

自由視点画像生成のための Eigen-Texture 法における係数の回帰

中島 悠太^{1,a)} 大倉 史生¹ 河合 紀彦² 川崎 洋³ 池内 克史⁴ ブロンコ アンブロジーオ⁴

概要: 物体やシーンの任意の視点における見えを生成する自由視点画像生成の一手法である 3次元形状情報を利用したイメージベースレンダリングは、通常のカメラで撮影した対象の物体やシーンの画像のみから自由視点画像を生成できることから、一般ユーザでも容易に利用可能である。この手法では、大まかな形状を表す 3次元メッシュのそれぞれの面に対して撮影方向が最も近い画像を選択し、対応する領域をテクスチャとして貼り付けることで自由視点画像を生成する。このとき、表面の反射などの影響により撮影方向によって物体の色が変化する場合、選択された画像が切り替わる際に色の不連続な変化が生じる。本研究では、3次元メッシュのある面に貼り付けられるメッシュを固有空間中の点で表現する Eigen-Texture 法を利用し、さらに任意視点の位置から固有空間中の点を回帰することにより、任意の視点移動に対して滑らかにテクスチャが変化する手法を提案し、複数の画像データセットを利用してその有効性を示す。

キーワード: 自由視点画像生成, Eigen-Texture 法, 回帰, ニューラルネットワーク

Regression of Eigen-Texture Coefficients for Novel View Synthesis

YUTA NAKASHIMA^{1,a)} FUMIO OKURA¹ NORIHIKO KAWAI² HIROSHI KAWASAKI³ KATSUSHI IKEUCHI⁴
AMBROSIO BLONCO⁴

Abstract: Realtime novel view synthesis, which generates a novel view of a real object or scene in realtime, enjoys a wide range of applications. Image-based rendering (IBR) with rough geometry can be done using only an off-the-shelf camera and thus can be used by many users. However, IBR from images in the wild is a tough problem due to color discontinuity; IBR with rough geometry picks up appropriate images for a given viewpoint, but the image used for a rendering unit (a face or pixel) switches when the viewpoint moves, which may cause noticeable changes in color. We use the eigen-texture technique, which represents images for a certain face using a point in the eigenspace. We propose to regress a new point in this space using small neural networks so that we can generate an image whose color smoothly changes according to the point. We demonstrate the advantages of our IBR approach using our own datasets as well as publicly available datasets for comparison.

Keywords: Novel view synthesis, eigen-texture technique, regression, neural networks

1. はじめに

リアルタイムでの自由視点画像生成 (Novel View Synthesis; NVS) は、物体やシーンの任意の視点からの見えを

リアルタイムで再現する手法であり、拡張現実感技術やテレプレゼンス、没入型遠隔コミュニケーションなどをはじめとするさまざまな応用が考えられる [1], [2].

自由視点画像生成手法は、モデルベースレンダリング (Model-Based Rendering; MBR) とイメージベースレンダリング (Image-Based Rendering; IBR) に大別される。MBR は、対象となる物体・シーンの 3次元形状や表面の反射率などの特性をモデルとして与え、自由視点画像を生

¹ 大阪大学
² 奈良先端科学技術大学院大学
³ 九州大学
⁴ Microsoft Research Asia
a) n-yuta@ids.osaka-u.ac.jp



図 1 VDTM に基づく自由視点画像の歪みの例.

Fig. 1 Example visual artifacts.

成する。写実的な画像の生成には精緻な形状と特性が必要となり、一般ユーザによるモデルの作成は困難である。

一方 IBR では、対象物体（もしくはシーン）をその物体からカメラに入射する光線のみによって表現し、対象物体の 3 次元形状は利用しない。IBR の初期の手法として、View Morphing [3], [4] が挙げられる。この手法では、2 枚の画像に含まれる光線を滑らかに補間することにより、2 枚の画像の間の見えを持つ自由視点画像を生成する。Plenoptic Function [5] と Lumigraph [6], Light Fields [7] はそれぞれ任意の 3 次元位置・方向における光線をモデル化する。これらの手法は、空間中の既知の位置に固定された多数のカメラを利用して光線を取得していたが、画像からカメラ位置を推定することで、自由に動きながら撮影された画像からでも光線をモデル化可能とする拡張がなされた [8], [9]。しかし、いずれの手法にしても、高品質な自由視点画像を生成するためには、大量の画像と高精度なカメラパラメータが必要となる。

MBR と IBR は、自由視点画像生成手法の両端に相当するものであり、これらの手法の間にはモデルと画像の双方を利用するさまざまな手法が提案されている。これらのうち、Debevec らによって提案された視点依存テクスチャマッピング [10], [11] (View-Dependent Texture Mapping; VDTM) は、対象物体・シーンの大まかな 3 次元形状を利用することにより、IBR に比べて少数の画像で写実的な自由視点画像を生成可能な手法である。この手法では、対象物体表面の 3 次元位置と画像を撮影したカメラの 3 次元位置、および自由視点画像を生成する視点の 3 次元位置について、物体表面の 3 次元位置を挟んで構成する角が最も小さくなるカメラの画像を選択し、この画像中の対応する領域を対象物体表面のテクスチャとして貼り付ける。これにより、物体表面の微細な形状は、適切に選択された画像によって表現されることから、大まかな形状からでも写実的な画像の生成が可能となる。このような優位性から、VDTM にはさまざまな拡張がなされている [12], [13], [14], [15], [16], [17]。

VDTM には上記のように選択された画像が視点位置の移動により切り替わる際に色が不連続に変わるといった問題がある。図 1 の 3 つの画像は、わずかに視点に異なる視点位置に対して生成された自由視点画像であり、中央の画像では明確な色の境界が確認できる。これは、物体表面が鏡

