

NAIST-IS-MT1551064

修士論文

自由視点動画画像生成における視点移動時の
テクスチャちらつきの低減

寺崎 希

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。

寺崎 希

審査委員：

横矢 直和 教授	(主指導教員)
向川 康博 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)
中島 悠太 客員准教授	(副指導教員/大阪大学)

自由視点動画画像生成における視点移動時の テクスチャちらつきの低減*

寺崎 希

内容梗概

複数地点から撮影した画像群を入力とし、任意視点（仮想視点）からの見えを再現する自由視点画像生成は、実シーンの仮想化のための基盤技術として注目されており、遠隔地のシーンを仮想的に体感できる VR アプリケーションなどへの応用が期待されている。従来の自由視点画像生成手法として定番となりつつある視点依存テクスチャマッピング法は、入力画像群からシーンの 3 次元形状を復元し、その 3 次元形状にマッピングするテクスチャとして自由視点画像の各画素に対して見え方が最も近い入力画像の画素を利用している。この手法を用いることで、3 次元復元の精度が不十分な場合でも詳細な形状を視覚的に再現することが可能である。しかし、従来の視点依存テクスチャマッピング法では 1 つの仮想視点から対象シーンを見たときの視覚的な違和感の低減を目的としており、視点移動が伴う動画画像生成においては各フレームの画像が独立に生成される。このため、入力画像取得時の照明条件の差異や 3 次元形状とカメラパラメータの誤差によって、フレーム間でマッピングされる入力画像が切り替わる際に、物体上の同一箇所のテクスチャが不連続に変化し、生成した動画画像がちらつく問題が生じる。

この問題に対し本研究では、視点移動を伴う自由視点動画画像の品質の向上を目的とし、フレーム間のテクスチャの整合性を考慮した自由視点動画画像生成手法を提案する。提案手法では、従来手法で用いられてきたテクスチャ選択の条件に加えて、前フレームで利用されたテクスチャと対象フレームで利用されるテクスチャ

*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-MT1551064, 2017 年 3 月 16 日.

との色の差異をコストとして用いたエネルギー関数を新たに設計し，グラフカットアルゴリズムを用いてこれを最小化することで，対象シーン内の同一箇所におけるフレーム間のテクスチャの差異が生じにくい入力画像を各画素に対して選択する．また，従来手法と同様に，異なる入力画像からのテクスチャの境界における空間的な輝度調整を各フレームで行うとともに，時間方向の輝度調整を行うことで，空間的・時間的な輝度の不連続を抑制する．実験では，小物体や屋外環境のシーンを対象として視点移動を伴う自由視点動画像を生成することで，追加したテクスチャ選択条件の効果と時間方向の輝度調整の効果を検証する．

キーワード

自由視点動画像生成，視点依存テクスチャマッピング，フレーム間のテクスチャの整合性，時間方向の輝度調整，グラフカット

Reducing Texture Flicker by Viewpoint Changes in Novel Viewpoint Video Generation*

Nozomi Terasaki

Abstract

Novel view synthesis (NVS) is a technique for synthesizing an image from an arbitrary viewpoint given by a user using multiple images taken from different viewpoints in a scene. NVS is a key technique to visually virtualize the real world and is expected to be used for VR applications that enable users to feel as if they are at remote sites. One of conventional NVS methods is View Dependent Texture Mapping (VDTM). In this method, given 3D geometry of a scene and camera poses restored from input images, textures mapped to 3D geometry are adaptively selected using the spatial relationship of 3D scene geometry and cameras. This method reproduces the visual details on the 3D geometry even if the reconstructed geometry is not sufficiently accurate. However, since the conventional VDTM aims to generate a novel view in high quality from a single viewpoint, flicker often occurs when generating videos from moving viewpoints because of the difference in the illumination condition of the input images and errors in the 3D geometry reconstruction.

In this study, to improve the quality of videos from arbitrary viewpoints, we propose a new method that considers the consistency in texture between consecutive frames. In the proposed method, to determine a pixel, which is used as a texture for each pixel in a novel view, from input images we design a new energy function based on the difference in colors between the current and previous

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1551064, March 16, 2017.

frames, in addition to the criteria used in the conventional methods. By minimizing the energy function with the graph cut algorithm, the proposed method selects such an input image for each pixel that is not likely to cause the significant discontinuities between consecutive frames. The proposed method also suppresses the spatial and temporal discontinuities by adjusting the color of a pixel using its temporally as well as spatially adjacent pixels. In our experiment, in order to verify the improvement of video quality, we compare videos with moving view-points generated by the proposed and conventional methods using input images of small objects and outdoor environments.

Keywords:

novel view synthesis, view-dependent texture mapping, consistency in textures between frames, temporal brightness adjustment, graph cut

目次

1. はじめに	1
2. 従来研究と本研究の位置づけ	3
2.1 Model-based Rendering	3
2.2 3次元形状を援用しない Image-based Rendering	6
2.3 3次元形状を援用する Image-based Rendering	9
2.4 自由視点動画生成	10
2.5 本研究の位置づけ	12
3. フレーム間のテクスチャの整合性を考慮した自由視点動画生成	14
3.1 提案手法の概要	14
3.2 エネルギー関数のデータ項	18
3.2.1 仮想視点と入力視点の光線の方向の近さに関する項	18
3.2.2 フレーム間のテクスチャの整合性に関する項	20
3.2.3 可視判定に関する重み	20
3.3 エネルギー関数の平滑化項	23
3.3.1 テクスチャ境界の画素値の不整合に関する項	23
3.3.2 3次元形状の空間的な変化の大きさに関する重み	23
3.4 空間的・時間的な輝度調整	26
3.4.1 空間的な輝度調整	28
3.4.2 時間的な輝度調整	29
4. 実験	32
4.1 実験の概要	32
4.2 小物体を対象とした映像生成	33
4.2.1 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ に関する検証	33
4.2.2 時間的な輝度調整のフレーム数 s に関する検証	47
4.3 屋外環境のシーンを対象とした映像生成	52
4.3.1 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ に関する検証	52

4.3.2 時間的な輝度調整のフレーム数 s に関する検証	62
4.4 被験者実験によるちらつき低減効果の確認と考察	68
5. まとめ	72
謝辞	73
参考文献	74

目 次

1	ToF 方式	4
2	光切断方式 ([1] より)	4
3	視体積交差法 ([2] より)	5
4	PMVS で復元した 3 次元形状を利用して生成した自由視点画像 ([3] より)	5
5	モーフィング ([4] より)	7
6	ライトフィールド ([5] より)	7
7	ライトフィールドレンダリングで生成した自由視点画像 ([6] より)	8
8	Debevec らの視点依存テクスチャマッピング ([7] より)	10
9	画素値の不連続が生じた自由視点画像 ([8] より)	11
10	鏡面反射を考慮した視点依存テクスチャマッピング ([9] より)	11
11	Kopf らのタイムラプス動画像生成 ([10] より)	12
12	提案手法の概要	15
13	仮想視点と入力視点の見え方の近さ	19
14	従来手法を用いた場合のフレーム間のテクスチャの不整合	21
15	フレーム間のテクスチャの整合性	21
16	可視判定の例	22
17	テクスチャの空間的な不連続	24
18	3 次元形状の空間的な変化が大きいシーンの例	25
19	法線取得	26
20	空間的な画素値の不連続が残ったフレーム	27
21	時間的な画素値の不連続が残ったフレーム	27
22	空間的な輝度調整	29
23	時間的な輝度調整	31
24	小物体 1 の入力画像例	35
25	小物体 2 の入力画像例	35
26	屋外環境シーン 1 の入力画像例	35
27	屋外環境シーン 2 の入力画像例	35

28	各シーンにおける仮想視点の動き	36
29	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (小物体 1) .	38
30	図 29 の一部分を拡大した画像	39
31	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解 消した例 (小物体 1)	40
32	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続が生 じた例 (小物体 1)	41
33	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ による時間的な不連続発生 の原因 (小物体 1)	42
34	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (小物体 2) .	43
35	図 34 の一部分を拡大した画像	44
36	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解 消した例 (小物体 2)	45
37	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続が生 じた例 (小物体 2)	46
38	時間的な輝度調整の効果の比較 (小物体 1)	48
39	図 38 の一部分を拡大した画像	49
40	時間的な輝度調整の効果の比較 (小物体 2)	50
41	図 40 の一部分を拡大した画像	51
42	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (屋外環境シー ン 1)	54
43	図 42 の一部分を拡大した画像	55
44	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解 消した例 (屋外環境シーン 1)	56
45	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続が生 じた例 (屋外環境シーン 1)	57
46	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ による時間的な不連続発生 の原因 (屋外環境シーン 1)	58

47	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (屋外環境シーン 2)	59
48	図 47 の一部分を拡大した画像	60
49	フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解消した例 (屋外環境シーン 2)	61
50	時間的な輝度調整の効果の比較 (屋外環境シーン 1)	63
51	図 50 の一部分を拡大した画像	64
52	時間的な輝度調整の効果の比較 (屋外環境シーン 2)	65
53	図 40 の一部分を拡大した画像	66
54	屋外環境シーン 2 に関するフレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ と時間的な輝度調整の効果の比較	67
55	被験者実験の結果 (小物体 1)	69
56	被験者実験の結果 (小物体 2)	69
57	被験者実験の結果 (屋外環境シーン 1)	70
58	被験者実験の結果 (屋外環境シーン 2)	70

表 目 次

1	実験で用いた PC の仕様	34
2	各シーンの入力画像の詳細	34
3	シーン別の出力フレーム数	34
4	共通の固定パラメータ	34
5	シーン別の固定パラメータ	34

1. はじめに

複数地点から撮影した画像群を入力とし、任意視点（仮想視点）からの見えを再現する自由視点画像生成 [1,3-8,10-27] は、実シーン仮想化のための基盤技術として注目されており、遠隔地のシーンを仮想的に体感できる VR アプリケーションなど [11,12] への応用が期待されている。

自由視点画像生成の従来手法として、Model-based Rendering (MBR) [1,3,13-19]、3次元形状を援用しない Image-based Rendering (IBR) [4-6,20-24]、三次元形状を援用した IBR [7-10,25,26,28] が提案されている。MBR では、シーンの3次元形状を復元し、複数の入力画像をブレンディングまたは取捨選択することで復元した各3次元部分形状に対応するテクスチャを生成し、それを各部分形状上にマッピングする。このようにして生成したテクスチャ付き3次元モデルを、設定した仮想カメラ上でレンダリングすることで、任意視点からの見えを再現する。この手法では、限られた入力画像から広範囲における自由視点画像を生成できる一方、特に画像のみを入力とした3次元形状の復元精度には限界があり、シーン内の詳細な形状を復元できない場合がある。このとき、生成した自由視点画像において写実性が損なわれるという問題がある。一方、三次元形状を援用しない IBR は、対象シーンを撮影した入力画像群を変形・合成することによって仮想視点からの見え方を再現する。この手法では、仮想視点に届く光線情報として多数かつ空間的に密な入力画像を用意することで、写実性の高い仮想視点からの見えを再現できる。しかし、広範囲における自由視点画像生成を実現するためには、膨大な入力画像が必要となる問題がある。

これらの問題に対して、比較的少ない入力画像から広範囲における写実的な自由視点画像を生成する手法として、3次元形状を援用した IBR が提案されている。そのなかで、仮想視点位置に応じて形状にマッピングするテクスチャを適応的に変化させる視点依存テクスチャマッピング法が、効果的に写実性を高める手法として広く利用されている。この手法は、入力画像群からシーンの3次元形状を復元し、その3次元形状にマッピングするテクスチャとして自由視点画像の各画素に対して見え方が最も近い入力画像の画素を利用している。この手法を用いることで、3次元復元の精度が不十分な場合でも詳細な形状を視覚的に再現すること

が可能である。また、入力画像の光線と仮想視点に届く光線が比較的遠い場合においても、形状を用いて画像上の見えを適切に変換することができるため、比較的広範囲においても写実的な自由視点画像を生成することが可能になる。しかし、従来の視点依存テクスチャマッピング法では1つの仮想視点から対象シーンを見たときの視覚的な違和感の低減を目的としており、視点移動が伴う動画生成においては各フレームの画像が独立に生成される。このため、入力画像取得時の照明条件の差異や3次元形状とカメラパラメータの誤差によって、フレーム間でマッピングされる入力画像が切り替わる際に、物体上の同一箇所のテクスチャが不連続に変化し、生成した動画像がちらつく問題が生じる。

この問題に対し本研究では、視点移動を伴う自由視点動画像の品質の向上を目的とし、フレーム間のテクスチャの整合性を考慮した自由視点動画生成手法を提案する。提案手法では、従来手法で用いられてきたテクスチャ選択の条件に加えて、前フレームで利用されたテクスチャと対象フレームで利用されるテクスチャとの色の差異をコストとして用いたエネルギー関数を新たに設計し、グラフカットアルゴリズムを用いてこれを最小化することで、対象シーン内の同一箇所におけるフレーム間のテクスチャの差異が生じにくい入力画像を各画素に対して選択する。また、従来手法と同様に、異なる入力画像からのテクスチャの境界における空間的な輝度調整を各フレームで行うとともに、時間的な輝度調整を行うことで、空間的・時間的な輝度の不連続を抑制する。実験では、小物体や屋外環境のシーンを対象として視点移動を伴う自由視点動画像を生成することで、追加したテクスチャ選択条件の効果と時間方向の輝度調整の効果を検証する。

本論文では、2章で自由視点画像生成の従来手法を紹介し、それらの研究との関連を踏まえて本研究の位置づけについて述べる。3章では、フレーム間のテクスチャの整合性を考慮した自由視点動画生成の手法について詳述する。4章では、小物体や屋外環境のシーンを対象として、従来手法と提案手法で生成した視点移動を伴う自由視点動画像を比較し、フレーム間の整合性を考慮したデータ項と時間的な輝度調整の効果を検証する。5章では、まとめと今後の展望について述べる。

2. 従来研究と本研究の位置づけ

本章では，自由視点画像生成の従来手法を，シーンの3次元形状を利用する Model-based Rendering (MBR) [1,3,13–19]，3次元形状を援用しない Image-based Rendering (IBR) [4–6,20–24]，3次元形状を援用する IBR [7–10,25,26,28] の3つに大別し，それぞれについて詳述する．また，これらを利用した自由視点動画画像生成について述べ，これらに対する本研究の位置づけについて述べる．

2.1 Model-based Rendering

Model-based Rendering (MBR) とは，対象のテクスチャ付きの3次元モデルを予め生成しておき，設定した視点の画像平面に3次元モデルをレンダリングすることで，仮想視点からの見えを再現する技術である．この手法では自由視点画像の品質が3次元形状復元の精度に大きく依存するため，3次元形状復元を自動で高精度に行う研究が行われている．以下では，レーザレンジファインダ [1,13,14] などを用いる能動的な手法，画像のみから復元を行う受動的な手法について詳述する．

能動的な手法には，Time of Flight(ToF) 法 [13]，光切断法 [14] などがある．ToF 法では，光源から投射されたレーザが対象で反射して戻るまでの時間から距離を算出し，その距離に基づいて対象の3次元位置を算出する (図1)．また，光切断法では，レーザを対象に照射してその反射光をイメージセンサで観測し，光源の位置，物体上におけるレーザの反射位置，イメージセンサの位置から三角測量の原理に基づいてレーザが反射した点の3次元位置を算出する (図2)．これらの手法は3次元形状を高精度に算出できるが，光を吸収したり透過する物体に対してはレーザの反射光を観測できないという問題がある．

受動的な手法には，視体積交差法 [15,16]，Multi View Stereo(MSV) [3,19] などがある．視体積交差法は，対象のシルエットを実空間に投影した視体積の中に対象物体が含まれるという制約 (シルエット制約) に基づいて，2次元のシルエット画像から対象物体の3次元形状を推定する方法である (図3)．視体積交差法では3次元形状の表現にボクセルモデルを使用し，入力画像を取得した全ての視点

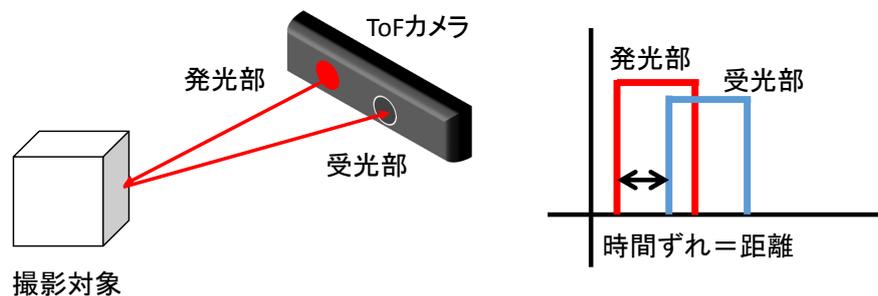


図1 ToF方式

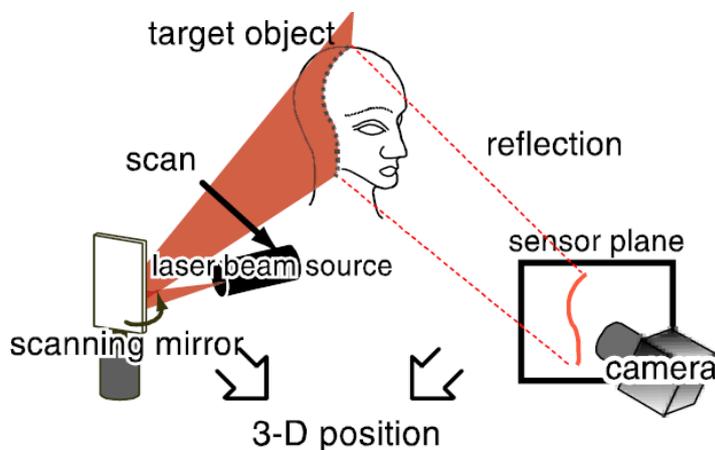


図2 光切断方式 ([1] より)

のシルエット画像を3次元空間に投影する。このときの視体積外のボクセルを削除することで、3次元形状を取得する。この手法では、対象の輪郭を利用することから、後述するMVSなどの手法と比較するとアルゴリズムが単純で高速に計算可能である。また、テクスチャがない対象でも輪郭を抽出できれば3次元形状を復元可能であり、色や照明による誤差が少ない。しかし、視体積交差法には、原理的に凹んだ形状を復元できないため、詳細な形状の復元精度が低いという問題がある。

MVSでは、はじめにStructure from Motion(SfM) [17]で入力画像取得時のカメラの内部・外部パラメータの推定を行い、これに基づいて密な3次元形状の復元を行う。SfMは、対象となるシーンを複数の地点から撮影し、これらの画像間

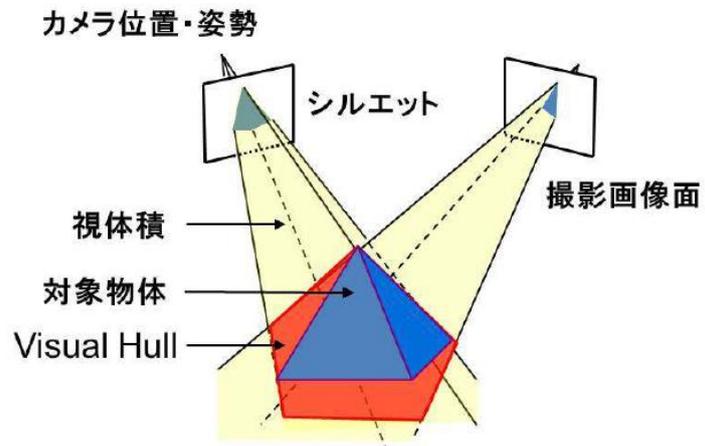


図 3 視体積交差法 ([2] より)



図 4 PMVS で復元した 3 次元形状を利用して生成した自由視点画像 ([3] より)

の特徴点の対応を求め、それぞれのカメラの位置姿勢情報や特徴点の 3 次元位置を推定する。MVS の一手法として、Kutulakos らは Space Carving を提案し

た [18]. この手法では, はじめに対象全体を内包するボクセルボリュームを用意し, ボクセルボリュームの表面上のボクセルを各入力視点の画像へ投影し, 投影箇所の画素値が異なればそのボクセルを削除する. この処理を収束するまで繰り返すことで, 視体積交差法では実現できなかった凹形状の表現も可能にした. また, Patch-based Multi-View(PMVS) では, 画像群から法線付きの 3 次元点を復元し, それをメッシュモデルに変換することで, 詳細な 3 次元形状を取得する [3]. しかし, この手法ではテクスチャが無い点で対応点が取れず, 欠損領域が生じる (図 4). これに対し, Jancosek らが提案した CMPMVS では, 視体積交差法の考え方を取り入れることで, テクスチャのない領域の 3 次元形状の復元を可能にしている [19].

以上のように, これまで様々な三次元復元手法が提案されているが, 植物や毛髪などの詳細な形状を持つ対象の高精度な復元は現状では困難である. このため, MBR において, テクスチャをマッピングした三次元形状をレンダリングした際, 形状の歪みや欠損によって写実性が損なわれる. また, 複数の入力画像をブレンディングまたは取捨選択し, 単一のテクスチャを三次元形状にマッピングするため, 鏡面反射の再現が困難だという問題がある.

2.2 3次元形状を援用しない Image-based Rendering

Image-based Rendering(IBR) とは, 対象のシーンを撮影した入力画像群を変形・合成することによって仮想視点からの見え方を再現する技術であり, 3次元形状を援用しない手法 [4-6, 20-24] と, 3次元形状を援用する手法 [7-10, 25, 26, 28] に分けることができる. IBR では, 仮想視点に届く光線情報として多数かつ空間的に密な入力画像を用意することで, 写実性の高い仮想視点からの見え方を再現できる. 以下では, 3次元形状を援用しない IBR の手法である, 入力画像を変形・合成するモーフィング [4, 20] と, 入力画像群からライトフィールドを取得するライトフィールドレンダリング [5, 6, 21-24, 29, 30] について概説する.

モーフィングとは, 2つの画像間で対応点を指定し, 各画像の混合比に基づいて一方の画像から他方の画像へ滑らかに変化させる技術である [4, 20]. この手法を用いて自由視点画像を生成する場合, 2つの画像の混合比の変化のみではシー

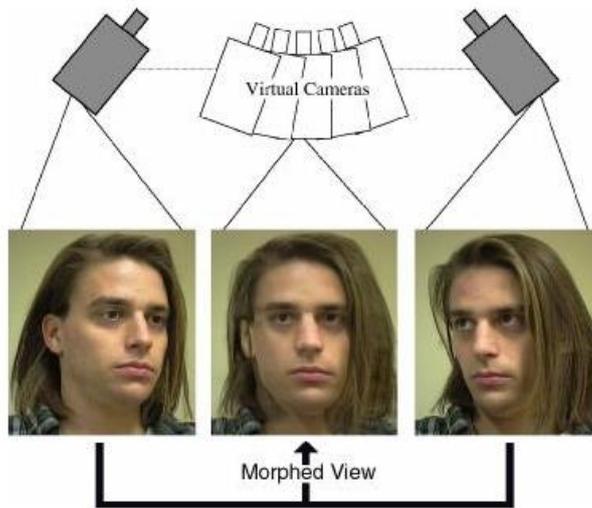


図 5 モーフィング ([4] より)

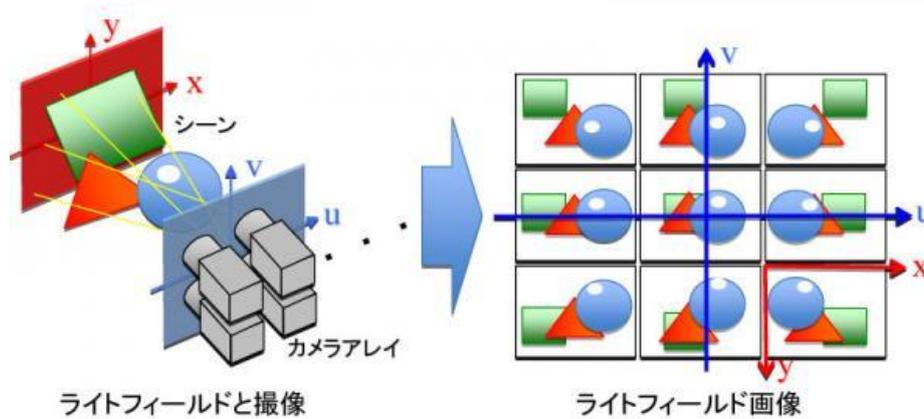


図 6 ライトフィールド ([5] より)

ンの幾何学的構造を再現することができない。そのため、Seitzらは入力画像を平行にしてモーフィングを行い、元の画像面に再変換するという方法で、入力視点間を繋ぐ直線上の仮想視点における幾何学的な構造を一定に保った自由視点画像生成を可能にした [4](図5)。また、Xiaoらは3つの入力視点の画像を用いる手法



図 7 ライトフィールドレンダリングで生成した自由視点画像 ([6] より)

を提案した [20]. しかし, これらの手法は, 仮想視点の位置の範囲が入力画像の視点の近くに限定されるという問題がある.

