

NAIST-IS-MT1151020

## 修士論文

# 拡張現実感のための視点依存テクスチャ・ジオメトリに 基づく仮想化実物体の提示

宇野 祐介

2013年 3月 15日

奈良先端科学技術大学院大学  
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に  
修士(工学) 授与の要件として提出した修士論文である。

宇野 祐介

審査委員：

|           |         |
|-----------|---------|
| 横矢 直和 教授  | (主指導教員) |
| 加藤 博一 教授  | (副指導教員) |
| 佐藤 智和 准教授 | (副指導教員) |
| 中島 悠太 助教  | (副指導教員) |

# 拡張現実感のための視点依存テクスチャ・ジオメトリに 基づく仮想化実物体の提示\*

宇野 祐介

## 内容梗概

近年、拡張現実感 (Augmented Reality; AR) 技術をスマートフォン等の携帯端末上に実装することで、実空間を撮影した映像上で家具や家電のような実物体を仮想的に配置するアプリケーションが一般に利用され始めている。このようなアプリケーションでは、仮想化された実物体 (仮想化実物体) の任意の視点からの見えを再現する必要がある。多くの場合において、設計図等に基づいて人手で作成されるテクスチャ付き三次元モデルが利用されている。一方、コンピュータビジョンの分野においては、実物体を撮影した画像群を入力として、自動で三次元復元を行うことで、人手によるモデリング作業を介さずにユーザ自身による実物体の仮想化を実現する手法が提案されている。このような手法により、画像群から自動生成される三次元モデルを用いてユーザ自身が所持する物体を仮想化し、拡張現実感アプリケーションに応用することが期待されるが、自動で生成される三次元モデルには、推定誤差に起因して実際とは異なる形状やテクスチャの不整合が視覚的に提示されるため、生成される画像の品質に問題がある。この問題に対し、本研究では視点依存テクスチャと視点依存ジオメトリを用いることで、実画像中に仮想化実物体を高品位に重畳合成する手法を提案する。具体的には、仮想視点に依存した適切なテクスチャを選択的に用いる視点依存テクスチャを利用することで、生成される画像上に生じるテクスチャの不整合を解消する。ただし、視点依存テクスチャを用いた場合においては、生成される自由視点画像の輪郭上

---

\* 奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1151020, 2013年3月15日.

において、不自然な形状の膨張や欠損が生じるという問題があり、これに対して、本研究では、仮想視点に依存して三次元モデルを変形する視点依存ジオメトリを併用することで、三次元モデルの欠損領域を補完し、これに加えて前景抽出した撮影画像群から視点位置に応じたテクスチャを選択して前景領域上の画像のみによるマッピングを行うことで、膨張領域を修復する。実験では、実時間で仮想化実物体の合成を行う拡張現実感システムを実装し、実画像に対して既存のマルチビューステレオ法を適用することで得られた三次元モデルと事前に前景抽出した画像群を入力として用いる提案手法を、モデルベースドレンダリングに基づく手法、及び視点依存テクスチャに基づく手法と比較することで提案手法の有効性を示す。

## キーワード

拡張現実感, 仮想化実物体, 自由視点画像生成, 視点依存テクスチャ, 視点依存ジオメトリ

# Generation of Augmented Reality Images with Virtualized Real Objects Using View-dependent Texture and Geometry\*

Yusuke Uno

## Abstract

Recently, augmented reality (AR)-based applications are emerging and are enabling us to virtually arrange a real object such as a home appliance and a piece of furniture. Such an application superimposes an image of a virtualized real object, which is a virtual object imitating a real object, with appropriate pose on a live video stream capturing the real space. One approach for generating such an image of virtualized real object from an arbitrary viewpoint is to build a three-dimensional (3D) CG model of the real object manually; e.g. , based on its design drawing, which is usually a laborious task and requires special skills for 3D modeling. Another possible approach is to use a computer vision technique called image based 3D reconstruction. This technique automatically builds a 3D model from multiple view images of a real object. This approach is preferable to the manual one, because it allows a user without the special skills to build a 3D model of a real object. However, automatic 3D reconstruction techniques raise a problem in the accuracy of 3D shape estimation, which leads such undesirable effects as unsmooth surfaces and blurred textures. This results in visual artifacts in generated arbitrary viewpoint images and degrades their visual quality.

---

\* Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1151020, March 15, 2013.

To overcome this problem, we propose a method that uses view-dependent texture mapping (VDTM) and geometry (VDG). Our method generates high-quality arbitrary viewpoint images from a 3D model built by an image-based 3D reconstruction technique and view images used for the 3D reconstruction. It firstly modifies the 3D model using VDG and applies textures extracted from the view images to the 3D model using VDTM. The modified 3D model reduces visual artifacts caused by missing region of 3D model, especially around the contour of the virtualized real object, and VDTM makes errors in 3D model less noticeable. Since our VDG algorithm may induce excessive dilation of the 3D model, which consequently results in applying textures containing background regions to the 3D model, we extract the regions of the real object in the real images as foreground regions and use them as textures. To evaluate our method, we have built a prototype AR system and applied our method to 3D model generated by an existing multi-view stereo technique with manually extracted foreground regions. Our experiments have successfully demonstrated that the proposed method improves the visual quality of generated arbitrary viewpoint images compared with baseline methods, i.e., a model-based and a VDTM-based methods.

**Keywords:**

Augmented reality, Virtualized real object, Free-viewpoint image generation, View-dependent texture, View-dependent geometry

# 目次

|                                   |    |
|-----------------------------------|----|
| 1. はじめに                           | 1  |
| 2. 関連研究と本研究の位置付け                  | 4  |
| 2.1 仮想化実物体の自由視点画像生成の関連研究          | 4  |
| 2.1.1 モデルベースドレンダリング               | 4  |
| 2.1.2 イメージベースドレンダリング              | 7  |
| 2.1.3 ハイブリッドレンダリング                | 10 |
| 2.2 本研究の位置付けと方針                   | 12 |
| 3. 視点依存テクスチャ・ジオメトリによる膨張領域と欠損領域の修復 | 14 |
| 3.1 提案手法の概要                       | 14 |
| 3.2 座標系の定義                        | 16 |
| 3.3 オフライン処理                       | 17 |
| 3.3.1 撮影画像のカメラ内部・外部パラメータの推定       | 17 |
| 3.3.2 三次元モデルの構築                   | 18 |
| 3.3.3 前景領域の抽出                     | 18 |
| 3.4 オンライン処理                       | 19 |
| 3.4.1 実環境を撮影するカメラの位置・姿勢推定         | 19 |
| 3.4.2 視点に依存した奥行き画像の生成             | 20 |
| 3.4.3 視点に依存した適切なテクスチャの選択          | 21 |
| 3.4.4 自由視点画像の生成と実環境画像への重畳         | 25 |
| 4. 実験                             | 28 |
| 4.1 実験の概要                         | 28 |
| 4.2 画像データセットと実装の詳細                | 29 |
| 4.3 生成画像の比較                       | 29 |
| 4.4 考察                            | 43 |
| 5. 提案手法のアプリケーション例                 | 45 |

|        |    |
|--------|----|
| 6. まとめ | 47 |
| 謝辞     | 48 |
| 参考文献   | 49 |

## 目 次

|    |   |    |
|----|---|----|
| 1  | CMPMVS [3] で生成された三次元モデルの例 . . . . .                 | 2  |
| 2  | 視点依存テクスチャを適用した仮想化実物体の自由視点画像の例 .                     | 3  |
| 3  | レーザレンジファインダによる三次元復元 [10] の例 . . . . .               | 5  |
| 4  | マルチビューステレオ法 [2] により生成された三次元モデルの例 .                  | 6  |
| 5  | Bastian ら [6] の手法 . . . . .                         | 7  |
| 6  | 変型ビルボードによる協調作業システム [20] . . . . .                   | 10 |
| 7  | 視点依存テクスチャ法 [7] . . . . .                            | 11 |
| 8  | Prince ら [25] の手法 . . . . .                         | 12 |
| 9  | 提案手法の概要 . . . . .                                   | 15 |
| 10 | 世界座標系, オブジェクト座標系, カメラ座標系, 仮想視点座標<br>系の関係 . . . . .  | 17 |
| 11 | 前景抽出した撮影画像の例 . . . . .                              | 19 |
| 12 | カメラ $C_E$ における画素 $n$ に対応する光線と三次元モデル $M$ . . .       | 20 |
| 13 | 奥行き画像 $D_E$ に対する拡張処理 . . . . .                      | 21 |
| 14 | カメラ $C_E$ に依存した三次元モデル $M$ の奥行き画像 . . . . .          | 22 |
| 15 | 点 $P_n$ に対する画像 $I_k$ を撮影したカメラ $C_k$ の選択基準 . . . . . | 23 |
| 16 | 自己遮蔽によって誤ったテクスチャが選択される状況 . . . . .                  | 24 |
| 17 | 生成された自由視点画像, 実環境画像, 拡張現実画像の例 . . . . .              | 27 |
| 18 | 画像データセット 1(DS1) . . . . .                           | 30 |
| 19 | 画像データセット 2(DS2) . . . . .                           | 30 |
| 20 | 画像データセット 3(DS3) . . . . .                           | 31 |
| 21 | CMPMVS [3] により DS1 を用いて生成した三次元モデル . . . . .         | 31 |
| 22 | CMPMVS [3] により DS2 を用いて生成した三次元モデル . . . . .         | 32 |
| 23 | CMPMVS [3] により DS3 を用いて生成した三次元モデル . . . . .         | 32 |
| 24 | DS1 を用いた従来手法 1 による生成結果 . . . . .                    | 34 |
| 25 | DS1 を用いた従来手法 2 による生成結果 . . . . .                    | 35 |
| 26 | DS1 を用いた提案手法による生成結果 . . . . .                       | 36 |
| 27 | DS2 を用いた従来手法 1 による生成結果 . . . . .                    | 37 |

|    |   |    |
|----|---|----|
| 28 | DS2 を用いた従来手法 2 による生成結果 . . . . .        | 38 |
| 29 | DS2 を用いた提案手法による生成結果 . . . . .           | 39 |
| 30 | DS3 を用いた従来手法 1 による生成結果 . . . . .        | 40 |
| 31 | DS3 を用いた従来手法 2 による生成結果 . . . . .        | 41 |
| 32 | DS3 を用いた提案手法による生成結果 . . . . .           | 42 |
| 33 | テクスチャとして選択される画像の切り替わりの例 (DS2) . . . . . | 43 |
| 34 | 提案手法による東塔の重畳例 . . . . .                 | 46 |

## 表 目 次

|   |   |    |
|---|---|----|
| 1 | 実験に用いた PC の仕様 . . . . .                 | 28 |
| 2 | 各データセットの提案手法における実験のパラメータ . . . . .      | 33 |
| 3 | DS1, DS2, DS3 における各手法のフレームレート . . . . . | 33 |

# 1. はじめに

スマートフォン等の携帯端末において、実空間を撮影した映像上に家具や家電などの実物体を仮想的に配置する拡張現実感 (Augmented Reality; AR) アプリケーションが一般に利用され始めている。このようなアプリケーションでは、仮想化された実物体 (仮想化実物体) の任意の視点における見えを再現する必要がある。多くの場合、設計図等に基づいて人手で作成されるテクスチャ付きの三次元モデルが利用されている。しかし、手作業による高品位な三次元モデルの生成は高度な技術が必要であるため、一般のユーザには困難である。一方、コンピュータビジョンの分野においては、図 1(a) に示すような実物体を撮影した画像群から、自動で図 1(b) のような三次元モデルを生成することで、人手によるモデリング作業を介さずにユーザ自身による実物体の仮想化を実現する手法が提案されている [1-6]。このような手法を用いることで、ユーザ自身が所持する物体をスマートフォン等のカメラを用いて手軽に仮想化し、AR アプリケーションで利用することができる。しかし、画像のみから自動で生成される三次元モデルには、推定誤差に起因して実際とは異なる形状やテクスチャの不整合が視覚的に提示される場合がある (図 1(b))。

テクスチャの不整合は、仮想視点に依存した適切なテクスチャを選択的に用いる視点依存テクスチャ[7]を利用することで低減することができる。ただし、視点依存テクスチャを用いて自由視点画像を生成した場合においては、仮想化実物体の輪郭上において、図 2 に示すような不自然な形状の膨張や欠損が生じることが多い。本研究ではこのような領域を、以下のように定義する。

