

隠消現実感における隠背景投影誤差の軽減 (3)

— 隠背景平面拘束の緩和と VDTM の利用 —

森 尚平^{*1} 酒井 章伸^{*1} 松見 優一^{*2} 柴田 史久^{*2} 木村 朝子^{*2} 田村 秀行^{*3}

Abstract --- This paper describes a Diminished Reality (DR) technique using View-Dependent Texture Mapping (VDTM) for visually removing real objects in a scene. Our previous work warps one of pre-observed hidden view images and overlays it to the source view on the assumption that the scene consists of planar surfaces. In this paper, the previous work is extended to be applied to 3D scenes. Our main contributions are the followings; 3D hidden scene recovery using VDTM and vision-based 6 degrees of freedom camera tracking using VDTM. Proposed method reconstructs a 3D scene and describes it as 3D surfaces, images, and camera poses corresponding to the images. Hidden views in the source view are recovered using a VDTM manner. To obtain input camera pose for VDTM, the system estimates camera pose by tracking feature points between a recovered view and the source view.

Keywords: Diminished Reality, Mixed Reality, VDTM, Camera Tracking

1 はじめに

現実世界と仮想世界の実時間融合及び表示を行う拡張現実感 (Augmented Reality; AR) や複合現実感 (Mixed Reality; MR) の概念的な発展系として、現実世界から実物体を実時間で視覚的に消去する隠消現実感 (Diminished Reality; DR) に関する研究が行われている。我々はこの技術を、除去対象で隠された背景領域(隠背景)の情報で、体験者視点におけるその除去対象を映像的に上書きする問題と捉え、体系的整理を行った[1]。

体験者視点から観測される実シーンに対して、同じく実シーンである隠背景の復元と合成を行う DR では、実シーンに対する CG の合成を行う AR/MR よりも実応用に対する障害が大きいことが知られている。こうした研究背景から、我々は対象シーンを限定することで問題解決を図ってきた。具体的には、隠背景が 2 つの平面にまで近似可能であるという拘束条件(隠背景平面拘束)を設け、その際の幾何学的及び光学的不整合の解消法を提案した[2]。また、こうした平面拘束から脱却する目的で、背景の幾何形状を用いることなく、代わりに系統的にかつ大量に撮影した視点依存画像群のみから任意視点画像を生成する Light Field Rendering を利用した DR 手法を試験的に実装した[3]。[3]では、事前に観測した背景画像と視点の位置姿勢情報を利用する Image-Based Rendering (IBR) の枠組みを DR 処理に応用している点を特徴とする。

本論文で提案する DR 手法は上記 2 つのアプローチ

*1 立命館大学大学院 情報理工学研究所

*2 立命館大学 情報理工学部

*3 立命館大学 総合科学技術研究機構

の中間に位置するものである。この位置づけにより、DR 技術に関する以下の 2 点に貢献する。

- ・ 隠背景平面拘束の緩和
- ・ カメラ位置姿勢推定機構の導入

本論文における第 1 の目的は、[2]での隠背景平面拘束を緩和することである。提案手法では、このために IBR の一手法である View-Dependent Texture Mapping (VDTM) [4]を利用する。これは、3次元幾何モデルと視点依存画像を組み合わせて実シーンを再構成する手法である。[3]程の高精度な結果は期待できないが、視点変化に起因する隠背景再構成の不整合を軽減できる。第 2 に、VDTM で必要とする情報を用いて画像ベースのカメラ位置姿勢推定を行う機構を導入する。[2, 3]のいずれの方法でもカメラ位置姿勢の推定にはセンサ等の何らかの位置姿勢計測器を利用していたが、提案手法ではこれを用いず DR を実現する。

2 関連研究

本論文で前提とするのは隠背景を実際に観測してその情報を利用する観測ベースの DR 手法である。この観測ベースの DR 手法は、観測を事前に行う手法と実時間で行う手法の 2 つに大別され、提案手法はこの内の事前観測を行う手法にあたる。この場合、事前観測した情報から、現在の体験時視点で隠された光景をできるだけ正確に再構成する必要がある。

提案手法に類似した手法として、Cosco ら[5]の例が挙げられる。Cosco らが提案する Visuo-Haptic システムでは、テーブルに設置された力覚デバイスを消して見せ、代わりに CG の工具を重畳することで、より現実に近い作業を可能にする。この手法では、想定された粗い