ライトフィールドとは, 3次元空間内の全ての光線場を指し, 視点の位置 (x, y, z) , 光線方向 (θ, ϕ) , 光線の波長 (λ) , 時間 (t) の7次元で定義される Plenoptic 関数で表現される. あるシーンのライトフィールドを取得することは, 対象シーンの全ての視点からの見えの情報を得ることと同義である (図 6). そのため, ライトフィールドを取得することができれば, 原理的には, 任意の視点に届く光線情報を抽出することで自由視点画像を生成可能である. しかし, 7次元の光線情報を密に取得するためには膨大な入力画像が必要となるため, 完全なライトフィールドを取得するのは困難である. そこで, 7次元で表現される光線情報の次元数を減らし, 光線情報の近似値を取得するための研究が行われている. McMillan らは, 光線の波長 (λ) を RGB 色空間などの画素値で代替し, 時間軸を無視した静止シーンを仮定することで, 視点の位置 (x, y, z) と光線の向き (θ, ϕ) のみで定義した 5D Plenoptic 関数を提案している [21]. また, Naemura らは, 光線が空間内を干渉や減衰などの影響を受けずに直進するものと仮定した場合に, 光線情報が2次元の視点位置 (x, y) と光線の向き (θ, ϕ) の4次元で表現できることを利用して次元数を減らした [22]. さらに, Levoy ら [23] や Gortler ら [24] は, それぞれ Light Field Rendering, Lumigraph を提案し, 4次元のライトフィールドを直線で表した光線が2つの平行した平面を通るときの交点 $(u, v), (s, t)$ の組を定義し

た。これにより、より効率的なレンダリングが可能となった(図7)。

ライトフィールドの取得に単眼カメラを用いる際、各入力視点の相対位置を知るために、複数のカメラを設置する方法やロボットアームを使う方法が用いられる。しかし、これらの方法は容易ではない。このため、カメラの撮像面に複数のレンズを規則的に並べることで一台での撮影を可能にする、ライトフィールドカメラが提案されている [30]。また、Davisらは携帯可能な単眼カメラを用いてライトフィールドを取得するために、最適な位置姿勢をユーザに提示するシステムを提案した [6]。

3次元形状を援用しないIBRでは、自由視点画像を生成可能な視点の位置や、生成される画像の質が入力画像に依存する。このため、広範囲における高品質な自由視点画像生成を実現するためには、膨大な入力画像が必要となる問題がある。

2.3 3次元形状を援用する Image-based Rendering

3次元形状を援用するIBR [7–10, 25, 26, 28]は、入力画像群等からシーンの3次元形状を復元し、その3次元形状にマッピングするテクスチャを入力画像群から適応的に選択することで、3次元形状の復元精度が不十分な場合でも詳細な形状を視覚的に再現可能にする技術である。3次元形状を援用しない手法に比べて、比較的少ない入力画像であっても、広範囲で写実的な自由視点画像を生成できる利点がある。以下では、この手法で広く用いられている視点依存テクスチャマッピングとその拡張について述べる。

視点依存テクスチャマッピングとは、MVSなどで復元した3次元形状に対してマッピングするテクスチャを、自由視点画像の画素毎に入力画像群から適応的に選択する方法である。Debevecら [7]は、マッピングするテクスチャの選択基準として、仮想視点と各入力視点における光線方向の類似度を利用している。この手法では、図8に示すように、仮想視点の投影中心と対象の3次元形状上の点を結ぶ直線と、さらにその点と入力画像の投影中心を結ぶ直線の成す角を算出し、この角度が最も小さくなる入力画像が選択される。源田ら [25]は、この選択基準に対象物体の3次元形状の法線情報に基づいた選択基準を加えることで、生成する画像の質を向上させた。また、高井ら [8]は、高ズームで撮影された入力画

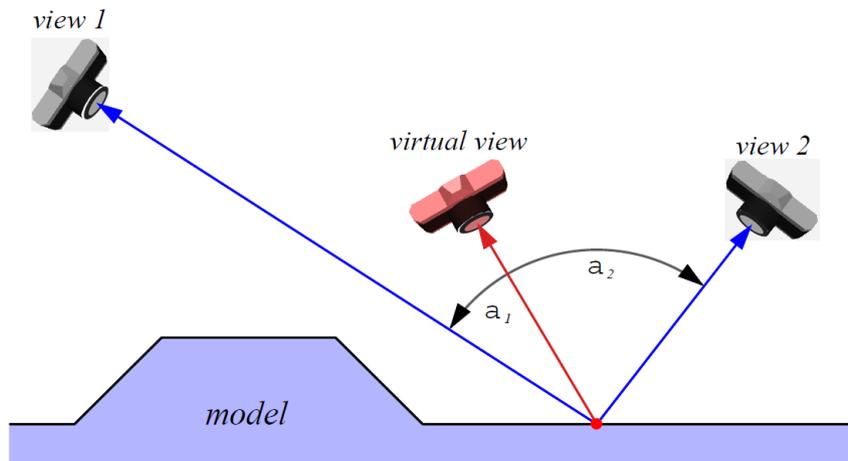


図 8 Debevec らの視点依存テクスチャマッピング ([7] より)

像を優先的に選択することで，高精細なテクスチャを生成する手法を提案した．しかし，視点依存テクスチャマッピングでは，入力画像取得時の照明条件の差異によって対象の同一箇所を投影した画素の値にばらつきが生じることや，3次元形状とカメラパラメータの誤差によって，図9に示すようにテクスチャの境界において空間的な画素値の不連続が生じる場合がある．この問題に対し，Debevec ら [26] はテクスチャの境界において画素値が滑らかに変化するように，複数枚の入力画像をブレンドすることで空間的な不連続を抑制している．しかしこの方法では，3次元形状とカメラパラメータの誤差がある場合にテクスチャの投影位置がずれてしまい，ブレンドによってテクスチャのぼけが生じる問題がある．この問題を低減する手法として，片桐ら [28] はブレンドを行わずにテクスチャの境界の空間的な画素値の不連続を解消する方法を提案した．また，Kopf [9] らは，鏡面反射があり，3次元形状の高精度な復元が困難なシーンにおいて高品質な画像を生成する手法を提案した (図 10)．

2.4 自由視点動画画像生成

前述の自由視点画像生成手法を利用して動画画像を生成する研究について述べる．Kopf [10] らは，撮影時のカメラパスに基づいてブレの少ない滑らかな仮想のカメラ



図 9 画素値の不連続が生じた自由視点画像 ([8] より)



図 10 鏡面反射を考慮した視点依存テクスチャマッピング ([9] より)

ラパスを設定し，撮影した動画像のフレームから適切なテクスチャを利用して動画像を生成する手法を提案した (図 11)．しかし，この手法では仮想視点の位置を入力視点の周辺に限定しているため，汎用的に用いることが困難である．また，事前に仮想カメラのパスを決定し，前述の 3 次元形状を援用する IBR を用いて自由視点動画像を生成することは可能であるが，これらの手法は単眼での高品質な自由視点画像生成を目的としており，視点移動が伴う動画像生成においては各フ



図 11 Kopf らのタイムラプス動画像生成 ([10] より)

フレームの画像が独立に生成される．このため，入力画像取得時の照明条件の差異や3次元形状とカメラパラメータの誤差によって，フレーム間でマッピングされる入力画像が切り替わる際に，物体上の同一箇所テクスチャが不連続に変化し，生成した動画像がちらつく問題が生じる．

2.5 本研究の位置づけ

本研究では，片桐らの視点依存テクスチャマッピングをベースとし，時間的なテクスチャの不連続を低減した自由視点動画像を生成する手法を提案する．視点依存テクスチャマッピングを採用することで3次元形状の復元精度が高くない場合でもシーンの見えを写實的に再現することが可能であるが，その中でも片桐らの手法はテクスチャの空間的な不連続を低減した高品質な自由視点画像生成を行うことが可能なため，これを採用する．提案手法では，フレームの前後で対象物体の同一箇所を投影する画素の画素値が大きく変化しないようテクスチャを選択することで，フレーム間のちらつきを抑制できるとの考えから，前フレームのテクスチャと類似したテクスチャを選択する条件を考慮する項を，片桐らが提案したエネルギー関数に加える．これをグラフカットアルゴリズムを用いて最小化することによって，空間的・時間的な画素値の不連続を抑制した高品質な自由視点動画像を生成する．また，エネルギー関数に基づく動画像生成では空間的・時間的な不連続を解消しきれない場合があるため，空間的・時間的な輝度調整を行う．

提案手法では、予め仮想カメラのパスを決定し、設計したエネルギー関数を用いて1フレーム目から逐次的に画像を生成し、これをつなぎ合わせることで動画像を生成する。

3. フレーム間のテクスチャの整合性を考慮した自由視点動画画像生成

本章では、片桐らが提案した視点依存テクスチャマッピング [28] を拡張した、動画画像のテクスチャちらつき（時間的なテクスチャの不連続）を抑制する自由視点動画画像生成手法について述べる。以下では、はじめに提案手法について概説する。次に、フレーム間のテクスチャの整合性を考慮したエネルギー関数について詳述する。最後に、エネルギー関数を用いた処理では抑制しきれない時間的・空間的なテクスチャの不連続を抑制するための輝度調整について述べる。

3.1 提案手法の概要

図 12 に提案手法の処理の流れを示す。提案手法では、まず、実空間のシーンを対象とし、その対象を複数の地点から撮影した画像群を入力し、入力画像を用いて前処理を行う。次に、ユーザが仮想カメラの動きを設定したうえで、各視点からの対象の見えを静止画像として 1 フレーム目から逐次的に生成する。最後に、それらをつなぎ合わせた動画画像を出力する。

前処理として、入力画像を取得したときのカメラの内部・外部パラメータ推定と対象の 3 次元形状の復元、入力視点のデプスマップ作成を行う。本研究では、まず Visual SfM [31] を用いて、カメラの内部・外部パラメータを推定する。具体的には、対象を複数の地点で撮影した画像 I_n からなる入力画像群 $S = \{I_n | n = 1, \dots, N\}$ から SIFT 特徴点 [32] を抽出し、Visual SfM により画像を撮影したカメラの内部パラメータ \mathbf{K}_n 、および任意の座標系におけるカメラの外部パラメータの回転行列 \mathbf{R}_n 、並進ベクトル \mathbf{t}_n を推定する。本研究では、Visual SfM で求められたカメラの内部・外部パラメータの座標系をワールド座標系とする。次に、入力画像群と得られたカメラの内部・外部パラメータから、CMPMVS [19] によって対象のシーンの 3 次元メッシュモデルを復元する。最後に、カメラの内部・外部パラメータと復元された 3 次元メッシュモデルを利用して、入力視点のデプスマップ作成を行う。



図 12 提案手法の概要

各視点における画像生成処理では、2章で紹介した片桐らの手法 [28] を拡張したエネルギー関数の最小化による各画素に対する最適な入力画像ラベル選択処理と、画素値の空間的・時間的な輝度調整を行う。以下では、対象フレーム k に対する自由視点画像生成について述べる。ここではまず、復元された3次元メッシュモデルから対象フレーム k におけるデプスマップを作成する。次に、対象フレーム k 中の画素 i_k に対応するデプスマップの値 d_{i_k} および、対象フレーム k における仮想カメラの内部パラメータ \mathbf{K}_{V_k} 、外部パラメータ \mathbf{R}_{V_k} 、 \mathbf{t}_{V_k} を利用して、画素 i_k に対応する入力画像群 S 内の画像中の画素を決定し、これらの中から単一の画素を選択したうえで、その値を割り当てることで、画素 i_k の値を決定することを基本とする。このときの画像選択の基準として、下記5点の条件を考慮する。

- (1) 3次元形状は視点位置によって見え方が異なるため、対象フレームの視点に対して見え方が類似していると考えられる、対象フレームから対象物体への光線に最も近い光線方向を持つ入力視点の画像を利用する。
- (2) 対象シーンの同一箇所ではフレーム間で大きな輝度変化が起こりにくいと考えられるため、3次元形状中の同一箇所に対応する対象フレームおよびその直前フレームの画素間で、値が大きく変化しないような入力画像を利用する。
- (3) 3次元形状上のある点は、自己遮蔽などが原因で入力視点の位置によって観測できていない可能性がある。このため、各画素に対して値を割り当てる際には、その画素に対応する3次元形状上の点を観測できている入力視点の画像のみを利用する。
- (4) 対象フレームは複数の入力画像の画素値から構成されるため、異なる入力画像の境界において不連続にテクスチャが切り替わる可能性がある。この問題を低減するため、対象フレームの隣接する画素に異なる入力画像が選択される場合には、隣接する画素それぞれに選択される入力画像上で、3次元形状上の同一箇所を投影している画素をそれぞれ対応付け、これらの対応する画素が類似している入力画像を選択する。

- (5) 3次元形状が空間的に大きく変化する場合、対象の材質の違いや光の当たり方により、視点の位置によって見え方が変化する可能性がある。このため、隣接する画素間に対応する3次元形状上の法線の向きが大きく異なる場合、条件(4)は緩和される。

提案手法では、対象フレーム k の画素 i_k への値の割り当てに利用する入力画像 x_{i_k} を選択する問題を、上記を考慮したエネルギー関数の最小化問題として定式化する。具体的には、 $X = \{x_{i_k} | x_{i_k} = 1, \dots, N\}$ の要素 x_{i_k} を i_k の画素値の決定に利用される画像のラベル $x_{i_k} \in \{1, \dots, N\}$ として、下記のエネルギー関数を定義する。

$$\begin{aligned}
J(X) = & \sum_{i_k \in A_k} \alpha_{i_k}(x_{i_k}) \{f_{i_k}(x_{i_k}) + K_h h_{i_k}(x_{i_k})\} \\
& + \sum_{(i_k, j_k) \in B_k} (K_\beta \beta_{i_k j_k} + K_g) g_{i_k j_k}(x_{i_k}, x_{j_k})
\end{aligned} \tag{1}$$

ここで、データ項の $f_{i_k}(x_{i_k})$ は条件(1)に対応しており、 i_k に届く光線と入力画像 x_{i_k} 上で i_k と3次元形状上の同一箇所を投影している画素に届く光線の角度が大きいほど大きな値となる。また、 $h_{i_k}(x_{i_k})$ は条件(2)に対応しており、 i_k に対応する直前フレームの画素 $D_{k-1}(i_k)$ との画素値の差が大きいほど大きな値となる。この項によって、3次元形状とカメラパラメータの誤差の影響による時間的なテクスチャの不連続を低減する。 $\alpha_{i_k}(x_{i_k})$ は条件(3)に対応しており、 i_k に対応する3次元形状上の点を観測できる場合は1、そうでない場合には大きな値となる。平滑化項では、 $g_{i_k j_k}(x_{i_k}, x_{j_k})$ が条件(4)に対応しており、隣接する画素 i_k, j_k 間で異なる入力視点の画像が選択される際、 i_k, j_k のそれぞれに選択される画像 x_{i_k}, x_{j_k} 間で i_k, j_k のそれぞれと同一箇所を投影する画素を対応付け、それらの画素間の値の差が大きいほど大きな値となる。また、 $\beta_{i_k j_k}$ は条件(5)に対応しており、隣接する画素間でそれぞれに対応する3次元形状上の法線方向の差が閾値より小さい場合は1、そうでない場合は0となる。式(1)中の A_k は対象フレーム k で画素値を割り当てる全ての画素のインデックスの集合、 B_k は対象フレーム k で隣接する画素のインデックスのペアの集合、 K_h, K_β, K_g はそれぞれ重みパラメー

タを表す。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 [33] で最小化することができる。

上記のエネルギー最小化により入力画像群から得られる画素値を直接利用した場合、入力画像取得時の照明条件の差異や3次元形状とカメラパラメータの誤差によって、対象フレームの自由視点画像中に生じる空間的な輝度の不連続や、フレーム間でマッピングされる入力画像が切り替わる際の、物体上の同一箇所の時間的なテクスチャの不連続を抑制しきれない可能性がある。このため提案手法では、対象フレームの各画素に対して入力画像ラベル選択処理を行った後、空間的・時間的な輝度調整を行う。これによって、入力画像ラベル選択処理で抑制しきれなかった空間的・時間的な不連続を低減する。

以降の節では、エネルギー関数のデータ項と平滑化項についてそれぞれ述べた後、空間的・時間的な輝度調整について詳述する。

3.2 エネルギー関数のデータ項

この節では、対象フレームの視点と入力視点の光線の方向の近さ、フレーム間のテクスチャの整合性、対象フレームの視点における各画素に対応する3次元形状上の点に関する各入力視点からの観測の可否に基づいたデータ項について詳述する。

3.2.1 仮想視点と入力視点の光線の方向の近さに関する項

視点依存テクスチャマッピングでは、3次元形状にマッピングする適切な入力画像の選択基準として、仮想視点画像と入力画像を取得したときの光線の方向の近さが利用される。これによって、見え方が近いと考えられる入力画像を仮想視点のテクスチャとして利用する。代表的なものに、2章で紹介した Debevec らの手法 [7] がある。Debevec らの手法では、図 13 に示すように、仮想視点の各画素からその画素に対応する3次元形状上の点へのベクトルと、その点に対応する入力視点の画素からその点へのベクトルが成す角度の大きさを基準としている。この角度を全入力視点について計算し、最小となる入力視点の画像を選択する。

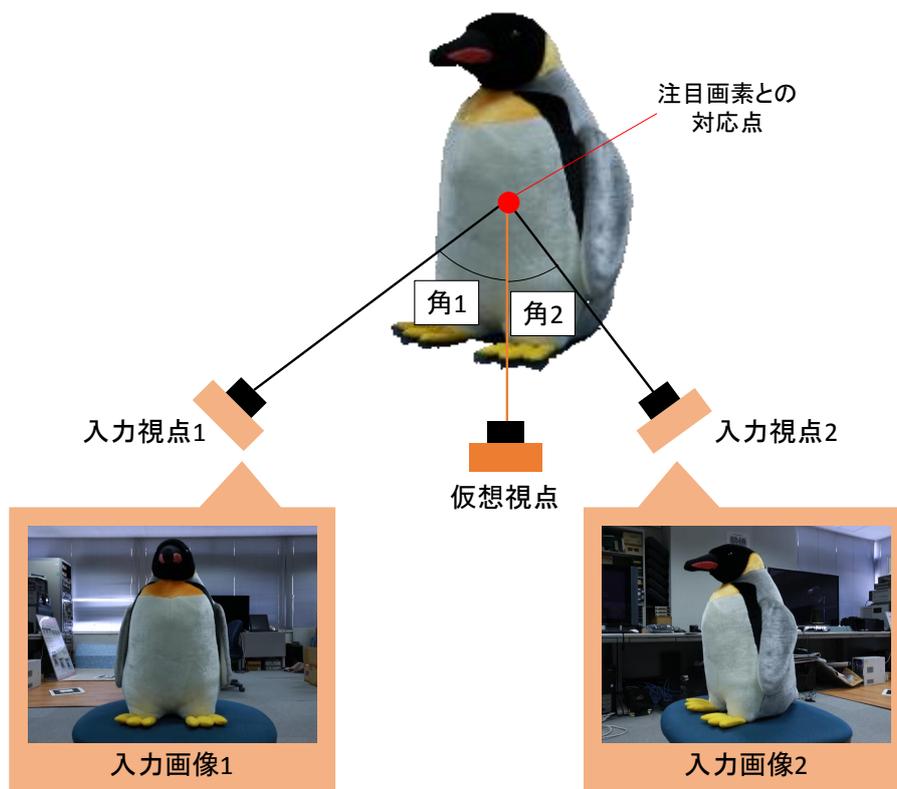


図 13 仮想視点と入力視点の見え方の近さ

提案手法でも Debevec らの手法の選択基準を利用する．入力画像 I_n の視点位置 $\mathbf{t}'_n = -\mathbf{R}_n^T \mathbf{t}_n$ ，対象フレーム k の仮想視点位置 $\mathbf{t}'_{V_k} = -\mathbf{R}_{V_k}^T \mathbf{t}_{V_k}$ ，および画素 i_k に対応するデプス値 d_{i_k} を用いて求められるワールド座標系における対象物体の3次元形状上の点 \mathbf{p}_{i_k} から，2つのベクトルが成す角 $\theta_{i_k n}$ は次式によって得られる．

$$\theta_{i_k n} = \cos^{-1} \frac{(\mathbf{t}'_{V_k} - \mathbf{p}_{i_k}) \cdot (\mathbf{t}'_n - \mathbf{p}_{i_k})}{\|\mathbf{t}'_{V_k} - \mathbf{p}_{i_k}\| \|\mathbf{t}'_n - \mathbf{p}_{i_k}\|} \quad (2)$$

この成す角を用いて，データ項 $f_{i_k}(x_{i_k})$ を次式により与える．

$$f_{i_k}(x_{i_k}) = \theta_{i_k x_{i_k}} \quad (3)$$

この項は成す角 $\theta_{i_k x_{i_k}}$ が小さい場合に小さくなるため，仮想視点画像と入力画像を取得したときの光線方向が近く，見え方が近いと考えられる入力視点の画像が選択されやすくなる．

3.2.2 フレーム間のテクスチャの整合性に関する項

3次元形状上の同一箇所テクスチャがフレーム間で不連続に変化することを抑制し、フレーム間のテクスチャの整合性を保持するため、前フレームで選択されたテクスチャとの類似度をマッピングする入力画像の選択基準として利用する。

図14に示すように、各フレームの画像を独立に生成した場合、3次元形状とカメラパラメータの誤差によって、フレーム間でマッピングされる入力画像が切り替わる際に物体上の同一箇所テクスチャに大きな差異が生じてしまう場合がある。図14の例では、対象物体の嘴の赤い部分や目のテクスチャの位置がフレーム間で異なっていることが分かる。これは、自由視点動画画像のテクスチャちらつきの原因となる。そこで、対象フレーム k と直前フレーム $k-1$ で3次元形状上の同一箇所を観測する画素 i_k と $D_{k-1}(i_k)$ を対応付け、これらの画素間の類似度をフレーム間のテクスチャの整合性に関する項 $h_{i_k}(x_{i_k})$ として次式で定義する。

$$h_{i_k}(x_{i_k}) = \|\mathbf{c}_{i_k x_{i_k}} - \mathbf{c}_{D_{k-1}(i_k) q(D_{k-1}(i_k))}\|_{\infty} \quad (4)$$

ここで、 $\mathbf{c}_{i_k x_{i_k}}$ は対象フレーム k の自由視点画像上の画素 i_k に対応する入力画像 $I_{x_{i_k}}$ 上の画素における画素値、 $\mathbf{c}_{D_{k-1}(i_k) q(D_{k-1}(i_k))}$ は前フレームの処理において、対応関係が既に決定済みの画素 $D_{k-1}(i_k)$ に対応する入力画像 $q(D_{k-1}(i_k))$ 上の画素における画素値を表す。 $\|\cdot\|_{\infty}$ 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、図15の1フレーム目と2フレーム目のように、3次元形状上の同一箇所の色が類似した入力画像ラベルが選択される場合には値が小さくなるが、2フレーム目と3フレーム目のように色が大きく異なる入力画像ラベルが選択される場合にはペナルティを与える。これによって、前フレームのテクスチャとの整合性を保持することが可能となり、動画画像のちらつきが低減される。

3.2.3 可視判定に関する重み

可視判定では、対象フレーム k の自由視点画像中の各画素に対応する3次元形状上の点が各入力画像 I_n の視点から観測されているか判定する。図16における青色の領域は I_n の入力視点から観測できない領域を示しており、このような領域

