

セグメンテーションおよびガウス過程回帰による深度画像の欠損値予測とその一般化 Prediction of Uncertain Values in Depth Images by Segmentation and Gaussian Process Regression Leading to Its Generalization.

梅田 裕輔[†] 花沢 明俊[†]
Yusuke Umeda Akitoshi Hanazawa

ガウス過程回帰を用いて深度画像の欠損部分の深度値予測を行う手法を提案する。セマンティックセグメンテーションを用いて RGB 画像のピクセルのカテゴリ尤度を抽出し、それに位置情報を加えた特徴量空間上でガウス過程回帰による深度推定を行った。その際、カテゴリの尤度における不確実性を考慮するために、共分散行列にペナルティ項を導入した。この提案手法を一般化することによって、説明変数の観測ノイズを考慮するガウス過程回帰を導くことができた。この一般化した手法は、全ての入力データを等価に扱わず、入力データに不確実性に基づいた優劣をつけて回帰を行うと思われる、汎用性が高い手法という結論を得た。

1. 背景

自動運転に必要なタスクに、オドメトリ推定という自己移動量を推定するタスクがある。その推定には深度情報（奥行き）が必要であり、欠損領域の深度をガウス過程回帰で求め、その推定値の信頼度に応じて自己移動量を求める目的関数を切り替えるという論文が 2021 年に発表された [1]。その論文では、画像の特徴点の深度をガウス過程回帰で推定するのだが、ガウス過程回帰へ入力する情報は、深度情報があるピクセルの座標情報(x,y)のみであった。

この手法では、位置的に近いピクセルの影響が大きく、遠いピクセルの影響は小さくなる。それよりも、物体種類のラベルが同じで位置的に近いピクセルに大きく影響される方がより良い推定ができるのではないかと考えた。つまり、図 1 の赤点のピクセルの深度を推定する際には、図 1 の左図の赤色の領域のピクセルに大きく影響されるよりも、図 1 の右図の青色の領域に大きく影響される方が良いと考えた。これは、位置的な近傍データには検査対象となるピクセルとは無関係なピクセルが含まれる可能性がある一方で、物体種類のラベルを考慮することによって、無関係なピクセルを参照する可能性が低くなるためである。

そこで、私はガウス過程回帰への入力データを高次元データに変換し、図 1 の青色の領域が近傍となるような空間で深度推定を行うことを考えた。

2. 目的

本論文では、車に搭載された LiDAR センサで深度情報を取得できなかった領域のセンサ値の推定をガウス過程回帰を用いて行う。論文 [1]では、LiDAR 画像のピクセル位置のみを入力として、深度推定を行っていた。本論文では、RGB 画像のピクセルの特徴量を抽出し、ピクセルの特徴量とピクセル位置を入力として、深度推定を行う手法を提案する。

ピクセルの特徴量抽出では、セマンティックセグメンテーションと呼ばれるピクセル単位で物体種類のカテゴリ分類を行うネットワークを用いて、19 次元データに変換する。変換された 19 次元データは、各次元が各カテゴリに所属

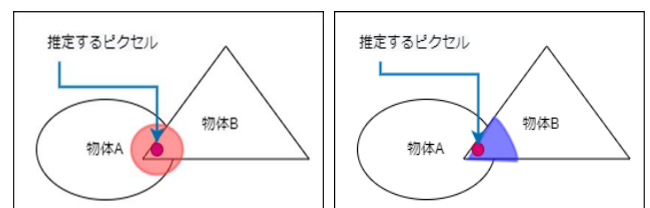


図 1. 推定する際に強く影響を受ける領域

赤色の領域は位置的に近いピクセルの領域を表し、青色の領域は物体種類のラベルが同じで位置的に近いピクセルの領域を表している。

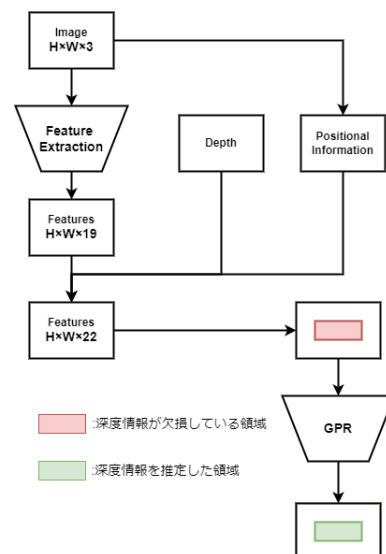


図 2. 提案手法の処理の流れ

する尤度に対応する。そして、そのデータにピクセルの位置情報を加えた 21 次元データを入力として、21 次元空間上で深度推定をするピクセルの特徴量の近傍データ 50 個のみを学習データとしてガウス過程回帰を用いて推定する。この流れを図 2 に示す。

