

隠消現実感における隠背景投影誤差の軽減 —モザイク法を用いた幾何学的不整合への対処—

Error Reduction for Hidden Area Projection in Diminished Reality
- An Approach to Geometric Consistency Using Image Mosaicing -

小向啓文¹⁾, 森尚平²⁾, 柴田史久¹⁾, 木村朝子¹⁾, 田村秀行¹⁾

Akifumi Komukai, Shohei Mori, Fumihisa Shibata, Asako Kimura, and Hideyuki Tamura

1) 立命館大学大学院 情報理工学研究科 (〒525-8577 滋賀県草津市野路東 1-1-1)

2) 立命館大学大学院 理工学研究科 (〒525-8577 滋賀県草津市野路東 1-1-1)

Abstract: In this paper, we propose an error reduction method for hidden area projection in Diminished Reality (DR) using image mosaicing. DR presents a view without a visual obstacle by projecting an image of its background captured beforehand or from a different viewpoint in real-time. While superimposing the image requires to estimate camera pose accurately, solely using existing attempts of camera tracking method in Mixed Reality (MR) does not satisfy the requirement due to lack of accuracy. Therefore, we propose to use image mosaicing to correct the inconsistency between an observed view and a projected background image. Results of quantitative evaluation using mean squared error showed effectiveness of our method.

Key Words: Diminished Reality, Mixed Reality, Image Mosaicing, Camera Registration

1. はじめに

Diminished Reality (DR) は、実在する物体を視覚的に隠蔽・消去する技術であり、現実世界に情報を追加する複合現実感 (Mixed Reality; MR) とは逆の概念である。我々は、この MR の発展的技術分野に対して「隠消 (いんしょう) 現実感」という日本語を充て、従来研究の調査と体系的な整理を行った[1]。

DR は、除去対象に隠された光景 (隠背景) を観測し、定義したモデルに即して体験者視点に 1 枚の画像 (隠背景投影画像) として重畳するという手順で実現できる。即ち、現実空間の光景に別視点の現実空間の光景を重畳することから発生する幾何学的・光学的不整合は、現実空間の光景に自由視点で描画可能な仮想物体のみを重畳する MR のそれよりも実現難易度が高い。それゆえ、DR における幾何学的・光学的整合性の同時達成はより一層の技術開発が必要となる。しかし、前者だけに注目しても、既存研究では MR トラッキングの単純な応用に留まっている。例えば、Cosco ら[2]は、Image Based Rendering を用いて物体の高精度な視覚的除去を実現したが、その位置合わせ手法は MR で一般的な人工矩形マーカを利用する手法が用いられているに過ぎない。

DR における隠背景投影画像を生成する処理までは、MR における仮想物体の重畳表示と同様の手順で実現されるため、シーンの復元やカメラ位置姿勢推定をいかに高精度に実現するかという問題に帰着する。しかし、DR においては、入力画像中の除去対象を隠背景投影画像の重畳により除去したように見せることが目的となる。よって、

本稿で着目し、提案するのは、隠背景投影画像生成後のその張り合わせ方法である。具体的には、隠背景投影画像と入力画像の合成時にイメージモザイク法[3]を用いた隠背景投影画像の重畳ずれを軽減する方法を提案する。

2. 提案手法

2.1 前提条件

本手法では DR の各構成要素に対して以下のような前提条件をおく。除去対象：幾何モデルを持っており、除去対象領域はそれにより決定できる。隠背景：幾何形状は 1 つの平面 (隠背景面) もしくはそれに近似でき、事前に撮影可能。体験者視点：自由視点で位置姿勢はトラックを用いて計測可能。

2.2 処理手順

本手法では、入力画像と事前に除去対象無しで撮影した画像とで検出される特徴点のマッチング結果から、両画像間のホモグラフィ行列を算出し張り合わせることで、隠背景投影画像重畳時の位置ずれを軽減する。本手法の処理手順を図 1 に示す。以降、この処理手順に則って説明する。

2.2.1 隠背景の撮影とカメラ位置姿勢の記録

本手法では、事前準備として、隠背景を様々な角度から撮影しその時の位置姿勢を保存しておく。これは、本手法でマッチングに用いる特徴点の検出器とその特徴量記述子として SURF[4]を採用していることに起因する。

