NAIST-IS-MT1051009

修士論文

自由視点画像生成手法を用いた
移動撮影した全方位動画像からの動物体除去

井上 直哉

2014年3月13日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

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。

井上 直哉

審査委員：

横矢 直和 教授　（主指導教員）
小笠原 司 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）

自由視点画像生成手法を用いた
移動撮影した全方位動画像からの動物体除去*

井上 直哉

内容梗概

遠隔地の実映像をユーザに提示することであたかもその場所にいるような感覚を与えることができる全方位テレプレゼンスシステムは, ナビゲーション, 娯楽, 医療, 教育など様々な分野への応用が期待されており, 近年, Google ストリートビューなどの商用サービスにも応用されている. このような全方位テレプレゼンスシステムにおいては, 全方位動画像中に写りこんだ人のプライバシー問題が生じる. また, 全方位動画像を入力とし, ユーザが撮影経路上で連続的に視点位置を変更可能なテレプレゼンスシステムにおいては, 提示される画像上の動物体の移動が視点移動と連動し, 時間とは連動しないため違和感が生じる. そこで本論文では, 移動撮影された全方位動画像から人などの動物体を除去し, 静止物体のみで構成される全方位動画像を生成する手法を提案する. 従来, 移動撮影された全方位動画像から動物体を除去する手法として, 複数回同一経路を撮影し, それらを統合することで動物体の存在しない全方位動画像を生成する手法が提案されている. この手法では一時的な駐停車中の車など一回の撮影中には動かない動物体に対しても, 別の時間に撮影された動画を用いて除去できるという利点はあるが, 撮影コストが大きいという問題がある. 他方, 動画像の背景に平面仮定をおいて前後のフレーム間の対応付けを行い, 動物体の除去を行う手法が提案されている. しかし, 平面仮定が適用できない一般的なシーンで用いることはできない. これらの問題に対して, 本研究では, 一回の移動撮影で得られた全方位動画像から, 自

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT1051009, 2014 年 3 月 13 日.

由視点画像生成手法を用いて複数フレームの画像を単一視点の画像に変形し, 変形された画像間の整合性を検証することで全方位動画像中の動物体領域を除去する．これにより, 従来研究の問題であった撮影コストの削減及び背景形状制約の緩和を図る．具体的には, まず Structure from Motion と Multi-view Stereo で環境の三次元形状を復元し, 各フレームにおいて密な全方位奥行き画像を生成する．次に, 全方位動画像中のあるフレームを対象フレームに設定し, 生成した全方位奥行き画像に基づきその前後複数のフレームを対象フレームでの見え方に変換する．最後に, 生成された対象フレーム視点の画像群を比較することで動物体の背景を取得する．これを, すべてのフレームを順に対象フレームに設定しながら繰り返す．実験では, 一回の移動撮影で得られた全方位動画像から動物体を除去することで提案手法の有効性を示す．

キーワード

全方位動画像, テレプレゼンス, 自由視点画像生成, 動物体除去

Removal of Moving Objects from Omnidirectional Video Taken by a Moving Camera Using a Novel-viewpoint Image Generation Technique*

Naoya Inoue

Abstract

Omnidirectional telepresence system enables us to experience a remote site, and it is expected to be used in a number of different fields such as navigation, entertainment, medical care and education. One implementation of omnidirectional telepresence is street-view system like Google-Street-View. In such an omnidirectional telepresence system, there exist privacy concerns of captured persons in omnidirectional video. One additional problem is that unexpected motion of moving objects, whose motions are not connected to the time but motion of user's viewpoint, often reduces the feeling of existance. In order to remove these problems, this thesis proposes a method to generate an omnidirectional video which consists of only static objects by removing moving objects such as people from omnidirectional video taken by a moving camera. Conventionally, methods for removing moving objects from multiple video sequences taken for the same route are proposed. Although such methods have an advantage that even temporarily static objects such as parked cars can be removed using video taken at a different time, the cost for taking multiple video sequences is bery high. Other conventional

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1051009, March 13, 2014.

methods remove moving objects by making correspondences between frames by assuming a planar background. However, these methods cannot be used in common scenes in which we cannot assume that the background is approximated by a plane. To solve these problems, this study removes moving objects by warping multiple frames to a single viewpoint with a novel-viewpoint image generation technique, and validating the consistency between the warped images. As a result, it is possible to relax the background shape and reduce the cost of taking videos. Specifically, we first restore the three-dimensional shape of the environment using structure from motion and multi-view stereo techniques and generate an omnidirectional dense depth image in each frame. Frames of around a target frame is then warped to the viewpoint of the target frame. Finally, moving objects are detected and the background is obtained by comparing the warped images. In experiments, the effectiveness of the proposed method is demonstrated by removing moving objects from an omnidirectional image sequence.

Keywords:

Omnidirectional video, Telepresence, Novel-viewpoint image generation, Removal of moving object

目 次

1. はじめに **1**

2. 動画像の欠損修復に関する従来研究および本研究の位置づけと方針 **3**

2.1 動画像の欠損修復に関する従来研究 3

2.1.1 一台のカメラを用いた一回の動画撮影による手法 3

2.1.2 一台のカメラを用いた複数回の動画撮影による手法 6

2.1.3 複数のカメラを用いる手法 8

2.2 本研究の位置付けと方針 10

3. 移動撮影した全方位動画像からの動物体除去 **11**

3.1 全方位動画像における動物体除去手法の概要 11

3.2 全方位カメラの位置・姿勢推定と三次元復元 14

3.3 全方位動画像の各フレームに対応する自由視点画像の生成 18

3.4 エネルギー最小化による動物体の除去 20

4. 実験 **22**

4.1 全方位動画像からの動物体除去実験 22

4.1.1 実験の概要 22

4.1.2 実験結果の比較 27

5. まとめと今後の課題 **36**

謝辞 **37**

参考文献 **38**

図 目 次

1 Google Street View の提示例 2
2 Shen ら [15] の手法による前景動物体除去の例 5
3 Fores ら [20] の手法による歩行者除去の例 6
4 Herling ら [24] の手法の静止物体の除去例 7
5 内山ら [27] の手法による動物体の除去例 7
6 榎本ら [33] の手法による前景物体の除去例 9
7 動物体が映った全方位動画像の例 11
8 移動撮影した全方位動画像から動物体を除去する処理の流れ . . . 12
9 VisualSFM[36] に入力する全方位動画像の例 14
10 図 9 の全方位動画像から作成した cube map 15
11 CMPMVS[37] で生成される環境の三次元モデルの例 17
12 自由視点画像生成例 19
13 全方位マルチカメラシステム (Ladybug3) 23
14 入力注目フレーム画像 D (図 18(a)) における自由視点画像例 (1/3) 24
14 入力注目フレーム画像 D (図 18(a)) における自由視点画像例 (2/3) 25
14 入力注目フレーム画像 D (図 18(a)) における自由視点画像例 (3/3) 26
15 注目フレーム画像 A における自由視点画像群に各手法を適用した結果 (1/2) 28
15 注目フレーム画像 A における自由視点画像群に各手法を適用した結果 (2/2) 29
16 注目フレーム画像 B における自由視点画像群に各手法を適用した結果 (1/2) 30
16 注目フレーム画像 B における自由視点画像群に各手法を適用した結果 (2/2) 31
17 注目フレーム画像 C における自由視点画像群に各手法を適用した結果 (1/2) 32
17 注目フレーム画像 C における自由視点画像群に各手法を適用した結果 (2/2) 33

18　注目フレーム画像 D における自由視点画像群に各手法を適用した結果 (1/2) 34
18　注目フレーム画像 D における自由視点画像群に各手法を適用した結果 (2/2) 35

1. はじめに

全方位動画像は遠隔地のテレプレゼンス [1, 2, 3, 4, 5, 6, 7] や景観のデジタルアーカイブ [8] などで利用されているように, 近年大きく普及が進んできており, そのようなアプリケーションへ全方位動画像を応用するための様々な研究 [9, 10, 11, 12, 13] がなされている. 全方位動画像を利用したアプリケーションの最も有名な例としては, Google Street View (図 1) を挙げることができる. Google Street View は全方位カメラを用いて撮影された市街地の動画像を, 地図上の位置とリンクさせたものであり, インターネットを介して閲覧することができる. 視点位置も撮影位置の範囲内で自由に変えることができ, ユーザが訪れたことがない場所であっても実際にその道を歩いているような感覚を得ることができる. しかし, サービスの広がりと共に人の顔や, 表札, 車のナンバープレートなどが写り込むことによるプライバシーの問題が持ち上がってきた. この問題を受けて, Google は顔やナンバープレートなどにぼかし処理を施すなどの対策を取ってはいるものの, そのために画像の見栄えの悪化が問題になっている. また, 移動撮影で得られた全方位動画像を入力とし, ユーザが撮影経路上で連続的に視点位置を変更可能なテレプレゼンスシステム [3, 4, 5, 6] が提案されているが, そのようなシステムにおいては, 提示される画像上の動物体の移動が視点移動と連動し, 時間とは連動しないため違和感が生じる. これらの問題点に対処する一つの手段として, 動画像中の動物体の除去が挙げられる.

従来, 移動撮影された全方位動画像から動物体を除去する手法として, 同一経路を複数回撮影し, それらを統合することで動物体の存在しない全方位動画像を生成する手法 [27, 29] が提案されている. これらの手法では一時的な駐停車中の車など一回の撮影中には動かない動物体に対しても, 別の時間に撮影された動画を用いて除去できるという利点はあるが, 撮影コストが大きいという問題がある. 他方, 経路の一度の移動撮影で得られた動画像を用いた動物体除去手法 [20, 22] も提案されている. これらの手法では, 背景に平面仮定をおいて前後のフレーム間の対応付けを行い, 動物体の除去を行うため, 同一経路を複数回撮影する手法に比べて撮影コストは小さいが, 平面仮定が適用できない一般的なシーンで用いることはできず, 汎用性が低い.

図 1: Google Street View の提示例

これらの手法の問題点に対して本論文では, 一回の移動撮影で得られた全方位動画像から, 自由視点画像生成手法を用いて複数フレームの画像を単一視点の画像に変形し, 変形された画像間の整合性を検証することで全方位動画像中の動物体領域を除去する手法を提案する. これにより, 従来研究の問題であった撮影コストの削減及び背景形状制約の緩和を図る. 提案手法では, 一回の移動撮影で得られた全方位動画像に対して Structure from Motion と Multi-view Stereo を用いて各フレームのカメラ位置姿勢推定, および環境の三次元形状復元を行い, これらに基いて各フレームにおいて密な全方位奥行き画像を生成する. 次に, それを用いて複数フレームの画像をある注目フレームの視点での見えに変換する. 最後に, 生成された注目フレームの視点の画像群から画素ごとに適切なフレームをエネルギー最小化により選択し画素値をコピーすることで, 明示的に動物体を特定することなく動物体を除去し, 背景画像を生成する.

