

# 装着型視覚センサを用いた対象の3次元形状復元

築澤 宗太郎 杉本 晃宏 松山 隆司

京都大学大学院情報学研究科 知能情報学専攻

日常生活のなかで人を自然に支援するシステムは、その入力情報を人に近い視点で得る必要がある。そこで本研究では、ウェアラブルビジョンという新たな視覚情報処理システムの実現を目指して、人が装着した2台のカメラから得られる情報を用い、至近距離にある物体の逐次的に3次元形状復元する手法を提案する。まず、(1) 装着型視覚センサで得た各距離画像に関して、位置情報に加えて明度情報、更にロバスト統計を用いることで精度の高い位置推定を行う。次に、(2) 座標変換した複数の距離画像の併合に際し、アウトライアー及び誤差の影響を受けない併合を行う。あらかじめ用意した距離画像に対して本手法を適用し、その有効性を確認した。

## 3D Shape Reconstruction by Wearable Vision Cameras

S. Tsukizawa A. Sugimoto T. Matsuyama

Department of Intelligence Science and Technology

Graduate School of Informatics, Kyoto University

Systems which support people automatically in everyday life increasingly need to obtain the input with the viewpoint near people. This research aims at the realization of a new vision information processing system called "Wearable Vision". The Wearable Vision technique of performing sequential objective 3D shape reconstruction in point-blank range is proposed using the information acquired from two cameras which the user worn. First, (1) in addition to position information, accurate position estimation is performed by brightness information and robust statistics about estimation of each depth map obtained by the "Wearable Vision Cameras". Next, (2) in case of merging with two or more depth maps with carried out coordinates change, merging which is not influenced of outliers and errors is performed. This technique was applied to the depth map prepared beforehand and the validity was checked.

### 1 はじめに

人と協調して何らかのタスクを行うシステムには、周辺の環境やタスクの対象、その人の意図、注意などを自動的に認識することが求められる。これらを認識する方法として、センサを外部に設置することが考えられるが、センサと人との距離が大きくは慣れる場合には、状況の認識結果とくい違いが生じることが多い。したがって、システムと人が協調して何らかのタスクを行うには、システムと人が同じ立場に立つ事が望まれ、そのため、センサが装着型である必要がある。装着型であれば、人とシステムが入力情報を共有することが可能になるため、装着者の意図や注意などの認識が容易になる。さらに、あらかじめセンサが設置されていないような施設や屋外において

もタスクを行うことが可能になり、人がより自由に行動できるようになる。ところで、人は、自分の周辺環境や動作対象などの情報を得る主な手段として視覚を用いている。したがって、周辺の環境やタスクの対象、その人の意図・注意などを自動的に認識するためのシステムのセンサとして、装着型視覚センサが有効であると考えられる。

装着型視覚センサに求められる基本的機能として、

1. 装着者の視線情報が得られること。
2. 3次元空間中での装着者の位置および運動を推定できること。
3. 注視対象の3次元形状およびその表面テクスチャに関する情報が得られること。

の3つが挙げられる。このうち、1,2についてはこれまでその手法が提案されている [6][7]。そこで本研

究では3に注目し、特に、対象の3次元形状復元の手法を提案する。

装着型視覚センサによる物体の3次元情報復元が実現されれば、遠隔地での発掘調査や持ちだし禁止のものの調査などの際、その場でそれを記録することができる。さらに、ネットワークとつながっていれば、対象の3次元情報を瞬時に送ることができる。これにより、遠隔地の資料を参照し、詳しい調査をすることが可能になる。したがって、一旦記録してしまえば、何度も現地に行く必要がなくなる。さらにこのとき、物体の全体を記録するために物体を手にとって回すことができれば、人が物を手に取ってみるような自然な状態で物体の3次元形状およびそのテクスチャを取得することができる。

装着型視覚センサは、上に述べたように、対象を手にとって眺め回すような状況で用いる場合、対象までの距離が非常に近くなる。このため、対象の一部しか見えないことがある。そこで、対象の全周囲3次元形状を復元することを考えると、各視点から得られた情報を統合する必要がある。また、得られた情報が対象の3次元形状復元に十分かどうかをその場で確認できるようにするため、逐次的に形状復元することが望ましい。

逐次的な対象の全周囲3次元形状復元は、(a) 距離画像の取得、(b) 距離画像の貼り合せ、(c) 併合という手順をこの順にくりかえすことで実現される。各手順において様々な手法が存在する。

