

NAIST-IS-MT0651005

修士論文

奥行き画像を利用した時空間超解像画像の生成

栗津 優作

2008年3月17日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

栗津 優作

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)

山澤 一誠 准教授 (副指導教員)

佐藤 智和 助教 (副指導教員)

奥行き画像を利用した時空間超解像画像の生成*

栗津 優作

内容梗概

実画像をユーザに提示することで遠隔地に居る感覚を与える技術はテレプレゼンスと呼ばれ、医療、娯楽、教育などの様々な分野への応用が期待されている。これらの分野で用いられるテレプレゼンスシステムでは、高い臨場感を再現することが求められる。高い臨場感を再現するためには、空間解像度および時間解像度の高い映像の提示が必要であり、近年、画像の空間解像度を向上させる超解像に関する研究や動画の時間解像度を向上させるフレーム補間に関する研究が盛んに行われている。前者の画像の超解像に関する研究では、一般に視点位置の異なる複数の地点で撮影された画像群に対して各画素をサブピクセル精度で対応付けることで、空間中の同一箇所に対する冗長な複数の観測データを用いて空間解像度を向上させる。しかし、サブピクセル精度で各画素を対応付けるためには、撮影対象の形状に関して平面仮定等の制約が必要となり、対象の形状が複雑な場合には良好な結果を得ることが難しいという問題がある。一方、後者の動画のフレーム補間に関する研究においても、隣接するフレーム間の画素の対応点を用いることで補間フレームを生成するが、カメラの視点位置・視線方向の変化に伴う撮影範囲の変更やオクルージョンの発生により、2フレーム間で対応点が存在しない場合には良好な結果を得ることが難しい。このような問題に対して本論文では、動画の各画像取得時のカメラパラメータを既知とした上で、各画素の奥行き値をパラメータとして多数の画像上での対応点を同時に決定することで時空間超解像画像を生成する手法を提案する。本手法では、各画像の撮影位置およびそ

* 奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT0651005, 2008年3月17日.

これらの中間視点位置に仮想的な高解像度カメラを設定し、仮想カメラの各画素に対する奥行きを周辺のフレームの複数の画像を用いてエネルギー最小化の枠組みで探索することで時空間超解像画像を生成する。本手法では、多数の画像に対する対応点を奥行きの1次元探索によって決定できるため、従来用いられていた平面仮定等を用いる必要がなく、空間解像度の向上において、複雑な形状の物体を含む動画像を扱うことができる。また、隣接するフレームの画像だけでなく多数の画像を利用することで、隣接するフレーム間で対応点が存在しない場合にも他のフレームの画像を用いて補間フレームを生成でき、空間解像度の向上と同一の枠組みで時間解像度を向上させる。実験では、仮想環境を用いたシミュレーション動画および実動画に対する時空間超解像画像の生成を行うことで提案手法の有効性を示す。

キーワード

時空間超解像, 奥行き画像, エネルギー最小化

Spatio-Temporal Super-Resolution Using Depth Maps*

Yusaku Awatsu

Abstract

Telepresence system enables users to experience a remote site virtually. The system can be widely used in a number of fields such as medical care, entertainment and education. In these fields, high spatial and temporal resolution images are required to give users high realistic sensation. Spatial and temporal super-resolution methods have been intensively investigated. Spatial super-resolution is based on establishing correspondences with subpixel accuracy between pixels in multiple images obtained from different viewpoints. However, in conventional methods, the planar assumption for the target scene is generally employed to achieve subpixel correspondences between images. Therefore, it is difficult to obtain good results for complex scenes. On the other hand, temporal super-resolution methods generate an intermediate frame by using the corresponding pixels in adjacent frames. When many pixels have no corresponding pixels in adjacent frames due to occlusions, intermediate frames can hardly be generated. In this thesis, in order to solve these problems, a spatio-temporal super-resolution method using depth maps is proposed. In the proposed method, the depth maps are used as the parameters for simultaneous determination of the corresponding pixels in multiple images by assuming that intrinsic and extrinsic camera parameters are known. Concretely, first, virtual cameras with high spatial resolution

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0651005, March 17, 2008.

are arranged at the captured positions and intermediate positions of adjacent frames. Then, the depth values are determined for all the pixels in the virtual frames with an energy minimization framework using multiple frames. The proposed method can determine the corresponding pixels in multiple frames by one-dimensional search for the depth values without the planar assumption, and thus spatial resolution can be increased even for complex scenes. In addition, since multiple frames can be used in the proposed method, temporal resolution can be increased even when large parts of the image are disappeared in the adjacent frame. In experiments, the validity of the proposed method is demonstrated by generating spatio-temporal super-resolution images in simulation and for real movies.

Keywords:

super-resolution, depth map, energy minimization

目次

1. はじめに	1
2. 時空間超解像の関連研究と本研究の位置づけ	4
2.1 空間解像度の向上に関する研究	4
2.1.1 データベースなどの事前知識を用いる手法	4
2.1.2 対象となる画像や動画のみを用いる手法	5
2.2 時間解像度の向上に関する研究	9
2.2.1 隣接するフレーム間の特徴点の動きを用いる手法	9
2.2.2 自由視点画像生成手法	10
2.3 時間および空間解像度を同時に向上させる研究	10
2.4 本研究の位置づけと方針	12
3. 奥行き画像を用いた時空間超解像画像の生成	14
3.1 画像超解像による空間解像度の向上	14
3.1.1 画像の整合性と奥行きの滑らかさに基づくエネルギー関数の定義	14
3.1.2 エネルギー最小化による奥行き値の最適化に基づく超解像画像の生成	19
3.2 新視点画像生成による時間解像度の向上	20
4. 実験	23
4.1 シミュレーション動画に対する超解像画像の生成実験	23
4.1.1 空間超解像に有効な画像枚数の検証	25
4.1.2 生成された時空間超解像画像の画質の定量的評価	28
4.2 実動画に対する超解像画像の生成実験	34
4.2.1 実験条件	34
4.2.2 超解像画像の生成結果と考察	34
5. まとめ	39

謝辭	40
參考文獻	41

目 次

1	Freeman ら [3] の手法を用いた空間解像度の向上の例	6
2	バイリニア補間による拡大の例	7
3	Ezra ら [11] の用いた位置制御が可能なカメラ	8
4	松延ら [26] の用いたカメラシステム	11
5	提案手法における入力と出力	15
6	画像の整合性に関するエネルギーの概念	16
7	超解像画像と観測画像の画素の対応関係	18
8	オクルージョンによる奥行き値の変化	19
9	補間フレームの設定	21
10	実験におけるカメラの動きと物体の配置	24
11	入力画像, 正解画像, 生成画像の比較	26
12	超解像に利用する画像枚数と PSNR の関係	27
13	PSNR の比較	29
14	超解像画像の比較 (41 フレーム目)	31
15	奥行きの変化	32
16	奥行きの滑らかさの影響	33
17	入力として用いた空撮動画画像の一部	35
18	超解像対象フレーム (第 16 フレーム) の奥行き値	36
19	入力画像と生成画像の比較	37
20	利用する画像枚数を変化させた時の生成画像	38

表 目 次

1	実験で用いたパラメータ	23
2	それぞれの手法で生成した動画画像の平均 PSNR	29

1. はじめに

実画像をユーザに提示することで遠隔地にいる感覚を与える技術はテレプレゼンスと呼ばれ、医療、娯楽、教育などの様々な分野への応用が期待されている。これらの分野で用いられるテレプレゼンスシステムでは、高い臨場感を再現することが求められており、臨場感向上のための様々な手法が提案されている [1, 2]。なかでも、人間の視覚から得られる臨場感に関する要素として、提示されるコンテンツが高空間解像度、高時間解像度であることが重要であると考えられている。しかし、一般に映像を取得するカメラの撮像素子の数は有限であり、また読み出し速度に制限があるため、これらの要素を同時に満たす映像を取得することは難しい。そこで、一般的なカメラによって取得された画像の空間解像度、時間解像度を向上させる研究が盛んに行われている。

空間解像度の向上に関する従来研究は、データベースなど事前知識を利用する手法 [3, 4, 5, 6, 7] と対象となる画像や動画のみを利用する手法 [8, 9, 10, 11, 12, 13, 14, 15, 16, 17] に大別できる。データベースを用いる手法では、予め低解像度画像と高解像度画像のペアから相関などを学習しておく手法 [3, 4, 5, 6] や対象となるシーンの高解像度画像を保持しておき、高解像度画像の画素と入力画像の画素を対応付けることで空間解像度を向上させる手法 [7] などがある。これらの手法は、対象となるシーンや物体が限定的である場合に有効であるが、生成される高解像度画像の品質がデータベースに大きく依存するため、対象シーンに応じたデータベースが必要となる。データベースなどの事前情報を用いずに空間解像度を向上させる手法には、対象となる 1 枚の画像のみから空間解像度を向上させる手法 [8, 9] と、動画など複数の画像を入力として用いて空間解像度を向上させる手法 [10, 11, 12, 13, 14, 15, 16, 17] がある。前者は、対象画像内のフラクタル性や周辺の画素値の統計量などを利用して補間する画素値を決定し、空間解像度を向上させる。これらの手法は、計算コストが低いという特長があるが、一般に複数の画像から空間解像度を向上させる手法に比べて、解像度の向上効果が低く、また対象となる画像の種類に大きく依存する。後者は、異なる視点から撮影された複数の入力画像を対象画像にサブピクセル精度で対応付けることで高解像度画像の画素値を推定し、空間解像度を向上させる。この手法では、複数の入力画像

上におけるシーンの同一箇所に対する画素値を推定に利用するため，利用する画像枚数が多いほど画質の向上が期待できる．しかし，サブピクセル精度での対応付けが必要となるため，従来手法では平面仮定など撮影対象に関する制約などを用いており，複雑な形状を含むシーンに対しては，良好な結果を得ることが難しいという問題がある．

時間解像度の向上に関する従来研究は，入力動画像で隣接するフレームの間に補間フレームを挿入し，何らかの手法によって補間フレームの画像を生成することで，時間解像度を向上させる．補間フレームを生成する手法として，隣接フレームの特徴点の動きを用いたモーフィングによる手法 [18, 19, 20, 21] が提案されている．この手法では，2枚以上の入力画像間で対応点を与え，補間フレームにおける対応点の位置に応じてテクスチャを変形させることで補間フレームを生成する．このアプローチでは，生成される補間フレームの画像の品質は与えられる対応点の数に大きく依存する．しかし，一般に二枚の画像上で正しい対応点を密に決定することは難しく，特にオクルージョンなどで画像間で対応点が存在しない場合には良好な結果を得ることは難しい．一方，自由視点画像生成と呼ばれる分野の研究 [22, 23, 24, 25] も中間視点の画像を生成するために有効である．このような研究には様々なアプローチが存在するが，多くの手法ではモーフィングによる手法と同様に対応点が必要であり，いかに正しい対応点を密に得るかが課題となる．

