

NAIST-IS-MT1351076

## 修士論文

# 移動カメラ映像からの視点移動機能を備える シネマグラフの生成

中島 章敬

2015年3月12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。

中島 章敬

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) |

# 移動カメラ映像からの視点移動機能を備える シネマグラフの生成\*

中島 章敬

## 内容梗概

近年, Google ストリートビューなど, 実写映像に基づいて構築した仮想化現実空間を探訪するアプリケーションが利用されている. このようなアプリケーションでは, 現状, シーン全体を静止画像で提示しているが, 動物体に関しては動きを提示した方がより高い臨場感をユーザに与えることができる. 動きの提示が可能な表現方法として, 動物体のみを違和感なくループさせ無限に動かし続けるシネマグラフと呼ばれる動画の提示手法がある. 従来, シネマグラフの生成は, 光学中心がほぼ固定された動画像を用いることを前提としている. しかし, Google ストリートビューなどの仮想化現実空間は広域な空間を対象とするため, 一般に移動カメラでの撮影により収集した画像群が用いられており, 視点位置に応じて撮影対象の見えが異なるため, 従来手法をそのまま適用することは難しい. 本論文では, これらの見えの違いを補償し, 動的シーンに対応したストリートビュー提示を実現することを目的として, 移動撮影した動画像からの各フレーム撮影位置でのシネマグラフ生成および, テクスチャの時間的連続性を考慮した撮影位置間の視点移動手法を提案する. ただし, 本研究では, 動物体として局所的な領域で動物体の見えの変化に周期性が存在する物体(水が流れる川や風に揺れる木の葉等)を対象とする. 本研究では, まず, 移動撮影で得られた動画から自由視点画像生成手法を用いて複数フレームを単一視点の画像に変形し, 複数の視点位置でのシネマグラフを生成する. 次に, 隣接する視点位置で生成されたシネマグラ

---

\*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-MT1351076, 2015年3月12日.

フ間で動物体の時間的な見えの変化が類似するシーケンスを探索する。最後に、類似したシーケンスがつながるように自由視点画像生成手法を用いてシネマグラフ間の中間視点画像を生成し視点移動を実現する。実験では、単眼カメラと全方位カメラで撮影した動画像を用いてシネマグラフを生成し、提案手法の有効性を評価する。また、移動撮影した動画像から視点移動可能なシネマグラフに基づくストリートビューのプロトタイプシステムを生成する。

## キーワード

シネマグラフ, 自由視点画像生成, 中間視点画像, 仮想化現実空間

# Viewpoint Movable Cinemagraph Generation from Video Captured with a Moving Camera \*

Akihiro Nakashima

## Abstract

Recently, applications such as Google Street View that enables us to explore the virtual space created based on real images are becoming common. Although such applications currently display still images of a target scene, it is expected to give users more highly realistic sensation if they show motions of moving objects. One of the projective methods that can reproduce motions in such applications is generation of a cinemagraph, which is an infinite looping video in which some objects continuously move. Conventional methods for cinemagraph generation assume that the optical center of a camera is fixed. However, the street view images for large scale environments are typically captured by a moving camera system. In this case, the appearance of a object changes according to the captured position. Therefore, it is difficult for conventional cinemagraph generation methods to generate a cinemagraph from a video captured with a moving camera. In order to create street view images that can represent dynamic scenes, we propose a method to create a cinemagraph from a video captured with a moving camera by using a novel view image generation technique. In this study, the dynamic objects whose motion can be reproduced are limited to those with periodic appearance changes in local areas (e.g. like trembling of leaves, flow of a river). For this propose, we generate a cinemagraph for each viewpoint using

---

\*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1351076, March 12, 2015.

the images that are generated by transforming input images to the appearance of a target frame with novel view synthesis. In order to enable smooth transition between adjacent viewpoints, we then generate images of intermediate viewpoints from an adjacent frame pair of cinemagraphs by searching for and connecting similar frames in them. In experiments, we demonstrate the effectiveness of the proposed method by generating cinemagraphs from videos captured with a standard monocular camera and an omni-directional camera. We also build a prototype system of the street view based on the proposed method.

**Keywords:**

Cinemagraph, novel view thensis, view interpolation, virtual space

# 目次

|  |    |
|--|----|
| 1. はじめに                                  | 1  |
| 2. 関連研究と本研究の位置づけ                         | 3  |
| 2.1 自由視点画像生成に関する研究                       | 3  |
| 2.1.1 Model-Based Rendering(MBR 法)       | 3  |
| 2.1.2 Image-Based Rendering (IBR 法)      | 6  |
| 2.1.3 Hybrid Rendering(HR 法)             | 9  |
| 2.2 シネマグラフ生成に関する研究                       | 10 |
| 2.3 本研究の位置づけ                             | 13 |
| 3. 視点移動可能なシネマグラフの生成                      | 14 |
| 3.1 提案手法の概要                              | 14 |
| 3.2 移動撮影した動画像からのシネマグラフ生成                 | 14 |
| 3.2.1 カメラ位置姿勢推定と三次元復元                    | 14 |
| 3.2.2 単一視点動画像の生成                         | 16 |
| 3.2.3 単一視点動画像からのシネマグラフ生成                 | 20 |
| 3.3 自由視点画像生成に基づく中間視点画像生成                 | 20 |
| 3.3.1 隣接する視点位置で生成したシネマグラフ間の類似シーケ<br>ンス探索 | 20 |
| 3.3.2 シネマグラフ間を接続する中間視点画像の生成              | 22 |
| 4. 実験                                    | 24 |
| 4.1 単眼動画像を用いたシネマグラフの生成実験                 | 24 |
| 4.1.1 実験環境                               | 24 |
| 4.1.2 実験結果およびその評価と考察                     | 26 |
| 4.2 シネマグラフに基づくストリートビューのプロトタイプシステム<br>の構築 | 33 |
| 4.2.1 プロトタイプシステムの構成                      | 33 |
| 4.2.2 実験環境                               | 33 |

|                                |    |
|--------------------------------|----|
| 4.2.3 実験結果およびその評価と考察 . . . . . | 36 |
| 5. まとめ                         | 42 |
| 謝辞                             | 43 |
| 参考文献                           | 44 |



## 目 次

|    |  |    |
|----|--|----|
| 1  | 欠損や歪みが生じた三次元形状 [1] . . . . .               | 4  |
| 2  | Furukawa らの手法 [7] により生成された三次元形状 . . . . .  | 5  |
| 3  | View morphing [10] . . . . .               | 7  |
| 4  | Town Digitizing [12] . . . . .             | 8  |
| 5  | 視点依存テクスチャ [20] . . . . .                   | 10 |
| 6  | Liao らの手法 [24] で決定した開始フレームとループ周期 . . . . . | 12 |
| 7  | Panoramic Video Texture [25] . . . . .     | 12 |
| 8  | 移動撮影した動画像からのシネマグラフ生成の流れ . . . . .          | 15 |
| 9  | 中間視点画像生成の流れ . . . . .                      | 15 |
| 10 | VisualSFM のカメラ位置姿勢推定結果と疎な三次元形状 . . . . .   | 17 |
| 11 | CMPMVS の出力結果例 . . . . .                    | 17 |
| 12 | カメラ位置姿勢と三次元形状による画素の対応付け . . . . .          | 18 |
| 13 | デプステストによる可視領域の判定 . . . . .                 | 18 |
| 14 | 対象フレーム視点への変換例 . . . . .                    | 19 |
| 15 | 物体の見えの変化を考慮した類似シーケンス探索 . . . . .           | 21 |
| 16 | 画素に対するシネマグラフの選択基準 . . . . .                | 23 |
| 17 | 誤った画素が選択される場合 . . . . .                    | 23 |
| 18 | 入力動画像の一部 (単眼動画像) . . . . .                 | 25 |
| 19 | シネマグラフを生成する視点 . . . . .                    | 25 |
| 20 | カメラ位置姿勢推定結果と三次元復元結果 (単眼動画像) . . . . .      | 27 |
| 21 | 指定した動領域 (単眼動画像) . . . . .                  | 27 |
| 22 | 単一視点動画像生成結果 (248 フレーム目視点) . . . . .        | 28 |
| 23 | 単一視点動画像生成結果 (268 フレーム目視点) . . . . .        | 29 |
| 24 | シネマグラフ生成結果 (248 フレーム目視点) . . . . .         | 30 |
| 25 | シネマグラフ生成結果 (268 フレーム目視点) . . . . .         | 31 |
| 26 | 中間視点画像生成結果 (単眼動画像) . . . . .               | 32 |
| 27 | 各カメラの歪み補正画像 (入力画像) . . . . .               | 35 |
| 28 | カメラ位置姿勢推定結果と三次元復元結果 (全方位動画像) . . . . .     | 37 |

|    |                                       |    |
|----|---------------------------------------|----|
| 29 | 指定した動領域 (全方位動画像) . . . . .            | 37 |
| 30 | 生成した 40 フレーム目における CubeMap . . . . .   | 38 |
| 31 | 対象フレーム視点 (40 フレーム目) への変換結果 . . . . .  | 38 |
| 32 | 生成した 100 フレーム目における CubeMap . . . . .  | 39 |
| 33 | 対象フレーム視点 (100 フレーム目) への変換結果 . . . . . | 39 |
| 34 | 動画の開始フレームと終了フレームの関係 . . . . .         | 40 |
| 35 | 中間視点画像生成結果 (全方位動画像) . . . . .         | 41 |

## 表 目 次

|   |                                    |    |
|---|------------------------------------|----|
| 1 | プロトタイプシステム構築に用いたコンピュータ仕様 . . . . . | 33 |
| 2 | 三次元復元に用いたコンピュータ仕様 . . . . .        | 34 |
| 3 | プロトタイプシステム構築の計算時間 . . . . .        | 36 |

## 1. はじめに

近年，現実空間の情報に基づいて構築された仮想空間は仮想化現実空間と呼ばれ，あたかもその場にいるような没入感を与えるテレプレゼンスや景観のデジタルアーカイブ等の様々なアプリケーションに応用されている．仮想化現実空間の例として，Google ストリートビューが挙げられる．このようなアプリケーションでは，事前に車載全方位カメラを用いて様々な地点を移動撮影することで収集した画像群から，ユーザが指定した場所・方向に応じて画像の選択および切り出しを行いユーザに提示することで，テレプレゼンスを実現している．しかし，現状のシステムでは臨場感に関して以下のような問題が挙げられる．

- (1) 歩行者，自動車，揺れる木の葉等の動物体も静止画像で提示される．
- (2) 撮影地点間を移動する時に映像に歪みが生じる．

(1) について，仮想化現実空間内で，動物体の動きを提示した方がより高い臨場感をユーザに与えることができる．動きを提示可能な表現手法として，動物体のみが無限に動き続けるシネマグラフと呼ばれる動画の提示手法がある．シネマグラフでは，単純に動画が繰り返し再生されるのではなく，画像中に静止画像の領域と動画像の領域が存在し，動画像の領域のみ動物体が周期的な動きを無限に繰り返す．このようなシネマグラフを生成する手法 [22–26] が従来から研究されている．従来手法は，共通して光学中心がほぼ固定された動画像を用いることを前提としている．しかし，Google ストリートビューのような広大な仮想化現実空間を構築する場合，一般に移動カメラで撮影した動画像が用いられるため，撮影対象の見えは統一されていない．このため，シネマグラフ生成の従来手法を移動撮影した動画像に対してそのまま適用するだけでは，良好な結果を得ることは難しい．(2) について，Google ストリートビューなどでは画像を歪ませ，かつ2視点の画像をブレンディングすることによって，あたかも移動しているかのような感覚を与える映像を提示しているが，自由視点画像生成の技術を応用することで中間視点を生成し，臨場感を向上させることが期待できる．従来，撮影地点間の滑らかな視点移動を実現する手法 [10–13, 15–19] が研究されてきたが，入力動画の撮影に特殊な機器が必要となる問題や物体の動きの再現が撮影時間によって

制限される問題，動的環境に対応していないという問題がある．また，2視点の画像が動画の場合の中間視点映像の生成手法 [26] も提案されているが，ブレンドによりぼけが発生する問題がある．

