生成 AI を用いた航空写真からの 3D 都市空間復元

志村燿平[†1]　川合康央[†1]

概要：本研究では，国土地理院が公開している過去の航空写真を情報源とし，生成 AI 技術を用いて整合性の取れた 3D 仮想空間を構築する手法を提案する．デジタルツイン技術の発展により，現存する空間の 3D 化は容易になったが，測量不可能な「過去の空間」の復元は依然として困難である．本研究では，超解像化・カラー化による前処理と，Large World Model (LWM) である「Marble」を組み合わせたパイプラインを構築し，単一の航空写真から 3D モデルを生成する手法を検証した．特に，本手法の有効性を客観的に証明するため，正解データ（Ground Truth）が存在する「現在の文教大学湘南キャンパス」を対象とした比較検証を行った．その結果，適切なプロンプトエンジニアリングを用いることで，単一画像からでも Google Earth 等の既存 3D マップと比較しうる整合性を持った 3D 空間が構築可能であることを示した．

1. はじめに

近年，現実空間をデジタルデータとして再現するデジタルツイン技術は，都市計画，防災シミュレーション，観光，エンターテインメントなど多岐にわたる分野で重要性を増している．これまでに著者はドローン（無人航空機）を用いた空撮画像と SfM（Structure from Motion）技術を組み合わせた 3D モデル構築手法の研究を行ってきた[1]．これらの技術は現存する空間の 3D 化において高い品質を実現する一方で，現地撮影が物理的に不可能な対象，すなわち「過去の空間」への適用は不可能である．失われた建物や地形を復元することは，歴史的資料の保存や災害検証の観点から意義深い．

過去の空間復元における一次資料としては，国土地理院等に保存されている「航空写真」が挙げられる．しかし，これらは現代のデータと比較して品質上の課題が多い．画質としてはフィルム撮影されたモノクロ画像であることからノイズが多く，視点としては真上からの「直下視」画像しか存在せず，側面情報が欠落している．この制約により，過去空間の復元には画像補完や視点生成を含む生成的アプローチが不可欠となる。

2. 先行研究

関連する先行研究として，Chen ら（2025）[2]は，歴史的な航空写真の低画質性が AI による建物認識を阻害する問題に対し，GAN（敵対的生成ネットワーク）を用いた画像強調パイプラインを提案している．彼らは，低品質な白黒写真を「DeOldify」でカラー化し，さらに「Real-ESRGAN」を用いて超解像化することで，建物検出精度の向上を報告しており，前処理プロセスの有効性を示唆している．また，単一の航空写真から直接 3D データを生成する試みとして，Bensedik ら（2025）の「AIM2PC」[3]が挙げられる．この研究では拡散モデルを用いることで，見えていない壁の形状までをノイズから予測・生成し，単一画像から建物の 3D 点群を再構築しており，1 枚の写真からの空間構築における重要な先行事例である．さらに Hua ら（2025）による「Sat2City」[4]は，単一の衛星画像から都市スケールの 3D シーンを自動生成する手法であり，広域モデルの生成可能性を示している．

しかし，これらの先行研究の多くは実行コードが公開されていないケースがあり，第三者による実証や応用が困難である．そこで本研究では，特定のブラックボックスなモデルを独自に開発するのではなく，Stable Diffusion や Gemini，Marble といった現在一般に利用可能な生成 AI ツールを体系的に組み合わせ，再現可能な空間構築フローを確立することを目的とする．また，本研究では生成 AI を完全自動化ツールとしてではなく，人間によるプロンプト設計を含むインタラクティブな復元プロセスとして位置付けた.具体的には，「現代のデータ（Google マップ）」を用いた検証実験と，「過去のデータ（国土地理院）」を用いた適用実験の 2 段階のアプローチを採用し，現代の画像で生成精度を検証した上で，過去のデータへと適用範囲を広げる構成とする．

3. 提案手法

本研究では，入力データの質に応じて処理を分岐させつつ，最終的に LWM で 3D 化する統合パイプラインを提案する．データの特性に合わせ，現代データによるベンチマーク検証（実験 1）と過去データを用いた復元（実験 2）の 2 つのルートを設定した．実験 1 では，入力画質が理想的な状態において，後段の 3D 生成 AI がどの程度の精度を出せるか，その上限値と最適なプロンプトを検証する．ここでは Google マップ等の現代の航空写真を入力とする．

†1 文教大学大学院　情報学研究科

一方，実験 2 の対象となる国土地理院の過去の航空写真は低解像度かつモノクロであるため，そのままでは 3D 生成に適さない．そのため，Stable Diffusion WebUI 上の「R-ESRGAN 4x+」を用いて解像度を 4 倍に拡張して輪郭を復元する超解像化と，「DeOldify」を用いた AI による着色（カラー化）を行う．