面反射成分を有するなどの場合に加えて、一般ユーザが撮影した画像などにおいて露光時間が画像ごとに異なる場合などに特に顕著となる。ユーザの入力などに従って視点位置が連続的に変化するリアルタイム自由視点画像生成のアプリケーションでは、不連続な色の変化によりちらつきが生じ、ユーザの知覚する画質が著しく下がると考えられる。

不連続に色が変化することを低減するために、VDTM のオリジナルの手法 [11] では、角度が最も小さくなる 1 枚の画像ではなく、角度が小さい複数の画像を選択し、それらを角度の大きさに応じた重みでブレンディングする。この方法は、3 次元形状の精度が十分に高い場合には不連続な色の変化を低減可能である。しかし、ブレンディングの際に利用する重みについては、ある画像が別の画像に切り替わる前に 0 になる必要があるなど、設計が困難である。Buehler らの手法 [8] では、複雑な重みを利用することで不連続な色の変化を低減している。この他の手法では、Graph Cut アルゴリズム [16], [18] やポアソン合成 [19] を利用するが、特に Graph Cut アルゴリズムはリアルタイムの自由視点画像生成には計算時間の点で不適である。

本報告では、上記のようにオリジナルの VDTM には不適な画像群からでも滑らかな色の変化を可能にするリアルタイム自由視点画像生成手法について述べる [20]。提案手法は、Nishino らによって提案された Eigen-Texture [21] を利用する。Eigen-Texture は画像パッチ群を、その固有空間の部分空間中の点として低次元で表現するものであり、部分空間中の 2 点を補間することにより、Debevec ら [11] や Buehler ら [8] のようにブレンディングのための重みを設計することなく、2 枚の画像パッチを滑らかに遷移させることができる。本研究では、この手法を発展させ、自由視点画像を生成する視点位置から直接部分空間中の点を直接回帰することにより、画像を選択することなく自由視点画像を生成する手法を提案する。

IBR [6], [7] においては、ある 3 次元点を通過する光線はその角度の関数である。ここでは 3 次元形状の表現として 3 次元メッシュを用い、その各面に対する自由視点画像を生成する視点（以降、仮想視点）の相対位置を入力として、ひとつの隠れ層を持つニューラルネットにより部分空間中の点を回帰する。本報告では、3 つの画像データセットを利用して提案手法により自由視点画像を生成する。

2. 提案手法の概要

本研究で提案する自由視点画像生成手法の概要を図 2 に示す。提案手法は、オフライン部とオンライン部で構成される。オフライン部では、まず VisualSFM [22] や CMPMVS [23] などの手法を用いて撮影された画像群からカメラパラメータ、および 3 次元形状を推定する。 $S = \{I_n | n = 1, \dots, N\}$ と $\{(K_n, R_n, t_n) | n = 1, \dots, N\}$ をそれぞれ入力画像群と対応するカメラパラメータとし、 M を推定された 3 次元形状

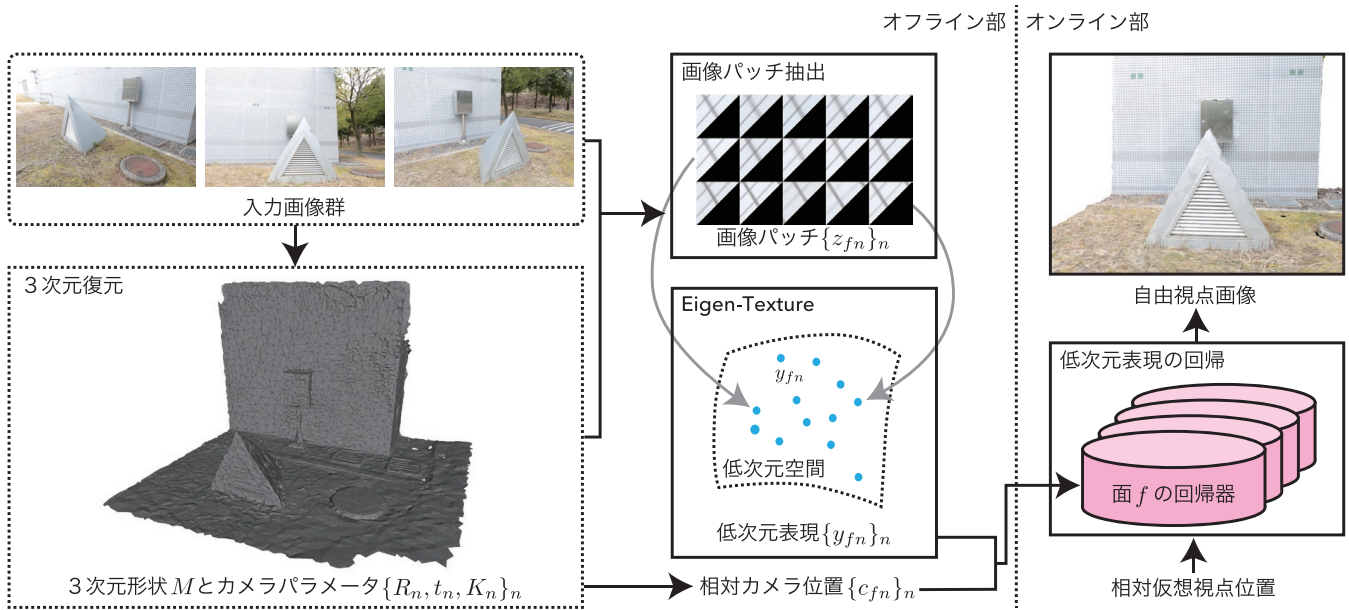


図 2 提案手法の概要.