以下, 2 章では, 動画像の欠損修復に関する従来研究および, 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, 移動撮影した全方位動画像から動物体の除去を行う手法について述べる. 4 章では, 提案手法と従来手法との比較実験を行い, 提案手法の有効性を示す. 最後に, 5 章ではまとめと今後の展望, 課題について述べる.

2. 動画像の欠損修復に関する従来研究および本研究の位置づけと方針

本章では, 動画像から動物体などの特定物体を除去し, その領域を修復する関連研究を概観し, 関連研究に対する本研究の位置付けを述べる.

2.1 動画像の欠損修復に関する従来研究

全方位カメラで撮影されたかどうかにかかわらず, 動画像から動物体などの特定の物体を検出, 除去する研究は盛んに行われている. このような手法は以下のように分類できる.

- 一台のカメラを用いた一回の動画撮影による手法
  - 前後のフレームを対応付け統合する手法
  - 周りのテクスチャから除去対象領域のテクスチャを補完する手法
- 一台のカメラを用いた複数回の動画撮影による手法
- 複数のカメラを用いる手法

以下, 各手法について詳述する.

2.1.1 一台のカメラを用いた一回の動画撮影による手法

一回の動画撮影による手法は, 前後のフレームを対応付け統合する手法と, 周りのテクスチャから除去対象領域のテクスチャを補完する手法に大別できる. 以下, それぞれの手法について詳しく述べる.

[**前後のフレームを対応付け統合する手法**]

動画像の前後のフレームを対応付け統合する手法は, 動画像の前後のフレームにおいて, フレーム間で撮影シーンの同一箇所の対応付けを行い, 複数の画像を統合することで動物体などの特定物の除去を行う. 以下では, 一般的なカメラで撮影された動画像と全方位動画像を対象とした場合に分け, 各手法を紹介する.

一般的なカメラで撮影された動画像に対する手法として, 譲田ら [14] は, 移動撮影したカメラの移動距離が小さく, 動画像におけるシーンがカメラから十分遠方にあるという仮定をおいて動物体を除去している. このような仮定の下で動画像のフレーム間の対応を射影変換であるとみなし, 前後のフレームを射影変換し統合することで, 除去対象である動物体領域を特定することなく, 動物体の存在しない動画像を生成している. また, Shen ら [15], 福地ら [16], 原田ら [17] は, パンチルトズームカメラで撮影された動画像を対象とし, 譲田ら [14] の手法と同様に対象シーンの静的な領域を射影変換で対応付け, それに加えて動的な背景も復元している. Shen ら [15] は, 初期フレームでユーザが動物体の領域を手動で指定し, Mean-Shift を用いて追跡する手法を提案している. この手法では, 除去対象の背景に存在する等速直線移動物体も復元できる (図 2) . 福地ら [16] や原田ら [17] は, 除去対象をカメラのレンズに付着した水滴とし, 一軸を時間, もう一軸を空間の水平軸とする時空間断面画像内で, 被写体とカメラのレンズに映った水滴との間における軌跡の違いを利用することで水滴領域を検出し, 時空間断面画像に対して静止画の欠損修復手法 [18] を適用することで背景情報を補間している. 一方, Matsushita ら [19] の手法は前述した手法とは異なり, 除去する対象を手動で指定する. この手法では, 背景形状やカメラの動きに制約をおかずオプティカルフローを用いて前後フレームを対応付け対象の除去を行うことができるが, 除去対象が文字列など画像上の動かない物体でない場合は, 手動で毎フレーム対象領域を指定する必要がある.

一台の全方位カメラを用いて撮影した全方位動画像を対象とした研究について述べる. Fores ら [20], 町北ら [22] は移動撮影された全方位動画像において, 除去対象の背景が平面であるという仮定をおくことで, 対象の除去を行う手法を提案

(a) 入力画像

(b) 前景動物体除去後の画像

図 2: Shen ら [15] の手法による前景動物体除去の例

している．Fores らの手法では Google Street View における歩行者の背景が平面であると仮定している．入力画像上で Leibe ら [21] の手法によって歩行者を検出し, その隣接したフレームの画像に射影変換を施して検出した歩行者の領域にコピーすることで動物体の存在しない全方位動画像を生成している (図 3) ．町北ら [22] は全方位動画像における全方位カメラの死角領域を除去対象とし, その対象が平面であると仮定することで, その死角領域を他のフレームのテクスチャを用いて補間する手法を提案している．他方, 堀ら [23] は固定された全方位カメラで撮影された全方位動画像を対象とした手法を提案している．この手法では, 輝度値の出現頻度を画素ごとに参照することで, 動物体の除去を行っている．しかし, この手法は固定カメラを用いることを前提とし, カメラの運動は考慮しないため, 移動カメラ画像にそのまま適用することは難しい.

(a) 入力画像　　(b) 歩行者除去後の画像

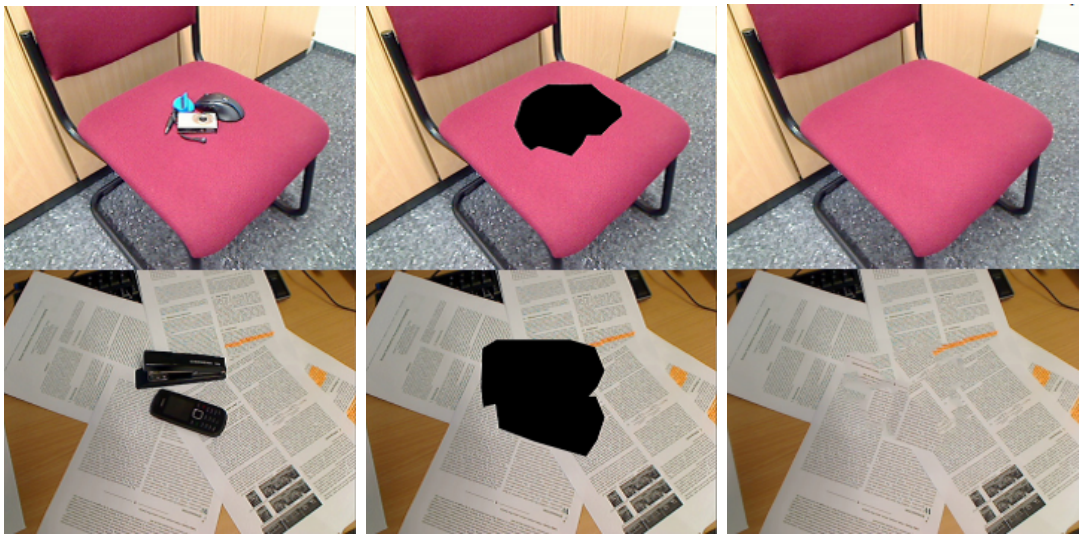
図 3: Fores ら [20] の手法による歩行者除去の例

[**周りのテクスチャから除去対象領域のテクスチャを補完する手法**]

実際の背景を観測せず周りのテクスチャから除去対象領域を補完する手法が提案されている．Herling ら [24] は移動カメラで撮影した動画像における除去対象に対して, 画像修復手法 [25] と類似パターンの探索手法 [26] を組み合わせることで, フレーム毎にテクスチャを生成する手法を提案している（図 4）．この研究は静止物体の実時間除去を目的としており, 動物体には対応していない.

2.1.2 一台のカメラを用いた複数回の動画撮影による手法

同一経路を複数回撮影する手法としては, 内山ら [27], 高橋ら [29] の手法が挙げられる．これらの手法は, 別々の時間に撮影された複数の動画像間の位置合わせを行い, それらを統合するという点で双方とも共通している．内山ら [27] の手法の例を図 5 に示す．この手法では, まず同一経路を複数回走行して得られた全方位動画像群の映像間で撮影位置が近い画像群を選択し, 非剛体レジストレーションにより画素を対応付ける．次に, 補正した同一位置での画像群にサブウィンドウ単位で, エネルギーを最小化するように部分画像を選択し, 統合することで動物体の存在しない全方位動画像を生成している．そのエネルギー関数は, ベクト



(a) 入力画像　(b) 除去対象指定後の画像　(c) 除去対象除去後の画像

図 4: Herling ら [24] の手法の静止物体の除去例

(a) 入力画像

(b) 動物体除去後の画像

図 5: 内山ら [27] の手法による動物体の除去例

ルメディアンフィルタ [28] に基づく部分画像に動物体が存在するかどうかの尤もらしさを表す項と, 隣接する部分画像間の連続性を考慮した項で構成されている. 高橋ら [29] は, 車載全方位カメラで複数の時間に密に撮影された入力画像群に対し, 線形濃度変換パラメータの推定処理と動物体候補領域の推定処理の二つの処理を交互に繰り返し行なうことで, 動物体除去を行いながらシーンの色調統一を行う手法を提案している. これらの手法は一時的な駐停車中の車など一回の撮影中には動かない動物体に対しても, 別の時間に撮影された動画を用いて除去できるという利点はあるものの, 撮影コストが大きい.

2.1.3 複数のカメラを用いる手法

複数のカメラを用いる手法は, 除去対象物体の映っているメインカメラの動画像とその対象物体の背景を撮影した他の位置にあるカメラ (隠背景撮影用カメラ) の動画像を, 特徴点や AR タグなどで対応付けて統合することで, 動物体などの除去対象領域の検出・除去を行う. これらのカメラに加えてデプスカメラを用いる手法もあり, デプスカメラを用いることで除去対象領域の検出精度を高めることができる. 以下, 各手法について詳述する.

