

# 3D Gaussian Splatting および赤外線タッチセンサを用いた 指の歩行ジェスチャーによる 3次元空間上歩行システムの制作

石井飛鳥<sup>†1‡2</sup> 白井雅也<sup>†1‡2</sup> 村田洋敏<sup>‡2</sup>

**概要:** 本研究では、指の歩行ジェスチャーを用いて3次元空間を歩行するシステムを開発した。赤外線タッチセンサを活用し、指先の動きをトラッキングすることで、ユーザーは指で歩く動作をするだけで、3D Gaussian Splatting によって再構成された建築物や自然環境のモデル内を探検できる。本システムは、ヘッドマウントディスプレイや AR デバイスといった技術を用いないことで、複数人での視点共有や簡単な操作性を実現しており、住宅や展示空間の模型、公共空間の案内板、アクセスが困難な地域のマップなど、様々な応用先が想定される。今後の課題としては、ジェスチャー認識の精度向上、赤外線タッチセンサの設置面の検討などが挙げられる。

## 1. はじめに

本研究は、図 1 に示した指先で人の歩く姿を模倣する動き（指人間の歩行ジェスチャー）を赤外線タッチセンサでトラッキングし、指人間のいる場所と向いている方向を推定した情報を元に、3D Gaussian Splatting[1]でキャプチャされた建築物や自然環境といった現実空間の再構成モデル内を探検するコンテンツのためのシステムを制作したものである。図 2 に示すように、ユーザーは、タッチセンサが取り付けられた地図上を指人間の歩行ジェスチャーで歩くという簡単な動作で、CG 上の 3次元空間を自由に移動することができる。VR および AR 技術を用いてある場所の再現をする取り組みはもはや一般的なものになりつつあるが、体験者が見ている視点を他の人と共有しづらい、一人一人がデバイス进行操作するため共通の体験を共有できないといった問題がある。本技術は、指先のジェスチャーから視点を計算し映像をディスプレイに描画するため、地図と指人間を俯瞰する 3人称視点と、指人間の 1人称視点を同時に複数の人が見ることができる。また、操作のためのデバイスを必要とせず、誰もが行ったことのある指のジェスチャーで操作ができるため、誰でも簡単に体験することができる。本技術は、住宅や展示空間などの建築物模型、公共空間における地図案内板、実際に歩くことが困難な環境保全区域や被災地の再現マップへの応用が想定されている。



図 1 指人間の歩行ジェスチャー



図 2 展示の全体図。

## 2. 関連研究

実世界を CG 上にモデルとして再構成し、中を探検できるようにする体験としては、歴史的建造物のデジタルアーカイブ研究などが代表的である。例えばギザのピラミッドを自由視点で探索する GIZA 3D[2]などが挙げられるが、こうした研究は建造物の形状に関して正確な計測を求めるものであり、非研究者や非専門職の人々は購入が難しいレーザースキャナやフォトグラメトリ装置を利用することが多い。本研究では、画像情報からボリュームデータを生成する深層学習モデルの 3D Gaussian Splatting を利用することによって、再構成のためのデータの獲得が容易な方法を採用した。

また、観光地や建造物や自然をリアルに探索するコンテンツとしては GIZA 3D の他にも、有馬温泉と草津温泉が参加する「湯めぐり VR」[3]や、長野県箕輪町の「360° VR でめぐる長野県箕輪町」[4]などが挙げられるが、いずれも VR や AR デバイスを用いることが想定されている。

<sup>†1</sup> 慶應義塾大学

<sup>‡2</sup> 電通クリエイティブ X

こうしたデバイスによる体験設計は没入感が演出できる一方で、複数人で体験を共有することが困難である。また、視点移動のための操作方法をユーザーが学習する必要がある。以上の問題に対して、本システムでは指のジェスチャーから推定されたカメラの位置と姿勢からレンダリングを行い、プロジェクターに投影することで、複数人での体験共有と、簡単な操作を実現している。

通常の平面を操作可能なインターフェイスにする研究としては、赤外線とステレオカメラによるタッチ認識を行う研究[5][6]や、レーザーを用いた接触認識[7]が挙げられる。本システムでは、赤外線によるマルチタッチ検出を採用し、押下された指の情報からカメラの情報を推定する。

空中での手のジェスチャー認識によるインターフェイス設計は多数の先行事例が存在する。Harrison et al. [8]では、平面と手が接触する音から、手のどの部位（指先、拳など）で接触を行なったか判定するシステム”TapSense”を開発した。本システムでは、指人間の歩くジェスチャーからカメラの位置および姿勢（指人間の目にあたる）を推定したい。従って、赤外線センサで指先の位置のみをキャプチャすることにした。

### 3. 提案するシステム

本研究では、建築物や自然環境といった空間をキャプチャした画像群を、3D Gaussian Splatting を用いてコンピューター・グラフィックスのソフトウェア上で扱えるボリュームとして3次元に再構成する。さらに、指先の歩くジェスチャーを、赤外線タッチセンサによってカメラの座標と姿勢に変換する。その情報を元にレンダリングを行い、プロジェクターに描画し、キャプチャされた元の空間を指人間の歩く動作によって探検する体験を提供する。

