

視点依存型アプローチによる超解像自由視点画像合成

高橋 桂太[†] 苗村 健[†]

[†] 東京大学 IRT 研究機構 〒 113-8656 東京都文京区本郷 7-3-1

E-mail: †keita.takahashi@ieee.org

あらまし 本稿では、多視点画像をもとに任意の視点から見た画像を合成する自由視点画像合成において、超解像による高解像度化を組み込んだ新しいアルゴリズムを提案する。本研究では、事前に被写体のモデルを構築することなく、ランタイムに合成対象視点の座標系において必要な処理のみを行う、視点依存型のアプローチを取る。提案手法では、まず合成対象視点から見た画素単位のデプスマップを推定し、そのデプス情報に基づいて再構成型超解像の枠組みを適用することで最終的な合成画像を得る。デプス推定には、セミグローバル・ステレオ法を基にした高精度なアルゴリズムを用いる。超解像合成では、デプスの信頼性を用いて画素ごとに適応的に正則化の強度を調整することで、ノイズを抑制しつつ解像感の高い画像合成を実現する。実写画像による実験により、提案手法の有効性を示す。

キーワード 自由視点画像, 超解像, セミグローバル・ステレオ法, 視点依存型

1. はじめに

自由視点画像合成とは、複数の視点で撮影された多視点画像をもとに、任意の新しい視点から見た画像を合成する技術である。この技術は、高度なリアリティを持つ三次元的な映像効果を可能にするものであり、次世代の放送・通信分野への応用が期待されている [1]。

自由視点画像合成は、一般に、(1) 被写体の形状推定、(2) 形状への着色、の 2 段階に分けて考えられる。(1)(2)に、どんなアルゴリズムを用い、どれほどコストをかけられるかは、主に想定しているアプリケーションに依存する。本研究では、単一のオブジェクトではなく被写体空間全体を処理の対象とする（全画面処理する）ケースを扱う。したがって、入力視点は、特定のオブジェクトを囲む形ではなく、平面格子に近い形に配置されるものとし、被写体空間側にブルーバック等の特殊な設定を想定しない。また、将来的に動画像通信等への応用を視野に入れるため、視点依存型、すなわち、合成対象視点に必要な処理のみをランタイムに行うアプローチ [2] を取る。言い換えると、事前に完全な形の三次元形状モデルと形状表面の光学モデルを構築しておくのではなく、ランタイムに合成対象視点の座標系で直接必要な部分のみに (1)(2) の処理を行う。

提案手法の特徴は、(2) の着色処理に再構成型超解像 [3] の枠組みを適用し、合成画像の高解像度化を目指す点にある。従来の自由視点合成手法では、入力画像の重み付け和（混色）によって着色を行うのが一般的であった。重み付け和は、画像をぼかす性質があるため、形状の不完全さ等に起因するノイズを目立たなくする効果がある一方で、合成画像の解像感を大きく低下させるという問題がある。この問題を解決する自然な手段は、混色と再構成型超解像を併用することである。

一般的な意味での超解像 [3] は、複数枚の低解像度画像のうちの一つを基準画像とし、基準画像を高解像度化するものであり、新しい視点から見た画像の合成を意図したものではない。超解像を自由視点画像合成に導入した例としては、入力多視点画像を超解像する手法 [4] や、被写体のテクスチャマップを超解像合成する手法 [5] があるが、いずれも単一オブジェクトの三次元モデルを生成するものであり、提案手法とは異なる。一方、文献 [6] の超解像 plane sweeping は、視点依存型アプローチで全画面処理をするが、デプス推定が原始的（ローカルなブロック関連法に等価）であり、オクルージョン等も考慮されないため、実シーンへの適用が困難である。

本研究は、文献 [6] を実用的な形態に発展させたものと位置づけられる。提案手法では、セミグローバル・ステレオ法 [9], [10] を基にした高精度なデプス推定と、オクルージョン処理、デプスの信頼性に基づく正則化を含めた超解像処理を導入する。提案する正則化手法は、混色合成と超解像合成を融合するものであり、ノイズを抑制しつつ解像感を高める画像合成を可能にする。

1.1 提案手法の概要

本稿で用いる設定を図 1 に示す。入力視点はほぼ同一平面上に配置されるとし、入力多視点画像を $I_{(m)}$ ($m = 1, \dots, M$) と表記する。カメラパラメータは既知とする。 z は入力視点平面からの距離を表す。提案手法では新たな視点（合成対象視点） t から見た画像を合成する。入力画像と同じ解像度の混色による合成画像を $I_{(t)}$ 、超解像による高解像度の合成画像を $I_{(t)}^{SR}$ と表記する。

提案手法は、合成対象視点 t の座標系において入力多視点画像の位置合わせを行い、再構成型超解像の枠組みを当てはめることで、高解像度の自由視点画像を得るのである。多視点画像の位置合わせは、デプス推定に等

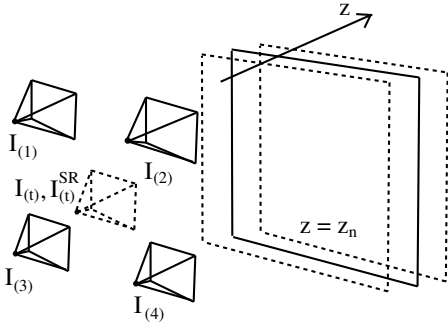


図1 基本設定.

