

2 回逆投影法による複数画像からの3次元情報の抽出

川戸 慎二郎

ATR人間情報通信研究所

本報告では、多数の視点からの画像を用いて、画像間の対応点探索を必要としない3次元情報抽出法を提案する。カメラの位置と姿勢が既知ならば、画像にあらわれた特徴点とレンズ中心を結ぶ直線を3次元空間中に引くことができ、その直線は物体上の特徴点を通過する。画像間での特徴点の対応関係が不明であっても、多数の視点からこのような直線を引き、3次元空間中の各点でその交差状況を評価すれば各特徴点の位置を抽出できることが予想される。先行研究と比較しつつその考え方を述べる。アルゴリズムを検証するため、視点を移動するのではなく回転ステージに載せた物体を一定角度で回転しつつ斜め上方から撮像した画像を用いて実験を行った。

Double Back Projection Method to Extract 3-D Information from Multiple Images

Shinjiro KAWATO

ATR Human Information Processing Research Laboratories

Seika-cho, Soraku-gun, Kyoto 619-02, JAPAN

e-mail:skawato@atr-hr.atr.co.jp

In this paper we propose a method to extract 3-D information from multiple images of different view points, avoiding the corresponding problem as in usual stereopsis. A characteristic point in the image and the lens center determines a back projection line in the 3-D space. The back projection lines from different images may cross in the space. Examining the crossing counts, we can extract the 3-D points appeared in the images. Comparing the previous researches, our new ideas are described. To demonstrate the idea, experimental results are also described.

1 はじめに

2次元画像から3次元の情報抽出するさまざまな方法が提案されている [1] [8]。それらはスリット照明やコード化照明などアクティブな照明を用いる方法と、そうでない方法に分けられるが、後者はさらに、シングル画像によるものと複数画像を利用するものに分けられる。

複数画像を利用する最もシンプルな方法は2眼立体視法であるが、画像間での対応付けの問題が難しく、まだ十分に解決されているとはいえない。誤対応を避けるために3眼立体視法が提案されているが [7]、本質的な問題はかわらない。

これに対し対応点の探索過程が不要になるとして、エビポーラ平面画像解析法と称される、カメラを直線移動させながら連続的に撮像した画像列（時空間画像）から3次元の情報抽出する方法が提案されている [2] [3] [12]。この手法によれば対応点探索問題はより容易な直線抽出問題に置き換えられる。

しかし浜野らは、カメラ運動が直線運動に限定されることは著しい制約であり、かつ直線運動だけでは効率よく視差を大きくできない、またエビポーラ平面画像に現れる直線の抽出もオクルージョンやノイズによって頻繁に分断されているためそれほど容易でないとし、時空間画像から3次元のボクセル空間に重みつきポーティングをおこなうという方法を提案している [5] [6]。これによれば任意のカメラ運動が可能で、かつ画像間での対応点探索や特徴点追跡を全く必要としない。

本研究は浜野らとは独立になされたものであるが、画像をボクセル空間に逆投影するという意味では類似の手法を提案するものである。ただし、浜野らが仮定している「画像の撮影頻度は、時空間画像中で連続な特徴点軌跡が形成されるのに十分なほど高い」 [5] という条件は必要としない。また2回逆投影を行うことにより、浜野らが誤対応点抑制のために工夫した時空間トンネル画像 [5] や、ボクセル空間から集積点を検出するための複雑な計算 [6] を必要としない。そのため直感的に理解し易いアルゴリズムとなっている。

実験環境としては、回転ステージに載せた物体を斜め上方から撮像するように1台のカメラを設定し、ステージを一定の角度で回転させた複数画像から、その

物体の3次元情報を抽出することを試みた。

2 画像の逆投影による3次元情報の抽出

カメラの位置と姿勢が既知ならば、画像にあらわれた特徴点とレンズ中心を結ぶ直線（以下、逆投影直線と呼ぶ）を3次元空間中に引くことができ、その直線は物体上の特徴点を通過する。複数の位置から同じ特徴点を観察すれば観察数に応じた逆投影直線が得られるが、それらは物体上の特徴点の位置で交差する（図1）。したがってこの交差点が抽出できれば特徴点の3次元位置が抽出できたことになる。画像間での特徴点の対応関係が不明であっても、画像に現れたすべての特徴点についてこのような逆投影直線を引き、その交差回数を3次元空間中の各点で評価すれば各特徴点を抽出できることが予想される。

