

GAN を用いた顔の RGB 画像と奥行画像の同時生成

Simultaneous generation of RGB image and depth image of face using GAN

小林 良輔[†] 中島 悠太[†] 馬場口 登[†]
Kobayashi Ryosuke Nakashima Yuta Babaguchi Noboru

1. はじめに

プライバシー保護やエンターテインメントなどへの応用に向けて、近年、GAN による顔画像の生成が広く研究されている[1]。多くの応用では、実際に撮影された画像中の顔を生成した顔画像で置き換える必要があり、多様な顔の向きに対応するためには生成された顔の 3 次元形状が必要となる。そのため的手段として、Aaron らの提案する Volumetric Regression Network (VRN) [2]などの手法を利用して、生成した顔画像から 3 次元形状を推定することが考えられる。このような二段階の処理をスマートフォンなどの携帯端末で行うことを想定すると、処理時間が問題となる可能性がある。

本研究では、3 次元形状を奥行画像で表現するものとし、顔の RGB 画像と奥行画像を同時生成する手法を提案する。具体的には、顔画像データセットに対して Aaron らの手法[2]を用いて奥行画像を推定し、RGB 画像の 3 チャンネルに加えて第 4 のチャンネルとして奥行画像を推定する顔画像生成器を用意し、GAN により生成器を学習する。実験では、生成した顔画像を再び Aaron らの手法を用いて形状を推定し、生成した奥行画像との整合性を評価する。

2. 提案手法

多くの GAN による画像生成手法では R, G, B の 3 チャンネルの画像をデータセットとして学習する。提案手法ではこの 3 チャンネルに加えて奥行画像を 4 チャンネル目に追加して学習を行う。これにより、RGB 画像と同時に奥行画像も併せて生成する。本手法の概要を図 1 に示す。本手法はデータセットの作成とそれを用いた学習の 2 つのフェーズにより実現する。それぞれの詳細を 2.1 節、および 2.2 節で述べる。

2.1 データセットの作成

提案手法における GAN の学習では、顔の RGB 画像と奥行画像をデータセットとして用いる。まず、学習を安定させるため、顔の位置を顔の目や鼻などのランドマークを用いて位置合わせし、また正面顔の画像のみを選択する。次に、それらの顔の奥行画像を推定し、RGB 画像と対応する奥行画像のペアを 1 つのデータとするデータセットを作成する。以下、各ステップの詳細を述べる。

2.1.1 顔の位置合わせ

データセットは、として VGGFace2[3]を基に作成する。このデータベースでは顔画像と併せて目、鼻、口の両端、それぞれの両端の画像内での座標が提供

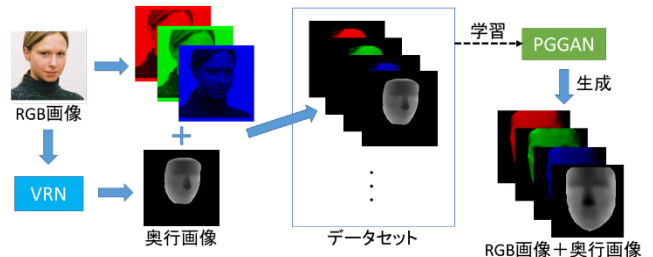


図 1: 提案手法の概要。

されている。これを用いて次の手順によりすべての顔画像の位置合わせを行う。

- ① 画像の回転: 左右の目の垂直方向の位置が等しくなるように画像を回転する。このとき、回転角 θ は以下の式で与えられる。

$$\theta = -\tan^{-1} \left(\frac{E_R^y - E_L^y}{E_R^x - E_L^x} \right)$$

ただし、 E_R^x, E_R^y, E_L^x および E_L^y はそれぞれ右目、左目の x 座標、 y 座標である（水平方向を x 、垂直方向を y とする）。

- ② 正面顔の選択: 正面顔において、鼻の水平方向の位置は左右の目の水平方向の位置の midpoint と概ね一致すると考える。そこで、以下の式を満たす顔画像のみをデータセットに含める画像として選択する。

$$\frac{|(E_R^x + E_L^x)/2 - N^x|}{|E_R^x - E_L^x|} < k$$

ただし、 N^x は鼻の x 座標、 k は閾値となる定数である。

- ③ 位置合わせ及びスケールリング: (2) で選出されたすべての画像の両目と鼻の位置の重心が一致するように、画像を平行移動する。加えて、すべての画像について、左右それぞれの目から鼻までの距離の合計が等しくなるようにスケール変換する。すなわち各画像に対するスケール係数 s を以下により求める。

$$s = \operatorname{argmin}_s (s \|E_R - N\|^2 + s \|E_L - N\|^2 - l)$$

ただし、 E_R, E_L および N はそれぞれ右目、左目、鼻の座標、 l はスケールの基準となる定数である。

2.1.2 奥行画像の推定

上記の手順で作成したデータセットの画像に VRN[2]を用いてそれぞれの奥行画像を推定する。次に推定した奥行画像を元の RGB 画像に 4 チャンネル目として加える。さらに、顔の生成には背景は不

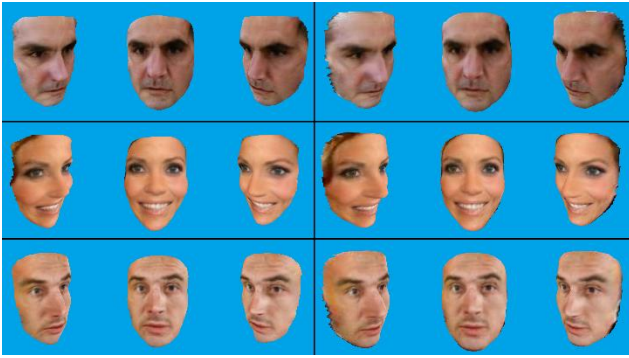


図 2: 生成した三次元モデルの例. 左が生成した奥行画像によるモデル, 右が VRN によって奥行を推定したモデル.

