

画像からの奥行連続性推定を利用した変分3次元点群超解像

八尾 泰洋^{1,a)} 安藤 慎吾^{1,b)} 倉田 夏菜^{1,c)} 新垣 仁^{1,d)} 嵯峨田 淳^{1,e)}

概要：疎な深度情報と画像を入力とし、密な深度情報を生成する技術は Depth Completion と呼ばれ、活発に研究されている。Light Detecting and Ranging(LiDAR) センサにより取得したデータを高解像度化する技術として実用化が期待される。Depth Completion については深層学習により高い精度が報告されているが、密な深度情報によって学習をする手法が主流である。学習用の密な深度情報の生成は容易ではなく、技術の実用化に向けての障壁となる。そこで本報告では、学習に密な深度情報を必要とせず、入手が容易なステレオ画像で学習が可能である手法を提案する。提案手法は、視差を推定するニューラルネットワークを用いて深度の連続性を推定し、深度の滑らかさと計測データからの乖離とのコストからなるエネルギー関数の最小化問題として Depth Completion を行う。提案手法について、公開データセットのベンチマークテストによって精度評価を実施し、提案手法が高密度深度で学習をしていないにもかかわらず従来の深層学習による深度超解像手法と同等の精度を確認した。

Variational Depth Completion Based on Depth Continuity Prediction

1. はじめに

Light Detecting and Ranging(LiDAR) センサにより計測した3次元点群情報は屋外設備の点検や構造劣化の診断への応用が期待されている。しかし、LiDARの計測データが疎であるため点検・診断に用いるにはデータを高密度化しなければならないことがある。そこで、画像と疎な深度情報から高密度な深度を生成する Depth Completion 技術の適用が注目されている。Depth Completion は、KITTI データセットのベンチマークタスクの1つでもあり、活発な研究が取り組まれている [1]。KITTI Depth Completion ベンチマークは LiDAR で計測した疎な深度情報と、深度情報とレジストレーションされた単眼画像を入力とし、高密度深度を出力するタスクであり、機械学習を行うための学習用のデータセットも提供されている。

KITTI ベンチマークにおいて精度の高い結果を示している手法は深層学習による。従来の深層学習による手法は、画像と疎な深度情報に対応する高密度深度情報が学習に必

要となる。しかし、画像とレジストレーションされた高密度深度を生成するのは容易ではなく、実応用のためにはより扱いやすい手法が求められる。

このような背景から、本報告では高密度深度を学習に用いない深度超解像手法を提案する。提案手法は、事前にステレオ画像により学習した視差推定をするニューラルネットワークを用いることで、単眼画像から奥行きの不連続性を推定し、その推定に基づき変分法により高密度深度情報を生成する。ニューラルネットワークは深度そのものを推定するのではなく、深度の変化が起こる位置を推定することを目的としている。そのため、必ずしも正確に視差を推定する必要はないため、4節の評価実験で示すように異なる環境のデータで事前学習したニューラルネットワークを適用することも可能である。提案手法の学習データたであるステレオ画像は容易に取得することができるため、提案手法は実応用の障壁が従来手法と比較して低いと考えている。

- 本報告のコントリビューションは以下のとおりである。
- ステレオ画像によって事前学習したニューラルネットワークによる視差推定により、密な深度情報による学習をしない、単眼画像と疎な深度情報からの深度超解像手法を提案する。
 - 視差推定の基づく深度連続性マスクを用いて、物体領

¹ NTT メディアインテリジェンス研究所
1-1 Hikarino-oka, Yokosuka, Kanagawa 239-0847, Japan
a) yasuihiro.yao.tc@hco.ntt.co.jp
b) shingo.andou.fv@hco.ntt.co.jp
c) kana.kurata.cb@hco.ntt.co.jp
d) hitoshi.niigaki.nt@hco.ntt.co.jp
e) atsushi.sagata.hw@hco.ntt.co.jp

