Vol.2019-CG-176 No.3
Vol.2019-DCC-23 No.3
Vol.2019-CVIM-219 No.3
2019/11/7

散乱媒体下での Multi-view Stereo のための Dehazing Cost Volume の提案

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，深層学習による Multi-view Stereo (MVS) を用いた三次元復元手法を提案する．深層学習による MVS 手法の一つである MVDepthNet は，cost volume (CV) を入力とし奥行き画像を出力するネットワークである．CV は，複数カメラ間の photometric consistency を評価するもので，物体がカメラに正対するある平面上に存在するとした場合にどの程度 consistency が保たれているかをコストとして与えたものである．しかしながら，散乱媒体下では観測した画像は光の散乱現象により劣化するため，通常の CV による photometric consistency の計算では精度が低下する．本研究では，散乱媒体下での新たな CV として dehazing cost volume (DCV) を提案する．DCV 内では，散乱光により劣化した画像の復元とコストの計算を同時に行うことができる．実験により，DCV が散乱媒体下での MVS に有効であることを示す．

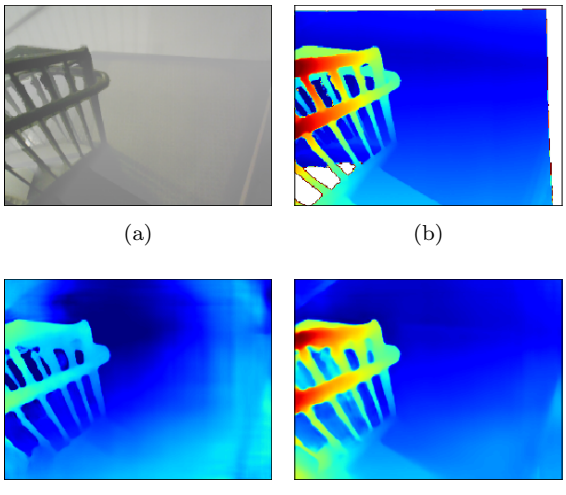
キーワード：散乱媒体，Multi-view Stereo，dehazing, cost volume

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．

本研究では，深層学習による Multi-view Stereo (MVS) を用いた散乱媒体下での三次元復元手法を提案する．MVS とは，複数のカメラで撮影した画像からシーンの三次元情報を復元する技術のことである [11]．近年，深層学習を用いた MVS が多く提案されており [14], [15], [34]，高い精度を達成している．本研究では，その中の一つである Wang と Shen [31] により提案された MVDepthNet を用いる．

MVDepthNet [31] は，ネットワークに cost volume (CV) [6] を入力することで，奥行き（depth）画像を推定する．CV 内では，複数カメラ間の photometric consistency に基づいてカメラからの奥行きに対するコストが計算される．しかしながら，先ほど述べたように，散乱媒体下では光の散乱現象により観測した画像が劣化するため，図 1(c) のように通常の CV を用いるだけでは三次元復元の精度が低下してしまうという問題がある．



(a)　(b)　(c)　(d)

図 1　(a) 散乱による劣化画像（合成）．(b) 正解の奥行き．(c) 劣化画像でファインチューニングした MVDepthNet [31] の出力．(d) 提案手法の出力．

この問題に対して，本研究では散乱媒体下で用いる新たな CV として，dehazing cost volume (DCV) を提案する．散乱媒体下では，シーンが遠くにあればあるほど，シーン

[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

で反射した光は減衰し，一方で観測される散乱光が大きくなる．つまり，散乱媒体下での画像の劣化はシーンの奥行きに依存する．DCV はこのような奥行き依存の劣化に対する画像復元と，photometric consistency の計算を同時に行うことができる CV であり，DCV を用いることで，散乱媒体下においてもロバストに奥行きの推定を行うことができる（図 1(d)）．

本研究では，複数の RGB 画像と奥行き画像，そしてそれらのカメラパラメータからなるデータセット [30] から，散乱媒体下における合成画像を作成した．実験では，単純なファインチューニングによるモデルや通常の画像復元とを組み合わせた手法に対して比較を行ない，DCV の有効性を検証した．

2. 関連研究

2.1 Multi-view Stereo

MVS [11] とは，複数のカメラで撮影した画像からシーンの三次元情報を復元する技術のことである．基本的には，画像間の密な対応点から三角測量の原理により三次元復元を行う．画像間の対応は，photometric consistency と呼ばれる画像間の輝度値の類似度から算出される．Photomtric consistency の計算を行う際に生じる主要な問題としては隠蔽があげられる．すなわち，一部の画像において対象とする物体表面が隠蔽されてしまうことにより，間違った対応付けが生じてしまう問題である．この問題に対して，MVS での三次元復元と photometric consistency を計算するのに効率的な画像選択を同時に行う手法が提案されており [23], [36]，高精度な三次元復元を可能としている．

