

日常生活における道具の三次元運動のステレオ画像計測

Stereo Image Measurement of 3D Object Motion in Human Daily Activities

○牛崎学, 出口光一郎

○Manabu Ushizaki, Koichirou Deguchi

東北大学大学院情報科学研究科

Graduate School of Information Sciences, Tohoku University

キーワード: ステレオ画像 (Stereo Image), 三次元運動計測 (3D-Motion Measurement), 色抽出 (Color Extraction)

連絡先: 〒980-8579 仙台市青葉区荒巻字青葉 東北大学 大学院 情報科学研究科
システム情報科学専攻 出口・岡谷研究室

E-mail: ushizaki@fractal.is.tohoku.ac.jp

1. はじめに

人が道具を操るとき、「うまい」人というのが存在する。包丁さばきの「うまい」人から、手品のように感動させられるような「うまい」技術を持つ人まで多くの「うまい」人がいる。残念ながら、そのような「うまい」技術というものは、その人が引退することで、失われてしまう。そのため、近年企業などでは、技術・技能の伝承・継承に対する取り組みが活発に行われている。「うまい」技術・技能の伝承・継承を行うには、その道具がどのように操られているかを分からなくてはならない。これは行っている本人ですら説明がうまく出来ない場合が多々あり、技術・技能の伝承・継承に時間がかかったり、途絶えてしまう原因の一つである。そのため、「うまい」操り方を解析することが求められている。しかし、人が操る道具とゆうものは複雑かつ繊細な動きをする。したがってその道具

の運動を計測するには、“人の行動に影響を与えず”，かつ“複雑な運動の計測に耐えるような計測方法”，が必要となる。

計測手法として、人の動きを計測し結果として道具の運動を計測する手法と、道具そのものの運動を計測する手法の、二通り考えられる。

前者の代表的な方法として、モーションキャプチャーがあげられる。これは人間の動きを捉えるものとして実用にいたっている。欠点としては、細かい動作は測れず、また計測にスタジオが必要となってしまう、大掛かりなものになってしまうこともある。

後者としては、道具に接触型のセンサを取り付ける方法や、画像を用いた計測方法がある。しかし、センサを用いる場合、人の運動を計測することを考えると、センサを取り付けることで、人に違和感を与えてしまい行動に影響を与えてしまう恐れ

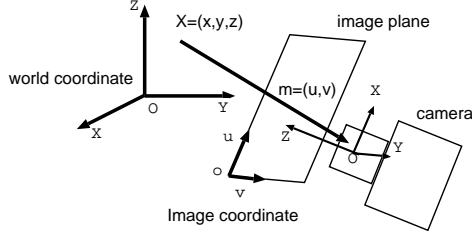


Fig. 1 透視投影変換

がある。一方画像を使うと、運動の計測を非接触できるので人の運動に影響を与えず、自然な運動を計測できる。画像を用いた運動計測方法は様々な研究されている^{1, 2)}。

本研究では、ステレオ画像を用いて道具の運動を計測する。実際の画像について計測を行い、ステレオ画像による道具の運動計測に対する有用性を調べた。

2. 計測の原理

2.1 ステレオの幾何学

Fig.1よりワールド座標系で、三次元上の点が $\mathbf{X} = \begin{pmatrix} X & Y & Z \end{pmatrix}^t$ で表されるとき、ワールド座標系から画像座標系への変換は

$$w \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{P} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (1)$$

と表すことができる。ここで、 w は0でない定数、 \mathbf{P} は、 3×4 の行列であり投影変換行列と呼ばれる³⁾。

ここで、 w を消去すると、

$$\mathbf{A}\mathbf{X} = \mathbf{b} \quad (2)$$

と表される。ただし、

$$\mathbf{A} = \begin{pmatrix} P_{31}u - P_{11} & P_{32}u - P_{12} & P_{33}u - P_{13} \\ P_{31}v - P_{21} & P_{32}v - P_{22} & P_{33}v - P_{23} \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} P_{14} - P_{34}u \\ P_{24} - P_{34}v \end{pmatrix}$$