複数台のカメラで同一視野を撮影した動画像を統合することで，時間および空間解像度を同時に向上させる手法 [26, 27, 28] も提案されている．これらの手法は，時間解像度の高いカメラを用いたり，複数のカメラを同期撮影することで動物体が存在する動画像にも対応が可能である．しかし，これらの手法は位置関係が固定された複数のカメラや特殊なカメラが必要になり，一般的なカメラで取得された動画像に対して手法を適用することはできない．

本論文では，時空間解像度の向上において，サブピクセル精度での密な対応点の決定が重要であることに着目し，新たに各画素値の奥行きをパラメータとして複数の画像間の対応点を1次元探索により決定することで，サブピクセル精度での密な対応点の決定を実現する手法を提案する．本研究では，各画像取得時のカ

メラパラメータを既知とした上で、複数の画像に対する対応点を奥行きの1次元探索によって決定することで対応点探索のパラメータの自由度を減らし、平面仮定などの撮影対象に関する強い制約を課すことなく、各画素を高精度に対応付ける。また、推定される奥行きに基づいてオクルージョンを考慮した対応点の探索を行うことで、従来手法では正しい対応が得られないシーンにおいても時間解像度を向上させる。ただし、本研究では静的なシーンのみを扱い、動画像中に動物体は存在しないものとする。

以降、2章では、時空間超解像に関する従来研究と本研究の位置づけについて述べ、3章では、時空間超解像画像の生成手法について詳述する。4章では、仮想環境を用いたシミュレーション動画および実動画に対する時空間超解像画像の生成を行うことで提案手法の有効性を示す。最後に5章で、本論文のまとめと今後の課題について述べる。

2. 時空間超解像の関連研究と本研究の位置づけ

本章では，本研究に関連する従来研究と本研究の位置づけについて述べる．まず，従来の時空間解像度の向上に関する研究を，(1) 空間解像度の向上に関する従来研究，(2) 時間解像度の向上に関する従来研究，(3) 時間および空間解像度を同時に向上させる従来研究に分類し，それらの特徴と問題点について述べ，最後に従来研究に対する本研究の位置づけについて述べる．

2.1 空間解像度の向上に関する研究

本節では，空間解像度の向上に関する従来研究を，データベースなどの事前知識を用いる手法 [3, 4, 5, 6, 7] と，対象となる画像や動画のみを用いる手法 [8, 9, 10, 11, 12, 13, 14, 15, 16, 17] に分類し，それぞれの特徴と問題点について述べる．

2.1.1 データベースなどの事前知識を用いる手法

データベースなどの事前知識を必要とする手法は，事前の学習によってデータベースを構築する手法 [3, 4, 5, 6] と対象となるシーンに直接対応する高解像度画像を用いる手法 [7] がある．

事前の学習によってデータベースを構築する手法では，高解像度画像と低解像度画像に関する相関や規則性を学習しておき，新たに撮影した低解像度画像の空間解像度を向上させる．Freeman ら [3] は画像を周波数領域に変換し，マルコフネットワークによって低解像度画像と高解像度画像のペアから予め特徴を学習しておき，入力された低解像度画像に対して，対応する高解像度画像を推定することで空間解像度を向上させる手法を提案している．溝内ら [4] は低解像度画像と高解像度画像の集合は，同じ構造をした多様体を持つと仮定し，多様体空間において，低解像度画像と高解像度画像を対応付けることで，空間解像度を向上させる手法を提案している．また，上村ら [5] は低解像度画像と高解像度画像の各画素を周波数特徴ベクトルで表し，それらを局所的なテクスチャ特徴量によって近

似して対応関係を学習しておくことで、計算コストを低く抑えた空間解像度の向上手法を提案している。一方で、Beginら [6] は、低解像度画像と高解像度画像の相関を学習した上で、画像の劣化過程を示す PSF (Point Spread Function) も推定することで、より正確な高解像度画像を推定する手法を提案している。これらの事前の学習を必要とする手法は、図 1 に示すように、生成される高解像度画像の品質がデータベースに依存するため、対象シーンに応じたデータベースが必要となる。

対象となるシーンに直接対応する高解像度画像を用いる手法では、予め対象シーンに関する高解像度画像を得ておき、入力画像の画素と高解像度画像の画素を対応付けることで空間解像度を向上させる。Bhatら [7] は、入力となる動画像と同一のシーンで撮影した高解像度な静止画を用い、高解像度画像の各画素を奥行き値を介して入力画像の画素に対応付けることで空間解像度を向上させる手法を提案している。この手法では、予めシーンの高解像度画像を保持しておく必要があるため、対象シーンが限定的である場合には有効であるが、一般的な動画像のみを入力として用いることはできない。

以上のように、データベースなどの事前知識を用いる手法は、対象となるシーンや物体が限定的である場合に有効であるが、生成される高解像度画像の品質がデータベースに依存するため、対象シーンに応じたデータベースが必要となるという問題がある。

2.1.2 対象となる画像や動画像のみを用いる手法

対象となる画像や動画像のみを用いて空間解像度を向上させる手法では、対象となる 1 枚の画像のみを用いて空間解像度を向上させる手法 [8, 9] と、動画像などの複数の画像を用いて空間解像度を向上させる手法 [10, 11, 12, 13, 14, 15, 16, 17] に大別できる。

- 対象となる 1 枚の画像のみを用いる手法

対象となる 1 枚の画像のみを用いる手法では、最近傍補間やバイリニア補間などのように、隣接する画素値や周辺の画素値の統計量を利用して補間



(d) 超解像画像 (トレーニングセットに (b) を使用)



(e) 超解像画像 (トレーニングセットに (c) を使用)

図 1 Freeman ら [3] の手法を用いた空間解像度の向上の例

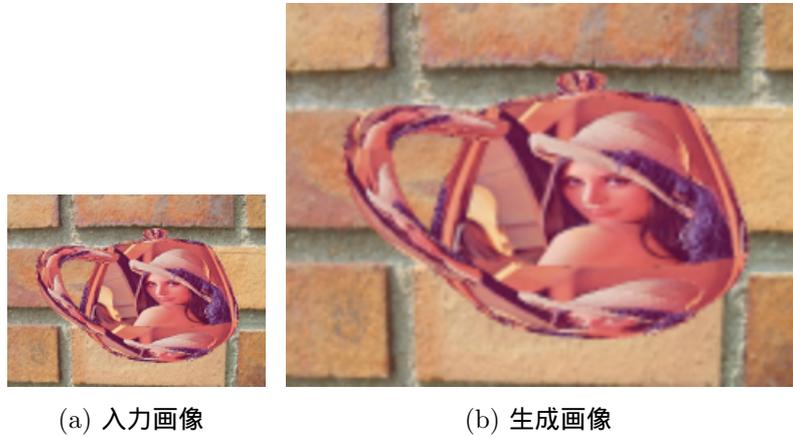


図 2 バイリニア補間による拡大の例

する画素値を決定し，空間解像度を向上させる．これらの手法は，人間の目が境界線などの不連続に対して敏感であることから，平滑化がベースとなっている．そのため，図 2 に示すように，生成される画像はエッジ部分や画像の細部でぼけてしまい，鮮明さに欠けるという問題がある．川合ら [8] は，画像の局所領域のウインドウの大きさを変えながら固有ベクトルを生成した場合に，この固有ベクトルがあまり変化しない性質を利用し，画像の局所領域を拡大することにより，空間解像度を向上させる手法を提案している．Hong ら [9] は，画像の局所的な統計量を用いて空間解像度を向上させる手法を提案している．このようなアプローチは，計算コストが低いという特長があり，市販ソフトでも広く利用されているが，一般に複数の画像から空間解像度を向上させる手法に比べて空間解像度の向上効果は低い．

- 動画像などの複数の画像を用いる手法

複数の画像を用いる手法では，異なる視点から撮影された複数の入力画像を対象画像にサブピクセル精度で対応付けることで高解像度画像の画素値を推定し，空間解像度を向上させる．高解像度画像の画素値の推定方法としては，観測画像の画素値を直接ブレンドする手法や推定した高解像度画

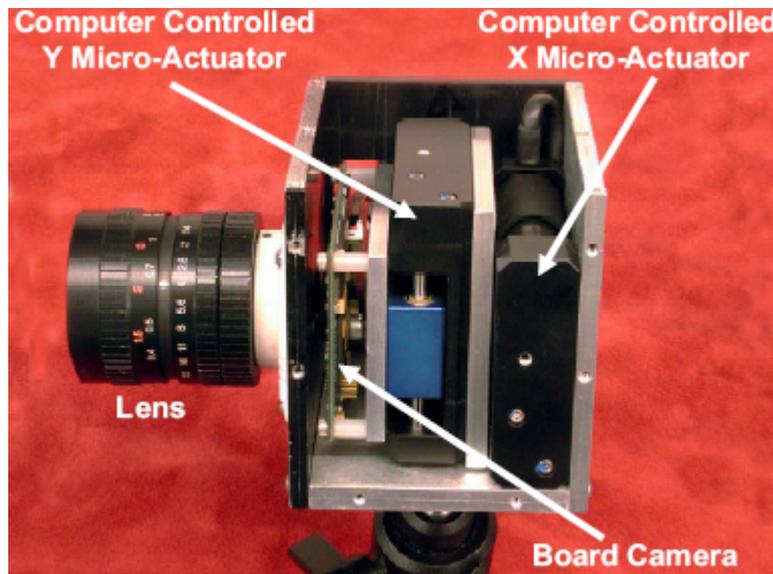


図 3 Ezra ら [11] の用いた位置制御が可能なカメラ

像から生成した低解像度画像と対応する観測画像の画素値の差を最小化する手法が用いられるが、どちらの手法においても生成される高解像度画像の画質は画像間の対応付けの精度に大きく依存する。従来、画像間の幾何学的な変形を既知とした上で、対象画像に他の観測画像を合成する手法 [10] や、図 3 に示すような位置制御可能な特殊なカメラを用いて、カメラの微小な動きを制御することで画素間の対応関係を与える手法 [11] のように、画像間の画素の動きが高精度に与えられていることを前提とした手法も存在するが、一般的には何らかの手法によって画像間の対応関係を求める必要がある。対応関係を自動で求める手法では、オプティカルフローを求める手法 [12, 13] などがあるが、従来手法では、高精度な位置付けを行うために以下に示すような仮定や限定が用いられる。

- フレーム間の動きが微小であると仮定する手法 [14]
- 動きを水平・垂直のみに限定した手法 [15, 16]
- 平面射影変形を仮定する手法 [17]

