拡散モデルを用いた LiDAR 点群投影ベースの歩容映像復元

アン・ジョンホ[1,a)]　中嶋 一斗[1,b)]　吉野 弘毅[1,c)]　岩下 友美[2,d)]　倉爪 亮[1,e)]

概要

近年，3D LiDAR 技術を使用した歩容認証の研究が増加しているが，遠い計測距離や低解像度による歩行者点群の欠損が生じ，識別性能の低下が主な課題となっている．そこで本稿では，拡散モデルに基づいて，欠損した LiDAR 点群の歩容映像を復元する手法を提案する．定性的・定量的評価の結果，歩容形状やフレームの欠損した歩容映像が高品質に復元され，歩容認証における識別性能が向上することを確認した．

1. はじめに

歩容認証は，歩く姿から個人を識別する生体認証の一種である．特に，対象者に特別な動作を行わせる必要がなく，顔を隠しても遠隔から識別が可能である点から，大規模な犯罪捜査，スムーズな広域管理システム，本人確認手段などに幅広く応用されている．歩容認証分野で使用される視覚センサとして RGB カメラが最も一般的であるが [4]，テキスチャ情報を活用するためデプスカメラを用いた研究も報告されている [15]．一方，デプスカメラと同様にレーザ光に基づく距離センサの一種である 3D LiDAR センサが存在する．特にデプスカメラと比べて，LiDAR センサは 360° の広い方位角を有しており，より遠い計測距離で周辺環境内の物体の位置や形状を点群データとして計測できる点から，高精度な自己位置同定や障害物検出といった環境認識を要する自律移動ロボットや自動運転車に広く利用されている．

しかし，この LiDAR センサを歩容認証手段として使用する際，遠い計測距離や低解像度のセンサによって歩行者点群がスパースに取得され，詳細・全体の人物の輪郭が崩れる問題がある．また，センサの遅い回転速度，障害物による遮蔽，不正確な人物検出などによって，歩容映像内のフレーム欠損が生じることも識別性能が低下する原因として挙げられる．

LiDAR センサを用いた歩容認証の先行研究の多くでは，歩行者点群を直接に投影した深度映像を識別手法の入力として用いている [1–3,12,14]．また，近年，画像生成分野では生成モデルの一種である拡散モデル [5]（以下，DDPM）を事前に学習させ，劣化した画像を線形逆問題として解く研究が盛んに行われている [7,8]．本稿では，学習ずみ DDPM を用いて，上述の解像度とフレームの二つの要素の観点から劣化した点群投影の歩容映像を復元する手法に着目する．実験では，LiDAR 点群の歩容認証の大規模ベンチマークである SUSTeck1K データセット [12] を用いて解像度やフレームの欠損を人工的に再現し，提案手法の復元品質を定性的・定量的に評価した．また，サブタスクである歩容認証に応用することで，点群歩容映像の復元によるベースライン [12] の識別性能向上を確認した．

2. 関連研究

2.1 LiDAR センサを用いた歩容認証

LiDAR センサを用いた歩容認証における深層学習手法では，主に GEI ベースの手法と深度映像ベースの手法の 2 つに分けられる．GEI ベースの手法 [3] では，歩行者点群を二値画像に投影し，これらの画像列の平均値を畳み込み処理に入力として用いる．この手法では，歩容映像を画像 1 枚に時系列情報を圧縮しているため，動的な歩容特徴を排除している問題が存在する．一方，深度映像ベースの手法ではセンサからの距離値を映像内のテキスチャ情報として扱っており，点群投影方式によって球面投影 [12,14] と平行投影 [1,2] の 2 つに分類される．特に，平行投影ベースの手法において，我々は計測距離の変化に対応し映像内の深度値を正規化しており，固定の 2 視点の歩容特徴を用いる識別手法を提案してきた．しかしながら，実空間上に歩行者点群をマッピングする平行投影方式では、より全体の歩容輪郭を保つものの，極めて解像度が低下した場合に歩容形状の維持が難しくなり，識別性能が低下する．

[1] 九州大学
[2] NASA/Caltech JPL
[a)] ahn@irvs.ait.kyushu-u.ac.jp
[b)] k_nakashima@ait.kyushu-u.ac.jp
[c)] yoshino@irvs.ait.kyushu-u.ac.jp
[d)] yumi.iwashita@jpl.gov
[e)] kurazume@ait.kyushu-u.ac.jp