しかしながら，テクスチャの無い物体や鏡面反射などの視点に依存した反射特性をもつ物体など，従来の MVS では困難なシーンはそのほかにも数多く存在する．これに対し，近年提案されている学習ベースの手法は，学習データからセマンティックな情報を自動で学習することにより，上に述べたようなシーンに対してもロバストな三次元復元を行うことを可能にしている．

学習ベースの MVS の多くでは CV を構築する方法が取られる．例えば，Wang と Shen [31] は，対象とするカメラから撮影された画像とそれ以外のカメラから撮影された画像から CV を生成し，直接ネットワークの入力に用いることで，対象とするカメラにおける奥行き画像を推定する手法を提案している．Huang ら [14] は，奥行き画像の推定対象であるカメラ以外のカメラについて，撮影された画像を CV と同様の方法で射影ををを行い，その後 patch matching ネットワークにより局所的なコストの計算を行っている．Yao ら [34] や Im ら [15] は，入力画像から直接 CV を作成するのではなく，一旦ネットワークに入力し，出力された特徴マップから CV を計算する手法を提案している．本研究は Wang と Shen [31] の MVDepthNet をベースラインとして用い，CV を散乱媒体下で適用できるように拡張する．

2.2 Dehazing

散乱媒体下では光の減衰と散乱光によって観測される画像のコントラストが低下する．コンピュータビジョンや画像処理の分野では，このような環境下で撮影された画像を入力し，劣化する前の画像を復元する研究が行われている [1], [8], [12], [20]．これらの手法は dehazing や defogging と呼ばれており，例えば He ら [12] や Berman ら [1] は，dark channel prior や haze-line prior などの prior ベースの dehazing 手法を提案している．また，近年は深層学習を用いた学習ベースの手法も多く提案されている [3], [21], [33], [35]．散乱除去と物体検出といった通常のタスクを組み合わせることにより，散乱媒体下でのタスクの精度が向上することが報告されている [16]．

2.3 散乱媒体下での三次元復元

散乱媒体下で画像復元ではなく，本研究と同様に直接シーンの三次元復元を行うことを目的とした研究も行われている．通常のカメラを使った手法としては，structured light を用いたもの [19] や，照度差ステレオ法を用いたもの [10], [18], [29] などが挙げられる．これらの手法は動的な光源を必要とするため，アプリケーションとしては限定的である．カメラではなくより特殊なセンサ，例えば Time-of-Flight (ToF) カメラや single photon avalanche diode (SPAD) を用いたものとしては [13], [22] があるが，これらは特殊なハードウェアのセッティングを要する．

本研究は通常のカメラを用いて，動的な光源を必要としないステレオにより三次元復元を行う．散乱媒体下で二眼ステレオを行う手法としては，Caraffa ら [4] の手法が挙げられる．この手法では MRF を用いて，画像復元とステレオによる三次元復元を同時にモデル化する．また，Song ら [25] は深層学習を用いて，画像復元と二眼ステレオをマルチタスクにして学習を行う手法を提案している．本研究に最も近いものとしては，Li ら [17] による散乱媒体下で MVS を行う手法が挙げられる．この手法は，MVS と画像復元を同時に定式化した上で，奥行きの正則化にラプラシアン平滑化と画像復元で得られる光学的深さの大小関係を用いる．しかしながら，MVS と画像復元を同時に定式化しているものの最適化は反復計算によって行われ，また各反復においてはグラフカットによって最適化が行われるため計算コストが大きいという問題がある．これに対し，本研究は深層学習を用いる手法であり，hand-crafted な正則化を用いることなく，end-to-end で学習し高速に推論を行うことができる．

