

単眼カメラで撮影した1枚の画像の収差情報に基づく高精度な距離計測手法

High-Accuracy Distance Measurement Method Using Image Captured by Monocular Camera Based on Aberration Maps

柏木 正子 KASHIWAGI Masako 三島 直 MISHIMA Nao

画像を用いる高精度な距離計測手法として、ステレオカメラを用いるものは、2台のカメラを使用するため小型化が困難であった。また、単眼カメラを用いるものには、深層学習による手法やカラーフィルターを挿入する手法などがあるが、被写体形状に対する汎用性と低コスト化を両立できない点が課題であった。

そこで、東芝は、市販の単眼カメラで撮影した1枚の画像だけを用い、低コストで汎用性が高い、高精度な距離計測手法を開発した。被写体の距離や画像の位置で異なるレンズ収差^(注1)による画像のぼけ情報(以下、収差マップと呼ぶ)を、深層ニューラルネットワークで解析して被写体までの距離を計測する。実際の画像を用いた評価実験により、被写体の形状に依存することなく、ステレオカメラを用いる手法に匹敵する精度で距離計測ができることを確認した。

Methods to perform distance measurement with high accuracy using images are generally classified into (1) methods using a stereo camera system equipped with two cameras, which have been widely used in various applications; and (2) methods using a monocular camera, including a monocular depth estimation method employing deep learning and a color-filter aperture method in which color filters are inserted into the optical path. However, there are significant issues with these approaches in terms of the need for downsizing of the camera systems and resolution of the trade-off between versatility in dealing with objects of various shapes and reduction of costs.

To rectify this situation, Toshiba Corporation has now developed a versatile method to precisely measure a distance at low cost using an image captured by a commercially available monocular camera. This method makes it possible to perform distance measurements based on blur information data related to the lens performance (hereafter referred to as aberration maps), which vary according to the object distance and image position, through analysis of the image applying deep neural networks. We have conducted evaluation experiments using actual images and confirmed that this method achieves distance measurement accuracy comparable to that attained by methods using a stereo camera system without dependence on object shapes.

1. まえがき

近年、ロボットによる物品のピックアップや、工場内の無人搬送車の自律移動、ドローンの遠隔操作によるインフラ点検など、様々な分野で画像センシングの重要性が増している。これらの用途では、被写体の画像を撮影するだけでなく、形状や距離などの高精度の3次元情報を把握できる、小型の装置が求められる。

このようなニーズに対し、幾つかの距離計測手法が検討・実用化されているが、小型化、高精度化、低コスト化を同時に満足する手法として、適切なものがなかった。

そこで、東芝は、市販の単眼カメラを用い、1枚の撮影画像から、被写体の距離や画像位置と、レンズ収差による画像のぼけ方との対応関係を深層ニューラルネットワークで解析し、被写体の距離を計測する距離計測手法を開発した⁽¹⁾

(注1) 物体の1点から出た光がレンズを通過したとき、1点に集まらずに広がりを持つことによる、像のぼけや色ずれのこと。

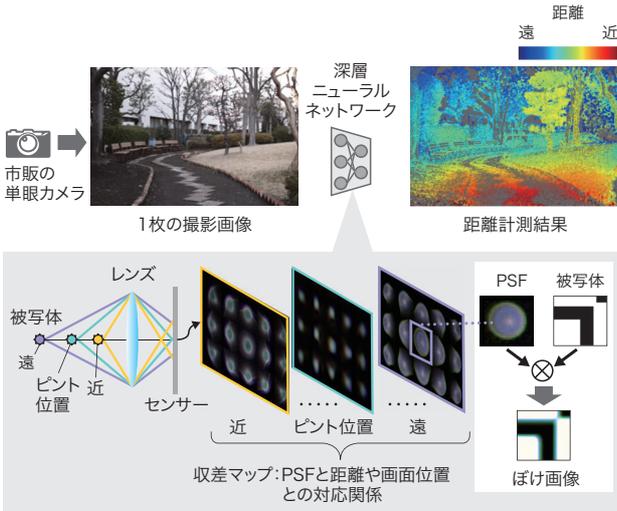
(図1)。ここでは、開発した手法の概要と評価実験について述べる。

2. 従来の距離計測手法

物理的な距離の手掛かりに基づき画像から距離を計測する手法として、2台のカメラを使用するステレオカメラ手法がよく知られている。この手法では、高精度に距離を計測するには、2台のカメラの間隔(眼間距離)を広げる必要があり、原理的に小型化と高精度化の両立が難しい。