Fig. 2 Overview of our system.

とする. $K_n \in \mathbb{R}^{3 \times 3}$ と $R_n \in \mathbb{R}^{3 \times 3}$, $t_n \in \mathbb{R}^3$ はカメラの内部パラメータ, 回転行列, 並進ベクトルである. さらに, M に含まれるそれぞれの面の各入力画像における可視性を判定する. 通常のレンダリングパイプラインにより, 入力画像 I_n を撮影したカメラ C_n のカメラパラメータを利用して 3次元形状 M のデプスマップ D_n を生成する. M の各面の I_n における可視性は D_n に基づいて判定する. 続いて, M の面 f についてその面が可視と判断された画像から図 3 のように画像パッチ $Z_f = \{z_{fn}\}$ (ただし, z_{fn} は I_n から抽出された f に対応する画像パッチ) を抽出し, Eigen-Texture 法による画像パッチの低次元表現 y_{fn} を算出する. 面 f について得られたすべての低次元表現を用いて, 面 f と仮想視点の相対位置から低次元表現 y を直接回帰するニューラルネットワークを学習する.

オンライン部では, 仮想視点の外部カメラパラメータ R と t , および内部パラメータ K に対して, 学習されたニューラルネットワーク面 f の Eigen-Texture 法における低次元表現を回帰により求め, 対応する画像パッチを再構成した上で, 面 f のテクスチャとして貼り付ける. オンライン部は, すべて GPU 上で実行可能である.

3. 画像パッチ抽出

本節では, 3次元形状 M とカメラパラメータから M の面 f に対応する画像パッチの抽出について述べる. 自己遮蔽などにより, 入力画像群 S に含まれる画像中で面 f が不可視となる場合があることから, Nakashima らの手法 [17] と同様の可視判定により, 面 f が可視となる画像のみから画像パッチを抽出する.

可視判定では, 通常のレンダリングパイプラインにより,

カメラ C_n のカメラパラメータ R_n , t_n , および K_n を用いて 3次元形状 M のデプスマップ D_n を生成する. このデプスマップにおいて, f を構成するすべての頂点が可視となる場合に, f は I_n で可視であると判定する. $v_{fi} \in \mathbb{R}^3$ を f を構成する頂点 $i \in \{0, 1, 2\}$ の世界座標系における 3次元位置とする. 次式により, この 3次元点をカメラ C_n 座標系に変換する.

$$v_{fi}^n = R_n v_{fi} + t_n \quad (1)$$

続いて, K_n を用いて v_{fi}^n を D_n に投影し, 対応するデプス値 d_{fi}^n を得る. この頂点が画像 I_n において可視の場合, d_{fi}^n は v_{fi}^n の第 3 成分 (デプス方向に対応する成分) に近い値となる. 本研究では, θ_{vis} をしきい値として, すべての頂点 $i \in \{0, 1, 2\}$ が次式を満たす場合に, v_{fi} は I_n において可視であると判定する.

$$|(0 \ 0 \ 1)v_{fi}^n - d_{fi}^n| < \theta_{\text{vis}} \quad (2)$$

θ_{vis} は M のスケールに依存して経験的に決定する.

次に, 面 f が可視と判定されたすべての画像から, 対応する画像パッチ z_{fn} を抽出し, 画像パッチ群 Z_f を構成する. 本研究では, 図 3 に示すように抽出された画像パッチを高さ L の直角二等辺三角形として記憶する. 画像パッチの抽出では, まず画像パッチ内の角画素に対応する面 f 中の 3次元点を求め, さらにその 3次元点を K_n によって I_n に投影することで, 当該の画素の色を取得する. α_1 と α_2 を, 画像パッチ内の画素 k の重心座標による表現とする. ただし, 画素 k の画像パッチ内での 2次元位置 $p'_k \in \mathbb{R}^2$ は, 直角二等辺三角形の等しい長さの 2 辺を表すベクトルを u'_1 と u'_2 とすると, 次式により与えられる.

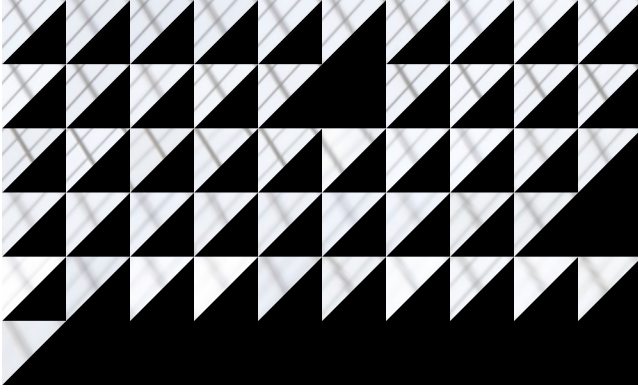


図 3 3次元形状のある面に対して抽出された画像パッチの例。
Fig. 3 Some examples of image patches for a certain face.

$$p'_k = \alpha_1 u'_1 + \alpha_2 u'_2 \quad (\alpha_1 + \alpha_2 \leq 1) \quad (3)$$

この点に対応する面 f 中の 3次元点は, u_1 と u_2 を u'_1 と u'_2 に対応する f の辺として, 次式で得られる.

$$p_k = \alpha_1 u_1 + \alpha_2 u_2 \quad (4)$$

得られた点 p_k を, C_n のカメラパラメータと式 (1) により I_n に投影することで, p'_k に対応する色を求める. 非整数の 2次元位置はバイリニア法により色を補間する.

図 3 に, 抽出された画像パッチの例を示す. 画像パッチ z_n はラスタスキャン順で n 番目のブロックに提示される. ブロックの左上三角形部分が黒いものは, 当該の画像パッチが I_n 内で不可視であることを示す. 画像パッチ内の線のパターンのずれや顕著なぼけが確認できる.

