ビデオ映像内対象物体の三次元トラッキングシステム

The 3D Object Tracking System for the Target Object in the Video Image

武　井　宏　将

要　約　日本ユニシスでは，単眼カメラで撮影したビデオ映像に映る対象物体の三次元空間内の動作をトラッキングするプロトタイプシステムを構築した．物体を三次元トラッキングする方法として，特別な計測装置を用いたモーションキャプチャーがよく知られている．本システムは，対象物体の三次元 CAD データを用いることで，市販の単眼カメラで撮影した映像から対象物体の三次元トラッキングを行う．そのため，特別な計測装置なしで三次元トラッキングシステムを構築することができる．

本システムを構築するにあたり，三次元 CAD データを利用して二次元画像空間と三次元空間を結びつける新たなアルゴリズムを開発した．本システムは，現物と同じ形状の CAD データを用いて二次元の情報から三次元の実世界の情報を得られる点で，これまでにない新たなシステムである．

Abstract　We developed the prototype system which is a 3D object tracking system for the target object in a video image recorded by a monocular camera. As a method to perform 3D tracking for a target object, a motion capture system utilizing a special measurement device is well known. By using the 3D CAD data of target object, our prototype system can track the 3D object by a monocular camera without using any special devices.

While developing this system, we also developed a new algorithm that connects the 2D image space to the 3D space by using the 3D CAD data. Our system is quite new system in connecting the 2D image space to the 3D space by using the 3D CAD data that represent the real object.

1. は　じ　め　に

日本ユニシスでは，これまで長年に亘り CAD/CAM システムや CG システムの開発および提供を行い，数多くの三次元データ処理技術を蓄積してきた[1][2][3]．一方で，近年の画像処理技術の発展は目覚ましく，多くの場面で画像処理技術が活躍している．日本ユニシスの保有する三次元データ処理技術と画像処理技術を組み合わせることで，これまでにない新たなシステムを提供できると考え，一例として，市販の単眼カメラで撮影した映像から対象物体の三次元空間内の動作をトラッキングするプロトタイプシステム（以降，本システム）を作成した（図 1）．

一般に物体の三次元トラッキングを行う方法として，特別な計測装置を用いたモーションキャプチャーが知られている．一方本システムは，対象物体の三次元 CAD データを用いることで，市販の単眼カメラだけで三次元トラッキングを行う．そのため，特別な計測装置を用いることなくシステムを構築できる．したがって，特別な計測装置を利用することが難しい実際の製造現場で，ものの動きをトラッキングするような場面への導入が可能となる．本システム開発において取り組んだ課題は，「二次元画像に映る対象物体の三次元空間位置・姿勢を定め

図 1　ビデオ映像内の対象物体をトラッキングして，三次元空間内の動作を抽出
（左：対象物体のトラッキング，右：抽出した三次元空間内の動作）

ること」である．本稿では，本システム開発に採用した技術について述べる．

本稿は以下の構成をとる．2 章において，本システムの仕様と技術的優位性について述べる．3 章において，二次元画像から三次元トラッキングを行う仕組みと採用した技術について述べる．4 章では，今後の課題とまとめについて述べる．

2.　システム仕様と技術的優位性

本章では，2. 1 節において，本システムの操作仕様および入出力仕様について述べる．2. 2 節において，既存研究と比較した本システムの技術的優位性について述べる．