このような、撮影対象に関する制約を用いた場合には、複雑な形状を含むシーンに対しては良好な結果が得ることが難しいという問題がある。

2.2 時間解像度の向上に関する研究

時間解像度の向上に関する従来研究は、入力動画で隣接するフレームの間に補間フレームを挿入し、何らかの手法によって補間フレームの画像を生成することで、時間解像度を向上させる。

補間フレームの画像の生成には、隣接するフレーム間の特徴点の動きを用いる手法 [18, 19, 20, 21] が提案されている。また、自由視点画像生成と呼ばれる分野の研究 [22, 23, 24, 25] も補間フレームの位置の視点画像を生成できるため、時間解像度の向上に有効である。以下では、それぞれの手法の特徴と問題について述べる。

2.2.1 隣接するフレーム間の特徴点の動きを用いる手法

隣接するフレームの特徴点の動きを用いる手法として、隣接する画像間で動きベクトルを求め、補間フレームにおける対応点の位置を算出し、テクスチャを変形させることで補間フレームの画像を生成する動き補償フレーム内挿法と呼ばれる手法 [18, 19, 20] が提案されている。これらの手法では、生成される補間フレームの画像の品質が与えられる対応点の数に大きく依存し、特にオクルージョンなどで画像間の対応点が存在しない場合には良好な結果を得ることが難しい。この問題に対して、大脇ら [21] は補間フレーム画像を領域分割し、動き推定によって前後フレームとの一致・不一致領域を判定することでオクルージョンの有無を判定し、オクルージョンが存在する場合は、さらに前後2フレームを参照することで、オクルージョンを考慮した補間画像を生成する手法を提案している。しかし、一般に二枚の画像上で対応点を密に決定することは難しく、また離れたフレーム間で対応点を正確に求めることは困難である。

2.2.2 自由視点画像生成手法

自由視点画像生成に関する手法としては、様々な手法が提案されているが、フレーム間の中間視点を対象として画像を生成することで、時間解像度の向上にも利用することができる。自由視点画像生成手法は、画像間の対応点を用いて自由視点画像を生成する手法 [22, 23] と対応点を用いずに自由視点画像を生成する手法 [24, 25] に分類できる。前者では、観測画像の画素と自由視点画像の画素を対応付けし、テクスチャを投影・変形することで自由視点画像を生成する。ステレオ画像から奥行きを求め、2.5次元シーンを表現し、自由な視点位置の画像を生成する手法 [22] や位置関係を固定したカメラで撮影を行うことで、画像間の対応関係を既知とし、自由な視点画像を生成する手法 [23] が提案されている。しかし、これらの手法は対応点を必要とするため、隣接するフレーム間の特徴点の動きを用いる手法と同様にオクルージョンなどで画像間の対応点が存在しない場合には良好な結果を得ることが難しい。

後者では、画像を光線情報の集合として捉え、光線を観測位置と方向によってモデル化し、自由視点位置において画像平面を通過する光線を画像化することで自由視点画像を生成する [24, 25]。しかし、この手法では任意視点位置に設定した仮想カメラから投影される最も近い光線を用いるため、フレーム間の中間視点を対象とした場合、一般に中間視点の画像は隣接するフレームの画像のみから構成され、特に隣接するフレーム間で視点位置や視線方向が大きく変わる場合には、良好な結果を得ることが難しい。

2.3 時間および空間解像度を同時に向上させる研究

従来提案されている時間および空間解像度を同時に向上させる手法 [26, 27, 28] は、複数台のカメラで同一視野を撮影した動画像を統合することで時空間解像度を向上させている。Shechtmanら [26] は、同性能のカメラを複数台使い、同一視野をサブピクセル・サブフレームずらして撮影し、それらを統合することで、時空間解像度の高い動画像を生成する手法を提案している。また、松延ら [27] や今川ら [28] は、高空間解像度かつ低時間解像度なカメラと低空間解像度かつ高時間

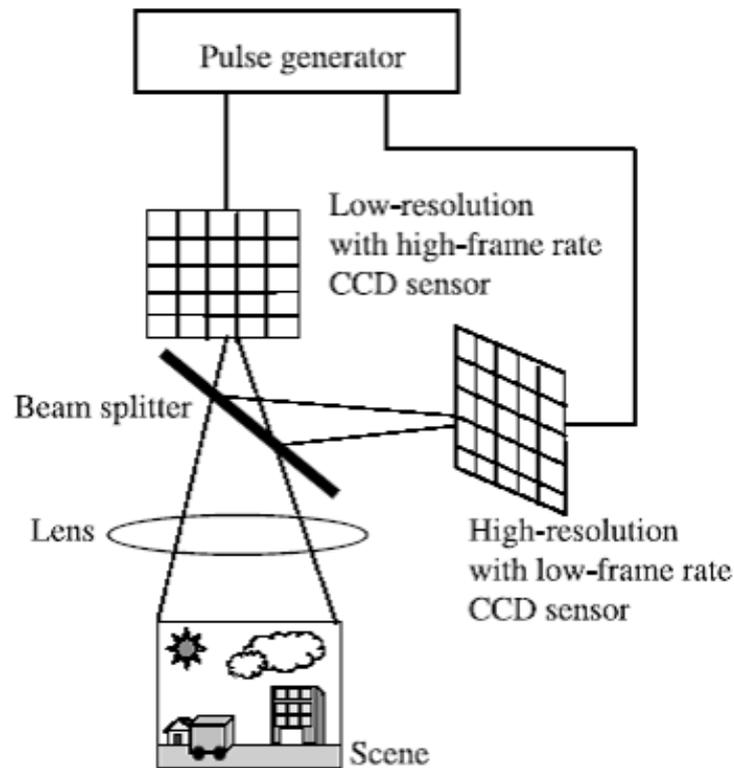


図 4 松延ら [26] の用いたカメラシステム

解像度なカメラで同一視野を同期撮影し，前者のカメラからはシーン情報を，後者のカメラからは動き情報を抽出し合成することで，時空間解像度の高い動画像を生成する手法を提案している．

これらの手法は，動物体が存在する動画像にも対応が可能であるという特長を持つが，位置関係が固定された複数のカメラや特殊なカメラ (図 4 参照) が必要になり，一般的なカメラで取得された動画像に対して手法を適用することはできない．

2.4 本研究の位置づけと方針

前節までに概観したように，空間解像度および時間解像度の向上に関する手法は既に多く提案されている．空間解像度の向上において，データベースを用いる手法では，予め低解像度画像と高解像度画像のペアから相関などを学習しておくか，対象となるシーンの高解像度画像を保持しておき，高解像度画像の画素と入力画像の画素を対応付けることで空間解像度を向上させる．これらの手法は，対象となるシーンや物体が限定的である場合に有効であるが，生成される高解像度画像の品質がデータベースに大きく依存するため，対象シーンに応じたデータベースが必要となる．データベースなどの事前情報を用いずに空間解像度を向上させる手法は，対象となる1枚の画像のみで空間解像度を向上させる手法と，動画像など複数の画像を入力として用いて空間解像度を向上させる手法がある．前者は，対象画像内のフラクタル性や周辺の画素値の統計量などを利用して補間する画素値を決定し，空間解像度を向上させる．これらの手法は，計算コストが低いという特長があるが，一般に複数の画像から空間解像度を向上させる手法に比べて，解像度の向上効果が低く，また対象となる画像の種類に大きく依存するという問題がある．後者は，異なる視点から撮影された複数の入力画像を対象画像にサブピクセル精度で対応付けることで高解像度画像の画素値を推定し，空間解像度を向上させる．この手法では，複数の入力画像上におけるシーンの同一箇所に対する画素値を推定に利用するため，利用する画像枚数が多いほど画質の向上が期待できる．しかし，サブピクセル精度での対応付けが必要となるため，従来手法では平面仮定など撮影対象に関する制約などを用いており，複雑な形状を含むシーンに対しては，良好な結果を得ることが難しいという問題がある．

時間解像度の向上においては，従来，隣接するフレーム間での特徴点の動きを用いて隣接するフレームの画像を変形させることにより補間フレームの画像を生成する手法や，シーンの3次元情報を推定し補間フレームと他のフレームを対応付けることで補間フレームの画像を生成する手法などが提案されている．これらの手法は，画像間の対応点を利用して補間フレームの画像を生成するため，生成される補間フレームの画質は，与えられた対応点の数や精度に依存する．一般に2枚の画像間において正しい対応点を密にかつ自動で与えることは難しく，特に

オクルージョンなどが存在する場合には、正しい対応点が得られず、良好な補間フレームの画像を生成することが難しい。

時間および空間解像度を同時に向上させる従来手法は、同一視野を複数台のカメラで撮影した動画像を統合することで時間解像度と空間解像度を同時に向上させる。複数台のカメラの同期撮影や時間解像度の高いカメラを用いて動き情報を抽出し合成することで、動物体を含む動画像にも対応が可能である。しかし、位置関係の固定された複数のカメラや特殊なカメラが必要となり、一般的なカメラで取得された動画像に対して手法を適用することができないという問題がある。

本論文では、動画像の空間解像度と時間解像度の向上には、サブピクセル精度での密な対応点の決定が重要であることに着目し、動画像の各画素値の奥行きをパラメータとして複数の画像間の対応点を1次元探索により決定することで、サブピクセル精度での密な対応点の決定を実現する手法を提案する。本研究では、動画像の各画像取得時のカメラパラメータを既知とした上で、複数の画像に対する対応点を奥行きによる1次元探索によって決定することで対応点探索のパラメータの自由度を減らし、平面仮定などの撮影対象に関する強い制約を課すことなく、各画素を高精度に対応付け、空間解像度を向上させる。また、推定される奥行きに基づいてオクルージョンを考慮した対応点の探索を行うことで、従来手法では正しい対応が得られないシーンにおいても時間解像度を向上させる。

3. 奥行き画像を用いた時空間超解像画像の生成

本章では，奥行き値を介して各フレームの画像の画素値を対応付けることにより，動画像の時空間超解像画像を生成する手法について述べる（図5参照）．ただし，本研究では静的シーンを撮影した動画像を対象とし，各フレームに対するカメラ位置・姿勢は既知とする．また，ステレオ法などにより求めた，ある程度信頼度の高い奥行き画像が初期値として与えられるものとする．提案手法では，空間解像度の向上のために，高解像度画像の尤もらしさに基づく画像の整合性に関するエネルギーと奥行きに関するエネルギーを定義し，これらの和を最小化することで空間解像度を向上させる．また，時間解像度の向上においても同様のエネルギーを用い，動画像上の補間フレームにおいて高解像度画像を生成することで時間解像度を向上させる．