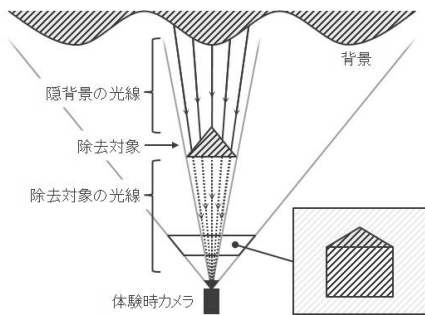


図1 体験時カメラとそこに入射される光線

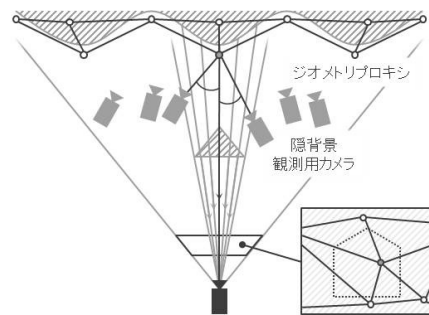


図2 隠背景から入射される光線の再構成

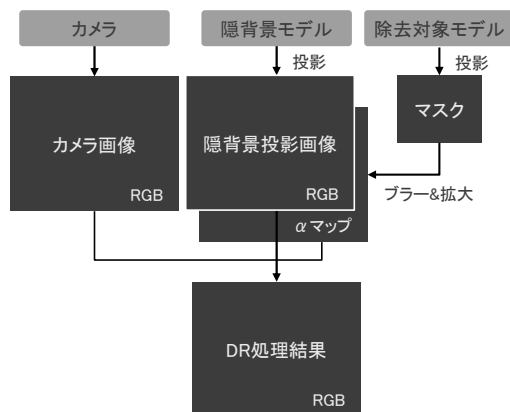


図3 隠背景の再構成に関するデータフロー

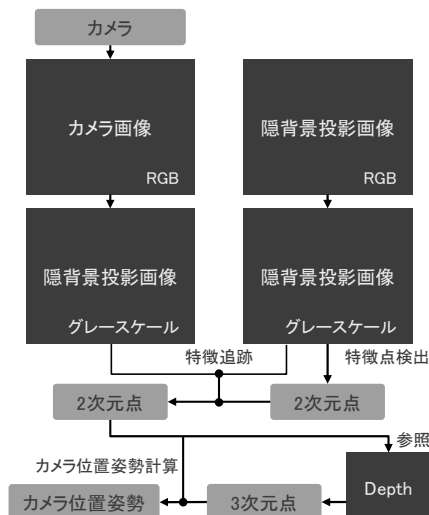


図4 カメラ位置姿勢推定に関するデータフロー

幾何形状を設定し、除去対象となる力覚デバイスを設置する前に、AR マーカを基準に複数枚の視点依存画像を撮影する。力覚デバイスが設置されてからは同様に AR マーカを用いて得られた視点位置姿勢を用いて、先の視点依存画像群から隠背景を再構成する。この手法では、隠背景の形状が手動で計測できる程度に単純なものである前提を設けている。加えて、目的とする除去対象を除去した後も AR マーカは残ってしまい、美観を損ねる。一方、提案手法では、こうした制限はなく、応用範囲がより広いと言える。

また、Liら[6]はインターネット上の写真を大量に利用することで、3次元的なシーンに対処可能な手法を提案した。しかしながら、この手法では最終的な隠背景の再構成の際には隠背景平面拘束を設けている。そのため、例えば、体験者視点位置が大きく変化する間に必要な情報がない場合、隠背景の再構成の精度は満足のないものになることは想像に難くない。よって、隠背景平面拘束を緩和する手法は依然として必要である。

3 提案手法

3.1 概要

提案手法では、体験者視点(カメラ)において隠背景を再構成し、カメラ画像上に重畳することで視覚的に除去する。これは概念的には図1, 2を用いて以下の通り説明できる。図1の状況では、体験時カメラ c^{SV} に映るのは除去対象とその領域以外の背景である。これは、隠背景から体験時カメラ c^{SV} に届く光線がなく、全て除去

対象によって遮断されていることを意味する。そこで、背景の光線を別視点 c_j^{VD} ($1 \leq j \leq N$) から取得しておき、体験時視点において、除去対象領域を通るものだけを再構成する(図2)。画像処理的には、除去対象領域を隠背景の再構成結果で上書きすればよい。

