

隠消現実感における隠背景平面拘束と 周辺参照領域の効果的利用

森 尚平*1 小向 啓文*2 柴田 史久*1 木村 朝子*1 田村 秀行*3

Effective Use of Planar Structure of Hidden Area and Surrounding Region of Interest in Diminished Reality

Shohei Mori*1, Akifumi Komukai*2,
Fumihisa Shibata*1, Asako Kimura*1, and Hideyuki Tamura*3

Abstract --- In this paper, we propose a diminished reality (DR) system by which target objects in front of planar structure are effectively removed from the user's view. Our approach is to use a view-dependent image set captured before target objects are placed in an environment and to use common region (surrounding region of interest; SROI) between the view-dependent images and the user's view. First, in the proposed method, a target object in the user's view is overlaid by a selected view-dependent image using homography warping. For accurately projecting the view-dependent image, feature points are matched to ones only in SROI. During real-time process, lightings can change and can be different from that in the advanced capture of the view-dependent image set. Therefore, the projected view-dependent image is compensated using color information in SROI. In experiments, we used a 6DOF robot arm to capture ground truth image sequences, and compared them with our DR results. The quantitative results showed success and limits of the proposed method.

Keywords: Diminished Reality, Mixed Reality, Image Mosaicing, Color Correction

1. はじめに

現実世界と仮想世界の実時間での融合・表示をする拡張現実感 (Augmented Reality; AR) や複合現実感 (Mixed Reality; MR) は、人工現実感 (Virtual Reality; VR) 技術の発展形として位置づけられ、本特集号に至るまでも、活発な研究が行われている [1]。とりわけ、利用者／体験者の目の前の光景に仮想物体を CG 描画で重畳できることは、新たな情報提示方法として、様々な応用が期待されている [2]。

同じ発想の更なる発展形として、目の前の現実世界に実在する物体を視覚的に消し去る技術が求められるようになってきた。我々はこの技術を、除去対象物体で隠された背景領域 (隠背景) の情報で、その物体を体験者の視点位置から映像的に隠蔽・消去・透過する技術と捉え、「隠消現実感」(Diminished Reality; DR) と名付けて、その体系的整理を行った [3]。単純化して考えれば、AR/MR が情報の加算であるのに対して、DR は減算であると言える。

概念的には逆の演算であるとはいえ、AR/MR に

比べて、DR の方が困難な課題であり、実応用に対する障害は大きい。隠背景情報の実時間重畳描画による DR は、図 1 及び図 2 のようにモデル化できる。この重畳描画自体は AR/MR と同様な演算であるが、重畳することによる周辺領域との僅かな違いが、却って違和感を引き起こすことが少なくない。問題を簡略化して言うならば、「復元ずれ」「投影ずれ」「観測時間差によるずれ」の 3 要因が複雑に絡みあって、複合的な視覚的違和感を引き起こす。

隠背景の完全な復元をめざすには、その幾何学的属性と光学的属性の両方を、なるべく正確に観測しておくことが望ましい。これは、隠背景情報を事前取得しておく場合も、実時間取得して DR 実行する場合も同じである。限られた対象に対する応用であれば、許容範囲を考慮して、観測装置や重畳描画の画質を選ぶこともできるが、対象に依存しない万能の方法を得ることは、かなり困難な課題である。

この目標に対して、一步一步着実に研究開発の努力を積み重ねて行く研究アプローチを採る。違和感の原因が複合的であるならば、その要因を 1 つずつ分離分解し、なるべく簡単な条件下 (即ち、厳しい制約条件下) で、安定して作動する手法 (近似解) を開発し、徐々により複雑な条件にこれを拡張して行く。その第一歩として、床面、壁面等、隠背景は

*1 立命館大学大学院 情報理工学研究所

*2 キヤノン株式会社

*3 立命館大学 総合科学技術研究機構

*1 Graduate School of Science & Engineering, Ritsumeikan University

*2 Canon Inc.

*3 Research Organization of Science and Technology, Ritsumeikan University

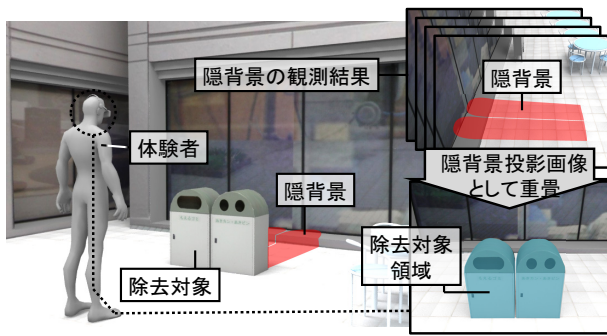


図 1 隠消現実感の実現

Fig. 1: Realization of diminished reality

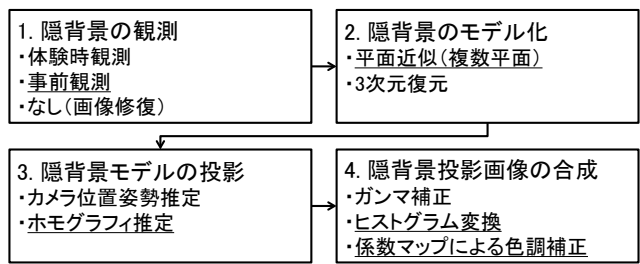


図 3 隠消現実感の実現手順

(下線は本手法でのアプローチ)

Fig. 3: DR implementation in steps

(Underlines show our approaches)

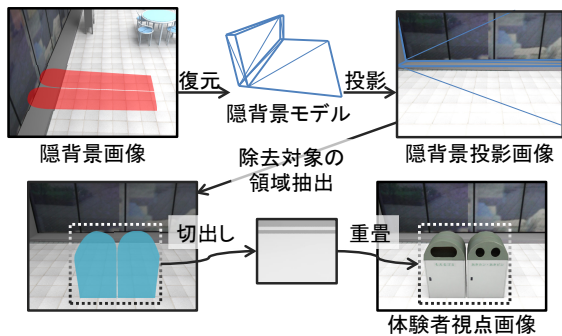


図 2 隠背景画像の取得・復元・体験者視点への投影と重畳

Fig. 2: Image acquisition of hidden area, reconstruction, projection and synthesis of the projected image

