

フレーム間のテクスチャの整合性を考慮した 視点依存テクスチャマッピング

寺崎 希^{1,a)} 河合 紀彦^{1,b)} 佐藤 智和^{1,c)} 横矢 直和^{1,d)}

概要：複数地点から撮影した画像群を入力とし、任意視点（仮想視点）からの見えを再現する自由視点画像生成は、遠隔地のシーンを仮想的に体感できる VR アプリケーションなどへの応用が期待されている。本研究では、従来の自由視点画像生成手法として定番となりつつある視点依存テクスチャマッピング法を用いて自由視点動画を生成する際の動画の品質向上を目的として、テクスチャ選択処理においてフレーム間のテクスチャの整合性を考慮し、かつ空間的・時間的な輝度の調整を行うことで、従来手法で生じていた自由視点動画におけるテクスチャのちらつきを低減する。実験では、小物体や屋外環境のシーンを対象として視点移動を伴う自由視点動画を生成することで、追加したテクスチャ選択条件の効果と輝度調整の効果を検証する。

キーワード：視点依存テクスチャマッピング、フレーム間のテクスチャの整合性、時間方向の輝度調整

1. はじめに

複数地点から撮影した画像群を入力とし、任意視点（仮想視点）からの見えを再現する自由視点画像生成は、実シーン仮想化のための基盤技術として注目されており、遠隔地のシーンを仮想的に体感できる VR アプリケーションなど [1], [2] への応用が期待されている。

自由視点画像生成の一手法として、比較的少ない入力画像から広範囲における写実的な自由視点画像を生成することが可能な、3次元形状を援用した IBR が提案されている。そのなかで、仮想視点位置に応じて形状にマッピングするテクスチャを適応的に変化させる視点依存テクスチャマッピング法 [3] が、効果的に写実性を高める手法として広く利用されている。この手法は、Structure-from-Motion(SfM)法 [4] および Multi-view stereo(MVS) 法 [5], [6], [7], [8], [9] 等により入力画像群からシーンの3次元形状を復元し、その3次元形状にマッピングするテクスチャとして自由視点画像の各画素に対して見え方が最も近い入力画像の画素を利用している。この手法を用いることで、3次元復元の精度が不十分な場合でも詳細な形状を視覚的に再現することが可能である。また、これを応用した自由視点画像生成手

法が多数提案されている [10], [11], [12], [13], [14]。例えば、Chaurasia ら [10] は、三次元形状を援用した IBR にとって問題となる三次元形状の欠損を尤もらしく補間したうえで、視点依存テクスチャマッピングを適用することでシーン全体で欠損のない自由視点画像生成を実現した。Okura ら [12] は、リアルタイムでのインタラクティブ性を重視した自由視点画像生成手法を提案した。具体的には、予め対象シーンにグリッドを設定し、その交点位置においてオフラインで自由視点画像生成を行い、オンライン処理においては、現在の仮想視点位置の近傍の交点位置で用いられた画像のみを用いて視点依存テクスチャマッピングを行うことで、高速な自由視点画像生成を実現した。一方、Katagiri ら [13] は、オフライン処理で高品質な自由視点画像を生成することを目的とし、見えの近さに加えて三次元形状の空間的な変化を考慮したテクスチャ選択手法を提案した。

ここで、従来のいずれの視点依存テクスチャマッピング法においても、1つの仮想視点から対象シーンを見たときの視覚的な違和感の低減を目的としており、視点移動を伴う動画生成においては各フレームの画像が独立に生成される。このため、入力画像取得時の照明条件の差異や3次元形状とカメラパラメータの誤差によって、フレーム間でマッピングされる入力画像が切り替わる際に、物体上の同一箇所のテクスチャが不連続に変化し、生成した動画がちらつく問題が生じる。

この問題に対し本研究では、視点移動を伴う自由視点動

¹ 奈良先端科学技術大学院大学 情報科学研究科

a) terasaki.nozomi.te2@is.naist.jp

b) norihi-k@is.naist.jp

c) tomoka-s@is.naist.jp

d) yokoya@is.naist.jp

画像の品質の向上を目的とし、フレーム間のテクスチャの整合性を考慮した自由視点動画生成手法を提案する。提案手法では、従来手法で用いられてきたテクスチャ選択の条件に加えて、前フレームで利用されたテクスチャと対象フレームで利用されるテクスチャとの色の差異をコストとして用いたエネルギー関数を新たに設計し、グラフカットを用いてこれを最小化することで、対象シーン内の同一箇所におけるフレーム間のテクスチャの差異が生じにくい入力画像を各画素に対して選択する。また、従来手法と同様に、異なる入力画像からのテクスチャの境界における空間的な輝度調整を各フレームで行うとともに、新たに複数フレームを用いた時間的な輝度調整を行うことで、空間的・時間的な輝度の不連続を抑制する。

2. フレーム間のテクスチャの整合性を考慮した視点依存テクスチャマッピング

図1に提案手法の処理の流れを示す。提案手法では、まず前処理として、対象シーンを複数の地点から撮影した画像群を入力し、SfM および MVS により、入力画像を取得したときのカメラの内部・外部パラメータ推定、対象シーンの3次元形状復元を行い、それらを利用して入力視点のデプスマップを作成する。次に、仮想カメラの動きを設定したうえで、各視点からの対象の見えを静止画像として1フレーム目から逐次的に生成する。最後に、それらをつなぎ合わせた動画画像を出力する。

各仮想視点における画像生成処理では、まずエネルギー関数の最小化により各画素に対する入力画像ラベルを選択する。次に、画素値の空間的・時間的な輝度調整を行う。以下では、 k フレーム目の自由視点画像生成について述べる。

