NAIST-IS-MT1451080

修士論文

運動視差を再現した事前生成型拡張現実感

西崎 優弥

2016 年 3 月 14 日

奈良先端科学技術大学院大学
情報科学研究科

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。

西崎 優弥

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）
河合 紀彦 助教　（副指導教員）

運動視差を再現した事前生成型拡張現実感*

西崎 優弥

内容梗概

可搬型デバイスを用いたモバイル拡張現実感 (Augmented Reality : AR) は，既に広く普及しているスマートフォンやタブレット端末・ゲーム端末を用いてどこでも手軽に現実シーンと仮想シーンを合成した AR 画像を提示できることから注目されている．一般に，臨場感の高いモバイル AR を実現するためには，実世界と仮想物体の位置合わせ問題である幾何学的整合性問題を解決することが重要となるが，小型デバイス上において複雑な位置合わせ処理や外部センサを利用することは難しく，従来のモバイル型拡張現実感システムでは位置合わせ精度が十分でないことに起因してジッタが生じ，高い臨場感が得られないという問題があった．近年，この問題を解決するために，事前生成型 AR と呼ばれる新たなモバイル AR の実現手法が提案されている．事前生成型 AR では，対象となるシーンを事前に撮影した全方位画像にあらかじめ仮想物体を合成しておき，端末内のセンサ等より取得した端末の方向に応じてその画像を切り出し提示することで，ジッタのない AR を実現している．ただしこの手法は，撮影された地点からの景観のみしか提示できず，端末の移動に伴う運動視差が再現されないため，ユーザの移動時には臨場感が低下するという問題が残されていた．

本論文では，事前生成型 AR に任意の視点からの映像を生成する自由視点画像生成技術を組み合わせることで端末の移動に応じた運動視差を再現する，新たな事前生成型 AR システムを提案する．提案システムは従来の事前生成型 AR システムと同様，必要なデータを事前に収集・生成するオフラインステージと，ユー

*奈良先端科学技術大学院大学 情報科学研究科 修士論文, NAIST-IS-MT1451080, 2016 年 3 月 14 日.

ザによるシステム利用時の提示処理を行うオンラインステージにより構成される．オフラインステージでは，対象となるシーンを全方位カメラにより複数地点から撮影し，Structure-from-Motion(SfM) 法および多視点ステレオ (MVS) 法を用いて対象シーンの密な三次元形状を復元する．オンラインステージでは，シーン内における端末の相対運動を推定し，推定された仮想視点位置に応じて適切なテクスチャを選択的に使用する視点依存テクスチャマッピングにより実時間で自由視点画像を生成しユーザに提示する．ここで本研究では，運動視差は画像上の相対的な物体の移動であることからカメラの相対運動により再現可能であることに着目し，処理コストが高くロバスト性に欠けるカメラの絶対位置姿勢の推定は行わず，フレーム間の相対的なカメラ運動を推定する．具体的には，カメラの姿勢情報の推定には従来の事前生成型 AR システムと同様に端末内蔵のジャイロセンサを用い，相対的な並進運動の推定には撮影画像上の少数の特徴点を用いる．これにより，低コストかつロバストにカメラ運動を推定し，これを用いて自由視点画像生成を行うことにより実時間で運動視差を再現する事前生成型 AR 画像の提示を実現する．

実験では，従来の事前生成型 AR を比較対象とした実験を行い，提案手法による運動視差再現効果およびユーザへの提示画像の品質評価を行うことにより，提案手法の有効性を検証する．

キーワード

モバイル拡張現実感，事前生成型拡張現実感, 自由視点画像生成, 運動視差, 幾何学的整合性

Indirect augmented reality with motion parallax*

Yuya Nishizaki

Abstract

There has been considerable interest in mobile augmented reality (AR) using portable devices such as smartphones and tablets because mobile AR can show images with the combination of real and virtual scenes anywhere. In AR applications, the geometric registration to accurately place virtual objects is one of the most essential requirements. However, it is difficult for small devices to conduct complex processes for alignment and to use the external sensors for accurate registration. Jitters between real and virtual scenes caused by such inabilities decrease realistic sensation in conventional mobile AR systems.

To solve the problem, a method called indirect AR which never produces jitters has been recently proposed. This method uses a pre-captured omnidirectional image and superimposes virtual objects on the image in advance. A part of the composite image is cropped according to the direction obtained by internal sensors of a device and is presented to a user. However, this method cannot represent the motion parallax caused by the motion of the device, which give users the feeling of low realistic sensation.

This thesis proposes a novel indirect AR method that can reproduce motion parallax depending on the movement of portable devices. It can be achieved by the combination of a conventional indirect AR technique and free viewpoint

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1451080, March 14, 2016.

image generation. Our proposed system is divided into an offline stage for prior data collection and an online stage for real-time video presentation. In the offline process, images are captured by an omnidirectional camera at multiple points in a target environment and its dence 3D model is reconstructed using Structure-from-Motion (SfM) and Multi-View Stereo (MVS) techniques. In the online process, the relative motion of the camera is estimated and the free-viewpoint image, which reproduces motion parallax, is generated and presented to users. The free-viewpoint images are generated by view-dependent texture mapping, which selects appropriate textures according to the estimated motion of the device.

This study estimates relative camera motion between frames instead of estimating absolute camera poses for which the computational cost is high and the robustness is low. Because the motion parallax can be produced by object movement in images caused by the relative camera motion. In particular, camera posture is estimated by a gyro sensor embedded in mobile device and camera translation is estimated by tracking a small number of feature points in images. The motion parallax in the image can be reproduced with real-time, low-cost and high-robustness of the proposed method.

In experiments, we compare the proposed method with the conventional indirect AR to verify the effectiveness of the reproduction of the motion parallax and to verify the quality of images presented to users.

Keywords:

mobile augmented reality, indirect augmented reality, free viewpoint image generation, motion parallax, geometric consistency

目 次

図 目 次

1. はじめに

スマートフォン等を用いたモバイル拡張現実感 (Augmented Reality：AR) は，博物館において展示物に仮想的な付加情報を与え新たな鑑賞体験を行うシステム [1] や都市計画における建物の建設シミュレーション [2] など，様々な応用分野での利用が期待されている．一般に，臨場感の高いモバイル AR を実現するためには，実世界と仮想物体の位置合わせ問題である幾何学的整合性問題を解決することが重要となる．幾何学的整合性問題を解決するためには，端末カメラの位置姿勢を推定する必要があり，端末に内蔵された加速度センサ，ジャイロセンサ等を組み合わせた手法 [2,3] や，撮影画像群を用いた手法 [4–10] などが提案されている．しかし，これらの手法においては，端末カメラの位置姿勢の推定誤差により撮影画像と仮想物体に位置ずれが発生し，ジッタ等により臨場感が損なわれるといった問題がある．

このような位置ずれの問題を解決するために，事前生成型 AR と呼ばれる新たなモバイル AR の実現手法が提案されている [11,12]．事前生成型 AR では，対象となるシーンを事前に撮影した全方位画像にあらかじめ仮想物体を合成しておき，端末カメラの内蔵センサ等より取得した端末カメラの方向に応じて合成画像を切り出し提示することで，仮想物体に位置ずれの生じない AR を実現している．しかし，事前生成型 AR では，あらかじめ撮影された地点からの景観のみしか提示できず，端末カメラの移動に伴う運動視差が再現されないことから相対的に大きな移動を伴うシーンでは臨場感が低下するという問題がある．

本研究では，任意の視点からの映像を生成する自由視点画像生成技術 [13] を事前生成型 AR に取り入れることで，端末カメラの移動に応じた運動視差を再現する新たな事前生成型 AR システムを提案する．提案システムは従来の事前生成型 AR システム [11,12] と同様，必要なデータを事前に収集・生成するオフラインステージと，ユーザによるシステム利用時の提示処理を行うオンラインステージにより構成される．オフラインステージでは，対象となるシーンを全方位カメラにより複数地点で撮影し，Structure-from-Motion(SfM) 法 [14] 及び多視点ステレオ (MVS) 法 [15] を用いて対象シーンの密な三次元形状を復元する．オンラインステージでは，まず，シーン内における端末カメラの相対運動を推定する．ここで，

