

詳細な形状モデルを用いたオブジェクトの3次元位置・姿勢推定

廣瀬 直子[†] 秋山 拓也[†] 有澤 博[‡]

[†] 横浜国立大学大学院環境情報学府 〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-5

[‡] 横浜国立大学環境情報研究院 〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-5

E-mail: [†] {naoko,takuya}@arislab.ynu.ac.jp, [‡] arisawa@ynu.ac.jp

あらまし 近年コンピュータを用いて実世界の情報を取得するために、ある方向から撮影された2次元画像を解析することで3次元情報を取り出すことが行われている。そのための手法として、3次元形状モデルを用いたモデルベースマッチングが行われている。しかしどの方向から撮影するかによって写るシルエットが大きく変わり、精度に大きく影響する。

したがって、そのような場合にどの方向から撮影するのが適切かを定量的に評価することが必要である。本研究では3次元形状モデルをコンピュータ上で様々な方向に投影し、そのシルエットに対してマッチングを行うことで適切な撮影方向を決定し、精度を向上させた。

キーワード 形状モデル, モデルベースマッチング, 画像処理

Position and posture estimation of 3D object with detailed shape model

Naoko HIROSE[†] Takuya AKIYAMA[†] and Hiroshi ARISAWA[‡]

[†] Graduate School of Environment and Information Sciences, Yokohama National University

[‡] Faculty of Environment and Information Sciences, Yokohama National University

79-7 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501 Japan

E-mail: [†] {naoko, takuya}@arislab.ynu.ac.jp, [‡] arisawa@ynu.ac.jp

Abstract Three dimension information has been taken out by analyzing two dimension image taken a picture from a certain direction to acquire information on the real world with a computer in recent years. Model based matching that uses three dimension shape model is done as a technique for that. However, the silhouette that reflects from which direction taking a picture changes greatly, and it influences accuracy greatly.

Therefore, it is necessary that taking a picture for that case from which direction evaluate quantitatively be appropriate. An appropriate direction of taking a picture is decided by projecting three dimension shape model in various directions on the computer in this research, and matching it to the silhouette, and accuracy has been improved.

Keyword Shape model, Model based matching, Image processing

1. はじめに

実世界のオブジェクトの3次元時空間情報(たとえば人体や機械, 物体の位置情報, 姿勢・形状などの変化情報)を計算機に取り込み, これを計算機に蓄積することで, オブジェクトの3次元空間上の振る舞いの定量的評価, イベントの発見, シミュレーション, 姿勢や動作の最適化などに役立てるシステムを構築することは非常に幅広い応用分野を持つと考えられる。我々は, この概念をリアルワールドデータベース(Real World Database)と称し, その構成要素技術の研究開発や, それを基にした三次元時空間情報取得システムの構築などを行ってきた[1].

実世界のオブジェクトの位置・姿勢の情報を計算機にとりこむ手段を考えたとき, 代表的なものとして位

置センサや角度センサなどを用いた計測装置や, 磁気センサ, マーカを用いたモーションキャプチャシステムなどが挙げられる。しかしこれらの手法では, 位置・姿勢を取得したいオブジェクトに対し, 特殊な装置やマーカを装着する必要がある, それらを装着することのできないオブジェクトに対しては適用することはできない。

これに対し, カメラなどから得られた画像データと, 位置・姿勢を取得したいオブジェクトの形状, 構造データを用い, 画像から抽出されたオブジェクトに対し形状, 構造データを照合させることでオブジェクトの位置・姿勢を決定するという手法(モデルベースマッチング)がある[2]. この手法は, オブジェクト自体に何らかの加工を施す必要がないため, 前述の手法では適