2.1 エネルギー最小化によるテクスチャ選択

提案手法では、 k フレーム目の自由視点画像上の画素 i_k への値の割り当てに利用する入力画像 x_{i_k} を選択する問題を、エネルギー関数の最小化問題として定式化する。具体的には、以下のようにエネルギー関数を定義する。

$$J(X) = \sum_{i_k \in A_k} \alpha_{i_k}(x_{i_k}) \{f_{i_k}(x_{i_k}) + K_h h_{i_k}(x_{i_k})\} + \sum_{(i_k, j_k) \in B_k} (K_\beta \beta_{i_k j_k} + K_g) g_{i_k j_k}(x_{i_k}, x_{j_k}) \quad (1)$$

ただし、 $X = \{x_{i_k} | i_k = 1, \dots, N\}$ (N は入力画像枚数)、 A_k は k フレーム目における画素値を割り当てる全ての画素のインデックスの集合、 B_k は隣接する画素のインデックスのペアの集合、 K_h , K_β , K_g はそれぞれ重みパラメータを表す。式 (1) はデータ項と平滑化項で構成されるエネルギー関数であり、グラフカットアルゴリズム [15] で最小化することができる。以下、データ項中の $f_{i_k}(x_{i_k})$, $h_{i_k}(x_{i_k})$, $\alpha_{i_k}(x_{i_k})$, および平滑化項中の $g_{i_k j_k}(x_{i_k}, x_{j_k})$, $\beta_{i_k j_k}$ について述べる。

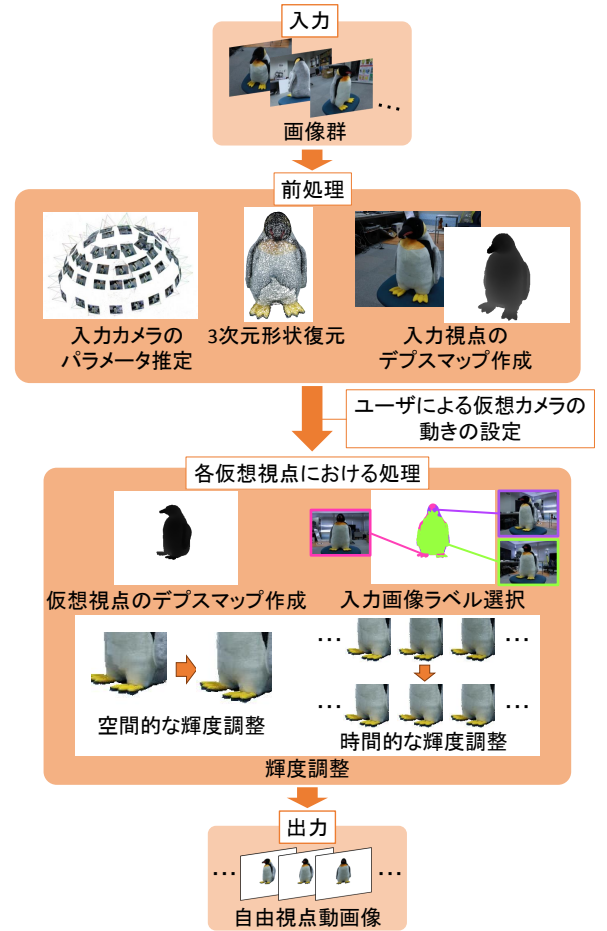


図1: 提案手法の概要

2.1.1 仮想視点と入力視点の光線方向の近さに関する項

Debevec ら [3] の手法に基づき、仮想視点における画素 i_k に対応する対象シーンの3次元形状上の点から、入力画像 x_{i_k} の視点位置および仮想視点位置への2つのベクトルが成す角 $\theta_{i_k x_{i_k}}$ を用いて $f_{i_k}(x_{i_k})$ を以下のように定義する。

$$f_{i_k}(x_{i_k}) = \theta_{i_k x_{i_k}} \quad (2)$$

この項により、仮想視点画像と入力画像を取得したときの光線の方向が近く、見え方が近いと考えられる入力視点の画像が選択されやすくなる。

2.1.2 フレーム間のテクスチャの整合性に関する項

提案手法では、3次元形状上の同一箇所のテクスチャがフレーム間で不連続に変化することを抑制するため、前フレームで選択されたテクスチャとの類似度を入力画像の選択基準として利用する。対象フレーム k と直前フレーム $k-1$ で3次元形状上の同一箇所を観測する画素 i_k と $D_{k-1}(i_k)$ を対応付け、これらの画素間の類似度をフレーム間のテクスチャの整合性に関する項 $h_{i_k}(x_{i_k})$ として次式で定義する。

$$h_{i_k}(x_{i_k}) = \|c_{i_k x_{i_k}} - c_{D_{k-1}(i_k) q(D_{k-1}(i_k))}\|_\infty \quad (3)$$

ここで、 $c_{i_k x_{i_k}}$ は k フレーム目の自由視点画像上の画

素 i_k に対応する入力画像 x_{i_k} 上の画素における画素値、 $\mathbf{c}_{D_{k-1}(i_k)q(D_{k-1}(i_k))}$ は前フレームの処理において、対応関係が既に決定済みの画素 $D_{k-1}(i_k)$ に対応する入力画像 $q(D_{k-1}(i_k))$ 上の画素における画素値を表す。 $\|\cdot\|_\infty$ は上限ノルムであり、ベクトルの要素の最大値を与える。この項は、3次元形状上の同一箇所の色が大きく異なる入力画像ラベルが選択される場合には値が大きくなるため、異なる入力画像への切り替えを許容しない。これによって、前フレームのテクスチャとの整合性を保持することが可能となり、動画像のちらつきが低減される。

2.1.3 可視判定に関する重み

注目画素 i_k に対応する三次元位置が、 x_{i_k} の入力視点から観測できない場合には、 x_{i_k} をテクスチャとして利用しないようにする。具体的には、画素 i_k に対応する3次元形状上の点 $\mathbf{p}_{i_k}$ を入力画像 x_{i_k} 上に投影し、投影画素における点 $\mathbf{p}_{i_k}$ から算出されるデプス値を $d'_{i_k x_{i_k}}$ とする。また当該画素における入力視点でのデプス値を $e_{x_{i_k}}$ とすると、可視判定に関する重みを次式により定義する。

$$\alpha_{i_k}(x_{i_k}) = \begin{cases} 1 & \text{if } |d'_{i_k x_{i_k}} - e_{x_{i_k}}| < \theta_d \\ K_\alpha & (\text{otherwise}) \end{cases} \quad (4)$$

