

動画像からのモデルに基づく3次元物体認識と運動復元

北川英裕美† 佐藤 嘉伸‡ 田村 進一‡

†大阪大学 基礎工学部 情報工学科

‡大阪大学 医学部 機能画像診断学研究所 565 吹田市山田丘 2-2

あらまし 本論文では、多面体モデルと動画像の照合により、動物体の認識および運動の復元を行う方法を述べる。従来手法では、動画像からの形状復元処理の後に物体モデルを利用したのに対して、本研究では、動きの抽出とセグメンテーションが行われていない段階で物体モデルを利用する。本手法では、まず、連続画像の各々の画像と物体モデルとの照合を Hough 変換により行い、各画像上にある物体の姿勢パラメータを推定する。しかし、ノイズや隠れが多い場合、単独の画像のみでは、誤認識や見逃しの危険性が高く、信頼性に問題がある。そこで、各時刻で推定された姿勢パラメータのうち、時間的になめらかに変化する系列のみを抽出することによって、認識の信頼性を向上させるとともに物体の運動を復元する。このとき、各時刻での見逃しに対処するため、系列内の補間も同時に行う。さらに、抽出された姿勢パラメータの系列に対して精密化処理を施し精度の向上をはかる。本手法では、従来手法で必要となる動画像からの形状復元という不安定な処理を伴わず、効果的に時間方向の統合を行うことができる。シミュレーション実験により、本手法による時間方向の統合の有効性を示す。

Model-Based Moving 3D Object Recognition and Motion Recovery from a Sequence of Images

Ayumi Kitagawa†, Yoshinobu Sato‡† and Shinichi Tamura‡†

†Faculty of Engineering Science, Osaka University

‡Division of Functional Diagnostic Imaging, Osaka University Medical School
Suita, Osaka, 565, Japan

Abstract The goal of this work is recognizing 3D objects and recovering their motion using a sequence of images and an object model. Unlike the previous method, which uses an object model only to match 3D data recovered from a sequence of images, we utilize an object model to estimate each object motion from a sequence of images which includes clutters and multiple moving objects. First, image data at each frame are matched against an object model using the Hough method. However, the Hough method cannot avoid detecting false peaks and overlooking true peaks. Secondly, false peaks are eliminated by selecting sequences of peaks which satisfy temporal smoothness of motion. Simultaneously, interpolation is performed at overlooked true peaks. Finally, extracted and interpolated sequences of peaks are further processed in order to improve accuracy. We present experiments involving image sequences including many clutters or multiple moving objects.

1 まえがき

コンピュータビジョンの技術分野への応用は広く、インパクトレンチ(ねじ締め機) [1]、ピンピッキング [2] や結晶生成過程の制御システム [3] などへの応用のための研究などがなされている。そのようなシステムにコンピュータビジョンの技術を応用し、システムに人間の目に相当する機能を加えることによってこれまで人間が絶えず監視し、制御することが必要であった技術分野を自動化することが可能になる。このような分野では、取り扱う物体の形状は一定である場合が多く、取り扱う物体の3次元形状は既知として、その物体領域の抽出や3次元運動の復元を行う手法の開発は、十分に意義がある。また、人間による視覚においては、ある瞬間の視覚情報だけではなく、時間的に連続した視覚情報が重要な役割を果たしていると考えられる。そこで、本研究では、モデルに基づく物体認識と動画像処理を統合させることにより、3次元物体の認識と運動復元を同時に行う手法を検討する。ただし、動物体の存在位置がある程度既知とする物体の追跡 [4] ではなく、雑音の多い画像や複数の動物体が存在する画像から動物体を発見することを目的とする。

モデルに基づく物体認識と動画像処理を統合させる試みの従来研究としては、Murray らが、多面体エッジの見かけの動きの抽出、セグメンテーション、3次元復元、および、その結果とモデルとの照合の4つの処理を順次行う方法を提案している [5]。しかし、2つ以上の物体が重なりあって運動するような場合、動きの抽出とセグメンテーションは、種々の手法 [6],[7] が検討されているものの、未だ非常に困難な問題とされている。この方法では、結局、誤りを多く含む動き情報を基に3次元復元を行った結果と物体モデルとの照合をとることが必要となり、よって、物体モデルと動画像の情報の有機的な統合が十分行われているとは言い難い。

