CH-006

Novel view synthesis のための Multiplane image の超解像

Super-Resolving Multiplane Image for Novel View Synthesis

佐藤 千幸 *　都竹 千尋 *　高橋 桂太 *　藤井 俊彰 *
Chisaki Sato　Chihiro Tsutake　Keita Takahashi　Toshiaki Fujii

1　はじめに

Novel view synthesis（NVS）とは，様々な視点から見た画像を基に新たな視点から見た画像（自由視点画像）を合成する技術である．NVS の問題は古くから研究されてきた [1, 2] が，近年では深層学習を活用した手法 [3, 4, 5, 6] に注目が集まる．本研究では，その中でも，multiplane image（MPI）と呼ばれる体積表現に基づく手法 [7, 8, 9, 10, 11, 12, 13] に着目する．MPI は，Fig. 1 に示すように，透明度を持つ複数の画像を積層したものである．各層を構成する画像は，観察方向に応じてシフトされつつ重なるため，方向に応じた見え方が実現できる．この描画処理は，コンピュータグラフィックスにおいて標準的な機能であるアルファ合成によって実装できるため，極めて高速である．さらに新しい方法論である neural radiance field の関連技術 [14, 15, 16, 17, 18] は描画に要する計算コストが高いため，描画処理の効率性は MPI のアドバンテージである．

一方，MPI では，生成される画像の解像度に原理的な限界がある．すなわち，MPI の各層は有限の解像度を持つ画像として表現されるため，生成される画像も同様に有限の解像度を持つ．そこで，本研究では，MPI の形式で表現されたデータそのものを直接的に超解像することで，高解像度での画像生成を可能にすることをめざす．MPI は多くの場合，被写体をさまざまな視点から撮影した多視点画像をもとに生成されるが，自由視点画像生成の際には MPI のみが必要である．したがって，実用上「元となる多視点画像が不明で MPI のみが入手可能」というケースが想定されるため，元画像に頼らずに MPI そのものを高解像度化する本研究の枠組みが有用である．

MPI の高解像度化を達成する安直な方法として，MPI を構成する各層を，“画像” として 2 次元の面内で超解像することが考えられる．しかし，後述するように，この方法では満足な結果が得られない．一方，提案手法では，MPI を 3 次元的に，すなわち，空間方向だけではなく奥行き方向にも同時に高解像度化する．この設計は，MPI から生成される光線空間（2 次元格子状に配列された多視点画像で構成される信号空間）におけるアンチエイリアス条件 [10] から導かれるものであり，生成画像の品質を左右する鍵を握る．

超解像の文脈においては，これまでに画像，動画像，および多視点画像を対象とする多数の手法 [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31] が提案されている．しかし，これらの技術は 2 次元的な高解像度化を行うものであり，奥行き方向も含めた 3 次元的な超解像には適用できない．そこで本研究では，低解像度の MPI を入力とし高解像度の MPI を出力とするような convolutional neural network（CNN）を新たに実装し，高解像度の光線空間を教師信号とすることで CNN の重みパラメータを最適化した．結果として，テストシーン（学習に用いておらず，高解像度の光線空間が未知であることを条件とする）に対して，低解像度の MPI を 3 次元的に超解像することに成功し，高解像度で高品質な画像生成を実現できた．

* 名古屋大学　大学院工学研究科　情報・通信工学専攻

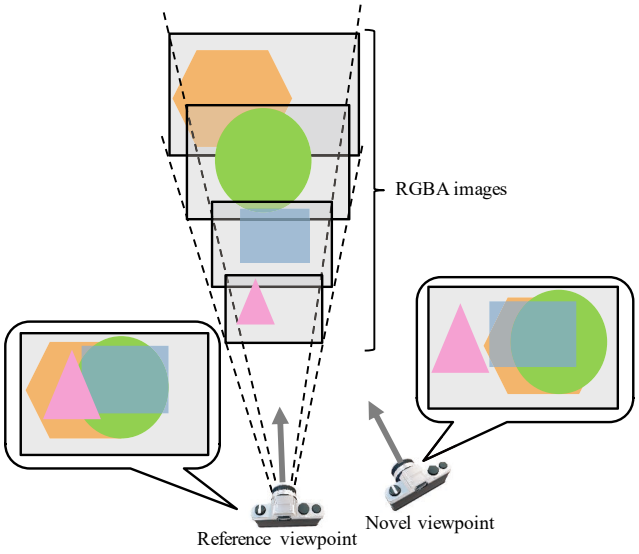

Fig.1: Schematic illustration of multiplane image (MPI).