その後共通のプロセスとして対象エリアのトリミングを行い,画像補正，視点変換,および 3D 生成を行う．まず，トリミングした画像に対して Gemini の Nano Banana Pro を用い，さらに高画質化を行う．次に，真上からの画像（直下視）だけでは LWM が建物の高さを正しく推論できない場合があるため，画像生成 AI を用いて直下視画像を「斜め上空から俯瞰した画像（鳥瞰図）」へと変換し，擬似的に側面情報を付与する．最後に，得られた俯瞰画像を LWM ツール「Marble」に入力し，3D モデルを生成する．この際，AI が建物の形状を正しく認識できるよう，プロンプトエンジニアリングによる幾何学的制約を与える．本手法の全体的な処理フローを以下(図 1)に示す．

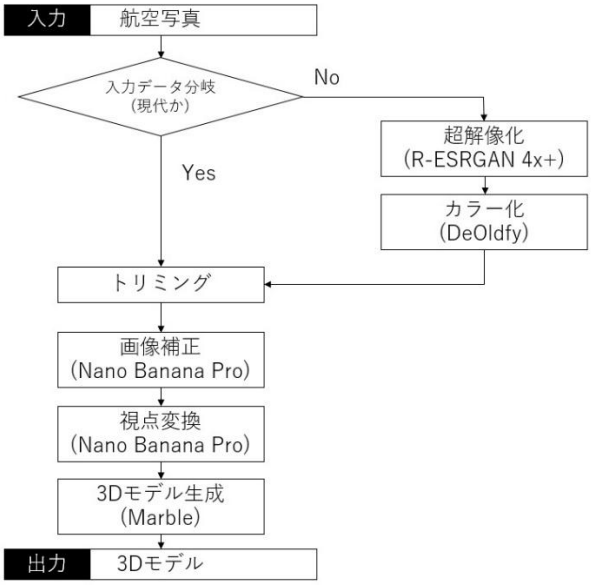

図 1　提案手法の処理フロー

4. 実験結果

4.1 実験 1：現代データによるベンチマーク検証

まず，提案手法の基礎能力と最適なプロンプト構成を確認するため，現代のデータを用いた実験を行った．対象は文教大学湘南キャンパスとし，Google マップの航空写真を入力とした． Google マップからトリミングした画像に対し，Gemini を用いて色褪せと歪みの補正，および高画質化を行った（図 2）．続いて，高画質化した画像から俯瞰視点画像の生成を行った（図 3）．

図 2　Gemini により高画質化を行った航空写真
(左:入力画像，右: 高画質化後)

図 3　Gemini で作成した俯瞰視点画像

この俯瞰画像を Marble に入力し，プロンプトを指定せずに 3D モデルを生成したところ，校舎の境界が曖昧になり，建物と地面が融解する現象が確認された（図 4）．これは，入力画像が高画質であっても，AI に対する形状の定義が不足していると，正確な立体構造が生成されないことを示唆している．

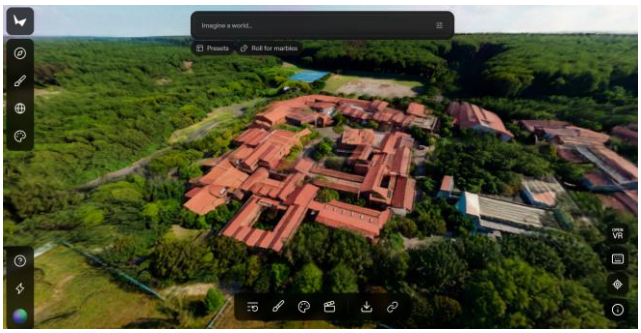

図 4　プロンプトなしで出力した 3D モデル

そこで，幾何学的整合性を保つため，「Solid, distinct geometry（硬固で，明確な幾何学形状）」や「Extremely sharp, clean architectural edges（極めて鋭く，綺麗な建築的なエッジ）」といった強い制約条件を含むプロンプトを適用した．その結果，品質の向上が見られ，複雑な校舎群が独立した構造物として分離された（図 5）．Google Earth の 3D 表示（図 6）と比較しても，一定の整合性が取れた 3D

モデルが生成されたと言える．

図 5　改善プロンプトを適用して出力した 3D モデル

図 6　Google Earth 上の文教大学湘南キャンパス

4.2　過去データへの適用検証

次に, 実験 1 で確立した手法を過去のデータに適用した．対象は 1968 年における神奈川県藤沢市善行の団地とし,国土地理院空中写真（整理番号 MKT685）を入力とした．400dpi のモノクロ画像を R-ESRGAN で 4 倍に拡大し，DeOldify（render factor: 35）でカラー化を行った（図 7）．その後対象エリアのトリミングを行い，Gemini を用いてさらなる高画質化（図 8）および俯瞰視点画像の生成を行った（図 9）．

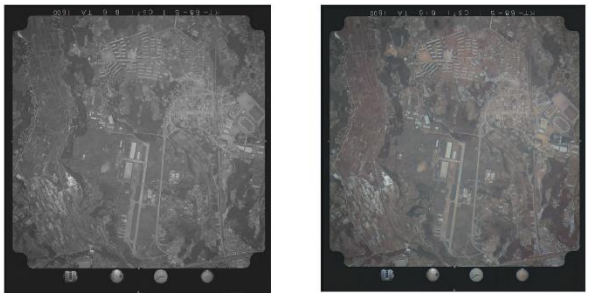

図 7　高画質化、カラー化した過去の航空写真
(左:入力画像，右:カラー化,高画質化後)