そこで、本研究では、最終的な照合段階でのみ物体モデルを用いるのではなく、動きの抽出とセグメンテーションが行われていない段階で、物体モデルを積極的に利用するアプローチを提案する。本研究では、まず、連続フレームの各画像と物体モデルから一般化 Hough 変換の考え方にに基づき、各時刻における物体の変換パラメータを推定する。しかし、ノイズや隠れなどの影響が大きい画像の場合、単独の画像での情報のみで判断することは、誤認識や見逃しの危険性が高く、信頼性が低いことが知られている [8]。本研究では、各時刻で推定された変換パラメータを基に、なめらかに変化する系列を補間および抽出することによって、認識の信頼性を向上させるとともに物体運動を復元することを試みる。本手法では、動画像からの3次元復元という不安定な処理を伴わず、また、各時間毎に画像特徴が見え隠れするような場合にも効果的に時間方向の統合を行うこ

とが期待できる。

以下、2. では解くべき問題を明確にし処理の概要を述べる。3. では、処理の各段階について詳しく述べる。4. では、実験結果により本手法の評価を行う。5. で、本研究のまとめを行い、今後の課題について述べる。

2 概要

2.1 処理課題

本研究では、運動物体が投影された連続フレームから、3次元モデルを用いて、各フレーム上の物体の3次元姿勢の復元を試みる。各フレームで3次元姿勢を得ることにより、物体の3次元運動の復元も同時に行える。

本研究で扱う処理課題の前提条件を以下に示す。物体のとりうる位置・姿勢は6自由度とする。投影はスケール変換つき平行投影を仮定し、取り扱う物体は多面体で剛体、また、物体モデルは1種類とする。画像中に認識可能な動物体が複数個投影されているものとし、隠れやノイズのある画像を仮定する。3次元モデルの情報としては、多面体のワイヤーフレームモデルを与える。

2.2 処理の概要

本節では、モデルを用いて、運動する剛体の連続画像から、その運動と3次元姿勢を復元する方法の概要について述べる。

まず、連続画像の各画像について、Hough 変換を用いて、その画像上の物体の3次元姿勢を計算する。本手法では回転成分を表す3パラメータのみを推定し、並進成分は考慮しない。1枚の画像のみからでは信頼性の高い結果を得ることは難しいので、この段階では、候補となる値と、その確からしさを計算する。 x, y, z はモデルを x 軸回り、 y 軸回り、 z 軸回りに回転させたときの回転角度とする。フレーム i における、確からしさを回転の3パラメータを3軸とする3次元配列 $w_i(x, y, z)$ によって表現する。3次元配列 $w_i(x, y, z)$ の各セルには、各3パラメータの確からしさが格納される。この確からしさは Hough 変換の投票値である。これをフレーム i の重みテーブルと呼ぶ。さらに、この重みテーブル内の確からしさのピークを検出する。本手法では、並進成分を扱わないが、物体の回転成分が推定できれば、フレーム上投影された物体の2次元形状が決定できる。また、本手法では、フレーム上の頂点とモデルの頂点の対応づけも行うので、本手法での処理結果から並進成分を推測することは、比較的容易な問題であると考えられる。また、位置・姿勢を表す6次元空間への一般化 Hough 変換を行えば、偽のピークは空間内で分散されるので、信頼性は高くなり、1枚のフレームからでも推測可能

であるとも考えられるが、必要なメモリ領域、計算時間の観点から実用性を欠く。

次に、各フレームでの処理結果を統合する。各フレームでの重みテーブル $w_i(w, y, z)$ のピークから、姿勢パラメータが時間的になめらかに変化する系列を抽出する。ただし、真のピークの見逃しに対処するため、系列を抽出する際、隣接するフレームのピークだけでなく、1枚飛ばしたフレームのピークも考慮に入れる。

最後に、抽出したピークの系列に対して精密化処理を施す。重みテーブルの3次元配列 $w_i(x, y, z)$ のセル値を参照する x, y, z の値は離散的であるため、抽出したピーク系列の各パラメータ値は、精度の上で問題がある。最終的に求めたい各フレームの姿勢パラメータは、重みテーブルの投票値が大きく、前後のフレームの姿勢パラメータとなめらかにつながるような値である。そこで、投票値と運動のなめらかさを考慮にいれた誤差関数を設定し、抽出された系列の値を初期値として、その関数を最小化するような繰り返し計算によりパラメータ値の精密化を行う。

3 動画像からの物体認識と運動復元

3.1 各時刻での変換パラメータの推定

3.1.1 画像とモデルとの対応付け

各時刻での物体の姿勢パラメータを推定する。

まず、入力画像から線分構造グラフを作成する。このグラフには、入力された線画から得た、頂点座標、隣接頂点情報、コーナーの角度の情報を格納しておく。

次に、モデルと画像との対応付けを行う。一般化 Hough 変換の考え方をを用いて、重みテーブルを作成する。このテーブルは、物体の姿勢パラメータを表す3次元空間からなる。この空間上の離散的に決めた座標上へ投票を行う。投票の方法は、物体をそのパラメータだけ回転した時にできる投影図と、入力画像で、角の角度の照合を行い、照合される角が多ければ、入力画像の姿勢パラメータである可能性が高いとしてその座標に大きな値が投票される。