用できないようなオブジェクトに対して非常に有用である。

このモデルベースマッチングにおいては、一度に多方向からの画像を取得し死角を少なくしたり、オブジェクトの色や模様などの付加的な情報を使ったりして、マッチングで得られた姿勢の精度を向上させる試みが一般的にはなされている。しかし工場のロボットなどの物体把握や、医療機器から得られる画像からのオブジェクトの位置姿勢の取得など、多方向からの画像を得ることが装置や環境の都合上困難であったり、色や模様が環境によって変化してしまうために付加的な情報としてまったく使えなかったりする場面が現実には非常に多く存在する。しかしそのような場面においても精度の向上を図りたいという要望は非常に多く、本稿ではそういった撮影環境でモデルベースマッチングの手法を適用した場合での位置姿勢の推定精度について考察する。

このような場面でモデルベースマッチングの手法を適用する際には、撮影されたシルエット形状（またはシルエットの輪郭）が推定する際に一番重要なファクターとなり、得られる位置・姿勢パラメータの精度に大きな影響を及ぼすだけでなく、形によっては位置・姿勢が一意に決定できない場合もある。しかしシルエットの形状や、オブジェクトの動作に対しての画像上での形状の変化量は、どの方向からオブジェクトを撮影するかによって大きく変化する。そして画像からの推定を行う場合は、奥行き方向に関する平行移動量の推定を高精度に行えず、特に単方向撮影では他方向からの補正も行えないという問題がある。したがって、高精度な位置姿勢推定を行うために、どの方向から撮影してシルエットを得るかという問題はモデルベースマッチングの精度を考える上で非常に重要な問題である。

高精度に位置姿勢推定を行うためには、オブジェクトの動作を高精度に画像上で捉えること、つまりオブジェクトの動作に応じた画像上でのシルエット形状の変化が必要である。そこで画像上でのシルエット形状の変化量を評価基準として用いた、撮影方向最適化システムを構築することを考え、そのプロトタイプシステムの構築を行った。

以降 2 章では一般的なモデルベースマッチングの手法と、本稿で対象とした単視点撮影での問題点についての考察、3 章で誤差・精度の定量的評価方法、4 章で撮影方向最適化システムのプロトタイプの詳細、5 章で実験結果、考察を述べ、6 章でまとめを行う。

2. モデルベースマッチング

2.1. 一般的手法

モデルベースマッチングの一般的手法について以下に順を追って述べる[2].

① 対象オブジェクトの 3 次元的な形状、構造モデルの取得

まず、対象オブジェクトの 3 次元的な形状、構造モデルを計算機上に構築する。オブジェクトの形状は、三角形パッチなどのポリゴンデータとして、また円筒、超楕円形など近似を用いた表現などが使われることが多い。また、近年では、容易に CAD データを扱えるようになったため、対象オブジェクトが CAD データを基に作成された場合などは、非常に正確な形状モデルが得られるようになった。加えて、人体など構造のあるオブジェクトの場合は、剛体が関節で接続された木構造をもつ多関節モデルでオブジェクトの構造を表現することが一般に良く用いられている。また、得られる画像からオブジェクトの色や模様のテクスチャ情報が取得できる場合は、対象オブジェクトのテクスチャ情報も合わせて取得し、付加的な情報としてマッチングの際に用いる。

② 画像からのオブジェクトのシルエット抽出

撮影された画像から、オブジェクトの領域を抽出する。背景が変化しない場合は、背景画像を事前に取得しておき、差分をとってシルエットを抽出する方法や、連続する画像間での差を取り、動きのある部分を抽出する方法、テクスチャ情報が使える場合は、オブジェクト固有の色を用いて抽出するなど、さまざまな方法が提案されている。また、医療画像など閾値の調整でオブジェクトのシルエットが抽出できる場合も存在する。

③ オブジェクトの投影、画像との照合(マッチング)によるオブジェクトの位置姿勢の取得

最終段階として、画像から抽出されたオブジェクトのシルエット、オブジェクトの形状、構造データを用いてオブジェクトの位置姿勢の取得を行う。

あらかじめ計算機上に、撮影時に使用したカメラなどの撮像系を反映したカメラモデルを構築しておく。一般的なカメラモデルとしては、ピンホールカメラモデル(後述)が良く用いられる。