提案手法では「背景の粗い幾何形状」と「複数の視点依存画像」を組み合わせて用いるVDTMをこの再構成に利用する。以降、これら隠背景の再構成に必要なデータを「隠背景モデル」と呼ぶ。VDTMをDRに応用することで体験時視点の変化に応じた隠背景の再構成が可能になる。加えて、提案手法では隠背景の再構成結果をカメラ位置姿勢推定にも利用し、体験時視点の変化に耐えうる高精度なDR処理結果を得る。

この隠背景の再構成とカメラ位置姿勢推定に関するデータフローを図3と図4に示す。提案手法の主な処理は以下に示す5つのステップから成る。尚、処理の詳細は図2~4を用いて次節以降に述べる。

- 背景の3次元サーフェス化
- 視点依存画像の隠背景モデルへの登録
- 背景の再構成
- 隠背景投影画像の合成
- 隠背景モデルを利用したカメラ位置姿勢推定

3.2 背景の3次元サーフェス化

提案手法では、「背景の粗い幾何形状」を3次元サ

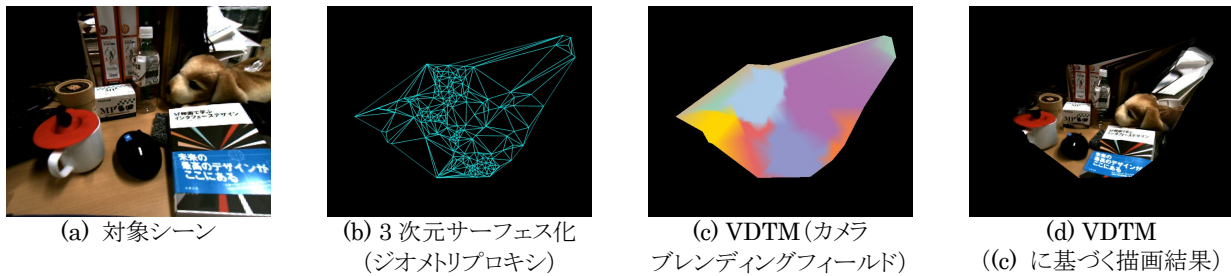


図 5 VDTM を用いた背景の再構成

ーフェス(3 つの 3 次元点とそれによって構成される面)として復元する. この 3 次元復元方法として手法を特に限定する必要はないが, 本論文では実験で用いたステレオ視による 3 次元復元の実装を例に説明する.

まず, 初期フレームにて検出した特徴点を継続して n フレーム目まで追跡する. この追跡には Lucas-Kanade トラッカ[7]を利用する. 初期及び n フレームでの特徴点の対応関係とカメラ内部パラメータから基本行列が求まるため, 三角測量の要領で追跡した特徴点の 3 次元位置 \mathbf{x}^{GP} が計算できる. 初期フレームでの特徴点検出結果から, 2D ドロネー三角形分割を行い, 対応付けられた 3 次元位置から 3 次元サーフェスを構成する. 図 5 (a) に示すシーンを対象とした場合, (b) のような 3 次元サーフェスが得られる. これを Buehler らの文献[8]に倣い, ジオメトリプロキシと呼ぶ.

3.3 視点依存画像の隠背景モデルへの登録

ある撮影地点でのカメラ画像を視点依存画像 I^{VDT} として隠背景モデルに登録する. この際のカメラ位置姿勢は, 後述する隠背景モデルを用いたカメラ位置姿勢推定法を利用して求める. ステレオ視を利用して 3 次元サーフェスを復元した場合, 初期フレームと n フレームでのカメラ位置姿勢が求まっているため, この 2 視点分の画像が視点依存画像として登録される. ユーザが任意に設定した必要視点数を登録し, この処理を終了する. 図 5 の例では 20 視点分の視点依存画像を登録した.

3.4 背景の再構成

作成した隠背景モデルを基に, 体験時視点 c^{SV} において隠背景を再構成する. これは, 図 2 における隠背景の光線を含む背景全ての光線を体験時視点 c^{SV} において再構成することに相当する.

