

対応という概念を使わないステレオビジョン

小栗 清^{1,a)} 柴田 裕一郎¹ 濱田 剛²

概要: ステレオカメラ画像に対して対応や視差という概念を使わずに、左右カメラの単一ピクセル同士の比較から直接、3次元空間上に表面存在仮説点を多数配置し、奥行き方向でどこかに表面が存在するはずだということ、表面はそんなに凸凹していないはずだということを使ってカメラに写っている対象の表面を検出する新しい手法を提案する。この手法では、対応という概念から派生するオクルージョンを考慮する必要がなく、また多視点画像からの3次元形状復元分野で使われるボクセルベースの手法のように過度に一般化しすぎておらず、将来のハードウェアによる直接実行との相性も良い。

1. はじめに

古くからのステレオビジョンの研究では、左右のカメラ画像の対応する点を見つければあとは三角測量の原理でその点の3次元位置が分かると考えてきた。対応という概念から視差やオクルージョンという概念が派生した。ステレオビジョンがステレオマッチングとも言われるのはこの対応を重視しているためである。対応を検出するために、領域ベースの手法では一方のカメラ画像のあまり大きくない領域と似た領域を他方のカメラ画像の中から探し、その領域の中心点同士を対応させる。領域の類似度の比較にはSAD(Sum of Absolute Difference)やSSD(Sum of square Difference)などが用いられた。また特徴点ベースの手法ではカメラ画像に対して特徴点をあらかじめ抽出しておいてから、その特徴量と位置関係から左右の特徴点同士を対応させる。特徴点の抽出手法としてはSIFT(Scale Invariant Feature)などが用いられた。しかしながら、領域ベースの手法も特徴点ベースの手法もカメラに投影された2次元の画像情報を処理の対象としている点では同じである。同じ対象を左右のカメラで撮影していても、左右のカメラの位置がわずかでも異なっていれば、2次元の画像情報は同じにはならない。この異なる画像情報に対して対応を求めようとしているところに困難があった。そこで左右のカメラを出来るだけ近づけて、すなわち2次元の画像情報の違いを出来るだけ小さくして、領域の比較を位相限定相関法(Phase Only Correlation)によりサブピクセル精度で行う青木らの手法[1], [2]も提案された。

また2次元の処理を行わずに左右のピクセルの対応をグ

ローバルに最適化するグラフカット[3]を使った石川らの手法[4]やDPマッチングを使った井関らの手法[5]が提案され、最近ではステレオマッチングは前者、すなわち2次元処理を行う局所マッチングと後者のグローバルマッチングの2つに分類されると言われるようになった[6]。

石川らの手法[4]などから始まった最近の性能向上が著しい多数画像を前提とした稠密ステレオビジョンの研究[6], [7], [8]では、直接3次元空間での処理を行うようになってきているが、視差や対応という概念が残っており、これが理論の枠組みに制約を与えているのではないだろうかという懸念がある。たとえばCVPR2008のWoodfordらの論文[7]において問題がどのように定義されているのかを見てみると以下のようになっている。

まず $N + 1$ 枚の入力画像 $\{I_i\}_{i=0}^N$ を考えたとき、2次元ベクトル \vec{x} が基準入力画像 I_0 上の位置を、 $I_0(\vec{x})$ がその位置の輝度値を、 $d = D(\vec{x})$ がその位置の視差(disparity)を表すものとする。 $\pi_i(\vec{x}, d)$ が基準入力画像上の位置 \vec{x} と視差 d から決定される3次元空間上の位置が入力画像 I_i のどの位置に対応するかを示す投影関数であるとする、 $I_i^\pi(\vec{x}, D(\vec{x})) = I_i(\pi_i(\vec{x}, D(\vec{x})))$ は入力画像 I_i の対応する点の輝度値となる。普通のステレオ視の場合は $N = 1$ で $\pi_1(\vec{x}, d) = \vec{x} + (d, 0)$ となる。ここから d はまさに視差であることが分かるが、 $N > 1$ で d がどのような意味を持つのかは簡単ではない。しかしながらここに矛盾がある訳ではない。問題はステレオ視をエネルギー関数を最小化するような視差 $D(\vec{x})$ を決定することだととらえている点である。この時、エネルギー関数は視差の関数として

$$E(D) = E_{photo}(D) + E_{smooth}(D)$$

$$E_{photo}(D) = \sum_{\vec{x}} \sum_{i=1}^N f(I_i^\pi(\vec{x}, D(\vec{x})) - I_0(\vec{x}), V_{\vec{x}}^i)$$