すべて平面近似できるという拘束条件を用いることにした。現実社会からの要請では、このような平面拘束の仮定を設け得る対象も少なくなく、またこの拘束下での既存研究もいくつか存在している。

本論文で述べる提案手法では、まず、事前に隠背景を多視点で観測し、視点依存画像を得ておくことで視点位置に依存して発生する光学的な復元ずれに関して対処する。つまり、提案手法では隠背景を事前に観測する方法をとるため、体験時と隠背景観測時との照明条件が異なる場合を想定し、重畳した隠背景投影画像に対して、後に定義する周辺参照領域を利用した色調補正を加え、観測時間差による光学的不整合の軽減を図る。投影ずれに関しては、隠背景投影画像と体験者視点画像中の周辺参照領域でホモグラフィ推定を行い、隠背景投影画像の重畳位置のずれを軽減することで解決を図る。

2. 関連研究

隠消現実感、隠背景の観測、隠背景のモデル化、隠背景モデルの投影、隠背景投影画像の合成の手順で実現できる (図 3)。ここでは、これらの手順における提案手法と先行研究との差異を述べる。

2.1 隠背景の観測

隠消現実感の実現方法は、まず隠背景の観測方法により、事前観測する方法と実時間観測する方法の

2 種に大別される。事前準備として除去対象を対象シーンから物理的に取り除き、隠背景を観測できることを想定するならば、事前に隠背景を観測しておく、体験時に除去対象を視覚的に除去する際に必要な情報を事前取得しておくことができる。そのため、効果的な視覚的除去が期待できる。ただし、事前準備時と体験時で対象シーンが幾何学的に変化したり、照明条件が変化したりする場合は、体験時にシーンと隠背景投影画像とで幾何学的及び光学的不整合が発生する要因となる。

Cosco ら[4]は、体験時に Image-based Rendering (IBR) により除去対象となる力覚デバイスを消して見せた。事前準備として除去対象を物理的に取り除いて隠背景を観測し、画像群とその撮影地点でのカメラ位置姿勢を保存し、視点依存画像として利用することで除去対象を除去した。一方、Barnum ら[5]は、壁に隠された隠背景を透視するため、また DR 体験時に変化する隠背景を観測するために隠背景観測用カメラを設置した。同様の方法は、津田ら[6]の See-Through Vision でも採用されている。

体験時に隠背景を観測していれば、事前準備で観測する手法よりも観測時間差が発生しない。しかし、体験者視点と隠背景観測用カメラが離れた位置に存在する場合、体験者視点とは見え方の異なる隠背景投影画像を得ることになるため、隠背景の復元ずれが発生する。そこで、提案手法では、Cosco らと同様に除去対象を取り除いて隠背景画像を撮影でき、隠背景の幾何形状が体験時に変化しないことを前提として、複数視点で隠背景画像を撮影し視点依存画像として保持することで復元ずれに対処する。更に、隠背景投影画像の合成時に色調補正処理を加えることで、観測時間ずれに対処する。

2.2 隠背景のモデル化

次に、隠背景のモデル化について述べる。これは対象シーンにより様々で、3 次元復元する手法、単純な幾何形状に近似する手法、そもそも隠背景を観測しない手法の 3 種に大別される。その内、特に本手法に関わる単純な幾何形状に近似する手法に関して

述べる。先述の Cosco らの手法では、隠背景の幾何形状は既知であるとして、机と壁の 2 平面程度の簡易な 3 次元モデルを用意している。Barnum らは、車載カメラを体験者視点とし、監視カメラを隠背景観測用カメラと想定している。この場合、隠背景は体験者視点から離れた建物の壁や人であり、それぞれを平面で近似できるという状況にあるため、隠背景は複数の平面で構成されているという前提を置いている。また、本田と斎藤[7]の手法では、壁面を隠背景、その前に立つ人等を障害物（除去対象）とする場合を想定している。そこで、平面である隠背景を 2 台のスマートフォンで観測し、ホモグラフィ（平面射影変換）により体験者視点に投影した後、メディア処理を用いて、体験者視点を含む合計 3 台のスマートフォンの画像から隠背景投影画像として適切な色情報を抽出している。

提案手法でもこれらの先行研究に倣い、隠背景は 1 枚の平面で近似できる、もしくは 2 平面の組み合わせによって構成されているとして除去対象を視覚的に除去することを対象にする。現実世界でこのような近似が可能な対象が数多く存在するだけでなく、美術館、博物館での展示、ショールームや商品展示会等では、人工的なこのような設定を無理なく行うことが可能な対象も存在している。

具体的には、提案手法により、ポスターや絵画の前に置かれた物体を除去したり、シーン中に置かれた AR マーカを除去したりといった応用が期待できる。提案手法と先行研究との最も大きな差異は、隠背景画像の取得方法及び後に定義する周辺参照領域を利用した隠背景投影画像の重畳と色調補正の枠組みにある。

2.3 隠背景モデルの投影

隠背景をモデル化した後は、それを体験者視点に投影する。この時の投影方法には大きく分けて 2 通りの方法が存在する。

1 つ目は、体験者視点のカメラ位置姿勢を取得し、そのカメラの画像平面に隠背景のモデルを投影する方法である。カメラ位置姿勢の取得方法には、センサを用いるものや、幾何マーカや Visual SLAM 等のビジョンベースのカメラ位置姿勢推定法を用いるものが存在する。例えば、Avery ら[8]の研究では GPS と姿勢センサを、Sandor ら[9]の研究では Visual SLAM の PTAM [10]を、Cosco らは幾何マーカによるカメラ位置姿勢推定法の ARToolKit [11]を用いている。1 章で指摘した通り、これらのカメラ位置姿勢取得方法は除去対象の除去結果の精度に直接的に影響を与える。

2 つ目は、平面間の射影変換であるホモグラフィ

を利用する方法である。DR においては、除去対象の視覚的除去が目的であるため、必ずしも体験者視点のカメラ位置姿勢を推定する必要がない。例えば、Barnum らの手法や本田と斎藤の手法では、隠背景を平面と仮定しているため、その平面と体験者視点画像平面間のホモグラフィを利用できる。特に、本田と斎藤の手法では、隠背景画像と体験者視点画像間で特徴点マッチングを行い、ホモグラフィを推定しているため、画像上の見た目のずれが少ないという特徴がある。