2　Multiplane Image の原理と実装

2.1　MPI の原理

Multiplane image（MPI）は，Fig. 1 に示すように，半透明な画像を積層させた構造を持つ．MPI を構成する各画像を，観察方向に応じてシフトしつつ重ね合わせることによって，自由な視点から見た画像が生成される．

この過程を具体的に述べる．奥行きのインデックス d を，奥から手前に向かう方向に定義する．インデックス d に対応する層のカラー画像を $c_d(x,y)$，アルファ画像を $\alpha_d(x,y)$ と表記する．積層された画像で構成される MPI を視点 $\mathbf{v}$ から見ると，以下の式で与えられる画像 $\hat{I}_{\mathbf{v}}(x,y)$ が観察される．

$$\hat{I}_{\mathbf{v}}(x,y) = \sum_{d} \mathcal{W}_{\mathbf{v}}\{c_d(x,y)\alpha_d(x,y)\} \prod_{d'>d} (1 - \mathcal{W}_{\mathbf{v}}\{\alpha_{d'}(x,y)\}) \quad (1)$$

ここで，$\mathcal{W}_{\mathbf{v}}$ は，視点 $\mathbf{v}$ と奥行きのインデックス d に応じて各層の画像をワーピング（ホモグラフィ変換）するオペレータである．

あるシーンを表現する MPI を求めることは，MPI を構成する $c_d(x,y)$ および $\alpha_d(x,y)$ をシーンに合わせて適切に決定することに等しい．そこで，対象シーンを様々な方向から撮影した画像（入力画像）を何枚か用意し，式 (1) にしたがって生成される画像が入力画像と一致するように，$c_d(x,y)$ および $\alpha_d(x,y)$ を求める．この処理を実現するため，convolutional neural network（CNN）を用いた学習ベースの手法 [7, 8, 9, 10, 11, 12, 13] がよく用いられる．すなわち，何枚かの画像を入力して MPI を出力するような CNN を構築し，出力された MPI から式 (1) にしたがって生成される画像が入力画像に一致するように CNN の重みパラメータを最適化する．学習済みの CNN を用いて推論を行うことで，学習セットに含まれないシーンを撮影した入力画像から対応する MPI を求めることが可能である．

Table.1: Network architecture for G_{LR}, $G_{2D\text{-}HR}$ and $G_{3D\text{-}HR}$. $Conv2_*$ indicates 2-D convolutional layer. Ch_{in}/Ch_{out}, k, s and d indicate the number of input/output channels, the kernel size, the stride and the dilation, respectively. n_{out} is $2D+3$ for G_{LR} and $G_{2D\text{-}HR}$ ($2D$: alphas, 3: RGB background colors), while it is $2aD+3$ for $G_{3D\text{-}HR}$ ($2aD$: alphas, 3: RGB background colors).

Layer	Input	Ch_{in}/Ch_{out}	k	s	d	Activation
input	PSVs					
$Conv2_{1\text{-}1}$	input	15D/64	3	1	1	ReLU
$Conv2_{1\text{-}2}$	$Conv2_{1\text{-}1}$	64/128	3	2	1	ReLU
$Conv2_{2\text{-}1}$	$Conv2_{1\text{-}2}$	128/128	3	1	1	ReLU
$Conv2_{2\text{-}2}$	$Conv2_{2\text{-}1}$	128/256	3	2	1	ReLU
$Conv2_{3\text{-}1}$	$Conv2_{2\text{-}2}$	256/256	3	1	1	ReLU
$Conv2_{3\text{-}2}$	$Conv2_{3\text{-}1}$	256/256	3	1	1	ReLU
$Conv2_{3\text{-}3}$	$Conv2_{3\text{-}2}$	256/512	3	2	1	ReLU
$Conv2_{4\text{-}1}$	$Conv2_{3\text{-}3}$	512/512	3	1	2	ReLU
$Conv2_{4\text{-}2}$	$Conv2_{4\text{-}1}$	512/512	3	1	2	ReLU
$Conv2_{4\text{-}3}$	$Conv2_{4\text{-}2}$	512/512	3	1	2	ReLU
$Conv2_{5\text{-}1}$	($Conv2_{3\text{-}3}$, $Conv2_{4\text{-}3}$)	1024/256	4	.5	1	ReLU
$Conv2_{5\text{-}2}$	$Conv2_{5\text{-}1}$	256/256	3	1	1	ReLU
$Conv2_{5\text{-}3}$	$Conv2_{5\text{-}2}$	256/256	3	1	1	ReLU
$Conv2_{6\text{-}1}$	($Conv2_{2\text{-}2}$, $Conv2_{5\text{-}3}$)	512/128	4	.5	1	ReLU
$Conv2_{6\text{-}2}$	$Conv2_{6\text{-}1}$	128/128	3	1	1	ReLU
$Conv2_{7\text{-}1}$	($Conv2_{1\text{-}2}$, $Conv2_{6\text{-}2}$)	256/64	4	.5	1	ReLU
$Conv2_{7\text{-}2}$	$Conv2_{7\text{-}1}$	64/64	3	1	1	ReLU
$Conv2_{out}$	$Conv2_{7\text{-}2}$	$64/n_{out}$	3	1	1	Hard Sigmoid
output	$Conv2_{out}$					