Vol.2019-CG-176 No.3
Vol.2019-DCC-23 No.3
Vol.2019-CVIM-219 No.3
2019/11/7

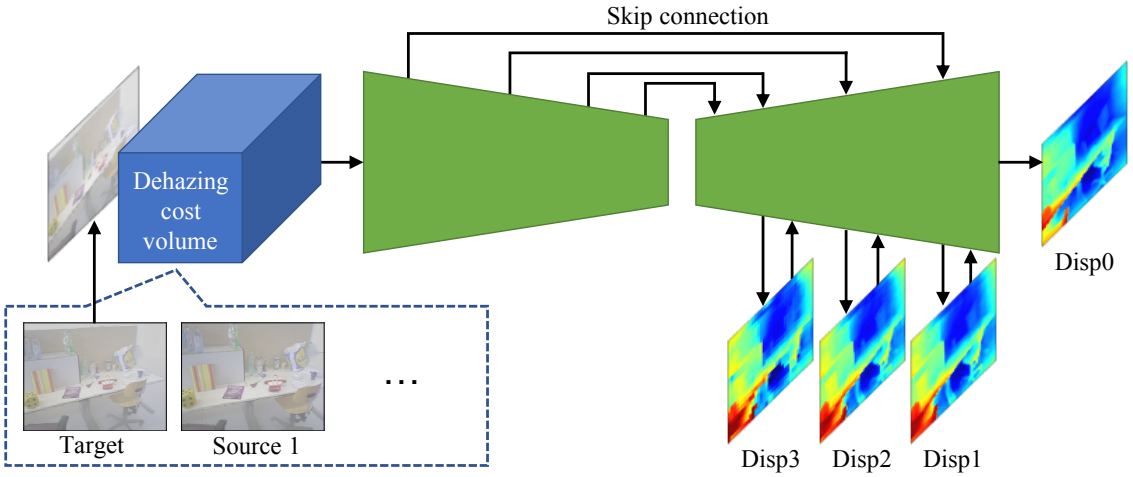
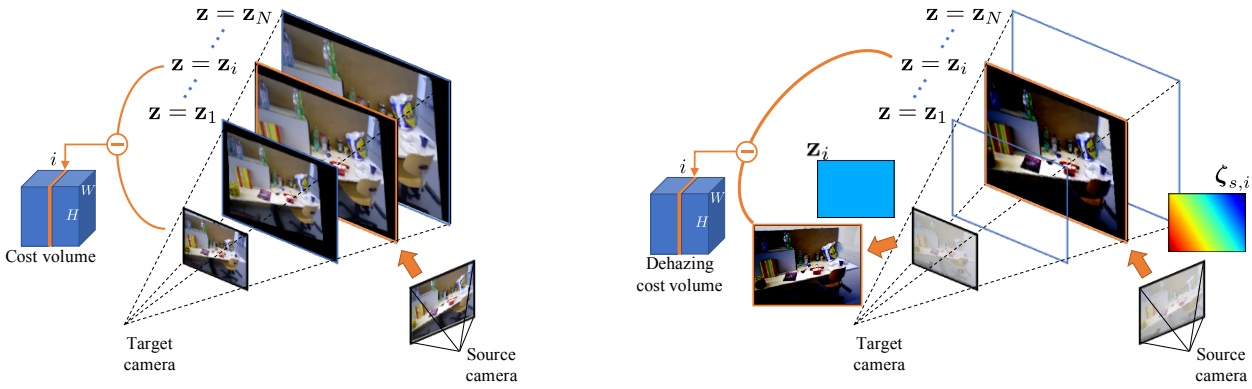


図 2　本研究で用いるネットワーク構造．ネットワークの入力はターゲット画像と DCV である．別々の層で異なる解像度の視差画像（奥行きの逆数）を 4 つ出力する．詳細は文献 [31] を参照．



(a)　(b)

図 3　(a) Cost volume．(b) Dehazing cost volume．

3. 散乱媒体下での Multi-view Stereo

本章では提案手法である散乱媒体下での MVS について説明する．最初に，大気散乱モデルを用いた散乱媒体下でのカメラの観測について述べる．その後，問題の定式化とネットワーク構造について説明を行う．最後に提案手法である DCV について説明を行う．

3.1 大気散乱モデル

多くの dehazing 手法では，散乱媒体下での画像の観測として大気散乱モデル [27] 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．本研究では各チャンネルの輝度値は 0 から 1 の値で与えられているとする．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．

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

ここで，$z(u,v)\in\mathbb{R}$ はピクセル (u,v) でのシーンの奥行き，$\beta\in\mathbb{R}$ は散乱媒体の濃度を表す散乱係数，$A\in\mathbb{R}$ は大気散乱光である．一項目はシーンで反射した光の成分であり，奥行き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，反射成分とは逆に奥行きに対して増大する．したがって，散乱による画像の劣化はシーンの奥行きに依存している．

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，β, A 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．一方で，A の推定については，例えば [12] や [2] などの従来手法が存在する．また，β については，[17] で述べられているように，カメラが多数あるという問題設定では推定することが可能である．したがって，本稿では以降，A と β はすでに得られているものと仮定する．

Vol.2019-CG-176 No.3
Vol.2019-DCC-23 No.3
Vol.2019-CVIM-219 No.3
2019/11/7

3.2 MVS の定式化

MVS には，対象とする三次元形状を点群ベースで復元したり，サーフェスベースで復元したりする手法が存在する．本研究はその中で，複数台のカメラのうちある一台のカメラを対象として，そのカメラにおける奥行き画像を推定するという問題を考える．ここで，対象とするカメラを本稿では以降ターゲットカメラ t，それ以外のカメラをソースカメラ $s \in \{1, \cdots, S\}$ とし，それぞれで撮影された画像をターゲット画像 I_t，ソース画像 I_s と定義する．この問題は一般に以下のように定式化できる．

$$\mathbf{z} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_s \sum_{u,v} \rho\left(I_t(u,v), I_s\Big(\pi_{t\to s}(u,v;z(u,v))\Big)\right) \tag{2}$$

