



# AR-GAN を用いた自然画像からの 奥行・ボケ効果の教師なし学習

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所  
金子卓弘

## 1. はじめに

ヒトは写真を見たとき、これまでに培った経験や知識などから、物体が奥行方向にどのように配置されているか、また、写真のどこにどのようなボケが生じているかなどの三次元世界に基づく情報（以下、「三次元情報」と呼ぶ）を推定することができる。一方、コンピュータにはそうした経験や知識がないため、写真から三次元情報を推定することは簡単ではない。

しかし、今後、ロボットが私たちの生活をサポートするようになったり、あるいは、仮想世界と現実世界の融合が進んでいったりした場合、コンピュータが三次元世界を理解し、現実世界と齟齬のない行動や情報提示ができるようになることは必要不可欠である。特に、私たちの最も身近にある記録・保存媒体は写真であることを考えると、写真のみから対応する三次元情報を推定できるようになれば、三次元情報を活用した様々な応用サービスの実現において有益であると考えられる。

このような写真から三次元情報を推定しようという研究は、コンピュータビジョンや機械学習などの分野で従来から活発に行われている。例えば、カメラと測距センサを用いて写真と三次元情報のペアデータを収集し、その対応関係を学習する手法<sup>1)</sup>や、ステレオカメラなどで集めた複数視点画像に対して幾何制約を考えることで三次元情報と写真の対応関係を学習する手法<sup>2)</sup>などが提案されている。

これらの手法の強みとしては、写真と三次元情報のペア、あるいは、画像のペアなどの強力な教師情報を用いて学習することができるため、学習が比較的容易で、高性能な三次元推定器（例えば、深度推定器やボケ合成器など）を獲得しやすい点が挙げられる。一方で、学習データを収集するためには測距センサやステレオカメラなどの専用機器が必要であり、データ収集コストが高い点が弱みとして挙げられる。

筆者らは、このような弱みを克服する方法として、教師なし学習の枠組みで、二次元画像から三次元情報を学習する方法について研究を進めている<sup>3,4)</sup>。具体的には、学習時には、普通のカメラで撮影した身近な画像や、ウェブ上で見かけるようなごく一般的な画像のみが利用可能であることを想定する。そして、この条件下で、画像のみから奥行情報や画像にどのようなボケ効果が付与されているかなどの三次元情報を学習できるようにすることをめざして研究を進めている。

このアプローチの強みは、学習に必要なデータは一般的な画像のみであるため、データ収集コストが低いことである。ここで留意したいことは、上述の複数視点画像を用いる手法とは異なり、当アプローチでは、いかなる画像ペアも必要ないということである。一方で、二次元画像のみから三次元情報を学習することは不良設定問題であるため簡単ではなく、挑戦的な課題である。この課題を緩和したことが、当研究の貢献であり、研究の肝でもある。

本稿では、まず第 2 章で、前提知識として当研究がベースとしている二つの技術、具体的には、Generative Adversarial Network (GAN)<sup>5)</sup>と、微分可能な絞りレンダラー<sup>6)</sup>について説明する。続いて、第 3 章では、教師なしで自然画像から奥行とボケ効果の学習を可能にした Aperture Rendering GAN (AR-GAN)<sup>3)</sup>について説明する。第 4 章では、AR-GAN の有効性を検証するために行われた実