

令和 2 年 6 月 12 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K12689

研究課題名（和文）ディープニューラルネットワークによる静止画像からの動画生成手法の開発

研究課題名（英文）Development of Video Generation Method from a Still Image using Deep Neural Networks

研究代表者

遠藤 結城（Endo, Yuki）

筑波大学・システム情報系・助教

研究者番号：00790396

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：動画は画像と比べて遥かに複雑な情報を含むことから、AIによる動画の理解・生成は画像以上に難しい問題として、数多くの研究者によって取り組まれてきた。今回開発した技術では、自然景観画像を対象に、複雑な変化を複数の単純な要素に分解して扱うことで、効率的な動画の学習を実現した。具体的には、雲のような細かい動きと昼夜のような全体的な色の変化を扱うニューラルネットワークを別々に学習させる。今回開発した技術により、インターネット上の景観画像から手軽に、従来技術よりも高品質な時間経過の動画を生成できるようになった。

研究成果の学術的意義や社会的意義

AIの代名詞とも言えるディープラーニングの進歩により、一枚の静止画像から未来の動きを予測することが可能となってきた。しかし現在の最先端の技術でさえ、出力となる動画の解像度が128×128画素程度という低解像度であったり、出力できるフレーム数が32程度とごく短かったりと、低品質な出力に留まっていた。今回開発した技術では、例えば640×360画素で1,000フレーム以上といった、従来技術よりも格段に高品質な動画を出力できるようになった。本技術は、エンターテインメント用途に加え、景観予測シミュレーションなどへの応用も期待される。

研究成果の概要（英文）：There have been many researches that tackle a problem of understanding and generating videos using AI, but this is still challenging because videos contain much more complicated information than images. Our technique can efficiently train videos by decomposing complicated components into simple ones. Specifically, we train two neural networks separately, each of which trains either motion such as moving cloud or appearance such as day-and-night variation. Our technique can easily create higher-quality videos from landscape images on the Web than prior work.

研究分野：コンピュータグラフィックス

キーワード：コンピュータグラフィックス 画像処理 深層学習 ニューラルネットワーク

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

我々人間が写真を眺めたとき、そのシーン中の物体や風景が、時間が経つにつれてどのように動き、変化していくかは、ある程度簡単に想像できる。例えば、湖や木々の映った風景は、風で水面が波打ち、枝葉が揺れる。早朝の街並みの空は朝日から夕焼け、そして夜空へと変化し、建物は街灯によって明るく照らされる。CG の分野においては、写真や絵画の静止画像にそのようなアニメーションの効果を付与する技術が古くから研究されている。しかし、実世界における物体や風景の挙動は膨大なパターンが存在するため、それらをコンピュータで扱えるようにモデリングし、再現することは、未だチャレンジングな課題である。

これまでの CG 分野における静止画像からの動画生成の取り組みは、主に二つのアプローチに分けられる。一つ目は、風に揺れる草花や水の流れなどの自然現象を、確率モデルを用いて表現するアプローチである。このアプローチは、特定の自然現象ごとに異なるモデルを作る必要があるため、その再現方法が確立されていない現象も多い。加えて、本物らしい挙動を再現するためにパラメータ設定に試行錯誤を要する。二つ目は、対象となる静止画像と類似した動画データを参照し、その動きを静止画像に自然に合成するアプローチである。このアプローチは、写実性の高い結果を生成できる一方で、対象画像と類似する参照動画をデータベースから検索し、その上で対象画像と参照動画との物体や背景領域を対応付ける必要があった。この工程の中には自動化が難しい処理も多く、それがユーザの大きな負担となっていた。さらに一つ目のアプローチと同様に、現象によって再現方法が異なるため、ユーザは複数の方法を使い分ける必要があった。

2. 研究の目的

本研究は、Web 上に膨大に存在する動画データを利用して、物体の挙動や風景の変化などを全自動でモデリングし、多様な静止画像に対して、その先の動画を予測・生成することを目的とする。提案手法では、畳み込みニューラルネットワークに基づく深層学習の枠組みで、動画のフレーム間の変化を効率的に捉えることを目指す。具体的には、昼夜による空や建物などの色味の変化や、水面の波などのテクスチャ模様の変化、人々の群衆や車の渋滞のような移動や回転を伴う挙動などを対象に、物体や景観変化の複雑性に応じて、それらに特化した動画生成モデルの実現可能性を検討する。そして、これら複数の要素を統合した複合的な動画生成手法を検討する。

