4D Gaussian Splattingによる博多祇園山笠の動的シーンレンダリング

松本伊織[1,a)]　張 ハンウェイ[2]　栗 達[1,b)]　川崎 洋[2,c)]　小野 晋太郎[1,d)]

概要：本研究では，博多祇園山笠の文化的価値を後世に伝えるための新たな記録手段を提供することを目的として，山笠の静的 3D モデルの生成と動的シーンの自由視点再現を行った．提案手法 (1) では，フォトグラメトリを利用して山笠の静的 3D モデルを生成し，正確な造形の保存に成功した．提案手法 (2) では，4D Gaussian Splatting（4DGS）を用いて，山笠の動きを記録し，自由視点での動的シーンを再現した．実験の結果，静的な 3D モデルの構築に成功し，自由視点での画像生成も達成された．しかし，動的シーンの再現においては，さらなる精度向上の余地があり，今後の課題として期待される．特に，背景の影響が大きいことが明らかとなり，これに対する対処が高精度な動的シーン再現に繋がると考えられる．

キーワード：自由視点画像生成，Gaussian Splatting，デジタルアーカイブ

Dynamic scene rendering of Hakata Gion Yamakasa by 4D Gaussian Splatting

Abstract: This study aims to preserve the cultural value of the Hakata Gion Yamakasa festival by generating a static 3D model using photogrammetry and reconstructing dynamic scenes with 4D Gaussian Splatting (4DGS). The static model was accurately preserved, and free-viewpoint image generation was achieved. However, there is room for improvement in dynamic scene reconstruction accuracy. Experimental results highlighted the significant influence of background interference, suggesting that addressing this issue will enhance reconstruction precision. Future work should focus on refining accuracy to improve the realism of dynamic scene reproduction.

Keywords: Free viewpoint image generation, Gaussian Splatting, Digital Archive

1. はじめに

博多祇園山笠は，福岡市で毎年 7 月に開催される伝統的な祭りであり，その中でも舁き山笠は祭りの象徴的な存在である．しかし，祭りが終わると舁き山笠はすぐに解体されてしまい，その姿や動きは映像や写真の記録に限られる．そのため，単なる造形としての保存だけでなく，山笠を実際に人が動かしている動的なシーンを含めた記録手法が求められている．

本研究の目的は，山笠の文化的価値を後世に伝えるための新たな記録手段を提供することである．そのため，初めに山笠の造形を詳細に保存することを目的として，撮影された複数の静止画像を用いたフォトグラメトリを活用し，静的な 3D モデルを構築する．次に，山笠の動きを含めた祭りの臨場感を記録するために，4D Gaussian Splatting（4DGS）[1] を用いて，時空間的に一貫した自由視点映像を生成し，動的なシーンの三次元再構築を行う．

2. 関連研究

文化財をはじめとした歴史的建造物のデジタルアーカ

[1] 福岡大学
Fukuoka University
[2] 九州大学
Kyusyu University
[a)] td232014@cis.fukuoka-u.ac.jp
[b)] lida@fukuoka-u.ac.jp
[c)] kawasaki@ait.kyushu-u.ac.jp
[d)] onoshin@fukuoka-u.ac.jp

イブ化については盛んに研究されている [2][3].「Digitally archiving cultural objects」[4] では，レーザーレンジセンサーから得られた距離画像などを用いてバイヨン寺院などの歴史的建造物の 3D モデル化を行っている．また，博多祇園山笠とよく比較される京都の祇園祭では，「祇園祭山鉾巡行」バーチャル体験システムを作成した研究 [5] がある．山鉾は設計図を元にモデルを作成し，撮影した画像をマッピングしている．担ぎ手については，モーションキャプチャを用いて動きを再現している．

紹介した研究では，センサーやモーションキャプチャーなどの機器を必要とするものであった．本研究では，特別な機材を必要としない複数枚の画像のみを用いた復元を試みる．

三次元表現に関する技術については，フォトグラメトリや NeRF[6]，3D Gaussian Splatting（3DGS）[7] などがある．

フォトグラメトリとは，複数枚撮影されたカメラ写真から 3DCG モデルを作成する一連のプロセスのことである．写真から特徴点の抽出，画像間の共通特徴点の認識，カメラの撮影位置の推定，三次元形状の復元まで行う．明示的な 3D メッシュや点群を直接生成できるため，3D モデルの編集に適している．