SURF は拡大縮小や回転、照明変化に不変な特徴量であり、高速なマッチングが可能であることが知られている。しかし、画像のアフィン変換に対しては脆弱であるため、

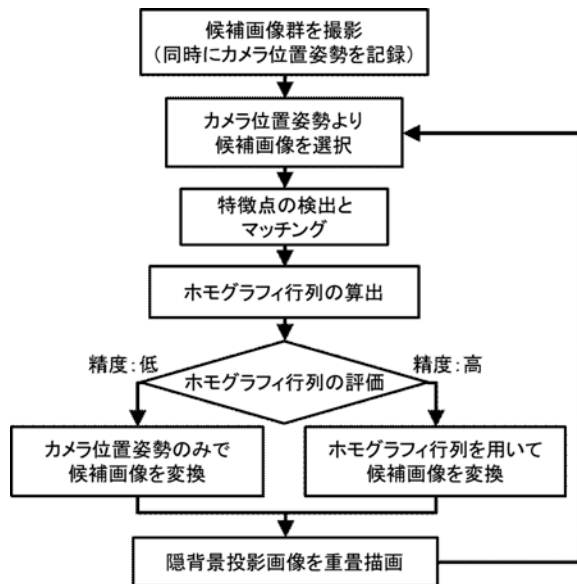


図1 提案手法の処理の流れ

隠背景投影画像と入力画像の撮影位置と姿勢が大きく異なる場合にマッチングが不可能、あるいは精度が大きく低下してしまうという問題がある。そこで本手法では、除去対象を設置する前のシーンを様々な方向から撮影し、隠背景投影画像の候補（候補画像群）として事前に取得しておく、トラックから得られるカメラ位置姿勢とともに記録しておく。そして、実行時に得られるカメラ位置姿勢との比較を行い、入力画像撮影時のカメラ位置と最も近い位置で撮影された画像を選択する。その時、カメラ位置は、注目点を原点とした球座標系において、その動径の長さを無視した2つの偏角で表す。こうすることで、視差による見え方の違いが見られない画像の不用意な保存を防ぐ。

球座標で表された3次元座標 \mathbf{q} は、その動径 r 、偏角 $\omega = [\theta, \phi]$ を用いて式 (1) のように表される。体験者視点位置とある候補画像のカメラの位置での偏角をそれぞれ ω_{cam} 、 ω_{bg} としたとき、2つの座標の偏角のユークリッド距離 d は式 (2) のように表される。よって、この d を最小化する候補画像を選択することで、特徴点マッチングのための適切な画像を選択することができる。

$$\mathbf{q} = [r \ \theta \ \phi]^T = [r \ \omega]^T \quad (1)$$

$$d = \|\omega_{\text{cam}} - \omega_{\text{bg}}\| \quad (2)$$

2.2.2 イメージモザイクによる隠背景投影画像の重畳

ここでは、選ばれた候補画像と入力画像とのマッチングを特徴点と SURF 特徴量を用いて行い、両画像間のホモグラフィ行列を求める。図2に示すのは、体験者視点、隠背景、選択した候補画像との幾何学的関係である。

特徴点マッチングの際に、画像中の特徴点同士の類似度を総当たりで計算し、最も類似度の高い特徴点同士をマッチングさせると、画像中に類似した特徴点が複数存在した場合、それらが誤ってマッチングされる可能性が高くなる。特徴点の誤対応が発生した場合、最終的に算出されるホモグラフィ行列の精度が低下し、隠背景投影画像の重畳描画

