

NAIST-IS-MT1551052

修士論文

複数の非同期定点撮影映像を用いた
仮想視点における全方位シネマグラフ生成

白川 誠

2017年3月16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。

白川 誠

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)

複数の非同期定点撮影映像を用いた 仮想視点における全方位シネマグラフ生成*

白川 誠

内容梗概

近年, Google ストリートビューに代表される, 実環境で撮影した全方位画像を入力として, ユーザが指定した自由な方向の映像提示を実現するアプリケーションが普及している. このようなアプリケーションでは全方位動画像を入力として用いることでシーン中の“動き”を再現し, ユーザにより高い臨場感を与えることができる. しかし, 動画像を用いたストリートビュー表示を実現するためには, 地点ごとに長時間の定点撮影を行うことが必要となり, 撮影コストが膨大になる問題がある. また, 動画像の再生完了時には冒頭からの繰り返し再生を行う必要が生じるが, 動画像の最終フレームから初期フレームへの切り替え時に映像の不連続が生じる場合があり, 臨場感が損なわれる. このような問題に対して本論文では, 動画像を用いたストリートビューシステムを構築するための撮影コストの低減および動画像の繰り返し再生における不連続の解消を目的として, 複数の地点で非同期に定点撮影した動画像から任意に定めた視点位置での不連続を抑えた繰り返し動画(シネマグラフ)を生成する手法を提案する. 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. 提案手法では, まず, 複数の地点で異なる時刻に定点撮影した動画像に対してそれぞれシネマグラフを生成する. 次に, 複数画像を用いて対象のシーン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生成したシネマグラフを対象視点からの見えにワーピン

*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-MT1551052, 2017年3月16日.

グする。最後に，ワーピング後のシネマグラフ群から，入力視点の見えるの近さとテクスチャの連続性を考慮してテクスチャ選択を行うことで任意地点でのシネマグラフ生成を実現する。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仮想視点のシネマグラフを生成し，提案手法の有効性を評価する。

キーワード

シネマグラフ，自由視点画像生成，グラフカット

Novel-viewpoint Omnidirectional Cinemagraph Generation Using Multiple Asynchronous Videos Captured at Fixed Points*

Makoto Shirakawa

Abstract

Recently, applications such as Google Street View have been widespread that enable us to see any directions of omnidirectional images taken in real environments. Such applications can enhance the sense of presence by reproducing motions in the scene using omnidirectional videos as input. However, it costs much time to use videos for street views because scenes have to be captured at each point for a long time. In addition, since this application needs to playback videos again from the first frame after the last frame, discontinuities which often occurs between the last and the first frames give users less realistic sensation. In order to reduce the cost for video capturing and ensure the visual continuity when looping videos, this thesis proposes a method to generate a looping video with less temporal discontinuity (referred as to cinemagraph) at an arbitrary viewpoint from videos asynchronously captured at multiple fixed viewpoints. 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 purpose, our method first generates a cinemagraph from each video asynchronously captured at a fixed point. On the basis of 3D shapes of the

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1551052, March 16, 2017.

target scene reconstructed using multiple images, we warp the generated cinemagraphs to the appearance of the target point of view by using novel-viewpoint image rendering. Finally, we generate a novel-viewpoint cinemagraph by selecting a cinemagraph for each pixel from warped cinemagraphs in consideration of the closeness of view directions and the continuity between textures. In experiments, we demonstrate the effectiveness of the proposed method by generating cinemagraphs from videos captured with monocular and omnidirectional cameras.

Keywords:

cinemagraph, novel viewpoint rendering, graph-cut

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)	3
2.1.2 Image-Based Rendering(IBR)	8
2.1.3 Hybrid Rendering(HR)	10
2.2 シネマグラフ生成に関する研究	11
2.3 本研究の位置づけ	14
3. 仮想視点におけるシネマグラフ生成	15
3.1 提案手法の概要	15
3.2 非同期に定点撮影した動画像群からの仮想視点におけるシネマグラフ生成	16
3.2.1 カメラ位置姿勢推定と三次元復元	16
3.2.2 各入力視点のシネマグラフ生成	17
3.2.3 シネマグラフの仮想視点からの見えへの変換	18
3.3 エネルギー関数最小化によるテクスチャ選択	21
3.3.1 仮想視点と入力視点の見えに関する項	21
3.3.2 可視判定に関する重み	22
3.3.3 空間方向の画素値の整合性を考慮した項	23
4. 実験	25
4.1 実験の概要	26
4.2 単眼動画像を用いたシネマグラフ生成実験	26
4.2.1 実験環境	26
4.2.2 実験結果およびその評価と考察	28
4.3 全方位動画像を用いたシネマグラフ生成実験	36
4.3.1 実験環境	36

4.3.2 実験結果およびその評価と考察	38
5. まとめ	43
謝辞	44
参考文献	45

目 次

1	欠損・誤差が生じた三次元モデルを用いた自由視点画像生成例 . . .	4
2	ライン光投影法 [1, 2]	5
3	ToF カメラを用いた測距法 [3, 4]	5
4	ステレオ法の原理	7
5	視体積交差法 [5]	7
6	View Morphing による視点内挿 [6]	9
7	光線情報の補間方法の違いによる生成画像の違い ((左)Davis ら [7] の手法, (右)Lipski ら [8] の手法).	10
8	debevec ら [9] の VDTM	11
9	Liao ら [10] の手法 ((左) 入力の 1 フレームと (右) 開始フレームと ループ周期)	12
10	Liao ら [11] の手法によるシームの低減	13
11	中島ら [12] の手法	13
12	提案手法概要	16
13	推定されたカメラ位置姿勢と対象シーンの密な三次元形状	17
14	カメラ位置姿勢と三次元形状による画素の対応付け	19
15	デプステストによる可視領域の判定	19
16	視点変換した画像例	20
17	見えの近さによる画素の選択	22
18	入力視点位置からの不可視領域	23
19	三次元形状復元のための入力画像群の一部 (噴水のシーン)	27
20	三次元形状復元のための入力画像群の一部 (木のシーン)	27
21	VisualSFM によるカメラ位置姿勢と推定された三次元位置 (噴水の シーン)	29
22	CMPMVS により復元した三次元形状 (噴水のシーン)	29
23	VisualSFM によるカメラ位置姿勢と推定された三次元位置 (木の シーン)	30
24	CMPMVS により復元した三次元形状 (木のシーン)	30

25	視点変換動画像 (噴水のシーン)	31
26	視点変換動画像 (木のシーン)	31
27	動画像中のある 1 フレームにおける空間的整合性の比較 (噴水のシーン)	32
28	動画像中のある 1 フレームにおける空間的整合性の比較 (木のシーン)	33
29	動画像中の前後 5 フレームにおける時間的整合性の比較 (噴水のシーン)	34
30	動画像中の前後 5 フレームにおける時間的整合性の比較 (木のシーン)	35
31	画角 120 度のキューブマップの展開例	36
32	三次元形状復元のためのキューブマップ画像群	37
33	VisualSFM によるカメラ位置姿勢と推定された三次元位置	39
34	CMPMVS により復元した三次元形状	39
35	視点変換後の全方位画像	40
36	エネルギー最小化による全方位動画像生成結果	41
37	生成したシネマグラフにおけるテクスチャの整合性	41
38	各ラベルにおける動物体の動きの大きさの違い	42

表 目 次

1	実験に用いた PC の仕様	25
2	三次元形状復元に用いた PC の仕様	25
3	入力画像画像群と撮影地点数	27

1. はじめに

近年，Google ストリートビューに代表される，景色の眺望や遠隔地の仮想的な訪問等を目的とした，ユーザに実際にその場にいるかのような没入感覚を与えるアプリケーションが普及している．このようなアプリケーションでは，車載全方位カメラを用いて様々な地点を移動撮影し収集した全方位画像を入力とすることで，ユーザが指定した自由な方向の画像提示を実現している．これらのシステムでは静止画像を用いているため，歩行者・自動車・揺れる木の葉などの動物体も静止画像として提示されるが，景観の眺望や観光地の仮想的な訪問などを目的とした場合，全方位動画像を入力として用いてシーン中の“動き”を再現し，ユーザに対してより高い臨場感を与えることが望ましい．しかし，動画像を用いたストリートビュー表示を単純な移動撮影のみで実現しようとした場合には，以下のような2つの問題点がある．

- (1) 映像を提示したい地点ごとにカメラを静止させ長時間の定点撮影を行うことが必要となり，撮影コストが膨大になる．
- (2) 動画像の再生完了時には冒頭からの繰り返し再生を行う必要が生じるが，動画像の最終フレームから初期フレームへの切り替え時に映像の不連続が生じる場合があり，臨場感が損なわれる．

これらの問題に対して本論文では，動画像を用いたストリートビューシステムを構築するための撮影コストの低減および動画像の繰り返し再生における不連続の解消を目的として，複数の地点で非同期に定点撮影した動画像から任意に定めた視点位置での不連続を抑えた繰り返し動画（シネマグラフ）を生成する手法を提案する．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．提案手法ではまず，複数の地点で異なる時刻に定点撮影した動画像に対してそれぞれシネマグラフ生成手法 [11] を適用する．次に，対象シーンを撮影した複数画像に対して三次元形状復元手法 [13] を適用し対象のシーン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生成したシネマグラフを対象視点からの見えにワーピングする．最後に，ワーピング後のシネマグラフ群から，入力視点

の見える近さとテクスチャの連続性を考慮してテクスチャ選択を行うことで任意地点でのシネマグラフ生成を実現する。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仮想視点のシネマグラフを生成し、提案手法の有効性を評価する。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、定点撮影した動画像群から任意視点でのシネマグラフを生成する手法について述べる。4章では、提案手法の有効性を示すために、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、動画像を用いたストリートビューシステムを構築するための撮影コストの低減および動画像の繰り返し再生における不連続の解消である。本章では他視点の画像情報を用いてコストを低減する自由視点画像生成に関する従来研究、動画像の繰り返し再生における不連続の解消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

自由視点画像生成手法に関する従来研究は、対象環境の三次元形状を復元し用いる Model-Based Rendering(MBR), 対象環境を撮影した画像群を変形・合成することで自由視点画像を生成する Image-Based rendering (IBR), MBR と IBR を組み合わせた Hybrid rendering(HR) に大別できる。

2.1.1 Model-Based Rendering(MBR)

MBR は、自由視点画像生成を行う対象環境のテクスチャ付き三次元モデルをあらかじめ生成し、その幾何形状およびテクスチャを任意に設定した視点位置へ射影することで画像を生成する手法である。対象環境の形状を利用するため、視点移動に関する自由度や、視点移動に伴う物体の隠蔽関係の表現に優れている。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、生成した画像に欠損や歪みが生じる(図1)。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。

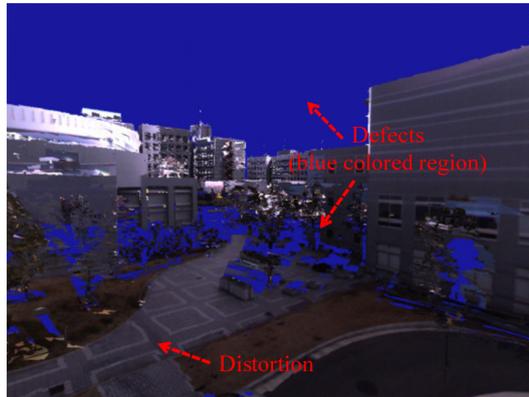


図 1 欠損・誤差が生じた三次元モデルを用いた自由視点画像生成例

【能動的な三次元形状復元手法】

投光器を用いた光の投影による見えの差異を計測し三次元推定を行うアクティブステレオ法 [1, 2] や、Time-of-Flight(ToF) カメラ等のレーザレンジファインダ(LRF) を利用するレーザ測距法 [3, 4] 等がある。アクティブステレオ法は、その投影する光の種類により、様々な手法が知られている。その中でスリット光を利用するライン光投影法について紹介する。ライン光投影法の原理を図2に示す。光を透過するスリット上のある位置対し、その位置に対する画像中のエピポーラ線と、画像から検出されるスリット光が投影された線の交点から、三次元の位置が計算される。さらにスリットの位置を左右にずらすことで、全体にわたる計測を行う。スリット光の投影方式として、実際に光源の前で移動するスリットを用る手法や、プロジェクタを用いて直線上の光を移動させながら投影する手法 [1, 2] がある。アクティブステレオ法は、受動的なステレオ法のように画像どうしのマッチングを必要としないため、より安定に三次元計測が行える一方、照射した光の反射光を利用するため、適用できる対象や環境が限定されている。