NeRF は，ニューラルネットワークを用いて 3D シーンの放射場を学習し，新しい視点の画像を高品質に合成する技術である．

GS は 3D 空間上にガウス分布を配置・最適化し，それをリアルタイムにレンダリングする手法である．NeRF はボリュームレンダリングを用いるため計算コストが高くリアルタイム性に欠けるが，GS はガウス分布の投影により高速なレンダリングが可能である．

本研究では，静的な 3D モデルの作成にフォトグラメトリを，動的シーンの三次元再現には GS を活用することとする．

3. 山笠の静的 3D モデルの作成

3.1 提案手法 (1)：静的 3D モデル作成の概要

複数枚の入力画像を元に (図 1) に示す手順で，静的シーンでの山笠の再現を試みる．フォトグラメトリで代表的なソフトウェアの 1 つである Metashape を活用し，再現する．

用いるデータは，2024 年に福岡市立博物館に展示されていた山笠を様々な視点から撮影した画像群，約 60 枚（図 2）である．

3.2 静的な山笠の再現結果

出力は，図 3 のようになる．静的なシーンであれば，3D モデルの作成が可能である．また，背景を除いて山笠のみの 3D モデルを作成した結果は図 4 のようになる．

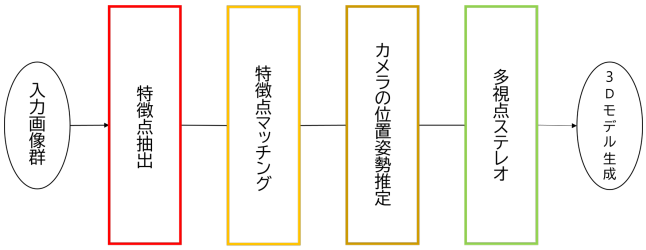


図 1 山笠の静的 3D モデルの作成手順

図 2 山笠を様々な視点から撮影した入力画像群

図 3 静的な山笠の再現

図 4 静的な山笠の再現 (山笠のみ)

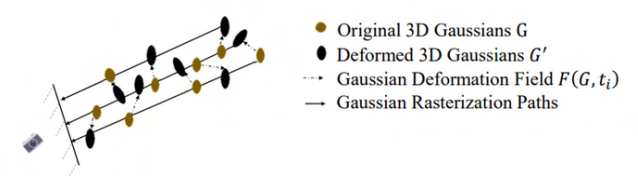


図 5 4DGaussian Splatting [1]

図 6 前景領域

図 7 背景領域

4. 山笠の動的シーンの再現

4.1 4DGaussian Splatting の概要

4D Gaussian Splatting（4DGS）は，動的なシーンを効率的に再構築するために設計された手法である．入力として，画像とカメラの位置姿勢推定の結果を活用し，静的な三次元シーンを再現する 3DGS の技術を動的シーンへと拡張したものである．GS であるために，従来のボリュームレンダリングに比べて計算効率が高く，軽量なレンダリングを実現している．

(1) 3DGS で，複数の視点画像とカメラの位置情報を元に，三次元空間を表現する「3D ガウス分布 G」を構築する．
(2) 4DGS は，3D ガウス分布 G と時刻 t を入力として受け取り，Gaussian 変形フィールドネットワーク F を使って時刻 t における変形 Δ G を予測する（図 5）．
(3) 時刻 t に対応するガウス分布 G′ を生成する．

4.2 動的シーンにおけるカメラの位置姿勢推定

4DGS では，カメラの位置姿勢情報が必要となる．カメラの位置姿勢は COLMAP を用いて推定する．しかし，COLMAP は静的シーンを対象に設計されており，動的シーンへの対応は困難である．