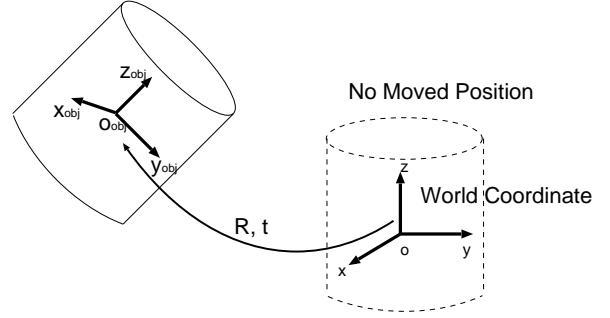


Fig. 2 三次元運動時の座標系(\mathbf{R}, \mathbf{t})

である。

ここで、Fig.2のような場合を考える。 \mathbf{X} がある物体上の点だとし、その物体がある三次元運動をしている。その三次元運動による物体上の点のワールド座標は式(3)で表される。

$$\mathbf{X} = \mathbf{R}_{obj}^{world} \mathbf{X}_{obj} + \mathbf{t}_{obj}^{world} \quad (3)$$

ここで、 \mathbf{X}_{obj} は物体座標系での位置、また、 \mathbf{R}_{obj}^{world} 、 \mathbf{t}_{obj}^{world} は物体座標系からワールド座標系に対する姿勢と平行移動を表す 3×3 の回転行列と3次元の並進ベクトルである。

このとき、式(2)、(3)から、 \mathbf{X} を消去すると、

$$\mathbf{A}(\mathbf{R}_{obj}^{world} \mathbf{X}_{obj} + \mathbf{t}_{obj}^{world}) = \mathbf{b} \quad (4)$$

という式が得られる。ここで未知数は \mathbf{R}_{obj}^{world} の回転を表す3パラメータと、 \mathbf{t}_{obj}^{world} の位置を表す3パラメータである。一方式は2本得られる。したがって、カメラの投影変換行列 \mathbf{P} が既知であり、物体座標系での位置 \mathbf{X}_{obj} と、それが画像上に投影された位置 $\begin{pmatrix} u & v \end{pmatrix}^t$ が3点以上対応付けられれば、式(4)を用いて \mathbf{R}_{obj}^{world} 、 \mathbf{t}_{obj}^{world} を計算することができる。実際には、式(4)を物体上の点の全てに適応し、

$$\min_{\mathbf{R}_{obj}^{world}, \mathbf{t}_{obj}^{world}} \sum_{k=0}^{n-1} \|\mathbf{A}_k(\mathbf{R}_{obj}^{world} \mathbf{X}_{obj,k} + \mathbf{t}_{obj}^{world}) - \mathbf{b}_k\|^2 \quad (5)$$

となるような、 \mathbf{R}_{obj}^{world} 、 \mathbf{t}_{obj}^{world} を求めている。ここで n は物体上の点の総数であり、 k は点の番号を表す。

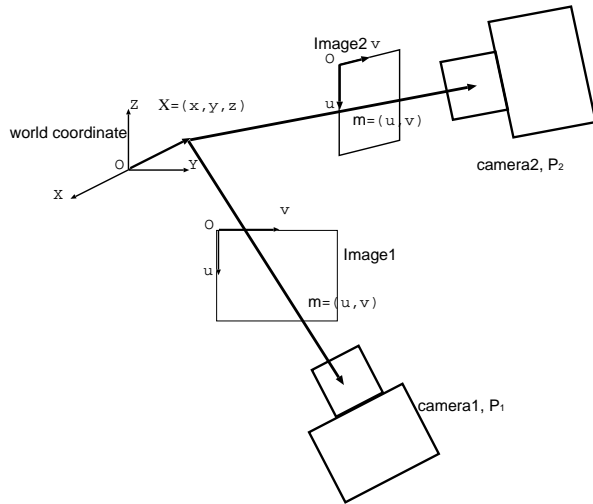


Fig. 3 ステレオカメラの構成

式(5)は、ステレオ画像に簡単に拡張できる。ステレオカメラにおいて、同一の点を見る場合、Fig.3のような関係が得られる。ステレオカメラをそれぞれカメラ1、カメラ2と呼ぶ。ここで、各カメラについて式(1)が成り立つので、カメラ1の投影変換行列を $\mathbf{P}^{(1)}$ 、またその画像上の点を $\mathbf{m}^{(1)}$ と置く。同様に、カメラ2についても投影変換行列を $\mathbf{P}^{(2)}$ 、またカメラ1の画像上の点と同じ点を表すような順にその画像上の点を $\mathbf{m}^{(2)}$ と置く。このとき式(2)より $\mathbf{A}^{(j)}\mathbf{X} = \mathbf{b}^{(j)}$, ($j = 1, 2$)が作られる。したがって、ステレオ画像の場合の運動計測の式は、