本論文では，(1)の問題に対して，移動撮影した動画からある視点位置での動きを再現するシネマグラフを生成する手法を提案する．具体的には，入力動画の各フレームのカメラ位置姿勢を推定し，撮影対象の三次元形状を復元する．次に，推定したカメラ位置姿勢と復元した三次元形状を用いて対象フレームを設定し，対象フレームの位置姿勢からの見えに統一された単一視点動画を生成する．生成した単一視点動画にシネマグラフ生成手法 [24] を適用し，ある視点位置でのシネマグラフを生成する．ただし，本研究では，局所的な領域で動物体の見えの変化に周期性が存在する物体（水が流れる川や風に揺れる木の葉等）の動きを対象とする．(2)の問題に対しては，移動撮影した動画から生成した各地点でのシネマグラフを用いて撮影地点間の滑らかな視点移動を実現する手法を提案する．具体的には，まず，隣接する視点位置で生成したシネマグラフの組で，物体の時間的な見えの変化が類似するシーケンスを探索する．次に，探索した類似シーケンスがつながるように，自由視点画像生成の技術を用いてシネマグラフ間の中間視点画像を生成し視点移動を実現する．これら提案手法を用いて，移動撮影した動画から物体の動きを再現しつつ視点移動可能なシネマグラフを生成し，かつこれに基づくストーリービューのプロトタイプシステムを構築する．

本論文では，2章でシネマグラフ生成に関する研究と自由視点画像生成に関する研究を概観し，本研究の位置づけと方針を述べる．3章では，移動撮影した動画からのシネマグラフ生成および視点依存テクスチャマッピングを用いた視点移動について述べる．4章では，提案手法の有効性を示すため，実環境において実験を行い，結果を考察する．5章では，本論文のまとめと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

本研究の目的は，移動撮影した動画像から物体の動きを再現した視点移動可能なシネマグラフの生成，およびそれに基づくストリートビューを構築することである．本章では視点移動を実現する自由視点画像生成に関する従来研究，物体の動きを再現するシネマグラフ生成に関する研究を概観し，本研究の位置づけを述べる．

### 2.1 自由視点画像生成に関する研究

従来，研究されてきた自由視点画像生成手法は，対象環境の三次元形状を復元し用いる Model-Based Rendering (MBR) 法，対象環境を撮影した画像群を変形と合成することで自由視点画像を生成する Image-Based rendering (IBR) 法，MBR 法と IBR 法を組み合わせた Hybrid rendering(HR) 法に大別できる．

#### 2.1.1 Model-Based Rendering(MBR 法)

MBR 法は，自由視点画像生成を行う対象環境の三次元形状にテクスチャマッピングをした三次元モデルを生成しておき，それを設定した視点に対して投影することで画像を生成する手法である．対象環境の形状を用いるため，視点移動における物体の隠蔽関係の表現に優れる．しかし，生成される自由視点画像の品質は三次元形状の精度に依存し，画像に基づく三次元復元手法などにより得られる欠損や誤差を含む精度の低い三次元形状を用いた場合には，図1のように生成した画像に欠損や歪みが生じる．このため，対象となる環境の三次元形状を精度良く取得するための手法が研究されてきている．このような手法は，撮影対象に光等を照射することで対象の形状を計測する能動的な手法と，対象環境を撮影した画像群から画像処理によって三次元形状を推定する受動的な手法に大別できる．以下では，各手法について述べる．

##### [能動的な三次元復元手法]

能動的な三次元復元手法として，距離センサを用いた手法 [1] やプロジェクタ等

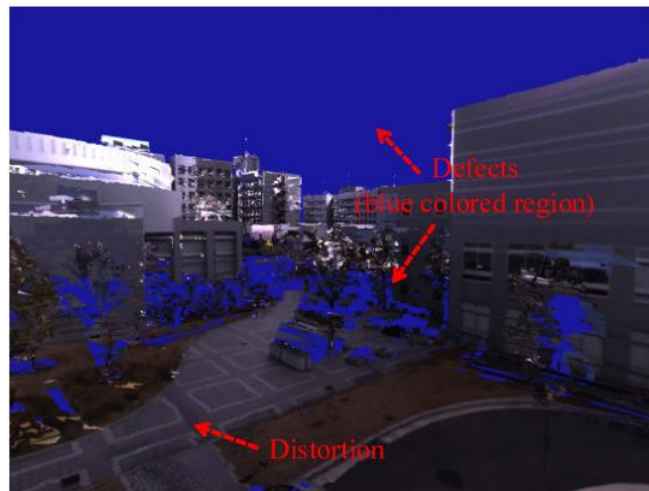


図 1 欠損や歪みが生じた三次元形状 [1]

を用いたアクティブステレオ法 [2,3] が提案されている。距離センサを用いた手法 [1] では、対象環境に光レーザを照射することで、レーザが反射して観測されるまでの時間を計ることでレンジデータを取得する。一方、アクティブステレオ法として、塚本ら [2] は、プロジェクタ等によってスリット光を対象環境に照射し、スリット光検出用のカメラでスリット光の位置を検出することで、三角測量の原理により対象環境の三次元形状を奥行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つつ計測するため時間を要する。この問題に対して、Kawasakiら [3] 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、計測時間の短縮に成功した。これらの手法は、精度の高い環境の三次元形状を取得可能であるが、プロジェクタや距離センサといった特殊な機材を用いる。そのため、一般のユーザが手軽に三次元復元を行うことは難しい。

#### [受動的な三次元復元手法]

受動的な三次元復元の代表的な手法として、多視点ステレオ法や視体積交差法がある。これらの手法は、画像撮影時のカメラ位置姿勢が既知の複数枚の画像から三次元復元を行う。これらの手法は、プロジェクタや距離センサといった特殊な機器を必要としないが、各カメラの位置関係を把握するためには、高精度なカ



図 2 Furukawa らの手法 [7] により生成された三次元形状

メラ外部パラメータの推定が必要となる。カメラ外部パラメータの推定手法として、三次元位置が既知であるマーカを用いる手法 [4] が代表的であるが、屋内環境を想定しており屋外環境を対象とした場合、視野角内に十分な数のマーカを配置することは難しい。これに対して、Structure from Motion(SfM) 法は、複数の画像を入力として与え、各画像から特徴点を抽出し、画像間でそれらの対応点を求めることで、各画像のカメラ位置姿勢および対応点の三次元座標を推定する。このようにして得られたカメラの位置姿勢を多視点ステレオ法や視体積交差法に用いることで、密な三次元モデルを復元する。

多視点ステレオ (Multi-view Stereo : MVS) 法 [5-7] は、カメラ位置姿勢が既知の画像群を入力とし、ステレオ法によってそれぞれの画像の奥行画像を推定することで対象環境の三次元形状を復元する手法である。MVS 法の代表的な方法として Furukawa ら [7] の Photo-Consistency を高めることでインターネット上に大量にある画像を用いて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、図 2 に示すように生成モデルに欠損が生じる。

視体積交差法 [8] 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形状を生成する手



法である。Kolev ら [9] 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。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形状を生成できる。ステレオ法とは異なり対応点探索の問題が発生しないため、復元できる三次元形状の品質がボクセル空間の解像度や使用する画像に左右されるが計算コストがステレオ法と比べて小さくなる。また、MVS 法と異なり欠損のない三次元形状を生成できる。しかし、原理的に対象物体の凹な部分については正しい三次元形状が得られないという制約がある。

### 2.1.2 Image-Based Rendering (IBR 法)

IBR 法は、対象とする環境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て、これらを変形・合成することで写実性の高い自由視点画像を生成する手法である。対象とする環境の三次元形状や反射特性を直接利用しないため、正確な幾何形状の復元が難しい複雑な形状を持つ木などの物体に関しても、写実性の高い描画が可能であり、MBR 法より自然な自由視点画像を生成できる。しかし、視点の変化量が増えるに応じて、自由視点画像に歪みが生じる。以下では、画像の変形に基づく手法と光線情報を用いる手法に大別して述べる。

#### [画像変形に基づく手法]

撮影された画像群を直接変形することで、未撮影視点である中間視点画像を生成する手法が提案されている。代表的な手法として、モーフィングやワーピングによって中間視点画像を生成する手法 [10,11] がある。これらの手法は、複数の入力画像間で対応点を与え、中間視点位置における対応点の位置を決定することで入力画像の変形を行う。さらに、対応点の位置関係に基づいて入力画像の色をブレンディングすることで中間視点位置における色を画素ごとに求め、中間視点画像を生成する (図 3)。これらの手法では、入力画像に対応点を与えることで中

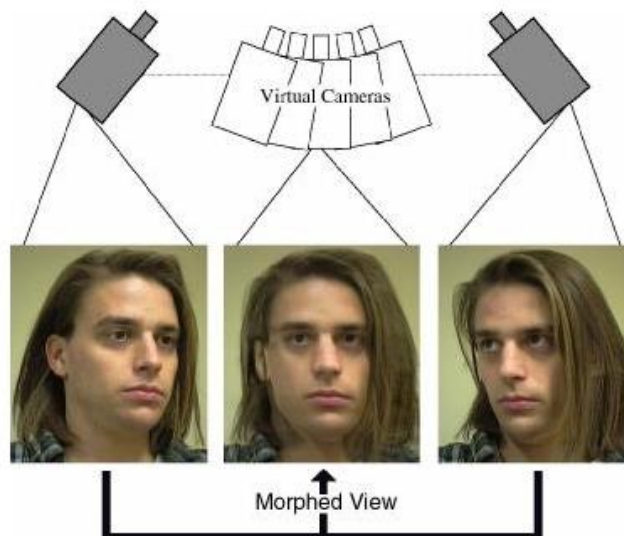


図 3 View morphing [10]

間視点画像を生成し，2 視点間に限定した視点移動を実現している．しかし，動的環境を対象とした場合には，対応点を正確かつ自動的に与えることは難しい．

これに対して，対応点を明示的に与えずに画像群を変形することで中間視点画像を生成する手法 [12, 13] が提案されている．離散的な視点位置で撮影された複数の全方位画像を用いる Town Digiting [12](図 4) では，ブレンディングと画像の拡大・縮小を用いて 2 枚の画像の視点位置の中間視点画像を生成している．このシステムでは静的環境を対象としている．一方，動的環境を考慮した手法として，密に設置した複数の非同期カメラで撮影した動画像を用いる手法 [13] が提案されている．この手法では，画素ごとのオプティカルフローに基づいて各カメラ位置における動物体間の対応関係を算出し，この対応関係から算出される平面射影変換画像群を用いてシーン全体の画像変形モデルを作成することで動的環境に対応した中間視点画像を生成する．物体の動きは再現されるが，複数のカメラを密に設置する必要があり，入力データの撮影コストが高いという問題がある．

これら画像群の変形に基づく手法においては，動的環境を対象とした仮想環境におけるユーザの自由な視点移動と物体の動きの再現を同時に実現する手法は提

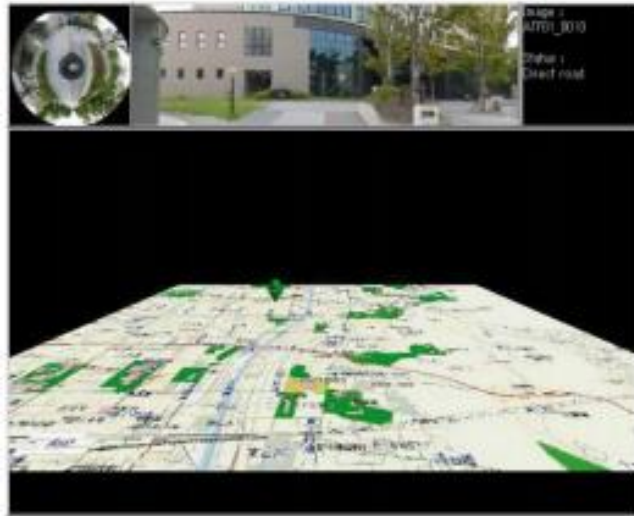


図 4 Town Digitizing [12]

案されていない。

#### [光線情報に基づく手法]

画像を光線情報が集合したものであるとし、任意の視点および視線の自由視点画像を生成する手法が提案されている。画像を、視点の座標  $(x, y, z)$ 、光線の向き  $(\theta_x, \theta_y)$ 、波長  $\lambda$  および時間  $t$  の集合からなる 7 次元の光線情報の集合と捉えることで、対象環境の空間を 7 次元の光線情報で記述することができる [14]。対象の環境に関して、7 次元のデータをすべて記録することができれば、対象の環境における任意視点から見た画像の各画素値を光線情報から決定することが出来る。しかし、実際には全ての 7 次元データをサンプリングすることは難しく、サンプリングされなかった光線情報は欠損となる。これに対して、Naemura ら [15] 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ている。この手法では、7 次元の光線情報から  $z$  座標、波長  $\lambda$  と時間  $t$  を削減し、情報量を 4 次元とすることで取得すべき画像の量を大幅に削減した。また、全方位カメラで一度に多数の方向の光線情報を収集する手法 [16, 17] も提案されている。山口ら [16] は、全方位画像センサを用いて直線上を移動しながら環境を撮影し、任意の視線方向に対してステレ