本手法でも本田と斎藤の手法と同様に、特徴点マッチングを行うアプローチを取ることで、隠背景の投影ずれを軽減する。ただし、特徴点マッチングの精度向上のため、また、隠背景の復元ずれを軽減するため、事前準備として隠背景観測結果とその時のカメラ位置姿勢とを組みにして複数用意しておくものとする。

2.4 隠背景投影画像の合成

特に、隠背景の事前観測を行う手法においては、隠背景投影画像の合成時に補正処理を加えることが重要となる。そのために先行研究では、体験者視点の画像と比較するための指標となるものを用意している。竹村ら[12]は HMD を視覚的に除去し、装着者の視線情報を復元する際、カラーチャート等で事前に取得した色情報を利用して隠背景投影画像の修正を行った。また、榎本ら[13]は、体験者のカメラ位置姿勢推定用に利用する ARTag [14]の白色部分を使って明度補正を行った。本手法では、カラーチャートや ARTag といった校正用の物体は利用せず、次章で定義する周辺参照領域を利用して隠背景投影画像の色調補正を行う。

Kawai ら[15]も提案手法と同様に、除去対象領域の周辺での色調変化を検出し隠背景投影画像に反映させる手法を提案している。ここでは、除去対象領域に影がかかるような場合に注目した処理も行われているが、本論文では、隠背景撮影前後での大まかな照明変化への対処法、及びその効果の実証に焦点を当てる。

3. 提案手法

3.1 想定シナリオと前提条件

本手法では、除去対象が壁の前に置かれている、もしくは壁と床といった 2 平面の前に置かれているとして、自由視点で移動する体験者の観測する光景から、その除去対象を視覚的に除去する。この際、平面と除去対象の位置関係や幾何形状は隠背景及び除去対象の 3 次元モデル（除去対象の場合はそれを覆うバウンディングボックス）により既知であると

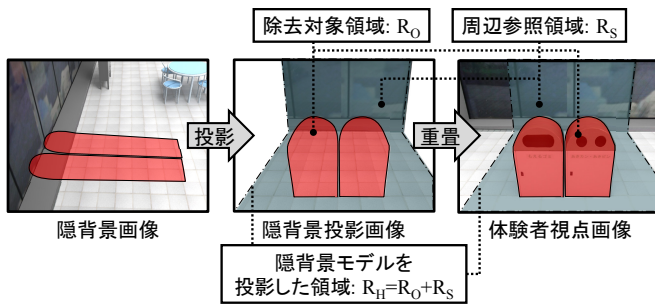


図4 周辺参照領域の定義

Fig. 4: Definition of surrounding region of interest

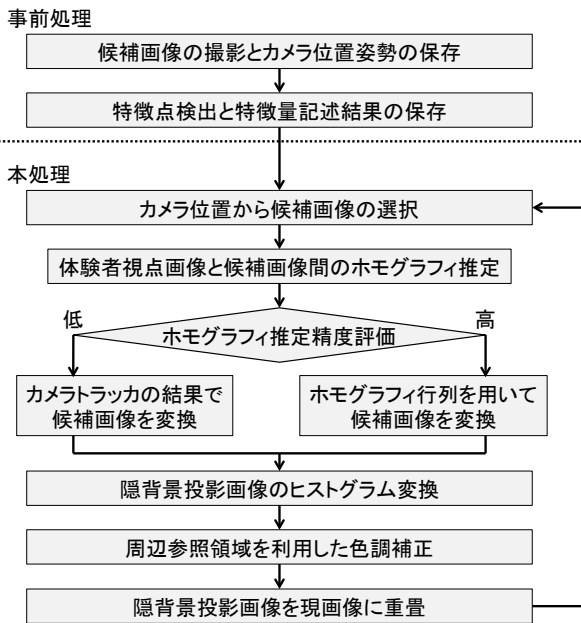


図5 提案手法のフロー図

Fig. 5: Flow of the proposed method

する。また、隠背景画像は除去対象を物理的に取り除くことで事前に取得しておく。そして、隠背景観測用カメラと体験者視点用カメラは同じであるとし、光学系ずれはないものとする。

3.2 周辺参照領域の定義

提案手法で重要な役割を果たす周辺参照領域とは、除去対象モデルを体験者視点画像及び隠背景投影画像に投影した除去対象領域 R_O と同様にして隠背景モデルを投影した領域 R_H があるとき、除去対象周りに現れる $R_H - R_O$ で表される領域 R_S である (図4)。

仮に、体験者のカメラ位置姿勢が正しく求まっており、除去対象領域の決定や隠背景のモデル化とそのモデルの投影が正しく行われていた場合、周辺参照領域において隠背景投影画像と体験者視点画像は一致する。よって、本手法では、この周辺参照領域において、特徴点マッチングを行う事でホモグラフィ行列を推定し、隠背景画像を体験者視点画像に重畳する。その後、周辺参照領域における体験者視点画像と隠背景投影画像の比較により隠背景投影画像

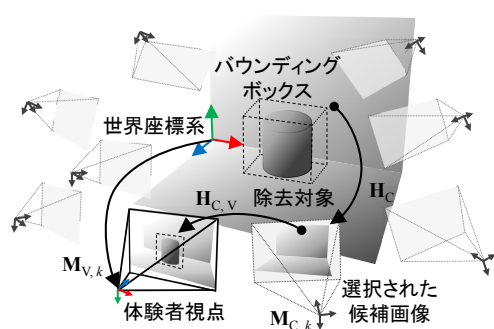


図6 候補画像群と体験者視点との幾何学的関係

Fig. 6: Geometric relations between a viewpoint and candidate images

の色調補正を行う。以降、図5に示す処理手順に則って説明する。

3.3 事前処理

提案手法では、除去対象を物理的に取り除いて事前に撮り溜めた N 枚の画像 (候補画像) $I_{C,k}$ ($1 \leq k \leq N$) とカメラトラックを用いて取得した撮影地点でのカメラ位置姿勢 $M_{C,k} = [R_{C,k} | t_{C,k}]$ ($1 \leq k \leq N$)、体験時に取得できる i フレーム目の体験者視点画像 $I_{V,i}$ とその撮影地点でのカメラ位置姿勢 $M_{V,i} = [R_{V,i} | t_{V,i}]$ を利用する。具体的には、DR 体験時のカメラ位置 $t_{V,i}$ に最も近い地点で撮影された候補画像 $I_{C,k}$ を選択し、体験者視点画像 $I_{V,i}$ に重畳する (図6)。

