

多視点ステレオのための 信頼度に基づく高精度デプスマップの推定と融合

池畑 諭† 山崎俊彦† 相澤清晴†

† 東京大学

E-mail: †{ikehata,yamasaki,aizawa}@hal.t.u-tokyo.ac.jp

あらまし 多視点ステレオ (MVS) において、正確なデプスマップやモデルを復元することは依然として困難な問題である。本研究ではデプスマップの推定と高精度化、および融合処理を組み合わせ、高精度なデプスマップと密な 3 次元モデルを得る手法を提案した。特にデプスマップを入力として、その高精度化のためにバンドル最適化を施すのは初めての試みであり、本手法で提案するもう一つの高精度化手法である奥行伝播と併せて既存のデプスマップ推定手法に容易に用いることができるだけでなく、その効果が非常に高い事を評価実験によって示した。

キーワード 多視点ステレオ, デプスマップ推定, イメージベースドモデリング, バンドルアジャストメント, 信頼度

1. はじめに

多視点ステレオ (Multi-View Stereo:MVS) に関する研究は、3D コンテンツの必要性の増加や Middlebury [1] や Strecha ら [2] が提供している共通のデータセットによって急速に進展している。Seitz ら [3] は、現存する MVS のアルゴリズムを体積ベース [4]、レベルセット等に基づく曲面進化 [5]、デプスマップの推定と統合 [6]~[10]、特徴抽出と曲面フィッティング [11], [12] という 4 つのアプローチに分類した。近年では特に 3 番目のカテゴリが注目されている。理由としては前景領域の抽出 [13] や物体追跡 [14] 等にも応用可能なデプスマップの汎用性や、他の表象と比較して非常に小さいデータ量で 3 次元を表象でき、安価なデプスカメラ等によってデプスマップを得る事が比較的容易になってきた事等が挙げられる。本研究においても、このアプローチに関して提案を行う。

既存のデプスマップ推定手法は大きく分けて局所的な特徴に基づいた手法 [6], [10] とエネルギーの最適化に基づいた手法 [7]~[9] の 2 つに分けられる。前者においては、局所的な見えの類似性に基づいて対応が探索されるため、計算コストとメモリにおいて効率的である反面、射影歪み等により対応検出精度が低くなるという欠点が存在する [6], [10]。一方、エネルギーの最適化に基づいた手法においては、奥行きの連続性や視点間の輝度の類似性等の様々な要因をエネルギー関数に反映させ、それが最適となるような解を求める。エネルギーを最適化する手法としては最急勾配法 [7], [8] や信念伝播法 [9] が用いられ、正しく収束されれば高精度のデプスマップを推定する事が可能である。反面、初期値が大きく誤っていたり、オクルージョンや奥行きの非連続性を正しく扱っていない場合においては、誤った解へと収束してしまう。またエネルギーは個々の視点に対して独立に最適化されるため、推定結果は視点間の空間的な整合性は考慮して

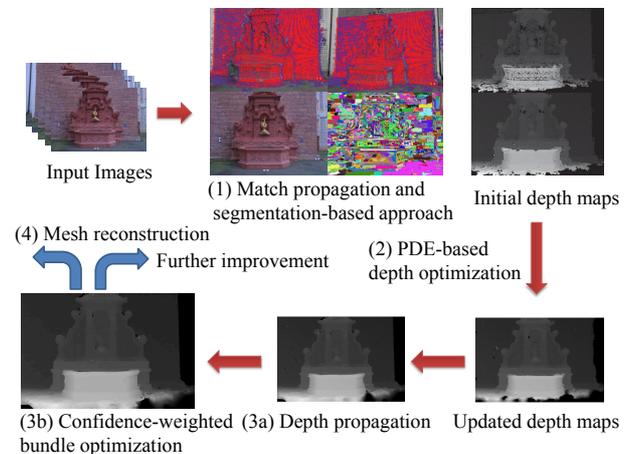


図 1 システムの概要。アルゴリズムは (1) 初期デプスマップの推定、(2) PDE に基づいたデプスマップの最適化、(3) 2 段階のデプスマップの高精度化、(4) メッシュ復元という 4 つのステージに分けられる。

いないという問題点も挙げられる。本論文では上述した課題を慎重に考慮しながら、高精度のデプスマップを推定・融合するための MVS アルゴリズムを提案する。図 1 にアルゴリズムの概要を示した。アルゴリズムは大きく 4 つのステージに分けられる。最初のステージでは対応伝播 [15] とセグメンテーションに基づく奥行き推定手法 [16] を組み合わせ、デプスマップの初期値を推定する。第二のステージにおいては、連続量としてのデプスマップの局所的な連続性や画素単位の色の整合性を考慮してエネルギー関数を構築し、偏微分方程式 (Partial Differential Equation :PDE) に基づいた最適化を施す。適切にオクルージョンを扱うために、推定された奥行きの信頼性を評定するための信頼度を導入する。また過度な平滑化を防ぐために、デプスマップにおける非連続境界の生起確率を推定する MRF(Markov Random Field) モデルを構築する。第三のステージでは、解の視点間の

整合性を考慮してデプスマップの高精度化を試みる。このステージはさらに奥行き伝播と信頼度に基づいたバンドル最適化という2つのステップに分けられる。前半部においては、信頼性が高い奥行き値を信頼性の低い奥行き値へと伝播させ、後半部においては、密な対応点列(トラック)を信頼度に基づいて抽出し、それらを重みづけされた再投影誤差に関して最適化する。得られた3次元点群は、セグメンテーションに基づいたデプスマップ推定法[16]の入力として与えて高精度化されたデプスマップを得るために用いたり、直接的に3次元モデル復元の入力として用いる事が可能である。3次元モデルは既存の曲面復元のアルゴリズムを用いる事によって得る事ができる。本論文の主な貢献は次のように要約される。

