

歩行者を含む実写映像からの4Dシーンモデリングの初期検討

A Preliminary Study on 4D Scene Modeling from Live-action Video Including Pedestrians

田中来樹¹
Raiki Tanaka

出口 大輔¹
Daisuke Deguchi

川西 康友¹
Yasutomo Kawanishi

村瀬 洋¹
Hiroshi Murase

名古屋大学¹
Nagoya University

1 はじめに

近年、より現実に即した自動運転技術の評価を目的として、写実的かつ現実的なドライビングシミュレーションの実現への期待が高まっている。その中でも、Epic Games社が開発したゲームエンジンである UnrealEngine4 [1] (以下、UE4) は、写実的な描画を可能にする高いレンダリング性能、マルチプラットフォーム対応、オープンソースなどの理由から、ドライビングシミュレータでも多く利用されている。このようなドライビングシミュレータ上で現実に即した環境を再現するためには、CG自体の写実性に加え、環境内に存在する物体の動きまでが詳細に記述された時間情報を含む3Dシーン(以下、4Dシーンと呼ぶ)が必要となる。しかし、現実環境に即した4Dシーンをゼロから手作業で構築するコストは非常に高い。そのため、写実的で実環境での物体の動きまでを再現する4Dシーンの安価な構築方法が求められている。

本発表では、歩行者を含む実写の単眼カメラ画像系列から歩行者の動きまでを含めた4Dシーンを構築する手法を提案する。また、UE4を用いて、構築した4Dシーンが再現可能なことを示す。

2 単眼カメラ画像系列からの4Dシーン構築手法

提案手法は、事前に構築した環境地図と単眼カメラ画像系列を入力とし、歩行者の動きを含めた4Dシーンを再現する手法であり、以下の4つの処理から構成される。

- (1) **環境地図の構築(事前処理)** シミュレーション対象の環境を撮影した複数枚の画像に対して Structure from Motion (SfM) を行い、各画像の撮影位置、カメラ姿勢、環境を表す低密度な3D点群を得る。次に、復元された3D点群とカメラ位置・姿勢を基にして Multi-View Stereo (MVS) を行い、高密度点群を推定する。そして、高密度点群から Mesh を生成し、テクスチャを加えた環境地図(3Dポリゴンモデル)を得る。これらの処理には、Agisoft社が開発した Metashape [2] を用いる。
- (2) **入力画像系列からの歩行者の3D位置・姿勢推定** 上記(1)で得られた環境地図と特徴点照合により、入力画像系列の各カメラ位置・姿勢を推定する。次に、入力画像系列の各画像に対して 3DMPPE [3] を適用し、各カメラ位置・姿勢と組み合わせて歩行者の3D位置・姿勢を得る。
- (3) **歩行者の3Dアニメーションの生成** 上記(2)で求めた歩行者の3D位置・姿勢を画像系列間で追跡す

ることで歩行者の3Dアニメーションを生成する。

- (4) **3Dアニメーションの配置** 事前に構築した環境地図と上記(3)で得られた3Dアニメーションデータを UE4 へと取り込む。

3 実験結果

提案手法の有用性を確認するために、名古屋大学豊田講堂を撮影した画像系列を入力として4Dシーンの構築実験を行った。実験に用いたカメラは Sony $\alpha 7$ III であり、レンズには Sigma Art 24-70 mm F2.8 DG DN を用いた。このカメラを用いて豊田講堂を複数視点から撮影した画像 205 枚を入力として環境地図を構築した。また、同じカメラを用いて撮影した動画(30 fps) から約 1,600 フレームを切り出して4Dシーン構築(歩行者の3D位置・姿勢の推定)の入力画像系列として用いた。

実験に用いた入力画像系列の一例と4Dシーンの構築結果を図1に示す。図1から分かるように、撮影シーン中の3D環境と人物位置を概ね正しく復元できていることを確認した。

4 むすび

本発表では、歩行者を含む実写の単眼カメラ画像系列から複数人の3D位置・姿勢推定を行い、UE4上で再現可能な4Dシーンの構築を行う手法を提案した。今後の課題として、大規模な環境を用いた実験、生成するアニメーションの質の評価、歩行者に合わせた3Dモデルの構築、などが挙げられる。

謝辞 本研究の一部は JSPS 科研費 17H00745 による

参考文献

- [1] UnrealEngine4, <https://www.unrealengine.com/>
- [2] Metashape, <https://www.agisoft.com/>
- [3] G. Moon, J. Y. Chang, K. M. Lee, "Camera Distance-aware Top-down Approach for 3D Multi-person Pose Estimation from a Single RGB Image", Proc. of ICCV2019, pp.10133-10142, July 2019.



(a) 入力画像 (b) 構築した4Dシーン

図1 4Dシーンの構築結果