(a) 距離画像の取得：各画素値が物体までの距離を表す距離画像は、コンピュータビジョンの世界では、3次元情報を表現する方法として一般的なものである[8]。これを取得する方法として、レーザーレンジセンサを用いる方法[9]、ステレオカメラを用いる方法[10]、光学ボケ特性を利用する方法[11]などがある。このうち、本研究では、ステレオカメラを用いる方法によって距離画像を取得することにする。この方法は、レーザーレンジセンサ等と比べると比較的疎な距離データしか得られないが、テクスチャ情報を同時に得られる、他の方法よりも対象の形状や材質の制限が比較的緩やかであるといった利点がある。また、撮影自体は一瞬で済むため、長時間位置を固定することが難しい装着型にも適している。

(b) 距離画像の貼り合せ：複数枚の距離画像から一つの対象形状を復元するには、各距離画像間の相対的な位置関係を推定し(以下、これを位置推定と呼ぶ)、それに基づいて複数の距離画像を一つの座標系で表す必要がある。これを距離画像の貼り合せと呼ぶ。逐次的に形状復元をする場合、2枚の距離画像を時系列順に位置推定することになるが、このとき2枚の距離画像には、共通して見える部分と片方にしか見えない部分が存在する。そこで、距離画像中の各点が2枚の距離画像中に存在するのかが片方にのみ存在するのかを識別しつつ、雑音の影響を排除して、ロバストに2枚の距離画像間の位置推定を行わなければ

ならない。これまでの研究[3][4][5]では、位置情報だけから距離画像の位置推定を行ってきた。しかし、位置情報だけを用いて、これを行うのが困難な形状をした物体も存在する。例えば、球や円柱、立方体などのように周期的な形状をしているものがそれにあたる。したがって、位置情報のみによって距離画像の位置推定を行おうとする限り、復元対象の形状に制限ができてしまう。この問題の解決策の一つとして、明度情報を用いることが挙げられる。明度情報は3次元形状を直接表すものではないが、物体の表面の情報を非常に多く含んでいる。したがって、これを用いれば対象の形状に関する制限を緩和することが可能となり得る。位置情報に加えて、明度情報を用いて評価することで、雑音の影響を排除しつつロバストに距離画像の位置推定を行い、それに基づいて座標変換をする手法が有効であると考えられる。

(c) 距離画像の併合：距離画像の併合とは、距離画像の貼り合せを行った後、本来同一の点群を一つの点に併合することである。雑音の無い距離画像で最適な位置推定を行えば、同一の点は完全に重なるはずである。しかし現実には、距離画像には雑音が含まれているので、必ずしも完全に重なるわけではない。そこで、完全に重ならない場合でも、同一の点とみなせる点同士は一つにまとめられるようにする必要がある。最も単純な併合方法の一つとして、近くにある点同士の重心を取るという方法があるが、この方法では点が集中しているような場合は適切な併合ができない。それに対し、各点に重みをつける方法[1]や、点単位ではなく、距離画像単位で併合する方法[2]などがある。各点に重みをつける方法は、点が比較的疎な場合や、雑音が含まれている場合に有効であるが、点が密な場合には正確な併合が難しい、処理に時間がかかるといった問題がある。一方、点単位ではなく距離画像単位で併合する方法は、レーザーレンジセンサで距離画像を得るときなど点が比較的密な場合に有効であるが、雑音に弱いといった問題がある。ステレオカメラを用いた方法で距離画像を得る場合には、距離画像中の点が比較的疎であることを考えると、各点に重みをつける方法が有効であると考えられる。

以上の議論に基づいて、本研究では、装着型ステレオカメラを用いた至近距離における物体の3次元形状復元を目指し、このうち、距離画像の貼り合せとその併合の逐次的アルゴリズムに焦点を当てる。

## 2 距離画像の貼り合せ

### 2.1 距離画像貼り合せの問題点

距離画像の貼り合せとは、基準となる座標系を設定し、それに各距離画像の座標系を変換し、一つの座標系で表現できるようにすることである。しかし、装着型視覚センサを用いて距離画像を取得する場合、各距離画像間の座標系の関係が既知でないこと、各距離画像が雑音を含んでいることといった問題が存

在する．このうち雑音は，大きく分けてガウス雑音とインパルス雑音の2つに分類することができる．ガウス雑音は量子化の際などに生じるもので，インパルス雑音はステレオ画像対から距離画像を取得する際に誤った対応付けをしてしまうことによって生じるものである．したがって，距離画像の貼り合せをするアルゴリズムは以下の条件を満たす必要がある．

- 距離画像の位置を正確に推定できること．
- データ処理が効率的であること．
- 距離画像に含まれるガウス雑音への対処がなされていること．
- インパルス雑音に対して頑健であること．