- **膨張領域**

自由視点画像上の仮想化実物体の輪郭周辺に膨らみが生じ、入力画像の背景領域がテクスチャマッピングされている領域 (図 2(b))。

- **欠損領域**

自由視点画像上の仮想化実物体の輪郭周辺にへこみが生じ、入力画像の背景領域がテクスチャマッピングされていない領域 (図 2(c))。

本論文ではこれらの問題に対して、視点依存テクスチャに加えて、仮想視点に

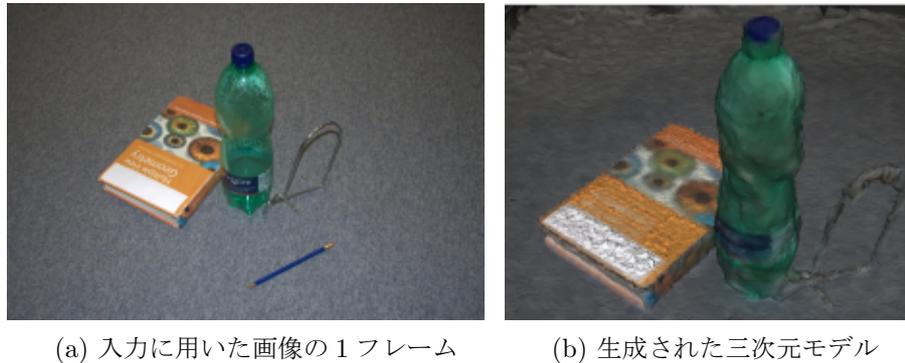


図 1: CMPMVS [3] で生成された三次元モデルの例

依存して三次元モデルを変形する視点依存ジオメトリを併用することで、膨張・欠損領域を修復し、高品位に仮想化実物体の自由視点画像を描画し、実環境の映像に合成する手法を提案する。本研究では、まず仮想視点における仮想化実物体の輪郭を広げることで欠損領域を補完し、次に予め前景抽出した撮影画像群から視点位置に応じたテクスチャを選択して前景領域のみをテクスチャマッピングに利用することで、膨張領域を修復する。

以下、2章では、仮想化実物体を対象とした自由視点画像生成と拡張現実感への応用についての関連研究及び、本研究の位置付けについて述べる。3章では、本論文の提案手法である、仮想化実物体の自由視点画像上における輪郭形状の修復について述べる。4章では、従来手法との比較実験により提案手法の有効性を示す。5章では、提案手法のアプリケーション例を示す。最後に、6章でまとめと今後の展望について述べる。

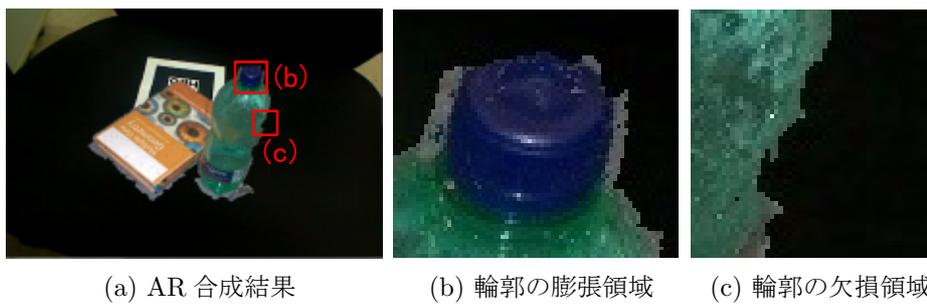


図 2: 視点依存テクスチャを適用した仮想化実物体の自由視点画像の例

## 2. 関連研究と本研究の位置付け

本章では，仮想化実物体の任意の視点位置における見えを表す自由視点画像の生成に関する従来研究を，拡張現実感での利用を念頭に置きつつ概観するとともに，本研究の位置付けを述べる．

### 2.1 仮想化実物体の自由視点画像生成の関連研究

仮想化実物体の自由視点画像生成手法は，(1) 形状・テクスチャが不変の三次元モデルを用いるモデルベースドレンダリング [1-6, 8-10]，(2) 対象物体を明示的に三次元復元することなく画像の変形・合成により自由視点画像を生成するイメージベースドレンダリング [11-21]，(3) モデルベースドレンダリングとイメージベースドレンダリングを組み合わせたハイブリッドレンダリング [7, 22-25] の3つに分類できる．以下では，これら三手法の特徴と，拡張現実感へ応用した場合の問題点について詳述する．

#### 2.1.1 モデルベースドレンダリング

一般的なARアプリケーションにおいて，仮想物体の提示は幾何形状が不変の三次元モデルを用いるモデルベースドレンダリングにより実現されている [26]．モデルベースドレンダリングは，対象物体の三次元形状を用いるため，視点移動に関する自由度や，視点移動に伴う物体の隠蔽関係の表現に優れている．また，仮想化された物体を追加表示することが容易であるという特徴がある．このようなモデルベースドレンダリングに基づくARアプリケーションにおいて，重畳対象として実物体を利用し，高品位な自由視点画像を生成するためには，対象物体の三次元形状を高精度に取得する必要がある．以下では，三次元復元手法を能動的，受動的な方法に大別し，それぞれについて述べる．

能動的な三次元復元手法として，プロジェクタ等を用いたアクティブステレオ法 [8, 9] や，レーザレンジファインダを用いた手法 [10] が提案されている．塚本ら [8] は，プロジェクタ等によってスリット光を対象に照射し，近接する位置に

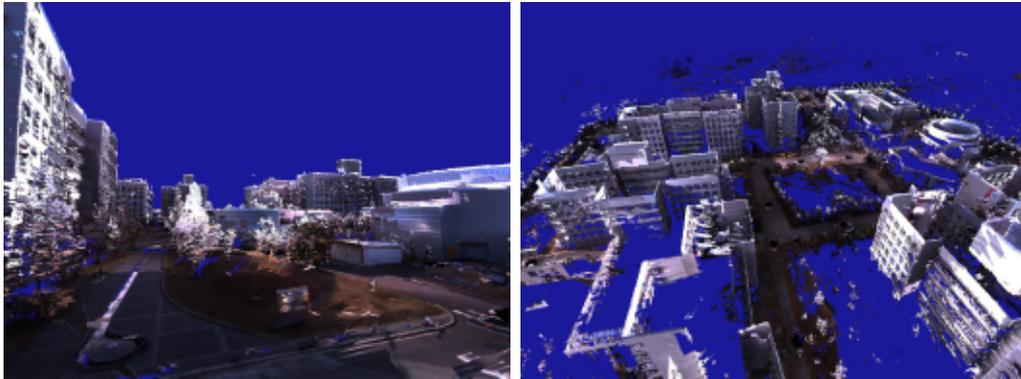


図 3: レーザレンジファインダによる三次元復元 [10] の例

設置したカメラでスリット光の位置を検出し，三角測量の原理によって対象環境の三次元形状を奥行き画像として取得する手法を提案した．この手法は，スリット光を動かしながら計測するので，計測に時間を要する．これに対して Kawasaki ら [9] の手法では，投影方向に関する情報を特殊な模様でコード化した光パターンを用いて，一回の投影で三次元形状を計測できる．time-of-flight 方式のレーザレンジファインダを用いた手法 [10] では，対象物体に光レーザを照射することによって，光が反射して計測されるまでの時間を計測することでレンジデータを取得する（図 3）．これらの手法は，物体の形状を高精度に取得可能である反面，プロジェクタやレーザレンジファインダ等の特殊な機材を用いるため，一般のユーザは手軽に実物体の仮想化を行うことが難しい．

三次元形状の推定をユーザ自身がスマートフォン等で手軽に行うことを想定した場合には，カメラにより取得可能な実環境画像による受動的な三次元復元手法が適している．受動的な三次元復元手法には，ステレオ法や視体積交差法が存在する．これらの手法は，撮影されたカメラの位置・姿勢が既知の複数枚の画像を用いる必要があり，カメラの位置・姿勢の推定には Structure from Motion 法 [27] が広く利用されている．

ステレオ法は，位置・姿勢が既知の二台以上のカメラによって撮影された画像に対し，撮影画像上で各画素の対応を決定することで，三角測量の原理により奥行き画像を決定する．その中でも，多数の画像上で対応点を見つけて奥行きを



図 4: マルチビューステレオ法 [2] により生成された三次元モデルの例

決定するマルチビューステレオ法 [1] は、二眼のステレオ法と比較してより精度良く奥行き値を求めることができる。マルチビューステレオ法の代表例として、Furukawa ら [2] の、多数の画像を入力として Photo-Consistency を高めることで、インターネット上の大量の画像による密な三次元復元を行う手法が挙げられる。ただし、テクスチャが少ない箇所では画像間での対応が付かず、生成モデルに欠損が生じる (図 4)。これに対して、Jancosek ら [3] は、推定された奥行きが疎な点群に対しても面を貼り付けて密な三次元復元を行い、欠損の少ないモデルを生成できる手法を提案した。しかし、この手法においてもテクスチャの少ない箇所では形状の推定精度が必ずしも良くない。

視体積交差法 [4] は、撮影空間をボクセルと呼ばれる立方体の集合として表現し、そのボクセル空間内で多地点でのカメラの光学中心とシルエット画像から成る視体積の積集合空間 (Visual Hull) を計算することで三次元モデルを生成する手法である。Kolev ら [5] は、ベイズ推定に基づくインタラクティブな前景抽出と視体積交差法による三次元復元を、GPU による実装で高速に復元可能な手法を提案した。また、Bastian ら [6] は、AR をインターフェースに用いたインタラクティブな前景抽出による視体積交差法に基づく実物体の三次元復元を行い、仮想化実物体を実環境に重畳する手法を提案した (図 5)。これらの手法では、対象物体を様々な方向から撮影したシルエット画像の量を増やすことによって対象物体の真の三次元形状に近い三次元モデルを生成できる。また、ステレオ法のような対応点探索の問題がないため、ボクセル空間の解像度や使用する画像数にもよ



(a) インタラクティブな前景抽出 (b) 復元した三次元モデルの重畳

図 5: Bastian ら [6] の手法

るが、ステレオ法による三次元復元に対して比較的計算コストが小さい。また、マルチビューステレオ法とは異なり欠損のない三次元モデルを生成できる。しかし、対象物体に凹な部分が存在する場合には、その部分は正しい三次元形状が得られないという制約がある。

以上をまとめると、モデルベースドレンダリングによる自由視点画像生成手法は、推定した三次元モデルに欠損が生じたり形状の推定精度が必ずしも良くないため、ARにそのまま用いることは難しい。また、視点位置に依存しないテクスチャを用いるため、提示する仮想化実物体の写実性が低いという問題もある。

### 2.1.2 イメージベースドレンダリング

モデルベースドレンダリングと比較して、より写実的な表現が可能なイメージベースドレンダリングが提案されている。イメージベースドレンダリングは、対象物体の三次元形状を明示的に復元することなく、物体やシーンを撮影した画像群を入力とし、これらを変形・合成することにより自由視点画像を生成する手法である。この手法では、対象物体の形状や反射特性を直接利用しないため、モデルベースドレンダリングでは正確な幾何形状の推定が難しい複雑な形状を持つ樹木などの物体に対しても、欠損を生じない自由視点画像を生成可能である。以下で

は、イメージベースレンダリングを、画像群を直接変形・合成することで自由視点画像を生成するアプローチと、光線情報を用いるアプローチに大別し、Virtual Reality(VR) 及び AR の分野におけるそれぞれの手法の特徴を述べる。

画像群を直接変形・合成することで自由視点画像を生成するアプローチとして、ワーピングやモーフィングと呼ばれる手法 [11,12] が提案されている。これらの手法は、入力となる複数の画像間で対応点を与え、仮想視点位置において対応点がどの位置にあらわれるかを計算し、入力画像を変形させることで自由視点画像を生成している。VR の分野において、冨手ら [13] は、複数の全方位画像からの自由視点画像をモーフィングによって生成し、ウォークスルー環境を構築した。一方、AR の分野において、稲本ら [14] は、事前に多視点カメラから撮影したサッカーシーンに対して自然特徴点の対応付けによるモーフィングで自由視点画像を生成し、高さ方向を考慮したフィールドと平行な平面に選手の画像をレンダリングすることで、フィールドモデルを写した映像に実際の選手を重畳する手法を提案した。この手法は、左右及び前後方向の視点移動のみを想定してユーザの視点の位置決定や選手の位置合わせを行っているため、ユーザ視点のカメラにおいてロールが発生した場合、視点位置の算出時に誤差が生じ、選手の見え方を正しく再現できない可能性がある。これらの手法は、手動で対応点を与えるだけで簡単に自由視点画像を生成できるという特徴があるが、三次元形状に関する情報を用いないため、物体の隠蔽関係を正しく再現することが困難である。また、三次元情報を用いない入力画像の変形により自由視点画像を生成するため、視点移動に伴って生成画像上で幾何学的な歪みが発生しやすい。