提案手法では Cosco らの手法と同様, c^{SV} , \mathbf{x}^{GP} , c_j^{VD} の 3 点により得られる角度を基準にし, ジオメトリプロキシの頂点 \mathbf{x}^{GP} における j 番目の視点依存カメラ c_j^{VD} の重み w_j を決定する. これは, [8]における 3 つの Penalty の内, Angular Penalty のみを計算に含めた構成と言える. 提案手法では, ある頂点 \mathbf{x}^{GP} に関して最小の角度 $\arg \min(w_j)$ を持つ視点の重みを最大, その頂点を含む 3 次元サーフェスの他の頂点での重みを最小とする.

図 5 の例において, 各視点に疑似色を与えて表示し

たカメラブレンディングフィールド[8]を (c) に, このカメラブレンディングフィールドに基づいて VDTM を実行した結果を (d) に示す. この結果を隠背景投影画像 I^{HV} (図 3) と呼び, 体験時カメラ c^{SV} の画像 I^{SV} に合成する.

3.5 隠背景投影画像の合成

前節では, カメラ c^{SV} の視野角に含まれる背景全てを再構成した. 実際の DR 処理においては隠背景の部分のみが必要であるため, これを切り出す. これは, 隠背景投影画像 I^{HV} に対するマスク処理を施すことによって実現できる(図 3).

提案手法では, 除去対象の形状を囲う程度の円筒形の幾何モデル(除去対象モデル)を 3 次元空間内に配置し, これを視点 c^{SV} に投影することで得られるマスク画像 I^{MK} を用意する. 除去対象モデルが円筒形なのは, 体験者が左右に移動しても見え方の変化が少なく, かつ除去対象を十分に囲える形であるためである. I^{MK} では, モデルが投影された画素値を 1, それ以外を 0 とおく. これを用いて式 (1, 2) に則って隠背景投影画像 I^{HV} を体験時画像 I^{SV} に合成する.

$$I^{\text{MK}}(\mathbf{x}) = \begin{cases} 1 & \text{if model is projected} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$I^{\text{DR}}(\mathbf{x}) = \begin{cases} I^{\text{SV}}(\mathbf{x}) & \text{if } I^{\text{MK}} = 1 \\ I^{\text{HV}}(\mathbf{x}) & \text{otherwise} \end{cases} \quad (2)$$

ただし, \mathbf{x} は c^{SV} における画像座標値を表す.

隠背景に相当する光線の切り出しの目的においては, この処理で十分である. しかし実際には, カメラ位置姿勢推定や隠背景の再構成に関するずれが発生するため, 合成箇所が目立つという問題が発生する. そのため, マスク画像 I^{MK} に対してボックスブラー処理を施し, これを α マップ I^α として隠背景投影画像 I^{HV} と体験時画像 I^{SV} を α ブレンディングにより合成する(式 (3)).

$$I^{\text{DR}}(\mathbf{x}) = I^\alpha(\mathbf{x})I^{\text{SV}}(\mathbf{x}) + (1 - I^\alpha(\mathbf{x}))I^{\text{HV}}(\mathbf{x}) \quad (3)$$

ただし, 処理負荷軽減の目的で, マスク画像 I^{MK} は縦横を半分にした 1/4 の画素数で生成し, ボックスブラー処理を施し, これを体験時視点画像の大きさにバイリニア補間を用いて拡大したものを利用する.

