Vol.2020-CG-180 No.29
Vol.2020-DCC-26 No.29
Vol.2020-CVIM-223 No.29
2020/11/6

未知散乱条件下での深層学習による Multi-view Stereo

藤村 友貴[1,a)]　薗頭 元春[2]　飯山 将晃[2]

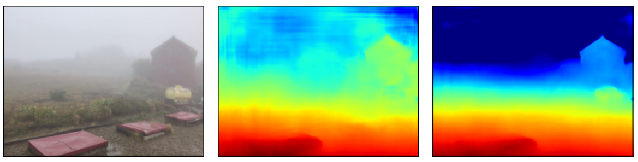
概要：本研究では霧や煙が充満した環境（散乱媒体）下での Multi-view Stereo (MVS) を提案する．従来提案されていた散乱媒体下における MVS では，散乱度合いを計算するための大気散乱光や散乱係数などの散乱パラメータを既知としていた．複数枚の画像とそれらから Structure-from-Motion (SfM) を用いて計算された三次元点群から散乱係数を計算することは，原理的には可能であるものの画像の輝度値を用いるため数値的に不安定である場合がある．本研究では散乱媒体下における深層学習ベースの MVS において，出力の深度画像が散乱パラメータの関数とみなせることを利用し，SfM で得られた三次元点群から散乱パラメータを最適化する．これにより，未知散乱条件下において散乱パラメータと深度画像の同時推定が可能になる．合成データによる実験で散乱パラメータと深度の推定精度を評価する．また，実際の霧がかったシーンに対しても適用した結果を示す．

キーワード：散乱媒体，Multi-view Stereo，深層学習, 散乱係数, 大気散乱光

1. はじめに

カメラで観測した画像からシーンの三次元情報を取得する三次元復元はコンピュータビジョンにおける重要なタスクの一つである．しかしながら，霧や煙が充満した環境（散乱媒体）下では，空間中に拡散した微粒子によって光の散乱と吸収が引き起こされ，観測した画像にコントラストが低下するなどの劣化が生じる（図 1(a)）．したがって，カメラで取得した画像の輝度値を直接利用する通常の三次元復元手法の多くは散乱媒体下では精度が低下してしまう．

このような散乱媒体下において，複数枚の画像から被写体の三次元形状を復元する Multi-view Stereo (MVS) を適用する研究が行われている [2], [5]．Li ら [5] は，散乱により劣化した画像の復元と MVS を同時に定式化した上で，各タスクを繰り返す事で画像復元と三次元復元を同時に行う手法を提案した．Fujimura ら [2] は深層学習ベースの MVS を用い，多視点間の幾何拘束に散乱媒体における画像の劣化モデルを組み合わせた dehazing cost volume と呼ばれるものをネットワークの入力に用いることで，散乱媒体下での深度推定の精度を向上させた．しかしながら，これらの手法はいずれも大気散乱光や散乱係数などの散乱パラメータを既知としているため，実際に適用する際はあらかじめそれらを推定しておく必要がある．Li ら [5] は複数枚の画像が与えられたときに散乱係数を推定する手法を提案しているが，画像の輝度値を用いるため数値的に不安定であるという問題がある．



(a)　(b)　(c)

図 1　(a) 霧がかったシーン．(b) MVDepthNet [11] の出力した深度画像．(c) 提案手法の出力した深度画像．

本研究では未知散乱条件下において，深層学習ベースの MVS の枠組みで散乱パラメータと深度を同時推定する手法を提案する．深層学習ベースの MVS で用いられている dehazing cost volume を計算するためには散乱パラメータが必要であるが，これはすなわち出力の深度は散乱パラメータの関数になっているとみなすことができる．したがって，あらかじめ Structure-from-Motion (SfM) で推定された散乱の影響を受けにくい疎な三次元点群と出力の深度が一致するように，散乱パラメータについて最適化を行う．画像の輝度値を用いることなく幾何的な情報を用いて最適化を行うため，より安定に散乱係数を推定することができ，その散乱係数を用いたときの出力の深度も幾何的に正しいものが出力されると期待できる．

本研究では散乱を合成した画像で学習された学習済みモ

[1] 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
[2] 京都大学学術情報メディアセンター
ACCMS, Kyoto University
[a)] fujimura@mm.media.kyoto-u.ac.jp

© 2020 Information Processing Society of Japan

Vol.2020-CG-180 No.29
Vol.2020-DCC-26 No.29
Vol.2020-CVIM-223 No.29
2020/11/6

デル [2] を用い，新たに作成した散乱合成画像で散乱係数と出力された深度について評価を行った．また，実際の霧がかったシーンについても適用を行い，その有効性を検証した．

2. 未知散乱条件下での Multi-view Stereo

本章では最初に，散乱媒体下での画像の劣化モデルについて説明したのち，散乱媒体下での深層学習を用いた MVS [2] について概説する．その後，画像の輝度値を用いた従来の散乱係数推定手法と，幾何的な情報を用いて最適化を行う提案手法について説明を行う．