光線情報を用いるアプローチとして、Plenoptic Function [15]、光線空間法 [16]、Light Field Rendering [17]、Lumigraph [18] 等の手法が提案されている。これらの手法は、空間中を伝搬する光線を記録し、自由視点位置において視界画像面を通過する光線をリサンプリングすることで自由視点画像の生成を行う。Adelsonら [15] は Plenoptic Function を提案し、視点の座標  $(x, y, z)$ 、光線の向き  $(\theta_x, \theta_y)$ 、波長  $\lambda$  及び時間  $t$  の七次元のデータを用いることで空間中の全ての光線を記述する手法を提案した。ただし、七次元のデータを全て記録するためには膨大な量の画像が必要となり、現実的に全てのデータを収集することは難しい。これに対し

てNaemuraら [16] は、より低い次元の情報で空間を記述することで任意視点の画像を効率的に再構築する光線空間法と呼ばれる手法を提案した。光線空間法は、空間を伝搬する光線は減衰・変移しないと仮定することで、Plenoptic Functionの七次元の光線情報から $z$ 座標、波長 $\lambda$ 及び時間 $t$ を減らし、情報量を四次元とすることで取得すべき画像の量を大幅に減らした。Levoyら [17] 及びGortlerら [18] は、同時期にLight Field Rendering, Lumigraph という手法を提案した。これらの手法では、光線空間法が平面 $(x, y)$ 及び光線の向き $(\theta_x, \theta_y)$ で四次元空間を表現したのに対して、この四次元を2つの平行した平面を横切る座標 $(u, v)$ 及び $(s, t)$ で表現することで、効率的なレンダリングを可能にした。これらの手法を応用し、特定の物体を対象にして自由視点画像を生成するImage-Based Visual Hull [19] が提案されている。この手法は仮想視点における画像面の各画素に対して、仮想カメラの光学中心（自由視点）とその画素を結ぶ直線が対象物体のVisual Hullを貫通するか否かを判定することにより自由視点画像を生成しており、視体積交差法 [4] におけるデータ量と計算量が増大する問題を解決している。

これらの光線情報を用いるアプローチをARへ応用した例として、南谷ら [20] の変型ビルボードを幾何形状に用いたイメージベースドレンダリングによる、遠隔地の人物との協調作業システムが挙げられる（図6）。この手法では、相手の映像をビルボード上にマッピングすることで、写実的な映像を提示できる。しかし、この手法は机を挟んで対面して座ることを前提としており、ユーザの視点位置に制約がある。これらの光線情報を用いる手法は、高品位な自由視点画像を生成するためには大量の画像が必要であり、そうでない場合には視点移動に伴って自由視点画像上で幾何学的な歪みが発生しやすいといった問題が存在する。

以上をまとめると、イメージベースドレンダリングでは、対象物体の三次元形状を直接用いず画像の変形・合成を行う。このため、視点移動が大きいARアプリケーションにおいて、広範囲で高品位な自由視点画像を生成するには大量の画像が必要であり、そうでない場合には自由視点画像上の仮想化実物体に幾何学的な歪みが生じるため、視点の移動に制約が生じる。



図 6: 変型ビルボードによる協調作業システム [20]

### 2.1.3 ハイブリッドレンダリング

モデルベースとイメージベースのハイブリッドな方法では、視点移動が広範囲な場合でも比較的高品位な仮想化実物体の自由視点画像を生成することができる。以下では、ハイブリッドレンダリングを視点に依存するテクスチャと形状が変化しない三次元モデルを用いる手法と、視点に依存するテクスチャとジオメトリを併用する手法に大別し、それぞれの特徴を述べる。

視点に依存するテクスチャと形状が変化しない三次元モデルを用いる手法として、VR の分野において、Debevec ら [7] は、事前に生成した簡易形状の三次元モデルに対して、実環境の画像から仮想視点に近い見え方の画像を選択してテクスチャマッピングを行う視点依存テクスチャ法を提案した (図 7)。これにより、三次元モデルの詳細な形状をテクスチャとして表現できるが、形状とテクスチャの輪郭を厳密に位置合わせする必要がある。

視点依存テクスチャ・ジオメトリの併用に基づく手法として、山崎ら [22] は、物体の幾何形状を視線方向に依存して方向が変化する微小面の集合で近似し、実画像をテクスチャとして貼り付けることにより、複雑な形状の物体を写実的かつ効率的に表示するための、微小面ビルボーディング法を提案した。また、高橋ら [23] は、固定された二次元カメラアレイを用いて取得した多眼画像を元に、対象環境の奥行き値を複数枚の平面で仮定した奥行きレイヤモデルを用いて推定すること

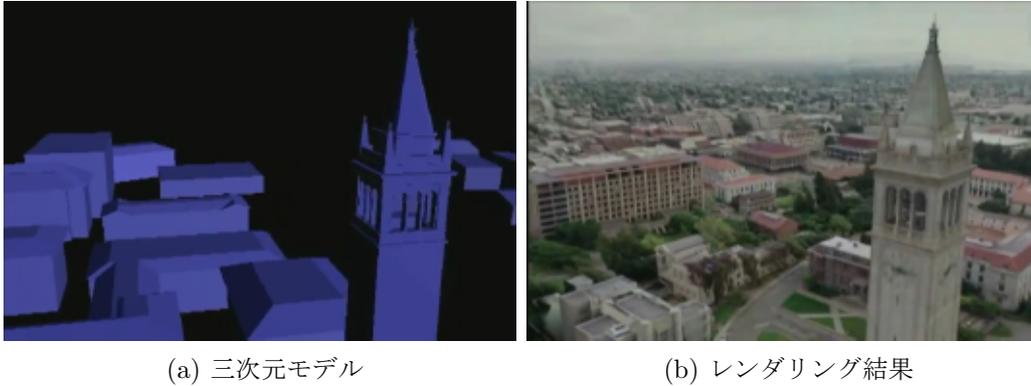


図 7: 視点依存テクスチャ法 [7]

で、自由視点画像の実時間生成が可能な手法を提案した。この手法は、オフライン計算を伴わず、動的なシーンを扱うことができる。伊吹ら [24] は、オフライン処理で全方位の映像と奥行き値を算出・蓄積しておき、オンライン処理で、三次元モデルの時間的・空間的な連続性とテクスチャの幾何学的・光学的な連続性を考慮した三次元モデルの生成とテクスチャの選択を行うことで、自由視点画像上に生じるテクスチャの不連続な変形や位置ずれに起因する歪みを低減した。これらの手法は視点移動に伴う画像上の歪みが少なく、高品位な自由視点画像が生成できる一方で、実環境中におけるシーン全体の自由視点画像を生成するため、家具などの単一物体のみを仮想化して重畳する AR にそのまま用いることは難しい。

ハイブリッドレンダリングを AR へ応用した研究として Prince ら [25] は、複数のビデオカメラから人物を撮影し、ユーザの視点位置の近傍のカメラによる Visual Hull と Photo-Consistency を計算した上で、任意視点での拡張現実画像の生成を実現している (図 8)。ただし、視体積交差法に基づいた三次元モデルを利用するため、対象形状が限定される。

以上で述べたように、VR 分野におけるハイブリッドレンダリングでは比較的高品位な自由視点画像を生成できるが、シーン全体の自由視点画像を生成するため、特定の物体のみの自由視点画像を生成する必要がある AR にそのまま適用することは難しい。AR を対象として、特定物体を抽出して自由視点画像を生成す

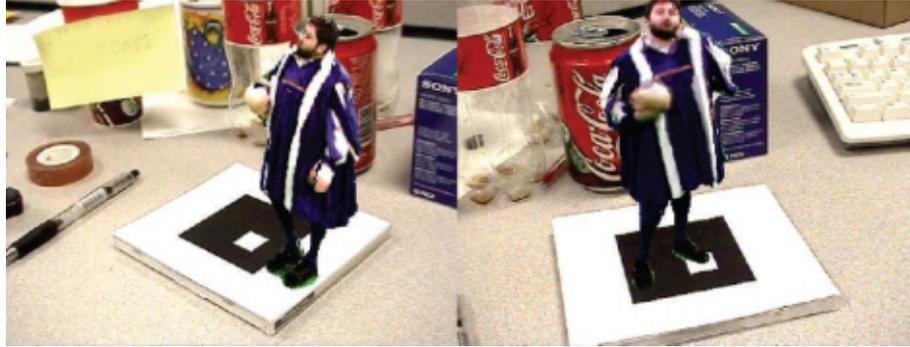


図 8: Prince ら [25] の手法

る視体積交差法による手法も提案されているが、この手法では対象形状が限定され、また、三次元モデルの生成時に用いた画像の撮影枚数が少ない場合には、撮影地点から離れた視点における自由視点画像上において、三次元モデルの輪郭が膨張してしまうという問題が残されている。

## 2.2 本研究の位置付けと方針

本研究は、仮想化実物体の高品位な自由視点画像を生成し、実環境映像へ重畳することを目的としている。既に述べた通り、仮想化実物体の自由視点画像を生成し実環境の映像に重畳する手法にはモデルベースドレンダリング、イメージベースドレンダリング、ハイブリッドレンダリングの三手法が存在する。モデルベースドレンダリングは視点に依存しないテクスチャを用いるため、提示する仮想化実物体の写実性が低い。また、能動的な三次元復元手法は、特殊な機材を用いるため、一般のユーザが手軽に実物体の仮想化を行うことが難しく、受動的な三次元復元手法は、形状の推定精度が必ずしも良くないため、推定結果をモデルベースドレンダリングにそのまま用いるのは難しい。イメージベースドレンダリングは対象物体の形状や反射特性を直接利用しないため、三次元復元手法では正確な幾何形状の推定が難しい複雑な形状の物体に対しても写実性の高い自由視点画像を生成可能であるが、入力画像の変形・合成によって新しい視点の画像を生成するため、モデルベースドレンダリングと異なり、視点移動に伴う仮想化実物体の自由視点画像上に幾何学的な歪みが生じ、写実性を保ったまま自由に広範囲を移

動するのは難しい。モデルベースレンダリングとイメージベースレンダリングを組み合わせたハイブリッドレンダリングは、比較的広範囲にわたり写実的な自由視点画像の生成が可能である。しかし、VRの分野におけるハイブリッドレンダリングの手法は、実環境中におけるシーン全体の自由視点画像を生成するため、本研究の目的である、家具などの単一物体のみを仮想化して重畳するARにそのまま用いることは難しい。また、ハイブリッドレンダリングに基づくARの分野においても、従来研究では視体積交差法に基づく手法で三次元モデルを生成しているため、対象形状が限定され、また、三次元モデルの生成時に用いた画像の撮影枚数が少ない場合には、撮影地点から離れた視点における自由視点画像上において三次元モデルの輪郭が膨張してしまう。

以上より、視点移動が大きいARアプリケーションにおける仮想化実物体の自由視点画像生成においては、比較的テクスチャの歪みが小さく、写実的な自由視点画像が生成可能なハイブリッドレンダリングによる手法が最も適していると考えられる。しかし、仮想化実物体を正しく実環境映像に重畳するには、重畳対象のみを正確に抽出した自由視点画像を生成する必要があり、従来研究ではARのための仮想化実物体の正確な抽出が実現されていない。そこで本研究では、重畳対象を単一、もしくは複数の物体とし、高品位な仮想化実物体の提示を目指して視点依存テクスチャ・ジオメトリに基づいたハイブリッドな手法を用いるとともに、AR特有の問題となる実環境映像と仮想化実物体の境界における膨張と欠損を修復する。具体的には、仮想視点における仮想化実物体の輪郭を拡張することで欠損領域を補完し、さらに前景抽出した撮影画像群から視点位置に応じたテクスチャを選択して前景領域のみをテクスチャマッピングに利用することで膨張領域を修復し、高品位な仮想化実物体の自由視点画像を実環境映像に合成する。

### 3. 視点依存テクスチャ・ジオメトリによる膨張領域と欠損領域の修復

本章では、仮想化の対象となる実物体（対象物体）を撮影した画像に基づき、仮想視点における仮想化実物体の膨張領域と欠損領域を実時間で修復する自由視点画像生成手法の詳細について述べる。以下ではまず、提案手法の概要を説明し、提案手法で用いる座標系を定義する。続いて、三次元モデルの生成等の前処理で構成されるオフライン処理と、実時間で自由視点画像を生成するオンライン処理について、詳述する。