理由として，主に以下の 2 つが挙げられる．

1 つは，複数視点間で整合性が保てないことである．動体シーンでは，同じ物体や特徴点が異なる画像間で異なる位置に現れるため，正確な特徴点のマッチングが難しくなる．結果，カメラ位置推定に誤差が生じやすくなる．

2 つ目は，時系列情報へ非対応であるためである．動体シーンでは，物体の動きを時系列として扱う必要があるが，COLMAP では時系列的な動きの変化を捉える機能が基本的にはない．そのため，動きを含む場面では一貫性のある再構築が難しい．

したがって，動的シーンにおける COLMAP を用いたカメラの位置および姿勢の推定は，現状において困難である．特に，被写体の動きが小規模な場合には一定の推定精度が期待できる可能性があるものの，大規模な動的シーンに対しては適用が難しいと考えられる．

4.3 前景領域・背景領域の分割

COLMAP をはじめとする SfM は静的なシーンを前提としており，動的シーンでは位置姿勢推定の結果が不完全であったり，推定できなくなったりすることがある．

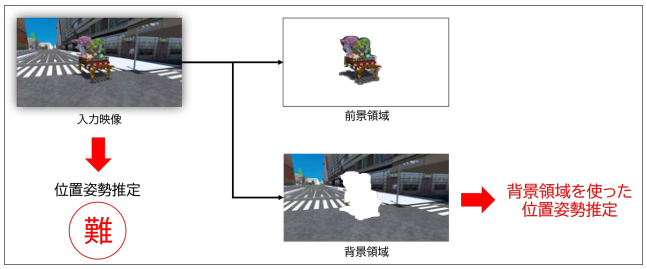


図 8 背景領域による推定

そこで，カメラ画像から動的な物体を「前景領域」，それ以外を「背景領域」として分ける．これにより，動体を含まない背景領域を COLMAP に与えれば，カメラの位置姿勢推定ができると考えられる（図 8）．

- 前景領域（図 6）：動体（人や動物，車など．本研究では山笠である．）
- 背景領域（図 7）：動体以外の静止した物体（建物や木など．人や動物でも静止していればこちらに分類される．）

4.4 深度情報とクラス情報を用いた分割

前景領域・背景領域に分割するための手法について考える．

4.4.1 STEGO

STEGO[8] とは，画像から深層学習モデルが抽出した「特徴点」を利用し，画像の各ピクセルや小領域を高次元特徴空間として表現する．これらを k-means，mean shift などのクラスタリングアルゴリズムで分類し，類似性に基づいてグループ分けを行うセマンティック・セグメンテーションの 1 つである．

セマンティック・セグメンテーションとは，画像内の各画素に対して人，建物，自動車といった事前に定義したラベル，カテゴリの中から，1 つを選択して関連付けるディープラーニングのアルゴリズムである．

4.4.2 MiDaS

MiDaS[9] は，単眼深度推定のモデルの一つである．絶対深度ではなく，相対深度の推定を行う．多様なデータセットで学習されており，屋内・屋外，昼夜問わず，あらゆるシーンで一貫した深度推定が可能で汎用性が高い．

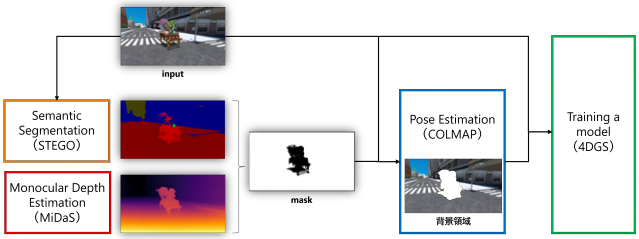


図 9　動的シーン再現までの流れ

実験	学習(入力)視点の数	出力視点の位置	出力視点の時刻	背景の有無	対象データ
①	8	学習視点と同一	学習視点の中間	あり	CG
		学習視点の中間	学習視点と同一		
②	2~8	同一	中間	あり	CG
		中間	同一		
③	2~8	同一	中間	なし	CG
		中間	同一		
④	1	同一	中間	あり	実映像
⑤	1	遠い	中間	あり	実映像