ここで、 θ_d は誤差を許容するための閾値である。デプス値の差が大きい場合は、何らかの遮蔽物があり観測できないことを示すため、十分に大きな値 K_α をペナルティとして与える。

2.1.4 隣接画素値の整合性に関する平滑化項

この項では、隣接する画素それぞれにテクスチャとして選択される2つの入力画像について、画像上の同一箇所を投影する画素間で値の類似度を算出し、これを利用することで空間的な不連続を抑制する。テクスチャの境界の画素値の不整合に関する項を次式で定義する。

$$g_{i_k j_k}(x_{i_k}, x_{j_k}) = \max(\|\mathbf{c}_{i_k x_{i_k}} - \mathbf{c}_{i_k x_{j_k}}\|_\infty, \|\mathbf{c}_{j_k x_{i_k}} - \mathbf{c}_{j_k x_{j_k}}\|_\infty) \quad (5)$$

ここで、 $\mathbf{c}_{i_k x_{i_k}}$ は画素 i_k に対応する入力画像 x_{i_k} 上の画素における画素値を表す。この項は、隣接する画素に異なる入力画像ラベルが割り当てられ、それらの同一箇所を投影する画素間で画素値が大きく異なる場合にペナルティが大きくなる。

2.1.5 3次元形状の空間的な変化の大きさに関する重み

この項では Katagiri ら [13] のアイデアを利用する。3次元形状が空間的に大きく変化する領域においては、テクスチャが大きく切り替わったとしても視覚的な違和感は少ない。このため、この項により、テクスチャの切り替わりを許容する重みづけを行う。具体的には、重みを以下のように定義する。

$$\beta_{i_k j_k} = \begin{cases} 1 & \text{if } \cos^{-1} \frac{\mathbf{n}_{i_k} \cdot \mathbf{n}_{j_k}}{\|\mathbf{n}_{i_k}\| \|\mathbf{n}_{j_k}\|} < \theta_n \\ 0 & (\text{otherwise}) \end{cases} \quad (6)$$

ここで、 $\mathbf{n}_{i_k}$ 、 $\mathbf{n}_{j_k}$ はそれぞれ画素 i_k 、 j_k に割り当てられたメッシュモデルに基づき算出した法線である。隣接する画素に割り当てられた法線の成す角が閾値 θ_n よりも大きければ3次元形状の変化が大きいと判断し、式(5)の項の影響を弱める。

2.2 空間的・時間的な輝度調整

式(1)のエネルギー関数を最小化することで、各フレーム k 中の各画素 i_k に対応する入力画像ラベル $q(i_k)$ が求まり、この入力画像の対応画素を用いることで対象フレームの空間的・時間的な画素値の不連続が起こりにくくなる。しかし、入力画像取得時の照明条件の差異や3次元形状とカメラパラメータの誤差によって、空間的・時間的不連続を抑制しきれない領域が存在する可能性がある。この問題に対し、空間的・時間的な輝度調整を行うことで、画素値の不連続を解消する。

2.2.1 空間的な輝度調整

空間的な画素値の不連続を解消する方法として、同一画素に対応する複数画素値をブレンディングする方法がある。しかし、このようなブレンディング処理を行うと、3次元形状やカメラパラメータの誤差によって異なる入力画像上の点が対象シーンの同一箇所に対応しない場合、表面の構造やテクスチャが二重になるなどの問題が生じる。そのため本研究では、従来のブレンディングを行わない Katagiri らの空間的な輝度調整手法 [13] を利用する。この手法では、注目画素周辺の局所的な輝度平均と、注目画素周辺で各入力画像ラベルが割り当てられた画素の割合に基づいて注目画素の輝度を調整する割合(補正值)を求める。

具体的には、注目画素 i_k に対してウィンドウを設定し、そのウィンドウ内に含まれる画素に割り当てられた入力画像ラベルの集合を Q_{i_k} とし、入力画像ラベル $n \in Q_{i_k}$ について、ウィンドウ内で n が割り当てられた画素の割合を $\gamma_{i_k n}$ とする。また、ウィンドウ内の全ての画素の入力画像ラベルを n としたときのウィンドウ内の平均輝度を $\bar{I}_{i_k n}$ とする。注目画素 i_k のラベルを $q(i_k)$ とすると、空間的な輝度調整の補正值 $\mathbf{b}_{i_k}$ は次式で与えられる。

$$\mathbf{b}_{i_k} = \frac{1}{\bar{I}_{i_k q(i_k)}} \sum_{n \in Q_{i_k}} \gamma_{i_k n} \bar{I}_{i_k n} \quad (7)$$

2.2.2 時間的な輝度調整

本研究では、対象フレームおよびその直前の複数フレームの対応画素の空間的な輝度補正值および対応づいた入力画像ラベルを用いることで、時間的な輝度調整を行う。具体的には、以下の式によって、対象フレーム k の画素 i_k の最終的な値 $\mathbf{c}'_{i_k}$ を算出する。

$$\mathbf{c}'_{i_k} = \frac{1}{s} \sum_{l=k-s+1}^k \mathbf{b}_{D_l(i_k)} \mathbf{c}_{i_k q(D_l(i_k))} \quad (8)$$



図 2: 入力画像例

ただし、時間的な輝度調整に利用するフレーム数を s 、輝度調整に利用するフレームを $k - s + 1 \leq l \leq k$ とする。また、 $c_{i_k q}(D_l(i_k))$ は、 k フレーム目の画素 i_k に対応する 3 次元形状上の点を入力画像 $q(D_l(i_k))$ (k フレーム目の画素 i_k に対応した l フレーム目の画素に対応づいた入力画像) に投影した際の投影画素における輝度値である。これにより、3 次元形状上にマッピングするテクスチャの輝度値の差がフレーム間で小さくなるため、時間的な画素値の不連続が解消される。