浜野ら [5] は観察空間を微小なボクセルに分割し、各ボクセルにおいて上に述べた逆投影直線が何回通過するかをカウント（ポーティング）し、そのカウント値が局所的なピークをなすボクセルを抽出するという「単純なポーティングアルゴリズム」（以下H1アルゴリズムと呼ぶ）をまず示し、これでは偽の局所ピークが生じる可能性があるとして、「時空間トンネル画像による重み付きポーティングアルゴリズム」（以下H2アルゴリズムと呼ぶ）なるものを提案している。これは画像に現れた特徴点の8近傍（時空間画像の場合26近傍）の画素で他の特徴点となっていないものについて、その画素が逆投影されるボクセルには-1のカウントを行うもので、これによって、偽の局所ピー

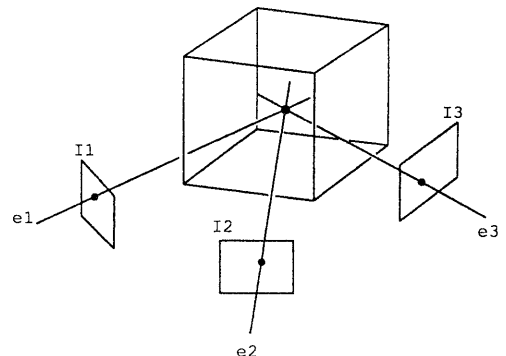


図1: 複数視点からの特徴点の観測

クの発生が抑制されるとしている。

ここで、H1 アルゴリズムの問題点をもう一度見直してみると、

(Q1) 一つの特徴点に対してボクセルのカウンタ値が広い範囲に分布し、別の特徴点による分布と重なって偽の局所ピークが発生することがある。

(Q2) ボクセルサイズが大きいと同じ画像の隣接する2本の逆投影直線が同じボクセルと交差することがあるが、これをどうカウントするか不明瞭である。

(Q3) オクルージョンが生じるとき、真の特徴点を含むボクセルであってもカウンタ値に差があり、(Q1)の問題と関連して簡単な固定閾値処理ではそれを抽出できない。

の3点があげられる。

(Q1) の問題点は図2で説明できる。(a) は近接した2つの特徴点を十分離れた、角度が5度づつことなる11の位置から観測して得た画像から逆投影直線を引いたようすを示したものである(特徴点によって実線と破線の区別をしている)。小四角形はボクセルを表し、(b) は各ボクセルを通過する直線のカウンタ値を表している。この図からわかるように真の特徴点位置を代表するボクセルの周囲になだらかなす野を引くようにカウンタ値が広く分布する。そして近接する2つの特徴点は互いに干渉しあって偽の局所ピーク(カウンタ値8のボクセル)が発生するのがわかる。

そこで浜野らは(Q1)の問題点を解決するとしてH2 アルゴリズムを提唱しているわけであるが、H2 アルゴリズムが有効であるためには時空間画像において特徴点の軌跡が連続になるように高頻度に撮像した画像が必要で[5]、処理すべき画像数が非常に多くなり、計算

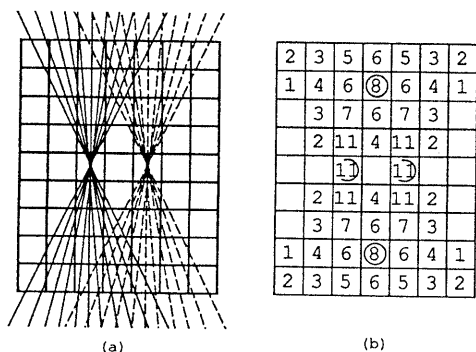


図2: H1 アルゴリズムと偽ピーク

量も膨大になる。逆にいえば、離散的に撮像した画像に対しては適用できない。

また、浜野らは明示的に述べていないが、H2 アルゴリズムの前提として画像分解能よりボクセル分解能のほうを相対的に高くしておく必要があり、このことよって(Q2)の問題を暗黙のうちに避けている。

