

全天球動画像を用いた屋外環境の新視点画像生成

New View Generation for Outdoor Scene Using Omnidirectional Video Images

池田 聖†
Sei IKEDA

佐藤 智和†
Tomokazu SATO

横矢 直和†
Naokazu YOKOYA

1. はじめに

予め取得した遠隔地の画像群から利用者の仮想視点画像を生成し提示することにより、擬似的に遠隔地で行動する感覚を与える技術はテレプレゼンスと呼ばれ、教育や娯楽などの分野への応用が期待されている。屋外のように広範囲かつ複雑な環境を仮想化し、自由に視点及び視線方向の変更が可能なテレプレゼンスシステムを構築するためには、利用者が視点移動する可能性のある全ての位置および方向の画像が必要となる。そこで、従来から実環境内の適当な位置で画像を取得しておき、新視点画像生成技術を用いて補間するというアプローチがしばしば用いられている [1, 2]。

従来提案されている新視点画像生成手法は、規模が小さく十分に制御された環境下だけで機能することが確認されているものが多い [3, 4, 5]。物体の反射に関するランバート拡散モデルに基づき 1 画素ずつ画像の輝度値を推定する手法 [4, 5] では、撮影対象がカメラの画角内に入るように画像が撮影されていることが前提であり、撮影地点の周囲に物体が存在する屋外環境への適用は難しい。また、このようなアプローチは幾何学的に正確な新視点画像が生成可能な反面、計算量が膨大となる問題がある。屋外環境のテレプレゼンスを想定した新視点画像生成手法 [1, 2] も提案されているが、複雑な形状を含む屋外環境の画像から見栄えの良い新視点画像を生成するために、画像間の対応点を手作業で与えなければならないという問題 [1] や、画像の取得位置によっては幾何学的に歪みのある画像が生成されるという問題がある [2]。

本稿では、屋外環境を対象としたテレプレゼンスにおける利用を目的とした新視点画像合成手法を提案する。提案手法では、効率よく広範囲に画像を取得するために、複数のカメラを放射状外向きに固定することにより高解像度に全天球の動画像を取得することができる全方位型マルチカメラシステム (以下、OMS) を用いる。従来の新視点画像生成手法 [4, 5] と同様のアプローチで、OMS の画像が扱えるように手法を拡張することで、撮影対象の位置や規模に関わらず幾何学的な歪みのない新視点画像の生成が可能になる。さらに、隣接画素の奥行き情報とフレーム間の差分領域を利用した 2 種類の手法により計算量削減を図る。

2. 全天球動画像からの新視点画像生成

提案手法では、あらかじめキャリブレーションされた OMS [6] を用いて取得した動画像と、動画像から推定したカメラ外部パラメータ [7] を用いて新視点画像を生成する。新視点画像を生成するために、基本アルゴリズムとして物体の反射に関するランバート拡散モデルとカメラの透視投影モデルに基づき 1 画素ずつ輝度値を決定する手法 [4, 5] と同様のアプローチをとる。ただし、基本アルゴリズムのみでは、計算量が膨大となるため、2 つ

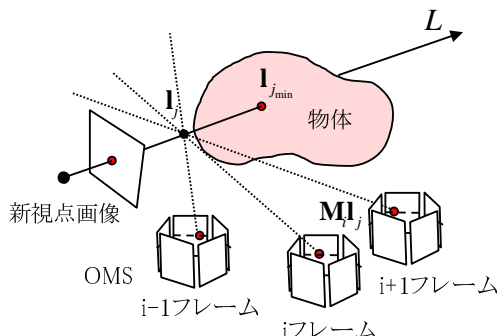


図 1: 基本アルゴリズムの概念図

の計算量削減手法を用いる。本章では、まず、新視点画像生成の基本アルゴリズムについて説明する。次に、生成画像上の隣接画素の奥行き情報と入力フレーム間の差分情報を利用した 2 つの計算量削減手法について述べる。

2.1 新視点画像生成の基本アルゴリズム

得られた動画像とそのカメラパラメータから新視点画像を生成する基本アルゴリズムについて説明する。基本アルゴリズムでは、Irani ら [4] や Fitzgibbon ら [5] の手法と同様に、環境の反射モデルとしてランバート拡散モデルを仮定し、新視点画像内の各画素値を入力画像内の対応するエピポーラ線上の画素から推定する。