2.1 大気散乱モデル

散乱媒体下での画像復元手法の多くでは，画像の劣化モデルとして大気散乱モデル [9] を用いる．大気散乱モデルは主に日中のシーンにおける散乱による画像の劣化をモデル化したものである．いま，散乱媒体下で観測した劣化画像のピクセル (u, v) での RGB の輝度値を $I(u, v) \in \mathbb{R}^3$，劣化する前の画像の輝度値を $J(u, v) \in \mathbb{R}^3$ とする．大気散乱モデルでは劣化した画像と劣化する前の画像の関係は以下の式で与えられる．

$$I(u, v) = J(u, v)e^{-\beta z(u,v)} + \mathbf{A}(1 - e^{-\beta z(u,v)}) \quad (1)$$

ここで，$z(u, v) \in \mathbb{R}$ はピクセル (u, v) でのシーンの深度，$\beta \in \mathbb{R}$ は散乱媒体の濃度を表す散乱係数，$\mathbf{A} \in \mathbb{R}^3$ は大気散乱光である．本研究では簡単のため，$\mathbf{A}$ は $\mathbf{A} = [A, A, A]^\top, A \in \mathbb{R}$，すなわち散乱媒体は灰色や白など無彩色であると仮定する．一項目はシーンで反射した光の成分であり，深度に応じて指数関数的に減衰する．二項目は観測した散乱光の成分であり，反射成分とは逆に深度に対して増大する．したがって，散乱による画像の劣化はシーンの深度に依存している．

画像復元の文脈においては，観測した画像 I から未知数 J，z，A，β，を推定するが，それらをすべて同時に推定することは一般に不良設定問題である．これに対し，A の推定については，例えば [3] や [1] などの従来手法が存在し，β についてはカメラが複数あるという問題設定では推定することが可能である [5]．したがって従来の散乱媒体下での MVS では A と β を既知としているが，これらの推定手法が必ずしも正確であるとは限らないため，その後の三次元復元の精度に影響を及ぼす．一方で，本研究では A と β についても，後段の深度推定の枠組みで同時に最適化を行う．

2.2 散乱媒体下での深層学習 Multi-view Stereo

本章では散乱媒体下での深層学習ベースの Mutli-view Stereo [2] について説明を行う．手法の概要を図 2 に示す．ネットワークの構造は MVDepthNet [11] と同じであり，入力は複数枚の画像（1 枚を reference 画像，その他を source 画像），出力は reference 画像の視差画像（深度の逆数）である．MVDepthNet との違いは，ネットワークの入力となる多視点間の幾何拘束に，通常の cost volume ではなく dehazing cost volume が用いられている点であり，これにより散乱による画像の劣化をモデル化することができる．

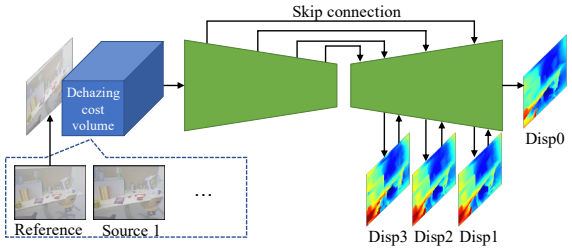


図 2 散乱媒体下での深層学習 Multi-view Stereo

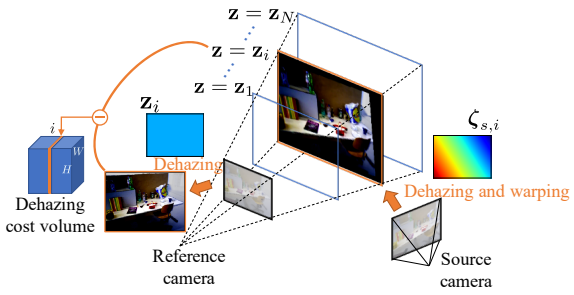


図 3 Dehazing cost volume

Dehazing cost volume の概要を図 3 に示す．通常の cost volume では，最初に reference カメラ座標系において空間を平面で走査する．その後，各平面上に source 画像を投影し，reference 画像との差分を取る．これはすなわち，シーンがその平面の深度にあるときの photometric consistency を計算したものに対応する．一方，dehazing cost volume では，図 3 に示したように，平面への投影と同時にその平面を用いて画像復元を行い，photometric consistency の計算と画像の劣化を同時にモデル化する．ここで画像復元には式 (1) を用いるが，画像復元を行うには z，A，β が必要である．z については画像を投影した平面の深度が用いられるが，A と β についてはあらかじめ推定しておく必要がある．

2.3 散乱パラメータ推定

2.2 で dehazing cost volume には A と β をあらかじめ推定する必要があることを述べた．本節ではこれらについて，特に β の推定の難しさについて述べたあと，深度推定の枠組みでこれらを同時に推定する手法について説明する．