3. 実験

提案手法の有効性を検証するため、図 2 に示す実空間の小物体と屋外環境のシーンを対象として実験を行った。実験ではまず、フレーム間のテクスチャの整合性に関するデータ項への重み K_h および、時間的な輝度調整に使用するフレーム数 s のパラメータに着目し、これらのパラメータの変化による結果への影響を検証した。次に、被験者実験によって提案手法の有効性を検証した。実験で用いた PC の仕様、および入力データの詳細をそれぞれ表 1、表 2 に示す。また、エネルギー関数の他のパラメータに関しては、表 3、表 4 に示すように経験的に設定した。

実験では、提案手法の前処理として、VisualSFM[16] を利用して各入力画像を取得した視点のカメラの内部・外部パラメータ推定を行い、CMPMVS[9] を利用して対象シーンの 3 次元形状復元を行った。続いて、仮想カメラの動きおよびそのフレーム数を図 3 および表 5 に示すように設定し、1 フレーム目から逐次的に自由視点画像を生成した。なお、各フレームの画像生成には、約 30 秒の時間を要した。

以下、パラメータの変化による結果への影響および、被験者実験について詳述する。

3.1 パラメータの変化による結果への影響

本実験では、フレーム間のテクスチャの整合性に関するデータ項への重み付けのパラメータ K_h および、時間的な

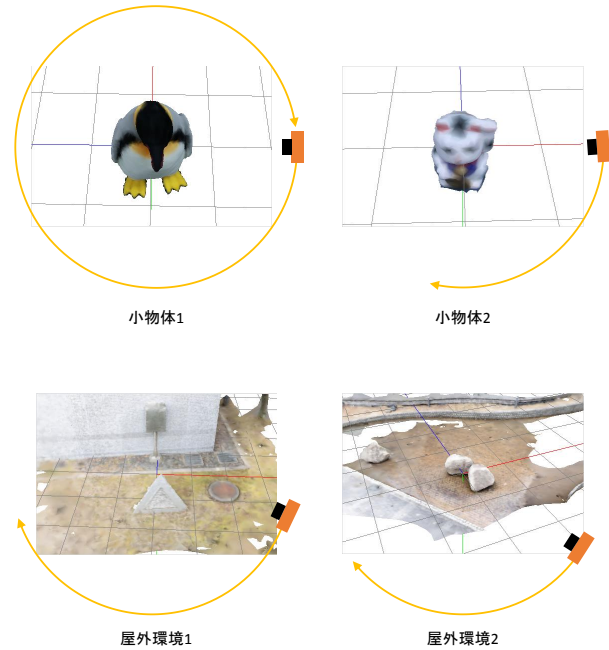


図 3: 各シーンにおける仮想視点の動き

輝度調整に使用するフレーム数 s をそれぞれ変化させながら生成した自由視点動画を比較する。なお、 $K_h = 0$ か

表 1: 実験で用いた PC の仕様

項目	詳細
CPU	Intel Corei7-4770 3.40GHz
メモリ	RAM 16.0GB
GPU	NVIDIA GeForce GTX TITAN Black
OS	Windows 8.1
グラフィック API	OpenGL

表 2: 各シーンの入力画像の詳細

対象	画像サイズ	画像枚数
小物体 1	1696 × 1280	61
小物体 2	1920 × 1280	52
屋外環境シーン 1	1920 × 1280	51
屋外環境シーン 2	1920 × 1280	35

表 3: 共通の固定パラメータ

パラメータ	値
K_β	400
K_g	200
K_α	20000
θ_n	5(deg)

表 4: シーン別の固定パラメータ

パラメータ	小物体 1	小物体 2	屋外環境 1	屋外環境 2
θ_d	0.1	0.1	0.5	0.5

表 5: シーン別の出力フレーム数

	小物体 1	小物体 2	屋外環境 1	屋外環境 2
フレーム数	360	67	141	96

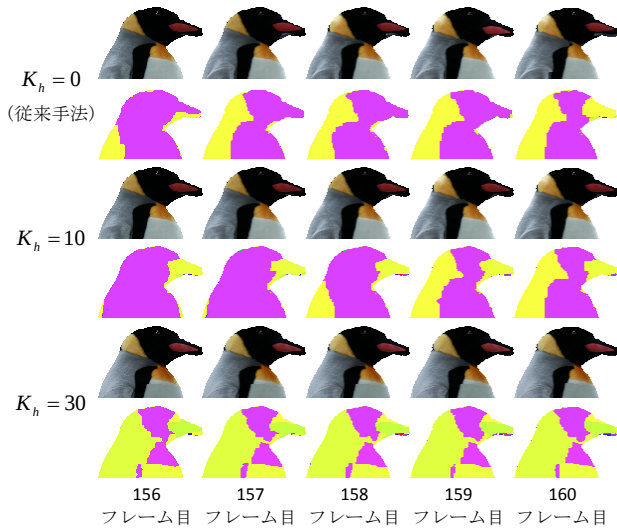


図 4: 小物体 1 における K_h の効果の比較

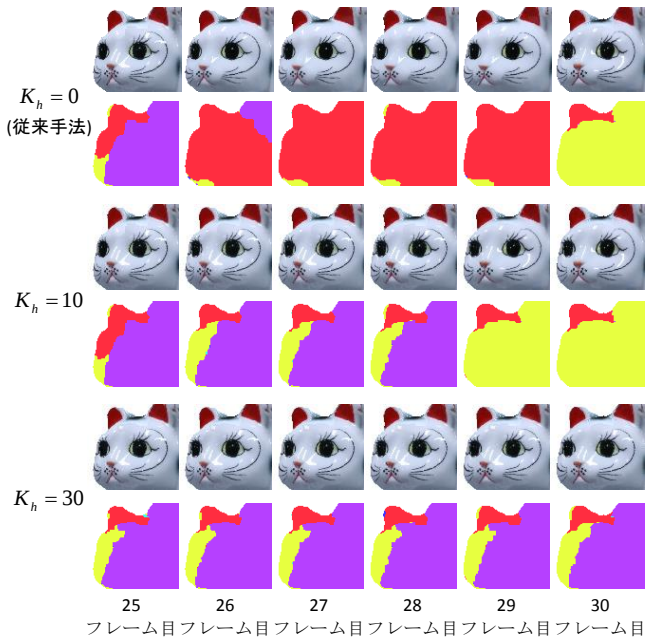


図 5: 小物体 2 における K_h の効果の比較

つ $s = 1$ の場合は, Katagiri らの手法を用いた場合の結果と同等のものとなる.