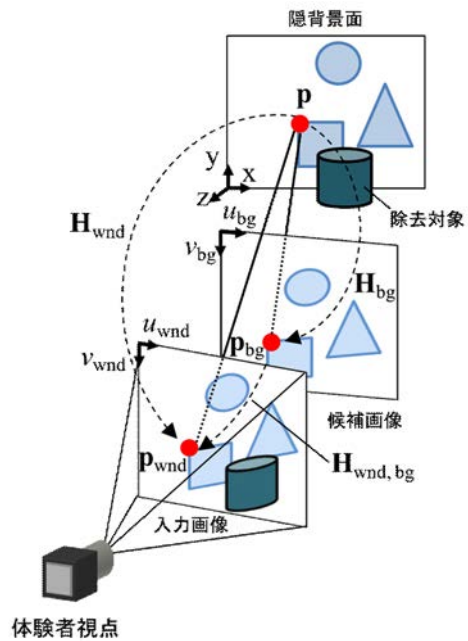


図2 各画像平面間の射影変換

時に幾何学的不整合が発生する可能性がある。そこで、トラックから得られるカメラ位置姿勢を用いて特徴点マッチングの範囲限定を行い、マッチング精度の向上を図る。

具体的には、まず、選択した候補画像の特徴点 $\mathbf{p}_{\text{bg}} = [u_{\text{bg}}, v_{\text{bg}}]^T$ を、画像とともに保存しておいたカメラ位置姿勢から得られる隠背景面と候補画像間のホモグラフィ行列 \mathbf{H}_{bg} と、トラックから取得した現在のカメラ位置姿勢から得られる隠背景面と入力画像間のホモグラフィ行列 \mathbf{H}_{wnd} により求まる $\mathbf{H}_{\text{wnd,bg}}$ を用いて入力画像に投影する。こうして投影された $\mathbf{p}_{\text{wnd}} = [u_{\text{wnd}}, v_{\text{wnd}}]^T$ 周りに探索窓を設けることで、探索範囲の限定を行う。 $\mathbf{H}_{\text{wnd,bg}}$ 、 \mathbf{H}_{bg} 、 \mathbf{H}_{wnd} の関係式を式 (3) に、 $\mathbf{H}_{\text{wnd,bg}}$ 、 \mathbf{p}_{bg} 、 \mathbf{p}_{wnd} の関係式を式 (4) に示す。

$$\mathbf{H}_{\text{wnd,bg}} = \mathbf{H}_{\text{wnd}} \mathbf{H}_{\text{bg}}^{-1} \quad (3)$$

$$\begin{pmatrix} u_{\text{wnd}} \\ v_{\text{wnd}} \\ 1 \end{pmatrix} = \mathbf{H}_{\text{wnd,bg}} \begin{pmatrix} u_{\text{bg}} \\ v_{\text{bg}} \\ 1 \end{pmatrix} \quad (4)$$

このように、特徴点マッチングの範囲限定を行った場合でも、完全に誤対応を抑制することはできない。そこで、ホモグラフィ行列を算出する際に、RANSAC アルゴリズム [5] を用いることによって可能な限り外れ値を除外し、最終的に算出されるホモグラフィ行列の精度向上を図る。

なお、 \mathbf{H}_{bg} と \mathbf{H}_{wnd} を求める手順は以下の通りである。世界座標系でのカメラ位置姿勢を表す行列を \mathbf{V} 、世界座標系の3次元座標をカメラのスクリーン座標系に投影する行列を \mathbf{P} とすると、行列 \mathbf{PV} を用いて世界座標系上の点 $[x, y, z]$ とカメラのスクリーン座標系上の点 $[u, v]$ の対応関係を表すことができる。ここで、前提として隠背景は平面であり、その位置姿勢は既知である。隠背景面を世界座標系における xy 平面と平行に設定し、その z 座標を 0 とすることで、 4×4 の行列 \mathbf{PV} から3行目と3列目の要素を取り除くこと

が可能となり、 3×3 の行列 \mathbf{H} として世界座標系とカメラのスクリーン座標系の対応関係を表すことができる。これをまとめたものが、式 (5) となる。行列 \mathbf{P} はカメラの画角などによって一意に定まり、行列 \mathbf{V} はカメラの位置姿勢によって決定する。カメラ位置姿勢は既知であるので、候補画像撮影時のカメラ位置姿勢と入力画像撮影時のカメラ位置姿勢を用いてそれぞれ \mathbf{H}_{bg} , \mathbf{H}_{wnd} を求めることができる。

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{P}\mathbf{V} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \mathbf{P}\mathbf{V} \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} = \mathbf{H} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (5)$$

2.2.3 ホモグラフィ行列の精度が低い場合の処理

本手法において最終的に算出されるホモグラフィ行列の精度は、最終的な重畳結果を左右する重要なものであるが、そのホモグラフィ行列の精度が著しく低くなる状況が2つ考えられる。1つ目は、画像中から特徴点が十分に検出されず、ホモグラフィ行列を算出するのに必要な4組の点のマッチングが得られない場合である。この場合はそもそもホモグラフィ行列を算出することができない。2つ目は、特徴点の誤対応が多い場合である。マッチング結果に誤対応が多く含まれていた場合も、算出されるホモグラフィ行列の精度が低くなる。このような状況では、場合によってはカメラ位置姿勢のみを用いて隠背景投影画像を重畳描画する場合よりも大きなずれが発生し、除去対象が露出してしまいう場合もある。そこでそのような場合には、カメラ位置姿勢を用いてホモグラフィ行列の精度評価を行い、精度が低いと判断された場合にカメラ位置姿勢のみを用いた重畳を行うことで、除去対象が露出するような極端な位置ずれを抑制する。

