

単眼カメラ画像を用いた距離推定に撮影環境が及ぼす影響

田村 悠風 大川 猛

東海大学情報通信学部組込みソフトウェア工学科

1 はじめに

ロボットが3次元空間内で行動するためには周囲の物体との位置関係を明確にするための距離計測が不可欠である[1]. 距離計測の方法にはレーザーセンサやステレオカメラなどを用いたものがあるが、これらの方法と比べて、1台のカメラを用いる単眼距離推定は軽量かつ安価であるため、低価格な計測系の構築や積載量に制限がある小型 UAV などに搭載するのに有利である[2]. また、近年のニューラルネットワークの進歩によりニューラルネットワークを用いた単眼カメラ距離推定手法が研究されている. この手法は制約が少なく幅広い場面で活用できる. しかし、ニューラルネットワークを用いた手法は既存の手法との比較に乏しい[3][4].

そこで、本研究では単眼カメラを用いた既存の距離推定法とニューラルネットワークを用いた距離推定法の性能を評価し、更に撮影環境の変化による推定結果の誤差について調査した.

2 距離推定手法

既存の単眼カメラによる距離推定手法とニューラルネットワークを用いた手法を比較するため、2つの手法を取り上げた. それぞれをマーカーによる推定法、ニューラルネットワークによる推定法と名づけ、以下に各方法を解説する.

(1) マーカーによる推定法

マーカーによる推定法は距離に応じて画像上の物体の大きさが変化する性質を利用して距離を推定する[5]. まず、ある距離を基準距離として定め、基準距離でのマーカーの直径を記録する. 記録した基準距離でのマーカーの直径と推定したい地点でのマーカーの直径を比較することで距離を求める.

(2) ニューラルネットワークによる推定法

ニューラルネットワークによる推定法は、単眼カメラで撮られた画像と距離データをもとにニューラルネットワークに学習させることで画像から奥行きを推定する. 実験では MiDas というオープンソースソフトウェアを利用した[4]. MiDas は単眼カメラの画像から深度を推定することが

The Effect of Shooting Environment on Distance Estimation Using Monocular Camera Images

Yunagi Tamura, Takeshi Ohkawa

Tokai University

できるオープンソースソフトウェアである.

3 評価

(1) 方式の比較

異なる2つの手法に関して処理時間の計測と推定結果の誤差を評価した.

性能評価には表1に示す環境で行った. 一つは一般的なPC環境. もう一つは組み込み環境での性能を評価するために、シングルボードコンピュータの Raspberry Pi を用いた.

処理時間の計測のために、カメラと物体との距離が異なる(10cm~100cm) 10種類の画像をテストデータとして用意した. (手法1)マーカーによる推定法と(手法2)ニューラルネットワークによる推定法のそれぞれで処理時間の計測を行った. 画像一枚あたりの処理時間を図1に示す.

表1: 評価環境

	PC	Raspberry Pi3 Model B
CPU	AMD Ryzen5 1600 3.2GHz	Broadcom BCM2837 900MHz
コア数	6	4
メモリ	16 GB	1 GB
OS	Ubuntu 18.04LTS	Raspberry Pi OS Buster

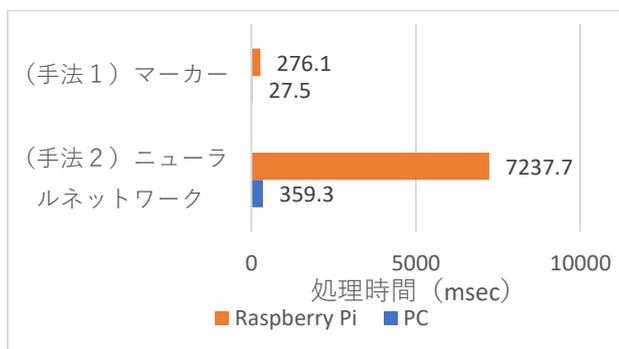


図1: 画像一枚あたりの処理時間

PC環境での画像一枚あたりの処理時間は手法1が手法2に対して13倍高速であるという結果が得られた. また、Raspberry pi環境では手法1が手法2に対して26倍高速であるという結果が得られた.