カメラの姿勢成分についてはセンサを用いるが，相対的な並進成分は撮影画像フレーム間のオプティカルフローによる少数の特徴点組から推定する．また，推定に RANSAC を用いることで低コストかつロバストなカメラ運動の推定を実現する．さらに，センサの誤差の蓄積を解消するため，位置補正処理を行う．これにより，アプリケーションの想定移動範囲において位置・姿勢推定が大きく破綻することを防ぐ．また，推定された仮想視点位置に応じて適切なテクスチャを選択的に使用する視点依存テクスチャマッピング [13] により実時間で自由視点画像を生成し，仮想物体を重畳合成した映像をユーザに提示する．

本論文では，2 章において従来のモバイル AR における幾何位置合わせに関する研究及び自由視点画像生成技術について概説し，本研究の位置づけを述べる．3 章では提案する運動視差を再現した事前生成型 AR システムの処理について述べる．4 章では，提案システムを複数の状況において従来の AR システムと比較するための実験の概要と結果について述べる．最後に，5 章でまとめ及び今後の展望について述べる．

2. 関連研究と本研究の位置づけ

本章では，モバイル端末を用いた拡張現実感における幾何位置合わせに関する従来研究を概説し，次に，本研究で用いる自由視点画像生成に関する従来研究を紹介する．最後に，それらの関連研究に対する本研究の位置づけを述べる．

2.1 モバイル拡張現実感における幾何位置合わせに関する研究

近年，スマートフォンやタブレット端末等のモバイル端末を用いたモバイル AR が広く利用されている．臨場感の高いモバイル AR を実現する上で，重畳される仮想物体に対してジッタやドリフトなど，現実物体と仮想物体の間の位置ずれがないこと，すなわち幾何学的整合性問題を解決することは重要な課題の一つである．

モバイル AR における幾何学的整合性問題を解決する手法に関する従来手法として，端末内蔵のセンサを用いて自己位置・姿勢を推定するセンサベースの手法と端末付属のカメラで撮影した画像から端末カメラの位置･姿勢を推定するビジョンベースの手法がある．センサベースの手法は，GPS やジャイロセンサ，加速度センサ等を用いたデッドレコニングにより得られる端末位置，ジャイロセンサ，電子コンパスにより得られた端末の姿勢から自己の位置姿勢を推定する [3]．センサを用いる手法は計算コストが小さいためモバイル端末で軽快に動作が可能であるが，計測誤差に起因して画素単位での位置合わせが困難であるといった問題点がある．従って，近年では撮影画像を用いたカメラの位置・姿勢推定を行うビジョンベースの手法に関する研究が主流となっている．

ビジョンベースの手法は，対象シーンに関する事前知識を用いない手法 [4–6] と，用いる手法 [7–10] に分けられる．以下，それぞれの手法の代表的な研究を紹介する．

2.1.1 事前知識を用いない手法

事前知識を用いない代表的な幾何位置合わせ手法に，Visual-SLAM (Simulataneous Localization and Mapping) がある．Visual SLAM は，入力画像中の自然特

徴点を追跡することにより，カメラの位置・姿勢推定，及び自然特徴点の三次元位置の推定・更新を繰り返すことによりシーンの三次元形状を取得・更新するものである．これらの代表的な手法として，PTAM [4], LSD-SLAM [5], SVO [6] などが挙げられる．PTAM [4] は，シーンの三次元形状のマッピングと特徴点のトラッキングを並列かつ非同期に行うことで実時間処理を実現している [4]．LSD-SLAM [5] はデプス推定と画像上の勾配から大規模なマップ環境を構築する手法である．SVO [6] は，単眼カメラを用いて高速で正確な位置推定を実現している．これら Visual-SLAM のアプローチによる AR の実現手法では，広域な環境を対象とした場合において，カメラ位置・姿勢の推定誤差が蓄積するという問題や，テクスチャの無いシーンでは処理が破綻してしまうという問題があり，位置合わせのロバスト性に欠けるという課題が残されている．

2.1.2 事前知識を用いた手法

事前知識を用いる手法では，人工マーカ，環境の三次元モデル・自然特徴ランドマーク等を事前知識として用い端末カメラの絶対位置を推定する．つまり，PnP 問題，事前知識としての特徴点の三次元座標とその特徴点の画像上での検出座標を用いて投影誤差を最小化するようなカメラの位置姿勢を推定する．人工マーカを用いる手法 [7] は，シーン内に配置されたマーカの三次元座標と画像中のマーカの二次元座標を対応付けることで，カメラの位置・姿勢推定を行う．この手法は，マーカが撮影できていれば，安定かつ高精度にカメラの位置・姿勢を推定することが可能である．しかし，広範囲にカメラが移動する場合，それに応じてマーカも広範囲に配置する必要があり，マーカにより景観が損なわれるという問題がある．

このような問題を解決するために，環境の三次元モデルを人工マーカの代わりに用いる手法 [8] が提案されている．この手法は，対象物体やシーンの三次元モデルと入力画像との対応関係を求めることでカメラの位置・姿勢を推定する．ただし，精度の高い三次元モデルを事前に生成するための人的コストが高いといった問題点がある．

一方，自然特徴ランドマークを用いる手法 [9, 10] は，Structure from Motion

法 [14] などを利用して事前に構築した三次元点群データベースに含まれる三次元点をカメラ画像中の特徴点を対応付けることによりカメラ位置姿勢を推定する．武富ら [10] らは，このアプローチにおいて，ランドマークに優先度を与えることで，実時間でのカメラ位置姿勢推定処理を実現した．しかし，このようなデータベースを用いる手法では，モバイル端末に大容量のランドマーク情報を格納する必要があるため取り扱いが難しい．

以上のように，事前知識を用いてオンラインでの位置合わせを行う従来手法では，位置推定処理の破綻が生じるため，ロバストかつ高品位な拡張現実感システムを実現することが難しい．

2.1.3 事前生成型 AR に関する研究

近年，制限された状況下において，図 1 に示すような位置ずれのない AR 映像の提示を実現可能な事前生成型 AR と呼ばれる手法 [11] が提案されている．本手法は，撮影画像に対しリアルタイムレンダリング処理を行うのではなく，事前に全方位画像を撮影しレンダリング処理を行った画像を用いる．つまり，ユーザの利用時においては，事前にレンダリングした全方位 AR 画像を，内蔵センサにより得られるモバイル端末の姿勢に応じて切り出して提示する．このような方式を採ることで，位置合わせ処理の破綻やジッタの問題が生じないモバイル AR を実現している．

ただし，事前生成型 AR には，事前撮影画像を用いることに起因するいくつかの問題点が発生する．すなわち，事前撮影したカメラの位置とユーザが体験する位置の差異や，事前に撮影した日時とユーザが体験する日時の違いに起因する提示映像と実シーンの差による違和感が生じる．

前者について，Wither ら [11] は被験者実験により，事前撮影時のカメラ位置とユーザの絶対位置の差が仮想物体の重畳対象地点とユーザ位置の距離の 10%以内であれば，実環境と撮影画像との位置ずれに対してユーザは違和感を持たないことを示している．ただし，この研究では端末位置の移動に伴う相対運動に起因する運動視差については言及されていない．後者について，Okura ら [12] は，事前に様々な日時で撮影した全方位画像群をデータベースとして保持し，ユーザに

図 1 事前生成型 AR[11] の提示映像比較 (左：理想重畳画像，中央：センサベース AR，右：事前生成型 AR)

提示する際にはデータベースの中からその場の照明条件に近いものを選択することで照明環境を考慮した事前生成型 AR を実現している．

以上のように，これまでの事前生成型 AR に関する研究では，事前撮影された画像を用いることを基本としており，撮影された地点からの景観のみしか提示できないことから，端末が移動した際に，その移動に伴う運動視差が再現されないことで，ユーザの臨場感が大きく損なわれるといった課題が残されている．