3.1.2 ピークの抽出

ある局所的な範囲内における最大値をピークとして抽出する。ここで得られたピークの座標値が、入力画像上の物体の姿勢パラメータを表す。しかし、このような単純な方法では、偽のピークの除去ができず、解が一意に定まらなかったり、正しい解が見逃されたりする危険性は避けられない。

3.2 系列の抽出

各画像での処理結果を時間軸方向に統合すると、入力画像は連続フレームであることから、正しいピークの系列は、時間的になめらかに変化する系列である。入力画像すべてについてピークの抽出をしたあと、なめらか、かつ、投票値の大きいピークの系列を抽出する。連続フレームの情報を使い、なめらかさの条件を加えることで、各フレームの偽のピークを取り除き、信頼性を高める。

1枚のフレームの計算結果で得られたピークの集合には、偽のピークも含まれている。画像によっては正しい値に対する投票値よりも、偽のピークへの投票値のほうが大きくなることもあり、1枚のフレームの処理結果の情報だけでは偽のピークは取り除けない。しかし、連続するフレームでは、物体の位置は急激に変化しないことを用いれば、前後のフレームのピークの集合に連続するようなピークの要素がないピークの要素は、偽のピークと判断できる。

しかし、すべてのフレームで、真のピークが必ず検出されているとは限らない。前述の考え方で、偽と判断されるピークを取り除いていくと、うまくピークを抽出できない1枚の画像があっただけで、その画像との連続性の拘束のために、そのフレームのみでなく、連続フレーム全体の計算が不可能になってしまう可能性がある。そこで、連続性の条件は、隣あう前後の画像で満たせない場合は、1枚とばした前後の画像で満たしていればよいものとし、前後のフレームで連続条件を満たし、そのフレームのピークの集合では連続条件を満たすピークの要素がないものは、*nil*としておく。*nil*となったフレームの要素は、前後のフレームから補間する。これにより、単独フレームでは真のピークが現れないような画像でも、物体の3次元姿勢が推定できる。第1フレームから第*N*フレーム（最終フレーム）まで、系列内での隣合うピークの位置、 x, y, z の差がそれぞれ*k*以下（角度では $k \times \Delta_\omega$ 度以下）であるか、1つとばしたフレームでのピーク位置の差が、 x, y, z それぞれについて $2k$ 以下（角度では $2k \times \Delta_\omega$ 度以下）であるように、前処理で抽出したピークの集合のなかから選ぶ。

すなわち、選ばれた系列のピークの位置は以下の条件をすべて満たす。

系列のなかで、第*i*フレームに対するピークの位置を (px_i, py_i, pz_i) とする。

$$|px_{i+1} - px_i| \leq k, \text{ または、} |px_{i+2} - px_i| \leq 2k \text{ かつ } px_{i+1} = nil$$

$$|px_i - px_{i-1}| \leq k, \text{ または、} |px_i - px_{i-2}| \leq 2k \text{ かつ } px_{i-1} = nil$$

$$|py_{i+1} - py_i| \leq k, \text{ または、} |py_{i+2} - py_i| \leq 2k \text{ かつ } py_{i+1} = nil$$

$$|py_i - py_{i-1}| \leq k, \text{ または、} |py_i - py_{i-2}| \leq 2k \text{ かつ } py_{i-1} = nil$$

$$|pz_{t+1} - pz_t| \leq k, \text{ または, } |pz_{t+2} - pz_t| \leq 2k \text{ かつ}$$

$$pz_{t+1} = nil$$

$$|pz_t - pz_{t-1}| \leq k, \text{ または, } |pz_t - pz_{t-2}| \leq 2k \text{ かつ}$$

$$pz_{t-1} = nil$$

なお、隣合う要素がともに *nil* となることはない。

nil の補間のしかたは、*nil* となったフレーム番号が 2 から $N-1$ の場合は、前後のフレームで選んだ値の平均値、フレーム 1 が *nil* の場合は、2 と N (N が *nil* の場合は $N-1$) から傾斜をもとめて 1 の値を推測、フレーム N が *nil* の場合は、1 (1 が *nil* なら 2) と $N-1$ から傾斜をもとめて N の値を推測する。

3.3 系列の精密化

求めたい系列の、第 t フレームにおけるピーク位置を (x_t, y_t, z_t) する。繰り返し計算の前段階で計算された系列の第 t フレームに対する値を $(\bar{x}_t, \bar{y}_t, \bar{z}_t)$ とする。