¹ 長崎大学工学部
² 長崎大学先端計算研究センター
a) oguri@cis.nagasaki-u.ac.jp

のようになる。ここで V_x^i はオクルージョンとなっているかどうかを表すフラグであり、 f は対応する位置が画像上に存在する（オクルージョンでない）場合にその輝度値が近いときに小さな値を持つような関数である。このエネルギー関数は視差と対応から出発しているために、オクルージョンを特別扱いしなければならない。また視差の滑らかさ $E_{smooth}(D)$ とはどのような物理的意味を持つのであろうか。

対応という考えを捨てて、最初から3次元の処理をするとうなるのだろうか。多視点画像からの3次元形状復元と呼ばれる研究領域では、視差という概念を使わずに問題を以下のように定義している。例えば、津田らの論文 [9] では、復元された3次元表面形状を S としたときに、 S 上の点 X を観測視点に投影して得られる輝度値の相違度 $\rho(X)$ の和が最小になる、すなわち

$$E(S) = \int_S \rho(X) dS$$

を最小にするような S を求める問題であると定式化している。具体的な計算をするために3次元空間をボクセルで分割し、そのボクセルが物体内にあるかどうかをグラフカットでラベル付けする久保らの研究 [10] や逆散乱問題として逆レイトレーシングを行うという Liu らの研究 [11] がある。

この問題設定は全く正しいが普遍化し過ぎてはいないだろうか。現実的な計算を行うために追加のアイデア、例えば視体積交差法などが必要となり理想から離れてしまっている。

そこで本論文ではあくまで左右画像からなるステレオビジョンに限定して、対応や視差という概念を排除した3次元空間上の問題設定とその処理手法を提案する。

2. 表面存在仮説点

外界の情報は光センサーが敷き詰められたカメラの撮像面に到着する。ここで光センサーの一個一個をピクセルと呼ぶ。カメラの配置や物理的仕様が分かれば、3次元空間の‘ある位置’から四方に発した光が、左カメラの‘あるピクセルの中心’と右カメラの‘あるピクセルの中心’に届くような位置を決めることが出来る場合がある。この位置を表面存在仮説点 (Hypothetical Surface Point) と呼ぶ。表面存在仮説点に鏡のようでない普通の乱反射する物体表面があれば、その位置からの光が到達する左右のピクセルには同じような輝度値が届くはずである。ある表面存在仮説点からの光が到達する左右のピクセルに全く違った輝度値が届いていればその表面存在仮説点には物体の表面がないということである。このように表面存在仮説点には左右のカメラのピクセルの輝度値からそこに表面がある尤もらしさ (尤度) を定義することが出来る。また左右のカメラのピクセルの輝度値の平均値をその表面存在仮説点の

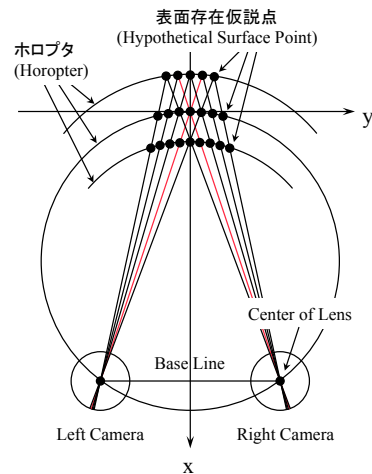


図 1 カメラの配置とホロプタ
Fig. 1 Camera Layout and Horopter

輝度値とする。

3. ホロプタ

表面存在仮説点に関する議論と計算を簡単にするためにカメラの配置や構造にいくつかの仮定を置く。

まず、図 1 のように左右のカメラを配置し、カメラの撮像面の中心に座標の原点から発した光が届くようにする。すなわち人と同じように左右のカメラは対象を注視するものとする。その意味で図 1 は人の頭部を上から見た図となっている。 $-x$ の方向を奥行き方向ということもある。撮像面がレンズの中心を中心とする球面上にあると仮定すると、左右のレンズの中心と原点を通るホロプタ (Horopter) と呼ばれる円の上から出た光は、同じ弧に対する円周角が等しくこの円周角が撮像面の中心角であるので、左右のカメラの撮像面の同じ位置に届く。従って表面存在仮説点はホロプタ上に存在する。