実際、図2の直線に隣接するように両側に-1をポーティングする直線を引いてみると、すべてのボクセルにおいてカウンタ値がマイナスになってしまふ。したがってH2 アルゴリズムが有効に働くには、隣接画素によって空間に引かれる2本の逆投影直線が同じボクセルと交差することがないほどにボクセルのサイズが小さくなくてはならない。これは画像分解能以上に空間分解能を考えることを意味しており、むだがおおいと思われる。ボクセルサイズが不適切に小さいと画像上ではエッジのように連なった特徴点が3次元空間では途切れ途切れのボクセルとして抽出される可能性もある。浜野らが提唱している錐体型視線のポーティング[6]は、この問題に対応するために考え出されたものと思われるが、これはさらに計算負荷を重くしている。

(Q3)の問題はH2 アルゴリズムによれば閾値をゼロにすれば良く、解決されている

3 2回逆投影法による3次元情報の抽出

観察空間を微小なボクセルに分割して、各ボクセルに物体の特徴点が含まれるかどうかの証拠を複数の画像から拾い出すことを考える。H1 アルゴリズムにおいて、画像にあらわれた特徴点の逆投影直線と交差するボクセルに+1のカウントをする意味は、その画像からそのボクセルに対して特徴点が含まれる可能性の一つの証言があったことである。それでは同じ画像にあらわれた近くの特徴点から引かれた逆投影直線が同じボクセルと交差したとき、証拠が増えたかというところではなく、これはボクセル分解能より画像分解能が高かったことを意味しているにすぎない。したがってこの場合さらに+1のカウントをしてはならない。

このことは画像に特徴が線分であらわれたとき、その線分をレンズ中心から空間へ逆投影したときにできる面と交差するボクセルに平等に+1カウントすることであると考えると理解し易い。図3は画像面の線分

を逆投影してボクセルと交差する様子を示したものであるが（見やすくするためボクセル上面で交差するもののみ表示）、これらのボクセルに平等に+1カウントするのである。このことによって画像分解能がボクセル分解能より大きい場合の問題はなくなる。

そこでH1アルゴリズムをすこしいかえて、すべての画像にあらわれたすべての特徴点の逆投影直線を引き、各ボクセルにおいては幾つの画像からの逆投影直線が交差するかをカウントすることにする。これによって得られた各ボクセルのカウント値をK1カウントと呼ぶ。

これで(Q2)の問題は解決するが、(Q1)(Q3)の問題が残っている。そこでもう一度画像にたちかえて、画像にあらわれた特徴点の由来を考えてみると、それはその画素の逆投影直線上に並んだボクセルのうちのどれか1つにある。どのボクセルかといえば、それはK1カウントが最高のボクセルであると考えるのが合理的であろう。

そこで、再度すべての画像にあらわれたすべての特徴点の逆投影直線を引き、その直線と交差するボクセルのなかで最大のK1カウントの値を持っているボクセルにマークをつけ、各ボクセルにおいて幾つの画像からの直線によってこのマークがつけられたかをカウントすることにする。このカウント値をK2カウントと呼ぶ。

H1アルゴリズムによって偽の局所ピークが発生した図2の例で説明すると、K1カウントは(b)と同じになる。しかし、K2カウントはK1カウント11のボクセルのみK2カウント11として残り、ほかのボクセルのK2カウントはゼロになる。したがって(Q1)の偽の局所ピーク発生の問題は解決される。