4. Eigen-texture

Eigen-Texture 法は, 画像パッチを固有値分解により同定される低次元空間で埋め込むものであり, Nishino らによって提案された [21]. 画像パッチ z_n をベクトル化したものを \tilde{z}_n , Z のすべての画像パッチで平均したものを \bar{z} とする (以降, 記述を簡単化するため面 f を表すサブスクリプトを省略する). 提案手法では, すべての画像パッチから平均値を減算し, 得られたベクトルを結合して行列 \tilde{Z} を構成する (\tilde{Z} の列ベクトルが減算結果 $\tilde{z}_n - \bar{z}$ に相当). $\tilde{Z}\tilde{Z}^\top$ は, 固有値分解により次式の通り分解できる.

$$\tilde{Z}\tilde{Z}^\top e_j = \lambda_j e_j \quad (5)$$

ただし, e_j と λ_j はそれぞれ j 番目の固有ベクトルと固有値である. 低次元固有空間は上位 J 本の固有ベクトルから張られる空間である.

画像パッチ z は次式により低次元空間表現に変換される.

$$y = E^\top(z - \bar{z}), \quad (6)$$

ただし, E は列ベクトルが上位 J 本の固有ベクトルからなる行列である. また, 低次元表現 y から z の再構成は次式

により与えられる.

$$z = Ey + \bar{z} \quad (7)$$

5. Eigen-Texture の低次元表現の回帰

式 (7) は, 低次元空間中の任意の点から画像パッチを生成可能であることを示す. 本研究では, 入力画像の撮影中は光源が固定されているものとし, 画像パッチは仮想視点と面の相対位置の関数であるものとする. この仮定の下では, 任意の仮想視点に対して低次元空間中の適切な点を回帰できれば, 適切な画像パッチを生成可能である. 本研究では, 仮想視点と面の相対位置を入力とした画像パッチの低次元表現の回帰を, 単一の隠れ層のみを持つ浅いニューラルネットにより実現する.

仮想視点と面の相対位置の表現として, 本研究では面 f に対して定義される座標系を用いる. 面 f の 2 辺に対応するベクトル u_1 と u_2 を用いて, まず f の法線ベクトル $a_0 \in \mathbb{R}^3$ を次式により求める.

$$a_0 = \frac{u_1 \times u_2}{\|u_1\| \|u_2\|} \quad (8)$$

ただし, 演算子 “ \times ” はクロス積を表す. 面 f の座標系は, a_0 , $a_1 = u_1/\|u_1\|$, および $a_2 = a_0 \times a_1$ を軸とし, f の頂点の重心位置 $o = \sum_i v_i/3$ を原点とする. この座標系において, 仮想視点 $c \in \mathbb{R}^3$ の位置は以下により与えられる.

$$c = A^\top(-R^\top t - o) \quad (9)$$

ただし, $A = (a_0 \ a_1 \ a_2)$ であり, R と t は世界座標系を仮想視点の座標系に変換する回転行列と並進ベクトルである.

提案手法で回帰に用いるニューラルネットは, 次式に示すとおり, 単一の隠れ層を持ち, 非線形性として双曲線正接を用いるものである.

$$\tilde{y}(c) = W_2 \tanh(W_1 c + b_1) + b_2 \quad (10)$$

ただし, $W_1 \in \mathbb{R}^{U \times 3}$, $W_2 \in \mathbb{R}^{J \times U}$ であり, U は隠れ層のユニット数を表す. 学習は M のそれぞれの面に対して独立に実施される. ロス関数は次式で定義される.

$$\ell(\{y_n\}, \{c_n\}) = \sum_n \|y_n - \tilde{y}(c_n)\|^2 \quad (11)$$

ただし, c_n はカメラ C_n の f に対する相対位置であり, 式 (9) で定義される. また, y_n は画像パッチ z_n の低次元空間での表現である. ニューラルネットは Weight Decay による正則化を含む確率的再急勾配法により学習される. c_n , y_{nf} はそれぞれの要素が分散 1 となるよう正規化する.

6. オンライン部

オンライン部では, 与えられた仮想視点 (回転行列 R , 並進ベクトル t , および内部パラメータ K) における対象

物体の見えをリアルタイムで描画する。まず、それぞれの面 f に対して、仮想視点の相対位置 c を式 (9) により算出し、式 (10) により対応する画像パッチの低次元表現 y を回帰する。このとき、学習時に y_{nf} に適用された正規化の逆変換を施すことにより、スケールが正しい低次元表現を復元する。この y を用いて、式 (7) により画像パッチ z を生成する。最後に、 z を面 f に対してテクスチャとして貼り付ける。提案手法では、回帰のために隠れ層 1 層のみからなる小さなニューラルネットを利用しているため、オンライン部はすべて GPU 上に実装可能である。

7. データセットと実装

本研究では Katagiri ら [18] の二つの画像データセット (Triangle Pyramid と Penguin) を利用して、提案手法を評価した。これらの画像データセットに含まれる画像から VisualSFM [22] と CMPMVS [23] を用いて対象物体、またはシーンの 3 次元形状をメッシュモデルとして復元した。また、得られたメッシュモデルに対して Quadratic Edge Collapse Decimation により面の数を 10,000 (Triangle Pyramid) と 5,000 (Penguin) に削減した。これら二つのデータセットに加えて、インターネット上で公開された Buddha データセット [24], [25] を利用した。Buddha データセットでは、メッシュモデルが合わせて提供されており、2,000,000 以上の面を持つことから、Triangle Pyramid データセットと同様にメッシュ数を 5,000 まで削減した。