次に右カメラでは撮像面の中心から一つ右のピクセル、左カメラでは撮像面の中心から一つ左のピクセルに光が到達する表面存在仮説点を考えると、これは一つ小さい内ホロプタ上にある。また右カメラでは撮像面の中心から一つ左のピクセル、左カメラでは撮像面の中心から一つ右のピクセルに光が到達する表面存在仮説点は一つ大きい外ホロプタ上にある。これらのホロプタに対しても同じ弧に対する円周角が等しく、この円周角が撮像面の中心角であるという関係が成立するので、この内ホロプタと外ホロプタの上にも複数の表面存在仮説点がある。さらにより内側のホロプタや外側のホロプタを考えることにより多数の表面仮説点を置くことが出来る。

ここまでは、左右のカメラのレンズの中心と原点を含む平面上のことであったが、一つのホロプタが円柱をなしていると近似することにより上下方向にも多数の表面存在仮説点を置くことが出来る。

カメラの撮像面のピクセルだけからホロプタ上に表面存

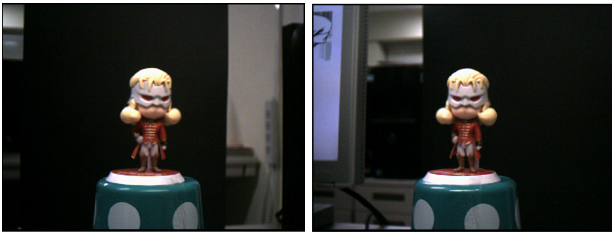


図 2 対象を注視した左カメラ画像と右カメラ画像カメラ
Fig. 2 Left Picture and Right Picture

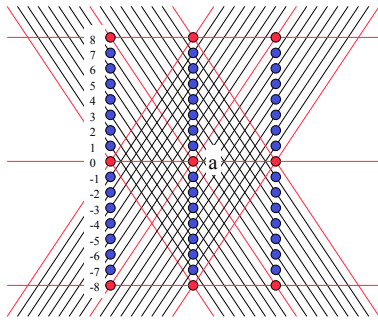


図 3 補間による表面存在仮説点の追加
Fig. 3 Insertion of Hypothetical Surface points by Interpolation

在仮説点をつくと図 1 から分かるように、ホロプタ間の表面存在仮説点の間隔はホロプタ上の表面存在仮説点の間隔の数倍にもなってしまふ。そこでピクセル間の輝度値を補間することにより、ホロプタの数を増やして 2 つの間隔がほぼ同じになるようにする。

図 2 は実験で用いたステレオ画像であるが、この画像では、対象までの距離と基線長からホロプタ間の表面存在仮説点の間隔はホロプタ上の表面存在仮説点の間隔のほぼ 8 倍となった。このため補間によってピクセルの数を 8 倍にしてホロプタの数を 8 倍にする必要があった。このときホロプタ上の表面存在仮説点は $\frac{1}{8}$ に間引きして使う。間引きしないとホロプタ上の表面存在仮説点の間隔まで $\frac{1}{8}$ になって意味がない。またこの補間がノイズの影響を受けないように入力画像は 1 ピクセルを σ とするガウス関数で前もって平滑化しておいた。また補間にはバイキュービック補間 (Bicubic Interpolation) を用いた。

図 3 は表面存在仮説点のこの追加の様子を示している。赤の点が元々の表面存在仮説点、青の点は補間により追加された表面存在仮説点である。

図 4 はこのようにして作られた多数の表面存在仮説点が 3D 空間上に配置される様子を示したものである。1 つの表面存在仮説点を球で表し、球の色は表面存在仮説点の輝度値、球の大きさは尤度に対応している。図 4 はほんの一部を示したにすぎず、実験では縦横 129 行 \times 129 列、奥行き 81 段、合計 1,347,921 個の表面存在仮説点を使用した。