また、深層学習の発展に伴い、1台のカメラで撮影した1枚の画像から、風景知識を使って距離を計測する単眼カメラ距離計測手法が、ここ数年で大きく進歩し、注目を集めている⁽²⁾。しかし、この手法は、学習したものとは異なる風景を計測するとき、精度が著しく低下するので、距離計測精度を向上させるためには膨大な学習が必要である。

一方、風景や物の形などの被写体形状ではなく、物理的な距離の手掛かりを利用する手段として、カラーフィルター



PSF: Point Spread Function (点像分布関数)

図1. 収差マップを用いた距離計測の概要

単眼カメラで撮影した1枚の画像から、被写体の距離や画像位置によって異なるPSFと被写体距離の対応関係を解析し、距離を計測する。

Overview of distance measurement by means of aberration maps of image captured by monocular camera

などのデバイスを光学系に追加する手法が提案されている⁽³⁾。これは、ぼけの大きさと色の組み合わせをモデル化して距離を計測する手法で、様々な種類の被写体形状に対して頑健であるが、特殊なデバイスの追加が必要のため、低コスト化が課題であった。

3. 収差マップ距離計測手法

1台のカメラで撮影した画像のぼけ情報を利用して距離を推定する手法としてDepth from Defocus (DfD)が知られている。DfDでは、ピント位置の前後を区別するために、2枚以上の画像が必要と考えられていた。

そこで、複数枚のレンズを組み合わせた市販のカメラレンズ(以下、カメラレンズと略記)には必ず色収差が存在することに着目し、この色収差を、ピント位置の前後を区別する手掛かりとして利用することで、1枚の画像から距離を計測できる手法を見いだした。

収差の特徴の解析には、点像分布関数 (PSF: Point Spread Function) を用いる。PSFは、被写体から生じる1点からの光 (点光源) がカメラレンズを通過した後どのような広がりを持って結像するか (ぼけの形や色づき) を表す関数である。PSFは、カメラレンズに対する被写体の距離や面内位置によって異なる。このPSFの、距離や面内位置との対応関係を、収差マップと呼ぶ。そして、様々な色や輪郭を持つ被写体の画像に、収差マップに従って現れるぼ

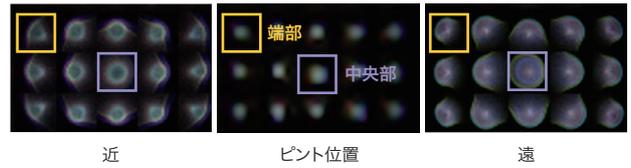


図2. 等間隔に配置された点光源から得られる収差マップの例

PSFは、距離と画面位置によって変化する。

Example of aberration maps of regularly arranged point light sources

けの色づき方や大きさなどの特徴と、距離との対応関係を、深層ニューラルネットワークで解析することにより、被写体までの距離を計測する (図1)。

以下では、収差マップによる距離計測の原理と収差マップを解析する深層ニューラルネットワークについて説明する。

3.1 収差マップによる距離計測の原理

ピント位置より近い・ピント位置・ピント位置より遠いの3か所でPSFを撮影した収差マップを、図2に示す。撮影には焦点距離50mmのカメラレンズを用いた。このPSF画像は、1画素を隣接する画素の光と混ざらない間隔で点灯させたモニターを撮影し、画像内の数か所から抜き出したものである。また、図3は、図2の画面中央部と端部のPSFを一つずつ抜き出し、図中のx方向において赤 (R)・緑 (G)・青 (B) の色別に輝度を数値化したグラフであり、R, G, Bの輝度分布を表す。

