

距離画像センサを用いた空間再現手法の一検討

A Study on Space Reproduction Method using Depth sensors

三功 浩嗣
Hiroshi Sankoh

内藤 整十
Sei Naito

1. まえがき

近年、視聴者が任意に選択する視点に合わせて3次元の空間を表示することが可能な自由視点映像の合成技術が注目されている[1][2]。筆者らは特に、人物等のオブジェクトが複数存在するシーンを対象に、通常はカメラを配置不可能な3次元空間中に入り込み、オブジェクト群の間を通り抜けるウォークスルーや、高さを自在に制御可能なフライスルーと呼ばれる視聴体験を映像合成によって実現することを目指してきた[2]。

自由視点映像の代表的な合成方式として、モデルベース方式[3][4]と画像ベース方式[5]の二つがあるが、後者の方式は主にカメラ間の内挿画像の合成を目的としたものであり、上記アプリケーションの実現を考えた場合は、一般的に前者の方式が用いられる。モデルベースの合成方式では、オブジェクトの3次元形状モデルを復元し、視聴者が選択する視点位置に応じて、多視点から得られる実写のテクスチャを適切な割合でマッピングすることで、任意の視点を再現することが可能である。最終的に生成される自由視点映像の合成精度は、オブジェクト3次元形状モデルの復元精度に大きく依存するため、高精度なモデル復元が重要となる。

複数カメラから得られる画像情報のみを用いてオブジェクト3次元形状モデルを復元する受動型計測の代表的な手法として、ステレオ法や、視体積交差法[3]が提案されている。これらの手法は、大量のカメラが使用可能である等、特定の撮影条件下における有効性が示されているものの、各手法の性能は、カメラ間の対応点マッチング精度、および各視点におけるオブジェクト抽出精度に大きく依存する。そのため、クロマキー等の特殊な撮影環境を持たない、一般的な撮影環境下においては高精度な復元には課題があった。一方、光やセンサ等をオブジェクトに照射し、伝搬時間の変位等を計測する能動型計測では、撮影カメラ以外に特殊なデバイスが必要であるものの、受動型計測に比べて高精度な3次元形状モデル復元が可能であることが知られている。特に、近年では、Microsoft社のKinect等、安価で使用可能な距離画像センサが急速に普及しており、カメラ画像と併せて使用することで、3次元形状モデルの高精度化が期待される。

以上の課題、およびデバイスの普及状況を踏まえ、本稿では、限られた台数の距離画像センサからオブジェクトの3次元形状モデルを復元し、自由視点映像の合成を実現する手法について提案する。

2. 関連研究

限られた台数のカメラのみを用いて自由視点映像を合成する手法として、主にサッカー映像を対象としたオブジェクトの単純化3次元モデルを用いる自由視点の合成方

† (株) KDDI 研究所

式が提案されている[6]。これらの手法では、3次元空間中の各オブジェクトをビルボードと呼ばれる一枚の長方形ポリゴンで近似し、ユーザが選択する視点位置に応じて、多視点画像から獲得したテクスチャ情報を適切にマッピングすることで、計算コストを抑えた合成を実現している。フィールド平面上に存在する各選手の3次元世界座標に関しては、フィールド平面上2次元世界座標として扱うことが可能であり、各カメラで取得されるテクスチャ領域の足下点を平面射影行列により投影することで算出可能であるため、カメラの強校正は不要である。また、オブジェクトの3次元空間における位置情報が算出されるため、カメラ間の内挿箇所だけでなく、フィールド内の仮想視点合成も違和感なく再現可能である。

一方で、ビルボード方式に基づく自由視点映像の合成では、ユーザが指定する仮想視点の最寄りの撮影カメラで取得されるオブジェクト毎のテクスチャをそのままマッピングする処理を基本とする。そのため、視点変化に伴う見え方の微小変化を再現できないという課題がある。また、各カメラ映像中で複数のオブジェクトが重なりあう、または一つのオブジェクトに含まれる腕や顔等のパーツどうしが重なりあうオクルージョンという現象が発生する場合、当該オクルージョン領域の形状を正確に復元することが困難である。そのため、合成画質が低下するという課題がある。特に、手前のオブジェクトに遮られる奥のオブジェクトのテクスチャに欠損が生じることを避けられないため、当該オブジェクトの正確なレンダリングは不可能であり、結果として合成精度が著しく低下するという課題がある。

3. 提案手法

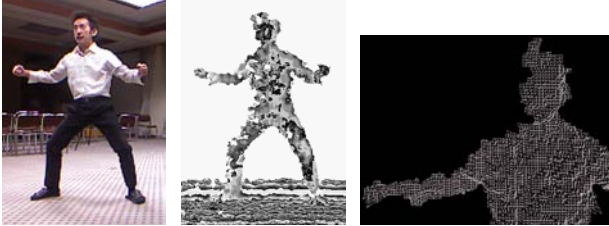
2章で述べたビルボード方式の課題を踏まえ、本稿では、距離画像センサを用いて、カメラ画像に加えて、同一視点における奥行き情報（以下、深度マップと表記）を活用した自由視点映像の生成手法について提案する。具体的には、距離画像センサで得られる画素毎の奥行き値に基づき深度マップをポリゴン化することで3次元形状モデルを復元する。さらに、当該3次元形状モデル、および画像情報との適切な対応付けに基づき、カメラが存在しない仮想視点における見え方をレンダリングする手法について述べる。提案手法の処理手順は以下の通りである。

1. 同期済みのRGB画像、および深度マップを取得
 2. RGB画像と深度マップ間の画素位置合わせを実施
 3. 画素位置合わせ済みの深度マップをポリゴン化することでオブジェクト3次元形状モデルを復元
 4. 各ポリゴンのテクスチャをRGB画像より取得
 5. ポリゴン情報に基づき仮想視点をレンダリング
- 以下、各処理の詳細について述べる。