2.3.1 大気散乱光 A の推定

最初に A の推定について述べる．一枚の画像から A を推定する手法は多くあるが，我々は深層学習による推定器を実装した．実装したネットワーク構造の詳細を表 1 に示

© 2020 Information Processing Society of Japan

Vol.2020-CG-180 No.29
Vol.2020-DCC-26 No.29
Vol.2020-CVIM-223 No.29
2020/11/6

表 1　大気散乱光 A を推定するネットワーク

Layer	Kernel	Channel	Input
conv1	7	3/16	I
conv2	5	16/32	conv1
conv3	3	32/64	conv2
conv4	3	64/128	conv3
conv5	3	128/256	conv4
conv6	3	256/256	conv5
glb_avg_pool	-	256/256	conv6
conv7	1	256/64	glb_avg_pool
conv8	1	64/1	conv7

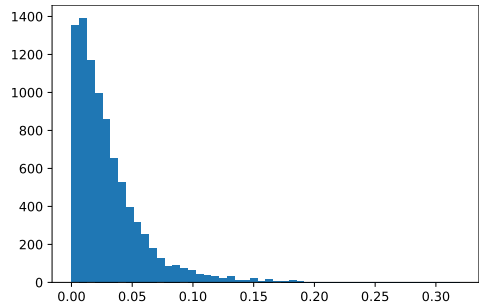


図 4　A の絶対誤差のヒストグラム

す．ネットワークには reference 画像を一枚入力し，ストライドが 2 である複数の畳み込み層ののち global average pooling により $256 \times 1 \times 1$ の特徴マップを生成する．その後，2 層の 1×1 convolution によって一次元の A を出力するようにした．なお，最終層（conv8）以外の畳み込み層には batch norm と ReLU が含まれている．学習は [2] の散乱合成画像データセットを用い，テストデータで評価を行った．正解との絶対誤差をヒストグラムにしたものを図 4 に示す．このデータセット中の A は $[0.7, 1.0]$ からランダムにサンプリングされており，したがって A については 1 枚の画像からでもある程度推定が可能であるということがわかる．

2.3.2　散乱係数 β の推定の困難さ

一方で，β はこのように 1 枚の画像から推定することは困難である．式 (1) に示したように，散乱による画像の劣化は $e^{-\beta z}$ を通じて散乱係数 β と深度 z に依存するが，この劣化にはスケールの不定性が存在する．つまり，任意の実数 $k \in \mathbb{R}$ に対して，散乱係数 $k\beta$，深度 $(1/k)z$ は同じ劣化度合いとなってしまう．一枚の画像からは深度のスケールを推定することはできず，したがって散乱係数を推定することは原理的に不可能である．

これに対し，Li ら [5] は，複数の視点から撮影された複数枚の画像から β を推定する手法を提案している．前提として，散乱により画像が劣化しているものの，エッジが顕著であるような部分から SfM によりカメラパラメータと疎な対応点，三次元点群が得られているとする．このとき，ある対応点のペア $I_1(u_1, v_1), I_2(u_2, v_2)$ に対して式 (1) から以下が得られる．

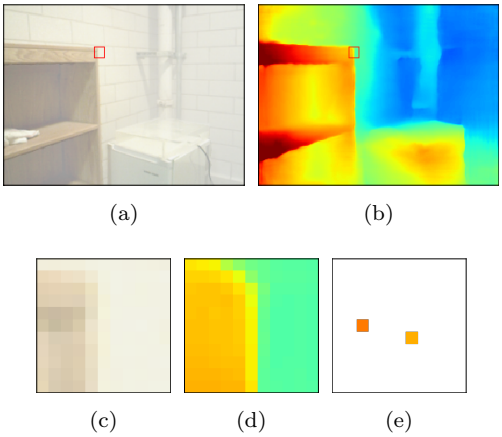


図 5　(a) 入力画像．(b) 散乱パラメータが既知であるときのネットワークが出力した深度画像．赤枠の箇所は深度が不連続である箇所（depth edge）であり，それぞれ (c) と (d) に拡大したものを示してある．(e) はこの箇所における SfM が出力した疎な三次元点群に対応する深度画像である．深度が不連続である箇所は SfM が推定した特徴点の位置が，手前側と奥側のどちらになるかは不確定であり，この図の右側のピクセルの深度のように，ネットワークが出力した深度と大きく異なってしまう可能性がある．

$$I_1(u_1, v_1) = J_1(u_1, v_1)e^{-\beta z_1(u_1, v_1)} + A(1 - e^{-\beta z_1(u_1, v_1)}) \tag{2}$$

$$I_2(u_2, v_2) = J_2(u_2, v_2)e^{-\beta z_2(u_2, v_2)} + A(1 - e^{-\beta z_2(u_2, v_2)}) \tag{3}$$