これらの図から、PSFの形状は、カメラから被写体までの距離や画面内の位置で色や形状が大きく異なることが分かる。例えば、被写体がカメラのピント位置より近い場合は、輝度分布の中央が凹になる。一方、ピント位置より遠い場合には輝度分布の中央が凸になる。

次に、カメラレンズの光路と色収差の関係について考察する。カメラレンズに色収差が生じている場合のR, G, Bの光線の進み方を、 u がカメラのピント距離 u_f より近い場合、同じ場合、遠い場合に分けて示したのが、図4(a)である。カメラレンズは、低分散の凸レンズと高分散の凹レンズを組み合わせることで、青色と赤色の光を補正して光の広がりを抑制している。しかし、収差を完全に除去するのは困難なため、全ての光が1点に集まることはなく、被写体がピント位置より遠い場合は緑のぼけ幅が最大になり、近い場合は赤と青が混ざった紫色のぼけ幅が最大となる。

図4(b)に、市販の高性能なカメラレンズで被写体を撮影した画像の、部分拡大図を示す。ぼけの色づきは、画像のエッジに現れる。被写体がピント位置より近い場合はパープルフリッジと呼ばれ、エッジが紫色に、ピント位置より遠い場合はグリーンフリッジと呼ばれ、エッジが緑色に見えるこ

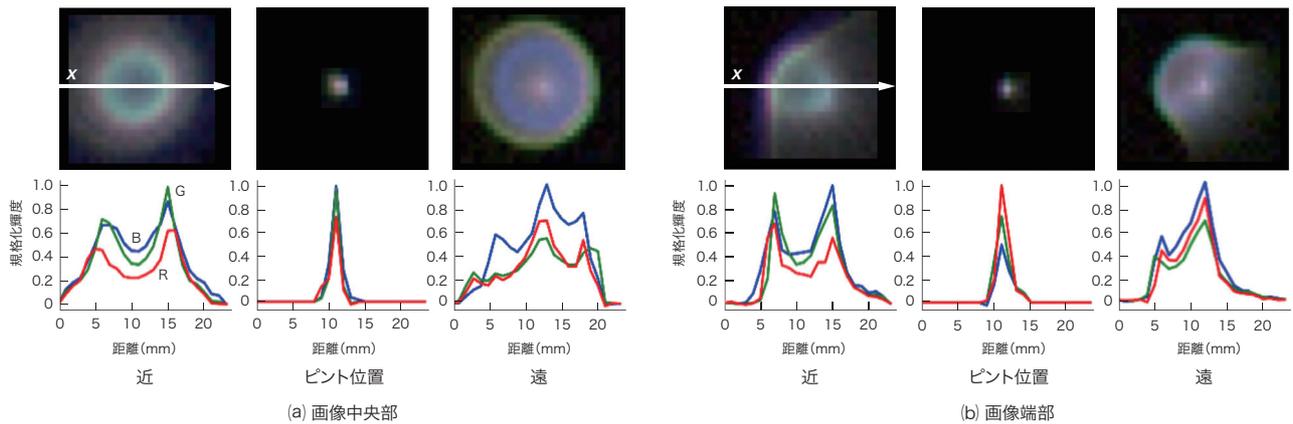


図3. 画面中央と端部のPSF拡大図とRGBの輝度分布

被写体がカメラのピント位置より近い場合は、輝度分布の中央が凹になり、遠い場合は凸になる。

Enlarged views of blurred shape and red-green-blue (RGB) brightness values at image center and edge parts