人のように左右のカメラで対象を注視することにより得られるホロプタを使った計算の単純化はハードウェアでの

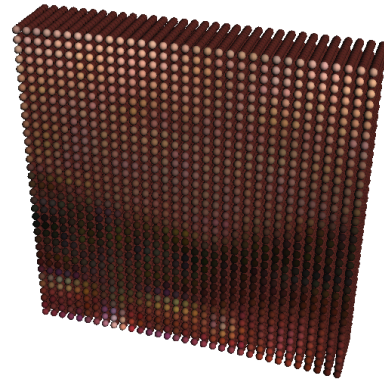


図 4 表面存在仮説点の立体配置
Fig. 4 Hypothetical Surface points in 3D Space

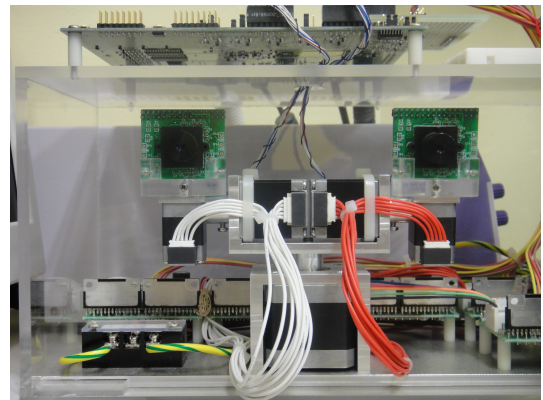


図 5 実験装置
Fig. 5 Experimental System

実装を行う場合に大きな意味を持つ。ピクセル単位で同期がとれた左右のカメラからのピクセルデータを直接比較すれば中心ホロプタの、ピクセル単位で遅延を加えたデータと比較すれば内ホロプタや外ホロプタの表面存在仮説点の尤度となる。すなわちリアルタイムでかつ小規模な回路で尤度計算が実行できる。これは画像のフレームバッファに対して処理を行うという従来の往復処理アーキテクチャから、高性能化のための今後の重要な計算アーキテクチャと考えられているストリーム処理 [12] への転換ができるということである。

図 5 は実験装置の外観である。5 台のモーター (首の回転、右カメラの上下回転、右カメラの左右回転、左カメラの上下回転、左カメラの左右回転) で向きが制御される左右のカメラは FPGA (Field Programmable Gate Array) からクロック単位で制御されて完全に同期がとれており、カメラからのピクセルデータは USB2.0 インタフェースにより PC にリアルタイム転送され尤度計算以降の処理は現状ではソフトウェアで行われている。

4. 表面の性質

要請 1 『真の表面はできるだけ尤度の高い表面存在仮説点の中から選ばれなくてはならない』けれども、左右の単

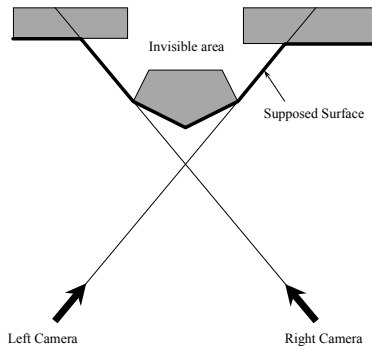


図 6 見えない領域
Fig. 6 Invisible Area

一ピクセル同士の輝度値の差の絶対値で定義される表面存在仮説点の尤度は全くわずかな違いでしかなく、これだけでは真の表面を決定することはできない。そこで表面はカメラからの奥行き方向でどこかにあるすなわち、要請2『奥行き方向でただ1つの表面存在仮説点がある』べきであるという要請と、表面は連続しているはずであるという要請を置く。孤立した物体のように実際は背景とは連続していない場合でも、それはいくつかの視点から見て分かるのであり、2つの視点からは図6に示すように孤立する原因となっている面は見えない。見えない部分では見えるぎりぎりまで表面は連続していることにすれば、要請3『表面は全体で連続かつ連結している』ことになる。見えない部分では見えるぎりぎりまで表面は連続しているということは図3において、aが真の表面であれば、その横は-8から8までのどれかが真の表面であることすなわち、要請4『見えるはずのない傾きの面は禁止される』ということである。図3は人の頭部を上から見た図となっているので人から見て左右方向の状況を示しているが、人から見て上下方向もほぼ同様に考えることができる。