2.2 自由視点画像生成に関する研究

自由視点画像生成は，対象シーンにおいて任意の地点からの見えを異なる地点で撮影された複数の画像から合成する手法である．従来手法は，シーンの三次元形状を用いるモデルベースドレンダリングに基づく手法 [15,16] と入力視点の画像群を用いるイメージベースドレンダリングに基づく手法 [17–19] およびそのハイブリッド手法 [13] に大別される．

2.2.1 モデルベースドレンダリング

モデルベースドレンダリングに基づく手法 [15,16] では，Multi-View Stereo(MVS) 等の三次元形状復元手法を用いて得られたテクスチャ付きの三次元モデルを用い，仮想視点からの見えを再現する．代表的な三次元復元手法である MVS では，Structure from Motion 法 [14] などで推定された入力画像群のカメラパラメータを基に密な三次元形状を推定する．しかし，テクスチャの無い領域では画像間の対応が

取れないため，三次元形状を推定できない領域に欠損が生じる．Jancosek [15] が提案した CMPMVS では，視体積交差法を取り入れることでテクスチャの無い領域の三次元形状復元を可能としている．また，この手法はフリーソフトウェアとして Web 上で公開されており，広く利用されている [15]．ただし，このようなアプローチにおいて生成される自由視点画像は，使用するモデルの精度に依存し，どれだけ高精度に形状復元を行うのかが重要な問題となる．例えば，植物などの複雑な形状をもつモデルを欠損や歪みなしに高精度に復元することは難しい．また，このアプローチでは鏡面反射を再現することが困難である．

2.2.2 イメージベースドレンダリング

イメージベースドレンダリングに基づく手法は，対象シーンを撮影した入力画像群を変形・合成することにより仮想視点からの見えを生成する技術である．代表的な手法に入力画像を変形・合成するモーフィング [18] と入力画像からライトフィールドを取得するライトフィールドレンダリング [19] がある．これらの手法により，生成される自由視点画像の品質は入力に用いる画像枚数に大きく依存し，高品質な画像を合成するためには，撮影コストの増大や大容量のメモリが必要となるという問題がある．

2.2.3 ハイブリッド手法

モデルベースドレンダリング及びイメージベースドレンダリングの手法の問題を解決する手法として，三次元モデルと複数地点の画像情報の双方を用いるハイブリッドな手法が提案されている．代表的な手法として，視点依存テクスチャマッピング法 [13] と視点依存デプスマップ法 [20] がある．前者は，対象シーンの三次元形状に仮想視点の視点位置に近い視点の入力画像上のテクスチャを適宜マッピングすることにより，三次元復元精度が不十分な場合でも詳細な見えを再現する手法である．仮想視点の見えに近い画像を決定する指標として，Debevec ら [13] は，図 2 に示すように仮想視点から注目画素に対するベクトルと各入力視点から注目画素に対するベクトルの成す角を用いている．後者は，仮想視点位置

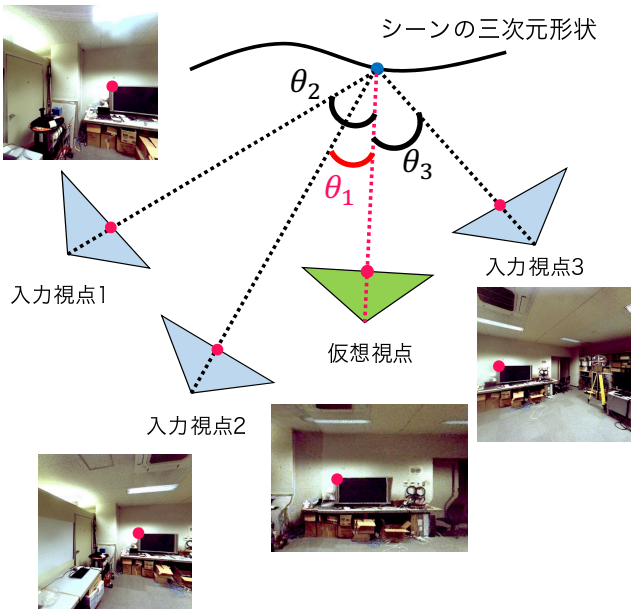


図 2 視点依存テクスチャマッピング

におけるデプス画像を推定・補完し，欠損のない仮想視点の映像を生成する手法である [20]．これらを組み合わせて用いることで，比較的少ない入力で欠損のない写実的な自由視点画像を生成できる．

2.3 本研究の位置づけ

これまでモバイル AR において幾何学的整合性問題を解決する手法が多く提案されてきたが，ジッタやドリフトが発生せず，かつモバイル端末という限られたリソースの中で実時間で動作するシステムの構築は困難であった．しかし近年，高品質な重畳画像合成と軽量に動作可能な事前生成型 AR [11,12] が提案され，ユーザの移動を伴わない限定されたアプリケーションにおいては高い臨場感と高いロバスト性を両立できることが示されている．ただし，このアプローチでは撮影された地点からの景観のみしか提示できず，ユーザの移動に伴う運動視差が再現できないという問題があった．本研究では，事前生成型 AR に自由視点画像生成技術を組み合わせることによって，端末の移動に応じた運動視差を再現する新たな

事前生成型 AR システムを提案する．本研究では，自由視点画像生成手法としてシーンの三次元形状を用いたハイブリッドなイメージベースドレンダリングを採用する．これは，リソースの制限されたモバイル端末上で高品質な自由視点画像を生成するために，少数の入力画像と三次元形状を組み合わせることで詳細な見えを視覚的に再現するためである．従来の事前生成型 AR においては，端末が移動した際においても事前撮影した同一視点からの画像を用いていたが，本研究では，事前撮影した複数地点の画像を用いて端末の移動に基づいた自由視点画像を生成することにより，生成画像において運動視差を再現することが可能となる．本論文では，従来の事前生成型 AR [11] との比較実験を行うことにより本研究の有用性を検証する．

3. 運動視差を再現した事前生成型拡張現実感

3.1 提案手法の概要

本研究では，任意の視点からの映像を生成する自由視点画像生成技術を事前生成型 AR に取り入れることで，端末カメラの動きに応じた運動視差を再現する新たな事前生成型 AR システムを提案する．提案システムは，従来の事前生成型 AR システム [11,12] と同様，必要なデータを事前に収集・生成するオフラインステージと，ユーザによるシステム利用時の提示処理を行うオンラインステージにより構成される (図 3 参照)．以下，オフラインステージ，オンラインステージにおける処理について詳述する．

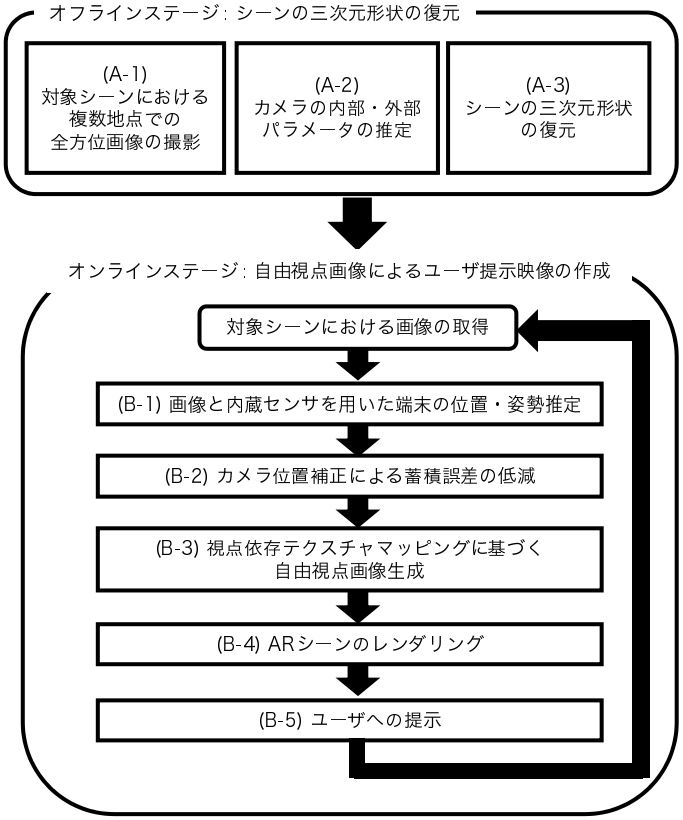


図 3 提案手法の概要

3.2 全方位画像群を用いたシーンの三次元形状の復元

オフラインステージでは，オンラインステージにおいて自由視点画像生成を生成する際に必要となる，(A-1) 対象とするシーンに対する複数地点での撮影画像，(A-2) その画像を撮影したカメラの位置・姿勢情報，(A-3) シーンの三次元形状，を取得・生成する．以下，各項目のデータの作成方法について述べる．