ここで，$J_1(u_1, v_1)$，$J_2(u_2, v_2)$ は元の劣化のない画像における輝度値であり，$z_1(u_1, v_1)$，$z_2(u_2, v_2)$ は SfM で得られた三次元点群に対応する各画像上での深度である．いま，元の劣化のない画像では対応点における輝度値が等しい，すなわち $J_1(u_1, v_1) = J_2(u_2, v_2)$ とすると，上式から $J_1(u_1, v_1)$，$J_2(u_2, v_2)$ を消去することで β が得られる．

$$\beta = \frac{1}{z_2(u_2, v_2) - z_1(u_1, v_1)} \ln \frac{I_1(u_1, v_1) - A(1 - e^{-\beta z_1(u_1, v_1)})}{I_2(u_2, v_2) - A(1 - e^{-\beta z_2(u_2, v_2)})} \tag{4}$$

しかしながら，この手法で β を求めるには元の劣化のない画像における輝度値 $J_1(u_1, v_1)$，$J_2(u_2, v_2)$ が等しい上で，観測した三次元点群に対応した各画像上での輝度値 $I_1(u_1, v_1)$ と $I_2(u_2, v_2)$ が十分異なる，すなわち深度 $z_1(u_1, v_1)$ と $z_2(u_2, v_2)$ が十分離れている必要がある．実際に深度が大きく離れたような対応点を取得するのは困難である．加えて，A があらかじめ正しく推定されていることも仮定しているため，このような輝度値を用いた方法では数値的に不安定であるという問題がある．

2.3.3　幾何的な情報を用いた推定

本研究では従来手法のような画像の輝度値を用いることなく散乱係数を推定する手法を提案する．深度推定の枠組みで同時に推定することにより，出力された深度の正しさも保証される．

まず最初に，従来手法と同様に本研究でもあらかじめ SfM

© 2020 Information Processing Society of Japan

Vol.2020-CG-180 No.29
Vol.2020-DCC-26 No.29
Vol.2020-CVIM-223 No.29
2020/11/6

により疎な三次元点群が得られていると仮定する．散乱媒体下での深層学習を用いた MVS の入力となる dehazing cost volume を計算するためには A と β が必要であるが，これはすなわちネットワークの出力である深度画像は A と β の関数であるとみなすことができる．したがって，ネットワークを $\mathcal{F}$ とすると，出力される深度画像は A と β の関数として $\mathbf{z}_{A,\beta} = \mathcal{F}(A, \beta)$ と書くことができる．なお，簡単のため関数の引数から入力画像は省略してある．ここで，SfM によって得られた疎な三次元点群に対応する深度画像を $\mathbf{z}_{sfm}$ とし，A と β を以下のようにして求める．

$$A^*, \beta^* = \underset{A,\beta}{\operatorname{argmin}} \sum_{u,v} m(u,v)\rho\Big(z_{sfm}(u,v), z_{A,\beta}(u,v)\Big) \tag{5}$$

ここで，$z_*(u,v)$ は深度画像 $\mathbf{z}_*$ のピクセル (u,v) での値である．$m(u,v)$ は SfM による深度が存在するピクセルでは 1，そうでない場合は 0 である．関数 ρ は与えられた二つの深度の差を計算する．つまり，上式の右辺を最小とする A と β を用いて推定された深度画像は，SfM で得られた疎な深度画像と一致する．これにより得られた A^*, β^* を用いて，最終的に密な深度画像 $\mathbf{z}^* = \mathcal{F}(A^*, \beta^*)$ を得ることができる．この手法では Li ら [5] の手法と異なり幾何的な情報を用いて最適化を行うため画像の輝度値が不要であり，最終的に出力された深度については少なくとも SfM で得られた深度に一致することが保証される．

関数 ρ については本研究では以下で定義した．

$$\rho\Big(z_{sfm}(u,v), z_{A,\beta}(u,v)\Big) = \min \left\{ \begin{array}{l} |z_{sfm}(u,v) - z_{A,\beta}(u,v)|, \\ |z_{sfm}(u,v) - z_{A,\beta}(u+\delta,v)|, \\ |z_{sfm}(u,v) - z_{A,\beta}(u-\delta,v)|, \\ |z_{sfm}(u,v) - z_{A,\beta}(u,v+\delta)|, \\ |z_{sfm}(u,v) - z_{A,\beta}(u,v-\delta)| \end{array} \right\} \tag{6}$$

図 5 に示すように，深度が不連続である箇所（depth edge）において SfM が特徴点を検出した場合，深度が手前と奥のどちらのピクセルを特徴点として検出するかは不確定である．したがって，SfM が推定した深度とネットワークが出力した深度が大きく異なってしまう可能性がある．この誤差による散乱パラメータ推定への影響を小さくするため，SfM が推定した深度とネットワークが推定した深度との差を単純に用いるのではなく，式 (6) のように各ピクセルにおいて水平と垂直方向に δ ピクセル離れた位置の深度の値も同時に参照し，最も差が小さかったものを最適化に用いるようにする．なお，本研究では $\delta = 5$ と設定した．