また、提案手法を一般化することによって、説明変数に不確実性がある際のガウス過程回帰の手法を提案する。この一般化した手法は、全ての入力データを等価に扱わず、

[†]九州工業大学工学府 Kyushu Institute of Technology
Graduate School of Engineering

入力データの不確実性に基づいて優劣をつけて回帰を行うと思われ、汎用性が高い手法であると考えられる。

3. 提案手法

センサ値が取得できなかった領域の深度を推定するにあたって、例えば車が写っているピクセルの深度を推定する際には、推定するピクセルからできるだけ近い車が写っているピクセルの情報を使用した方がより良い精度で推定できるであろうと考えた。そこで、以下のようにガウス過程回帰への入力データを高次元データに変換し、深度推定を行う手法を考えた。

3.1 提案手法 1

1. セマンティックセグメンテーションの学習を行い、ネットワークを作成する。
2. 作成したネットワークの出力層を省いたネットワークを特徴量抽出器とする。
3. 特徴量抽出器を通して得たデータに、画素の位置情報と深度を加えて、ピクセル毎の特徴量を作成する。
4. 特徴量空間上で、深度推定を行うピクセルの特徴量と距離が近い特徴量を持つピクセル 50 個を学習データとして、ガウス過程回帰により深度を推定する。

本論文では 1 ピクセルあたりの深度を求めるため、ピクセル単位で特徴量抽出を行う必要がある。そこで、車載カメラから得られる道路の情報の特徴量抽出器を作成するための事前タスクとしてセマンティックセグメンテーションを設定した。セマンティックセグメンテーションは、ピクセル単位で分類問題を解いてくれるため、最終層付近ではピクセル単位の特徴量が得られる。この手順の流れを図 2 に示す。

3.2 提案手法 2

提案手法 1 は、データをニューラルネットワークで特徴量空間に射影し、ガウス過程回帰を特徴量空間上の近傍データのみを用いて行う。ただし、この手法は「特徴量抽出が正確に行われている」という仮定が前提になっている。これは全てのデータが正確に特徴量空間に射影されている場合は問題がない。しかし、提案手法 1 ではニューラルネットワークで特徴量空間に射影をしており、ニューラルネットワークはパラメータを最適化することによって得られる近似関数である。したがって、射影を正確に行うことはできない。

そこで、特徴量空間への射影の正確さをエントロピーを計算することによって定量化し、ガウス過程の共分散行列にペナルティ項を導入することによって提案手法 1 を改善する手法を提案する。エントロピーは乱雑さ、不規則さを表す指標で、この値が大きければ大きいほど、その確率分布が不規則であることを表す。セマンティックセグメンテーションはピクセル単位で分類問題を解くため、特徴量抽出で得られる特徴量には偏りがあればより正確に特徴量空間に射影されていると考えられる。したがって、エントロピーが小さいほど正確に射影できていると考えられる。

また本論文では、このガウス過程回帰を一般化することによって、説明変数の不確実性を考慮したガウス過程回帰を提案する。

3.3 提案手法の一般化

入力に不確実性がある際に、ガウス過程の共分散行列にペナルティ項を導入することを提案する。入力 x の不確実性を u_x とすると以下のように定式化できる。

$$f \sim N(k_*^T K^{-1} Y, k_{**} - k_*^T K^{-1} k_*) \quad (1)$$

$$k_* = (k(x_*, x_1) \dots k(x_*, x_N)) \quad (2)$$

$$k_{**} = k(x_*, x_*) \quad (3)$$

$$K = \begin{pmatrix} k(x_1, x_1) + w(u_{x_1} \times u_{x_1}) & \dots & k(z_1, z_N) + w(u_{x_1} \times u_{x_N}) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) + w(u_{x_N} \times u_{x_1}) & \dots & k(x_N, x_N) + w(u_{x_N} \times u_{x_N}) \end{pmatrix} \quad (4)$$

ただし、 w は重みパラメータを、 $k(x, x')$ はカーネル関数を表す。ここでは u の値が小さければ x は正確な入力、 u の値が大きければ x は不正確な入力と考えることができる。そして、本提案手法ではペナルティ項を導入することによって、全ての入力データを等価に扱わず、正確な入力は確実なデータとして扱い、不正確なデータは間違っているかもしれないデータとして扱うことができると思われる。