まず、図 1 に示すように、新視点位置から新視点画像内の注目画素を通る半直線 $L = \{l_j \mid j = 0, 1, \dots\}$ を考え、 L 上の各点 l_j を各入力画像上に投影する。このとき、 i フレーム目のカメラの外部パラメータを 3×4 の投影行列 M_i で表すと、点 l_j は画像上の位置 $M_i l_j$ に投影される。ただし、画像上及び空間中の位置ベクトルは、齊次座標で表現する。また、入力画像は予め歪み補正されており、OMS を構成する複数のカメラのうち、投影座標と画像中心の距離が最小となるカメラが選択されているものとする。

次に、全ての入力画像上における点 $M_i l_j$ の輝度 $I_i(M_i l_j)$ の分散 σ^2 が閾値以下かつ j が最小の物体表面上の点、すなわち新視点に最も近い点 $l_{j_{min}}$ を探索する。このとき $I_i(M_i l_{j_{min}})$ の i に関する中間値を新視点画像上の各画素の輝度値として算出する。

一般の単眼カメラを用いた場合、常に L がカメラの画角内に収まるように画像を撮影しなければ、点 $l_{j_{min}}$ を探索することができない。提案手法では、OMS を用いることにより、複数のカメラで L 全体を写すことができ、カメラの位置及び姿勢や環境の規模に関わらず $l_{j_{min}}$ の探索が可能である。このように基本アルゴリズムでは、点 $l_{j_{min}}$ が見つかるまで全ての点 l_j の全フレーム上での輝度値を調べる必要があり、屋外環境のように規模の大きな環境では計算量が膨大になる。

†奈良先端科学技術大学院大学
(Nara Institute of Science and Technology)

2.2 隣接画素の奥行きを利用した計算量の削減

屋外環境で取得された画像の場合、建物や道路など、生成画像内の各画素の奥行きが隣接画素で類似していることが多い。これを利用して、推定する画素についても隣接の画素の奥行き値付近に物体表面 $l_{j_{min}}$ があると仮定し、半直線 L 上の探索範囲を以下のように間隔 $2r$ に限定する。

$$j_0 < j < j_1 \quad (|l_{j_0} - l_j| = |l_j - l_{j_1}| = r) \quad (1)$$

限定された範囲内での探索において輝度の分散 σ^2 が閾値を超える場合は、全ての範囲を探索する。

2.3 フレーム間差分を利用した計算量の削減

異なる位置で撮影された画像間で視差が発生しないような遠方の領域は、その画像位置付近の新視点画像と入力画像の間にも視差が発生しないことを利用する。提案手法では、自由に移動しながら撮影した動画をを入力とするため、カメラの姿勢の変化を考慮してフレーム間の差分をとることにより計算を省略できる画素を判定することができる。まず前処理として、(i) 選択された2フレーム間での差分とり、差分が発生している領域(差分領域)を登録する。次に、生成する画像の画素毎の処理として、(ii) 差分領域に基づいて画素の輝度を更新すべき領域(更新領域)を判定する。(iii) 更新領域においては基本アルゴリズムを用い視線方向の輝度値を決定する。更新領域以外は、選択された2フレーム上の対応する画素の輝度値を使用する。

(i) 差分領域の登録：カメラの姿勢の変化を吸収したフレーム間差分を行うために、世界座標系における絶対方位により、フレーム間の画素を対応付ける。まず、新視点位置に最も近い2フレーム A, B を選択する。次に、撮影時のカメラの回転に対応するため、予め推定されたカメラ外部パラメータ M_i を使用し、画素毎に対応する絶対方位 $d = [d_x, d_y, d_z, 0]^t$ の差分：

$$|I_A(M_A d) - I_B(M_B d)| \quad (2)$$

を算出する。閾値以上の差分が検出される方位 d を差分領域として登録する。

(ii) 更新領域の判定：フレーム間の差分は、近景に物体が存在する場合に生じることから、新視点画像上でその物体に対応する画素は新たに推定する必要がある。図2に、選択されたフレーム A, B の視点位置、新視点位置、物体の位置、及びそのとき差分が検出される絶対方位との関係を示す。近景の位置 p に物体が存在する場合、その物体は各フレームの画像上には $M_i p$ ($i = A, B$) の位置に投影され、2フレーム間の差分画像上には印で示す2つの方位 d_1, d_2 に差分領域が検出される。このことから、逆に複数の方位 d_k ($k = 0, 1, 2, \dots$) に差分領域が検出されている場合は、差分領域と各フレームの投影中心を結ぶ直線群が交差する空間中の点 p に物体が存在する可能性があると考えられる。そこで、新視点の投影中心から注目画素を結ぶ半直線 L 上にその物体が存在するとき、すなわち以下の式を同時に満たす k, k', j が存在する場合、その画素を更新領域に加える。