3.2.1 対象シーンにおける複数地点での全方位画像群の撮影

まず，対象とするシーンにおいて，全方位カメラを複数地点に移動設置しながら全方位画像の撮影を行う．図 4 に屋外環境において撮影した全方位画像を示す．また，以降の処理における扱いを容易とするために，取得した全方位画像を図 5 のようなキューブマップに投影する．具体的には，キューブマップを用いることで，Structure from Motion において同一視点における画像間で類似の特徴点を検出しやすくする．

図 4 屋外環境で撮影した全方位画像

図 5 投影処理された屋外環境のキューブマップ

3.2.2 カメラの内部・外部パラメータの推定

処理 (A-1) で取得した撮影画像群に対し，撮影カメラの内部・外部パラメータを推定する．本研究では，フリーソフトとして公開されている VisualSFM [21] を用いた Structure from Motion により推定を行う．図 6 に屋外環境の撮影画像群から推定したカメラの位置・姿勢の例を示す．なお，本研究では安定したカメラ位置の推定を実現するために，図 5 に示したキューブマップを用いる際に各面に対応する仮想カメラの画面を 120 ℃に設定した画像を作成し，これを VisualSFM の入力として用いる．

3.2.3 対象シーンの三次元形状の復元

撮影画像群と推定されたカメラの内部･外部パラメータから，オンラインステージにおける視点依存テクスチャマッピングで利用する対象シーンの三次元形状を CMPMVS [15] 等のマルチビューステレオ法を利用し復元する．図 7 に CMPMVS を用いて復元された屋外環境シーンの三次元形状の例を示す．生成されたモデル

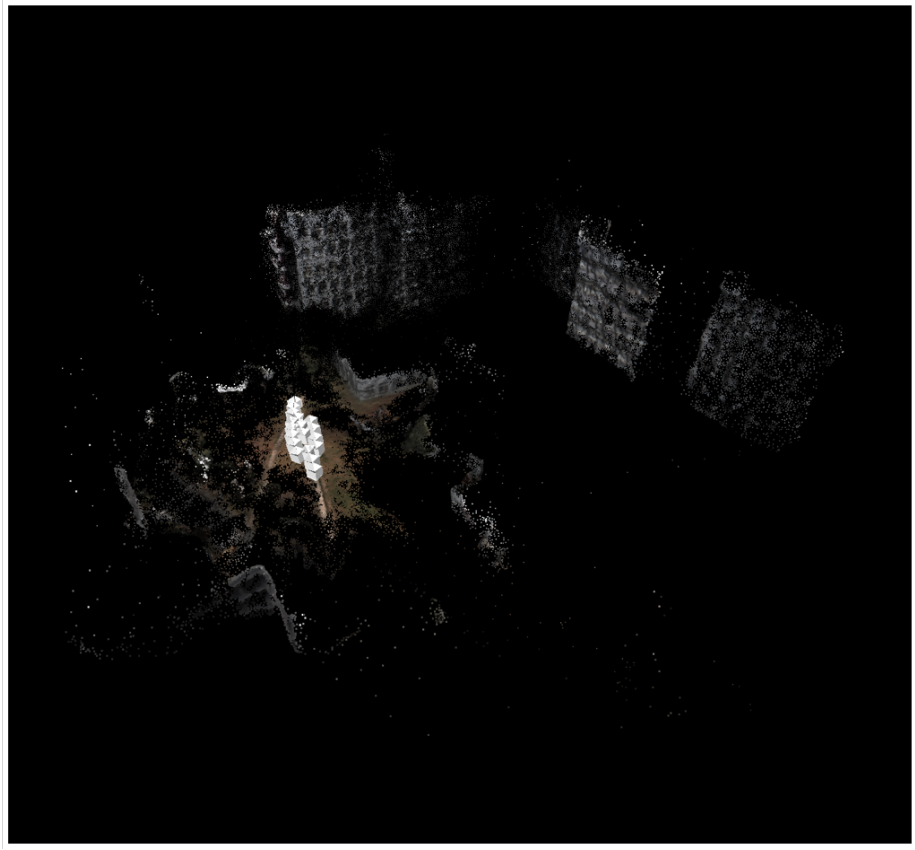


図 6 VisualSFM で推定されたカメラの位置・姿勢および三次元点群 (白物体が推定されたカメラ)

において三次元点の存在しない穴が生じているが，視点依存デプスマップによる奥行き補完処理は比較的処理コストが高いため，ここでは利用を想定するシーンにおいて空以外の穴については撮影範囲に含まれる部分に対して点群データを編集可能なフリーソフトウェアである MeshLab [22] を用いて補間処理を行っている．

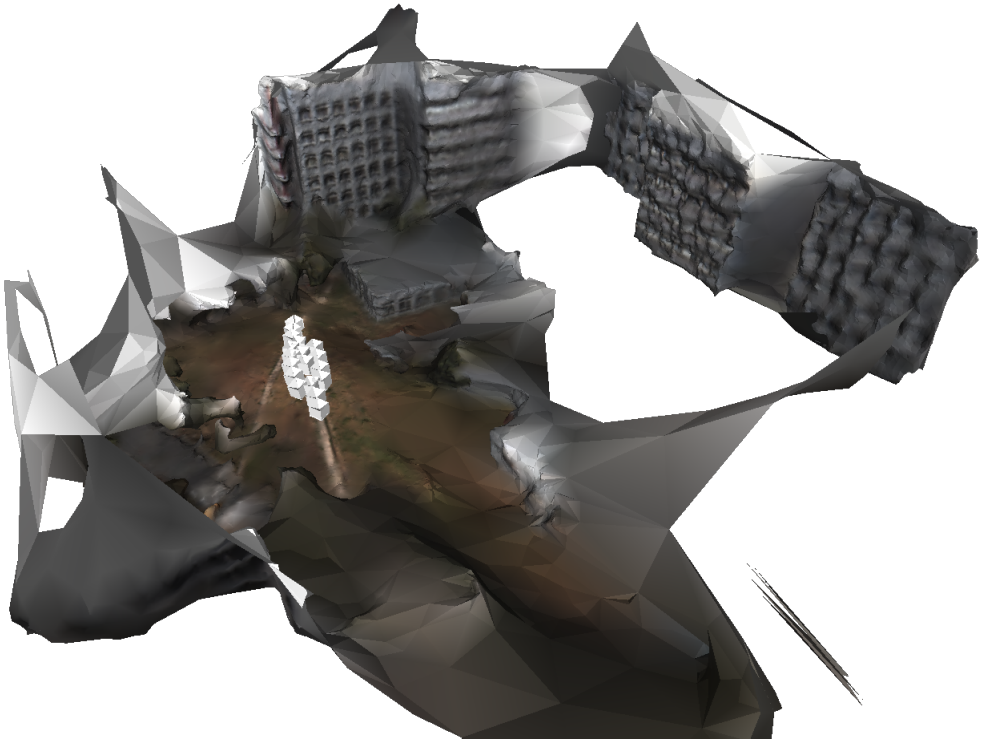


図 7 CMPMVS で復元されたシーンの三次元形状

3.3 自由視点画像によるユーザ提示画像の作成

オンラインステージでは，オフラインステージにおいてあらかじめ取得したデータ，及び対象シーン内において端末カメラより撮影される実画像を入力として，端末カメラ位置の推定と自由視点画像の生成をフレーム毎に繰り返す．以下では，処理 (B-1) から (B-4) について記述する．