Zokai ら [30], 橋本ら [31] は隠背景撮影用カメラとメインカメラを明確に分けて除去対象の除去を行っている. Zokai ら [30] は固定されたメインカメラと隠背景撮影用カメラを用意し, 手動で指定した対象を除去する手法を提案している. この手法では, 立体的な隠背景を複数の平面の集合と仮定することで, 3 次元的な背景に対応している. 橋本ら [31] は事前にキャリブレーションされた複数台の固定カメラを用いることで野球の試合映像における審判側のカメラから審判とキャッチャーを消去し, ピッチャーを表示する手法を提案している. この手法では背景が平面であるという仮定のもと, 除去対象である動物体領域を Garett ら [32] のグラフカットアルゴリズムおよび背景差分を用いて特定し, 隠背景カメラで得られた映像をメインカメラに対して射影変換することで動物体領域の補間を行っている. 他方, 榎本ら [33] と本田ら [34] は複数のカメラを用意し, それぞれがメインカメラと隠背景撮影用カメラの役割を兼任し, 前景の対象物体を除去する手法を提案している. これらの手法も, 除去対象の背景に対する平面仮定を用いており, 榎

(a) 入力画像　　(b) 前景物体除去後の画像

図 6: 榎本ら [33] の手法による前景物体の除去例

本ら [33] は図 6 のように AR タグを用いて, 本田ら [34] はカメラ間のシーンの特徴点を BRISK を用いて, 複数動画像を対応付けている. この特徴点に応じて動画像を射影変換し, それらの動画像群のメディアンを取ることで除去対象の除去を行っている. 清水ら [35] は, 橋本ら [31] の手法における除去対象である動物体領域の指定を, デプスカメラを用いてさらに高精度に行う手法を提案している. これらの手法では, 基本的に複数のカメラを用いることで物体の隠背景が撮影されているという前提が必要であり, 広域な屋外環境の移動撮影で得られた全方位動画像からの動物体除去を対象とした場合, カメラの配置位置が難しいという問題がある.

2.2 本研究の位置付けと方針

前節までに概観したように, 全方位動画像を様々なアプリケーションに利用する上でプライバシーなどの問題点を解決するために, 人などの動物体を動画像上から除去することが要求されており, それを実現するための手法がすでに多く提案されている. 全方位動画像から動物体を除去する従来研究の中で, 動画像を複数回撮影する手法は一時的な駐停車中の車など一回の撮影中には動かない動物体を, 別の時間に撮影された動画を用いて除去できるという利点はあるものの, 撮影コストが大きい. 一回の撮影による手法は同一経路を複数回撮影する手法に比べて撮影コストは小さいが, 多くの手法で平面仮定を用いたり, カメラ移動を制限しており, 汎用性が低い. 平面仮定やカメラの移動制限を用いない場合には動物体領域を手動で指定する必要がある. このように従来手法では汎用性や撮影コストはトレードオフの関係として成り立っていた.

これに対し, 本論文では, 一回の移動撮影で得られた全方位動画像から, 環境の三次元形状を復元した上で自由視点画像生成手法を用いて複数フレームの画像を単一視点の画像に変形し, 変形された画像間の整合性を検証することで全方位動画像中の動物体領域を除去する手法を提案する. これにより, 従来研究の問題であった撮影コストの削減および背景形状制約の緩和を同時に達成する. ただし, 動物体の背景が動画像中のいずれかのフレーム内に映っていることを前提とする.

3. 移動撮影した全方位動画像からの動物体除去

本章では, 移動撮影した全方位動画像から動物体を除去する提案手法の具体的な内容について説明する.

3.1 全方位動画像における動物体除去手法の概要

全方位カメラで移動撮影した動画像には, 一般的に図 7 のように人などの動物体が写り込んでいる. 本論文では, 自由視点画像生成手法を用いてこのような全方位動画像から動物体を除去する手法を提案する. 提案手法の流れを図 8 に示す. まず, 一回の移動撮影で得られた全方位動画像に対し, Structure from Motion と Multi-view Stereo を用いて各フレームのカメラ位置推定, および環境の三次元復元を行う (a). (a) に基づいて全方位動画像の各フレームにおいて密な全方位奥行きを生成する (b). 全方位動画像中のあるフレームを注目フレームに設定し, 生成した全方位奥行き画像に基づいてその前後複数フレームの画像を注目フレームの視点での見えに変換する (c). 生成された注目フレームの視点の画像群から, グラ

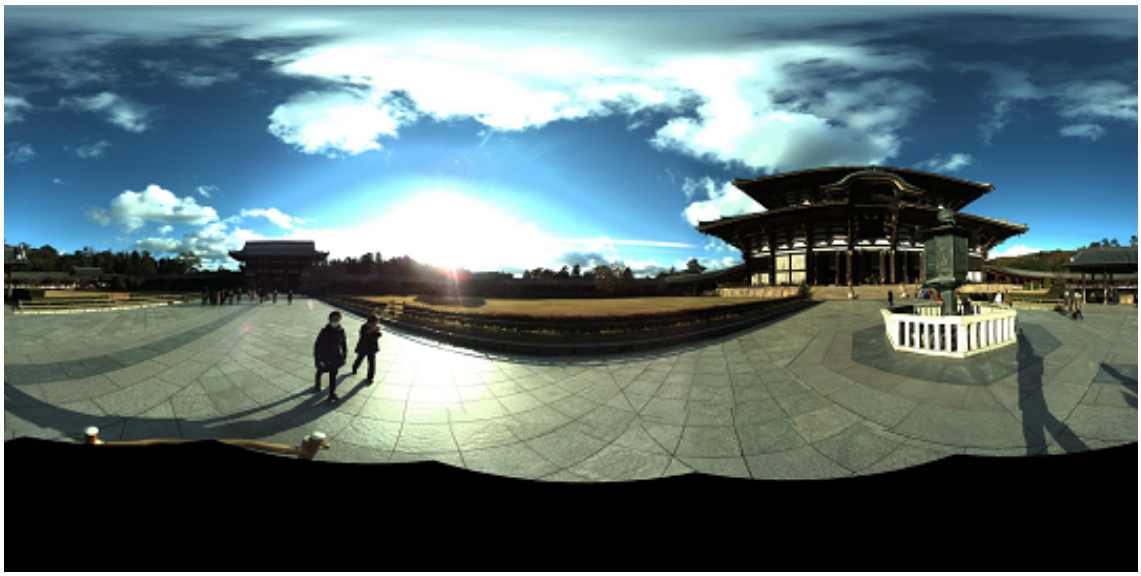


図 7: 動物体が映った全方位動画像の例

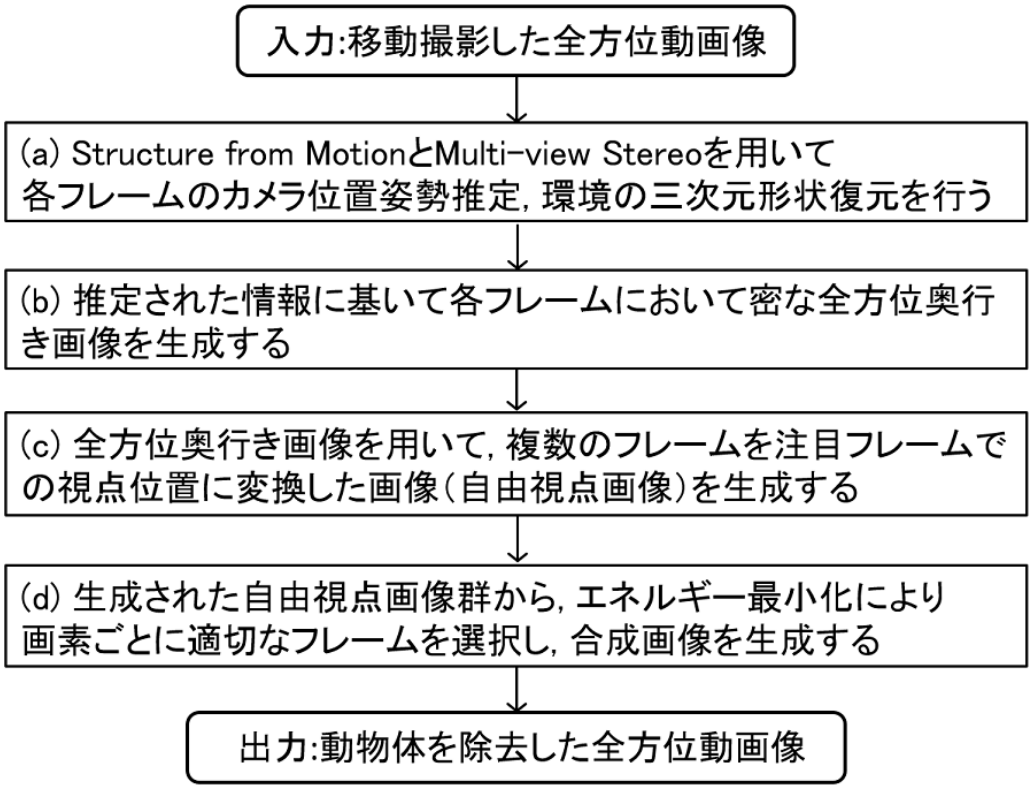


図 8: 移動撮影した全方位動画像から動物体を除去する処理の流れ

フカットを用いたエネルギー最小化により画素ごとに適切なフレームを選択し, 画素値をコピーすることで, 動物体が除去された合成画像を生成する (d).

以下, 3.2 節では全方位カメラの位置・姿勢推定と三次元復元について述べ, 3.3 節で全方位動画像の各フレームに対応する自由視点画像の生成について述べる. また, 3.4 節でエネルギー最小化による動物体の除去について述べる.

3.2 全方位カメラの位置・姿勢推定と三次元復元

本研究では, 全方位カメラの位置姿勢推定を Structure from Motion 法である VisualSFM[36], 環境の三次元復元を Multi-view Stereo 法である CMPMVS[37] を用いて行う. VisualSFM[36] では, 自然特徴点を全方位動画像の各フレーム間で対応付けることによって, 撮影時の全方位カメラのパラメータおよび自然特徴点の三次元位置を推定する. ここで得られたカメラパラメータと全方位動画像群を CMPMVS[37] に入力することで, 環境の三次元モデルを生成する.

VisualSFM[36] への入力に使用する全方位動画像の例を図 9 に示す. このようなパノラマ画像では, 特に画像の上部および下部でテクスチャが大きく歪んでいるため, このまま VisualSFM[36] の入力として用いると, 自然特徴点の対応付けが適切にできない場合がある. このため, 図 10 に示すような cube map を作成し, 各画像を VisualSFM[36] に入力することで撮影時のカメラパラメータおよび自然特徴点を推定する.