2. 1　システム仕様

はじめに，本システムの操作仕様を図 2 に示す．本システムは図中の番号に対応する以下の 1)から 4)の四つのステップからなる．

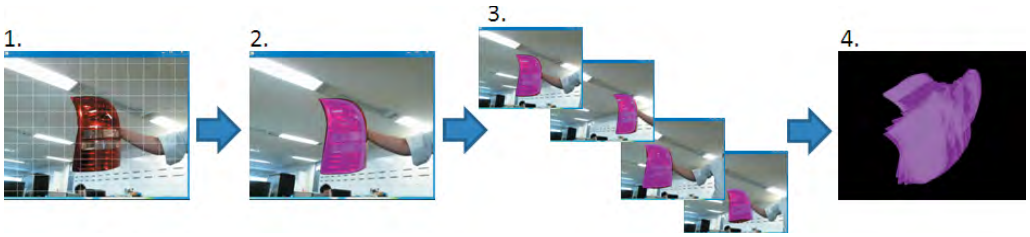

図 2　システム操作仕様

1)　ビデオ映像内に対象物体（図 2 では赤いテールランプ）を映し出す．
2)　ビデオ映像内の対象物体から三次元位置・姿勢を認識して，映像に三次元 CAD データを重ね表示する．
3)　ビデオ映像内の対象物体の動きに合わせて対象物体をトラッキングし，三次元 CAD データを重ね表示する．トラッキング時に各映像における対象物体の三次元位置・姿勢を認識する．
4)　3)のトラッキングにより抽出された三次元空間内の動作を他のビューアで再生する．

3)の処理において，リアルタイムで対象物体の三次元トラッキング処理を行っている．そのため，3)の処理終了と同時に，他のビューアを用いて対象物体の三次元空間内の動作を再生することができる．これは，ビデオ映像から対象物体の三次元空間内の動作を認識できているこ

とを意味する.

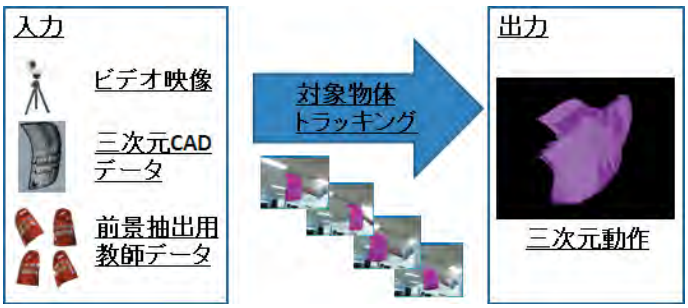

図3　システム入出力仕様

図3は，本システムの入出力仕様を表している．本システムは単眼のビデオカメラで撮影した映像に対して，三次元CADデータおよび映像から対象物体を抽出する処理に利用するデータ（前景抽出用教師データ）を用いることで，対象物体の三次元トラッキングを行う．本システムは，カメラは固定しておき対象物体が動くものとする．また，事前にカメラのキャリブレーションを行っておき，オートズームは無効にしておく（物体とカメラの両方が動く場合およびオートズームが有効の場合については4章で述べる）.

カメラのパラメータが既知であり対象物体の形状がわかっている場合，原理的に対象物体の三次元位置・姿勢は一意に定まる（詳細については3.1節で述べる）．ビデオ映像を，静止画像が連続的に表示されているものとみると，各静止画像に対して対象物体の三次元位置・姿勢が定まれば，連続的に位置・姿勢を取得することで三次元トラッキングができる.

2.2　システムの技術的優位性

本節では，既存研究と比較した本システムの技術的優位性について述べる.

単一のビデオ映像内の対象物体をトラッキングする手法は，画像処理の分野で多数提案されている．特に，パーティクルフィルターを用いた対象物体のトラッキングは非常によく知られた手法である[4]（図4）.

図4　パーティクルフィルターを用いた対象物体のトラッキング

これらの対象物体のトラッキング手法の多くは二次元画像内における対象物体のトラッキングであり，本システムで行っている三次元トラッキングは実現できない.

マーカーレス AR の研究において，映像内の三次元空間を認識するシステムに関する研究がいくつか存在する．特に単眼カメラを用いて三次元空間を認識する研究としてPTAM（Parallel Tracking and Mapping）[5]がよく知られている（図 5）．

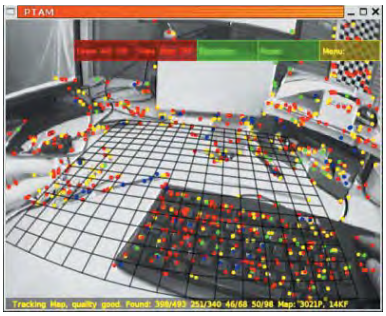

図 5　PTAM による三次元空間の認識

PTAM は，映像内の画像特徴点から平面に近い部分を認識することで三次元座標系を定め，各フレーム間の画像特徴点の相対的な位置関係から三次元位置を認識する．PTAM は空間追跡を目的として設計されており，物体追跡に利用するには三次元空間の認識精度や追跡精度の面で十分でない．

