

単一の全方位画像からの空の動き生成

角方 元紀 池林 ハキーム 河合 紀彦

大阪工業大学情報科学部 〒573-0196 大阪府枚方市北山 1-79-1

E-mail: norihiko.kawai@oit.ac.jp

あらまし 昨今、ネットワーク経由での展示を行うバーチャルミュージアムがある。バーチャルミュージアムでは、仮想空間を構築し、その中で様々なコンテンツを展示するが、コンテンツによってはそれに関連した現実世界を再現した仮想空間を展示場所とすることで、より展示物の魅力が引き立てられると考える。現実世界を再現した仮想空間の構築方法の一つとして、全方位画像によるパノラマ画像アプローチがあるが、Google Street Viewのように移動撮影して画像収集を行った場合には、ある地点においては静止画像が提示されるため、臨場感が高くないという問題がある。そこで本研究では、全方位画像の空の領域に動きを持たせ全方位動画を生成することで、より臨場感の高い現実世界を再現した仮想空間を構築する手法を提案する。

キーワード 全方位画像・動画生成

Sky Motion Generation from a Single Omnidirectional Image

Motoki KAKUHO Hakim Ikebayashi and Norihiko Kawai

Faculty of Information Science, Osaka Institute of Technology, 1-79-1 Kitayama, Hirakata, Osaka 573-0196, Japan

E-mail: norihiko.kawai@oit.ac.jp

Abstract There have been virtual museums that exhibit via networks. In virtual museums, various contents are exhibited in a virtual space. One of the methods that construct a virtual space imitating the real world is the panoramic image approach using omnidirectional images. In this approach, when images are captured while moving, such as Google Street View, a static image is presented at certain point, which does not give a sense of realism. In this study, we propose a method to construct a virtual space that reproduces the real world with a higher sense of realism by adding motion to the sky area in the omnidirectional image and generating an omnidirectional video.

Keyword Panoramic Image, Video Generation

1. はじめに

昨今、ネットワーク経由での展示を行うバーチャルミュージアム^[1]がある。バーチャルミュージアムでは、仮想空間を構築し、その中で美術品や遺物をはじめ様々なコンテンツを展示するが、コンテンツによってはそれに関連した現実世界を再現した仮想空間を展示場所とすることで、より展示物の魅力が引き立てられると考える。現実世界を再現した仮想空間の構築方法の一つとして、全方位画像によるパノラマ画像アプローチがあるが、Google Street View^[2]のように移動撮影して画像収集を行った場合には、ある地点においては静止画像が提示されるため、臨場感が高くないという問題がある。そこで本研究では、全方位画像の空の領域に動きを持たせ全方位動画を生成することで、より臨場感の高い現実世界を模倣した仮想空間を構築する手法を提案する。

単一画像からの動画生成を行う従来手法として、一枚の透視投影画像の風景画像から畳み込みニューラルネットワークを使用して空や川の動きを再現した動画を生成する Animating Landscape^[3]が提案されている。この研究では通常の透視投影画像を学習データとして使用しているため、公開されているモデルを正距円筒図法により展開された全方位画像に適用した場合にも、透視投影画像であることを前提とした動きが再現される。このため、空方向を透視投影画像として見た場合には、違和感のある空の動きになるという問題が発生する。また、パラメータに大きく依存し、通常動きがない領域でも動きが発生する問題がある。

これに対して、提案手法では、従来のニューラルネットワークを用いる手法とは異なり、3次元空間で動きをモデル化して、それを2次元画像上での動きに変換する。また、セマンティックセグメンテーション^[4]