図 9: VisualSFM[36] に入力する全方位動画像の例

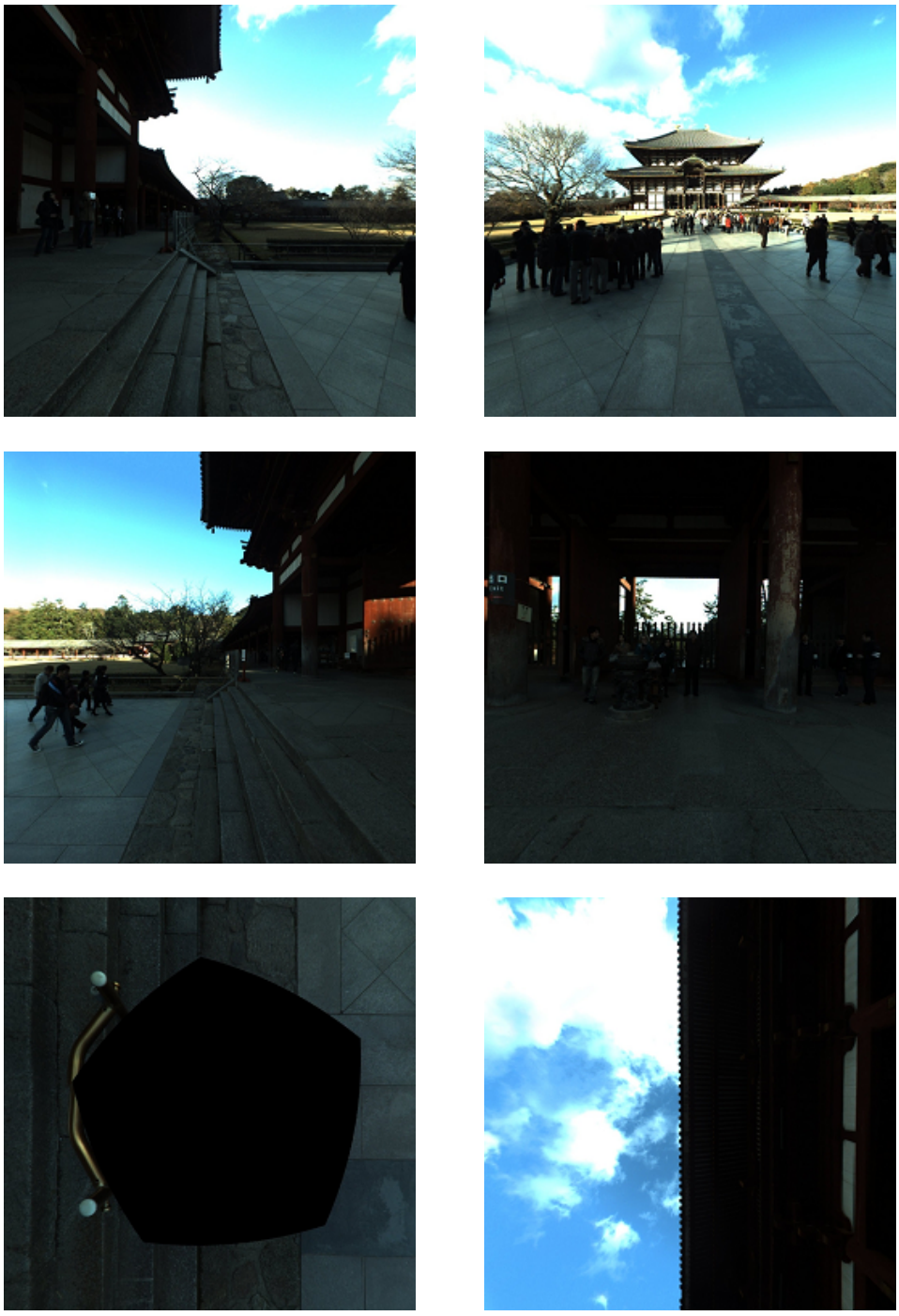


図 10: 図 9 の全方位動画像から作成した cube map

次に, VisualSFM[36] で得られたカメラパラメータと入力全方位動画像の cube-map 群を CVPMVS[37] に入力することで環境の三次元モデルが生成される．図 11 が生成される環境の三次元モデルの例である．

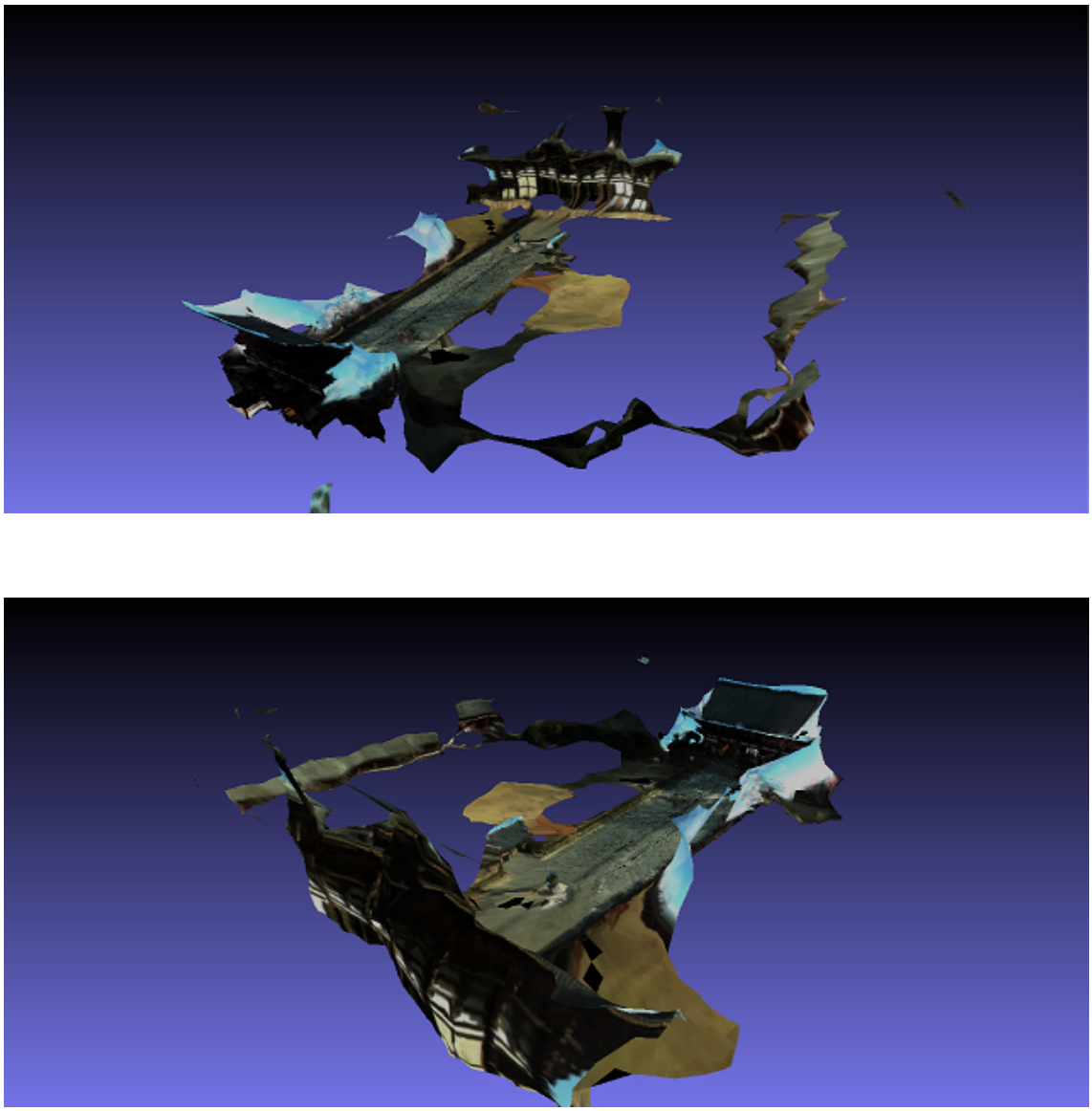
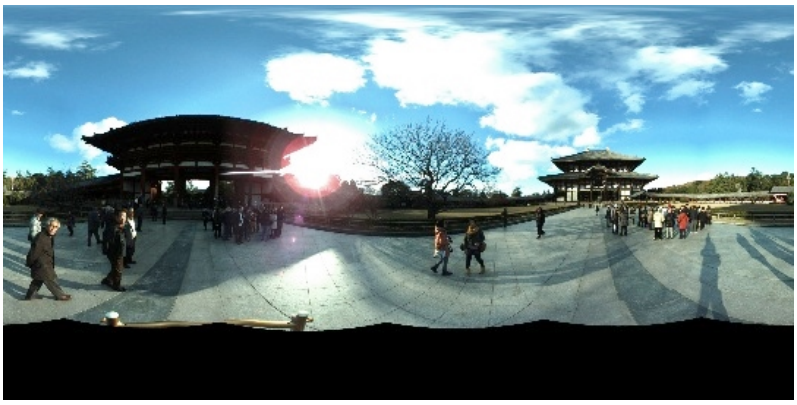


図 11: CMPMVS[37] で生成される環境の三次元モデルの例

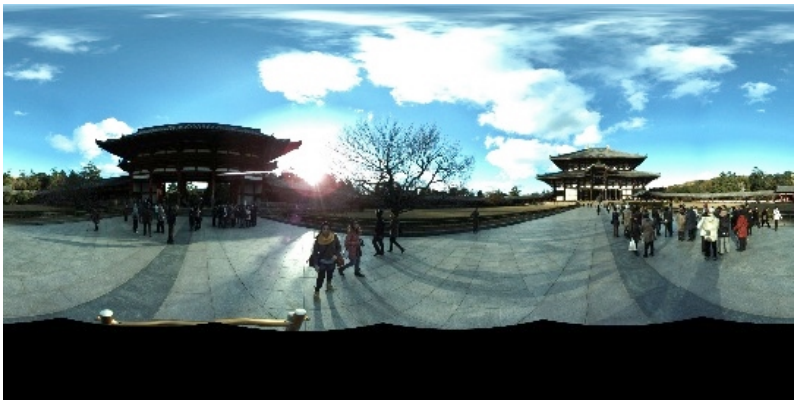
3.3 全方位動画像の各フレームに対応する自由視点画像の生成

本節では, 全方位動画像の各フレームに対応する自由視点画像を生成する方法について説明する. まず, 全方位カメラの位置・姿勢と環境の三次元形状に基づき全方位動画像の各フレームにおいて密な全方位奥行き画像を生成する. ただし, 図 11 の空の領域のような三次元形状が生成されない場所に対応する画素では, 奥行きが無限遠にあるとする. 次に, 注目フレームに対して生成された全方位奥行き画像と他フレームのカメラ位置・姿勢に基づき, 注目フレームの各画素を他フレームに投影することで, 対応付けを行い, 他フレームの画素値をコピーする. それを注目フレーム前後の複数フレームに対して個々に行うことで, 注目フレーム視点での見えを再現した画像群を生成する.