ここで，$\mathbf{z}$ はターゲットカメラにおける奥行き画像で，ピクセル (u,v) での値は $z(u,v)$ で与えられる．関数 $\rho(f,g)$ は photometric consistency，すなわち f と g の類似度を計算する．$\pi_{t\to s} : \mathbb{R}^2 \to \mathbb{R}^2$ は，ターゲットカメラのピクセル (u,v) を奥行きを用いてソースカメラ上に射影する写像であり，以下で与えられる．

$$\begin{bmatrix} \pi_{t\to s}(u,v;z) \\ 1 \end{bmatrix} \sim z\mathbf{K}_s\mathbf{R}_{t\to s}\mathbf{K}_t^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} + \mathbf{K}_s\mathbf{t}_{t\to s} \tag{3}$$

$\mathbf{K}_t$ と $\mathbf{K}_s$ はターゲットカメラとソースカメラの内部パラメータ，$\mathbf{R}_{t\to s}$ と $\mathbf{t}_{t\to s}$ はターゲットカメラ座標系からソースカメラ座標系への回転行列と並進ベクトルである．

3.3 ネットワーク構造

MVDepthNet [31] では 3.2 章での定式化のもと，深層学習を用いて奥行き画像の推定を行う．本研究では MVDepthNet と同じネットワーク構造を用いる（図 2）．ネットワークの入力は，奥行き画像の推定対象であるターゲット画像と DCV である．DCV は後述するように，ターゲット画像と複数枚のソース画像から計算される．ネットワークの出力はターゲットカメラにおける視差画像（disparity，奥行きの逆数 $1/\mathbf{z}$）である．なお，本研究では [31] と同様に別々の層で異なる解像度のものを 4 つ出力する．学習する際は，これらそれぞれに対して正解との $L1$ 誤差を計算し，それらの和を誤差関数として用いる．

3.4 Dehazing cost volume

本研究ではネットワークの入力にターゲット画像と DCV を用いる．DCV を用いることによって，散乱による劣化を考慮したコストを計算することが可能になる．

最初に通常の CV について説明を行う．CV の計算を図 3(a) に示す．まず初めにターゲットカメラの座標系において，空間を平面で走査し奥行き方向にサンプリングを行う．その後，各奥行きに対応する平面上にソースカメラで撮影された画像を射影する．そして，ターゲットカメラで撮影された画像と射影された画像間で輝度値の差分を取ることで，その奥行きに対する各ピクセルのコストを計算する．したがって，画像のサイズを $W \times H$，奥行きのサンプリング数を N とすると，CV は $\mathcal{V} : \{1, \cdots, W\} \times \{1, \cdots, H\} \times \{1, \cdots, N\} \to \mathbb{R}$ で与えられ，各要素は以下のようになる．

$$\mathcal{V}(u,v,i) = \frac{1}{S}\sum_s \|I_t(u,v) - I_s(\pi_{t\to s}(u,v;z_i))\|_1 \tag{4}$$

ここで z_i は i 番目の平面の奥行きの値である．CV は各ピクセルにおいて，サンプリングした奥行きに対応する photometric consistency を計算していることになり，正しい奥行きに対応する CV の要素は理想的には 0 になる．

散乱媒体下では式 (1) にしたがって画像が劣化するため，式 (4) で定義される通常の CV はこの劣化の影響を受けてしまう．提案手法である DCV では，劣化した画像の復元を行いながらコストを計算する．3.1 章で述べたように，散乱による画像の劣化はシーンの奥行きに依存するので，各平面でシーンを走査する際，その平面での奥行きを用いて画像の復元を行う．

DCV の計算を図 3(b) に示す．ターゲット画像に対しては，シーンを走査した平面の奥行きの値を用いて画像の復元を行う．ソース画像に対しては，ソースカメラ座標系における平面の奥行きの値を計算し，それを用いて画像の復元を行なったのちにターゲットカメラ座標系への射影を行う．したがって，DCV を $\mathcal{D} : \{1, \cdots, W\} \times \{1, \cdots, H\} \times \{1, \cdots, N\} \to \mathbb{R}$ と定義すると，各要素は以下の式で与えられる．

$$\mathcal{D}(u,v,i) = \frac{1}{S}\sum_s \|J_t(u,v;z_i) - J_s(\pi_{t\to s}(u,v;z_i))\|_1 \tag{5}$$

ここで，$J_t(u,v;z_i)$ と $J_s(\pi_{t\to s}(u,v;z_i))$ は復元されたターゲット画像とソース画像であり，式 (1) から

$$J_t(u,v;z_i) = \frac{I_t(u,v) - A}{e^{-\beta z_i}} + A \tag{6}$$

$$J_s(\pi_{t\to s}(u,v;z_i)) = \frac{I_s(\pi_{t\to s}(u,v;z_i)) - A}{e^{-\beta \zeta_{s,i}(\pi_{t\to s}(u,v;z_i))}} + A \tag{7}$$