4. 実験

4.1 データセット

セマンティックセグメンテーションの学習データとしては、Cityspaces というドイツのダイムラー社、マックス・プランク研究所、ダルムシュタット工科大学のチームが公開しているデータセットを用いて学習を行った。ドイツの 50 都市の画像にセマンティックセグメンテーション情報が付与されている。クラスは 19 クラス含まれている。

深度推定を行うデータセットとして、KITTI というデータセットを用いた。KITTI はドイツのカールスルーエ工科大学とアメリカ・シカゴにある豊田工業大学シカゴ校のチームによる、自動車ビジョンのための画像データセットである。

4.2 実験方法

本論文では、深度画像の欠損値予測を行うので、深度情報があるピクセルをマスクして、そのピクセルの深度予測を行った。ピクセルのマスク範囲は、(450, 180), (450, 220), (600, 180), (600, 220) を頂点とする 40×150 の長方形の領域とした。また、深度推定は KITTI データセットの 300 枚の画像に対して行った。そして、ピクセルの座標情報のみを入力として深度推定した時 (従来手法) と提案手法 1 と提案手法 2 を比較して、平均相対誤差と予測分布の分散と相対誤差の相関を調べた。

4.3 結果

それぞれの手法の平均相対誤差と予測分布の分散と相対誤差の相関を調べた結果、表 1 のようになった。

表 1. 各手法の平均相対誤差と相関係数

	従来手法	提案手法 1	提案手法 2
平均相対誤差	0.1669	0.1542	0.1641
相関係数	0.3137	0.3467	0.3705

また、従来手法と提案手法 1 で深度推定を行ったときのヒートマップを図 3 に示す。そして、提案手法 1 と提案手法 2 の予測分布の分散のヒストグラムを図 4 に示す。

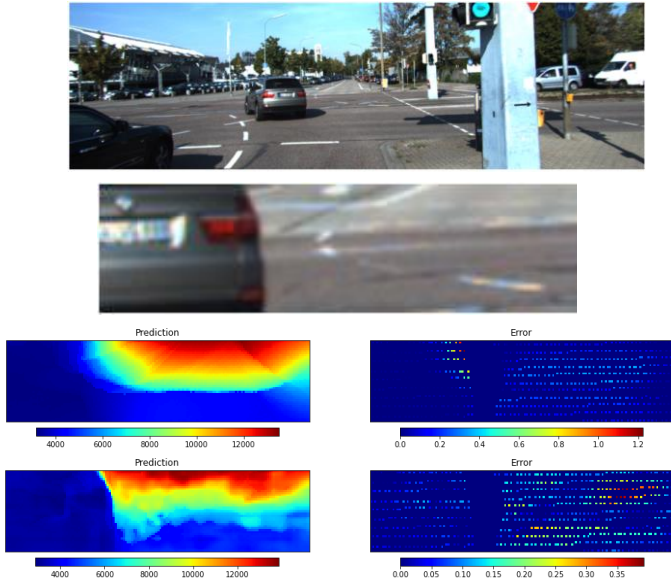


図 3. 従来手法と提案手法 1 の深度推定の様子

1 段目：深度を推定する画像。2 段目：深度推定を行う領域を切り抜いた画像。3 段目：位置情報のみを入力としたときの深度推定の結果。4 段目：画素の特徴量と位置情報を入力としたときの深度推定の結果。