3.3.1 画像と内蔵センサを用いた端末の位置姿勢推定

オンラインステージにおけるユーザへの提示映像である自由視点画像を生成するために，対象シーン内においてユーザの保持する端末カメラの位置・姿勢を推定することが必要となる．本研究では，ロバストなカメラの相対運動推定を実現するために，ジャイロセンサと電子コンパスにより得られる姿勢成分はそのまま利用し，並進運動は画像特徴点を用いて推定する．これにより，画像内にごく少数の特徴点しかない場合にもロバストな推定を実現する．

図 8 に i フレーム目における端末カメラの位置姿勢推定の概要を示す．ここで端末カメラの相対運動は，フレーム間での端末カメラの相対的な回転成分 $\boldsymbol{R}$，及び並進成分 $\boldsymbol{t}$ により表されるが，本研究では，相対回転成分 $\boldsymbol{R}$ はセンサから取得した値を用いるため，相対並進成分 $\boldsymbol{t}$ を推定する問題に帰結する．フレーム間での対応点 (u_{i-1}, v_{i-1}), (u_i, v_i)，端末カメラの焦点距離 f，及び対応点の奥行き値 d_{i-1}, d_iとすると，フレーム間の相対位置姿勢の関係は以下の式で表される．

$$\begin{pmatrix} \frac{d_i u_i}{f} \\ \frac{d_i v_i}{f} \\ d_i \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{d_{i-1} u_{i-1}}{f} \\ \frac{d_{i-1} v_{i-1}}{f} \\ d_{i-1} \\ 1 \end{pmatrix} \tag{1}$$

本研究では，この式を用い，以下の (a) から (e) の流れで相対並進成分 $\boldsymbol{t}$ を算出する．ただし，奥行き値 d_{i-1} は $i-1$ フレーム目でのカメラの位置姿勢とシーンの三次元形状から既知とする．

(a) $i-1$ フレーム目の画像において KLT 法 [23] により特徴点を検出する．

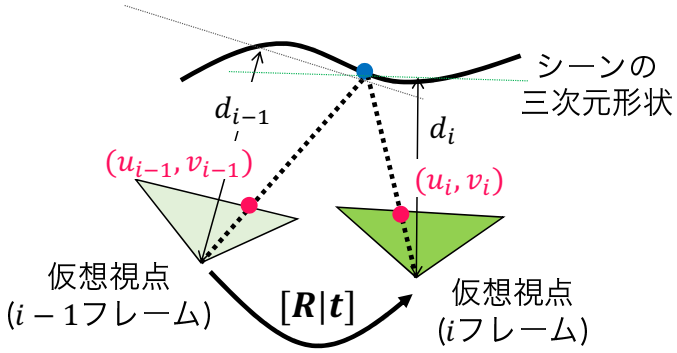


図 8　端末カメラの位置・姿勢推定手法の概要

(b) Lucas-Kanade 法 [24] により特徴点における $i-1$ フレーム目から i フレーム目へのオプティカルフローを推定し，特徴点の対応点を決定する．

(c) 特徴点の中からランダムに 2 点を選び，式 (1) により，並進成分 $\boldsymbol{t}$ 及び d_i を算出する．

(d) $\boldsymbol{R}$ 及び (c) により算出された $\boldsymbol{t}$ を用いて $i-1$ フレーム目における残りの特徴点を i フレーム目に投影し，投影点とオプティカルフローによって求まっている i フレーム目での特徴点間のユークリッド距離を算出する．この距離が，閾値以下の特徴点の数を数える．

(e) (c) と (d) の処理を繰り返し，距離が閾値以下の特徴点が最多となったときの $\boldsymbol{t}$ を最終的な相対並進成分として決定する．

3.3.2　カメラ位置補正による蓄積誤差の低減

提案手法では，処理 (B-1) による端末カメラの相対的な位置姿勢を行うことで，端末の位置姿勢情報を更新する．ここで，システムを長時間実行し続けた場合，センサの誤差と運動推定の誤差の蓄積が発生する．結果として，提示映像の視点位置と実際の端末の位置にずれが生じ，ユーザに違和感を与える．この問題に対して，本研究では，図 9 に示すように，対象シーンにおいて定点を設定し，(B-1)

で得られた端末の位置を定点に近づけるよう端末の移動量に依存した補正を毎フレーム行うことで，アプリケーションの想定移動範囲内において位置・姿勢推定が破綻することを防ぐ．具体的には，定点の位置を $\boldsymbol{x}_0$, $i-1$ フレーム目でのカメラ位置を $\boldsymbol{x}_{i-1}$, i フレーム目での位置補正前のカメラ位置を $\boldsymbol{x}'_i$，移動量に基づいた重みを w としたとき，補正後のカメラ位置 $\boldsymbol{x}_i$ を以下の式で決定する．

$$\boldsymbol{x}_i = \begin{cases} \boldsymbol{x}'_i + w(\boldsymbol{x}_0 - \boldsymbol{x}'_i) & \text{(端末が運動状態のとき)} \\ \boldsymbol{x}_{i-1} & \text{(端末が静止状態のとき)} \end{cases} \tag{2}$$

$$w = \alpha(|\boldsymbol{x}'_i - \boldsymbol{x}_{i-1}|) \tag{3}$$

ここで，端末が静止状態であるとき，位置補正処理を行うとユーザが静止しているにも関わらず提示映像上に運動視差が生じるためユーザに違和感を与える．そこで式 (2) では処理 (B-1) でオプティカルフローにより計算されたフレーム間での全特徴点組のユークリッド距離の平均値がある閾値以上であれば運動状態であるとして位置補正を行い，閾値より小さければ静止状態であるとして $i-1$ フレーム目でのカメラ位置 $\boldsymbol{x}_{i-1}$ を i フレーム目の補正後のカメラ位置 $\boldsymbol{x}_i$ として用いる．ここで，静止状態において前フレームのカメラ位置を用いる理由は，端末が静止状態であっても推定によって微小な移動量が検出されてしまい提示映像の各フレームにおいて画像の微妙なゆれ (ジッタ) が発生してしまう現象を解消するためである．

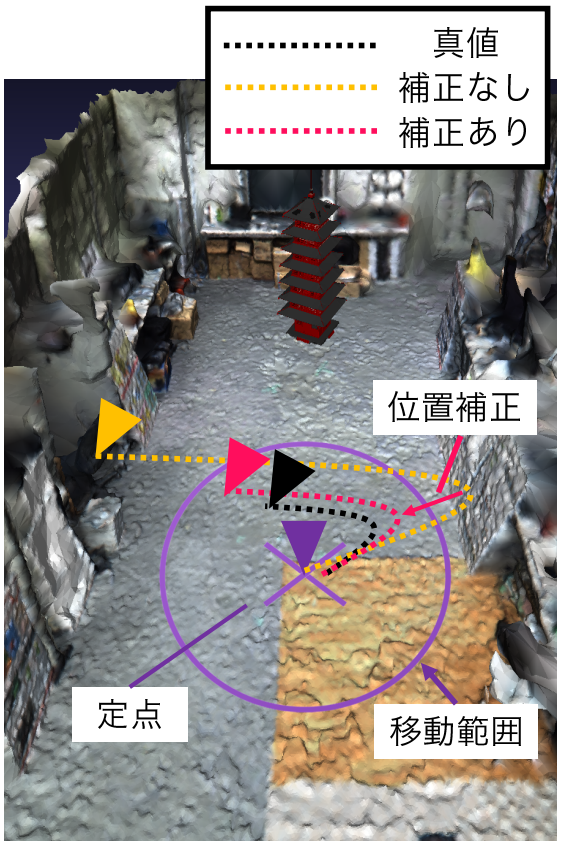


図 9　位置補正処理のイメージ

3.3.3　視点依存テクスチャマッピングに基づく自由視点画像生成

処理 (B-1), (B-2) による端末カメラの位置・姿勢の推定結果に基づき，自由視点画像生成を行う．提案手法では，オフラインステージにおいて取得した対象シーンの三次元形状を用いた視点依存テクスチャマッピング [13] に基づき自由視点画像を生成する．視点依存テクスチャマッピング [13] は，図 2 に示すように，ある仮想視点が定まった場合，その仮想視点から注目画素に対応する三次元点への光線ベクトルと事前取得した各入力視点から同三次元点に対するベクトルの成す角に着目し，その角度が最も小さい入力視点の画素値をマッピングする手法である．図 2 に示した例では，角度が最小となる入力視点 2 の画素がマッピングされる．