これまでの研究では位置推定を行う際に位置情報のみを利用してきたが，本研究では，位置情報に加えて明度情報を利用することで，上記の条件を満たすアルゴリズムの実現を目指す．

## 2.2 ワールド座標系の設定

各距離画像の座標変換のためには，基準となる座標系，すなわちワールド座標系を新たに設定する必要がある．まず， $t$  枚の距離画像を用意し，時系列順に  $M_1, \dots, M_i, \dots, M_t$  とする．また， $M_i$  を生成する際に使用した2枚の画像を  $\text{Image}_{i1}, \text{Image}_{i2}$  とする．そして， $M_1, \dots, M_i, \dots, M_t$  の座標系をそれぞれ  $W_1, \dots, W_i, \dots, W_t$  とし，最初の距離画像の座標系  $W_1$  をワールド座標系とする．ここで，各距離画像の座標系とは，基準としているカメラ，すなわち左カメラの各時刻におけるカメラ座標系を表す．このとき， $W_i$  から  $W_1$  への座標変換とは， $W_i$  の3つの基底ベクトルを  $W_1$  の3つの基底ベクトルに一致させることである．したがって， $W_i$  から  $W_1$  への座標変換は，回転行列  $R$  および並進ベクトル  $T$  を用いて実現される．これ以降，各距離画像の座標系を  $W_i$  の座標系へ変換していく．

## 2.3 各距離画像のワールド座標系への変換

距離画像の座標変換は，そのまま探索すると計算量は膨大になり，しかも局所解に陥る可能性が高い．そこで，まずおおまかな位置推定し，次に詳細に探索し，最適解を求めるという2段階のアプローチをとる．これにより，計算量を軽減し，局所解に陥ることを回避する．

### 2.3.1 ランダムサンプリングによる変換候補の絞り込み

$W_{i-1}$  までの座標系の座標変換が終った状態での  $W_i$  から  $W_{i-1}$  への変換を考える．つまり， $W_i$  から  $W_{i-1}$  へ座標系を変換する回転行列  $R$  および並進ベクトル  $T$  を求める．また， $W_1$  への変換は， $W_{i-1}$  までの座標変換が既知であるので，これを用いることで実現される．

回転行列  $R$  は，Hamilton の四元数 ( Quaternion ) を用いて表すこととする．四元数を用いることで任意軸を中心とした回転を表せるため，イメージが持ちやすいことや，線形補間が可能であるため，細かな探索に適しているといった利点を持っているからである．Hamilton の四元数を用いると，回転行列  $R$  は次のように表せる．

$$R = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & -2(q_0q_3 - q_1q_2) & 2(q_0q_2 + q_1q_3) \\ 2(q_0q_3 + q_1q_2) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & -2(q_0q_1 - q_2q_3) \\ -2(q_0q_2 - q_1q_3) & 2(q_0q_1 + q_2q_3) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \quad (1)$$

ただし， $q_0, q_1, q_2, q_3$  は，

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1 \quad (2)$$

を満たす． $q_0$  は回転を表す成分， $q_1, q_2, q_3$  はそれぞれ回転軸を表す単位ベクトルの  $x, y, z$  成分に対応する．

回転行列の自由度は3であり，並進ベクトルの自由度も3である．したがって，合計6の自由度を決定することで座標変換は一意に行われる．この決定に必要な  $M_i$  と  $M_{i-1}$  間の対応点数は，3以上である．そのため，理想的には  $M_i$  と  $M_{i-1}$  に共通する3点を選び，連立方程式を解いてその3点が完全に一致するような  $R$  と  $T$  を求めればよいことになる．しかし，各距離画像にはガウス雑音及びインパルス雑音が含まれているので，実際には，完全に3点が一致するような  $R$  と  $T$  が必ずしも存在するわけではない．そこで，回転行列の初期解と，並進ベクトルの初期解を求め，その周辺を探索して行く．

初期解の求め方であるが， $\text{Image}_{i1}$  と  $\text{Image}_{(i-1)1}$  のテンプレートマッチングによって  $M_i$  と  $M_{i-1}$  に共通する3点を選ぶ． $M_i$  の座標系と  $M_{i-1}$  の座標系が平行になるように  $W_i$  の基底ベクトルを回転させれば，後は平行移動することで  $M_i$  の座標系と  $M_{i-1}$  の座標系が一致する．そこで，はじめに  $M_i$  と  $M_{i-1}$  がほぼ平行になるような回転行列を求め，これを回転行列の初期解とする．次に，回転後の  $M_i$  と  $M_{i-1}$  がほぼ重なるような並進ベクトルを求め，これを並進ベクトルの初期解とする．