価である。特に、合成対象視点の座標系において画素単位のデプスが分かれば、所望の位置合わせに必要な情報が得られる。そこで、提案手法では最初に、合成対象視点から見たデプスマップを求める。デプスマップを求める解像度は入力画像と同一とするが、 $I_{(t)}^{SR}$ のサイズに拡大して超解像に用いる。さらに、超解像においては、デプス推定の過程で得られた信頼度情報に基づいて正則化の強度を画素ごとに変化させることで、混色合成と超解像合成を適応的に組み合わせた高品質な画像合成を実現する。以下、デプス推定については第2章で、超解像合成については第3章で述べる。

2. デプス推定

2.1 では、本研究で用いるセミグローバルステレオ法 [9], [10] について述べる。この手法は、ステレオペア画像のうち一方（あるいは多視点画像のうちの一つ）を基準画像とし、その基準画像の座標系においてデプスを求める、通常ステレオ処理を対象としている。

2.2 では、セミグローバル法を本研究の設定に適用し、合成対象視点の座標系においてデプスマップを推定する提案手法を示す。提案手法と通常ステレオ処理とは、デプスマップを求める座標系が異なるため、デプスの量子化やマッチングコストの定義等に違いが生じる。提案手法では、さらに、二次関数の当てはめ [11] によってデプス値を高精度化する。ここで推定するデプスマップの解像度は、入力多視点画像と同一とする。

2.1 セミグローバル・ステレオ法

ステレオ法 [7] では、基準画像と参照画像の画素ごとの対応を見つけることで、各画素のデプスを求める。一般的には、ステレオ対応の妥当性を表す項とデプスの平滑性を表す項を加算したコスト関数を定義し、コスト関数の最小値に対応する最適解としてデプスマップを得る。最適化には、belief propagation や multi-label graph-cut がよく用いられるが、これらは基本的に反復解法であり、収束にかなりの計算時間を要する。これに対して、セミグローバル法 [9], [10] は、反復計算を要さず低オーダで最適解を得るものであり、精度と速度を両立できる優れたアルゴリズムである。

セミグローバル法のコスト関数は以下の通りである。

$$E_{sm}(D) = \sum_{\vec{p}} \left\{ C(\vec{p}, D(\vec{p})) + \lambda_1 \sum_{\vec{q} \in N_{\vec{p}}} T[|D(\vec{p}) - D(\vec{q})| = 1] + \lambda_2 \sum_{\vec{q} \in N_{\vec{p}}} T[|D(\vec{p}) - D(\vec{q})| > 1] \right\} \quad (1)$$

ここで、 D は基準画像における画素単位のデプスマップ、 \vec{p} , \vec{q} は離散化された画素位置を表す二次元ベクトルである。画素 \vec{p} に対するデプス値 $D(\vec{p})$ は、物理的なデプスそのものではなく、視差（基準画像と参照画像との対応画素の位置ずれ）を表し、整数単位で定義される。

第1項は、画素 \vec{p} が $D(\vec{p})$ を取る時のコストの合計である。この項は、ステレオペア間での画素対応の妥当性を評価するものであり、マッチングコストと呼ぶことにする。マッチングコストとして、文献 [9], [10] では画素ごとに定義される相互情報量 [12] が用いられたが、BT コスト [8] や、画素単位ではなくブロック単位のマッチングも取り入れられている [15], [16]。ブロック単位のマッチングは、細部構造や物体境界の扱いに難があり、foreground fattening effect などを生じるが、対応の曖昧性 (ambiguity) を減少させて結果を安定化する効果があるため、現実的に好まれる選択肢である。

第2項および第3項の $N_{\vec{p}}$ は、画素 \vec{p} に隣接する8近傍の画素の集合を表している。 $T[*]$ は、 $[]$ の内部が真なら1を、偽なら0を返す。これらの項により、隣接画素同士のデプスの変化が1ステップの場合には λ_1 、2ステップ以上の場合には λ_2 のコストが加算される。ただし、 $\lambda_2 \geq \lambda_1 > 0$ である。デプスの不連続部分は画像のエッジと一致しやすいため、 λ_2 の値は対応する画素ペアの輝度差に反比例するように設定される。

(1) 式のコストに対する準最適解を求める手順は、Dynamic Programming (DP) に類似する。まず、画面上をある方向 \vec{r} に沿って走査しながら、次のルールにしたがってコストを加算してゆき、各画素 \vec{p} の各視差 n に対応する累積コスト $L_{\vec{r}}(\vec{p}, n)$ を得る。

$$L_{\vec{r}}(\vec{p}, n) = C(\vec{p}, n) - \min_k L_{\vec{r}}(\vec{p} - \vec{r}, k) + \min\{L_{\vec{r}}(\vec{p} - \vec{r}, n), L_{\vec{r}}(\vec{p} - \vec{r}, n-1) + \lambda_1, L_{\vec{r}}(\vec{p} - \vec{r}, n+1) + \lambda_1, \min_k L_{\vec{r}}(\vec{p} - \vec{r}, k) + \lambda_2\} \quad (2)$$

ここで、 $\vec{p} - \vec{r}$ は、方向 \vec{r} に沿うパス上で、画素 \vec{p} の一つ前に位置する画素である。次に、全方位（8または16方向に量子化）のパスの累積コストを加算する。

$$S(\vec{p}, n) = \sum_{\vec{r}} L_{\vec{r}}(\vec{p}, n) \quad (3)$$

最後に、画素ごとに、 $S(\vec{p}, n)$ を最小にする視差を選択することで、(1) 式に対して準最適なデプスマップを得る。

$$D(\vec{p}) = \arg \min_n S(\vec{p}, n) \quad (4)$$

(3)(4) 式のように、全方向のパスを同時に考慮することにより、従来の DP ベースの手法 [8], [13] で問題だったライン間の不整合 (streaking effects) を回避できる。

2.2 合成対象視点におけるデプス推定

2.2.1 デプスの量子化

通常のステレオ処理 [7] では、視差、すなわち基準画像と参照画像の位置ずれ量によってデプスを表現する。提案手法でも同様に、視差空間 (デプスの逆数) を等分割するように離散化された N 階調のデプス $\{z_1, z_2, \dots, z_N\}$ の中から、合成対象視点画像の各画素に対して最も適切なデプスを求める。デプスの最大値、最小値を z_{\max} , z_{\min} とすると、 z_n は、

$$\frac{1}{z_n} = \frac{1}{z_{\max}} + \frac{n-1/2}{N} \left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \right) \quad (5)$$

と定められる。このような視差空間を等分するデプス量子化の根拠は、文献 [14] にも見られる。

2.2.2 セミグローバル法によるデプス推定

コスト関数は (1) 式と同様である。ただし、提案手法の D は、合成対象視点から見たデプスマップである。 $D(\vec{p})$ は物理的なデプスそのものではなく、デプスのインデックス $\{1, \dots, N\}$ を表す。物理的なデプスとの対応は (5) 式で与えられるが、特に画素 \vec{p} における物理的なデプスを表す場合には $z_{D(\vec{p})}$ と表記する。(1) 式の第 1 項の定義は、提案手法と通常のステレオ法とで大きく異なるため、後ほど詳述する。第 3 項の係数 λ_2 は定数とする。なぜなら、合成対象視点から見た画像が存在しない (その画像を合成することが提案手法の目的である) ため、エッジを用いた定式化が直接には不可能だからである。