#### 3.1 提案手法の概要

提案手法は、対象物体を撮影した複数の画像（撮影画像群）をもとに、既存手法により得られた三次元モデルから仮想化実物体の高品位な自由視点画像を生成し、実環境画像へ重畳することを目的とする。前述のように、既存手法により得られた三次元モデルには誤差が含まれるため、生成された自由視点画像には膨張領域や欠損領域が生じる（図2）。提案手法ではこれらの問題に対して、視点依存ジオメトリのアプローチを利用して、誤差を含む三次元モデルの輪郭を拡張することで欠損領域を補完し、さらに視点依存テクスチャのアプローチを利用して、前景抽出した撮影画像群から視点位置に応じたテクスチャを選択して前景領域のみをテクスチャとして適用することで膨張領域を修復する。

提案手法は図9に示すように、オフライン処理(A)とオンライン処理(B)から構成される。オフライン処理(A)では、対象物体を撮影した撮影画像群 $S'$ を入力として、撮影画像群 $S'$ に含まれる撮影画像 $I_k$ を撮影したカメラの内部・外部パラメータを推定、歪みを補正し、提案手法に適した枚数に間引いた画像からなる画像群 $S$ を生成する(A-1)とともに、既存手法を用いて対象物体の三次元モデル $M$ を構築する(A-2)。また、撮影画像 $I_k \in S$ に含まれる対象物体領域を前景として抽出する(A-3)。オンライン処理(B)では、ユーザが取得した実環境画像に対して、カメラの位置・姿勢を推定する(B-1)。続いて、三次元モデル $M$ をもとに、視点依存ジオメトリのアプローチに基づいて実環境画像の視点に依存した奥

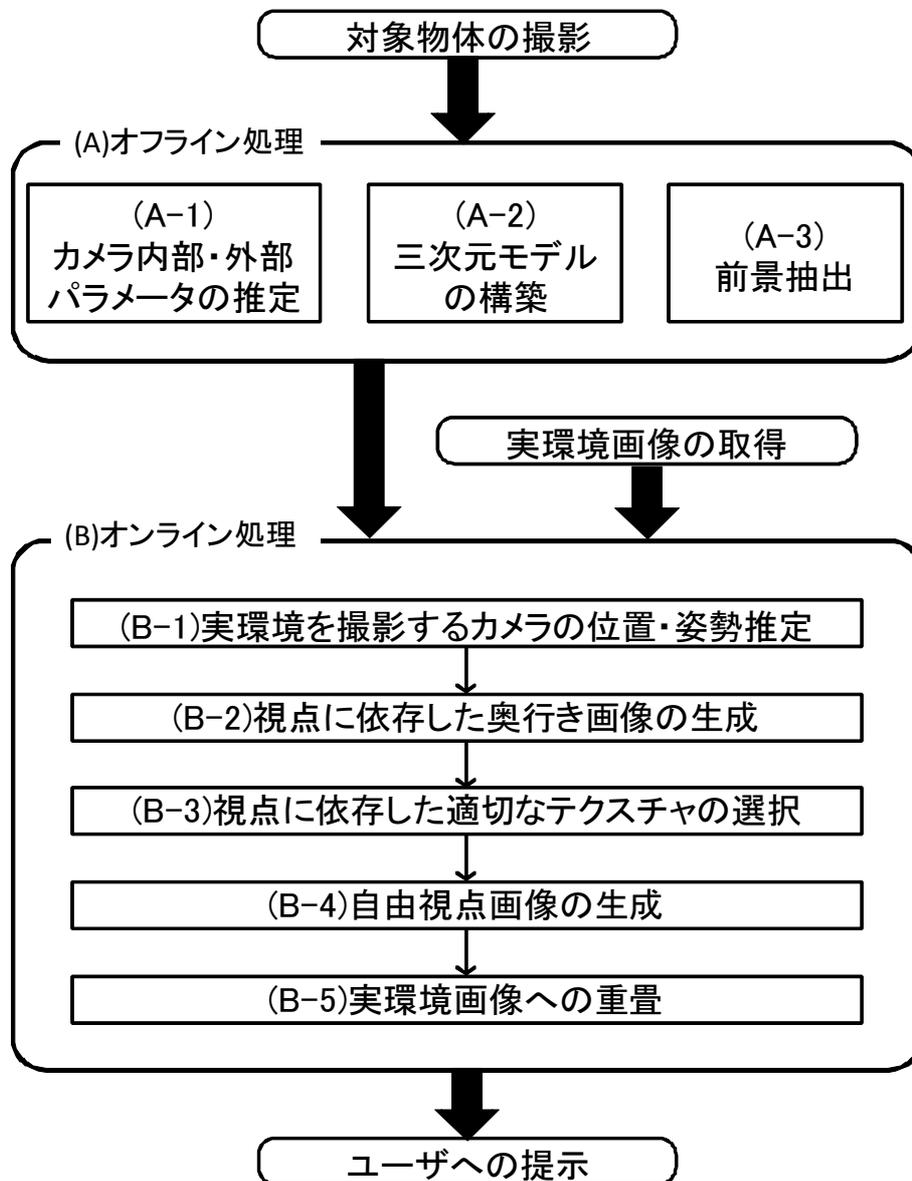


図 9: 提案手法の概要

行き画像を生成する (B-2). さらに, この奥行き画像に基づいて, 自由視点画像の色情報を取得する画像を視点依存テクスチャのアプローチを利用して撮影画像群  $S$  から選択し (B-3), 選択した画像から色情報を各画素に付与することで, 自由視点画像を生成する (B-4). 最後に, 生成した自由視点画像を実環境画像に重畳し, ユーザに提示する (B-5). 提案手法では, 実環境画像を取得するたびに上記の処理を繰り返す.

### 3.2 座標系の定義

提案手法では, 図 10 に示すように, オフライン処理で用いる撮影画像群  $S$  に含まれる撮影画像  $I_k$  を撮影するカメラ  $C_k$  と, 仮想化実物体を重畳するための実環境画像  $I_E$  を撮影するカメラ  $C_E$  が存在する. これらのカメラや仮想化実物体の位置・姿勢を記述するために, 1) アプリケーションの利用者であるユーザが任意に決定する世界座標系, 2) 対象物体を撮影するカメラ  $C_k$  の位置・姿勢と三次元モデル  $M$  の記述に用いられるオブジェクト座標系, 3) 対象物体を撮影したカメラ  $C_k$  のカメラ座標系, 4) 実環境を撮影するカメラ  $C_E$  のカメラ座標系の 4 種類の座標系を用いる. 本研究では,  $C_k$  のカメラ座標系と  $C_E$  のカメラ座標系を区別するために, 実環境を撮影するカメラ  $C_E$  が自由視点画像生成における仮想空間中の仮想視点に対応することを考慮し,  $C_E$  のカメラ座標系を仮想視点座標系と呼ぶ.

次に, これら 4 種類の座標系を変換する変換行列のうち, 提案手法で利用するものについて説明する. まず, オブジェクト座標系から  $C_k$  のカメラ座標系への変換は, 対象物体撮影時の対象物体とカメラ  $C_k$  の位置・姿勢によって決定される. この変換行列は, 手順 (A-1) で得られたカメラ  $C_k$  の外部パラメータ  $[\mathbf{R}_k \mathbf{t}_k]$  により与えられる. 同様に, 世界座標系から仮想視点座標系への変換行列は, 手順 (B-1) で推定されるカメラ  $C_E$  の外部パラメータ  $[\mathbf{R}_E \mathbf{t}_E]$  である. オブジェクト座標系から世界座標系への変換行列は, 世界座標系のどこに仮想化実物体を配置するかを決定する. この変換行列  $\mathbf{O} \in \mathbb{R}^{4 \times 4}$  は, 三次元モデルの作成者が任意に設定可能である.

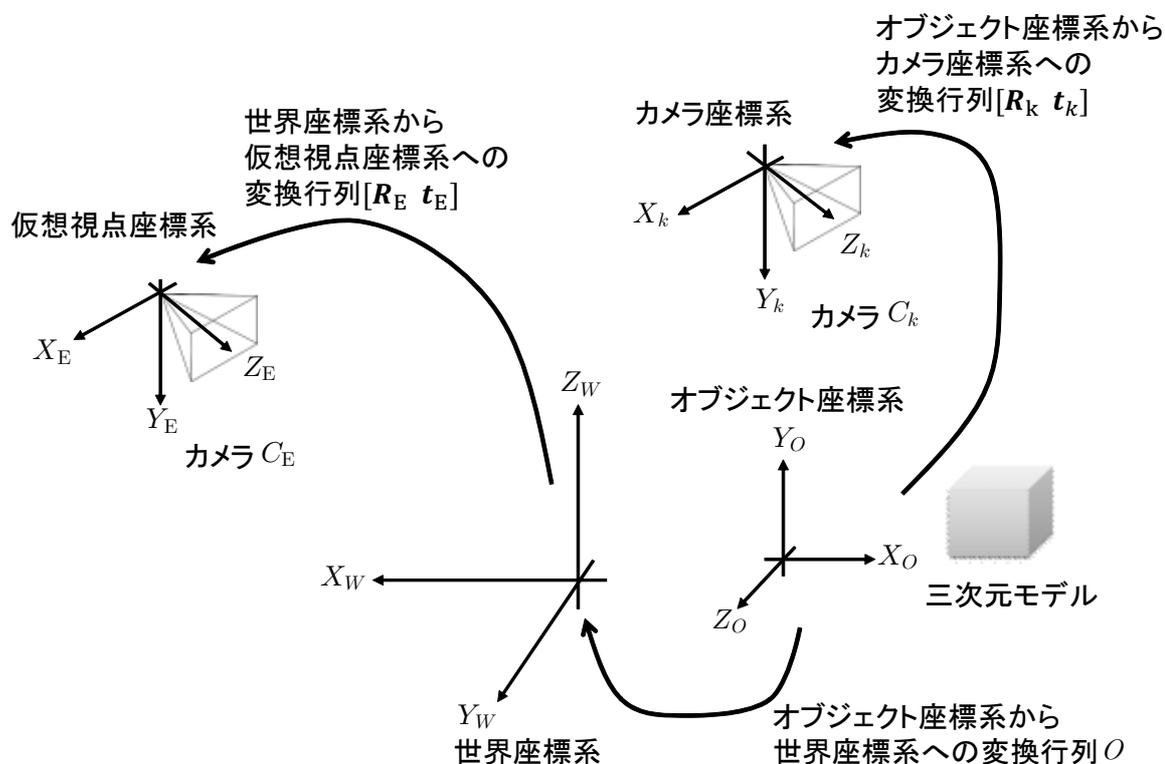


図 10: 世界座標系, オブジェクト座標系, カメラ座標系, 仮想視点座標系の関係

### 3.3 オフライン処理

提案手法におけるオフライン処理では, 撮影画像のカメラ内部・外部パラメータの推定 (A-1), 三次元モデルの構築 (A-2), 及び, 撮影画像の前景領域の抽出 (A-3) を行う. 本節では, これらの処理について述べる.