最適な系列は、投票値の大きい系列であるから、

$$W = \sum_{t=1}^N w_t(x_t, y_t, z_t, t) \quad (1)$$

を最大にするはずである。

しかし、3.2 の処理で抽出したピークの系列は、 x, y, z 成分ともに Δ_ω 度刻みに値を設定したテーブルを基に抽出しているため、*nil* となって前後から補間した値以外は Δ_ω の倍数の離散的な値となっている。物体の 3 次元姿勢は任意の値をとるので、このことは不都合である。また、物体の真の 3 次元姿勢が Δ_ω の倍数でない場合は、真の位置付近の格子点上に投票値が分散しているものと考えられる。

そこで、繰り返し重み付け最小 2 乗法を用いて、系列の値を修正する [9]。すなわち、あるエラー関数を設定し、そのエラー関数が最小値になるような系列を繰り返し計算によって探す。重みとして、各格子点への投票値を用いる。

また、真の値から離れた場所にも、偽のピークが存在したように、ノイズ等の影響で関係のない場所の投票値が大きくなっており、その部分も含めて重み付け最小 2 乗法を用いるのは誤った値に収束する原因となる。そこで、ここでは、前処理の結果値から一定範囲内の点と投票数だけを考慮にいった、重み付け最小 2 乗法とする。

以上の理由から、(1) 式を最小 2 乗法を適用できる形に変形する。

$$E_1 = \sum_{t=1}^N \sum_{\text{all}(x_i, y_j, z_k)} e^{-\frac{dis^2}{2\sigma^2}} w_t(x_i, y_j, z_k, t)$$

$$((x_t - x_i)^2 + (y_t - y_j)^2 + (z_t - z_k)^2) \quad (2)$$

ただし、 $dis = \sqrt{((x_t - \bar{x}_t)^2 + (y_t - \bar{y}_t)^2 + (z_t - \bar{z}_t)^2)}$ とし、 σ はガウス関数の幅を決定するパラメータとする。

また、なめらかさの拘束条件を加える。連続するフレームでは、物体の位置は急激に変化しないと仮定し、その評価関数を E_2 とする。

$$E_2 = \sum_{t=1}^{N-1} ((x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2 + (z_{t+1} - z_t)^2) \quad (3)$$

さらに、運動速度は急激に変化しないと仮定し、その評価関数を E_3 とする。

$$E_3 = \sum_{t=2}^{N-1} ((x_{t+1} - 2x_t + x_{t-1})^2 + (y_{t+1} - 2y_t + y_{t-1})^2 + (z_{t+1} - 2z_t + z_{t-1})^2)$$

最小にすべき評価関数は、(2) 式、(3) 式、および (4) 式を考慮にいった

$$ERROR = E_1 + \alpha E_2 + \beta E_3 \quad (4)$$

とする。

このような問題を解く場合、ある初期値を与えて、関数値が最小になるまで、求めたい値を更新することをくり返して計算するが、計算結果が初期値に左右されることが多い。しかし、ここでは、処理 3.2 で得られた系列を初期値として用いることができるので問題とはならない。

処理 3.2 の結果、複数の系列が得られた場合は、すべての系列を初期値として与えて計算する。複数の系列が得られる場合は、1. フレーム上に複数の物体がある場合、2. 同一物体に対する初期値が複数個得られる場合の 2 とおりの可能性がある。1. の場合は処理結果が全く別の系列になり、2. の場合では近い値に収束する。このことから、フレーム上の物体の数を推定することができる。

4 実験結果

計算機により発生した線画を入力画像とし、提案手法を用いてシミュレーションを行った。入力画像は、計算機により作成した多面体の線画とし、ノイズにより余分な線分の付加、あるいは、見逃される状況をシミュレートした。また、線分の端点のずれを乱数により発生させた。

4.1 モデル

シミュレーションに用いた物体の 3 次元モデルを図 1 に示す。対象物体の対称性が大きい場合、物体の姿勢の認識を行うことは困難をとまなう。しかし、インパクトレンチやピンピック等の工業応用においては、取り扱う物体は対称性が大きい形状の物が多い。そこで、本研究ではとくに対称性が大きい物体を選びシミュレーションを行った。用いたモデルは、砂糖結晶を模擬したものである。

4.2 実験結果

ここでは、連続画像は7枚のフレームを使った。モデルの対称性を考え、全く同じ投影図が得られる姿勢パラメータは省略した。すなわち、処理3.1でのモデルの回転範囲は $0 \leq x_i \leq 90$ (度)、 $0 \leq y_j \leq 90$ (度)、 $-180 \leq z_k \leq 180$ (度) とした。処理3.2におけるパラメータは、 $k=1$ とした。処理3.3の α, β の値は、初期値として与える系列より決定した。すなわち、初期系列の要素に対する投票値の平均値を計算し (= av とする) $\alpha = av, \beta = av \times 10, \sigma = 1$ とした。 α はなめらかさの拘束の重みを表すパラメータであるが、運動物体の連続フレームを取り扱っているため、正しい速度値に対しても $E_2 = 0$ とはならないので、 α の値は小さくした。