オ視可能な画像を生成する手法を提案した。また, Takahashi ら [17] は, 全方位画像センサを搭載した自動車道路を走行し, 移動経路付近の自由視点画像を生成する手法を提案している。これらの光線情報を用いた手法は, 主に静的環境を対象としており, 多くの光線情報を必要とする動的環境を対象として画像の合成に必要な光線情報を収集することは難しい。

### 2.1.3 Hybrid Rendering(HR 法)

MBR 法と IBR 法のハイブリッドな方法 (HR 法) では, 対象環境の形状と画像群を用いることで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。HR 法では, 視点に応じて使用するテクスチャあるいは画素値を選択する視点依存テクスチャ (View-Dependent Texture Mapping:VDTM) と呼ばれる手法と視点依存ジオメトリ (View-Dependent Geometry:VDG) と呼ばれる視点に応じた奥行画像から復元出来る三次元形状を用いる手法がある。

Debevec ら [20] は, 事前に生成した簡易な三次元形状に対して, 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 VDTM を提案した。これにより, 詳細な三次元形状の見えをテクスチャとして再現できるが, 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。Matusik ら [18] は, 動画像を入力とし自由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とその画素を結ぶ直線が Visual Hull を貫通するか否かを判定することにより三次元形状にマッピングするテクスチャをリアルタイムで選択し, 仮想視点から見た自由視点画像を生成する手法を提案している。石川ら [19] は, 複数の全方位カメラによって同期撮影された動画像を基に, 動的環境において自由視点画像を生成する手法を提案している。この手法では, まず, 撮影環境を静的領域と動的領域に分離する。静的領域に対してはあらかじめ与えられた対応点から対象環境の疎な三次元形状を復元し, 三次元形状を考慮したモーフィングにより高速な自由視点画像生成を行う。また, 動的領域に対しては Visual Hull を用い



(a) 三次元形状

(b) レンダリング結果

図 5 視点依存テクスチャ[20]

た自由視点画像生成を行い，それぞれの領域の自由視点画像を合成することにより，環境全体の自由視点画像を生成している．対象環境の幾何形状を併用することで，動的環境における自由視点画像生成を実現しているが，幾何形状を取得するためには複数の同期撮影された動画像が必要である．VDTMとVDGを併用して用いる手法として，伊吹ら[21]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，三次元形状の時間的・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形状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歪みを低減した．しかし，この手法は，動的環境に対応していない．

## 2.2 シネマグラフ生成に関する研究

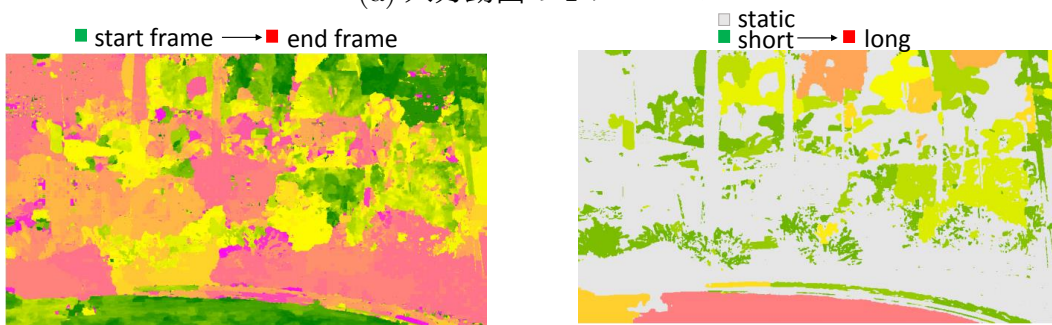
静止画像の領域と動画像の領域を併せ持つ新しい表現であるシネマグラフ生成はSchödlらによるビデオテクスチャ[22]と呼ばれる手法が始まりである．この手法では，周期的な動きを繰り返す物体を一定時間撮影した動画像を入力とし，時間的な連続性を考慮することであるフレームと類似したフレームを動画像中から探索し，それらのフレームをつなげることでループビデオを生成する手法である．Kwatraら[23]は，グラフカットを用いて画素ごとに開始フレームとそれに類似す

るフレームを，各画素の周期が同一という仮定のもとに決定する手法を提案した．これらの手法では，全画素について同一のループ周期が設定されるため，動きの周期が異なる動物体が動画像内に複数存在する場合，繰り返し時に動きが不連続なる物体が発生する可能性がある．この問題に対して，Liao ら [24] は，フレーム間での時間的整合性とフレーム内での空間的整合性を考慮して，画素ごとにループ周期と開始フレームを決定する手法を提案した (図 6)．これにより，周期の異なる複数物体が存在する動画にも対応した．また，この手法では 1 パラメータの調整によりシネマグラフ中の動画像領域の範囲の大きさを任意に設定することができる．しかし，これらの手法は固定カメラにより撮影された動画像を入力として想定している．これに対して，Agarwala ら [25] は，カメラを横に振りながら撮影した動画像を入力として，パノラマのシネマグラフを生成する手法を提案した (図 7)．これにより，カメラの動きに関する制約は緩和されたが，それでもなおカメラの光学中心がほぼ固定されている動きに限定される．

一方，馬場口ら [26] は，異なる視点位置で独立にシネマグラフを作成しておき，その間の視点移動を実現する手法を提案している．しかし，複数地点で固定カメラにより撮影した動画像が必要であり，また動画像間で対応点を手動で与える必要がある．また中間視点を各画像のブレンディングにより生成しているため，結果画像がぼける問題がある．



(a) 入力動画の1フレーム



(b) 開始フレーム

(c) ループ周期

図 6 Liao らの手法 [24] で決定した開始フレームとループ周期



図 7 Panoramic Video Texture [25]

## 2.3 本研究の位置づけ

本研究の目的は、移動撮影した動画像から物体の時間的な見えの変化を表現する視点移動可能なシネマグラフを生成し、それに基づくストリートビューを構築することにある。現状の Google ストリートビューなどで広く利用されている仮想化現実空間の問題点として以下の2点があげられる。

- (1) 歩行者，自動車，揺れる木の葉等の動物体も静止画像で提示される。
- (2) 撮影地点間を移動するとき映像に歪みが生じる。

(1)の問題について，動物体の動きをシネマグラフとして再現することが考えられる。しかし，従来手法では光学中心がほぼ固定されたカメラで撮影された動画像が入力であることが前提とされており，自動車の搭載されたカメラなどで移動撮影した動画像に対してシネマグラフ生成の従来手法を適用しても良好な結果を得ることは難しい。(2)の問題については，撮影視点間の滑らかな視点移動を実現できる自由視点画像生成の技術を応用することで中間視点を生成し，臨場感を向上させることが期待される。しかし，従来手法の多くは，静的なシーンを対象としており，シネマグラフ間の視点移動にそのまま適用することは難しい。一方，2地点で撮影した動画像の中間視点映像の生成手法 [26] も提案されているが，ブレンディングによりぼけが発生する問題がある。本論文では，(1)の問題に対して，撮影対象の三次元復元およびそれに基づく自由視点画像生成により，単一視点動画像を生成することで，移動撮影画像からのシネマグラフ生成を実現する。(2)の問題に対しては，自由視点画像により中間視点を生成するが，この際に隣接する視点位置でのシネマグラフ間で，物体の時間的な見えの変化が類似するシーケンスを探索し，それがつながるように中間視点画像を生成する。これにより，動画像間の滑らかなテクスチャの変化を実現する。本研究では，移動撮影した動画像から物体の動きを再現しながら視点移動可能なシネマグラフを生成し，それに基づくストリートビューのプロトタイプシステムを作成する。ただし，本研究では，局所的な領域で動物体の見えの変化に周期性が存在する物体（水が流れる川や風に揺れる木の葉等）の動きを対象とする。



### 3. 視点移動可能なシネマグラフの生成

本章では、視点移動可能なシネマグラフに基づいたストリートビューを実現するために、移動撮影した動画像からのシネマグラフ生成手法とシネマグラフ間の視点移動手法について説明する。

#### 3.1 提案手法の概要

本研究では、移動撮影した動画像を入力として、各フレームでのシネマグラフの生成および生成したシネマグラフ間の視点移動を実現する。移動撮影した動画像からシネマグラフを生成する手法の流れを図 8 に示す。まず、移動撮影した動画像に対して SfM 法を適用することでカメラ位置姿勢推定し、推定したカメラ位置姿勢と入力画像群から MVS 法により対象環境の三次元復元を行う。次に、シネマグラフを生成する対象フレームを設定し、入力動画像を推定したカメラ位置姿勢と対象の三次元形状を用いて対象フレーム視点の見えに変換する。次に、生成した単一視点動画像にシネマグラフ生成の従来手法である Liao らの手法 [24] を適用することで、対象フレームのシネマグラフを生成する。ただし、本研究では、局所的な領域で動物体の見えの変化に周期性が存在する物体 (水が流れる川や風に揺れる木の葉等) をシネマグラフ内で動く対象とする。隣接する視点位置で生成したシネマグラフ間を接続する中間視点画像を生成する手法の流れを図 9 に示す。まず、隣接する視点位置で生成されたシネマグラフを入力とし、動物体の時間的な見えの変化が類似するシーケンスを探索する。次に、探索した類似シーケンスがつながるように自由視点画像生成手法に基づき中間視点画像を生成することで視点移動を実現する。以下では、各処理について詳述する。

#### 3.2 移動撮影した動画像からのシネマグラフ生成

##### 3.2.1 カメラ位置姿勢推定と三次元復元

本研究では、移動撮影した動画像に対して、SfM 法の最新手法の実装である VisualSFM [27] を用いて各フレームのカメラ位置姿勢推定を行う。VisualSFM

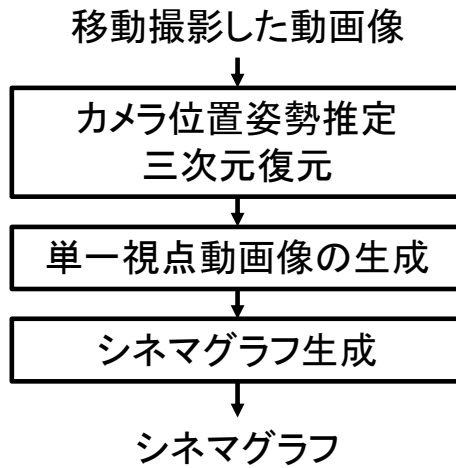


図 8 移動撮影した動画像からのシネマグラフ生成の流れ

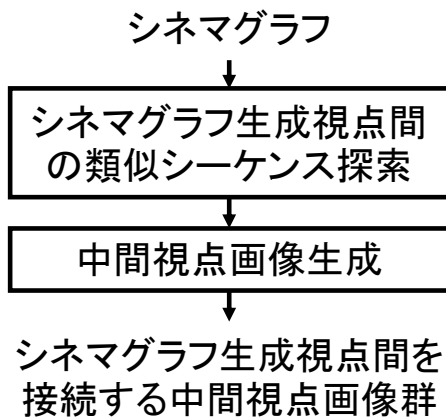


図 9 中間視点画像生成の流れ

は、Scale-Invariant Feature Transform(SIFT) を用いて自然特徴点を画像ごとに検出し、フレーム間で特徴点の対応を付けることで、撮影時のカメラ位置姿勢と自然特徴点の三次元位置を推定する (図 10). これらの情報を MVS 法の最新手法である CMPMVS [28] に入力することで、図 11 に示すような対象環境の密な三次元形状を復元する.