図 12(a) が注目フレーム画像の例, 図 12(b) は注目フレームの近くの他フレーム画像の例である. ここで, 他フレーム画像を注目フレーム視点画像に変換することで, 図 12(c) のような自由視点画像が生成される.



(a) 注目フレーム画像



(b) 他フレーム画像

(c) 他フレーム画像 (b) を注目フレーム (a) での視点に変換した画像

図 12: 自由視点画像生成例

3.4 エネルギー最小化による動物体の除去

本節では, 生成された注目フレームの視点の画像群から, グラフカットを用いたエネルギー最小化により画素ごとに適切なフレームを選択し, 画素値をコピーする方法について説明する.

エネルギー関数 E は以下のように定義する.

$$E = \sum_{p \in A} E_1(f_p) + \lambda \sum_{(p,q) \in B} E_2(f_p, f_q) \tag{1}$$

ここで, f_p, f_q は合成画像上における画素 p および q の画素として用いるフレーム番号, A は合成画像内の画素の集合, B は合成画像内の隣り合う画素ペアの集合, λ は重みである.

エネルギー関数 E の第一項 E_1 は, 選択されるフレーム上の画素が動物体上になく, かつ注目フレームに近いフレームが選択されるよう以下のように定義する.

$$E_1(f_p) = \varsigma(|f_p - t|) + \alpha \sum_{g \in G_{f_p}} SSD(f_p, g_p) \tag{2}$$

t は注目フレーム番号, α は重み, G_{f_p} はフレーム f_p と他フレーム間における画素 p を中心とする一定範囲の画素の相違度 SSD(Sum of Squared Difference) の値が下位半分のフレームの集合である. $\varsigma(\cdot)$ はシグモイド関数であり, 以下の式で定義される.

$$\varsigma(x) = \frac{1}{1 + e^{-a(x - d_x)}} \tag{3}$$

ただし, a, d_x はシグモイド関数の形を決定するパラメータである. 注目フレームから一定以上離れたフレームの自由視点画像は解像度が低くなる傾向があるため, シグモイド関数を用いることで注目フレームから一定以上離れた自由視点画像のエネルギーは大きくなり, 選ばれにくくなる.

エネルギー関数 E の第二項 E_2 は, 画素間の繋ぎ目が目立たないようにフレームが選択されるよう以下のように定義する.

$$E_2(f_p, f_q) \quad = \quad \delta(f_p, f_q)\varsigma(\|\mathbf{I}_p(f_p) - \mathbf{I}_q(f_q)\|^2) \tag{4}$$

ただし, $\delta(f_p, f_q)$ は $f_p = f_q$ のとき 0, $f_p \neq f_q$ のとき 1 となる関数である. $\mathbf{I}_p(f_p)$, $\mathbf{I}_q(f_q)$ はそれぞれフレーム f_p, f_q における画素 p および q の画素値ベクトル (RGB 色空間) である. この項では, 隣接した画素が同一フレームから選ばれる場合は, 重みが 0 となり, エネルギーが小さくなる. 隣接した画素が別のフレームから選ばれる場合は, 画素間の画素値の差が小さいほどエネルギーが小さくなる.

単一フレーム視点画像群を入力とし, 定義したエネルギーを $\alpha - \beta$ 交換によるグラフカットを用いて最小化することで, 画素ごとにフレームを選択し, 画素値をコピーすることで, 動物体が除去された合成画像を生成する.

4. 実験

本章では, 提案手法の有効性を示すために, 屋外環境で撮影された人などの動物体が写り込んだ全方位動画像を用いて, 提案手法と一般的な手法により動物体を除去し, その結果を比較する.

4.1 全方位動画像からの動物体除去実験

4.1.1 実験の概要

入力として用いる全方位動画像の注目フレームの前後 10 フレーム, 計 20 フレームをそれぞれ注目フレーム視点画像に変換し, 注目フレーム画像を含む 21 枚の画像群を用いて動物体を除去する. これらの画像群に提案手法を含む 3 手法を適用し, 画素ごとに合成に用いるフレームを選択することで合成画像を生成し, 比較を行う. 提案手法と比較する手法は以下の通りである.

- 手法 1
  画素のフレーム選択に画素値のメディアンを用いる.

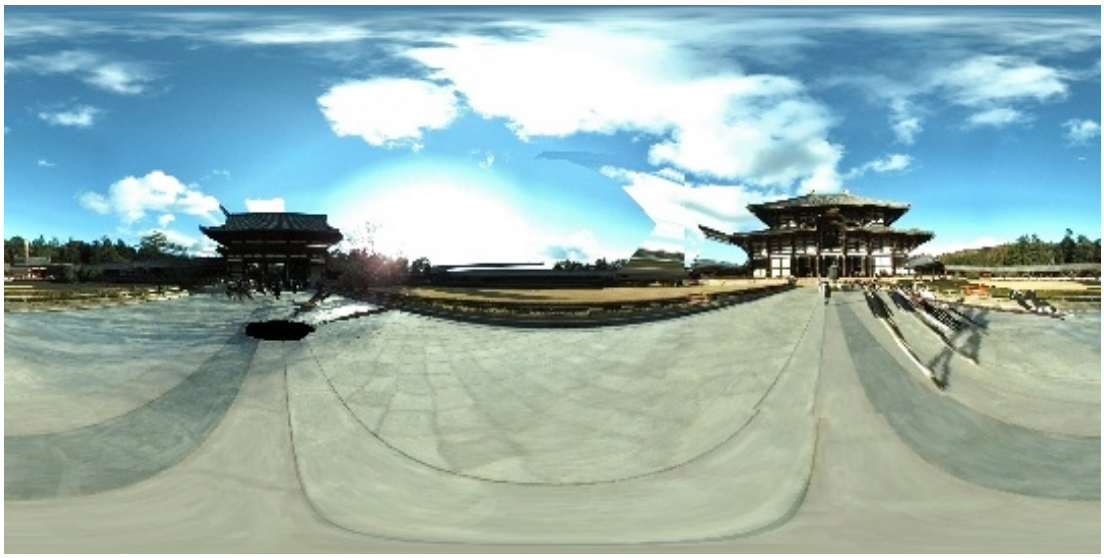
- 手法 2
  画素のフレーム選択時にグレースケールの画素値を用いた Mean-Shift クラスタリング [38] を行い, 最も頻度の大きいクラスタの元の画素値にメディアンを適用する. なお, Mean-Shift では, 平均を算出するための画素値の差の閾値を 30 とする.

提案手法で用いるエネルギー関数 E のパラメータは以下のように設定する. 第一項 E_1 で用いるシグモイド関数は a を 1, d_x を 4, 重み α は 50, SSD のウィンドウサイズは 15×15（画素）とした. 第二項 E_2 への重み λ は 0.25, E_2 で用いるシグモイド関数は a を 0.006, d_x を 300 と設定した.

図 13: 全方位マルチカメラシステム (Ladybug3)

本実験では, 全方位マルチカメラシステム Ladybug3 (図 13) を用いて直線的な移動により撮影された全方位動画像を入力とし, その解像度を 540 × 270 にリサイズして用いた. 入力動画像のうち, 40 フレーム目 (図 18(a)) を注目フレームとした場合において, その前後のフレームを注目フレームでの視点に変換した画像例を図 14 に示す. 図 14(a), (c), (e) が前後フレームの入力画像で, それぞれの画像を注目フレーム視点に変換した画像が図 14(b), (d), (f) である. いずれのフレームにおいても, 動物体領域以外は大きな歪みが生じることなく変換されていることが確認できる. また, 図 14(b), (d), (f) 上において, 入力画像を撮影したカメラ位置付近の地面に黒い円の領域が存在し, これは図 14(a), (c), (e) の下部の黒い領域に対応している. これらの領域は全方位カメラの死角領域のため, テクスチャが取得できていない. 本研究では, このような死角領域も他フレームで対応箇所が撮影されているため, 動物体と同様の扱いとなる.

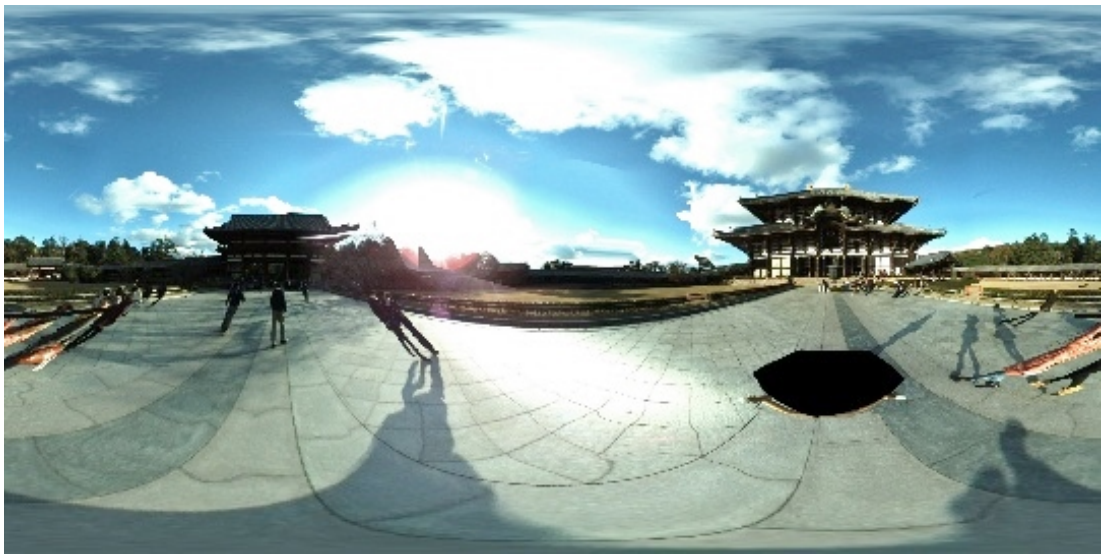
(a) 注目フレーム画像 D (図 18(a)) の 6 フレーム前の画像



(b) (a) を注目フレーム画像 D の視点に変換した画像

図 14: 入力注目フレーム画像 D (図 18(a)) における自由視点画像例 (1/3)