最適化では、パスの方向を 8 とし、(2)–(4) 式を用いる。結果として、各画素がデプスのインデックス $\{1, \dots, N\}$ を取る離散値デプスマップ D が得られる。文献 [10] では、 D から孤立ノイズ等を除去する後処理を施すが、本稿の段階では後処理は省略している。

2.2.3 マッチングコストの定義

複数の入力画像 $I_{(m)}$ ($m = 1, \dots, M$) の対応を同時に評価するため、(1) 式の第一項を以下のように定義する。

$$C(\vec{p}, D(\vec{p})) = \frac{1}{Z} \sum_{\vec{q} \in B_{\vec{p}}} \left\{ \sum_{m \neq m'} C_{(m, m')}(\vec{q}, D(\vec{q})) \right\} \quad (6)$$

ここで、 $C_{(m, m')}$ は、入力画像 $I_{(m)}$, $I_{(m')}$ のペアに対して計算されたマッチングコストであり、これをすべてのペアについて平均し、さらに画素 \vec{p} を中心とするブロック $B_{\vec{p}}$ 内で平均を取ったものを $C(\vec{p}, D(\vec{p}))$ とする。したがって、ブロックマッチングに相当する処理をしていることになる。 Z はブロックサイズと入力画像のペア数の積で決まる正規化定数である。

$C_{(m, m')}$ の具体的な形式は以下の通りである。

$$C_{(m, m')}(\vec{p}, D(\vec{p})) = \text{diff}(I_{(m)}(\vec{p}_{(m)}), I_{(m')}(\vec{p}_{(m')})) \quad (7)$$

$$\vec{p}_{(m)} = \text{mapping}(\vec{p}; I_{(t)} \rightarrow I_{(m)}, z_{D(\vec{p})}) \quad (8)$$

(8) 式の関数 mapping は、合成対象画像 $I_{(t)}$ 上の画素 \vec{p} を、デプス $z_{D(\vec{p})}$ を仮定して、入力画像 $I_{(m)}$ 上の画素 $\vec{p}_{(m)}$ に対応させる。この関数は、 $I_{(t)}$, $I_{(m)}$ のカメラパラメータ、およびデプス $z_{D(\vec{p})}$ から、ホモグラフィ行列による座標変換として定式化されるが、基礎的な幾何の問題なので導出は割愛する。このマッピングにより、合成対象画像 $I_{(t)}$ の座標系でデプスマップを求められる。また、(7) 式の diff を以下のように定義する。

$$\text{diff}(a, b) := \min\{\|a - b\|^2, \text{diff}_{\max}\} \quad (9)$$

この二乗差分は簡単に計算でき、後述する高精度化処理とも親和性が高い。 diff_{\max} は、 $\text{diff}(a, b)$ の上限を定めるものであり、オクルージョンに対処する効果がある。すなわち、複数の画像ペアを同時に評価する場合、あるペアでオクルージョンのため対応が取れなくても、別のペアで対応が成立することがある。前者のペアのコストの上限を抑えておけば、後者のペアでコスト最小となる真のデプスを見つけれられる可能性が高まる。

(8) 式でデプスの値 $z_{D(\vec{p})}$ を任意に決めた場合には、変換前の \vec{p} が整数座標であっても、変換後の $\vec{p}_{(m)}$ は一般に整数座標にならない。提案手法では、 $I_{(m)}(\vec{p}_{(m)})$ を読み出すときに、画素間を bicubic 補間することにした。平行化を前提とした通常のステレオ法では、基準画像と参照画像の視差が整数値になるデプスが規則的に定義されるため、画素位置の整数/非整数の違いを注意深く扱うことが可能である。しかし、平行化を前提としない場合や、本研究のように基準座標系が任意の視点位置に設定される場合には、一般にそのような扱いは困難である。

2.2.4 デプス情報の高精度化

セミグローバル法で得られた離散値デプスマップ D に対して、デプス情報の高精度化を行う。具体的には、各画素 \vec{p} に対して、(3) 式のコスト $S(\vec{p}, n)$ の最小値付近に二次関数を当てはめて極値を求めることで、浮動小数精度でデプスを求める。同様の処理は、ステレオマッチングに限らず、画像の位置合わせ一般に用いられる [11]。

画素 \vec{p} に対してコスト $S(\vec{p}, n)$ ($n \in \{1, \dots, N\}$) の最小値を与える奥行きインデックス、およびその前後のインデックスを $D(\vec{p})$, $D^+(\vec{p})$, $D^-(\vec{p})$, それぞれに対応するコストの値を $S_{\min}(\vec{p})$, $S_{\min}^+(\vec{p})$, $S_{\min}^-(\vec{p})$ と表記する。これらの 3 点について、二次関数を当てはめて極値を求めると、高精度化された (浮動小数に拡張された) デプスのインデックス $\hat{D}(\vec{p})$ と、対応するコスト関数の最小値 $\hat{S}_{\min}(\vec{p})$ を以下のように得られる。

$$\hat{D}(\vec{p}) = D(\vec{p}) + \frac{S_{\min}^-(\vec{p}) - S_{\min}^+(\vec{p})}{2(S_{\min}^-(\vec{p}) - 2S_{\min}(\vec{p}) + S_{\min}^+(\vec{p}))} \quad (10)$$

$$\hat{S}_{\min}(\vec{p}) = S_{\min}(\vec{p}) - \frac{(S_{\min}^-(\vec{p}) - S_{\min}^+(\vec{p}))^2}{8(S_{\min}^-(\vec{p}) - 2S_{\min}(\vec{p}) + S_{\min}^+(\vec{p}))} \quad (11)$$

デプスのインデックスは視差（奥行き）の逆数）と相似なので、物理的には、視差-コスト空間を二次補間して最小値を求めたとみなせる。 $\hat{D}(\vec{p})$ は浮動小数を取るが、物理的な奥行きへの対応は同様に (5) 式によって与えられる。