3.6 隠背景モデルを利用したカメラ位置姿勢推定

体験時視点カメラの位置姿勢は, 体験時カメラ画像

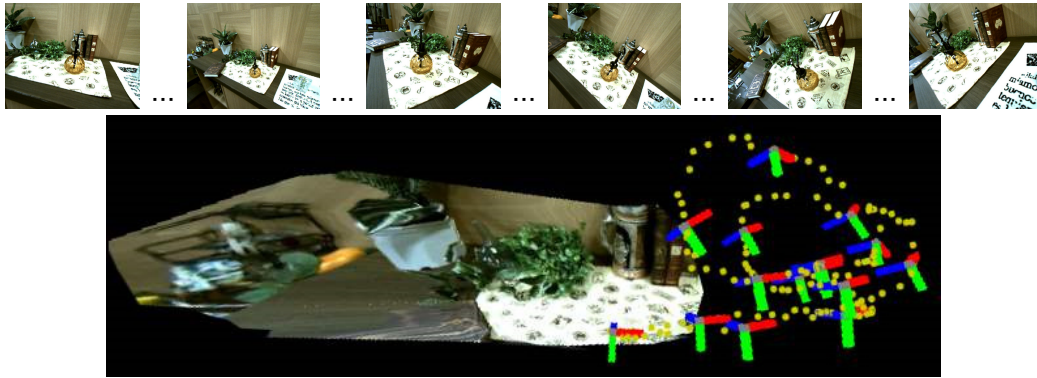


図6 トラッキングしたフレームの例と推定されたカメラ位置姿勢。
 除去対象となる物体(この場合、霧吹き器)は復元された背景に含まれない。

I^{SV} と隠背景投影画像 I^{HV} 生成時に得られるグレースケール画像及び奥行情報 I^D を用いて推定する。前フレームでのカメラ位置姿勢 M_{i-1} が既知の場合、現在のカメラ位置姿勢 M_i は以下に示す方法で追跡可能である。尚、初期位置姿勢は背景の3次元サーフェス推定時に得られているものとする。

まず、前フレーム $i-1$ での隠背景投影画像 I_{i-1}^{HV} と現フレーム i での体験時視点画像 I_i^{SV} 間でオプティカルフローを計算し、特徴点の2D-2D対応を得る。具体的には、グレースケール化した隠背景投影画像 I_{i-1}^{HV} に対して特徴点検出器を用いて特徴点を検出し、同じくグレースケール化した体験時視点画像 I_i^{SV} に対してオプティカルフローを計算する。尚、本論文の実装では、特徴点検出器に Harris 特徴点検出器[9]、オプティカルフローの計算に Lucas-Kanade トラッカを用いた。

次に、隠背景投影画像 I_{i-1}^{HV} に対応する特徴点の3次元位置を隠背景モデルの3次元サーフェスから得られる奥行画像 I^D から取得することで2D-3D対応を取得する。こうして得られた3D-2D対応を用いて PnP 問題を解くことで現フレームでのカメラ位置姿勢 M_i を計算する。図6に示すのは、実際に推定されたカメラ位置姿勢を可視化した結果である。

4 実験と考察

4.1 実験内容とシステム構成

提案手法が有効に機能することを示すために下記の屋内外4シーンを対象に提案手法を実行する。内2シーンに関しては他手法との比較評価実験を行う。

【棚上】除去対象、背景ともに任意の物体を選択し配置。Cosco らと同様、除去対象を取り除いた状態で隠背景の観察。除去対象には霧吹き器を、背景には幾何形状が複雑な造花や鏡面反射する瓶等を配置。屋内。

【スタジオ】実在するスタジオ内。除去対象のスタンドは任意の位置に配置。ただし、除去対象の裏に回り込むことで隠背景を観測。「棚上」よりも広いシーン。屋内。

【喫煙所】実在するシーン。隠背景は除去対象を取り除

いた状態で観察可能。除去対象に灰皿を配置。ベンチや壁等、背景の構造は比較的単純。屋外。

【植え込み】実在するシーン。隠背景は除去対象を取り除いて観察可能。除去対象にスタンドを配置。背景には草木が存在し、構造が比較的複雑。屋外。

提案システムの実装には、プログラミング言語として C++、描画用 API に OpenGL (GLSL 3.3 以上での動作を確認)、統合開発環境に Visual Studio 2013 を用いた。システムの実行には、Windows 8.1 64-bit、Intel Core i7-4500U (1.8 GHz)、8GB メモリ、Intel HD Graphics 4000 を搭載したラップトップ PC を用いた。カメラには富士フィルム社製 DV3.4x3.8SA-1 レンズを付けた PointGrey 社製 Flea 3 (FL3-U3-13E4C-C) を用いた。画像サイズは 640×512 [画素] である。

尚、処理速度向上のため、図3に示す処理及び図4中の奥行き画像生成は OpenGL を用いて GPU 上で処理するよう実装した。

4.2 他手法との比較評価実験

本実験では「棚上」と「喫煙所」のシーンを対象に他手法との比較評価を行う。除去対象を撮影した1枚の画像に対して下記の3手法により DR 処理を実行し、結果を比較する。尚、DR 処理結果の参考画像を得るため、除去対象を物理的に取り除いて同一視点で撮影した画像も用意する。

- Barnesらによる PatchMatch [10]
- テクスチャ付3次元モデル:提案手法 ($k=1$)
- 提案手法 ($k>1$)