### 3.2.2 単一視点動画像の生成

移動撮影した動画像から推定したカメラ位置姿勢と三次元形状を用いた単一の対象フレーム視点に対する動画像の生成方法について説明する. まず, 復元した三次元形状と推定したカメラ位置姿勢から入力動画像の各フレームのデプスマップを求める. 次に, シネマグラフを生成する対象フレームを設定した上で, 対象フレームのデプスマップと他フレームのカメラ位置姿勢に基づき, 対象フレームの画素を他フレームに投影して, 対応する画素値をコピーすることで, 対象フレームの見えに統一された単一視点動画像を生成する (図 12). このとき, 対象フレームの視点から対応する世界座標系における点  $\mathbf{P} = (x, y, z)$  が観測できるかをデプステストにより判定する. 図 13 に示すように, フレーム  $A$  のカメラ位置姿勢からであれば点  $\mathbf{P}$  を観測できるが, このときフレーム  $A$  のカメラ座標系における点  $\mathbf{P}$  の  $Z$  座標の値  $Z_A$  は, フレーム  $A$  のカメラ位置姿勢に対して生成したデプスマップの点  $\mathbf{P}$  が投影される座標における値  $D_A$  と一致する. 一方, フレーム  $B$  のカメラ位置姿勢からは点  $\mathbf{P}$  を観測できず, このような場合には, 点  $\mathbf{P}$  の  $Z$  座標の値  $Z_B$  はフレーム  $B$  のカメラ位置姿勢に対して生成したデプスマップの点  $\mathbf{P}$  が投影される座標における値  $D_B$  より大きくなる. 以上のことから, デプスマップに投影された座標における値  $D$  とフレームのカメラ座標系における  $Z$  座標の値について

$$D + k \leq Z$$

を満たすとき点  $\mathbf{P}$  を観測できると判定し画素値をコピーする. ただし,  $k$  は定数である. 図 14 に対象フレーム視点への変換例を示す.

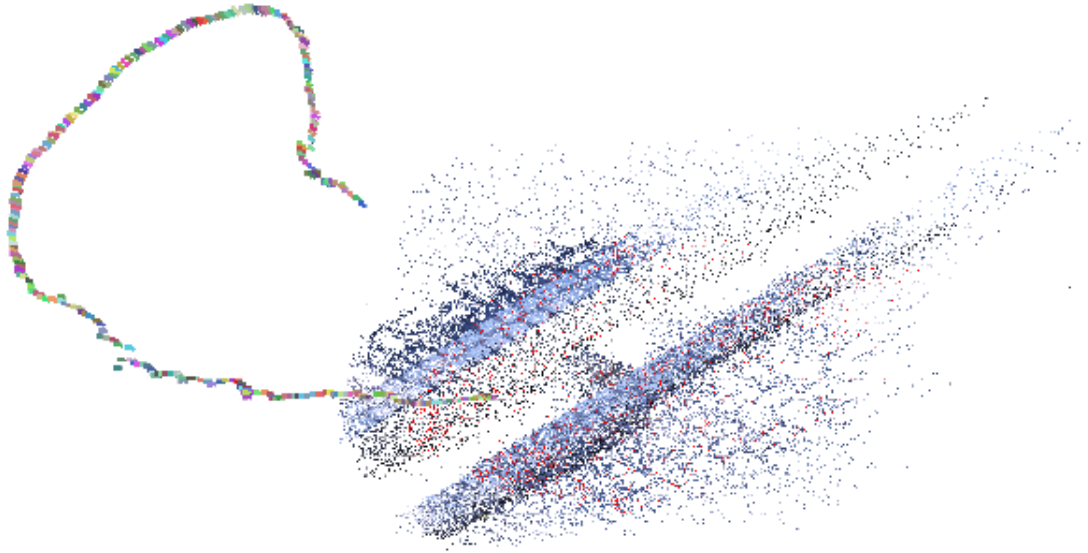


図 10 VisualSFM のカメラ位置姿勢推定結果と疎な三次元形状

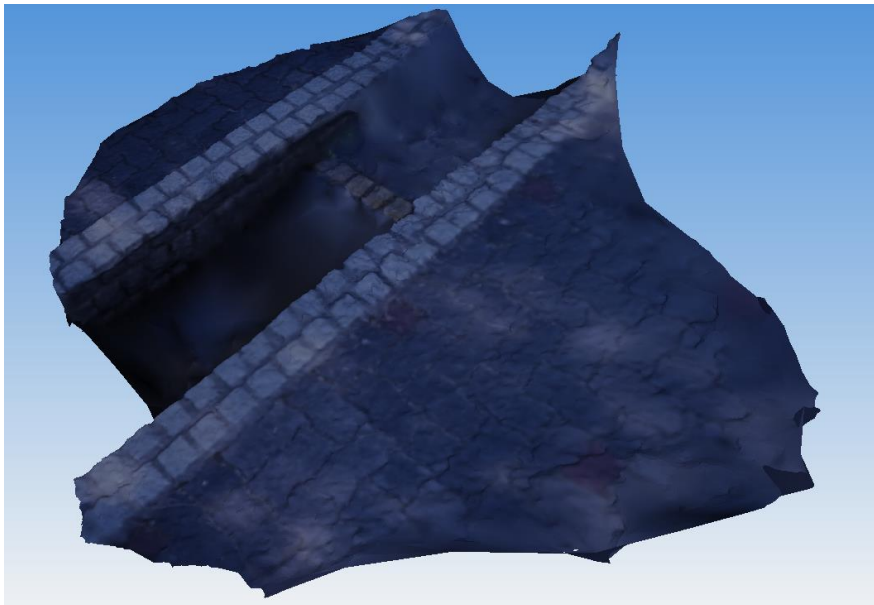


図 11 CMPMVS の出力結果例

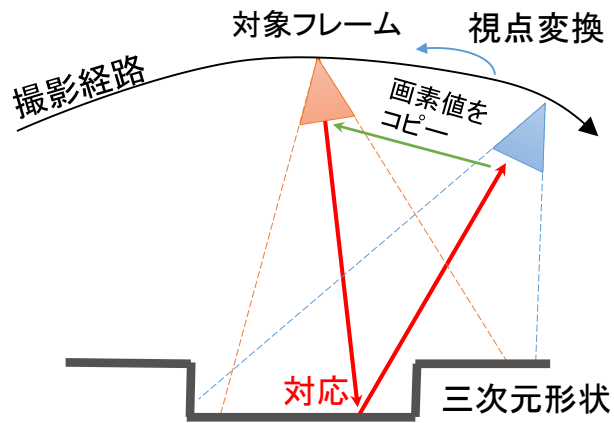


図 12 カメラ位置姿勢と三次元形状による画素の対応付け

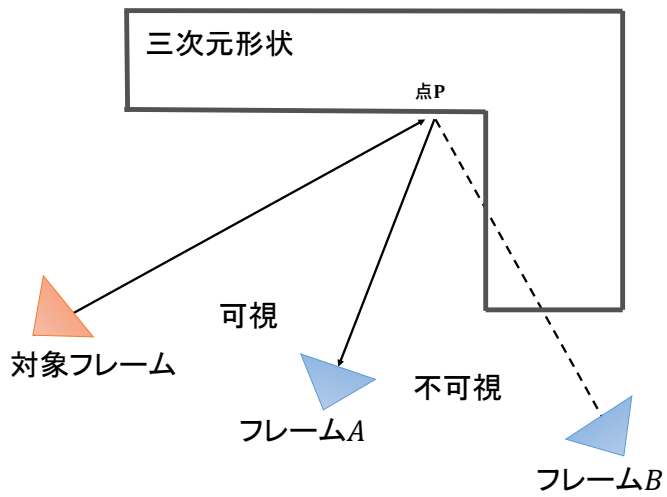
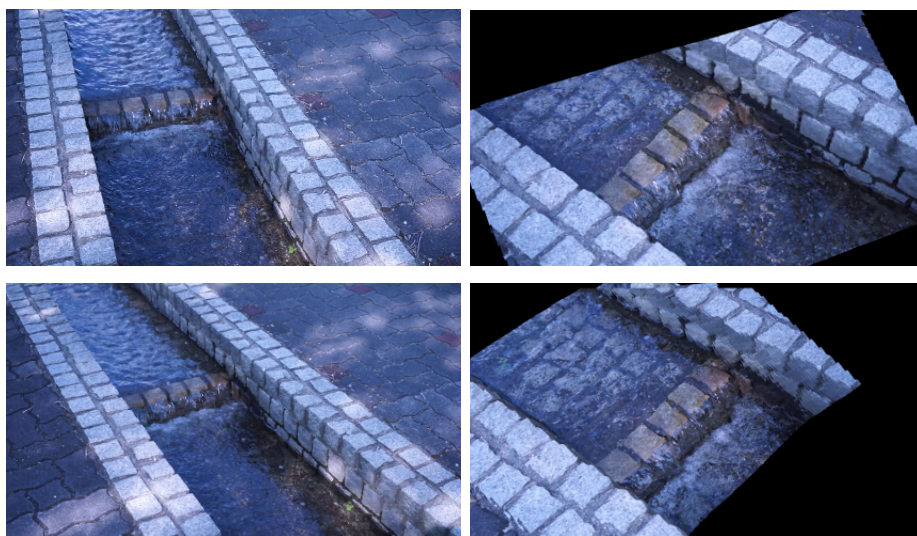


図 13 デプステストによる可視領域の判定



対象フレームの見え



(左) 入力フレーム

(右) 視点変換画像

図 14 対象フレーム視点への変換例

### 3.2.3 単一視点動画からのシネマグラフ生成

ここでは、対象フレームに隣接するフレーム群が対象フレームの撮影位置近辺で撮影されたと仮定し、隣接するフレーム群を用いてシネマグラフの生成を行う。シネマグラフの生成は、フレーム間での時間的整合性とフレーム内での空間的整合性を考慮して、画素  $\mathbf{x}$  ごとにループ周期  $P_x$  と開始フレーム  $S_x$  を決定する Liao らの手法 [24] を適用する。この手法は、時間的整合性として開始フレームの画素値  $I_{S_x}(\mathbf{x})$  と終了フレームの画素値  $I_{S_x+P_x-1}(\mathbf{x})$  の差を、空間的整合性として画素  $\mathbf{x}$  の注目画素と 4 近傍画素の画素値の差を用い、これにより表現されるエネルギー式が最小となる  $(S_x, P_x)$  を画素ごとに決定する。

ただし、入力である単一視点動画は少なからず復元誤差を含む三次元形状を用いて生成されるため、対象フレーム視点への視点変換は完全ではなく、生成したシネマグラフにおいて実際には動きのない静止領域に細かな動きが発生する可能性がある。本研究では、復元した三次元形状中の動領域を人手で指定することで、各画像上において動きを再現したい動領域を特定する。これを用いて、指定された動領域以外の領域に、対象フレームの画素値を適用することで静止領域の細かな変化を抑止する。

## 3.3 自由視点画像生成に基づく中間視点画像生成

隣接する視点位置で生成されたシネマグラフ間の滑らかな視点移動を実現するために、隣接する視点位置で生成したシネマグラフ間の類似シーケンス探索とこれを用いたシネマグラフ間を接続する中間視点画像の生成を行う。

### 3.3.1 隣接する視点位置で生成したシネマグラフ間の類似シーケンス探索

隣接する視点位置で生成したシネマグラフ  $(A, B)$  間で、見えの時間的な変化が類似するシーケンスを探索する方法について説明する。ここでは、対象の見え方の変化を考慮した探索を行うために、フレーム  $A$  視点で生成したシネマグラフ  $A$  をフレーム  $B$  視点で生成したシネマグラフ  $B$  の見えに変換したシネマグラフ  $A'$  を用いて類似シーケンスの探索を行う (図 15)。シネマグラフ  $A'$  の生成は、単

一視点動画像を生成した方法と同様に、フレーム  $B$  のデプスマップとフレーム  $A$  のカメラ位置姿勢に基づき、フレーム  $B$  の画素をフレーム  $A$  の画素に投影し、対応する画素値をコピーすることにより  $A$  を  $B$  の視点の画像に変換することで行う。ここでは、シネマグラフ  $(A', B)$  を入力として、以下の式により類似シーケンスの画素値の差の絶対値の合計である SAD(Sum of Absolute Diffrence) を算出し、これを最小にする開始フレーム対  $(S_A, S_B)$  を求めることで類似シーケンスを探索する。

$$(S_A, S_B) = \operatorname{argmin}_{s_a, s_b} \sum_{t=0}^{N-1} \sum_{\mathbf{x} \in (\Omega_M \cap \Omega_P)} |I_{A'_t+s_a}(\mathbf{x}) - I_{B_t+s_b}(\mathbf{x})|$$

ただし、 $N$  は探索する類似シーケンスのフレーム枚数、 $I_{A'_t}$  および  $I_{B_t}$  ( $0 \leq t \leq N-1$ ) はシネマグラフ  $(A', B)$  の各フレーム、 $I_{A'_t}(\mathbf{x})$  および  $I_{B_t}(\mathbf{x})$  は画素  $\mathbf{x}$  の画素値を表し、類似シーケンスの探索にはシネマグラフ  $B$  の動領域  $\Omega_M$  かつデプステストにおいて可視であると判定された領域  $\Omega_P$  に含まれる画素  $\mathbf{x}$  のみを用いる。