とが多い。

これらのように、ぼけの色づきの特徴を解析することで、ピント位置の前後のどの位置に被写体が存在するかが推定できる。

ぼけの大きさから距離を求める方法について説明する。レンズの焦点距離を f 、直径を a とし、レンズ中心からイメージセンサーが設置されている撮像面までの距離を v_i 、イメージセンサーの画素ピッチを p とする。イメージセンサーで撮像されるぼけの大きさに相当する画素数 b は、レンズの結像公式により、式(1)で決まる。これにより、 b から被写体の距離 u を求めることができる。

$$b = \frac{av_i}{2p} \left[\frac{1}{f} - \frac{1}{u} - \frac{1}{v_i} \right] \quad (1)$$

3.2 収差マップを解析する深層ニューラルネットワーク

図3に示したように、PSFの形状は複雑で、数値的なモデル化が難しい。そこで、この収差マップの特徴と、被写体までの距離の関係とを解析する深層ニューラルネットワークを開発した。

深層ニューラルネットワークは、事前学習のために、カメラから被写体までの距離にひも付いた学習用データとしての画像が必要となる。そこで、全自動の学習用データ収集システムを構築した。このシステムは、140型モニター、カメラを自動で移動させる12mのスライドステージ、及び自然画像データベースで構成される(図5)。ランダムに大型モニターで自然画像を表示してカメラで撮影し、そのときの距離にひも付けて記録することで、学習用データが全自動で収集できる。学習用データは、カメラのピント位置1.5mに対

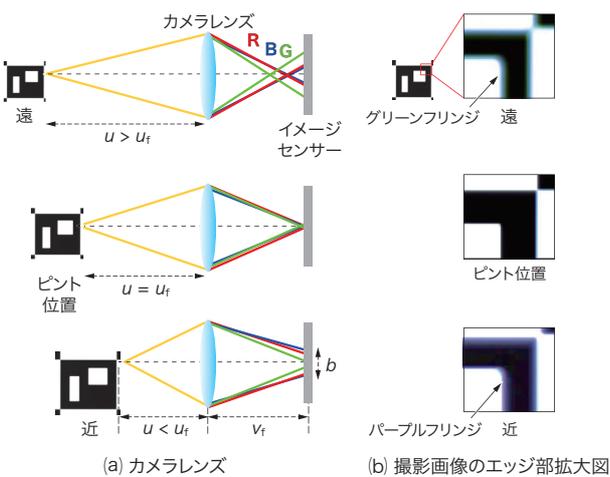


図4. カメラレンズに色収差がある場合のエッジのフリッジの違い

被写体がピント位置より遠い場合は、エッジにグリーンフリッジが、近い場合はパープルフリッジが現れる。

Differences in color fringes depending on distance to object due to chromatic aberration of achromatic lens