図 4 に、低次元空間を構成する次元数 J が 1, 5, および 10 の場合の累積寄与率のヒストグラムを示す。これは、3 次元形状の各面の画像パッチが、異なる J でどの程度保存されているかを示すものであり、ヒストグラムが 1 付近に集中している場合、多くの面において高い精度で画像パッチを再構成可能であることを表す。このデータセットでは、 $J = 1$ の場合、ヒストグラムは分散しており、累積寄与率が低い面が多いことがわかる。 J が増加するにつれて、ヒストグラムは 1 付近に集中し、 $J = 10$ では各面における画像パッチのバリエーションに対応できていると考えられる。本研究では、十分な画質を確保するため $J = 10$ を用いるが、対象物体によってはより小さな J でも十分であろう。いくつかの面では、画像パッチ数が 10 に満たず、10 次元の低次元空間を構成することができなかった。この場合、 E を構成する列ベクトルの不足分を 0 で充填した。

ニューラルネットの隠れ層のユニット数は 10 とし、Weight Decay の重みは 0.1 に設定した。学習アルゴリズムとして、確率的最急勾配法の一つである Adam [26] を用いるが、画像パッチ群 Z_f に含まれる画像パッチ数が小さいため、ミニバッチに分割せずにすべての画像パッチから勾配を算出した。ここでは、ニューラルネットのパラメータの更新回数を 30,000 回とした (エポック数は 30,000)。

ベースライン手法のひとつとして、本報告では最も適し

た画像から得られた色を自由視点画像の対応する画素に適用する単純な VDTM を用いた。この手法では、生成された画像に空間的・時間的な色の変化が生じる。第 2 のベースライン手法として、テクスチャのブレンディングを行う VDTM を用いる。これはオリジナルの VDTM [11] に最も近いものである。各面に最も適した 3 つの画像を入力画像群から選出し、(i) 3 次元形状上の点から入力画像を撮影したカメラ C_n までと、(ii) 同じ点から仮想視点までのふたつのベクトルの間のコサイン類似度に基づいてブレンディングの際の各画像に対する重みを決定する。第 3 のベースライン手法は、第 2 のベースライン手法に加えて、時間方向でのブレンディングを利用するものである。 g_t を時刻 t に生成された自由視点画像のある画素の画素値であるとし、 g_{t-1} をその画素に対応する時刻 $t-1$ のブレンディング後の画素値とする。ただし、対応する画素は 3 次元形状に基づいて決定する。時刻 t におけるブレンディング後の画素値は次式により与えられる。

$$g_t' = \frac{g_t + g_{t-1}}{2} \quad (12)$$

8. 実験結果と比較

図 5, 6, および 7 は、移動する仮想視点において生成された自由視点画像例である。それぞれ左上から右下に、提案手法、平均画像パッチ、提案手法と平均画像パッチによる生成結果の差分画像、単純な VDTM、テクスチャのブレンディングを行う VDTM、時間方向のブレンディングを行う VDTM の結果を示す。提案手法による生成結果と合わせて、画像パッチの平均値 $\{\bar{z}_f\}_f$ をテクスチャとして貼り付けたもの、および提案手法と平均値との差分画像 (低次元表現により再構成された成分) を示す。

Triangle Pyramid データセットでは、後方の壁面に鏡面反射成分を持つ銀色の箱が取り付けられている。単純な VDTM では、画像が切り替わる際の表面の不連続な色の変化が顕著で、写実的ではないと感じられる。一方で、提案手法では不連続な色の変化は知覚できない。差分画像では、中央の三角形ピラミッド型のオブジェクトの表面の色が大きく変化することが確認できる。これは、当該オブジェクトの表面が弱い鏡面反射成分を持つためである。この鏡面反射成分は、ベースライン手法においても色の不連続な変化の原因となっている。

Penguin データセットで撮影されたぬいぐるみには、表面の起毛に起因する微細なテクスチャがある。提案手法は、低次元空間から再構成された画像パッチにもかかわらず、このテクスチャを再現している。単純な VDTM では、ペンギンの背面部分に画像の切り替わりによる境界が明確に見て取れる。この問題は、テクスチャのブレンディング、およびテクスチャと時間方向のブレンディングを利用したベースライン手法では見られないが、移動する仮想視点か

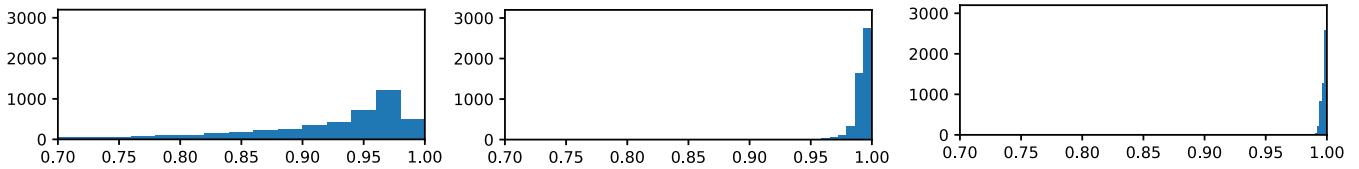


図 4 Penguin データセットにおいて、 J を 1 (左), 5 (中), 10 (右) とした場合の累積寄与率のヒストグラム.

Fig. 4 Histograms of contribution ratios for the penguin dataset for $J = 1$ (left), 5 (middle), and 10 (right).