本システムは，対象物体を二次元画像内でトラッキングするのみではなく，三次元空間におけるトラッキングを実現している点で，多くの画像処理におけるトラッキング技術と異なる．また，実現する三次元空間の認識精度および追跡精度においてマーカーレス AR の技術を上回る．

3.　二次元画像から三次元トラッキングを行う方法

本章では，三次元 CAD データを用いて二次元画像から対象物体の三次元位置・姿勢を定める原理と本システムで採用したアルゴリズムについて述べる．3. 1 節において，三次元 CAD データを用いて二次元画像から対象物体の三次元位置・姿勢を定める原理について述べる．3. 2 節以降で，対象物体の三次元位置・姿勢を定めるアルゴリズムについて述べる．

3. 1　二次元画像から三次元位置・姿勢を定める原理

本システムではピンホールカメラモデルを用いて二次元画像と三次元空間を結び付ける．はじめにピンホールカメラモデルについて説明する．

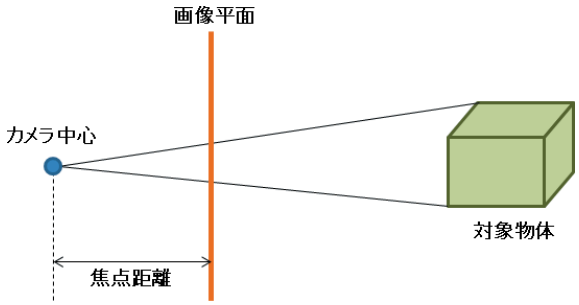

図 6　ピンホールカメラモデル

ピンホールカメラモデルは，カメラ中心と画像平面を持つ（図 6）．対象物体上のある点は，その点とカメラ中心を結んだ直線と画像平面の交点上に映し出される．カメラ中心と画像平面との距離を焦点距離と呼ぶ．また単位距離あたりのピクセル数を解像度と呼ぶ．本システムでは，事前にキャリブレーションを行い，カメラの焦点距離および解像度を算出しておく．

次に，二次元画像と三次元空間の座標系の関係（図 7）を次のように定める．二次元画像平面上の座標系を，幅方向を w 軸，高さ方向を h 軸とする．三次元空間の座標系を，カメラ中心を原点，カメラ中心から画像平面に垂直に下した方向を z 軸，z 軸と直交し画像平面の w 軸と平行な方向を x 軸，z 軸と直交し画像平面の h 軸と平行な方向を y 軸とする．また z 軸と画像平面の交点を画像中心とする．

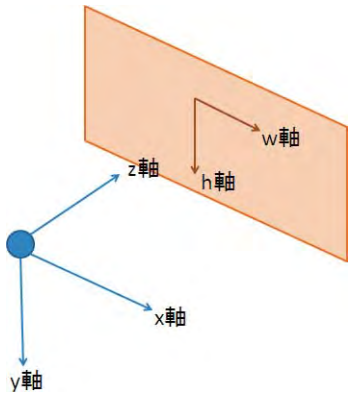

図 7　二次元画像平面と三次元空間の関係

本システムの開発にあたって解決すべき課題は，「二次元画像に映る対象物体の三次元空間位置・姿勢を定めること」である．三次元空間位置・姿勢が定まった対象形状の二次元画像上における像は，画像平面と三次元空間の対応関係を用いれば簡単に求まる．しかし，本システムで解決すべき課題はその逆問題であり，解決は単純ではない．

この課題に対して，筆者らは，「二次元画像に映る対象物体の境界」と「二次元画像平面に投影した対象形状の像の境界」が一致する三次元空間位置・姿勢を算出するという方針を取った（図 8）．

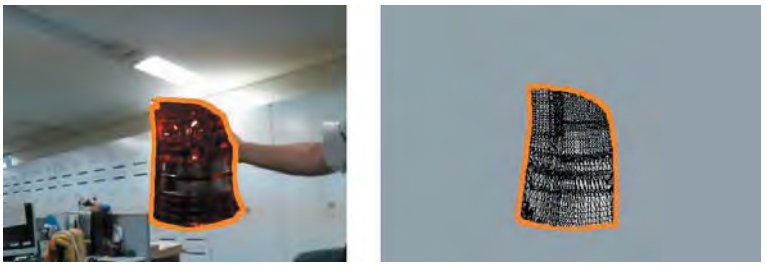

図 8　二次元画像の対象物体の境界と二次元画像平面に投影した対象物体の像の境界
（左：二次元画像の対象物体の境界，右：二次元画像平面に投影した対象物体の像の境界）