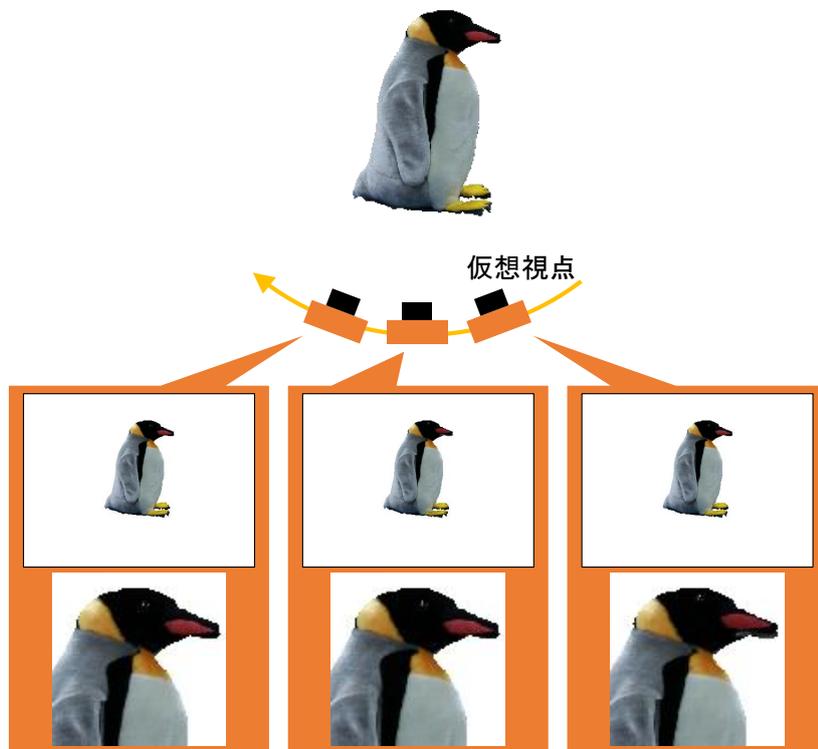


図 14 従来手法を用いた場合のフレーム間のテクスチャの不整合

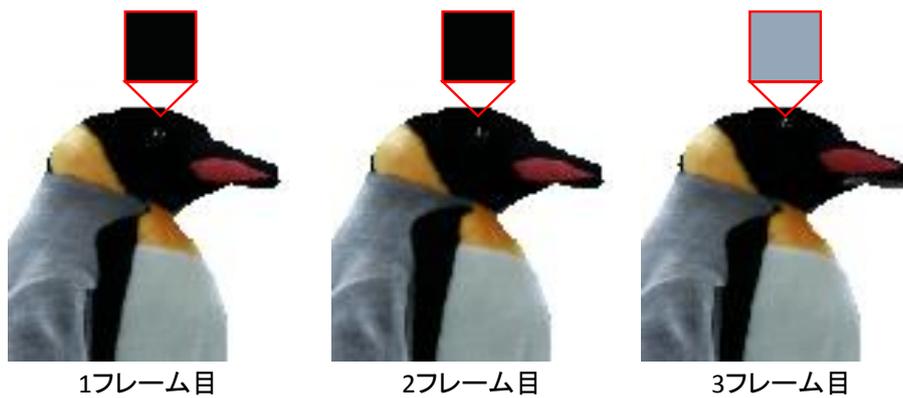


図 15 フレーム間のテクスチャの整合性

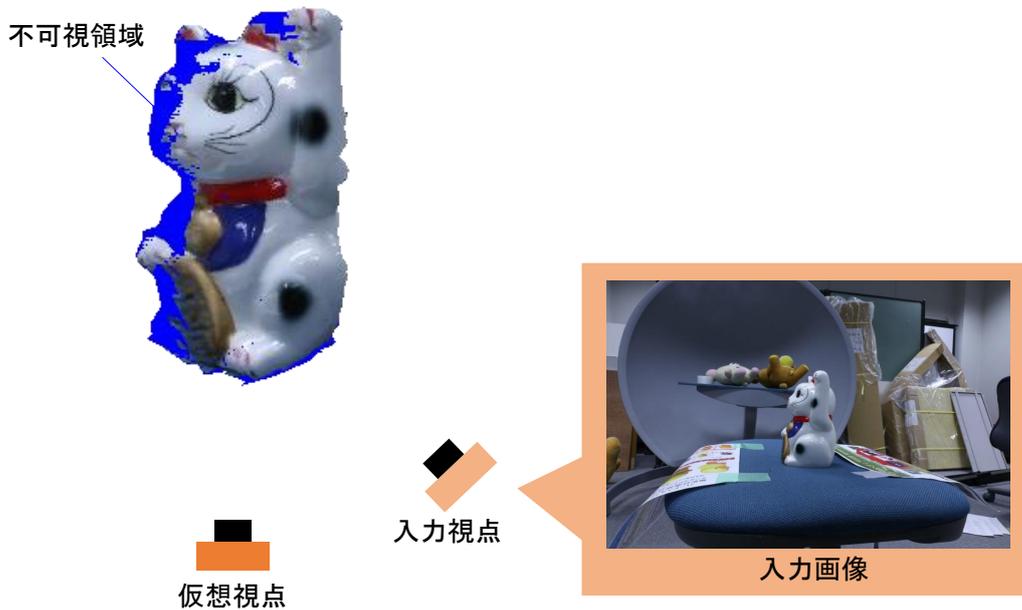


図 16 可視判定の例

には I_n を対象フレーム k の自由視点画像のテクスチャとして利用しないようにする。

可視判定では、3.2.1 節で述べた手法と同様に対象フレーム k における画素 i_k に対応する 3次元形状上の点 \mathbf{p}_{i_k} を求め、さらに \mathbf{p}_{i_k} に対応する入力画像上の画素を求め、これらの画素に対応するデプス値の比較を行う。仮想視点と入力視点それぞれのデプスマップから、 i_k とこれに対応する入力画像の画素のデプス値 $d'_{i_k n}$, e_n を算出すると、これらのデプス値の差が小さい場合に点 \mathbf{p}_{i_k} が I_n の視点から観測されたと考えることができる。一方、デプス値の差が大きい場合は、何らかの遮蔽物があり、点 \mathbf{p}_{i_k} が I_n の視点から観測されていないと考えられる。入力画像 $I_{x_{i_k}}$ における点 \mathbf{p}_{i_k} の可視判定に関する重み $\alpha_{i_k}(x_{i_k})$ を次式により定義する。

$$\alpha_{i_k}(x_{i_k}) = \begin{cases} 1 & \text{if } |d'_{i_k x_{i_k}} - e_{x_{i_k}}| < \theta_d \\ K_\alpha & (\text{otherwise}) \end{cases} \quad (5)$$

ここで、 θ_d は誤差を許容するための閾値であり、 K_α は \mathbf{p}_{i_k} が $I_{x_{i_k}}$ の視点から観測されていない場合のペナルティである。可視判定の結果、 \mathbf{p}_{i_k} が $I_{x_{i_k}}$ の視点から観測されていないと判定された場合には $I_{x_{i_k}}$ は利用できないため、 K_α には十

分に大きな値を与える。

3.3 エネルギー関数の平滑化項

この節では、片桐らの手法 [28] を利用した、対象フレームの自由視点画像の各画素に割り当てられた入力画像ラベルの境界における画素値の連続性 (空間的な画素値の連続性) と、メッシュモデルの法線に基づいた空間的な形状の変化の大きさに基づく平滑化項について詳述する。

3.3.1 テクスチャ境界の画素値の不整合に関する項

図 17 に示すように、対象フレームの隣接する画素間において異なる入力画像が選択され、それぞれの画素のテクスチャとして選択された入力画像間で、撮影対象の同一箇所を投影する画素の値が大きく異なる場合、テクスチャの空間的な不連続が生じる。そこでこの項では、対象フレーム上で隣接する画素それぞれにテクスチャとして選択される 2 つの入力画像について、画像上の同一箇所を投影する画素間で値の類似度を算出し、これを利用することで空間的な不連続を抑制する。テクスチャの境界の画素値の不整合に関する項を次式で定義する。

$$g_{i_k j_k}(x_{i_k}, x_{j_k}) = \max(\|\mathbf{c}_{i_k x_{i_k}} - \mathbf{c}_{i_k x_{j_k}}\|_\infty, \|\mathbf{c}_{j_k x_{i_k}} - \mathbf{c}_{j_k x_{j_k}}\|_\infty) \quad (6)$$

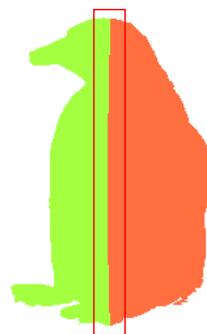
ここで、 $\mathbf{c}_{i_k x_{i_k}}$ は対象フレーム k の画素 i_k に対応する入力画像 $I_{x_{i_k}}$ 上の画素における画素値を表す。 $\|\cdot\|_\infty$ は上限ノルムであり、ベクトルの要素の最大値を与える。この項は、対象フレーム k の隣接する画素に異なる入力画像ラベルが割り当てられ、それらの同一箇所を投影する画素間で画素値が大きく異なる場合にペナルティを与える。

3.3.2 3次元形状の空間的な変化の大きさに関する重み

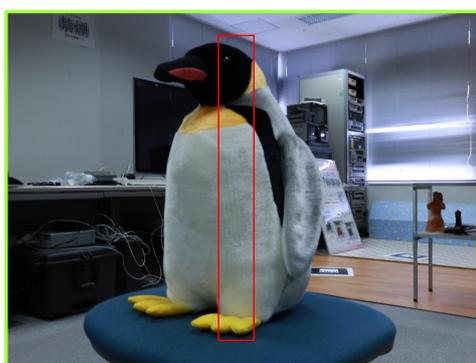
図 18 の赤線で示す領域のように、対象の 3次元形状が空間的に大きく変化している箇所では、その表面を構成する材質や光の当たり方が大きく異なると考え



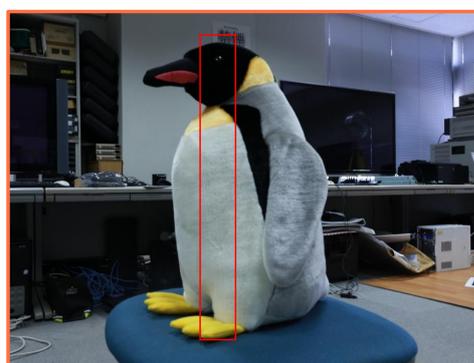
自由視点画像



入力画像ラベル選択結果画像



選択された入力画像(黄緑)



選択された入力画像(橙)

図 17 テクスチャの空間的な不連続

られる。このような領域では、画素値が大きく異なるテクスチャへ切り替わった場合でも視覚的な違和感が生じにくいと考えられる。そのため、この項では3次元形状が空間的に大きく変化する領域におけるテクスチャの切り替わりを許容する重みづけを行い、画素値の空間的な不連続を抑制することにより生じる歪みを緩和する。3次元形状の空間的な変化を表す尺度として、対象フレームの各画素に対して割り当てられた法線方向の差を利用する。具体的には、隣接する画素に割り当てられた2つの法線を利用し、これらの法線の成す角が大きい場合に、3次元形状の空間的な変化が大きいとする。

具体的な重みを決定するにあたって、まず、CMPMVS で得られた対象のメッシュモデルを構成する面の法線を求める。図 19 に示すように、ある面を構成する頂点を p_1 , p_2 , p_3 とし、これらの頂点が面の表側から見たときにこの順番で



図 18 3次元形状の空間的な変化が大きいシーンの例

反時計周りに割り当てられているものとした場合，この面の法線 \mathbf{n} は次式により求められる．

$$\mathbf{n} = (\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p}_3 - \mathbf{p}_2) \quad (7)$$

次に，対象フレーム k の画像平面上に投影されたメッシュモデルを構成する面の法線を各画素に割り当てる．このとき，メッシュモデルを復元する際の誤差によって，実際の形状が滑らかであっても隣接する画素間で法線方向の差が大きくなる場合がある．そのため，事前に対象フレーム k に対する法線マップを生成し，このマップに対してガウシアンフィルタを用いた平滑化処理を行う．最後に，3次元形状の空間的な変化の大きさに関する重みを次式で算出する．

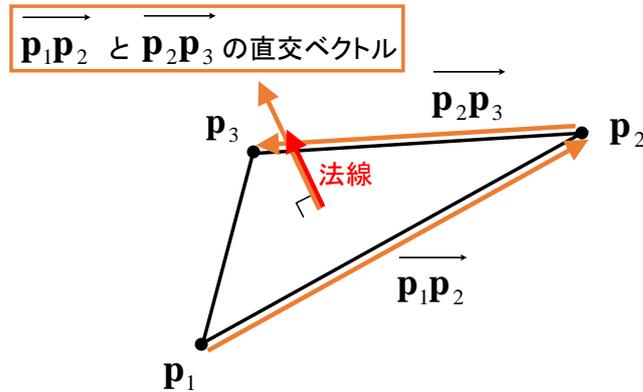


図 19 法線取得

$$\beta_{i_k j_k} = \begin{cases} 1 & \text{if } \cos^{-1} \frac{\mathbf{n}_{i_k} \cdot \mathbf{n}_{j_k}}{\|\mathbf{n}_{i_k}\| \|\mathbf{n}_{j_k}\|} < \theta_n \\ 0 & \text{(otherwise)} \end{cases} \quad (8)$$

ここで、 \mathbf{n}_{i_k} 、 \mathbf{n}_{j_k} はそれぞれ画素 i_k 、 j_k に割り当てられたガウシアンフィルタ適用後の法線である。隣接する画素に割り当てられた法線の成す角が閾値 θ_n よりも大きければ3次元形状の変化が大きいと判断する。3次元形状の変化が大きいと判断された箇所では、入力画像ラベルの切り替えを許容するためこの重みは0になり、入力画像ラベル選択に対する式(6)の項の影響が弱まる。

3.4 空間的・時間的な輝度調整

式(1)のエネルギー関数をグラフカットアルゴリズムにより最小化することで、対象フレームの空間的・時間的な画素値の不連続は低減できると考えられる。しかし、図20や図21に示すように、入力画像取得時の照明条件の差異や3次元形状とカメラパラメータの誤差によって、空間的・時間的不連続を抑制しきれない領域が存在する場合がある。例えば、図20では対象の足元に明るいテクスチャと暗いテクスチャがマッピングされており、このテクスチャの境界で空間的な不連続が発生していることが分かる。また、図21では、シーン中に含まれる三角形の物体の表面にマッピングされているテクスチャが1フレーム目では明るい灰色であるのに対し、2フレーム目で暗い灰色になっており、フレーム間で時間的な

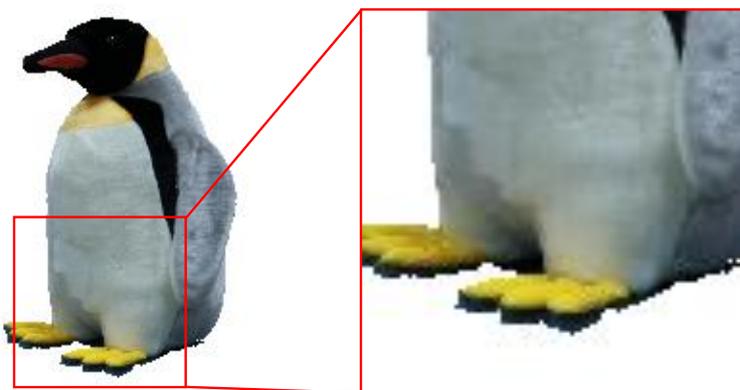


図 20 空間的な画素値の不連続が残ったフレーム

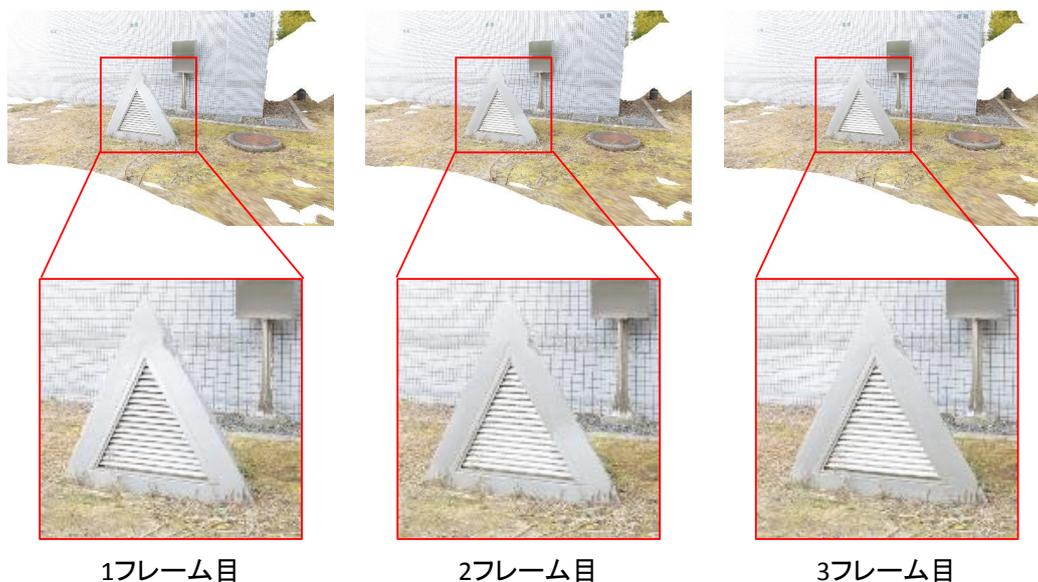


図 21 時間的な画素値の不連続が残ったフレーム

不連続が発生していることが分かる。これらの問題に対し、空間的・時間的な輝度調整を行うことで、画素値の不連続を解消する。

具体的には、対象フレームと過去の数フレームで、予め入力画像ラベル選択処理とラベル選択結果に対する各画素の空間的な輝度調整の補正值の算出を行い、

この補正値を対象フレームの時間的な輝度調整に利用することで、空間的・時間的な輝度調整を行う。また、空間的な輝度調整の補正値は各フレームの入力画像ラベル選択処理後に算出し、過去のフレームで算出した輝度調整の補正値を対象フレームの時間的な輝度調整に利用する際には、対象フレームで同一の箇所を投影する画素位置に対応付けて利用する。本節では、空間的な輝度調整および、空間的な輝度調整の補正値を利用した時間的な輝度調整について詳述する。

3.4.1 空間的な輝度調整

空間的な画素値の不連続を解消する方法として、同一画素に対応する複数画素値をブレンディングする方法がある。しかし、このようなブレンディング処理を行うと、3次元形状やカメラパラメータの誤差によって異なる入力画像上の点が対象シーンの同一箇所に対応しない場合、表面の構造やテクスチャが二重になるなどの問題が生じる。そのため本研究では、従来のブレンディングを行わない片桐らの空間的な輝度調整手法 [28] を利用する。この手法では、注目画素周辺の局所的な輝度平均と、注目画素周辺で各入力画像ラベルが割り当てられた画素の割合に基づいて注目画素の輝度を調整する割合 (補正値) を決定し、テクスチャの境界部分の画素値の差を小さくしていくことで、空間的な画素値の不連続を解消する。

具体的には、まずフレーム k の注目画素 i_k に対してウィンドウを設定し、そのウィンドウ内に含まれる画素に割り当てられた入力画像ラベルの集合を Q_{i_k} とし、入力画像ラベル $n \in Q_{i_k}$ について、ウィンドウ内で n が割り当てられた画素の割合を $\gamma_{i_k n}$ とする。また、ウィンドウ内の全ての画素の入力画像ラベルを n としたときのウィンドウ内の平均輝度を $\bar{I}_{i_k n}$ とする。注目画素 i_k のラベルを $q(i_k)$ とすると、空間的な輝度調整の補正値は次式で与えられる。

$$b_{i_k} = \frac{1}{\bar{I}_{i_k q(i_k)}} \sum_{n \in Q_{i_k}} \gamma_{i_k n} \bar{I}_{i_k n} \quad (9)$$

この補正値は、ウィンドウ内の全ての画素に同じラベルが割り当てられている際には、注目画素の画素値をそのまま出力する。また、図 22 に示すように、テクスチャが切り替わる領域で画素 i_k に入力画像 (黄緑) のように平均輝度が低い入

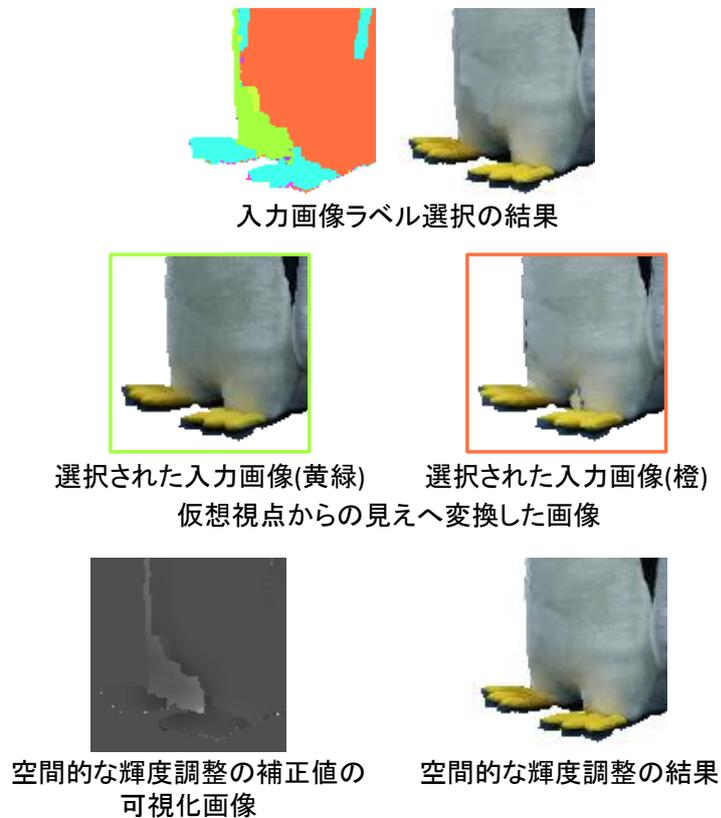


図 22 空間的な輝度調整

力画像が選択され、そのときのウィンドウ内に入力画像 (橙) のように平均輝度が高い入力画像が選択されている画素が多い場合、 b_{i_k} によって補正された輝度値は画素 i_k の値よりも大きくなる。反対に、画素 i_k に入力画像 (橙) が選択され、そのときのウィンドウ内に入力画像 (緑) が選択されている画素が多い場合、 b_{i_k} によって補正された輝度値は画素 i_k の値よりも小さくなる。

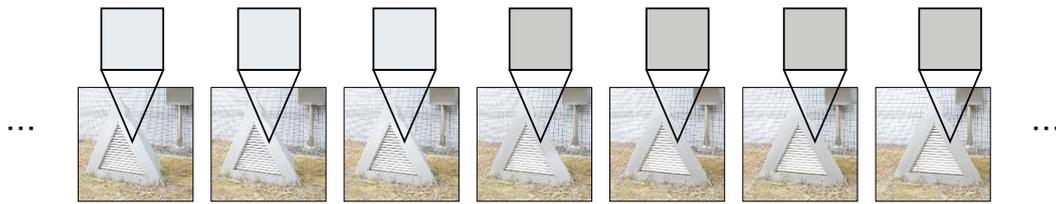
3.4.2 時間的な輝度調整

時間的な画素値の不連続を解消する方法として、時系列画像間に対応する画素値を用いて時間的なブレンディングを行う方法がある。しかし、時間方向に対して単純なブレンディング処理を行うと、3次元形状やカメラパラメータの誤差によって対応点が正しく与えられない場合、生成した自由視点画像をブレンディ

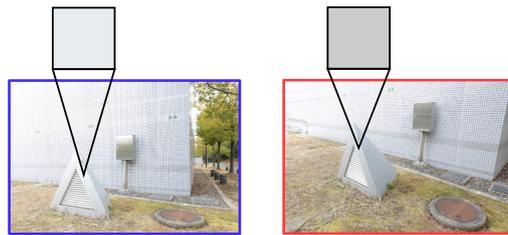
ングすることによってテクスチャがぼけたり二重像が生じるという問題がある。そのため、本研究では過去フレームで生成した自由視点画像を直接用いる代わりに、対象フレームと過去の数フレームの入力画像ラベル選択結果を利用して時間的な輝度調整を行う。これにより、照明条件の差異による時間的な画素値の不連続を解消する。また、各フレームの空間的な輝度調整の補正值を利用することで、空間的な輝度調整も同時に行う。具体的には、以下の式によって、対象フレーム k の注目画素 i_k に対して空間的・時間的な輝度調整を行った値 \mathbf{c}'_{i_k} を算出する。

$$\mathbf{c}'_{i_k} = \frac{1}{s} \sum_{l=k-s+1}^k b_{D_l(i_k)} \mathbf{c}_{i_k q(D_l(i_k))} \quad (10)$$