(Q3)の問題であるが、K1カウントはそのボクセルを通過する逆投影直線のもとになった特徴点が何枚の画像にあらわれたかを示しており、K2カウントはK1カウントを超えることはない。オクルージョンがあればk1カウントはそれだけ小さくなり、ノイズや演算誤差のない理想的な場合には特徴点を含むボクセルにおいてK2カウントはK1カウントに一致する。したがってK2カウントをK1カウントで正規化したボクセル評価値

$$p = (K2 \text{ カウント}) / (K1 \text{ カウント})$$

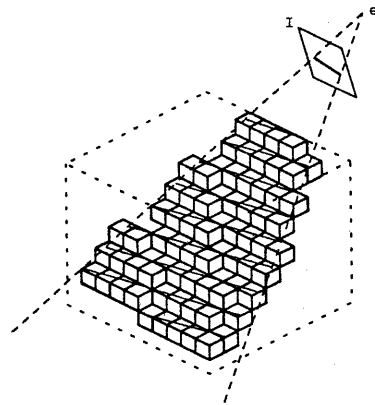


図3: 線分のボクセル空間へ逆投影

で各ボクセルを評価するようにすれば(Q3)の問題は解決されるはずである。

以上をまとめて、

- Step1: 全画像の特徴点からその逆投影直線を引く
- Step2: 各ボクセルの K1 カウントを計算する
- Step3: 各ボクセルの K2 カウントを計算する
- Step4: 各ボクセルのボクセル評価値 p を計算する
- Step5: 閾値以上の評価値 p のボクセルを抽出する

によって、3次元情報が抽出できる。これを2回逆投影法（略してDBP法：Double Back Projection Method）と呼ぶことにする。

4 実験

実画像を用いてDBP法の実験をおこなった。

カメラが移動する代わりに回転テーブルを用い、カメラは適当にテーブルを見おろす角度で設置し（キャリブレーション結果で26.5度と判明）、レンズはズームレンズを用いて物体が適当な大きさに写るように調節した。物体は立方体とし回転テーブルのほぼ中央において、20度づつ角度の異なる18枚の画像を撮像した。一枚の画像は640×480の画像データとなったが、メモリ節約のためその中央部分400×300を切り出して処理した。切り出された画像に対してエッジ抽出オペレータを施し、抽出されたエッジ点を特徴点とした。このようにして得られた画像が図4である。側面が暗い色だったので、輪郭になる縦の稜線はでているが、両側の面が見えている中央の稜線はエッジとして抽出できていない。

図5はK1カウントの値がどのような分布をしているかを示そうとしたもので、見やすくするため、図4の画像から1頂点でつながっている3つの稜線以外の部分を人為的に消して処理した結果である。ボクセルは図示されている大きな立方体を $64 \times 64 \times 64$ に分割したものである。ボクセルの形が視認できるように64分割でとめた。

(a)はK1カウントが17以上のボクセルを抽出したもので、上面に属する2つの稜線は抽出されているが、縦の稜線はまだあらわれていない。これは当然のことで、前者の2つの稜線は18枚のすべての画像で見えているが、後者の稜線は図4の1、2、3、8、9、10、11、18の8枚の画像でしか見えていないからである。

(b)は縦の稜線まで抽出されるように閾値をさげて、K1カウントが6以上のボクセルを抽出したものである。閾値を下げたことによって他の2つの稜線は逆にその形状が不明確になっている。

(c)はK1カウントが2以上のボクセルを抽出したものである。

(d)は比較のために、同じ画像、同じボクセルサイズに対してDBP法を適用して $p > 0.33$ 以上のボクセルを抽出したものである。0.33は画像の3枚に1枚以上の割合でここが特徴点に見えたという程度の意味

である。 p に対する固定の閾値で3つの稜線が抽出されていることがわかる。

図6は図4の18枚の画像に対してDBP法を適用した結果である。ボクセルは図5と同じ立方体を $256 \times 256 \times 256$ に分割したもので、 $p > 0$ のボクセル、すなわちK2カウントが1回でもあったものを抽出して異なった角度からの図を表示した。対応点ができるように1頂点にAの記号を付してある。

この図からわかるように立方体の形状が良く抽出されている。ただし、底面に属する稜線は元の画像にも良くあらわれていないこともあって抽出できていない。

また4本の縦方向の稜線はいずれも2重の稜線のように抽出されている。この理由については未解明であるが、カメラキャリブレーションの誤差に原因するものか、あるいは図4にあらわれたエッジが右の面から見たものと左の面から見たものと異なる位置に出ているのではないかと思われる。図5に示したボクセル分解能のレベルではこの現象は隠されている。ちなみに図6を得た処理結果において、K1カウントの最大値は16であった。上面の4つの稜線は18枚の画像すべてにあらわれていることを考えると、画像分解能にくらべてボクセル分解能が高すぎるといえるかも知れない。