要であることから, 奥行画像に基づいて RGB 画像の顔以外の部分の画素は全て 0 (黒色) とした.

2.2 RGB 画像と奥行画像の同時学習

顔画像生成器の学習には Progressive Growing of GANs [4] (PGGAN) を用いる. データセットは 192 画素 × 192 画素のサイズの RGB 画像と奥行画像のセットを 61757 枚用いる. 入力に 100 次元の乱数とし, 出力として 128 画素 × 128 画素の RGB 画像と奥行画像が生成されるように学習を行う.

画像生成器により得られた奥行画像にはノイズが含まれることから, 3 次元モデルとしての自然さを重視し, 7×7 のボックスフィルタを適用する.

3. 評価実験

提案手法による生成結果を視覚的に検証するために, 生成した RGB 画像と奥行画像を利用してレンダリングした顔の例を図 2 に示す. このとき, 2.1.1 節で述べたデータセットの作成において, 顔の位置合わせの際の定数 k および l はそれぞれ $k = 1/15$, $l = 50$ とした. 比較のために, 生成した RGB 画像から VRN により奥行画像を推定したときの三次元モデルも併せて示す. 画像のサイズは 128 画素 × 128 画素である. 提案手法による奥行画像は, VRN によるものに比べて凹凸が抑制されていることがわかる. 原因の一つとしてボックスフィルタの適用が考えられる.

定量的に性能を評価するために, 生成された RGB 画像と奥行画像の整合性を, VRN によるものと比較する. 生成した RGB 画像から VRN を用いて奥行画像 D_{vrn} を推定し, それと提案手法により生成した奥行画像 D との SAD (Sum of Absolute Distance) を求め, それを画素数で割った値, すなわち 1 画素ごとの誤差の平均値 t を求める.

$$t = \frac{1}{A} \sum_i^A |D_i - D_{vrn}^i|$$

ただし, A は提案手法および VRN の推定で共通して奥行の値が得られた画素の数, D_i , D_{vrn}^i はそれぞれ

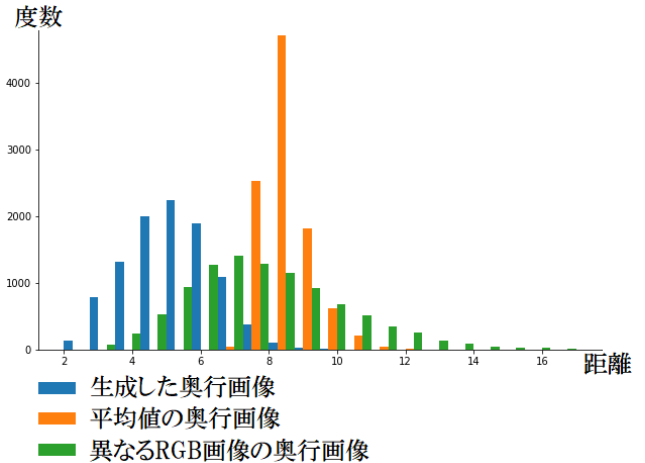


図 3: VRN により推定した奥行画像との比較.

れ D , D_{vrn} の i 番目の画素とする.

比較対象として, VRN による奥行画像の平均値 $M_{vrn} = \sum_i^A D_{vrn}^i / A$ で奥行の値が得られた画素の値をすべて置き換えた奥行画像 D_{ave} と, ランダムに選んだ 1 つの異なる RGB 画像と同時に生成された奥行画像 D_{other} を同様に検証する. 生成した約 10000 件の画像について距離のヒストグラムを生成したところ, 図 3 のような結果が得られた.

図 3 より, 生成した奥行画像は比較対象の二種類の奥行画像と比べて, VRN により生成した奥行画像との整合性が高いことがわかる.

4. おわりに

本稿では, GAN を用いて顔の RGB 画像と奥行画像を同時に生成する手法を提案した. 実験の結果, 視覚的に自然な顔の RGB 画像とそれに対応する奥行画像を生成できた. 今後は生成した顔画像を写真の顔に自然に置き換える手法を検討する.

本研究は JSPS 科研費 16H06302 の助成による.

参考文献

- [1] A. Radford et al., “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” arXiv:1511.06434, 2015
- [2] A. Jackson et al., “Large Pose 3D Face Reconstruction From a Single Image via Direct Volumetric CNN Regression,” IEEE International Conference on Computer Vision, 2017
- [3] Q. Cao et al., “Vggface2: A dataset for recognising faces across pose and age,” 2018 13th IEEE International Conference on Automatic Face Gesture Recognition, 2018
- [4] T. Karras et al., “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” arXiv:1710.10196, 2017

‡ 大阪大学
Osaka University