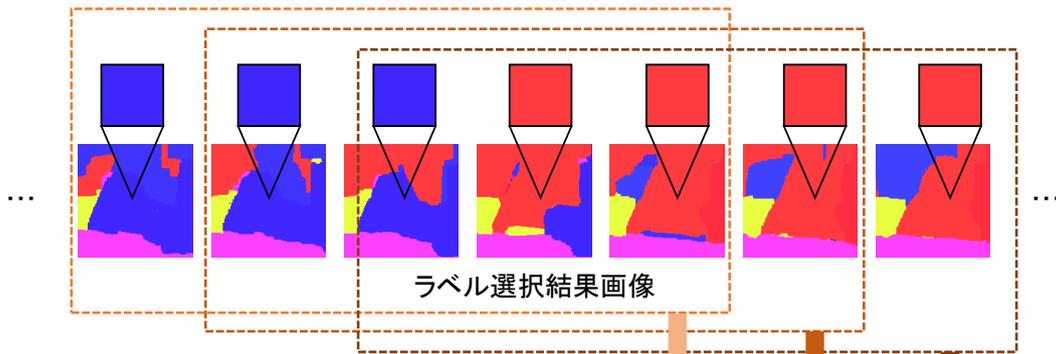
ただし、空間的な輝度調整に利用するフレーム数を s 、輝度調整に利用するフレームを $k-s+1 \leq l \leq k$ とする。また、 i_k と 3次元形状上の同一箇所を投影するフレーム l 中の画素を $D_l(i_k)$ 、 $D_l(i_k)$ で選択された入力画像ラベルを $q(D_l(i_k))$ 、 $D_l(i_k)$ の空間的な輝度調整の補正值を $b_{D_l(i_k)}$ 、入力画像 $I_{q(D_l(i_k))}$ 中の i_k と対象の同一箇所を投影する画素の値を $\mathbf{c}_{i_k q(D_l(i_k))}$ とする。この処理では、図 23 に示すように、入力画像ラベルが選択された割合に基づいて入力画像の画素値を合成するため、フレーム間で対象の同一箇所を投影している全ての画素に同じラベルが割り当てられている場合には、空間的な輝度調整の結果のみが出力に反映される。また、3次元形状上にマッピングするテクスチャの輝度値の差がフレーム間で小さくなるため、時間的な画素値の不連続が解消される。



時間的な輝度調整なし

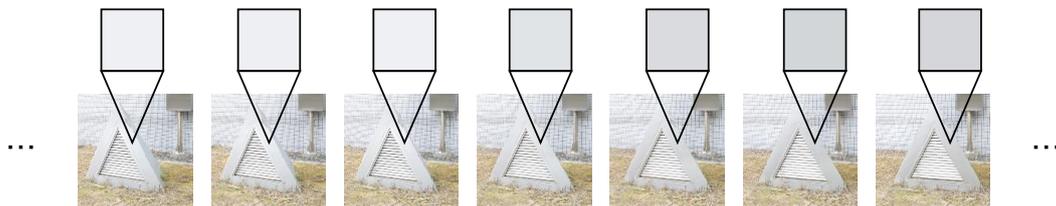


選択された入力画像



ラベル選択結果画像

選択されたラベルの割合に基づいた
入力画像の画素の合成



時間的な輝度調整の結果

図 23 時間的な輝度調整

4. 実験

提案手法の有効性を検証するため、実空間の小物体と屋外環境のシーンを対象とし、従来手法との比較実験を行う。以下では、実験内容を概説し、フレーム間のテクスチャの整合性に関するデータ項への重み付けのパラメータ、時間的な輝度調整のパラメータを変えて自由視点動画像を生成した結果を比較し、フレーム間のテクスチャの整合性に関するデータ項と時間的な輝度調整処理それぞれの効果について検証する。

4.1 実験の概要

本実験では、2つの小物体と2つの屋外環境のシーンを複数の地点(入力視点)から撮影した画像群を入力とする。実験で用いたPCの仕様、および入力データの詳細をそれぞれ表1、表2に示す。また、各入力画像の例を図24-27に示す。

前処理として、Visual SfM [31] を利用して各入力画像を取得した視点のカメラの内部・外部パラメータ推定を行い、CMPMVS [19] を利用して対象シーンの3次元形状復元を行った。続いて、仮想カメラの動きおよびそのフレーム数を各シーンで図28および表3に示すように設定し、各仮想視点位置で以下の3つの処理を行い、対象の見えを静止画像として1フレーム目から逐次的に生成した。

- Boykov らのグラフカットアルゴリズム [33] を利用したエネルギー関数(式(1))の最小化による入力画像ラベル選択
- 入力画像ラベル選択の結果に基づいた自由視点画像生成と、空間的な輝度調整の補正值の算出
- 対象フレームと過去数フレームの入力画像ラベル選択結果および空間的な輝度調整の補正值を利用した、空間的・時間的な輝度調整

各フレームの画像生成には、約30秒の時間を要した。最後に、全フレームの出力画像をつなぎ合わせて、動画像として出力した。

本実験では、自由視点動画像のテクスチャちらつきの抑制効果の有効性を示すため、エネルギー関数におけるフレーム間のテクスチャの整合性に関するデータ

項への重みパラメータ K_h と，時間的な輝度調整に使用するフレーム数 s をそれぞれ変化させ，出力画像がどのように変化するか比較する．エネルギー関数の他のパラメータに関しては，表 4，表 5 に示すように経験的に設定した．

以下の節では，小物体と屋外環境のシーンそれぞれを対象として自由視点動画を生成した結果を比較し，テクスチャの整合性に関するデータ項とそれに対する考察を述べる．また，被験者実験による評価に対する考察を述べる．

4.2 小物体を対象とした映像生成

本節では，2つの小物体を撮影した画像群を入力データとしたときの実験結果を示し，考察する．これら2つの物体は，それぞれ以下に示す特徴を持つ．

小物体 1

高周波成分を含むテクスチャが少なく，鏡面反射する素材ではない．

小物体 2

小物体 1 と比較すると，低周波成分を含むテクスチャだけでなく高周波成分を含むテクスチャも含まれており，鏡面反射する素材でできている．

フレーム間の整合性を考慮したテクスチャ選択，時間的な輝度調整それぞれの処理の効果を検証するため，まず，それぞれに関するパラメータ K_h および s のうち一方を， $K_h = 0$ または $s = 1$ に固定し，もう一方のパラメータを変化させて動画を生成した． $K_h = 0$ かつ $s = 1$ の場合には，片桐らの手法を用いた場合の結果と同様になる．以下，それぞれのシーンの動画を比較した結果について述べる．

4.2.1 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ に関する検証

フレーム間の整合性に関するデータ項 $h_{i_k}(x_{i_k})$ の効果を検証するために，輝度調整に使用するフレーム数 s を 1 に固定し， $h_{i_k}(x_{i_k})$ への重みパラメータ K_h を 0, 10, 20, 30, 40 と変化させて動画を生成した結果の例を図 29 と図 34 に，図 29 と図 34 の画像の一部分を拡大したものを図 30 と図 35 に示す．図 30 と図 35 の赤枠は，フレーム間でテクスチャの変化が大きい箇所を示す．また，図 31, 図 36 に

表 1 実験で用いた PC の仕様

項目	詳細
CPU	Intel Corei7-4770 3.40GHz
メモリ	RAM 16.0GB
GPU	NVIDIA GeForce GTX TITAN Black
OS	Windows 8.1
グラフィック API	OpenGL

表 2 各シーンの入力画像の詳細

対象	画像サイズ	画像枚数
小物体 1	1696 × 1280	61
小物体 2	1920 × 1280	52
屋外環境シーン 1	1920 × 1280	51
屋外環境シーン 2	1920 × 1280	35

表 3 シーン別の出力フレーム数

	小物体 1	小物体 2	屋外環境 1	屋外環境 2
フレーム数	360	67	141	96

表 4 共通の固定パラメータ

パラメータ	値	式番号
K_β	400	(1)
K_g	200	(1)
K_α	20000	(5)
θ_n	5(deg)	(8)

表 5 シーン別の固定パラメータ

パラメータ	小物体 1	小物体 2	屋外環境 1	屋外環境 2	式番号
θ_d	0.1	0.1	0.5	0.5	(5)



図 24 小物体 1 の入力画像例



図 25 小物体 2 の入力画像例



図 26 屋外環境シーン 1 の入力画像例



図 27 屋外環境シーン 2 の入力画像例

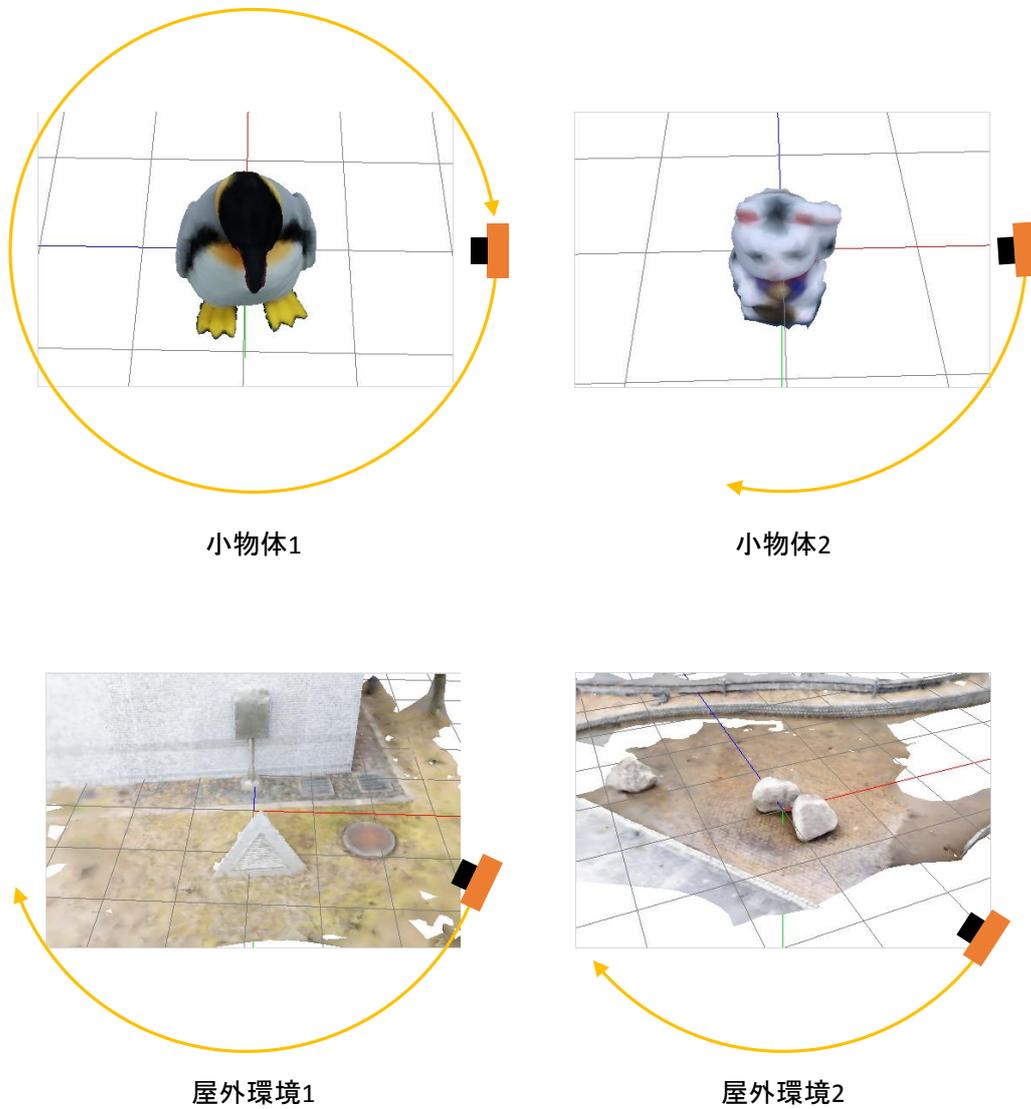


図 28 各シーンにおける仮想視点の動き

は、 K_h を 0, 10, 30 に設定することで、従来手法と比較して時間的な不連続を解消できた事例を示す。図 32, 図 37 には、図 31, 図 36 の例と反対に、従来手法と比較して時間的な不連続が発生している例を示す。図 31, 図 32, 図 36, 図 37

における各パラメータの上段の画像群は出力画像であり，下段の画像群は入力画像ラベル選択結果を可視化したものである．

小物体1を対象とし， s の値を1に固定して K_h の値を変化させた動画像を比較した結果について述べる． K_h を0に設定した場合は，フレーム間の画素値の整合性が全く考慮されず，他の条件のみで入力画像ラベル選択処理を行っており，カメラパラメータ推定や3次元復元の誤差によって，図30の赤枠で示す部分のテクスチャの位置がフレーム間でずれているように見えることが分かる．また， K_h を10，20に設定した場合は，0の場合と比較するとテクスチャちらつきが低減されていることが確認できたが，図31中に示すように，テクスチャが切り替わるタイミングが遅れただけのように見える領域もあった．しかし， K_h を30に設定した場合は，図31からも分かるようにテクスチャの不自然な切り替わりが少ない結果となった． K_h を40に設定した場合には，図32に示すように頻度は少ないがテクスチャが急激に切り替わるタイミングがあることを確認した．これは， K_h の重みが大きく，一つのテクスチャが選択されたときに他のテクスチャに切り替わりにくくなったためだと考えられる．これにより，図33中の K_h を30に設定した場合は徐々にテクスチャの切り替わりが発生しているのに対し，40に設定した場合はテクスチャの切り替わりが急激なものとなっている．これがテクスチャの不整合が発生したような印象を与えると考えられる．

小物体2の動画像を比較した結果について述べる．小物体2については，図35や図36に示すように，対象の目や鼻の部分などのテクスチャの不連続が小物体1の実験結果と同様に解消されていることが分かる．しかし，図37で示すように，対象の脚の部分に注目すると， K_h が0の場合にはなかった脚の模様の不連続が発生している．これは，カメラパラメータや3次元形状復元の誤差が大きく，それによってフレーム間の物体の同一箇所を投影する画素間で適切な対応関係が得られなかったためと考えられる．

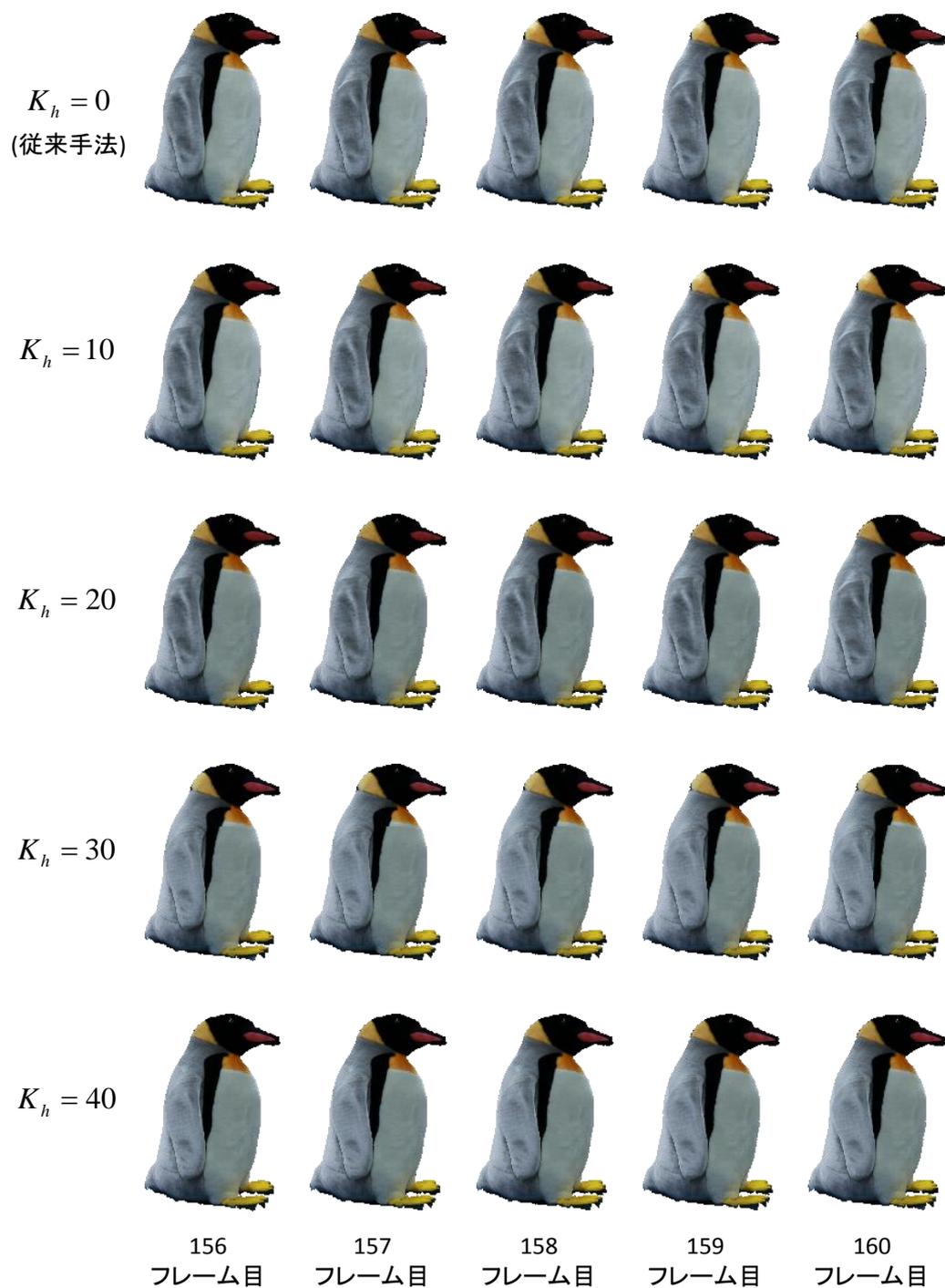


図 29 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (小物体 1)

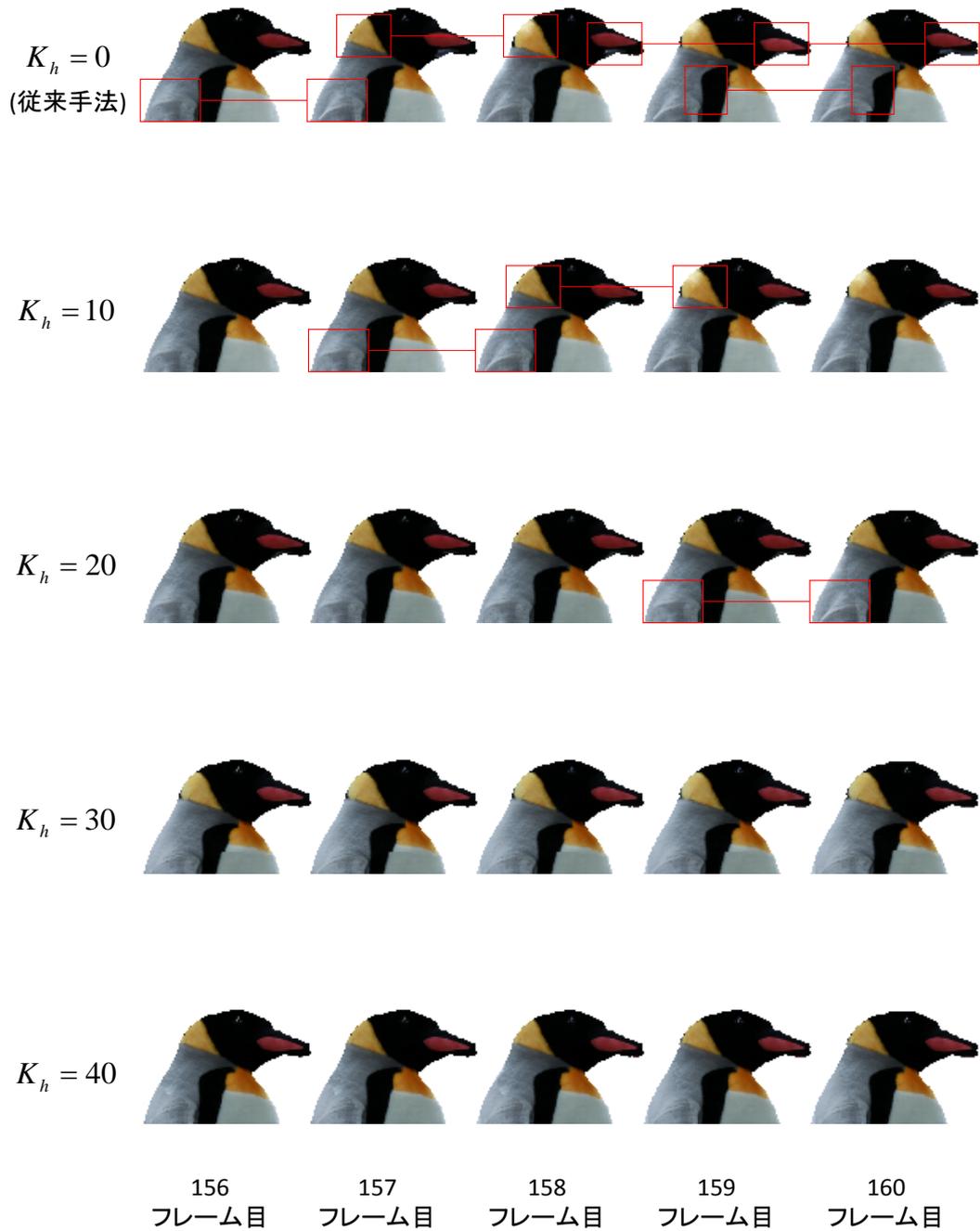
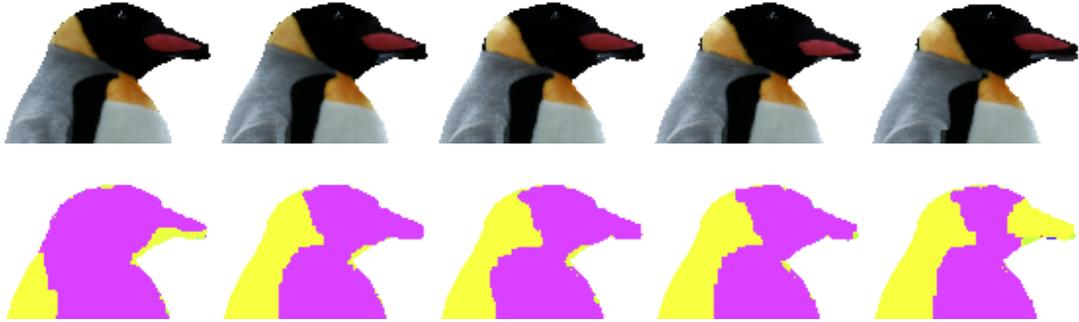
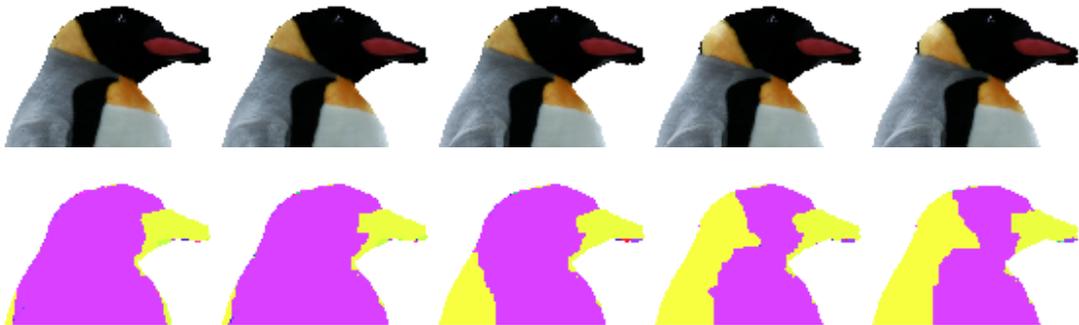


図 30 図 29 の一部分を拡大した画像



$$K_h = 0$$



$$K_h = 10$$



$$K_h = 30$$

図 31 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解消した例 (小物体 1)

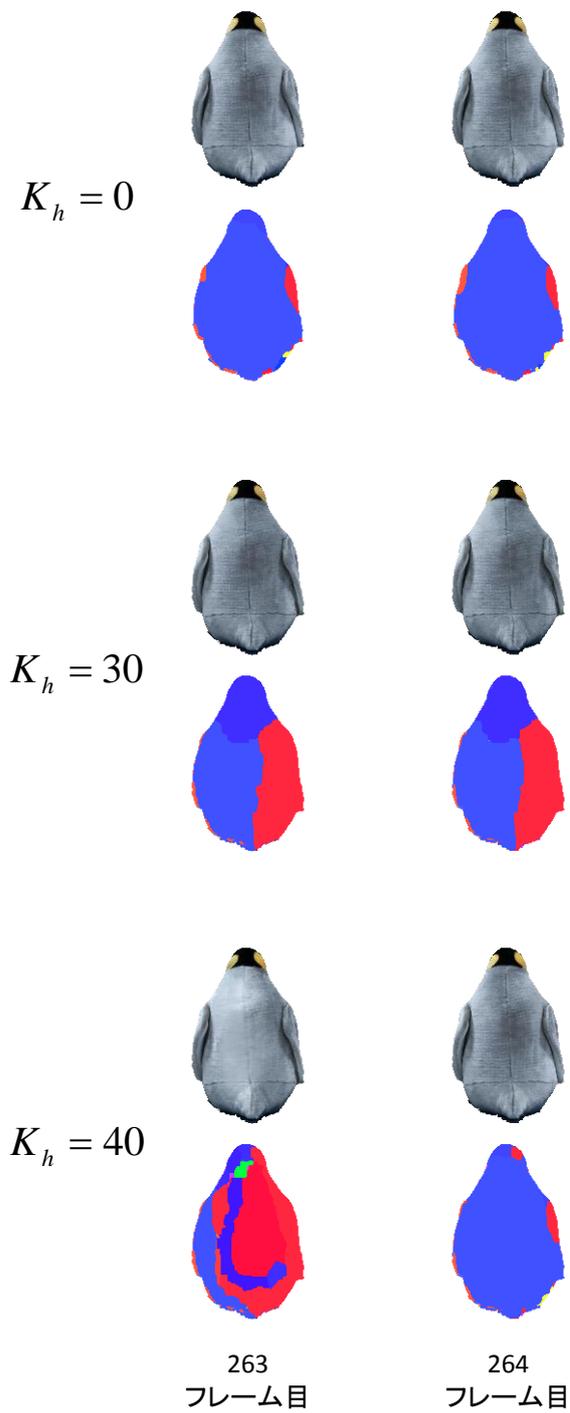


図 32 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続が生じた例 (小物体 1)