次に、オブジェクトの形状、構造モデルをもとに、オブジェクトにさまざまな位置、姿勢をとらせ、カメラモデルを使って、仮想的にオブジェクトを画面に投影し、シルエットを得る。これと、実際に撮影されたシルエットの一致度を比較し、もっとも一致した位置姿勢をそのときのオブジェクトの位置姿勢として出

力する。(図1)

一般的に用いられる一致度としては、実際のシルエットと、仮想的なシルエットが重なった領域と2つの和集合領域との比や、仮想的なシルエットの輪郭上の点に対し、実際のシルエットの輪郭上で最も近い点の距離を全体にわたって平均したものなどが使われる。多視点の画像が得られる場合には、これをすべての画像に適用して、解の信頼性を向上させる。

また、さまざまな位置姿勢をとらせる際、自由度が大きいためあらゆる姿勢に対して評価するのは不可能に近い。そのため、最急降下法などの最小化アルゴリズムや、遺伝的アルゴリズム (GA) などを利用して解を求める。

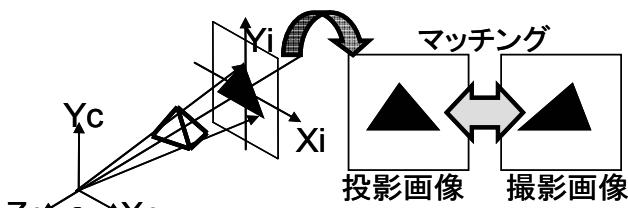


図1 投影

2.2. 単視点撮影によるモデルベースマッチングの問題点

本節では、本稿で対象とする単視点撮影によるモデルベースマッチングの問題点について考察する。

前節で述べたように、モデルとオブジェクトのマッチングは画像どうしで行われるため、画像に写ったモデルあるいはオブジェクトのシルエット形状はもっとも重要な要素である。しかし、画像は、3次元空間上のオブジェクトを2次元平面に投影したものであり、これは1次元分の情報の欠落を意味する。これによって解析の精度や信頼度は大きく影響を受ける。多方向からの撮影の場合は、欠落を補うことができるが、単方向撮影では、与えられる情報は2次元の画像のみである。このことから起こる問題として、例えば、カメラの死角に入っているオブジェクトの部分は解析を行えないが、単視点では死角を少なくすることはできない、またオブジェクトがカメラの奥行き方向に沿った動作を行ったとき、図2の左側の投影平面のように画像上でのシルエットの形状変化は小さくなるが、単視点撮影ではこれを図2の右側のような別角度からの撮影によって補正することができないなどが挙げられる。

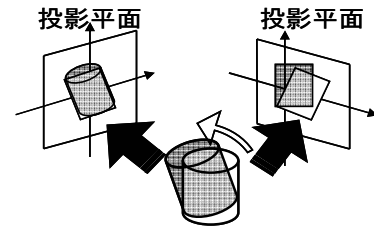


図2 方向による変化

3. 単視点撮影におけるモデルベースマッチングの精度の定量的評価手法

本章では、単視点撮影におけるモデルベースマッチングの精度の定量的評価手法について述べる。

これまでの議論で、単視点撮影では1次元分の情報が欠落した2次元の画像の情報から3次元のオブジェクトの位置姿勢を求められなければならない、そのためにカメラの奥行き方向に対して精度が悪くても、それを補正できないことを述べた。また、モデルベースマッチングの具体的な手法についても述べたが、モデルベースマッチングとは、シルエットの2次元的な変化の情報を用いて、オブジェクトの位置姿勢を決定しているともいえる。ここで言う変化の情報とは、同一な撮影方向から撮影された、ある位置姿勢のシルエットと、別の位置姿勢のシルエットが異なる、その変化ということである。別な位置姿勢であってもシルエットが変化しなければ、正確な位置姿勢は取得できない。

そこで、単視点撮影における精度の評価基準のひとつとして、ある位置姿勢と別の位置姿勢で得られるシルエットの変化量が使えるのではないかと考えた。精度を高めたい移動方向、回転などの限定した動きを与えれば、その動きに対して、ある撮影方向での精度の良否が評価できる。