2.2 本研究における実装

本研究では，MPI は $2D+1$ 枚の層で構成されるとする．z 軸を奥から手前に向かう方向に定義し，MPI の各層が一定の間隔 Δ_z で $z = d\Delta_z$ ($d \in \{-D, -D+1, \ldots, D-1, D\}$) に配置されているとする．また，MPI の各層を構成する画像のサイズを同一とし，その MPI から平行投影によって画像が生成されると仮定する．このとき，視点 $\mathbf{v}$ は方向ベクトル (u, v) で表現される．$(u, v) = (0, 0)$ を MPI の正面に対応させる．この条件においては，式 (1) におけるワーピングオペレータ $\mathcal{W}_{u,v}$ は以下の単純な平行移動として定義される．

$$\mathcal{W}_{u,v}\{\zeta_d(x, y)\} = \zeta_d(x + ud\Delta_z, y + vd\Delta_z) \quad (2)$$

また，本研究では，2 次元格子状に配列された多視点画像 $I_{u,v}(x, y)$（文脈に応じて光線空間とも呼ぶ）を表示対象とする．ここで，(u, v) は視点位置を表す．これら多視点画像は実際には透視投影のカメラで撮影されたものであるが，これを平行投影された角度画像とみなすことで MPI における方向と対応づける．すなわち，MPI を (u, v) 方向から観察したときに生成される画像を $\hat{I}_{u,v}(x, y)$ と表すと，$\hat{I}_{u,v}(x, y) \simeq I_{u,v}(x, y)$ を満たすように，$2D+1$ 枚の層を最適化する．

多視点画像から MPI を生成する処理については，先行研究 [7] と同一構成の CNN（ただし，入出力チャネル数は異なる）によって実装した．CNN のアーキテクチャを Table 1 に示す．CNN への入力は，多視点画像のうち，中央と四隅の視点から生成された 5 視点の画像の plane sweep volumes（PSVs）とした．CNN からの出力は背景画像（3 ch）とアルファ画像（$2D$ ch）とした．そして，文献 [11] の手順にしたがって，$2D+3$ ch を MPI（$2D+1$ 枚の RGBA 画像）に変換した．学習時には，すべての視点に対して MPI から生成された画像と正解画像の間で誤差を求め，それらの平均二乗誤差をロス関数とした．