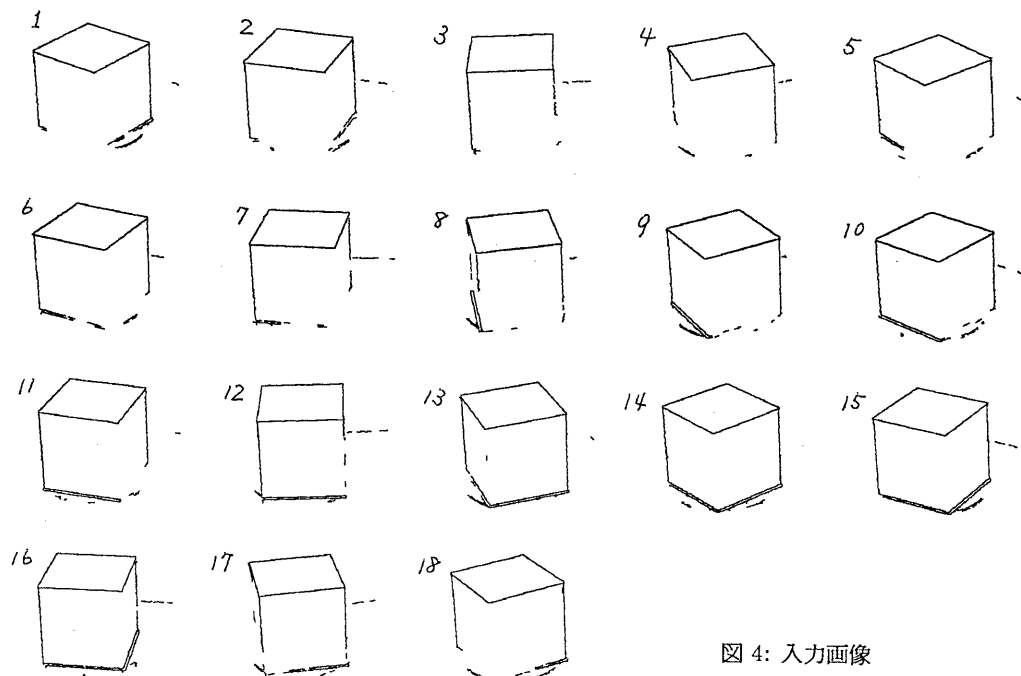


図4: 入力画像

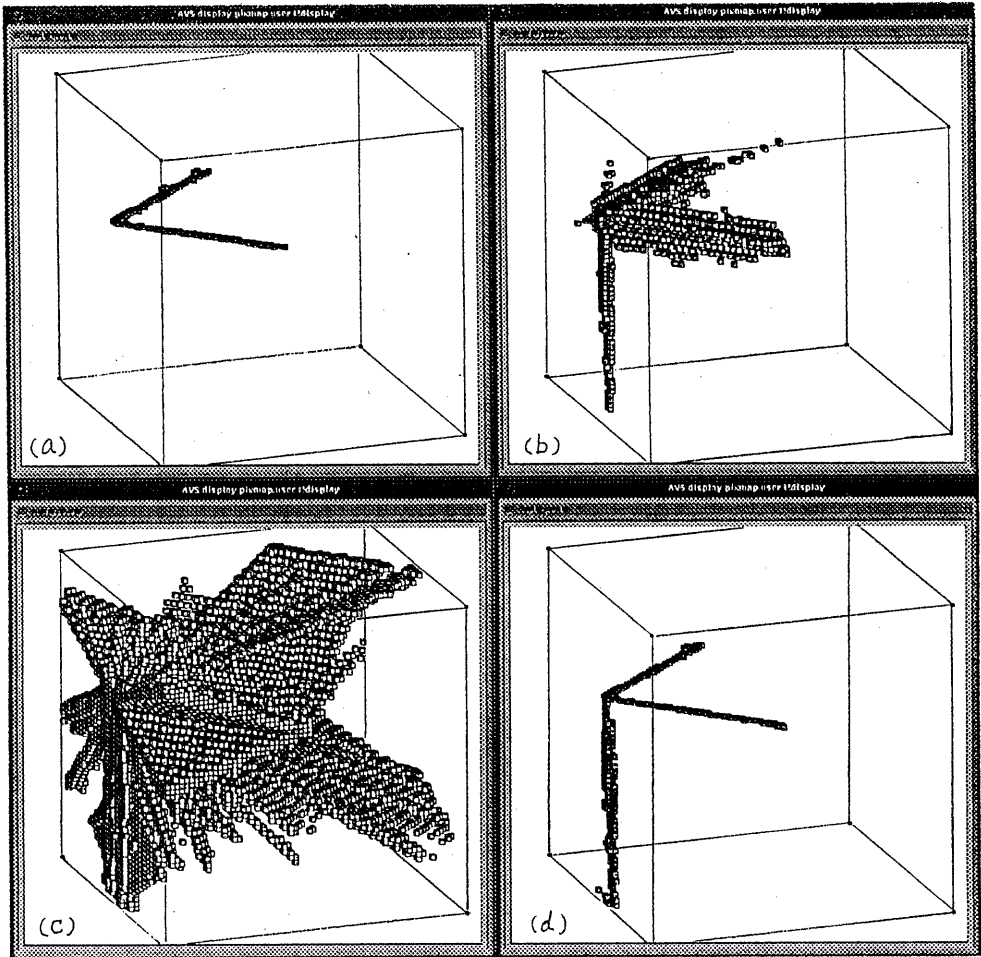


図 5: K 1 カウントの分布例

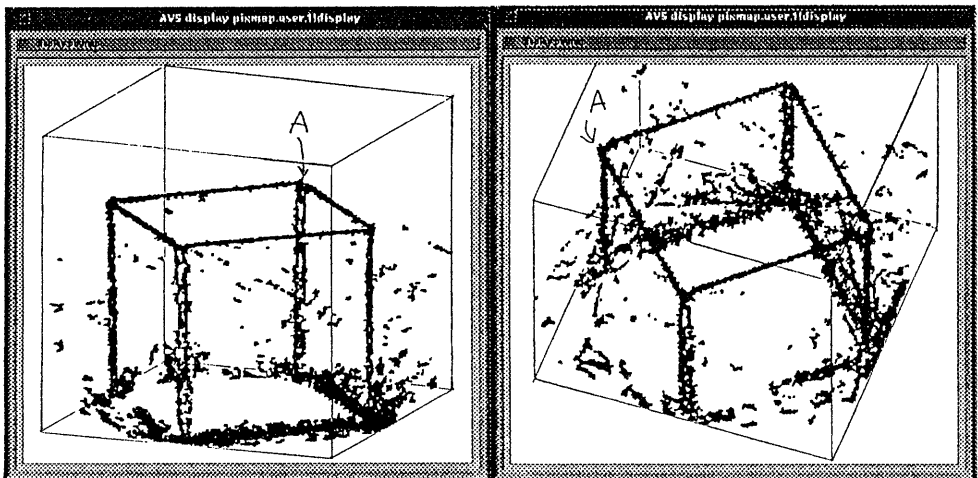


図 6: DBP アルゴリズムの結果例

5 他の研究との関連

(1) Shape from Rotation

回転する物体を観測してその形状を復元しようとする研究はすでいくつか報告されている。Szeliski は本研究と同様の実験環境で物体のシルエット像を利用し、オクトツリーのデータ構造を用いて空間中に物体像をいわば彫り出していくようなアルゴリズムを示している [10]。また Zheng は人物の 3 次元顔モデル生成のため、回転軸とカメラ光軸が直交して画像も正射影を仮定しているが、やはりシルエット像を利用して人物の胸像を復元している [14]。シルエット像を利用したときの欠点はシルエット像にあらわれない凹部の情報が得られないことで、そのため Szeliski は同じ実験環境でオプティカルフローを利用するアルゴリズムも実験している [11]。しかしオプティカルフローを精度良く求めることは困難で、その結果は余りかばしくない。本研究は特徴点の幾何学的位置を利用しようとするもので、これらの研究とは別のアプローチをめざしている。

