H-001

手持ち撮影した RGB-D 映像中の動的領域の切り出し
Image Segmentation in a Handheld RGB-D Movie

柚木 玲士† 北原 格† 大田 友一†
Reiji Yunoki Itaru Kitahara Yuichi Ohta

1. はじめに

物体を複数方向から撮影した画像群を計算機内部で統合することでその３次元モデルを復元し，任意の方向からの観察を可能にする Image-Based Modeling に関する研究が盛んに行われている[1]．生成された３次元モデルは，物体形状認識[2]，運動シミュレーション[3]など様々な分野での応用が考えられる．Structure from Motion（以下 SfM）は，１台のカメラで物体の周りを移動しながら撮影した映像を用いて，被写体の３次元形状とカメラ移動を同時に推定することができるが，撮影対象が静止物体に限定される[4]．また，被写体の表面に画像特徴が乏しい場合，推定される３次元形状精度が低下するという問題も存在する．Kinect Fusion は，カラー（RGB）と奥行き（D）を有する RGB-D 画像を用いることで，この問題を解決しているが，被写体は静止物体に限定される．同期撮影可能な複数の固定カメラで撮影した多視点画像を用いることにより，動的な物体の３次元形状の復元が可能であるが[1]，撮影装置が大掛かりになる．

我々は，１台カメラを用いて動的な物体の３次元モデルを生成することを目的とした研究に取り組んでいる．上述したように，静止物体であれば，従来研究を用いることで，３次元モデルの生成が可能である．動的な領域の３次元モデルを復元するためには，撮影映像において，動的な領域と静的な領域を分離し，別途３次元形状推定処理を施す必要がある．

本研究では，３次元モデル生成処理の前半に着目し，手持ちの RGB-D カメラ１台を用いて，動的な被写体の周囲を移動しながら撮影した映像から，静的領域と動的領域を分割する手法について述べる．

2. 手持ち撮影した RGB-D 映像における動的領域分割手法

図１に示すように，提案手法は大きく二つのステップに分かれる．一つ目は Kinect Fusion による静的領域の３次元モデル生成である．二つ目は，生成した静的領域の３次元モデルを用いた動的領域の切り出しである．

先述したように，RGB-D 映像に対して Kinect Fusion を適用した結果生成される３次元モデルは，静的領域に限定される．図２に示す復元された３次元モデルでは，撮影中に被写体が肘から先を動かしたため，その領域が欠落している．ある視点において RGB-D カメラで撮影した深度画像と，同一視点に仮想カメラを設置し，復元された３次元モデルをレンダリングする際のデプスマップ（z-buffer）を比較すると，動的な領域で大きな差異が発生する．本手法ではこの特徴に着目し，深度画像と z-buffer の差分画像に対し，適切な閾値処理を施すことにより，動的領域の切り出しを実現する．