以下，3.1節で奥行き画像を用いた空間解像度の向上手法について述べ，3.2節で新視点画像生成による時間解像度の向上手法について述べる．

3.1 画像超解像による空間解像度の向上

本節では，奥行き画像を用いて動画像の空間解像度を向上させる手法について述べる．空間解像度の向上処理では，生成される超解像画像の尤もらしさに基づくエネルギー関数を定義し，そのエネルギー関数を最小化することによって空間解像度を向上させる．以下，本研究で用いるエネルギー関数の定義とその最小化手法について詳述する．

3.1.1 画像の整合性と奥行きに関するエネルギー関数の定義

本研究では，図6に示すような超解像処理の対象となる第 f フレームの超解像画像の画素値と各フレームの観測画像の画素値との整合性に関するエネルギー E_{If} （以下，画像の整合性に関するエネルギー）と奥行きに関するエネルギー E_{Df} を用い，以下のように，エネルギー関数を定義する．

$$E_f = E_{If} + wE_{Df} \quad (1)$$

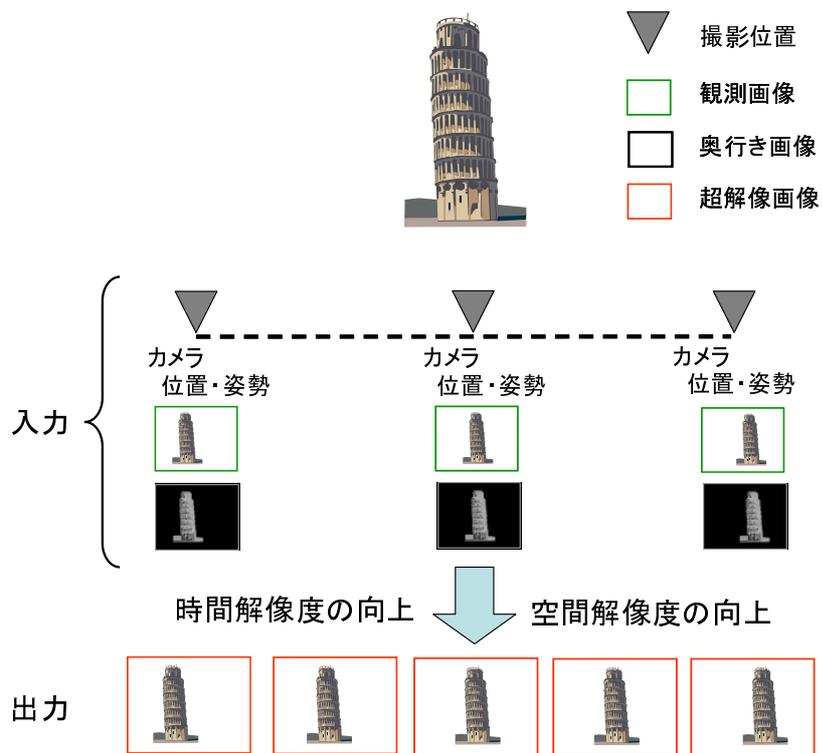


図 5 提案手法における入力と出力



- ▼ 撮影位置
- 観測画像
- 超解像画像
- シミュレート画像

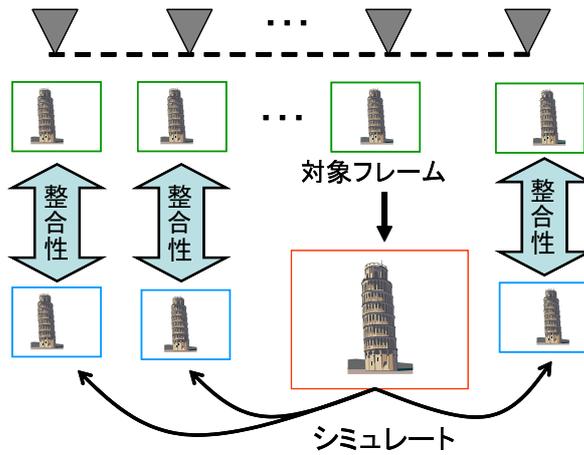


図 6 画像の整合性に関するエネルギーの概念

ここで、 w は重み係数である．以下では、エネルギー E_{If} 、 E_{Df} について順に詳しく述べる．

エネルギー E_{If} は、第 f フレームの超解像画像の尤もらしさを表すエネルギーであり、第 a フレームから第 b フレーム ($a \leq f \leq b$) までの観測画像を用い、次のように定義する．

$$E_{If} = \frac{\sum_{n=a}^b |\mathbf{N}(\mathbf{O}_n)(\mathbf{g}_n - \mathbf{m}_{nf})|^2}{\sum_{n=a}^b |\mathbf{O}_n|^2} \quad (2)$$

ただし、 $\mathbf{g}_n = (g_{n1}, \dots, g_{np})^T$ は第 n フレームの観測画像の画素値を p 次元のベクトルとして表記したものであり、 $\mathbf{m}_{nf} = (m_{nf1}, \dots, m_{nfp})^T$ は推定された第 f フレームの超解像画像と奥行き値に基づいて第 n フレームの観測画像をシミュレートした画像（以下、シミュレート画像）である（図 7 参照）．また、 $\mathbf{N}(\mathbf{O}_n)$ はベクトル \mathbf{O}_n の各要素を対角要素に持つ $p \times p$ の対角行列である． E_{If} は、基本的

に観測画像 g_n とシミュレート画像 m_{nf} の差分に基づいて計算されるが、シミュレート画像 m_{nf} の画素値には、オクルージョンや画面外への投影によって、第 f フレームの画像上の画素と直接対応付かないものが含まれる。本研究では、0 または 1 を要素に持つマスク画像 $O_n = (O_{n1}, \dots, O_{np})$ を用いて、式 (2) に示すように、直接対応しない画素をエネルギー算出時に除外する。なお、シミュレート画像 m_{nf} は次式により算出する。

$$\mathbf{m}_{nf} = \mathbf{H}_{fn}(\mathbf{z}_f) \mathbf{s}_f \quad (3)$$

ここで、 $\mathbf{s}_f = (s_{f1}, \dots, s_{fq})^T$ は超解像画像の画素値を、 $\mathbf{z}_f = (z_{f1}, \dots, z_{fq})^T$ は超解像画像 \mathbf{s}_f の各画素に対応する奥行き値を表す q 次元ベクトルである。 $\mathbf{H}_{fn}(\mathbf{z}_f)$ は奥行き値 \mathbf{z}_f を用いて、第 f フレームの超解像画像から第 n フレームの観測画像をシミュレートする変換行列を表しており、次式により算出する。

$$\mathbf{H}_{fn}(\mathbf{z}_f) = \left[\alpha_1 \mathbf{h}_1, \dots, \alpha_i \mathbf{h}_i, \dots, \alpha_p \mathbf{h}_p \right]^T \quad (4)$$

ただし、 α_i は後述する正規化要素であり、 \mathbf{h}_i は次式に示す q 次元ベクトルである。

$$\mathbf{h}_i = \left(h_{i1}, \dots, h_{ij}, \dots, h_{iq} \right)^T \quad (5)$$

ここで、 h_{ij} は超解像画像の j 番目の画素と観測画像の i 番目の画素の対応の有無を示す 0 または 1 のスカラー値であり、推定された奥行き値に基づいて、次式によって算出する。

$$h_{ij} = \begin{cases} 0; & d_n(\mathbf{p}_{fj}) \neq i \text{ または } z'_{fj} > z_{ni} + C \\ 1; & \text{otherwise} \end{cases} \quad (6)$$

ただし、 \mathbf{p}_{fj} は図 7 に示すように、超解像画像の j 番目の画素に対応するシーンの 3 次元座標を表し、 $d_n(\mathbf{p})$ は、 \mathbf{p} を第 n フレームに投影した座標に対応する画素のインデックスを表す。また、図 8 に示すように、 z'_{fj} は第 f フレームの奥行き値 z_{fj} を第 n フレームの奥行き値に変換した値を表し、 z_{ni} は対応する第 n フレームの奥行き値を表す。 C はオクルージョンの判定に用いる閾値である。

なお、式 (4) における α_i はシミュレート画像 m_{nf} の i 番目の画素に投影される超解像画像の画素数による正規化要素であり、上記の方法で決定された \mathbf{h}_i を用い

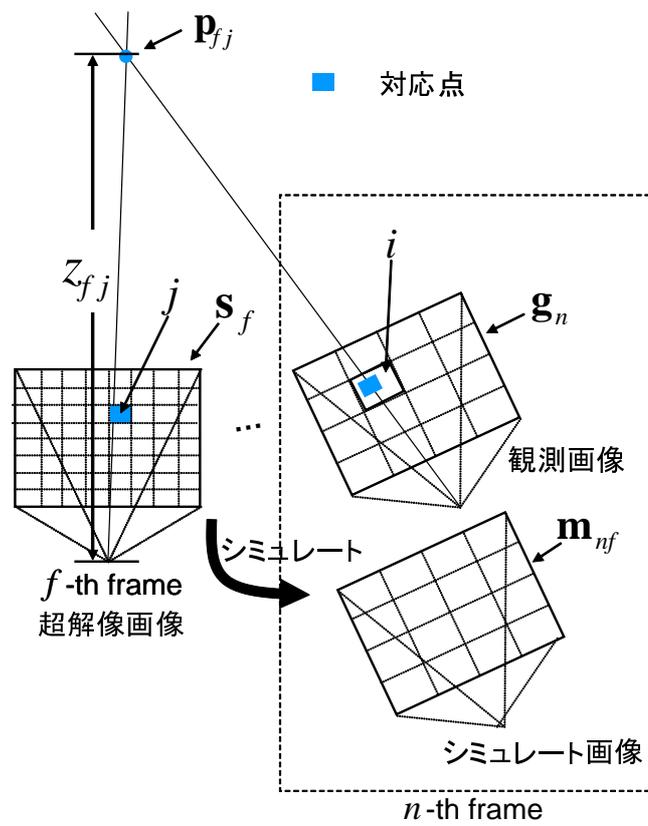


図 7 超解像画像と観測画像の画素の対応関係

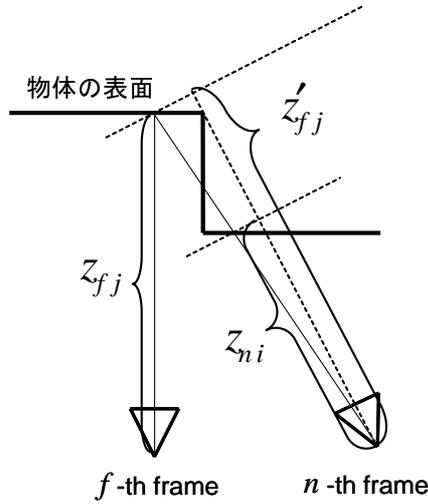


図 8 オクルージョンによる奥行き値の変化

て次式で算出する．

$$\alpha_i = \begin{cases} 0 & ; |\mathbf{h}_i| = 0 \\ \frac{1}{|\mathbf{h}_i|^2} & ; |\mathbf{h}_i| > 0 \end{cases} \quad (7)$$