となる．図 3(b) で示すように，ターゲット画像では奥行き z_i の平面の奥行き画像 $\mathbf{z}_i$ を用いて画像復元が行われる．対して，ソース画像では奥行き画像 $\boldsymbol{\zeta}_{s,i}$ を用いて画像復元が行われる．$\boldsymbol{\zeta}_{s,i}$ は，ターゲットカメラ座標系における奥行き z_i の平面をソースカメラからみたときの奥行き画像である．ターゲット画像のピクセル (u,v) におけるコストを計算する際は，それに対応するソース画像のピクセル $\pi_{t\to s}(u,v;z_i)$ での奥行き $\zeta_{s,i}(\pi_{t\to s}(u,v;z_i))$ が用いられる．DCV では劣化した画像ではなく，コントラストの復

元された画像を用いることで，より効果的なコストを計算することができる．また，明らかなように，劣化のない場合の画像における photometric consistency が保証される．

DCV はサンプリングしたすべての平面の奥行きの値を用いて画像復元を行う．したがって，正しい奥行きから大きく離れた平面を用いて画像復元を行うことで，画像復元の結果が本来の劣化する前の画像と大きく異なってしまう場合がある．極端な例でいえば，画像復元によって画像の輝度値が負の値を取るといったことが考えられる．このような場合，式 (5) によって計算されたコストが非常に大きな値となってしまう可能性がある．本研究ではこの問題に対処するため，式 (5) を以下のように修正する．

$$\mathcal{D}(u,v,i) = \frac{1}{S}\sum_{s}\begin{cases} \|J_t(u,v;z_i) - J_s(\pi_{t\to s}(u,v;z_i))\|_1 \\ \quad if\ 0 \le J_t^c(u,v;z_i) \le 1\ and \\ \quad 0 \le J_s^c(\pi_{t\to s}(u,v;z_i)) \le 1\ \forall c \in \{r,g,b\} \\ \gamma \quad otherwise \end{cases} \tag{8}$$

ここで，$J_t^c(u,v;z_i)$ と $J_s^c(\pi_{t\to s}(u,v;z_i))$ は復元された画像のチャンネル $c \in \{r,g,b\}$ の輝度値である．復元された画像の輝度値が定義域から外れた場合に，コストにペナルティとして定数 γ を与えるようにする．これにより，DCV の要素の値に上限値を与えることができ，学習を安定させることができる．さらに，明示的に画像復元におけるペナルティを与えることで，奥行きの探索空間を小さくすることもできる．本研究では，ペナルティの値として $\gamma = 3$ を用いた．これは，RGB 画像の各チャンネルの輝度値の定義域が 0 から 1 の場合の，式 (4) で定義される通常の CV の要素の最大値である．

4. 実験

本研究ではベースラインとして MVDepthNet [31] を用いる．提案手法は 3.3 章で述べたように，MVDepthNet の CV を DCV に置き換えたものである．実験では提案手法に対して，ベースラインを劣化画像でファインチューニングしたモデルに加えて，通常の画像復元とを組み合わせた手法との比較を行なった．

4.1 データセット

本研究ではモデルの学習に DeMoN データセット [30] を用いる．DeMoN データセットは実画像からなる SUN3D [32]，RGB-D SLAM [26]，MVS [9] データセットと，合成画像からなる Scenes11 [5], [30] で構成されており，時系列の RGB 画像と奥行き画像，そしてカメラパラメータが与えられている．実データの奥行き画像については，センサの計測限界から欠損値が含まれている．あとで述べるように，提案手法の学習には散乱による劣化の合成画像を用いるが，合成画像を作成するためには欠損の無い密な奥行き画像が必要となる．そこで本研究では前処理として，奥行きの補完手法である bilateral filtering [24] を用いて欠損の修復を行なった．しかしながら，この手法を用いても欠損の大きな箇所については正しく復元することが難しい．したがって，修復後の欠損が画像全体の 10 パーセント以下であるものだけを学習とテストに用いた．学習時に誤差を計算する際は，もともと欠損していた領域の誤差を含めないようにした．最終的に，学習とテストに RGB 画像をペアにしたものをそれぞれ 387,120，8,343 サンプル生成した．各サンプルで，一枚をターゲット画像，もう一枚をソース画像に用いた．画像サイズは 256×192 にリサイズして用いた．

提案手法を学習する際は，先ほどのデータセットから散乱による劣化を加えた合成画像のデータセットを作成した．合成画像は式 (1) を用いて作成する．大気散乱光 A は各サンプルに対して $A \in [0.7, 1.0]$ からランダムに生成した．散乱係数 β は各サンプルに対して，各データセット，SUN3D，RGB-D SLAM，MVS ，Scenes11 において，$\beta \in [0.4, 0.8], [0.4, 0.8], [0.6, 1.0], [0.05, 0.15]$ からランダムに生成した．