また、シルエットの変化を定量的に表すために、オブジェクトを構成する各面の頂点の位置に注目した。オブジェクトの姿勢変化にともなう、投影平面上での頂点の移動量が多いほど、頂点が形成する面の移動量も多くなり、面によって構成されるシルエット形状の変化量も多くなると考えられる。したがってシルエット形状の変化量を頂点の移動距離の総和として表現する。また、オブジェクトの姿勢変化は、オブジェクト座標系と世界座標系間の変換パラメータを時系列に与えることで、カメラの撮影方向はカメラ座標系と世界座標系の変換パラメータで与えることで表現する。したがってある撮影方向の評価は、一連の動作におけるシルエットの変化量となるため、時系列に与えられた変換パラメータを順次適用したときの変化量の総和となる。

以降4章で具体的な評価式を示す。

4. 撮影方向最適化システム

3章で単視点撮影におけるモデルベースマッチングの精度の定量的評価手法について述べたが、これを用いて、撮影前にシミュレーションを繰り返し、あるオブジェクトのある動きに対してもっとも適切な撮影方向を提示するシステムがあれば非常に有用であると考え、撮影方向最適化システムを提案する。本稿では、実際に構築した撮影方向最適化システムのプロトタイプの詳細について述べる。

座標系の設定

はじめに、本稿で用いる座標系を設定する。

- (1) 世界座標系 (X_w, Y_w, Z_w)
- (2) カメラ座標系 (X_c, Y_c, Z_c)
- (3) 画像座標系 (u, v)
- (4) オブジェクト座標系 (X_o, Y_o, Z_o)

異なる座標系間の変換は、回転パラメータ $R(\theta, \phi, \varphi)$ と位置パラメータ $T(dX, dY, dZ)$ によって行われ、オブジェクト座標系から世界座標系の変換パラメータ (R_{ow}, T_{ow}) は、世界座標系からカメラ座標系への変換パラメータは (R_{oc}, T_{oc}) とする。

例として、座標系1から座標系2への変換を、変換パラメータ (R_{12}, T_{12}) によって次式で表す。

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} = R_{12} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} + T_{12}$$

4.1. システム構成

システム構成図を図3に示す。

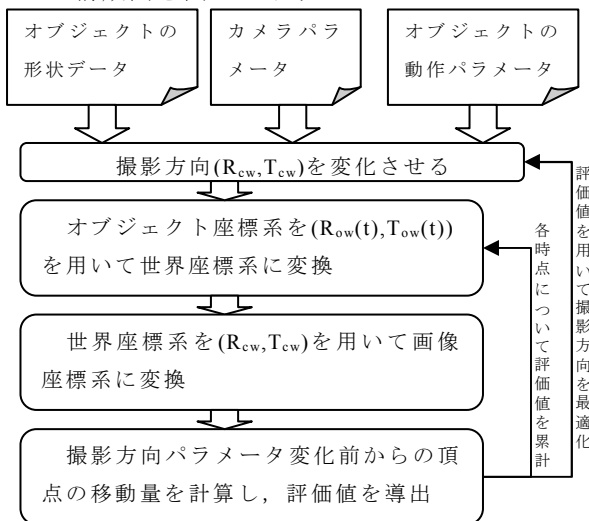


図3 システム構成図

入力データとしてオブジェクトの3次元形状データ、投影に用いられるカメラのキャリブレーションデータ、オブジェクトの動作パラメータである時系列の

$(R_{ow}(t), T_{ow}(t))$ を用いる。最も評価値の少ない撮影方向

(R_{wc}, T_{wc}) を求めるため、推定したい変数を撮影方向のパラメータ、評価関数をシルエット形状の変化量とし、Powellの方法で最適化を行う[5]。