域で深度が連続でありかつ LiDAR 計測値との誤差が小さくなるような深度に関するエネルギー関数を導出し、その凸最適化に基づく最小化を示す。

- 提案手法を KITTI Depth Completion ベンチマークに適用した結果により、精度評価および従来手法との比較を示す。

2. 関連研究

Depth Completion は疎な深度情報と画像を入力として、密な深度情報と画像を生成する技術である。様々な手法が提案されているが、そのアプローチにより大きくは Joint Bilateral Upsampling、大域的エネルギー最小化、深層学習に分類できる。

2.1 Joint Bilateral Upsampling

Joint Bilateral Filtering は画素の対応がついている高解像度の画像と低解像度の深度情報を入力とし、「画像平面での距離が近く、高解像度画像において近い画素値を持つ点同士は深度も近い」という経験則に基づき深度情報を超解像する手法である。Kopf らにより提案された Joint Bilateral Upsampling[2] は、画像平面での距離と高解像度画像における画素値の近さに基づき、式 1 のように低解像度の深度情報を足し合わせることで高解像度深度を生成する。なお、式 1 で \uparrow は高解像度データを、 \downarrow は低解像度データを意味する。

$$U\uparrow_p = U\downarrow_q f(\|p - q\|) f(\|I\uparrow_p - I\downarrow_q\|) \quad (1)$$

Kopf らの手法から発展して、高解像度データにおけるノイズへの耐性の向上 [3]、異なるフレームレートのセンサからの情報の統合 [4]、経路探索に基づいたサンプリングによるノイズ耐性とエッジ保存性の向上 [5]、頑健なエッジ検出に基づくフィルタリング [6] などの取り組みより、精度向上と適用可能シーンが拡大してきている。

Joint Bilateral Upsampling の利点は、学習が不要である、最適化によらないため高速であるという点である。欠点としては、低解像度の深度情報が均一に分布し、また正解値であるという前提を課している点である。上記の 2 点の前提は、Kinect などのカメラ一体型のデプスカメラの深度情報を対象とする場合には合理的である。しかし、LiDAR を対象とする場合、疎で均等でないデータが計測されること、カメラと LiDAR の位置が離れておりオクルージョンの影響が無視できないことによりその前提が崩れる。Joint Bilateral Filtering はそのような観点から LiDAR データの Depth Completion には適していない。

2.2 大域的エネルギー最小化

大域的エネルギー最小化は、高解像度の深度情報が従うべきエネルギー関数を定義し、その最小化によって解を導

出するアプローチである。提案手法はこのカテゴリに分類される。基本的には、エネルギー関数は深度の滑らかさを評価する項と計測データからの乖離を評価する項により構成され、画像の勾配が大きい点では深度が不連続でもエネルギーが増大しない定式化が提案されてきた。Diebel らはマルコフ確率場により [7]、Ferstl らは Total Generalized Variation により [8]、Barron らは超解像深度の画像平面での滑らかさではなく Bilateral 空間における滑らかさを評価することにより [9]、超解像深度が従うべきエネルギー関数を導き、その最適化手法を提案してきた。

従来の大域的エネルギー最小化による Depth Completion の課題は、入力が密度が均一な深度マップだとみなしてきた点である。これは 2.1 で述べたように、デプスカメラのような計測機器の場合は正しいが、LiDAR では正しくない前提である。LiDAR 計測を入力とする Depth Completion に大域的エネルギー最小化の適用が少ない要因の一つとしては、画像の勾配を利用する際に物体の領域を過分割することを前提にしていることが考えられる。低解像度ながら密な深度情報があれば、物体領域を過分割してもそれぞれの領域内に深度情報を確実に割り当てられるため、大域的エネルギー最小化の枠組みを当てはめることが容易である。しかし、LiDAR 計測データのような疎な深度情報を対象とした場合、分割された領域内に深度情報が存在しないような状況が発生する。