対象物体上の点が画像平面に投影された像の境界線上の点になるとき，対象物体はその点でカメラ中心とその点を結ぶ直線と接している（図 9）．つまり，二次元画像上に投影した対象物体の像の境界となる点は，カメラ中心と二次元画像の対象物体の境界上の点を結んだ直線上に存在する．

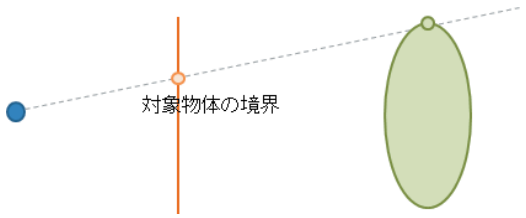

図 9　二次元画像平面上の境界

そこで，以下 1)～ 4)のステップ（図 10 の(1)～(4)に対応）により対象物体の三次元位置・姿勢を算出できると考えた[*1]．

1)　二次元画像から対象物体の境界を抽出する．
2)　カメラ中心と境界線のピクセルを結んでできる直線の集まりにより放射形状を作成する．
3)　前のフレームの画像情報を基に対象物体の位置を補正する．
4)　放射形状上に対して位置合わせをする．

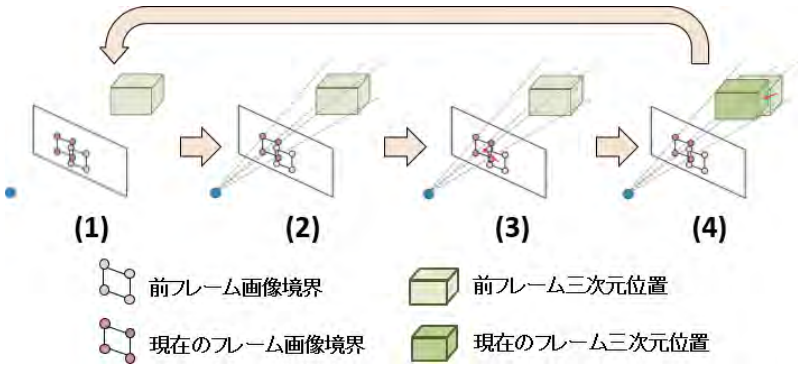

図 10　放射形状を用いた対象形状の位置・姿勢算出

この処理を行うためには，以下の三つの要素技術が必要となる．

- 二次元画像から対象物体の境界を抽出する技術（図 10 の(1)）．
- 前のフレームの対象物体位置を用いた位置補正技術（図 10 の(3)）．
- カメラ中心と二次元画像の境界線上の点から生成される放射形状に対して，対象物体の三次元位置合わせを行う技術（図 10 の(4)）．

以降，3. 2 節において二次元画像から対象物体の境界を抽出する処理について，3. 3 節において前のフレームの対象物体位置を用いた位置補正処理について，3. 4 節において対象物体の三次元位置合わせを行う処理について述べる．

3. 2　二次元画像からの対象物体境界の抽出

本システムでは，二次元画像から前景抽出処理により対象物体を抽出し，抽出した対象物体から境界を抽出する．前景抽出処理および境界抽出処理はリアルタイムに行っている．前景抽出を行う方法として，グラフカットを用いた前景抽出がよく知られている[6]．この方法は，抽出精度が高い半面で，最小カット問題と呼ばれる問題を解く必要があるため処理コストが高い．そのため，本システムで必要とされるリアルタイム処理には適さない．

そこで，抽出精度は多少劣るがリアルタイムに適した高速な前景抽出を本システムでは採用した．そして，後処理である位置合わせ処理において，前景抽出により発生したノイズを許容するような処理を行う方針を取った．

本システムにおける前景抽出処理として，教師あり機械学習を用いた処理を採用した．教師あり機械学習とは，事前に教師データと呼ばれる正解を表すデータから学習データを作成し，処理時に入力データと学習データを用いる処理を指す（図 11）．