$$\begin{pmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \end{pmatrix} (\mathbf{R}_{obj}^{world} \mathbf{X}_{obj} + \mathbf{t}_{obj}^{world}) = \begin{pmatrix} \mathbf{b}^{(1)} \\ \mathbf{b}^{(2)} \end{pmatrix} \quad (6)$$

と表される。これを用いると式(4)より精度の高い運動計測が期待できる。式(6)の場合も式(4)と同様に物体上の点がどこに投影されたかを3点以上対応付けられないと運動計測が出来ない。ステレオの場合にも実際の計算には、式(5)のときと同様に、式(6)を物体上の点の全てに適用し、

$$\min_{\mathbf{R}_{obj}^{world}, \mathbf{t}_{obj}^{world}} \sum_{k=0}^{n-1} \left\| \begin{pmatrix} \mathbf{A}_k^{(1)} \\ \mathbf{A}_k^{(2)} \end{pmatrix} (\mathbf{R}_{obj}^{world} \mathbf{X}_{obj,k} + \mathbf{t}_{obj}^{world}) - \begin{pmatrix} \mathbf{b}_k^{(1)} \\ \mathbf{b}_k^{(2)} \end{pmatrix} \right\|^2 \quad (7)$$

となるような $\mathbf{R}_{obj}^{world}, \mathbf{t}_{obj}^{world}$ を求めている。

2.2 画像上の点と物体座標上の点の対応付け

ここまでで、物体上の点が画像上で計測できてかつその物体座標系での位置との対応が、三点以上取れば物体の運動を計測できることが分かった。そこで、どういった特徴を使えばステレオ画像上の点を計測できて、かつその物体座標系での位置が分かるかが問題となる。

画像上の特徴抽出の方法としては、HarrisのCorner Detector⁴⁾や、SUSAN⁵⁾などがある。しかし、これらの特徴抽出法では、物体上の特徴が抽出できる保障が無く、また物体上に合ったとしても、物体のどこが抽出できたのかが分からない。

そのため、本研究では色抽出によってステレオ画像上の位置と、物体座標系での位置の対応取得を行った。

背景に無く、物体上にだけある色の抽出を行うことで、物体上の特徴を抽出できる。ただし、物体上に都合よく色が分布していることはまれである。また、似た色は区別するのが難しいので、何種類もの色を使うのは無理がある。そこで、本研究では色抽出用に板状で円形にくりぬいたマーカーを使用した、また、同色のマーカーを複数枚使用している。マーカーの画像上の位置は色抽出したときの連結された領域の重心とした⁶⁾。マーカーの物体座標上の位置は、物体に取り付けた面の中心とした。(例 Fig.4).

背景の色を一部抽出してしまうことがあるが、まばらに分布するような点は、連結された数に閾値を設けることでそのような点は除去できる。数で除去できない場合でも計測する物体上のマーカーを画像に投影することで、それらしい点かどうかの確認ができる。

ここで本研究で用いた、画像上の色抽出された

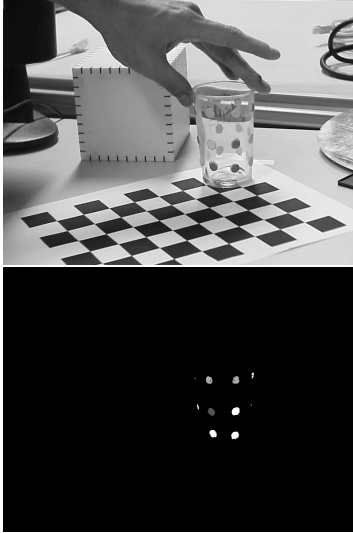


Fig. 4 色抽出の例. 対象の画像(上)と, 結果の画像(下)

マーカーの位置と物体上の位置との対応付け, 及び運動計算の手順を書く.