エネルギー E_{Df} は，超解像対象フレームの奥行き値の滑らかさを表し，同一物体上では奥行き値の変化は滑らかであるという仮定の下，次のように定義する．

$$E_{Df} = \sum_j \left(\left(\frac{\partial^2 z_{fj}}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z_{fj}}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z_{fj}}{\partial y^2} \right)^2 \right) \quad (8)$$

ただし， $\frac{\partial}{\partial x}$ ， $\frac{\partial}{\partial y}$ は画像面上での x ， y 方向の偏微分を表す．

3.1.2 エネルギー最小化による奥行き値の最適化に基づく超解像画像の生成

超解像画像の画素値と奥行き値を変数とし，前項で定義したエネルギー関数を最小化することで，超解像画像を生成する．式(2)に示したようにエネルギー E_{If} は観測画像 g_n とシミュレート画像 m_{nf} の差分に基づき算出される．ここで， g_n は不変であるが，式(3)に示したように， m_{nf} は超解像画像の画素値 s_f と奥行き値 z_f に依存する．計算コストの問題から超解像画像の画素値と奥行き値を同時

に最適化することは難しいため，本研究では次の2つの処理をエネルギーが収束するまで繰り返すことで，エネルギー E を最小化する．

(i) 超解像対象フレームの奥行き値 z_f を固定し，超解像画像の画素値 s_f を更新

(ii) 超解像画像の画素値 s_f を固定し，超解像対象フレームの奥行き値 z_f を更新

処理 (i) では，超解像対象フレームの奥行き値 z_f を固定するため，超解像画像と観測画像の画素間の対応関係 $H_{fn}(z_f)$ は変化しない．また，奥行きの滑らかさに関するエネルギー E_{Df} も定数として扱えるため，式 (1) を最小化するには，画像の整合性に関するエネルギー E_{If} が最小となるように超解像画像の画素値を更新すればよい．ここでは，Irani らの手法 [14] と同様の方法で，次式に従い超解像画像の画素値 s_{fj} の更新を行う．

$$s_{fj} \leftarrow s_{fj} + \frac{\sum_{n=a}^b ((g_{ni} - m_{nfi})O_{ni})}{\sum_{n=a}^b O_{ni}} \quad (9)$$

処理 (ii) では，超解像画像の画素値 s_f を固定し，奥行き値 z_f の最適化を行う．本研究で用いるエネルギー E において，シミュレート画像 m_{nf} の各要素は奥行き z_f に対して不連続に変化するため， E の z による微分値を代数的に求めることは不可能である．そのため，本研究では，各奥行き値を微小範囲内で離散的に動かし，エネルギー E が最小となる値に奥行きを繰り返し更新することでエネルギーを最小化する．

3.2 新視点画像生成による時間解像度の向上

本研究では，図9に示すように，観測フレーム間に仮想的なフレームを設定し，そのフレームに対して前節で述べた空間解像度の向上処理を適用することで，補間画像を生成する．ただし，補間フレームでは，空間解像度の向上処理で利用するカメラ位置・姿勢と奥行き情報が無いため，これらの値を設定する必要がある．

新視点画像生成処理では，はじめに補間フレームに隣接する観測フレームのカメラ位置・姿勢を用いて，式 (10)，(11) に従い，補間フレームのカメラ位置・姿

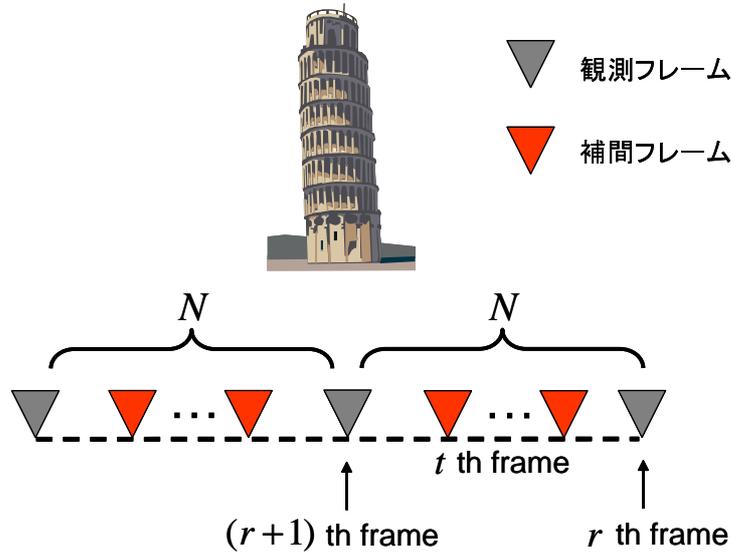


図 9 補間フレームの設定

勢を設定する．

$$\mathbf{P}'_t = \frac{(N-t)}{N}\mathbf{P}_r + \frac{t}{N}\mathbf{P}_{r+1} \quad (1 \leq t < N) \quad (10)$$

$$\mathbf{Q}'_t = \frac{(N-t)}{N}\mathbf{Q}_r + \frac{t}{N}\mathbf{Q}_{r+1} \quad (1 \leq t < N) \quad (11)$$

ここで， N は時間解像度の倍率を示しており， \mathbf{P}_r は r 番目の観測フレームを， \mathbf{P}'_t は， r 番目と $r+1$ 番目の観測フレームを N 分割したときの t 番目の補間フレームのカメラ位置を表す．同様に， \mathbf{Q}_r ， \mathbf{Q}'_t はそれぞれ観測フレームと補間フレームのカメラ姿勢を表す．次に，設定した補間フレームのカメラ位置・姿勢を用いて，観測フレームの奥行き値を変換することで補間フレームの奥行き値の初期値を設定する．ここで本研究では，補間フレームにおいて，オクルージョンの発生や観測フレームの奥行き値に誤差があることを考慮し，多数の観測フレームの奥行き値を用いて補間フレームの奥行き値を設定することで，補間フレームの奥行き値の初期値の信頼度を向上させる．具体的には，次式に従い．補間フレームの奥行き

きの初期値を設定する .

$$\mathbf{z}_t = \underset{l \in \Psi}{\text{median}}(\mathbf{M}_{lt} \mathbf{z}_l) \quad (12)$$

ここで , \mathbf{z}_t は補間フレーム t の奥行きを表し , \mathbf{M}_{lt} はカメラ位置・姿勢情報を用いて , 第 l フレームの奥行き画像を補間フレーム t へと変換する行列を表す . また , Ψ は一定範囲内の近傍フレームを表す .

4. 実験

提案手法の有効性を検証するため，仮想環境を用いたシミュレーション動画および実動画に対する超解像画像の生成を行い，画質の評価を行った．以下，4.1節でシミュレーション動画に対する超解像画像の生成実験について述べ，4.2節で実動画に対する超解像画像の生成実験について述べる．

4.1 シミュレーション動画に対する超解像画像の生成実験

シミュレーション動画に対する超解像画像の生成実験では，図 10 に示すような仮想環境において，対象物体を移動しながら撮影した動画像を入力として用いた．ここでは，観測地点のカメラ位置・姿勢として真値をそのまま用い，奥行き
の初期値には，奥行きの真値に画像上での平均投影誤差 1 画素に相当するガウスノイズを加えたものを用いた．表 1 に本実験で用いたパラメータを示す．上記の条件のもと，PC(CPU: Xeon 3.4GHz, Memory: 3GB) を用いて超解像画像を生成し，以下に示す 2 種類の実験を行った．

- (1) 空間超解像に有効な画像枚数の検証
- (2) 生成された時空間超解像画像の画質の定量的評価

ただし，各実験では正解画像を画素平均によって縦横それぞれ $\frac{1}{2}$ 倍したものを入力画像として用いた．

表 1 実験で用いたパラメータ

入力画像の空間解像度	320 × 240[画素]
出力画像の空間解像度	640 × 480[画素]
正解画像の空間解像度	640 × 480[画素]
重み w	100
閾値 C	1[m]

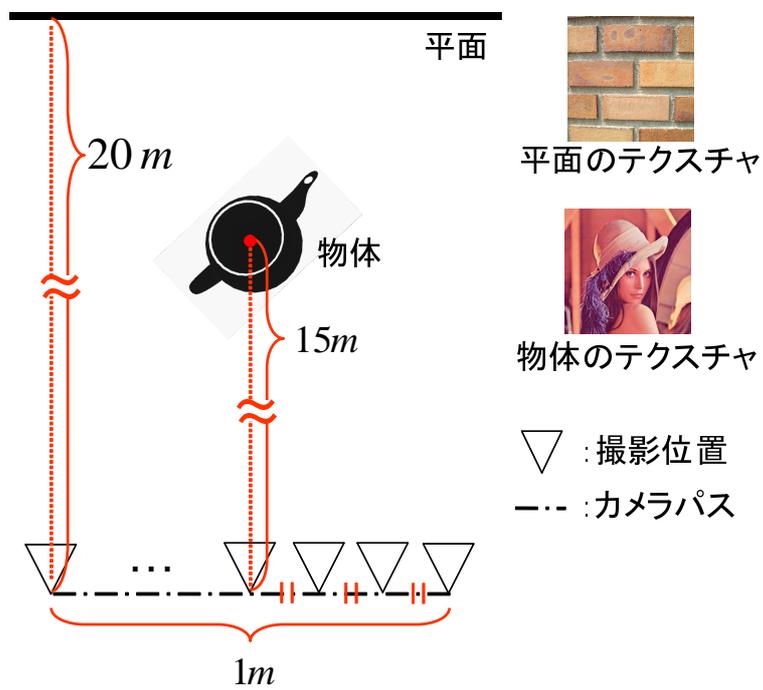


図 10 実験におけるカメラの動きと物体の配置

4.1.1 空間超解像に有効な画像枚数の検証

超解像処理では、生成される超解像画像の画質の向上効果は利用する画像枚数に依存し、一般に利用できる画像の枚数が多いほど画質が向上する。本実験では、カメラの動きを等速直線運動とし、 $1m$ の距離を任意のフレーム数で等間隔に撮影することで入力となる動画像を取得し、これを用いて生成した空間超解像画像の真値に対するPSNR (Peak Signal-to Noise Ratio) を算出することで、画質の向上効果を検証した。ここでPSNRは、生成した超解像画像が正解画像に対してどの程度ノイズが混入しているかを示す指標であり、次式によって算出する。

$$PSNR[dB] = 10 \log_{10} \frac{255^2}{S} \quad (13)$$

ただし、

$$S = \frac{1}{q} \sum_{j=1}^q (C_j - S_j)^2 \quad (14)$$

ここで、 C と S は、それぞれ正解画像と生成画像の画素値を表す q 次元ベクトルであり、 C_j と S_j は、それぞれ正解画像と生成画像の j 番目の画素値を表す。