上記の二次関数の当てはめは、画素値の二乗差分で定義されるコストに対して理論的に正当化される [11]。 (3) 式は、全体としてはこの定義に当てはまらないものの、内部に含まれるマッチングコスト $C(\vec{p}, D(\vec{p}))$ は二乗差分で構成される。したがって、画面内のデプスが平坦な領域（デプスの変化に対するコストが生じない）に関しては理論的正当化が成立すると考えられる。また、二次関数の当てはめには、最終的な位置が整数画素位置に偏る問題（pixel locking）が伴うが、マッチングコストが二乗差分であれば偏りは軽微である [11]。

3. 超解像合成

2章により、合成対象視点から見たデプスマップ \hat{D} ((10) 式) が得られる。本章では、 \hat{D} を用いて入力多視点画像を合成対象視点の座標系に対応付け、その視点から見た高解像度画像 $I_{(t)}^{SR}$ を得る手法を示す。このとき、デプス推定の過程で得られた最小コストマップ \hat{S}_{\min} を正則化に用いることで、混色と超解像を融合し、ノイズを抑制しつつ解像度の高い画像合成を実現する。

準備として、いくつかのパラメータを定義する。まず、入力画像と同一の解像度を持つ合成画像 $I_{(t)}$ は、入力画像の対応画素の混色によって以下のように合成する。

$$I_{(t)}(\vec{p}) = \frac{1}{M} \sum_m I_{(m)}(\vec{p}_{(m)}) \quad (12)$$

$$\vec{p}_{(m)} = \text{mapping}(\vec{p}; I_{(t)} \rightarrow I_{(m)}, z_{\hat{D}(\vec{p})}) \quad (13)$$

\hat{D} , $I_{(t)}$, および (11) 式の \hat{S}_{\min} は、それぞれ、入力画像と同一の解像度を持つが、これを bicubic 法により拡大して高解像度画像 $I_{(t)}^{SR}$ のサイズに合わせたものを、それぞれ、 \hat{D}_{\uparrow} , $I_{(t)\uparrow}$, および $\hat{S}_{\min\uparrow}$ と表記する。

一般の再構成型超解像の枠組みを当てはめれば、合成対象視点における高解像度画像 $I_{(t)}^{SR}$ は、以下のコストを最小化する画像として定義される。

$$E_{sr}(I_{(t)}^{SR}) = E_{sr}^{(1)}(I_{(1)}, \dots, I_{(M)} | I_{(t)}^{SR}) + \lambda E_{sr}^{(2)}(I_{(t)}^{SR}) \quad (14)$$

ここで、 $E_{sr}^{(1)}$ は尤度に基づくコストであり、入力多視点画像 $I_{(1)} \dots, I_{(M)}$ と高解像度合成画像 $I_{(t)}^{SR}$ の関係の妥当性を評価する。 $E_{sr}^{(2)}$ は $I_{(t)}^{SR}$ そのものの事前知識に基づくコストであり、正則化に対応する。 λ は重み付け係数である。一般に、 λ を小さくすれば尤度項の寄与が大きくなるため解像感が高まるが、 λ を大きくすれば正則化が強まってノイズが抑制される。

以下、尤度項 (3.1)、正則化項 (3.2) の定義を述べた後、具体的な最適化手法 (3.3) に言及する。

$$I'_{(m)}(\vec{p}) = 0 \text{ for all } \vec{p} \in I'_{(m)} \dots \text{(i)}$$

for each $\vec{p} \in I_t^{SR}$

$$\vec{p}_{(m)} = \text{mapping}(\vec{p}; I_{(t)}^{SR} \rightarrow I_{(m)}, z_{\hat{D}_{\uparrow}(\vec{p})}) \dots \text{(ii)}$$

get integer pixel positions $\vec{p}_{(m),i}$ ($i = 1, 2, \dots$) ... (iii)

for each $\vec{p}_{(m),i}$

$$\text{if } \|D_{(m)}(\vec{p}_{(m),i}) - \hat{D}_{\uparrow}(\vec{p})\| \leq 1 \dots \text{(iv)}$$

get r_i based on $|\vec{p}_{(m)} - \vec{p}_{(m),i}| \dots \text{(v)}$

$$I'_{(m)}(\vec{p}_{(m),i}) = I'_{(m)}(\vec{p}_{(m),i}) + r_i I_{(t)}^{SR}(\vec{p}) \dots \text{(vi)}$$

end

end

end

図2 Eq. (16) の疑似コード

3.1 尤度項

まず、(14) 式の第一項を以下のように定義する。

$$E_{sr}^{(1)} = \sum_m \sum_{\vec{p} \in I_{(m)}} \|I_{(m)}(\vec{p}) - I'_{(m)}(\vec{p})\|^2 \quad (15)$$

$$I'_{(m)} = f_{(m)}(I_{(t)}^{SR}, \hat{D}_{\uparrow}, PSF) \quad (16)$$

(15) 式の関数 $f_{(m)}$ は、デプスマップ \hat{D}_{\uparrow} を用いて、高解像度合成画像 $I_{(t)}^{SR}$ を入力画像 $I_{(m)}$ の座標系に変換する処理を表す。この変換では、視点だけではなく、解像度も変換される。 PSF は点広がり関数であり、すべての入力画像で一様であると仮定する。(15) 式では、関数 $f_{(m)}$ によって生成される画像 $I'_{(m)}$ と入力画像 $I_{(m)}$ との差分を、 $I_{(m)}$ のすべての画素について合計する。

以下、関数 $f_{(m)}$ の詳細を述べる。まず、オクルージョンの影響を無視すれば、高解像度合成画像上の座標 \vec{p} と、入力画像 $I_{(m)}$ 上の座標 $\vec{p}_{(m)}$ は、

$$\vec{p}_{(m)} = \text{mapping}(\vec{p}; I_{(t)}^{SR} \rightarrow I_{(m)}, z_{\hat{D}_{\uparrow}(\vec{p})}) \quad (17)$$