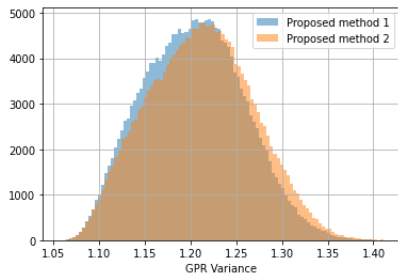


図 4. 提案手法 1 と提案手法 2 の予測分布の分散のヒストグラム

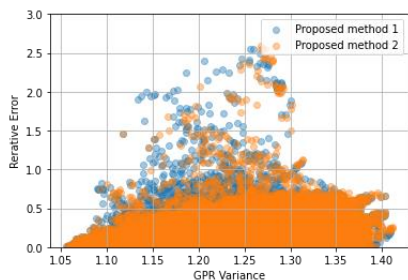


図 5. 提案手法 1 と提案手法 2 の予測分布の分散と相対誤差の散布図

4.4 考察

図 3 から分かるように、画素の位置情報のみを入力として深度推定を行ったときは、平坦で車と道路の境界が曖昧な予測を行っている。これは推定を行う際に用いる入力データが位置的な近傍データであるためである。つまり、車が写っているピクセルの深度を予測する際に道路が写っているピクセルの深度を参照するように、物体の境界付近を推定するには不要なデータを参照してしまうためである。一方、画素の特徴量と位置座標を入力として深度推定を行ったときは、予測に起伏がある。これは予測を行う際に、特徴量空間上の近傍データ、つまり自分と特徴量的に似ているデータを参照しているためである。また、従来手法では、車と道路の境界付近の予測の誤差が大きいたことが分かる。したがって、表 1 から分かるように画素の特徴量と位置座標を入力として深度推定を行った提案手法 1 の方が精度が良かった。

また、表 1 から分かるように、提案手法 2 は提案手法 1 と比較して、平均相対誤差は悪くなっているが、相対誤差と予測分布の分散との相関係数が強くなっていることが分かる。これは図 4 から分かるように、予測分布の分散が小さい予測をしたピクセル数が減少し、予測分布の分散が大きい予測をしたピクセル数が増加したためであると思われる。そして、図 5 より提案手法 2 は、予測分布の分散が小さい領域では、相対誤差が大きい予測をしているピクセル数が減少している。つまり、ペナルティ項を追加することによって、予測分布の分散が小さい予測のロバスト性が向上したと思われる。

5. ペナルティ項有りのガウス過程回帰の考察

5.1 トイデータでの実験

入力 $x=[1,4,10,16,19]$ 、不確実性 $u=[0.1, 0.1, 0.8, 0.1, 0.1]$ としたときに、 $y=0.5x$ の回帰をペナルティ項を導入したガウス過程と普通のガウス過程で行ったときの例と入力 $x=[1,4,16,19]$ としたときに普通のガウス過程で回帰した例を図 6 に示す。そして、表 2 に $x=10$ の時の標準偏差の値を示す。

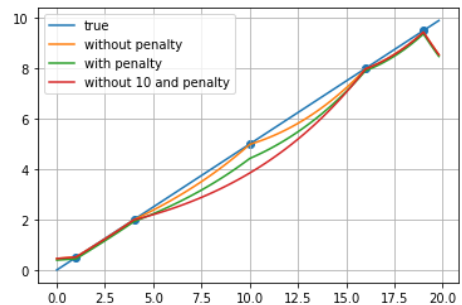


図 6. トイデータでの回帰の様子

表 2. $x = 10$ の標準偏差

	ペナルティ項なし	ペナルティ項あり
標準偏差	0.14096	0.2522

図 6 から分かるように、ペナルティ項有りのガウス過程で回帰した結果が、 $x=10$ があるときと無いときに普通のガウス過程で回帰した結果の中間にある。これは $x=10$ の不確実性が 0.8 とかなり大きく、他のデータに比べて学習データとしての有効度が低いためである。その結果 $x=10$ に多少はフィッティングしているが、他のデータに比べるとフィッティングしていない。また表 2 から、 $x=10$ の不確実性が大きいという情報が反映されていることが分かる。

5.2 理論の考察

ペナルティ項を導入することによって、入力データの優劣を考慮する手法は共分散行列にゆがみを作ることによって、上手くいっていると思われる。ここで、 x の標準偏差を S_x 、 y の標準偏差を S_y 、 x と y の共分散を S_{xy} とすると、相関係数 r と共分散行列 S は以下のように表される。

$$r = \frac{S_{xy}}{S_x S_y} \quad (6.5)$$

$$S = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{pmatrix} \quad (6.6)$$

次に x の不確実性を U_x 、 y の不確実性を U_y とするとペナルティ項を導入した共分散行列 S' とそのときの相関係数 r' は以下のように表される。

$$r' = \frac{S_{xy} + U_x U_y}{\sqrt{(S_x^2 + U_x^2)(S_y^2 + U_y^2)}} \quad (6.7)$$

$$S' = \begin{pmatrix} S_x^2 + U_x^2 & S_{xy} + U_x U_y \\ S_{xy} + U_x U_y & S_y^2 + U_y^2 \end{pmatrix} \quad (6.8)$$

