深層学習を用いた全天球パノラマ画像からの自己位置推定

梅田 将孝*・伊達 央**

Spherical Panoramic Image-based Localization by Deep Learning

Masataka UMEDA* and Hisashi DATE**

The goal of our research is to construct intelligence for autonomous robots which navigate through the real world using images from attached cameras. Localization is one of the key elements for autonomous navigation. In this paper, we propose a grid-based localization from a single image using deep learning. By using a grid map, uncertainty of localization can be defined in a natural way, which is useful for fusion with other sensors. The network takes spherical panoramic images with position and orientation of the robot as training data. Position and orientation are expressed by multi-dimensional grid. The network behaves as a classifier, where each grid corresponds to individual class and its probability represents that of position and orientation. The experimental results support the effectiveness of the proposed method.

Key Words: deep learning, localization, mobile robot, CNN

1. 序 論

任意の目的地へ自律的に移動するロボットにはできるだけ正確な自己位置推定が求められる。そのためには、初期値が不明なときや地図とのマッチングから外れたときに、現在得られるセンサデータのみから自己位置を推定する、大域的な自己位置推定が必要である。自己位置推定は測域センサやカメラを用いて行なわれているが、カメラ画像は測域センサよりも多くの情報を含んでいるにもかかわらず処理が複雑になるためこれまで十分に活用されていなかった。

しかし近年、画像認識において深層学習が、物体認識¹⁾や領域分割²⁾、距離画像の推定³⁾などにおいて高い認識率を達成しており、この技術を用いることでカメラ画像からの大域的な自己位置推定ができる可能性がある。また、測域センサよりも安価な web カメラを使うことはコストダウンにもつながる。

そこで本研究では、深層学習を用いて市販の全天球カメラにより取得される画像 1 枚から自己位置を推定する方法を提案する。既報⁴⁾において筆者らは深層学習を用いた全天球パノラマ画像からの自己位置推定に関する基礎実験を行なっ

おり、全天球パノラマ画像を用いることで方位の推定性能が向上することを確認している。本論文では、ネットワークの出力における位置や方位をグリッドとして表現することにより、自己位置推定を分類問題として定式化する。これにより、推定値の分布を見ることによって推定の不確実性を求めることができ、オドメトリなどのほかの手法により求めた推定値との融合が可能になる。また学習時に、教師信号として分布そのものを与えることができるため、学習用のデータを取得した経路から遠い地点での推定精度が向上すると考えられる。屋内および屋外環境における実験により、これらの妥当性を確認するとともに、大域的な自己位置推定がどの程度まで可能なのかを明らかにする。

2. 関連研究

移動ロボットの自己位置推定に関する研究は盛んに行なわれており、MCL (Monte-Carlo-Localization)⁵⁾は最も有名なアルゴリズムの 1 つである。こうした研究の多くでは測域センサが用いられているが、画像を用いる場合もあり、局所特徴量を用いてデータベースを作成し、取得した画像とのマッチングにより自己位置を推定している。たとえば、全天球パノラマ画像を用いる手法⁶⁾や、Structure from Motion を用いる手法^{7), 8)}、ランドマークのデータベースを構築することで大域的な自己位置推定のエラーを減らす手法⁹⁾がある。ほかにも、画像内に含まれる顕著な特徴表現の生成モデルを自己位置推定に取り入れる手法¹⁰⁾がある。ここでは SIFT¹¹⁾のような位置や回転の変化、照明の変化などに対して頑健な特徴量がよく用いられている。局所特徴量を用いた大域的な自己位置推定において推定精度を上げるためには、日照条件や人

* 筑波大学大学院システム情報工学研究科
つくば市天王台 1-1-1

** 筑波大学システム情報系 つくば市天王台 1-1-1

* Graduate School of Systems and Information Engineering,
University of Tsukuba, 1-1-1 Tennodai, Tsukuba

** Faculty of Engineering, Information and Systems, Uni-
versity of Tsukuba, 1-1-1 Tennodai, Tsukuba

(Received July 2, 2017)

(Revised October 4, 2017)

物やオブジェクトの移動を考慮してより多くの画像を取得する必要があるが、結果的に取得した分だけデータベースが大きくなり、推定をする上での処理時間が単調に増加する。一方で、本研究における提案手法は画像のデータに依らず、推定時の処理時間が増えることはない。ただし、現状ではリアルタイムで動作させるには GPU を用いた並列計算が必要である。