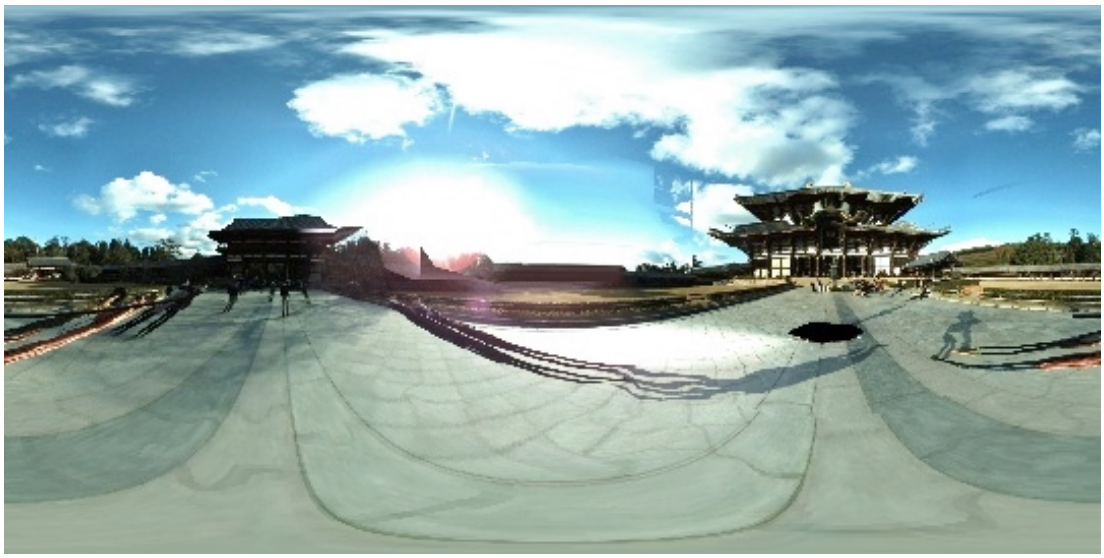
(c) 注目フレーム画像 D (図 18(a)) の 3 フレーム後の画像



(d) (c) を注目フレーム画像 D の視点に変換した画像

図 14: 入力注目フレーム画像 D (図 18(a)) における自由視点画像例 (2/3)

(e) 注目フレーム画像 D (図 18(a)) の 6 フレーム後の画像



(f) (e) を注目フレーム画像 D の視点に変換した画像

図 14: 入力注目フレーム画像 D (図 18(a)) における自由視点画像例 (3/3)

4.1.2 実験結果の比較

本節では, 動画像のうち 4 フレームに対する提案手法と比較手法の結果を示し, 比較を行う. 入力として用いた各全方位画像群に対する実験結果を図 15〜図 18 に示す. それぞれ, 14 フレーム目, 20 フレーム目, 30 フレーム目, 40 フレーム目であり, 以下では画像 A, 画像 B, 画像 C, 画像 D とする. まず, 画像 A (図 15(a)) の自由視点画像群に各手法を適用した結果を比較する. このフレームではどの手法においても動物体を除去しきれておらず, 中央に動物体のテクスチャが残っている. この理由として, 人などの動物体が密に集まり, かつ一時的に静止している場合, 注目フレームの前後のフレームにおいても動物体の背景部分のテクスチャを得ることができないためである. 次に画像 B (図 16(a)) , 画像 C (図 17(a)) , 画像 D (図 18(a)) の各自由視点画像群に各手法を適用した結果を比較する. 各図 (b) に画像 A, B, C の自由視点画像群に対して手法 1 を適用した結果を示す. これらは全体的にぼけた画像になっている. この理由として, 手法 1 では画素値のみを基準として用いているため, 選択される画素値が注目フレームから遠いフレームの解像度の低い自由視点画像から選択されることが多くなるためである. 各図 (c) に, 画像 A, B, C の自由視点画像群に対して手法 2 を適用した結果を示す. これらの結果においても, 手法 1 の場合と同様に画像のぼけが目立つ. また, この結果では手法 1 に比べて動物体のテクスチャが多く残ってしまっている. この理由は, 手法 1 では, 動物体領域が自由視点画像群の過半数に存在していれば動物体上の画素が選ばれるが, 手法 2 は最頻のクラスタが動物体の画素値付近であれば, 動物体領域が自由視点画像群の過半数に存在していなくても動物体上の画素が選ばれてしまうためである. 各図 (d) に画像 A, B, C の自由視点画像群に対して提案手法を適用した結果を示す. これらの結果においては, 手法 2 のような動物体の残存が見られず, かつ手法 1 よりも高精細に動物体除去画像が生成されていることが確認できる. ただし, 図 17(d) では一部で画素間における画素値の不連続が見られた. エネルギー関数の設計上, 隣接した画素は同じフレームまたは類似した画素値のフレームが選ばれやすいが, これだけでは不十分な場合も確認した.

(a) 注目フレーム画像 A（入力全方位動画像の 14 フレーム目）

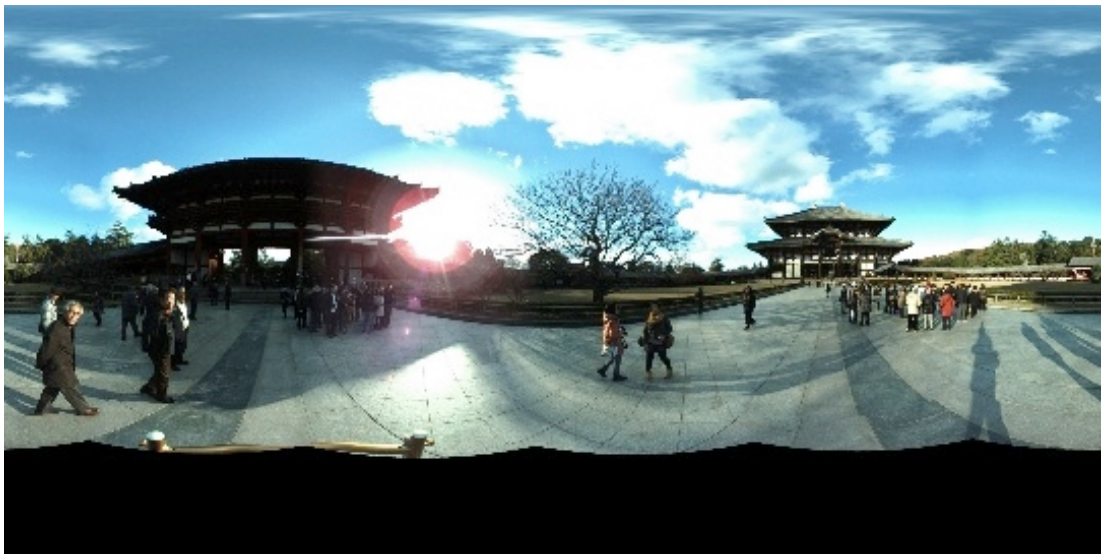
(b) 手法 1 を適用した結果

図 15: 注目フレーム画像 A における自由視点画像群に各手法を適用した結果 (1/2)

(c) 手法 2 を適用した結果

(d) 提案手法を適用した結果

図 15: 注目フレーム画像 A における自由視点画像群に各手法を適用した結果 (2/2)



(a) 注目フレーム画像 B（入力全方位動画像の 20 フレーム目）

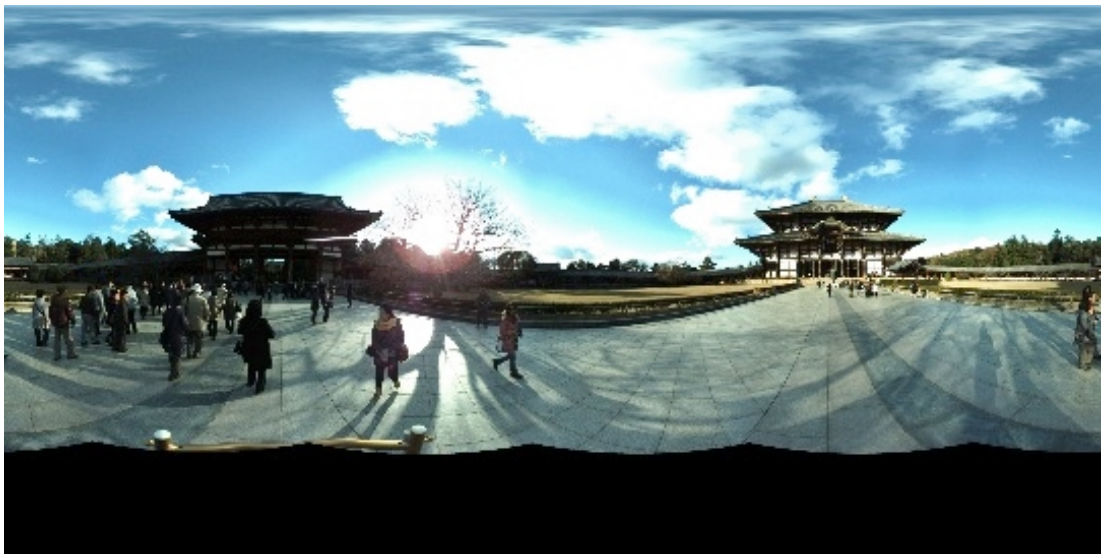
(b) 手法 1 を適用した結果

図 16: 注目フレーム画像 B における自由視点画像群に各手法を適用した結果 (1/2)

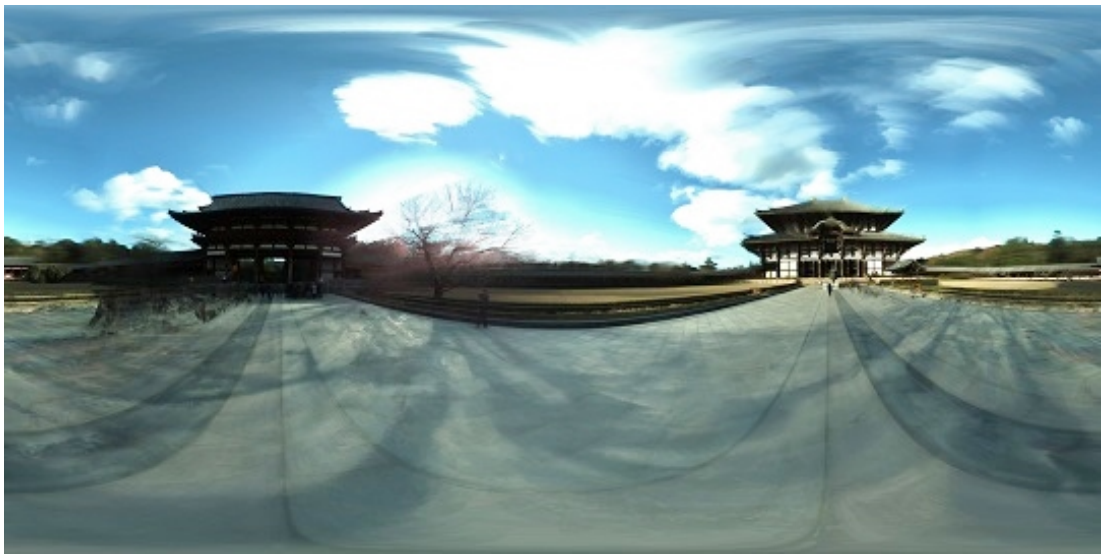
(c) 手法 2 を適用した結果

(d) 提案手法を適用した結果

図 16: 注目フレーム画像 B における自由視点画像群に各手法を適用した結果 (2/2)