(a) ブレンディングなし (b) ブレンディング領域 (c) ブレンディングあり

図 10 ブレンディングの有無による生成結果

ただし，単純にテクスチャを貼り付けていくと，異なる視点における入力画像間の輝度値が異なるため，入力視点が切り替わる領域において不連続な輝度を持つ画像が生成されてしまう．そこで，本研究では，切り替え領域においては，二つの視点からの画像をブレンディングすることで違和感の小さい画像を生成する．具体的には，視点依存テクスチャマッピング処理において，仮想視点のある画素のマッピングを行うとき，選択された最小の角度を θ_{1st}，二番目の角度を θ_{2nd} としたとき，二つの入力視点のテクスチャを $\alpha : 1-\alpha$ の比率でブレンディングする．α は以下の式により決定する．

$$\alpha = \begin{cases} \frac{0.5}{1.0-T_b}\left(\frac{\theta_{1st}}{\theta_{2nd}} - T_b\right) & \left(\frac{\theta_{1st}}{\theta_{2nd}} > T_b\right) \\ 0.0 & \left(\frac{\theta_{1st}}{\theta_{2nd}} < T_b\right) \end{cases} \tag{4}$$

図 10 にブレンディング処理の有無による結果を示す．このように，ブレンディング処理を行うことで，切り替え領域において違和感の小さい映像を提示できている．

なお，本研究では，モバイル端末を用いることから，CPU やメモリ等のリソースが限られており，視点依存テクスチャマッピングを行う際，例えば，全入力視点について角度計算を行うと実時間で画像を生成することが難しい．そこで，本研究では，視点依存テクスチャマッピングにおいていくつかの処理を簡略化することによって，リアルタイム性を実現する．一つ目に，処理 (B-1)，(B-2) により仮

想視点が求められたとき，角度計算を全ての入力視点に対し行うのではなく，仮想視点近傍の三つの視点に対してのみ行うこととした．また，二つ目に，従来の視点依存テクスチャマッピングでは全画素に対して処理が行われるが，本研究では，出力画像をグリッド分割し，各グリッドの中心画素についてのみ角度計算を行うこととした．同一グリッド内の周辺画素については，中心画素の角度計算で求められた入力視点の画素の周辺画素をマッピングすることで処理を高速化する．

3.3.4 AR シーンのレンダリング

本処理では，生成した自由視点画像上に事前に位置を与えた仮想物体を重畳合成する．推定された端末カメラの位置姿勢に基づく自由視点画像が生成された後，画像内に仮想物体を重畳する．なお，自由視点画像生成と仮想物体の重畳は，共に処理 (B-1)，(B-2) で推定された位置姿勢を用いるため，対象シーンの自由視点画像と仮想物体間に幾何学的な位置ずれは発生しない．

4. 実験

本節では，3 章で示したシステムの有効性を検証するため，屋内外環境において従来の事前生成型 AR [11] との比較実験を行った．初めに予備実験として，特徴点トラッキングが容易な屋内環境において運動視差の比較実験を行った．また，応用アプリ―ケーションを想定した場合である屋外環境における実験においては，処理 (B-2) の位置補正処理による蓄積誤差の低減効果の検証，及び提案システムと従来の事前生成型 AR との臨場感の比較を行った結果について述べる．

4.1 実験 1 : 屋内環境における動作確認

4.1.1 実験条件

本実験では，オフラインステージにおいて，全方位マルチカメラシステム Ladybug 3 を用いて屋外環境における 28 か所で全方位画像を撮影し，それを各画像が 600×600 画素のキューブマップに展開し，VisualSFM [14] 及び CMPMVS [15] の入力とした．図 11 に CMPMVS により復元された屋内環境の三次元モデルを示す．また，仮想物体については図 12 に示す 7 重の塔の CG モデルを用いる．オンラインステージでは，端末として Surface Pro3(CPU:Core i7 1.7GHz, メモリ: 8GB) を使用し，入力画像，出力画像の解像度は共に 640×480 画素とした．また，図 12 に示した仮想物体は，図 13 に示すような位置にあらかじめ手動で配置した．

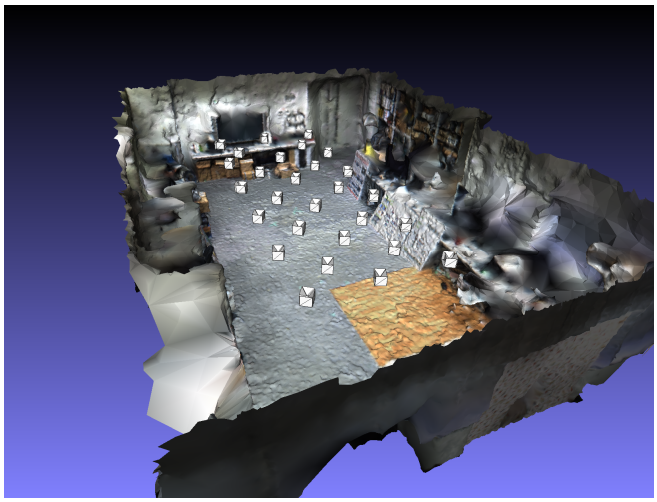


図 11　屋内環境の三次元モデル

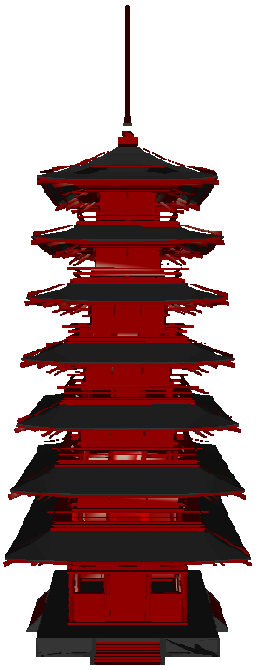


図 12　AR シーンに用いられる仮想物体

図 13　屋内環境における仮想物体との位置合わせイメージ

4.1.2　運動視差の再現による臨場感向上効果の検証

本実験では，従来の事前生成型 AR [11] と提案する事前生成型 AR を同じ条件下で実行した際に，提示映像にどのような差異があるのかを検証する．実験は屋内環境において，図 14 のように地点 A(0 フレーム目) を初期地点として，地点 B(75 フレーム目), 地点 C(180 フレーム目) と順に移動した際，各システムにおける地点毎の提示映像を比較する．まず，従来の事前生成型 AR [11] においては，図 15 のように，地点 A から地点 B，地点 C へユーザが移動した場合であっても提示映像は回転成分しか考慮されておらず，臨場感のない映像提示となっていることが分かる．一方，提案する事前生成型 AR においては，地点 A から地点 B へユーザが移動した場合，仮想物体と背景の位置関係が変化しており運動視差が再現されていることを確認した．

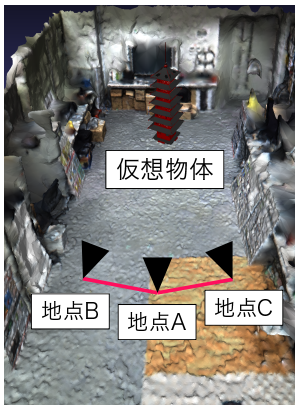


図 14　実験 1：実験環境の俯瞰図

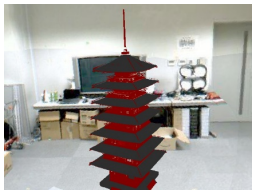
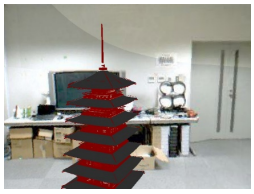
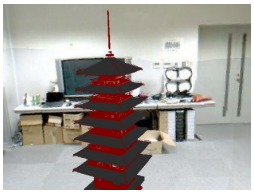
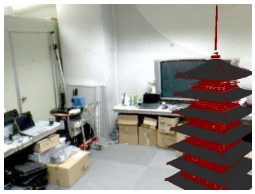
	0フレーム目(地点A)	75フレーム目(地点B)	180フレーム目(地点C)
入力画像			
出力画像(運動視差なし)			
出力画像(運動視差あり)			