- 1) 各カメラの画像に色抽出を行いマーカーの画像座標の候補を得る. 抽出できなかった場合にそなえ候補に, 見えていない場合の座標を加える. 以下の手順で見えていない場合の座標を使った場合は, 計算時その座標の重みを0にする.
- 2) 一方の画像から得られた画像座標の候補について同じ色の中から, ランダムに物体上の点との対応を仮定して式(5)を用いて運動計測する.
- 3) 手順2)の計測結果を使って物体上の点を各カメラの画像にそれぞれ投影する. 投影点と抽出された点との間で, 同じ色でかつ最も近い点との間の距離がある閾値以下になる数を数える.
 - (a) 前回の数より多くなれば, そのときの現在の画像について投影した画像上の点の配列を保存する.
 - (b) 前回の数と等しい場合, 前回より距離が小さければ, そのときの現在の画像について

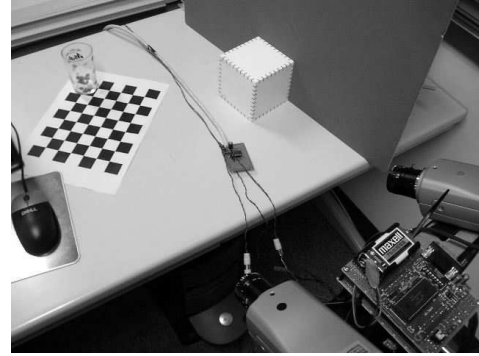


Fig. 5 計測の様子. ステレオカメラ(右下). 計測道具”コップ”(左上)

投影した画像上の点の配列を保存する.

- 4) 手順2),3)を一定回数繰り返す.
- 5) もう一方の画像から得られた画像座標についても手順2),3),4)を行う.
- 6) 以上の手順で得られた各画像上の点の配列を使って式(7)により運動計算を行う.

3. 実験と考察

ステレオ画像はSony製のデジタルカメラ”DFW-X700”を二台用いて撮影した. キャリブレーションは別途行い, 各カメラの投影変換行列 P を計算しておいた. また, 二台のカメラの間の距離は約 $0.2(m)$. カメラとワールド座標の原点との距離は約 $0.75(m)$ である. 計測する対象には, コップを選んだ. コップのサイズは, 高さ $9.3(cm)$, 口の直径 $6.0(cm)$, 底の直径 $4.5(cm)$ であり, 円錐台の形状をしている. マーカーは直径 $0.9(cm)$, 合計9点取り付けた.

3.1 物体が動かない場合の位置と姿勢の計測

本研究で行う計測方法は, 画像一組についての位置と姿勢を計測する方法なので, 画像一組からはどれくらいの精度で得られるのかが, 問題となる. そこで, 動かないものを計測した場合, どれ

Table 1 固定時の計測結果

種類	成分	平均(cm)	標準偏差(cm)
原点座標	x	0.082	0.053
	y	-0.190	0.074
	z	-0.217	0.037
Z軸方向の 単位ベクト ルの座標	x	-0.011	0.010
	y	0.016	0.012
	z	0.9999	0.0003

だけ位置，姿勢がずれるかを調べる．計測の様子を，Fig.5に示す．

姿勢の評価はそのままでは難しいので，物体座標系のz軸方向の法線ベクトルの振る舞いで，姿勢を評価する．11枚の画像について調べた．計測結果はTable. 1. Table.1より平均座標の値に偏りがあることが分かる．キャリブレーションの精度を変えても同程度の偏りを持つので，マーカーを取り付けたときマーカーの物体座標系の位置の計測に誤差が乗っており，それが偏りを生じたためと考えられる．計測する道具が複雑な形状のものになるにつれこの誤差は大きくなると考えられるので，この改善方法は今後の課題である．

3.2 実際に人が道具を扱う場合の位置と姿勢を計測する際の問題点

ここでは，実験を通して問題となった事柄を述べる．これは，実際に人が道具を扱う場合の計測時に考える，本研究の手法の問題点である．

1) 人がマーカー部分を持ってしまう場合

このとき，計測の精度が下がってしまうことが考えられる．さらに，マーカーを持ったために人に違和感を与えてしまい，人の自然な運動を計測するという趣旨に反してしまう恐れがある．このため，マーカーの取り付けには道具の使い方に応じた位置を探す必要がある．この様子の例は，Fig.6.