2.2 拡散モデルを使用した画像復元

拡散モデルを用いた復元手法では，主に problem-specific 手法と problem-agnostic 手法の 2 つに分けられる．problem-specific 手法 [9,10] では，綺麗な画像と劣化した画像の両方を拡散モデルの入力として用いる．この手法では，学習データに含まないノイズパターンに性能低下が生じることだけでなく，学習時に計算コストが飛躍的に上昇する問題が存在する．一方，problem-agnostic 手法 [7,8] では，学習済み DDPM を用いて生成ステップの中間表現を経由することで，劣化した画像を逆問題 $\mathbf{y} = \boldsymbol{H}\mathbf{x} + \mathbf{n}$ として解く．この中でも，疑似逆行列を介して観測画像 $\mathbf{y}$ と推定画像 $\mathbf{x}$ の一致度を誘導スコアとして定義する ΠGDM [8] は，観測画像がスパースな場合でも密な誘導スコアを計算が可能であり，他手法と比べてより高品質な画像復元に優れている．本研究では，点群投影のモダリティとして前述の平行投影方式 [1,2] を使用し，歩容映像内の欠損箇所を二値映像のマスクに変換し ΠGDM の疑似逆行列に用いる．また，歩容映像を入力として用いるため，映像向けの DDPM の一種である video diffusion model（VDM）[6] を ΠGDM の逆過程のパラメータとして学習させる．

3. 手法

3.1 学習済み VDM を用いた ΠGDM の歩容映像復元

本手法では，平行投影方式 [1,2] を用いて獲得した時系列の歩行者点群を深度映像 $\mathbf{y} \in \mathbb{R}^{F\times C\times H\times W}$ に変換する．また，歩容映像 $\mathbf{y}$ から欠損マスク $\boldsymbol{H} \in \mathbb{R}^{F\times C\times H\times W}$ に変換し，本研究で使用する ΠGDM の既知の劣化作用素として入力に用いる．

$$H^{(f,c,h,w)} = \begin{cases} 1, & \text{if} \quad y^{(f,c,h,w)} > 0, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

ただし，F はフレーム数，C はチャンネル数，H は縦の画素数，W は横の画素数である．この歩容映像 $\mathbf{y}$ は逆問題 $\mathbf{y} = \boldsymbol{H}\mathbf{x} + \mathbf{n}$ でのランダムノイズ $\mathbf{n}$ を含まないため，ΠGDM によって条件付き誘導項 $\nabla_{\mathbf{x}_t}\log(\mathbf{y}|\mathbf{x}_t)$ は以下の通りに近似できる．

$$\nabla_{\mathbf{x}_t}\log(\mathbf{y}|\mathbf{x}_t) \approx \sqrt{\alpha_t}((\boldsymbol{H}^{\dagger}\mathbf{y} - \boldsymbol{H}^{\dagger}\boldsymbol{H}\hat{\mathbf{x}}_t)^{\top}\frac{\partial\hat{\mathbf{x}}_t}{\partial\mathbf{x}_t})^{\top} \tag{2}$$

ここで，$\boldsymbol{H}^{\dagger}$ は欠損マスク $\boldsymbol{H}$ の疑似逆行列，α_t は各生成ステップ t のスケーリング係数を示す．この誘導項 $\nabla_{\mathbf{x}_t}\log(\mathbf{y}|\mathbf{x}_t)$ は，歩容映像を生成する際に DDIM [13] のサンプリングに学習済み VDM の事前分布項 $\log p(\mathbf{x}_t)$ と組み合わせられる．

3.2 分散マスクによる歩容輪郭のフィルタリング

観測の歩容映像 $\mathbf{y}$ では，歩行者内の点群欠損と背景の値が両方 0 となっているため，欠損マスク $\boldsymbol{H}$ における欠損部分と背景の区分できないため，映像内の背景部分にも歩容形状が生成される可能性を有する．この不確実性を抑制するため，先行研究 [11] と同様に生成された推定映像 $\mathbf{x} \in \mathbb{R}^{B\times F\times C\times H\times W}$ 中でバッチ数 B の分散 $\mathbf{x}_{\text{var}}$ を算出し，以下のように復元映像 $\tilde{\mathbf{x}}$ を求める．

$$\mathbf{x}_{\text{mean}} = \frac{1}{B}\sum_{b=1}^{B}\mathbf{x}^{(b)}, \tag{3}$$

$$\mathbf{x}_{\text{var}} = \frac{1}{B}\sum_{b=1}^{B}(\mathbf{x}^{(b)} - \mathbf{x}_{mean})^2, \tag{4}$$

$$\tilde{\mathbf{x}} = \boldsymbol{M}\cdot\mathbf{x}_{\text{mean}} \tag{5}$$

ただし，分散マスク $\boldsymbol{M} \in \mathbb{R}^{F\times C\times H\times W}$ は以下の通りに閾値 λ を用いて計算する．

$$M^{(f,c,h,w)} = \begin{cases} 0, & \text{if} \quad x_{\text{var}}^{(f,c,h,w)} > \lambda, \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