(1) 対応の信頼性と奥行き値の信頼性を正確に評定する信頼度を提案し、オクルージョンの検出やデプスマップの高精度化・融合の枠組みにおいて活用する。

(2) デプスマップの過度な平滑化を防ぐために、非連続境界を抽出するためのMRFモデルを提案する。従来のように画素単位ではなく、セグメント境界に対してラベル付けを行う事により過度な非連続境界の検出を防ぐ。

(3) 信頼度とバンドル最適化を組み合わせたデプスマップの高精度化・融合法を世界で初めて提案する。この高精度化法は、他の手法やデプスカメラ等によって得られた精度の低いデプスマップを高精度化する目的においても用いる事が可能な汎用的な手法である。

2. 関連研究

最初にデプスマップ推定法について言及する。Strechaら[8]は、オクルージョン検出のために信頼度を導入し、デプスマップを連続値として得るためにPDEを構築した。Liuら[7]は複数のスケールで量子化された画像の推定結果を統合して最終的なデプスマップを得る手法を提案した。これらは前章で述べたエネルギー最適化に基づく手法である。対照的に、Lhuillierら[15]は、対応伝播と呼ばれるエネルギー最適化を要しないアルゴリズムを提案した。これは最初に与えられた疎な対応点の近傍において更なる対応点を探索するという手法である。またJancosekら[17]は短いベースラインのステレオで一般的に用いられているセグメンテーションに基づいたアプローチ[16]をMVSに導入した。このアプローチはある程度密なサンプルが与えられれば、テクスチャに乏しい領域における奥行きを推定する事ができる。我々の提案する手法は、これらと密接に関わっているが、4つの点で異なっている。第一に、我々はエネルギー最適化に基づく手法とそうではない手法を組み合わせ、両者の欠点を補う。第二に我々はStrechaら[8]が用いたforward-backward誤差よりも頑健な信頼度を提案し、信頼度をより広い目的で用いる。第三に我々はデプスマップの非連続性を保持するために、セグメント境界のラベル付けを行うMRFモデルを構築する。第四に我々は、

デプスマップを推定した後にそれらを高精度化、融合するための2段階の処理を導入する。

次にデプスマップの融合法について言及する。多くのアルゴリズムは信頼度が最も高い点を対応として与えるWinner-Takes-All(WTA)戦略[6],[7]に基づいている。近年Liら[10]は、structure-from-motion法[18]で一般的な3次元点とカメラの運動を同時に修正するための手法であるバンドル最適化をデプスマップの融合に対して用いる事を試みた。Liらは、推定された2視点間対応を結びつけて密なトラックを抽出し、それらを再投影誤差に関して最適化した。我々が提案する高精度化・融合手法もこの方法と非常に密接に関係しているが、大きく異なる点として、我々は2視点間の対応をではなくデプスマップを入力しているという点、トラックの抽出に対して信頼度を用いているという点が挙げられる。

第三に信頼度について言及する。本研究で述べる信頼度とは、「対応あるいは奥行きの正しさを評価するための指標」である。信頼度は元来ステレオにおいて局所的に対応点を探索するために用いられてきたが、近年はそれ以外の目的で用いられるようになってきている。例えば、Gongら[19]は、信頼度をdynamic programmingにおけるコストを評定するために導入し、Strechaら[8]は、信頼度を最適化問題におけるオクルージョンのコストとして用いた。信頼度としては様々な指標が提案されており、Huら[20]は、13種類の信頼度を屋内屋外両方のデータセットを用いて評定し、彼らが提案したLeft-Right-Difference(LRD)がその中で特に優れている事を示した。本論文では、LRDを参考に新たな信頼度を提案し、それらをエネルギー最適化とデプスマップの高精度化および統合処理の枠組みの中に導入する。

最後に、局所特徴量について言及する。SIFT[21]は恐らく最も一般的に用いられている局所特徴である。SURF[22]はSIFTを高速化したものだが、やはり計算コストが高く、反復計算には向いていない。一方DAISY[23]はSIFTと同様に輝度変化や射影歪みに対して頑健であるという性質を持っていながら、SURFと比較しても計算コストが非常に小さい。Liら[10]はこのDAISYをMVSのシステムに導入し、それが高い精度と効率性の両方を与える事を示した。我々のアルゴリズムにおいても、DAISYを信頼度の計算の一部に導入している。

3. 信頼度

本論文では、対応と奥行き信頼性を評定するために、次式で定義される2種類の信頼度 c_{ij} と C_i を導入する。

$$c_{ij}(x_i) = \frac{1}{1 + \frac{|x_i - \hat{x}_i|}{\eta_a}} \frac{1}{1 + \frac{|F(x_i, x_j) - \min_{p_i} F(x_j, p_i)|}{\eta_b}} \quad (1)$$

$$C_i(x_i) = \max_j c_{ij}(x_i) \quad (2)$$

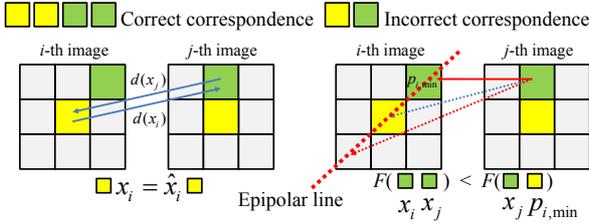


図 2 $d(x_i)$ と $d(x_j)$ が誤っていても，対応が閉じていれば ($x_i = \hat{x}_i$) forward-backward 誤差は高い信頼度を与える (左)．しかし第二項において局所的に最適な対応の類似性を評価する事によってペナルティを与える事ができる (右)．

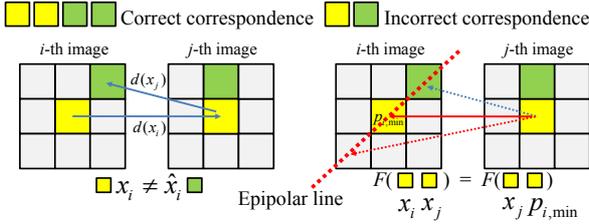


図 3 $d(x_i)$ が正しくても $d(x_j)$ が誤っていれば ($x_i \neq \hat{x}_i$) forward-backward 誤差は低い信頼度を与える (左)．しかし最適な解がその近傍に存在していれば，第二項によって高い信頼度を与える事ができる (右)．