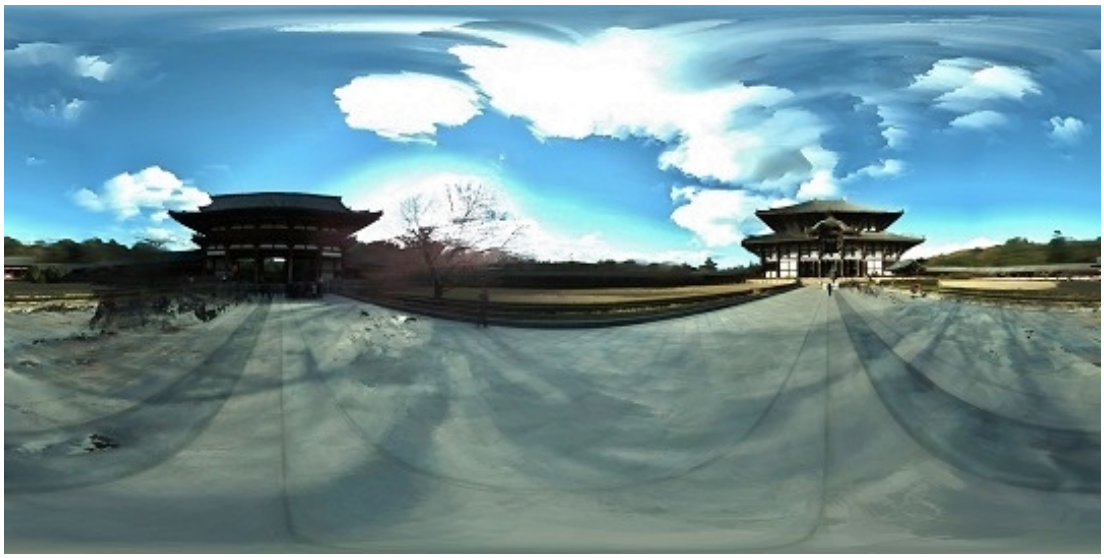


(a) 注目フレーム画像 C（入力全方位動画像の 30 フレーム目）

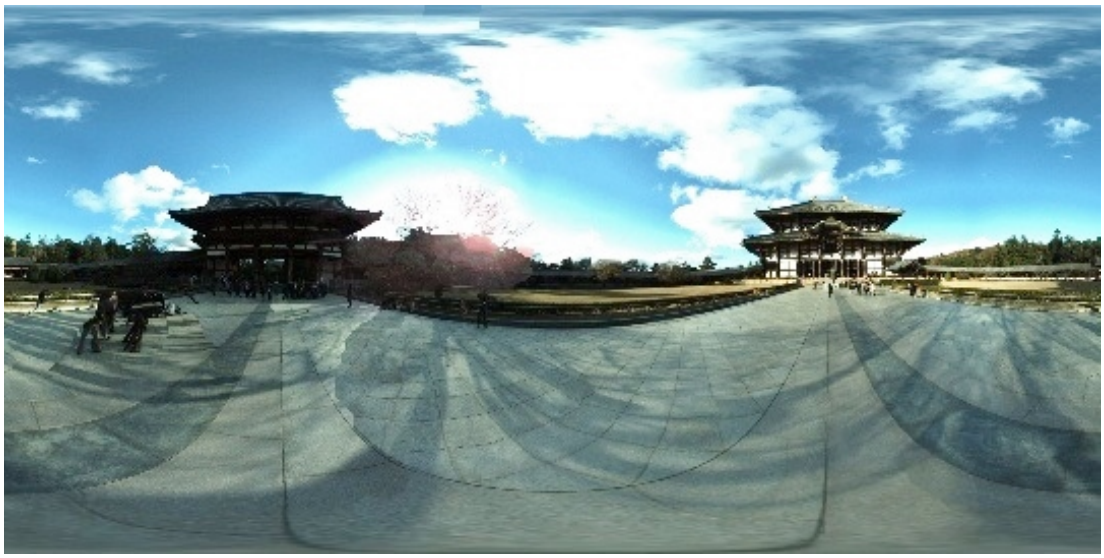


(b) 手法 1 を適用した結果

図 17: 注目フレーム画像 C における自由視点画像群に各手法を適用した結果 (1/2)



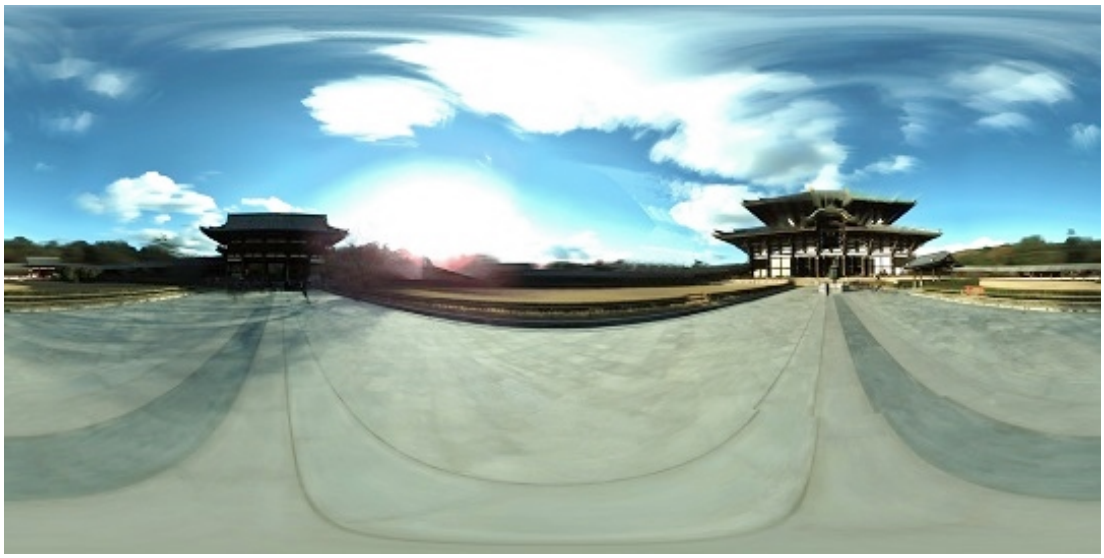
(c) 手法 2 を適用した結果



(d) 提案手法を適用した結果

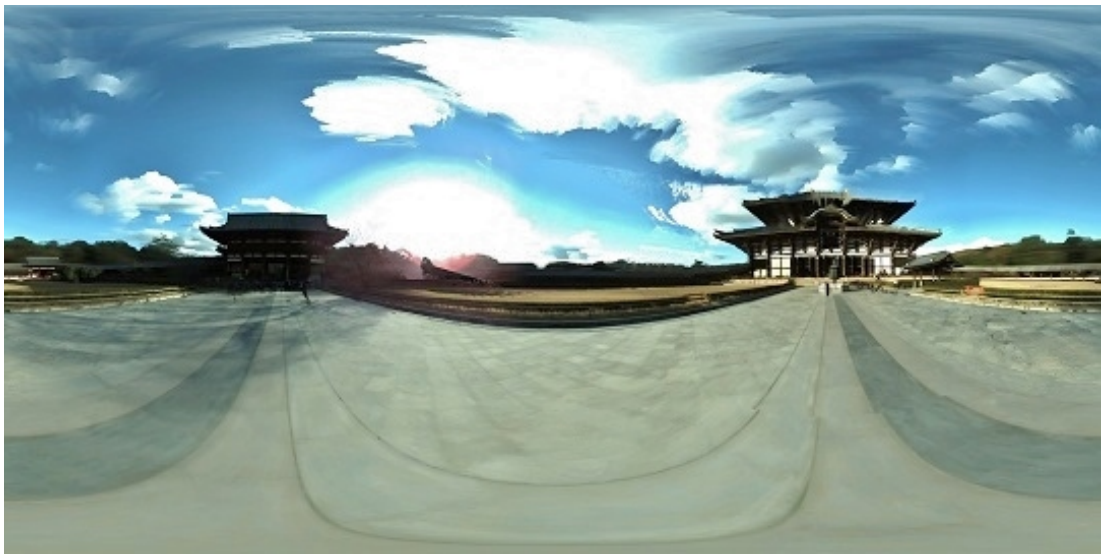
図 17: 注目フレーム画像 C における自由視点画像群に各手法を適用した結果 (2/2)

(a) 注目フレーム画像 D（入力全方位動画像の 40 フレーム目）



(b) 手法 1 を適用した結果

図 18: 注目フレーム画像 D における自由視点画像群に各手法を適用した結果 (1/2)



(c) 手法 2 を適用した結果

(d) 提案手法を適用した結果

図 18: 注目フレーム画像 D における自由視点画像群に各手法を適用した結果 (2/2)

5. まとめと今後の課題

本論文では, 移動撮影された全方位動画像を様々なアプリケーションに応用する上で, 人の写り込みによるプライバシーの問題などを解決することを目的とし, 一回の移動撮影で得られた全方位動画像から人などの動物体を除去し, 静止物体のみで構成される全方位動画像を生成する手法を提案した. 提案手法では, 一回の移動撮影で得られた全方位動画像に対して Structure from Motion と Multi-view Stereo を用いて各フレームのカメラ位置・姿勢推定および環境の三次元形状復元を行い, これらに基づき各フレームにおいて密な全方位奥行き画像を生成した. 次に, それらを用いて複数フレームの画像をある注目フレームの視点での見えに変換した. 最後に, 生成された注目フレームの視点の画像群から, グラフカットを用いたエネルギー最小化により画素ごとに適切なフレームを選択し, 選択されたフレームから画素値をコピーすることで動物体が除去された背景画像を生成した. これにより, 動物体の背景が注目フレームの前後のフレームで取得できていることを前提に, 従来手法で問題となっていた撮影コストの大きさや, 動物体の背景の形状の制約を緩和した上で動物体の除去を行えることを確認した.