具体的には，以下のような手法で距離画像の座標系の変換を行う．

$M_i$  と  $M_{i-1}$  にある特徴点の数をそれぞれ  $m, n$  とする． $M_i$  の各点を  $C_{i1}, \dots, C_{ij}, \dots, C_{im}$ ， $M_{i-1}$  の各点を  $C_{(i-1)1}, \dots, C_{(i-1)k}, \dots, C_{(i-1)n}$  とする．そして，各点の3次元座標をそれぞれ  $C_{i1}, \dots, C_{ij}, \dots, C_{im}$ ， $C_{(i-1)1}, \dots, C_{(i-1)k}, \dots, C_{(i-1)n}$  とする．また  $C_{ij}$  に対応する  $\text{Image}_{i1}$  上の点を中心とした  $a \times a$  ピクセルのテクスチャーを  $\text{Texture}_{ij}$  とする． $a$  は  $\text{Image}_{i1}$  の解像度に応じて適宜設定する．

まず， $M_i$  の特徴点の中からランダムサンプリングで3点を選ぶ．これらを  $C_{ip}, C_{iq}, C_{ir}$  とする．

次に， $C_{ip}, C_{iq}, C_{ir}$  に対応する点を  $M_{(i-1)}$  上の特徴点から探索する． $M_{i-1}$  と  $M_i$  は時系列的に連続で

あり、撮影間隔が十分短ければ視点は大きく変わらないので、対応する点周辺の明度分布はほぼ同じであると考えてよい．そこで、 $M_{(i-1)}$  の各点に対して  $\text{Texture}_{ip}$  と  $\text{Texture}_{(i-1)j}$  ( $j = 1 \cdots n$ ) でテンプレートマッチングを行い、各画素の明度差の絶対値が最も少ない点を探索する．その点を  $C_{ip}$  に対応する点  $C_{(i-1)p}$  とする． $C_{iq}$ 、 $C_{ir}$  についても同様に行う．

回転行列を求めるためには、 $C_{ip} - C_{iq}$  と  $C_{(i-1)p} - C_{(i-1)q}$  が等しく、かつ  $C_{ip} - C_{ir}$  と  $C_{(i-1)p} - C_{(i-1)r}$  が等しくなるような回転行列  $R$  を探索すればよい．そこで、回転行列  $R$  を探索するための評価関数として、次のような  $F(R)$  を考える．

$$F(R) = \sum_{\alpha=q,r} \|(C_{(i-1)p} - C_{(i-1)\alpha}) - R(C_{ip} - C_{i\alpha})\|^2 \quad (3)$$

ただし、 $\|\cdot\|$  はベクトルのノルムを表す．そして、 $F(R)$  が最小となる  $R$  を求め、これを回転行列  $R$  の初期解  $R_{init}$  とする．なお、この時  $F(R)$  に閾値を設け、最小となる  $F(R)$  が閾値以上であれば誤った対応付けがなされたとみなし、 $M_i$  から 3 点をリサンプリングしてもう一度  $R_{init}$  を求める．

理想的には、並進ベクトル  $\mathbf{T}$  は  $C_{(i-1)p} - R_{init}C_{ip}$  であるが、距離画像に含まれている雑音を軽減するために、 $C_{(i-1)p} - R_{init}C_{ip}$ 、 $C_{(i-1)q} - R_{init}C_{iq}$ 、 $C_{(i-1)r} - R_{init}C_{ir}$  の平均を取る．つまり、

$$\mathbf{T}_{init} = \frac{\sum_{\alpha=p,q,r} \{C_{(i-1)\alpha} - R_{init}C_{i\alpha}\}}{3} \quad (4)$$

となるような  $\mathbf{T}_{init}$  を並進ベクトルの初期解とする．

### 2.3.2 ロバスト統計を用いた最適変換の導出

第 2.3.1 節で求めた  $R_{init}$ 、 $\mathbf{T}_{init}$  を中心とした最適解の探索を M-estimator を用いて行う [8]．M-estimator は最もよく利用されるロバスト推定手法の一つであり、インパルス雑音の影響をなるべく受けたくないような評価基準を用いる手法である．これまでの研究では、この評価基準に位置情報のみを使用してきたが、本研究では、これに加えて明度情報を利用する．つまり、対応する点同士の距離が近く、かつ、その点周辺の明度分布が似ていれば良い値を返す評価関数を設定することで、最適な変換を探索する．これにより、明度情報を利用しない場合に比べ、変換精度が上がるのが期待される．以下に、具体的な最適変換の導出方法を示す．