図11に、超解像対象フレームの入力画像と正解画像および超解像処理に利用する画像枚数を変化させた場合の生成画像を、図12に、超解像処理に利用する画像枚数と生成結果のPSNRの関係を示す。図11、図12より、提案手法では、利用する画像枚数が多いほど生成される超解像画像の画質が向上していることが分かる。また、71枚以上ではPSNRの伸びが頭打ちになっていることが確認でき、これ以上多くのフレームを用いても、画質の向上が見込めないことが分かる。



(a) 対象フレームの入力画像



(b) 正解画像



(c) 生成画像 (入力 11 枚)



(d) 生成画像 (入力 31 枚)



(e) 生成画像 (入力 51 枚)



(f) 生成画像 (入力 81 枚)

図 11 入力画像，正解画像，生成画像の比較

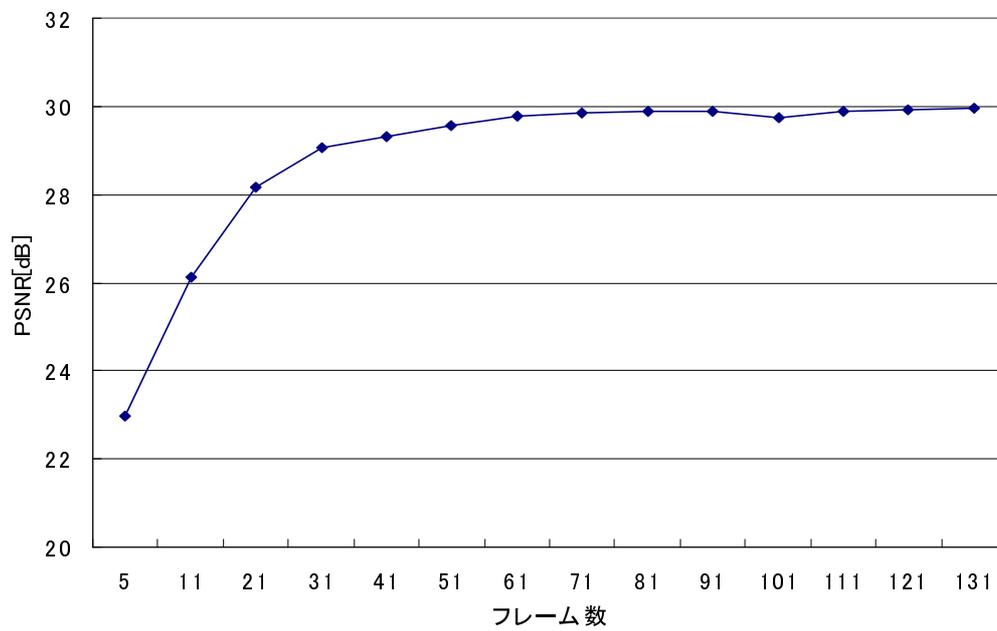


図 12 超解像に利用する画像枚数と PSNR の関係

4.1.2 生成された時空間超解像画像の画質の定量的評価

提案手法の有効性を確認するため、真値に対する PSNR を用いて以下に示す 4 つの動画像を比較した。なお、本実験では入力動画像の時間解像度は 31 フレームとし、空間解像度の向上処理では全ての観測フレームを利用した。

- (A) 入力動画像に対して同一の空間解像度かつ 2 倍の時間解像度で観測を行い、バイリニア補間によって空間解像度を向上させた動画像
- (B) 入力動画像に対して縦横それぞれ 2 倍の空間解像度かつ同一の時間解像度で観測を行い、直前のフレームの画像を補間フレームに挿入することで、時間解像度を向上させた動画像
- (C) 真の奥行き値を用い、提案手法による奥行き最適化なしに時空間解像度の向上処理を行って生成した動画像
- (D) 奥行き初期値にノイズを付加したものを用い、提案手法によって時空間解像度を向上させた動画像

図 13 にそれぞれの手法で生成した動画像の真値に対する PSNR を示す。また、表 2 にそれぞれの手法の平均 PSNR を示す。なお、動画像 (B) における観測フレームは正解画像と同一であるため、補間フレームにおける PSNR のみを示している。また、動画像 (C) と (D) について、観測フレームにおける生成画像 ((C-1), (D-1)) と補間フレームにおける生成画像 ((C-2), (D-2)) の PSNR をそれぞれ示している。図 13 および表 2 より、動画像 (A) と比較して、提案手法で生成した超解像画像はどのフレームにおいても PSNR が高く、画質が向上していることが分かる。また、動画像 (B) と比較して、提案手法で生成した超解像画像は補間フレームにおける PSNR の落ち込みが小さいことが分かる。ただし、提案手法では、開始フレームと最終フレーム周辺において画質の向上効果が小さいことも確認できる。これらは対象フレームの観測位置から空間的に近い位置で撮影されたフレームが少ないことが原因であると考えられる。また、動画像 (C) と比較して、奥行き初期値にノイズを付加したものをを用いた場合でも、フレームによっては奥行き値に真値を用いた場合と近い値の PSNR が得られていることが分かった。図 14

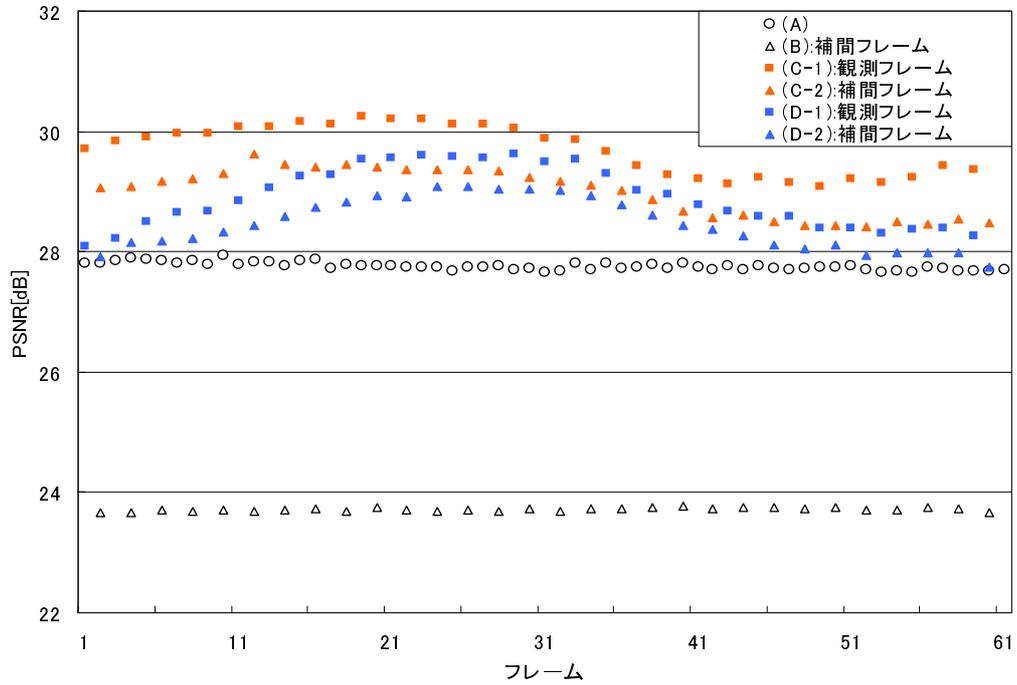


図 13 PSNR の比較

表 2 それぞれの手法で生成した動画像の平均 PSNR

対象画像		PSNR[dB]
(A)		27.76
(B):補間フレーム		23.21
(C)	(C-1):観測フレーム	29.71
	(C-2):補間フレーム	28.99
	全フレーム	29.35
(D)	(D-1):観測フレーム	28.91
	(D-2):補間フレーム	28.46
	全フレーム	28.68

に示すように，動画像 (C) と動画像 (D) の 41 フレーム目の生成画像では，大きな差は感じられない．

また，奥行き最適化効果を確認するため，図 11(d) に示した画像を生成した場合に用いられた，奥行きの初期値と最終的な奥行きの値を図 15 に示す．同図より，背景の平面と物体の表面上において不連続であった奥行きが滑らかに更新されていることが確認できる．なお，画像上で隣接していた物体の胴や蓋と背景の平面の境界部分も滑らかに更新されており，図 16 に示すように，生成された画像上の輪郭付近に，正解画像には本来出現しない画素値が表れ，動画像として再生した場合には，物体の輪郭部分に多少のちらつきが見られた．

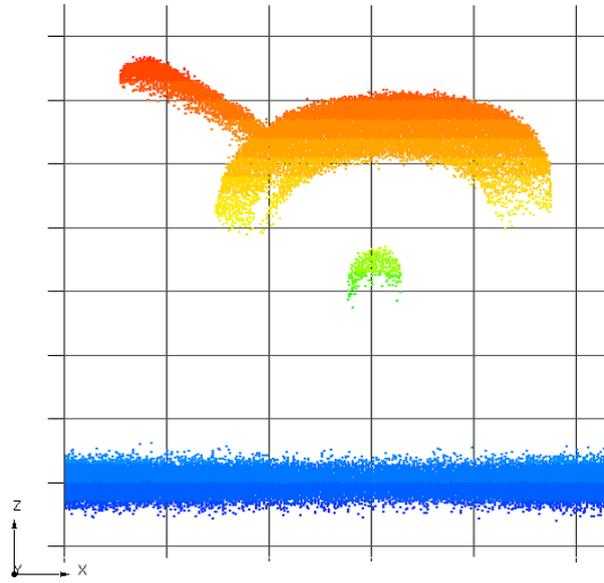


(a) 動画像 (C) の生成画像

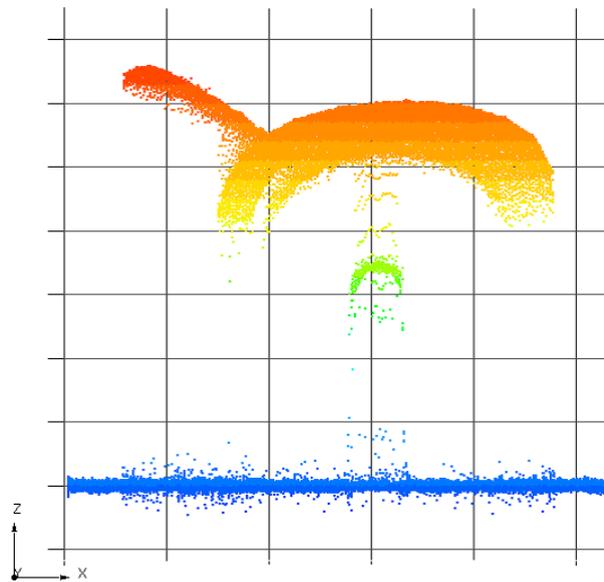


(b) 動画像 (D) の生成画像

図 14 超解像画像の比較 (41 フレーム目)