2.3.4 最適解探索

ネットワーク自体は A と β について微分可能であるため，式 (5) の最小化問題には勾配降下法を用いることができる．しかしながら，誤差逆伝播を用いた方法では容易に局所解となってしまうことがわかったため，あらかじめ探索範囲を定めグリッドサーチを行うことにより最適解を求めるようにする．図 6 は正解の A が与えられたときに β だけを探索した例である．図 6(a) は入力した画像で，(b) は SfM で得られた疎な深度画像である．(c) の横軸は β で，各 β で式 (5) をプロットしてある．緑の破線は正解の β の位置を示しており，これが大域最小値と一致していることがわかる．(d) はこの最小となる β を用いてネットワークが出力した深度画像である．

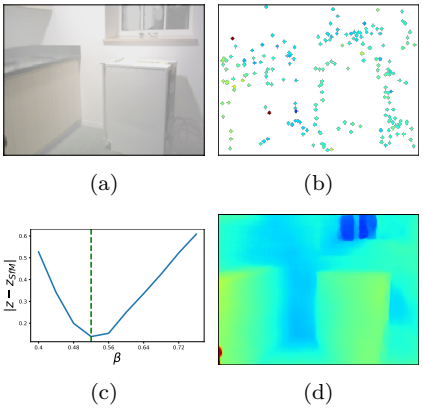


図 6 (a) 入力画像．(b) SfM で得られた疎な深度画像．(c) β を探索したときの誤差．(d) 最終的に出力した深度画像

2.3.1 で述べたように，A については深層学習を用いて一枚の画像からでもある程度推定が可能である．したがって，この推定をパラメータ探索の初期値に用いることができる．ネットワークによって推定された値を A_0 とし，本研究ではまず最初に β の探索範囲を $[\beta_{min}, \beta_{max}]$ として β を探索する．

$$\beta_0 = \underset{\beta}{\operatorname{argmin}} \sum_{u,v} m(u,v)\rho\Big(z_{sfm}(u,v), z_{A_0,\beta}(u,v)\Big) \tag{7}$$

その後，A と β の探索範囲を $[A_0 - \Delta_A, A_0 + \Delta_A]$，$[\beta_0 - \Delta_\beta, \beta_0 + \Delta_\beta]$ として，式 (5) を満たす A^*，β^* を求める．

3. 実験

3.1 学習モデル

提案手法には [2] での学習済みモデルを用いる．これは 2.2 で述べたように，MVDepthNet [11] の cost volume を dehazing cost volume に置き換えたものである．この学習には，RGB 画像と深度画像の時系列データである DeMoN データセット [10] から散乱を合成した画像が用いられている．

3.2 合成データでの実験

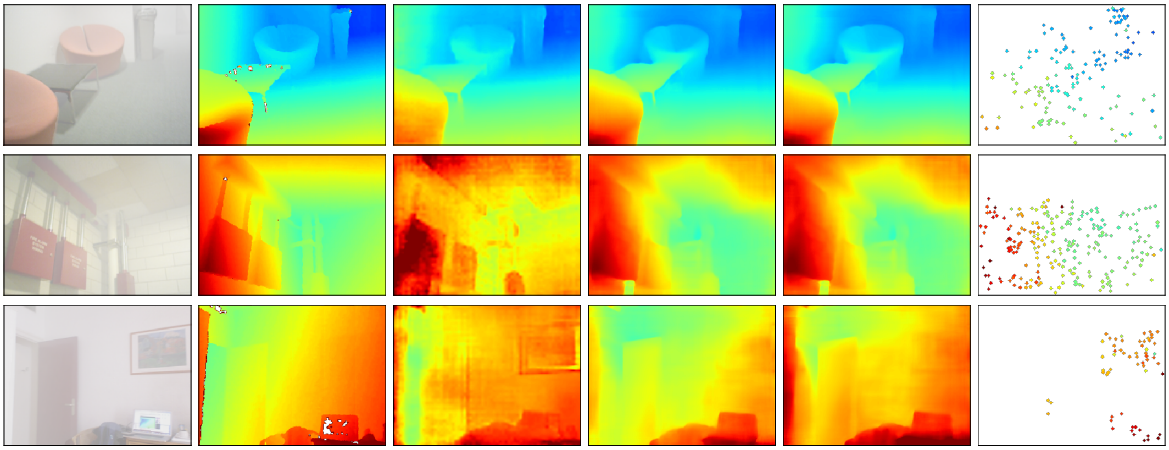
最初に，散乱を合成したデータを用いて推定された散乱パラメータと深度の評価を行う．合成データの作成には SUN3D データセット [12] を用いた．このデータセットも RGB 画像と深度画像の時系列データになっている．本研究ではこのデータセットから 68 シーンを取り出した．ま