次に異なる2つの手法の精度を比較した。しかし、手法1はcm単位で推定距離を出力するのに対し、手法2ではdepth画像が出力されるのみで定量的な比較が行えない。そこで、depth画像の明度値から距離に変換する式(1)を作成した。

式(1)は次のように求めた。まず、距離が異なる(10cm~100cm)10種類の画像を6つの地点で撮影し、その画像を手法2で推定する。手法2から推定結果として出力されたdepth画像から明度値を取得する。取得した明度値の平均を取り、近似を取ると10~100cmの距離での明度の近似式となる。この近似式を距離について求めるように変形することで式(1)を求めた。

式(1)は明度を x 、距離を y とする。

$$y = \sqrt[3]{\frac{x}{721.45}} \quad (1)$$

そして、精度を比較するために誤差率を求めた。ここでは誤差率を推定値から正解値を引いた値を正解値で割った値とした。2つの手法の誤差率を図2に示す。

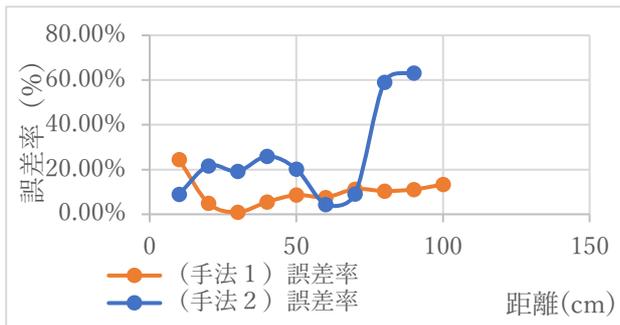


図 2: 2つの手法の誤差率

手法1は誤差率の平均が9.83%、手法2の平均は25.75%となった。また、手法2の誤差率を調べていく中で10種類の画像を比較すると撮影環境の変化によって推定結果に偏りが見られた。

(2) 撮影環境による影響

推定結果に偏りが見られる画像と通常の画像を比較すると、物体と背景の距離が推定に大きく影響を与えている傾向がみられた。検証のために図3の環境でテストした。結果は図4に示す。



背景近 背景遠

図 3: 背景近, 背景遠の具体例

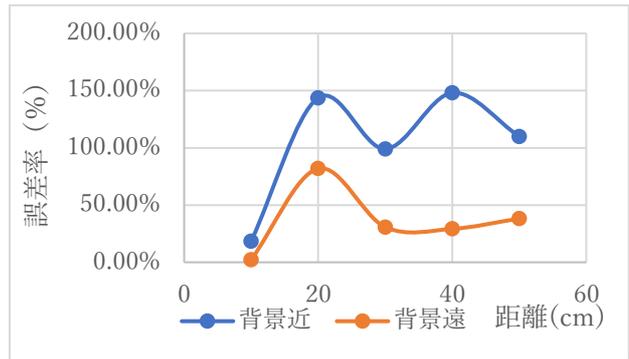


図 4: 背景による誤差率の違い (手法2)

背景との距離がゼロ距離の場合(背景近)は誤差率が平均して103.87%、背景との距離が1mの場合(背景遠)は36.49%となった。これはこのニューラルネットワークがカメラからの深度を推定しているためであると考えられる。物体と背景の位置が離れているほど深度を正確に測れるため、物体と背景が遠いほど誤差が小さくなると考えられる。

4 おわりに

単眼カメラの手法ごとの特性を得ることを目的として、異なる2つの手法の比較と撮影環境による影響の調査を行った。マーカーによる推定法は処理速度が短く、誤差率が平均して9.83%と低い。しかし、マーカーがなければ距離を推定することができないため汎用性が低いことがわかった。一方ニューラルネットワークによる推定法は処理速度が長く、誤差率が平均して25.75%と高いが幅広い場面で距離を推定することが可能なため汎用性が高いことがわかった。さらにニューラルネットワークによる推定法では撮影環境が推定結果に影響を及ぼすことがわかった。背景と物体の距離関係で物体が背景と近い場合は誤差率の平均が103.87%、遠い場合は36.49%であった。これは背景と物体との距離が近い場合、背景と物体が同化しているように誤認しているためだと考えられる。

なお、実際にロボットに搭載した際の性能評価は今後の課題である。

参考文献

- [1] 出澤正徳. ロボットのための距離検出法. 計測と制御, 26.2, pp.103-110, 1987.
- [2] 小林祐一. 単眼カメラによる Visual SLAM の原理と3次元再構成の実装例. システム/制御/情報, 64.2, pp.57-62, 2020.
- [3] BHOI, Amlaan. Monocular depth estimation: A survey. arXiv preprint arXiv:1901.09402, 2019.
- [4] RANFTL, René, et al. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. arXiv preprint arXiv:1907.01341, 2019.
- [5] 出口光一郎, et al. コンピュータビジョンのための幾何学: 1. 画像の生成過程を記述する. 情報処理, 37.6, pp.549-556, 1996.