図 8　Gemini によりさらに高画質化した画像
(左:入力画像，右:高画質化後)

図 9　Gemini で作成した過去の航空写真の俯瞰視点

この俯瞰画像を Marble に入力し，実験 1 の知見に基づき，団地の形状特性（均一な白い直方体）を定義する強い制約条件を含むプロンプト（"perfect, simple, solid rectangular prisms", "untextured white material"等）を適用して 3D モデルを生成した（図 10）．

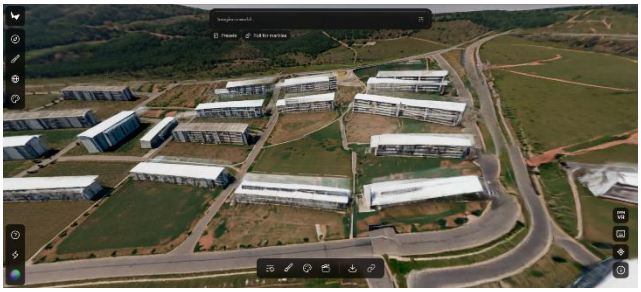

図 10　上記のプロンプトで出力した過去の 3D モデル

4.3　考察と他モデルとの比較

実験 2 の結果，Marble に対し強い制約条件を与えたにもかかわらず，生成された 3D モデルには建物の「残像」や「重なり」といった破綻が見られた（図 10）．

本手法の妥当性を検証するため，同一の俯瞰画像を他社の 3D 生成サービス「Hitem3D」に入力し，生成結果の比較を行った（図 11）．なお，Hitem3D はプロンプトによるテキスト指示に対応しておらず，画像情報のみから 3D モデルを生成する仕様である．

図 11　Hitem3D で出力したモデル

Hitem3D による生成結果を確認すると，壁面のテクスチャにはノイズが表れているものの，Marble で見られたような建物の融解や多重化（残像）は発生しておらず，団地の配置構造自体は正しく認識されている． この事実は，画像からだけでも，AI 自体は対象物の基本的な構造を認識する能力を持っていることを示唆している．

それにもかかわらず，Marble において構造的な破綻（残像）が生じた原因は，人間が与えたプロンプトにあると考えられる．本実験では，入力画像の 3D モデル生成に対し，「欠陥のない完全な幾何学立体」という理想的な状態を強要した．AI が画像から読み取った情報を，人間がプロンプトで無理に上書きしようとした結果，AI の認識プロセスに矛盾が生じ，二つの情報が整合せずに重なり合う「残像」として出力された可能性が高い.つまり，画像の実情を無視した的外れなプロンプトは，AI の推論を助けるどころか，かえって本来の正しい認識を阻害するノイズとして作用してしまったと考えられる．

5. まとめと考察

本研究では，国土地理院が公開する過去の航空写真を対象とした 3D 都市空間の復元を目的とし，汎用生成 AI ツールを体系的に統合したワークフローを提案した．

現代の高画質画像を用いた検証では，強い制約プロンプトが有効に機能し，高精度なモデル構築に成功した．一方，過去のデータを用いた実験では，プロンプトを用いない Hitem3D が構造の再現に成功した一方で，強い制約を与えた Marble では構造が破綻するという逆説的な結果が得られた．

この結果から導き出される重要な知見は，プロンプトは常に品質を向上させるわけではなく，入力画像との適合性が低ければ阻害要因になり得るということである．AI は画像からだけでもある程度の構造を推論可能である．しかし，そこに人間が画像の状態を無視した的外れな理想をプロンプトとして強制すると，AI の画像認識とテキスト指示が衝突し，結果として出力の破綻を招く．したがって，過去の空間復元において重要なのは，単に理想的な指示を与えることではない．入力画像の特性を人間が正しく分析し，AI の認識と矛盾しないよう，画像の現実に即した適切なプロンプト設計を行うことなのではないかと考えられる．

結論として，本研究は生成 AI による復元プロセスにおいて，人間が「AI の認識と指示の乖離」を埋める調整役として機能する必要性を示した点に意義があると考えられる．画像の実情に即したプロンプト設計こそが，AI の能力を正しく引き出し，整合性を担保する鍵である．

謝辞　本研究は JSPS 科研費 JP 23K11728 の助成を受けたものです．

参考文献

[1] 志村燿平，櫻井 淳.. SfM による大規模構造物 3D モデル構築のためのドローン空撮方法に関する研究，情報処理学会第 87 回全国大会，2025.
[2] Chen, P., et al.. A GAN-Enhanced Deep Learning Framework for Rooftop Detection from Historical Aerial Imagery, International Journal of Remote Sensing, 2025.
[3] Bensedik, S., et al.. AIM2PC: Aerial Image to 3D Building Point Cloud Reconstruction, arXiv preprint, 2025.
[4] Hua, T., et al.. Sat2City: 3D City Generation from A Single Satellite Image with Cascaded Latent Diffusion, ICCV, 2025.