ここで， x_i は画像 i 上の点， x_j は視点 j における x_i の対応点であり， x_i における奥行き値によってユニークに与えられる．また \hat{x}_i は x_j がその点の奥行き値によって視点 i に対して逆投影された点である． $p_i \in N(\hat{x}_i)$ は x_j に対応するエピポーラ線上に存在する \hat{x}_i の近傍の点の集合であり， $F()$ は次のように与えられる．

$$F(v, w) = \sum_{k=1}^m |D^{[k]}(v) - D^{[k]}(w)|^2 \quad (3)$$

$D^{[k]}()$ はある点における DAISY 特徴 [23] の k 番目のヒストグラムである．DAISY 記述子の 4 つのパラメータ (局所領域の半径 R ，円環の数 Q ，一つの円環に含まれるヒストグラムの数 T ，一つのヒストグラムにおけるビンの数 H) にはそれぞれ 15, 3, 8, 8 を与えた． η_a と η_b は重みであり，それぞれ 1 と 0.2 を与えた．

Strecha ら [8] は，伝統的な forward-backward 誤差 (c_{ij} の前半部) のみを信頼度として導入した．これは頑健ではあるものの信頼性を誤って推定してしまう次の二つのケースが挙げられる．すなわち (1) $d(x_i)$ と $d(x_j)$ が共に誤っているが，対応が閉じているため信頼度を高く推定してしまう場合 (図 2 参照)，(2) x_i における奥行き値 $d(x_i)$ は正しいのに x_j における奥行き値 $d(x_j)$ が誤っているために信頼度を低く推定してしまう場合 (図 3 参照) である．我々はこれらの問題に対して Hu ら [20] が提案した LRD と類似した性質を持つ c_{ij} の後半部を導入する事によって解決する．直観的には，真に対応している点は局所的に最も類似性が高いという仮定に基づいている

(詳細は図 2, 図 3 を参照)．類似性の評価には高速かつロバストな DAISY 特徴を用いた．オクルージョンによって対応が存在しない可能性を考慮し，奥行き信頼度 C_i としては c_{ij} の最大値を用いた．

4. デプスマップの初期化

エネルギー最適化に基づくデプスマップの推定は大域的な制約を反映させる事ができるが，誤った解への収束を防ぐためにはある程度正確な初期値を必要とする．一方で，最適化を伴わない手法は初期値問題に煩わされる事は無いが，解の密度や精度が不十分である．本論文では双方の欠点を補うために，最初に最適化を伴わない手法でデプスマップを推定し，その結果をエネルギー最適化の枠組みで用いるというフレームワークを提案する．特に最適化を伴わないデプスマップの推定法として，既存手法である対応伝播 [15] とセグメンテーションに基づいた手法 [16], [17] を適切に組み合わせる．まず入力画像の連続する 2 視点に対して対応伝播を適用する．対応伝播に関しては Lhuillier ら [15] を参照していただきたいが，オリジナルと異なり，入力の疎な対応点群として ZNCC ではなく SIFT [21] を用いた．SIFT を用いたのは，Tola ら [23] が言及したように ZNCC や DAISY よりも SIFT が頑健であり，デプスマップの初期化は反復を伴わないので計算コストが問題にならないためである．また本研究ではカメラパラメータが与えられている事を前提として，再投影誤差が一定の閾値 (本実装では 1.0) を上回っている点を誤対応として除去した．得られた精密な点群は，直ちにセグメンテーションに基づいたデプスマップ推定法 [16], [17] の入力として用いられる．画像をミーンシフトセグメンテーション法 [24] により分割し，各セグメントにおけるデプスマップを平面に近似して，そのパラメータを投票法 [16] によって推定する．本手法による推定結果例を図 4 に示した．

5. PDE に基づいたデプスマップの最適化

5.1 連続量としての奥行き推定

デプスマップの最適化は，次式を PDE に変換し，最急勾配法により最小化する事で行われる．

$$E_D(d_i) = \int \sum_{j \neq i}^N (1 - o_{ij}(x_i)) |I_i(x_i) - I_j(x_j)|^2 dx_i + \lambda \int \sum_{x_k \in N(x_i)} b_{ik} |d(x_i) - d(x_k)|^2 dx_i \quad (4)$$

d_i は画像 i に対応する連続量としてのデプスマップ， I は RGB 空間で表される色値， $N(x_i)$ は x_i の 4 近傍である．また o_{ij} と b_{ik} はそれぞれオクルージョンと非連続境界に関する 2 値ラベルであり，理想的には， o_{ij} は x_i が視点 j から観察されていない場合， b_{ik} はデプスマップ

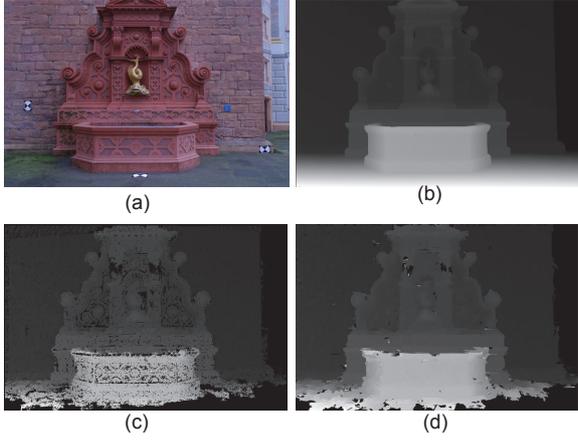


図4 初期化結果 (a) 入力画像 (b) ground truth (c) 対応伝播の結果 (d) セグメンテーションに基づいたデプスマップ推定の結果。

プが x_i と x_k の間で断続的に変化している場合にそれぞれ1を示す。最急勾配法を適用するためにコスト関数をPDEに変換する方法や、陰的スキーマにおいて解を逐次的に更新する方法に関してはStrechaら[8]と同様であるためここでの説明は割愛する。