図 11　教師あり機械学習

本システムの前景抽出処理では，対象物体の画像および背景の画像を教師データとして用いた．これらから抽出したデータを混合正規分布によりモデル化することで学習データを作成した．混合正規分布とは，複数の正規分布に和が 1 となる重みを掛けて，その総和を取ることで作成した確率分布であり，複数の山をもつ（図 12）．混合正規分布は，データ内に頻度のピークが複数あるデータをモデル化する場合に用いられる確率分布の一つである．

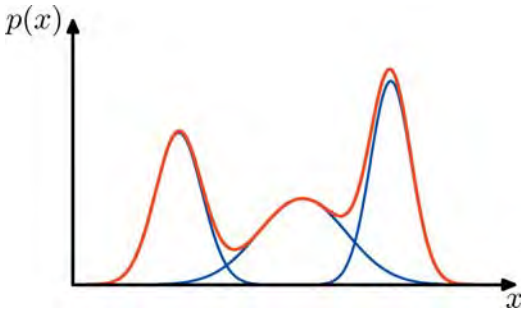

図 12　混合正規分布

次にモデルの作成方法について述べる．モデル作成のために正解データの各ピクセルの色を三次元の座標値で表現する（ここでは，この座標値を色座標値と呼ぶことにする）．色の表現方法として RGB 値や HSV 値がよく知られている．本システムでは，人間の感覚に近い色の表現である HSV 値を用いた．HSV 値は円錐形状内の一点として表される（図 13）．

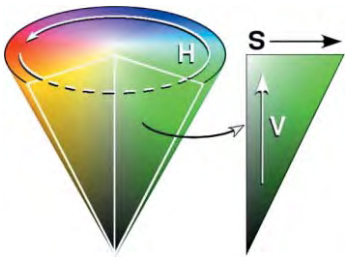

図 13　HSV 色空間

各ピクセルの RGB 値を HSV 値の値に変換し，さらに HSV 値を表す円錐を三次元空間に埋め込んだときの対応する座標値に変換することで，各ピクセルの RGB 値を三次元の色座標値に変換する．

前景および背景の各ピクセルを色座標値に変換し，これらの色座標値の確率分布を混合正規分布の最尤推定に算出する．最尤推定とは，与えられたデータの発生確率が最も高くなるような確率分布を算出する手法である．混合正規分布の最尤推定については，EM アルゴリズムを用いた方法がよく知られている[7]．算出した混合正規分布を学習データとして使用する．

この混合正規分布の各正規分布に，「前景」・「背景」のいずれかをラベリングすることができる．直観的には，正規分布の中心付近に前景の色座標値が多く集まっている場合はその正規分布を「前景」としてラベリングし，背景の色座標値が多く集まっている場合はその正規分布を「背景」としてラベリングする．図 14 では左側の正規分布の付近には前景のデータが集まっており，右側の正規分布の付近には背景のデータが集まっている．そのため，左側の正規分布は「前景」を表す正規分布，右側の正規分布は「背景」を表す正規分布としてラベリングされる．

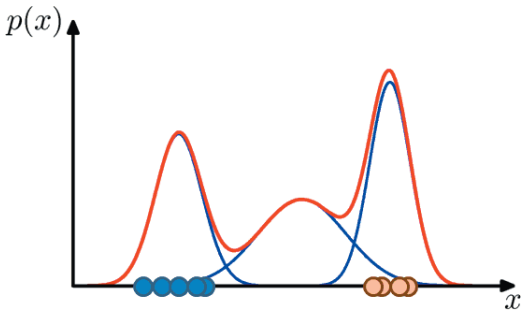

図 14　混合正規分布を用いた前景・背景のラベリング
（青：前景のデータ，赤：背景のデータ）

次に作成した学習データの前景抽出方法について述べる．本システムの前景抽出処理は入力された画像の各ピクセルについて前景または背景の判定を行い，前景と判定されたピクセルを抽出することで前景抽出を行う．

各ピクセルを色座標値に変換し，学習データの各正規分布の中心とのマハラノビス距離を算出する．マハラノビス距離とは，分布の分散を考慮した距離である．マハラノビス距離が最小となる正規分布のラベル（「前景」または「背景」）をそのピクセルのラベルとすることで，各