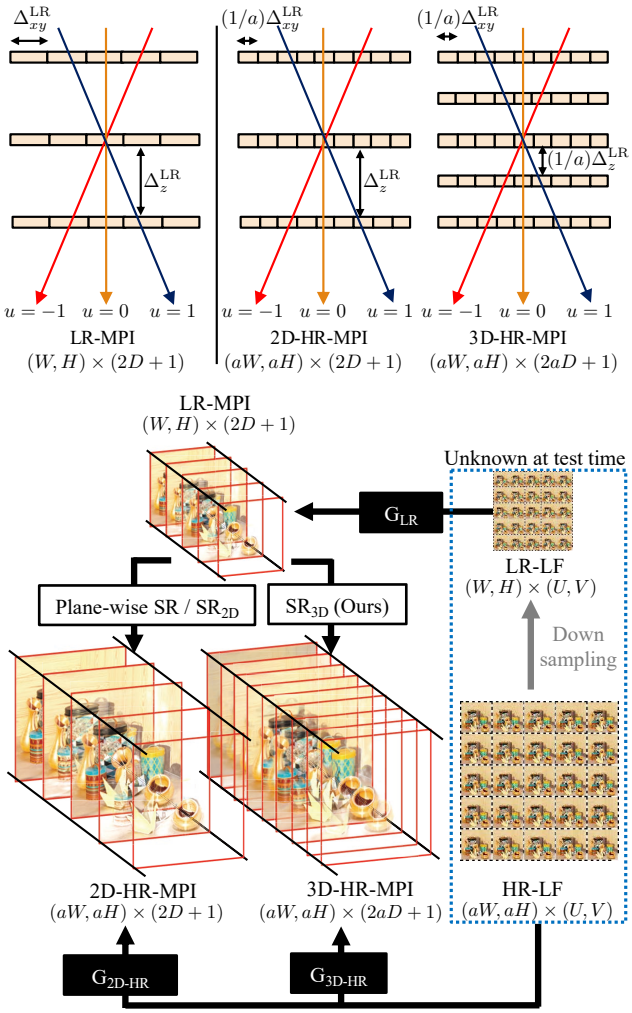

Fig.2: Geometric illustration of MPIs (top) and the framework of MPI super resolution (bottom).

本研究の目的は，MPI のための超解像手法を構築することである．研究の土台として，Fig. 2 に示す 3 つの CNN のインスタンス（G_{LR}, $G_{2D\text{-}HR}$, $G_{3D\text{-}HR}$）を用意して学習した．G_{LR} は，低解像度の光線空間（LR-LF：$W \times H$ 画素，$U \times V$ 視点）から低解像度の MPI（LR-MPI：$W \times H$ 画素，$2D+1$ 層）を生成する．この低解像度の MPI は，後述する提案手法の入力として用いる．一方，$G_{2D\text{-}HR}$ および $G_{3D\text{-}HR}$ は，高解像度の光線空間（HR-LF：$aW \times aH$ 画素，$U \times V$ 視点）から高解像度の MPI（HR-MPI：$aW \times aH$ 画素）を生成するが，$G_{2D\text{-}HR}$ では層の数が $2D+1$，$G_{3D\text{-}HR}$ では $2aD+1$ である．ここで $a > 1$ は高解像度化の倍率を表す．本稿では，空間解像度のみが高くなる場合には “2D”，空間解像度とともに奥行き方向の解像度も高くなる場合には “3D” を付して表記する．$G_{2D\text{-}HR}$ および $G_{3D\text{-}HR}$ はいずれも HR-LF を入力とするので，それぞれの構造の MPI において達成可能な性能の上限を知るための指標となる．次章で述べる MPI の超解像は LR-MPI を入力とするため，原理的には HR-LF を入力とする場合を上回ることはできないからである．

3 提案手法：MPI の超解像

3.1 高解像度 MPI の設計条件

低解像度の MPI（LR-MPI）は，$W \times H$ 画素，$2D+1$ 層で構成され，そこから $W \times H$ 画素，$U \times V$ 視点の多視点画像が生成されると仮定する．LR-MPI の画素間隔と層間隔をそれぞれ Δ_{xy}^{LR}, Δ_z^{LR} と表記す

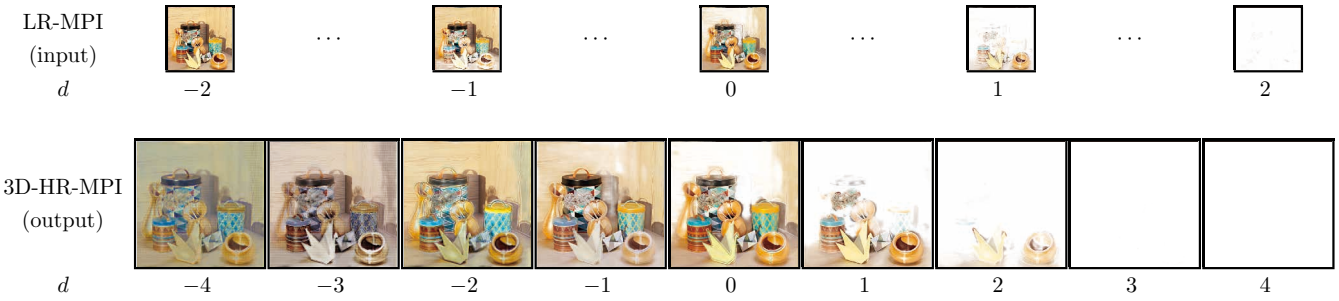

Fig.3: Example of input/output MPIs.

Table.2: Network architecture for $\mathrm{SR_{3D}}$. $\mathrm{Conv3_*}$ indicates 3-D convolutional layer with $3 \times 3 \times 3$ kernel. $\mathrm{Res_*}$ is residual connection. $\mathrm{U_{W,H}}$ and $\mathrm{U_D}$ indicate pixel shuffle operators that change shape of tensors from $(64, 2D, H, W)$ to $(16, 2D, 2H, 2W)$ and from $(16, 2D, 2H, 2W)$ to $(8, 4D, 2H, 2W)$, respectively. Elements presented in parentheses applies $\mathrm{SR_{2D}}$.