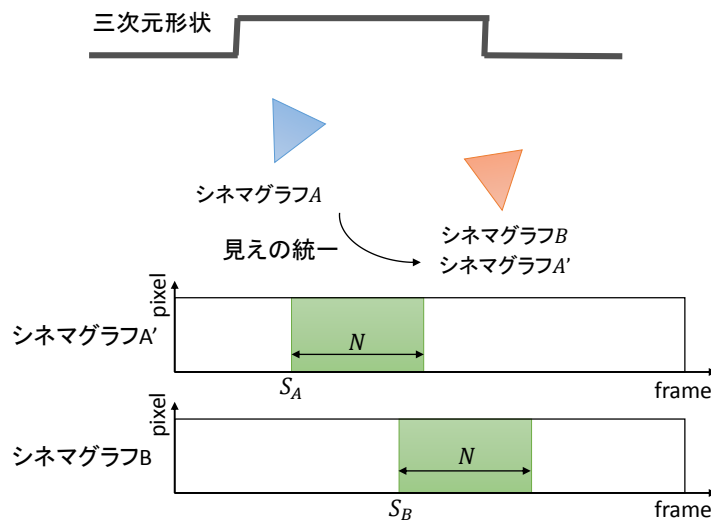


図 15 物体の見えの変化を考慮した類似シーケンス探索



### 3.3.2 シネマグラフ間を接続する中間視点画像の生成

シネマグラフ間を接続する中間視点画像生成について説明する。シネマグラフ間の滑らかな視点移動を実現するために、シネマグラフ間で見えの変化が類似するシーケンスを用い、自由視点画像生成手法に基づいて中間視点画像を生成する。中間視点画像  $M_n (1 \leq n \leq N-1)$  の位置姿勢は、シネマグラフ  $(A, B)$  を生成した位置姿勢を線形補間することで決定する。位置については、単純に線形補間をすることで決定できるが、一般に用いられる姿勢を表現する回転行列は正規直行系にならないため正しく線形補間することはできない。提案手法では、クォータニオンによる姿勢表現で線形補間を行う。クォータニオンは、姿勢  $\mathbf{Q}$  を4次元で表現する。中間視点画像  $M_n$  の姿勢  $\mathbf{Q}_n$  は、シネマグラフ  $(A, B)$  を生成したフレームの姿勢  $(\mathbf{Q}_A, \mathbf{Q}_B)$  を用いて次式で求められる。

$$\mathbf{Q}_n = \frac{\sin(1 - \frac{n}{N})\omega}{\sin \omega} \mathbf{Q}_A + \frac{\sin \frac{n}{N}\omega}{\sin \omega} \mathbf{Q}_B$$

姿勢  $(\mathbf{Q}_A, \mathbf{Q}_B)$  の角度  $\omega$  はクォータニオンの内積から

$$\omega = \cos^{-1}(\mathbf{Q}_A \cdot \mathbf{Q}_B)$$

と計算できる。なお、演算子  $\cdot$  は内積を意味する。

次に、線形補間によって決定した中間視点画像  $M_n$  の位置姿勢とシネマグラフ  $(A, B)$  を生成したフレームの位置姿勢情報を用いて中間視点画像  $M_n$  を生成する。中間視点画像は、生成する中間視点に応じてシネマグラフ  $(A, B)$  で見え方の近い画素を選択する VDTM [20] により生成する。具体的には、シネマグラフ  $(A, B)$  を生成した位置姿勢と中間視点画像  $M_n$  の位置姿勢から注目画素へ引いたベクトルの成す角  $(\theta_A, \theta_B)$  が最も小さい画素の画素値を選択する (図 16)。ただし、ベクトルの成す角が小さいという判断基準のみでは、オクルージョンの影響により注目画素の三次元点がある位置姿勢から観測できない場合があり、誤った画素が選択されることがある (図 17)。そこで、単一視点動画生成におけるデプステストと同様に、中間視点画像  $M_n$  の視点から点  $\mathbf{P}$  における可視性を判断し中間視点画

像を生成することで、シネマグラフ間の接続を行う。ただし、中間視点画像  $M_n$  の生成に用いるシネマグラフ  $(A, B)$  のフレームは、3.3.1 項で探索した類似シーケンスのフレーム  $(I_{A_{S_A+n}}, I_{B_{S_B+n}})$  である。

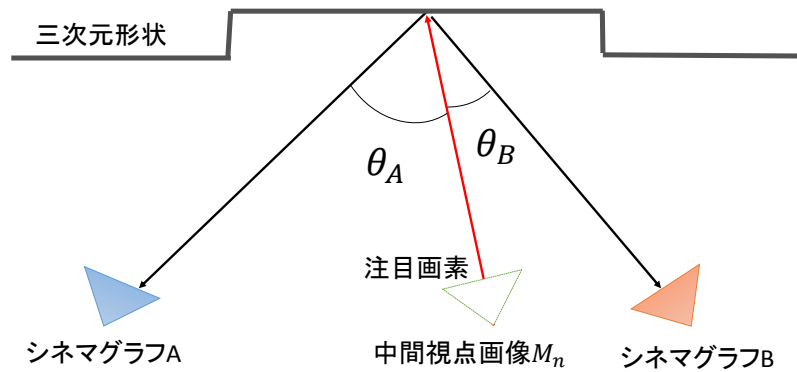


図 16 画素に対するシネマグラフの選択基準

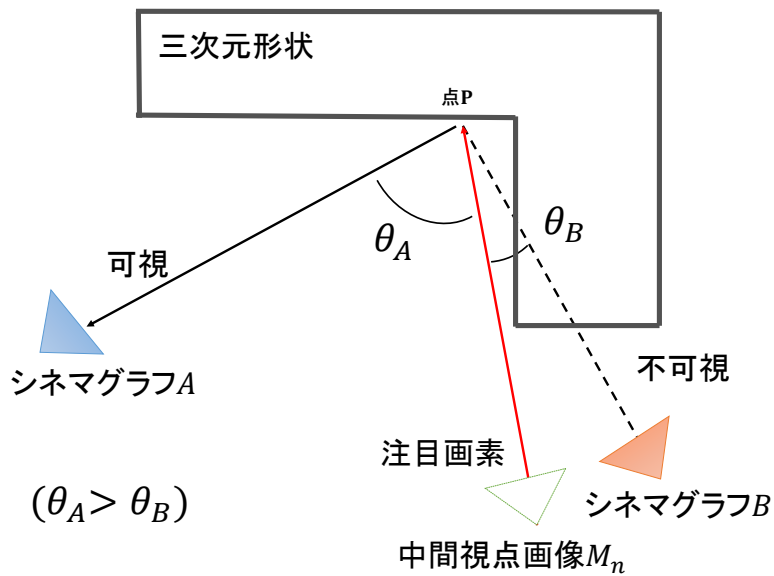


図 17 誤った画素が選択される場合

## 4. 実験

本章では，屋外環境で移動撮影した動画像を用いたシネマグラフの生成およびシネマグラフ間を接続する中間視点画像の生成結果を確認し，提案手法の有効性を示す．以下では，まず単眼動画像を用いた実験について述べ，次に，全方位カメラ映像を用いて生成したシネマグラフに基づくストーリービューのプロトタイプシステムについて述べる．

### 4.1 単眼動画像を用いたシネマグラフの生成実験

#### 4.1.1 実験環境

入力動画像は，水の流れる水路を移動撮影した解像度が  $640 \times 360$  画素の 300 フレームの画像群からなる．図 18 に，入力動画像の一部を示す．また，図 19 に示す 248 フレーム目と 268 フレーム目の画像を対象フレームとして設定し，これらのフレームの視点に変換した単一視点動画像を用いてシネマグラフを生成した．シネマグラフの生成に用いたフレームは，対象フレームとその前後 30 フレーム（計 61 フレーム）とした．また，シネマグラフ間の視点移動の際に生成する中間フレーム数を 20 枚と設定した．

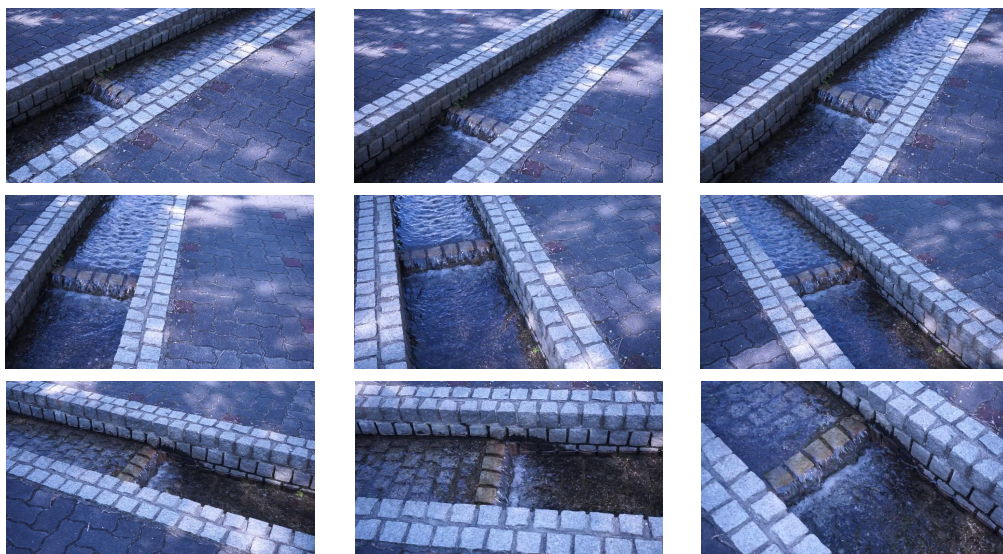


図 18 入力動画像の一部 (単眼動画像)



(a) 248 フレーム目

(b) 268 フレーム目

図 19 シネマグラフを生成する視点

#### 4.1.2 実験結果およびその評価と考察

図 20 に VisualSFM と CMPMVS を用いて復元されたカメラ位置姿勢と三次元形状を、図 21 にシネマグラフ内で動領域とする箇所を人手で指定した領域 (図中: 赤) を示す。図 22 および図 23 に入力動画像を 248 フレーム目および 268 フレーム目の視点に視点変換した単一視点動画像をそれぞれ示し、図 24 および図 25 にこれらの視点で生成したシネマグラフの 1 フレームと図中の一部領域の拡大図を時系列に並べたものを示す。

図 23 より、遮蔽によってテクスチャが得られなかった領域以外については、移動撮影した動画像から単一視点動画像が概ね違和感なく生成できていることが分かる。ただし、255 フレーム目の視点変換画像においては、実際には水路となる領域にタイルのテクスチャがマッピングされていることが確認できる (図 23(b))。これは、復元した三次元形状に復元誤差が含まれることが原因である。図 25 は、このような不自然なテクスチャがマッピングされたフレームを含む単一視点動画像を用いてシネマグラフを生成した結果であるが、生成されたシネマグラフでは、実際には水路である領域にタイルのテクスチャが選択された領域を含むフレームが存在した場合でも、水の流れる水路のテクスチャが選択された。また、泡の流れがループのつなぎ目において不自然な動きをせず、違和感のない動画像が生成されていることが確認できる。

図 26 に 248 フレーム目の視点と 268 フレーム目の視点に対して生成したシネマグラフ間をつなぐ中間視点画像の生成結果と、VDTM で選択されたシネマグラフを表したマップを示す。同図中、赤は 248 フレーム目の視点で生成したシネマグラフを用いた領域、緑は 268 フレーム目の視点で生成したシネマグラフを用いた領域である。生成した動画像より、248 フレーム目および 268 フレーム目において撮影範囲内の領域については、物体の運動を再現した滑らかな視点移動が再現されたことを確認した。ただし、中間視点においてどちらのフレームにも撮影されていない領域がフレーム内に存在する場合には、図 26(b), (c) に示すような黒色の未撮影領域が中間視点に現れるという問題が生じた。一般的な単眼動画像からのシネマグラフ生成において、この問題を解決するためには、生成したシネマグラフに含まれない他のフレームを用いた補間が必要となる。