ToF カメラを利用するレーザ測距法 [3, 4] は、光が光源からカメラに到達する時間から物体の形状を計測する、光の位相差を利用した手法である(図3)。具体的には、投光パルスを高速で点滅させ、反射光の位相遅れの程度を計測することで距離計測を行う。これらの手法は、物体の形状を高精度に取得可能であるが、LRF 等の特殊な機材を用いるため非常に高価である。また、広範囲において計測

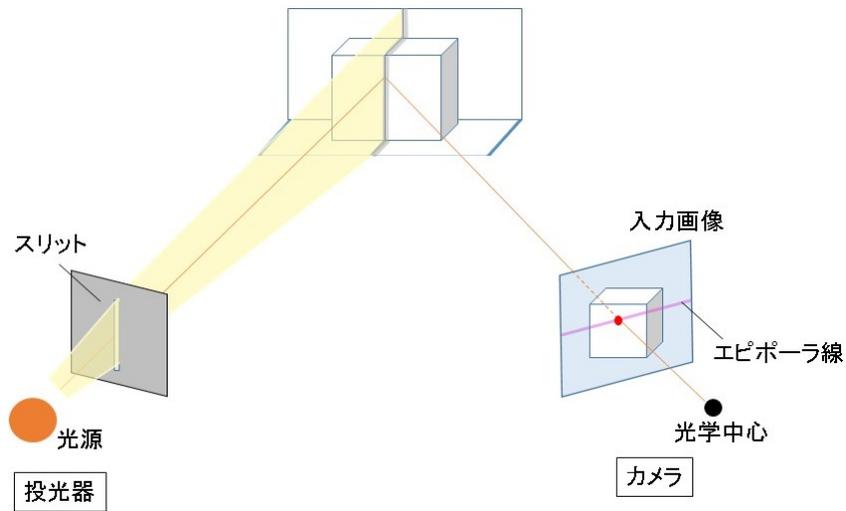


図 2 ライン光投影法 [1, 2]

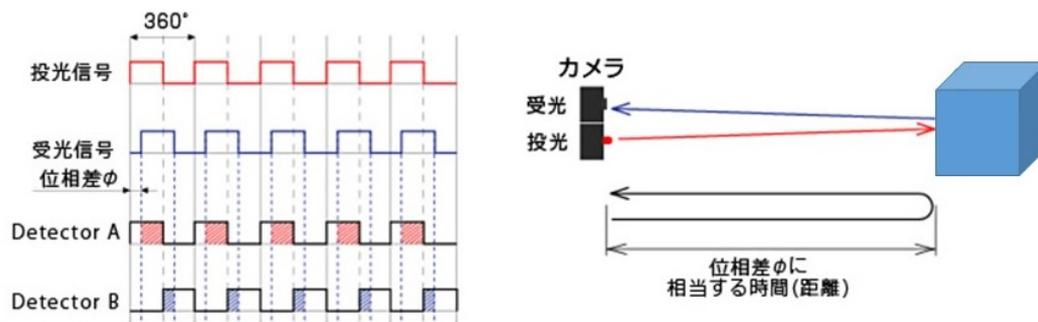


図 3 ToF カメラを用いた測距法 [3, 4]

を行う場合，何回も機材を設置し直す必要があり，撮影コストが掛かる．

【受動的な三次元形状復元手法】

対象環境を撮影した画像群から三次元形状を推定する受動的な手法として、ステレオ法 [13, 14] や視体積交差法 [5] 等がある。これらの手法では、撮影されたカメラの外部パラメータが既知である複数の画像を用いる必要があり、推定には Structure from Motion(SfM)[15] が広く用いられている。

ステレオ法は、異なる複数の視点から撮影された二次元画像上で物体表面上の点の対応づけを行い、三角測量によって対象までの距離を求め、三次元形状を復元する手法である (図 4)。利用する画像が 3 枚以上の場合は特に多視点ステレオ法と呼ぶ。代表例として、Furukawa ら [14] の、多数の画像を入力として光学的整合性を高めることで、インターネット上の大量の画像による密な三次元復元を行う手法挙げられる。ただし、テクスチャが少ない箇所では画像間での対応が付かず、生成モデルに欠損が生じる (図 4)。これに対して、Jancosek ら [13] は、推定された奥行きが疎な点群に対して面を貼り付けて密な三次元復元を行い、欠損の少ないモデルを生成できる手法を提案した。しかし、この手法においてもテクスチャの少ない箇所では形状の推定精度が必ずしも良くない。

視体積交差法 [5] は、撮影空間をボクセルと呼ばれる立方体の集合として表現し、そのボクセル空間内で多地点でのカメラの光学中心とシルエット画像から成る視体積の積集合空間 (Visual Hull) を計算することで三次元モデルを生成する手法である (図 5)。対応づけを行わないために誤対応が生じず、安定な形状復元が行うことができるため、多視点ステレオ法とは異なり欠損のない三次元モデルを生成できる。しかし、対象物体に凹な部分が存在する場合には、その部分は正しい三次元形状が得られないという制約がある。よって、毛や植物などの細かい形状の復元精度は低い。

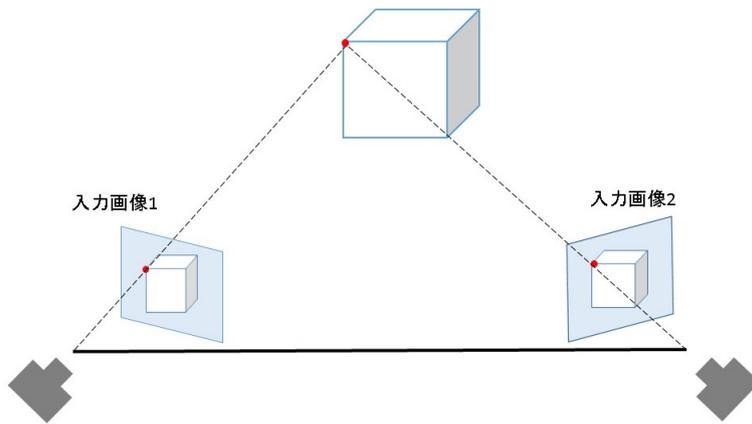


図 4 ステレオ法の原理

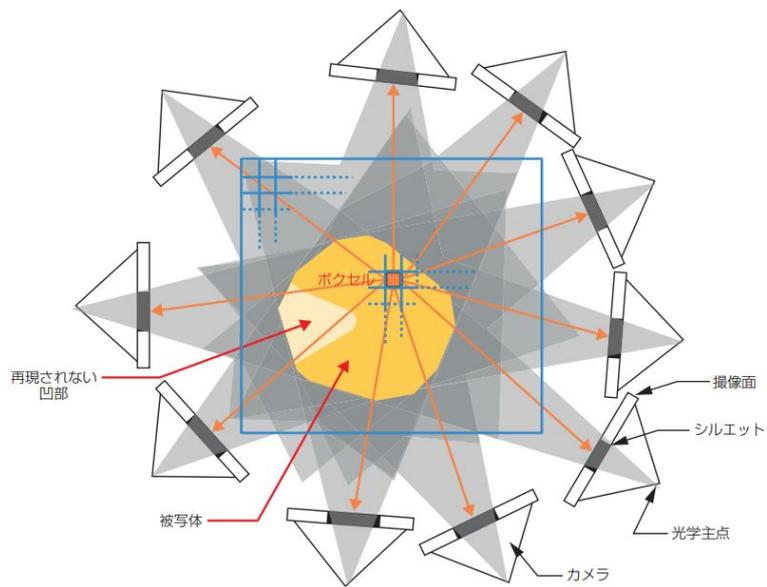


図 5 視体積交差法 [5]

2.1.2 Image-Based Rendering(IBR)

IBR は、対象とする環境の三次元形状を復元することなく、物体やシーンを撮影した複数画像を入力として、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。IBR は対象環境の三次元モデルや反射特性を必要としないため、正確な幾何形状の再現が困難な複雑な形状を持つ樹木等の物体に対しても、自然な自由視点画像を生成可能である。一方で、視点変更量に応じて、生成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、それぞれの特徴を述べる。

[変形による手法]

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 [6] 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 6)。これらの手法では、入力画像に対応点を与えることで中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 [16, 17] 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 Digitizing[16] 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 [17] 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

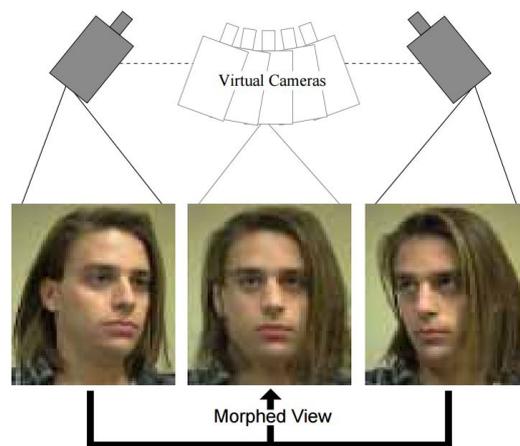


図 6 View Morphing による視点内挿 [6]

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．

[光線情報を用いる手法]

対象環境を撮影した画像を，視点位置の座標 (x, y, z) ，光線の向き (θ_x, θ_y) ，波長 λ 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，対象環境である空間を 7次元の光線情報で記述することが出来る．対象環境に関して，この 7次元のデータを全て記録することができれば，対象環境に関して任意視点における画像の各画素値を光線情報から一意に決定することができる．しかし，実際には全ての 7次元データを全て記録するためには膨大な量の画像が必要となり，現実的に全てのデータを収集することは難しい．サンプリングされなかった光線情報は欠損となる．

欠損を補償する方法として，Lipski ら [8] は線形補間によって最近傍の光線情報から補間することで欠損を補償する手法を提案した．しかし，最近傍の光線情報が切り替わる画素の周辺で，切り替わる地点を境に急激に見えが変化する．Davis ら [7] は見えが滑らかに変化するように，重み係数を用いた N 個の近傍の光線情報のブレンディングにより欠損を補償する手法を提案している (図 7)．

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

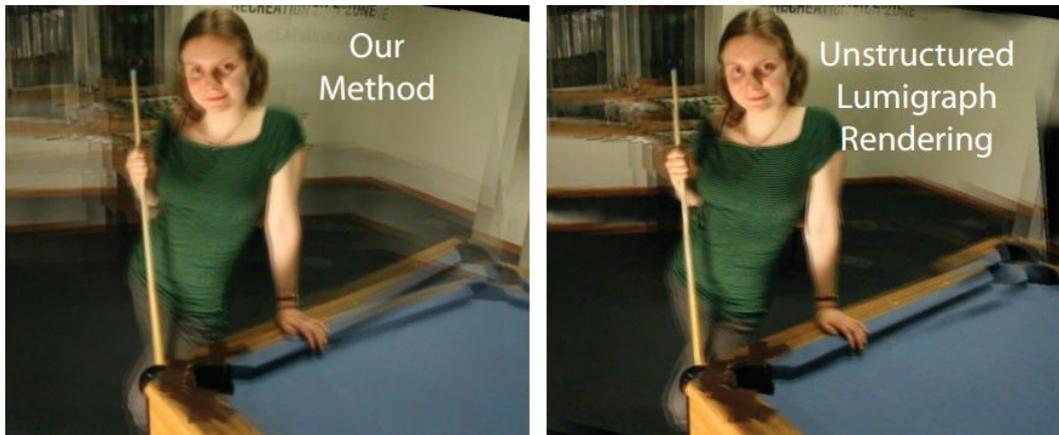


図 7 光線情報の補間方法の違いによる生成画像の違い ((左)Davis ら [7] の手法, (右)Lipski ら [8] の手法).