図 10　実験表

図 11　CG シーン

図 12　実映像シーン

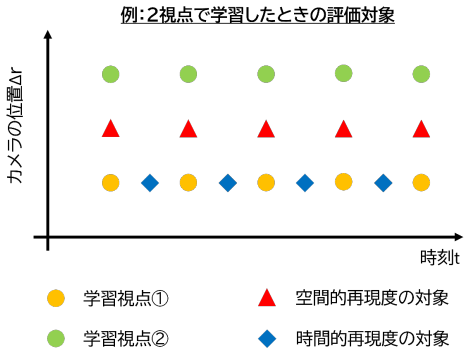


図 13　評価対象

4.4.3　分割処理の概要

＜前提条件＞ 動体が中心に映っていること．

(1) 1 フレーム目で以下を調整する．

- 中心の動体を囲むように長方形の枠を作成する．
- STEGO：枠内のクラスを上位 5 つまで列挙，動体の該当クラスの選択する．
- MiDaS：枠内のデプスの平均，中央値を取り，任意の範囲を決める．
- 任意の深度範囲にある該当クラスを前景領域の範囲とする．

(2) 1 で指定した範囲からすべてのフレームに対し，マスクを作成する．

(3) 作成したマスクから，前景領域と背景領域を分割する．

この手法は，動体（山笠）が中心に映っていることが前提条件ではあるが，カメラの移動がある場合にも使用できるため，こちらを使用する．

4.5　提案手法 (2):4DGS による動的な山笠の再現

まず，背景領域を使った COLMAP による位置姿勢推定を行う．その後，推定結果とカメラ画像を入力として，4DGS による動的シーンの再現を行う．（図 9）

5.　実験

動的シーンの再現にあたり，適切な入力条件を探すため，図 10 に示す 5 つの実験を行う．

5.1　実験条件

5.1.1　対象シーン

Unity で作成した CG シーンと実映像シーンの 2 つを使い，動いている山笠のレンダリング結果を評価する．

- CG シーン (図 11)
  Unity によって作成したシーン．入力視点の配置や数，背景の有無を変えることが出来る．
- 実映像シーン (図 12)
  実際に山笠を動かしているシーン．1 視点のみの映像．カメラの移動や山笠の移動だけでなく，多数の動的物体がある．

上記の 2 つのシーンを使い，出力視点を変えながら結果を確認する．

5.1.2　評価方法

空間的再現度と時間的再現度の 2 つの観点から評価を行う．空間的再現度は，NeRF や 3DGS における視点の補完への評価と同義である．時間的再現度とは，物体の動き（時間変化）への評価である．それぞれの評価対象（図 13）は，次のとおりである．

空間的再現度の評価対象は，複数のカメラの中間視点での出力である．入力にない視点を出力を確認する．学習（入力）カメラが 1 つの場合は，厳密に中間視点を取ることが難しいため，カメラの移動を xyz それぞれに 1m 程度平行移動した視点を評価の対象とする．また，比較のために学習視点での出力も同時に提示する．

時間的再現度への評価は，入力の画像枚数に対し，出力をその倍の枚数出力することで，時間的な補完画像を出力させ，これを評価の対象とする．

5.2　実験 (1)

実験 (1) では，4DGS に用いられていたデータを参考にカメラ配置として，8 視点を図 14 のように動体の正面に配置している．入力画像は図 15 のようになる．

図 14 8 視点のカメラ配置

図 15 8 視点から得られる画像の例

図 16 2〜8 視点のカメラ配置

5.3 実験 (2)

学習（入力）に使われるカメラ視点数を 2，3，4，8 と変化させたときの再現度を調査する．カメラ視点の配置は図 16 に示すとおりである．

- 8 視点
  4DGS で用いられたシーンでのカメラ配置を参考に，正面に 8 つのカメラを配置し，撮影する．
- 4 視点
  両側の沿道からの 2 視点とドローンや監視カメラなどの上からの視点を合わせた 4 視点を想定．
- 3 視点
  両側の沿道からの 2 視点とドローンや監視カメラなどの視点を合わせた 3 視点を想定．
- 2 視点
  両側の沿道からの 2 視点を想定．