Kuschik らは大域的エネルギー最小化による深度超解像を LiDAR 点群に適用しているが、彼らはステレオ画像を入力として使うことで物体の境界の検出を視差によって推定した [10]。ステレオ視差により過分割を行わず物体境界のみを正確に捉えることができるため、Kuschik らの手法は LiDAR 計測データに適用可能である。しかし、本検討の問題設定では単眼画像が入力となるため、Kuschik らの手法は適用することができない。

そこで、提案手法では物体の境界をステレオ視差を単眼画像から推定する深層学習の手法に着目し、大域的エネルギー最適化を疎な LiDAR 点群の超解像に適用可能とした。

2.3 深層学習

深層学習は画像と疎な深度情報を入力として密な深度情報をディープニューラルネットワークで学習し、深度情報の推定器を構築するアプローチである。Joint Bilateral Upsampling や大域的エネルギー最小化手法と異なり、疎な LiDAR による計測データを入力とする手法が多く提案されている。また、KITTI Depth Completion ベンチマークにおいて、最も高精度な結果が報告されているのはこのカテゴリに入る手法である。

Uhrig ら、Huang らは入力データの疎なデータを畳み込み可能なネットワーク構造を [1]、[11]、Jaritz らはエンコーダデコーダ構造による深度超解像とセマンティックセグメ

ンテーションの同時学習を [12]、Riegler らは大域的エネルギー最小化による深層学習を [13]、Chodosh らは深層学習による辞書作成を [14] を提案している。

深層学習手法の課題は入力となる画像と深度情報に対して正解となる密な深度情報が学習のために必要となる点、学習した環境と同一な撮影環境・機材であるシーンにしか適用できない点である。正確で密な深度情報を生成することは容易ではなく、KITTI ベンチマークでは GPS、ステレオ画像のセンサ情報に加え、人による確認によりデータセットを作成している [1]。

Ma らは自己学習、すなわち密な深度情報を用いずに撮影した画像と LiDAR データから深度超解像を学習できる手法を提案した [15]。しかし、Ma らの手法も学習したシーンと同等な撮影環境・機材の環境でしか適用できない。

提案手法は深層学習を用いているがステレオ視差推定にのみ用いて深度情報の値の導出には直接利用しないため、4 節で示すように事前に他のデータセットで学習したニューラルネットワークを利用することも可能である。

3. 提案手法

図 1 に提案手法への入力、および提案手法が生成する中間出力、最終出力を示す。提案手法は以下のステップからなる。

- 事前にステレオ画像データセットにより学習したニューラルネットワークにより単眼画像 I から推定視差 D を導く。推定視差 D と画像 I を用いて、深度の連続性を示すマスク M を生成する。
- 生成したマスク M と LiDAR の計測値 V から深度の滑らかさに対するコストとデータからの乖離に対するコストから構成されるエネルギー関数を導く。
- エネルギー関数を最小化することにより、高解像度の深度情報 \hat{U} を得る。

以下、上記のそれぞれのステップについて説明をする。

3.1 視差推定・マスク生成

提案手法では Godard らの提案したステレオ画像から視差を教師なし学習するニューラルネットワーク [16] により、単眼画像から視差推定を行う。我々は教師データを作成する必要なく、学習データセットの入手が容易であるという観点で Godard らの手法に着目した。なお、提案手法ではニューラルネットワークの目的は奥行きが不連続となる可能性の高い位置を抽出することであり、その推定値自体は問題とならない。そのため、事前に他のデータセットで学習することが可能である。例えば、4 節で示す評価実験ではニューラルネットワークの学習には Cityscapes データセット [17] を用い、評価には KITTI ベンチマークデータセット [1] を用いた。

グレースケールの幅 W 、高さ H の入力画像を I とす

ると、Godard らの手法により I と等しいサイズの視差推定 D が得られる。視差と深度は反比例の関係にあるので、 $D' = 1/D$ として、 D' と I より以下に示すようにマスク M を生成する。

$$M = \max \left\{ e^{-\left(\alpha_I \|\nabla I\|_2^{\beta_I} + \alpha_D \|\nabla D'\|_2^{\beta_D}\right)} + \gamma, 1.0 \right\} \quad (2)$$