によって対応づけられる。解像度は $I_{(t)}^{SR}$, $I_{(m)}$ それぞれのカメラパラメータ行列に内包されるため、縮小も自動的に考慮される。この対応付けにおいては、一般に、 \vec{p} が整数座標であっても、 $\vec{p}_{(m)}$ は整数座標とはならないため画素間の補間が必要である。

補間を考慮した処理の原理を述べる。まず、(17) 式の対応関係を使って $I_{(t)}^{SR}$ の各画素を移動した変形画像を生成する。この変形画像は、視点位置は $I_{(m)}$ と同一だが、画素のサンプリング位置は異なる。次に、この変形画像を補間して連続信号を生成し、点広がり関数 PSF を畳み込む。本研究では、連続信号の生成に線形補間、点広がり関数として入力画像の画素サイズと同一のサポートを持つ方形関数を用いた。最後に、この連続信号を $I_{(m)}$ に対応する画素位置でサンプリングし、 $I'_{(m)}$ を得る。

上記の原理的なプロセスを、オクルージョンの前後判定と組み合わせると図2のようなアルゴリズムが導かれる。最初に (i) で $I'_{(m)}(\vec{p})$ の全ての画素はゼロで初期化される。(ii) では、(17) 式の対応関係を使って、 $I_{(t)}^{SR}$ 上の座標 \vec{p} に対応する $I_{(m)}$ 上の座標 $\vec{p}_{(m)}$ を求める。 $\vec{p}_{(m)}$ は一

$D_{(m)}(\vec{p}) = 0$ for all $\vec{p} \in D_{(m)} \dots$ (i)
 for each $\vec{p} \in I_t^{SR}$
 $\vec{p}_{(m)} = \text{round}(\text{mapping}(\vec{p}; I_t^{SR} \rightarrow I_{(m)}, z_{\hat{D}_\uparrow(\vec{p})})) \dots$ (ii)
 if $D_{(m)}(\vec{p}_{(m)}) \leq \hat{D}_\uparrow(\vec{p}) \dots$ (iii)
 $D_{(m)}(\vec{p}_{(m)}) = \hat{D}_\uparrow(\vec{p}) \dots$ (iv)
 end
 end

図3 $D_{(m)}$ を求める処理の疑似コード

般に整数座標ではないため、(iii)において、 $\vec{p}_{(m)}$ 近傍の整数座標 $\vec{p}_{(m),i}$ を求める。 $\vec{p}_{(m),i}$ は、補間関数および点広がり関数の形状を考慮して、 $I_t^{SR}(\vec{p})$ が寄与し得る範囲で選ばれよ。本研究では、有限なサポートを持つ補間関数と点広がり関数を仮定しているため、求める $\vec{p}_{(m),i}$ の数も有限である。(iv)では、オクルージョンの前後判定に相当する処理をする。 $D_{(m)}$ は、入力視点画像 $I_{(m)}$ と同一視点から見た画素単位のデプスマップ（求め方は後述）である。変換元のデプス $\hat{D}_\uparrow(\vec{p})$ と変換先のデプス $D_{(m)}(\vec{p}_{(m),i})$ が一致する場合にのみ、以降の処理を行う。(v)で求める重み $r_i \in [0, 1]$ は、 $I_t^{SR}(\vec{p})$ の $I'_{(m)}(\vec{p}_{(m),i})$ に対する寄与の割合を表わす。 r_i は、 $\vec{p}_{(m),i}$, $\vec{p}_{(m)}$ 間の距離の関数であり、関数の形状は、補間関数と点広がり関数によって定められる。最後に(vi)で、重み r_i を乗じた $I_t^{SR}(\vec{p})$ の値を $I'_{(m)}(\vec{p}_{(m),i})$ に加算する。

$I_{(m)}$ と同一視点から見た画素単位のデプスマップ $D_{(m)}$ は、図3の手順で求められる。まず、(i)で $D_{(m)}$ の全画素を無限遠に相当する0で初期化しておく。 I_t^{SR} 上の各画素 \vec{p} について、(ii)で $I_{(m)}$ 上での対応点 $\vec{p}_{(m)}$ を求め、(iii)で前後判定をしたうえで、(iv)で $\hat{D}_\uparrow(\vec{p})$ のデプスを $D_{(m)}(\vec{p}_{(m)})$ に割り当てる。デプスマップの変換に関しては、(ii)のように小数精度の画素位置ずれを丸めても、超解像の結果にそれほど影響はない。

3.2 正則化項

(14)式の第二項（正則化項）を以下のように定める。

$$E_{sr}^{(2)} = \sum_{\vec{p} \in I_t^{SR}} w(\vec{p}) \|I_t^{SR}(\vec{p}) - I_{(t)\uparrow}(\vec{p})\|^2 \quad (18)$$

$$w(\vec{p}) = \max\{\|\hat{S}_{\min\uparrow}(\vec{p})\|^4, w_{\min}\} \quad (19)$$

この項は、拡大補間された混色による合成画像 $I_{(t)\uparrow}$ が、目的とする画像 I_t^{SR} の近似になっているという仮定に基づく。(14)式で λ を大きくすれば、最適解は $I_t^{SR} \sim I_{(t)\uparrow}$ となることから、混色合成の拡張という意味でも自然な定式化となっている。これに対して、コストの最小値 $S_{\min\uparrow}(\vec{p})$ に基づく重み $w(\vec{p})$ をつけたのは、画素 \vec{p} ごとに正則化の強度を調整するためである。 $\hat{S}_{\min\uparrow}$ は、オクルージョン境界や鏡面反射等のためにデプスが不安定な部分で大きな値になり、デプス推定の信頼度を良く表していると考えられる（図6(a)参照）。したがって、デプスが信頼できる部分では相対的に第一項の比重を大きく

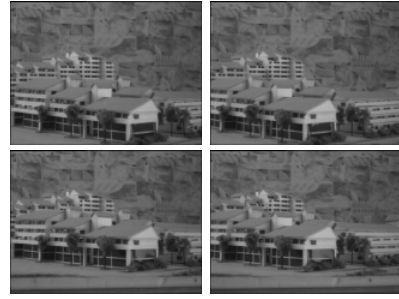


図4 入力画像（筑波大学多視点画像データベースより）

表1 パラメータのデフォルト値一覧

Eq. (1)	$\lambda_1 = 100, \lambda_2 = 400$
Eq. (5)	$z_{\min}=250\text{mm} (21.00^*), z_{\max}=1900\text{mm} (2.76^*)$ $N = 40$
Eq. (6)	size of $B_{\vec{p}}$: 3×3 pixels
Eq. (9)	$\text{diff}_{\max} = 150$
Eq. (14)	$\lambda = 5.0 \times 10^{-13}$
Eq. (19)	$w_{\min} = 10$