5.4 実験 (3)

実験 (2) で用いた CG シーンを背景がないもの（図 17）に変えたときの再現度を調査する．

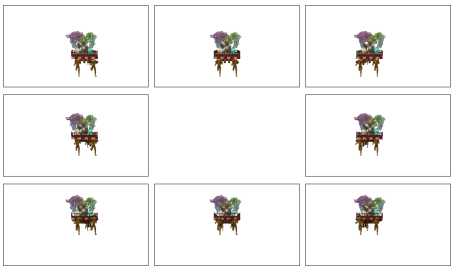


図 17 背景のないシーンで 8 視点から得られる画像の例

図 18 出力視点の移動における軸の向き

PSNR[dB]	8視点
空間的再現度（学習視点）	27.38
空間的再現度（中間視点）	18.40
時間的再現度	27.04

表 1 実験 (1): 結果

5.5 実験 (4)

実映像シーンでの再現度を調査する．

5.6 実験 (5)

実映像シーンでの出力視点を xyz それぞれに 1m 程度平行移動した場合 (図 18) の再現度を調査する．

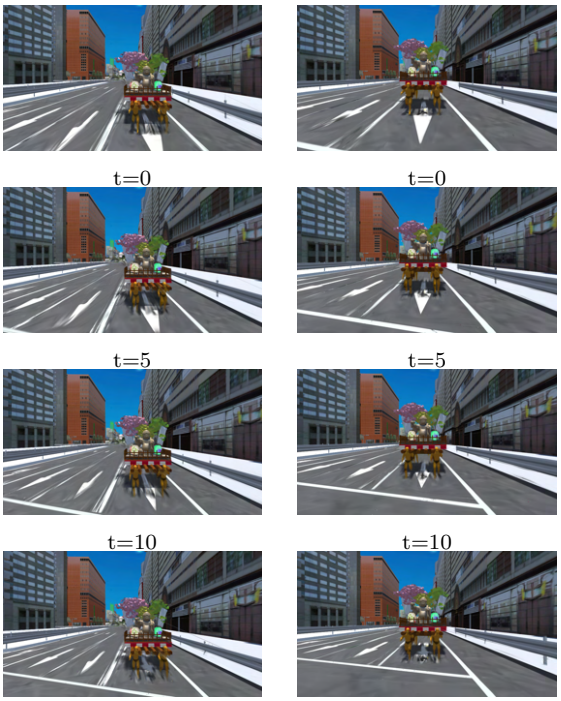
6. 実験結果

実行結果について，動的シーンの結果画像全てを載せるのは難しい．そのため，空間的再現度については一部のフレーム (t=0,5,10,15) を載せ，時間的再現度については，t=0.5 のフレームを例として載せる．また，PSNR による結果は，グラフ又は表を用いて示す．

6.1 実験 (1): 結果

4DGS に用いられていたデータを参考に配置した 8 視点の入力による再現では，表 1 のような結果となった．学習視点での PSNR で 27.38，中間視点での PSNR は 18.40 である．中間視点での値が低いが，図 19 では大きく劣化している様子は見られない．

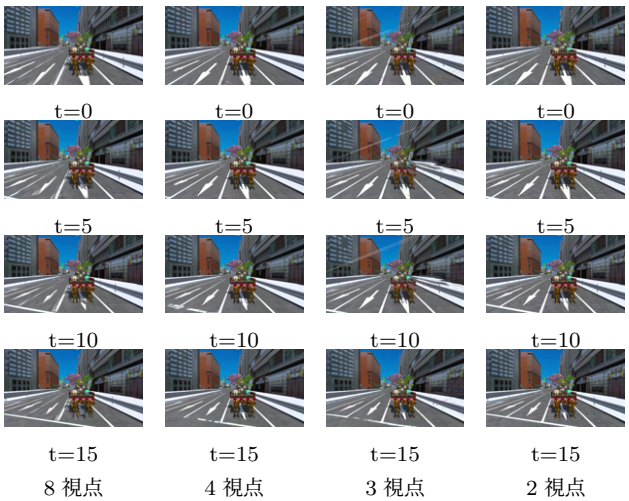
時間的再現度に関しては，PSNR の値が学習視点と比べて大幅に低下していないことから，動体の時間変化が適切に表現されていると考えられる．