Layer	Input	$\mathrm{Ch_{in}/Ch_{out}}$	Activation
input	LR-MPI		
$\mathrm{Conv3_{in}}$	input	4/64	ReLU
$\mathrm{Conv3_{1a}}$	$\mathrm{Conv3_{in}}$	64/64	ReLU
$\mathrm{Conv3_{1b}}$	$\mathrm{Conv3_{1a}}$	64/64	
$\mathrm{Res_1}$	$\mathrm{Conv3_{in}} + \mathrm{Conv3_{1b}}$		ReLU
$\mathrm{U_{W,H}}$	$\mathrm{Res_1}$	64/16	
$\mathrm{Conv3_{2a}}$	$\mathrm{U_{W,H}}$	16/16	ReLU
$\mathrm{Conv3_{2b}}$	$\mathrm{Conv3_{2a}}$	16/16	
$\mathrm{Res_2}$	$\mathrm{U_{W,H}} + \mathrm{Conv3_{2b}}$		ReLU
$\mathrm{Conv3_{3a}}$	$\mathrm{Res_2}$	16/16	ReLU
$\mathrm{Conv3_{3b}}$	$\mathrm{Conv3_{3a}}$	16/16	
$\mathrm{Res_3}$	$\mathrm{Res_2} + \mathrm{Conv3_{3b}}$		ReLU
$\mathrm{U_D}$	$\mathrm{Res_3}$ (none)	16/8 (none)	
$\mathrm{Conv3_{4a}}$	$\mathrm{U_D(Res_3)}$	8/8 (16/16)	ReLU
$\mathrm{Conv3_{4b}}$	$\mathrm{Conv3_{4a}}$	8/8 (16/16)	
$\mathrm{Res_4}$	$\mathrm{U_D(Res_3)} + \mathrm{Conv_{4b}}$		ReLU
$\mathrm{Conv3_{out}}$	$\mathrm{Res_4}$	8/4 (16/4)	Hard Sigmoid
output	$\mathrm{Conv3_{out}}$		

る．本研究の目的は，LR-MPI を入力として，$aW \times aH$ 画素，$U \times V$ 視点の多視点画像を生成できるような，高解像度の MPI（HR-MPI）を得ることである．本稿では以後，$a = 2$ とするが，一般性を保つため引き続き a と表記する．

HR-MPI として，Fig. 2 に示す 2 つの構成を考える．2D-HR-MPI は，MPI を xy 次元のみ高解像度化したものであり，画素数 $aW \times aH$，層数 $2D+1$ となる．この場合，画素間隔および層間隔はそれぞれ $\Delta_{xy} = (1/a)\Delta_{xy}^{\mathrm{LR}}$，$\Delta_z = \Delta_z^{\mathrm{LR}}$ となる．一方，3D-HR-MPI は，MPI を xyz 次元全てにおいて高解像度化したものであり，画素数 $aW \times aH$，層数 $2aD+1$ となる．画素間隔および層間隔がそれぞれ，$\Delta_{xy} = (1/a)\Delta_{xy}^{\mathrm{LR}}$，$\Delta_z = (1/a)\Delta_z^{\mathrm{LR}}$ である．

本研究の提案は，HR-MPI の構造を 3D-HR-MPI（$aW \times aH$ 画素，$2aD+1$ 層）とすることである．これは，HR-MPI から生成される光線空間（多視点画像の集合で構成される信号空間）のアンチエイリアス条件 [10] から導出される．具体的には，生成される多視点画像の視点（角度）間隔を Δ_{uv} としたとき，HR-MPI における層間隔 Δ_z は，以下を満たす必要がある．

$$\Delta_z \leq \Delta_{xy}/\Delta_{uv} \tag{3}$$

まず，LR-MPI について，アンチエイリアス条件が満たされていること，すなわち，$\Delta_z^{\mathrm{LR}} \leq \Delta_{xy}^{\mathrm{LR}}/\Delta_{uv}$ を仮定する．HR-MPI においては，$\Delta_{xy} = (1/a)\Delta_{xy}^{\mathrm{LR}}$ となるが，視点間隔 Δ_{uv} は変化しない．3D-HR-MPI においては，$\Delta_z = (1/a)\Delta_z^{\mathrm{LR}}$ となるため，式 (3) のアンチエイリアス条件が必ず満たされる．一方，2D-HR-MPI においては，$\Delta_z = \Delta_z^{\mathrm{LR}}$ であるため，式 (3) を満たすことが保証できない．