#### 3.3.1 撮影画像のカメラ内部・外部パラメータの推定

提案手法では, マルチビューステレオ法 [1-3] 等の三次元復元に基づく三次元モデルの構築 (A-2), 及び視点依存テクスチャの選択 (B-3) などに際して, 撮影画像群  $S'$  に含まれる画像  $I'_k$  ( $k = 1, 2, \dots, |S'|$ , ただし  $|S'|$  は  $S'$  に含まれる画像数) を撮影したカメラ  $C_k$  の内部・外部パラメータを利用する. 本研究では, カ

カメラ内部パラメータを Zhang の手法 [28] により推定し、カメラ外部パラメータを Structure from Motion 法 [29] により推定する。推定されたカメラ内部パラメータを  $\mathbf{A}_k \in \mathbb{R}^{4 \times 4}$  とし、カメラ内部パラメータ  $\mathbf{A}$  を次式で定義する。

$$\mathbf{A} = \begin{pmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

ただし、 $f$ 、 $c_x$ 、及び  $c_y$  はそれぞれ焦点距離、画像の横方向、縦方向に対する投影中心座標である。また、Zhang の手法 [28] により内部パラメータと同時に推定される歪みに関するパラメータを利用して画像  $I'_k$  のレンズ歪みを補正し、補正後の画像を提案手法に適した枚数に間引いた  $I_k$  からなる撮影画像群  $S = \{I_k | k = 1, 2, \dots, |S|\}$  を生成する。3.2 節で述べたように、カメラ外部パラメータは  $[\mathbf{R}_k \ \mathbf{t}_k]$  として与えられる (図 10)。

### 3.3.2 三次元モデルの構築

画像データセット  $S$  と推定されたカメラの内部・外部パラメータから、視点依存ジオメトリなどの処理の元となる対象物体の三次元モデル  $M$  を構築する。これには、CMVS [2] や CMPMVS [3] 等のマルチビューステレオ法を利用する。生成された三次元モデルを図 1(b) に例示した。

### 3.3.3 前景領域の抽出

マルチビューステレオ法 [1-3] により生成された三次元モデル  $M$  には、形状の推定誤差により欠損が生じることから、提案手法では手順 (B-3) において、三次元モデル  $M$  から奥行き画像  $D_E$  を生成し、輪郭を拡張させることで欠損を補完する。一方で、これにより図 2(b) に示したように、撮影画像中の対象物体ではない領域がテクスチャとして適用される可能性がある。そこで、対象物体の領域のみをテクスチャとして利用するために、撮影画像  $I_k$  に対して対象物体の領域を前景領域、それ以外の領域を背景領域として前景抽出を行い、画像中の全画素に対



図 11: 前景抽出した撮影画像の例

して前景領域・背景領域を表すラベルを付与する．手順 (B-4), (B-5) では，前景領域の画素のみをテクスチャとして用いることで自由視点画像を生成することにより，膨張領域を修復する．前景抽出には GrabCut [30] 等の手法が利用可能である．図 11 に前景抽出された撮影画像を例示する．

### 3.4 オンライン処理

提案手法におけるオンライン処理では，オフライン処理で得られた三次元モデル  $M$  と各カメラの内部・外部パラメータ，前景抽出した撮影画像群  $S$  を入力として膨張領域と欠損領域を修復した自由視点画像を生成する．本節では，実環境を撮影するカメラ  $C_E$  の位置・姿勢推定 (B-1)，視点に依存した奥行き画像  $D'_E$  の生成 (B-2)，視点に依存した適切なテクスチャの選択 (B-3)，自由視点画像の生成 (B-4) と実環境画像への重畳 (B-5) について説明する．

#### 3.4.1 実環境を撮影するカメラの位置・姿勢推定

本研究では，AR での利用を想定し，実環境を撮影したカメラ  $C_E$  から得られる実環境画像  $I_E$  に対して，仮想化実物体の自由視点画像を生成し，重畳する．これには，カメラ  $C_E$  の外部パラメータの実時間推定が必要となる．本研究では，ARToolKit [31] 等を用いて外部パラメータを推定する．カメラ  $C_E$  の内部パラメータは Zhang の手法 [28] 等により事前に算出する．また，3.2 節で述べたように，

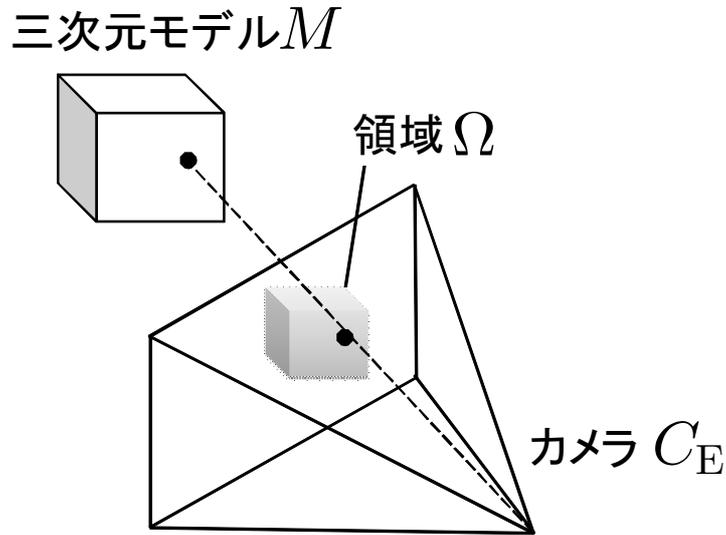


図 12: カメラ  $C_E$  における画素  $n$  に対応する光線と三次元モデル  $M$

カメラ  $C_E$  の外部パラメータの基準となる世界座標系については，事前にアプリケーションの利用者であるユーザが任意に決定するものとする．ここで推定されるカメラ  $C_E$  の内部パラメータを  $\mathbf{A}_E \in \mathbb{R}^{4 \times 4}$ ，外部パラメータを  $[\mathbf{R}_E \mathbf{t}_E]$  とする (図 10) .

### 3.4.2 視点に依存した奥行き画像の生成

マルチビューステレオ法 [1–3] 等により得られる三次元モデル  $M$  には形状推定の誤差が含まれるため，カメラ  $C_E$  における自由視点画像の輪郭に膨張や欠損が生じる．ここでは欠損領域の補完を目的として，視点依存ジオメトリのアプローチに基づいて三次元モデル  $M$  のカメラ  $C_E$  上における輪郭に対応する部分を拡張することにより奥行き画像  $D'_E$  を構築する．

提案手法ではまず，カメラ  $C_E$  により撮影された実環境画像  $I_E$  に対応する三次元モデル  $M$  の奥行き画像  $D_E$  を生成する．三次元モデル  $M$  に対する奥行き画像  $D_E$  は，図 12 に示すように，画像  $I_E$  上において三次元モデル  $M$  が投影される画素の集合  $\Omega$  上で定義される．本稿では画素  $n \in \Omega$  における奥行き値を  $d_{E,n}$  と表記する．次に，奥行き画像  $D_E$  に対して拡張処理を反復して適用することで， $D'_E$

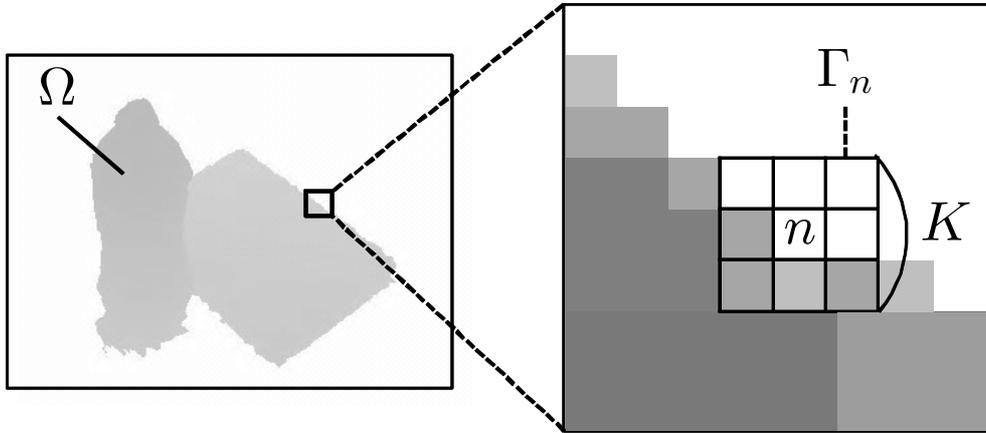


図 13: 奥行き画像  $D_E$  に対する拡張処理

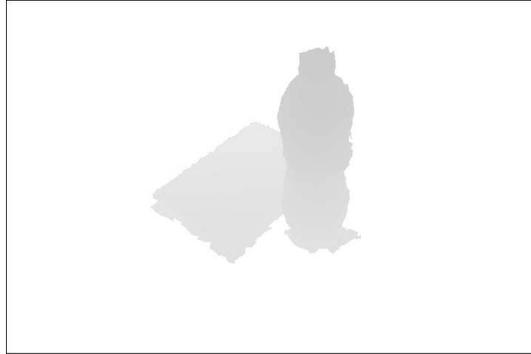
の奥行き値  $d'_{E,n}$  を求める. 提案手法では, 図 13 に示すように画素  $n$  を中心とする  $K \times K$  の正方領域に含まれる画素の集合を  $\Gamma_n$  とし,  $\Gamma_n$  と  $\Omega$  が共通領域を持つ場合に画素  $n$  に対して奥行き値を与えることで  $D_E$  を拡張する. 具体的には,  $\Gamma_n$  と  $\Omega$  が共通領域を持つ画素  $n$  の集合を  $\Omega'$  として, 拡張された奥行き画像  $D'_E$  の奥行き値  $d'_{E,n}$  は次式により算出される.

$$d'_{E,n} = \begin{cases} \frac{1}{|\Gamma_n \cap \Omega|} \sum_{n' \in \Gamma_n \cap \Omega} d_{E,n'} & (n \in \Omega' \setminus \Omega) \\ d_{E,n} & (\text{otherwise}) \end{cases} \quad (1)$$

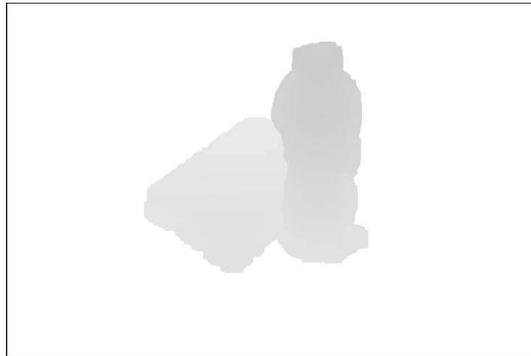
ただし,  $A \setminus B$  は集合  $A$  から集合  $B$  を引いた差集合を表す.  $n \in \Omega' \setminus \Omega$  に対して, 式 (1) は奥行き値  $d'_{E,n}$  の平滑化処理である. この拡張処理では, 正方領域の大きさを決定するパラメータ  $K$  と, 拡張処理の反復回数に依存して仮想化実物体の輪郭が拡張される. 図 14 に  $720 \times 480$  画素の奥行き画像に対して,  $K = 3$  での拡張処理を 10 回行った結果と 30 回行った結果を示す. これらの結果より, 反復回数に応じて, 輪郭が拡張されていることが分かる.

### 3.4.3 視点に依存した適切なテクスチャの選択

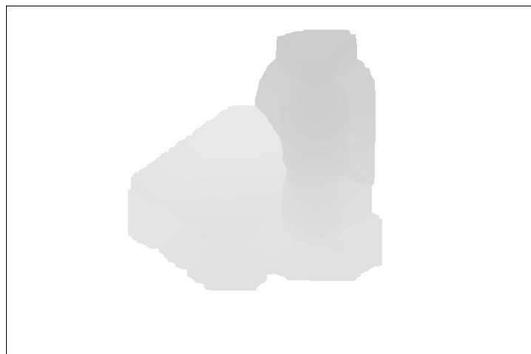
手順 (A-2) で生成された三次元モデル  $M$  は, 形状の推定誤差により表面の詳細な形状が失われ, 対象物体とは異なる形状が生じる. この問題に対して, 視点



(a) 拡張処理前の奥行き画像  $D_E$



(b) 拡張処理後の奥行き画像  $D'_E$ (反復回数 10 回)



(c) 拡張処理後の奥行き画像  $D'_E$ (反復回数 30 回)

図 14: カメラ  $C_E$  に依存した三次元モデル  $M$  の奥行き画像

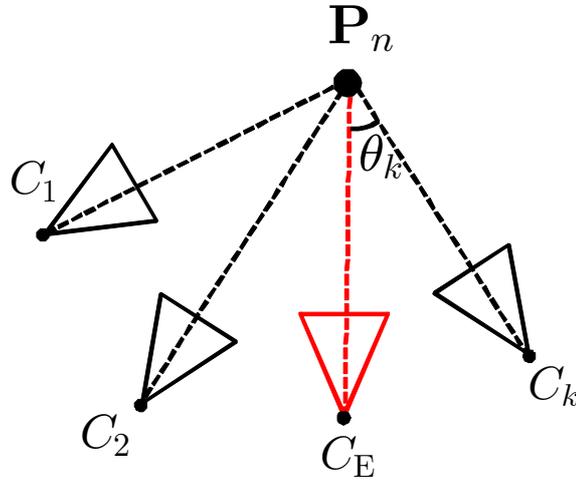


図 15: 点  $P_n$  に対する画像  $I_k$  を撮影したカメラ  $C_k$  の選択基準

依存テクスチャ[7]では，仮想視点からの見えに近い画像をテクスチャとして利用することで，推定誤差の影響を視覚的に低減している．提案手法では，奥行き画像  $D'_E$  上の画素  $n$  に対応する世界座標上の点  $P_n$  に付与する色情報の取得のために，実環境画像を撮影したカメラ  $C_E$  に対して適切な画像を画像群  $S$  から選択する．ここでは，文献 [7] にならい，下記の条件を満たすものを適切な画像として用いる．

- カメラ  $C_E$  と点  $P_n$  を結ぶ直線，及びカメラ  $C_k$  と点  $P_n$  を結ぶ直線のなす角が最小の画像（図 15）．見えの近い画像は，対象物体を近い方向から撮影したものである．
- 画像  $I_k$  に対応するカメラ  $C_k$  から点  $P_n$  が可視である．カメラ  $C_E$  と  $C_k$  は一般に異なる位置となることから，図 16 に示すように自己遮蔽などにより  $C_k$  から点  $P_n$  が不可視となる場合がある．点  $P_n$  における適切な画像は， $C_{k'}$  のように点  $P_n$  が可視な画像である．