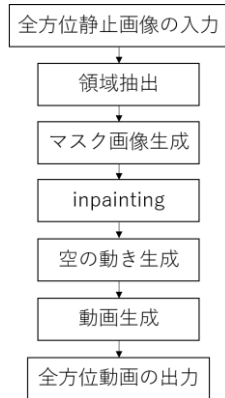


図 1 提案手法のフローチャート



図 2 入力画像例

による領域分割を行い、対象領域以外は動かないようにすることで、全方位画像上の空領域のみに動きを再現した全方位画像を生成する。

2. 提案手法

2.1. 概要

提案手法のフローチャートを図 1 に示す。提案手法では、一枚の空を含む風景画像を入力し、まず図 2 に示すような正距円筒図法で展開された全方位画像にセマンティックセグメンテーション^[4]を用いて領域分割を行い、マスク画像を生成する。その後、マスク画像を用いて、地平線より上部の空領域でない領域に対して Inpainting^[5]を行い、地平線より上部が全て空となる画像を生成する。次に、その画像の上半分の領域に対して動きを加え、空領域以外の部分を静止画の入力画像に置き換えることで、空領域のみに動きがある動画画像を生成する。以下、各処理について詳述する。

2.2. 領域抽出

領域抽出では、全方位画像にニューラルネットワークを用いたセマンティックセグメンテーションを適用して、空・建物・木などの領域を抽出する。図 2 に手法^[3]を適用した結果である図 3 では、空や建造物がなどの物体が青・緑・赤などの色に領域分割されている。この領域抽出ではそれぞれの画素に領域を示すラベルが付与されている。

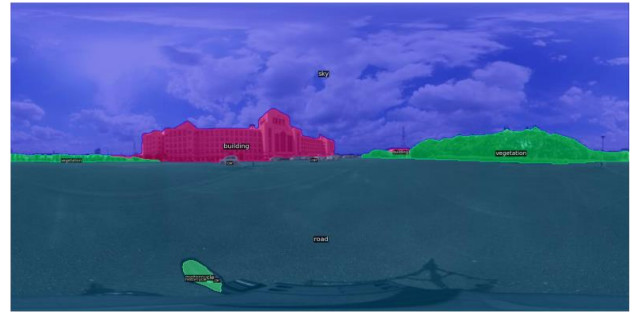


図 3 領域抽出の結果



図 4 inpainting 用のマスク画像

2.3. マスク画像の生成

次に、以降の処理である Inpainting および動領域と静止領域の合成で用いるマスク画像を生成する。まず、Inpainting を行うために図 3 の領域抽出の結果を用いて空領域と地面の領域の間にある建物と木の領域のマスク画像を生成する。ここでは、セマンティックセグメンテーションで得られるマスク領域を膨張させることで、空以外の物体領域を高い確率で消去対象となるマスク領域に含めるようにする。図 4 に Inpainting 用の膨張後のマスク画像例を示す。

また、最後の動領域と静止領域を合成するためのマスク画像も生成する。静止領域について、セマンティックセグメンテーションの結果をそのまま使用した場合、境界領域での精度の不十分さから合成時に建物や木が欠損する可能性がある。このため、空以外の静止画像の領域を少し膨張させてマスク画像を生成する。図 5 は静止領域を膨張させたマスク例である。ただし、図 5 のように膨張させると欠損は防ぐことができるが、空領域の本来動く場所が動かなくなる問題が発生する。このため、図 6 の膨張で得られた赤色で示す動領域と静止領域の境界領域に対して、生成した動画画像と入力画像の画素値を混ぜ合わせる Alpha blending を施すことで違和感を低減する。ここでは、動領域に近いほどに生成した動画画像の画素値の割合を大きく、静止領域に近いほど入力画像の画素値の割合を大きくしてブレンディングを行う。



図 5 静止領域のマスク画像

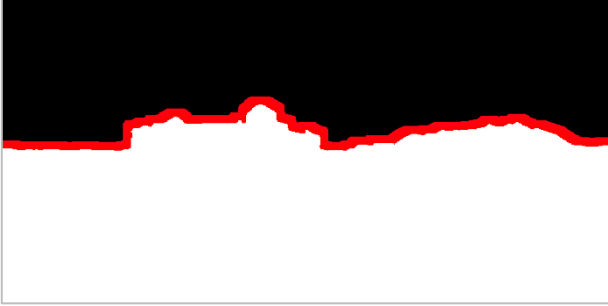


図 6 alpha blending のマスク画像



図 7 inpainting の結果

2.3. Inpainting による画像の生成

後述する動画生成時において、オプティカルフローで対応付けられる画素が建物などの物体上の画素になる場合がある。このため、物体の背景となる空のテクスチャを Inpainting により生成しておき、その画素値を利用して動画を生成する。Inpainting の手法として、機械学習を用いた Yu らの手法[4]を用いて図 4 に示すマスク領域を修復する。図 2 の入力画像例に図 4 のマスク画像の白領域を欠損領域として修復を行った結果例を図 7 に示す。