しかしこれだけでは表面のなめらかさとも言うべき面としての性質を要請できていない。微分可能であるという意味でのなめらかさの前に、要請5『真の表面の奥行きがあまり凸凹しないこと』という要請が必要である。これは自然界に存在するものが持っている性質である。極端な凸凹は構造として弱いからであろう。次に微分可能という意味での滑らかさ、これは曲面の場合は寄っていくと平面に見えるということであるので、隣接する表面存在仮説点のうち平面に含まれる複数の表面存在仮説点の尤度が共に高いときは、そこにその向きの表面がある尤度がより高いと考えるべきであろう。これは、領域の類似度の比較で使われるSSDやSADを2次元の画像上ではなく、3次元の物体表面上で行っていることに相当する。

この要請は今回は、図7に示す8通りの傾きの向きと図3に示す8通りの傾きの強度を考えて得られる64通りの傾きに傾きがない場合を加えた合計65通りの平面に対してその面に含まれる9個の表面存在仮説点の尤度の平均

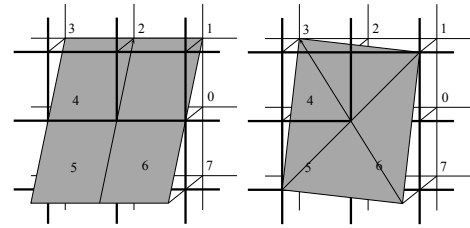


図 7 傾きの方向
Fig. 7 Direction of Gradient

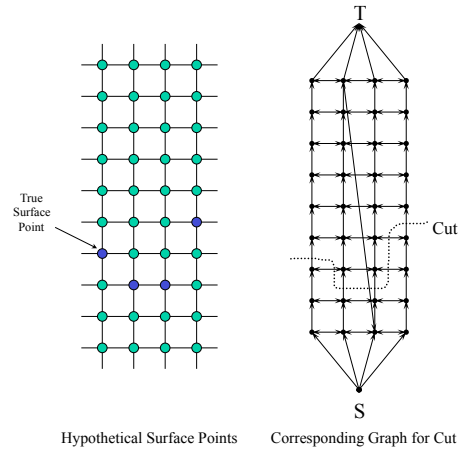


図 8 表面存在仮説点と対応するグラフ
Fig. 8 Corresponding Graph Cut to True Surface

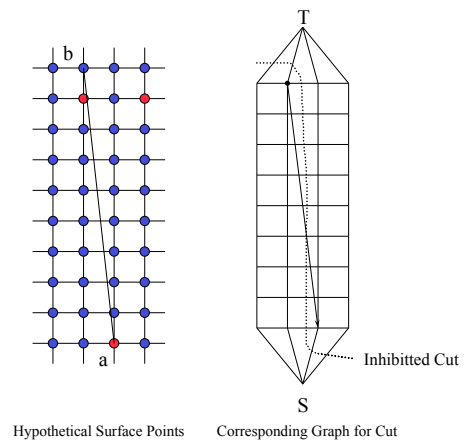


図 9 鋭角の禁止
Fig. 9 Inhibit Too Sharp Angle

を計算してその中で最も大きいものを、中心にある表面存在仮説点の新たな尤度とすることで反映させた。なお、隣接する表面存在仮説点は8個あるので自分を含めて9個となる。また図7は奥行き方向に見た図となっており、左は方向2の傾き、右は方向3の傾きを示している。これに対して図3は上から見た図となっている。