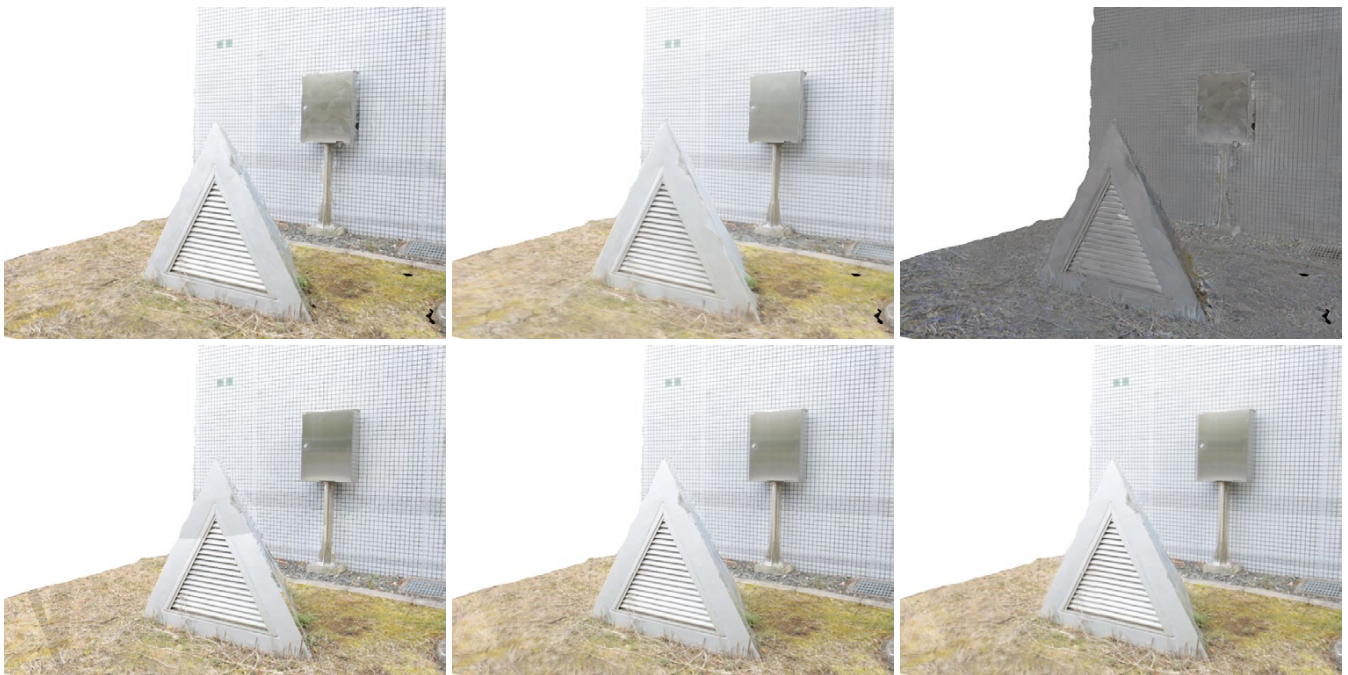


図 5 Triangle Pyramid データセットに対する自由視点画像生成結果.

Fig. 5 Novel views from the triangle pyramid dataset.

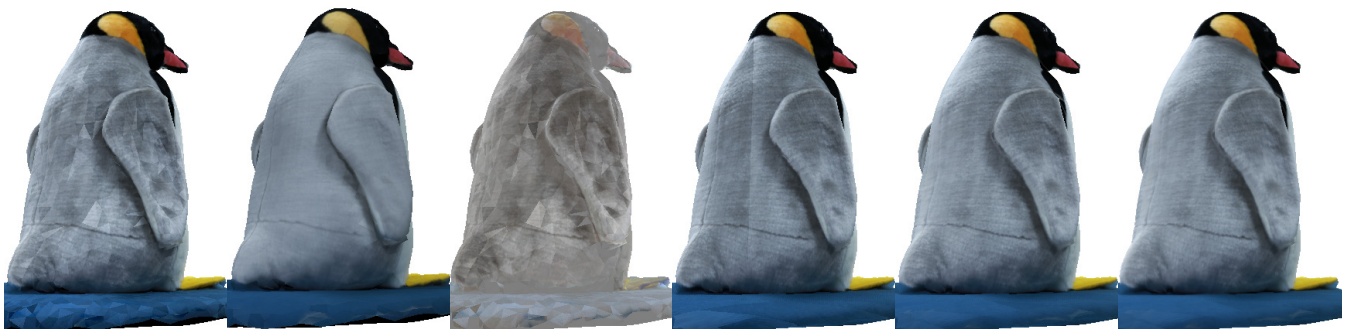


図 6 Penguin データセットに対する自由視点画像生成結果.

Fig. 6 Novel views from the penguin dataset.

らの画像を連結して生成した映像では色の急激な変化が確認できる。また提案手法では、起毛によって生じる非等方向性を持つ反射成分が保持されているように見えるが、画像の切り替え、またはブレンディングによるベースライン手法ではそのような効果は得られなかった。

Buddha データセットは、上記ふたつのデータセットと比較して大量の画像が含まれており、表面がほぼ拡散反射

成分のみを持つため、容易に高品質な生成結果が得られる。このデータセットでは、入力画像を撮影するカメラを動かすと同時に光源も移動した（もしくは、対象物体をターンテーブル上に設置して撮影した）ものと考えられ、上腕と胴体の付近の輝度がカメラの動きに従って変化する。図7の差分画像からも明らかなように、提案手法ではこの輝度値の変化を再現している。これは、光源位置がカメラ位置



図 7 Buddha データセットによる自由視点画像生成結果
Fig. 7 Novel views from the Buddha dataset.