2.4. 空の動きの算出

空の動きについては、空の雲が三次元空間での上空での平面上を真っすぐ動くという仮定を用いて、その 3 次元の動きを全方位画像上での 2 次元オプティカルフローとして表現する。

具体的には、全方位画像は球の中心に向かって投射される光線の集合であるため、球表面の 3 次元位置に

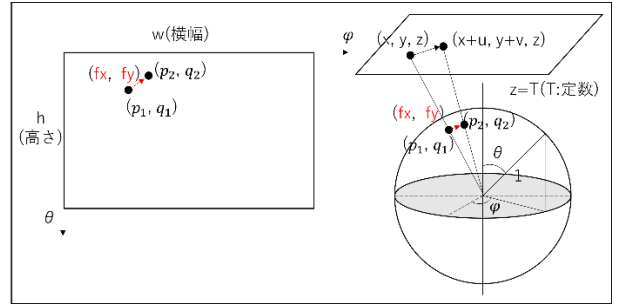


図 8 球と平面の 3 次元位置の関係

対応する全方位画像上の画素と雲の位置を表す平面上の 3 次元座標との画素の関係を求める。ここではまず、図 8 に示す z 軸と原点及び球表面上の点を結ぶ線分間の角度 θ_1 と z 軸周りの角度 φ_1 は、全方位画像における画素の座標 (p_1, q_1) と画像の縦横の画素数 h, w を用いて以下の式(1)(2)によって算出する。

$$\theta_1 = \frac{\pi q_1}{h} \quad (1)$$

$$\varphi_1 = \frac{2\pi p_1}{w} \quad (2)$$

次に、求めた 2 つの角度 $\theta_1 \cdot \varphi_1$ で表される球の表面を通る光線が通過する空を表す平面上の点の 3 次元座標 (x, y, z) を求める。ここでは、球の半径を 1 とし、空の高さを表す z 座標を T (定数) とすると、平面上の 3 次元座標は以下のように算出される。

$$x = T \tan \theta \cos \varphi \quad (3)$$

$$y = T \tan \theta \sin \varphi \quad (4)$$

$$z = T \quad (5)$$

次に、求めた平面上の点の x, y 座標を、図 8 に示すようにそれぞれある一定距離 u, v 移動させ、平面上の点と球の原点を結ぶ光線上にある全方位画像の画素に変換する。移動させた点 $(x+u, y+v, z)$ に対応する球表面上の点を表す角度 $\theta_2 \cdot \varphi_2$ は以下のように算出できる。

$$\theta_2 = \cos^{-1} \frac{z}{\sqrt{(x+u)^2 + (y+v)^2 + z^2}} \quad (6)$$

$$\varphi_2 = \tan^{-1} \frac{y+v}{x+u} \quad (7)$$

さらに、求めた θ_2 と φ_2 と、画像の縦横の画素数 h, w により、移動させた点に対応する画素の座標 (p_2, q_2) は以下のように求められる。

$$p_2 = \frac{\varphi_2 w}{2\pi} \quad (8)$$

$$q_2 = \frac{\theta_2 h}{\pi} \quad (9)$$

最後に変換された画素(p_2, q_2)から最初の変換される前の画素(p_1, q_1)の差を計算することでその画素におけるオプティカルフロー(f_x, f_y)を算出する.

$$f_x = p_2 - p_1 \quad (10)$$

$$f_y = q_2 - q_1 \quad (11)$$

以上の処理を画像の全画素に対して行うことで全画素のオプティカルフローを算出する.