3. 研究の方法

まず基礎となるニューラルネットワークのモデルを決定した上で、そのモデルをベースに手法の検討を進めた。当初の基本アイデアは、畳み込みニューラルネットワークを基にしたものである。畳み込みニューラルネットワークによる処理の過程で、時間変化のパターンが類似する各領域に対して、異なる役割を担う別のネットワーク層で扱うことを検討していた。この際、予測のための畳み込み層の設計や、最適化方法などの個別の課題が生じてくることが想定されていたため、まずはパターンの変化の複雑性が比較的単純なものに限定したうえで、これらの課題の解決方法を検討した。そのあとは、より複雑なパターンや複合的な現象を扱えるように、さらなるモデルの拡張の検討を進めた。大規模なデータセットを用いた実験を実施し、従来手法との正解画像との定量評価やユーザスタディを通して、提案手法の有効性を評価した。得られた成果は国際会議やプレスリリースなどを通じて発表した。

4. 研究成果

今回開発した技術では、深層ニューラルネットワークを用い、1枚の静止画像から高品質な動画を生成する。本研究は空の雲や湖の水の動き、1日の昼夜変化などを含む風景のシーンを対象に、Web上で比較的容易に手に入るタイムラプス動画を訓練データとして教師なし学習を行う。自然景観は、1日のうちに数秒単位で雲や水面が動くだけでなく、数分から数時間単位で夜になるにつれて色も変化していく。これらすべての要素を考慮するには長期間のフレームを小刻みに予測する必要があるが、予測できる長さには限界がある。そこでまず一つの工夫点は、動画中のアニメーションを動き（モーション）と外見の色（アピアランス）の変化の二つの要素に分解して考え、2種類のニューラルネットワークモデルにこれらを学習させることである。図1に示す通り本技術によって、動きと色の変化を複合的に扱うことができる。



図1: 本技術により静止画像から予測した動きと色の変化

二つ目の工夫点として、モーション用とアピアランス用の各畳み込みニューラルネットワーク（CNN）には動画を直接生成させず、入力画像を変換するためのモーションと色変換マップを予測させる。そしてこれらのマップに正則化をかけることでモデルの汎化性能の向上を図る。この際モデルの複雑さを減らし学習コストを削減するために、従来研究において主流であった3D畳み込み層を使うのではなく、2D畳み込み層を使い複数枚の画像を生成するアプローチを採用する。2D畳み込みによって計算された各マップにもとづき、入力画像にワーピングと色変換を適用することで、各フレームの出力を生成する。図2は動画を直接生成したベースラインと、本工夫により生成した提案手法によって得られる結果を比較したものである。アーティファクトによるちらつきが生じるベースラインに対して、提案手法はアーティファクトの少ない安定した結果を得られる。



図2: 動画を直接生成したベースラインと色変換マップを使って生成した提案手法との比較

最後に三つ目の工夫点は、各モデルに出力を制御するための潜在変数を教師なしで学習させることである。これにより入力画像に対する未来の不確定性に対処し学習の安定化を図るとともに、この潜在変数をベースにした少数のコントロールパラメータを用いることで一対多（one-to-many）の画像を生成できるようにする。ユーザはモーションとアピランスの二つの要素について、別々に結果をコントロールできる。さらに、移動方向や色の簡易的な指定にもとづいて潜在変数を最適化することで、ユーザは直感的に結果を制御できる。図3 に例を示す。ユーザが雲を動かしたい方向を粗く矢印で指定すると、指定された領域において矢印の方向に雲が動くという条件で、提案手法が潜在変数を最適化する。潜在変数は画像全体の動きに左右するため、図に示す流れ場のように、ユーザ入力によって異なる方向へ雲が移動する動画を生成できていることがわかる。