5.2 非連続境界の検出

過度の平滑化は正確なデプスマップ推定の妨げになる。既存研究においては、輝度や信頼度が大きく変化する領域では平滑化を行わないといった工夫がなされてきた[8]。しかしこれらはいずれも画素単位で局所的に非連続境界の有無を判断するもので、境界の連続性や非連続境界の幾何的特性に関する進んだ考察は為されていない。本研究では、非連続境界をより自然環境の性質に適した形で検出するためのモデルを構築する。すなわち非連続境界は物体境界のように断続的に出現するという性質、そして非連続境界はオクルージョン領域の境界において発生するという性質である。エネルギー関数を次のように定義しグラフカット法により最適化する。

$$\begin{aligned}
E_B(b_i) &= \sum_{(s_i^l, s_i^m) \in S(i)} [(1 - b_i^{lm})\rho_b(SL(s_i^l, s_i^m)) + \eta_b b_i^{lm}] \\
&+ \sum_{(s_i^l, s_i^m) \in S(i)} \beta_o |b_i^{lm} - O(s_i^l, s_i^m)| \\
&+ \sum_{(s_i^l, s_i^m) \in S(i)} \sum_{(s_i^p, s_i^q) \in N(s_i^l, s_i^m)} \beta_s |b_i^{lm} - b_i^{pq}|
\end{aligned} \quad (5)$$

このモデルのサイトは初期化処理で与えられたセグメント間の境界である。 $S(i)$ は共通の境界を持つセグメント対の集合であり、 s_i^k を画像 i の k 番目のセグメント ($k < S_i$)、 $b_i^{lm} \in \{0, 1\}$ を s_i^l と s_i^m の境界に付与されるラベルとする (非連続境界であれば1)。 β_o 、 β_s 、 η_b は重

みであり、本実装では全て1.0が与えられる。また ρ_b はロバスト関数であり、次のように定義される。

$$\rho_b(x) = -\ln((1 - e_b) \exp(-|x|/\sigma_b) + e_b) \quad (6)$$

σ_b と e_b はロバスト関数の形状を規定するパラメータで本実装では0.01と0.003を各々に与えた。 $SL(s_i^l, s_i^m)$ は隣接するセグメントに対応するデプスマップの関係性に応じてコストを与える項であり、便宜上傾き関数 (Slant Function, SL) と呼ぶ。 (a_l, b_l, c_l) と (a_m, b_m, c_m) を s_i^l と s_i^m に対応する平面のパラメータ、 V_i^{lm} と H_i^{lm} を s_i^l と s_i^m の共通セグメント境界に属する直交・並行の境界断片 ($v_{i,k}^{lm}, h_{i,k}^{lm}$) の集合であるとする。この時、 $SL()$ は次のように定義される。

$$\begin{aligned}
SL(s_i^l, s_i^m) &= \frac{1}{S} \left[\sum_{v_{i,k}^{lm} \in V_i^{lm}} \left(\frac{|a_l x_{v,k}^l + b_l y_{v,k}^l + c_l - a_m x_{v,k}^m - (b_m y_{v,k}^m + c_m)|}{\sqrt{1 + a_m^2}} \right. \right. \\
&+ \left. \frac{|a_m x_{v,k}^m + b_m y_{v,k}^m + c_m - a_l x_{v,k}^l - (b_l y_{v,k}^l + c_l)|}{\sqrt{1 + a_l^2}} \right) \\
&+ \sum_{h_{i,k}^{lm} \in H_i^{lm}} \left(\frac{|a_l x_{h,k}^l + b_l y_{h,k}^l + c_l - b_m y_{h,k}^m - (a_m x_{h,k}^m + c_m)|}{\sqrt{1 + b_m^2}} \right. \\
&+ \left. \left. \frac{|a_m x_{h,k}^m + b_m y_{h,k}^m + c_m - b_l y_{h,k}^l - (a_l x_{h,k}^l + c_l)|}{\sqrt{1 + b_l^2}} \right) \right] \quad (7)
\end{aligned}$$

S は V_i^{lm} と H_i^{lm} に含まれる断片の総数である。直観的には $SL()$ は隣接する一方の平面の境界からもう一方の平面に下ろした垂線の和を表し、セグメント境界において各々の平面に対応する奥行き値の差が大きく、その平面が平行である場合に高い値を示す。

第二項はデプスマップの非連続境界がオクルージョン領域の境界において発生するという性質を反映させた項である。ただしオクルージョン境界は注目視点の物体境界のみならず、他視点における物体境界の射影である可能性もあり得るので、その境界がどちらの物体の非連続境界に属しているのかを判断するのは難しい(図5参照)。そこで、本研究ではセグメント境界がオクルージョン境界に属している場合においてある一定のペナルティを与えるという手法を提案した。具体的には対象とするセグメント境界を s_i^l と s_i^m とした時、それらのセグメントの中心点 \bar{x}_i^l と \bar{x}_i^m におけるオクルージョンラベル、あるいはそれらの点の他視点への投影 $\bar{p}_{ij}^l, \bar{p}_{ij}^m$ におけるオクルージョンラベルを比較し、2値マップ $O(s_i^l, s_i^m) \in \{0, 1\}$ を $o_{ij}(\bar{x}_i^l) \neq o_{ij}(\bar{x}_i^m)$ か $o_{ji}(\bar{p}_{ij}^l) \neq o_{ji}(\bar{p}_{ij}^m)$ が満たされている場合に1を与えるものとして定義する。

5.3 オクルージョンの検出

オクルージョンが発生している対応を最適化の系に含む事は誤推定の原因となる。提案手法ではオクルージョンを正確に検出するために、次のコスト関数をグラフカット法により最適化する。