2.5. 動画生成

動画生成では、まず、図7に示す Inpainting を用いて生成された空と地面だけの画像の空領域を含む上半分の領域に対して作成されたオプティカルフローを適用し、1枚の移動後の画像を作る。その画像に対して図6のマスク画像の白部分を入力画像の対応する画素値に、黒部分に移動後画像の画素値を合成する。また、赤領域については、2.3節で述べた通り、動領域からの距離に基づき、移動後画像と入力画像の画素値をアルファブレンドし、動画の1フレームを生成する。

この処理を空を表す平面上の移動量を変えながら2フレーム目以降を生成する。ここでは、空は一定速度で動いているものと仮定し、1フレーム目から*i*フレーム目までの平面上の移動量(u_i, v_i)を以下の(12)(13)の式により決定する。

$$u_i = i \times G \quad (12)$$

$$v_i = i \times H \quad (13)$$

ただし、 G, H はそれぞれ空を表す3次元平面上の*x*方向、*y*方向の移動の大きさを決定するパラメータである。これを指定のフレーム数繰り返し行い、それらの画像を時間的に繋ぎ合わせることで空が一定方向に動く動画を生成する。

3. 実験と考察

3.1. 実験内容

空の動きを生成する実験を行った。実験対象として全方位カメラ(RICOH THETA Z1)を用いて撮影し、1600×800の解像度にリサイズした正距円筒図法で展開された全方位画像である図2の画像を使用した。なお本実験では、定数*T*を2、画素ごとの移動量(u_i, v_i)は*G*を0.0015、*H*を0とする*x*方向に動かすものと、*G*を0、*H*を0.0015とする*y*方向に動かすものの2通りの180フレームの動画を生成した。また、マスク

画像の膨張は5回繰り返し、Alpha blendingする境界領域の幅は20画素とした。また、これらの結果を、透視投影画像に対してニューラルネットワークで動きを生成する手法[3]を用いて生成した動画と比較した。

3.2. 実験結果

生成した全方位画像の*x*方向と*y*方向それぞれ1, 60, 120, 180フレーム移動した画像を図9と図10に示す。また、それらを特定の方向の透視投影画像に変換した結果を図11と図12に示す。実験結果から空領域が回転することなく一方向に向かって真っすぐに動くようになっていることが分かる。また、真上の雲は遠くにある雲よりも早く動くようになっていることから遠近感が出ていることが分かる。また、空領域以外の領域が止まっており、空と建物などの境界領域に対しても違和感なく動きが付与されていることが確認できる。

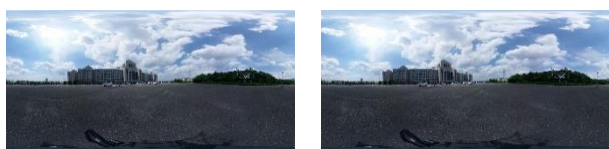
次に従来手法との比較を行う。提案手法と従来手法の結果画像の90フレーム目と180フレーム目を図13に示す。また、それらをそれぞれ異なる特定の方向の透視投影画像に変換した結果を図14と図15に示す。図14より、実験結果から従来手法では、建物の輪郭が崩れて雲の形が不自然に歪んでいるのに対して提案手法では、輪郭が保たれ雲も違和感なく一方向に動いていることが分かる。また、図15は全方位画像の左端と右端の方向の透視投影画像であるが、実験結果から従来手法では、画像の両端の境界部分で動きが異なり不自然な線や歪みができているのに対して提案手法では、境界部分に線や歪みが形成されることなく動画が生成されていることが確認できる。これらのことから、通常動きがない領域が動くという問題と空の違和感のある動きになるという問題が解消されていることが分かる。

4. まとめ

本研究では、単一の全方位画像から空の動きのある動画の生成を行う手法を提案した。実験を通して提案手法が有効的であることを確認した。現在の実験では空領域にとどまっているが、水や木などに動きを加えることでさらに臨場感の高い動画の生成が行えると考えられる。