この時、付録より $r' \leq r$ である。つまり、ペナルティ項を導入すると元の相関係数よりも必ず小さい相関係数となる。ガウス過程は入力が似ていれば出力も似るという性質がある。これは似た入力は相関係数が高くなるためである。つまり、ガウス過程の共分散行列から作られる相関行列の非対角要素はデータ間の関連度合いを表す値であると言える。そして、相関係数が小さくなるということは、データ間の関連度合いが小さくなることを意味する。したがって、ペナルティ項を導入したガウス過程では、不確実性が大きいデータは参照度合いが小さくなり、不確実性が小さいデータは参照度合いが大きくなるということが言えると思われる。

6. 結論

本論文では、深度推定をガウス過程回帰で行う際に、入力データを高次元化し、特徴量が似ているデータ 50 個のみを学習データとして扱うことで予測精度の向上を提案手法 1 で目指した。そこで提案手法 1 に加えて、ガウス過程の共分散行列にペナルティ項を導入する提案手法 2 を提案した結果、精度は少し劣化したが、予測分布の分散が小さい予測のロバスト性が向上した。

さらに、提案手法 2 を一般化することにより、全ての入力データを等価に扱わず、入力データに優劣をつけて回帰を行うガウス過程回帰を導くことができた。この手法は適用範囲が広く、良質なデータと捨てるほどではない程度の

質が悪いデータがある際にガウス過程回帰を用いて分析をしたいときに、使用できると思われる。

参考文献

- [1] 川. 出. 井. 村. 清水政行, “画像特徴点の対応を制約とした疎な LiDAR 点群からの Odometry 推定,” 精密工学会誌, 2021.
- [2] Y. B. R. N. ., S. S. S. Jaehoon Lee, “DEEP NEURAL NETWORKS AS GAUSSIAN PROCESSES,” ICLR 2018, 2018.
- [3] R. M. Neal, “Bayesian Learning for Neural Networks,” Number 118 in Lecture Notes in Statistics, 1996.
- [4] L. K. S. Youngmin Cho, “Kernel Methods for Deep Learning,” 2009.
- [5] Y. Z. G. Gal, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning,” 2016.
- [6] A. Z. Karen Simonyan, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,” ICLR 2015, 2015.
- [7] DeepAge, “Residual Network(ResNet)の理解とチューニングのベストプラクティス,” 30 11 2016. [オンライン]. Available: https://deepage.net/deep_learning/2016/11/30/resnet.html. [アクセス日: 28 1 2022].
- [8] G. P. F. S. H. A. Liang-Chieh Chen, “Rethinking Atrous Convolution for Semantic Image Segmentation,” 2017.
- [9] X. Z. S. R. a. J. S. Kaiming He, “Spatial Pyramid Pooling in Deep Convolutional”.

付録

$r' \leq r$ を証明する。 r' を両辺 2 乗すると、

$$\begin{aligned} r'^2 &= \frac{(S_{xy} + U_x U_y)^2}{(S_x^2 + U_x^2)(S_y^2 + U_y^2)} \\ &= \frac{S_{xy}^2 + 2S_{xy}U_x U_y + U_x^2 U_y^2}{S_x^2 S_y^2 + U_x^2 S_y^2 + U_y^2 S_x^2 + U_x^2 U_y^2} \end{aligned}$$

ここで、 $\Delta_1 = U_x^2 S_y^2 + U_y^2 S_x^2 + U_x^2 U_y^2$ 、 $\Delta_2 = 2S_{xy}U_x U_y + U_x^2 U_y^2$ とおくと

$$r'^2 = \frac{S_{xy}^2 + \Delta_2}{S_x^2 S_y^2 + \Delta_1}$$

である。そして、この時

$$\Delta_1 - \Delta_2 \geq 0 \Rightarrow r' \leq r$$

である。

$$\begin{aligned} \Delta_1 - \Delta_2 &= (U_x^2 S_y^2 + U_y^2 S_x^2 + U_x^2 U_y^2) - (2S_{xy}U_x U_y + U_x^2 U_y^2) \\ &= U_x^2 S_y^2 + U_y^2 S_x^2 - 2S_{xy}U_x U_y \\ &= S_x S_y \left(\frac{S_y}{S_x} U_x^2 + \frac{S_x}{S_y} U_y^2 - 2r U_x U_y \right) \quad \left(\because r = \frac{S_{xy}}{S_x S_y} \right) \\ &\geq S_x S_y (2U_x U_y - 2r U_x U_y) \quad (\because \text{相加相乗平均}) \\ &= 2S_x S_y U_x U_y (1 - r) \geq 0 \quad (\because r \leq 1) \end{aligned}$$

$$\therefore r' \leq r$$