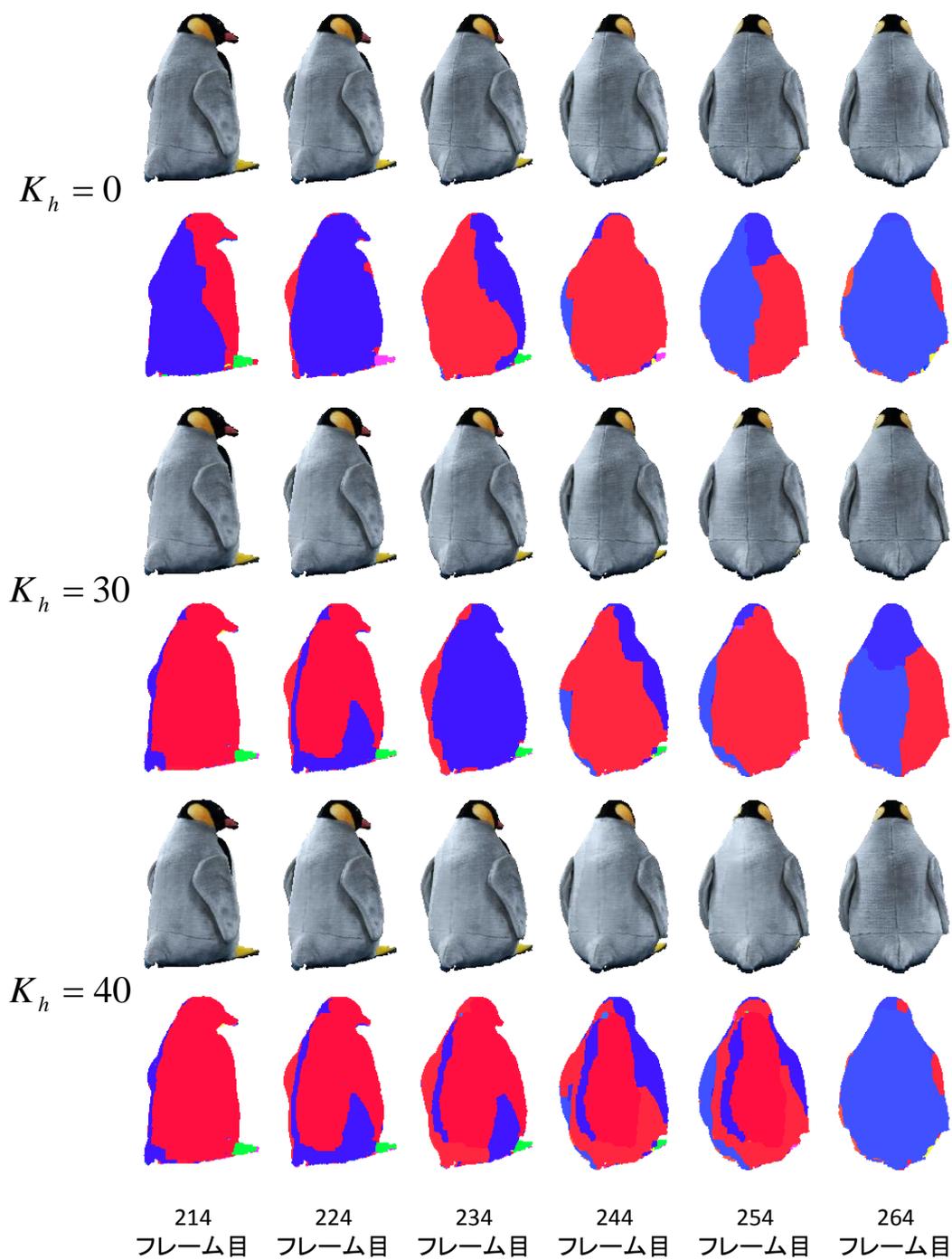


図 33 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ による時間的な不連続発生の原因 (小物体 1)

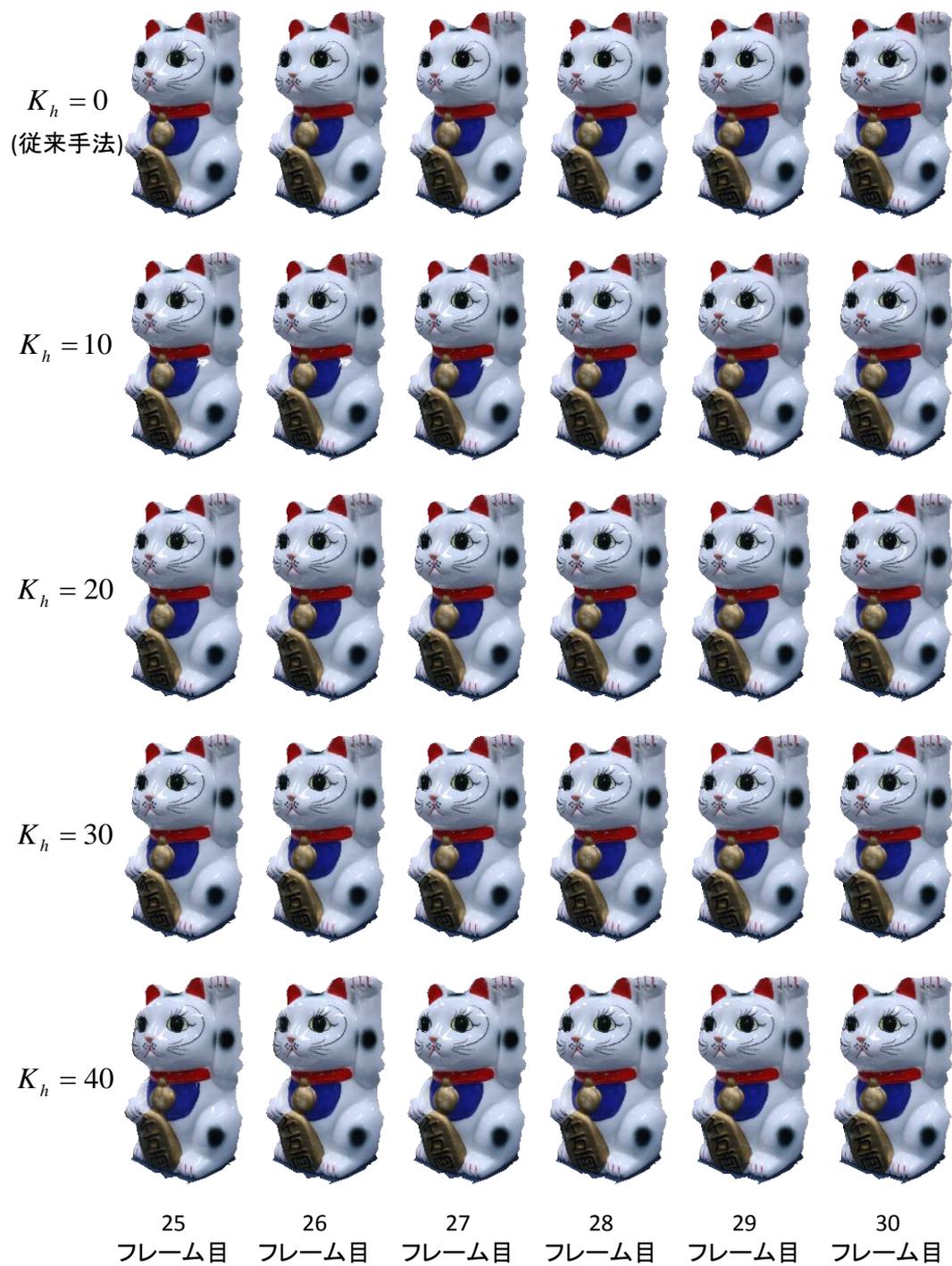


図 34 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (小物体 2)

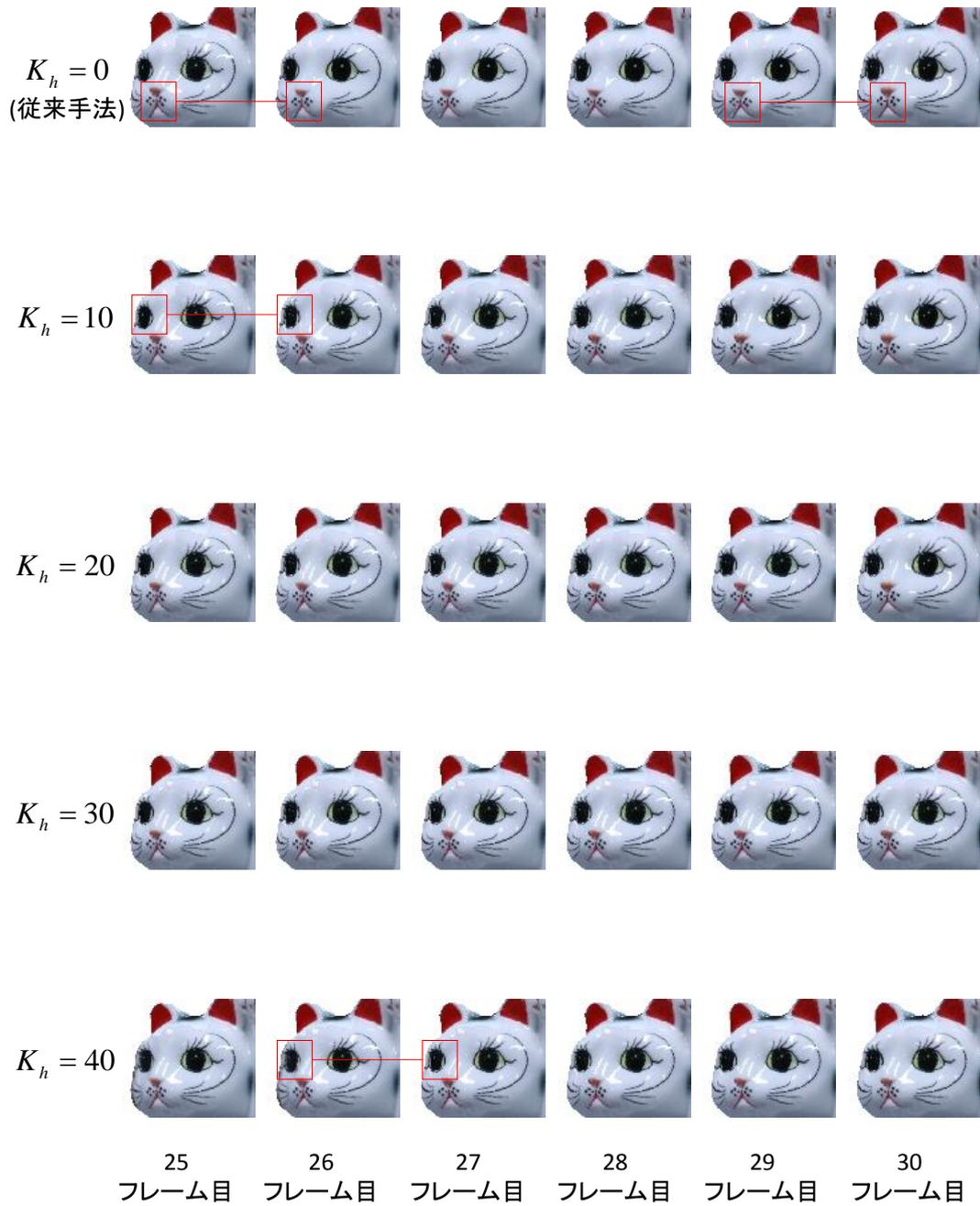


図 35 図 34 の一部分を拡大した画像

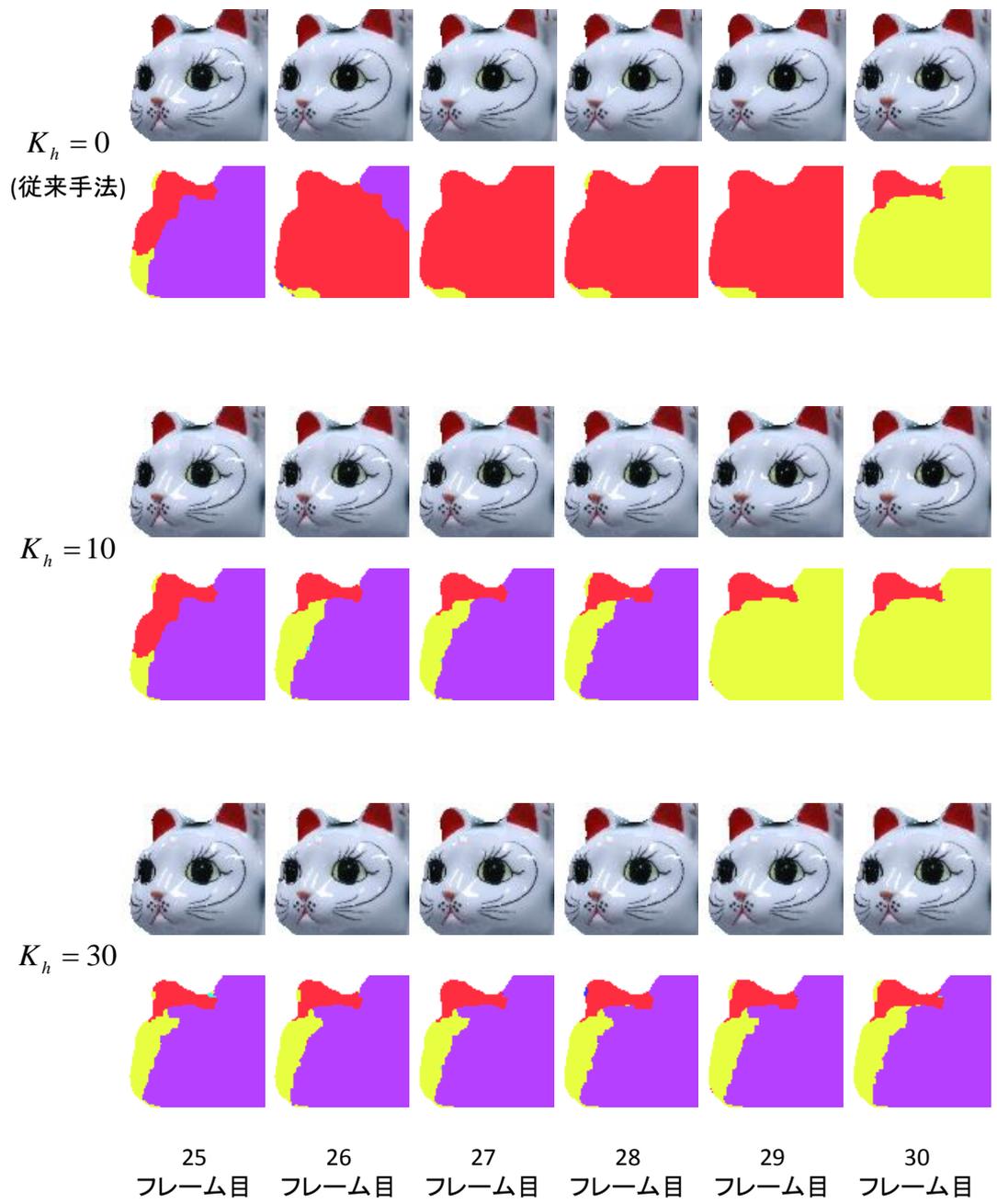


図 36 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解消した例 (小物体 2)

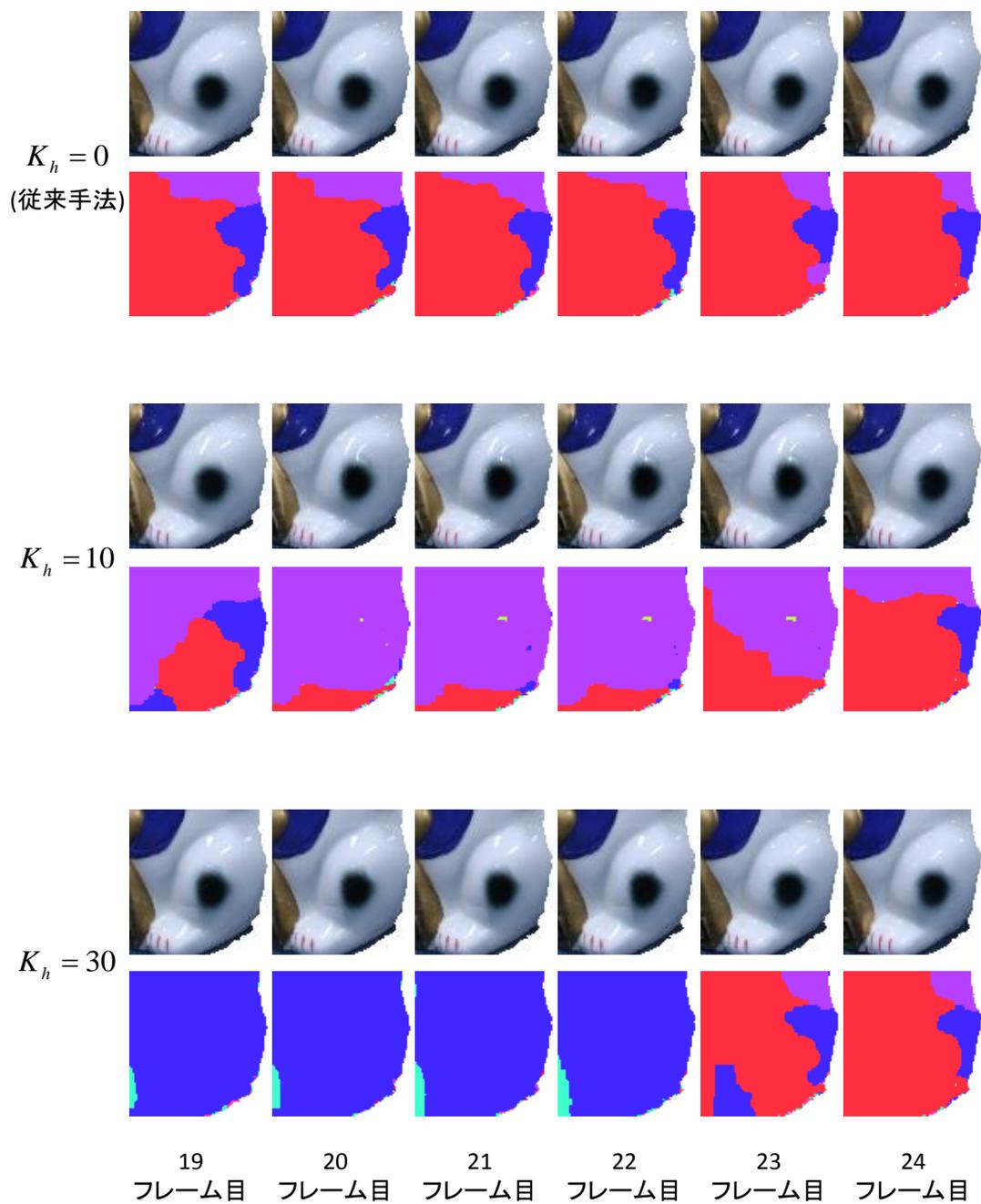


図 37 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続が生じた例 (小物体 2)

4.2.2 時間的な輝度調整のフレーム数 s に関する検証

時間的な輝度調整に使用するフレーム数 s の効果を検証するために、 K_h を 0 に固定し、 s を 1, 3, 5, 7, 9 と変化させて動画像を生成した結果の例を図 38 と図 40 に、図 38 と図 40 の画像の一部を拡大したものを図 39 と図 41 に示す。また、赤枠は s が 1 の場合に時間的な不連続が生じる箇所を示す。図 39 と図 41 から分かるように、時間的な輝度調整によって画素値の不連続を抑制できていることが分かる。しかし、 s の値が小さい場合には輝度調整の効果が小さく、不連続を抑制しきれない場合がある。例えば、図 41 に示す小物体 1 の嘴は、 s を 3 に設定した場合に不連続を抑制しきれずに二重像が生じていることが分かる。

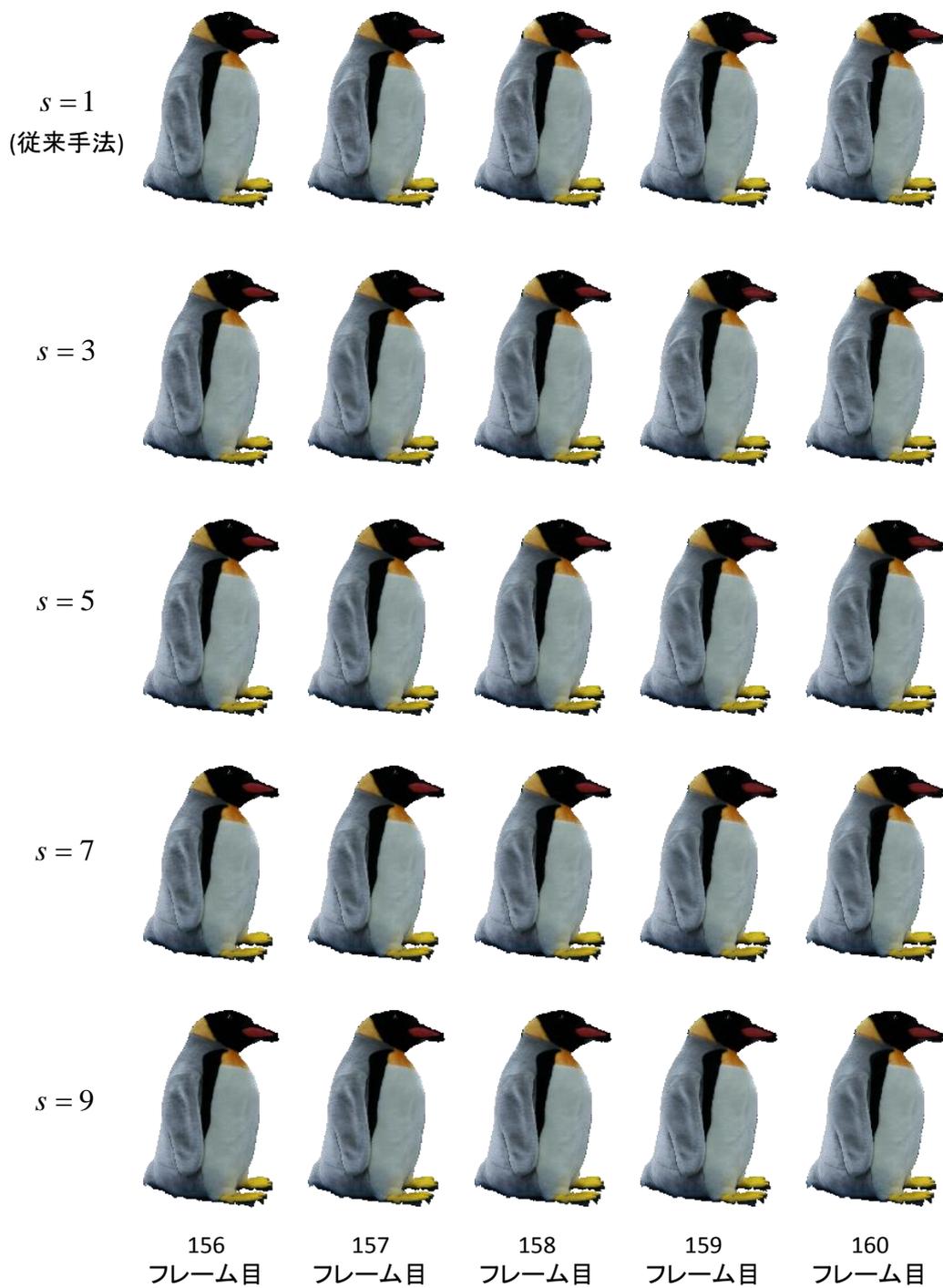


図 38 時間的な輝度調整の効果の比較 (小物体 1)