謝辞

本研究の一部は、JSPS 科研費 JP21H03483 の助成を受けて実施した。



Frame 1

Frame 60



Frame 120

Frame 180

図 9 x 方向の動きを生成した実験結果



Frame 90 (従来手法)



Frame 180 (従来手法)

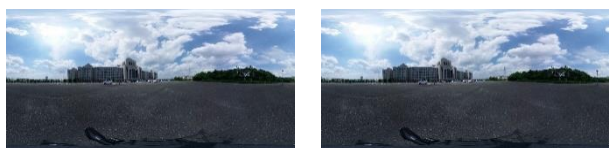


Frame 90 (提案手法)



Frame 180 (提案手法)

図 13 従来手法と提案手法の比較



Frame 1

Frame 60



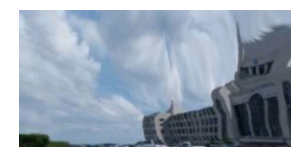
Frame 120

Frame 180

図 10 y 方向の動きを生成した実験結果



Frame 90 (従来手法)



Frame 180 (従来手法)

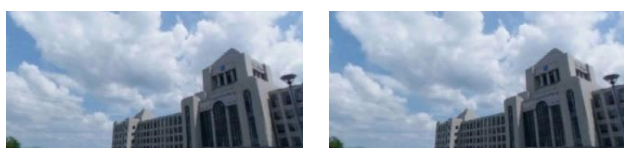


Frame 90 (提案手法)



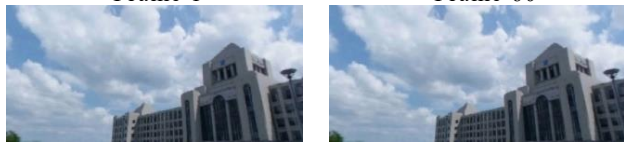
Frame 180 (提案手法)

図 14 従来手法と提案手法の比較を透視投影画像として表示した場合



Frame 1

Frame 60



Frame 120

Frame 180

図 11 x 方向の動きを生成した結果を透視投影画像として表示した場合



Frame 90 (従来手法)



Frame 180 (従来手法)

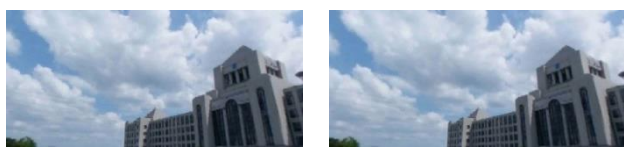


Frame 90 (提案手法)



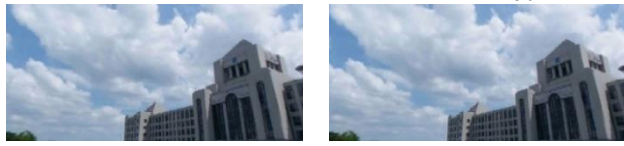
Frame 180 (提案手法)

図 15 従来手法と提案手法の比較を透視投影画像として表示した場合



Frame 1

Frame 60



Frame 120

Frame 180

図 12 y 方向の動きを生成した結果を透視投影画像として表示した場合

- [2] Google Street View, <https://www.google.co.jp/maps> (参照 9 月 30 日)
- [3] Yuki Endo, Yoshihiro Kanamori, Shigeru Kuriyama, "Animating Landscape: Self-Supervised Learning of Decoupled Motion and Appearance for Single-Image Video Synthesis", ACM Transactions on Graphics, Vol.38, No.6, pp.175:1-175:19, 2019.
- [4] John Lambert, Zhuang Lie, Ozan Sener, James Hays, Vladlen Koltun, "MSeg : A Composite Dataset for Multi-domain Semantic Segmentation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [5] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S. Huang, "Free-Form Image Inpainting With Gated Convolution", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4471-4480, 2019.

文 献

- [1] Stlaiou Styliani, Liarokapis Fotis, Kotsakis Kostas, Patias Petros, "Virtual Museums, a Survey and Some Issues for Consideration", Journal of Cultural Heritage, Vol.10, No.4, pp.520-528, 2009.