ることは難しい. IBR によって自由視点画像生成を行うには, 膨大な量の光線情報が必要となる. 特に, サンプルした光線情報からより離れた地点で自由視点画像生成を行うほど, 欠損の補償の不完全さからより大きな歪みが生じる. また, MBR と異なり対象環境の三次元モデルを有していないため, 仮想環境中の道路の幅や曲率を変更するような, シーン構造の変形を実現することは困難である.

2.1.3 Hybrid Rendering(HR)

HR は, MBR, IBR それぞれの利点・欠点を補うために, MBR と IBR を組み合わせた手法である. HR 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される三次元形状を用いる手法がある.

Debevec ら [9] は, 事前に生成した三次元形状に対して, 実環境で撮影した画像群を入力として, 仮想視点に近い見え方の入力画像を選択してテクスチャマッピングを行う手法を提案した. これにより, 粗な三次元形状を利用した場合においても詳細な形状の見えをテクスチャとして再現できる. しかし, 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.

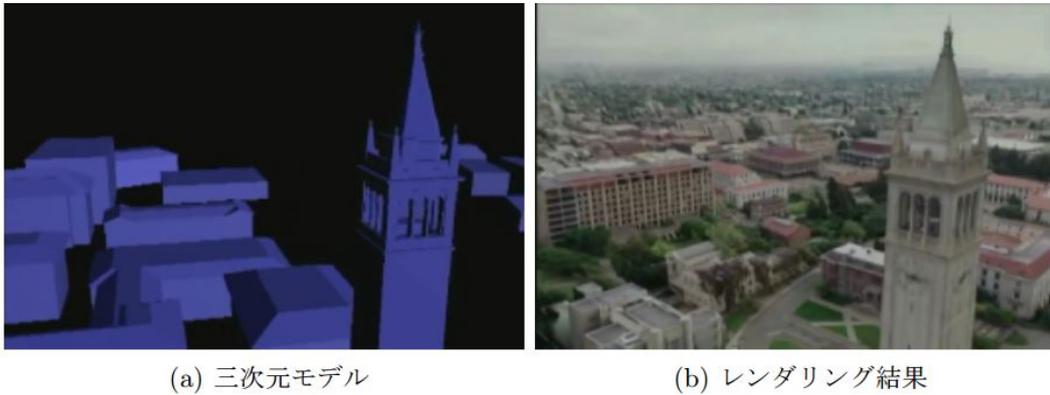


図 8 debevec ら [9] の VDTM

VDG を用いる手法として，伊吹ら [18] は，対象シーンの奥行画像を入力とした自由視点画像生成手法を提案した．任意視点における奥行画像と入力視点の奥行画像群との整合性を考慮し，任意視点における奥行を推定する．そのため，生成画像の品質は，入力である奥行画像群の枚数および精度に依存する．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

静止領域と動領域を併せ持ち，繰り返しを感じさせないループ動画の生成は，Schödl らの Video Textures [19] と呼ばれる手法から始まり，“シネマグラフ”という言葉は，2011 年に Jamie Beck と Kevin Burg により提唱された [20]．Schödl ら [19] は，周期的な動きを持つシーン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してあるフレームに類似したフレームを探索し，それらのフレーム間をつなぐことで繰り返しを感じさせないループ動画を生成した．また，Kwatra ら [21] は，各画素の周期が同一という仮定のもとに，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であると仮定しているため，動画像内に動きの周期が異なる物体が複数存在した場合，動きの不連続を感じる動領域が発生する場合がある．この問題に対して，Liao ら [10] は，時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

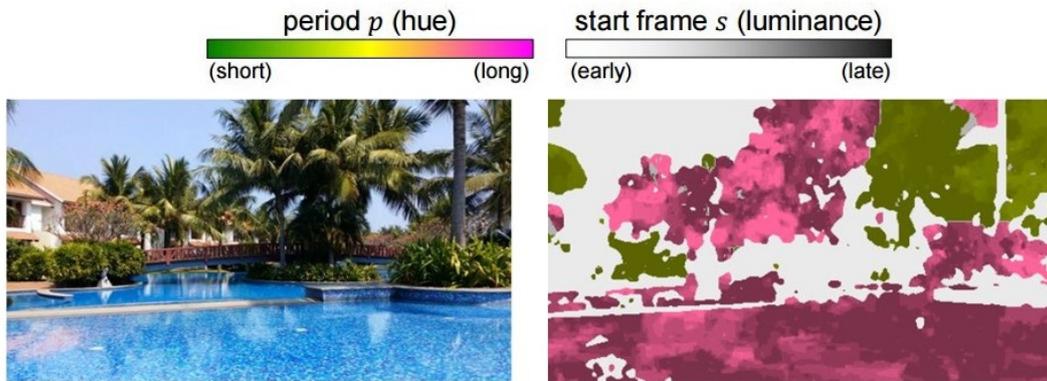


図9 Liaoら[10]の手法((左)入力の1フレームと(右)開始フレームとループ周期)

案した(図9). これにより, 周期の異なる動物体が複数存在する動画像にも対応した. しかしこの手法では, 時間的に画素値が大きく変化しないいくつかのシーンにおいてシームが発生する場合があった. これに対してLiaoら[11]は, Liaoらの手法[10]に基づき, シネマグラフを生成する過程でポアソンプレンディングを適用することによって, シームを低減した(図10). しかし, これらの従来手法では, 入力として光学中心がほぼ固定された動画像が想定されている.

この問題に対してBaiら[22]は, 顔を撮影した光学中心が固定されていないハンドヘルドカメラ映像から, ユーザが指定した領域情報とその特徴点を利用し, 対象フレームにワーピングすることでシネマグラフを生成している. ただし, 入力動画像のカメラ位置が大きく動くことは想定していない. これに対して, 中島ら[12]は, 移動撮影した動画像から対象シーンの三次元形状を復元し, それを利用して他のフレームを対象フレームにワーピングすることで, シネマグラフを生成する手法を提案した(図11). しかし, 移動撮影において, 対象シーンの撮影フレーム数が少ない場合には, 十分な精度で三次元形状復元を行うことが難しく, また利用するテクスチャの枚数が限定的であるため, 最終的に生成されるシネマグラフの品質が低下する問題がある. また, 上記のいずれの手法においても, 任意視点におけるシネマグラフの生成は実現されていない.



(a)Liaoら [10] の手法 (シーム (赤))

(b)Liaoら [11] の手法

図 10 Liaoら [11] の手法によるシームの低減

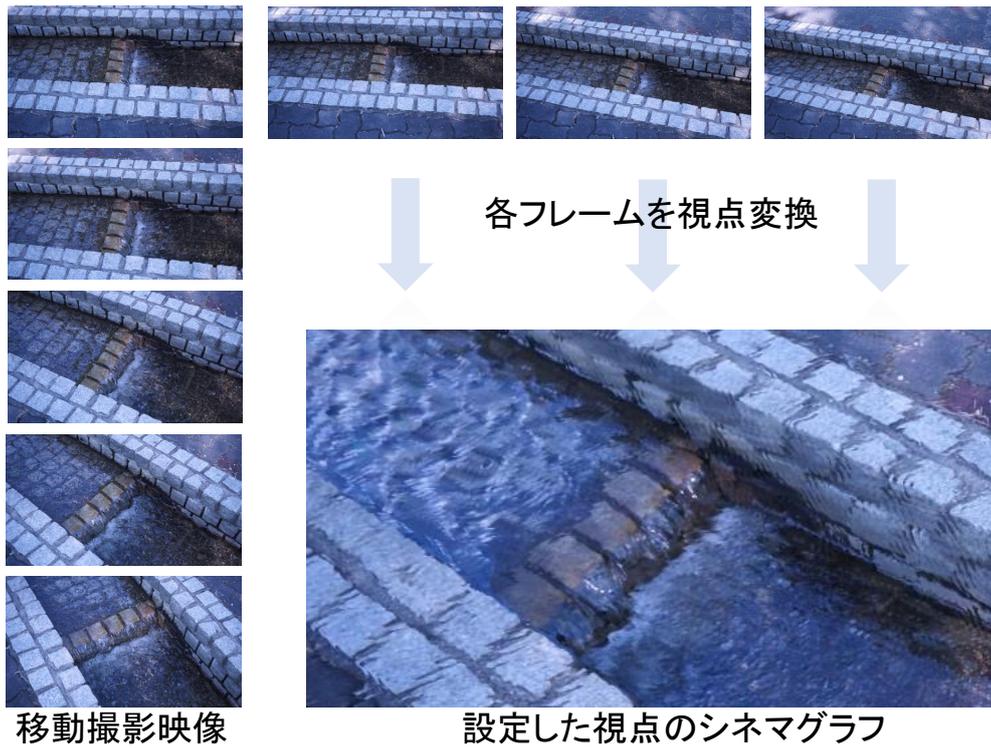


図 11 中島ら [12] の手法

2.3 本研究の位置づけ

本研究では、動物体を含む動画像を用いたストリートビューシステムを構築するための撮影コストの低減および動画像の繰り返し再生における不連続の解消を目的としている。これを実現するために、撮影地点を減らし、撮影視点の動画像から未撮影の仮想視点位置における動画像を生成（撮影地点の補間）をすることで撮影コストを低減させる。従来、動物体を対象とした自由視点画像生成には、時間同期のとれた複数台のカメラを密に設置して撮影した動画像を入力として、対象シーンの三次元形状に基づいて動物体と背景を領域分割することで屋内などの広域でない空間における仮想視点の動画像を生成する手法がある [23]。しかし、本研究では広域を対象するため、同期のとれた多くのカメラを密に設置することは現実的ではない。この問題に対して中島ら [12] は、一台のカメラを用いて移動撮影を行い、撮影した動画像から復元した対象シーンの三次元形状を用いて、動画像をワーピングすることで対象フレームにおける動画像を生成する手法を提案している。しかし、この手法では、対象シーンの撮影フレーム数が少ない場合には、高品質なシネマグラフ生成にとって十分な精度で三次元形状復元を行うことが難しく、また、移動撮影では遮蔽物により”動き”を再現するために必要なテクスチャが不足する場合がある。

これに対して、本研究では一台のカメラにより複数の地点で異なる時刻に定点撮影を行い取得した動画像に加えて、多くの視点からの静止画像も利用することで、対象シーンの三次元形状を十分な精度で復元する。また、定点撮影した動画像を用いることで、テクスチャの不足なく予めその地点でのシネマグラフを生成したうえで、それらを自由視点画像生成技術により仮想視点での見えに変換し、エネルギー最小化による最適なテクスチャ選択を行うことで任意地点における高品質なシネマグラフ生成を実現する。ただし、動画像中の対象とする動物体として、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）とする。

3. 仮想視点におけるシネマグラフ生成

本章では、動画像を用いたストリートビューを比較的低コストで実現するために、非同期に定点撮影した動画像から仮想視点におけるシネマグラフ生成をい実現する手法について説明する。

3.1 提案手法の概要

提案手法では、対象シーンの三次元形状と複数の地点で異なる時刻に定点撮影した動画像群を入力として、自由視点画像生成を行うことで未撮影視点からの見えを再現したシネマグラフを出力する。手法の流れを図12に示す。まず、あらかじめ撮影した対象シーンのモデル生成用画像群と非同期に定点撮影した動画像群の各1フレームの双方を入力として、これらの画像群にSfMを適用することで各画像のカメラ位置姿勢を推定し、推定したカメラ位置姿勢と入力画像群からMVSにより対象シーンの三次元復元を行う。また、定点撮影動画像群それぞれに対してシネマグラフ生成手法を適用しシネマグラフに変換する。次に、推定された各視点におけるカメラ位置姿勢・三次元形状を入力として、シネマグラフ群を仮想視点における見えにワーピングする。最後に、各入力視点のシネマグラフ群から、空間方向の画素値の連続性を考慮したテクスチャ選択を行うことで仮想視点におけるシネマグラフを生成する。以下では、各処理について詳述する。