4.2 実装と学習の詳細

CV を計算するには，対象とする奥行きの空間とサンプリング間隔を決定する必要がある．本研究では，視差について 0.02 から 2 の値を $N = 256$ 分割し，それに対応する奥行きの値で空間をサンプリングした．

最初にベースラインの学習を行なった．パラメータの更新には Adam を用いた．ミニバッチのサイズは 32 で，学習率は最初の 100×10^3 回は 1.0×10^{-4} で，その後 20×10^3 回の更新ごとに 0.8 倍にした．トータルで約 240×10^3 回パラメータの更新を行なった．

ベースラインの学習の後は，提案手法の学習と，比較手法としてベースラインを劣化画像のデータセットを用いてファインチューニングしたモデルの学習を行う．どちらもベースラインのパラメータを初期値として学習を行う．これらは学習率を 1.0×10^{-4} とし，20×10^3 回の更新ごとに 0.8 倍にした．それぞれトータルで約 120×10^3 回パラメータの更新を行なった．

もう一つの比較手法としてベースラインと通常の画像復元手法を組み合わせたもの，すなわち散乱による劣化を復元したあとベースラインに入力するという手法を用いた．画像復元手法としては Li ら [16] のモデルを用い，提案手法の学習に用いた劣化画像のデータセットで学習を行なった．

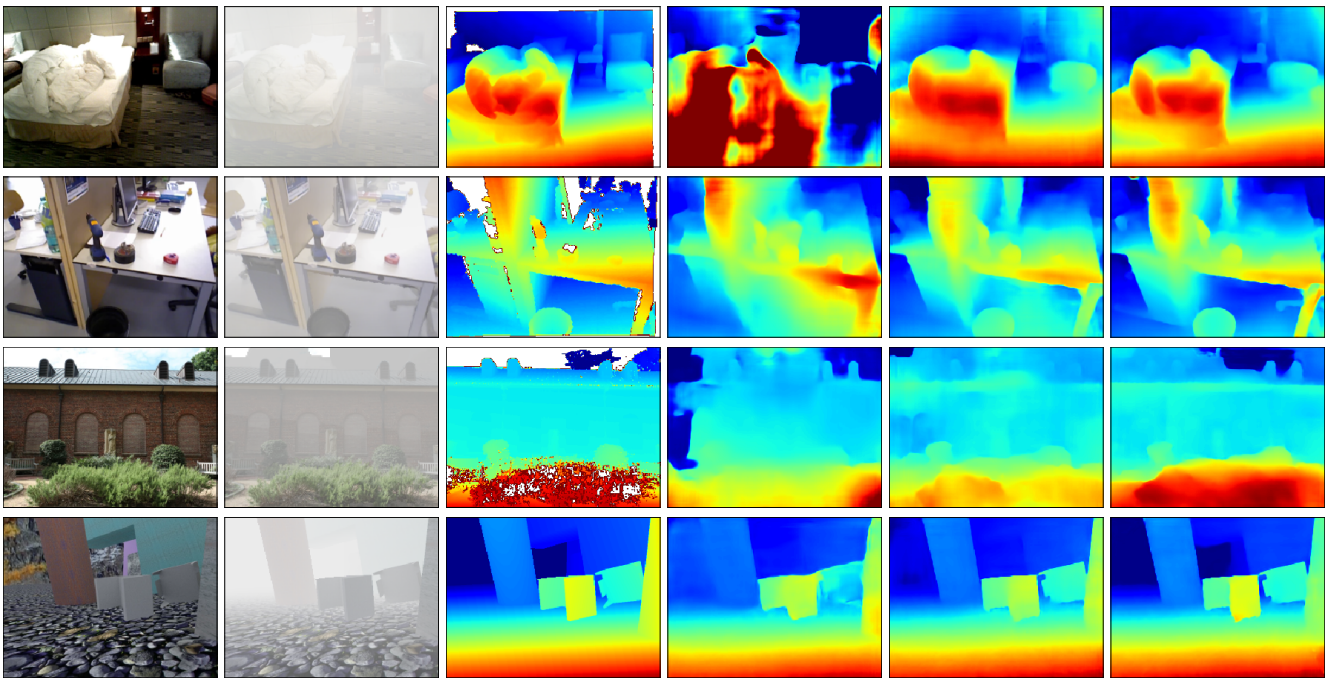
4.3 実験結果

表 1 に結果の定量評価を示す．精度評価として 3 つの評価指標を用いた．L1-rel は，推定した奥行きの相対誤差を

Vol.2019-CG-176 No.3
Vol.2019-DCC-23 No.3
Vol.2019-CVIM-219 No.3
2019/11/7

表 1　定量評価．各データセットに対して，ファインチューニングしたベースライン [31]，画像復元 [16] とベースライン [31] を組み合わせた手法，提案手法の結果を示してある．