適切なテクスチャを選択するために，まず，奥行き画像  $D'_E$  上の画素  $n \in \Omega'$  に対応する点  $P_n$  の世界座標系での位置を算出する．画素  $n \in \Omega'$  に対応する点  $P_n$

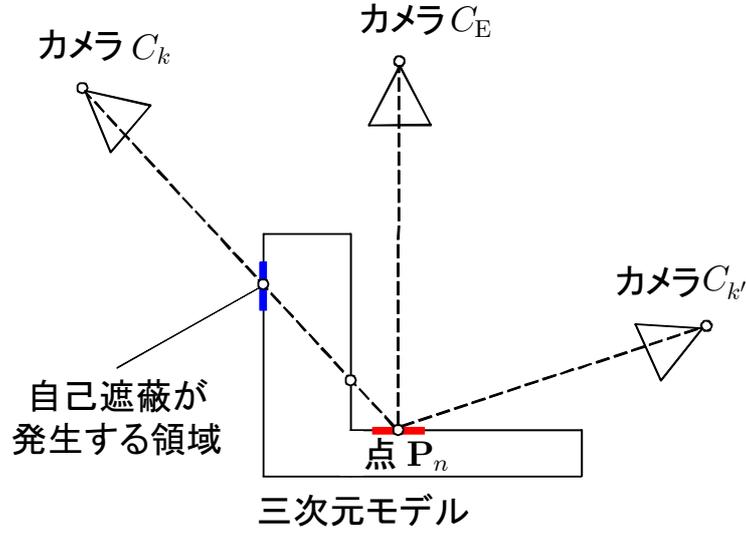


図 16: 自己遮蔽によって誤ったテクスチャが選択される状況

の位置の計算では、画素  $n$  の座標  $(u_{E,n}, v_{E,n})$  とこの座標に対応する奥行き値  $d'_{E,n}$  より得られる同次座標ベクトル  $\mathbf{p}_{E,n} = (u_{E,n}d'_{E,n}, v_{E,n}d'_{E,n}, d'_{E,n}, 1)^\top$  と、 $C_E$  の内部パラメータ  $\mathbf{A}_E$ 、および外部パラメータ  $[\mathbf{R}_E \mathbf{t}_E]$  を用いて、次式により世界座標系における点の位置  $\mathbf{P}_n = (x_n, y_n, z_n)^\top$  を算出する。

$$\mathbf{P}'_n = \begin{bmatrix} \mathbf{R}_E^\top & -\mathbf{R}_E^\top \mathbf{t}_E \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{A}_E^{-1} \mathbf{p}_{E,n} \quad (2)$$

なお、 $\top$  は転置を表し、 $\mathbf{P}'_n = (\mathbf{P}_n^\top 1)^\top$  とする。

続いて、世界座標系での位置を基準に、式 (2) で算出した点  $\mathbf{P}_n$ 、カメラ  $C_E$ 、 $C_k$  の位置  $-\mathbf{R}_E^\top \mathbf{t}_E$ 、 $-\mathbf{O}\mathbf{R}_k^\top \mathbf{t}_k$  を利用して、2つのベクトルのなす角  $\theta_{k,n}$  に対応する単純類似度を次式により算出する。

$$\cos \theta_{k,n} = \frac{(\mathbf{P}_n + \mathbf{R}_E^\top \mathbf{t}_E) \cdot (\mathbf{P}_n + \mathbf{O}\mathbf{R}_k^\top \mathbf{t}_k)}{|\mathbf{P}_n + \mathbf{R}_E^\top \mathbf{t}_E| |\mathbf{P}_n + \mathbf{O}\mathbf{R}_k^\top \mathbf{t}_k|} \quad (3)$$

なお、演算子  $\cdot$  は内積を意味する。

可視判定では、点  $\mathbf{P}_n$  がカメラ  $C_k$  から可視であることを判定する。図 16 のように、点  $\mathbf{P}_n$  がカメラ  $C_k$  から可視の場合、点  $\mathbf{P}_n$  よりカメラ側に遮蔽物が存在しな

いため、 $C_k$  のカメラ座標系における点  $\mathbf{P}_n$  の  $Z$  座標の値  $w_{k,n}$  は、カメラ  $C_k$  の位置・姿勢に対して生成した奥行き画像  $D_k$  の  $\mathbf{P}_n$  が投影される座標  $(u_{k,n}, v_{k,n})$  における値  $d_{k,n}$  と一致する。一方で、点  $\mathbf{P}_n$  が不可視になる場合、 $d_{k,n}$  は  $w_{k,n}$  より小さくなる。以上を考慮し、まずカメラ  $C_k$  における奥行き画像  $D_k$  を三次元モデル  $M$  から生成する。

続いて、次式により点  $\mathbf{P}_n$  を  $I_k$  に投影し、座標  $(u_{k,n}, v_{k,n})$  と  $C_k$  のカメラ座標系における  $z$  座標の値  $w_{k,n}$  を求める。

$$\begin{cases} u_{k,n} = u'_{k,n}/w_{k,n} \\ v_{k,n} = v'_{k,n}/w_{k,n} \end{cases} \quad (4)$$

$$\mathbf{u}'_{k,n} = \mathbf{A}_k \begin{bmatrix} \mathbf{R}_k & \mathbf{t}_k \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{O}^{-1} \mathbf{P}'_n \quad (5)$$

ただし、 $\mathbf{u}'_{k,n} = (u'_{k,n}, v'_{k,n}, w_{k,n}, 1)^\top$  とする。奥行き画像  $D_k$  の座標  $(u_{k,n}, v_{k,n})$  における奥行き値  $d_{k,n}$  と  $w_{k,n}$  について、

$$d_{k,n} = w_{k,n} \quad (6)$$

かつ、 $\mathbf{P}_n$  が  $C_k$  の視野錐体に含まれるとき、点  $\mathbf{P}_n$  は  $C_k$  から可視であると判定する。

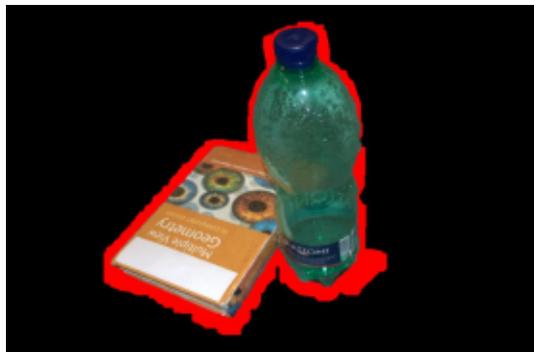
最後に、画素  $n \in \Omega'$  に対応する  $\mathbf{P}_n$  に対して、可視と判定されたカメラのインデックスの集合を  $V_n$  とし、次式により得られる  $\hat{k}_n$  を適切な画像とする。

$$\hat{k}_n = \arg \max_{k \in V_n} \cos \theta_{k,n} \quad (7)$$

#### 3.4.4 自由視点画像の生成と実環境画像への重畳

カメラ  $C_E$  における仮想化実物体の自由視点画像  $F_E$  は、画素  $n \in \Omega'$  に対して手順 (B-3) で選択された画像  $I_{\hat{k}_n}$  の対応する座標  $(u_{\hat{k}_n,n}, v_{\hat{k}_n,n})$  の色を付与することにより生成する。このとき、 $I_{\hat{k}_n}$  の前景・背景を表すラベルをあわせて  $F_E$  の各画素に付与する。これにより、三次元モデル  $M$  の誤差や手順 (B-2) での拡張処理に起因する  $F_E$  の膨張領域に背景ラベルが付与される。ここでは、 $F_E$  を実環境画像

$I_E$ に重畳する際に，前景ラベルが付与された画素のみを対象とすることで，膨張領域の問題を低減する．生成された自由視点画像，実環境画像，拡張現実画像の例を図17に示す．



(a) 自由視点画像（赤色の画素は背景ラベルが付与された画素）



(b) 実環境画像



(c) 拡張現実画像

図 17: 生成された自由視点画像，実環境画像，拡張現実画像の例

## 4. 実験

### 4.1 実験の概要

提案手法の有効性を示すために、表1に示すPC上にARシステムを実装し、後述する3つの画像データセットを入力として、提案手法に加え、下記の2つの手法による自由視点画像の実環境映像への重畳結果を比較した。

- **従来手法 1:**

文献 [3] の手法により生成した誤差を含む三次元モデルを用いたモデルベースの手法であり、三角形パッチの集合で構成される三次元モデルによるモデルベースドレンダリングで自由視点画像を生成する。仮想化実物体の色は、文献 [3] の手法が三次元モデルの各頂点に付与したものをを用いる。

- **従来手法 2:**

文献 [7] の手法に基づいて、三次元モデルの各三角形パッチに対して適切なテクスチャをマッピングした自由視点画像を生成する。

- **提案手法:**

誤差を含む三次元モデルと前景抽出画像を用いた視点依存ジオメトリ・テクスチャによる手法により自由視点画像を生成する。

表 1: 実験に用いた PC の仕様

| 項目         | 詳細                         |
|------------|----------------------------|
| CPU        | Intel Corei7-2600K 3.40GHz |
| メモリ        | RAM 8.0GB                  |
| GPU        | NVIDIA GeForce GTX 560 Ti  |
| OS         | Windows 7 64bit            |
| グラフィック API | OpenGL                     |

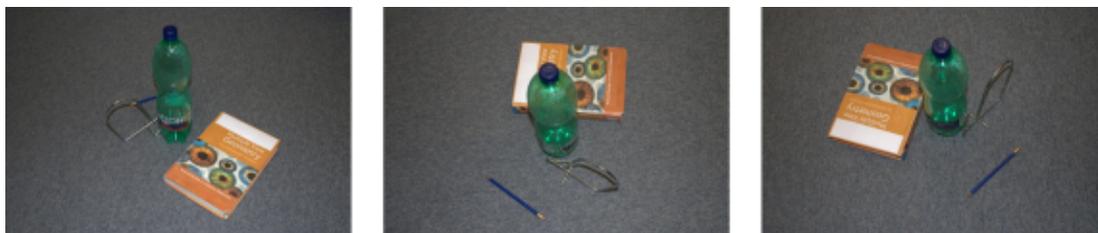
## 4.2 画像データセットと実装の詳細

本実験では，図 18(a) に示すペットボトルと本を撮影した画像データセット (DS1)，図 19(a) に示す東大寺の東塔を復元した模型を撮影した画像データセット (DS2)，及び図 20(a) に示す棚を撮影した画像データセット (DS3) を用いて各手法の比較を行った．DS1 は  $1950 \times 1308$  画素の画像 24 枚，DS2 は  $1920 \times 1080$  画素の画像 571 枚（間引いた後の画像 59 枚），DS3 は  $1920 \times 1080$  画素の画像 89 枚（間引いた後の画像 22 枚）から構成される．

手順 (A-1) では，データセット中の撮影画像のカメラ内部パラメータ，及びレンズ歪み係数を Zhang の手法 [28] により算出し，カメラ外部パラメータは VisualSFM [29] を用いて算出した．また，三次元モデルは画像データセットと求めたカメラパラメータから CMPMVS [3] を用いて生成した（図 21，図 22，図 23）．手順 (A-2) における三次元モデル生成の結果，各データセットに対して生成された三次元モデルのパッチ数はそれぞれ 36,051，19,408，及び 144,493 となった．手順 (A-3) の撮影画像の前景抽出には，GrabCut [30] を用い，対象物体の領域に対して前景ラベルを，それ以外の領域に対して背景ラベルを付与した．前景抽出した撮影画像の例を図 18(b)，図 19(b)，図 20(b) に示す．手順 (B-1) で推定する実空間を撮影するカメラの位置・姿勢は，ARToolKit [31] により取得した．なお，本実験では世界座標系は ARToolKit [31] のマーカ上に設定した．

## 4.3 生成画像の比較

前節で示した画像データセットを用いて，前述の 3 つの手法により自由視点画像を生成した．なお，オンライン処理に入力する実環境映像は，比較のために事前に撮影し，各データセットで異なる物 (DS1: 470 フレーム，DS2: 460 フレーム，DS3: 300 フレーム) を利用した．実環境映像の取得には， $720 \times 480$  画素の解像度の画像を 30fps で撮影可能なカメラを用いた．表 2 に本実験で用いた提案手法のパラメータの設定を示す．パラメータ  $K$  と反復回数は，提案手法の生成結果が良好となる値を設定した．DS1 に対して従来手法 1，従来手法 2，及び提案手法を用いて生成した自由視点画像をそれぞれ図 24，図 25 及び図 26 に，DS2 に