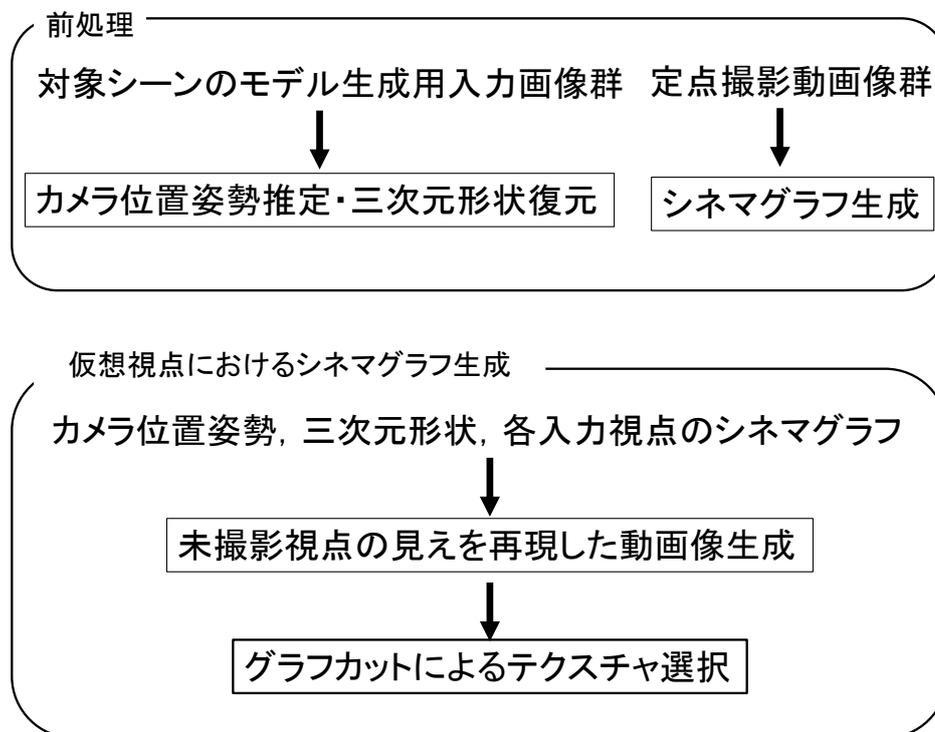


図 12 提案手法概要

3.2 非同期に定点撮影した動画画像群からの仮想視点におけるシネマグラフ生成

3.2.1 カメラ位置姿勢推定と三次元復元

推定されたカメラ位置姿勢と三次元形状を図 13 に例示する。本研究では、対象シーンの三次元形状を十分な精度で得るために、撮影した画像群と非同期に定点撮影した動画画像群の各 1 フレームに対して、VisualSfM[15] を用いることでカメラ位置姿勢推定を行う。VisualSfM は、SIFT 特徴量を用いて自然特徴点を画像ごとに検出し、画像間で特徴点の対応をつけることで、撮影カメラの内部パラメータ・外部パラメータと自然特徴点の三次元位置を推定する。これらの情報を CMPMVS[13] と呼ばれる多視点ステレオ法に入力することで、対象シーンの密な三次元形状を復元する。

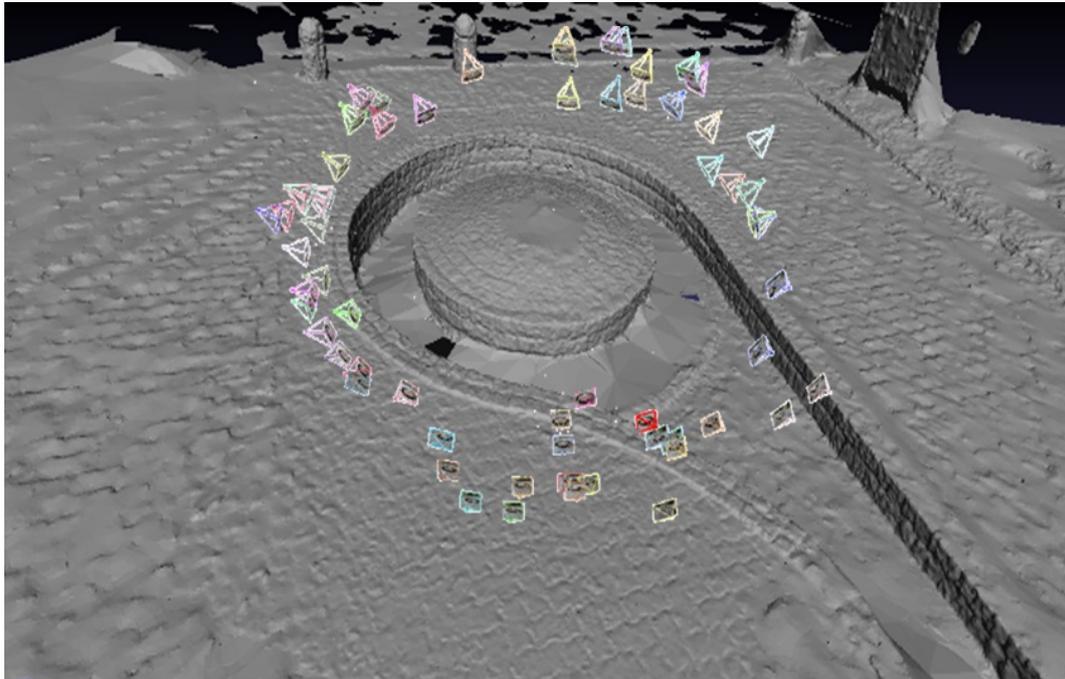


図 13 推定されたカメラ位置姿勢と対象シーンの密な三次元形状

3.2.2 各入力視点のシネマグラフ生成

各地点で定点撮影した動画像に対して Liao ら [11] のシネマグラフ生成手法を適用し、シネマグラフを生成するが、生成された各シネマグラフでフレーム数が異なるため、各画像のフレーム数の最小公倍数を用いてあらかじめフレーム数が同一のシネマグラフを生成しておく。

Liao ら [11] の手法では、時空間的整合性を考慮したエネルギー関数を最小化し、画素ごとにループ周期と開始フレームを決定することで、動画像からシネマグラフを生成する。

3.2.3 シネマグラフの仮想視点からの見えへの変換

複数の地点で非同期に定点撮影した動画像から生成したシネマグラフとそのカメラ位置姿勢および対象シーンの三次元形状を用いて仮想視点の見えにワーピングを行う。まず、復元した三次元形状と推定したカメラ位置姿勢に基づき、定点撮影した動画像群の各視点におけるデプスマップを求める。次に、シネマグラフを生成する仮想視点位置を設定した上で、仮想視点のデプスマップを復元した三次元形状から求める。それらのデプスマップと入力視点のカメラ位置姿勢を利用し、仮想視点と入力視点における画素を対応付け、入力視点の画素値をコピーすることで、シネマグラフの見えの変換を行う（図 14）。このとき、世界座標系において、仮想視点の注目画素に対応する 3 次元形状上の点 P が、参照する入力視点から観測できるかをデプステストを行うことで判定する（図 15）。ここではまず、仮想視点から観測される点 P の座標を求め、入力視点のカメラ位置姿勢を用いて入力画像に投影する。このとき、点 P が入力画像の画素 i に投影されたとすると、点 P から求められるその画素のデプス値 z_i と入力視点の画素 i で観測されたデプス値 d_i を比較し、この差が大きければ遮蔽物が存在すると判定する。また、この差が小さければ、入力視点からも点 P が観測されると判定する。具体的には、

$$|z_i - d_i| < \theta_d \quad (1)$$

を満たすとき、点 P を観測できると判定し画素値をコピーする。ただし、 θ_d は閾値である。また、図 16 に視点変換した画像例示する。

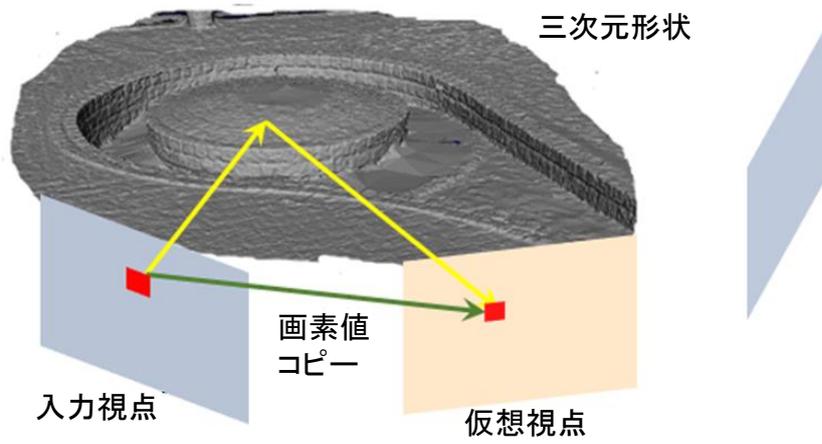


図 14 カメラ位置姿勢と三次元形状による画素の対応付け

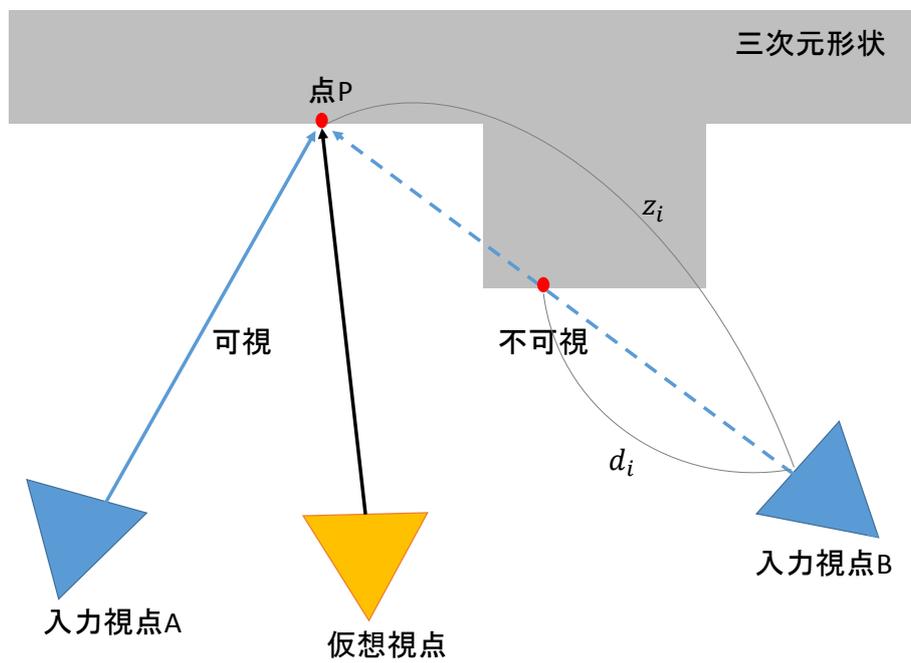


図 15 デプステストによる可視領域の判定



入力視点



対象視点



入力視点から対象視点の見えに変換後

図 16 視点変換した画像例

3.3 エネルギー関数最小化によるテクスチャ選択

複数視点でのシネマグラフ群からのテクスチャ選択では、画素ごとに入力視点を決定する問題を、空間的な画素値の連続性を考慮したエネルギー最小化問題として定式化する。

$$J(X) = \sum_{i \in A} \alpha_i(x_i) f_i(x_i) + K_h \sum_{(i,j) \in B} h_{ij}(x_i, x_j) \quad (2)$$