PatchMatch は Adobe Photoshop CS5 に実装されている静止画像に対する画像修復法である。PatchMatch を用いる場合は手作業で除去対象領域を沿うように切り取った。残りの2つは提案手法であるが、一方は視点依存画像を1枚のみ用い、もう一方は k 枚用いる。視点依存画像を1枚のみ用いる場合は、隠背景の再構成のために、テクスチャマッピングした3次元幾何モデルを用いることと同義である。

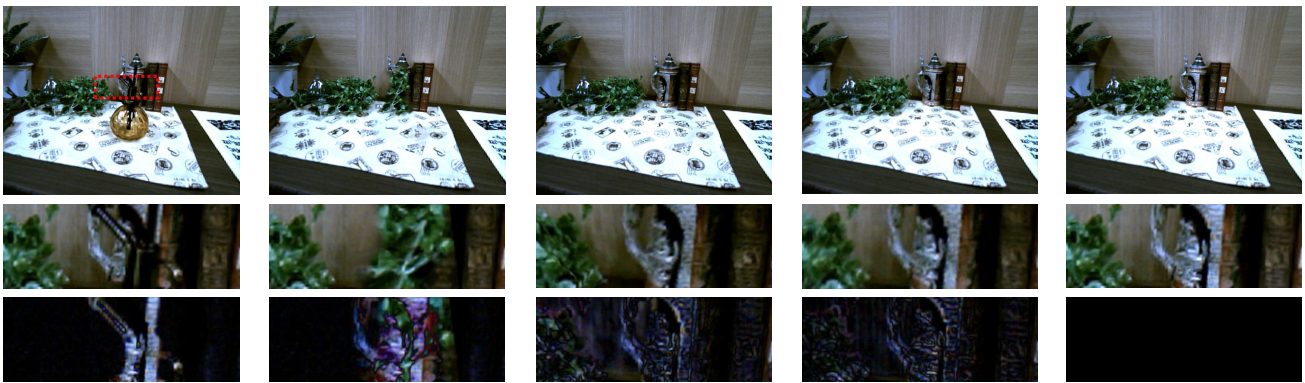


図 7 「棚上」での 3 手法の比較結果 (左から, カメラ画像, PatchMatch, テクスチャ付 3 次元モデル, 提案手法 ($k = 19$), 参考画像. 上段より, 処理結果, 赤枠で囲われた部分を拡大した処理結果, 同じ部分の参考画像との差分画像 (差の絶対値))

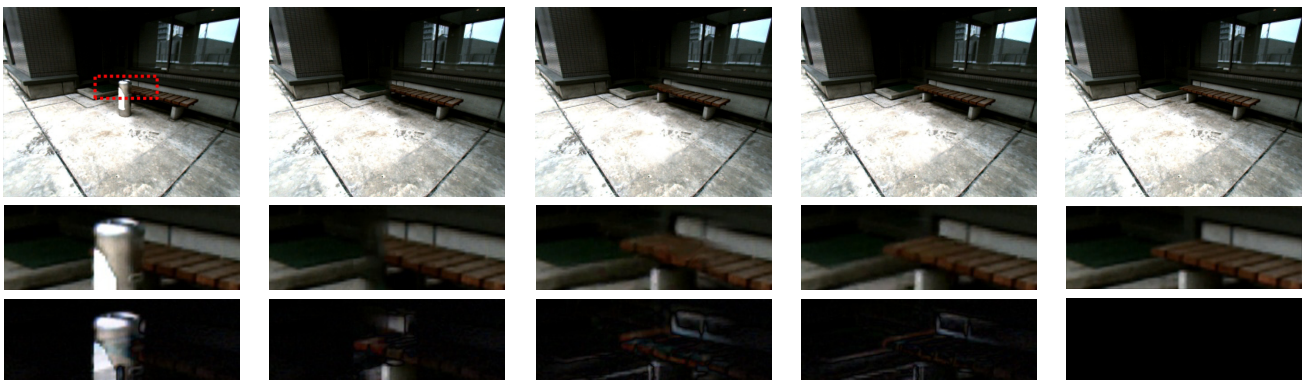


図 8 「喫煙所」での 3 手法の比較結果 (左から, カメラ画像, PatchMatch, テクスチャ付 3 次元モデル, 提案手法 ($k = 22$), 参考画像. 上段より, 処理結果, 赤枠で囲われた部分を拡大した処理結果, 同じ部分の参考画像との差分画像 (差の絶対値))

4.3 実行結果と実行速度の計測

本実験では先述の 4 シーンにおいて画像シーケンスを撮影し, その画像シーケンスに対して提案手法を用いることで, 実在する物体を除去する. その結果を示すとともに, 各処理での処理時間を計測し, 全フレームの平均の処理速度を示す.