3.2 MPI 超解像の実装

MPI の構造を 2D-HR-MPI とした場合には，各層を個別に高解像度化してもよい．具体的には，2 次元画像の超解像手法（例えば文献 [20]）を各層のカラー画像とアルファ画像に適用することで，目的を達成できる．一方，MPI の構造を 3D-HR-MPI とした場合には，画像や動画像を対象とした従来の 2 次元的な超解像技術をそのまま適用することができない．

そこで，本研究では，低解像度の MPI から高解像度の MPI を生成するための CNN を新たに実装した．具体的には，3D-HR-MPI を生成する CNN（$\mathrm{S_{3D}}$）と 2D-HR-MPI を生成する CNN（$\mathrm{S_{2D}}$）をそれぞれ構築した．$\mathrm{S_{3D}}$ のアーキテクチャを Table 2 に示す．ネットワークは 3D convolution と residual connection で構成され，pixel shuffle 操作により空間方向と奥行き方向の解像度を増加させる．$\mathrm{S_{2D}}$ においても，アーキテクチャも概ね同様であるが，奥行き方向の解像度が保たれる．$\mathrm{S_{2D}}$ では，アンチエイリアス条件は保証されないものの，層ごとの画像を個別に高解像度化する場合とは異なり，層をまたいだインタラクションが許容される．$\mathrm{S_{2D}}$ および $\mathrm{S_{3D}}$ のいずれにおいても，$\mathrm{G_{LR}}$ によって生成された LR-MPI を入力として与えると，CNN から HR-MPI が出力される．学習時には，その HR-MPI から生成された多視点画像と正解画像（HR-LF）との平均二乗誤差をロス関数として，CNN の重みパラメータを最適化した．

Figure 3 に，提案手法（$\mathrm{S_{3D}}$）の入出力の例を示す．入力の MPI は 248×248 画素，5 層であるが，出力は 496×296 画素，9 層であり，空間方向および奥行き方向において解像度が 2 倍になっている．

4 実験

実験の全体像を Fig. 2 を用いて整理する．データセットに含まれるオリジナルデータを高解像度の多視点画像（HR-LF）とみなし，文献 [20] と同様に，HR-LF を構成する各画像にバイキュービックダウンサンプリングを適用して低解像度の多視点画像（LR-LF）を生成した．多視点画像から MPI を生成する CNN として，$\mathrm{G_{LR}}$, $\mathrm{G_{2D\text{-}HR}}$, $\mathrm{G_{3D\text{-}HR}}$ を用意した．MPI の層数および超解像倍率は，それぞれ $D = 2$, $a = 2$ とした．すなわち，$\mathrm{G_{LR}}$ と $\mathrm{G_{2D\text{-}HR}}$ は 5 層，$\mathrm{G_{3D\text{-}HR}}$ は 9 層の MPI を生成する．MPI を超解像する CNN として，$\mathrm{SR_{3D}}$（LR-MPI から 3D-HR-MPI を生成），および $\mathrm{SR_{2D}}$（LR-MPI から 2D-HR-MPI を生成）を用意した．また，LR-MPI から 2D-HR-MPI を生成する別の

(a) G_{LR} (b) SR_{3D} (Ours) (c) HR-LF

Fig.4: Top left views rendered from (a) LR-MPI (G_{LR}) and (b) SR_{3D}, followed by (c) ground truth (HR-LF).

(a) G_{LR} (b) SR_{2D} (c) SR_{3D} (Ours)

Fig.5: Top left views rendered from (a) LR-MPI (G_{LR}), (b) SR_{2D} and (c) SR_{3D}.

Table.3: PSNR scores.

Method	Input/output	PSNR (dB)					
		INIRIA Lytro	HCI old	Stanford Gantry	EPFL	HCI new	ALL
Downsample	HR-LF/LR-LF	27.47	33.55	27.00	26.39	28.00	28.48
G_{LR}	LR-LF/LR-MPI	26.37	31.96	24.93	25.41	26.25	26.96
$G_{2D\text{-}HR}$	HR-LF/2D-HR-MPI	29.68	35.09	26.93	29.16	28.08	29.79
$G_{3D\text{-}HR}$	HR-LF/3D-HR-MPI	**30.92**	**39.72**	**31.14**	**30.54**	**31.59**	**32.78**
RCAN [20]	LR-MPI/2D-HR-MPI	28.69	34.07	26.35	27.79	27.50	28.88
SR_{2D}	LR-MPI/2D-HR-MPI	29.06	34.69	27.03	28.16	27.92	29.38
SR_{3D} (Ours)	LR-MPI/3D-HR-MPI	29.46	36.74	29.05	28.67	29.44	30.67