まず  $R_{init}$ 、 $\mathbf{T}_{init}$  を用いた  $C_{ij}$  ( $j = 1, 2, \dots, m$ ) の変換後の座標  $R_{init}C_{ij} + \mathbf{T}_{init}$  に最も近い  $M_{(i-1)}$  の点の座標を  $C_{(i-1)\beta}$  とする．また、 $\text{Texture}_{ij}$  と  $\text{Texture}_{(i-1)\beta}$  の各画素の明度差の絶対値の和を求め、それを  $\text{Diff}_{j\beta}$  とする．

次に、 $R_{init}$ 、 $\mathbf{T}_{init}$  を中心に  $R_{init}$  を表す四元数、および、 $\mathbf{T}_{init}$  を変化させ、最適となる  $R$ 、 $\mathbf{T}$  を求める．

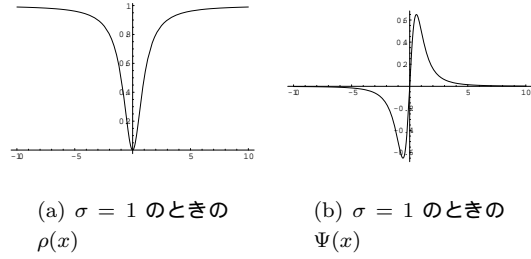


図 1: Geman と McClure の  $\rho$  関数

このとき、評価関数として Geman と McClure の  $\rho$  関数

$$\rho(x) = \frac{x^2}{\sigma + x^2} \quad (5)$$

を用いる．ここで  $x$  はモデルとの誤差であり、 $\rho(x)$  は  $x = 0$  で唯一最小値を持つ正定値偶関数である．図 1(a) に  $\sigma = 1$  のときの  $\rho(x)$  のグラフを示す．このように、 $\rho(x)$  は  $x = 0$  に近くなるほど小さな値を返す関数である． $\rho(x)$  がモデルからはずれたデータに対してどれくらいの重みを与えるかを見るには、 $\rho(x)$  の微分

$$\Psi(x) = \frac{\partial \rho(x)}{\partial x} \quad (6)$$

をとればよい．図 1(b) に  $\Psi(x)$  の  $\sigma = 1$  のときのグラフを示す．

このように、この関数はデータがモデルからある程度離れると、その影響力がほとんどなくなる関数であるため、距離画像中のインパルス雑音の影響を取り除くのに適している．本研究では、誤差として変換後の  $C_{ij}$  とそれに最も近い  $M_{i-1}$  上の点との距離にそれぞれの点周辺の各画素の明度差の絶対値の和をかけたものを用いる．つまり、評価関数として、

$$E(R, \mathbf{T}) = \sum_{j=1}^m \frac{\{\|R(C_{ij} + \mathbf{T}) - C_{(i-1)\beta}\| \text{Diff}_{j\beta}\}^2}{\sigma + \{\|R(C_{ij} + \mathbf{T}) - C_{(i-1)\beta}\| \text{Diff}_{j\beta}\}^2} \quad (7)$$

という  $E(R, \mathbf{T})$  を置き、 $E(R, \mathbf{T})$  が最小になるような  $R$ 、 $\mathbf{T}$  を、 $M_i$  から  $M_{i-1}$  への最適な回転行列及び並進ベクトルとする．ただし、 $\sigma$  は定数で、距離画像のスケールに応じて適宜設定する．

以上により、これまで位置推定が困難であった物体も信頼性の高い位置推定が可能となるのが期待される．この位置推定の基づいて  $W_i$  から  $W_{i-1}$  への座標変換を行う．以上の手順を繰り返すことで、逐次的な距離画像の貼り合せが実現される．

## 3 距離画像の併合

全周囲 3 次元形状復元のためには、第 2.3 節で述べた距離画像の座標系変換後、対応する点同士の併合を

行わなくてはならない．本研究では逐次的な処理を目標としているので，距離画像を時系列順に1枚ずつ，逐次併合して行く．つまり，距離画像  $M_1$  に対し  $M_2, \dots, M_i, \dots, M_t$  を順に併合し，更新していく．このとき，雑音の影響をなるべく受けないように，各点に重みをつけて併合を行う．ただし，この重みは，モデルに近い程高い値を返し，遠くなるほど小さな値を返すものでなければならない．

$M_{i-1}$  までを併合した  $M_1$  に， $M_i$  を併合する場合を考える．なお， $M_i$  には  $m$  個の特徴点が存在するとしておく．