図 15　実験 1：運動視差の有無による出力画像の違い

4.2 実験 2：屋外環境における本システムの有効性の検証

4.2.1 実験条件

本実験では，オフラインステージにおいて，全方位マルチカメラシステム Ladybug 3 を用いて屋外環境における 30 か所で全方位画像を撮影した．図 16 に CMPMVS により復元された屋外環境の三次元モデルを示す．また，仮想物体及び使用端末は屋内環境の実験と同じものを利用している．図 12 に示した仮想物体は，図 17 に示すような位置にあらかじめ手動で配置した．

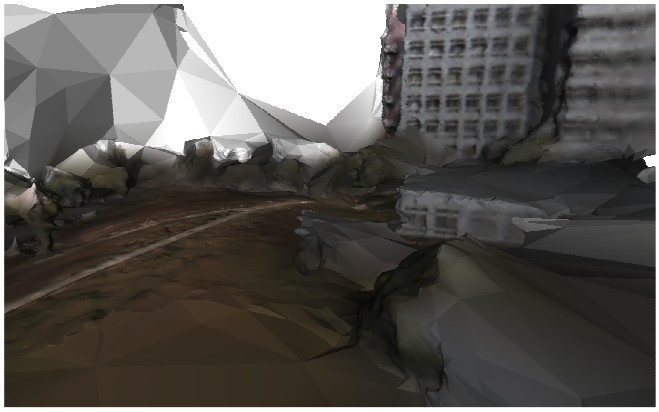


図 16 屋外環境の三次元モデル

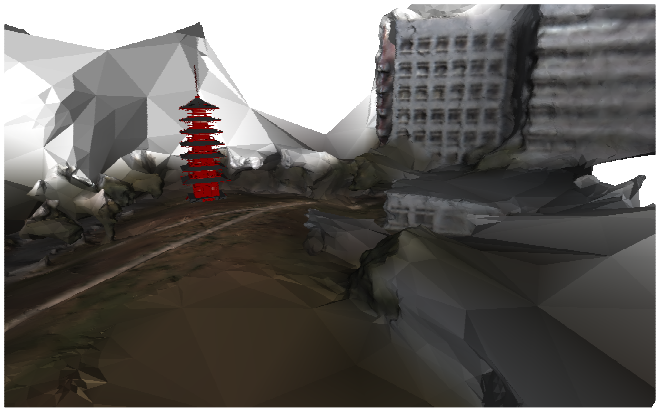


図 17 屋外環境における仮想物体との位置合わせイメージ

4.2.2 カメラ位置補正による蓄積誤差低減効果の確認

図 18 に，処理 (B-2) による補正を行わないカメラ位置姿勢推定結果に基づく出力画像と，処理 (B-2) によるカメラ位置補正を行った場合のカメラ位置姿勢推定結果に基づく出力画像を示す．システム起動直後の初期フレームでは，位置補正処理の有無に関わらず提示映像は同一のものとなる．しかし，図 19 に示すように，230 フレーム経過後において位置補正処理を適用していない場合には，端末カメラの位置が入力画像とは大きく異なってしまっていることが分かる．一方，位置補正処理を適用した場合には，長時間推定を行っても入力画像との多少のずれはあるもののユーザに違和感を感じさせるほどではないことが分かる．以上より，位置補正処理を行うことで蓄積誤差が低減され提示映像の見えを改善できることが確認できた．

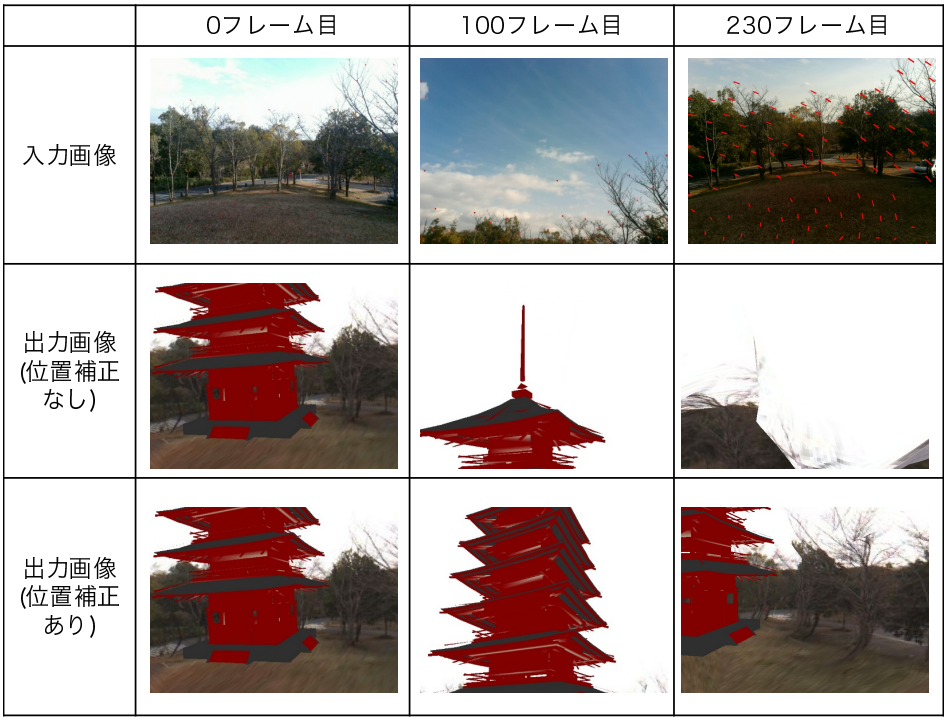


図 18 実験 1: 位置補正処理の有無による出力画像の違い

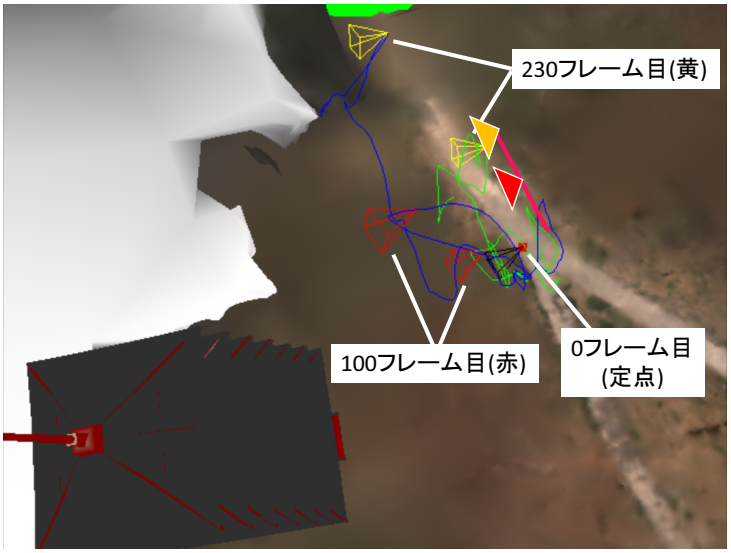


図 19　実験 1：位置補正処理の有無により推定されたカメラパス (赤線：実際のカメラパス，青線：位置補正なし，緑線：位置補正あり)