(a) 学習視点　(b) 中間視点

図 19　8 視点の空間的再現度

図 20　8 視点の時間的再現度

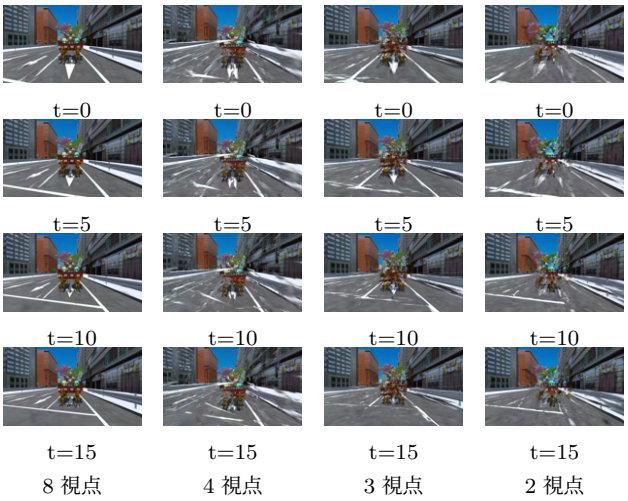


8 視点　4 視点　3 視点　2 視点

図 21　2〜8 視点の空間的再現度（学習視点）

6.2　実験 (2): 結果

実験 (2) の結果は図 24 のようになった．学習視点の数を変化させたときの出力を見ると，学習視点の数が多いほど学習視点での PSNR は低くなり，中間視点での PSNR は高くなる傾向があると言える．学習視点での PSNR は



8 視点　4 視点　3 視点　2 視点

図 22　2〜8 視点の空間的再現度（中間視点）

8 視点　4 視点　3 視点　2 視点

図 23　2〜8 視点の時間的再現度

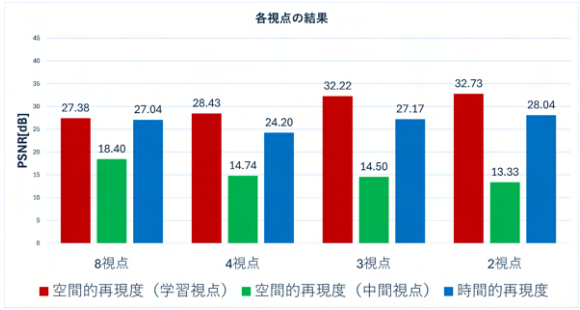


図 24　実験 (2): 結果

正解画像が増えることで，相対的に学習視点 1 つあたりの PSNR が下がっていると考えられる．中間視点では PSNR が増加していることから，視点数が多い方が自由視点表現という面での空間的再現度は高いと言える．

時間的再現度は，空間的再現度に依存している．また，視点数が多いほど学習視点での PSNR に対する時間的再現度の PSNR の劣化が抑えられている．

6.3　実験 (3): 結果

背景のないシーンでは，図 28 のような結果となった．背景ありのシーンであった図 24 と比較すると，学習視点の数が多いほど学習視点での PSNR は低くなり，中間視点での PSNR は高くなる傾向は実験 (2) と変わらない．しかし，背景がないシーンでの結果の方が背景があるシーンよりも PSNR が 5 近く高いことが分かった．