の関数となっており、光源が入力画像の撮影時に変化する場合でも提案手法が対応可能な特殊な例である。VDTMに基づく手法では、当該部位で不連続な変化が確認できる。

図 8 は、回帰により得られた画像パッチの低次元表現 \tilde{y} の各要素を、物体周囲を円軌道に従って移動するカメラの動きに従ってプロットしたものを示す。この図から、Penguin データセットにおける画像パッチの変化の多くが、第 1 固有ベクトル (青線に対応) によって記述されることがわかる。これは、図 4 で $J=1$ の場合でも累積寄与率の分布が 1 付近に集中していることと一致する。また、 \tilde{y} の各要素が滑らかに変化することから、再構成される画像パッチも滑らかに変化する。図 9 は、図 8 の 110 番目付近で提案手法によって再構成されたある面の画像パッチと、ひとつ前の時刻において再構成された画像パッチとの差分を示す。提案手法では、次第に色が変化する一方で、VDTM に基づく手法では急激な色の変化があり、これが自由視点画像生成における不連続として知覚される。

フレームレート オンライン部は、NVIDIA Titan X Pascal (ビデオメモリ 12GB, コア数 3584) 上で CUDA により実装されており、Buddha, Penguin, および Triangle Pyramid のそれぞれのデータセットで、60fps (ディスプレイのリフレッシュレート), 56fps, 21fps で自由視点画像を生成した。計算時間の多くは回帰と画像パッチの再構成で占められていることから、計算時間は 3 次元形状を表現するメッシュモデルの面の数にほぼ比例するものと考えられる。

制限 提案手法では、特に入力画像を撮影したカメラ位置から離れた仮想視点において、画像パッチに対応する形状を見ることができる。これは、Eigen-Texture 法とニューラルネットワークによる回帰において隣接する画像パッチ間の空間的連続性を考慮していないためである。空間的制約を考慮したニューラルネットワークの学習は、ロス関数に新たな正則化項を導入するなどにより可能であると考えられるが、多数のパラメータを同時に最適化することになることからその実装は単純ではないと予想される。また、オンライン部においてポワソン合成 [19] などの援用により、フレームレートは下がるものの空間的連続性が保たれると考えられる。

補外 (近辺に入力画像が存在しない仮想視点位置での自由視点画像の生成) も、多くの IBR に基づく手法に共通の問題点である。提案手法では、例えばニューラルネットワークの学習の際に、入力画像が存在しない領域では平均画像パッチ (低次元表現としてはすべての要素が 0 の点) を出力するように、データ拡張を施すことにより全く異なる色の出力を制限し、画質としては向上すると期待できるものの、実際の見えに忠実な自由視点画像の生成は困難である。

9. まとめ

本報告では、Eigen-Texture 法に基づく自由視点画像生成手法を提案した。元の Eigen-Texture 法 [21] とは異なり、提案手法では任意の仮想視点における画像パッチをニューラルネットワークによる回帰を利用することで可能とした。実験では、金属表面の鏡面反射成分や、起毛による非等方性の反射成分、さらに視点位置に依存した光源による見えの変化を再現可能であることを示した。仮想視点の滑らかに移動する際の時間方向の連続性は、提案手法の特筆すべき利点であり、VDTM に基づく手法ではこの実現は困難であった。今後は、ニューラルネットワークによる回帰の学習の際に空間的連続性を担保するための項の導入や、ポワソン合成などを利用することにより、提案手法の現在の実装の問題である画像パッチの形状の顕在化を低減し、さらに動物体に対する提案手法の適用についても検討する。

謝辞 本研究はマイクロソフト・リサーチ CORE プロジェクトの補助による。

参考文献

- [1] Azuma, R.: A survey of augmented reality, *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4, pp. 355–385 (1997).
- [2] Tanimoto, M., Tehrani, M. P., Fujii, T. and Yendo, T.: Free-viewpoint TV, *IEEE Signal Processing Magazine*, Vol. 28, No. 1, pp. 67–76 (2011).
- [3] Seitz, S. M. and Dyer, C. R.: View morphing, *Proc. ACM SIGGRAPH*, pp. 21–30 (1996).
- [4] Xiao, J. and Shah, M.: From images to video: View morphing of three images, *Proc. Vision, Modeling and*

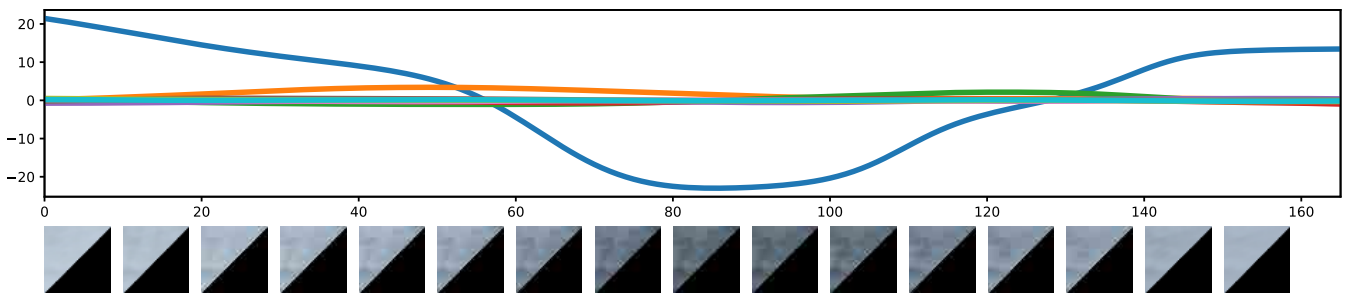


図 8 仮想視点移動時の \tilde{y} の変遷 (上) とそのとき生成される画像パッチの例 (下).

Fig. 8 Evolutions of \tilde{y} when viewpoint moves (top) and image patches (bottom).