4.2.3　運動視差の再現による臨場感向上効果の検証

本実験では，従来の事前生成型 AR [11] と提案する事前生成型 AR を同じ条件下で実行した際に，提示映像にどのような差異があるのかを検証する．実験は屋外環境において，図 20 のように地点 A(0 フレーム目) を初期地点として，地点 B(50 フレーム目), 地点 C(200 フレーム目) と順に移動した際，各システムにおける地点毎の提示映像を比較する．まず，従来の事前生成型 AR [11] においては，図 21 のように地点 A から地点 C へユーザが移動した場合，提示映像はその移動を反映したものとなっておらず，景観に対する運動視差が再現されていないことが分かる．一方，提案する事前生成型 AR においては，地点 A から地点 C へユーザが移動した場合において，地点 A では仮想物体の塔に遮蔽されていた背景が地点 C では見えており，移動に伴う運動視差が再現されているため，臨場感のある映像を提示することができている．

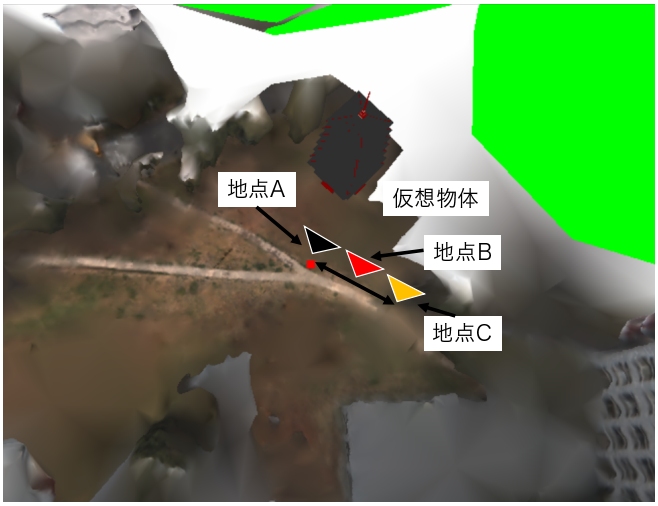


図 20　実験 2：実験環境の俯瞰図

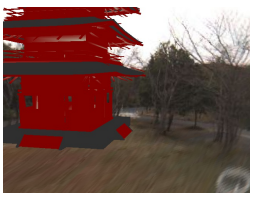
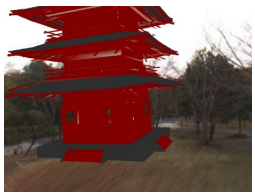
	0フレーム目	50フレーム目	200フレーム目
入力画像			
出力画像 (運動視差なし)			
出力画像 (運動視差あり)			

図 21　実験 2：運動視差の有無による出力画像の違い

5. まとめ

従来の事前生成型 AR システム [11,12] では，モバイル端末を保持するユーザの移動に基づいた運動視差を再現できず，臨場感が損なわれるという問題点が存在した．これに対し，本論文では，端末カメラの相対運動を推定し，その推定結果に基づいた自由視点画像を生成することで運動視差を再現し，かつ現実環境と仮想物体間で位置ずれのない AR 映像を提示する新たな事前生成型 AR システムを提案した．また，端末カメラの相対運動を推定する際に蓄積する誤差の解消手法についても提案した．実験において，誤差の解消法の有効性を検証し，また提案システムと従来の事前生成型 AR の提示映像との比較による提案システムの優位性を検証した．本研究の成果により，応用先として観光向けの AR アプリ―ケーションなどへの適用が考えられる．そのため，今後の展望として，提示映像の品質向上や被験者実験による本手法の有効性を検証する必要があると考えられる．

謝辞

本研究を進めるに当たり，終始暖かくご指導，ご鞭撻頂いた視覚情報メディア研究室 横矢直和 教授に心より感謝申し上げます．また，本研究の遂行に当たり，有益なご助言，ご鞭撻を頂いたインタラクティブメディア研究室 加藤博一 教授に厚く御礼申し上げます．さらに，本研究を進めるに当たり，終始細やかなご指導，ご助言頂いた視覚情報メディア研究室 佐藤智和 准教授に厚く御礼申し上げます．

本研究へのご助言，ご協力を頂いた視覚情報メディア研究室 河合紀彦 助教，中島悠太 助教に深く感謝致します．また，研究室での生活を支えて頂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．あらゆる面において，多大なるご助言を頂いた大阪大学産業科学研究所 大倉史生 助教に深く感謝いたします．

参考文献

[1] A Damala, I Marchal, and P Houlier. Merging augmented reality based features in mobile multimedia museum guides. In *Proc. Int'l Symp. on Anticipating the Future of the Cultural Past (CIPA'07)*, pp. 259–264, 2007.

[2] G A Lee and M Billinghurst. CityViewAR: A mobile outdoor AR application for city visualization. In *Proc. IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12)*, pp. 57–64, 2012.

[3] R Azuma, B Hoff, Howard Neely III, and R Sarfaty. A motion-stabilized outdoor augmented reality system. In *Proc. IEEE Virtual Reality (VR'99)*, pp. 252–259, 1999.

[4] G Klein and D Murray. Parallel tracking and mapping on a camera phone. In *Proc. IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'07)*, pp. 83–86, 2009.

[5] J Engel, T Schöps, and D Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conf. on Computer Vision (ECCV'14)*, pp. 834–849. 2014.

[6] C Forster, M Pizzoli, and D Scaramuzza. SVO: Fast semi-direct monocular visual odometry. In *Proc. IEEE Conf. Robotics and Automation (ICRA'14)*, pp. 15–22, 2014.

[7] H Kato and M Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proc. IEEE and ACM Int'l Workshop on Augmented Reality (IWAR' 99)*, pp. 85–94, 1999.

[8] V Lepetit, L Vacchetti, D Thalmann, and P Fua. Fully automated and stable registration for augmented reality applications. In *Proc. IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'02)*, pp. 03–102, 2002.

[9] 大江統子, 佐藤智和, 横矢直和. 幾何学的位置合わせのための自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定. 日本バーチャルリアリティ学会論文誌, Vol. 10, No. 3, pp. 295–304, 2005.

[10] T Taketomi, T Sato, and N Yokoya. Real-time and accurate extrinsic camera parameter estimation using feature landmark database for augmented reality. *Int'l Journal of Computers and Graphics*, Vol. 35, No. 4, pp. 768–777, 2011.

[11] J Wither, Y-T Tsai, and R Azuma. Indirect Augmented Reality. *Int'l Journal of Computers and Graphics*, Vol. 35, No. 4, pp. 810–822, 2011.

[12] F Okura, T Akaguma, T Sato, and N Yokoya. Indirect augmented reality considering real-world illumination change. In *Proc. IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'14)*, pp. 287–288, 2014.

[13] P Debevec, C Taylor, and J Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *Proc. ACM SIGGRAPH'98*, pp. 11–20, 1996.

[14] C Wu. Towards linear-time incremental structure from motion. In *Proc. Int'l Conf. on 3D Vision*, pp. 127–134, 2013.

[15] M Jancosek and T Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, pp. 3121–3128, 2011, "http://ptak.felk.cvut.cz/sfmservice/websfm.pl?menu=cmpmvs".

[16] Y Furukawa and J Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362–1376, 2010.

[17] J Xiao and M Shah. From images to video: View morphing of three images. In *Proc. Vision, Modeling and Visualization*, pp. 495–502, 2003.

[18] S M Seitz and C R Dyer. View morphing: Uniquely predicting scene appearance from basis images. In *Proc. Image Understanding Workshop*, pp. 881–887, 1997.

[19] E H Adelson and J R Bergen. *The plenoptic function and the elements of early vision.* Computational Models of Visual Processing, MIT Press, Cambridge, 1991.

[20] 高橋桂太, 苗村健. 視点依存奥行きマップの実時間推定に基づく多眼画像からの自由視点画像合成. *The Journal of The Institute of Image Information and Television Engineers*, Vol. 60, No. 10, pp. 1611–1622, 2006.

[21] C Wu. Visualsfm: A visual structure from motion system. 2011.

[22] P Cignoni, M Corsini, and G Ranzuglia. Meshlab: an open-source 3d mesh processing system. *Ercim news*, Vol. 73, No. 45-46, p. 6, 2008.

[23] J Shi and C Tomasi. Good features to track. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593–600, 1994.

[24] B Lucas and T Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int'l Joint Conf. on Artificial Intelligence*, pp. 674–679, 1981.