事前処理では、視差による隠背景の見え方の違いが見られない候補画像の不用意な記録を防止するために、カメラ位置 $t_{C,k}$ を、除去対象の中心におく注目点を原点とした極座標系において、その動径の長さ r を無視した2つの偏角 $\omega_{C,k} = [\theta_{C,k} \ \varphi_{C,k}]^T$ に変換して、一定の撮影間隔を隔てて記録する。

最後に、撮り溜めた候補画像全てに対して特徴点検出及び特徴量記述を行う。本手法では、回転不変性及びスケール不変性が特徴の SURF 特徴量[16]を用いる。

3.4 候補画像の選択

本処理の始めには、事前処理で記録した候補画像群から、体験者視点位置に最も近い地点で撮影された候補画像1枚を選択する。これには、候補画像撮影地点でのカメラ位置の極座標表現 $\omega_{C,k}$ と体験者視点位置の極座標表現 $\omega_{V,i}$ のユークリッド距離を最小にするものを選択する (式 (1))。

$$\arg \min_k (\|\omega_{C,k} - \omega_{V,i}\|) \quad (1)$$

これにより、体験者視点と候補画像の撮影位置の差異を減らし、隠背景の復元ずれを軽減する。

3.5 ホモグラフィ推定

3.5.1 候補画像との特徴点マッチング

ここでは、選ばれた候補画像と体験者視点画像と

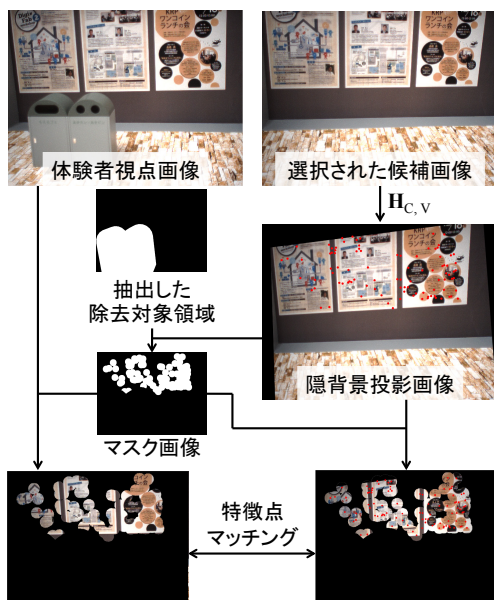


図7 体験者視点画像と候補画像との特徴点マッチング
Fig.7: Feature point matching between a viewpoint image and a selected candidate image

の間でSURF特徴量を用いた特徴点マッチングを行い、両画像間のホモグラフィ行列を求める。一般に特徴量記述子はアフィン変換に脆弱とされるが、本手法では体験者視点位置に近い位置で撮影された候補画像と体験者視点画像とのマッチングを行うためこうした問題は発生しにくいと考える。

また、特徴点マッチングの際に、画像中の全特徴点を対象として類似度が最大になる特徴点を探索した場合、計算量が増加する上、画像中での幾何学的な拘束を与えないため誤対応の発生確率が高まる。特徴点マッチングでの誤対応は最終的に算出されるホモグラフィ行列の精度の低下につながり、隠背景投影画像の重畳描画時に幾何学的不整合が発生する投影ずれの原因となる。そこで、カメラトラッカから得られる体験者視点でのカメラ位置姿勢 $\mathbf{M}_{V,k}$ と候補画像撮影地点でのカメラ位置姿勢 $\mathbf{M}_{C,k}$ を用いてマスク画像を作成し、特徴点マッチング時の探索範囲限定を行うことで、マッチング精度の向上を図る (図7)。

マスク画像の生成には、まず、体験者視点画像と候補画像間のホモグラフィ行列 $\mathbf{H}_{C,v}$ (式(2)) を用いて、候補画像中の特徴点の位置 $\mathbf{x}_c = [u_c \ v_c]^T$ を体験者視点画像に投影し、体験者視点画像上の点 $\mathbf{x}_{c,v} = [u_{c,v} \ v_{c,v}]^T$ を求める (式(3))。

$$\mathbf{H}_{C,v} = \mathbf{H}_v \mathbf{H}_c^{-1} \quad (2)$$

$$\mathbf{x}_{c,v} = \mathbf{H}_{C,v} \mathbf{x}_c \quad (3)$$

尚、 \mathbf{H}_v は体験者視点でのカメラ位置姿勢より求まるホモグラフィ行列、 \mathbf{H}_c は選択された候補画像のカメラ位置姿勢より求まるホモグラフィ行列である。



図8 ホモグラフィ推定精度が低い場合の対処法 (推定精度が低い場合 (左) カメラトラッカのみを用いて隠背景投影画像を重畳 (右))

Fig.8: Failure case of homography estimation (left) and the result using homography calculated by sensor pose (right)

この点 $\mathbf{x}_{c,v}$ を中心に半径 s [画素] で囲まれる領域を特徴点マッチングの探索範囲とする。ただし、除去対象モデルを体験者視点に投影して作られる領域は、この探索範囲から除外する。

しかし、このような特徴点マッチングにおける探索範囲の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、ロバスト推定法のRANSACアルゴリズム[17]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。

3.5.2 ホモグラフィ行列の推定精度が低い場合

特徴点を用いたホモグラフィ行列の推定では、推定精度が著しく低くなる2つの状況が考えられる。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。2つ目は、特徴点の誤対応が多い場合である。このような状況では、場合によってはカメラトラッカから得られたカメラ位置姿勢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隠背景の投影ずれが発生してしまう場合がある (図8左)。

そこで、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は、カメラ位置姿勢のみを用いた重畳を行うことで極端な投影ずれを防ぐ。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(2)で求めた候補画像を体験者視点画像に射影変換するホモグラフィ行列を $\mathbf{H}_{C,v}$ 、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 を式(4)のように表す。

$$\mathbf{H}_E = \mathbf{H}_{C,v}^{-1} \mathbf{H}_M \quad (4)$$

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、 $\mathbf{H}_{C,v}$ と \mathbf{H}_M は理想的には一致するため、 \mathbf{H}_E は単位行列となる。