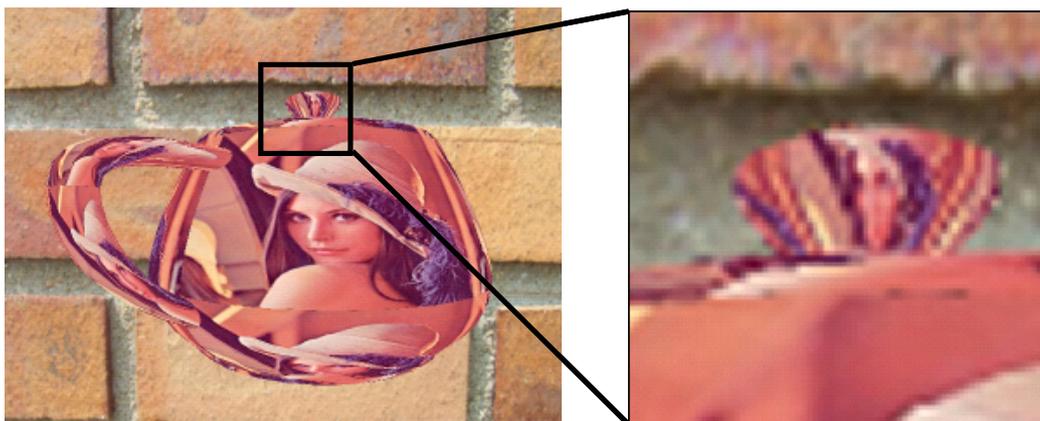


(a) 初期値 (XZ 平面)

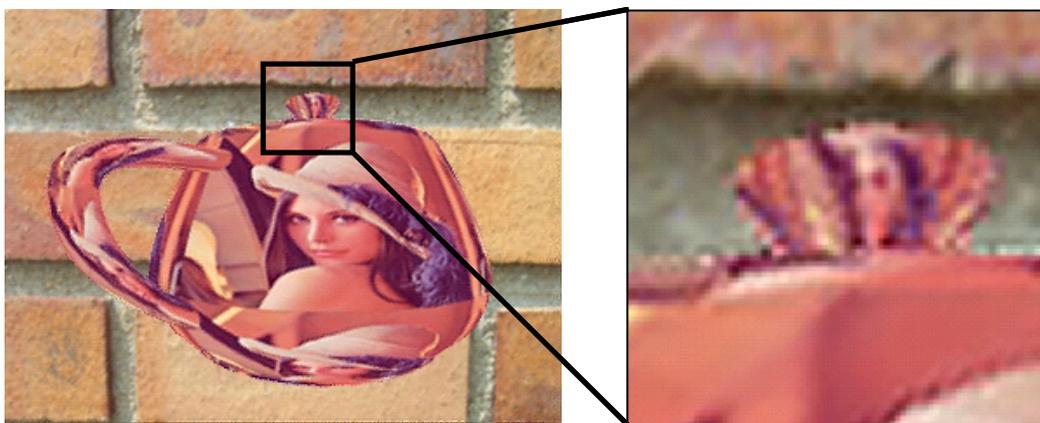


(b) 最終的な奥行き値 (XZ 平面)

図 15 奥行きの変化



(a) 正解画像



(b) 動画像 (D) の生成画像

図 16 奥行きのある滑らかさの影響

4.2 実動画に対する超解像画像の生成実験

本節では，実動画に対して超解像画像の生成実験を行い，生成された画像の画質の評価と考察を行う．以下では，まず実験条件を示し，次に生成された超解像画像の画質について考察する．

4.2.1 実験条件

本実験では，Sony HDR-FX1 (1920 × 1080 画素) にワイドコンバージョンレンズ VCL-HG0872 (垂直画角 42 度，水平画角 69 度) を取り付けたものを使用し，上空から本学を撮影した動画像を用いた．ここでは，撮影された動画像を画素平均によって 320 × 240 画素に縮小したものを入力動画像として用いた．図 17 に入力画像の一部を示す．各画像取得時のカメラ位置・姿勢については自然特徴点の追跡による structure from motion 法 [29] を用いて入力動画像から推定した値を用いた．また，奥行き初期値は，特徴点の数え上げによるマルチベースラインステレオ法 [30] によって推定した奥行き値を用いた．ただし，特徴点の数え上げによるマルチベースラインステレオ法では画像特徴点以外の画素に対する奥行き値は算出されないため，最近傍の奥行き値を用いて全ての画素の奥行き値を補間したものを初期値として用いた．なお，本実験では観測画像のみを対象として超解像画像の生成を行い，出力画像の解像度は 640 × 480 画素とした．

4.2.2 超解像画像の生成結果と考察

図 18 に，入力動画像の第 16 フレームに対して用いた奥行き初期値と最適化後の奥行き値を示す．また，図 19 に前後 5 フレームを用いて生成した第 16 フレームの超解像画像と入力画像の拡大画像を示す．図 19 から，入力画像と生成画像を比較して，画質が向上している箇所と画質が劣化している箇所を確認できる．これは，初期値として与えられた奥行き値の精度が低いことが原因と考えられる．本実験では，特徴点の数え上げによるマルチベースラインステレオ法によって求められなかった奥行き値については最近傍補間によって値を与えたため，図 19 の青枠で囲まれた箇所のように，奥行き値が様々に変化し，複雑なテクスチャを含



(a) 第 1 フレーム



(b) 第 5 フレーム



(c) 第 11 フレーム



(d) 第 16 フレーム



(e) 第 21 フレーム



(f) 第 31 フレーム

図 17 入力として用いた空撮動画像の一部

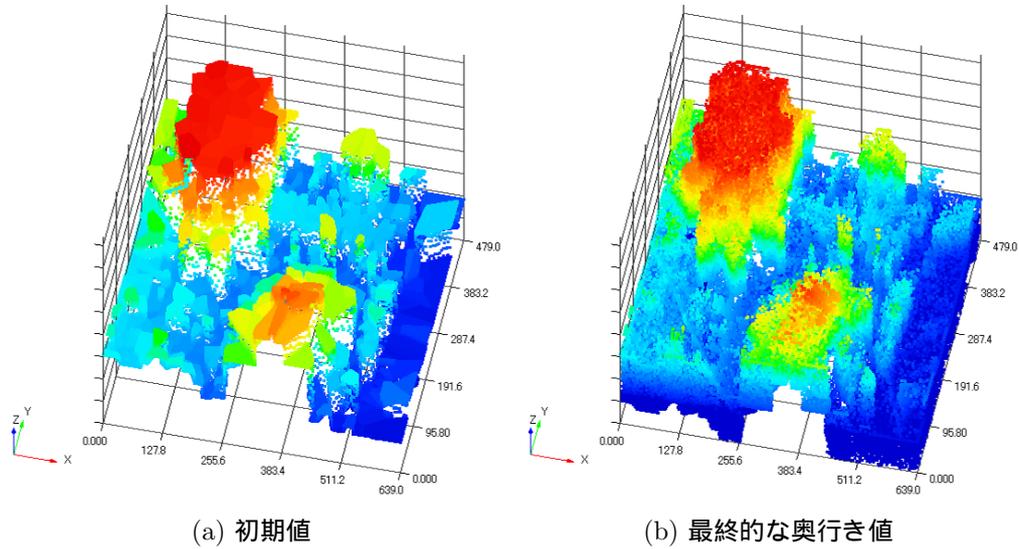
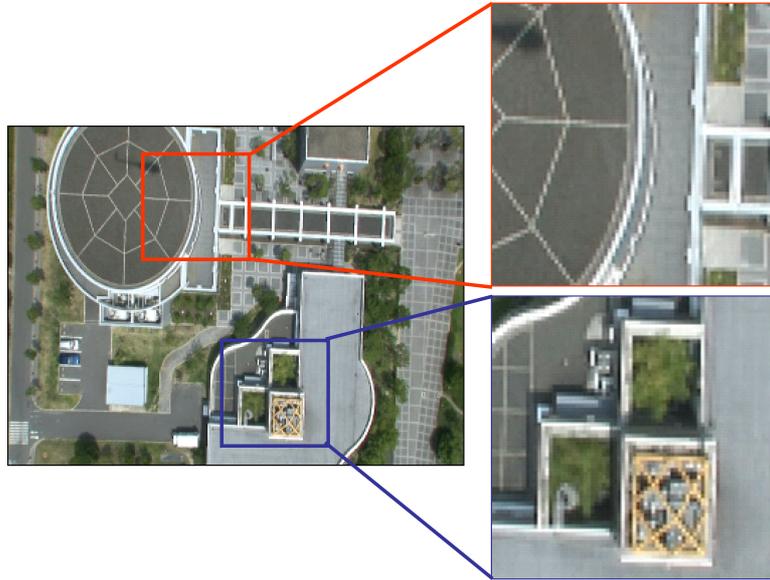
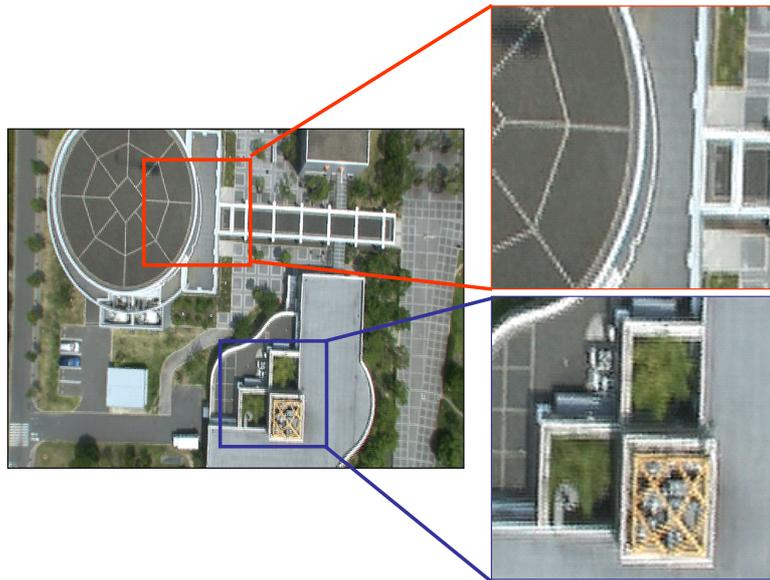