深層学習を用いた自己位置推定に関する研究もいくつかあり、その 1 つとして、大域的な自己位置推定の誤差を減らすために深層学習を用いて物体認識を行ない、その結果を MCL のパーティクルの重みの更新に利用する手法¹²⁾がある。しかし、これはあくまで補助的な役割として深層学習を利用しているに過ぎない。ほかにも、Alex らは 1 枚のカメラ画像からその撮影位置を推定する PoseNet と呼ばれるネットワークを提案している¹³⁾。さらに Alex らは、1 枚の画像から一意でない複数の推定値を出力できるように拡張している¹⁴⁾。複数の推定値を求め、その分散を見ることで推定における不確実性を推定できると Alex は述べているが、複数の推定値を求めるために 1 枚の画像を何度もネットワークに入力する必要があるため、計算量が増えてしまうという問題がある。また、使用しているカメラはスマートフォンに搭載されているような普通のカメラであるため、本研究で使用する全天球カメラに比べて画像 1 枚に含まれる情報量が少ないといえる。PoseNet や提案手法は 1000 種類の物体を識別するネットワークをベースにしているため、さまざまな環境に適応できるスケーラビリティに優れている。自己位置推定に限らずさまざまなタスクに適用可能なものとして、深層学習を用いた特徴量抽出の手法もあり、ネットワークの中間層の情報を利用する手法¹⁵⁾や、特徴抽出器そのものをネットワークに学習させる手法¹⁶⁾が提案されている。深層学習は学習の過程でデータセットに適応した、より優れた局所特徴量の抽出ができると考えられる。深層学習を特徴抽出器として捉えているこれら 2 つの手法と本論文で述べる提案手法との違いについて、提案手法は自己位置推定の結果と真値との誤差に応じて学習が行なわれるため、自己位置推定に特化していることが挙げられる。したがって、データセットだけでなくタスクにも適応した特徴抽出ができるようになると考えられる。

深層学習ではなく、ランダムフォレストを用いて画像からの自己位置推定を回帰で行なう手法¹⁷⁾も提案されているが、こちらは RGB 画像と Depth 画像を併用して推定している。

本実験では既存の自己位置推定手法を用いずに、深層学習をベースとした大域的な自己位置推定がどこまでできるのかを明らかにする。本論文では PoseNet を用いた実験および、新たに提案する自己位置推定のネットワークを用いた実験についての比較を行なったので、それについて報告する。

3. PoseNet

PoseNet では自己位置推定を回帰問題として解いている。物体認識のネットワークとして有名な GoogLeNet¹⁸⁾の出力層を変更したもので、1 枚のカメラ画像を入力し、位置 (x, y)

Fig. 1 PoseNet used in this experiment

と方位 $(\cos \theta, \sin \theta)$ を出力する。学習におけるパラメータの更新方法としては、推定値と真値との差分を用いた二乗誤差関数を損失関数として、誤差逆伝播法を適用する。具体的にはネットワークが推定した位置を $\mathbf{p} = (x, y)$ 、推定方位を $\mathbf{q} = (\cos \theta, \sin \theta)$ 、位置と方位の真値をそれぞれ $\hat{\mathbf{p}}, \hat{\mathbf{q}}$ とすると、損失関数は以下の式で表わされる。

$$Loss(\mathbf{I}) = \|\hat{\mathbf{p}} - \mathbf{p}\|_2 + \beta \|\hat{\mathbf{q}} - \mathbf{q}\|_2 \quad (1)$$

ここで、 $\mathbf{I}$ は入力画像を示しており、 β は方位誤差の位置誤差に対する重みを決定するハイパーパラメータとなる。この損失関数の値が小さくなる方向にパラメータを更新していくことで、自己位置推定器を獲得する。

このネットワークは主に畳み込み層と全結合層で構成されており、畳み込み層ではフィルタをかける演算が行なわれることで、局所的な特徴を抽出することができ、全結合層ではその抽出された特徴を統合することで、入力画像に対応した位

Fig. 2 The concept of the proposed network output. This network estimates position and orientation of the robot from a single spherical panoramic image. This network behaves as a classifier, which each grid corresponds to individual class and its probability represents that of position and orientation

位置と方位を推定する。本実験ではあまり多くのデータを集めることができなかったため、比較的層の浅い AlexNet¹⁾を改造することで PoseNet と同様の性質をもつネットワークを構築した。ネットワークの構造を **Fig. 1** に示す。input および output の横に記載されている (None, 300, 300, 3) といった数字は InputLayer や Convolution2D などの各レイヤーにおける入力次元および出力次元のことである。None は、引数として任意の正の整数を期待する表現であり、Fig. 1 において、一度のネットワークの順伝播において入力する画像の枚数を表わしており、本実験では 16 とした。InputLayer は画像の読み込みを行なう層で、Convolution2D は畳み込み層、LRN2D は Local Response Normalization と呼ばれる層で、値が飽和しないように正規化する役割を担う。MaxPooling2D はプーリング層である。Flatten はつぎの Dense (全結合層) で演算を行なうためにデータの表現方法を変える層である。学習時に β を用いて位置誤差と方位誤差に重み付けをする処理を簡潔に記述するために、途中で枝分かれさせている。以下では、この AlexNet を改造したものを PoseNet と呼ぶことにする。AlexNet もまた物体認識のネットワークとして有名であり、GoogLeNet²⁾ほどではないが高い認識精度を有する。

PoseNet の特徴として、推定値に対する不確実性を求めることができるという点がある。ネットワークのパラメータの一部を一時的に 0 に置き換える Dropout と呼ばれる手法を用いて、1 つの入力値に対して複数の出力値を得られる枠組みになっている。本来 Dropout は、推定を行なう上である特定のパラメータに依存するのを防ぐことで過学習を抑制するために用いられる。PoseNet ではこの複数の出力値の分散を推定における不確実性としている。