4.4 結果と考察

他手法との比較評価実験の「棚上」での結果を図 7 に示す. PatchMatch の結果では, 隠背景にある水差しと類似した模様が画像中になかったことから造花が重畳されてしまった. テクスチャ付 3 次元モデルと提案手法 ($k = 19$) の差は, 水差しの模様の部分に表れた. より多くの視点依存画像を用いた後者の結果がより高精度だということが分かる. 両手法では, カメラ位置姿勢推定誤差が除去対象領域の推定に影響を及ぼし, 除去対象が露出するのを防ぐため, 大きめの除去対象モデルを用意した. これによって, 実際の隠背景ではない造花の部分で復元結果の誤差が露出してしまった. しかし, この部分でも提案手法の結果の方が高精度であった.

他手法との比較評価実験の「喫煙所」での結果を図 8 に示す. 図 7 と同様の理由で, PatchMatch ではベンチの脚が消えてしまっている. この結果ではテクスチャ付 3 次元モデルと提案手法との差はあまり見られない. これは, このシーンが視点位置の変化に対して見た目

に表れる変化が少なかったためであると考ええる.

4.2 節の実験結果を図 9 に示す. 復元した背景が体験時カメラの画角に収まっている限り, 視点位置, 屋内外問わず, 高精度に除去対象を視覚的に除去できていることが分かる. ただし, 本論文での実装のような背景の復元方法では, 「スタジオ」の例のように, 体験時との位置関係から除去対象領域に対しての背景の復元範囲が狭くなる傾向にあり, 除去対象を完全に除去できない. この問題を解決するにあたって, 単純に背景の復元範囲を広くする方法が考えられるが, ジオメトリプロキシの頂点数が増えてしまうことから, 実行速度の低下を招くと考えられる. 他の方法として, 画像修復法と提案手法を組み合わせることも考えられる (図 10). 各シーンでの隠背景モデルの構築に利用したデータと 1 フレームあたりの処理時間の平均を以下の表に示す.

表 隠背景モデルのデータと処理時間

シーン	頂点数	画像枚数	処理時間 [ms]
棚上	307	19	116.3
スタジオ	346	21	105.0
喫煙所	352	22	114.2
植え込み	238	23	102.5