© 2020 Information Processing Society of Japan

Vol.2020-CG-180 No.29
Vol.2020-DCC-26 No.29
Vol.2020-CVIM-223 No.29
2020/11/6

表 2　合成データでの深度推定と散乱パラメータ推定の定量評価

Dataset	Method	L1-rel	L1-inv	sc-inv	C.P. (%)	MAE_A	MAE_β
L1-rel ≤ 0.1 #1364	FFA-Net [6] + MVDepthNet [11]	0.141	0.104	0.152	57.0	-	-
	DPSNet [4]	0.109	0.069	0.125	65.2	-	-
	MVDepthNet [11]	0.130	0.090	0.135	59.9	-	-
	MVDepthNet w/ dcv [2]	0.069	0.043	0.104	80.7	-	-
	MVDepthNet w/ dcv, pe	0.081	0.050	0.116	76.3	0.028	0.043
L1-rel ≤ 0.2 #2661	FFA-Net [6] + MVDepthNet [11]	0.154	0.102	0.172	52.4	-	-
	DPSNet [4]	0.120	0.072	0.138	61.1	-	-
	MVDepthNet [11]	0.138	0.088	0.152	56.0	-	-
	MVDepthNet w/ dcv [2]	0.077	0.044	0.116	78.4	-	-
	MVDepthNet w/ dcv, pe	0.092	0.053	0.132	72.9	0.028	0.042
L1-rel ≤ 0.3 #3157	FFA-Net [6] + MVDepthNet [11]	0.162	0.103	0.182	50.7	-	-
	DPSNet [4]	0.124	0.072	0.144	59.9	-	-
	MVDepthNet [11]	0.143	0.089	0.158	54.7	-	-
	MVDepthNet w/ dcv [2]	0.079	0.045	0.120	77.6	-	-
	MVDepthNet w/ dcv, pe	0.100	0.056	0.141	70.3	0.027	0.044



(a)　(b)　(c)　(d)　(e)　(f)

図 7　合成データでの出力結果．(a) 散乱合成画像．(b) 正解の深度画像．(c) DPSNet [4] の出力．(d) 散乱パラメータ既知での [2]．(e) 提案手法．(f) SfM で得られた疎な深度画像

た，各シーンについて 80 フレームを選択した．画像はもとの 680×480 から 512×384 の部分を切り出し，その後 256×192 にリサイズして用いた．散乱を合成するためには密な深度画像が必要であるが，このデータセットの深度画像には欠損した部分が存在する．したがって，最初に劣化のない画像で学習した MVDepthNet [11] の出力を用いて欠損部分を修復した．散乱を合成するためのパラメータは，各シーンについて A と β をそれぞれ $[0.7, 1.0]$，$[0.4, 0.8]$ の範囲でランダムにサンプリングを行った．ネットワークへの入力には，各シーンから二枚ずつ画像を取り出し，一枚を reference 画像，もう一枚を source 画像として用いた．

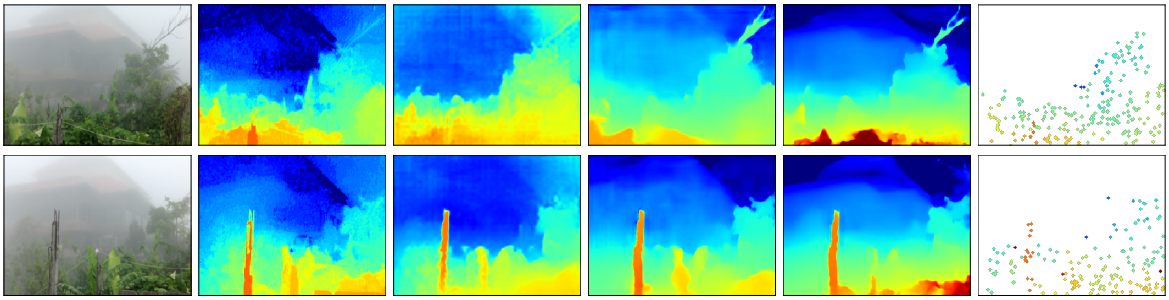
疎な深度画像を作成するため，最初に各シーンの 80 フレームを SfM [7], [8] に入力した．元の SUN3D データセットの深度画像を用いて定量評価を行うため，得られた疎な深度画像のスケールを元の深度画像に合わせた．また，カメラパラメータについても元のデータセットに含まれているものを用いた．

A と β の探索範囲は，最初の β の探索については $\beta_{min} = 0.4$，$\beta_{max} = 0.8$ でステップ数を 10 とし，その後 $\Delta_A = 0.05$，$\Delta_\beta = 0.05$ としてステップ数を 4×4 とした．したがって，ネットワークの forward の計算は合計で 26 回であり，約 15 秒ほどで推定が可能である．