4. 提案手法

提案手法では自己位置推定を分類問題として解いている。

PoseNet と同様に AlexNet¹⁸⁾の出力層を変更したものであるが、直接位置 (x, y) と方位 $(\cos \theta, \sin \theta)$ を推定するような構造ではなく、**Fig. 2** で示すような構造になっている。位置推定については正方形のグリッドを敷き詰めた xy 平面で表現される環境を離散化し、各グリッドにおけるロボットの存在確率を出力する。色が濃いほどそのグリッドにいる確率が高いことを表わしている。方位推定については、 -180 度から 180 度までの方位空間を離散的に区切ってグリッド状にした上で、方位とグリッドとの対応関係に基づいてロボットの方位を推定する枠組みになっている。Fig. 2 の方位推定における縦軸は確率値を示している。ネットワークの構造を **Fig. 3** に示す。PoseNet と異なる点として、Softmax 層の追加が挙げられる。これは出力層の各ユニットの値が確率の公理を満たすように調整する層であり、グリッドそのものをなしている。出力層の次元は環境の広さやグリッド 1 辺の長さまたは方位の範囲によって変化するため、Fig. 3 の値に一致するとは限らない。グリッドの確率分布から共分散行列を求めることができ、推定の不確実性を定義できる。この不確実性が実際の誤差を反映するならば、オドメトリなどの融合において有用である。

学習を行なうためには、ネットワークに教師信号として正しい自己位置 $(\hat{x}, \hat{y})$ および $\hat{\theta}$ をグリッドマップの形で与える必要がある。提案手法では、各グリッドの中心の座標値と真値との距離から正規分布に従って確率値を与えている。仮に正しい自己位置に対応する 1 つのグリッドのみに確率値として 1 を割り当てると、提案手法における損失関数 (2) では推定値と真値との距離を考慮することができない。ロボットが実際にいるグリッドの位置から 1 つずれたグリッドにいと誤認識する場合と、遠く離れたグリッドにいと誤認識する場合を考えたとき、一般的に前者は後者よりも許容されるべきである。そこで、正規分布に従って隣接するグリッドを含めて確率値を与えることでグリッドに連続性をもたせることができるので、結果として推定精度が向上すると考えられる。ネットワークが出力した位置および方位における各グリッドの確率値をそれぞれ p_i, q_j とし、教師信号を $\hat{p}_i, \hat{q}_j$ とすると、損失関数は以下の式で表わされる。 $\mathbf{I}$ と β の意味は (1) 式で用いたものと同じである。また、log は自然対数である。

$$Loss(\mathbf{I}) = -\left(\sum_i \hat{p}_i \log(p_i) + \beta \sum_j \hat{q}_j \log(q_j)\right) \quad (2)$$

教師信号として与える各グリッドの確率値 $\hat{p}_i, \hat{q}_j$ は、グリッドの中心の点の位置座標を $(\hat{x}_i, \hat{y}_i)$ 、同様に中心の方位を $\hat{\theta}_j$ 、位置の教師信号に適用する正規分布の共分散行列を $\Sigma_{position}$ 、方位の教師信号に適用する正規分布の標準偏差を σ_{angle} とすると、以下の式で求められる。

$$\mathbf{d}_i = \begin{pmatrix} \hat{x} - \hat{x}_i \\ \hat{y} - \hat{y}_i \end{pmatrix} \quad (3)$$

$$\hat{p}_i = \frac{1}{Z_p} \exp\left(-\frac{1}{2} \mathbf{d}_i^T \Sigma_{position}^{-1} \mathbf{d}_i\right) \quad (4)$$

Fig. 3 Proposed network

$$\hat{q}_j = \frac{1}{Z_q} \exp \left(-\frac{(\hat{\theta} - \tilde{\theta}_j)^2}{2\sigma_{angle}^2} \right) \quad (5)$$

Z_p や Z_q は位置と方位のそれぞれにおいて確率値の総和が 1 になるようにするための正規化定数である。

5. 実験

5.1 データセットの構築

2 種類の実環境からカメラ画像と自己位置・方位の値のペアによって構成されるデータセットを構築した。1 つ目は研究室で、その広さは約 7.0 m × 5.0 m である。2 つ目は屋外で、筑波大学にある体育館の周囲を利用した (Fig. 4)。広さは約 90.0 m × 90.0 m である。どちらも、コントローラを用いてロボットに 3, 4 周させたうちの何周かを訓練用、残りをテスト用というように振り分けた。使用するカメラとして、全

Fig. 4 The route of the robot when acquiring training data. The left figure shows laboratory environment. The right figure shows outdoor environment

Fig. 5 Robots used in this experiments. A robot in the left figure was used in the laboratory environment. A robot in the right figure used in the outdoor environment

天球の画像 (Fig. 5) を取得できる (株) リコー製の RICOH THETA を用いた。また、データセット構築時の自己位置の真値の取得と、そのために必要な地図の生成を行なうために、研究室環境では北陽電機 (株) 製の測域センサ URG-04LX を用い、屋外環境では多層型レーザスキャナである Velodyne の VLP-16 を用いた。これらのセンサを搭載したロボットの外観を Fig. 6 に示す。両方のロボットともカメラの高さは地面から約 1.1 m 程度である。RICOH THETA は 1 枚の画像から全方位の情報が得られるため、1 枚の画像に含まれる情報量が多く、位置座標を変えずに方位を変えても正しく推定できると考えられる。しかし、取得される画像は魚眼画像になっており、同じ物体でも写る位置によって形状が異なって見えるという歪みの問題が生じる。効率良く学習を行なうためには、この歪みに対する何らかの補正が必要であり、今回は射影変換を行なうことで解決を図った。具体的には、画像に写る 2 つの魚眼レンズからの景色をそれぞれ切り出して、正距円筒図法に従って個別に射影した 2 つの画像をつなぎ合わせて、1 つの画像にした (Fig. 7)。また、訓練用の画像をすべて足し合わせて、画像の数で割ることによって平均画像を求めた。