* equivalent disparities (pixels) between the input images

することで解像感を高め、デプスが信頼できない部分では第一項の比重を下げて混色合成と同程度の結果を得ようというのが、上記の定式化の意味である。さらに、 w_{\min} によって重み w の最小値を定め、正則化の重みがゼロになるのを防ぐ。この正則化により、混色と超解像を適応的に組み合わせた着色処理が行われる。

3.3 最適化

I_t^{SR} , $I_{(t)\uparrow}$, および $I_{(m)}$ を一次元のベクトルに展開したものを X , \bar{X} , Y_m とする。(16)式の関数 $f_{(m)}$ の入出力関係を表現した行列を A_m , (19)式の w を一次元のベクトルに展開して対角行列としたものを W と表記する。すると、(14)式は、以下のようにも表せる。

$$E_{sr}(X) = \sum_m \|Y_m - A_m X\|^2 + \lambda W \|(X - \bar{X})\|^2 \quad (20)$$

最適化手法としては、(20)式に直接、勾配法等を適用する方法と、(20)式を微分して零と置いた線形方程式を適当な数値計算ライブラリで解く方法がある。本稿では、前者のアプローチで、以下の反復解法を用いた。

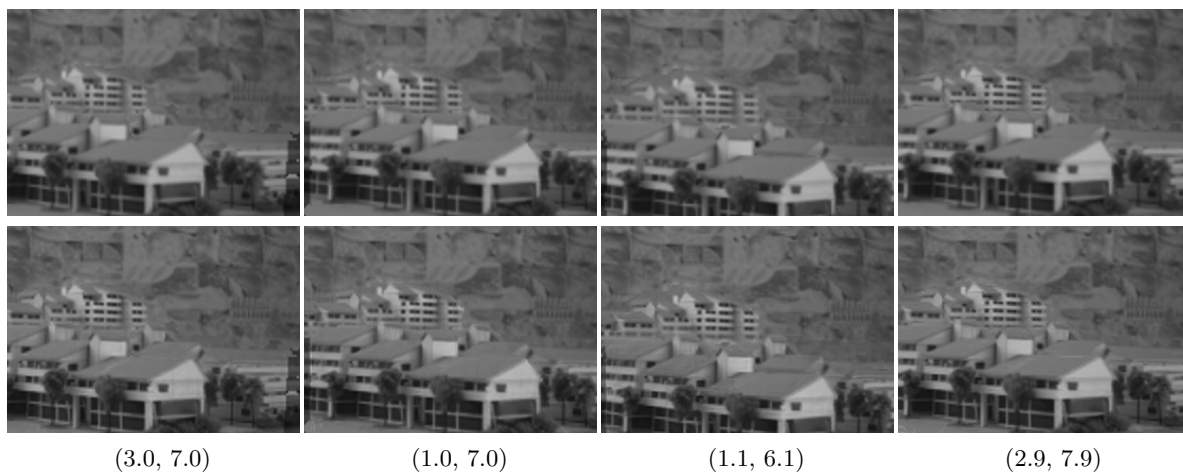
$$X_{j+1} = X_j - \alpha_j \nabla E_{sr}(X_j) \quad (21)$$

$$\alpha_j = \frac{\|\nabla E_{sr}(X_j)\|^2}{\nabla E_{sr}(X_j)^T (\nabla^2 E_{sr}) \nabla E_{sr}(X_j)} \quad (22)$$

ただし、 X の初期値は $X_0 = \bar{X}$ とした。

4. 実験

入力画像として「筑波大学多視点画像データベース」より街のジオラマの4枚の画像（図4）を用いた。入力画像は一辺を16mmとする正方形の頂点（データベースの表記で、(1, 6), (3, 6), (3, 8), (1, 8)の位置）に配列されている。元データは 640×480 画素のカラー画像であ

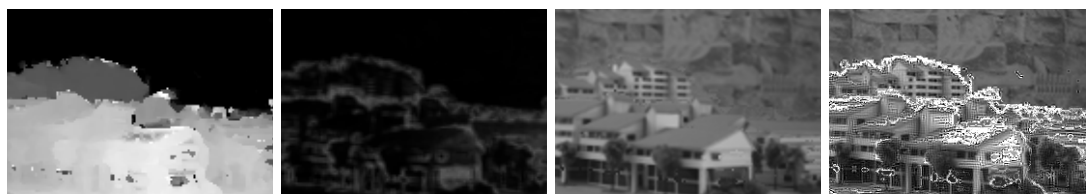


(3.0, 7.0) (1.0, 7.0) (1.1, 6.1) (2.9, 7.9)

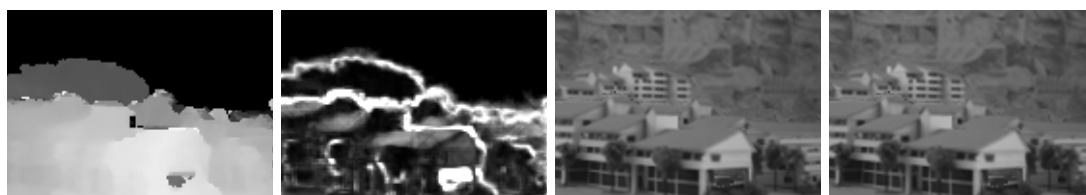
図5 様々な視点に対する合成. 上段: 混色合成して拡大, 下段: 超解像合成. 括弧内は視点位置を表す. 電子版での閲覧を推奨. 添付ビデオも併せて参照されたい.