実験では, 一般的な手法と提案手法の除去結果の比較を行い, 提案手法では一般的な手法に比べて高精細な合成画像を生成できることを確認した. ただし, 人が密に集まり, かつ一時的に静止している場面では他の手法と同様に不十分な動物体除去結果となった. また, 画素間の画素値の不連続が見られる場合も確認した.

今後の課題としては, 人などが一時的に静止している場面で除去しきれなかった動物体の領域を特定し, 特定した領域に対して画像修復手法 [24] などを用いて背景のテクスチャを生成することで, 不要な動物体のない全方位動画像を生成することが考えられる. また, テクスチャの明度を調整し, 画素値の不連続を抑制する必要がある.

謝辞

本研究を進めるにあたり, その全過程において懇切なる御指導, 御鞭撻を賜りまし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. 本研究の遂行にあたり, 有益な御助言, 御鞭撻を頂いたロボティクス研究室の 小笠原 司 教授に厚く御礼申し上げます, 本研究の全過程を通して, 終始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. そして, 本研究を行うにあたり, 多大なる御助言, 御鞭撻を賜りました視覚情報メディア研究室 河合 紀彦 助教に心より感謝致します. また, 本研究の遂行に適切な御助言を頂きました視覚情報メディア研究室 中島 悠太 助教に心より感謝いたします. 特に 河合 紀彦 助教には, 本論文の執筆, その他の論文の添削に至るまで細やかな御指導を頂きました. 本研究に関する貴重な御助言や御指摘を頂きました視覚情報メディア研究室 大倉 史生 氏に深く感謝いたします. 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. 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さまに深く感謝致します. 最後に, 両親をはじめ, 私の大学院生活に関わった全ての方々に感謝の意を表します.

参考文献

[1] H. Kawasaki, M. Murao, K. Ikeuchi and M.Sakauchi, “Enhanced navigation system with real images and real-time information”, World Congress on Intelligent Transport Systems(ITSWC), 2001.

[2] 遠藤 隆明, 片山 昭宏, 田村 秀行, 広瀬通孝, “写実的な広域仮想空間構築のための画像補間手法”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 185-192, 2002.

[3] T. Pintaric, U. Neumann and A. Rizzo, “Immersive panoramic video”, Proc. ACM International Conference on Multimedia, pp. 493-494, 2000.

[4] C. J. Taylor, “Videoplus: A method for capturing the structure and appearance of immersive environments”, IEEE Transactions on Visualization and Computer Graphics, Vol. 8, No. 2, pp. 171-182, 2002.

[5] 池田 聖, 佐藤 智和, 横矢 直和, “全方位型マルチカメラシステムを用いた高解像度な全天球パノラマ動画像の生成とテレプレゼンスへの応用”, 日本バーチャルリアリティ学会論文誌, Vol. 8, No. 4, pp. 443-450, 2003.

[6] M. Uyttendaele, A. Griminisi, S. Winder SB Kang, R. Szeliski and R. Hartley, “Image-based interactive exploration of real-world environments”, IEEE Computer Graphics and Application, Vol. 24, No. 3, pp. 52-63, 2004.

[7] 横矢 直和, “時空を越える拡張テレプレゼンス〜フライスルー MR 平城京〜”, JACIC 情報, Vol. 26, No. 3, pp. 62-67, 2011.

[8] 山本 和仁, 池谷 友介, 伊藤 喜輝, 三宅 美博, “地域の多様性を考慮した空間デジタルアーカイブシステム”, 第 12 回計測自動制御学会システムインテグレーション部門講演会論文集, pp 144-147, 2011.

[9] 藤川 和利, “自動車を用いた全周型テレプレゼンスシステム”, 情報科学技術フォーラム (FIT) 2004, pp. 335-336, 2004.

[10] 何 書勉, 田中 克己, “被写体追跡による全方位映像のメタデータ生成”, 第 16 回データ工学ワークショップ (DEWS) 論文集, 6A-i9, 2005.

[11] 堀 磨伊也, 神原 誠之, 横矢 直和, “低自由度モーションベースと没入型ディスプレイを用いた慣性力の再現によるテレプレゼンスシステムの構築”, 日本バーチャルリアリティ学会論文誌, Vol. 16, No. 2, pp. 283-292, 2011.

[12] 田中 佳樹, 大倉 史生, 堀 磨伊也, 神原 誠之, 横矢 直和, ”IBR テレプレゼンスのための提示映像評価に基づく画像獲得手法”, 電子情報通信学会 技術研究報告, MVE2011-128, 2012.

[13] 大倉 史生, 神原 誠之, 横矢 直和, “無人飛行船に搭載された 2 台の全方位カメラを用いた不可視領域のない全天球 HDR ビデオの生成”, 日本バーチャルリアリティ学会論文誌, Vol. 17, No. 3, pp. 139-149, 2012.

[14] 譲田 賢治, 坪内 貴之, 菅谷 保之, 金谷 健一, “移動ビデオカメラ画像からの運動物体の抽出”, 情報処理学会研究報告, CVIM-143, pp. 41-48. 2004.

[15] Y. Shen, F. Lu, X. Cao and H. Foroosh, “Video completion for perspective camera under constrained motion”, Proc. International Conference on Pattern Recognition(ICPR), vol. 3, pp. 63-66. 2006.

[16] 福地 功, 山下 淳, 金子 透, 三浦 憲二郎, “ 時空間画像処理による雨天時画像からの視野妨害ノイズ除去 ”, 映像情報メディア学会誌, Vol. 62, No. 5, pp. 771-777, 2008.

[17] 原田 知明, 山下 淳, 金子 透, “カメラの方向変化を利用した動的シーンからの視野妨害ノイズ除去”, 情報処理学会研究報告, CVIM-144, pp. 9-16, 2004.

[18] M. Bertalmio, L. Vesa, G. Sapiro and S. Osher, “Simultaneous structure and texture image inpainting ”, IEEE Transactions on Image Processing, Vol. 12, No. 8, pp. 882-889 2003.

[19] Y. Matsushita, E. Ofek and H. Shum “Full frame video stabilization with motion inpainting”, IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI), pp. 1150-1163, 2006.

[20] A. Flores and S. Belongie, “Removing pedestrians from Google Street View images”, IEEE International Workshop on Mobile Vision, 6 pages, 2010.

[21] B. Leibe, A. Leonardis and B. Schiele, “Robust object detection with interleaved categorization and segmentation”, International Journal of Computer Vision Vol. 77, No. 1-3, pp. 259-289, 2008.

[22] 町北 幸大郎, 河合 紀彦, 佐藤 智和, 横矢 直和, “動画像の欠損修復による全方位カメラを用いた不可視領域のない全天球テレプレゼンスの実現”, 画像の認識・理解シンポジウム (MIRU) 講演論文集, pp. 391-397, 2009.

[23] 堀 磨伊也, 河合 紀彦, 神原 誠之, 新井 イスマイル, 西尾 信彦, 横矢 直和, “パノラマビューシステムのための死角領域修復とプライバシー保護を行った全天球画像生成”, 画像の認識・理解シンポジウム (MIRU) 講演論文集, pp. 923-929, 2011.

[24] J. Herling and W. Broll, “Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments”, Proc. IEEE International Symposium on Mixed and Augmented Reality(ISMAR), pp. 207-212, 2010.

[25] D. Simakov, Y. Caspi, E. Shechtman and M. Irani, “Summarizing visual data using bidirectional similarity”, Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2008.

[26] C. Barnes, E. Shechtman, A. Finkelstein and D. B. Goldman, “PatchMatch: a randomized correspondence algorithm for structural image editing”, ACM Transactions on Graphics, 2009.

[27] 内山 寛之, 高橋 友和, 出口 大輔, 井手 一郎, 村瀬 洋, “複数画像系列の部分画像選択に基づく移動物体を含まない車載カメラ映像の生成”, 電子情報通信学会論文誌, Vol. J94-D, No. 12, pp. 2093-2104, 2011.

[28] J. Astola, P. Haavisto, and Y. Neuvo, “Vector median filters”, Proceedings of the IEEE, Vol. 78, No. 4, pp. 678-689, 1990.

[29] 高橋 英之, 堀 磨伊也, 神原 誠之, 横矢 直和, “全天球画像データベース作成のための動物体除去と色調統一”, 画像の認識・理解シンポジウム (MIRU) 講演論文集, pp. 1933-1940, 2010.

[30] S. Zokai, J. Esteve, Y. Genc, and N. Navab, “Multiview paraperspective projection model for diminished reality”, Proc. IEEE International Symposium on Mixed and Augmented Reality(ISMAR), pp. 217-226, 2003.

[31] 橋本 昂宗, 植松 裕子, 斎藤 英雄, “多視点カメラ撮影による野球のシースルー映像生成”, 映像情報メディア学会誌, Vol. 65, No. 4, pp. 505-513, 2011.

[32] Z. Garett and H. Saito, “Real-Time online video object silhouette extraction using graph cuts on the GPU”, Proc. International Conference on Image Analysis and Processing(ICIAP), pp. 985-994, 2009.

[33] 榎本暁人, 斎藤英雄, “複数のハンディカメラを利用したDiminished Reality ”, 画像の認識・理解シンポジウム (MIRU) 講演論文集, pp. 1277-1282, 2007.

[34] 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”, 日本バーチャルリアリティ学会論文誌, Vol. 17, No. 3, pp. 181-190, 2012.

[35] 清水 直樹, 橋本昂宗 植松裕子, 斎藤 英雄, “デプスカメラを用いたリアルタイム領域抽出による隠消現実感映像生成”, 映像情報メディア学会誌 Vol. 66, No. 12, pp. 549-552, 2012.

[36] C. Wu, VisualSFM: A visual structure from motion system, 2011. http://ccwu.me/vsfm/.

[37] M. Jancosek, T. Pajdla, “Multi-View reconstruction preserving weakly-wupported wurfaces”, Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 3121-3128, 2011.

[38] Y. Cheng, ”Mean shift, mode seeking, and clustering”, IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI), Vol. 17, No. 8, pp. 790-799, 1995.