まず, フレーム間の整合性に関するデータ項 $h_{ik}(x_{ik})$ の効果を検証するために, 輝度調整に使用するフレーム数 s を 1 に固定し, $h_{ik}(x_{ik})$ への重みパラメータ K_h を 0, 10, 20, 30, 40 と変化させて生成した動画像を検証した. 各シーンの出力動画像の一部と選択されたテクスチャインデックスを可視化したものを図 4~7 に示す. なお, これらの画像は各フレームの一部を拡大したものである. また, 各パラメータの下段はテクスチャとして利用した入力画像の番号を可視化した画像である.

K_h を 0 に設定した場合, 各シーンのフレーム間で対象

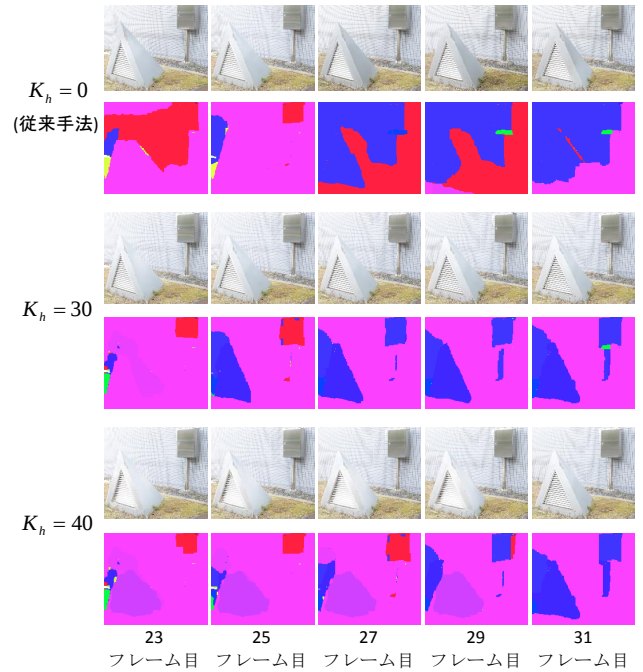


図 6: 屋外環境シーン 1 における K_h の効果の比較

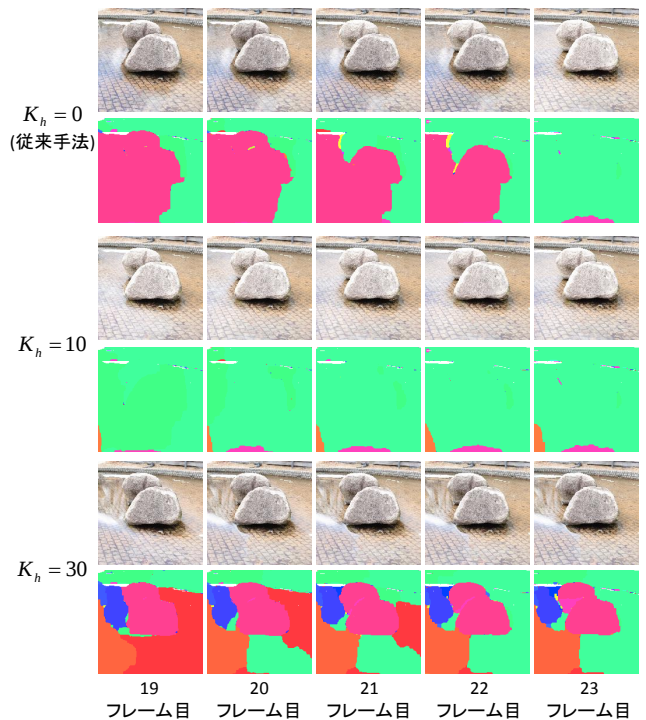


図 7: 屋外環境シーン 2 における K_h の効果の比較

上の同一箇所に関して, 選択されるテクスチャインデックスが頻繁に切り替わっていることが分かる. これにより, 例えば小物体 1 ではフレーム 159, 160 間において口ばしの赤い領域の位置ずれが生じている. また, 屋外環境シーン 1 では三角錐の物体の側面の色が, 屋外環境シーン 2 では石の色が急激に変化している. これらの現象によりテクスチャのちらつきが観測されることが分かる. このような時間的な不連続は, K_h の重みを大きくした場合に低減さ