(a) 撮影した画像の例



(b) 前景抽出した画像の例

図 18: 画像データセット 1(DS1)



(a) 撮影した画像の例

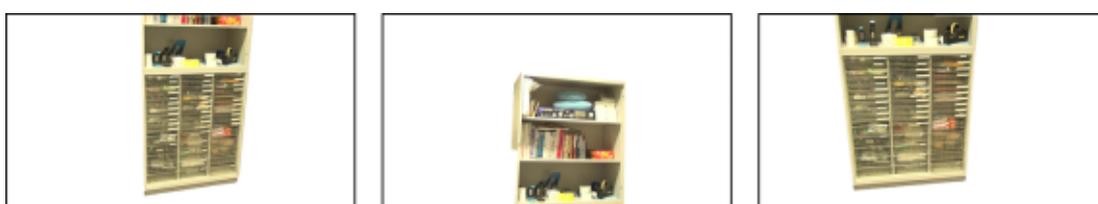


(b) 前景抽出した画像の例

図 19: 画像データセット 2(DS2)



(a) 撮影した画像の例



(b) 前景抽出した画像の例

図 20: 画像データセット 3(DS3)

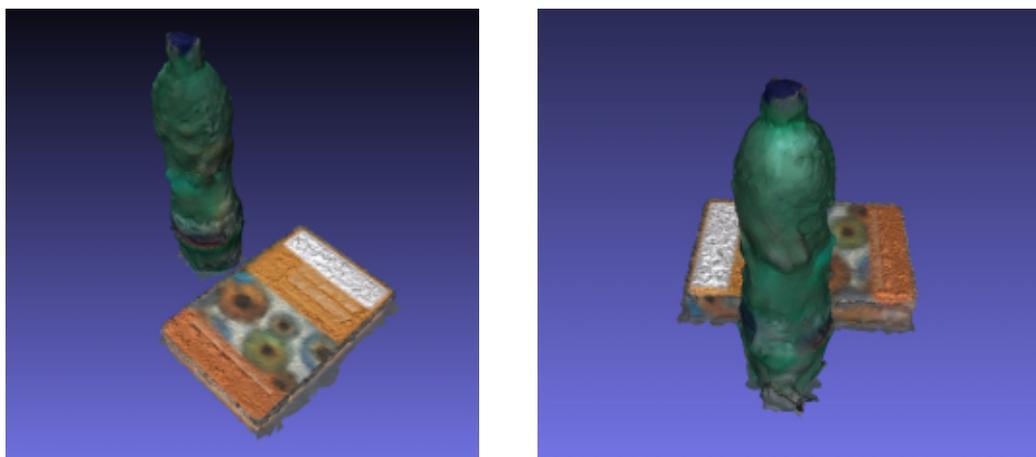


図 21: CMPMVS [3] により DS1 を用いて生成した三次元モデル



図 22: CMPMVS [3] により DS2 を用いて生成した三次元モデル

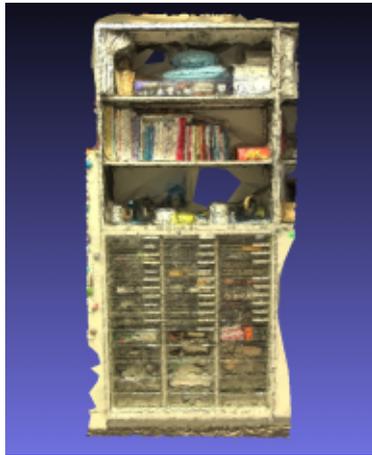


図 23: CMPMVS [3] により DS3 を用いて生成した三次元モデル

表 2: 各データセットの提案手法における実験のパラメータ

|                               | DS1        | DS2        | DS3        |
|-------------------------------|------------|------------|------------|
| 視点依存ジオメトリの生成に用いる<br>奥行き画像のサイズ | 720×480 画素 | 720×480 画素 | 720×480 画素 |
| 拡張処理に用いるフィルタサイズ $K$           | 3          | 3          | 3          |
| 拡張処理の反復回数                     | 10         | 20         | 20         |

表 3: DS1, DS2, DS3 における各手法のフレームレート

| 手法     | DS1      | DS2      | DS3      |
|--------|----------|----------|----------|
| 従来手法 1 | 34.33fps | 37.36fps | 32.80fps |
| 従来手法 2 | 8.80fps  | 5.91fps  | 3.03fps  |
| 提案手法   | 3.69fps  | 1.44fps  | 2.51fps  |

対して従来手法 1, 従来手法 2, 及び提案手法を用いて生成した自由視点画像をそれぞれ図 27, 図 28, 及び図 29 に示す. DS3 に対して従来手法 1, 従来手法 2, 及び提案手法を用いて生成した自由視点画像をそれぞれ図 30, 図 31, 及び図 32 に示す. また, DS1, DS2, DS3 における各手法のフレームレートを表 3 に示す.

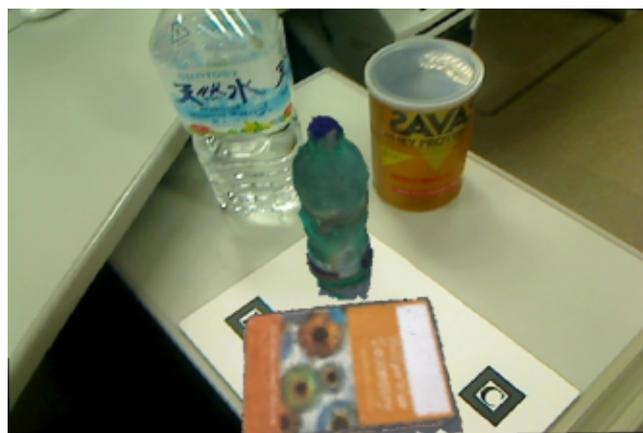


図 24: DS1 を用いた従来手法 1 による生成結果

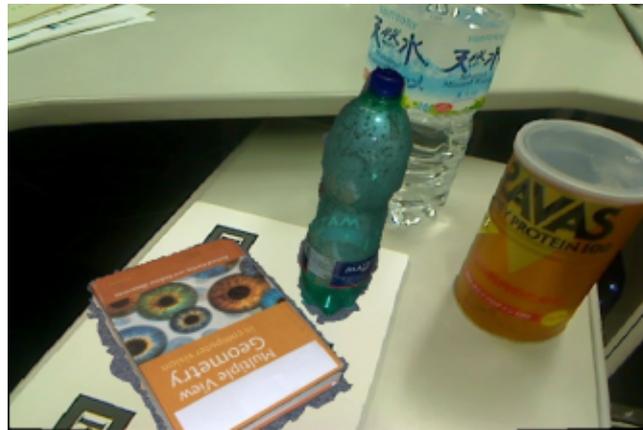


図 25: DS1 を用いた従来手法 2 による生成結果

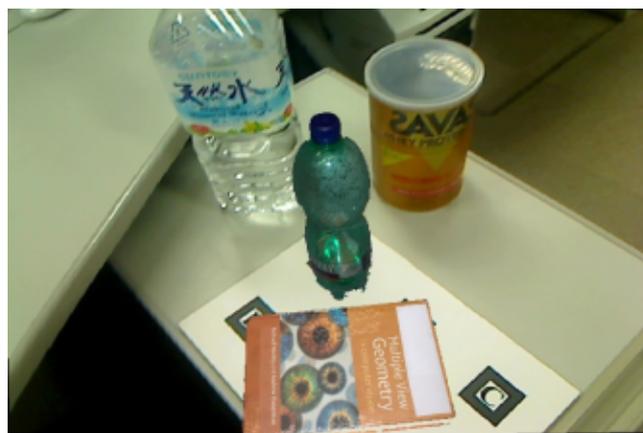
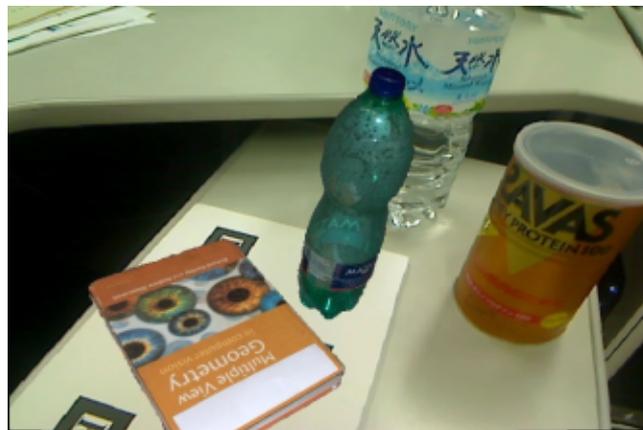


図 26: DS1 を用いた提案手法による生成結果

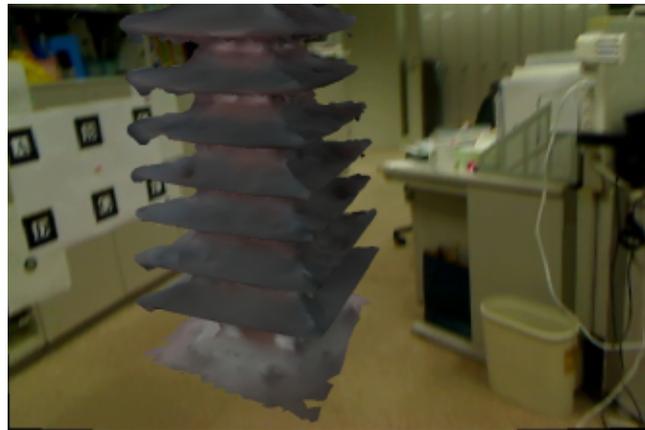


図 27: DS2 を用いた従来手法 1 による生成結果

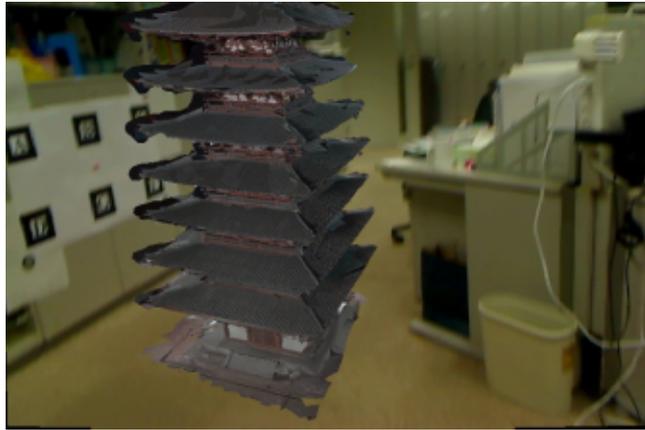


図 28: DS2 を用いた従来手法 2 による生成結果

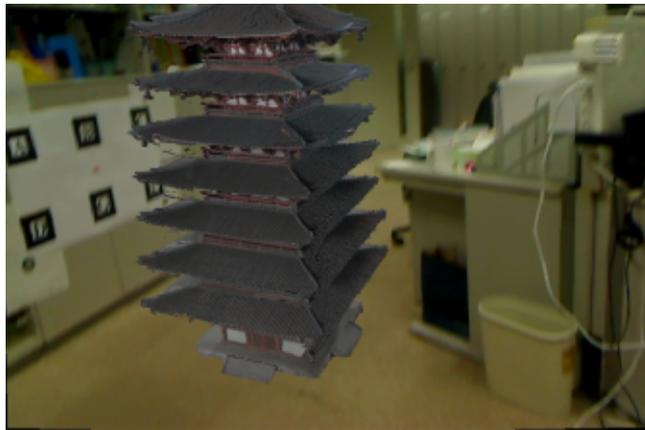
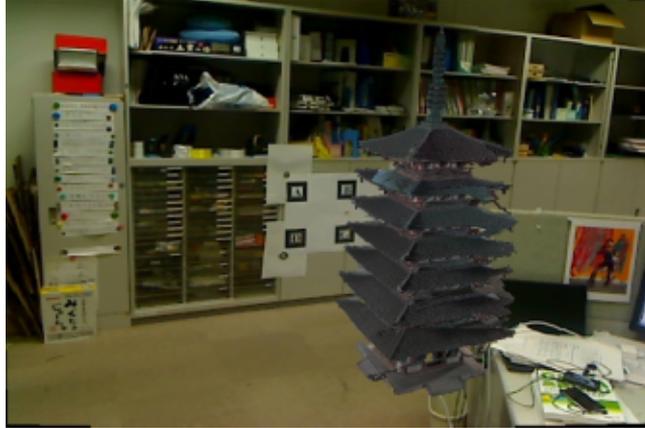