実験結果を表 2 に示す．深度推定手法として，dehazing (FFA-Net [6]) を行ってから綺麗な画像で学習した MVDepthNet を適用したもの，深層学習ベースの MVS である DPSNet [4] と MVDepthNet をそれぞれ散乱合成画像で学習したもの，dehazing cost volume を用いた散乱媒体下での MVS [2]，散乱パラメータの推定を追加した提案手法を比較した．[2] については，正解の A と β を入力に与えてある．推定された深度については，4 つの指標 (L1-rel, L1-inv, sc-inv, C.P. [11]) を用いている．A と β については，絶対誤差によって評価を行っている．なお，評価に用いた 3 つのデータセットについては，入力となる SfM で出力された疎な深度画像の相対絶対誤差の平均 (L1-rel) が

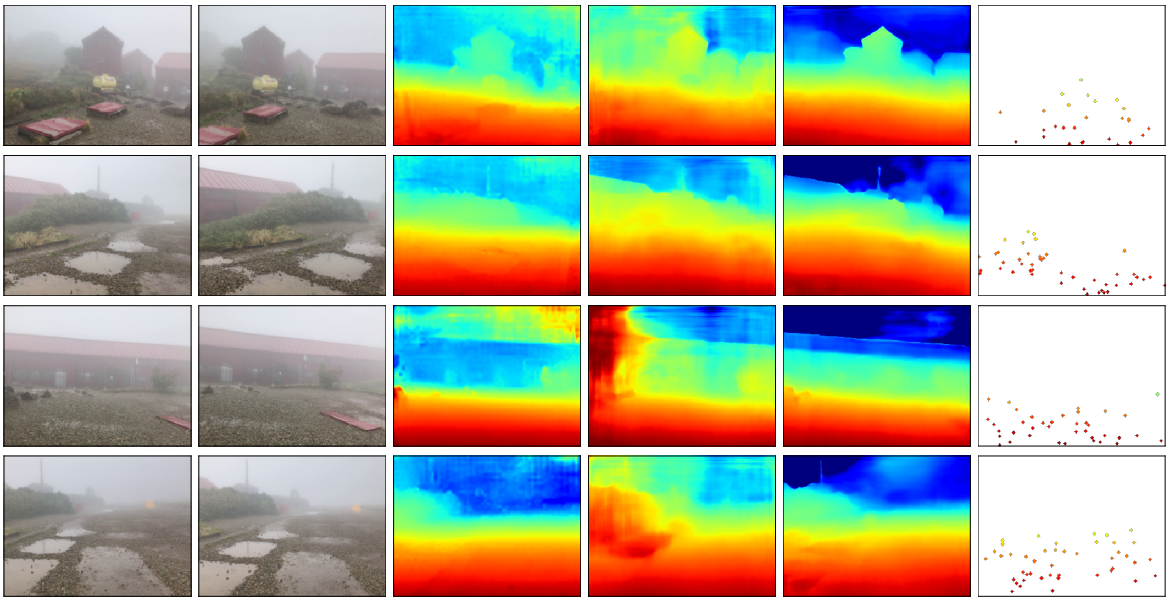
© 2020 Information Processing Society of Japan

Vol.2020-CG-180 No.29
Vol.2020-DCC-26 No.29
Vol.2020-CVIM-223 No.29
2020/11/6



(a) (b) (c) (d) (e) (f)

図 8 実際の霧がかったシーン（*bali* [5]）での出力結果．(a) 入力画像．(b) Li ら [5] の出力．(c) DPSNet [4] の出力．(d) MVDepthNet [11] の結果．(e) 提案手法．(f) SfM で得られた疎な深度画像



(a) (b) (c) (d) (e) (f)

図 9 実際の霧がかったシーンでの出力結果．(a) Reference 画像．(b) Source 画像．(c) DPSNet [4] の出力．(d) MVDepthNet [11] の結果．(e) 提案手法．(f) SfM で得られた疎な深度画像

それぞれ，0.1，0.2，0.3 以下のものだけで構成されており，表中に各データセットのサンプル数を示してある．この表から，正解の A と β が与えられた場合の [2] が最も深度推定の精度が良いことがわかる．一方で，散乱パラメータ推定を追加した場合でも，それ以外の比較手法よりも深度推定の精度が良い．推定された A と β についても，SfM で得られた疎な深度画像に多少の誤りが含まれていてもロバストに推定できていることがわかる．

出力した深度画像の例を図 7 に示す．散乱パラメータを未知とした場合でも，既知の場合とほぼ同じ結果が得られていることがわかる．一方で，三段目の画像の左側のような，SfM の段階で特徴点がまったく得られなかった箇所については若干の精度の低下がみられる．

3.3 実データでの実験

次に実際の霧がかったシーンに対して提案手法を適用した結果を示す．最初に，Li ら [5] の手法で用いられたデータ（*bali*）に提案手法を適用した．このデータは屋外のシーンで撮影されたビデオであり，約 200 フレームほどからなる．合成データでの実験と同様，疎な深度画像を得るためすべてのフレームを SfM [7], [8] に入力した．提案手法には，reference 画像と source 画像のペアと，SfM で推定された深度画像とカメラパラメータを入力した．散乱パラメータの探索範囲は $\beta_{min} = 0.01$，$\beta_{max} = 0.1$，$\Delta_A = 0.05$，$\Delta_\beta = 0.01$ とした．ステップ数は合成データでの実験と同じである．