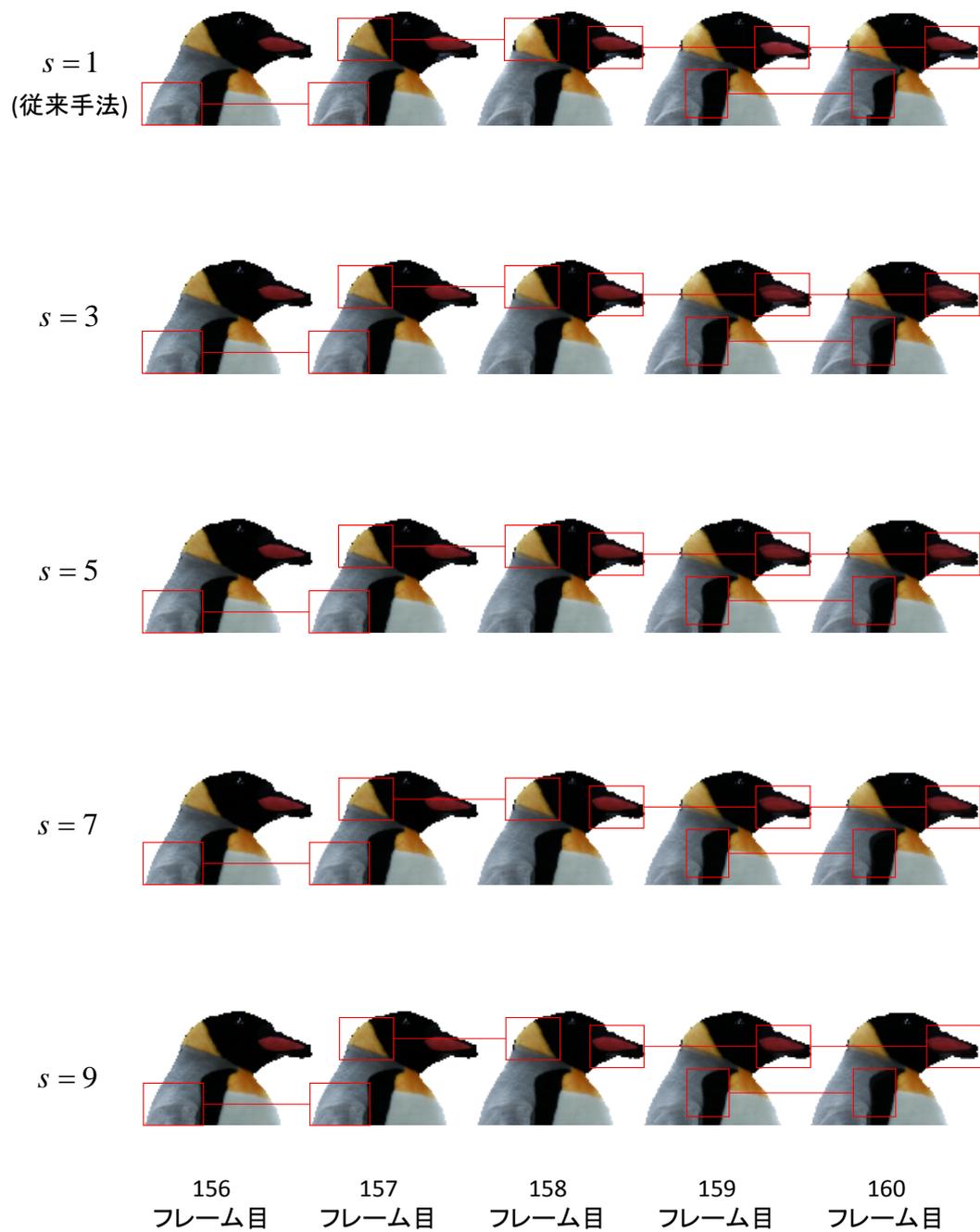


図 39 図 38 の一部分を拡大した画像

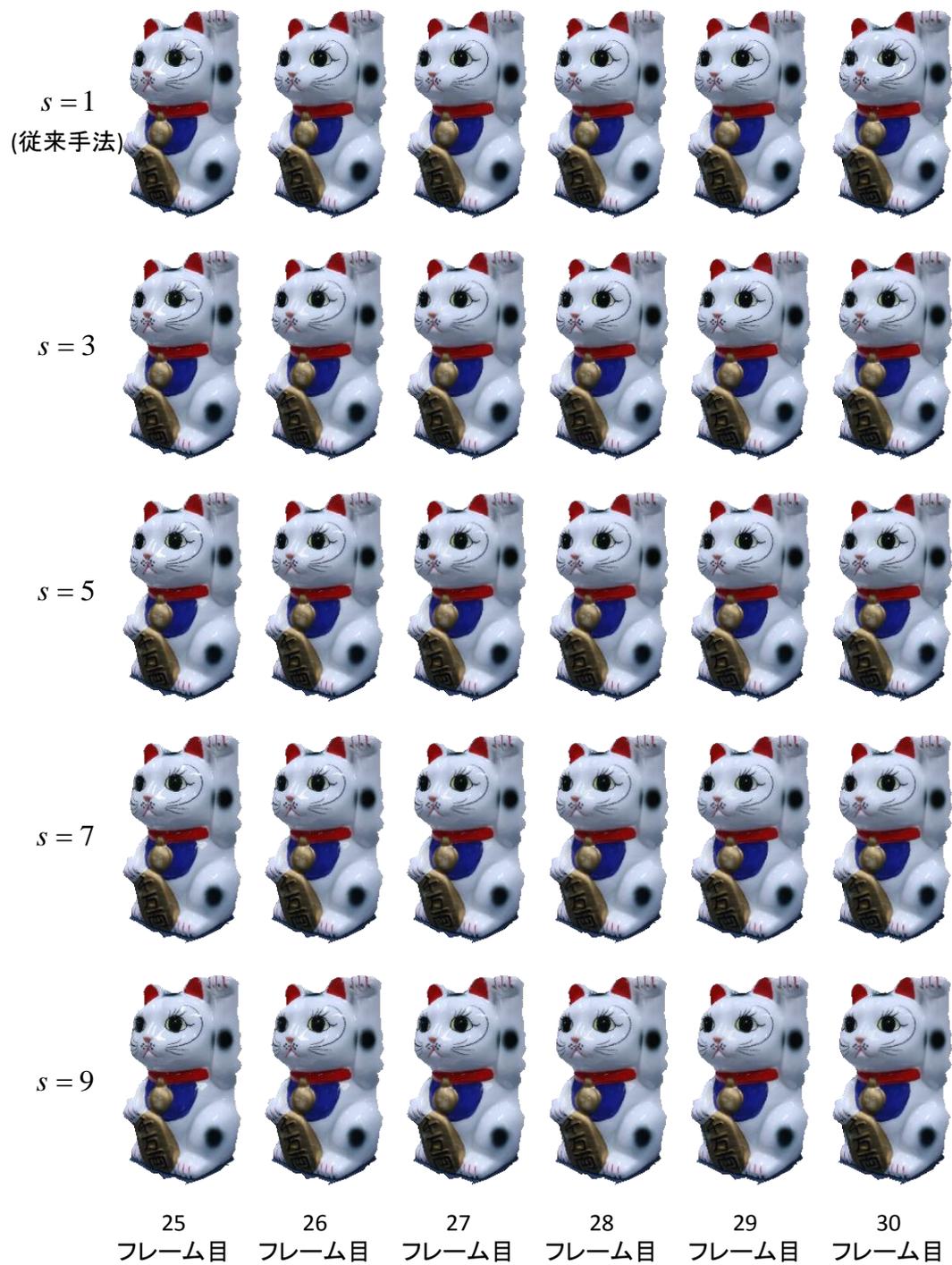


図 40 時間的な輝度調整の効果の比較 (小物体 2)



図 41 図 40 の一部分を拡大した画像

4.3 屋外環境のシーンを対象とした映像生成

本節では、2つの屋外環境のシーンを撮影した画像群を入力データとして、4.2節と同様に実験を行った結果を示し、考察する。これら2つのシーンは、それぞれ以下に示す特徴を持つ。

屋外環境 1

空間的な形状の変化が大きい物体が含まれている。

屋外環境 2

水面が写っており、水面からの鏡面反射光が含まれている。

フレーム間の整合性を考慮したテクスチャ選択、時間的な輝度調整それぞれの処理の効果を検証するため、まず、それぞれに関するパラメータ K_h および s のうち一方を、 $K_h = 0$ または $s = 1$ に固定し、もう一方のパラメータを変化させて動画像を生成した。 $K_h = 0$ かつ $s = 1$ のときは、片桐らの手法を用いた場合の結果と同様になる。以下、それぞれのシーンの動画を比較した結果について述べる。

4.3.1 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ に関する検証

フレーム間の整合性に関するデータ項 $h_{i_k}(x_{i_k})$ の効果を検証するために、輝度調整に使用するフレーム数 s を 1 に固定し、 $h_{i_k}(x_{i_k})$ への重みパラメータ K_h を 0, 10, 20, 30, 40 と変化させて動画像を生成した結果の例を図 42 と図 47 に、図 42 と図 47 の画像の一部分を拡大したものを図 43 と図 48 に示す。各図の赤枠は、フレーム間でテクスチャの変化が大きい箇所を示す。また、図 44 と図 49 には、 K_h を 0, 10, 30 に設定し、従来手法と比較して時間的な不連続を解消できている例を示す。図 45 には、図 44 の例と反対に、従来手法と比較して時間的な不連続が発生している例を示す。図 44, 図 45, 図 49 における各パラメータの上段の画像群は出力画像であり、下段の画像群は入力画像ラベル選択結果を可視化したものである。

屋外環境シーン 1 では、 K_h を変化させたときに図 43, 図 44 に示すように、三角形の物体領域の時間的な不連続を従来手法よりも抑制できていることが分かる。

また、4.2節における実験結果と同様に、 K_h の値が小さい場合に効果が小さいことが分かった。 K_h が40の場合には全体的なテクスチャちらつきが一番少なかったが、テクスチャ切り替わりの際に不連続が生じる箇所があることが分かった(図45)。これは、小物体1の実験の場合と同様に、 K_h の重みが大きく、一つのテクスチャが選択されたときに他のテクスチャに切り替わりにくくなったためだと考えられる。これにより、図46中の K_h を30に設定した場合と比較すると、40に設定した場合はテクスチャの不整合が発生したような印象を与えると考えられる。

屋外環境シーン2では、シーン中の岩に注目すると、図48、図49に示すように、テクスチャの時間的な不連続を抑制できていることが分かる。また、それ以外の領域についても、従来手法と比較するとテクスチャの時間的な不連続を抑制できていることを確認した。しかし、このシーンは水面の鏡面反射が多いシーンであることや、このシーンを撮影する際に、同じく鏡面反射がある小物体2を撮影したときよりも光線情報を密に取得できなかったことから、入力画像間で鏡面反射光による見え方が大きく異なる領域が多く、図54に示すように、フレーム間で鏡面反射光による見え方が視覚的に不自然に変化する場合があった。

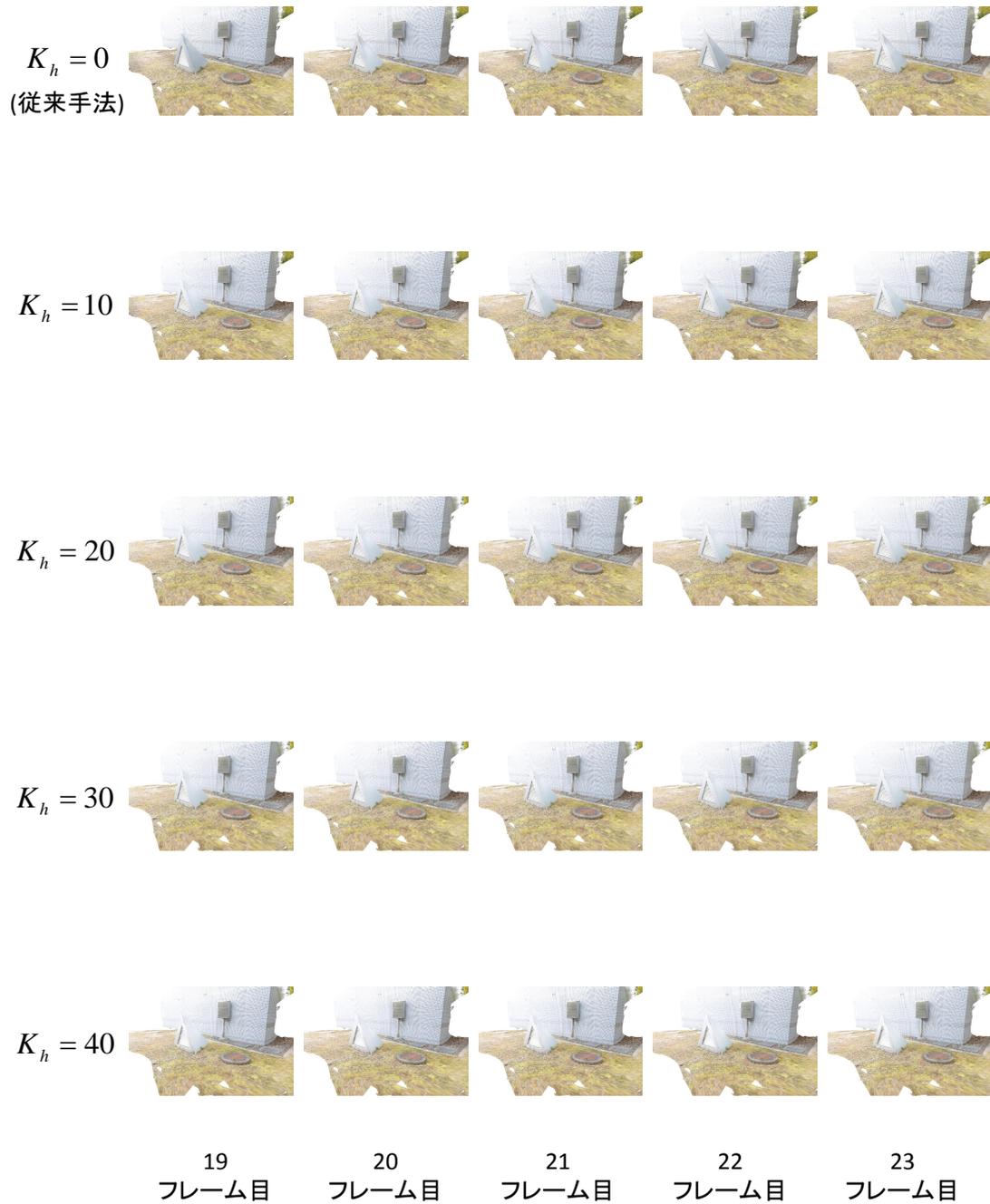


図 42 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (屋外環境シーン 1)

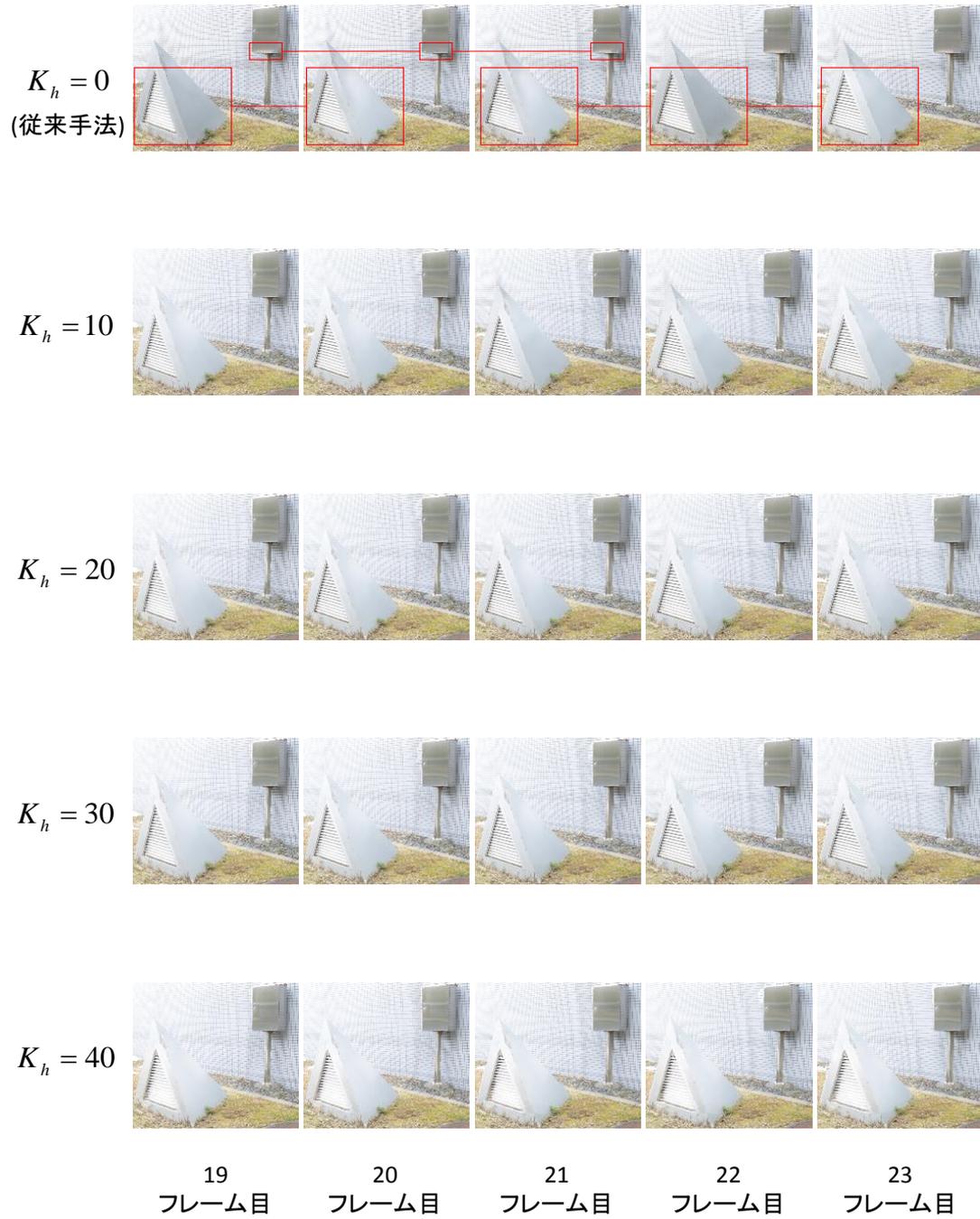


図 43 図 42 の一部分を拡大した画像

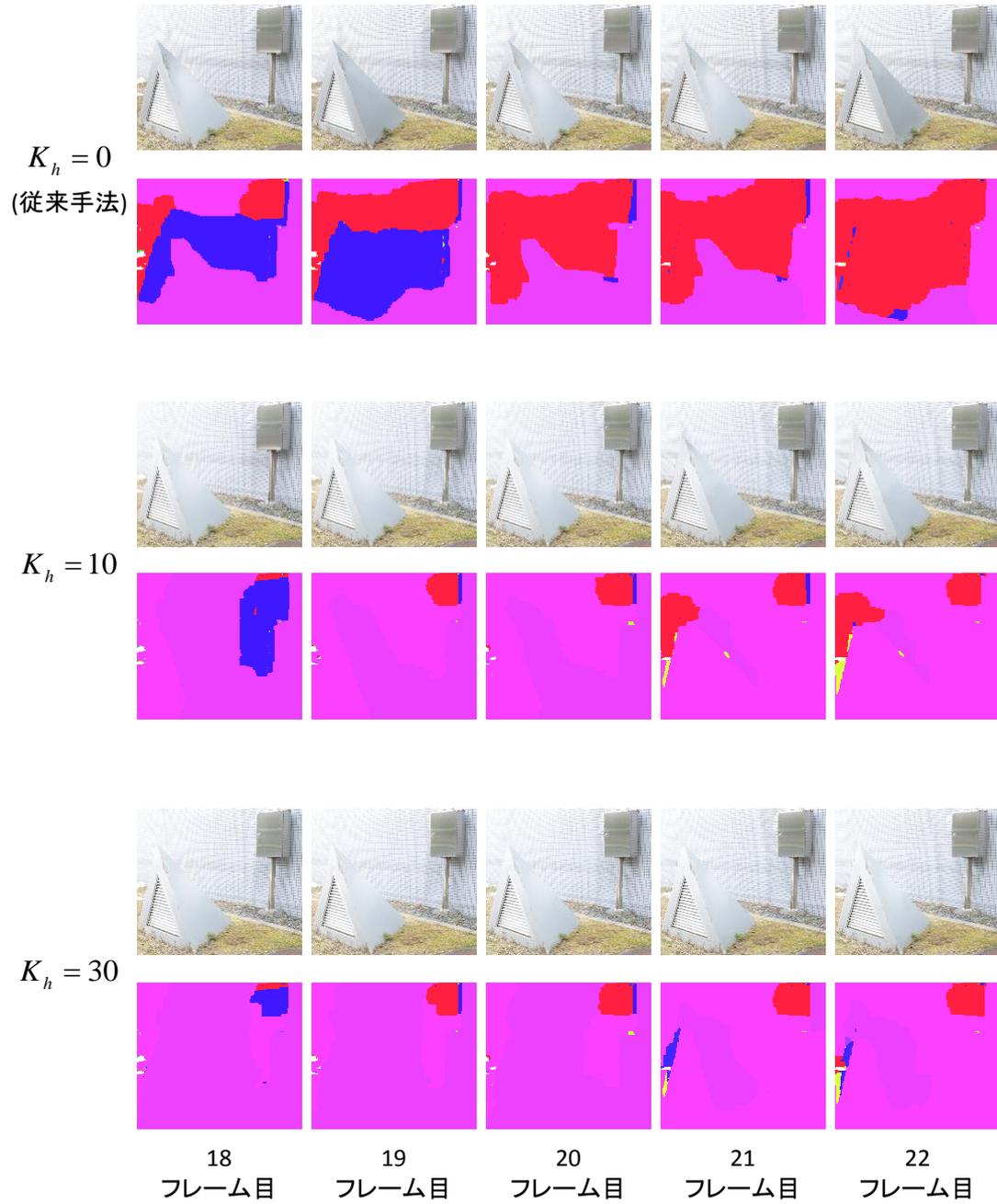


図 44 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解消した例(屋外環境シーン 1)

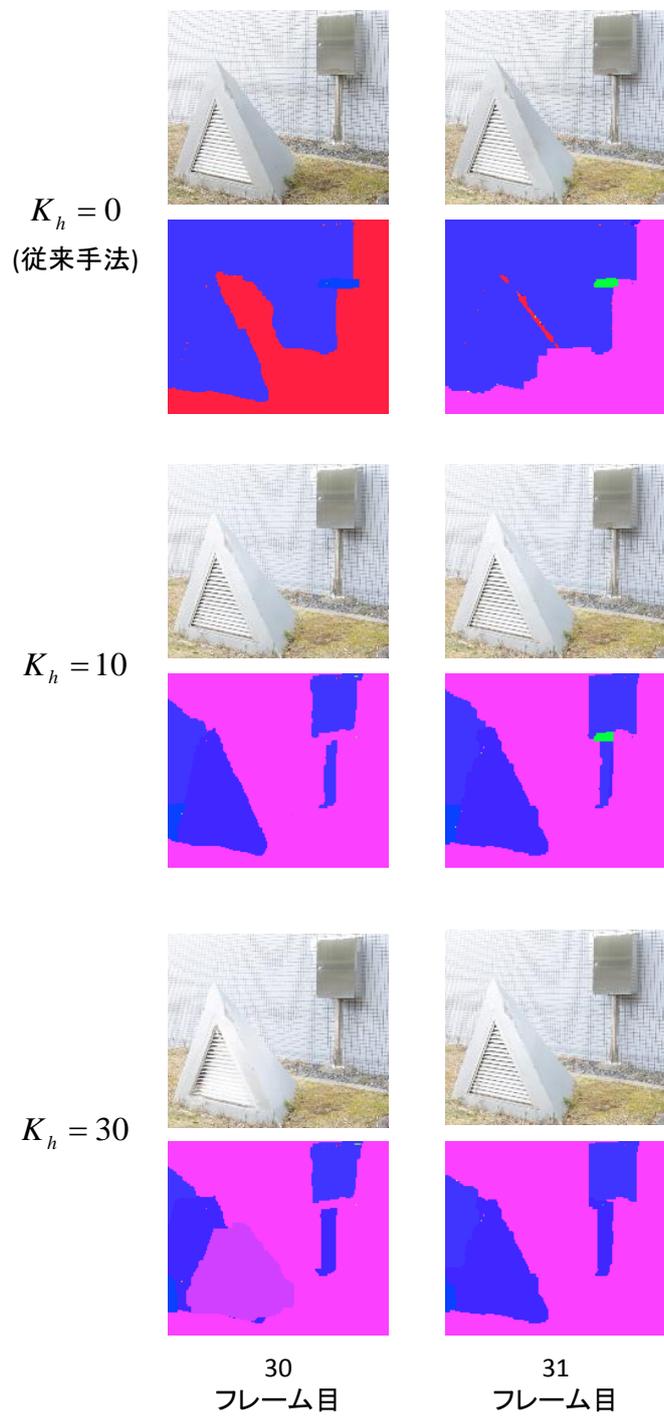


図 45 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続が生じた例 (屋外環境シーン 1)