データセットに多くのバリエーションがあることは、学習における汎化性能の向上に大きな影響を与える。しかし、大量のデータを取得することは大きな負担を要する作業であり、その対策の 1 つとして画像処理によって擬似画像を生成することで、データセットにより多くのバリエーションをもたせるデータ拡張という手法がある。今回の自己位置推定というタスクにおいては、全天球のパノラマ画像を水平方向にシフ

Fig. 6 Image before projective transformation

Fig. 7 Image after projective transformation

トさせて、異なる方位から見た景色を人工的に生成することでデータのバリエーションを増やすことができる。以前に著者らがこのデータ拡張⁴⁾を行ない、1枚の画像に対して45度刻みで向きを変えた画像を新たに7枚生成して比較実験を行なったところ、データ拡張による推定精度の向上が見られたため、本実験においても同様の条件でデータ拡張を行なった。データセットの内訳を **Table 1** に示す。

Table 1 Training and test dataset

environment	training	training with data augmentation	test
laboratory	773	6184	793
outdoor	916	7328	434

5.2 提案手法における教師信号の与え方

4章で示した教師信号の求め方に従って、位置と方位に関する教師信号を生成した。提案手法においてユーザが決める必要のあるパラメータに、教師信号を生成する際に用いる正規分布の分散とグリッド1辺のサイズがある。正規分布の分散を位置については x 方向と y 方向の両方とも同じ値に設定し、研究室環境においては教師信号生成時の正規分布の分散を変えた場合の違いを見るために 0.5m^2 と 1.0m^2 の2種類、屋外環境においては 1.0m^2 の1種類で実験を行なった。どの種類においても x と y の共分散は 0.0m^2 にして、方位については 10.0deg^2 とした。なお、位置推定におけるグリッド1辺の長さは研究室環境では 0.5m で、屋外環境では 5.0m にした。ただ、6.3節において、グリッドサイズの違いによる影響を調べるために研究室環境のグリッド1辺の長さを 1.0m にして学習を行ない、比較を行なった。方位推定におけるグリッドは両環境のすべての条件において10度の刻み幅に設定した。一例として、研究室環境内のとある画像に対する位

Fig. 8 An example of data labeled position information in laboratory environment

置の教師信号としてのグリッドマップを **Fig. 8** に示す。各グリッドに書かれている数値は、そのグリッド内にロボットがいる確率値を表わしている。

5.3 検証項目

本実験では、PoseNet および提案手法を用いた自己位置推定の精度を比較するとともに、共分散行列を求めることによって推定に対する不確実性を測ることができるかを検証した。訓練用データセットを取得したときのロボットの走行経路(以下、訓練時走行経路と呼ぶ)に近い位置で取得した画像は正しい推定値を求めやすい反面、遠い位置で取得した画像はネットワークの学習時に与えた画像との違いが大きくなっているため、推定が外れやすいと考えられる。このとき、訓練時走行経路との近さに応じて、出力されるグリッドマップにおける分布も異なると考えられる。すなわち訓練時走行経路から遠ざかるほど、グリッドマップから求まる推定値の分散が大きくなり、推定される自己位置の誤差楕円や方位の標準偏差が大きくなると考えられるため、その検証も行なった。

なお、各条件における β の値を **Table 2** に示す。訓練用データセットのすべての画像に対してパラメータの更新を行なう操作を1 epoch とするとき、すべての条件において50 epoch だけ回して学習を行なった。学習および評価において、GPUはNVIDIA社製のGeForce GTX TITAN Xを、CPUはIntel社製のCore i7-6700Kを使用した。

Table 2 Value of β

environment	network	β
laboratory	PoseNet	3.0
	Proposed Network	1.0
outdoor	PoseNet	500.0
	Proposed Network	1.0

提案手法におけるグリッド1辺のサイズを変えた場合も違いを見るために、研究室環境において1辺の長さを 1.0m にした場合における学習も行なった。

6. 実験結果

6.1 位置座標の推定結果

研究室および屋外環境のテスト用データセットを用いて、位置座標に関する推定値と真値との比較を行なった。PoseNet

Fig. 9 The comparison of position estimation results in the laboratory environment. The top left figure shows the result by PoseNet. The top right figure shows the result by the proposed method in case that variance is 1.0 m^2 . The lower figure shows that in case that variance is 0.5 m^2

Fig. 10 The comparison of position estimation results in the outdoor environment. The left figure shows the result by PoseNet. The right figure shows the result by the proposed method