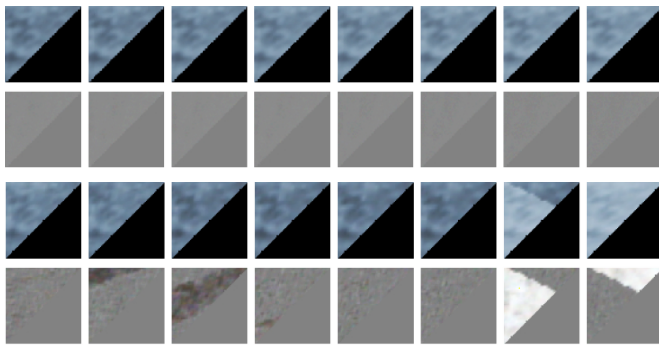


図 9 提案手法 (上図) と単純な VDTM (下図) における, 仮想視点移動する場合に生成された画像パッチの例 (上) と 1 時刻前の画像パッチとの差分.

Fig. 9 Image patches when viewpoint moves and the differences from the previous one for ours and simple VDTM.

Visualization (VMV), pp. 495–502 (2003).

[5] McMillan, L. and Bishop, G.: Plenoptic modeling: An image-based rendering system, *Proc. ACM SIGGRAPH*, pp. 39–46 (1995).

[6] Gortler, S. J., Grzeszczuk, R., Szeliski, R. and Cohen, M. F.: The lumigraph, *Proc. ACM SIGGRAPH*, pp. 43–54 (1996).

[7] Levoy, M. and Hanrahan, P.: Light field rendering, *Proc. ACM SIGGRAPH*, pp. 31–42 (1996).

[8] Buehler, C., Bosse, M., McMillan, L., Gortler, S. and Cohen, M.: Unstructured lumigraph rendering, *Proc. ACM SIGGRAPH*, pp. 425–432 (2001).

[9] Davis, A., Levoy, M. and Durand, F.: Unstructured light fields, *Computer Graphics Forum*, Vol. 31, No. 2, pp. 305–314 (2012).

[10] Debevec, P. E., Taylor, C. J. and Malik, J.: Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach, *Proc. ACM SIGGRAPH*, pp. 11–20 (1996).

[11] Debevec, P. E., Yu, Y. and Borshukov, G.: Efficient view-dependent image-based rendering with projective texture-mapping, *Proc. Eurographics Workshop on Rendering (EGWR)*, pp. 105–116 (1998).

[12] Porquet, D., Dischler, J.-M. and Ghazanfarpour, D.: Realtime high-quality view-dependent texture mapping using per-pixel visibility, *Proc. Int. Conf. Computer Graphics and Interactive Techniques (GRAPHITE)*, pp. 213–220 (2005).

[13] Chaurasia, G., Duchêne, S., Hornung, O. S. and Dretakis, G.: Depth synthesis and local warps for plausible

image-based navigation, *ACM Trans. Graphics*, Vol. 32, No. 3, pp. 30:1–30:12 (2013).

[14] Nobuhara, S., Ning, W. and Matsuyama, T.: A real-time view-dependent shape optimization for high quality free-viewpoint rendering of 3D video, *Proc. Int. Conf. 3D Vision (3DV)*, pp. 665–672 (2014).

[15] Zhou, Q.-Y. and Koltun, V.: Color map optimization for 3D reconstruction with consumer depth cameras, *ACM Trans. Graphics*, Vol. 33, No. 4, pp. 155:1–155:10 (2014).

[16] Kopf, J., Cohen, M. F. and Szeliski, R.: First-person hyper-lapse videos, *ACM Trans. Graphics*, Vol. 33, No. 4, pp. 78:1–78:10 (2014).

[17] Nakashima, Y., Uno, Y., Kawai, N., Sato, T. and Yokoya, N.: AR image generation using view-dependent geometry modification and texture mapping, *Virtual Reality*, Vol. 19, No. 2, pp. 83–94 (2015).

[18] Katagiri, K., Nakashima, Y., Sato, T. and Yokoya, N.: Novel view synthesis based on view-dependent texture mapping with geometry-aware color continuity, *Trans. Virtual Reality Society of Japan*, Vol. 21, No. 1, pp. 153–162 (2016).

[19] Pérez, P., Gangnet, M. and Blake, A.: Poisson image editing, *ACM Trans. Graphics (Proc. ACM SIGGRAPH)*, Vol. 22, No. 3, pp. 313–318 (2003).

[20] Nakashima, Y., Okura, F., Kawai, N., Kawasaki, H., Ikeuchi, K. and Blonco, A.: Realtime novel view synthesis with eigen-texture regression, *Proc. British Machine Vision Conference*, pp. 1–12 (2017).

[21] Nishino, K., Sato, Y. and Ikeuchi, K.: Eigen-texture method: Appearance compression and synthesis based on a 3D model, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 11, pp. 1257–1265 (2002).

[22] Wu, C.: Towards linear-time incremental structure from motion, *Proc. Int. Conf. 3D Vision (3DV)*, pp. 127–134 (2013).

[23] Jancosek, M. and Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 3121–3128 (2011).

[24] Technische Universität Darmstadt: *Image-Based Modeling and Rendering Benchmark* (2011, accessed May 1, 2017). <https://ibmr-benchmark.gcc.informatik.tu-darmstadt.de/>.

[25] Waechter, M., Beljan, M., Fuhrmann, S., Moehrle, N., Kopf, J. and Goesele, M.: Virtual Rephotography: Novel View Prediction Error for 3D Reconstruction, *ACM Trans. Graphics*, Vol. 36, No. 1, pp. 8:1–8:11 (2017).

[26] Kingma, D. and Ba, J.: Adam: A Method for Stochastic Optimization, *Proc. Int. Conf. Learning Representations (ICLR)* (15 pages, 2015).