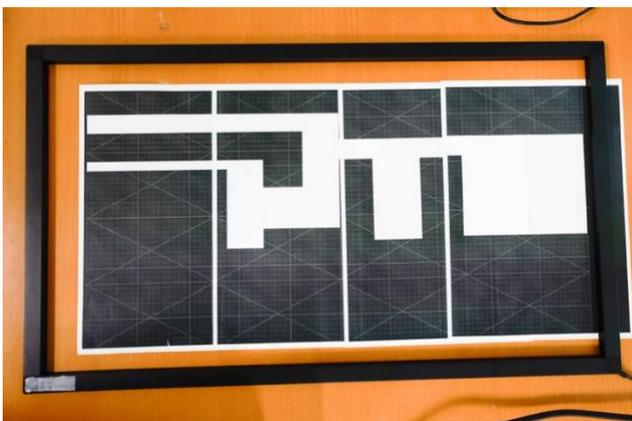


図3 赤外線タッチセンサーとマップ。白色の領域内を歩行する。

#### 3.1 キャプチャデバイスの設計

市販されているディスプレイ用赤外線センサを用いて、キ

ャプチャ可能領域内で押下、押上された指先の位置をトラッキングすることにより、指人間が空間上のどこに立っているのかを推定する。このデバイスは本来ディスプレイの上から取り付けて、タッチ入力を擬似的に可能にする製品である。しかし、今回は複数の指の同時トラッキングが可能な単なる赤外線タッチセンサとしてこれを用いて、図3のように紙に印刷されたマップを取り囲むように設置することにより、マップを歩きまわるといったインターフェイス設計をおこなっている。

マップには、歩行可能エリアと歩行不可能エリアがそれぞれ白黒でゾーニングされている。これは、3D Gaussian Splatting によって再構成されたボリュームの中を自由に動き回った結果、本来は壁の内側であったような場所にカメラが入るような、製作者が意図しないユーザーの操作を回避するためである。ユーザーは、歩行可能エリアの中に指を置いて、歩行のジェスチャーをすることによって、マップ内の探索を行う。

#### 3.2 歩行ジェスチャーの認識

赤外線タッチセンサで取得された指の状態（押下、押上、位置の $x, y$ 座標）から、3次元空間内のカメラ座標と姿勢を推定する。そこで、リアルタイムレンダリングソフトの TouchDesigner[9]を用いて図4のような状態監視を行なった。



図4 カメラ映像とは別に、TouchDesignerでの状態監視と体験のインストラクションが表示される。

##### 3.2.1 カメラの座標計算

TouchDesigner に入力されたセンサの値から、押下されている全ての指の $x, y$ 座標の平均をカメラの $x, y$ 座標とする。 $z$ （高さ）座標は1.4mで固定する。

##### 3.2.2 カメラの姿勢計算

TouchDesigner に入力されたセンサの値から、最後に押下された指 $t$ 、一個前に押下された指 $t-1$ 、2個前に押下された指 $t-2$ の $x, y$ 座標から、カメラの $z$ 軸の回転角 $\theta_t$ を計算する。つまり、2歩前に地面につけた足の座標から1歩前に地面につけた足の座標に進むベクトルと、1歩前に地面につけた足の座標から最後に地面につけた足の座標に進むベクトルの加重平均を取る。数式では以下の通りである。

$$\theta_t = \alpha \text{Arctan2}(x_t - x_{t-1}, y_t - y_{t-1}) + (1 - \alpha) \text{Arctan2}(x_{t-1} - x_{t-2}, y_{t-1} - y_{t-2})$$

ただし、 $0.5 \leq \alpha \leq 1$ のハイパーパラメータである。現在時刻 $t$ における $z$ 軸回転量と、時刻 $t-1$ の $z$ 軸回転量の加重平均をとることによって、視線の極端な回転を抑制した。

本システムでは、 $|\theta_t - \theta_{t-1}|$ （ただし、 $0 \leq |\theta_t - \theta_{t-1}| \leq 180$ ）が $90$ に近づくと、 $\alpha$ が $1$ に近づくように、動的な重みづけを行なった。こうすることで、歩くジェスチャーをしている際に $90$ 度に曲がるようなターンが発生した場合でもスムーズな方向転換が実現される。

### 3.3 レンダリング

本システムでは、実世界の建築物や自然環境をモデル化するために3D Gaussian Splattingを用いた。3D Gaussian Splattingとは、複数の視点の画像から3D空間を再構成する手法の一つである。複数視点の画像からポイントクラウドを作成し、機械学習によって各ポイントをパラメータのあるガウシアンに変換したデータとして記録する。そのため、レンダリング時は記録されたパラメータを用いての高速なレンダリングを可能にしている。