† 筑波大学，University of Tsukuba

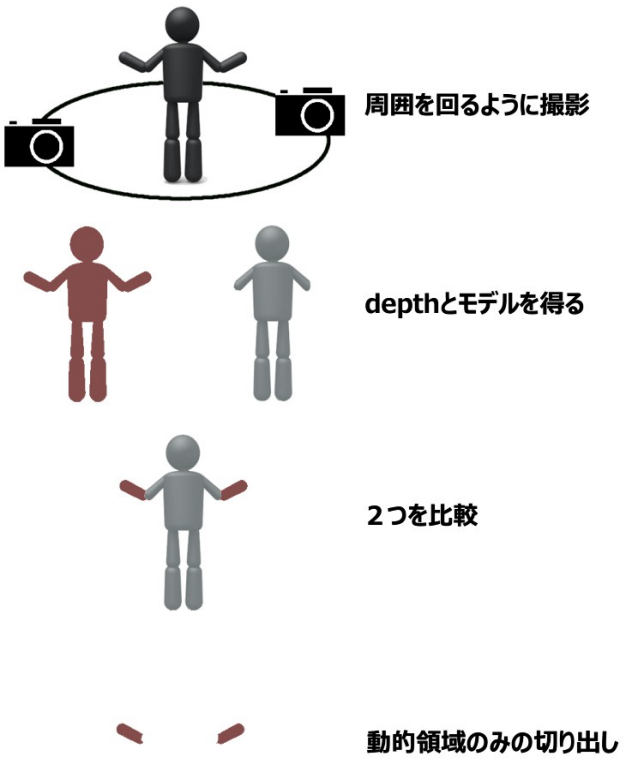

図１ 提案手法のおおまかな流れ

3. 各ステップの詳細

3.1 RGB-D 映像の撮影と Kinect Fusion による静的領域の３次元復元

本研究では，撮影機材として RGB-D カメラ（Kinect）を１台使用する．RGB-D カメラは，人間の手によって把持されている．被写体全周の３次元モデルを生成するために，対象物体を中心に，周囲を取り囲むように移動しながら RGB-D 映像を撮影する．これらの画像群に対して Kinect Fusion を適用し，３次元モデルを生成する．生成される３次元モデルの一例を図２に示す．この図に示すように，被写体の腕から先の３次元モデルが欠損している．Kinect Fusion では，映像のフレーム間で３次元的整合性が取れない動的領域は復元対象から除外されるためである．裏を返すと，静的領域のみを３次元モデル化することができていると考えることができる．

3.2　深度画像と静的領域の z-buffer の比較による動的領域の分割

前節で生成した静的領域モデルを，撮影映像中のあるフレームにおけるカメラ位置に設置した仮想カメラで撮影（レンダリング）を行い，それに伴い生成される z-buffer と，実際に撮影した深度画像の差分画像を生成する．３次元モデルが生成されている静的領域では，差分値が 0 に近い値となる．一方，深度画像より z-buffer の深度が大きくなる領域は，実際には物体が存在するが，動きによってモデル化されなかったと考えられる．このような領域を切り出すことで，動的領域の分割を実現する．

4.　実験

本研究では，実際に撮影された深度画像と同一視点において３次元モデルをレンダリングする際の z-buffer を比較するため，可能な限り精度の高いモデルが必要である．そこで Kinect Fusion によるモデル生成については，Point Cloud Library の kinfu よりも高精度なモデル生成が可能な Microsoft Kinect SDK を用いる．RGB 画像と深度情報は処理可能な最高速のフレームレートで撮影され，それらに対して Kinect Fusion が適用され，３次元モデルが更新される．撮影開始直後の３次元モデルの復元精度は高くないが，一定時間撮影を継続することにより，精度が向上する．

実験では肘から先のみを動かしている人間をおよそ 1m 離れた位置から手持ちのカメラで 10 秒間程度撮影し，肘から先の情報が欠落した３次元モデルを得た．

RGB-D 映像を撮影したカメラの位置姿勢の情報を用いて，同一の位置姿勢に仮想カメラを設置し，３次元モデルをレンダリングする，その際，仮想カメラから３次元モデルまでの距離マップを z-buffer として獲得する．図３のように深度画像と z-buffer を比較することで差分マップを生成し，差分値により動的領域を切り出す．

5.　おわりに

本研究では，手持ちの RGB-D カメラで撮影した映像シーケンスから動的領域を切り出す手法について述べた．撮影の準備に手間を掛けることがなく，また一般的な環境での撮影が可能となっており，物体の動作状況を手軽に取得することができる．将来的には，この手法で分割した動的領域情報を元に 1 台の RGB-D カメラによる撮影映像から動的領域を含めた３次元モデルを生成することを目指す．

図２　動的領域が欠落した３次元モデル

図３　モデルと深度の比較例

参考文献

[1]Benjamin Petit, Jean-Denis Lesage, Clément Menier, Jérémie Allard, Jean-Sébastien Franco, Bruno Raffin, Edmond Boyer, and François Faure, "Multicamera Real-Time 3D Modeling for Telepresence and Remote Collaboration", International Journal of Digital Multimedia Broadcasting, Volume 2010 (2010)

[2]Blanz, V., Vetter, T., "Face recognition based on fitting a 3D morphable model" , Pattern Analysis and Machine Intelligence, Volume.25 (2003)

[3]Bing Chen, Yu Guang Fan, "3D Modeling and Open-Close Motion Simulation of the Triple Eccentric Butterfly Valve", Advanced Materials Research, Volume.215 (2011)

[4]山下 淳, 原田 知明, 金子 透: "全方位カメラ搭載移動ロボットによる Structure from Motion を用いた 3 次元環境モデリング", 日本機械学会論文集, Vol.73 (2007)