4.2. カメラモデル

オブジェクトを2次元平面に投影するためには、カメラでの撮影をコンピュータ上でモデル化し、再現できなければならない。本稿で使用したモデルは一般的に広く用いられているカメラモデルであるピンホールカメラモデルを用いた[6]。(レンズ歪の考慮などはTsai[7]参照)カメラモデルのパラメータは、実際に撮影に使われる装置でカメラキャリブレーションを行って求めた値を入力する。

(x, y) を画像座標、 (X_w, Y_w, Z_w) を世界座標、 f を焦点距離とすると次式が成り立つ。

$$x = f \frac{X}{Z}$$

$$y = f \frac{Y}{Z}$$

またカメラ座標系と世界座標系との関係は

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + T$$

と表すことができる。レンズの半径方向のゆがみを考慮した座標 (X_d, Y_d) は

$$x = x_d(1 + k_1 r^2 + k_2 r^4 + \dots)$$

$$y = y_d(1 + k_1 r^2 + k_2 r^4 + \dots)$$

ただし $r = \sqrt{x_d^2 + y_d^2}$

画像座標系からカメラ座標系への変換は

$$u = S_x \cdot \frac{x_d}{d'_x} + C_x$$

$$v = \frac{y_d}{d'_y} + C_y$$

ただし

$$d'_x = d_x \cdot \frac{x_d}{N_{ix}}$$

S_x …スケールファクタ(CCD素子の縦横比)

d_x, d_y …受光素子の中心間距離(mm)

N_{ix} …1走査線のサンプリング数

C_x, C_y …画像中心の x, y (mm)

以上をまとめると、

$$\frac{d'_x}{S_x} \cdot (u - C_x) \cdot (1 + k_1 r^2 + \dots) = f \frac{r_1 X_w + r_2 Y_w + r_3 Z_w + T_x}{r_7 X_w + r_8 Y_w + r_9 Z_w + T_z} \quad (\ast)$$

$$d'_y \cdot (u - C_y) \cdot (1 + k_1 r^2 + \dots) = f \frac{r_4 X_w + r_5 Y_w + r_6 Z_w + T_y}{r_7 X_w + r_8 Y_w + r_9 Z_w + T_z}$$

(※) 式を用いて3次元空間上のオブジェクトを2次元の画像に投影できる。

本稿では、この式に含まれている未知変数を求めるため、Tsaiのキャリブレーション手法[7]を用いた。

4.3. オブジェクトの投影

オブジェクトの投影は、オブジェクト座標系で記述された各頂点を画像座標系まで変換することで行われる。変換のステップを以下に示す。

- (1) (R_{ow}, T_{ow}) を用いてオブジェクト座標系を、世界座標系に変換する。
- (2) (R_{wc}, T_{wc}) を、(※)式に代入する。
- (3) (※)式を用いて世界座標系から画像座標系に変換する。

オブジェクトの一連の動作は、オブジェクト座標系と世界座標系の変換行列を、時系列に変化させていくことで表現される。したがってまずオブジェクト座標系を世界座標系へ変換し、動作を反映させる。次に撮影方向を設定するため、カメラ座標系と世界座標系の変換パラメータ (R_{wc}, T_{wc}) を(※)式に代入する。そして世界座標系へ変換された各頂点座標を、撮影方向を反映した(※)式に代入することで、画像座標へ変換する。

4.4. 評価値の決定

3章で述べた評価値は、以下の変換ステップに従って、オブジェクト座標系で記述されたオブジェクトの各頂点 (X_{oi}, Y_{oi}, Z_{oi}) を、時系列に変化する変換パラメータ $(R_{ow}(t), T_{ow}(t))$ 、最適化の過程で変化する撮影方向のパラメータ (R_{wc}, T_{wc}) によって、画像平面上に投影し

た頂点の画像座標 $(u_i(t), v_i(t))$ を用いて導出される。

- (1) $(R_{ow}(t), T_{ow}(t))$ を用いてオブジェクト座標系から世界座標系に変換する。
- (2) (R_{wc}, T_{wc}) を用いて世界座標系から画像座標系に変換する。