(a) 提案手法



(b) グローバル最適化なし ($\lambda_1, \lambda_2 = 0$)



(c) マッチングコスト上限値なし ($\text{diff}_{\max} = \infty$)



(d) ブロックマッチングなし ($|B_p| = 1$)



(e) デプス高精度化なし ($\hat{D}(\vec{p}) = D(\vec{p})$)

図6 デプス推定手法による合成結果の違い: 左よりデプスマップ \hat{D} , 最小コストマップ \hat{S}_{\min} , 混色による合成画像の拡大 $I_{(t)\uparrow}$, 超解像による合成画像 $I_{(t)}^{SR}$. 電子版での閲覧を推奨.

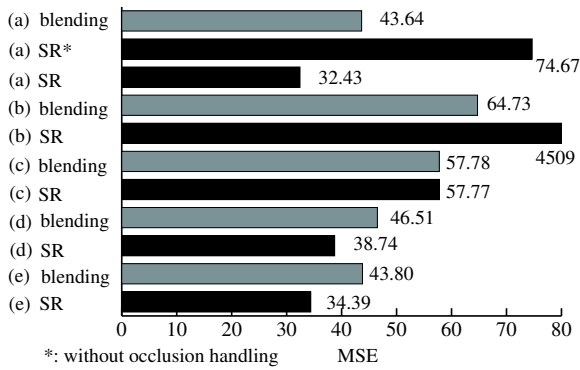


図7 デプス推定手法, およびオクルージョン処理による合成品質の比較. (a)–(e) は図6に対応.

るが, これをグレースケール変換し, 160×120 画素に縮小したものを入力画像として用いた. 提案手法では, 与えられた新しい視点において, 入力画像と同じ 160×120 画素でデプスマップを推定したのち, 超解像合成により 320×240 画素の自由視点画像を得る. 合成対象視点は, 入力視点を頂点とする正方形内部に配置した. 本実験のパラメータのデフォルト値一覧を表1に示す.

まず, 様々な視点位置に対して提案手法で合成した自由視点画像 (I_t^{SR}) を図5に示す. 比較のため, 同一の視点に対して混色のみに基づいて合成した画像 ($I_{(t)\uparrow}$) も併せて掲載する. 提案する超解像処理によって高精細化が達成できていることは明らかである.

次に, 提案手法を構成する要素技術の有効性を検証するため, 詳細な定量評価の結果を示す. 評価では, 合成対象視点を正解画像が存在する中央の位置 ((2, 7) に相当) に設定し, 平均二乗誤差 (MSE) で品質を評価した. 対応がとれない領域が含まれる画像の端部 (上下左右の端部 24 画素) は評価から除外した. 以下, 4.1 ではデプス推定について, 4.2 では超解像合成について述べる.

4.1 デプス推定の評価

提案手法のデプス推定の有効性を確認するため, 手法を構成する要素技術の一つずつ無効にして, デプス推定および画像合成を行った結果を図6に示す. 各行において, 左から順に, デプスマップ \hat{D} , 最小コストマップ \hat{S}_{\min} ($1/10$ にスケール), 混色による合成画像を拡大したもの $I_{(t)\uparrow}$, および超解像による合成画像 $I_{(t)}^{SR}$ を掲載する. (a) では提案手法においてデフォルトパラメータを用いた. (b) では, $\lambda_1, \lambda_2 = 0$ としてセミグローバル最適化を無効にした. (c) では $\text{diff}_{\max} = \infty$ としてマッチングコストの上限を無効にした. (d) では, マッチングコストを累積するブロックサイズを1とした. (e) では, デプス情報の高精度化 ((10)(11) 式) を無効にした.

まず, (a)–(e) のデプスマップを比較する. (a) では, 主観的に良好な品質のデプスマップが得られた. (b) では, かなり誤差が多いことから, セミグローバル最適化の効果は明らかである. (c) では, オクルージョン境界に誤差

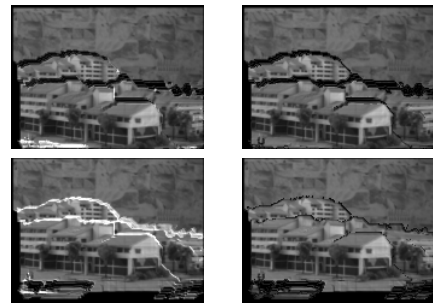


図8 超解像モデルにおける前後判定の効果

が出ており, マッチングコストに上限を設けることがオクルージョン対策に有効であることがわかる. (d) では, 画素単位でマッチングコストを得るため, 粒状感のあるデプスマップになっている. (e) では, 高精度化をしていないため, デプスの階調が明確に量子化されている.

次に (a)–(e) の合成画像を比較する. 図7に定量品質を示す ((a) SR* については後述). 混色合成して拡大した画像 (blending) および超解像による合成画像 (SR) のどちらにおいても, (a) が最も優れている. デプスの精度が高い方が, 混色合成だけではなく, 超解像合成にも有利であるのは当然であるが, 超解像合成には最小コストマップ \hat{S}_{\min} による正則化も含まれることに注意を要する. \hat{S}_{\min} は, デプス推定の手法に依存する. 例えば, (c) のケースでは, ある程度の品質のデプスが得られても, \hat{S}_{\min} の性質が異なるため, 超解像が効いていない. ここでは, 提案手法の品質を最適化するように超解像の正則化パラメータ λ を定めている.

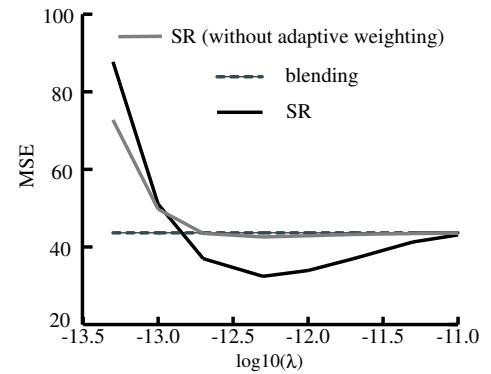
4.2 超解像合成の評価

まず, (16) 式の画像変換の妥当性を調べる. 合成対象視点の正解画像を2つの入力視点に変換した結果 ($I'_{(m)}$ に対応) を図8に示す. (a) は, 前後判定処理 (図2の (iv)) を省略した場合の結果である. 画像中で黒く抜けている領域は, 合成対象画像 $I_{(t)}^{SR}$ のどの画素にも対応しない. この領域は, 関数 $f_{(m)}^{-1}$ の零空間に相当するため, どんな値を取ろうが $I_{(t)}^{SR}$ には影響しない. しかし, 画像中で白とびしている領域には注意が必要である. この領域は, 本来, オクルージョンの関係で背後に隠れるべき画素値が重ねられており, 合成対象画像 $I_{(t)}^{SR}$ と入力画像 $I_{(m)}$ の対応関係が正しく表現されていない. (b) では, 前後判定処理を導入した結果, この問題が解消されていることがわかる. (a) の変換を用いて超解像合成した場合, オクルージョン境界付近に劣化が発生し, MSE も 74.67 と大きくなった (図7の (a) SR*).