$M_1$  に含まれる特徴点の数を  $l$  とし， $M_{i-1}$  まで併合したときの  $M_1$  上の点を  $C_{1\gamma}(i-1)$  ( $\gamma = 1 \dots l$ ) とする．また，座標変換後の  $M_i$  の各点  $C_{ij}$  ( $j = 1, 2, \dots, m$ ) から  $C_{i\gamma}$  への距離を  $x_i(\gamma j)$  ( $i = 2 \dots t, \gamma = 1 \dots l$ ) とする．このとき， $M_i$  上の各点  $C_{ij}$  の  $C_{1\gamma}(i-1)$  に対する重みは，

$$g_i(\gamma j) = \frac{\tau}{\tau + x_i^2(\gamma j)} \quad (8)$$

で表される関数とする．ここで  $\tau$  は定数で，距離画像のスケールに応じて適宜設定する．図2に  $\tau = 1$  のときの  $g_i(\gamma j)$  をグラフを示す．このように， $g_i(\gamma j)$  は， $x_i(\gamma j)$  が0から離れる程，小さな値を返す関数であるため，併合の際，遠くにある点の影響を取り除くのに適してる．

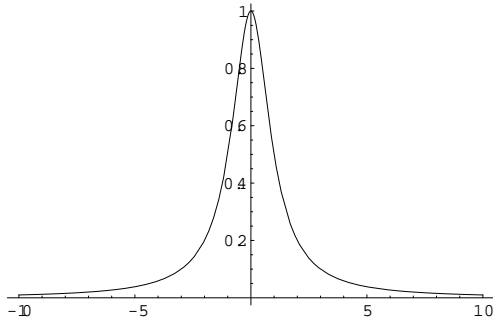


図 2:  $\tau = 1$  のときの  $g$

$M_i$  まで併合したときの  $M_i$  に含まれる  $C_{1\gamma}(i)$  の重み  $G_i(\gamma)$  は次の式で表される．

$$G_i(\gamma) = \sum_{\delta=1}^i \sum_{j=1}^m g_\delta(\gamma j) \quad (9)$$

で表される．

したがって， $M_{i-1}$  まで併合したときの  $C_{1\gamma}(i-1)$  は， $M_i$  を併合したとき，

$$C_{1\gamma}(i) = \frac{G_{i-1}(\gamma)C_{1\gamma}(i-1) + \sum_{j=1}^m g_i(\gamma j)C_{ij}}{G_{i-1}(\gamma) + \sum_{j=1}^m g_i(\gamma j)} \quad (10)$$

というベクトルを持つ点に更新される．これにより， $M_1$  に存在する点に関しては，点の併合が行われる．しかし， $M_i$  上には， $M_1$  上に存在しない点もあるので，併合の際には，これを新たに  $M_1$  に追加しなくてはならない．そこで，以下のような関数  $P_i(j)$  ( $j = 1 \dots m$ ) を考える．

$$P_i(j) = \sum_{\gamma=1}^l g_i(\gamma j) \quad (11)$$

この関数に適当な閾値を設定する．この関数の値が閾値以下であれば， $M_1$  上のあらゆる点から一定以上離れているということを意味するので， $C_{ij}$  を， $M_1$  に  $M_{i-1}$  まで併合した時点では存在しない点とみなせる．そこで， $C_{ij}$  をあらたに  $M_1$  に追加する．これにより， $M_i$  上にある， $M_{i-1}$  まで併合した時点で  $M_1$  上に存在しない点についても併合することができる．

以上の手順を繰り返すことで，逐次的な距離画像の併合が実現される．

## 4 実験

提案手法の有効性を示すために，まず，2枚の距離画像の位置推定および併合が正確に行えるかどうかを実験により確認し，その後，多数枚の距離画像を用いた場合について検証する．なお，この実験で用いる全ての距離画像は，ステレオ画像対に対して手動による対応点付けを行うことで取得した．



図 3: 装着型視覚センサ

システムは，PC と2台の視点固定型カメラが一体化されたヘルメット (図3) からなる．視点固定型カメラ2台から得られた画像を，PCに取り付けられた2枚のビデオキャプチャボードでそれぞれ取り込み，取り込んだ画像を解析・処理する．また，PCからのパン・チルト角による視線制御，ズーム，フォーカスの制御が可能である．用いた主な機器は，以下の通りである．

- PC : INTEL XEON 1.7GHz Dual × 1
- カメラ : Sony EVI-G20 × 2

- ビデオキャプチャボード: Canopus WINDVR PCI New Edition × 2

#### 4.1 2枚の距離画像の貼り合せ

はじめに、2枚の距離画像の貼り合せおよび併合が正確にできるかどうかを実験で確認する。復元対象には図4(a)のヒートシンクを用いた。明度情報を得るためのテンプレートマッチングを容易に行うためにシールを張り付けた。