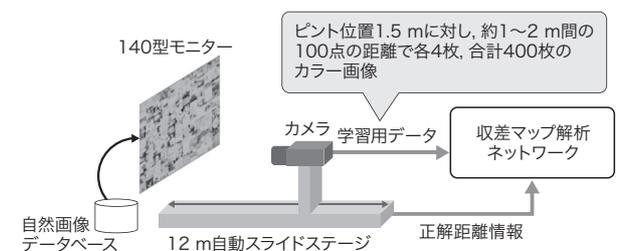


図5. 学習用データ収集システム

スライドステージ上のカメラを自動で移動させながら、様々な色や形状のカラー画像を撮影して学習用データを収集する。

Automatic calibration data acquisition system

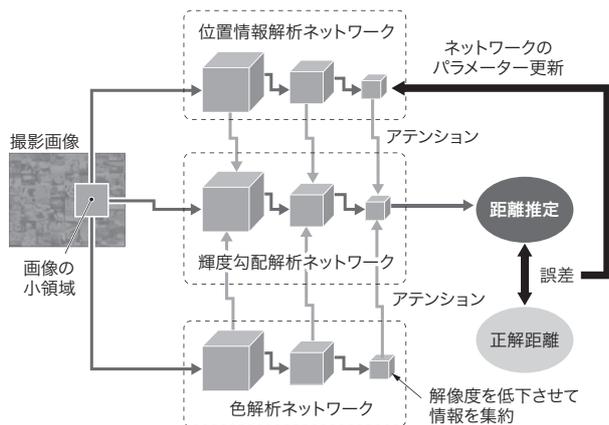


図6. 収差マップ解析用の深層ニューラルネットワークの概要

収差に基づくぼけの位置や距離依存性などの特徴を効率良く解析するために、位置情報、輝度勾配、及び色を解析する三つのネットワークで構成する。

Outline of deep neural networks for aberration map analysis

し、約1～2 m間の100点の距離で各4枚撮影した、合計400枚のカラー画像である。このようにして準備した学習用データを、深層ニューラルネットワークに入力して、学習を実行する。

図6に、開発した深層ニューラルネットワークの概要図を示す。3.1節に述べたように、収差に基づくぼけの特徴を詳細に解析するため、エッジの特徴を解析する輝度勾配解析ネットワーク、位置に依存したぼけの特徴を解析する位置情報解析ネットワーク、及びぼけの色づきを詳細に解析する色解析ネットワークで構成する。風景知識に依存せず、ぼけ情報だけを解析する工夫として、ランダムに選択した16×16ピクセルサイズの小領域の画像をそれぞれのネットワークに入力する。各ネットワークにおいて、入力画像にフィルター処理を行って特徴量を抽出し、特徴を大まかに整理するために解像度を低下させて情報を圧縮する処理を繰り返しながら特徴量を集約していく。また、位置情報解析と色解析のネットワークからの出力は、アテンションと呼ばれる重み付け機構を経て輝度勾配解析ネットワークに入力される。アテンション機能は、当該撮影画像の位置と色において、輝度勾配のどこを重視すればぼけ形状変化の影響を小さくできるかを制御するために活用される。学習の過程では、深層ニューラルネットワークで推定した距離と学習用データ収集システム(図5)で収集した正解距離との誤差が小さくなるようにネットワークのパラメーターを更新する。

4. 評価実験

市販の一眼レフカメラに焦点距離50 mmのカメラレンズ

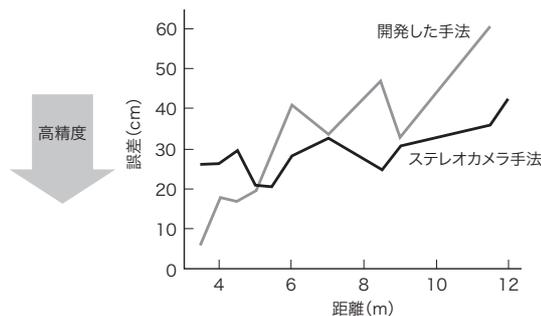


図7. 距離精度の評価実験結果

開発した手法で得られる距離精度の平均誤差は、ステレオカメラ手法と同程度であった。

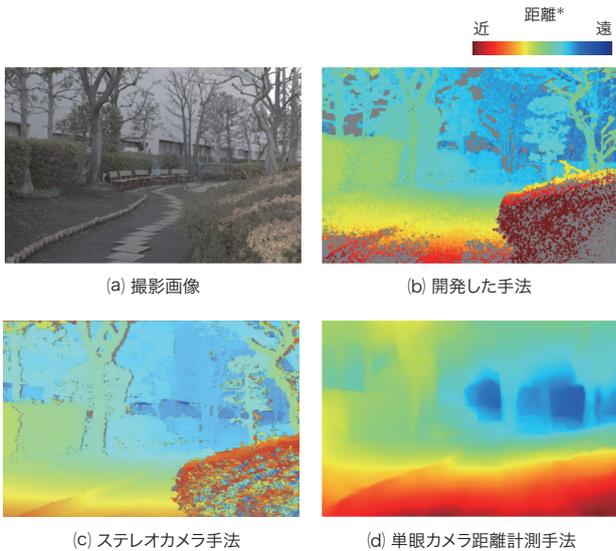
Results of evaluation experiments on accuracy of distance measurements obtained by conventional method using stereo camera and newly developed method using monocular camera