以上の要請は視差が滑らかであるという要請に比べて物理的意味が明白である。

5. グラフカットによる表面抽出

従来はピクセル毎に定義される視差の全体（すなわち視

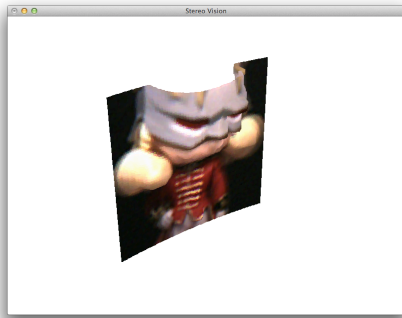


図 10 抽出された表面 左上から
Fig. 10 Extracted Surface

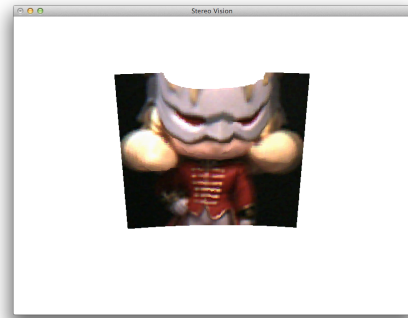


図 11 抽出された表面 上方から
Fig. 11 Extracted Surface

差マップ) に対して定義されるあるエネルギー関数を最小化する, そのような視差マップをグラフカットなどを使って求めていた. そこで用いられているグラフはカットの位置が視差 (ラベル) に対応する. すなわちラベル付けの最適化として捉えられていた. ここではグラフカットをラベル付けではなく, まさに面の切り出しに使用する. エネルギー関数も定義しない. 図 8 の左は表面存在仮説点群を上から見たものである. 図の下が人の目に近い方, 上が遠くの方を表す. 図 8 の右はグラフカットを適用するグラフである. ソースからシンクに向かう方向のエッジが表面存在仮説点に対応する. 一番下の表面存在仮説点はソースからのエッジに対応する. 一番上の表面存在仮説点はシンクへのエッジに対応する. 表面存在仮説点の中から真の表面を選ぶことはまさにグラフをある面でソース側とシンク側に分けることに対応する. 出来るだけ尤度の高い表面存在仮説点を選ぶべきであり, また表面の性質を出来るだけ満足しなければならない. ある面でソース側とシンク側に分けることで要請 3 は自動的に満足される. 次にソースからシンクに向かう方向のエッジに表面存在仮説点の尤度の大小を反転させたコストを割り付ける. これで尤度の高い表面存在仮説点に対応するエッジがカットで選ばれるようになり要請 1 が反映される. 次に図 9 右に示す斜めのエッジ (向きに注意すること) に非常に大きなコストを割り付けることで要請 4 が満足される. なぜならばこのエッジは図 9 左の傾きが大きすぎる面の切り出し a-b を禁止する働きをするからである. またこの結果回り込むような面は出現しなくなり, 奥行き方向でただ一つの面でカットされるようになる. すなわち要請 2 が満足される. 次に図 8 右のソースからシンクへの方向に垂直なエッジには両方向に対して小さなコストを与える. このコストにより激しい凸凹が押さえられ要請 5 が満足される.

6. 実験結果

実験は OpenGL の環境を使い, カメラの撮像面のピクセルや表面存在仮説点が実寸で 3D 表示されるようにして, 視

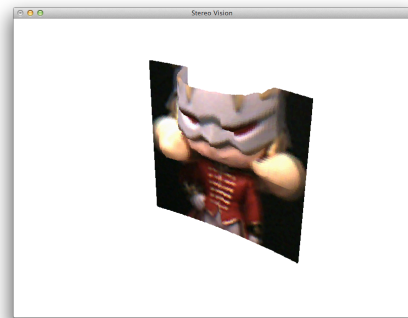


図 12 抽出された表面 右上から
Fig. 12 Extracted Surface



図 13 実際の写真 右上から
Fig. 13 Real Picture

点を変えて任意の方向から眺められるようにした. グラフカットは <http://pub.ist.ac.at/~vnk/software.html> からダウンロードできる Kolmogorov[13] の maxflow-v3.02.src.tar.gz を使用した. 図 10, 図 11, 図 12 はグラフカットにより真の表面となった表面存在仮説点のみを視点を変えて表示させた結果である. 図 13 は同じような視点から対象を撮影したものである. もともと表面存在仮説点は輝度値を持っているので, 画像を貼付けるなどの特別な処理をしなくても自然な物体として見ることが出来る. 視点を近づけすぎると表面存在仮説点の 1 つ 1 つが球となって現れるが, 視点を離せば自然なテクスチャーとなる. 背景部分は左右ともに黒で尤度差が

ないため、できるだけ凸凹がない方がよいという要請5と全体が連結した面であるはずであるという要請3,そして見えるはずのない傾きの面は禁止されるという要請4から物体の後ろにつながった平面となっている。目視では現物とほぼ同じ形が復元されているように見えるが、定量的な評価は今後の課題である。2.66GHz Intel Core i5 CPUと4GB MEMを使用して初期化に5秒、グラフカットに3秒ほどかかる。