距離画像の取得: 対象をカメラから約300mmの距離に置いた。まず、異なった2つの方向から対象を撮影し、2組のステレオ画像対から距離画像を取得した。

距離画像の貼り合せおよび併合: 第2.3節に示した方法に沿って距離画像の位置推定を行った。位置推定の結果に基づいて座標変換した結果、対応する点同士の距離は、最大4.0mm、最小0.072mm、平均0.8mmであった。この処理に要した時間は、初期解の導出に約3秒、最適解の導出に約2時間である。次に、2枚の距離画像の併合を行った。併合の結果を図4(b)に示す。

#### 4.2 多数の距離画像の貼り合せ

次に、逐次的な処理による3枚以上の距離画像の貼り合せおよび併合を行う。復元対象にはタバコの箱(図5)を用いた。

距離画像の取得: 対象をカメラから約500mmの距離に置き、回転させながら撮影し、10組のステレオ画像対を取得した。図6、図7に時系列順に並べたときの1番目と2番目のステレオ画像対を示す。

次に、各ステレオ画像対から距離画像を取得した。図8(a)に図6から取得した距離画像を、図8(b)に図7から取得した距離画像を示す。

距離画像の位置推定および併合逐次的に、10枚の距離画像を貼り合せ、併合した結果を図9に示す。10枚の距離画像併合にかかった時間は約20時間であった。

図9を見ると、大まかな形は出ているが、本来長さが同じ辺同士の長さが違うことが分かる。例えば、図9(b)では、手前の方の辺は奥の方の辺に比べて約1.5倍長い。また、本来直線であるはずの部分が歪んでいることも分かる。

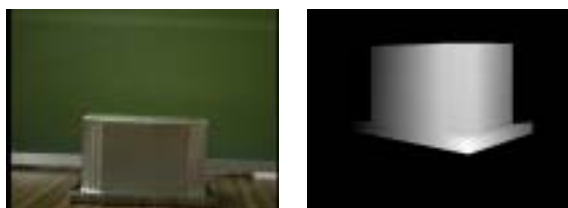
#### 4.3 考察

距離画像の貼り合せの精度および、距離画像の併合が適切に行われたかどうかの2点について考察する。

距離画像の貼り合せについて: 第4.1節の実験より、2枚の距離画像間の位置推定は最適に行われたと考えられる。しかし、多数の距離画像の位置推定では、今回提案した考え方の場合、ワールド座標系への変換は2枚距離画像間の座標変換を繰り返すことで実現しているので、変換を繰り返すほど、誤差が蓄積される。そのため、時系列で後になるほどワールド座標系への変換は、正解からずれていくと考えられる。

距離画像の併合について: 併合時に、評価関数の一つとして、第3節の式(8)を用いたが、この関数は $\tau$ に大きく影響される。すなわち、 $\tau$ が大きすぎると、本来同一でない点の影響を大きく受けてしまう。逆に、 $\tau$ が小さすぎると、本来同一である点まで違う点であると認識されてしまう。しかし、最適な $\tau$ の値は距離画像によって違うため、適切な値の推定は難しい。第4.2節の実験では、本来同一である点は全て同一であると認識されているが、歪みがあることから、 $\tau$ の値が、最適値よりも大きすぎたと考えられる。

以上、本稿で示した提案手法を用いて対象の3次元形状復元を行った。一部精度に問題があるものの、ある程度の結果を得ることができた。



(a) 復元対象に用いたヒートシンク (b) 第4.1節における併合後の距離画像

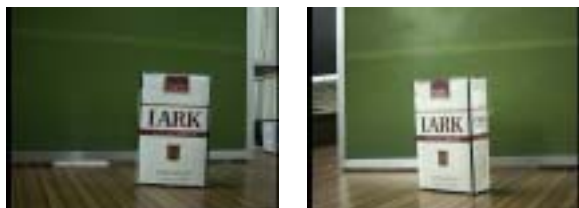
図4: 多数の距離画像貼り合せのためのステレオ画像対1組目



図5: 復元対象に用いたタバコの箱

## 5 結論

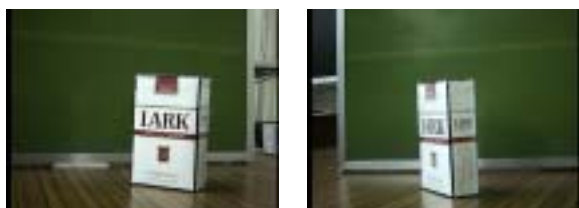
本研究では、装着型ステレオカメラを用いた至近距離における物体の3次元形状復元を目指した。対象の3次元形状復元に関して現在提案されている手法は、対象の位置情報しか利用していないため、対象の形状、表面テクスチャに制限があるなど、さまざまな問題を抱えている。そこで本研究では、位置情報に加えて対象の明度情報を用いることでこれらの問題の解決をはかった。まず、距離画像の位置推定の際に