Dataset	Method	L1-rel	sc-inv	C.P. (%)
SUN3D	Fine-tuned [31]	0.123	0.132	59.8
	Dehazing [16] + Baseline [31]	0.223	0.230	44.2
	Ours	**0.072**	**0.094**	**78.1**
RGB-D SLAM	Fine-tuned [31]	0.108	0.163	68.3
	Dehazing [16] + Baseline [31]	0.158	0.268	60.0
	Ours	**0.100**	**0.139**	**69.9**
MVS	Fine-tuned [31]	0.234	0.239	40.2
	Dehazing [16] + Baseline [31]	0.515	0.388	36.7
	Ours	**0.132**	**0.172**	**54.1**
Scenes11	Fine-tuned [31]	0.136	0.260	67.9
	Dehazing [16] + Baseline [31]	0.228	0.536	54.9
	Ours	**0.108**	**0.181**	**75.5**



(a)　(b)　(c)　(d)　(e)　(f)

図 4　実験結果．(a) 劣化の無い RGB 画像．(b) 散乱による劣化合成画像．(c) 正解の奥行き画像．(d) 画像復元 [16] したものを入力したときのベースライン [31] が出力した奥行き画像．(e) ファインチューニングしたベースライン [31] が出力した奥行き画像．(f) 提案手法が出力した奥行き画像．各行はそれぞれ，SUN3D, RGB-D SLAM, MVS, Scenes11 の画像である．

すべてのピクセルで平均したものである．sc-inv は，Eigen ら [7] が提案したスケール不変である奥行きの誤差である．C.P.（correctly estimated depth percentage）[28] は，奥行きの相対誤差が 10 パーセント以下であるピクセルの割合である．表 1 には各データセットにおける，ベースラインをファインチューニングしたモデル，dehazing とベースラインを組み合わせた手法，提案手法についてのこれら 3 つの評価指標を示してある．表 1 より，すべての指標において提案手法はファインチューニングしたモデルを上回っており，DCV が散乱媒体下での MVS において有効であることがわかる．また，通常の dehazing とベースラインを組み合わせた手法についても提案手法が精度で上回っており，散乱による画像の劣化と MVS を同時にモデル化することができる DCV が有効であることがわかる．

それぞれの手法を用いて出力された奥行き画像を図 4 に示す．それぞれ，(a) 劣化の無い RGB 画像，(b) 散乱による劣化合成画像，(c) 正解の奥行き画像，(d) 画像復元 [16] したものを入力したときのベースライン [31] が出力した奥行き画像，(e) ファインチューニングしたベースライン [31] が出力した奥行き画像，(f) 提案手法が出力した奥行

き画像である．各行はそれぞれ，SUN3D, RGB-D SLAM, MVS, Scenes11 の画像である．定性的にも提案手法が最も良い結果であることがわかる．特に，Scenes11 の結果に示すように，提案手法は DCV 内で画像のコントラストを復元するため，大きく劣化する遠くのシーンについても奥行きを推定することができる．

5. まとめ

本研究では散乱媒体下での MVS を実現するため，DCV という新たな枠組みを提案した．DCV 内では，photomeric consistency のコストの計算と，散乱による画像の劣化の復元を同時に行うことができる．合成画像を用いた実験において，通常の CV を用いたモデルをファインチューニングした手法や，MVS と画像復元とを組み合わせた手法との比較を行い，DCV が有効であることを示した．

今後の課題としては，現在は大気散乱光 A と散乱係数 β を既知として扱っているが，これらも end-to-end で推定できるように拡張することなどが挙げられる．また，合成データではなく実際に散乱媒体下で撮影された画像に対しても手法の有効性を検証する予定である．本研究で提案した DCV は，奥行きに依存した画像の劣化に対して効果的な MVS を実現するものであり，散乱以外の奥行き依存の劣化に対しても応用することが可能であると考えている．