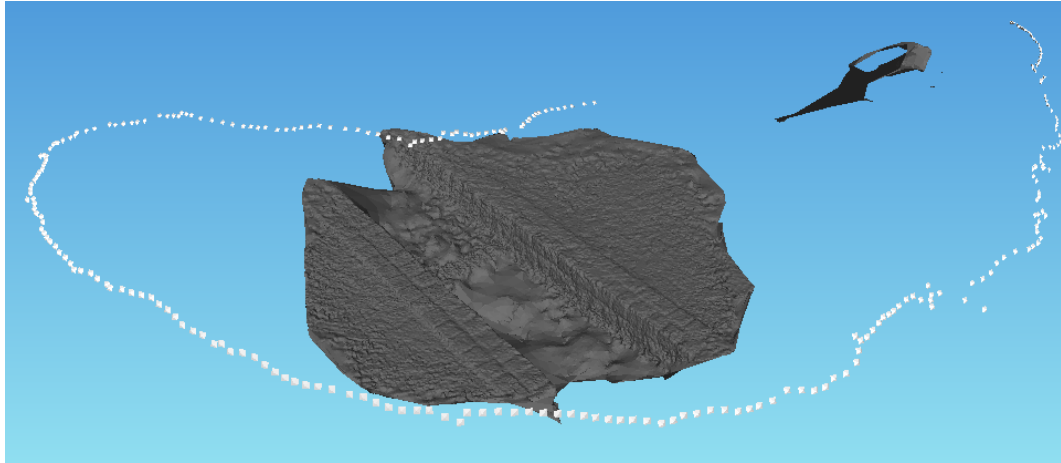


図 20 カメラ位置姿勢推定結果と三次元復元結果 (単眼動画像)

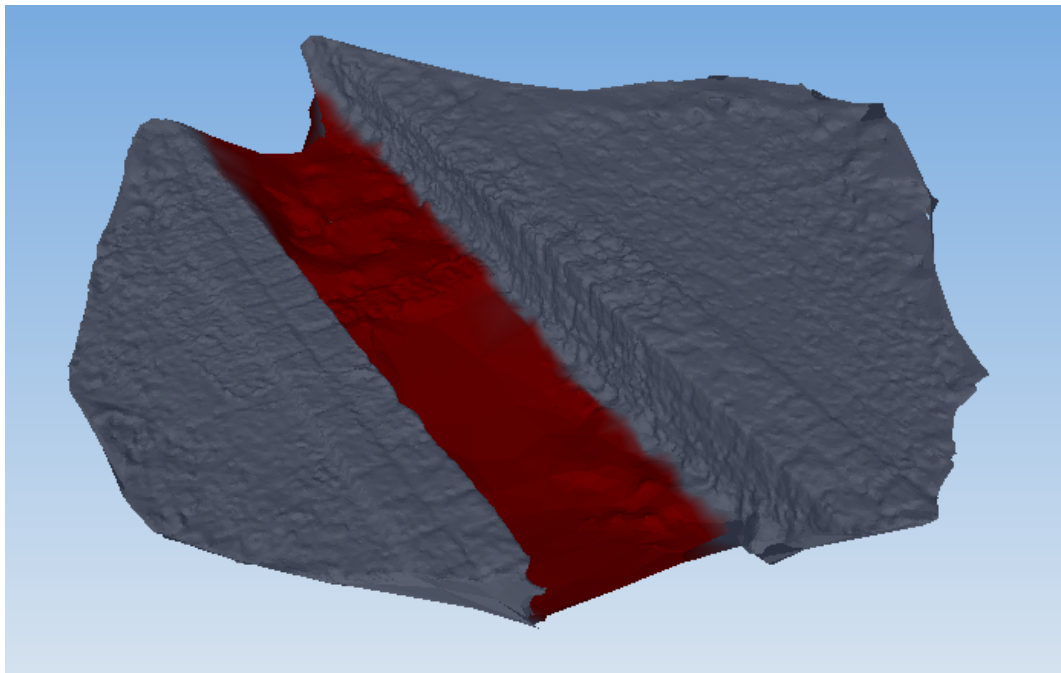


図 21 指定した動領域 (単眼動画像)



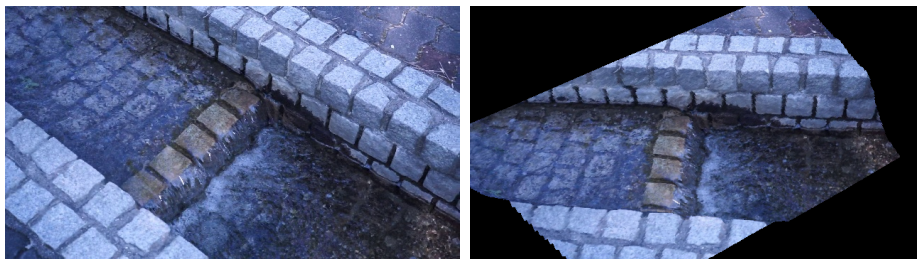
(a) 対象フレーム (248 フレーム目)



(b) 235 フレーム目



(c) 253 フレーム目



(d) 269 フレーム目

(左) 入力フレーム

(右) 視点変換画像

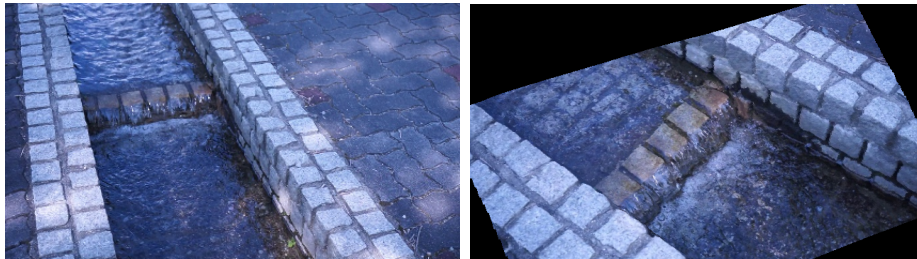
図 22 単一視点動画像生成結果 (248 フレーム目視点)



(a) 対象フレーム (268 フレーム目)



(b) 255 フレーム目



(c) 280 フレーム目



(d) 296 フレーム目

(左) 入力フレーム

(右) 視点変換画像

図 23 単一視点動画像生成結果 (268 フレーム目視点)



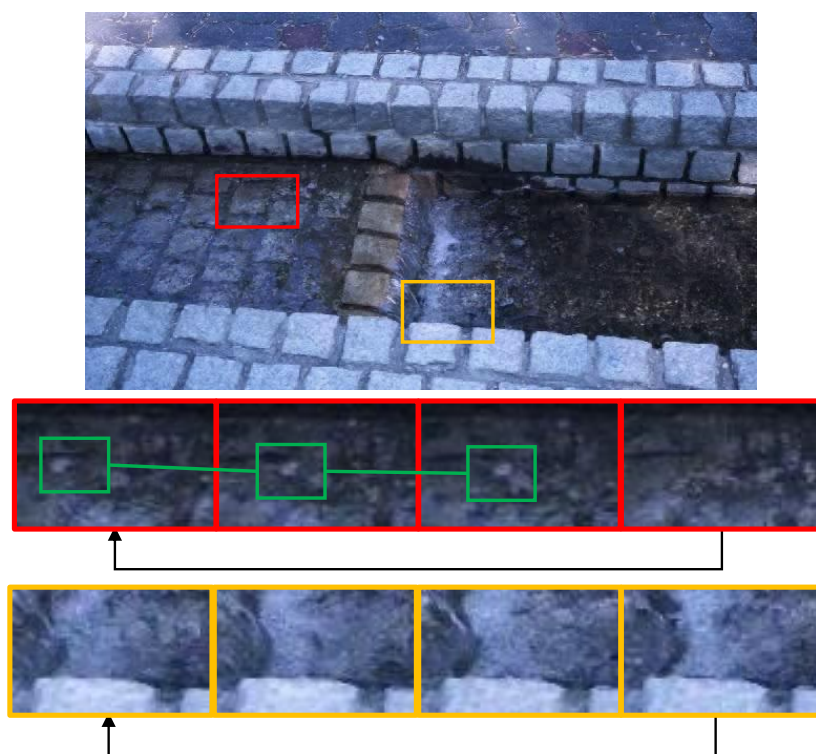


図 24 シネマグラフ生成結果 (248 フレーム目視点)

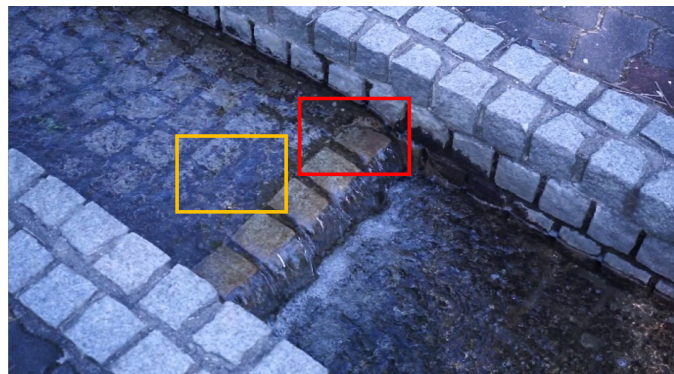


図 25 シネマグラフ生成結果 (268 フレーム目視点)



(a) 1 フレーム目



(b) 9 フレーム目



(c) 13 フレーム目



(d) 19 フレーム目

(左) 中間視点画像

(右) 使用されたシネマグラフ

図 26 中間視点画像生成結果 (単眼動画像)

## 4.2 シネマグラフに基づくストリートビューのプロトタイプシステムの構築

移動撮影した動画像からのシネマグラフ生成手法および自由視点画像生成に基づく中間視点画像によるシネマグラフ間の視点移動手法を全方位カメラで撮影した動画像に適用してシネマグラフを生成し，それに基づくストリートビューのプロトタイプシステムを構築し，評価を行った。

### 4.2.1 プロトタイプシステムの構成

本プロトタイプシステムは，移動撮影した全方位動画像に提案手法を適用することで物体の動きを再現しながら視点移動可能なシネマグラフを生成し，それに基づくストリートビュー提示を行う．ストリートビュー提示として，シネマグラフを生成した視点位置での全方位の見渡しおよび，隣接する視点位置で生成したシネマグラフへ視線固定状態での視点移動が本プロトタイプシステムで可能である．表 1 にプロトタイプシステムの構築に用いたコンピュータの仕様を示す。

表 1 プロトタイプシステム構築に用いたコンピュータ仕様

|        |   |
|--------|---|
| OS     | Windows 7 Professional (64bit)          |
| CPU    | intel(R) Core(TM) i7-4770 CPU @ 3.40GHz |
| GPU    | -                                       |
| Memory | 16GB                                    |

### 4.2.2 実験環境

本実験では，全方位カメラ Ladybug3 を用いて，奈良県佐保川の桜が揺れる道を移動撮影した全方位動画像の 200 フレームを入力とした．図 27 に示す全方位カメラの各カメラが撮影した画像のレンズ歪みを補正した解像度が  $616 \times 808$  画素の画像 1200 枚を VisualSFM および CPMVMS の入力としてカメラ位置姿勢推定と三次元復元を行った．表 2 に三次元復元に用いたコンピュータの仕様を示す。

表 2 三次元復元に用いたコンピュータ仕様

|        |   |
|--------|---|
| OS     | Windows 7 Professional (64bit)          |
| CPU    | intel(R) Core(TM) i7-4770 CPU @ 3.40GHz |
| GPU    | NDIVA GeForce GTX770                    |
| Memory | 16GB                                    |

入力画像の 40 フレーム目と 100 フレーム目を対象フレームとして設定し、対象フレームの前後 30 フレーム（計 61 フレーム）を対象フレーム視点の全方位画像に変換する。ここでは、全方位カメラの各カメラ画像を  $512 \times 512$  画素の 6 枚の正方形画像をつなげた Cubemap に変換したものを全方位画像とする。これに対して、Liao らの手法 [24] を適用することで、全方位シネマグラフを生成した。また、シネマグラフ間の視点移動に要するフレーム数を 60 枚として、中間視点画像を生成した。使用したコンピュータ仕様は表 1 と同様である。また本実験では、シネマグラフに基づいたストリートビューのプロトタイプシステムを評価するために、シネマグラフによる映像提示を行った場合と対象フレームの単一視点動画の繰り返し再生による映像提示を行った場合においてプロトタイプシステムの比較を行った。