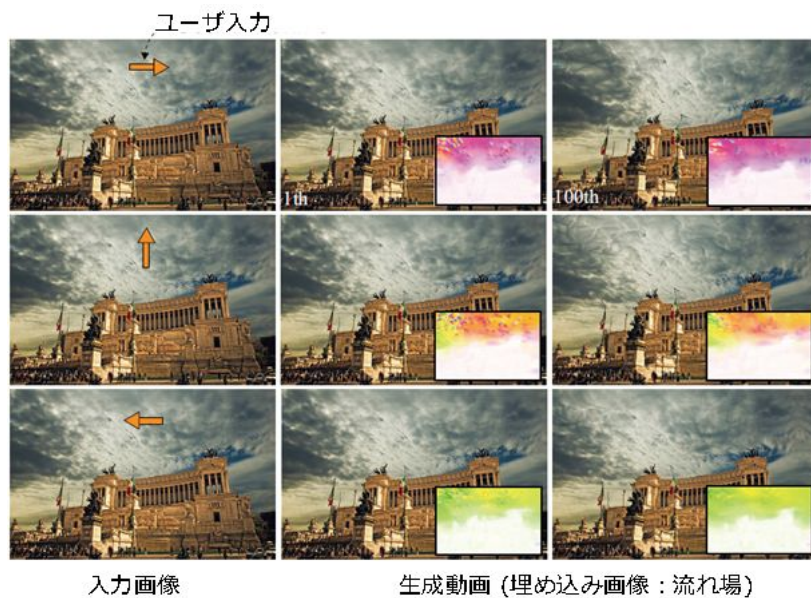


図3: ユーザ入力に基づく動画の生成例

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Endo Yuki, Kanamori Yoshihiro, Kuriyama Shigeru	4. 巻 38
2. 論文標題 Animating landscape	5. 発行年 2019年
3. 雑誌名 ACM Transactions on Graphics	6. 最初と最後の頁 1~19
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1145/3355089.3356523	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takazumi Kikuchi, Yuki Endo, Yoshihiro Kanamori, Taisuke Hashimoto, and Jun Mitani	4. 巻 -
2. 論文標題 Transferring pose and augmenting background for deep human-image parsing and its applications	5. 発行年 2018年
3. 雑誌名 Journal of Computational Visual Media	6. 最初と最後の頁 1-12
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/s41095-017-0098-0	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Yuki Endo, Yoshihiro Kanamori, Jun Mitani	4. 巻 36
2. 論文標題 Deep Reverse Tone Mapping	5. 発行年 2017年
3. 雑誌名 ACM Transactions on Graphics (Proc. of SIGGRAPH ASIA 2017)	6. 最初と最後の頁 177:1-177:10
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1145/3130800.3130834	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 1件 / うち国際学会 1件）

1. 発表者名 鬼塚慎吾, 遠藤結城, 栗山繁
2. 発表標題 流れ場に基づくワーピングによる深層生成モデルを用いた人物動画生成
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019年

1. 発表者名 遠藤 結城, 金森 由博, 栗山 繁
2. 発表標題 色と動き情報の学習による静止画像からのシネマグラフ生成
3. 学会等名 CVIM・CGVI・DCC 合同研究会
4. 発表年 2018年

1. 発表者名 遠藤 結城, 金森 由博, 三谷 純
2. 発表標題 Deep Reverse Tone Mapping
3. 学会等名 Visual Computing 2018 国際会議 / SIGGRAPH Asia 2017 採択論文 招待講演 (招待講演)
4. 発表年 2018年

1. 発表者名 Takazumi Kikuchi, Yuki Endo, Yoshihiro Kanamori, Jun Mitani
2. 発表標題 Transferring Pose and Augmenting Background Variation for Deep Human Image Parsing
3. 学会等名 Pacific Graphics 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 菊池 敬済, 遠藤 結城, 金森 由博, 橋本 泰輔, 三谷 純
2. 発表標題 ポーズ情報の考慮と背景データの拡張によるCNNを用いた人物画像の意味的領域分割
3. 学会等名 Visual Computing/グラフィクスとCAD合同シンポジウム 201
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

研究成果に関するWebページ
<http://www.cgg.cs.tsukuba.ac.jp/~endo/projects/AnimatingLandscape/>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----