式 2 において、 $\alpha_I, \alpha_D, \beta_I, \beta_D$ は I と D' の微分値のマスク M における影響度を定めるパラメータであり、正の値である。 γ はマスクの最低値であり $0 \leq \gamma < 1$ の値を定める。 M は深度が連続であると推定される点で 1 に近い値を取り、深度が不連続であると推定させる点で γ に近い値を取る、 I, D' とサイズの等しい配列である。

3.2 エネルギー関数の導出

提案手法はエネルギー関数の最小化により超解像深度を生成する。エネルギー関数は、深度の連続性を評価する項と、LiDAR の測定値からの乖離を評価する項からなり、式 3 で表される。

$$E(U) = \sum_{ij} M_{ij} \|\nabla U_{ij}\|_{\epsilon} + \lambda \sum_{V_{ij} > 0} |U_{ij} - V_{ij}| \quad (3)$$

式 3 において、 U は導出する超解像深度、 V は入力される疎な深度でありどちらも入力画像 I とサイズが等しい配列である。 λ は正の定数であり、 λ が大きいほどエネルギー関数に占める測定値との乖離を評価する項の影響が増大する。また、 $\|\cdot\|_{\epsilon}$ は Huber Norm であり、式 4 によって定義される。

$$\|x\|_{\epsilon} = \begin{cases} \frac{\|x\|_2^2}{2\epsilon} & \text{if } \|x\|_2 < \epsilon \\ \|x\|_1 - \frac{\epsilon}{2} & \text{otherwise} \end{cases} \quad (4)$$

深度連続性のコストを Huber Norm によって定めることにより、 U について深度の不連続な変化を許容することができる。超解像深度は式 3 を最小化する $\hat{U} = \arg \min_U E(U)$ として導かれる。

3.3 エネルギー関数の最小化

式 3 の最小化は凸最適化問題であり、Chambolle らによって提案された First-Order Primal-Dual Algorithm [18] によって解く。まず、式 3 のエネルギー関数の最小化問題を Legendre-Fenchel 変換 [19] により式 5 のように変換する。

$$\begin{aligned} \min_{\mathbf{u}} E(\mathbf{u}) &= \min_{\mathbf{u}} \sup_{\mathbf{q}_k, \|\mathbf{q}_k\|_{\infty} \leq 1} \sum_k m_k \left(\langle \mathbf{q}_k, L_k \mathbf{u} \rangle - \frac{\epsilon}{2} \|\mathbf{q}_k\|_2^2 \right) \\ &\quad + \lambda \sum_{\mathbf{v}_k > 0} |\mathbf{u}_k - \mathbf{v}_k| \end{aligned} \quad (5)$$

式 5 では式 4 の U, V, M をベクトル形式の $\mathbf{u}, \mathbf{v}, \mathbf{m} \in \mathcal{R}^{WH}$

で、また微分演算子 ∇ を各 $k(0 \leq k < WH)$ ごとに $L_k \in \mathcal{R}^{2 \times WH}$ と行列形式で表記している。 $\mathbf{q}_k \in \mathcal{R}^2(0 \leq k < WH)$ は \mathbf{d} の凸共役の要素である。式 5 において $\langle \cdot, \cdot \rangle$ はベクトルの内積であり、ここでは $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ である。

本変換の狙いは U についての最適化と ∇U についての最適化を分離可能とすることである。具体的には、式 6 のような式変換を考えることで、微分演算子 L_k が \mathbf{u} から分離できることが示される。

$$\langle \mathbf{q}_k, L_k \mathbf{u} \rangle = \mathbf{q}_k^T L_k \mathbf{u} = (L_k^T \mathbf{q}_k)^T \mathbf{u} = \langle L_k^T \mathbf{q}_k, \mathbf{u} \rangle \quad (6)$$