手段として，単一画像を超解像する SOTA 手法である RCAN [20] を MPI を構成する各画像に適用する手法も実装した．学習およびテストには BasicLFSR データセット[*1]に含まれる 5×5 視点の多視点画像を用いた．このデータセットは，5 つのデータセットを統合したものであり，学習用の多視点画像が 144 組，テスト用の多視点画像が 23 組含まれる．品質評価には，MPI から生成された多視点画像と正解の HR-LF との PSNR（全視点，全カラーにわたって平均二乗誤差を求め，デシベル表記にしたもの）を用いた．

Table 3 にそれぞれの手法により得られた PSNR の平均値を示す．ここで，LR-MPI および LR-LF については，正解の HR-LF とは空間解像度が異なるため、多視点画像を最近傍補間法でアップサンプリングして評価した．提案手法である SR_{3D} は，$G_{3D\text{-}HR}$（3D-HR-MPI の上限性能）には及ばないものの，入力である LR-MPI（G_{LR}）をはるかに上回る品質を達成できた．これは，MPI に対する超解像が期待通り動作したことを示している．また，SR_{3D} は，ほとんどの場合において，2D-HR-MPI を出力する手法（$G_{2D\text{-}HR}$，RCAN，SR_{2D}）を大きく上回った．これは，「MPI の超解像においては，xy 次元だけでなく，z 次元についても同時に高解像度化する必要がある」，という筆者らの主張をサポートする結果である．

結果の一部を Figs. 4，5 に可視化する．Figure. 4 では，左から，超解像前の MPI（LR-MPI）からの生成画像，提案手法（SR_{3D}）で超解像後の MPI からの生成画像，および正解画像（HR-LF）を示す．これらはいずれも 5×5 視点の左上視点から観察した画像である．これらの例では，オリジナルデータをダウンサンプリングした LF から LR-MPI を生成したため，高解像度の正解画像が存在する．提案手法 (SR_{3D}) では，細かい模様や文字などの高周波成分が正確に復元できていることがわかる．一方，Figure. 5 には，左から，超解像前の MPI（LR-MPI）からの生成画像，2 次元的に超解像（SR_{2D}）された MPI からの生成画像，および提案手法（SR_{3D}）による超解像後の MPI からの生成画像を示す．これらの例では，オリジナルの LF から生成した MPI を LR-MPI を見なしたため，高解像度の正解画像が存在しない．SR_{2D} では，特に物体の輪郭部分において二重像が目立つ．この原因は，MPI における奥行き方向の量子化が粗いことに帰着される．一方，SR_{3D} では，奥行き方向が同時に高解像度化された結果として，二重像が軽減されるとともに，より鮮明なテクスチャが復元されている．

最後に，計算時間に触れる．提案手法（SR_{3D}）によって，248×248 画素，5 層の MPI を 496×496 画素，9 層へと超解像した場合の所要時間は 39.97 msec であった．これは，496×496 画素，9 層の MPI を多視点画像から生成した場合（$G_{3D\text{-}HR}$）の所要時間（41.89 msec）と同程度であった．なお，この計測には，Intel Core i9-10900K および NVIDIA GeForce RTX 3090 を搭載した PC を用いた．

[*1] https://github.com/ZhengyuLiang24/BasicLFSR

5 まとめ

本研究では，高品質な自由視点画像生成を実現するため，MPI を直接的に超解像する手法を提案した．まず，MPI から生成される光線空間におけるアンチエイリアス条件を念頭に，3 次元的な超解像（MPI の空間方向と奥行き方向を同時に高解像度化する）が必要であることを示した．また，MPI に対して 3 次元的な超解像を実現する CNN を構築した．実験では，提案手法により高解像度で高品質な画像生成を実現できることを示した．今後は，異なる条件（層数など）において提案手法の評価を進めるとともに，他の手法 [7, 8, 9, 10, 11, 13] で生成された MPI を対象として，提案手法の有効性を検証したい．

参考文献

[1] Heung-Yeung Shum, Sing Bing Kang, and Shing-Chow Chan, “Survey of image-based representations and compression techniques,” *IEEE TCSVT*, vol. 13, no. 11, pp. 1020–1037, 2003.