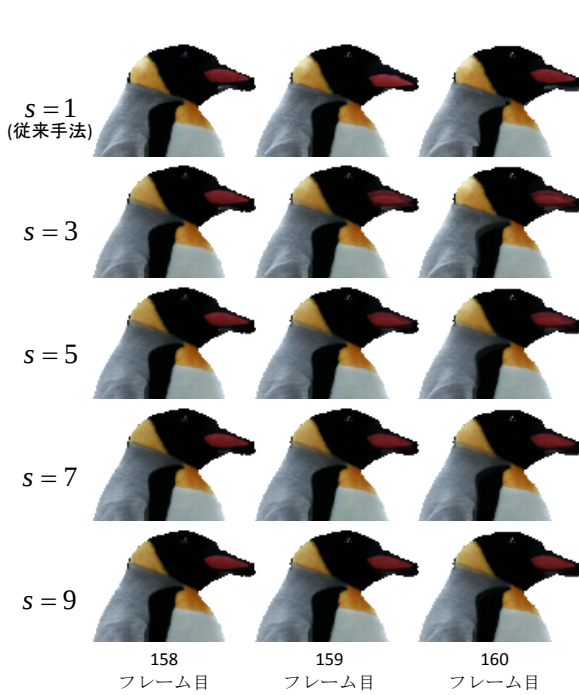


図 8: 小物体 1 における s の効果の比較

れていることを確認した。また、 K_h を 10, 20 に設定した場合は、テクスチャ切り替えの不連続を低減する効果が小さかったが、 K_h を 30 または 40 に設定した場合は、テクスチャの不自然な切り替わりがより低減されていた。しかし、小物体 1 や屋外環境シーン 1 において K_h を 40 に設定した場合には、頻度は少ないがテクスチャが急激に切り替わることがあり、テクスチャの不整合が発生したような印象を与えることを確認した。これは、 K_h の重みが大きく、一つのテクスチャが選択されたときに他のテクスチャに切り替わりにくくなったためと考えられる。また、小物体 2 に関しては、 K_h が 0 の場合にはなかった不連続が発生した箇所もあった。これは、カメラパラメータや 3 次元形状復元の誤差が大きく、それによってフレーム間の物体の同一箇所を投影する画素間で適切な対応関係が得られなかったためと考えられる。屋外環境シーン 2 では、図 7 に示すようにシーン中の岩に注目すると、テクスチャの時間的な不連続を抑制できていることが分かるが、このシーンは水面の鏡面反射が多いシーンであることや、このシーンを撮影する際に、同じく鏡面反射がある小物体 2 を撮影したときよりも光線情報を密に取得できなかったことから、入力画像間で鏡面反射光による見え方が大きく異なる領域が多く、フレーム間で鏡面反射光による見え方が視覚的に不自然に変化する場合があった。

次に、時間的な輝度調整に使用するフレーム数 s の効果を検証するために、 K_h を 0 に固定し、 s を 1, 3, 5, 7, 9 と変化させて生成した動画像を検証した結果について述べる。各シーンの出力動画像の一部を図 8~11 に示す。本実