ホモグラフィ行列の精度評価を行うために、誤差を表すホモグラフィ行列を算出する。式 (3) で求めた候補画像を入力画像に射影変換するホモグラフィ行列を $\mathbf{H}_{wnd,bg}$ 、特徴点マッチングの結果算出されたホモグラフィ行列を \mathbf{H}_m とするとき、誤差を表すホモグラフィ行列 \mathbf{H}_e は式 (6) のように表すことができる。

$$\mathbf{H}_e = \mathbf{H}_{wnd,bg}^{-1} \mathbf{H}_m \quad (6)$$

カメラ位置姿勢が正確に計測されており、特徴点マッチングの結果算出されたホモグラフィ行列 \mathbf{H}_m が正確である場合、 $\mathbf{H}_{wnd,bg}$ と \mathbf{H}_m は一致するため、 \mathbf{H}_e は単位行列となる。この \mathbf{H}_e によってホモグラフィ行列の精度評価を行う。

ホモグラフィ行列は投影の成分と回転・並進の2つの行列に分解することが可能である。そこで、 \mathbf{H}_e を投影成分と回転・並進成分に分解し、回転・並進成分から得られた平行移動量と回転角にそれぞれ閾値 t_p , t_θ を設けてホモグラフィ行列の精度を評価する。一定以上の平行移動量と回転角があった場合には精度の低いホモグラフィ行列とみなし、そのような場合にはカメラの位置姿勢のみから算出した $\mathbf{H}_{wnd,bg}$ を用いて隠背景投影画像の変換・重畳を行う。

3. 実験と考察

3.1 実験と評価方法

提案手法の有効性を確認するために、6自由度センサを用いてカメラ位置姿勢推の計測を行い、そのカメラ位置姿勢のみに基づいて隠背景投影画像の重畳を行った場合と、提案手法を用いて隠背景投影画像の重畳を行った場合の比較を行う。その際、目視による確認だけでなく、平均二乗誤差 (MSE) を用いた定量的な評価も行う。

実験で用いる PC の仕様は、CPU: Intel Core i7 940 (2.93GHz), RAM: 4GB, GPU: NVIDIA GeForce GTX 280 である。カメラは Point Gray Research 社製 Flea2 (640×480[px], 30FPS) を使用し、SPACE 社製レンズ S4.8 1.8 を装着する。カメラ位置姿勢取得用のトラックには磁気センサである Polhemus 社製 Fastrak を用いる。

定量的な評価の方法は以下の通りである。まず三脚にカメラを固定し、除去対象が存在する場合と存在しない場合の2枚の画像を撮影する。次に除去対象が存在する画像に対して、提案手法とセンサから得られるカメラ位置姿勢のみを用いて隠背景投影画像を重畳する手法の2つの手法を用いて物体を除去した画像を生成する。最後に除去対象の存在しない画像を真値として、物体を除去した2枚の画像それぞれとの MSE を算出し、その値を評価基準とする。MSE の算出方法は式 (7) の通りである。MSE は非負の数であり、同一の画像間の MSE は 0 となることから、算出した MSE が 0 に近い程真値に近い結果が生成できたことになる。ここで $\mathbf{I}_{a,i}$, $\mathbf{I}_{b,i}$ はそれぞれ2枚の画像をグレースケール化した際の画素値、 N は画素数を表す。

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\mathbf{I}_{a,i} - \mathbf{I}_{b,i})^2 \quad (7)$$

3.2 実験シーンの構築

実験を行うシーンは手法の前提条件に従い、幾何形状が既知である物体を除去対象として卓上に設置し、その背後に本を並べ、背景を平面で構成したシーンとする (図 3)。手法の評価は幾何学的不整合の軽減に主眼を置くため、幾何学的整合性に関連しない要素は極力排除した実験シーンを構築する。

まず、候補画像撮影時と物体除去時のカメラは同一であり、シーンの照明変化は無いものとする。また、光源を除去対象の真上に設置することで、除去対象が背景に落とす影はないものとする。この条件によってシーンの光学的な差異および光学的不整合を考慮しない。次に、除去対象と体験者視点の間には遮蔽物が存在しないものとし、それらのオクルージョンは考慮しない。また背景の色や模様は、ホモグラフィ行列を算出するための特徴点が十分に検出できるものであるとする。