本システムにおける実装方法としては、まず撮影した映像からFFmpeg[10]を用いて動画から2フレーム/秒の連番画像を書き出す。その後、書き出された画像群をもとに3D Gaussian Splattingを用いて映像として記録した空間の特徴をボリュームデータとして復元する。

Luma AIの提供するプラグイン[11]を用いて3D Gaussian SplattingのデータをUnreal Engine 5[12]内で描画し、3.2で計算したカメラの情報はOpen Sound ControlをもちいてTouchDesignerからUnreal Engine 5に送信される。Unreal Engine 5側は、受信した情報を元にカメラの位置と姿勢を決定し、レンダリングを行う。レンダリングされた画面を図5に示す。



図5 Unreal Engine 5でレンダリングされた画像

## 4. 今後の展望と課題

### 4.1 指が3本以上押下されていた場合の判定

本作品は指人間による歩くジェスチャーが体験として想定されているため、指が三本以上押下されている場合、カメラの座標は全ての指における $x, y$ 座標の平均となる。指を故意に三本以上押下していない場合でも、指を押したり離したりする瞬間における計測の誤差により指が三本以上押された判定になる場合がある。本課題への対処としては、三本以上の押下が検出された時にどの指が人差し指と中指なのかをwebカメラ入力と、OpenPose[13]やMediaPipe[14]などのハンドトラッキングシステムによって判定する方法が考えられる。

### 4.2 指のジェスチャー認識

本システムは、指が接地している座標を赤外線タッチセンサによる取得しているものであり、歩行動作をおこなっているかの認識を行うものではない。しかし、歩いている状態の判定が可能になると、誤ってセンサに触れてしまった場合の誤作動の防止や、複数人での利用が可能となる。

### 4.3 赤外線タッチセンサの設置面の検討

本システムでは、赤外線タッチセンサの設置面に3.1にて説明した移動可能範囲が描かれた紙のマップを置いた。しかし、赤外線タッチセンサそれ自体は設置面が平面であればどこでも設置ができる。従って本システムは、建築物模型、地図案内板、実際に歩くことが困難な環境保全区域や被災地の再現マップへの応用が見込まれる。また、ディスプレイ上に設置することにより地面で動的な情報の表示が可能となる。例えば、地図上の特定の地点に近づくと、そこに関する詳細が地面に表示される情報の提示や、リアルタイムでの天候や交通情報を反映させることも可能である。

## 参考文献

- [1] Kerbl, Bernhard, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. "3D Gaussian Splatting for Real-Time Radiance Field Rendering." arXiv [cs.GR]. arXiv. <http://arxiv.org/abs/2308.04079>.
- [2] "Digital Giza |." n.d. Accessed December 22, 2023. <http://giza.fas.harvard.edu/giza3d/>.
- [3] "湯めぐり VR." 2020. 湯めぐり VR. Accessed December 22, 2023. April 20, 2020. <http://onsenvr.com/>.
- [4] "360° VRで巡る長野県箕輪町." n.d. Accessed December 22, 2023. <https://minowa-vr.nagano.jp/>.
- [5] Wilson, Andrew D. 2005. "PlayAnywhere: A Compact Interactive Tabletop Projection-Vision System." In Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, 83–92. UIST '05. New York, NY, USA: Association for Computing Machinery.
- [6] Agarwal, Ankur, Shahram Izadi, Manmohan Chandraker, and Andrew Blake. 2007. "High Precision Multi-Touch Sensing on Surfaces Using Overhead Cameras." In Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07), 197–200. IEEE.

- [7] Roeber, Helena, John Bacus, and Carlo Tomasi. 2003. "Typing in Thin Air: The Canesta Projection Keyboard - a New Method of Interaction with Electronic Devices." In CHI '03 Extended Abstracts on Human Factors in Computing Systems, 712–13. CHI EA '03. New York, NY, USA: Association for Computing Machinery.
- [8] Harrison, Chris, Julia Schwarz, and Scott E. Hudson. 2011. "TapSense: Enhancing Finger Interaction on Touch Surfaces." In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, 627–36. UIST '11. New York, NY, USA: Association for Computing Machinery.
- [9] "Derivative." n.d. Derivative. Accessed December 22, 2023. <https://derivative.ca/>.
- [10] "FFmpeg." n.d. Accessed December 22, 2023. <https://ffmpeg.org/>.
- [11] "Luma Unreal Engine Plugin (0.4)." n.d. Luma Unreal Engine Plugin (0.4). Accessed December 22, 2023. <https://lumaai.notion.site/Luma-Unreal-Engine-Plugin-0-4-8005919d93444c008982346185e933a1>.
- [12] "Unreal Engine 5." n.d. Unreal Engine. Accessed December 22, 2023. <https://www.unrealengine.com/ja/unreal-engine-5>.
- [13] Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." arXiv [cs.CV]. arXiv. <http://arxiv.org/abs/1812.08008>.
- [14] Lugaresi, Camillo, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, et al. 2019. "MediaPipe: A Framework for Building Perception Pipelines." arXiv [cs.DC]. arXiv. <http://arxiv.org/abs/1906.08172>.