4.2.1 実験1：画像中の物体が1個の場合

実験1に用いた図のフレーム1,4,7を図2に示す。エッジの一部は欠けているものとし、各線分の端点の x 座標値、 y 座標値には、乱数により発生させた、3以下のノイズを加えた。

表1に、処理1の結果の、各フレームでのピークを示す。各フレームのデータにおいて、ピークになったときの位置パラメータ、 x, y, z を回転角度と投票値を示す。で表している。

表2に、処理2の結果得られたピークの系列を示す。2通りの系列が得られたが、これは nil の与えが2通りあるため2種類の系列が得られたが、実際には同じ系列を表していると考えられる。

上から、フレーム1、フレーム2... に対する要素を表し、各フレームでは、左から位置パラメータ x, y, z の値を表す。各値は回転各度で表した。

系列1、2のフレーム7、系列1のフレーム5、系列2のフレーム4では、適切なピークが抽出されていないので、どちらの場合も nil が選ばれて、補間されている。とくにフレーム7については、ピークの集合に正しい3次元姿勢に近い値のものが含まれていなかったが、連続フレームの情報から適切な値を見つけている。

また、精密化処理の結果を表3に、最終結果と正しい値をフレーム1,4,7について図3に示す。また、処理3.2で得られた2種類の系列は、精密化処理により似た系列になった。

この実験では、1枚のフレームだけでは、位置パラメータを正しく推測できないような場合でも、前後から推測が可能であることを示した。1枚の画像だけを処理した結果では、位置パラメータの候補として、偽のピークも含み、ピークが複数ででくともある。正しい位置パラメータへの投票数よりも、間違っただけの位置パラメータへの投票数のほうが大きくなる場合もあり、1枚のフレームからでは正しい値が決定できない。しかし、前後のフレームでの計算結

果を考慮に入れることで、間違っただけの値を除去することができる。また、正しいピークが、含まれていない場合もある。このような場合も前後のフレームから推測が可能である。

4.2.2 実験2：画像中の物体が2個の場合

実験2では、各フレームに複数の物体が存在する場合についてシミュレーションを行った。実験に用いた図のフレーム1,4,7を図4示す。それぞれ2個の物体があり、不完全な線画の状態を与えられているものとする。1つの物体(フレーム1で(15,42,0)と表されている方)を、 x 軸方向に100、 y 軸方向に100平行移動させたが、この値は、このシミュレーションには無関係である。

各線分の端点の x 座標値、 y 座標値には、乱数により発生させた、2以下のノイズを加えた。

処理3.1の結果では、フレーム7では一方の物体の姿勢パラメータに関して真の値に近いピークが得られなかった。フレーム4では、2つの物体の位置パラメータに関係のないピークでの投票数が大きくなった(真のピーク(15,45,0)への投票値1.533、(70,40,65)への投票値1.564)に対して偽のピーク(65,45,100)への投票値2.420)。また、他のフレームについても、真の値から5度程度離れた位置にピークが抽出されたものがあり、前後のフレームとの連続性が得られず、処理3.2の段階で補間を行った。

図5に最終結果と正しい値を示す。

1枚の画像だけからでは、複数のピークが抽出された場合、真のピークが決定できないことは既に述べたが、そのため、1枚のフレーム上に存在する物体の個数さえ決定できない。しかし、連続フレームの情報を用いると、各フレームでの真のピークは連続性をなすことから、真のピークだけが系列として抽出される。フレーム上の物体の個数は系列の個数として表れる。

4.2.3 実験3：物体の一部しか見えない場合

実験3では、物体の一部しか見えない場合についてシミュレーションを行った。

実験に用いた図のフレーム1,4,7を図6に示す。各線分の端点の x 座標値、 y 座標値には、乱数により発生させた、3以下のノイズを加えた

フレーム1では、正しい値付近にピークが得られなかった。また、フレーム4では、偽のピークへの投票値の方が真のピークへの投票値より大きくなった(真のピーク(50,50,20)への投票値0.216)に対して、偽のピーク(25,40,155)への投票値0.251)が、処理結果を時間軸方向へ統合することにより、正しい系列を得た。