図 25 背景のないシーンでの 2〜8 視点の空間的再現度（学習視点）

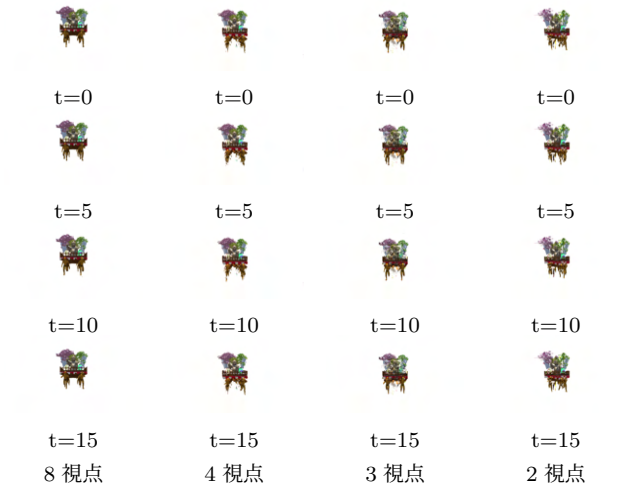


図 26 背景のないシーンでの 2〜8 視点の空間的再現度（中間視点）

図 27 背景のないシーンでの 2〜8 視点の時間的再現度

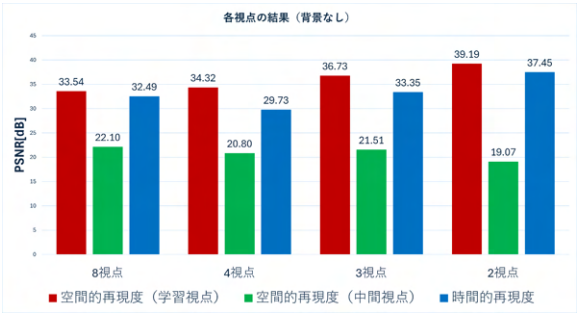


図 28 実験 (3): 結果

6.4 実験 (4): 結果

実映像での実行結果は，図 29 のようになった．表 2 を見ると，PSNR としては 21.73 である．しかし，PSNR の値以上に実映像シーンでの結果画像は全体的にぼんやりとしており，不鮮明な印象を受ける．視点数が 1 であることや CG シーンと異なり，山笠以外の多くの動的物体がシーンに含まれていることが原因として考えられる．

図 29 実映像シーンでの実行結果

PSNR[dB]	実映像
空間的再現度（学習視点）	21.73
時間的再現度	21.26

表 2 実験 (4): 結果

図 30 学習視点から平行移動した視点での出力画像

6.5 実験 (5): 結果

実映像シーンは 1 視点のみでの学習であったため，中間視点ではなく，学習視点から xyz にそれぞれ 1m 程度の移動を加えた視点を出力した．結果は図 30 のようになった．視点数が 1 であることや山笠以外の多くの動的物体がシーンに含まれていることに加え，視点移動が大きいため品質の劣化が確認された．

7. 考察

時間的再現度は，正面の 8 視点から得られた結果画像だけでなく，他の視点であっても時間的変化を捉えていた．時間的再現度は空間的再現度に依存しており，時間的再現

度の向上には，空間的再現度の向上が必要不可欠と言える．

8 視点から得られた結果では，比較的高い精度で視点の再現が可能であった．しかし，現実的なカメラ配置数と考えられる 4 視点，3 視点，2 視点では，情報量の不足により空間の再現が困難となり，中間視点において顕著な劣化が確認された．特に，今回使用した山笠のシーンにおける空間的再現度については，4DGS の splatting による細部描写精度の低下と背景の影響により，再現度が十分に確保されないことが分かった．

8 視点を用いた結果画像では，中間視点を比較的高い精度で再現できていることから，さらにカメラの台数を増加させることで再現精度を向上させることは可能と考えられる．しかし，その場合，撮影フレーム数の増加に伴い，カメラの位置・姿勢推定やトレーニングに要する計算負荷が著しく増大するため，実運用におけるコストが課題となる．したがって，空間的再現度の向上には，シーンに応じた適切なカメラ台数と配置の最適化が不可欠である．

さらに，同じ視点数であっても，背景のないシーンでは再現度が高い一方で，背景を含むシーンでは動体再現の難易度が上昇することが明らかとなった．特に，背景の有無が再現の精度に及ぼす影響は大きく，背景処理の改善によって大幅な精度向上が期待できる．背景のあるシーンにおいて再現度が低下する要因の一つとして，画像内で動体の境界を正確に識別することの難しさが挙げられる．この課題の解決策として，画像中の動体の位置情報を明示的に入力として与える手法や，本研究で適用した前景・背景領域を用いて，個別に処理した上で結果を統合する手法を導入することで，より高精度な再現が可能となると考えられる．