[2] Cha Zhang and T. Chen, “A survey on image-based rendering - representation, sampling and compression,” *Signal Processing: Image Communication*, vol. 19, pp. 1–28, January 2004.

[3] John Flynn, Ivan Neulander, James Philbin, and Noah Snavely, “Deep stereo: Learning to predict new views from the world’s imagery,” *IEEE CVPR*, pp. 5515–5524, 2016.

[4] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi, “Learning-based view synthesis for light field cameras,” *ACM Trans. Graph.*, vol. 35, no. 6, nov 2016.

[5] Gaochang Wu, Mandan Zhao, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu, “Light field reconstruction using deep convolutional network on EPI,” *IEEE CVPR*, pp. 1638–1646, 2017.

[6] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow, “Deep blending for free-viewpoint image-based rendering,” *ACM Trans. Graphics*, vol. 37, no. 6, pp. 257:1–257:15, 2018.

[7] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely, “Stereo magnification: Learning view synthesis using multiplane images,” *ACM Trans. Graphics*, vol. 37, no. 4, pp. 65:1–65:12, 2018.

[8] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker, “Deepview: View synthesis with learned gradient descent,” *IEEE CVPR*, 2019.

[9] Pratul P. Srinivasan, Richard Tucker, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, and Noah Snavely, “Pushing the boundaries of view extrapolation with multiplane images,” *IEEE CVPR*, 2019.

[10] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Trans. Graphics*, vol. 38, no. 4, pp. 29:1–29:14, 2019.

[11] Richard Tucker and Noah Snavely, "Single-view view synthesis with multiplane images," *IEEE CVPR*, 2020.

[12] Suttisak Wizadwongsa, Pakkapon Phongthawee, Jiraphon Yenphraphai, and Supasorn Suwajanakorn, "Nex: Real-time view synthesis with neural basis expansion," *IEEE CVPR*, 2021.

[13] Yuemei Zhou, Gaochang Wu, Ying Fu, Kun Li, and Yebin Liu, "Cross-MPI: Cross-scale stereo for image super-resolution using multiplane images," *IEEE CVPR*, pp. 14842–14851, 2021.

[14] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger, "Differentiable volumetric rendering: Learning implicit 3D representations without 3D supervision," *IEEE CVPR*, 2020.

[15] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein, "Scene representation networks: Continuous 3D-structure-aware neural scene representations," *NeurIPS*, 2019.

[16] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman, "Multiview neural surface reconstruction by disentangling geometry and appearance," *NeurIPS*, vol. 33, 2020.

[17] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *ECCV*, 2020.

[18] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron, "NeRV: Neural reflectance and visibility fields for relighting and view synthesis," *IEEE CVPR*, 2021.

[19] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," *IEEE CVPR*, July 2017.

[20] Yulun Zhang, Kunpeng Li, Kai Li, Lichenand Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," *IEEE CVPR*, 2018.

[21] Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan, "Closed-loop matters: Dual regression networks for single image super-resolution," *IEEE CVPR*, 2020.

[22] Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren, and Ce Zhu, "Real-world single image super-resolution: A brief review," *Inf. Fusion*, vol. 79, pp. 124–145, 2021.

[23] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *IEEE CVPR*, pp. 1874–1883, jun 2016.

[24] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Real-time video super-resolution with spatio-temporal networks and motion compensation," *IEEE CVPR*, pp. 2848–2857, jul 2017.

[25] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu, "TDAN: Temporally-deformable alignment network for video super-resolution," *IEEE CVPR*, June 2020.

[26] Tom E Bishop, Sara Zanetti, and Paolo Favaro, "Light field superresolution," *IEEE ICCP*, pp. 1–9, 2009.

[27] Sven Wanner and Bastian Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE TPAMI*, vol. 36, no. 3, pp. 606–619, 2014.

[28] Zhen Cheng, Zhiwei Xiong, Chang Chen, and Dong Liu, "Light field super-resolution: A benchmark," *IEEE CVPR*, pp. 1804–1813, 2019.

[29] Shuo Zhang, Youfang Lin, and Hao Sheng, "Residual networks for light field image super-resolution," *IEEE CVPR*, pp. 11038–11047, 2019.

[30] Nan Meng, Hayden K.-H. So, Xing Sun, and Edmund Y. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," *IEEE TPAMI*, vol. 43, no. 3, pp. 873–886, Mar 2021.

[31] Zhen Cheng, Zhiwei Xiong, Chang Chen, Dong Liu, and Zheng-Jun Zha, "Light field super-resolution with zero-shot learning," *IEEE CVPR*, pp. 10010–10019, June 2021.