次に, 正則化の効果を図9に示す. 提案手法では, 全体の正則化のパラメータ λ ((14) 式) に加えて, 最小コストマップ \hat{S}_{\min} により画素ごとに適応的に正則化の強度を調整している ((18)(19) 式) ため, 適応的な重み付けを無効にした (すべての画素に対して, $\hat{S}_{\min} = 2000$



$\lambda = 5.0 \times 10^{-14}$ $\lambda = 5.0 \times 10^{-13}$ $\lambda = 1.0 \times 10^{-11}$
 (a) 合成結果 (上段: 提案手法, 下段: 適応的重み付けなし)



(b) 定量評価

図9 超解像における正則化の効果. 電子版での閲覧を推奨.

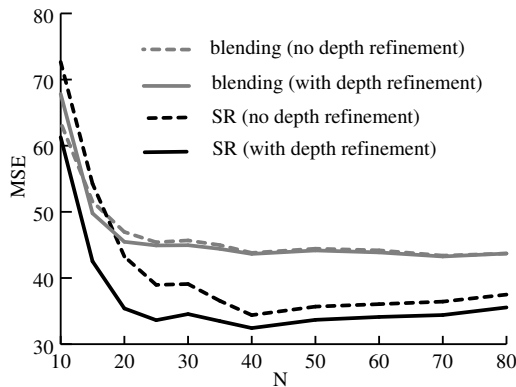


図10 奥行きステップ数と合成品質

の定数とした) 場合も併せて評価する. λ が小さい場合, 解像感は向上するが, 物体の境界付近にノイズが出る. 逆に λ を大きくすると, ノイズが抑制される代わりに解像感が失われ, 混色による合成結果に漸近する. 両者をバランスするところで品質が最良となる. また, 適応的な重みづけにより, ノイズが出やすい物体境界に選択的に強い正則化を施すことができ, 結果として画質向上が達成できていることもわかる. 適応的重み付けを無効にした場合, MSE は混色合成と高々同程度である.

最後に, デプス推定における奥行きのステップ数 (N) を変化させたときの合成品質を図10に示す. ここでは, 混色と超解像のそれぞれについて, デプスの高精度化 ((10)(11) 式) をする場合としない場合を評価した. グラフより, 多少の振動はあるものの, N の増加に伴って MSE が減少する傾向があり, $N = 30$ 程度で品質が飽和することがわかる. また, 混色よりも超解像合成のほうが明らかに優位なこと, 特に超解像合成で N が小さいときに, デプスの高精度化が有効に働くこともわかる.

5. むすび

本稿では, 視点依存型アプローチで, 自由視点映像生成と超解像の枠組みを融合した新しいアルゴリズムを提案した. 今後は, より多数の入力視点画像への対応や,

高速実装に取り組みたい.

謝辞: 本研究は総務省による戦略的情報通信研究開発推進制度 (受付番号: 101703007) の支援を受けました.

文 献

- [1] A. Kubota et al.: "Multiview Imaging and 3DTV", IEEE Signal Processing Magazine, Vol. 24, No. 6, Special Issue, pp. 10–111, 2007.
- [2] 高橋, 苗村: "自由視点画像合成のための視点依存奥行きマップの実時間推定法", MIRU, pp. 197–204 (2005).
- [3] S.-C. Park et al.: "Super-resolution Image Reconstruction: A Technical Overview", IEEE Signal Processing Magazine, Vol. 20, No. 3, pp. 21–36, 2003.
- [4] T. Tung, S. Nobuhara, T. Matsuyama: "Simultaneous Super-Resolution and 3D Video Using Graph-Cuts", IEEE CVPR, pp. 1–8, 2008.
- [5] B. Goldluecke, D. Cremers: "Superresolution Texture Maps for Multiview Reconstruction", IEEE ICCV, pp. 1677–1684, 2009.
- [6] 石井, 高橋, 苗村: "超解像 plane-sweeping による自由視点画像合成", MIRU2010, pp. 1483–1490, 2010.
- [7] <http://vision.middlebury.edu/stereo/>
- [8] S. Birchfield, C. Tomasi: "A Pixel Dissimilarity Measure That is Insensitive to Image Sampling," IEEE TPAMI, Vol. 20, No. 4, pp. 401–406, 1998.
- [9] H. Hirschmuller: "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," IEEE CVPR, pp. 807–814, 2005.
- [10] H. Hirschmuller: "Stereo Processing by Semiglobal Matching and Mutual Information," IEEE TPAMI, Vol. 30, No. 2, pp. 328–341, 2008.
- [11] 清水, 奥富: "画像のマッチングにおける高精度なサブピクセル推定手法", 信学論, Vol. J84-D-II, No. 7, pp. 1409–1418, 2001.
- [12] J. Kim, V. Kolmogorov, R. Zabih: "Visual Correspondence Using Energy Minimization and Mutual Information," IEEE ICCV, pp. 1033–1040, 2003.
- [13] V.-M. Vergauwen et al.: "A Hierarchical Symmetric Stereo Algorithm using Dynamic Programming," IJCV, vol. 47, pp. 275–285, 2002.
- [14] J.-X. Chai, X. Tong, S.-C. Chan, H. Shum: "Plenoptic Sampling," ACM SIGGRAPH, pp. 307–318 (2000).
- [15] S.-K. Gehrig, U. Franke: "Improving Stereo Sub-Pixel Accuracy for Long Range Stereo," IEEE ICCV, pp. 1–7, 2007.
- [16] OpenCV v2.1 Documentation, StereoSGBM.