図5 画像 i の視点 j に対する可視性マップを画像 i に重畳表示した．明るい領域は視点 j から観察不可能な領域を表し，図に描かれた曲線はオクルージョン領域の境界と他視点における対応を示す．

$$E_O(o_{ij}) = \sum_{x_i} [(1 - o(x_i))(1 - c_{ij}(x_i)) + o(x_i)\eta_o] + \sum_{x_i} \sum_{y \in N(x_i)} \eta_s |o_{ij}(x_i) - o_{ij}(y)| \quad (8)$$

o_{ij} は5.1節で定義した2値のオクルージョンラベルである． η_o と η_s は重みであり本実装ではそれぞれ0.8, 0.1を与えた．このモデルは，低信頼度領域においてオクルージョンが生起しており，またオクルージョン領域は空間的に連続しているという仮定に基づいている．Strechaらのモデル[8]はforward-backward誤差を連続量としてモデルに導入しオクルージョンを検出したが，我々の提案する信頼度は2視点間対応の幾何的かつ視覚的な局所対称性を考慮しているのど，むしろSunら[25]が提案した，2視点間の見えの対称性を考慮したオクルージョン検出モデルに類似した働きをすると捉える事ができる．

6. 2段階のデプスマップ高精度化と融合

特徴ベースとエネルギー最適化ベースの手法の融合により誤差の少ないデプスマップが得られる．一方最適化は各視点独立に施されるため，解の視点間の整合性は一切考慮されていない．即ちそれを考慮する事により，さらに精度を向上させる事が可能である．本章ではその知見に基づき，2段階の処理によってデプスマップを高精度化・融合し，最終的に3次元点群を得る手法を提案する．

6.1 低信頼度領域に対する奥行き伝播

デプスマップ高精度化の最初の処理は奥行き伝播である．概要を述べると，対象視点におけるデプスマップを改善するために，その視点に対する仮想的なデプスマップを他視点のデプスマップの融合として生成し，対象視点と仮想視点のデプスマップの信頼度を比較してより高い方を採用するという手法である．これは，複数の異なるスケールの画像に対して推定されたデプスマップを融合して高精度のデプスマップを得るLiuらの手法[7]と類似しているが，仮想的なデプスマップを合成するために他の視点の情報を用いるという点および用いる奥行き値の選択に信頼度を利用するという点が異なる．

提案手法をより形式的に述べる．対象視点 i における点 x_i を対応点として持ち，かつ $F(x_i, x_j) < \varepsilon_a$ が満たされている点 x_j を他の全ての視点から抽出する(対応は $d(x_j)$ に基づいて抽出する)．ここで ε_a は閾値であり，本実装では2.0を与えた．次に抽出された点群の中から，最も信頼度の高い点を次のように選択する．

$$j_{max} = \operatorname{argmax}_j C(x_j) \quad (9)$$

対象視点の全ての画素に対して同一の処理を施す事により信頼度が高い点によって対象視点の仮想的なデプスマップが生成される．この時 $C_i(x_i)$, $C(x_j)$ を比較し，次の条件が満たされている場合において，三角測量の原理に従って $d(x_j)$ を $d(x_i)$ に伝播する．

$$C(x_{j_{max}}) > \varepsilon_b, C(x_{j_{max}}) - C_i(x_i) > \varepsilon_c \quad (10)$$

ここで ε_b と ε_c は閾値であり，本実装ではそれぞれ0.5と0.2を与えた．対象視点と仮想視点のうち信頼度が高い奥行き値を採用する事により，デプスマップにおける低信頼度領域の多くが高精度化される．またこれにより，仮想視点を生成するために用いた視点と対象視点の間で空間的な整合性が保持される．ただし高精度化はこれだけでは不十分である．なぜならばこの手法は，第一に低信頼度領域のみしか高精度化されず，第二にサブピクセル精度の対応を保証せず，そして第三に2視点間の幾何的整合性のみしか考慮されていないためである．それゆえ，第二段階として，次節で述べるバンドル最適化を施す．

6.2 デプスマップに対するバンドル最適化

2段階目の処理は信頼度に基づいた密なトラックの抽出，バンドル調整法によるトラックの最適化，デプスマップの再推定(任意)という3つのステップによって成り立っている．最初にトラックの定義について述べる．本論文ではトラックを対応点の集合として次のように定義する $t_k = \{x_1^k, x_2^k, \dots, x_{N(t)}^k\}$ ．密な2視点対応が得られていれば，トラックは密に定義する事が可能であり，この意味において密なトラックであると表現する．これとは対照的にstructure-from-motion法における一般的なバンドル調整は疎なトラックに対して用いられる手法であると言える．密なトラックに対するバンドル調整を初めて行ったのはLiら[10]であるが，彼らは特徴ベースの手法によって2視点間対応を抽出し，それらを結合して密なトラックを生成しており，デプスマップに対してバンドル調整を施してはいない．他方，我々は推定されたデプスマップを入力としてそれに対してバンドル調整を適用してより高精度なデプスマップを得る事を目的として，デプスマップから密なトラックを抽出する手法を提案する．表1にトラック抽出のアルゴリズムを示した．直観的には，トラックを高い信頼性を持った点同士の結合として抽出する．全視点の全画素からトラックに含まれていない最も C_i が高い点を選択し，それを新たなト

表 1 信頼度に基づいた密なトラック抽出のアルゴリズム

0.	$List$, $Match$ と $Track$ を作成する. $List$ と $Match$ は画素を要素とするスタック, $Track$ は $Match$ を要素とするスタックである.
1.	全ての画素を信頼度 C_i , c_{ij} と共に $List$ に追加する.
2.	$List$ を C_i の降順で並び替える.
3.	$List$ の先頭から点を取り出し, x_i とする. もし x_i が $Track$ に含まれていなければ, $Match$ に追加する.
4.	次の 3 つの条件が満たされれば, 画像 j における x_j を $Match$ に追加する. (1) x_j は x_i の対応点である. (2) $j = \operatorname{argmax}_k c_{ik} C_k$. (3) 視点 j は $Match$ に含まれていない. もし全ての視点が $Match$ に含まれていたり, $c_{ij} C_j$ が閾値を下回っていければ 4 へ進む. そうでなければ x_j を x_i にセットして 3 を繰り返す.
5.	もし $Match$ のサイズが 3 未満であれば 6 へ進む. そうでなければ 5 へ進む.
6.	$Match$ を $Track$ に追加する.
7.	$List$ が空で無ければ, $Match$ を初期化して 2 に戻る.
8.	密なトラック群を $Track$ に得る.