ここで、 $X = \{x_i | i = 1, \dots, I\}$ の要素 x_i を画素 i の画素値の決定に利用されるシネマグラフのラベル $x_i \in \{1, \dots, I\}$ とする。また、各シネマグラフのフレーム数を N とする。データ項の $f_i(x_i)$ は、仮想視点へ入射する光線と x_i に対応するシネマグラフ生成前の動画を撮影したカメラへ入射する光線の角度が大きくなる場合に大きな値を与える。 $\alpha_i(x_i)$ は可視領域の判定に対応しており、画素 i が x_i に対応する画像から観測される場合には 1 を、そうでない場合は大きな値を与える。平滑化項 $h_{ij}(x_i, x_j)$ では、隣接する画素でラベルが変化する際、その境界で画素値に整合性がない場合に大きな値を与える。ただし、 A は自由視点画像で画素値を割り当てるすべての画素のインデックスの集合、 B は隣接する画素のインデックスの集合を表す。式 (1) は、データ項と平滑化項で構成されるエネルギー関数であり、グラフカットアルゴリズム [24] によって最小化することができる。ここで、 K_h は重みパラメータを表す。

3.3.1 仮想視点と入力視点の見えに関する項

VDTM では、対象シーンの三次元形状に対する適切な入力画像の選択基準として、仮想視点と入力画像を撮影したカメラ（入力視点）位置から対象シーンの物体の三次元位置への光線の類似度が広く利用されている。

本手法では、Debevec ら [9] の手法に基づき、仮想視点から注目画素に対応する三次元形状上の点に向かうベクトルと各入力視点から同じ点に向かうベクトルの成す角の大きさを選択基準として利用している (図 17)。仮想視点の注目画素 i に対応するなす角 θ_{ix_i} は、世界座標系における点 P の座標 \mathbf{p}_i と、仮想視点の力

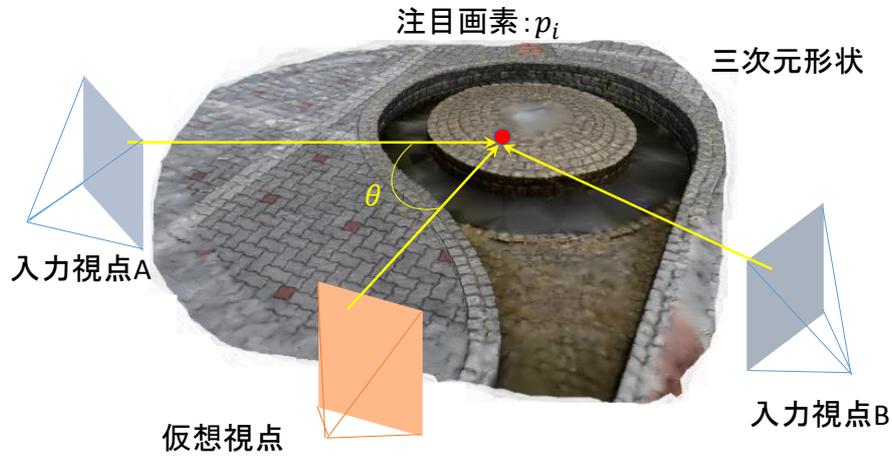


図 17 見えの近さによる画素の選択

メラ位置 \mathbf{t}'_v , 入力視点のカメラ位置 \mathbf{t}'_{x_i} を用いて, 以下の式のように表せる.

$$\theta_{ix_i} = \cos^{-1} \frac{(\mathbf{t}'_v - \mathbf{p}_i) \cdot (\mathbf{t}'_{x_i} - \mathbf{p}_i)}{\|\mathbf{t}'_v - \mathbf{p}_i\| \|\mathbf{t}'_{x_i} - \mathbf{p}_i\|} \quad (3)$$

このなす角を用いて, データ項 $f_i(x_i)$ を次式により与える.

$$f_i(x_i) = \theta_{ix_i} \quad (4)$$

この項において, 角度 θ_{ix_i} が小さい値をとるほどエネルギーが小さくなり, ラベル x_i が選択されやすくなる.

3.3.2 可視判定に関する重み

3.2.2 節で詳述した通り, 自由視点画像中の各画素に対応する物体上の点 that 各入力視点から観測できない場合が存在する. 例えば, 図 18 中の赤色の領域は, 入力視点 A から観測できない領域を例示している. また, 青色の領域は, 入力視点 B から観測できない領域を例示している. このため, 仮想視点におけるテクスチャ

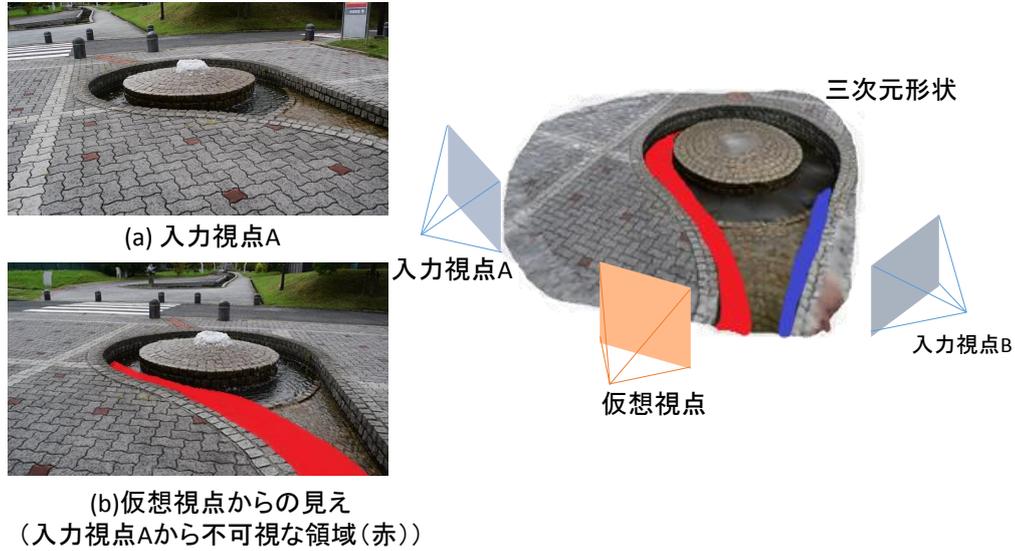


図 18 入力視点位置からの不可視領域

選択の際，デプステストにより不可視と判定された場合にペナルティを設ける．式 (1) を用いて，可視判定に関する重みは以下の値をとるように設定する．

$$\alpha_i(x_i) = \begin{cases} 1 & (|z_i - d_i| < \theta_d) \\ K_\alpha & (\text{上記以外}) \end{cases} \quad (5)$$

ただし K_α は仮想視点上の画素 i に対応する点が入力画像 x_i から不可視の場合のペナルティであり，本研究では K_α に十分に大きな値を設定する．

3.3.3 空間方向の画素値の整合性を考慮した項

空間的なテクスチャの不連続は，隣接する画素間において異なる入力画像ラベルが割り当てられた際に，入力画像間の対応する画素の画素値が大きく異なる場合に生じる．そこで，この項では生成する自由視点画像上で隣接する画素の画素値の不整合を抑止する．画素値の整合性に関する項 $h_{ij}(x_i, x_j)$ を次式で定義する．

$$h_{ij}(x_i, x_j) = \frac{1}{N} \sum_{n=1}^N \max(\|c_{nix_i} - c_{nix_j}\|_\infty, \|c_{njx_i} - c_{njx_j}\|_\infty) \quad (6)$$

ただし, c_{nix_i} は自由視点画像上の画素 i に対応するシネマグラフの n フレーム上における画素値, N はフレーム数を表す. $\|\cdot\|_\infty$ は上限ノルムであり, ベクトルの要素の最大値を与える. この項は, 隣接する画素が異なる入力画像ラベルを割り当てられ, 境界画素において入力画像間の対応する画素値が大きく異なる場合にペナルティを与える.

4. 実験

本章では，屋外環境において複数の地点で異なる時刻に定点撮影した動画像群を用いて，仮想視点におけるシネマグラフの生成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，全方位動画像を用いて生成したシネマグラフの生成結果について述べる．また，以下にシネマグラフ生成，ワーピング，テクスチャ選択で用いた PC，三次元形状復元に用いた PC の仕様について表に示す．

表 1 実験に用いた PC の仕様

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

表 2 三次元形状復元に用いた PC の仕様

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

4.1 実験の概要

本実験では，提案手法の有効性を示すために，単眼動画像に対して以下の3つのラベルの決定手法と提案手法と適用して生成した結果を比較し，主観的に評価した．

(手法1)

Debevecら [9]の手法に基づき，見えの近さによりラベルを決定する．(提案手法において $K_h = 0$ とする．)

(手法2)

提案手法 ($N = 1$) を初期フレームに適用してラベルを決定し，得られた初期フレームのラベルを全フレームで用いる．

(手法3)

提案手法 ($N = 1$) を1フレームごと独立に適用し，ラベルをフレームごとに決定する．

また，全方位動画像に対して提案手法を適用し，任意視点における全方位シネマグラフを生成した．

4.2 単眼動画像を用いたシネマグラフ生成実験

4.2.1 実験環境

単眼カメラを用いて撮影した水の流れる噴水，風で揺れる木の2つのシーンを用いて実験を行なった．定点撮影してシネマグラフ生成手法を適用した各入力動画像は，解像度が 1920×1080 画素の150フレームからなる．図19, 20に対象シーンの三次元形状復元に用いた画像群の一部を示す．また，三次元形状復元に用いた画像枚数と定点撮影地点数を以下の表に示す．生成した仮想視点におけるシネマグラフは，解像度が 1920×1080 画素の150フレームを出力した．

表 3 入力画像画像群と撮影地点数

	復元に用いた画像枚数	定点撮影地数
噴水	67 枚	3
木	30 枚	5



図 19 三次元形状復元のための入力画像群の一部 (噴水のシーン)



図 20 三次元形状復元のための入力画像群の一部 (木のシーン)

4.2.2 実験結果およびその評価と考察

図 21-24 に VisualSFM[15] と CMPMVS[13] を用いて復元したカメラ位置姿勢と三次元形状を示す。また、図 25, 26 に仮想視点として三次元形状復元に用いた画像群のうちの一視点を設定し、シネマグラフ群からそれぞれワーピングを行い仮想視点の見えに視点変換した結果を示す。また、図 27, 28 に、手法 1, 手法 2, 提案手法による結果の比較を、図 29, 30 に、手法 3, 提案手法による結果の比較を示す。

噴水のシーンの三次元復元について、動領域において三次元形状は概ね復元できている。しかし、水が湧き出る箇所において形状が復元出来なかったため、ワーピング後のシネマグラフでは、湧き出る水のテクスチャは本来存在すべき場所に現れていない。これは、画像間で見えが大きく異なり対応付けが適切にできなかったためだと考えられる。また、木のシーンについては、動領域全体においては大まかな形状が復元できている。

図 27(a) の Debevec ら [9] に基づく手法 1 により生成された噴水のシーンに対するシネマグラフでは、カメラ位置姿勢推定または三次元形状復元の誤差が原因で視点 B のワーピング後の画像において本来水路上にないテクスチャが水路上に存在しているが、そのテクスチャを選択し、テクスチャ切り替えの境界において大きなテクスチャの不連続が存在する結果となった。図 28(a) の木のシーンにおいても、三次元形状の誤差によりテクスチャ間の境界が不連続となった。図 27(b) の手法 2 では、初期フレームのみを用いてラベルを決定しているため、噴水のシーンにおいては、初期フレームから時間的に画素値が大きく変化し、フレームによってテクスチャの境界においてテクスチャが不連続になる場合があった。これらと比較し、提案手法による結果では、動画像全体において空間的に大きくテクスチャが切り替わるフレームが比較的少ないことを確認した。

また、手法 3 では、各フレーム独立にラベルを決定して生成した結果をつなぎ合わせて動画像を生成しているため、図 30, 31 のように前後のフレーム間でテクスチャが切り替わる位置が大きく変わり、ユーザにちらつきを感じさせる。これに対して、提案手法では、動画像全体に対して 1 つのラベルを用いているため、前後のフレーム間において、テクスチャの切り替わる位置が変化せず、そのよう



図 21 VisualSFM によるカメラ位置姿勢と推定された三次元位置 (噴水のシーン)

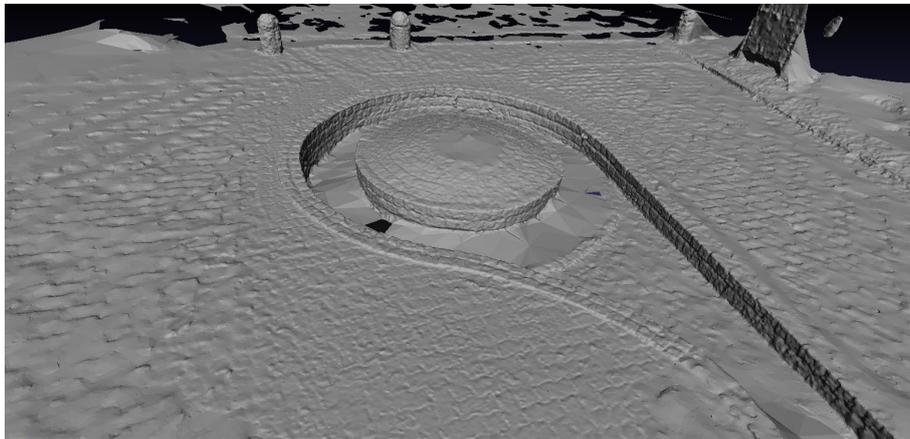


図 22 CPMVS により復元した三次元形状 (噴水のシーン)

なちらつきが生じないことを確認した.