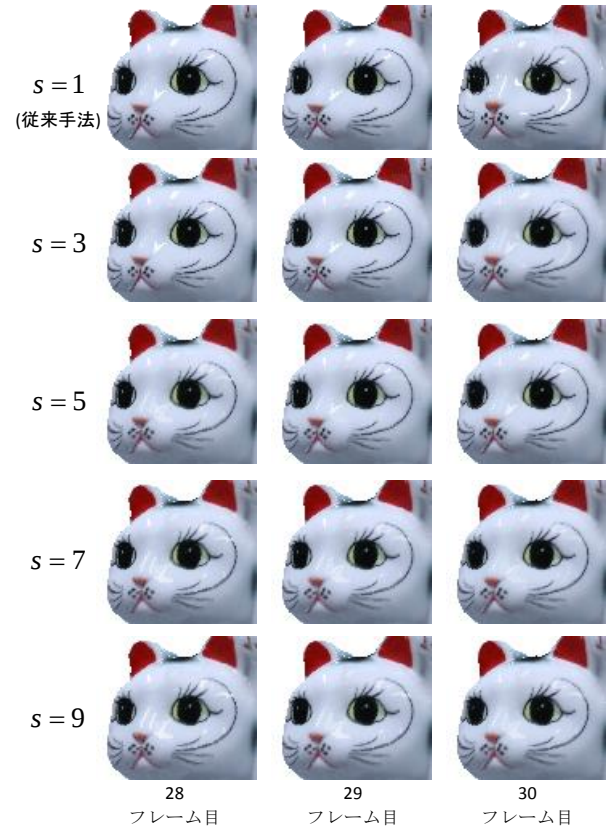


図 9: 小物体 2 における s の効果の比較

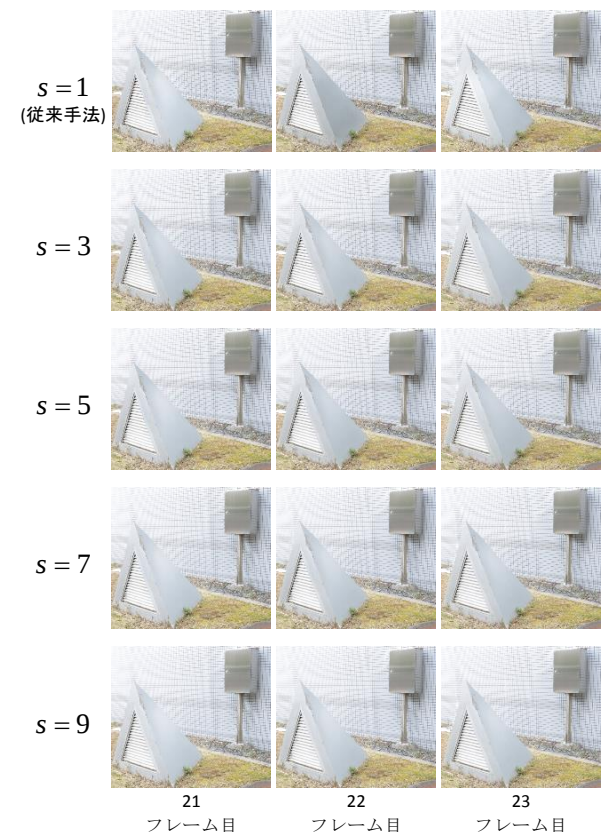


図 10: 屋外環境シーン 1 における s の効果の比較

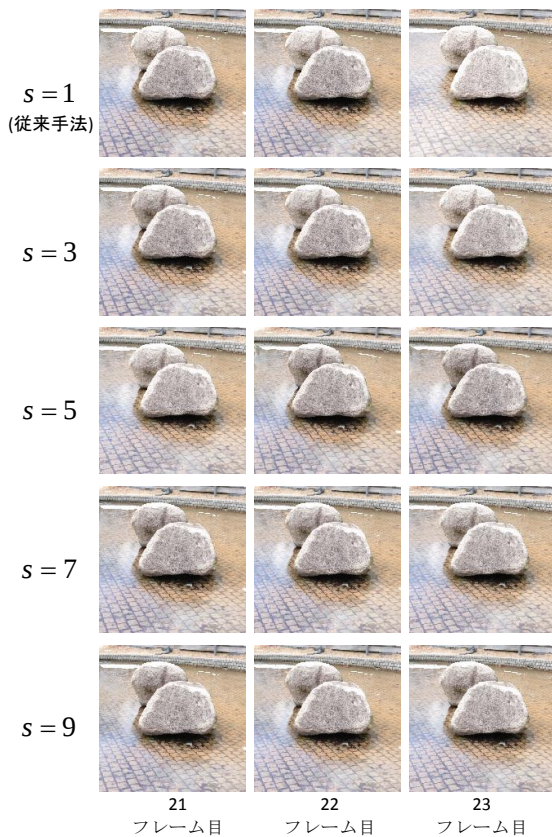


図 11: 屋外環境シーン 2 における s の効果の比較

験では、 s の値を 7 や 9 など十分に大きな値を設定した場合には、不連続を十分に低減できていることを確認した。例えば、 s が大きくなるにしたがって、小物体 1 の肩の色、小物体 2 の口元の色、屋外シーン 1 の三角錐三角錐の側面の色が滑らかに変化することを確認した。また、屋外環境シーン 2 の水面の領域に関しては、 $K_h = 0$, $s = 1$ の出力結果と $K_h = 0$, $s = 7$ の出力結果を比較すると、 $K_h = 0$, $s = 7$ のほうが水面の鏡面反射による見え方の変化を小さくできていることが確認できた。

3.2 被験者実験によるちらつき低減効果の確認

提案手法の有効性を検証するための被験者実験について述べる。本実験では、提案手法における各パラメータの設定に関する検証を行った。パラメータ設定に関する検証では、 K_h と s の一方の数値を固定し、もう一方の数値を変化させた動画像を評価データとして用いる。このパラメータの固定値は、事前実験に基づき経験的に決定した。具体的には、 $s = 9$ に固定し、 K_h を 0, 10, 20, 30, 40 に変化させた。また、小物体では $K_h = 30$ に、屋外環境では $K_h = 40$ に固定し、 s を 1, 3, 5, 7, 9 に変化させた。また、Katagiri らの手法 [13] との比較のために $s = 0$ かつ $K_h = 1$ の場合の動画像も比較対象に加えた。これら 10 通りの動画像をランダムに配置したものを被験者に同時に提示し、テクスチャちらつきの少なさと動画の品質について

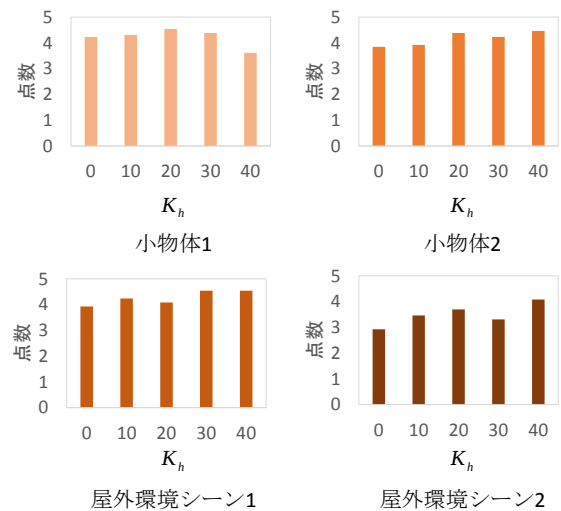


図 12: K_h の設定に関する被験者実験の結果 ($s = 9$)

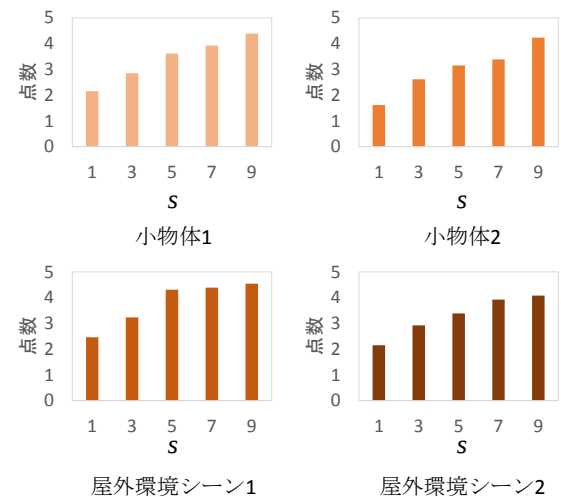


図 13: s の設定に関する被験者実験の結果 (小物体: $K_h = 30$, 屋外環境: $K_h = 40$)

最低値 1, 最高値 5 の 5 段階で評価してもらった。本実験には 20 代学生の男女 13 人の被験者が参加した。

図 12 に $s = 9$ に固定し K_h を変化させた場合の評価結果を示す。図より、ある程度大きな値を設定した場合の評価が高いことが確認できる。しかし、小物体 1 の結果に関して K_h が 40 のときに評価が低くなっていることが分かる。これは、前述のように、 K_h の重みが大きいことにより、一つのテクスチャが選択されたときに他のテクスチャに切り替わりにくくなり、テクスチャの切り替えが起こったときに不連続が生じたためであると考えられる。図 13 に K_h を 30 または 40 に固定し s を変化させた場合の評価結果を示す。図より、 s が大きくなるほどより高品質な動画像を生成できていることが分かる。