トラックの始点とする. 次にその点の奥行き値を用いて, 次に結合する点を探索する. より長いトラックを生成するために, 現在の視点と次の視点に対する対応の信頼度 (c_{ij}) と次の視点からその次の視点に対する対応の存在の可能性 (C_j) の積が最も高い点を次の対応点とする. もし $c_{ij} C_j$ が閾値を下回ったならば (本実装では 0.5) トラックの抽出を終了する. 閾値を設定する事により, 信頼度が高い点同士でトラックが生成されるためバンドル調整において誤推定点に解が牽引される事が抑制される. 抽出されたトラックを最適化するために次のコスト関数を定義する.

$$\min E = \sum_{t=1}^T \sum_{i=1}^{N_t} w_{i,t}^2 |\tilde{x}_{i,t} - P_i \tilde{X}_t|^2 \quad (11)$$

$$\begin{aligned} w_{i,t} &= C_{1,t} & (i = 1) \\ w_{i,t} &= C_{1,t} \prod_{k=2}^i c_{(k-1,k),t} & (else) \end{aligned} \quad (12)$$

ここで, T はトラックの数, N_t はトラックに含まれる点の数, P は投影行列である. $\tilde{x}_{i,t}$ はトラック t における i 番目の点, \tilde{X}_t はトラック t に対応する 3 次元点である. $w_{i,t}$ は対応の信頼性とその点がトラックに導入された順序を反映させた重みである. これは, 先述した閾値と同様に誤推定点に収束解が牽引されるのを防ぎ, 誤った点よりも精度の高い点の恩恵を強く受けるために与えられる. コスト関数は Levenberg-Marquardt 法によって最小化され, 出力として高精度な 3 次元点群を得る. これらの点群を用いてメッシュ復元を行う事が可能であるが, これらの点を再度各視点に投影し, セグメンテーションに基づいたアプローチ [16], [17] の入力として用いる事によって高精度化されたデプスマップを得る事も可能である. トラックの最適化は並列化可能であり, それにより大幅に高速化する余地がある事に言及しておく.

7. メッシュ復元

3 次元点群を入力としてメッシュによって表現された 3 次元モデルを生成するためには様々な既存手法を適用可能であるが, 本論文では poisson surface reconstruction 法 [26] を用いた. この手法はノイズや定率化されていないサンプリングレートに対して頑健である.

8. 評価実験

8.1 実装とデータセット

提案手法を C/C++ プラットフォームで実装した. DAISY の実装においては Tola ら [23] が提供しているライブラリを, Levenberg-Marquardt 法の実装にはオープンソースのライブラリ levmar [27] を利用した. 全ての実験は 3.06 GHz Intel Core2 Duo CPU and 4GB RAM マシンで行われた. 計算コストの大半は, 対応伝播および DAISY を用いた信頼度の計算に費やされた (それぞれ反復一回につき各画像に対して 1 分程度であった). アルゴリズムを 2 種類のデータセットを用いて評価した. 用いたデータセットは, Strecha ら [2] が提供している屋外データセットと Middlebury [1] の屋内データセットである. 本章では最初に Strecha ら [2] のデータセットを用いてデプスマップ推定における提案手法の性能を示す. 続いて Middlebury データセット [1] を用いて 3 次元メッシュモデルの復元結果を示す.

8.2 屋外画像を用いたデプスマップ推定性能評価

Strecha ら [2] によって提供されている屋外データセットに対しては, Tola ら [23] がデプスマップとオクルージョンマップの ground truth データを提供しており, 推定されたデプスマップの精度を量的に評価する事が可能である. そこで本提案手法をデプスマップベースの MVS アルゴリズムにおいて特に優れたアルゴリズムの一つである Strecha らの手法 [8] と比較した. 用いたのは fountain-P11 の 11 枚の画像のうち, 5 枚の画像をオリジナルの解像度 $3,072 \times 2,048$ から 768×512 にダウンサンプリングしたものである. 奥行き値はシーンによって分布が大きく異なるため, 絶対量として評価するのは不適切である. そこで次のように定義される相対的奥行き誤差と呼ばれる指標を導入した.

$$h_k \propto \sum_{ij} \delta_k (|D_l^{ij} - D_s^{ij}| 0.01 D_\sigma^j) \quad (13)$$

D_l^{ij} と D_s^{ij} は視点 j の座標 i に対する奥行きの実値と推定値であり, D_σ^j は視点 j における奥行きの実値の標準偏差である. $\delta_k()$ は $|D_l^{ij} - D_s^{ij}|$ が $[k \times 0.01 D_\sigma^j, (k+1) \times 0.01 D_\sigma^j]$ に含まれる場合に 1 を返し, それ以外は 0 を返す. 得られた相対誤差のヒストグラム h_k を少なくとも 2 視点から観察された点の総数によって正規化する. ただし誤差が $k > 20$ の範囲に含まれるならば, その全ての

点を $k = 20$ に含めるものとする．さらにより視覚的に誤差を表現するためのエラーマップを生成する．エラーマップにおいては誤差に応じて色が青色 ($k = 0$) から赤色 ($k = 20$) へと変化する．緑色のエリアは真値が与えられていなかったり、2 視点以上から観察されていないといった理由で今回の評価において用いなかった点を表している．