図7に最終結果と正しい値を示す。この実験では、物体の一部しか見えないような場合でも、本手法が有効であることを示した。本手法では、認識のための手がかりとして、

頂点の角度情報を用いているので、物体の一部しか見えていないような場合でも、特徴的な部分が見えていればモデルとの照合で判断できる。

図7に最終結果と正しい値を示す。

4.2.4 実験4：ノイズによる線分がある場合

実験4では、ノイズによる余分な線分が多数ある場合についてシミュレーションを行った。実験に用いた図を、図のフレーム1,4,7を図8に示す。

各線分の端点の x 座標値、 y 座標値には、乱数により発生させた、3以下のノイズを加えた。さらに、エッジは欠けているものもあるとする。また、乱数によって発生させた、認識物体とは関係のない線分が24本のっているものとする(物体の認識では角の角度を用いるため、ノイズの線分も2本で1つの角をなすようにした)。

処理3.1の結果では、特にノイズの大きかったフレーム4で、6箇所偽のピークへの投票値の方が真のピークへの投票値より大きくなった。また、フレーム5でも、真のピークより偽のピークの方が投票値が大きかった。また、各フレームで得られた真のピークも、正しい姿勢パラメータからのずれの大きいものもあり(5度程度)連続性が満たされず、処理3.2で補間があった。図9に最終結果と正しい値を示す。

また、実験結果4では、ノイズによる、物体とは直接関係のない線分が描かれている場合でも有効であることが判った。ノイズによって線分が発生しても、線分のみでは、誤認識の情報は作り出さない。また、そのノイズによる線分とノイズによる線分、または、物体を表している線分とノイズによる線分の組合せで角ができて、ノイズによってできた角による投票が1ヵ所に集中するような可能性は低いので、認識の妨げになりにくいからである。

5 むすび

本論文では、多面体モデルと動画像の照合により、動物体の認識および運動の復元を行う方法を提案した。ノイズや隠れなどの大きい画像での誤認識や見逃しの危険を避け、信頼性を向上させるために、まず、連続画像の各々の画像と物体モデルの照合を行い、各画像での3次元姿勢を表すパラメータを推定したあと、なめらかに変化する系列のみを抽出する方法を提案した。

シミュレーション実験では、連続画像における各フレームの結果を統合することにより、偽のピークを取り除けること、1枚のフレームでは真ピークが抽出できない場合でも前後から補間できることを示した。また、1枚のフレームに複数の物体がある場合、ノイズによる偽のエッジを含んだ画像、物体の一部しか見えていないような画像に対し

ても有効であることを示した。しかし、3次元姿勢を表すパラメータの値が近い2つの物体が同一フレーム上にあるような場合では、複数物体の識別が困難であった。

検討課題としては、本手法では取り扱わなかった平行移動成分も含めた3次元姿勢の認識を行うことがあげられる。しかしながら、処理3.1.2での投票空間を6次元にすることは、メモリの容量や計算時間などへの影響がある。2.1節で、回転パラメータを推定したあと、モデルの頂点と入力画像の頂点を対応させることで並進成分を計算することが可能だと考え、回転成分のみの認識を試みたが、それでは姿勢パラメータ値が近い物体が複数ある場合に、認識できなかった。そこで、頂点对応に印を付けて回転成分の3次元空間に投票する方針で検討したい。

今後の課題としては、提案手法ではバッチ型処理をしているが、オンライン型処理に修正したほうが、実用的であると思われる。また、モデルが複数種類ある場合や、モデルの物体がフレーム上でない場合の対処も今後の課題である。

謝辞

日頃、ご討論頂く医学部機能画像診断学講座の多くの方々に心より感謝の意を表します。

参考文献

- [1] 鳥野、江尻、徳永 “移動物体の実時間視覚認識方式とその応用” 電気学会論文誌, 96-C, 3, pp.49-55 (1976).
- [2] 池内、長田、B. ホーン、k. ニシハラ “照度差ステレオとプリズムステレオによるビンピッキングタスク中のグリッパの姿勢決定” 信学論, vol. J68-D, no.4, pp.546-553 (1985).
- [3] 衣笠、千種 “砂糖結晶缶工程の解析と制御” 計測技術, pp.40-52, (1984).
- [4] D.G.Lowe, “Integrated Treatment of Matching and Measurement Errors for Robust Model-Based Motion Tracking”, Proc. 3rd ICCV, pp.436-440, (1990).
- [5] D.W.Murray, D.A.Castelow and B.D.F.Buxton, “From Image Sequences to Recognized Moving Polyhedral Objects” International Journal of Computer Vision 3, pp181-208, (1987).
- [6] M.Shizawa and K.Mase, “Simultaneous Multiple Optical Flow Estimation” Proc.ICPR'90, pp.274-277, (1990).
- [7] J.R.Bergen, P.J.Burt, R.Higorani and S.Peleg, “Computing Two Motions from Three Frames” Proc. 3rd ICCV, pp.27-32, (1990).
- [8] W.E.Grimson, D.P.Hattenlucher, and T.D.Alter “Recognizing 3D Objects from 2D Images: An Error Analysis” Proc.CVPR'92, pp.316-321, (1992).
- [9] P.J.Besl, J.B.Birch, and L.T.Watson “Robust Window Operators” Proc.ICCV, pp.591-600, (1988).