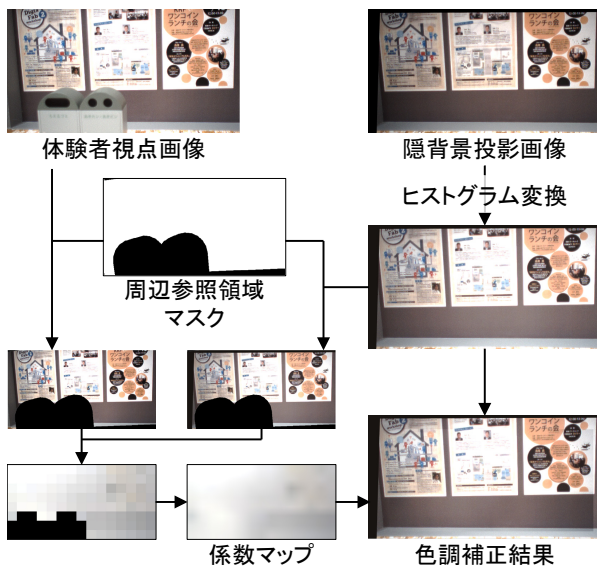


図9 隠背景投影画像の色調補正

Fig.9: Color correction in the proposed method

ホモグラフィ行列は投影の成分及び回転と並進の2つの行列に分解することが可能である。そこで、 H_E をそれらに分解し、回転と並進成分から得られた回転角と平行移動量にそれぞれ閾値 th_θ, th_l を設けてホモグラフィ行列の精度を評価する。閾値以上の平行移動量と回転角があった場合には、ホモグラフィ行列の精度が低いとみなし、カメラトラッカの結果から得られる $H_{C,v}$ を用いて隠背景投影画像の変換・重畳を行う (図8右)。

3.6 周辺参照領域を用いた色調補正

3.6.1 ヒストグラム変換

候補画像撮影時と体験時とでは照明条件が異なる場合があり、その差が光学的不整合の発生要因となり得る。本手法では、この光学的不整合を軽減するために、周辺参照領域を利用する。図9に示すのは、本節の周辺参照領域を利用した隠背景投影画像の色調補正処理の流れである。

まず始めに、隠背景投影画像全体を体験者視点画像の色味に近づけるため、隠背景投影画像と体験者視点画像の周辺参照領域全体でのヒストグラムを求め、色調補正を行う (式(5))。

$$I_v(x_v, y_v) = \frac{\sigma_v(I_c(x_c, y_c) - m_c)}{\sigma_c} + m_v \quad (5)$$

ただし、 $I_v(x, y)$ は I_v の画像座標 (x, y) 上の画素値、 $I_c(x, y)$ は I_c の画像座標 (x, y) 上の画素値、 m_v, σ_v はそれぞれ I_v での周辺参照領域中の画素値の平均と標準偏差、 m_c, σ_c はそれぞれ I_c での周辺参照領域中の画素値の平均と標準偏差である。

3.6.2 係数マップによる色調補正

前項の処理により隠背景投影画像全体の色調補正

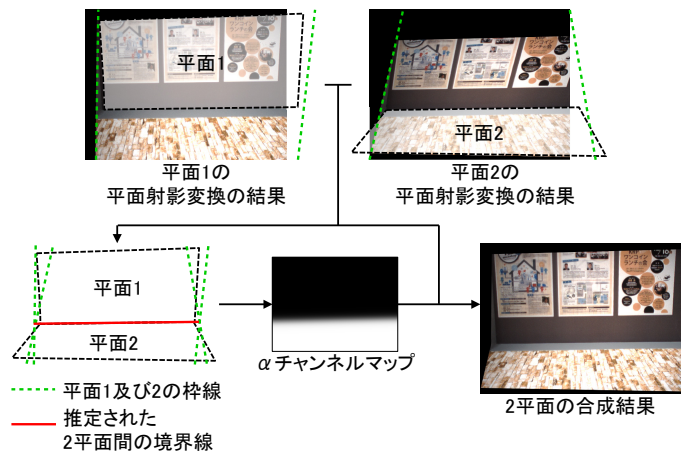


図10 2平面の合成

Fig.10: Synthesis of two planes

を施した後、隠背景投影画像のより局所的な色調変化に対処するため、係数マップを作成しそれを用いて色調補正を施す。まず、体験者視点画像と隠背景投影画像を $m \times n$ のグリッドに分割し、周辺参照領域に含まれる各小領域中の画素値の平均を算出し、グリッド上での位置を (x_G, y_G) として、それぞれを $I_{v,M}(x_G, y_G), I_{c,M}(x_G, y_G)$ と表す。この時、グリッド上での位置 (x_G, y_G) での係数値 $k(x_G, y_G)$ を次の式で表す。

$$k(x_G, y_G) = \frac{I_{v,M}(x_G, y_G)}{I_{c,M}(x_G, y_G)} \quad (6)$$

ただし、隠背景投影画像と体験者視点画像の除去対象領域での係数値は計算できない。そこで、この領域が係数値で埋まるまで、8近傍における有効なグリッドの平均値をとる処理を繰り返す。最後に、係数マップを隠背景投影画像のサイズに合わせるため、内挿により補間する (図9中の係数マップ)。こうして作成された係数マップを隠背景投影画像の画素値に乗算することで、隠背景投影画像の色調補正を行う。

3.7 2平面への拡張

これまで背景が壁等の1平面で構成されている場合を述べてきた。ここでは、背景が壁と床で構成されたような2平面で構成されている場合について述べる。本手法では、各平面でこれまで述べてきた処理を独立に実行し、両結果を合成することで背景が2平面で構成されている場合に対処する (図10)。

まず、シーン中の2平面各々に対してホモグラフィ推定を行い、候補画像を体験者視点画像に合わせて変換する。次に、それぞれの画像の枠線の交点を画像の両端で求め、それをつなぐ線分を平面間の境界線とする。そして、この境界線を境に両変換結果を合成する。合成の際は、その境界が目立たぬよう、