4. 実験

4.1 データセットと実験設定

本実験には，生成品質評価と識別性能評価の両方に SUSTeck1K データセットを使用した．このデータセットでは，垂直 128 ラインの高詳細な解像度を有る VLS-128 の LiDAR センサが使用され，1,050 人分の歩行者点群データが含まれる．また，各歩行者ごとに 8 種の歩行角度と 8 種のバリエーションの組み合わせがある．このデータセットを使用して ΠGDM の事前分布として用いられる VDM を学習する際，フレーム数 F を 10 と設定し、歩行者 1,050 人のうち 250 人分の平行投影方式からの綺麗な歩容映像を訓練に用いた．テスト時，学習データに含まれない被験者４０人分の歩容映像を使用した．提案手法の復元品質を評価する際に点群投影映像の劣化を再現するため，pepper ノイズと vertical 欠損とフレーム欠落の 3 種のマスクを組み合わせた．復元時，生成ステップ数は 100，分散マスクの閾値 λ は 0.15 と設定した．

LiDAR を用いた歩容認証の評価には Shen らの識別手法 [12] をベースラインとして用いた．ベースラインを学習する際，SUSTeck1K データセットの 250 人分の欠損なしの点群投影映像を学習に使用し，残りの 40 人分を評価に用いた．評価時には，特徴量空間のユークリッド距離に基づき，最近傍法（kNN）を用いて事前に保存した辞書データ（gallery）とクエリデータ（probe）の照合を行った．

4.2 生成品質評価

提案手法を用いて欠損した歩容映像を復元した結果と，その定量的評価を図 1-7 と表 1 にそれぞれ示す．生成品質の評価指標には Mean Squared Error（MSE）と Mean Structural Similarity Index（MSSIM）を用いた．表 1 の結果から，どちらのノイズマスクパターンでも提案手法が

表 1: 生成品質の定量的評価（%）.

pepper ノイズ	vertical 欠損	フレーム欠損	提案手法	MSE ↓	MSSIM ↑
0.2	1/2	0/10		0.013	0.660
			✓	0.003	0.964
0.4	2/3	0/10		0.017	0.461
			✓	0.003	0.965
0.4	2/3	2/10		0.018	0.438
			✓	0.003	0.958

欠損映像を高品質に復元していることがわかる．特に，元映像（図 1）と比べても復元された歩容映像（図 3，5，7）が時系列かつ深度情報の一貫性を保っていることが確認できる．

4.3 識別性能評価

表 2: 識別手法の精度評価（%）.

gallery	probe	提案手法	kNN	平均値
欠損なし	pepper ノイズ (0.2) ＋ vertical 欠損 (1/3)		Rank1	34.24
			Rank5	70.61
		✓	Rank1	44.09
			Rank5	87.69
欠損なし	pepper ノイズ (0.4) ＋ vertical 欠損 (2/3)		Rank1	23.52
			Rank5	50.49
		✓	Rank1	38.05
			Rank5	80.28

提案手法を用いた識別性能の定量的評価を表 2 に示す．ここで表 2 の平均値は，8 種の cross-view と 8 種の cross-variance からの識別精度の平均を計算したものである．表 2 の結果から，probe の歩容形状がスパースになった場合でも，提案の復元手法を用いることで識別性能が向上していることがわかる．

5. まとめと今後の予定

本稿では，拡散モデルに基づいた画像復元手法を LiDAR 点群の歩容映像に拡張することで，歩行者点群が欠損した場合でも識別性能の向上が可能であることを確認した．今後は，時系列の LiDAR 点群からの歩行角度を拡散モデルの条件として用いることで，極めてスパースな歩行者点群に対する高品質な映像復元の実現に取り組む．

謝辞

本研究では，JST 次世代研究者挑戦的研究プログラム JPMJSP2136，JSPS 科研費 JP20H00230 の助成を受けたものである．

参考文献

[1] Ahn, J., Nakashima, K., Yoshino, K., Iwashita, Y. and Kurazume, R.: 2V-Gait: Gait Recognition using 3D LiDAR Robust to Changes in Walking Direction and Measurement Distance, *Proceedings of the IEEE/SICE International Symposium on System Integration (SII)*, pp. 602–607 (2022).

[2] Ahn, J., Nakashima, K., Yoshino, K., Iwashita, Y. and Kurazume, R.: Learning Viewpoint-Invariant Features for LiDAR-Based Gait Recognition, *IEEE Access*, Vol. 11, pp. 129749–129762 (2023).