3.3 実験結果と考察

図 4 (a) は隠背景投影画像を重畳する前の入力画像、図 4 (b) は図 4 (a) にカメラ位置姿勢のみを用いて隠背景投影画像を重畳した結果、図 4 (c) は提案手法を用いて隠

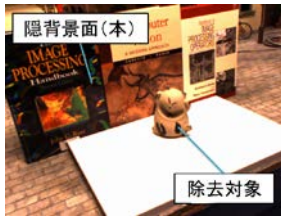


図3 物体を除去するシーン



(a) 重畳前



(b) カメラ位置姿勢のみを用いた重畳結果



(c) 提案手法による重畳結果

図4 カメラ位置姿勢のみを用いた手法との比較 (撮影箇所1)



(a) 撮影箇所2 (重畳前)



(b) 撮影箇所2 (重畳後)



(c) 撮影箇所3 (重畳前)



(d) 撮影箇所3 (重畳後)

図5 提案手法による重畳結果

表1 MSEの算出結果

撮影箇所	1	2	3	4	5	6	7	8	9	10	平均
カメラ位置姿勢のみ	656.60	473.68	347.92	332.82	324.11	613.04	362.14	47.54	361.56	302.31	382.17
提案手法	49.40	13.11	15.32	12.24	15.25	39.15	41.90	12.38	55.88	8.81	26.34

景投影画像を重畳した結果である。図4(b)ではカメラ位置姿勢の計測誤差によって隠背景投影画像の重畳箇所に位置ずれが発生し、物体を違和感なく除去することができていない。一方提案手法によって隠背景投影画像を重畳した図4(c)では、位置ずれが補正されて幾何学的整合性を達成することができており、目視で確認しても違和感なく物体を除去できていることが分かる。図5に、図4とは異なる位置から物体を撮影し、隠背景投影画像を重畳した結果を示す。様々な位置姿勢で物体を撮影し隠背景投影画像を重畳した場合でも幾何学的不整合は発生せず、違和感なく物体の除去が実現できていることが確認できる。

動画において目視で確認をした場合も、カメラ位置姿勢のみを用いた手法ではほとんどのフレームで幾何学的不整合が発生しており、物体を違和感なく消すことはできていない。一方、提案手法を利用した場合は概ね幾何学的整合性を達成することができている。

次に、MSEを用いた定量的な比較結果を述べる。シーンを計10ヶ所から撮影した画像に対してMSEを求めた結果を表1に示す。いずれの場合においても、提案手法の方が良い結果が得られた。このことから提案手法では、カメラ位置姿勢のみを用いて隠背景投影画像を重畳する手法と比較して、幾何学的不整合を軽減した違和感の少ない重畳が可能であるといえる。

しかし、いくつかのフレームにおいて幾何学的不整合が発生する場合があります。全く違和感なく物体を消去するまでは至っていない。これはカメラの動きによって画像にモーションブラーが発生し、特徴点が検出できないことが原因として考えられる。

4. むすび

本稿では、物体を視覚的に除去するDRにおいて、隠背景投影画像の重畳時に発生する幾何学的不整合の軽減手法として、MRの一般的な位置合わせ手法であるカメラ位置姿勢推定法に加え、複数枚の画像を継ぎ目なく合成する画像処理技術であるイメージモザイクングを応用した手法を提案した。実際に、提案手法を用いて物体の除去を行い、目視による確認とMSEを用いた定量的な評価を行った。今後は、カメラの動きに対する頑健性の向上に取り組む予定である。

参考文献

- [1] 森尚平 他：“隠消現実感の技術的枠組と諸問題～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] F. I. Cosco *et al.*: “Augmented touch without visual obtusion,” Proc. ISMAR 2009, pp. 99 - 102, 2009.
- [3] 千葉直樹 他：“画像特徴に基づくイメージモザイクング”，電子情報通信学会論文誌，pp. 1581 - 1589, 1999-10-25.
- [4] H. Bay *et al.*: “Speeded-Up Robust Features (SURF),” Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346 - 359, 2008.
- [5] M. A. Fischler *et al.*: “Random sample con-sensus: a paradigm for model fitting with application to image analysis and automated cartography,” Communication Association and Computing Machine, Vol. 24, No. 6, pp. 381 - 395, 1981.