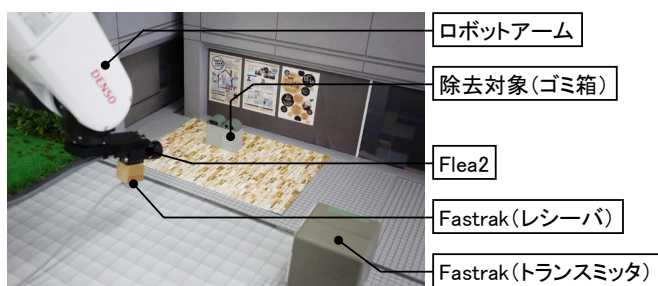


図 11 実験の様子

Fig.11: Qualitative experiment system configuration

シグモイド関数を利用して作成した α チャンネルマップを用いて、2 平面の α ブレンディングを行う。

4. 実験と考察

4.1 実験内容とシステム構成

本手法が有効に機能することを確認するため以下の実験を行う。

実験 1: 候補画像取得時と体験時とで照明変化のない対象シーン (ミニチュアセット (1/12 スケール)) において、除去対象を除去できるか

実験 2: 候補画像取得時と体験時とで照明変化のある実験的な対象シーンにおいて、除去対象を除去できるか

全ての実験においてのシステム構成は以下の通りである。DR 処理には Windows XP OS, CPU に Intel Core i7 2.93 GHz, 3.50 GB メモリ, GPU に NVIDIA GeForce GTX 280 を搭載する PC を用いた。候補画像取得及び体験時のカメラ位置姿勢取得 (カメラトラッカ) には、磁気式位置姿勢取得用センサの Polhemus 社製 Fastrak を利用した。カメラとそのレンズには PointGrey 社製 Flea2 とミュートロン社製 FV0622 (水平画角 40.5 度) を用いた。画像サイズは 640×480 画素である。統合開発環境には Microsoft Visual Studio 2010, 開発言語には C++ を用いた。また、グラフィックス API には DirectX 9 を用いた。

実験 1 及び実験 2 において、除去対象が消えているかどうかを定量的に示すため、位置繰り返し精度 ± 0.03 [mm] のエンコーダを備えた垂直多関節ロボットであるデンソーウェーブ社製 VS-087 (以降、ロボットアーム) を用いる。具体的には、除去対象を置いたシーンにて撮影した体験者視点画像シーケンス $I_{v,i}$ と除去対象を物理的に取り除いてシーンを撮影した画像シーケンス $I_{R,i}$ を同じカメラパスでフレーム数 M だけ撮影し、除去結果 $I_{DR,i}$ との平均自乗誤差 (MSE) を計算することで評価する (式 (7))。

$$MSE_i = \frac{1}{M} \left(I_{R,i} - I_{DR,i} \right)^2 \quad (7)$$

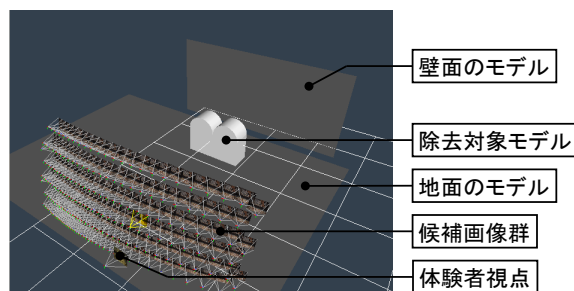


図 12 撮影された候補画像の可視化

Fig.12: Input data for DR in the quantitative experiments

太陽光によって照らされる屋外シーンを想定し、照明には、太陽光にあたるキー照明として ARRI 社製 HMI575W を 1 基、環境光にあたるアンビエント照明として ARRI 社製 650 Plus を 2 基用いた。尚、アンビエント照明には色変換フィルタを付け、キー照明と同等の色温度に変換した。実験 2 での照明条件の変更は、キー照明のみの位置を変更することで行う。これは、太陽の位置が変化したことを想定している。図 11 に実験の様子を示す。

4.2 実験 1: 除去対象の除去(照明変化なし)

実験 1 では、ミニチュアセットにて、候補画像取得時と体験時とで照明変化がないという条件のもと、除去対象を視覚的に除去できるかを確認する。尚、本実験では、キー照明はセット右手に置いた。

まず、事前準備として、除去対象設置前のシーンにおいて、体験時に除去対象が置かれる箇所に注視点を置き、その位置を取り囲むように 35 cm 程度離れたところから 200 枚の候補画像を撮影した (図 12)。次に、除去対象から約 35 cm 離れた位置を初期位置とし、体験者視点カメラを 6 自由度で 300 枚の体験者視点画像シーケンスを撮影した。最後に、定量評価を行うため、体験者視点画像シーケンス撮影時と同じカメラパスで、除去対象を物理的に取り除いたシーンを 300 枚撮影した。図 13 に示すのは、100, 200, そして 300 フレーム目の体験者視点画像と除去対象の除去結果の参照画像である。

尚、候補画像の撮影枚数が多いほど高精度な DR 処理結果が期待できる。処理負荷に関しては、候補画像の選択基準である式 (1) に示す程度のみであり、むしろメモリ容量と DR 処理結果の精度とのトレードオフという側面が強い。

本実験では、提案手法、提案手法の色調補正処理を除いた場合 (以降、提案手法 (色調補正なし)), カメラトラッカのみを用いる場合 (以降、基本手法) の 3 つの手法で除去対象を視覚的に除去した結果を比較する。図 14 に示すのは、上記の 3 つの手法で体験者視点画像群から除去対象を除去した結果である。処理時間の平均はそれぞれ約 809 ms, 約 618 ms,



図 13 実験 1 での体験者視点画像シーケンス (上段) と除去結果の参照画像 (下段) (左から, 100, 200, 300 フレーム目)

Fig. 13: Viewer's images (1st row) and reference images of visual removal (2nd row) in Exp. 1 (from left: 100th, 200th, and 300th frame)



図 14 実験 1 での提案手法 (上段), 提案手法 (色調補正なし) (中段), 基本手法 (下段) の除去結果の比較 (左から, 100, 200, 300 フレーム目)