ピクセルが前景であるか背景であるかを判定する．この処理は，各ピクセルに対してマハラノビス距離を計算するのみであるため，非常に高速である．また，ピクセル間の隣接関係を考慮しないため，すべてのピクセルを独立に計算できる．そのため，複数 CPU や GPGPU を用いた並列処理による高速化と非常に相性がよい方法である．

図 15 は，本処理を用いて対象物体を抽出した結果である．対象物体が抽出されていることが確認できると同時に，対象物体とは異なる場所も残っていることがわかる．

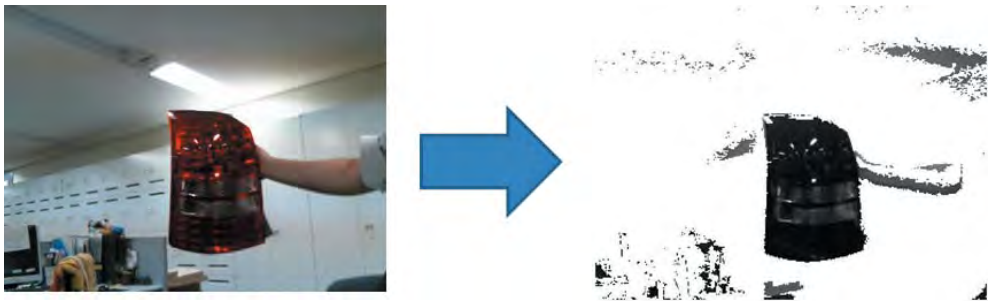

図 15　対象物体抽出処理

3. 3　前フレームの対象物体位置を用いた位置補正処理

対象物体の位置を定めるために，対象物体の三次元位置の初期位置を定めて，初期位置を基に三次元位置合わせ処理を行う．3. 2 節における対象物体抽出処理は，高速に処理を行う代わりにノイズが残るため，ノイズを考慮した位置合わせ処理を行う必要がある．

対象物体のトラッキング時は，前のフレームの映像に対して，対象物体の三次元位置が定まっている．そこで，前フレームの対象物体の三次元位置を用いて，現在のフレームの映像に対する対象物体の初期位置を定める．最初のフレームの位置については，投機的にいくつかの向きを持つ対象物体に対して位置合わせを行い，もっとも当てはまりのよい位置を初期位置としている．

位置補正処理は，前のフレームの画像と現在のフレームの画像からそれぞれ対象物体の中心を算出し，その差分を対象物体の中心が乗る xy 平面上における差分量に変換して，前のフレームの三次元位置に差分量を足すことで補正する．

ここで画像上の対象物体の中心位置の算出方法が課題となる．3. 2 節における対象物体抽出処理はノイズを含んでおり，対象物体のピクセルを単純に平均するとノイズの影響により中心位置がずれてしまう．そのため，ノイズに対して頑健な中心位置推定が必要となる．

そこで本システムでは，コーシー分布を用いた中心推定を採用した．単純な平均位置の算出は，正規分布の最尤推定により求まる中心位置と一致する．一般に，正規分布の最尤推定はノイズに対して敏感であることが知られており，そのため単純平均はノイズに弱い．

図 16 はコーシー分布と正規分布を比較した図である．コーシー分布は正規分布と比較して裾の部分の確率が大きく，ノイズに対して頑健な性質をもつことが知られている．対象物体のピクセルに対して，コーシー分布を用いた最尤推定により中心を推定する．

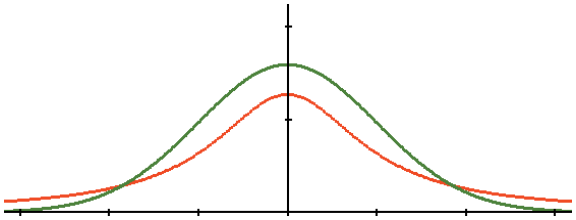

図 16　コーシー分布と正規分布の比較
（赤：コーシー分布，緑：正規分布）

単純平均で算出した中心位置とコーシー分布の最尤推定により算出した中心位置の比較を図 17 に示す．緑の点が単純平均により算出した中心位置，赤の点がコーシー分布を用いて算出した中心位置である．図 17 より赤の点の方が緑の点と比べてより物体の中心を捉えていることがわかる．