図 29: DS2 を用いた提案手法による生成結果



図 30: DS3 を用いた従来手法 1 による生成結果



図 31: DS3 を用いた従来手法 2 による生成結果



図 32: DS3 を用いた提案手法による生成結果



図 33: テクスチャとして選択される画像の切り替わりの例 (DS2)

#### 4.4 考察

従来手法1では、図24に示すように、ペットボトルのふたの部分に、本来の青色とは異なる色が含まれている。この問題は、文献[3]の手法における三次元モデルの推定誤差と頂点の色の決定方法に起因するものである。文献[3]の手法では、三角形パッチを構成する各頂点を入力である撮影画像に投影し、撮影画像中の投影した位置における色を付与することで頂点の色を決定する。このとき、パッチの頂点位置や外部パラメータの誤差により投影した座標が本来の位置からずれている場合に、頂点に本来の色とは異なる色が付与されるためである。

従来手法2では、図25に示すように、撮影画像中の仮想化実物体に対応する領域に基づき色が決定されるため、仮想化実物体の本来の色が忠実に再現されている。しかし、図25において、ペットボトルのふたの輪郭付近で推定誤差による膨張が生じ、オフラインでの画像取得時の背景の色が輪郭付近に合成される。これは、図28の下段においても、DS2の撮影画像(図19)との比較から確認できるように、塔の先端に輪郭の膨張が生じている。加えて、図25の上段と下段においては、ペットボトルの左側の輪郭付近の欠損により、本来のペットボトルの形状が再現できていない。これは、図31の中段においても、上から三段目の棚に壁の欠損が生じている。また、図28の上段に示すように、塔の屋根付近において、色が不連続に変化する部分が存在する(図33(a)参照)。これは、式(3),(7)によ

りテクスチャとして選択される画像が切り替わるためである。

提案手法では、図 26 に示すように、従来手法 2 と同様に、撮影画像中の仮想化実物体に対応する領域に基づき色が決定されるため、画像中の本来の色が忠実に再現されている。加えて、図 26 や図 29 の下段に示すように、提案手法では輪郭付近の膨張を視覚的に低減している。提案手法では撮影画像の前景を抽出しているため、仮想化実物体以外の領域の自由視点画像生成への利用を防ぐことができる。また、欠損領域に関しても提案手法では視点依存ジオメトリにより輪郭を拡張させるため、前景抽出による膨張領域の修復との組み合わせにより、撮影画像中の仮想化実物体に忠実な輪郭を再現可能である。しかし、図 29 の塔の屋根付近において従来手法 2 と同様に、色が不連続に変化する部分が存在する（図 33(b) 参照）。

なお、提案手法の処理速度は、画像データセットに含まれる画像数に依存した結果となった。これは、テクスチャの選択処理において計算回数が増加するためである。

## 5. 提案手法のアプリケーション例

提案手法のアプリケーションとして、現在では失われた文化遺産の提示が挙げられる。日本各地には、多くの文化遺産が存在している。しかし、これらの文化遺産では、当時の形をそのまま残しているものは少なく、観光客などが当時の様子を想像することは難しい。このため、近年、文化遺産のARによる提示が試みられている [32]。文化遺産は当時の様子を再現した模型が存在する場合があります、このような場合、模型の画像から提案手法によってその模型を仮想化実物体として実環境に合成できる。これにより、観光客は遺構上にかつて存在した文化遺産の当時の様子を知ることができる。

図 34 は、奈良県の東大寺において AR による東塔の復元を行った画像の例である。東大寺には、当時の東塔を復元した模型が存在し、この模型を撮影した画像データセットを入力として提案手法を用いた。ただし、仮想化実物体の位置合わせは手動で行った。



図 34: 提案手法による東塔の重畳例

## 6. まとめ

本研究ではARでの応用を想定し、視点依存テクスチャ・ジオメトリによる仮想化実物体の自由視点画像上での輪郭形状の修復手法を提案した。提案手法では、自由視点画像上で三次元モデルにおける輪郭に対応する部分を拡張させることにより、形状の欠損の影響を低減させている。さらに、テクスチャとして用いられる画像の前景領域を事前に抽出し、視点依存テクスチャにより選択された画像の前景領域のみを用いることで、自由視点画像上での膨張領域を修復する。提案手法は、現在広く研究されている三次元復元手法等で自動生成された誤差を含む三次元モデルから高品位な自由視点画像を生成可能にするものであり、ARのコンテンツとして利用される実物体の自由視点画像生成に係る人的コストを削減することで、ARの実利用の促進に大いに貢献すると考える。

実験では、実時間で仮想化実物体の合成を行う拡張現実感システムを実装し、モデルベースドレンダリングに基づく手法、及び視点依存テクスチャに基づく手法と提案手法を比較することで、提案手法による撮影画像に忠実な色の再現、及び膨張領域と欠損領域の修復が可能であることを確認した。ただし、提案手法は視点に応じてテクスチャを切り替えるので、生成した自由視点画像中にテクスチャの不連続が生じた。また、提案手法の処理速度は実時間での利用が前提のARアプリケーションを考慮すると不十分な結果となった。

今後の課題として、テクスチャの滑らかな切り替わりの実現や、拡張処理におけるパラメータ  $K$ 、反復回数の適切な設定方法の検討が挙げられる。また、GPUによる処理の高速化を行う。さらに、Cosegmentation [33] を前景抽出処理に適用することで、撮影画像に対する効率的な前景抽出を図る。

## 謝辞

本研究を進めるにあたり，その全過程において細やかな御指導，御鞭撻を頂いた視覚情報メディア研究室 横矢 直和 教授に心より感謝いたします。そして，本研究の遂行にあたり，有益な御助言，御鞭撻を頂いたインタラクティブメディア設計学研究室 加藤 博一 教授に厚く御礼申し上げます。さらに，本研究の全過程を通して，始終温かい御指導をして頂いた視覚情報メディア研究室 佐藤 智和 准教授に深く感謝申し上げます。また，研究に関して的確な御助言を頂いた視覚情報メディア研究室 河合 紀彦 助教と 中島 悠太 助教に厚く御礼申し上げます。また，研究室での生活を支えていただいた視覚情報メディア研究室 石谷 由美 女史に心より感謝いたします。実験データの収集に協力して下さった，東大寺の方々に深く感謝申し上げます。さらに，研究活動だけでなく日々の生活においても大変お世話になった視覚情報メディア研究室の皆様にも深く感謝いたします。最後に，両親をはじめ，私の二年間の大学院生活にかかわったすべての方々に感謝の意を表します。

## 参考文献

- [1] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. 2006 IEEE Conf. Computer Vision and Pattern Recognition*, CVPR '06, pp. 519–528, 2006.
- [2] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Proc. 2010 IEEE Conf. Computer Vision and Pattern Recognition*, CVPR '10, pp. 1434–1441, 2010.
- [3] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Proc. 2011 IEEE Conf. Computer Vision and Pattern Recognition*, CVPR '10, pp. 3121–3128, 2011.
- [4] H. Baker. Three-dimensional modeling. In *Proc. Int. Joint Conf. Artificial Intelligence*, Vol. 2, pp. 649–655, 1977.
- [5] K. Kolev, T. Brox, and D. Cremers. Fast joint estimation of silhouettes and dense 3D geometry from multiple images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 34, No. 3, pp. 493–505, 2012.
- [6] J. W. Bastian, B. Ward, R. Hill, A. V. D. Hengel, and A. R. Dick. Interactive modelling for ar applications. In *Proc. IEEE Int. Symposium on Mixed and Augmented Reality*, ISMAR '10, pp. 199–205, 2010.
- [7] P. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. In *Proc. 9th Eurographics Rendering Workshop*, 13 pages, 1998.
- [8] 塚本壮輔, 呉本堯, 古賀和利, 三池秀敏. 階層化位相シフト法による高精度な奥行計測. 電子情報通信学会論文誌 D-II, Vol. J83-D-II, No. 9, pp. 1962–1965, 2000.

- [9] H. Kawasaki, Y. Ohsawa, R. Furukawa, and Y. Nakamura. Dense 3d reconstruction with an uncalibrated active stereo system. In *Proc. 7th Asian Conf. Computer Vision, ACCV '06*, Vol. 2, pp. 882–891, 2006.
- [10] 浅井俊弘, 神原誠之, 横矢直和. 全方位距離画像と全方位カラー画像の統合による屋外環境の三次元モデル化. *画像電子学会誌*, Vol. 34, No. 5, pp. 529–538, 2005.
- [11] S. E. Chen and L. Williams. View interpolation for image synthesis. In *Proc. ACM SIGGRAPH '93*, pp. 279–288, 1993.
- [12] S. M. Seitz and C. R. Dyer. View morphing. In *Proc. ACM SIGGRAPH '96*, pp. 21–30, 1996.
- [13] 冨手要, 山澤一誠, 横矢直和. 複数の全方位画像を用いた広範囲な任意視点画像の生成. *電子情報通信学会 技術研究報告*, Vol. 26, No. 9, pp. 65–70, 2002.
- [14] 稲本奈穂, 斎藤英雄. 多視点スポーツ映像からの自由視点映像合成と提示. *電子情報通信学会論文誌 D-II*, Vol. J88-D-II, No. 8, pp. 1693–1701, 2005.
- [15] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pp. 3–20. 1991.
- [16] T. Naemura, T. Takano, M. Kaneko, and H. Harashima. Ray-based creation of photo-realistic virtual world. In *Proc. 1997 Int. Conf. Virtual Systems and MultiMedia, VSMM '97*, pp. 59–68, 1997.
- [17] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. ACM SIGGRAPH '96*, pp. 31–42, 1996.
- [18] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. ACM SIGGRAPH '96*, pp. 43–54, 1996.
- [19] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. In *Proc. ACM SIGGRAPH '00*, pp. 369–374, 2000.

- [20] 南谷真哉, 北原格, 亀田能成, 大田友一. 変型ビルボードを用いた人物像の提示による複合現実感卓上作業の遠隔共有. *日本バーチャルリアリティ学会論文誌*, Vol. 13, No. 3, pp. 363–373, 2008.
- [21] T. Ishikawa, K. Yamazawa, and N. Yokoya. Novel view generation from multiple omni-directional videos. In *Proc. 4th IEEE/ACM Int. Symposium on Mixed and Augmented Reality*, ISMAR '05, pp. 166–169, 2005.
- [22] 山崎俊太郎, 佐川立昌, 川崎洋, 池内克史, 坂内正夫. 視点依存の微小面を用いた複雑な表面形状を持つ物体の表示方法. *電子情報通信学会論文誌 D-II*, Vol. J86-D-II, No. 10, pp. 1441–1449, 2003.
- [23] 高橋桂太, 苗村健. 視点依存奥行きマップ実時間推定に基づく多眼画像からの自由視点画像合成. *映像情報メディア学会誌*, Vol. 60, No. 10, pp. 1611–1622, 2006.
- [24] 伊吹拓也, 佐藤智和, 松下康之, 横矢直和. 視点位置に依存して変形する三次元メッシュモデルを利用した自由視点画像生成における違和感の低減. *電子情報通信学会 技術研究報告*, Vol. 109, No. 470, pp. 437–442, 2010.
- [25] S. Prince, A. D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato. 3D live: Real time captured content for mixed reality. In *Proc. 1st Int. Symposium on Mixed and Augmented Reality*, ISMAR '02, pp. 7–13, 2002.
- [26] R. Azuma. A survey of augmented reality. *Presence: Teleoperators and virtual environments*, Vol. 6, No. 4, pp. 355–385, 1997.
- [27] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, Vol. 80, No. 2, pp. 189–210, 2008.
- [28] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330–1334,

2000.

- [29] W. Changchang. *VisualSFM: A visual structure from motion system*, 2011.  
<http://www.cs.washington.edu/homes/ccwu/vsfm/>.
- [30] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *Proc. ACM SIGGRAPH '04*, Vol. 23, pp. 309–314, 2004.
- [31] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proc. 2nd IEEE/ACM Int. Workshop on Augmented Reality, IWAR '99*, pp. 85–94, 1999.
- [32] 角田哲也, 大石岳史, 池内克史. 高速陰影表現手法を用いた飛鳥京 MR コンテンツの開発とその評価. Vol. 62, No. 9, pp. 1466–1473, 2008.
- [33] C. Rother, V. Kolmogorov, T. Minka, and A. Blake. Cosegmentation of image pairs by histogram matching — incorporating a global constraint into MRFs. In *Proc. 2006 IEEE Conf. Computer Vision and Pattern Recognition, CVPR '06*, pp. 993–1000, 2006.