5 まとめ

本論文では, 事前観測ベースの DR 手法に関する提案を行った. 提案手法では, 当研究グループでこれま

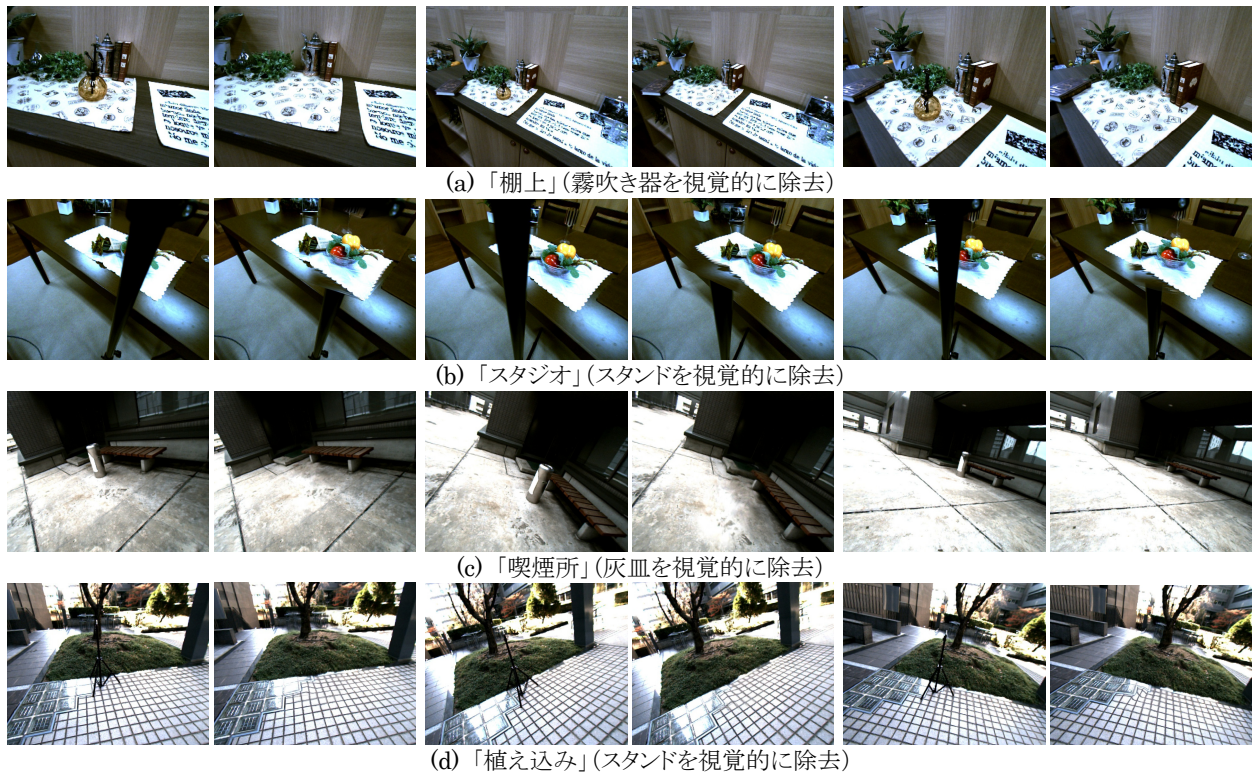


図9 体験者視点のカメラ画像(左側)とDR処理結果(右側)



図10 「スタジオ」の結果に PatchMatch を組み合わせた例

でに取り組んできた隠背景平面拘束を緩和する目的で VDTM を利用した。また、ここから得られる情報を基にカメラ位置姿勢を推定する機構を導入した。提案手法では、これらの組み合わせにより高精度なDR処理結果を得ることができることを実験により示した。具体的には、画像修復法やテクスチャ付3次元モデルを利用するDR手法との比較評価実験を行い、複数の屋内外での適用結果を示した。今後は、照明条件が変化することで発生する光学的不整合への対処法の開発等に取り組む予定である。

謝辞

本研究の一部は、科研費・基盤研究(S)「複合現実型情報空間の表現力基盤強化と体系化」と学振・特別研究員奨励費(課題番号 25・9193)「複合現実空間の表現力向上のための2つのアプローチ」による。

参考文献

[1] 森尚平, 一刈良介, 柴田史久, 木村朝子, 田村秀行: “隠消現実感の技術的枠組みと諸問題～現実世界に実在する物体を視覚的に隠蔽・消去・透視する技術について～”, 日本VR学会論文誌, Vol.

16, No. 2, pp. 239–250, 2011.

- [2] 小向啓文, 森尚平, 柴田史久, 木村朝子, 田村秀行: “隠消現実感における隠背景投影誤差の軽減(2) – 周辺参照領域の効果的利用 –”, 電子情報通信学会パターン認識・メディア理解研究会, 信学技報, Vol. 113, No. 402, MR2014-1, pp. 147–152. 2014.
- [3] K. Yanagida, Y. Nakanishi, S. Mori, F. Shibata, A. Kimura, and H. Tamura: “Image-based recovery of geometrically complex backgrounds in diminished reality,” Proc. KJMR, 2014.
- [4] P. E. Debevec, C. J. Taylor, and J. Malik: “Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach,” Proc. SIGGRAPH, pp. 11–20, 1996.
- [5] F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M. A. Otaduy: “Augmented touch without visual obstruction,” Proc. ISMAR, pp. 99–102, 2009.
- [6] Z. Li, Y. Wang, J. Guo, L.-F. Cheong, and S. Z. Zhou: “Diminished reality using appearance and 3D geometry of internet photo collections,” Proc. ISMAR 2013, pp. 11–19, 2013.
- [7] J.-Y. Bouguet: “Pyramidal implementation of the Lucas Kanade Feature Tracker: Description of the algorithm,” Retrieved from http://robots.stanford.edu/cs223b04/algo_tracking.pdf, 2000.
- [8] C. Buehler, M. Bosse, L. McMillan, S. J. Gortler, and M. F. Cohen: “Unstructured Lumigraph Rendering,” Proc. SIGGRAPH, pp. 425–432, 2001.
- [9] C. Harris and M. Stephens: “A combined corner and edge detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.