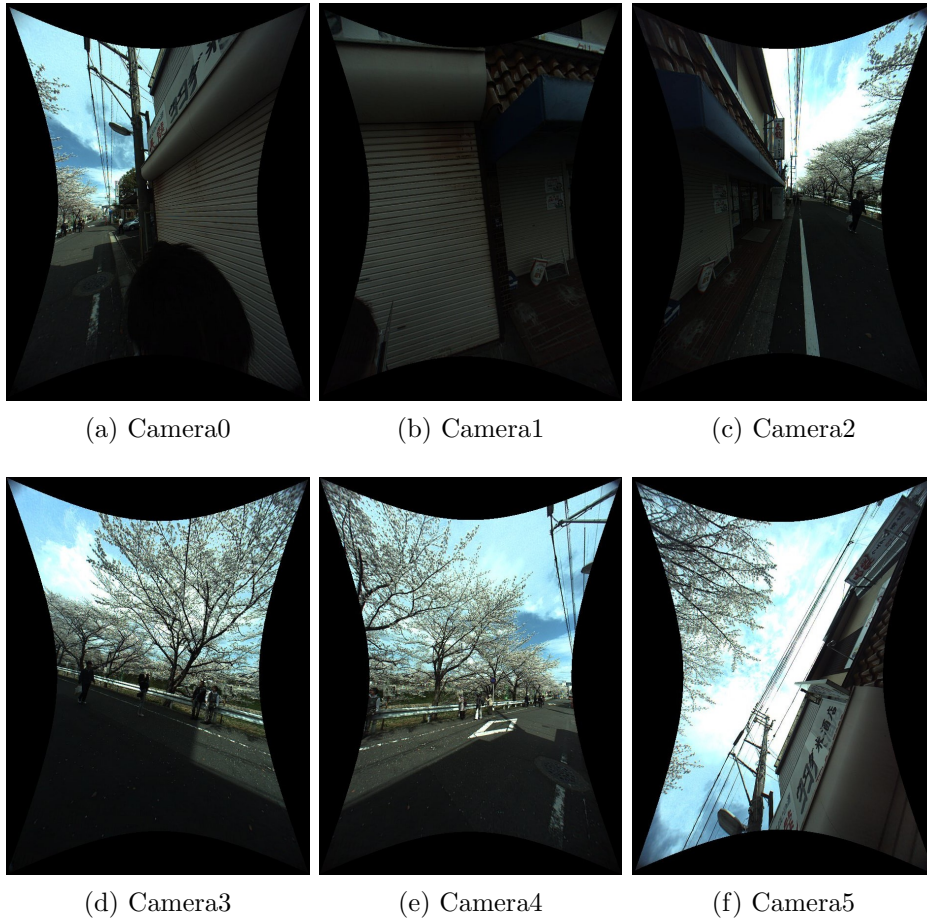


図 27 各カメラの歪み補正画像 (入力画像)

### 4.2.3 実験結果およびその評価と考察

表 3 に提案手法を実装した本実験環境でのプロトタイプシステムの生成に要した計算時間を示す。

表 3 プロトタイプシステム構築の計算時間

|                      |                 |
|----------------------|-----------------|
| 移動撮影した動画像からのシネマグラフ生成 | 10 時間 41 分 42 秒 |
| 自由視点画像生成に基づく中間視点画像生成 | 15 分 24 秒       |

図 28 にカメラ姿勢推定結果と三次元復元結果を，図 29 に指定した動領域を示す。図 30 に 40 フレーム目の視点位置において生成した CubeMap，図 32 に 100 フレーム目の視点において生成した CubeMap をそれぞれ展開した図を示す。また，図 31 および図 33 に他フレームを対象フレーム視点の CubeMap に変換し展開した結果をそれぞれ示す。底面については，全方位カメラで撮影できない領域のため表示していない。図 34 に生成したストリートビューのプロトタイプシステムにおける物体の動きの再現結果を示す。

図 34 より，対象フレームの単一視点動画を繰り返し映像提示をしたストリートビューとシネマグラフによる映像提示をしたストリートビューを主観的に比較した結果，シネマグラフによる映像提示をしたストリートビューの方が物体の動きの時間的連続性を保った映像を提示しており，違和感のない映像を生成できたことが確認できる。これにより，シネマグラフはストリートビューにて動きを再現する有効な表現方法であると言える。一方，視点移動を行ったとき，テクスチャに不連続が生じる問題があることを確認した。図 35 において，静止領域であるガードレールが歪んでいることが分かる。これは，三次元形状を復元できていない領域があることが原因だと考えられる。また，桜が揺れる動領域では，VDTM によって選択されるシネマグラフが切り替わるタイミングで物体の動きに不連続が生じた。移動撮影した全方位カメラ映像からある視点位置でのシネマグラフを生成することには成功したが，視点移動についてはテクスチャに不連続が発生する課題が残った。

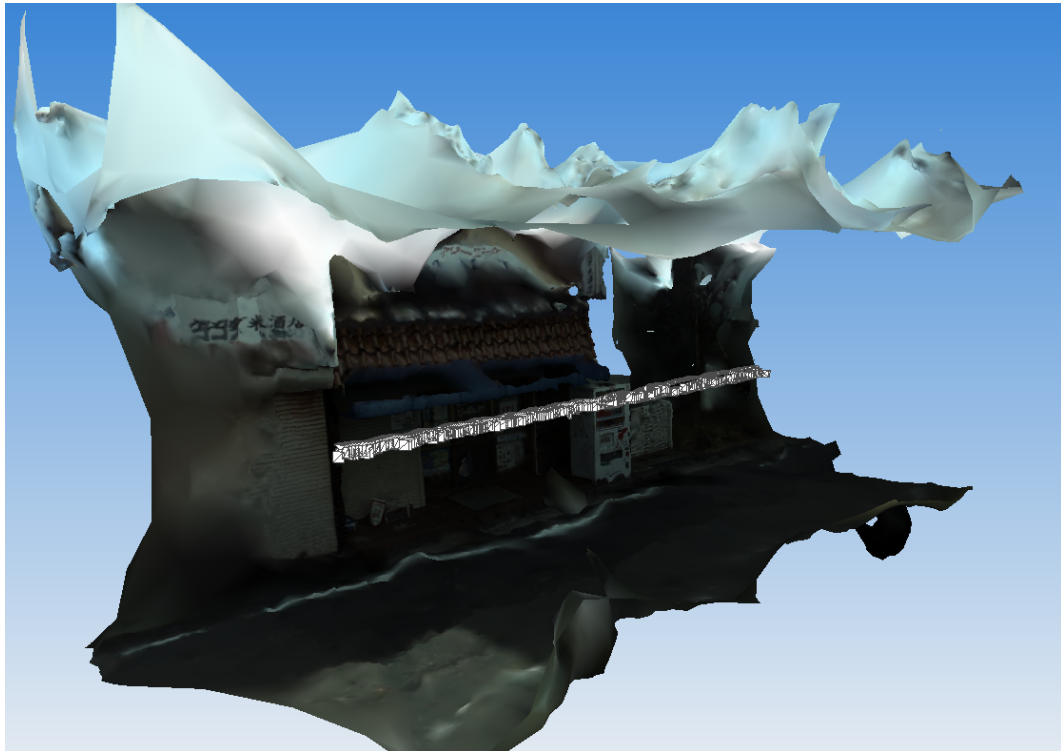


図 28 カメラ位置姿勢推定結果と三次元復元結果（全方位動画像）

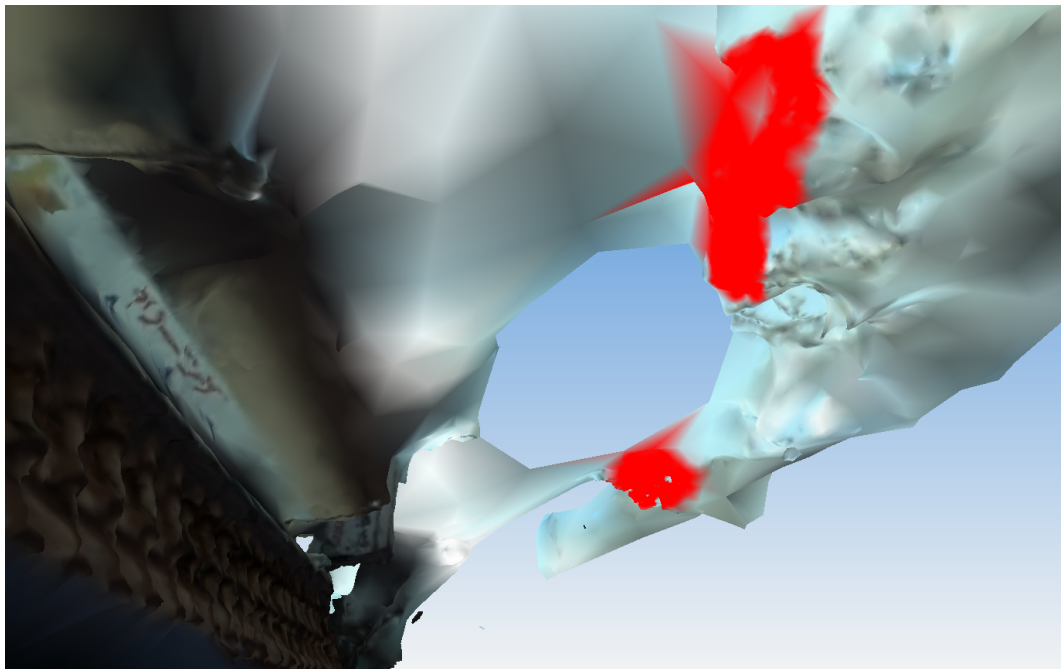


図 29 指定した動領域（全方位動画像）



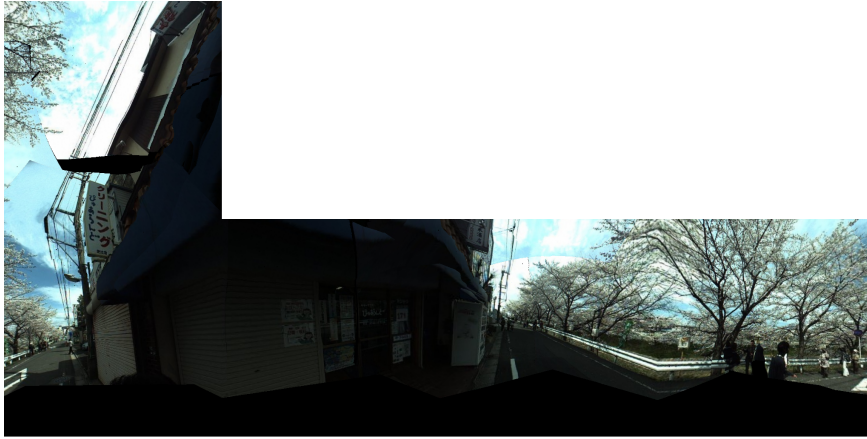
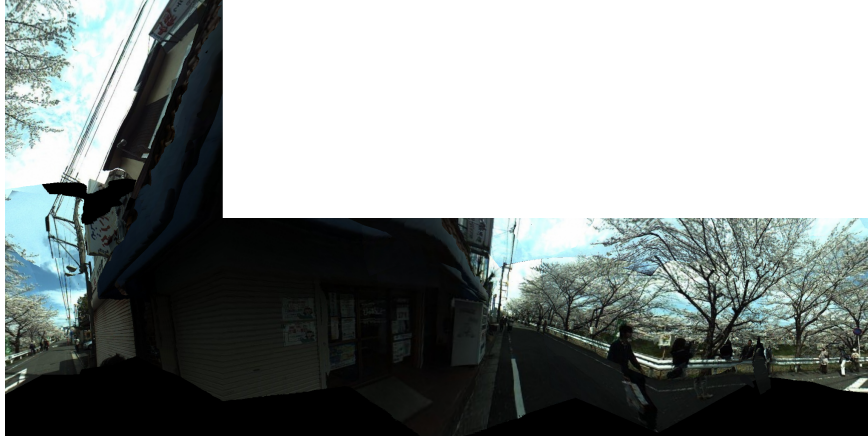
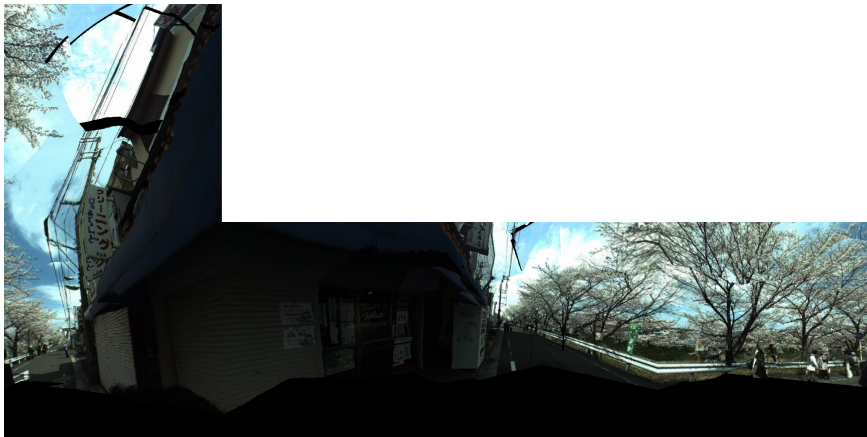