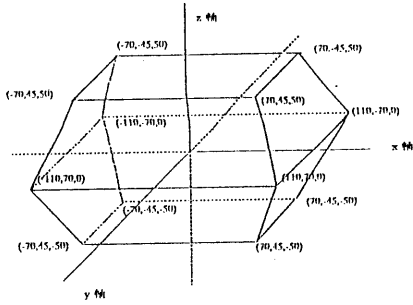


図 1: 物体モデル

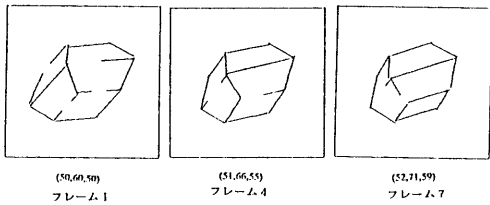


図 2: 実験結果 1: 入力画像

frame	x	y	z	frame	x	y	z
1	50.0	60.0	50.0	1	50.0	60.0	50.0
2	50.0	60.0	50.0	2	50.0	60.0	50.0
3	55.0	60.0	50.0	3	55.0	60.0	50.0
4	50.0	65.0	55.0	#4	57.5	62.5	50.0
#5	52.5	67.5	55.0	5	60.0	65.0	50.0
6	55.0	70.0	55.0	6	55.0	70.0	55.0
#7	56.0	72.0	56.0	#7	56.0	72.0	56.0

補間のおこったフレーム

表 2: 実験結果 1: 処理 3.2 結果

フレーム	真の値			処理 3.2 の結果 (精密化前)			処理 3.3 の結果 (精密化後)		
	x	y	z	x	y	z	x	y	z
1	50	60	50	50.0	60.0	50.0	49.5	58.6	50.1
2	50	62	52	50.0	60.0	50.0	50.1	60.6	51.3
3	51	64	53	55.0	60.0	50.0	50.9	62.6	52.6
4	51	66	55	50.0	65.0	55.0	51.5	64.5	53.8
5	52	68	56	52.5	67.5	55.0	52.2	66.5	55.0
6	52	69	58	55.0	70.0	55.0	52.9	68.4	56.2
7	52	71	59	56.0	72.0	56.0	53.5	70.4	57.4

表 3: 実験結果 1: 処理 3.3 結果

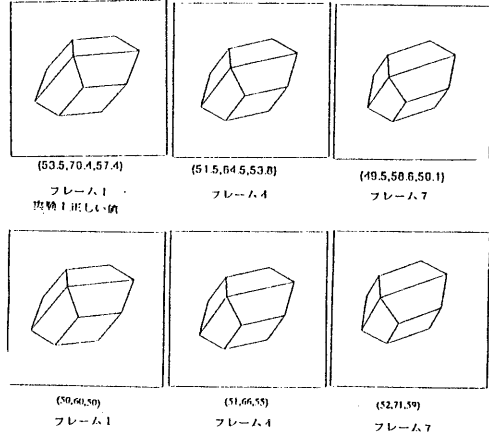


図 3: 実験結果 1: 最終結果

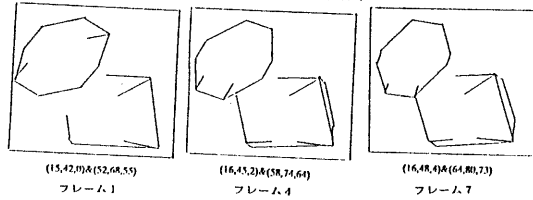


図 4: 実験結果 2: 入力画像

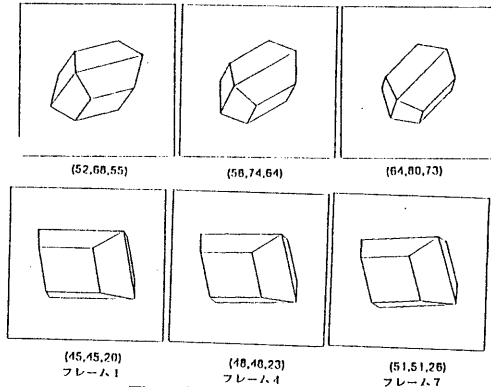
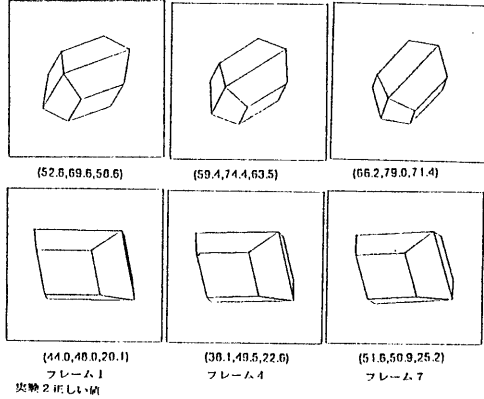
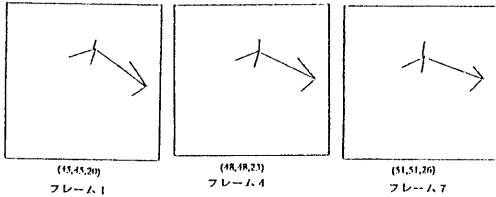