の場合は推論時にも Dropout を適用することで 1 枚の画像から複数の推定値を求めることができる。今回は 100 個の推定値を求め、平均ベクトルと共分散行列を求めた。その平均ベクトルを最終的な推定値とみなし、真値との距離 (RMS) および最大距離を算出した。提案手法では環境をグリッド状に区切った上で、各グリッドに対してそこにロボットがいる確率が求められ、さらに各グリッドの確率の和が 1 になる性質がある。したがって、各グリッドに対応する位置ベクトルと確率との重み付きの和をとることで推定値の期待値を求め、その期待値を用いて共分散行列を算出した。期待値を最終的な推定値とみなし、真値との誤差 (RMS) および最大誤差を算出した。その結果を **Table 3** に示す。両手法とも同程度の精度で推定ができていることが確認できた。提案手法における教師信号の分散として 1.0 と 0.5 の 2 つの条件で実験を行なった研究室環境においては、位置の教師信号として与える分散が小さければ推定精度も上がるというわけではなく、むしろ最大誤差が大きくなる結果となった。

つぎに、推定値と真値とをグラフにプロットしたものを

Fig. 11 The position estimation and the error ellipse in the laboratory environment. The top left figure shows the result by PoseNet. The top right figure shows the result by the proposed method in case that variance is 1.0 m^2 . The lower figure shows that in case that variance is 0.5 m^2

Fig. 12 The position estimation and the error ellipse in the outdoor environment. The left figure shows the result by PoseNet. The right figure shows the result by the proposed method

Fig. 9, Fig. 10 に示す。なお、見やすさのためにデータをいくつか間引いている。丸い点が推定値で、×の点が真値である。2 点を結ぶ黒い線は、推定値と真値との対応関係を表わしている。また、求めた共分散行列をもとにして、ある信頼区間におけるロボットの自己位置座標として取りうる範囲をプロットしたものを **Fig. 11, Fig. 12** に示す。推定値を表わす丸い点を中心として描かれている楕円は、長半径および短半径がネットワークから出力されるグリッドマップの分布を正規分布で近似したときの標準偏差に相当している。PoseNet では全体的に楕円が小さく、推定誤差が大きい地点でも確信をもって誤った推定をしていた。提案手法では全体的に楕円が大きくなり、推定誤差が大きいほど楕円も大きくなった。また、位置の教師信号として与える分散が小さければ、全体として楕円も小さくなった。

6.2 方位の推定結果

PoseNet の場合は位置推定と同様に 100 個の推定値から求めた平均ベクトルと真値 $(\cos \theta, \sin \theta)$ から方位 θ をそれぞれ算出し、その方位の値を用いて誤差を算出した (**Table 3**)。推定値と真値が一致すれば 0 度、完全な不一致の場合は 180 度

Table 3 Position and orientation estimation error

network	environment	step size of one grid (m)	variance of position (m^2)	positional error		orientational error	
				mean error (m)	max error (m)	mean error (deg)	max error (deg)
PoseNet	laboratory	N/A	N/A	0.253	1.845	5.844	57.840
GridPoseNet	laboratory	0.5	1.0	0.359	1.538	5.566	59.324
GridPoseNet	laboratory	0.5	0.5	0.383	2.835	5.566	55.584
PoseNet	outdoor	N/A	N/A	2.341	8.434	2.835	9.586
GridPoseNet	outdoor	5.0	1.0	4.036	9.593	1.600	9.793

Fig. 13 Relationship between distance from the route of the robot when acquiring training data and position estimation error on PoseNet and the proposed method **Fig. 15** Relationship between distance from the route of the robot when acquiring training data and area of the error ellipse on PoseNet and the proposed method **Fig. 14** Relationship between distance from the route of the robot when acquiring training data and position estimation error on the proposed methods in different variance **Fig. 16** Relationship between distance from the route of the robot when acquiring training data and area of the error ellipse on the proposed methods in different variance

の誤差になる。また、提案手法についても位置推定と同様に各グリッドに対応する方位と確率との重み付きの和をとることで推定の期待値を求め、真値と比較した。位置推定と同様に、両手法とも同程度の精度で推定できていた。今回、研究室環境のデータセットの教師信号に異なる分散を与えているのは位置に対してであり、方位に対してはどちらも 10.0 deg^2 で固定にしていたので、方位推定の誤差についてはそれほど大きな違いが見られなかった。

6.3 位置推定における不確実性

テスト用データセットにおいて、訓練時走行経路から遠い位置にて撮影した画像については推定精度が下がる傾向にあると考えられる。これを確かめるために研究室環境のデータセットに対して縦軸を推定誤差、横軸を訓練時走行経路からの距離としてデータをプロットしたものを **Fig. 13**, **Fig. 14** に示す。Fig. 13 では、PoseNet と提案手法 (位置の教師信号における分散が 1.0) の比較をした。両手法において訓練時走行経路からの距離が大きいくほど推定誤差も増加する傾向にあることが見て取れる。また、Fig. 14 では、提案手法における

