

同一歩行姿勢に NeRF を適用した側方からの歩行映像生成 Side-View Gait Video Generation by Applying NeRF on Same Posture Images

平岡 翼^{1*}
Tsubasa Hiraoka¹

原田 翔太¹
Shota Harada¹

満上 育久¹
Ikuhisa Mitsugami¹

¹ 広島市立大学

¹ Hiroshima City University

Abstract: 本研究では、固定カメラ映像から任意視点の歩行映像を生成する手法を提案する。歩容解析では人物を一定方向から観測することを前提とするが、固定カメラで移動する歩行者を一定の方向から観測することは難しい。提案手法は、固定カメラで撮影された映像中の同じ歩行姿勢を、異なる位置での同一の姿勢で撮影された画像群とみなし Neural Radiance Fields を学習することで、固定カメラ映像から任意視点の歩行映像の生成を可能とする。

1 はじめに

歩容解析とは、人の歩行を撮影した映像から、歩幅や腕の振りなどの歩き方の特徴を解析する技術であり、個人認証や犯罪捜査、医療など幅広い分野で応用されている。このような歩容解析では歩行する人物を常に一定の方向から観測することを前提としている。しかし、実際には移動する歩行者を常に一定の方向から観測することは困難である。例えば固定カメラで人物の歩行を撮影すると、歩くごとに観測方向が変化してしまい、その前提が成り立たなくなる。

そこで本研究では、固定カメラで撮影した人物の観測方向が変化する歩行映像から、歩行する人物を常に真横から観測する映像を生成することを目的とする。そのアイデアとしてまず、他の視点から人物を観測するために、複数視点の画像から対象物の3次元形状を復元する手法である Neural Radiance Fields (NeRF) [1] を学習する。しかし、基本的に NeRF は静的なシーンに対して学習するものであるため、歩行中の人物などの動的なシーンには適用できない。この問題を解決するために、歩行の周期性に着目する。基本的に人の歩行運動は2歩を1周期とした繰り返し運動であるため、固定カメラで複数周期の歩行を観測すれば、位置は異なるものの同じ歩行姿勢をとっているフレームが映像中に複数存在することになる。これらのフレームは、実際には異なる時刻に撮影された異なる位置の人物を捉えたものであるが、姿勢が同じであることを踏まえて見方を変えれば、対象人物の同一の歩行姿勢を異なる

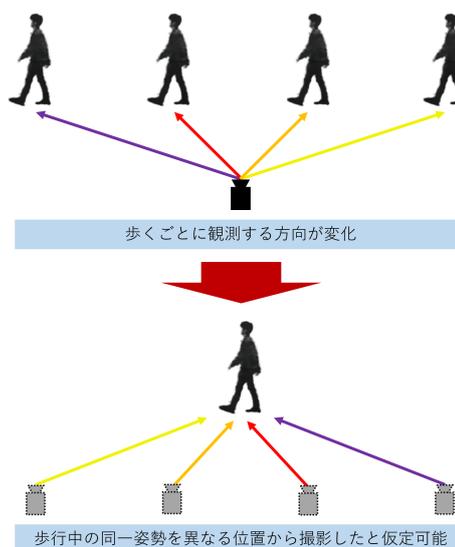


図 1: 固定カメラで擬似的に複数視点から撮影する手法

位置から撮影した画像群とみなすことができる (図 1)。この画像群に対して NeRF を適用することで、この歩行姿勢の任意視点画像が生成可能となり、真横視点から見た画像を生成することもできる。本研究では、このアイデアに基づき、歩行中の人物を真横視点から見た画像をすべての歩行姿勢において取得し、固定カメラで撮影される時々刻々と人物の観測方向が変化してしまう映像から、その人物を常に真横から撮影したような映像を生成できる手法を提案する。

*連絡先: 広島市立大学
〒731-3194 広島市安佐南区大塚東三丁目4番1号
E-mail: f20137@e.hiroshima-cu.ac.jp

2 Neural Radiance Fields

Neural Radiance Fields (NeRF) [1] とは、対象物を複数の視点から撮影した画像をもとに、物体の3次元的な形状を復元し、撮影していない視点を含む任意の視点からの画像を生成する手法である。