図 23 VisualSFM によるカメラ位置姿勢と推定された三次元位置 (木のシーン)

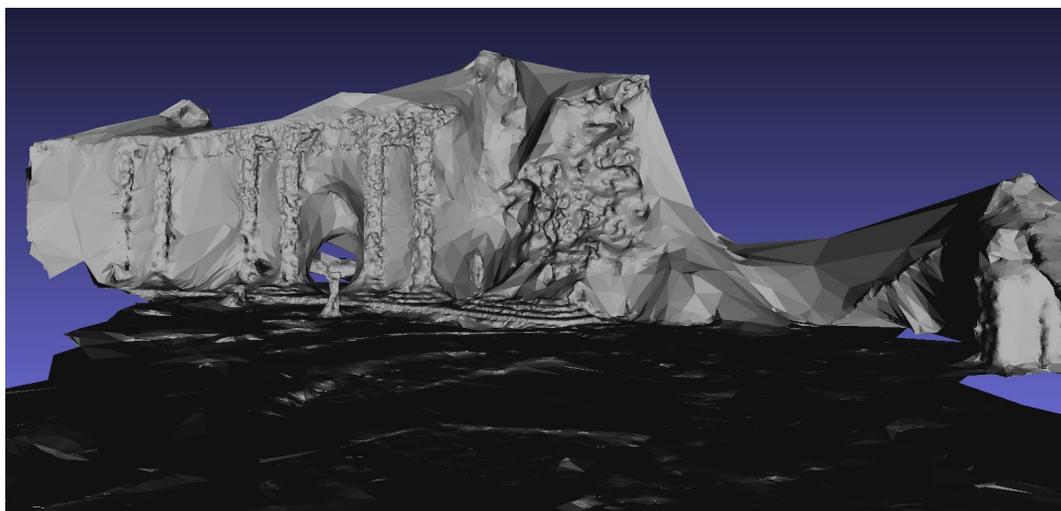


図 24 CMPMVS により復元した三次元形状 (木のシーン)



図 25 視点変換動画像 (噴水のシーン)

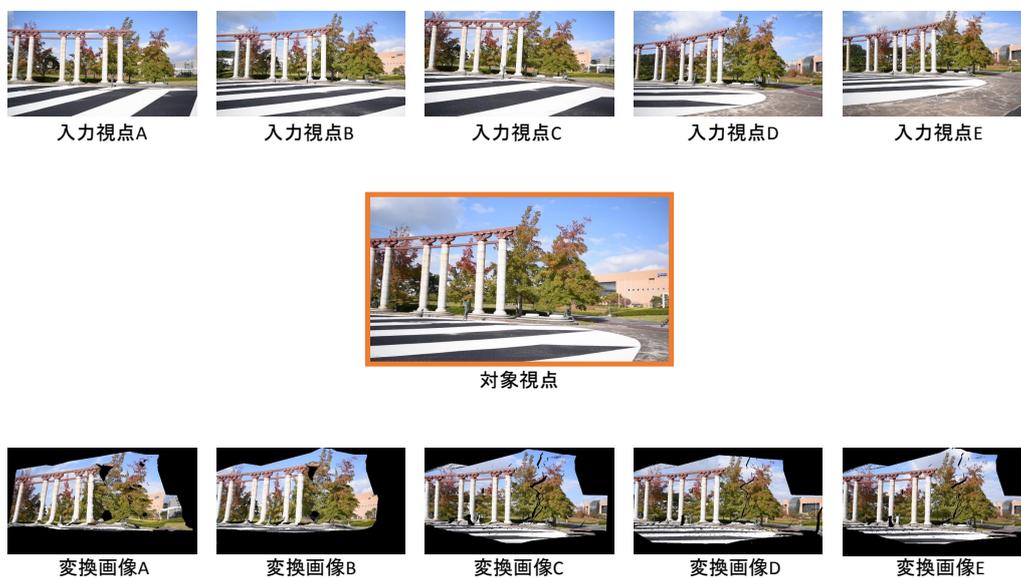


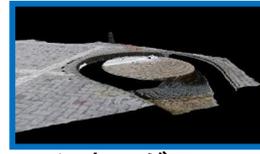
図 26 視点変換動画像 (木のシーン)



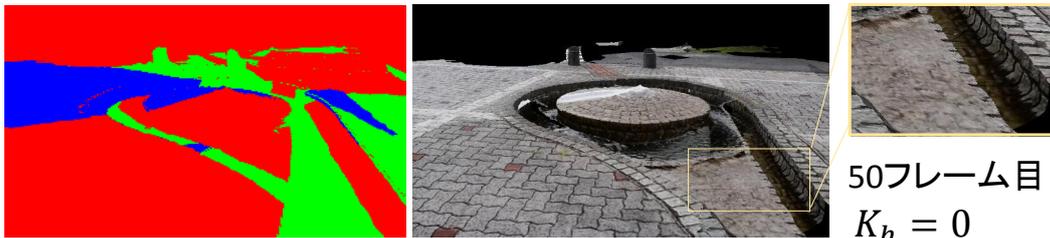
シネマグラフィA



シネマグラフィB



シネマグラフィC



50フレーム目
 $K_h = 0$

(a) Debevecら[9]の手法により選択されたテクスチャ



50フレーム目
 $K_h = 500$

(b) 初期フレームの空間的整合性により選択されたテクスチャ



50フレーム目
 $K_h = 500$

(c) 提案手法により選択されたテクスチャ

図 27 動画像中のある 1 フレームにおける空間的整合性の比較 (噴水のシーン)



(a) Debevecら[9]の手法により選択されたテクスチャ



(b) 初期フレームの空間的整合性により選択されたテクスチャ



(c) 提案手法により選択されたテクスチャ

図 28 動画像中のある 1 フレームにおける空間的整合性の比較 (木のシーン)

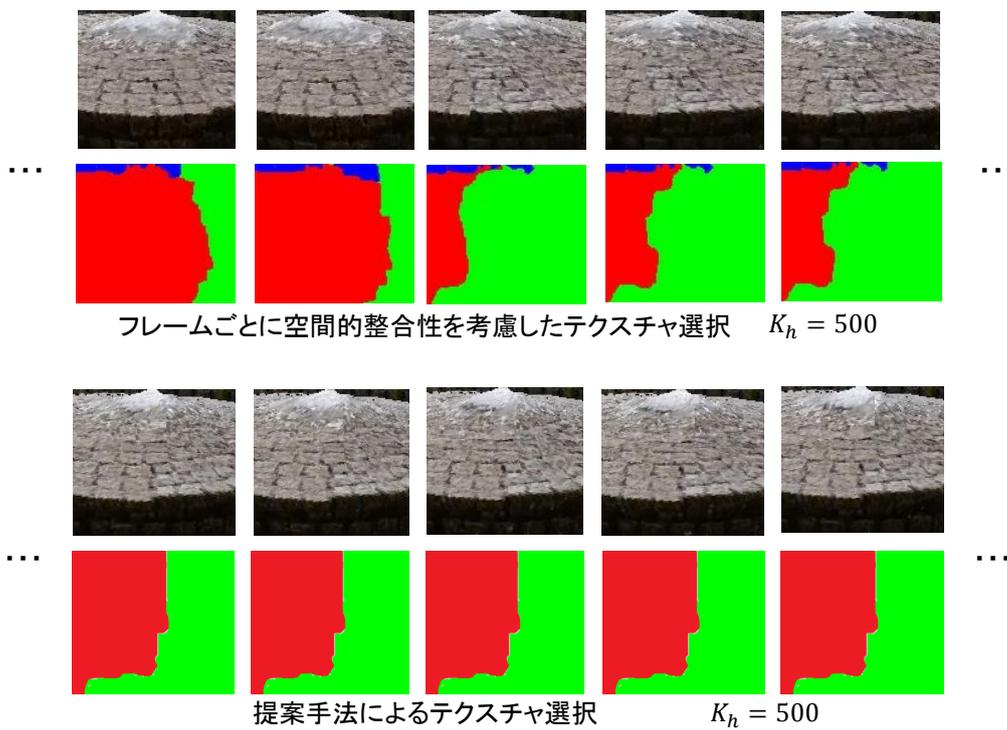


図 29 動画像中の前後 5 フレームにおける時間的整合性の比較 (噴水のシーン)

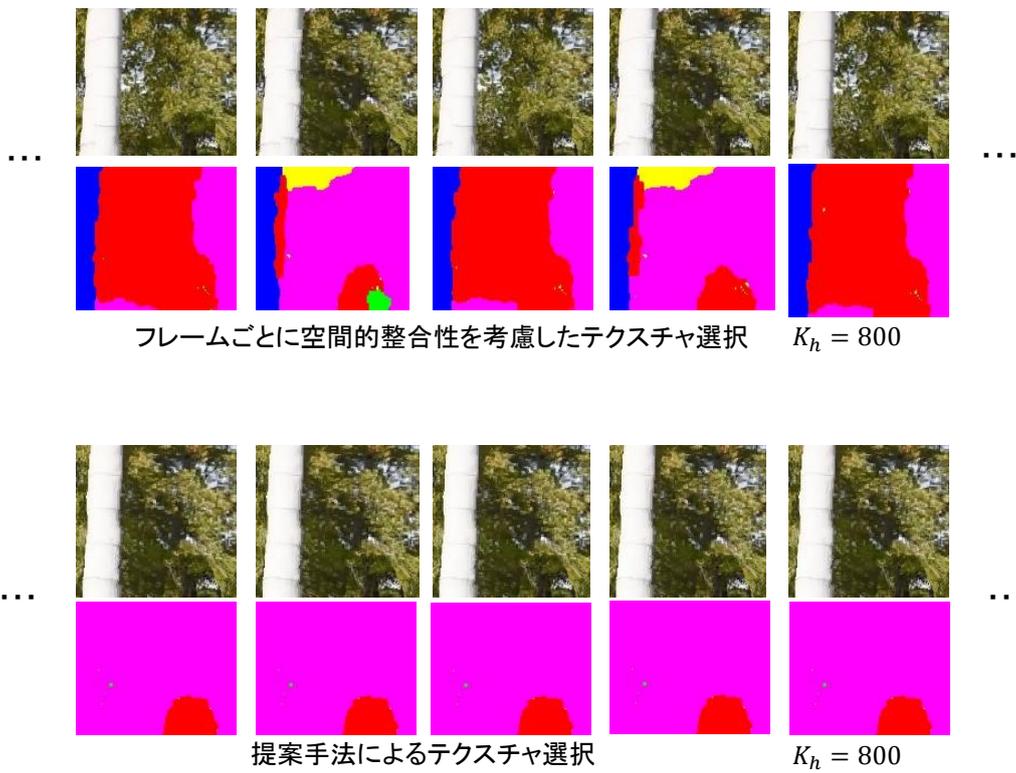


図 30 動画像中の前後 5 フレームにおける時間的整合性の比較 (木のシーン)