具体的には画像平面上でのオブジェクトの頂点の移動距離を、すべての頂点(N)、すべての時間(T)について和を取ることで、一連の動作におけるシルエット画像の変化量として次式から得る。

$$E(R_{wc}, T_{wc}) = \sum_{t \in T} \sum_{i \in N} \{(u_i(t) - u_i(t-1))^2 + (v_i(t) - v_i(t-1))^2\}$$

5. 実験結果・考察

システムを実装し、検証を行った。使用したPC環境は次のとおりである。

CPU : Xeon(TM) CPU 2.40GHz

メモリ : 1.00GB RAM

OS : Microsoft Windows XP Professional

開発環境 : Microsoft Visual C++ .NET

結果を以下に示す。図4のような縦100[mm]×横120[mm]×高さ150[mm]の四角錐を、XとZの2つの軸を中心に、 $0^\circ \sim 180^\circ$ まで 10° ずつ回転させたときに、1050[mm]離れた視点から撮影した場合の最適方向評価を行い、最適だと判断された方向から動作中のオブジェクトの投影を行った結果画像を図5、図6に示す。処理にかかった時間は5秒であった。

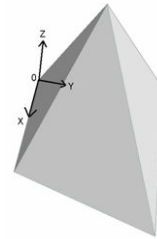


図4 オブジェクトの形状データ

いずれの結果においても、不適切な撮影方向からの投影ではオブジェクトの動作がほとんどシルエット形状に変化がないのに対し、最適撮影方向からの投影では大きく変化しているのが明らかであり、今回の実験に関してはシステムの有用性が示された。同様に適用すれば任意のオブジェクトについても最適撮影方向評価が行えるが、頂点数をnとすると計算時間は $O(n)$ となるため、シルエットの形状変化の評価値の導出方法に工夫が必要になってくると考えられる。

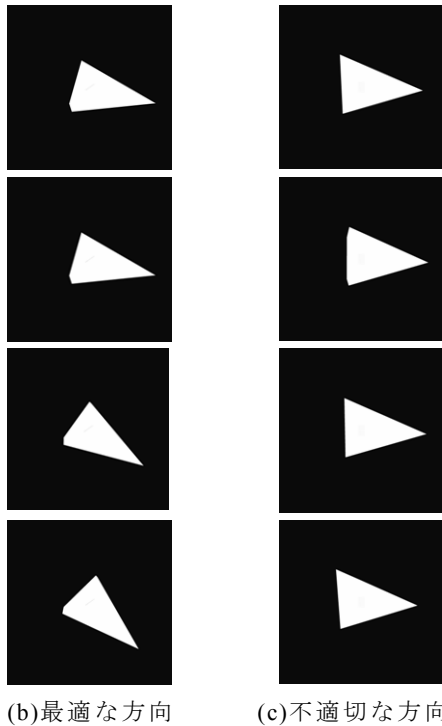
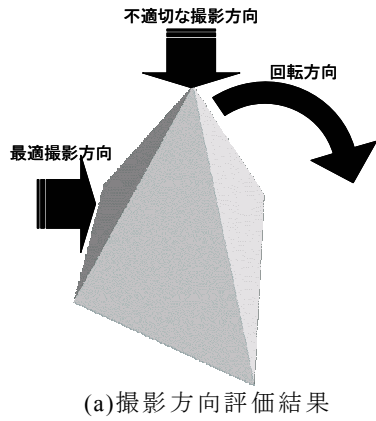