(a) 左カメラからの画像

(b) 右カメラからの画像

図 6: 多数の距離画像貼り合せのためのステレオ画像対 1 組目



(a) 左カメラからの画像

(b) 右カメラからの画像

図 7: 多数の距離画像貼り合せのためのステレオ画像対 2 組目



(a) 図 7 の視点から見た  
併合後の距離画像

(b) 別方向から見た併合  
後の距離画像

図 9: 併合後の距離画像



(a) 図 6 から得られた距  
離画像

(b) 図 7 から得られた距  
離画像

図 8: 第 4.2 節における距離画像

位置情報に加えて明度情報を用いることで、初期解の導出および正確な位置推定を行うアルゴリズムを提案した。次に、逐次的な距離画像の併合を行うアルゴリズムを提案した。最後に、これらのアルゴリズムを実証するための実験を行った。この結果、本研究で提案したアルゴリズムは現段階では精度に問題はあつたものの、対象の3次元形状の逐次的な復元が可能であることを確認した。

今後の課題として、次のことが挙げられる。

距離画像の自動取得：本研究では手動による対応点付けをしているため、距離画像の取得に時間がかかる。また、手動での対応点付けはインパルス雑音は無くなるものの、ガウス雑音が多くなるため、正確な形状が得にくくなる。したがって、信頼性の高い対応点付けの自動化が望まれる。

処理速度の高速化：現在の手法では、位置推定時の最適解導出に時間がかかりすぎるため、逐次的な処理をする際の利点が失われている。そこで、最適化手法を改善し高速化する必要がある。

表面テクスチャの復元：装着型視覚センサを実用化するためには、3次元形状だけでなく、表面テクスチャの復元も求められる。ステレオ画像の明度情報や復元された形状情報を用いて光源の位置を推定するなどして、表面テクスチャを復元することが望まれる。

## 謝辞

本研究は科学研究費補助金特定領域研究 13224051の補助を受けて行われた。

## 参考文献

- [1] Brian Curless, Marc Levoy, "A Volumetric Method for Building Complex Models from Range Images", Computer Graphics (SIGGRAPH 96 Proceedings), 1996.
- [2] G. Turk, M. Levoy, "Zippered polygon meshes from range images", Proceedings of SIGGRAPH '94 (Orlando, FL, July 24-29, 1994), pages 311-318. ACM Press, July, 1994.
- [3] 西野 恒, 池田克史, "大規模レンジデータのロバストな同時位置合わせ", コンピュータビジョンとイメージメディア, No.127, 2001.
- [4] 金子俊一, 近藤友紀, 宮本敦, 五十嵐悟, "M 推定を導入したロバスト ICP 位置決め法", 精密工学会誌, vol.67, no.8, pp.1276-1280, 2001.
- [5] 鎌田 大樹, 相澤 清晴, 望月 研二, 齋藤 隆弘, "シーンの複数視点レンジデータの統合に関する検討", 電子情報通信学会 マルチメディアと仮想環境基礎研究会, MIV2000-124, pp.95-102, Mar.22 2001.

- [6] 中山 章宏, 視点情報とステレオ視を用いた注視対象の抽出, 修士論文, 京都大学, 2001.
- [7] 長友 渉, 装着型能動視覚センサを用いた人物の位置及び運動の推定, 修士論文, 京都大学, 2002.
- [8] 松山隆司, 久野義徳, 井宮淳. "コンピュータビジョン 技術評論と将来展望. 新技術コミュニケーションズ", 1998.
- [9] M. Rivox, "Laser range finder based on synchronized scanners", Applied Optics, vol.23, no.21, pp.3837-3844, 1984.
- [10] T. Kanade, A. Yoshida, K. Oda, H.Kano, M. Tanaka, "A stereo machine for video-rate dense depth mapping and it new application", Proc. CVPR, pp.192-202, 1996.
- [11] S.K.Nayar, Y.Nakagawa, "Shape from focus: Aneffective approach for rough sarfaces", Proc. ICRA, pp.218-225, 1990.