この微分演算子を分離可能な性質により、 \mathbf{u} と \mathbf{q} を L_k を交互に勾配法により更新することで、各イテレーションで微分値を切り離してエネルギー関数の最適化が可能となる。すなわち、 \mathbf{u} の更新では式 6 の右辺の形にし、 \mathbf{q} を固定して勾配法に更新し、 \mathbf{q} の更新では式 6 の左辺の形にし、 \mathbf{u} を固定して勾配法により更新すれば良い。

ここではさらに、式 5 のエネルギー関数について、収束性を向上するために、Steinbrucker ら [20]、Newcombe ら [21] と同様、変数 \mathbf{a} を導入し、Quadratic Relaxation により式 7 に変換する。

$$\begin{aligned} \min_{\mathbf{u}} E(\mathbf{u}) &= \min_{\mathbf{u}, \mathbf{a}} \sup_{\mathbf{q}, \|\mathbf{q}_k\|_\infty \leq 1} \sum_k m_k \left(\langle \mathbf{q}_k, L_k \mathbf{u} \rangle - \frac{\epsilon}{2} \|\mathbf{q}_k\|_2^2 \right) \\ &\quad + \lim_{\theta \rightarrow +0} \frac{1}{2\theta} \|\mathbf{u} - \mathbf{a}\|_2^2 \\ &\quad + \lambda \sum_{v_k > 0} |\mathbf{a}_k - v_k| \end{aligned} \quad (7)$$

式 7 は First-Order Primal Dual Algorithm により、式 8~10 に示す更新式によって、及び \mathbf{a} を式 11、 θ を式 12 に示す更新式によって最小化する。本更新式の意味は勾配法により交互に \mathbf{u} と \mathbf{q} を更新することである。

$$\mathbf{q}_k^{n+1} = \Pi \left(\frac{\mathbf{q}_k^n + \sigma_q L_k \bar{\mathbf{u}}^n}{1 + \sigma_q \epsilon} \right) \quad (8)$$

$$\mathbf{u}^{n+1} = \frac{\mathbf{u}^n - \sigma_u \left(\sum_k m_k L_k^T \mathbf{q}_k^{n+1} - \frac{1}{\theta^n} \mathbf{a}^n \right)}{1 + \frac{\sigma_u}{\theta^n}} \quad (9)$$

$$\bar{\mathbf{u}}^{n+1} = 2\mathbf{u}^{n+1} - \mathbf{u}^n \quad (10)$$

$$\mathbf{a}_k^{n+1} = \begin{cases} u_k^{n+1} & \text{if } v_k = 0. \\ v_k & \text{if } |u_k^n - v_k| \leq \theta \lambda. \\ u_k^{n+1} + \text{sign}(v_k - u_k^n) \theta \lambda & \text{otherwise.} \end{cases} \quad (11)$$

$$\theta^{n+1} = (1 - \beta \theta) \theta \quad (12)$$

ここで、 σ_d 、 σ_u は勾配法での更新における係数であり、十分に小さい値とする。また、 β は $0 < \beta < 1$ の定数である。終了条件は様々な方法によって定めることができるが、本検討では予め定められた δ に対し、 $\theta^n < \delta$ をとることとした。

表 1: KITTI Depth Completion ベンチマーク結果
Table 1 Result of KITTI Depth Completion Benchmark

Method	Input	RMSE[mm]	MAE[mm]
below on Test Set			
Ma et al. [15]	rgb	814.73	249.95
Jaritz et al. [12]	rgb	917.64	234.81
Huang et al. [11]	d	937.48	258.48
Ma et al. (w/o gt) [15]	rgb	1299.85	350.32
Chodosh et al. [14]	d	1325.37	439.48
Urig et al. [1]	d	1601.33	481.27
below on Validation Set			
Ma et al. [15]	rgb	878.56	260.90
Ma et al. (w/o gt) [15]	rgb	1384.85	358.92
Ours	rgb	1649.70	398.01