位置の教師信号における分散の違いに関して比較をした。分散が小さいほど、推定誤差の増加量が大きくなる傾向にあり、訓練時走行経路から離れた地点における推定誤差が大きくなることを示している。訓練時走行経路からの距離と、6.1 節で定義した誤差楕円の面積との関係をプロットした **Fig. 15**, **Fig. 16** から、提案手法では訓練時走行経路から遠ざかるほど推定の不確実性が大きくなっていくことが確認できた。それに対して、PoseNet では不確実性の増加を確認できなかった。Fig. 16 より、位置の教師信号における分散が小さいと、訓練時走行経路に近い地点でも誤差楕円の面積が大きくなる場合があることが確認できた。これは正しい自己位置に対応するグリッドの周囲にも 0 より大きい確率値を割り当てる本手法の有効性を示唆している。

推定誤差が大きい場合においても、訓練時走行経路からの距離が大きいくときにその推定に対する不確実性も大きくなるということがネットワークの出力から求められると、その指標を用いて、その推定値をどれだけ信用するかといった重み付けに利用することができる。そのような利用に向けた可能性を

Fig. 17 Relationship between distance from the route of the robot when acquiring training data and mahalanobis distance in position on PoseNet and the proposed method

Fig. 18 Relationship between distance from the route of the robot when acquiring training data and mahalanobis distance in position on the proposed methods in different variance

確かめるために研究室環境のデータセットに対して縦軸をマハラノビス距離、横軸を訓練時走行経路からの距離としてデータをプロットしたものを **Fig. 17**, **Fig. 18** に示す。Fig. 17 では、PoseNet と提案手法 (位置の教師信号における分散が 1.0) の比較をした。Fig. 18 では、提案手法における位置の教師信号における分散の違いに関して比較をした。本論文におけるマハラノビス距離は、ネットワークの推定から求めた平均ベクトル μ と共分散行列 Σ から 2 次元の正規分布を近似したときに、位置の真値 t における確率密度の小ささを表わす距離のことであり、以下の式で表わされる。

$$D(t) = \sqrt{(t - \mu)^T \Sigma^{-1} (t - \mu)} \quad (6)$$

訓練時走行経路からの距離に依らずマハラノビス距離の分散が 1 に近ければ、推定器の出力する分布の標準偏差が実際の誤差の大きさを忠実に再現することを意味し、推定器として有用である。

PoseNet では推定時にも Dropout を適用することで 1 枚の画像から複数の推定値を求め、それを用いて分散を算出し、推定の不確実性としている¹⁴⁾が、Fig. 17 より、訓練時走行経路から遠ざかるほどマハラノビス距離が大きくなっており、推定誤差と PoseNet による分散には大きな相関が見られなかった。一方、提案手法では訓練時走行経路からの距離に依らずマハラノビス距離がほぼ横ばいの状態になっている。これはネットワークが誤った推定をしているときは、分散が大

Fig. 19 The network output at large error (above figure) and at small error (lower figure)

Fig. 20 The difference of output maps in position estimation by the difference of giving label data. The top left figure shows the result in case of the length of one side of a square grid is 0.5 m and the variance of position label data is 1.0 m². The top right figure shows the result in case of 0.5 m and 0.5 m². The lower figure shows the result in case of 1.0 m and 1.0 m²

きくなる傾向があることを示している。Fig. 18 より、位置の教師信号における分散が異なる場合でも同様の傾向が見られるが、分散が大きいほど全体的にマハラノビス距離は小さくなる傾向にある。

Fig. 9 において、推定値が真値から大きく外れているエリアがあることがわかる。提案手法では推定位置がグリッドマップとして得られるので、この推定誤差の大きいエリアにおける

位置推定のグリッドマップと、推定誤差の小さい位置推定をしているときのグリッドマップとの比較を行なった (Fig. 19). 丸い点が推定値で、×の点が真値である。色が濃いほどそのグリッドにいる確率が高いことを表わしており、各グリッド内に書かれている数値はそのグリッドにロボットがいる確率を表わしている。推定誤差が小さいときに比べて、推定誤差が大きいときのグリッドマップのほうが確率値のばらつきが多く、多峰性がより際立って見える。

位置の教師信号を生成する際の正規分布の分散を変えた場合や、グリッド1辺のサイズを変えた場合におけるネットワークの出力を比較するために、それぞれ個別に学習を行ない、それぞれのネットワークに対して同じ画像を1枚入力したときのグリッドマップを Fig. 20 に示す。楕円の大きさが違うことから、位置の教師信号における分散を変えることで、出力されるグリッドマップから求まる誤差楕円の大きさが変わることがいえる。また、グリッド1辺のサイズが大きいと、誤差楕円も大きくなった。

6.4 方位推定における不確実性

6.3 節と同様に、研究室環境のデータセットにおける方位推定について、縦軸を推定誤差、横軸を訓練時走行経路からの距離としてデータをプロットしたものを Fig. 21 に示す。位置推定の場合とは違い、訓練時走行経路からの距離が方位の推定精度に影響を与えないことがわかるとともに、両手法の結果が類似していることも見て取れる。Fig. 22 では、訓練時走行経路からの距離と推定した方位のマップから求めた標準偏差との関係を示している。PoseNet では訓練時走行経路からの距離に関係なく、標準偏差が一定であるのに対し、提案手法では標準偏差がやや大きくなる傾向にあることが確認できた。また、縦軸をマハラノビス距離、横軸を訓練時走行経路からの距離としてデータをプロットしたものを Fig. 23 に示す。全体的に提案手法のほうがPoseNet よりもマハラノビス距離が小さいのは、提案手法のほうが推定した方位の分布における分散が大きいからである。方位推定に関しては位置推定のととは違い、訓練時走行経路からの距離に依らず、ネットワークから出力される分布は Fig. 24 のように単峰性を有していた。