4.3 全方位動画像を用いたシネマグラフ生成実験

4.3.1 実験環境

全方位カメラ Ladybug3 を用いて、流れる水・揺れる木を含むシーンを撮影し、実験を行なった。定点撮影した各入力動画像は、解像度が 2000×1000 画素の 150 フレームからなる。また、定点撮影を行った地点数は 4 箇所とし、モデル生成には全方位画像 100 枚を用いた。ワーピングの前処理としてカメラ位置姿勢推定・三次元形状復元を行うときは、全方位パノラマ画像をキューブマップに展開し、これを入力とした (図 31)。展開したキューブマップの画角は、入力画像間の対応点を探索するための視野領域の重複を考慮して画角 120 度とした。ただし、下方向カメラ画像を除く。図 32 に対象シーンの三次元形状復元に用いた画像群の一部を示す。三次元形状復元に用いた画像枚数は、300 フレーム \times 5 方向でキューブマップはそれぞれ 1024×1024 画素の画像とした。また、生成した仮想視点におけるシネマグラフは、解像度が 2000×1000 画素の 150 フレームを出力した。

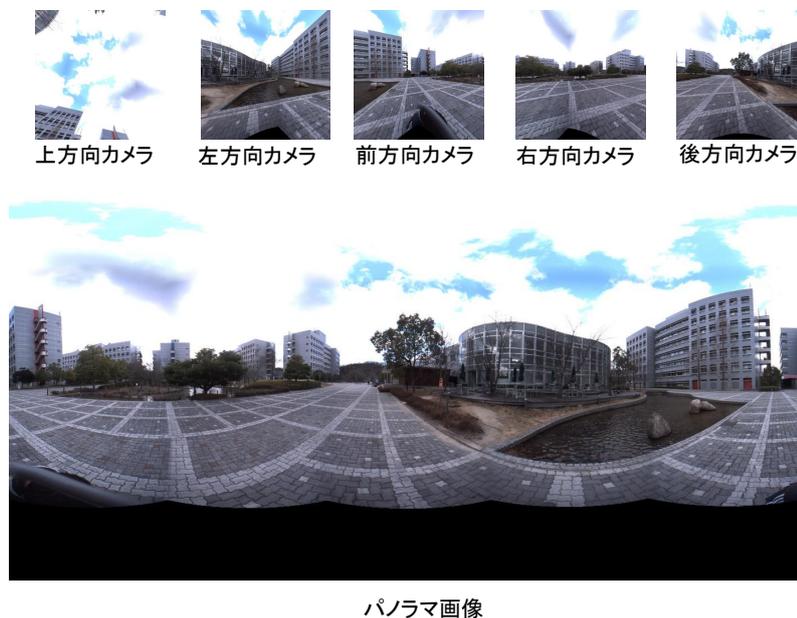


図 31 画角 120 度のキューブマップの展開例

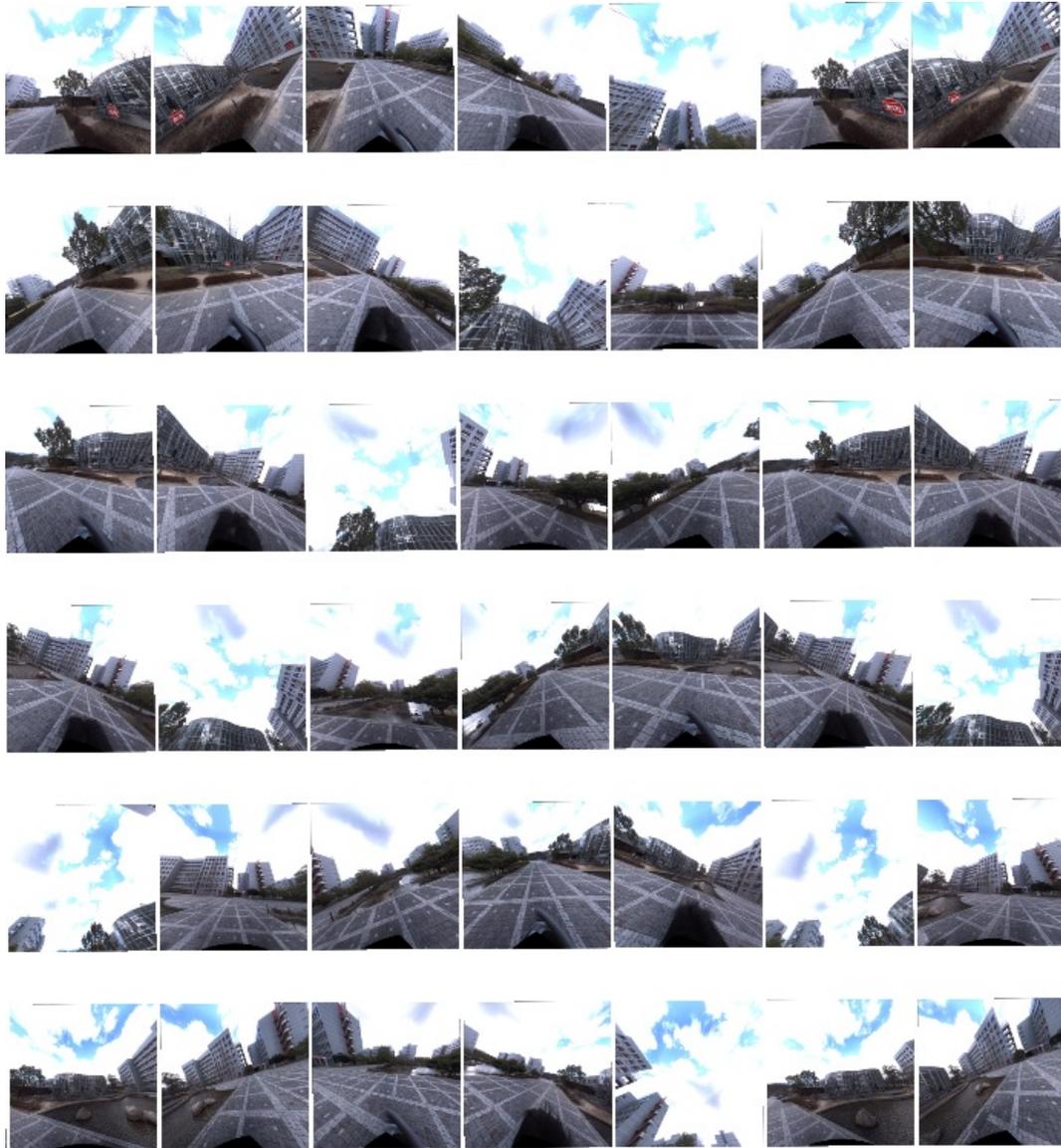


図 32 三次元形状復元のためのキューブマップ画像群

4.3.2 実験結果およびその評価と考察

図 33, 34 に VisualSFM と CMPMVS を用いて復元したカメラ位置姿勢と三次元形状を示す。また, 図 35 に仮想視点として三次元形状復元に用いた画像群の一視点を設定し, シネマグラフ群からそれぞれワーピングを行い仮想視点の見えに視点変換した結果を示す。ここでは, 空等の三次元形状の欠損箇所においては, ワーピングを適用することができないため, テクスチャを選択することができないことを考慮し, 設定した仮想視点位置で撮影した画像を用いてこれを欠損位置におけるテクスチャとした。各視点変換後のシネマグラフ群から提案手法を用いてテクスチャ選択を行った結果を図 36 に示す。

本実験において, 木や池などの動物体の三次元形状を概ね復元することができたため, 全方位動画像に対して提案手法を適用できることを確認した。また, 提案手法により生成された動画像において, 木や池などの動物体に対してある程度違和感のないテクスチャを選択することができた。しかし, ワーピングを行う際に入力視点と仮想視点の間の距離が大きいと, ワーピング後の画像において解像度の低い動画像を生成した。よって, 仮想視点におけるテクスチャ選択をする際に, 入力視点位置に距離の制限を設ける必要がある。また, 三次元形状の欠損に対応するテクスチャを選択する手法が必要となる。

図 37 に提案手法を適用したシネマグラフの数フレームの拡大図を示す。一見, ラベルの境界において, 空間方向のテクスチャの切り替えに対する整合性が概ねとれているように見える。ただし, このシーンにおいて, 図 38 に示すような各ラベルに対応する動物体の動きの大きさの違いが原因で, 木全体の動領域においてシネマグラフ再生時にラベルの境界が目立つことが分かった。