4. 評価実験

提案手法を KITTI Depth Completion ベンチマーク [1] に適用した。本ベンチマークは Training Set、Validation Set、Test Set からなる。Training Set と Validation Set については正解の高密度深度が公開されているが、Test Set については正解は公開されておらず、Test Set での精度検証はベンチマークへの公式エントリーとともに Test Set についての精度検証が可能である。本報告では提案手法については Validation Set での精度確認を実施し、結果が公開されている従来手法の Test Set での結果を参考値として比較した。なお、Validation Set 及び Test Set の数は 1000 組である。

なお、提案手法の各パラメータについては $\alpha_I = 0.01$ 、 $\beta_I = 1.0$ 、 $\alpha_D = 0.1$ 、 $\beta_D = 1.0$ 、 $\gamma = 0.1$ 、 $\epsilon = 1.0$ 、 $\lambda = 1.0e2$ 、 $\sigma_u = 0.3$ 、 $\sigma_q = 0.4$ 、 $\theta_0 = 1.0$ 、 $\beta = 1.0e - 31$ とした。また、最適化の終了条件は $\theta_n < 1.0e - 6$ とした。視差推定に用いるニューラルネットワークは事前に Cityscapes[17] の 22973 組のステレオ画像データにより学習した。提案手法では Validation Set のみを検証に利用し、Training Set については学習にも検証にも用いてない。

表 1 に、KITTI Depth Completion ベンチマークでの提案手法、従来手法の結果を示す。評価は Rooted Mean Squared Error (RMSE)、Mean Absolute Error (MAE) により実施している。表 1 の Input という項目はそれぞれの手法に対する入力の種類であり、rgb は画像と疎な深度情報を、d は疎な深度情報のみを入力を示す。表 1 に示す提案手法以外の手法はすべて Training Set により学習をした手法である。提案手法は他の手法と比較して RMSE は劣るが、MAE については Training Set により学習をした手法に匹敵をする精度を示した。RMSE よりも MAE の精度が優れているのは、提案手法が式 3 のエネルギー関数で示されているように LiDAR との計測誤差については絶対値をコストとして評価するためであると考えら

れる。

図1に深度超解像の結果の例を示す。提案手法が自転車の車輪などの複雑な構造の深度を推定できていることがわかる。また、図1に示す推定誤差の図からわかるように、提案手法はセンサから近い物体については比較的精度よく深度推定ができているが、センサから遠い物体については推定誤差が大きくなっていることがわかる(推定誤差の画像中で色が赤い領域が目立つ)。これについて、以下に示すいくつかの要因が考えられる。

- センサから遠い被写体は画像中では小さくなり、複数の物体が小さい領域で撮影されるため、深度がなめらかでなくなる。それに対し提案手法では深度をなめらかにしようとはたらくため、結果として誤差を大きくする。
- 視差と深度は反比例の関係にあるため、センサから遠い被写体については視差の変化が小さい場合でも深度が大きく変化する。Godardらの手法[16]はステレオ視差を学習しているが、視差の微小変化については学習におけるコストに及ぼす影響が小さいため、結果としてセンサから遠い対象について深度を正確に推定できない。

5. 結論

本報告ではDepth Completion技術の実応用を視野に入れ、密な深度情報を学習に必要としない深度超解像手法を提案した。提案手法はKITTI Depth Completionベンチマークにおいて従来の深層学習による手法にMAEの観点で匹敵する精度を示した。今後は、一般的な精度向上を目指すのではなく、屋外設備点検など実利用における利用環境に照準を合わせて技術検討・課題解決を進めていくことで、技術の実用化を目指して研究に取り組んでいく。