6.5 推定に要する時間

GPU として NVIDIA 社製の GeForce GTX TITAN X を、CPU として Intel 社製の Core i7-6700K を用いた本実験環境における、提案手法を用いて1枚の画像から自己位置・方位のグリッドマップを得るために、CPU のみを用いた場合と GPU を用いた場合においてそれぞれ要した時間を Table 4 に示す。

GPU を用いた場合が CPU のみを用いた場合よりもやや計算にかかる時間が少ないことがわかる。しかし、取得した魚眼画像に対する射影変換の処理や、グリッドマップから自己位置および方位を点推定する処理が加わることを考えると、リアルタイムで推定をする上ではこの処理時間は長いといえる。

Fig. 21 Relationship between distance from the route of the robot when acquiring training data and orientation estimation error

Fig. 22 Relationship between distance from the route of the robot when acquiring training data and standard deviation of orientation estimation

Fig. 23 Relationship between distance from the route of the robot when acquiring training data and mahalanobis distance in orientation

Fig. 24 An example of orientation distribution from the proposed network

Table 4 Time taken for estimation

environment	CPU only(s)	GPU(s)
laboratory	0.52	0.46
outdoor	0.52	0.47

7. 考 察

位置推定においては PoseNet のほうが提案手法よりも推定精度が高く、方位推定においてはその逆の結果となったが、両手法とも位置および方位に関してある程度正しく推定できており、学習そのものが破綻していることはないと考えられる。したがって、深層学習において重要とされている重みの初期化方法や、入力値および出力値のスケーリング、ハイパーパラメータの調整、損失関数 (1) (2) における位置誤差と方位誤差の重み β の設定、そして提案手法においては 1 つのグリッドの大きさの調整などを行なうことで両手法とも推定精度の向上が見込める可能性がある。

本手法で提案した環境をグリッド状に区切ることで自己位置推定を行ない、各グリッドにおけるロボットの存在する確率値に基づいて共分散行列を求めるやり方は、6.3 節より、位置推定に関しては訓練時走行経路からの距離が大きいほど分散が大きくなる、すなわち推定の不確実性が大きくなることから確かめられた。対して PoseNet では、訓練時走行経路からの距離に応じてマハラノビス距離が大きくなることから、推定の不確実性が正しく求められておらず、ネットワークの出した推定値に対してどれだけ信用すればよいかを求めるのが困難になると考えられる。さらに、訓練時走行経路から遠い地点で撮影した画像に対する位置推定の分布は多峰性を有することがあり、PoseNet によって推定値を求める手法や提案手法で行なった全グリッドの確率値を用いた期待値を推定値とする手法では、この多峰性を考慮することができていないと考えられる。提案手法の場合は、たとえば全グリッドの中で最も高い確率値を出力するグリッドとその近傍のグリッドのみを参照して重み付き平均を算出するアプローチをとることでこの問題を解決することができる。方位推定に関しても訓練時走行経路からの距離が大きくなると推定した分布の分散も大きくなる傾向が見られたが位置推定ほど顕著ではなかった。この理由としては、訓練時走行経路に近い地点からの画像を入力したとしても、そのときのロボットの方位が訓練時とは違う方角を向いていれば、訓練画像からは遠いものとなるからだと考えられる。ただ、このようなロボットの向きが異なるために推定精度が下がることを防ぐために、すでに本実験ではデータ拡張を行っていた。これはロボットの位置座標を変えずにロボットの向きだけを変えた場合に得られる画像を擬似的に生成することで方位に対するデータのバリエーションを増やす手法であり、今回は 1 枚の画像に対して 45 度刻みで向きを変えて新たに 7 枚だけ擬似画像を生成している。この刻み幅が大きすぎた可能性があり、今後は刻み幅を変えて学習を行ない、評価する必要があると考える。

提案手法では、学習時に真の自己位置に対応するグリッドだけでなく、その周囲のグリッドにも正規分布に基づいてある程度の確率値を出力するように教師信号を生成した。研究室環境のデータセットにおいて、位置の教師信号生成時の正規分布の分散を変化させたり、グリッド 1 辺のサイズを変化

させたりしたところ、出力されるグリッドマップの推定精度や分散に違いが出ることがわかった。正規分布の分散が小さいと、特に訓練時走行経路から遠い地点における推定誤差が大きくなるため、正規分布に基づいて周囲のグリッドにもある程度の確率値を割り当てる手法は精度向上に有効であると考えられる。また、グリッドが粗いと推定における分散が大きくなり、推定精度が悪化することが示唆された。今後はグリッドが細かすぎる場合についても調査する必要がある。さらに、このような比較実験を方位推定に対しても行ない、方位推定における提案手法の有効性を示す必要がある。

ネットワークによる推定にかかる時間は両環境ともほとんど同じであった。これは環境の広さに応じてグリッドマップのサイズを変える際に、ネットワークの出力層のサイズだけを変更しているからであると考えられる。したがって、より広い環境に適用する場合においても計算時間の大幅な増加や、必要とするメモリの増加が問題になることは考えにくい。ただし、より広い環境に対しても本提案手法による自己位置推定が正しく行なえるかは別問題であり、検討すべき課題である。