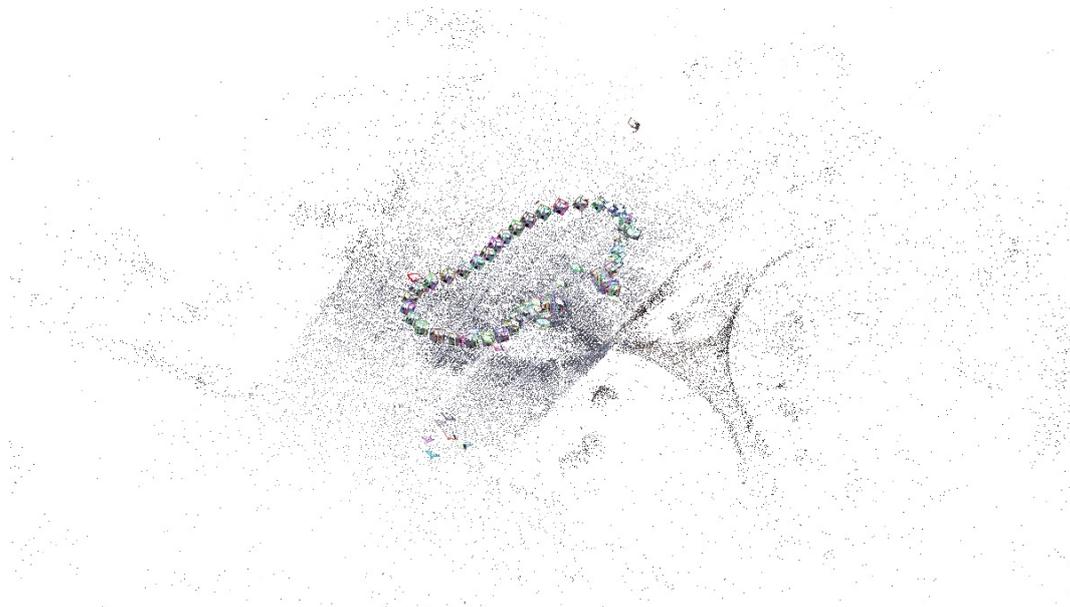


図 33 VisualSFM によるカメラ位置姿勢と推定された三次元位置

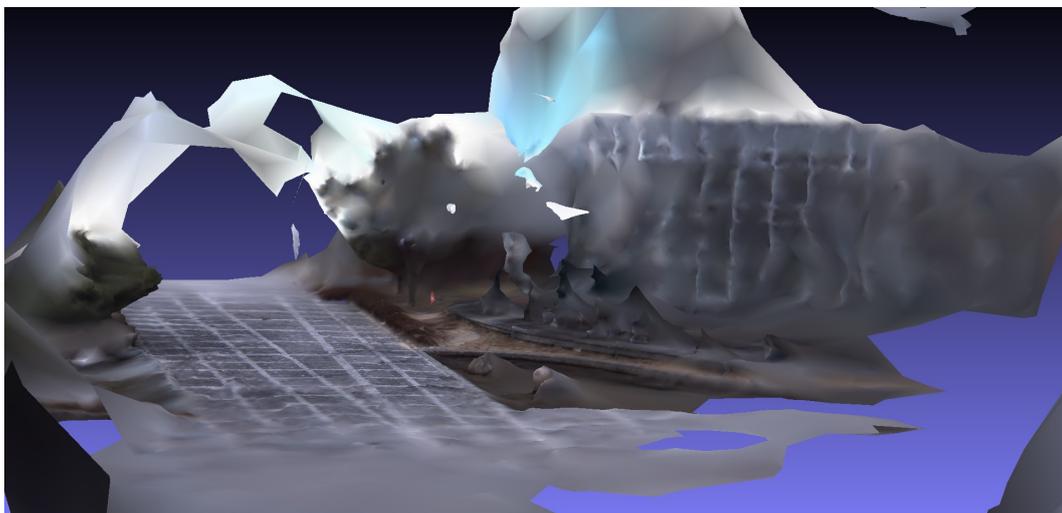


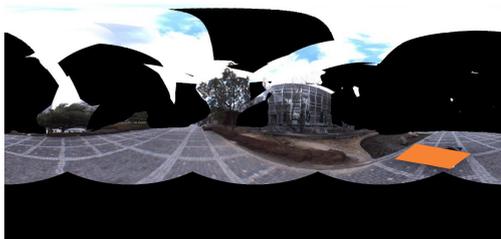
図 34 CPMVS により復元した三次元形状



入力視点のパノラマ画像



対象視点のパノラマ画像



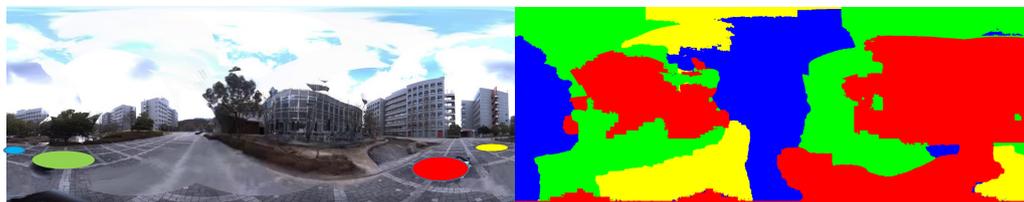
対象視点位置のテクスチャなし



対象視点位置のテクスチャあり

入力視点から対象視点の見えに変換後
(入力視点位置: **オレンジ**)

図 35 視点変換後の全方位画像



提案手法を全方位動画像に対して適用した結果
(各色はラベルの入力視点位置)

提案手法により決定したラベル $K_h = 500$



選択された動物体のテクスチャ



選択された動物体のテクスチャ



選択された動物体のテクスチャ

図 36 エネルギー最小化による全方位動画像生成結果

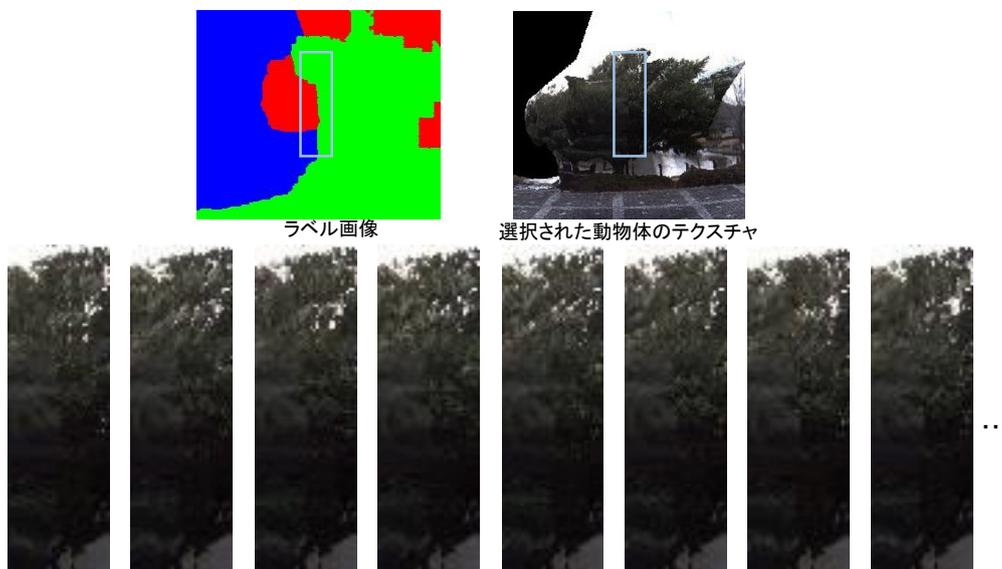


図 37 生成したシネマグラフにおけるテクスチャの整合性

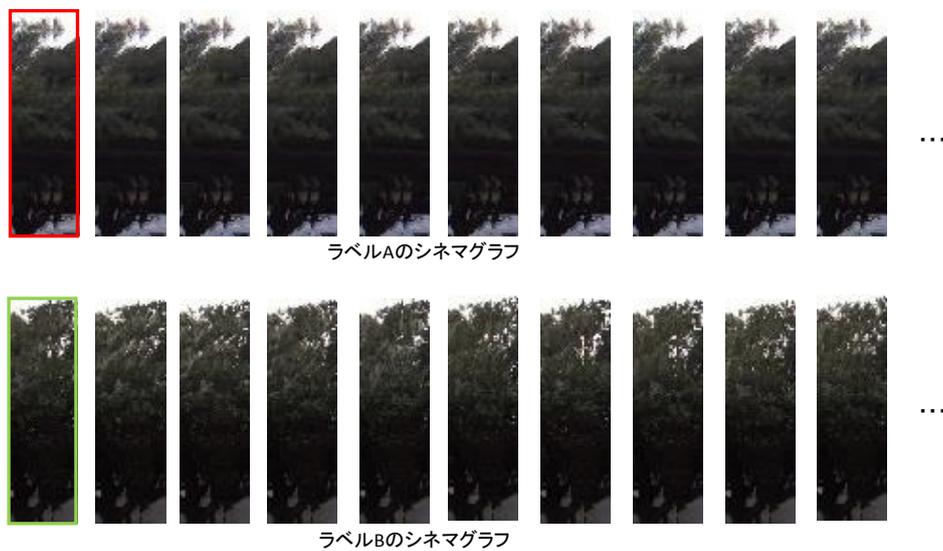


図 38 各ラベルにおける動物体の動きの大きさの違い

5. まとめ

本論文では、複数の地点で異なる時刻に定点撮影した動画像から、ある視点位置での動物体の”動き”を再現するシネマグラフを生成する手法を提案した。提案手法では、復元した対象シーンの三次元形状とカメラ位置姿勢を用いてワーピングしたシネマグラフ群から、隣接する画素間における入力画像の対応画素の値の整合性を考慮したエネルギー関数を最小化することにより各画素に対して画素値を取得するテクスチャを選択する。

実験では、提案手法において他のテクスチャ選択手法と比較し、提案手法の有効性について主観的に評価した。全方位カメラを用いた実験においても、提案手法により設定した仮想視点の見えを再現したシネマグラフを実現できることを示した。結果として、形状の復元やカメラ位置姿勢推定に大きな誤差がある場合において質の高いシネマグラフを生成できないが、三次元形状復元の誤差が小さい場合には、提案手法は不連続なテクスチャを生じさせにくく有効であることが分かった。しかし、空などの遠方においてワーピングに必要な三次元形状が復元されず、テクスチャを選択することができなかった。また、提案手法では、動画像全体に対して1つのラベルを用いているため、複数のテクスチャを含む動領域において、動きの大きさが異なる木などの動物体に対応できなかった。

今後の課題として、時間方向におけるテクスチャの切り替わりについて考慮したエネルギー関数の設計や、三次元形状における欠損の補間方法および三次元形状の欠損箇所におけるテクスチャ選択手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，御鞭撻を頂いたロボティクス研究室 小笠原 司 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，研究に関する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 客員准教授（現 大阪大学 データビリティフロンティア機構 准教授）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，南 あずさ 女史，中村 美奈 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] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, Vol. 37, No. 4, pp. 827–849, 2004.
- [2] S. Zhang and P. Huang. High-resolution, real-time 3d shape acquisition. *In Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW)*, pp. 28–37, 2004.
- [3] M. Hansard, S. Lee, O. Choi, and R. P. Horaud. Time-of-flight cameras: Principles, methods and applications. *Springer Science & Business Media*, 2012.
- [4] L. Li. Time-of-flight camera—an introduction. Technical report, Texas Instruments - Technical White Paper, 2014.
- [5] J. H. Woodhouse and A. M. Dziewonski. Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms. *Journal of Geophysical Research: Solid Earth*, Vol. 89, No. B7, pp. 5953–5986, 1984.
- [6] S. M. Seitz and C. R. Dyer. View morphing. *In Proc. ACM Conf. on SIGGRAPH*, pp. 21–30, 1996.
- [7] A. Davis, M. Levoy, and F. Durand. Unstructured light fields. *Computer Graphics Forum*, Vol. 31, No. 2, pp. 305–314, 2012.
- [8] C. Lipski, C. Linz, K. Berger, A. Sellent, and M. Magnor. Virtual video camera: Image-based viewpoint navigation through space and time. *Computer Graphics Forum*, Vol. 29, No. 8, pp. 2555–2568, 2010.
- [9] P. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. *Rendering Techniques*, pp. 105–116, 1998.

- [10] Z. Liao, N. Joshi, and H. Hoppe. Automated video looping with progressive dynamism. *ACM Trans. on Graphics (TOG)*, Vol. 32, No. 4, p. 77, 2013.
- [11] J. Liao, M. Finch, and H. Hoppe. Fast computation of seamless video loops. *ACM Trans. on Graphics (TOG)*, Vol. 34, No. 6, pp. 197:1–197:10, 2015.
- [12] 中島章敬, 河合紀彦, 佐藤智和, 横矢直和. 自由視点画像生成に基づく移動カメラ映像からのシネマグラフ生成. 映像情報メディア学会 2014 年冬季大会講演予稿集, Vol. 12, No. 2, pp. 4–11, 2014.
- [13] 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)*, pp. 3121–3128, 2011.
- [14] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 32, No. 8, pp. 1362–1376, 2010.
- [15] C. Wu, et al. Visualsfm: A visual structure from motion system. 2011.
- [16] S. Koizumi and H. Ishiguro. Town digitizing: Omnidirectional image-based virtual space. *In Proc. IEEE Computer Society Conf. on International Digital Cities Workshop(IDCW)*, pp. 247–258, 2003.
- [17] G. Albuquerque M. Magnor T. Stich, C. Linz. View and time interpolation in image space. *Pacific Graphics*, Vol. 27, No. 7, 2008.
- [18] 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, PRMU2009-370, pp. 437–442, 2010.
- [19] A. Schödl, R. Szeliski, D. H. Salesin, and I. Essa. Video textures. *In Proc. ACM Conf. on SIGGRAPH*, pp. 489–498, 2000.
- [20] Cinemagraphs. <http://cinemagraphs.com/>.

- [21] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. *ACM Trans. on Graphics(TOG)*, Vol. 22, No. 3, pp. 277–286, 2003.
- [22] J. Bai, A. Agarwala, M. Agrawala, and R. Ramamoorthi. Automatic cinemagraph portraits. *Computer Graphics Forum*, Vol. 32, No. 4, pp. 17–25, 2013.
- [23] 稲本奈穂, 斎藤英雄. 多視点スポーツ映像からの自由視点映像合成と提示. 電子情報通信学会論文誌 D-II, Vol. J88-D-II,, No. 8, pp. 1693–1701, 2005.
- [24] Y. Boykov and G. Funka-Lea. Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision*, Vol. 70, No. 2, pp. 109–131, 2006.