参考文献

- [1] Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T. and Geiger, A.: Sparsity Invariant CNNs, *2017 International Conference on 3D Vision (3DV)*, (online), DOI: 10.1109/3dv.2017.00012 (2017).
- [2] Kopf, J., Cohen, M. F., Lischinski, D. and Uyttendaele, M.: Joint Bilateral Upsampling, *ACM Trans. Graph.*, Vol. 26, No. 3 (online), DOI: 10.1145/1276377.1276497 (2007).
- [3] Thrun, D. C. H. B. C. T. S.: A Noise-Aware Filter for Real-Time Depth Upsampling (2008).
- [4] Dolson, J., Baek, J., Plogemann, C. and Thrun, S.: Upsampling range data in dynamic environments, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1141–1148 (online), DOI: 10.1109/CVPR.2010.5540086 (2010).
- [5] Liu, M.-Y., Tuzel, O. and Taguchi, Y.: Joint Geodesic Upsampling of Depth Images, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2013).
- [6] Xie, J., Feris, R. S. and Sun, M.: Edge-Guided Single Depth Image Super Resolution, *IEEE Transactions on Image Processing*, Vol. 25, No. 1, pp. 428–438 (online), DOI: 10.1109/TIP.2015.2501749 (2016).
- [7] Diebel, J. and Thrun, S.: An application of markov random fields to range sensing, *Advances in neural information processing systems*, pp. 291–298 (2006).
- [8] Ferstl, D., Reinbacher, C., Ranftl, R., Ruether, M. and Bischof, H.: Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation, *The IEEE International Conference on Computer Vision (ICCV)* (2013).
- [9] Barron, J. T. and Poole, B.: The Fast Bilateral Solver, *Lecture Notes in Computer Science*, pp. 617–632 (2016).
- [10] Kuschik, G., Boi, A. and Cremers, D.: Real-time variational stereo reconstruction with applications to large-scale dense SLAM, *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1348–1355 (online), DOI: 10.1109/IVS.2017.7995899 (2017).
- [11] Huang, Z., Fan, J., Yi, S., Wang, X. and Li, H.: HMS-Net: Hierarchical Multi-scale Sparsity-invariant Network for Sparse Depth Completion (2018).
- [12] Jaritz, M., Charette, R. D., Wirbel, E., Perrotton, X. and Nashashibi, F.: Sparse and Dense Data with CNNs: Depth Completion and Semantic Segmentation, *2018 International Conference on 3D Vision (3DV)*, pp. 52–60 (online), DOI: 10.1109/3DV.2018.00017 (2018).
- [13] Riegler, G., Ferstl, D., Rother, M. and Bischof, H.: A Deep Primal-Dual Network for Guided Depth Super-Resolution (2016).
- [14] Chodosh, N., Wang, C. and Lucey, S.: Deep Convolutional Compressed Sensing for LiDAR Depth Completion (2018).
- [15] Ma, F., Cavalheiro, G. V. and Karaman, S.: Self-supervised Sparse-to-Dense: Self-supervised Depth Completion from LiDAR and Monocular Camera (2018).
- [16] Godard, C., Mac Aodha, O. and Brostow, G. J.: Unsupervised Monocular Depth Estimation with Left-Right Consistency, *CVPR* (2017).
- [17] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B.: The Cityscapes Dataset for Semantic Urban Scene Understanding, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).
- [18] Chambolle, A. and Pock, T.: A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging, *Journal of Mathematical Imaging and Vision*, Vol. 40, No. 1, pp. 120–145 (online), DOI: 10.1007/s10851-010-0251-1 (2011).
- [19] Boyd, S. and Vandenberghe, L.: *Convex Optimization*, Cambridge University Press, New York, NY, USA (2004).
- [20] Steinbrcker, F., Pock, T. and Cremers, D.: Large displacement optical flow computation without warping, *2009 IEEE 12th International Conference on Computer Vision*, pp. 1609–1614 (online), DOI: 10.1109/ICCV.2009.5459364 (2009).
- [21] Newcombe, R. A., Lovegrove, S. J. and Davison, A. J.: DTAM: Dense tracking and mapping in real-time, *2011 International Conference on Computer Vision*, pp. 2320–2327 (online), DOI: 10.1109/ICCV.2011.6126513 (2011).