Fig. 14: Comparison of the proposed method (1st row), the proposed method (w/o color correction) (2nd row), and basic method (3rd row) in Exp. 1 (from left: 100th, 200th, and 300th frame)

約 534 ms であった。

図 14 に示されるように, 提案手法を用いた場合が最も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。300 フレーム目では, これら 3 手法の違いが特に顕著に現れている。基本手法では, カメラ位置姿勢計測が低精度であることから隠背景モデルの投影ずれが発生しており, これが原因で隠背景投影画像の重畳箇所が目立つ結果となった。また, 提案手法 (色調補正なし) では, 特徴点ベースの隠背景投影画像の重畳処理により隠背景モデルの投影ずれこそ見られないが, 隠背景投影画像の重畳箇所の右端で光学的不整合が発生し, 隠背景投影画像の重畳箇所がその周りよりも明るく写ってしまっている。一方, 提案手法では, こうしたずれが軽減され, 除去結果の参照画像に近い結果が得られていることが分かる。

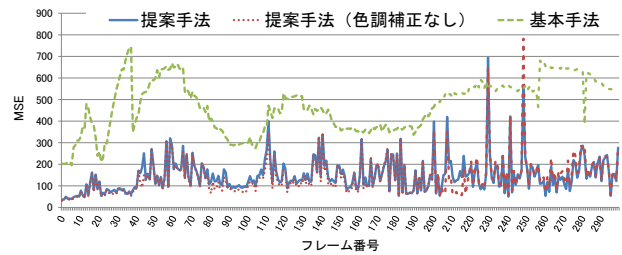


図 15 実験 1 における各種法での MSE の比較

Fig. 15: Comparison of MSE b/w methods in Exp. 1



図 16 色調補正処理の失敗例

Fig. 16: Failure cases of color correction process

また, 図 15 に示す MSE の推移からも分かる通り, 提案手法や提案手法 (色調補正なし) を用いる場合, 高精度な除去対象の視覚的除去が実現できていることが分かる。尚, 提案手法, 提案手法 (色調補正なし), 基本手法の MSE の平均はそれぞれ約 147.83, 140.64, 473.00 であった。MSE の平均値において提案手法が提案手法 (色調補正なし) に劣るのは, そもそも隠背景画像撮影時と DR 体験時での照明変化がないシーンであったため, 特徴点ベースの隠背景投影画像の重畳処理を行っても解消しきれなかった隠背景モデルの投影ずれが原因で, その後の色調補正処理が上手く機能せず, 除去結果に悪影響を与えていることが原因であると考えられる。例えば, 図 16 左では中央のポスターの右下が青白く変色しており, 同図右では黒く変色しているように見える。

4.3 実験 2: 除去対象の除去 (照明変化あり)

実験 2 では, 候補画像取得時と体験時とで照明が変化する実験的な対象シーン (ミニチュアセット) において, 除去対象を視覚的に除去できるかを確認する。本実験は, 実験 1 でのミニチュアの配置と同一のシーン及び同一の候補画像群を用いて行うが, DR 体験時には候補画像撮影時に対して時間が経過したことを想定し, 照明条件を変更する。これにより, 意図的に観測時間差による光学的不整合が発生しやすい状況を作り出し, 提案手法がこの不整合の軽減に貢献できるかを確認する。照明条件を変更するため, 候補画像撮影時にセット右手に設置していたキー照明を, 本実験ではセット左手に設置した。

本実験でも, 実験 1 と同じカメラパスで体験者視点画像シーケンスを撮り溜め, その後, 設置された

除去対象を物理的に取り除き、体験者視点画像撮影時と同じカメラパスで除去結果の参照値となる画像シーケンスを撮影した（図 17）。尚、本実験でも実験 1 と同様に、提案手法、提案手法（色調補正なし）、基本手法を用いる場合の 3 つの方法で除去対象を視覚的に除去した結果を比較する。

図 18 に示すのは、上記の 3 つの方法で体験者視点画像シーケンスから除去対象を除去した結果である。尚、処理時間の平均はそれぞれ約 837 ms, 約 646 ms, 約 532 ms であった。図 18 に示されるように、提案手法を用いた場合が最も高精度に除去対象の視覚的除去を達成できていることが分かる。具体的には、基本手法で発生している隠背景モデルの投影ずれ、そして、基本手法と提案手法（色調補正なし）で見受けられる観測時間差による光学的不整合が提案手法では軽減されていることが見て取れる。しかし、実験 1 と同様の理由で、色調補正が上手く機能せず、疑似色のようなものが見受けられるフレームも存在した。

図 19 に示すのは、本実験における MSE の推移である。尚、提案手法、提案手法（色調補正なし）、基本手法での MSE の平均は、それぞれ 261.44, 398.23, 833.53 であった。このことから、提案手法を用いることで、カメラトラッカが低精度であるために発生する幾何学的不整合を軽減し、隠背景画像取得時と DR 体験時での照明条件の違いによって発生する光学的不整合を軽減できることが分かった。

5. まとめ

本論文では、壁や地面といった 2 平面の前に置かれた物体を視覚的に除去する手法を提案した。具体的には、除去対象の写らないシーンをカメラ位置姿勢とともに候補画像群として記録しておき、DR 体験時に利用することで、効果的に除去対象を除去した。この際、背景が平面であるという幾何的拘束のもと、周辺参照領域として定義した隠背景投影画像と体験者視点画像とで共通する領域において、両画像間の特徴点マッチングとホモグラフィ推定によって隠背景投影画像の重畳時の位置ずれを補正した。また、候補画像撮影時と DR 体験時には照明条件が異なることを想定し、周辺参照領域を利用した色調補正処理を施すことで、光学的不整合に対処した。

ミニチュアセットとロボットアームを利用した実験では、DR 体験時と同じカメラパスで除去対象が写っていないシーンを撮影して参照値を取得し、提案手法を含 3 つの手法を比較することで、提案手法が 3 手法の中で最も有効に機能することを示した。今後の発展としては、磁気センサ等のカメラ位置姿勢取得用センサを用いない手法への拡張や隠背景が



図 17 実験 2 での体験者視点画像シーケンス（上段）と除去結果の参照画像（下段）（左から、100, 200, 300 フレーム目）
Fig. 17: Viewer's images (1st row) and reference images of visual removal (2nd row) in Exp. 2 (from left: 100th, 200th, and 300th frame)