図 17　コーシー分布を用いた平均と単純平均の比較
（赤：コーシー分布を用いた平均　緑：単純平均）

3. 4　対象物体の三次元位置合わせ

補正された位置を初期位置とした対象物体の三次元位置合わせについて述べる．三次元位置合わせには，ICP（Iterative Closest Point）アルゴリズム[8]を用いる．ICP アルゴリズムとは，サンプル点と位置合わせ基準との対応関係を作成し（一般的には最近点を用いる），対応の差異を最小化する変換を繰り返し施すことで位置合わせを行う手法である（図 18）．

図 18　ICP アルゴリズム

はじめに，対象物体を画像平面に投影したときに，二次元画像として境界となる点をサンプル点として取得する．次にこれらのサンプル点と放射形状の対応を最近点により求め，対応を最小化する変換を施す処理を繰り返す（ICP アルゴリズム）．対象物体は位置が補正されており，かつ放射形状とは最近点により対応付けられるため，対象物体から見て遠くの点は変換に

影響を与えない．そのため，対象物体の抽出処理でノイズが残っていても頑健に位置合わせを行うことが可能となる．

3.5　三次元トラッキングのまとめ

本章では，三次元 CAD データを用いて二次元画像から対象物体の三次元位置・姿勢を定める原理と本システムで採用したアルゴリズムについて述べた．筆者らは，この課題に対して「二次元画像に映る対象物体の境界」と「二次元画像平面に投影した対象形状の像の境界」が一致する三次元空間位置・姿勢を算出するという方針を取ることで，リアルタイム三次元トラッキングを実現した．

4. お　わ　り　に

本システムでは，カメラは固定であり，オートズームを無効とすることを仮定した．これらの仮定を外すためにはフレーム毎のカメラモデルから定まる座標系の差異を補正できればよい．ビデオ映像のフレーム間の座標系の補正に関する研究は SfM（Structure from Motion）と呼ばれ，多くの研究成果がある[9]．本システムと SfM を組み合わせることで，これらの仮定への対応が可能であると考えている．また，位置合わせ精度や処理速度に関する課題への対応策として，GPGPU やマルチ CPU を用いた並列化の適用を予定している．

本稿では，単眼カメラで撮影したビデオ映像から対象物体の三次元トラッキングを行うプロトタイプシステムについて述べた．本システムは，これまで日本ユニシスが蓄積してきた三次元データ処理と画像処理を組み合わせることで実現した新しい仕組みの一例である．特別な計測装置なしで三次元トラッキングのシステムを構築することができるので，計測装置の使用が難しい実際の製造現場で，ものの動きをトラッキングするような場面への導入が可能となる．本技術により，これまでにない新たな価値をユーザに提供できるようになると考えている．

＊ 1　本手法は日本ユニシス株式会社にて特許出願中の手法である．

参考文献
［1］「特集：CADCEUS」，ユニシス技報，日本ユニシス，Vol.14 No.4，通巻 44 号，1995 年 2 月．
［2］「特集：デジタルエンジニアリング」，ユニシス技報，日本ユニシス，Vol.23 No.3，通巻 79 号，2003 年 11 月．
［3］「特集：自動車産業」，ユニシス技報，日本ユニシス，Vol.23 No.4，通巻 80 号，2004 年 2 月．
［4］ M. Isard A. Blake, “CONDENSATION – conditional density propagation for visual tracking”, Int. J. Computer Vision, Vol.29-1, pp.5-28, 1998.
［5］ G. Klein D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces”, In Proc. of ISMAR’07, pp.225-234, 2007.
［6］ C. Rother V. Kolmogorov A. Blake, “GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts”, ACM Trans. Graph., vol.23, pp.309-314, 2004.
［7］ C. M. ビショップ，“パターン認識と機械学習 下（ベイズ理論による統計的予測）第 9 章 混合モデルと EM”，丸善出版，2012 年 2 月
［8］ Z. Zhengyou, “Iterative point matching for registration of free-form curves and surfaces”, Int. J. Computer Vision, Vol.13-12, pp.119-152, 1994.
［9］ R. Hartley A. Zisserman, “Multiple View Geometry in Computer Vision”, Cambridge University Press, March 2004.

執筆者紹介　武　井　宏　将（Hiromasa Takei）
2004年 日本ユニシス(株)入社．入社時よりCAD/CAM分野のシステム開発業務に従事．2013年より，画像処理・三次元形状処理の研究開発を担当．