NeRFの入力には複数視点からシーンを撮影した画像のほかに、各画像を撮影したカメラの位置・姿勢のデータが必要となる。そのため、通常 NeRF ではまず前処理として、様々な視点からシーンを撮影した画像に対し、Structure-from-Motion (SfM) と呼ばれる手法を行うことで、カメラ位置・姿勢と焦点距離等のカメラの内部パラメータを推定する。そして、撮影した画像と SfM で推定したカメラ位置・姿勢、カメラの内部パラメータを基に NeRF を学習することで、シーンをニューラルネットワークを用いて3次元的に表現し、任意の視点から見ることのできる画像を生成する。

3 提案手法

提案手法の概要を図2に示す。提案手法ではまず、歩行路の側方に設置したカメラで人物の歩行を複数周期撮影する。撮影した映像の各フレームから歩行周期を検出し、各周期から同じ歩行姿勢をとっているフレームを集めて同一姿勢画像セットを作る。そして、画像セットに対して SfM を行い、各画像を撮影したカメラの位置・姿勢を推定する。そのカメラ位置・姿勢と画像セットを用いて NeRF を学習し、真横から人物の歩行姿勢を観測した画像を生成する。この処理を、歩行中の各姿勢に対して行うことで、すべての歩行姿勢における真横視点画像を取得でき、人物を常に真横から見た映像を生成できる。

なお、提案手法では上述の同一姿勢画像セットから人物領域のみを抽出した画像に SfM を適用し、カメラ位置・姿勢を推定する。通常 SfM では、物体が静的であるという仮定のもとカメラ位置・姿勢を推定するため [2]、SfM の入力は同一のシーンを複数視点から撮影した画像群でなければならない。しかし、同一姿勢画像セットの画像は同一の歩行姿勢だが、異なる時刻に撮影された異なる位置にいる人物を撮影した画像であるため、人物を基準に見ると背景が移動していることになり、SfM で正しくカメラ位置・姿勢を推定することはできない。そのため、この画像群にセグメンテーションを適用して人物領域のみを抽出した画像を生成し、背景情報を用いずに SfM を行うことで、人物に対する相対的なカメラ位置・姿勢を推定する。

4 実験

本研究では、固定カメラで撮影した歩行映像から同一の歩行姿勢を取っている画像セットを作成し、それを

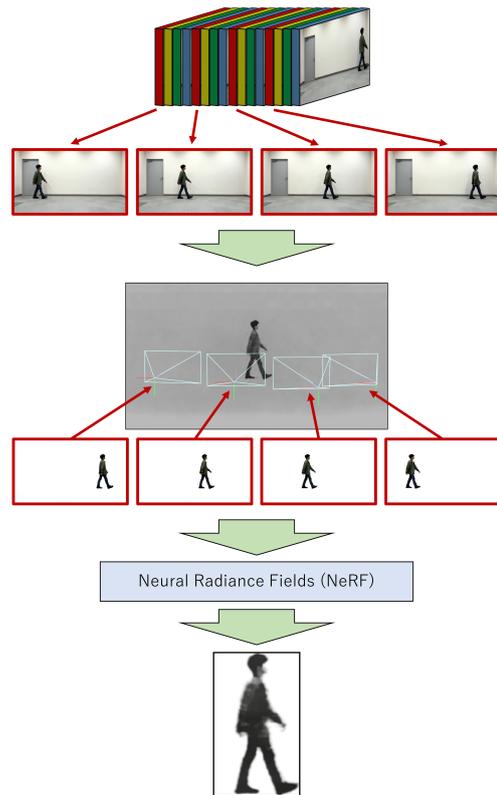


図 2: 提案手法の概要

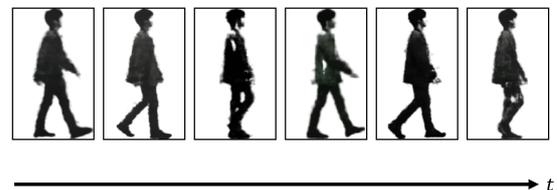


図 3: 映像生成結果

もとに SfM によるカメラ位置・姿勢推定を行い NeRF を学習して各姿勢を真横から観測した画像を取得することで、歩行中の人物を常に真横から観測する映像を生成する実験を行った。

映像の生成結果を図3に示す。映像は各姿勢を真横から見た画像を時系列で表示することで生成した。この結果より、歩行している人物を常に真横から撮影した映像が生成できていることが確認できる。

参考文献

- [1] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tan-cik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” European Conference on Computer Vision (ECCV), 2020.
- [2] 紺野隆志, 西田健次, 糸山克寿, 中臺一博, 「視聴覚統合による動的環境下における三次元再構成の提案」, 第 55 回人工知能学会 AI チャレンジ研究会, 2019.