また、従来手法 ($K_h = 0$, $s = 1$)、整合性の項のみを考慮した場合 ($K_h = 30$ または 40 , $s = 1$)、輝度調整のみを行った場合 ($K_h = 0$, $s = 9$)、提案手法 ($K_h = 30$ または 40 , $s = 9$) の評価結果をまとめたグラフを図 14 に示す。

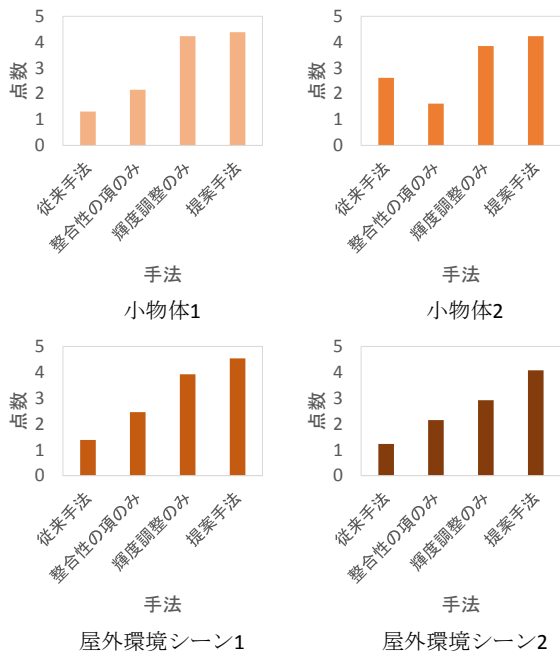


図 14: 従来手法も含めた被験者実験の結果

図より、ほとんどシーンにおいて、整合性の考慮および時間的な輝度調整がそれぞれ有効であり、また両方を合わせることでさらに高品質な自由視点動画像を生成できていることが分かる。この結果より、提案手法が有効であることを確認した。ただし、小物体 2 に関しては整合性のみを考慮する手法の結果は Katagiri らの手法で生成した結果よりも点数が低くなっており、この項単独では有効でない場合があることも確認した。これは、前述のように、カメラパラメータや 3 次元形状復元の誤差が大きく、それによってフレーム間の物体の同一箇所を投影する画素間で適切な対応関係が得られなかったためではないかと考えられる。また、このような対象では、低周波成分と高周波成分のテクスチャ両方を多く含んでいるため、テクスチャの位置ずれが視覚的に大きく影響したのではないかと考えられる。しかし、このようなフレーム間のテクスチャの不整合の影響は輝度調整処理によって十分に低減できている。

4. まとめ

本稿では、フレーム間のテクスチャの整合性を考慮したテクスチャ選択および、空間的・時間的な輝度調整により、時間的なテクスチャの不連続が生じにくい自由視点動画像生成手法を提案した。実験では、実空間の小物体や屋外環境のシーンを対象とし、設計したエネルギー関数と時間的な輝度調整のパラメータに様々な値を設定し生成された動画像を比較することで、提案手法の有効性を示した。しかし、各フレームの画像生成に時間を要することから、リアルタイム性が要求されるアプリケーションへの応用が困難である。そのため、処理コストの削減が今後の課題となる。

謝辞 本研究の一部は、科学研究費補助金 (No.26330193)

および総務省戦略的情報通信研究開発推進事業 (SCOPE)(No.162107002) の支援による。

参考文献

- [1] 大倉史生, 神原誠之, 横矢直和. 無人飛行船からの空撮全方位動画像を用いた蓄積再生型拡張テレプレゼンス. 日本バーチャルリアリティ学会論文誌, Vol. 16, No. 2, pp. 127–138, 2011.
- [2] M. Dellepiane, R. Marroquim, M. Callieri, P. Cignoni, and R. Scopigno. Flow-based local optimization for image-to-geometry projection. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 18, No. 3, pp. 463–474, 2012.
- [3] P. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. *Proc. Eurographics Rendering Workshop*, pp. 105–116, 1998.
- [4] S. M. Seitz and C. R. Dyer. View morphing. *Proc. ACM SIGGRAPH*, pp. 21–30, 1996.
- [5] K. Kolev, T. Brox, and D. Cremers. Fast joint estimation of silhouettes and dense 3D geometry from multiple images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 3, pp. 493–505, 2012.
- [6] S. M. Seitz N. Snavely and R. Szeliski. Modeling the world from internet photo collections. *Proc. ACM SIGGRAPH*, Vol. 80, No. 2, pp. 189–210, 2008.
- [7] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *Int. Journal of Computer Vision*, Vol. 38, No. 3, pp. 199–218, 2000.
- [8] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362–1376, 2009.
- [9] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly supported surfaces. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3121–3128, 2011.
- [10] G. Chaurasia, S. Duchene, O. S. Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. on Graphics*, Vol. 32, No. 3, pp. 30:1–30:12, 2013.
- [11] J. Kopf, M. F. Cohen, and R. Szeliski. First-person hyper-lapse videos. *ACM Trans. on Graphics*, Vol. 33, No. 4, pp. 78:1–78:10, 2014.
- [12] F. Okura, M. Kanbara, and N. Yokoya. Mixed-reality world exploration using image-based rendering. *ACM Journal on Computing and Cultural Heritage*, Vol. 8, No. 2, pp. 9:1–9:26, 2015.
- [13] K. Katagiri, Y. Nakashima, T. Sato, and N. Yokoya. Novel view synthesis based on view-dependent texture mapping with geometry-aware color continuity. *Trans. of the Virtual Reality Society of Japan*, Vol. 21, No. 1, pp. 153–162, 2016.
- [14] P. Hedman, T. Ritschel, G. Drettakis, and G. Brostow. Scalable inside-out image-based rendering. *ACM Trans. on Graphics*, Vol. 35, No. 6, pp. 231:1–231:11, 2016.
- [15] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 11, pp. 1222–1239, 2001.
- [16] C. Wu. Visualsfm: A visual structure from motion system. <http://www.cs.washington.edu/homes/ccwu/vsfm/>, 2011.