図 18 超解像対象フレーム（第 16 フレーム）の奥行き値

む箇所については良好な結果が得られなかった．しかし，同図の赤枠で囲まれた箇所のように，補間された奥行き値と真の奥行き値が大きく異ならないような箇所では，画質の向上効果を確認することができた．また，図 20 に利用するフレーム数を変化させて生成した場合の超解像画像を示す．図 20 から，超解像に利用する画像枚数を増やすほど生成される超解像画像にぼけが発生していることが分かる．この原因として，利用する画像枚数が増えるほど空間的に離れた地点から撮影された画像を利用する割合が増えるため，奥行きやカメラ位置・姿勢に誤差が含まれている場合，誤差の影響を大きく受けることが考えられる．提案手法では複数の入力画像中で対応する画素値が一致するように超解像画像の画素値を更新するため，位置合わせの精度が低い場合にはぼけたような画像が生成される．今後は，推定されたカメラパラメータについても最適化などを行い位置合わせの精度向上をはかることで，生成される超解像画像の画質を向上させる必要がある．



(a) 入力画像



(b) 生成画像（入力 5 枚）

図 19 入力画像と生成画像の比較



(a) 対象フレームの入力画像



(b) 生成画像 (入力 5 枚)



(c) 生成画像 (入力 7 枚)



(d) 生成画像 (入力 11 枚)



(e) 生成画像 (入力 21 枚)



(f) 生成画像 (入力 31 枚)

図 20 利用する画像枚数を変化させた時の生成画像

5. まとめ

本論文では、動画像の各画像取得時のカメラパラメータを既知とした上で、各画素の奥行き値をパラメータとして多数の画像上での対応点を同時に決定することで時空間超解像画像を生成する手法を提案した。提案手法では、複数の画像に対する対応点を奥行きによる1次元探索によって決定することで対応点探索のパラメータの自由度を減らし、平面仮定などの撮影対象に関する強い制約を課すことなく、各画素を高精度に対応付けし、空間解像度を向上させる。また、推定される奥行きに基づいてオクルージョンを考慮した対応点の探索を行うことで、従来手法では正しい対応が得られないシーンにおいても時間解像度を向上させる。シミュレーション実験において、生成した超解像画像を観測画像と比較し、提案手法による画質の向上効果を確認した。また、正解画像に対するPSNRを算出することで、超解像画像の画質の定量的評価を行い、バイリニア補間による空間解像度の向上処理などと比較して、提案手法の解像度の向上効果が高いことを確認した。しかし、実動画における超解像画像の生成実験では、奥行き初期値によっては良好な超解像画像を得ることが困難であった。今後、推定されたカメラパラメータについても最適化を行うことで位置合わせの精度向上をはかり、生成される超解像画像の画質を向上させる必要がある。

謝辞

本研究を行うにあたり，多大なる御指導をいただきました視覚情報メディア講座横矢直和教授に厚く御礼を申し上げます．また，本研究の遂行にあたり，有益な御助言，御鞭撻をいただいたロボティクス講座小笠原司教授に感謝致します．そして，本研究の全過程を通して御指導いただいた視覚情報メディア講座山澤一誠准教授に深く感謝致します．さらに，本研究の遂行に的確な御助言をいただいた視覚情報メディア講座神原誠之助教に深く御礼申し上げます．研究活動の全過程を通して多くの御助言，御指導賜りました視覚情報メディア講座佐藤智和助教に心より感謝致します．特に佐藤智和助教には本研究のテーマ設定から本論文の執筆，その他の発表論文の添削，発表練習に至るまで細やかな御指導を頂きました．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講座事務補佐員高橋美央女史に心より感謝申し上げます．そして，本研究の全過程を通して無数の御助言，御指導を賜った視覚情報メディア講座の河合紀彦氏に厚く御礼申し上げます．

最後に，研究活動において常に御支援していただいた視覚情報メディア講座の諸氏に深く感謝致します．

参考文献

- [1] 池田聖, 佐藤智和, 横矢直和: “全方位型マルチカメラシステムを用いた高解像度な全天球パノラマ動画の生成とテレプレゼンスへの応用”, 日本バーチャルリアリティ学会論文誌, Vol. 8, No. 4, pp. 443–450, 2003.
- [2] 和田浩明, 池田聖, 佐藤智和, 横矢直和: “複数経路上を歩行移動可能なテレプレゼンスシステムにおける違和感の低減に関する検討”, 日本バーチャルリアリティ学会 サイバースペースと仮想都市研究会研究報告, Vol. 11, No. 1, pp. 29–34, 2006.
- [3] W. T. Freeman, T. R. Jones and E. C. Pasztor: “Example-based super-resolution,” IEEE Computer Graphics and Applications, Vol. 22, pp. 56–65, 2002.
- [4] 溝内義章, 末松信朗, 林朗: “多様体学習に基づく単フレーム超解像”, 情報科学技術フォーラム 一般講演論文集, pp. 65–66, 2006.
- [5] 上村健二, 津村徳道, 中口俊哉, 菅谷隆, 三宅洋一: “Texton 置換に基づく画像の高解像度化手法”, 映像情報メディア学会誌, Vol. 60, No. 10, pp. 1655–1657, 2006.
- [6] I. Begin and F. P. Ferrie: “Blind super-resolution using a learning-based approach,” Proc. Int. Conf. on Pattern Recognition, Vol. 2, pp. 85–89, 2004.
- [7] P. Bhat, C. L. Zitnick, N. Snavely, A. Agarwala, M. Agrawala, M. Cohen, B. Curless and S. B. Kang: “Using photographs to enhance videos of a static scene,” Eurographics Symposium on Rendering, pp. 327–338, 2007.
- [8] 川合康裕, 天野敏之: “固有空間を用いた bplp による super resolution”, 電子情報通信学会総合大会講演論文集, Vol. 2004 年 情報・システム, No. 2, p. 91, 2004.

- [9] M. C. Hong, T. Stathaki and A. K. Katsaggelos: “Iterative regularized image restoration using local constraints,” Proc. IEEE Workshop on Nonlinear Signal and Image Processing, pp. 145–148, 1997.
- [10] M. C. Chiang and T. E. Boult: “Efficient super-resolution via image warping,” Image and Vision Computing, pp. 761–771, 2000.
- [11] M. Ben-Ezra, A. Zomet and S. K. Nayar: “Jitter camera: High resolution video from a low resolution detector,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 135–142, 2004.
- [12] Z. Jiang, T. Wong and H. Bao: “Practical super-resolution from dynamic video sequences,” Proc IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 549–554, 2003.
- [13] W. Y. Zhao: “Super-resolving compressed video with large artifacts,” Proc. Int. Conf. on Pattern Recognition, Vol. 1, pp. 516–519, 2004.
- [14] M. Irani and S. Peleg: “Improving resolution by image registration,” Graphical Models and Image Processing, Vol. 53, No. 3, pp. 231–239, 1991.
- [15] R. R. Schulz and R. L. Stevenson: “Extraction of high-resolution frames from video sequences,” IEEE Trans. Image Processing, Vol. 5, No. 6, pp. 996–1011, 1996.
- [16] 後藤知将, 奥富正敏: “画像復元とレジストレーションの同時最適化の実験的検証”, 電子情報通信学会論文誌 (D), Vol. J90-D, No. 6, pp. 1632–1635, 2007.
- [17] 田中正行, 奥富正敏: “再構成型超解像処理の高速化アルゴリズムとその精度評価”, 電子情報通信学会論文誌 (D-II), Vol. J88-D-II, pp. 2200–2209, 2005.
- [18] T. Kuo, J. W. Kim and J. Kuo: “Motion-compensated frame interpolation scheme for h.263 codec,” IEEE Int. Symposium on Circuits and Systems, Vol. 4, pp. 491–494, 1999.

- [19] J. W. Kim, Y. Kim, H. Song, T. Kuo, Y. J. Chung and J. Kuo: “Tcp-friendly internet video streaming employing variable frame-rate encoding and interpolation,” *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 10, No. 7, pp. 1164–1177, 2000.
- [20] S. E. Chen and L. William: “View interpolation for image synthesis,” *Proc. Int. Conf. on Computer Graphics and Interactive Techniques*, Vol. 1, pp. 279–288, 1993.
- [21] 大脇一泰, 三島直, 伊藤剛: “陰面領域を考慮したフレーム補間技術 (framebooster®)”, *電子情報通信学会総合大会講演論文集*, Vol. 2006 年情報・システム, No. 2, p. 96, 2006.
- [22] J. Shimamura, H. Takemura, N. Yokoya and K. Yamazawa: “Construction and presentation of a virtual environment using panoramic stereo images of a real scene and computer graphics models,” *Proc. Int. Conf. on Pattern Recognition*, pp. 463–467, 2000.
- [23] 石川智也, 山澤一誠, 横矢直和: “複数の全方位動画像からの自由視点画像生成”, *画像の認識・理解シンポジウム講演論文集*, pp. 1396–1403, 2005.
- [24] M. Levoy, P. Hanrahan, N. Yokoya and K. Yamazawa: “Light field rendering,” *Proc. of SIGGRAPH*, pp. 31–42, 1996.
- [25] S. J. Gortler, R. Grzeszczuk and M. F. Cohen: “The lumigraph,” *Proc. of SIGGRAPH*, pp. 43–54, 1996.
- [26] E. Shechtman, Y. Caspi and M. Irani: “Space-time super-resolution,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, pp. 531–545, 2005.
- [27] 松延徹, 長原一, 岩井儀雄, 谷内田正彦, 鈴木俊哉: “モーフィングによる高解像度高フレームレート動画像の生成”, *電子情報通信学会論文誌 (D)*, Vol. J90-D, No. 4, pp. 1073–1084, 2007.

- [28] 今川太郎, 吾妻健夫, 佐藤智和, 横矢直和: “解像度と露出時間の異なる画像を用いた高時空間解像度画像の生成”, 電子情報通信学会 技術研究報告, PRMU-278, pp. 137–142, 2007.
- [29] 佐藤智和, 神原誠之, 横矢直和, 竹村治雄: “マーカと自然特徴点の追跡による動画像からのカメラ移動パラメータの復元”, 電子情報通信学会論文誌 (D-II), Vol. J86-D-II, No. 10, pp. 1431–1440, 2003.
- [30] 佐藤智和, 横矢直和: “画像特徴点の数え上げによるマルチベースラインステレオ法”, 画像の認識・理解シンポジウム講演論文集, pp. 189–196, 2005.