(2) 複数画像データの融合

Quinio は、画像による計測エラーには、カメラパラメータなどの imprecision によるものと、観測プロセスにおけるノイズなどに起因する uncertainty によるものがあるとし、その両方を統一的に扱う数学的手段として RACS 理論 (Random Closed Set Theory) を適用することを提唱している [9]。筆者は RACS 理論を深く理解するところではないが、示された実験内容は次のようなものである。ある視点からステレオペアの画像をとり特徴点の視差を計算する。画素に大きさをもたせ、カメラパラメータにもある幅をもたせる (imprecision) と各視差から特徴点の存在する可能性のある多面体が空間中に決定される。そこでその中のボクセルに観測が一度あったとして Hit Count を +1 インクリメントする。一方共通視野全体のボクセルには Visibility Count を +1 インクリメントする。画像に特徴点が出たり出なかったり、あるいは誤対応を生じたりするのは確率的であると考え。そして多数の視点からの観測を融合したとき、RACS 理論に従えば、各ボクセルに特徴点が存在する確率は (hit count)/(visibility count) で計算されるとして、サンプル画像にたいして得られた確率の分布状態を示している。

本研究に当てはめて考えてみると、ステレオ画像かシングル画像かの違いはあるが、多面体は直線に、Hit Count は K1 カウントに相当し、Visibility Count は処理画像枚数に対応する。そして特徴点が存在する確率の分布は図 5(a)(b)(c) に示したものになる (ただし図 5 は正規化がなされていない)。しかし K2 カウントに相当するものは考えられない。その点で Quinio が計算している確率と、DBP 法で計算するボクセル評価値 p とは全く異なったものである。

(3) ハフ変換

ボクセル空間がハフ変換のパラメータ空間に対応すると考えると、K1 カウントをカウントしていくプロセスはハフ変換に類似している。ハフ変換においてもパラメータ空間に投票したあとその中のどの点を抽出するかの判断は難しいものがあり、その点で K2 カウントに相当する手順を追加することが考えられる。

Gerig はすでにこのことに気づいていて、ハフ変換のあと、Backmapping という処理を追加して、2 次元形状の抽出・認識に非常に有効だという例を示している [4]。Gerig のいう Backmapping はハフ変換のあと、実空間の各特徴点に対してパラメータ空間に引かれる曲線上で最大投票値を持つ点を関係づけて、パラメータ空間から実空間へのマッピングデータを持つというもので、DBP 法は逆のマッピングデータをもつというものではないが、その考え方は似ている。ただ 2 次元データを対象とする限り、DBP 法で考慮しているセルフオクルージョンは発生することがなく、その点で DBP 法は Gerig の Backmapping の考え方を少しだけ拡張したものになっている。

6 あとがき

多数の観察点からの画像を空間に 2 回逆投影することにより 3 次元情報を抽出する方法を提案し、実画像を用いた実験例をしめした。

課題を Shape from Rotation としてみたときの他の研究との関連、複数画像の融合と見たときの類似研究との関連、処理内容をハフ変換アルゴリズムと見たときの類似研究との関連についても述べた。

画像にあらわれた特徴点とレンズ中心を結ぶ直線 (逆投影直線) を空間に引き、何本の直線がボクセルと交差するかでなく、何枚の画像からの直線がボクセルと

交差するかをカウントすることにより、画像分解能とボクセル分解能の差の問題を解消できた。ボクセル分解能を画像分解能より高くすることはあまり意味がないと思われる。具体的アルゴリズムは示さなかったが、ボクセルのデータ構造としてオクトツリーを用い、順次分解能を上げながらボクセルと交差する1枚の画像からの直線の数の最大値が2ないし1となった時点で、画像分解能にみあう適当なボクセル分解能になったといえる。