$$M_A d_k = \lambda_A M_A l_j, \quad M_B d_{k'} = \lambda_B M_B l_j \quad (3)$$

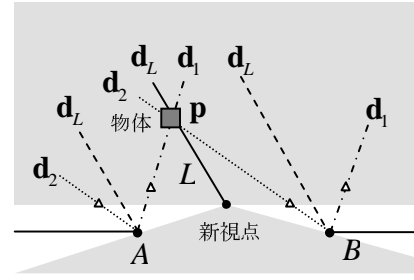


図 2: 更新領域の判定

ただし、 λ_A, λ_B は、比例係数である。また、半直線 L の方位 d_L に差分が検出される場合も、更新領域に d_L を加える。

理想的な場合でも、線分 AB を除く直線 AB 上の点のみはフレーム間差分で検出されない。このため線分 AB 上に新視点を設定する場合は、全天球画像の生成が可能であるが、図2に示す位置に新視点を設定した場合は、提案手法により新視点画像が生成可能な範囲は灰色の領域で表される方向になる。

(iii) 視線方向の輝度値の決定：更新領域に含まれる画素の輝度値は、全てのフレームを使用し、基本アルゴリズムと隣接画素の奥行きを利用した計算量の削減手法を用いて視線方向の輝度値を計算する。更新領域以外では、2フレームの輝度の平均値 $\{I_A(M_A d) + I_B(M_B d)\}/2$ を使用する。

3. 実験

実際に屋外で撮影された全天球動画をを用いて、新視点画像を生成することにより、その生成精度と計算量を評価する。生成精度の評価では、入力として使用していない画像と、その画像のカメラ位置で生成された新視点画像を比較する。実験では、屋外で歩きながら撮影した全天球動画像 200 フレームのうち 4 フレーム間隔の計 50 フレームを用いた。そのうち 49 フレームの画像を入力として使用し、ある 1 フレームの画像の撮影位置での全天球画像を生成した。画像の取得には、6 個のカメラからなり 15fps の動画像が取得できる Ladybug (Point Grey Research) を用いた。図3は、画像の取得経路を表しており、四角錐は OMS の 6 個のカメラのうち 1 個のカメラの視錐台を表す。図4に Ladybug により取得した動画像の 1 フレームを示す。入力画像の解像度は各カメラにつき 768×1024 (XGA) である。

図5に差分領域の検出で選択された2フレームの全天球画像、図6(左)にそれらの差分領域、図6(右)に更新領域、図7に生成した新視点における解像度 500×250 の全天球画像を示す。図6(右)に示すように、差分領域として検出されていない画素が、更新領域として登録されている。これは、近景に物体が存在する可能性のある方向に対応する画素を、全て更新領域としているためである。各全天球画像は球面上に投影された画像をサンソン図法により展開したものである。新視点画像生成に要した時間は約35時間であり、基本アルゴリズムのみ用いて生成した場合の54.7%に相当する。また、図8に、新視点位置と同じ位置で取得された画像との比較結果を示す。右の画像は、生成画像の誤差として評価用画像との輝度差の絶対値を示しており、真黒の画素は階調数の50%、

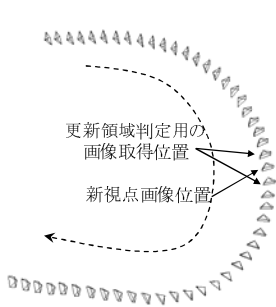


図 3: 画像の取得経路

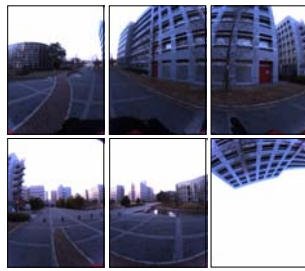


図 4: Ladybug による取得画像 (上方向 (右下) と水平方向 (その他))

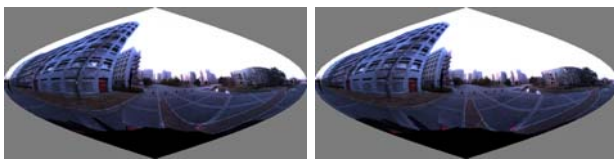


図 5: 差分領域の登録で選択された全天球画像

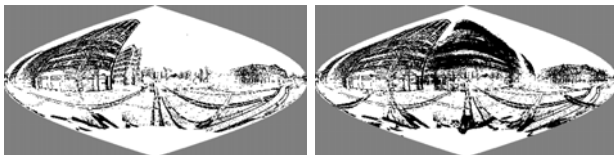


図 6: 差分領域 (左) と更新領域 (右)