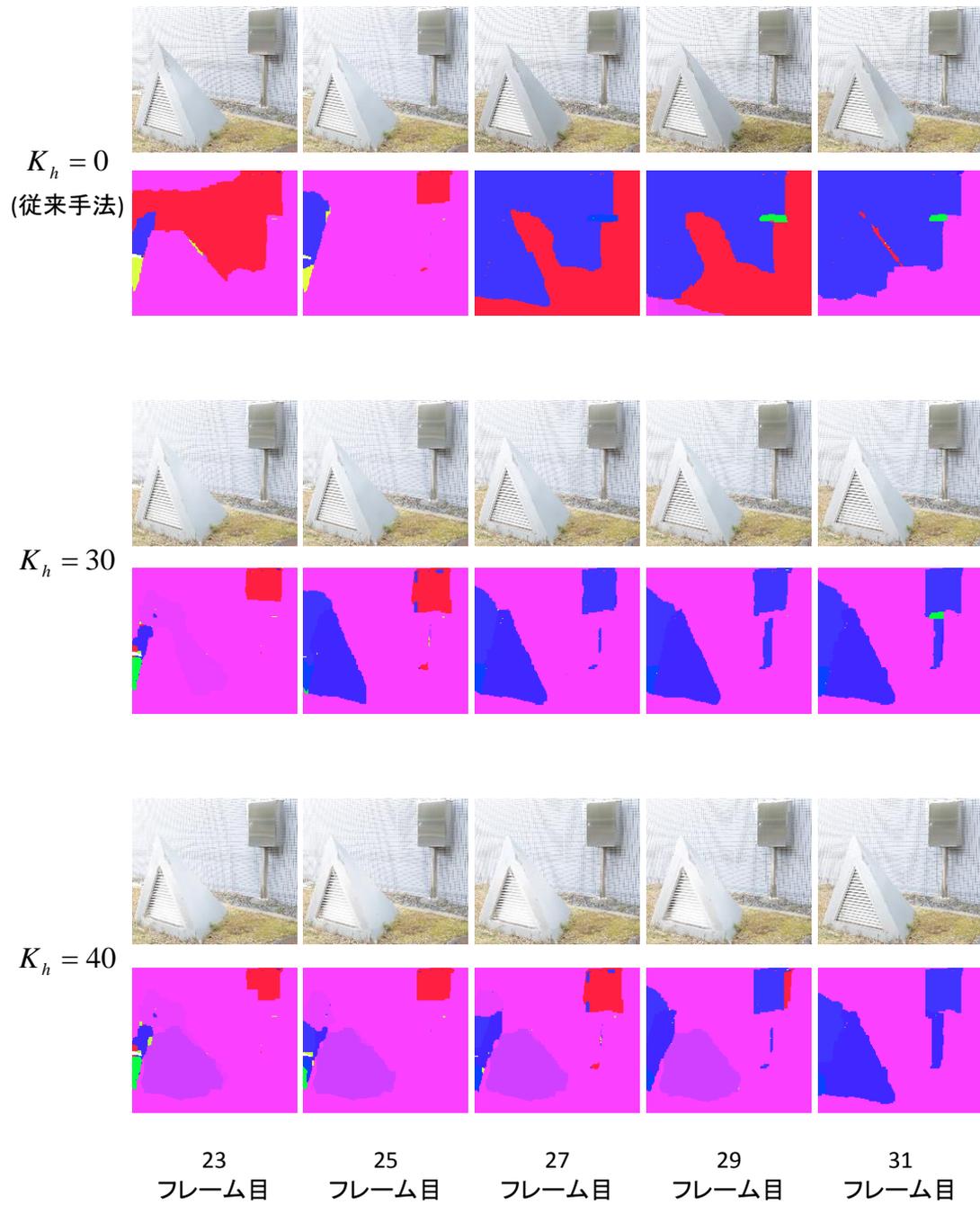


図 46 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ による時間的な不連続発生の原因 (屋外環境シーン 1)

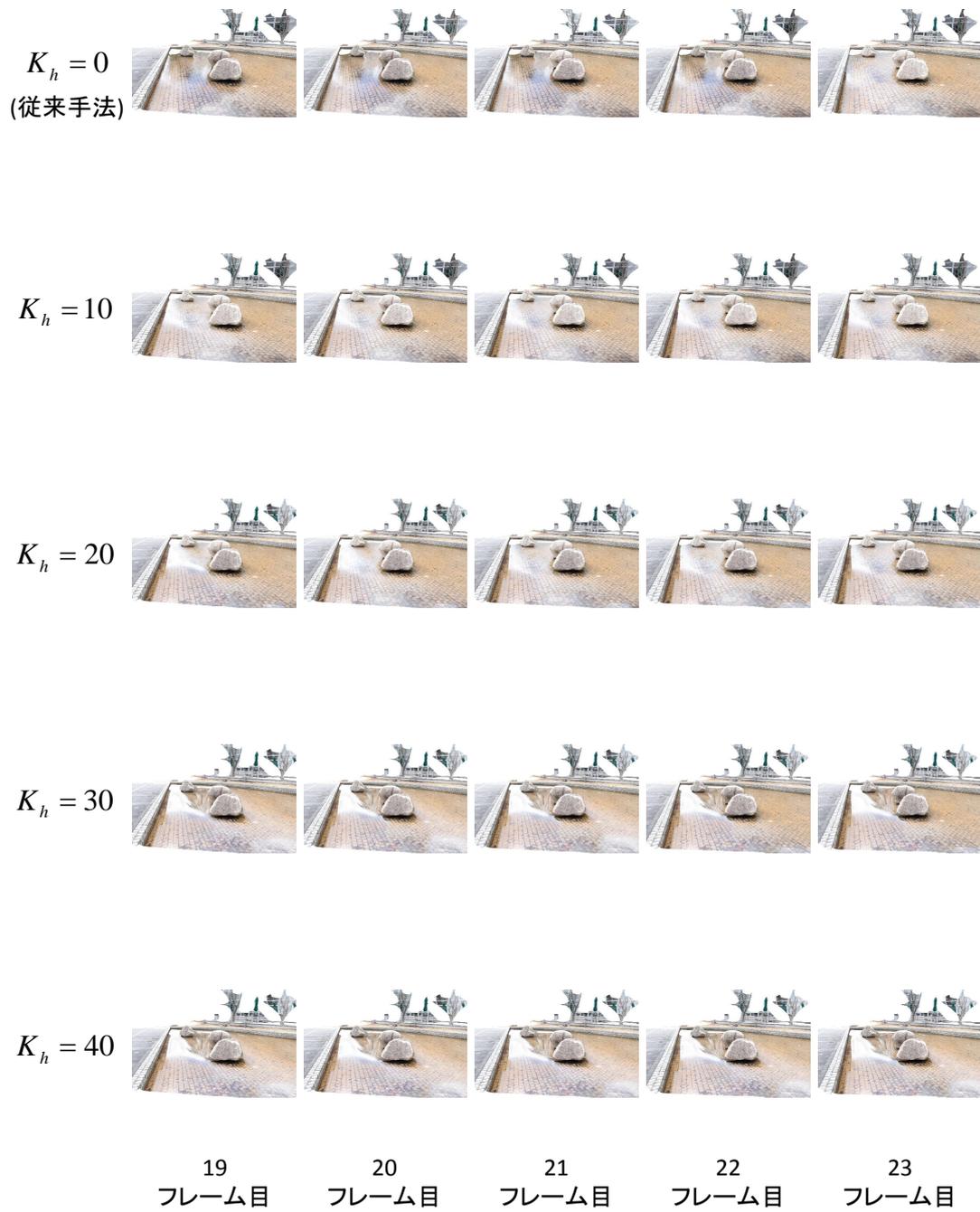


図 47 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ の効果の比較 (屋外環境シーン 2)

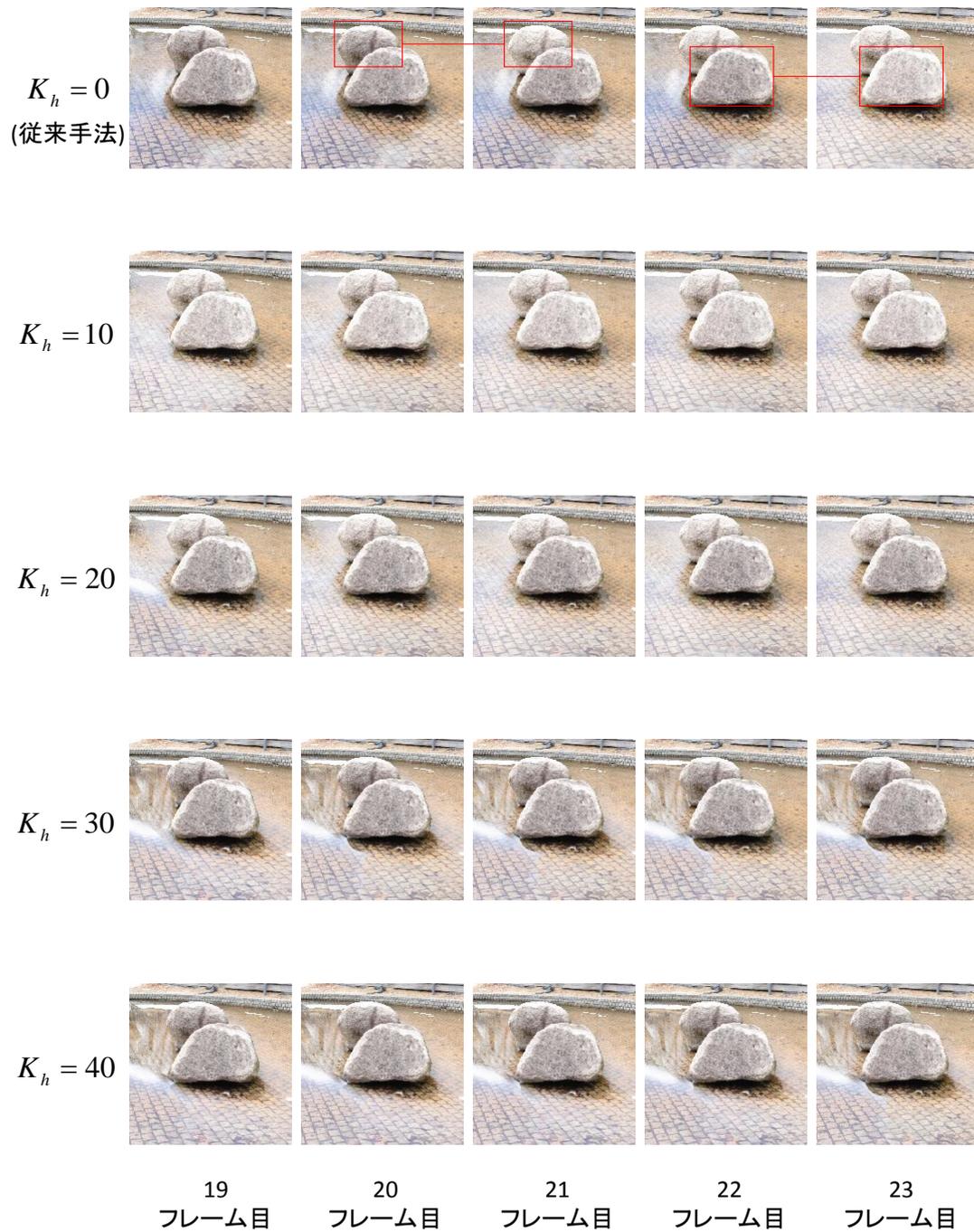


図 48 図 47 の一部分を拡大した画像

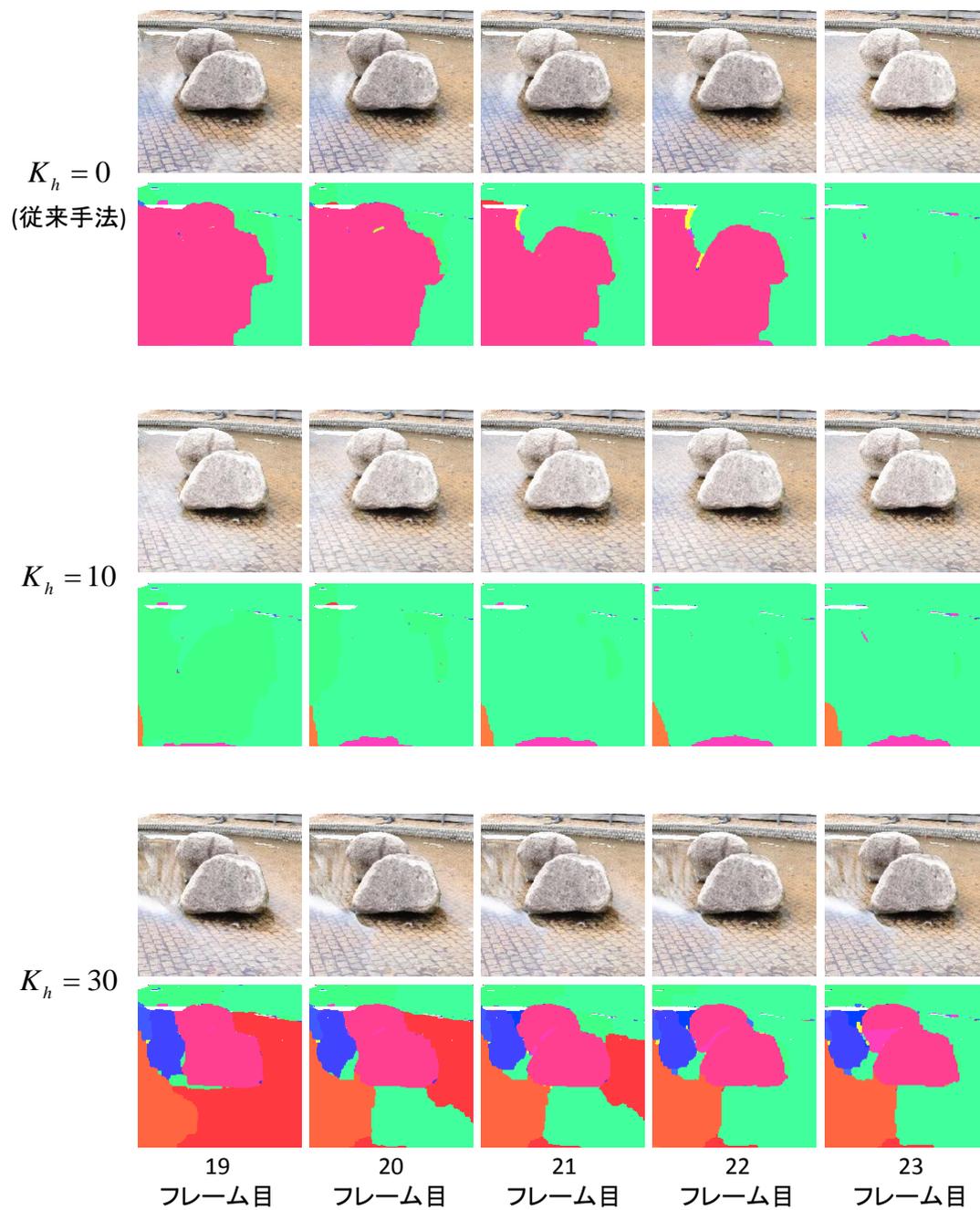


図 49 フレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ により時間的な不連続を解消した例 (屋外環境シーン 2)

4.3.2 時間的な輝度調整のフレーム数 s に関する検証

時間的な輝度調整に使用するフレーム数 s の効果を検証するために、 K_h を 0 に固定し、 s を 1, 3, 5, 7, 9 と変化させて動画像を生成した結果の例を図 50 と図 52 に、図 50 と図 52 の画像の一部分を拡大したものを図 51 と図 53 に示す。赤枠は、 s が 1 の場合に時間的な不連続が生じる箇所を示す。図 51, 図 53 から分かるように、時間的な輝度調整によって画素値の不連続を抑制できていることが分かる。また、屋外環境シーン 2 の水面の領域に関しては、図 54 中の $K_h = 0, s = 1$ の出力結果と $K_h = 0, s = 7$ の出力結果を比較すると、 $K_h = 0, s = 7$ のほうが水面の鏡面反射による見え方の変化を小さくできていることが分かる。前述のフレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ に関するパラメータ K_h を変化させた場合の検証結果と合わせると、屋外環境シーン 2 ではフレーム間の整合性を考慮した入力画像ラベル選択と輝度調整を組み合わせた場合に、より自然な見えを再現できると考えられる。

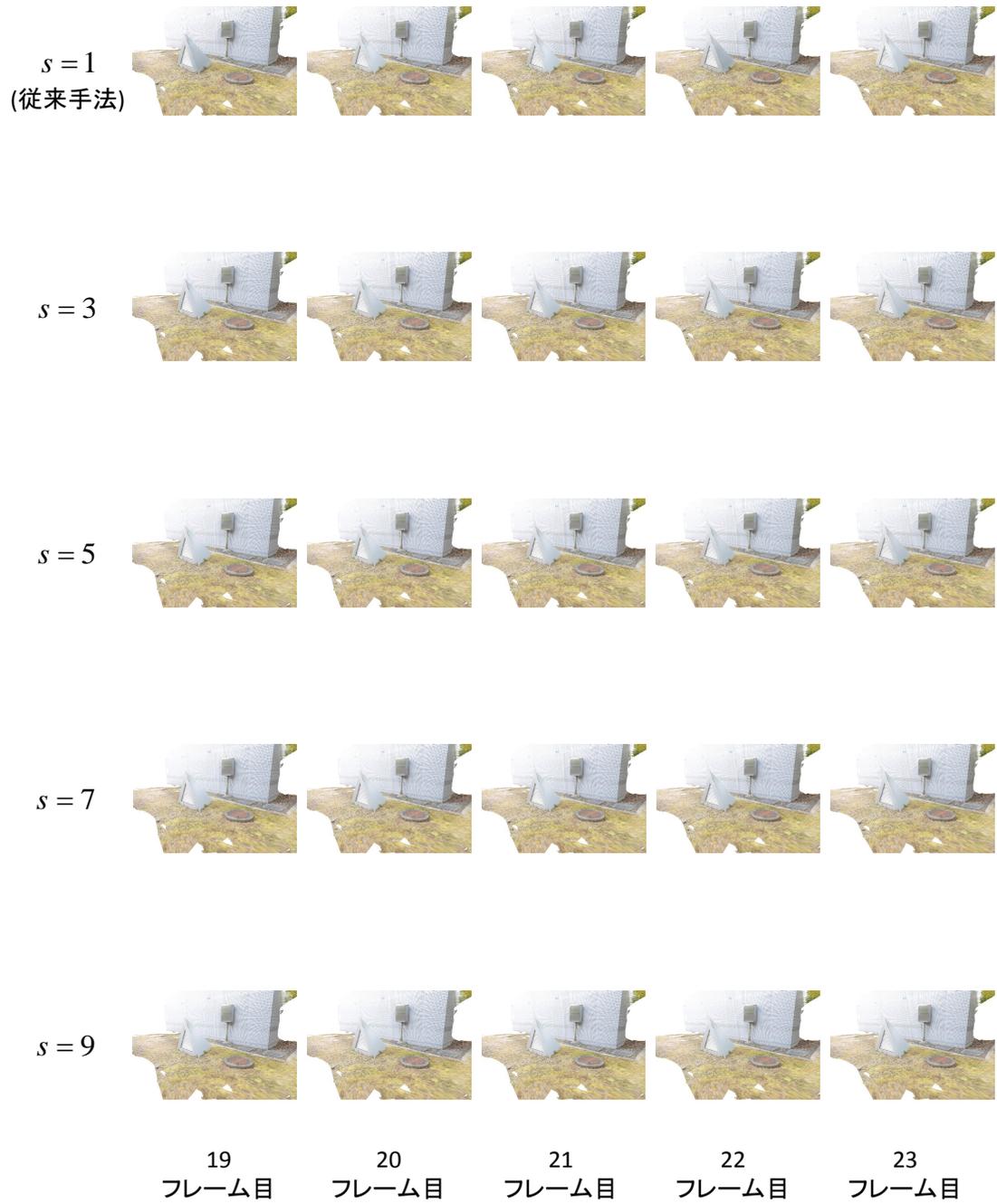


図 50 時間的な輝度調整の効果の比較 (屋外環境シーン 1)

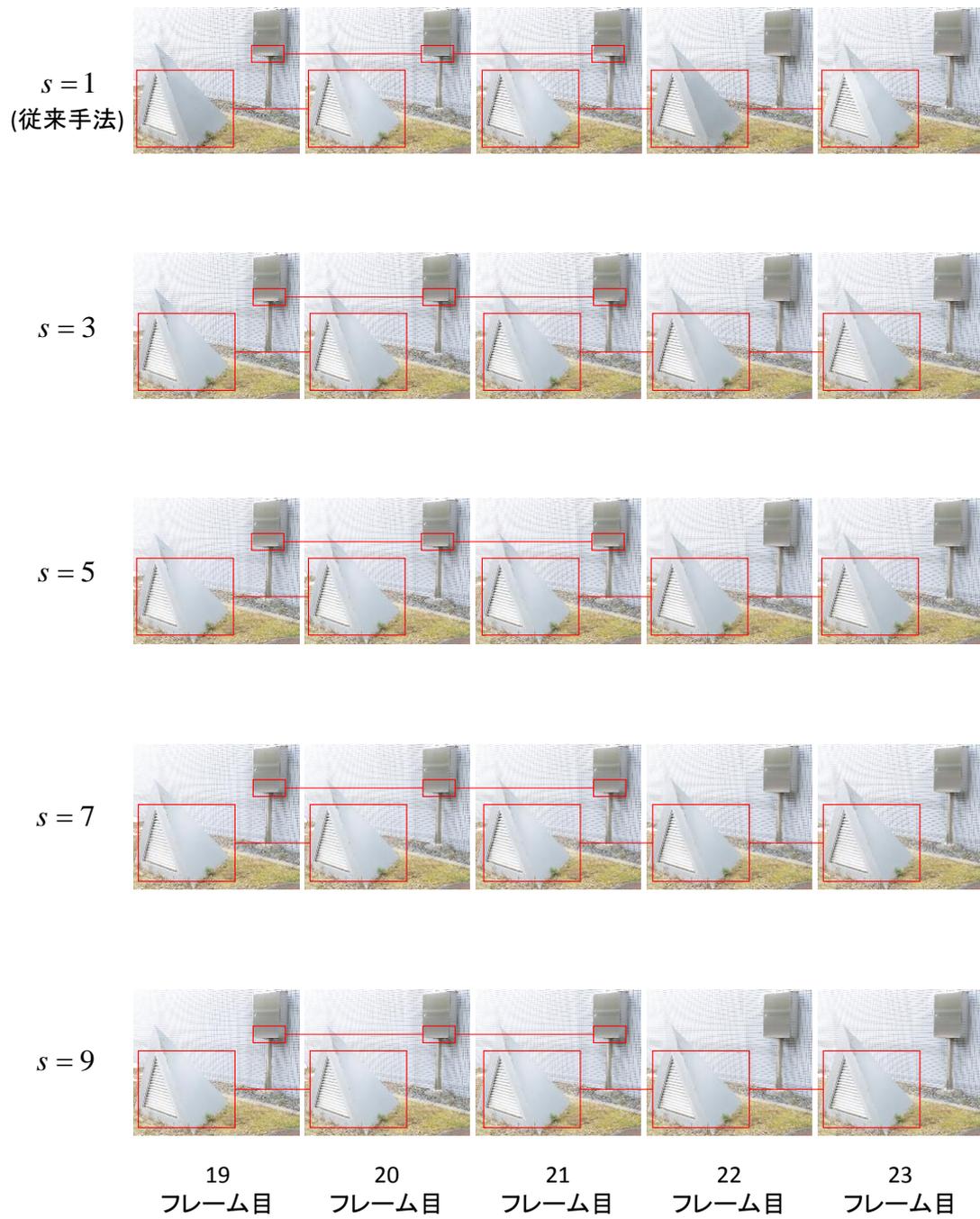


図 51 図 50 の一部分を拡大した画像



図 52 時間的な輝度調整の効果の比較 (屋外環境シーン 2)

