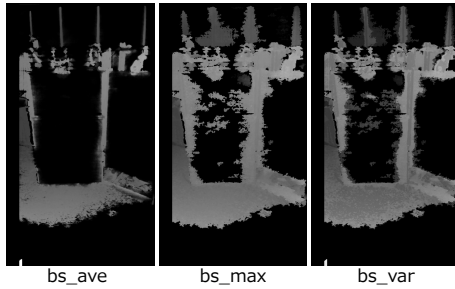


図 1 作成した背景の視差マップ例 (R1-L2 のペア)。

固定閾値を利用したマスク作成法と提案手法の映像オブジェクトの抽出結果例を図??に示す。

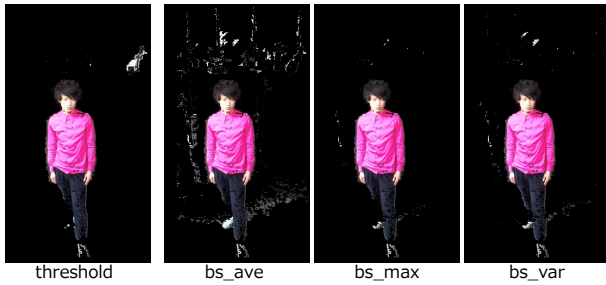


図 2 抽出結果例 (R1-L2 のペア)。

各手法の *F-measure* のみを抜き出したものを図??に示す。

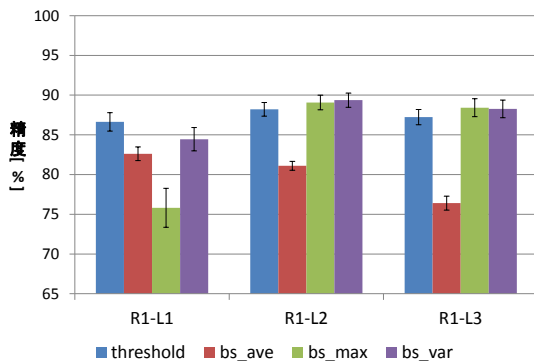


図 3 *F-measure* の比較。

3.3 考察

基線長が短いほど、カメラからの距離に近いオブジェクトに対するマッチング精度は高くなるが、カメラからの距離が遠いオブジェクトに対するマッチングに弱くなる。R1-L1 のステレオマッチングでは、出力される視差マップの濃淡がはっきりしておらず、提案手法では下半身

と床の差分があいまいとなったり、ノイズの影響を受けて前景のオブジェクトよりも背景にかかったノイズのほうが前景だと認識されたりしてしまった。そのため、提案手法を用いた場合、他の基線長と比較しても *F-measure* が大きく劣るという結果が得られた。

しかしながら、R1-L2 と R1-L3 のステレオマッチングにおいては、提案手法を用いると、わずかではあるが *F-measure* の値が改善された。ただし提案手法の中でも、平均値によるマスク作成法は、ノイズに対応することができないため、今回使用した OpenCV 内のステレオマッチングのクラスを利用して視差マップを出力にする限り、この手法の有用性はないことが分かった。

視差マップから固定閾値を用いてマスクを作成する際には、カメラの基線長が変われば、閾値を変えなければならない。提案手法では、ステレオカメラの基線長を変更しても同じシステムを流用できるため、*F-measure* の数値上は劇的な改善は見られなかったものの、有用性があると考えられる。

また、キャリブレーションにより、ステレオカメラで撮影した画像間で異なる歪みが起こり、ミスマッチングが生じていた。このミスマッチングは画像の端でミスマッチングが起こることが多い。

4 まとめと今後の展望

本論文では、t-Room における映像エコー問題を解決するために必要な映像エコー・キャンセリング法としてステレオマッチングによる深度情報を用いる背景差分法に基づく映像オブジェクトの抽出法を提案した。実際に提案したシステムを実装し、ステレオマッチングから得られる深度情報を基にマスク処理を行う手法との比較実験により、映像オブジェクトの抽出精度を用いて、その性能の向上を評価した。各手法を適用した結果をピクセル単位で定量的に精度評価した結果、基線長による制限はあるものの、オブジェクトの抽出精度に改善が見られた。また、視差マップから値が一定の閾値を用いてマスクを作成する際には適切な閾値を探索しなければならなかったが、提案手法では基線長の変更によるシステムの変更は必要ないため、汎用性が高い。しかしながら、リアルタイムの観点からはまだ実用とはほど遠いため、GPGPU の導入を考えている。また、映像オブジェクトの抽出精度にもまだまだ改善の余地がある。主にノイズの問題やオブジェクトの輪郭付近のピクセルを余分に抽出する問題など、使用した OpenCV 内のステレオマッチングのクラスに依存した問題もあるため、他のステレオマッチング手法の利用も検討したい。さらに本論文のシステムでは USB ウェブカメラを用いて評価実験を行ったが、実際の t-Room 環境では IP カメラを使用しているため、t-Room への導入のために IP カメラの導入とその評価を行う。

参考文献

- 1) Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemitsu Aoyagi, Yoshinari Shirai, Naomi Yamashita, and Junji Yamato; "The t-Room - Toward the Future Phone", NTT Technical Review, Vol. 4, No. 12, pp. 26-33(2006 12).
- 2) 和田理, 片桐滋, 大崎美穂, “空間重量型遠隔コラボレーション支援システムにおける利用可能領域の拡大”, IPSJ SIG Technical Report, 2015.