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.: Non-Local Image Dehazing, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1674–1682 (2016).
[2] Berman, D., Treibitz, T. and Avidan, S.: Air-Light Estimation Using Haze-Lines, *The IEEE International Conference on Computational Photography (ICCP)* (2017).
[3] Cai, B., Xu, X., Jia, K., Qing, C. and Tao, D.: DehazeNet: An End-to-End System for Single Image Haze Removal, *IEEE Transaction on Image Processing*, Vol. 25, No. 11, pp. 5187–5198 (2016).
[4] Caraffa, L. and Tarel, J.: Stereo Reconstruction and Contrast Restoration in Daytime Fog, *Asian Conference on Computer Vision (ACCV)*, pp. 13–25 (2012).
[5] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L. and Yu, F.: ShapeNet: An Information-Rich 3D Model Repository, *arXiv:1512.03012* (2015).
[6] Collins, R. T.: A space-sweep approach to true multi-image matching, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 358–363 (1996).
[7] Eigen, D., Puhrsch, C. and Fergus, R.: Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, *Twenty-eighth Conference on Neural Information Processing Systems (NeurIPS)* (2014).
[8] Fattal, R.: Dehazing Using Color-Lines, *ACM Transactions on Graphics (TOG)*, Vol. 34, No. 1 (2014).
[9] Fuhrmann, S., Langguth, F. and Goesel, M.: MVE: a multi-view reconstruction environment, *Eurographics Workshop on Graphics and Cultural Heritage*, pp. 11–18 (2014).
[10] Fujimura, Y., Iiyama, M., Hashimoto, A. and Minoh, M.: Photometric Stereo in Pariticipating Media Considering Shape-Dependent Forward Scatter, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7445–7453 (2018).
[11] Furukawa, Y. and Hernández, C.: Multi-view stereo: A tutorial, *Foundations and Trends®in Computer Graphics*, Vol. 9, No. 1–2, pp. 1–148 (2015).
[12] 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).
[13] Heide, F., Xiao, L., Kolb, A., Hullin, M. B. and Heidrich, W.: Imaging in scattering media using correlation image sensors and sparse convolutional coding, *Optics Express*, Vol. 22, No. 21, pp. 26338–26350 (2014).
[14] Huang, P., Matzen, K., Kopf, J., Ahuja, N. and Huang, J.: DeepMVS: Learning Multi-View Stereopsis, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2821–2830 (2018).
[15] 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).
[16] Li, B., Peng, X., Wang, Z., Xu, J. and Feng, D.: AOD-Net: All-in-One Dehazing Network, *The IEEE International Conference on Computer Vision (ICCV)*, pp. 4770–4778 (2017).
[17] 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).
[18] Murez, Z., Treibitz, T., Ramamoorthi, R. and Kriegman, D. J.: Photometric Stereo in a Scattering Medium, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 9, pp. 1880–1891 (2017).
[19] Narasimhan, S. G., Nayar, S. K., Sun, B. and Koppal, S. J.: Structured Light in Scattering Media, *Proceedings of the Tenth IEEE International Conference on Computer Vision*, Vol. I, pp. 420–427 (2005).
[20] Nishino, K., Kratz, L. and Lombardi, S.: Bayesian Defogging, *International Journal of Computer Vision*, Vol. 98, No. 3, pp. 263–278 (2012).
[21] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X. and Yang, M.: Single Image Dehazing via Multi-scale Convolutional Neural Networks, *European Conference on Computer Vision (ECCV)*, pp. 154–169 (2016).
[22] Satat, G., Tancik, M. and Rasker, R.: Towards Photography Through Realistic Fog, *The IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10 (2018).
[23] 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).
[24] Silberman, N., Hoiem, D., Kohli, P. and Fergus, R.: Indoor Segmentation and Support Inference from RGBD Images, *The European Conference on Computer Vision (ECCV)*, pp. 746–760 (2012).
[25] Song, T., Kim, Y., Oh, C. and Sohn, K.: Deep Network

for Simultaneous Stereo Matching and Dehazing, *British Machine Vision Conference (BMVC)* (2018).

[26] Sturm, J., Engelhard, N., Endres, F., Burgard, W. and Cremers, D.: A Benchmark for the Evaluation of RGB-D SLAM Systems, *The International Conference on Intelligent Robot Systems (IROS)* (2012).

[27] 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).

[28] Tateno, K., Tombari, F., Laina, I. and Navab, N.: CNN-SLAM: Real-Time Dense Monocular SLAM With Learned Depth Prediction, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6243–6252 (2017).

[29] Tsiotsios, C., Angelopoulou, M. E., Kim, T. and Davison, A. J.: Backscatter Compensated Photometric Stereo with 3 Sources, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2259–2266 (2014).

[30] 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).

[31] Wang, K. and Shen, S.: MVDepthNet: real-time multiview depth estimation neural network, *International Conference on 3D Vision (3DV)*, pp. 248–257 (2018).

[32] 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).

[33] Yang, D. and Sun, J.: Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing, *The European Conference on Computer Vision (ECCV)*, pp. 702–717 (2018).

[34] Yao, Y., Luo, Z., Li, S., Fang, T. and Quan, L.: MVSNet: Depth Inference for Unstructured Multi-view Stereo, *The European Conference on Computer Vision (ECCV)*, pp. 767–783 (2018).

[35] Zhang, H. and Patel, V. M.: Densely Connected Pyramid Dehazing Network, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3194–3203 (2018).

[36] Zheng, E., Dunn, E., Jojic, V. and Frahm, J.: PatchMatch Based Joint View Selection and Depthmap Estimation, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1510–1517 (2014).