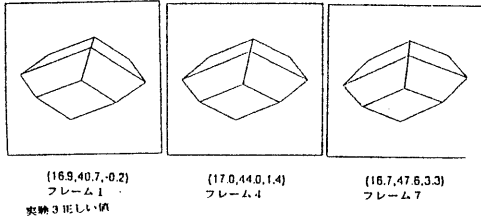
図 5: 実験結果 2: 最終結果

frame1	frame3	frame 4	frame5	frame6
x y z val	x y z val	x y z val	x y z val	x y z val
20 40 -130 0.236	10 55 -90 0.244	10 55 -85 0.387	5 40 -75 0.180	55 70 55 1.029
25 20 -130 0.242	10 85 -90 0.234	15 30 95 0.282	15 55 15 0.160	
25 35 5 0.242	15 35 90 0.250	25 20 -85 0.318	25 35 90 0.193	frame7
25 75 5 0.297	20 25 -90 0.262	30 70 75 0.293	25 40 -165 0.251	x y z val
25 75 70 0.263	25 35 -30 0.253	50 65 55 0.554	25 40 -90 0.226	
45 85 50 0.434	25 75 10 0.252	60 5 95 0.505	25 65 80 0.212	25 5 -160 0.175
50 60 50 0.438	25 80 75 0.324	60 5 95 0.505	30 25 15 0.240	25 65 -160 0.249
55 25 -115 0.302	30 60 10 0.220	85 45 20 0.297	30 35 -30 0.228	25 65 125 0.228
60 55 85 0.221	35 15 90 0.245		30 40 15 0.249	30 20 -100 0.181
	45 30 -105 0.435		30 65 -105 0.156	30 35 90 0.178
frame2	50 25 70 0.237		45 75 -165 0.151	35 70 75 0.342
x y z val	50 85 50 0.227		50 50 -160 0.182	45 65 20 0.178
	50 60 50 0.329		60 65 50 0.292	50 50 20 0.192
50 60 60 1.291	60 5 90 0.284		60 75 120 0.146	55 25 80 0.173
			70 5 15 0.156	60 40 -120 0.183
			75 15 -165 0.162	70 70 20 0.206
			85 40 -165 0.176	75 15 -160 0.184
			85 75 20 0.181	

表 1: 実験結果 1: 処理 3.1 結果



実験の結果 図 6: 実験結果 3: 入力画像



実験 3 正しい値 図 7: 実験結果 3: 最終結果

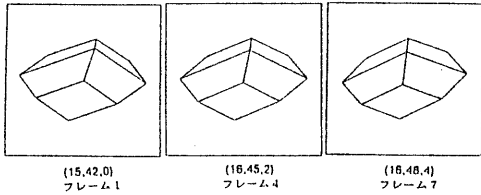


図 8: 実験結果 4: 入力画像

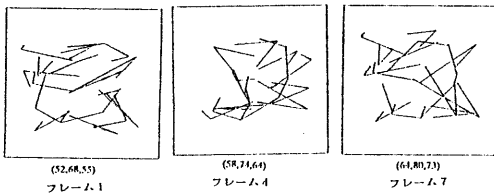
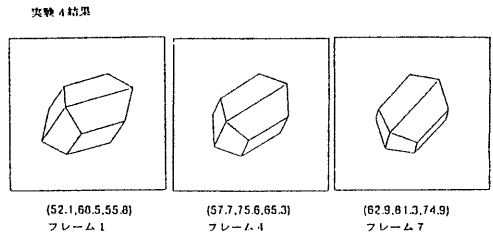


図 9: 実験結果 4: 最終結果



実験 4 正しい値

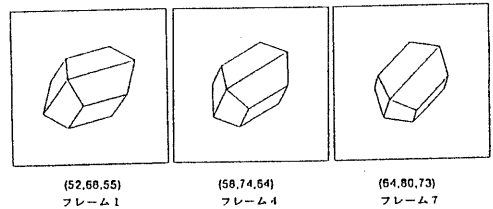


図 9: 実験結果 4: 最終結果