図 6, 図 7 に実験結果を示す．それぞれ次の 4 つの条件に対してデプスマップを推定した結果である．条件は (1) 奥行き伝播とバンドル最適化の両方を適用, (2) 奥行き伝播のみを適用, (3) 高精度化を行わない, (4) Strecha ら [8] の手法を適用である．但し, Strecha らの手法に対しても, 我々の PDE に基づいたデプスマップ最適化とデプスマップの高精度化法の性能を評価するために, 提案手法によって同一のデプスマップの初期値を与えた．比較結果は本提案手法がデプスマップの精度において Strecha らの手法 [8] より大きく優れている事を示しており, 特に 2 段階の高精度化の寄与が大きい事が示されている．より詳細に結果を見ると, 奥行き伝播は比較的高いエラーレベルの領域 ($k = 20$ 付近) に強く寄与しており, 信頼度に基づくバンドル最適化は比較的低いエラーレベルの領域 ($k < 3$) において劇的に推定精度を向上させている．これらの結果は 2 段階の高精度化が互いを生かしながら機能している事を示唆している．参考として図 9 において *Herz-Jesu P8* データセットに対して本手法を適用した結果を示す．

8.3 屋内画像を用いた 3 次元モデル復元性能評価

次に, 提案手法を Middlebury データセット [1] に適用した結果を示す．ここでは 16 枚のワイドベースライン画像によって構成されている *dinoSparseRing* データセットを用いた．このデータセットは先述した Strecha ら [2] のデータセットと比較して, 全周画像であるため視点間の距離が非常に大きく, 物体表面にテクスチャが殆ど存在しないため復元が困難である．本節ではデプスマップ推定に留まらず, それらを用いた 3 次元モデル復元結果を示す．また大きく異なる点として, 屋外データセットにおいて一度しか施さなかった高精度化を 5 回繰り返し, 反復して適用する事の効果を示す．図 10 に高精度化を最初の 1 度だけ施した時点でのデプスマップと, 5 回繰り返した後の結果を, 最後の結果によって得られた 3 次元点群を図 11 に示し, それらを用いてメッシュを復元した結果を図 12 に示した．高精度化は明らかに複数回適用した方が優れた結果を得る事ができた．これは Li ら [10] と異なり, 入力と出力の双方がデプスマップである事が本手法の大きな利点である事を示唆している．

9. 結 論

本論文では, 信頼度に基づいたデプスマップの推定と融合に関するアルゴリズムの提案を行った．特徴ベース

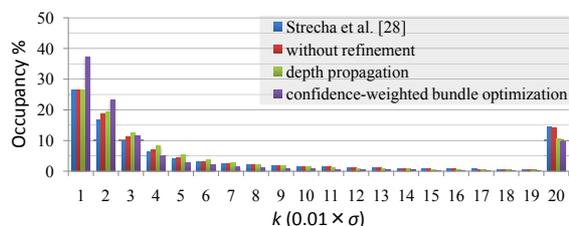


図 6 *fountain-P11* データセットに対するエラーヒストグラム．視点 { 3,4,5,6,7 } 番目の画像に対する結果を平均して表示している．

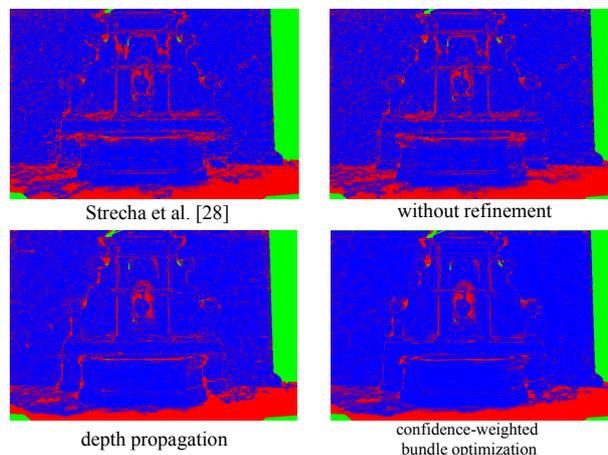


図 7 エラーマップの例．*fountain-P11* における 5 番目の画像に対する結果を示す．

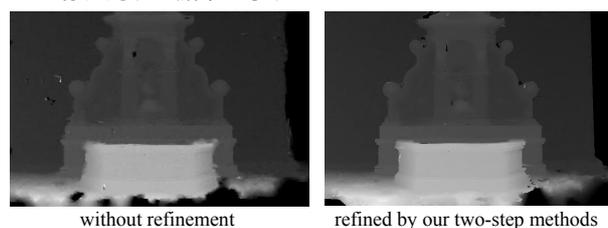


図 8 デプスマップの例．左側は高精度化を施していない場合, 右側は 2 段階の高精度化を適用した結果である．

の手法とエネルギー最適化ベースの手法を有効に組み合わせさせて密なデプスマップを推定し, さらにそれらを 2 段階の高精度化を施す事によって, 実験結果が示すように非常に高精度なデプスマップおよび 3 次元モデルを生成する事ができた．将来の課題としては, パラメータを自然画像を元にした機械学習によって推定したり, デプスマップから 3 次元モデル表面を生成するためにより適した方法を模索する事が挙げられる．

文 献

- [1] <http://vision.middlebury.edu/mview/>.
- [2] C. Strecha, W. V. Hansen, L. V. Gool, P. Fua and U. Thoennessen: "On benchmarking camera calibration and multi-view stereo for high resolution imagery", In CVPR (2008).
- [3] Seitz: "A comparison and evaluation of multi-view stereo reconstruction algorithms", In CVPR (2006).
- [4] S. Seitz and C. Dyer: "Photorealistic scene reconstruction by voxel coloring", IJCV, **35**, 2, pp. 151–173 (1999).