出力結果の例を図 8 に示す．ここで，(b) は深層学習を用いない散乱媒体下での MVS である Li ら [5] の手法の結果である．使用したカメラパラメータが異なるため，出力された深度画像は SfM で得られた深度画像 (f) にスケールを合わせてある．また，(d) は図 7(d) とは異なり，MVDepthNet を散乱合成画像で学習したものの出力結果

© 2020 Information Processing Society of Japan

である．この結果より，提案手法ではカメラに近い位置にある物体の詳細が失われてしまっているものの，散乱による劣化の大きいカメラから離れた箇所については提案手法による推定が上手くいっていることがわかる．

最後に，霧がかった悪天候下で我々が実際に撮影したデータについても適用してみた．このデータについてもビデオになっており，さきほど同様すべてのフレームに対して SfM [7], [8] を適用して疎な深度画像とカメラパラメータを推定した．提案手法には reference 画像と source 画像のペアを入力し，散乱パラメータの探索範囲，ステップ数は *bali* の場合と同じである．

出力結果の例を図 9 に示す．(a)(b) はそれぞれ入力した reference 画像と source 画像である．先ほどと同様カメラから離れた劣化の大きい箇所では提案手法が最も精度が良いことがわかる．

4. まとめ

本研究では，大気散乱光や散乱係数といった散乱条件が未知な散乱媒体下における深層学習を用いた MVS を提案した．散乱媒体下における深層学習を用いた MVS が散乱パラメータの関数とみなせることを利用し，SfM で得られた疎な三次元点群に一致するように散乱パラメータを最適化した．合成データを用いて散乱パラメータと深度の推定精度の定量評価を行い，加えて実際の霧がかったシーンについても深度が正しく復元できることを示した．

自動運転等の実応用ではリアルタイムで深度推定ができることが望ましい．現在の散乱パラメータの推定法では一度の推定に十数秒必要であるが，シーンの散乱パラメータが均一であると仮定すれば，あるフレームで推定した散乱パラメータを別のフレームにそのまま適用することが可能である．また，現在は 1 フレームごとに散乱パラメータを推定しているが，複数フレームにおいて単一の散乱パラメータを推定することによる精度向上なども今後の展望としてあげられる．

謝辞 本研究は JSPS 科研費 18H03263，19J10003 の助成を受けたものである．

参考文献

[1] Berman, D., Treibitz, T. and Avidan, S.: Air-Light Estimation Using Haze-Lines, *The IEEE International Conference on Computational Photography (ICCP)* (2017).
[2] Fujimura, Y., Sonogashira, M. and Iiyama, M.: Dehazing Cost Volume for Deep Multi-view Stereo in Scattering Media, *Asian Conference on Computer Vision (ACCV)* (2020).
[3] He, K., Sun, J. and Tang, X.: Single Image Haze Removal Using Dark Channel Prior, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 12, pp. 2341–2353 (2011).
[4] Im, S., Jeon, H., Lin, S. and Kweon, I. S.: DPSNet: End-to-end Deep Plane Sweep Stereo, *International Conference on Learning Representations (ICLR)* (2019).
[5] Li, Z., Tan, P., Tang, R. T., Zou, D., Zhou, S. Z. and Cheong, L.: Simultaneous Video Defogging and Stereo Reconstruction, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4988–4997 (2015).
[6] Qin, X., Wang, Z., Bai, Y., Xie, X. and Jia, H.: FFA-Net: Feature Fusion Attention Network for Single Image Dehazing, *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, pp. 11908–11915 (2020).
[7] Schönberger, J. L. and Frahm, J. M.: Structure-from-Motion Revisited, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113 (2016).
[8] Schönberger, J. L., Zheng, E., Pollefeys, M. and Frahm, J.: Pixelwise view selection for unstructured multi-view stereo, *The European Conference on Computer Vision (ECCV)*, pp. 501–518 (2016).
[9] Tan, R. T.: Visibility in Bad Weather from a Single Image, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8 (2008).
[10] Ummenhofer, B., Zhou, H., Uhrig, J., Mayer, N., Ilg, E., Dosovitskiy, A. and Brox, T.: DeMoN: Depth and Motion Network for Learning Monocular Stereo, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5038–5047 (2017).
[11] Wang, K. and Shen, S.: MVDepthNet: real-time multiview depth estimation neural network, *International Conference on 3D Vision (3DV)*, pp. 248–257 (2018).
[12] Xiao, J., Owens, A. and Torralba, A.: SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels, *The IEEE International Conference on Computer Vision (ICCV)*, pp. 1625–1632 (2013).

© 2020 Information Processing Society of Japan