距離画像センサにおいて、RGB画像の取得センサと、深度マップの取得センサ位置が異なることから、各画像

間において画素ずれが生じる。Kinect では、画素位置合わせを行う API が用意されており、深度マップの各画素座標を当該関数で射影することで、RGB 画像の各画素と位置合わせを行うことが可能である。

位置合わせ済み深度マップの各画素において、8 近傍の隣接関係をもとに、ポリゴン化を行う。各ポリゴンの頂点毎に、RGB 画像中の対応画素を特定し、テクスチャ情報を取得する。仮想視点に応じて、当該 3 次元形状モデル、およびテクスチャをマッピングすることで、レンダリングを実施する。RGB 画像と位置合わせ済みの深度マップ、およびポリゴンモデルの一部を拡大した例を図 1 に示す。



(a) RGB 画像 (b) 深度マップ (c) ポリゴンモデル
図 1. 距離センサ 1 の取得データ, およびポリゴンモデル

4. 実験結果

提案手法の有効性を検証する目的で、5 メートル四方程度の室内環境において 2 台の Kinect を用いて撮影した画像データを対象とした実験を実施した。オブジェクトは、1 名の人物とし、Kinect の設置間隔は約 2 メートルとした。RGB 画像、および深度マップの解像度は 640×480 で取得した。これらのデータに対して提案手法を適用し、2 セットの Kinect の中央付近で、オブジェクトの正面方向に設定した仮想視点 1、仮想視点 2、および、オブジェクトの側面方向に設定した仮想視点 3 においてレンダリングを行った。また、比較方式として、ビルボード方式[6]を実装し、提案手法と同じ仮想視点におけるレンダリングを実施した。各仮想視点における提案手法、および比較手法のレンダリング結果を図 2, 図 3, および図 4 に示す。

図 2, および図 3 より、提案手法では、オブジェクトの正面化がある程度自然に実現できていることがわかる。図 4 より、側面方向についても、3 次元形状に基づく自然な見え方を再現できていることが確認された。一方、比較手法のビルボード方式では、カメラ画像から取得されるテクスチャをそのままマッピングしているため、オブジェクトの正面化や、オクルージョン再現が実現できていない。特に、オブジェクトの腕や手首の重なりによって隠れるセルフオクルージョンの一部（各図の赤枠部分）においてレンダリングの差が表れることを確認した。一方で、今回の提案手法では、オクルージョン領域のテクスチャを他カメラの対応領域から取得して補正する処理は適用できておらず、レンダリングできない領域が欠損として存在することによる画質劣化が確認された。

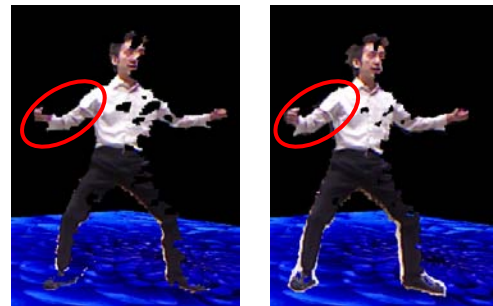
5. まとめ

本稿では、限られたカメラセットからオブジェクトの高精度な 3 次元形状モデルを復元し、仮想視点のレンダリングを行う手法として、距離画像センサを用いた手法を提案した。各センサで取得される RGB 画像と、深度マップの画素位置合わせを行い、隣接画素の奥行き値に基づ

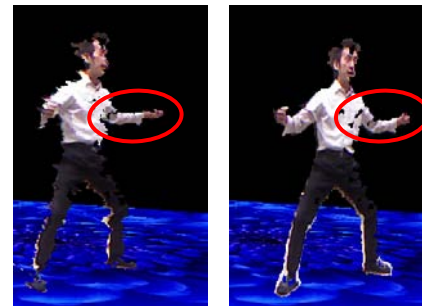
きポリゴン化し 3 次元形状モデルを復元することで、仮想視点において、オクルージョン再現の観点でより自然なレンダリングが可能であることを示した。今後は、複数カメラ間の情報を統合することで、オクルージョンにおけるテクスチャの補間について検討を行う予定である。



(a) 提案手法 (b) 比較手法
図 2. 仮想視点 1 における合成結果



(a) 提案手法 (b) 比較手法
図 3. 仮想視点 2 における合成結果



(a) 提案手法 (b) 比較手法
図 4. 仮想視点 3 における合成結果

参考文献

- [1] T. Kanade, P. W. Rander, and P. J. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes", *IEEE MultiMedia*, 4, 1, pp. 34-47 (1997).
- [2] 三功 浩嗣, 石川 彰夫, 内藤 整, 酒澤 茂之, "確率尤度を用いた 3 次元形状モデル投影型背景分離法", *映像学誌*, Vol.64, No.11, pp.1685-1697(2010)
- [3] A. Laurentini, "How far 3d shapes can be understood from 2d silhouettes", *IEEE Trans. PAMI*, vol. 17, no. 2, pp. 188-195 (1995).
- [4] K. N. Kutulakos and S. M. Seitz, "A Theory of Shape by Space Carving", *IJCV*, vol. 38, no. 3, pp. 192-218 (2000).
- [5] S. M. Seitz and C. R. Dyer, "View Morphing", In *Proc of ACM SIGGRAPH 1997*, pp. 21-30 (1997).
- [6] 三功 浩嗣, 石川 彰夫, 内藤 整, 酒澤 茂之, "大空間カメラ配置におけるカメラの強校正を前提としない自由視点画像生成法の一検討", *FIT2010*, I-054 (2010).