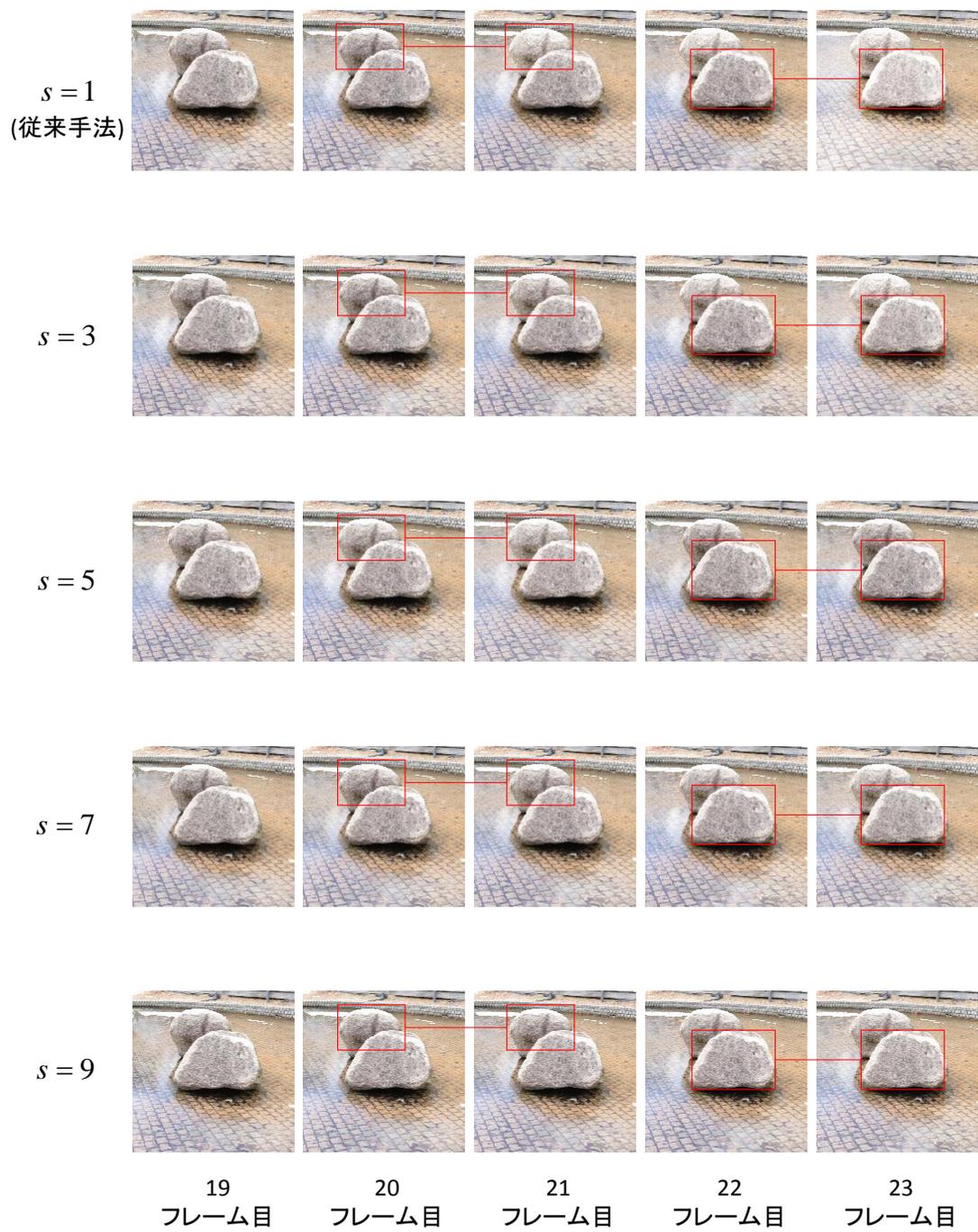


図 53 図 40 の一部分を拡大した画像

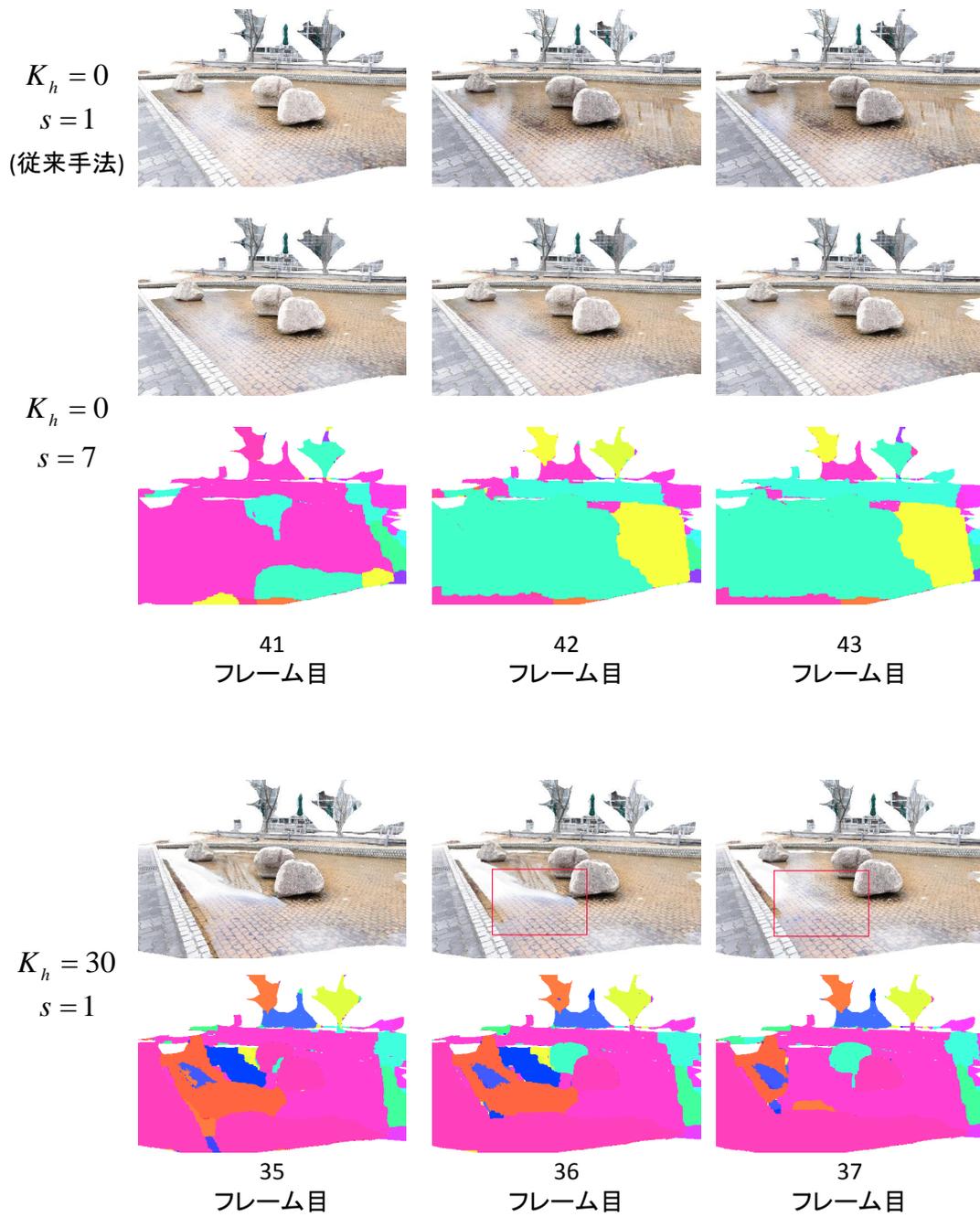


図 54 屋外環境シーン 2 に関するフレーム間の整合性に関する項 $h_{i_k}(x_{i_k})$ と時間的な輝度調整の効果の比較

4.4 被験者実験によるちらつき低減効果の確認と考察

本節では、提案手法の有効性を検証するための被験者実験について述べる。本実験では、提案手法における各パラメータの設定に関する検証を行った。パラメータ設定に関する検証では、 K_h と s の一方の数値を固定し、もう一方の数値を変化させた動画像を評価データとして用いる。このパラメータの固定値は、事前実験に基づき経験的に決定した。具体的には、 $s = 9$ に固定し、 K_h を0, 10, 20, 30, 40に変化させた。また、小物体では $K_h = 30$ に、屋外環境では $K_h = 40$ に固定し、 s を1, 3, 5, 7, 9に変化させた。また、Katagiriらの手法 [28] との比較のために $s = 0$ かつ $K_h = 1$ の場合の動画像も比較対象に加えた。これら10通りの動画像をランダムに配置したものを被験者に同時に提示し、テクスチャちらつきの少なさと動画の品質について最低値1, 最高値5の5段階で評価してもらった。本実験には20代学生の男女13名の被験者が参加した。以下では、小物体1の動画像に関する評価結果を図55, 小物体2の動画像に関する評価結果を図56, 屋外環境シーン1の動画像に関する評価結果を図57, 屋外環境シーン2の動画像に関する評価結果を図58に示し、結果に関する考察を述べる。

各シーンの、従来手法($K_h = 0, s = 1$), 整合性の項のみを考慮した場合($K_h = 30$ または $40, s = 1$), 輝度調整のみを行った場合($K_h = 0, s = 9$), 提案手法($K_h = 30$ または $40, s = 9$)の評価結果をまとめたグラフ(図55-図58の右図)から、ほとんどのシーンにおいて、整合性の考慮および時間的な輝度調整がそれぞれ有効であり、また両方を合わせることでさらに高品質な自由視点動画像を生成できてきていることが分かる。ただし、小物体2に関しては整合性のみを考慮する手法の結果はKatagiriらの手法で生成した結果よりも点数が低くなっており、この項単独では有効でない場合があることも確認した。これは、前述のように、カメラパラメータや3次元形状復元の誤差が大きく、それによってフレーム間の物体の同一箇所を投影する画素間で適切な対応関係が得られなかったためではないかと考えられる。また、このような対象では、低周波成分と高周波成分のテクスチャ両方を多く含んでいるため、テクスチャの位置ずれが視覚的に大きく影響したのではないかと考えられる。しかし、このようなフレーム間のテクスチャの不整合の影響は輝度調整処理によって十分に低減できている。また、図55-図58

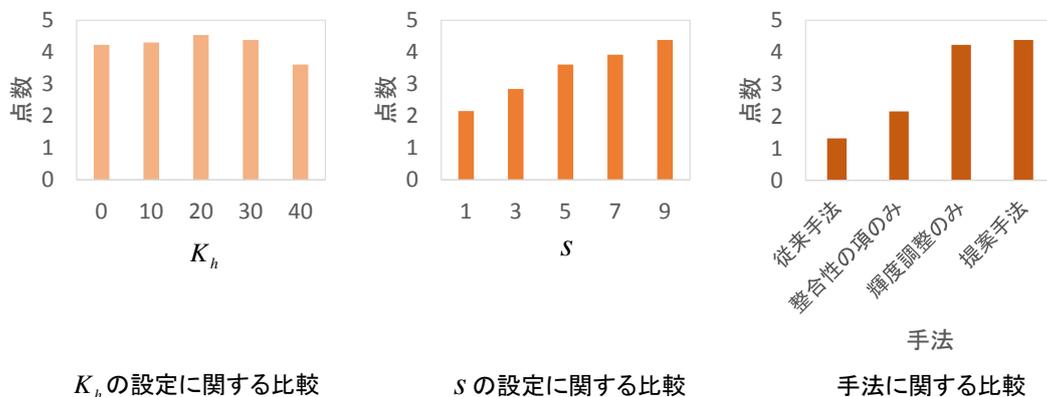


図 55 被験者実験の結果 (小物体 1)

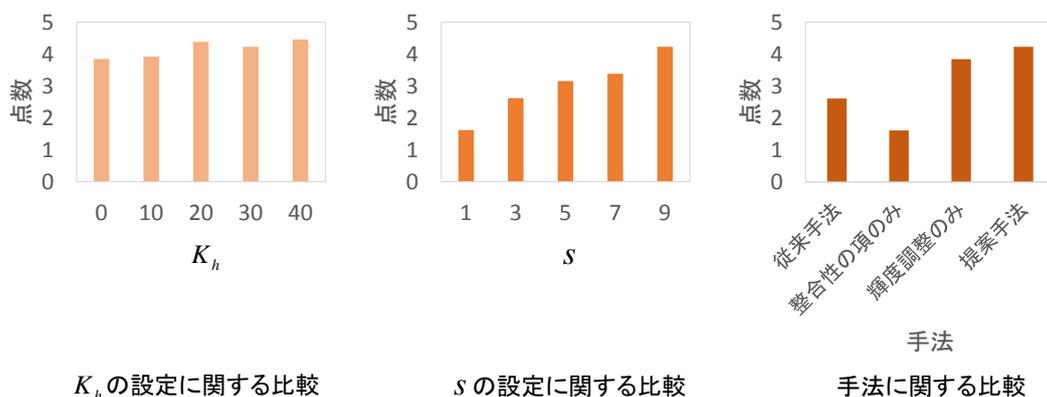


図 56 被験者実験の結果 (小物体 2)

の中央の図より、各シーンで s が大きくなるほどより高品質な動画像を生成できていることが分かる。

小物体 1 の結果に関して、図 55 の左のグラフに注目すると、 K_h が 40 のときに評価が低くなっていることが分かる。これは、4.2 節で述べたように、 K_h の重みが大いことにより、一つのテクスチャが選択されたときに他のテクスチャに切り替わりにくくなり、テクスチャの切り替えが起こったときに不連続が生じたためであると考えられる。また、屋外環境シーン 2 の評価結果に注目すると、従来手法よりも結果が良くなっているものの、全体的に他のシーンよりも評価が低

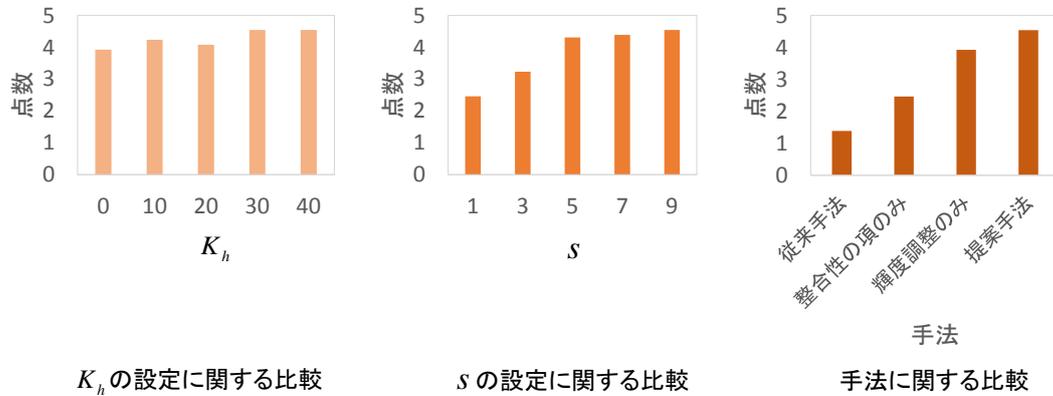


図 57 被験者実験の結果 (屋外環境シーン 1)

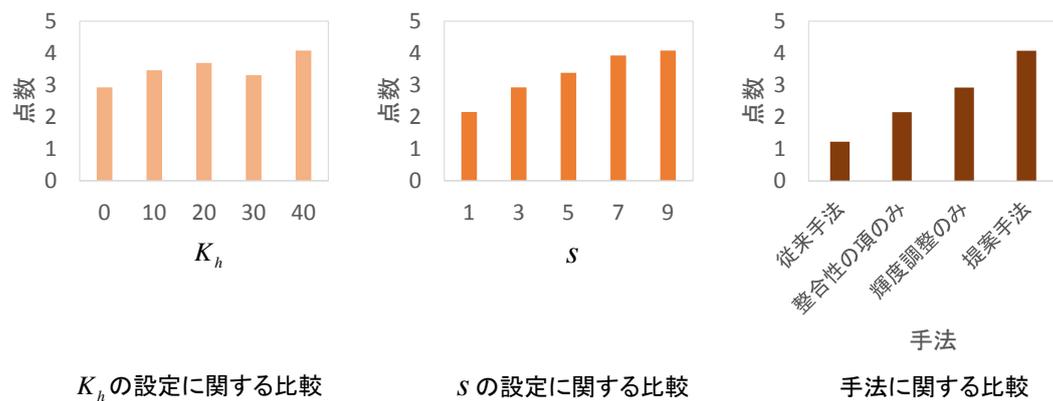


図 58 被験者実験の結果 (屋外環境シーン 2)

くなっていることが分かる。これは、4.3 節でも述べたように、シーン中に水面の鏡面反射が多くみられるため、他のシーンに比べて動画像がちらついて見えてしまったのではないかと考えられる。

以上の実験結果をまとめると、いずれのシーンにおいても、適切なパラメータを設定して生成した動画像は従来手法で生成した動画像よりも高い評価を得られ

たため、提案手法の有効性が確認されたと言える。

5. まとめ

本論文では、フレーム間のテクスチャの整合性を考慮したテクスチャ選択により、時間的なテクスチャの不連続を低減した自由視点動画画像生成手法を提案した。提案手法では、対象物体の3次元形状上の同一箇所を投影する対象フレームの画素とその前フレームの画素を対応付け、前フレームで選択された入力画像中の画素と類似した画素を持つ入力画像を選択するエネルギー関数を設計し、グラフカットアルゴリズムを用いてこの関数を最小化することで、対象フレームの各画素に対して適切な入力画像を選択する。また、この処理で抑制しきれなかった空間的・時間的な不連続に対し輝度調整処理を行うことで、不連続を低減する。

実験では、実空間の小物体や屋外環境のシーンを対象とし、設計したエネルギー関数と時間的な輝度調整のパラメータに様々な値を設定して生成された動画画像を比較することで、提案手法の有効性を示した。また、被験者実験の結果からも、提案手法が有効であることを確認できた。

提案手法では、時間的な画素値の不連続を低減した、視覚的に違和感の少ない動画画像を生成可能である。しかし、各フレームの画像生成に時間を要することから、リアルタイム性が要求されるアプリケーションへの応用が困難である。そのため、処理時間の短縮が今後の課題となる。

謝辞

本研究を進めるにあたり，終始温かく御指導，御鞭撻を頂いた視覚情報メディア研究室 横矢直和 教授に心より感謝申し上げます。副指導教員として有益な御助言を頂きました光メディアインタフェース研究室 向川康博 教授に厚く御礼申し上げます。本研究を進めるにあたり細やかな御指導，御助言を頂いた視覚情報メディア研究室 佐藤智和 准教授に厚く御礼申し上げます。本研究への多大なる御助言，御協力を頂いた視覚情報メディア研究室 河合紀彦 助教に厚く御礼申し上げます。本研究への的確な御助言，御協力を頂いた，視覚情報メディア研究室 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，研究室生活において様々な支援をして頂いた，視覚情報メディア研究室秘書 中村美奈 氏，南あずさ 氏，石谷由美 氏に厚く御礼申し上げます。そして，研究のみならず研究室での生活全般においてお世話になりました視覚情報メディア研究室の皆様深く感謝致します。最後に，長い学生生活を温かく見守り，支えてくださった家族に心より感謝の意を表します。

参考文献

- [1] 新宅宏彰. ハードウェア制御による高速三次元計測システムの構築と評価. 博士論文, 東京大学大学院工学系研究科, 2006.
- [2] 野口悠. 視覚的な自然さを考慮したカメラ位置指示機能を有する実物体のインタラクティブ仮想化システム. 修士論文, 奈良先端科学技術大学院大学情報科学研究科, 2014.
- [3] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362–1376, 2009.
- [4] S. M. Seitz and C. R. Dyer. View morphing. *Proc. ACM SIGGRAPH*, pp. 21–30, 1996.
- [5] 九州大学大学院システム情報科学研究院 情報知能工学部門イメージ・メディア理解研究室. ライトフィールドビジョン. <http://limu.ait.kyushu-u.ac.jp/project/project003.html>.
- [6] A. Davis, M. Levoy, and F. Durand. Unstructured light fields. *Journal of Computer Graphics Forum*, Vol. 31, No. 2, pp. 305–314, 2012.
- [7] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. *Proc. conf. on Computer graphics and interactive techniques*, pp. 11–20, 1996.
- [8] 高井勇志, 延原章平, 田辺太一, 松山隆司, ウ小軍. 視点依存形状最適化による高精細自由視点画像生成. 画像の認識・理解シンポジウム (MIRU2012), 2012.
- [9] J. Kopf, F. Langguth, D. Scharstein, R. Szeliski, and M. Goesele. Image-based rendering in the gradient domain. *ACM Trans. on Graphics*, Vol. 32, No. 6, pp. 199:1–199:9, 2013.

- [10] J. Kopf, M. F. Cohen, and R. Szeliski. First-person hyper-lapse videos. *ACM Trans. on Graphics*, Vol. 33, No. 4, pp. 78:1–78:10, 2014.
- [11] 大倉史生, 神原誠之, 横矢直和. 空撮画像を用いた拡張テレプレゼンス-無人飛行船の自動操縦と全方位カメラによる AR システム-. 画像の認識・理解シンポジウム (MIRU2010) 講演論文集, pp. 1183–1189, 2010.
- [12] M. Dellepiane, R. Marroquim, M. Callieri, P. Cignoni, and R. Scopigno. Flow-based local optimization for image-to-geometry projection. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 18, No. 3, pp. 463–474, 2012.
- [13] 浅井俊弘, 神原誠之, 横矢直和. 全方位距離画像と全方位カラー画像の統合による屋外環境の三次元モデル化. 画像電子学会誌, Vol. 34, No. 5, pp. 529–538, 2005.
- [14] 大池祐輔, 新宅宏彰, 池田誠, 浅田邦博. 光切断法による高精細・実時間 3次元撮像システム. 映像情報メディア学会誌, Vol. 57, No. 9, pp. 1149–1151, 2003.
- [15] H. Baker. Three-dimensional modeling. *Proc. Int. Joint Conf. on Artificial Intelligence*, Vol. 2, pp. 649–655, 1977.
- [16] K. Kolev, T. Brox, and D. Cremers. Fast joint estimation of silhouettes and dense 3D geometry from multiple images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 3, pp. 493–505, 2012.
- [17] S. M. Seitz N. Snavely and R. Szeliski. Modeling the world from internet photo collections. *Proc. ACM SIGGRAPH*, Vol. 80, No. 2, pp. 189–210, 2008.
- [18] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *Int. Journal of Computer Vision*, Vol. 38, No. 3, pp. 199–218, 2000.

- [19] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly supported surfaces. *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 3121–3128, 2011.
- [20] J. Xiao and M. Shah. From images to video: view morphing of three images. *Proc. Vision, Modeling and Visualization*, pp. 495–502, 2003.
- [21] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Proc. conf. on Computer graphics and interactive techniques*, pp. 39–46, 1995.
- [22] 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.
- [23] M. Levoy and P. Hanrahan. Light field rendering. *Proc. conf. on Computer graphics and interactive techniques*, pp. 31–42, 1996.
- [24] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumi-graph. *Proc. conf. on Computer graphics and interactive techniques*, pp. 43–54, 1996.
- [25] 源田大輔, 向川康博, 尺長健. 視点と法線を組み合わせた任意視点映像の生成法. *情報処理学会研究報告コンピュータビジョンとイメージメディア*, Vol. 2003, No. 36, pp. 53–60, 2003.
- [26] P. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. *Proc. Eurographics Rendering Workshop*, pp. 105–116, 1998.
- [27] G. Chaurasia, S. Duchene, O. S. Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. on Graphics*, Vol. 32, No. 3, pp. 30:1–30:12, 2013.

- [28] 片桐敬太. 視点依存テクスチャマッピングにおけるテクスチャの連続性を考慮した自由視点画像生成. 修士論文, 奈良先端科学技術大学院大学情報科学研究科, 2015.
- [29] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, pp. 3–20, 1991.
- [30] 蚊野浩. ライトフィールドカメラ lytro の動作原理とアルゴリズム. 光アイアンス, Vol. 24, No. 7, pp. 1–6, 2013.
- [31] C. Wu. Visualefm: A visual structure from motion system. <http://www.cs.washington.edu/homes/ccwu/vsfm/>, 2011.
- [32] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. journal of computer vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [33] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on pattern analysis and machine intelligence*, Vol. 23, No. 11, pp. 1222–1239, 2001.