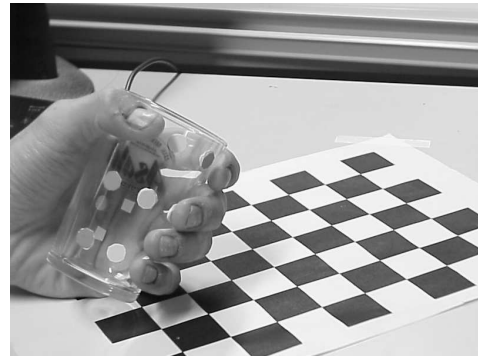


Fig. 6 人がマーカー部分を持ってしまう場合の例．マーカーをもたれている

2) 道具を早く動かされる場合

道具周辺の画像がぼやけてしまい色抽出がうまくいかないことがある．これについては，カメラのシャッタースピードを上げることである程度解決できる．この様子の例は，Fig.7.

3) 道具を大きく傾けられた場合

各カメラから見えるマーカーの面が異なってしまったり，複数のマーカーの画像上の座標が重なり，画像上のマーカーが物体ではどれに対応するのかが物体座標系上の点を投影するだけでは判定しづらいことがある．これについては，画像上に投影する以外のアプローチが必要と考えられる．この様子の例は，Fig.8.

4. まとめ

ステレオ画像中の道具の三次元運動の計測法を述べた．まず，透視投影変換の式と物体の運動の式より，運動計測時に最小化すべき式を導いた．次に，画像上の点と物体上の点の対応のとり方を述べた．

また，実験を通してこの計測方法の問題点が様々判明した．これらの問題点を解決させることで，この計測法がより有用になると考えられる．

参考文献

- 1) Tom Drummond and Roberto Cipolla: Real-Time Visual Tracking of Complex Structures. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, **24**(7): 932-946, JULY 2002.
- 2) Luca Vaahetti, Vincent Lepetit, and Pascal Fua: Stable Real-Time 3D Tracking Using Online and Offline Information. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, **26**(10): 1385-1391, OCTOBER 2004.
- 3) 出口光一郎(著): ロボットビジョンの基礎, コロナ社, pages 12-79, (2000)
- 4) C. G. Harris and M. Stephens: A Combined Corner and Edge Detector. In Proc. of the 4th Alvey Vision Conference, pages 147-151, 1988.
- 5) S. M. Steve and M. Brady: SUSAN — a new approach to low level image processing. Int'l J. Comput. Vision, **23**(1):45-78, 1997.
- 6) James Bruce: Real-Time Machine Vision Perception and Prediction. <http://www-2.cs.cmu.edu/jbruce/cmvision/JBThesis00.pdf>, pages 17-18, 2000.

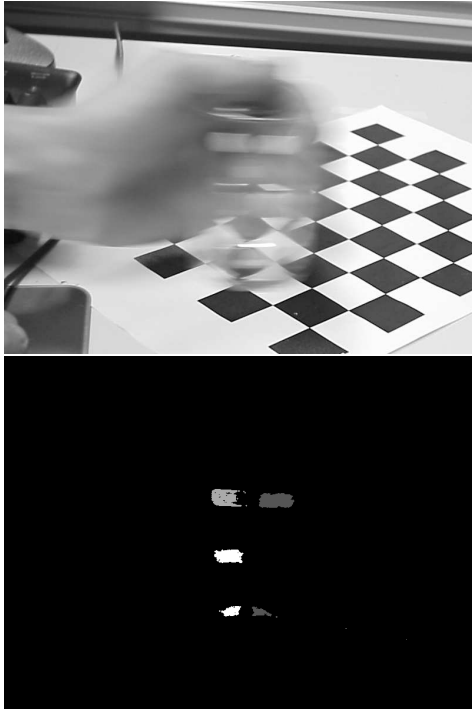


Fig. 7 道具を早く動かされる場合の例. マーカーの座標を正しく抽出できない. 抽出前(上). 抽出後(下).



Fig. 8 道具を大きく傾けられた場合の例. 同色のマーカーがほぼ同じ座標にあり, 物体座標系上の点を投影しても区別がつかない