図5 X軸を中心とした回転動作の投影

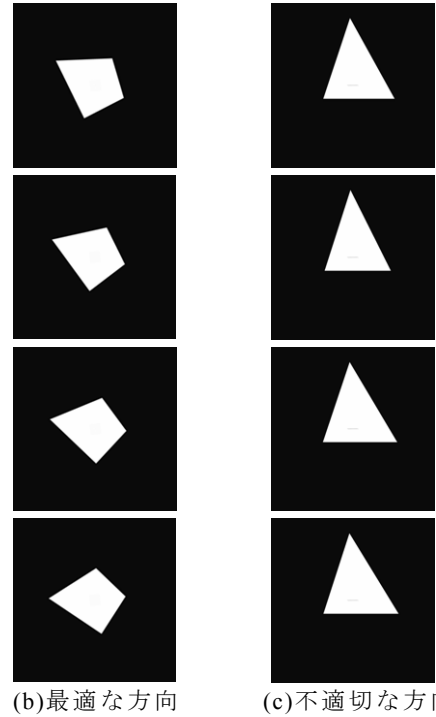
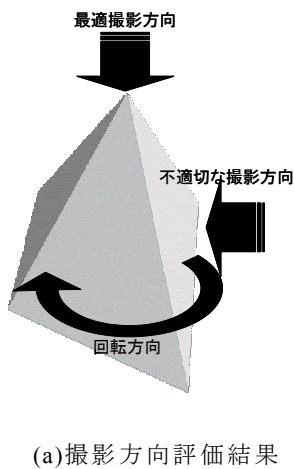


図6 Z軸を中心とした回転動作の投影

6. まとめ

本稿では単方向撮影画像を用いたモデルベースマッチングの精度向上のため、適切なシルエット形状を撮影できる撮影方向の検討を行った。オブジェクトが動作したときのシルエット形状の変化量を評価関数として撮影方向パラメータの最適化を行うシステムのプロトタイプを作成し、結果を得た。

今後は

- ・処理の高速化
- ・撮影方向やオブジェクトの動作範囲に制約がある場合の処理方法
- ・自由度の高い動作に対する評価方法

といった問題に取り組みシステムの性能を向上させるとともに、複雑な形状のオブジェクトに対して方向評価システムを適用させていく。その具体例の一つとして、X線透視撮影装置で取得した画像を用いて人工膝関節のモデルベース動作解析を考えている。これは患者の膝内部の様子を知るだけでなく、最適な人工関節設計に役立てられ、非常にその需要も多い。この動作解析の特徴として、例えば図7に示すような人工膝関節のデータは、形状の記述が詳細である点、撮影を単方向からしか行えない点、得られた画像がモノクロである点などが挙げられる。この環境では、X線透視撮影装置から得られるシルエットがその解析精度に非常に強い影響を及ぼすと考えられ、本稿で提案した手法と組み合わせると、更なる精度向上が見込めると期待できる。

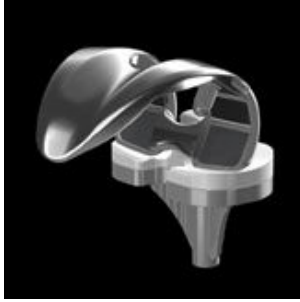


図 7 人工膝関節の形状

文 献

- [1] H.Arisawa, "Real World Database System - Design and Implementation", ADVANCED TECHNOLOGY AND SCIENCE, Proceedings of THE JAPAN-KOREA JOINT Symposium, pp.35-41, 1996
- [2] "新編 画像解析ハンドブック", 高木幹雄, 下田陽久 (監修) pp.323-343, 東京大学出版会, 東京, 2004
- [3] Banks, S.A, and W.A.Hodge, "Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy," IEEE Transactions in Biomedical Engineering, no.43(6), pp.638-649.
- [4] 山崎隆治, 渡邊哲, 中島義和, 菅本一臣, 富田哲也, 前田大助, 佐藤嘉伸, 吉川秀樹, 田村進一, "X線透視画像を用いた人工膝関節の三次元動態解析システムの開発," 日本放射線技術学会雑誌, 61(1), pp.79-87, 2005.
- [5] William H.Press et al. "NUMERICAL RECIPES in C 日本語版", 丹慶勝市他訳, pp.282-338, 技術評論社, 東京, 2004
- [6] 徐剛, 辻三郎, "3次元ビジョン", 共立出版, 東京, 2004
- [7] R.Y.Tsai "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", IEEE Journal of Robotics and Automation, RA-3(4) pp.323-344, (Aug)1987.