図 18 実験 2 での提案手法（上段）、提案手法（色調補正なし）（中段）、基本手法（下段）の除去結果の比較（左から、100, 200, 300 フレーム目）
Fig. 18: Comparison of the proposed method (1st row), the proposed method (w/o color correction) (2nd row), and basic method (3rd row) in Exp. 2 (from left: 100th, 200th, and 300th frame)

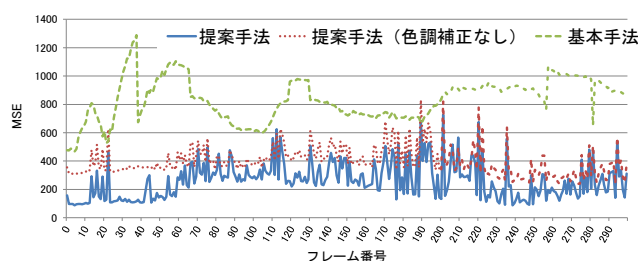


図 19 実験 2 における各種法での MSE の比較
Fig. 19: Comparison of MSE b/w three methods in Exp. 2

より複雑な場合への対処法に取り組む。

謝辞

ミニチュアセットの建設に関してご協力を頂いた東映（株）京都撮影所次長 竹村寧人氏、ミニチュアの照明に関する助言を頂いた映画撮影監督 坂本誠吾氏に感謝する。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」による。

fitting with applications to image analysis and automated cartography,” *Comm. of the ACM*, Vol. 24, pp. 381 - 395, 1981.

(2013年12月9日受付)

参考文献

[著者紹介]

- [1] 特集号「複合現実感 1~6」, 日本 VR 学会論文誌 (1999 - 2014)
- [2] 特集 拡張現実感 (AR), 情報処理, Vol. 51, No. 4, 2010.
- [3] 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.
- [4] F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M. A. Otaduy: “Augmented touch without visual obtrusion,” *Proc. ISMAR 2009*, pp. 99 - 102, 2009.
- [5] P. Barnum, Y. Sheikh, A. Datta, and T. Kanade: “Dynamic seethroughs: Synthesizing hidden views of moving objects,” *Proc. ISMAR 2009*, pp. 111 - 114, 2009.
- [6] 津田崇博, 山本治由, 亀田能成, 大田友一: “死角を透視表示する屋外型複合現実感システムにおける提示手法の比較検討”, 信学技報, MVE, Vol. 105, No. 256, pp. 41 - 47, 2005.
- [7] 本田俊博, 斎藤英雄: “複数のスマートフォンカメラの協調利用による実時間隠消現実感”, 日本 VR 学会論文誌, Vol. 17, No.3, pp. 181 - 190, 2012.
- [8] B. Avery, C. Sandor, and B. H. Thomas: “Improving spatial perception for augmented reality x-ray vision,” *Proc. IEEE VR 2009*, pp. 79 - 82, 2009.
- [9] C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila: “An augmented reality x-ray system based on visual saliency,” *Proc. ISMAR 2010*, pp. 27 - 36, 2010.
- [10] G. Klein and D. Murray: “Parallel tracking and mapping for small AR workspaces,” *Proc. ISMAR 2007*, pp. 225 - 234, 2007.
- [11] H. Kato and M. Billinghurst: “Marker tracking and HMD calibration for a video-based augmented reality conferencing system,” *Proc. IWAR 1999*, p.85 - 94, 1999.
- [12] 竹村雅幸, 北原格, 大田友一: “MR Face 映像における視覚的特性を考慮した光学的整合性の実現”, 画像ラボ, Vol. 19, No. 2, pp. 48 - 53, 2008.
- [13] 榎本暁人, 斎藤英雄: “複数のハンディカメラの協調利用による遮蔽物体除去映像のオンライン生成”, 映情学誌, Vol. 62, No. 6, pp. 901 - 908, 2008.
- [14] M. Fiala: “ARTag, a fiducial marker system using digital techniques,” *Proc. CVPR 2005*, Vol.2, pp. 590 - 596, 2005.
- [15] N. Kawai, M. Yamasaki, T. Sato, and N. Yokoya: “Diminished reality for AR marker hiding based on image inpainting with reflection of luminance changes,” *ITE Trans. on Media Technology and Applications*, Vol. 1, No. 4, pp. 343 - 353, 2013.
- [16] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool: “Speeded-Up Robust Features (SURF),” *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346 - 359, 2008.
- [17] M. A. Fischler, and R. C. Bolles: “A paradigm for model

森 尚平 (学生会員)



2011年立命館大学情報理工学部メディア情報学科卒。2013年同大学院理工学研究科博士前期課程修了。現在、同大学院情報理工学研究科博士後期課程在学中。複合現実感及び隠消現実感の研究に従事。

小向 啓文 (学生会員)



2012年立命館大学情報理工学部メディア情報学科卒。2014年、同大学院情報理工学研究科博士前期課程修了。同年4月、キヤノン(株)入社。2011年より2014年まで、隠消現実感における幾何学的不整合及び光学的不整合の解消に関する研究に従事。

柴田 史久 (正会員)



1999年阪大・基礎工・博士後期課程了。現在、立命館大学情報理工学部教授。博士(工学)。経歴の詳細は本号の他論文参照。

木村 朝子 (正会員)



1998年阪大・基礎工・博士前期課程了。現在、立命館大学情報理工学部教授。博士(工学)。経歴の詳細は本号の他論文参照。

田村 秀行 (正会員)



1970年京都大学工学部電気工学科卒。工業技術院電子技術総合研究所、キヤノン(株)等を経て、2003年4月より立命館大学理工学部教授。現在、同大学総合科学技術研究機構(特別招聘)教授。工学博士。1997年より2001年まで、MRシステム研究所にて「複合現実感研究プロジェクト」を率いた。本学会フェロー、元理事、現在、評議員、複合現実感研究委員会顧問。編著書「Mixed Reality」(Ohmsha & Springer)「コンピュータ画像処理」(オーム社)など。電子情報通信学会フェロー、IEEE、ACM、情報処理学会、人工知能学会、映像情報メディア学会等の会員。本学会及び情報処理学会論文賞、人工知能学会功労賞等を受賞。