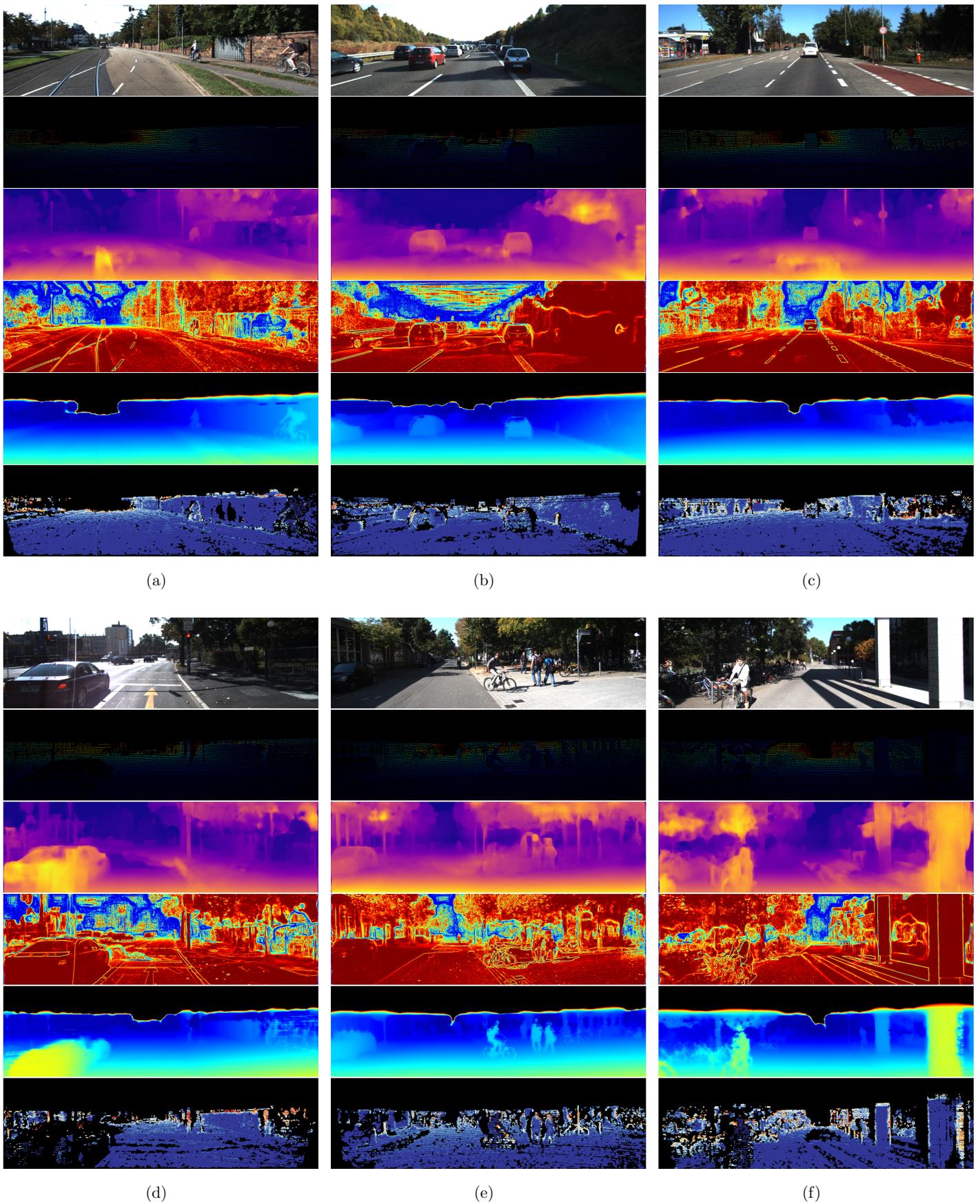


図 1: 上から入力画像 I 、入力深度 V 、視差推定 D 、マスク M 、超解像深度 \hat{U} 、推定誤差
Fig. 1 From the top, image I , depth V , disparity prediction D , mask M , result \hat{U} , error