姚は浜野の研究を引き継いで、ポーティングの計算時間を削減するため画像とボクセル空間の多重解像度化によるアルゴリズムを提案し、両解像度がマッチする関係式を導いているが[13]、カメラから対象までの距離を必要としている。

第1回目の逆投影でK1カウントをもとめ、ひきつづき第2回目の逆投影でK2カウントをもとめて、 $p = K2/K1$ で評価することにより、オクルージョンが発生する場合の問題を解消するとともに、K1カウントだけでは偽の局所的ピークが発生する問題を回避した。

実際のアルゴリズムとしては、ボクセルのデータ構造としてオクトツリーを採用し、 $p > 0$ となったボクセルのみさらに分割することにより、メモリーを節約し無駄なボクセルに対する計算を大幅に減らすことができる。

今後、カメラを2台にして、さらに精度をあげることを検討している。浜野も述べているように[5]、カメラの移動方向に平行な線としてあらわれる特徴点はその3次元位置を抽出できない。本研究実験例でいえばテーブル回転軸を中心とする円弧がオクルージョンのため一部の方向からしか見えないうとき、その位置を精度よくは抽出できない。そのため回転テーブルを異なった俯角で臨む2台以上のカメラが必要であるが、別々に調整されたカメラの画像の融合という新たな課題がある。

謝辞

本研究をまとめるにあたり、助言をいただいた当研究室の赤松茂室長に感謝します。Gerigの論文を紹介していただいたATR通信システム研究所の田中弘美氏に感謝します。またボクセルの3次元表示の環境作りに援助いただいた、元ATR視聴覚機構研究所の客

員研究員 Philippe Quinio 氏に感謝します。

参考文献

- [1] J.K.Aggarwal and C.H.Chien: 3-D structure from 2-D images, in "Advances in Machine Vision" edited by J.Sanz, pp.64-121, Springer-Verlag, 1989
- [2] H.H.Baker and R.C.Bolles: Generalizing epipolar plane image analysis on the spatiotemporal surface, IJCV, vol.3, no.1, pp.33-49, 1989
- [3] R.C.Bolles, H.H.Baker and D.H.Marimont: Epipolar plane image analysis: An approach to determining structure from motion, IJCV, vol.1, no.1, pp.7-55, Jun.1987
- [4] G.Gerig: Linking image-space and accumulator-space: A new approach for object-recognition, Proc. ICCV'87, pp.112-117, Jun.1987
- [5] 浜野輝夫、安野貴之、石井健一郎: 空間への Voting による 3次元環境情報抽出手法、信学論、vol.J75-D-II, no.2, pp.342-350, 1992
- [6] 浜野輝夫、安野貴之、石井健一郎: 錐体型視線のポーティングによる3次元情報抽出法、信学技報、PRU91-128, pp.17-24, 1992-02
- [7] 北村喜文、谷内田正彦: 三眼視による三次元情報の計測、ロボット誌、vol.5, no.2, pp.47-54, 1987
- [8] 大田友一: 画像を用いた三次元計測技術の動向、電学論C、vol.107, no.7, pp.608-612, 1987
- [9] P.Quinio: A random set approach to 3D scene reconstruction by stereoscopic vision, Proc. SPIE vol.1669, pp. -, Feb.1992
- [10] R.Szeliski: Real-time octree generation from rotating objects, Tech.Rep. 90/12, DEC Cambridge Research Lab., p.30, Dec.1990
- [11] R.Szeliski: Shape from rotation, Tech.Rep. 90/13, DEC Cambridge Research Lab., p.38, Dec.1990
- [12] 山本正信: 連続ステレオ画像からの3次元情報の抽出、信学論、vol.J69-D, no.11, pp.1631-1638, 1986
- [13] 姚偉明、堀越力、安野貴之、浜野輝夫: 多重解像度空間への Voting による3次元情報の抽出、信学技報、PRU9-17, pp.25-32, 1992-05
- [14] 鄭緯宇、永嶋美雄、岸野文郎: 人物の回転画像からの3次元顔モデルの再構成、信学技報、IE91-87, pp.45-50, 1991