8. 結 論

本研究では、深層学習を用いた全地球パノラマ画像からの大域的な自己位置推定に関する手法の提案を行なった。ネットワークの出力する位置や方位の推定結果をグリッドとして表現することで、提案手法を分類問題として定式化した。これにより、推定値の分布が得られるため、その分布から推定の不確実性を求めることができた。深層学習を用いた自己位置推定の一手法である PoseNet との比較実験を行なったところ、推定精度そのものは同程度であったが、位置推定の不確実性については提案手法がより正しく求められているといえる。

本研究では、学習においてパラメータの更新を行なう回数を 50 epoch と固定して行っており、その回数の決め方については特に考慮していない。今後は一定の回数ごとに訓練誤差とテスト誤差を算出し、パラメータが収束しつつ、過学習も発生していないようなモデルを抽出できるようにして評価を行なう必要がある。

提案手法では、グリッドのサイズと教師信号を生成する際に用いる正規分布の分散をユーザが設定しなければならない。これらはロボットに自己位置推定を行なわせたい環境の大きさやユーザが求める精度によって変化するものであり、これらのパラメータをどう設定すべきかといったガイドラインを示すことも重要な課題である。

推定値に対する不確実性を定義できると、カルマンフィルタなどのアルゴリズムを用いて移動ロボットのオドメトリと組み合わせた自己位置推定器を実装することができる。このように現在得られるセンサデータのみを用いる大域的な自己位置推定と、過去に推定した自己位置を利用して逐次的に推定値を更新し続けていく局所的な自己位置推定とを適切に組み合わせることで推定精度の向上が期待できる。

参 考 文 献

- 1) A. Krizhevsky, I. Sutskever and G.E. Hinton: ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, 1097/1105 (2012)
- 2) J. Long, E. Shelhamer and T. Darrell: Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431/3440 (2015)
- 3) B. Li, C. Shen, Y. Dai, A. van den Hengel and M. He: Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1119/1127 (2015)
- 4) 梅田, 伊達: 深層学習を用いた全天球パノラマ画像からの自己位置推定, 第22回ロボティクスシンポジウム予稿集, 97/98 (2017)
- 5) F. Dellaert, D. Fox, W. Burgard and S. Thrun: Monte Carlo Localization for Mobile Robots, *Proceedings of the IEEE ICRA*, **2**, 1322/1328 (1999)
- 6) 井上, 董, 田平, 鳥居, 奥富: 全球パノラマ画像を用いた SfM による 3 次元復元と自己位置・方位推定への応用, *精密工学会誌*, **81-12**, 1173/1179 (2015)
- 7) 山崎, 持丸, 金出: 一人称ビジョンシステムのための自己位置推定法, *電子情報通信学会信学技報*, **110-27**, 73/78 (2010)
- 8) A. Irschara et al.: From structure-from-motion point clouds to fast location recognition, *Computer Vision and Pattern Recognition*, 2599/2606 (2009)
- 9) Z. Zhu et al.: Real-time global localization with a pre-built visual landmark database, *Computer Vision and Pattern Recognition*, 1/8 (2008)
- 10) M. Cummins and P. Newman: FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance, *The International Journal of Robotics Research*, **27-6**, 647/665 (2008)
- 11) D.G. Lowe: Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, **60-2**, 91/110 (2004)
- 12) 石伏, 谷口, 高野, 谷口: Convolutional Neural Network による物体認識の自己位置推定への統計的活用, *人工知能学会全国大会論文集*, **29**, 1/4 (2015)
- 13) A. Kendall, M. Grimes and R. Cipolla: PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization, *Proceedings of the International Conference on Computer Vision*, 2938/2946 (2015)
- 14) A. Kendall and R. Cipolla: Modelling Uncertainty in Deep Learning for Camera Relocalization, *Proceedings of the International Conference on Robotics and Automation*, 4762/4769 (2016)
- 15) A. Babenko, A. Slesarev, A. Chigorin and V. Lempitsky: Neural codes for image retrieval, *European Conference on Computer Vision*, 584/599 (2014)
- 16) K.M. Yi et al.: Lift: Learned invariant feature transform, *European Conference on Computer Vision*, 467/483 (2016)
- 17) J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi and A. Fitzgibbon: Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2930/2937 (2013)
- 18) C. Szegedy et al.: Going deeper with convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1/9 (2015)

[著 者 紹 介]

梅 田 将 孝

2016 年筑波大学理工学群工学システム学類卒業。同年筑波大学大学院システム情報工学研究科知能機能システム専攻博士前期課程に進み現在に至る。移動ロボットの自己位置推定の研究に従事。

伊 達 央 (正会員)

2003 年東京工業大学大学院理工学研究科機械制御システム専攻博士課程修了。同年防衛大学校情報工学科助手, 2007 年同助教, 2013 年同講師, 2015 年筑波大学大学院システム情報系知能機能工学域准教授となり現在に至る。蛇ロボットおよび移動ロボットなどのロボットの制御の研究に従事。博士 (工学)。IEEE, 日本機械学会, 日本ロボット学会の会員。