8. おわりに

本研究では，博多祇園山笠の文化的価値を後世に伝えることを目的として，静的な 3D モデルの構築と動的シーンの自由視点再現の両面からアプローチを行った．

まず，静的な 3D モデルの構築では，フォトグラメトリと SfM を活用し，高精細な山笠の三次元復元を実現した．これにより，山笠の造形を正確に保存できることを確認した．次に，動的なシーンの再現では，4D Gaussian Splatting（4DGS）を用いて，時空間的に一貫した自由視点映像を生成した．実験の結果，提案手法が山笠の動きを含めたシーンの再構築に有効であることを示した．

しかし，本研究にはいくつかの課題が残されている．特に，山笠の動きの再現精度についてはカメラの配置や数，背景の有無により精度が大きく影響を受けることが確認された．また，4DGS はレンダリングが高速である一方，トレーニングに時間を要するため，大規模なシーンの処理にはさらなる計算効率の改善が求められる．

今後の展望として，まず，カメラの配置最適化や高精細なデータ取得手法を導入することで，動的シーンの再現精度を向上させることが考えられる．背景の影響が大きかったことを踏まえ，前景・背景領域を個別に処理し，最終的に統合することで精度の向上が期待できる．また，実映像シーンでは多視点での検証や複数の動体への対処が求められ，これらを踏まえたより精緻なアプローチが必要となると考えられる．

参考文献

[1] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, Xinggang Wang, ”4D Gaussian Splatting for Real-Time Dynamic Scene Rendering.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June, 2024, pp. 20310-20320.

[2] Nadia Magnenat-Thalmann, Alessandro Enrico Foni, George Papagiannakis, Nedjma Cadi-Yazli, ”Real Time Animation and Illumination in Ancient Roman Sites.” Int. J. Virtual Real., vol. 6, no. 1, pp. 11–24, 2007.

[3] 野口淳, モバイルスキャン協会, 高田祐一, 中村良介, 金澤舞, 中島将太, 木翼郎, 山田暁, 石井淳平, 宮本利邦, 仲林篤史, 廣瀬覚, 清水直哉, 今井邦彦, 中尾智行, 堀木真美子, 古川淳一, 伊藤由美子, 大矢祐司, 桑山優樹, 村陸, 長谷川浩, 国武貞克, 川崎志乃, 矢内一正, 数藤雅彦, 上山敦史, 鬼塚勇斗, 武内樹治, 津田富夢, 林亮太, 溝口泰久, 中鉢賢治, 中野純, 中村耕作, 樋上昇, 三好清超, ”デジタル技術による文化財情報の記録と利活用.” 奈良文化財研究所研究報告, vol. 5, no. 37, 独立行政法人国立文化財機構奈良文化財研究所, 2023.

[4] Katsushi Ikeuchi, Daisuke Miyazaki, ”Digitally archiving cultural objects.” Springer Science & Business Media, 2008.

[5] Liang Li, Woong Choi, Kozaburo Hachimura, Keiji Yano, Takanobu Nishiura, Hiromi T. Tanaka, ”Virtual yamahoko parade experience system with vibration simulation.” ITE Transactions on Media Technology and Applications, vol. 2, no. 3, pp. 248–255, 2014.

[6] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, ”NeRF: representing scenes as neural radiance fields for view synthesis.” Commun. ACM, vol. 65, no. 1, pp. 99–106, Jan. 2022, doi: 10.1145/3503250.

[7] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, ”3D Gaussian Splatting for Real-Time Radiance Field Rendering.” ACM Transactions on Graphics, vol. 42, no. 4, July 2023.

[8] Mark Hamilton, Zhoutong Zhang, Bharath Hariharan, Noah Snavely, William T. Freeman, ”Unsupervised Semantic Segmentation by Distilling Feature Correspondences.” International Conference on Learning Representations, 2022.

[9] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, Vladlen Koltun, ”Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 3, 2022.