- [5] O. Faugeras and R. Keriven: “Variational principles, surface evolution, pde ’s, level set methods and the stereo problem”, IEEE TIP (1998).
- [6] S. Goesele, B. Curless and S. Seitz: “Multi-view stereo revisited”, In CVPR (2006).
- [7] Y. Liu, X. Cao, Q. Dai and W. Xu: “Continuous depth estimation for multi-view stereo”, In CVPR (2009).
- [8] C. Strecha, T. Tuytelaars and L. V. Gool: “Dense matching of multiple wide-baseline views”, In ICCV (2003).
- [9] M. F. Tappen and W. T. Freeman: “Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters”, In ICCV (2003).
- [10] J. Li, E. Li, Y. Chen, L. Xu and Y. Zhang: “Bundled depth-map merging for multi-view stereo”, In CVPR (2010).
- [11] Y. Furukawa and J. Ponce: “Accurate, dense, and robust multiview stereopsis”, In CVPR (2007).
- [12] M. Habbecke and L. Kobbelt: “A surface-growing approach to multi-view stereo reconstruction”, In CVPR (2007).
- [13] G. Zhang, J. Jia, W. Xiong, T. T. Wong, P. A. Heng and H. Bao: “Moving object extraction with a hand-held camera”, In ICCV (2007).
- [14] A. Ess, B. Leibe, K. Schindler and L. V. Gool: “A mobile vision system for robust multi-person tracking”, In ICCV (2007).
- [15] M. Lhuillier and L. Quan: “Match propagation for image-based modeling and rendering”, IEEE PAMI, **24**, 8, pp. 1140–1146 (2002).
- [16] Z. Wang and Z. Zheng: “A region based stereo matching algorithm using cooperative optimization”, In CVPR (2008).
- [17] M. Jancosek and T. Pajdla: “Segmentation based multi-view stereo”, Computer Vision Winter Workshop (2009).
- [18] M. Pollefeys, F. Verbiest and L. V. Gool: “Surviving dominant planes in uncalibrated structure and motion recovery”, In ECCV2002, Lecture Notes in Computer Science, **2351**, pp. 837–851 (2002).
- [19] M. Gong and Y. Yang: “Fast unambiguous stereo matching using reliability-based dynamic programming”, IEEE PAMI, **27**, 6, pp. 998–1003 (2005).
- [20] X. Hu and P. Mordohai: “Evaluation of stereo confidence indoors and outdoors”, In CVPR (2010).
- [21] D. Lowe: “Distance image features from scale-invariant keypoints”, International Journal of Computer Vision, **60**, 2, pp. 91–110 (2004).
- [22] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool: “Surf: Speeded up robust features.”, CVIU, **110**, 3, pp. 346–359 (2008).
- [23] E. Tola, V. Lepetit and P. Fua: “A fast local descriptor for dense matching”, In CVPR (2008).
- [24] D. Comaniciu and P. Meer: “Mean shift: A robust approach toward feature space analysis”, IEEE PAMI, **24**, 5, pp. 603–619 (2001).
- [25] J. Sun, Y. Li, S. Kang and H. Shum: “Symmetric stereo matching for occlusion handling”, In CVPR (2005).
- [26] M. Kazhdan, M. Bolitho and H. Hoppe: “Poisson surface reconstruction”, Eurographics on Geometry Processing, pp. 61–70 (2006).
- [27] M. I. A. Lourakis: “levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++” (Jul. 2004). Accessed on 31 Jan. 2005.

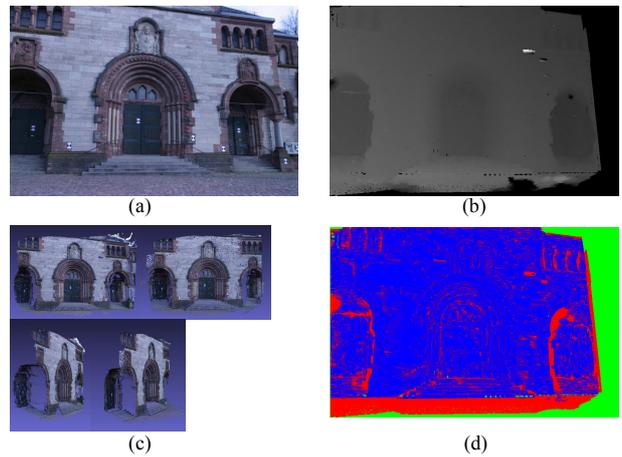


図9 Herz-Jesu-P11 データセットを用いた結果例。(a) 入力画像 (内 6 番目の画像を示した), (b) デプスマップ推定結果, (c) 3次元点群 (左が推定結果, 右が真値), (d) エラーマップ。

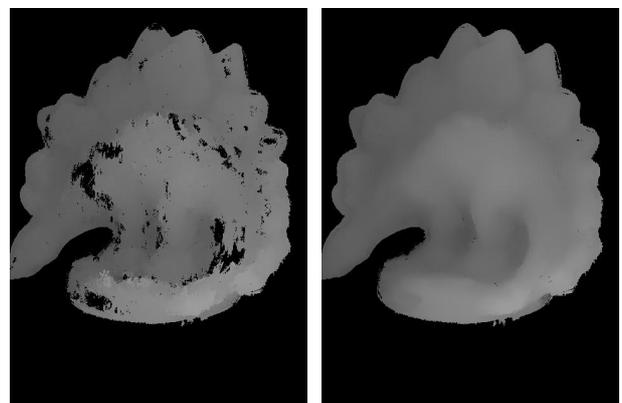


図10 デプスマップ推定結果の例。左図:高精度化を1度だけ適用した結果。右図:高精度化を5回繰り返した結果



図11 3次元点群復元結果の例 (頂点数はおよそ 30 万)

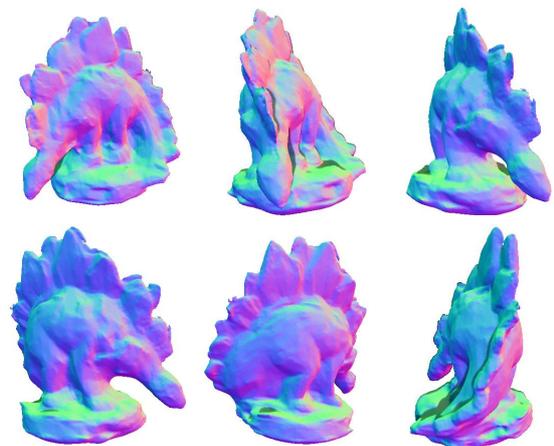


図12 3次元モデル復元結果