7. 残された課題

今回はグラフカットにより大域的な最適化を計算したが、局所的な計算でどれほどの表面が抽出できるのかは非常に興味があるところである。リアルタイムでの処理を行うには現状のグラフカットでは少し時間がかかりすぎる。また、今回は尤度が高いという情報のみを利用しているが、尤度が低いという情報は確信を持ってそこに表面がないということなので、これを利用しないで手はないと思われる。また激しい凸凹を押さえるためのエッジのコストは正対平面を優先するように働くので、その影響がどれくらいか、あるいはこれを避けるグラフの構造を検討する必要がある。さらにカメラが動いて次々と入力されるステレオ画像からより実際に近い形状認識が行える手法の研究が必要である。

8. むすび

ステレオビジョンの新しい問題設定を提案した。対応や視差、オクルージョンなどの概念を排除して、最初から3次元空間に表面存在仮説点を多数配置して、物理的な意味を考慮しながら真の物体表面となる表面存在仮説点を選び出すというアプローチをとった。このために人と同じように対象を両目で注視するカメラ配置を採用し、グラフカットはエネルギー関数の最小化ではなく、物理的な面の切り出しとして使用した。従来研究との定量的な比較はこれからであるが、ボクセルベースのより一般化された3次元形状復元に関する研究よりもステレオビジョンに特化した手法として価値があるものと考えられる。

参考文献

- [1] 柴原琢磨, 沼徳 仁, 長嶋 聖, 青木孝文, 中島 寛, 小林孝次: 一次元位相限定相関法に基づくステレオ画像の高精度サブピクセル対応付け手法, 電子情報通信学会論文誌 D, Vol. J91-D, No. 9, pp. 2343-2356 (2008).
- [2] 青木孝文, 伊藤康一, 柴原琢磨, 長嶋 聖: 位相限定相関法に基づく高精度マシンビジョン, *Fundamentals Review*, Vol. 1, No. 1, pp. 30-40 (2007).
- [3] 石川 博: グラフカット, CVIM 研究会チュートリアル (2007).
- [4] Ishikawa, H. and Geiger, D.: Occlusions, Discontinuities, and Epipolar Lines in Stereo, *Proceedings of Fifth European Conference on Computer Vision (ECCV '98)*, pp. 232-248 (1998).

- [5] 井関健太, 矢口勇一, 岡 隆一: 画像間のピクセル最適対応を用いた小数枚数画像による3次元形状復元, CVIM 研究報告 (2008).
- [6] Smith, B. M., Zhang, L. and Jin, H.: Stereo Matching with Nonparametric Smoothness Priors in Feature Space, *CVPR2009* (2009).
- [7] Woodford, O. J., Torr, P. H. S., Reid, I. D. and Fitzgibbon, A. W.: Global Stereo Reconstruction under Second Order Smoothness Priors, *CVPR2008* (2008).
- [8] Bleyer, M., Rother, C., Kohli, P., Scharstein, D. and Sinha, S.: Object Stereo - Joint Stereo Matching and Object Segmentation, *CVPR2011* (2011).
- [9] 津田佳行, 延原章平, 松山隆司: 連結性を考慮したグラフカットによる多視点画像からの3次元形状復元, 画像の認識・理解シンポジウム (MIRU2009) (2009).
- [10] 久保健介, 富山仁博, 片山美和, 岩館祐一: グラフカットを用いた多視点映像からの3次元形状復元法の検討, 映像情報メディア学会冬季大会 (2008).
- [11] Liu, S. and Cooper, D. B.: A Complete Statistical Inverse Ray Tracing Approach to Multi-View Stereo, *CVPR2011* (2011).
- [12] 濱田 剛, 横田理央, 似鳥啓吾, 成見 哲, 泰岡顕治, 泰地真弘人, 小栗 清: 42 TFlops Hierarchical N-body Simulations on GPUs with Applications in both Astrophysics and Turbulence, *Super Computing 2009 (SC09)* (2009).
- [13] Kolmogorov, V. and Zabih, R.: What Energy Functions Can Be Minimized via Graph Cuts?, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp. 147-159 (2004).