すなわち 256 階調の輝度表現で輝度差 128 の誤差を表す。本実験では、平均絶対誤差は 7.0%であった。

生成された全天球画像とその差分領域の登録で選択された 2 フレームの全天球画像を用いてアニメーション表示することにより、地面の模様や建物の位置が 2 つのフレームを補間するように描画されていることを確認し、これにより新視点の全天球画像の生成が可能であることを確認した。しかし、建物の側面や木など誤った輝度値が選ばれている箇所も存在する。この主な原因としては、提案手法が物体表面の反射としてランバート拡散モデルが仮定されていることが挙げられる。実際の屋外環境に合わせて鏡面反射光を含む環境への対応が必要である。

4. まとめ

屋外環境のテレプレゼンスのための提示画像の生成手法として、全天球動画から新しい視点の全天球画像を生成する手法を提案した。提案手法では、全方位型マルチカメラシステムにより効率よく画像を取得し、推定されたカメラの外部パラメータを用いて撮影されていない地点での画像を生成するものである。さらに、隣接画素の奥行き情報、及びフレーム間差分を用いて計算量を削減する。

実験では、実際に屋外環境で撮影した動画に対して提案手法を適用し、動画のフレームを補間するような視点位置での新視点画像の生成が可能であることを確認した。また、提案手法は、画像生成における計算量を従来手法に比べて半分程度に削減可能である。しかし、生成した全天球画像をテレプレゼンスシステムの提示画像

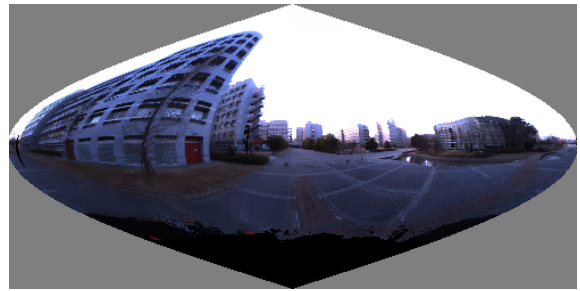


図 7: 生成された全天球画像



図 8: 評価用画像と生成画像との比較

として利用するためには、さらなる計算量の削減が必要である。今後は、複数の全天球画像を生成する場合に、1 つの全天球画像を生成する過程で得られる各画素の奥行き情報を利用した計算量の削減手法について検討する。また、3 フレーム以上の間でフレーム間差分を検出することにより、より広い範囲で新視点画像が生成可能なように拡張する予定である。

本研究の一部は、平成 15 年度文部科学省科学研究費補助金 (特別研究員奨励費) による。

参考文献

- [1] 遠藤, 片山, 田村, 廣瀬: “写実的な広域仮想空間構築のための画像補間手法”, 日本バーチャルリアリティ学会論文誌, vol.7, no.2, pp.185-192, 2002.
- [2] 高橋, 川崎, 池内, 坂内: “全方位画像を用いた広域環境の自由視点レンダリング”, 情報処理学会論文誌, vol. 42, no. SIG13, pp. 99-109, 2001.
- [3] A. Zomet, D. Feldman, S. Peleg and D. Winshall: “Mosaicing new views: the crossed-slits projection”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 741-754, 2003.
- [4] M. Irani, T. Hassner, and P. Anandan: “What does the scene look like from a scene point?”, Proc. 7th European Conf. on Computer Vision, vol. 2, pp. 883-897, 2002.
- [5] A. Fitzgibbon, and Y. Wexler, and A. Zisserman: “Image-based rendering using image-based priors”, Proc. IEEE 9th Int. Conf. on Computer Vision, pp. 1176-1183, 2003.
- [6] 池田, 佐藤, 横矢: “全方位型マルチカメラシステムを用いた高解像度な全天球パノラマ動画の生成とテレプレゼンスへの応用”, 日本バーチャルリアリティ学会論文誌, vol. 8, no. 4, pp. 443-450, 2003.
- [7] T. Sato, S. Ikeda, and N. Yokoya: “Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system”, Proc. European Conf. on Computer Vision, Vol. 2, pp. 326-340, 2004.