図 30 生成した 40 フレーム目における CubeMap



(a) 20 フレーム目

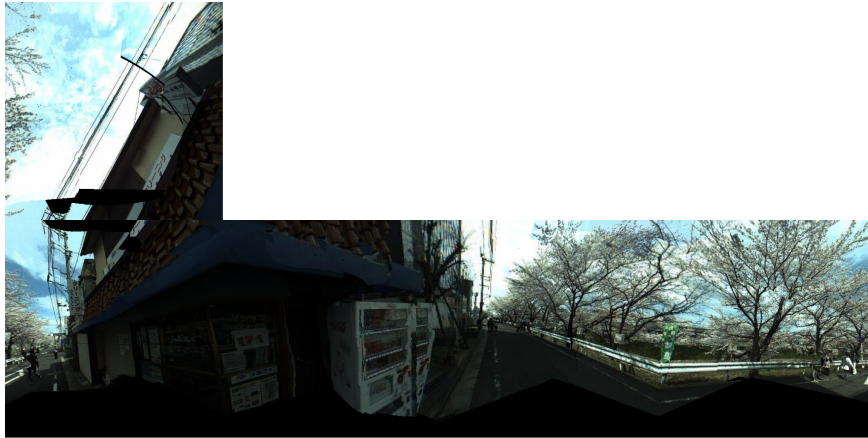


(b) 60 フレーム目

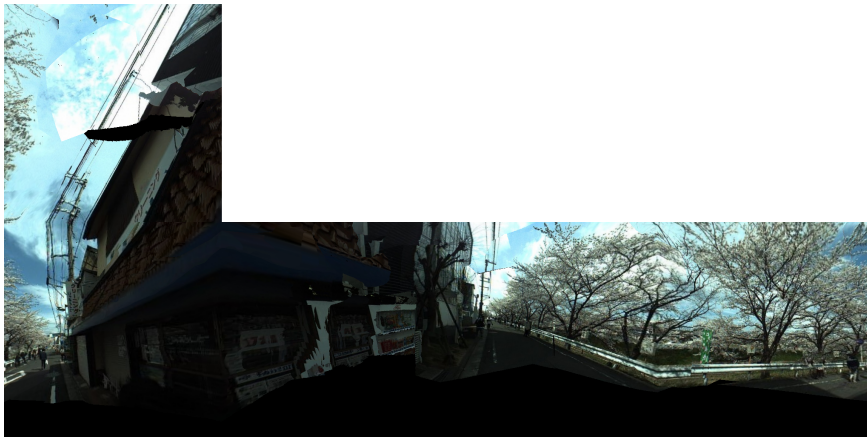
図 31 対象フレーム視点 (40 フレーム目) への変換結果



図 32 生成した 100 フレーム目における CubeMap



(a) 80 フレーム目



(b) 120 フレーム目

図 33 対象フレーム視点 (100 フレーム目) への変換結果

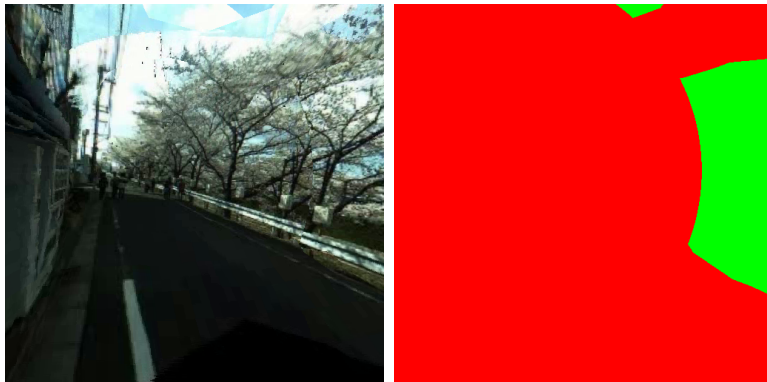


(a) シネマグラフ

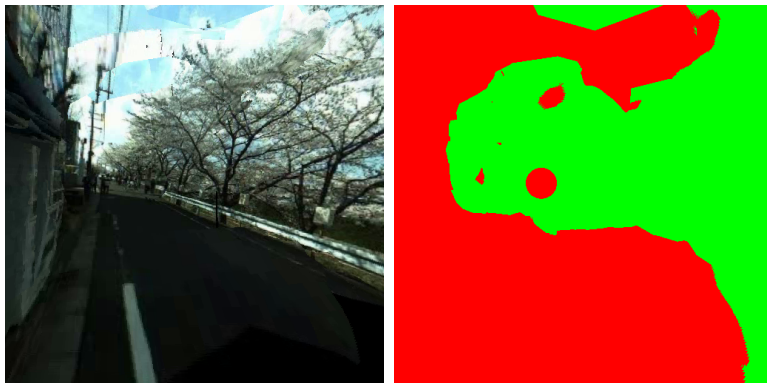


(b) 単一視点動画像

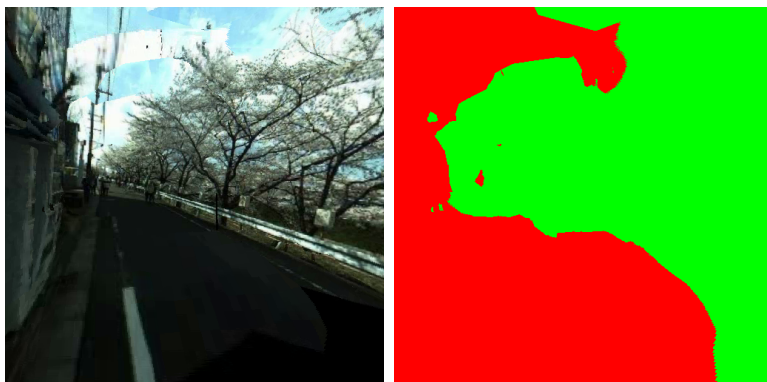
図 34 動画の開始フレームと終了フレームの関係



(a) 32 フレーム目



(b) 34 フレーム目



(c) 36 フレーム目

(左) 中間視点画像      (右) 使用されたシネマグラフ

図 35 中間視点画像生成結果 (全方位動画像)

## 5. まとめ

本論文では、移動撮影した動画像からある視点位置での動きを再現するシネマグラフを生成する手法および、移動撮影した動画像から生成したある地点でのシネマグラフを用いて撮影地点間の滑らかな視点移動を実現する手法を提案した。また、提案手法を用いて、移動撮影した動画像から物体の動きを再現しながら視点移動可能なシネマグラフに基づくストーリービューのプロトタイプシステムを構築し評価した。単眼動画像を用いた実験では、移動撮影した動画像からシネマグラフを生成した上で、隣接するシネマグラフ間の視点移動に対応する中間視点画像を生成し、違和感のないループ画像の生成と視点移動が実現できることを示した。また、提案手法を用いて動きを再現するストーリービューのプロトタイプシステムを全方位カメラ映像から構築した結果、特定の視点位置でのシネマグラフによる動きを再現したストーリービューの構築が可能であることを示した。ただし、隣接する視点位置で生成したシネマグラフ間の接続においては、動物体に対応するシーンの形状が正しく復元されなかったことに起因して、生成した中間視点画像にテクスチャの不連続が生じ、十分な結果を得られなかった。

今後の課題として、三次元形状が復元されなかった部分の補間方法、中間視点画像中のテクスチャの不連続を抑止する方法の検討が考えられる。

## 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，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。研究遂行にあたりデータ収集などに協力していただいた視覚情報メディア研究室の方々に深く感謝いたします。

## 参考文献

- [1] 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形状”, 画像電子学会誌, Vol. 34, No. 5, pp. 529-538 (2005)
- [2] 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”, 電子情報通信学会論文誌 D-II, Vol. J83-D-II, No. 9, pp. 1962-1965 (2000)
- [3] H. Kawasaki, Y. Ohsawa, R. Furukawa, and Y. Nakamura. ”Dense 3d Reconstruction with an Uncalibrated Active Stereo System”, *Proc. 7th Asian Conf. Computer Vision, ACCV' 06*, Vol. 2, pp. 882-891(2006)
- [4] R. Y. Tsai”An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision”, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 364-374 (1986)
- [5] M. Goesele, B. Curless, and S. M. Seitz.”Multi-View Stereo Revisited”, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 2402-2409 (2006)
- [6] C. Hernandez, G. Vogiatzis, and R. Cipolla.”Probabilistic Visibility for Multi-View Stereo”, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 8 pages (2007)
- [7] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski.”Towards Internet-scale Multi-view Stereo”, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 1434-1441 (2010)
- [8] H. Baker.”Three-dimensional Modeling”, *Proc. Int. Joint Conf. Artificial Intelligence*, Vol. 2, pp. 649-655 (1977)

- [9] 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)
- [10] S. Chen and L. Williams. "View Interpolation for Image Synthesis", *Proc. SIGGRAPH' 93*, pp. 279-288 (1993)
- [11] S. Seitz and C. Dyer. "View Morphing", *Proc. SIGGRAPH' 96*, pp. 21-30 (1996)
- [12] K. Kato, H. Ishiguro and M. Barth. "Town Digitizing: Recording of Street Views by Using ODVS & GPS", *Proc. IEEE Intelligent Vehicles Symposium*, pp. 183-188 (2001)
- [13] T. Stich, C. Linz, G. Albuquerque and M. Magnor. "View and Time Interpolation in Image Space", *Computer Graphics Forum*, Vol. 27, No. 7, pp. 1781-1787 (2008)
- [14] E. Adelson and J. Bergen. "The Plenoptic Function and The Elements of Early Vision", *In Computational Models of Visual Processing*, pp. 3-20 (1991)
- [15] T. Naemura, T. Takano, M. Kaneko and H. Harashima. "Ray-based Creation of Photo-realistic Virtual World", *Proc. Int. Conf. on Virtual Systems and MultiMedia*, pp. 59-68 (1997)
- [16] 山口 晃一郎, 山澤 一誠, 竹村 治雄, 横矢 直和. "全方位動画像を用いた両眼ステレオ画像の実時間生成によるテレプレゼンス", *電子情報通信学会論文誌*, Vol. 84, No. 9, pp. 2048-2057 (2001)
- [17] T. Takahashi, H. Kawasaki, K. Ikeuchi and M. Sakauchi. "Arbitrary View Position and Direction Rendering for Large-scale Scenes", *Proc. Computer Vision and Pattern Recognition*, Vol. 2, pp. 296-303 (2000)



- [18] W. Matusik, C. Buehler, R. Raskar, S. Gortler and L. McMillan. "Imagebased Visual Hulls", *Proc. SIGGRAPH' 00*, pp. 369-374 (2000)
- [19] 石川 智也, 山澤 一誠, 横矢 直和. "複数の全方位動画像を用いた自由視点テレプレゼンス", *映像情報メディア学会誌*, Vol. 60, No. 10, pp. 1599-1610 (2006)
- [20] P. Debevec, Y. Yu, and G. Borshukov. "Efficient View-dependent Image-based Rendering with Projective Texture-mapping", *Proc. 9th Eurographics Rendering Workshop*, 13 pages (1998)
- [21] 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*, PRMU2009-307 (2010)
- [22] A. Schödl, R. Szeliski, D. Salesin and I. Essa. "Video Textures", *Proc. SIGGRAPH' 00*, pp. 489-498 (2000)
- [23] V. Kwatra, A. Schödl, I. Essa, G. Turk and A. Bobick. "Graphcut Textures: Image and Video Synthesis Using Graph Cuts", *ACM Trans. Graphics (TOG)*, Vol.22, No. 3, pp. 277-286 (2003)
- [24] Z. Liao, N. Joshi and H. Hoppe. "Automated Video Looping with Progressive Dynamism", *ACM Trans. Graphics (TOG)*, Vol. 32, No.4, pp. 77:1-77:10 (2013)
- [25] A. Agarwala, K. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin and R. Szeliski. "Panoramic Video Textures", *ACM Trans. Graphics (TOG)*, Vol. 24, No. 3, pp. 821-827 (2005)
- [26] 馬場口 豊, 桑 秀行, 佐藤 智和, 横矢 直和. "物体運動の時間的・空間的な連続性を保持した視点移動を実現するビデオテクスチャ", *電子情報通信学会技術研究報告*, PRMU2010-262, (2011)
- [27] C. Wu. "VisualSFM: A Visual Structure from Motion System", <http://ccwu.me/vsfm/>, (2011)

- [28] M. Jancosek and T. Pajdla. "Multi-view Reconstruction Preserving Weakly-supported Surfaces", *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 3121-3128 (2011)