を取り付け、ピント距離を6.5 m、カメラの絞り(F値)を4に設定し、画素数を1/4に縮小した120万画素の画像を用いて定量評価を行った。学習用データ収集システムの自動スライドステージを活用し、設定距離4～12 mに置いた平面画像を撮影して距離を計測し、自動スライドステージの距離と比較して誤差を算出した。また、参考のため、眼間距離25 cmのステレオカメラ手法でも同様の評価を行った。

各距離での距離計測結果の誤差を、図7に示す。設定距離範囲内の距離計測誤差の絶対値は、開発した手法では平均22.3 cmであったのに対し、ステレオカメラ手法では平均27.0 cmであった。ステレオカメラ手法は、開発した手法に比べ、遠方での距離精度は高いが、近傍は2台のカメラで撮影できる範囲が狭いため、距離精度が低い。

更に、被写体を屋外で撮影し、3種類の手法に対して初めて解析するシーンを用いて距離を計測した結果を、図8に示す。単眼カメラ距離計測手法は、ほかの二つの手法に比べて、距離推定値の解像度が低く、被写体の輪郭部周辺の誤差が大きい。また、ステレオカメラ手法は遠方の精度は高いが、近くでは高精度に解析できる範囲が狭くなり、距離の誤差が大きくなる。一方、開発した手法は、ステレオカメラ手法と同様に細かい輪郭を持つ被写体であっても、近くから遠くまで高精度に計測ができている。

これらの結果から、開発した手法は、ステレオカメラ手法に匹敵する距離計測精度であることが分かった。また、単眼カメラ距離計測手法では不得手な、空や地面がない画像や未学習のシーンについても、被写体の形状によらず距離計測ができる。開発した手法は、ぼけ情報を解析するため、輝度変化がない領域ではぼけが生じないため計測は不得手であるが、ステレオカメラ手法や単眼カメラ距離計測手法に



*赤色が近距離、青色が遠距離を表し、距離推定に適さない計測対象外の領域は灰色で示す

図8. 屋外撮影画像による距離計測結果

開発した手法は、被写体の形状によらず、近場から遠方まで満遍なく距離計測ができる。

Distance estimation results using outdoor images captured by stereo and monocular cameras

比べると、近場から遠方まで満遍なく距離計測ができることを確認した。

5. あとがき

単眼カメラで撮影した1枚の画像から距離情報を取得できる、取差マップを利用した距離計測手法を開発した。開発した手法は、特殊なデバイスの追加が不要で、低コストかつ高精度に距離計測が実現できる。精度評価実験では、開発した手法の計測精度が高いこと、学習していない画像でも計測が可能であること、及び近くから遠くまで広い範囲を計測できることが確認できた。

今後、小型カメラに適用するとともに、計算量削減の検討を進め、早期の実用化を目指していく。

文献

- (1) Kashiwagi, M. et al. "Deep Depth from Aberration Map". International Conference on Computer Vision 2019 (ICCV 2019). Seoul, Korea, 2019-10, IEEE. The Computer Vision Foundation (CVF), 2019, p.4070-4079.
- (2) Z. Li, Z.; Snavely, N. "MegaDepth: Learning Single-View Depth Prediction from Internet Photos". Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018). Salt Lake City, UT, 2018-06, IEEE. 2018, p.2041-2050.
- (3) 三島 直, 佐々木貴之. 単眼カメラで撮影した1枚の画像から精度良く距離計測できるカラー開口撮像技術. 東芝レビュー. 2018, **73**, 1, p.39-43. <https://www.toshiba.co.jp/tech/review/2018/01/73_01pdf/t02.pdf>, (参照 2020-01-07).



柏木 正子 KASHIWAGI Masako
研究開発センター 知能化システム研究所
メディア AI ラボラトリー
Media AI Lab.



三島 直 MISHIMA Nao
研究開発センター 知能化システム研究所
メディア AI ラボラトリー
日本液晶学会会員
Media AI Lab. ー