[3] Benedek, C., Gálai, B., Nagy, B. and Jankó, Z.: Lidar-Based Gait Analysis and Activity Recognition in a 4D Surveillance System, *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, Vol. 28, No. 1, pp. 101–113 (2018).

[4] Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y. and Yu, S.: OpenGait: Revisiting Gait Recognition Toward Better Practicality, *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9707–9716 (2023).

[5] Ho, J., Jain, A. and Abbeel, P.: Denoising Diffusion Probabilistic Models, *Advances in Neural Information Processing Systems*, Vol. 33, pp. 6840–6851 (2020).

[6] Ho, J., Jain, A. and Abbeel, P.: Video Diffusion Models (2022).

[7] Kawar, B., Elad, M., Ermon, S. and Song, J.: Denoising Diffusion Restoration Models, *arXiv* (2022).

[8] Kawar, B., Elad, M., Ermon, S. and Song, J.: Pseudoinverse-Guided Diffusion Models for Inverse Problems (2022).

[9] Saharia, C., Chan, W., Chang, H., Lee, C. A., Ho, J., Salimans, T., Fleet, D. J. and Norouzi, M.: Palette: Image-to-Image Diffusion Models, *arXiv* (2021).

[10] Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J. and Norouzi, M.: Image Super-Resolution via Iterative Refinement, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1867–1876 (2022).

[11] Shan, T., Wang, J., Chen, F., Szenher, P. and Englot, B.: Simulation-based lidar super-resolution for ground vehicles, Vol. 134, p. 103647 (2020).

[12] Shen, C., Chao, F., Wu, W., Wang, R., Huang, G. Q. and Yu, S.: LidarGait: Benchmarking 3D Gait Recognition with Point Clouds, pp. 1054–1063 (2023).

[13] Song, J., Meng, C. and Ermon, S.: Denoising Diffusion Implicit Models, *International Conference on Learning Representations* (2021).

[14] Yamada, H., Ahn, J., Mozos, O. M., Iwashita, Y. and Kurazume, R.: Gait-based person identification using 3D LiDAR and long short-term memory deep networks, *Advanced Robotics*, Vol. 34, No. 18, pp. 1201–1211 (2020).

[15] Ye, M., Yang, C., Stankovic, V., Stankovic, L. and Kerr, A.: Gait analysis using a single depth camera, *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 285–289 (2015).

図 1: Ground Truth

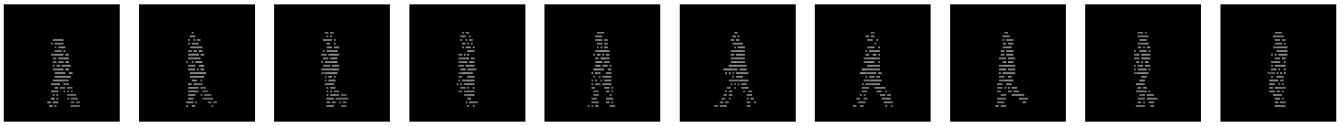

図 2: pepper ノイズ（0.2）+verical 欠損（1/2）

図 3: pepper ノイズ（0.2）+verical 欠損（1/2）+提案手法使用

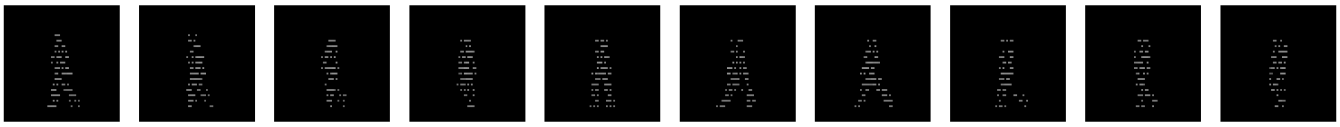

図 4: pepper ノイズ（0.4）+verical 欠損（2/3）

図 5: pepper ノイズ（0.4）+verical 欠損（2/3）+提案手法使用

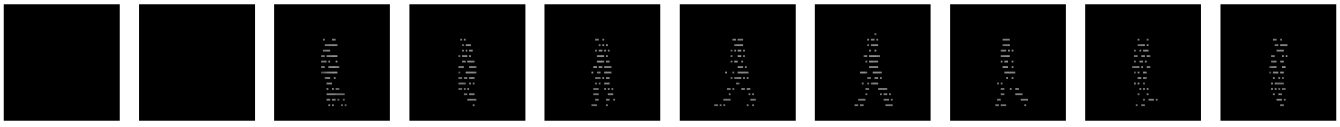

図 6: pepper ノイズ（0.4）+verical 欠損（2/3）+フレーム欠損（2/10）

図 7: pepper ノイズ（0.4）+verical 欠損（2/3）+フレーム欠損（2/10）+提案手法使用