

## 実映像中のランドマーク情報を利用した拡張現実感の1生成手法

1W-2

河村 貴弘 館村 純一 坂内 正夫

東京大学 生産技術研究所

e-mail: kawamura, tatemura, sakauchi@sak.iis.u-tokyo.ac.jp

## 1 はじめに

AR(Augmented Reality, 拡張現実感)という研究分野が注目を集めはじめています。ARは狭義では、シースルー型の頭部搭載型ディスプレイ(HMD:Head Mounted Display)を使って、コンピュータが生成する画像を現実の背景にオーバーラップさせるシステムを指すが、最近では、コンピュータがモノやユーザの行動を観察し、それに基づいて情報検索などの処理を実行するシステムも包含するようである。前者の例<sup>1)</sup>では、トラッキングシステムを利用することで、実映像と3次元CG(Computer Graphics)の3次元的合成を行っており、現実感に注意が払われている。しかし、トラッキングシステムは手術室のような限られた空間の中でしか利用できない。後者の例<sup>2)</sup>では、被写対象物に添付したカラーコードを認識することによってデータベースの検索を行ない、結果を実映像に重ねて表示しており、現実を拡張することに注意が払われている。しかし、このような人為的な指標を建築物などの大きな対象物に添付することはできないし、前者のような3次元的合成を念頭においていない。

以上のような考察から、両者の目的を共に満たすには立体認識が必要があると考え、より広範な範囲で利用するために実映像中のランドマーク情報を利用したARシステムの構築を行なった。この方法だと、謂わばテント式(Wherever You Want)の利用が可能になり、キャリブレーションを行なえばどこでもARシステムを構築できる。

## 2 PnP問題

立体認識の手法には様々なものがあるが、単眼であり、受動的な認識手法である透視n点問題(Perspective n-Point Problem:PnP問題)をARに利用した。カメラの焦点距離などの撮影パラメータが既知であると仮定すると、カメラの姿勢は最大3個の回転パラメータ

と最大3個の平行移動パラメータによって決定される。画像中の1点の3次元座標からは、2個のパラメータを求めることができるため、一般には3点以上の点の3次元座標が既知である必要がある。ただし、精度の向上や計算の平易化を目的として4点以上の点を用いるのが一般的である。PnPに関する論文は数多く提出されているが、その中でもカメラの歪み補正をも考慮しているTsaiのアルゴリズム<sup>3)</sup>が最も優れていると考え、実装を行なった。

## 3 Tsaiのアルゴリズムの概要

以下のStepを逆算し、カメラの位置と向きを求める。特徴点が同一平面上にある場合は5点以上、そうでない場合は7点以上必要である。詳しくは、文献<sup>3)</sup>を参照されたい。

Step 1: ワールド座標系における物体上の点Pの座標 $(x_w, y_w, z_w)$ から、カメラ座標系における点Pの座標 $(x, y, z)$ に変換する。 $R$ を回転行列、 $T$ を並進行列とすると以下のように表される。

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \quad (1)$$

Step 2: カメラをピンホールカメラとみなした場合の撮像画面座標系上の点Pの理想的な(レンズのゆがみを受けていない)投影点を $(X_u, Y_u)$ とする。カメラの焦点距離を $f$ とすると、 $(x, y, z)$ から、 $(X_u, Y_u)$ への変換は以下のように表される。

$$X_u = f \frac{x}{z}, \quad Y_u = f \frac{y}{z} \quad (2)$$

Step 3: 撮像画面座標系上の点Pの実際の(レンズのゆがみの受けた)投影点 $(X_d, Y_d)$ を以下のように求める。

$$X_d + D_x = X_u, \quad Y_d + D_y = Y_u \quad (3)$$

$$D_x = X_d(\kappa_1 r^2 + \kappa_2 r^4 + \dots) \quad (4)$$

$$D_y = Y_d(\kappa_1 r^2 + \kappa_2 r^4 + \dots) \quad (5)$$

$$r = \sqrt{X_d^2 + Y_d^2} \quad (6)$$

Step 4: 撮像画面座標系上の座標 $(X_d, Y_d)$ から、フレームメモリ座標系上の座標 $(X_f, Y_f)$ に変換する。CCD

センサの  $X$ (スキャンライン) 方向の間隔を  $d_x$ 、 $Y$  方向の間隔を  $d_y$ 、CCD センサの  $X$  方向の数を  $N_{cx}$ 、フレームメモリの水平方向の画素数を  $N_{fx}$ 、中心画素の座標を  $(C_x, C_y)$ 、主にカメラとコンピュータのスキャンのタイミングのずれを補正するパラメータを  $s_x$  とすると以下のように表される。

$$X_f = s_x d_x'^{-1} X_d + C_x \quad (7)$$

$$Y_f = d_y^{-1} Y_d + C_y \quad (8)$$

$$d_x' = d_x \frac{N_{cx}}{N_{fx}} \quad (9)$$

#### 4 結果

視点同定に用いたカメラ (Sony UVW-100L) の仕様は、有効画素数 768×492(水平/垂直)、撮像面積 6.4×4.8mm、焦点距離 7.5mm である。図 1,2 の特徴点は表 1 のようになる。このとき、カメラの視点、参照視点、視軸に関する回転角 (参照視点を見て右回りが正、 $y_w$  軸の正方向を上とする) は表 2 のようになる。特徴点は、原画像を 2 値化した後に Hough 変換を施し、直線の抽出を行ない、それらの交点から 5 つを選んだ。視点同定には、SGI Indy R4000SC 48MB メモリ上で、0.068[sec] の計算時間を必要とする。実行時間の大部分は特徴点の抽出が占めており、640×480 画素の場合、1 フレームあたり 10~20 秒の計算時間が必要である。画面はクリックプルであり、CG 物体ばかりでなく、実物体も WWW の URL (Uniform Resource Locator) と対応づけられている。

表 1 : 図 1,2 における特徴点の座標

	$x_w$	$y_w$	$z_w$	$X_f$	$Y_f$
特徴点 1	0.0	0.0	0.0	341	404
特徴点 2	450.0	0.0	0.0	476	300
特徴点 3	0.0	450.0	0.0	193	302
特徴点 4	450.0	450.0	0.0	331	225
特徴点 5	0.0	900.0	0.0	87	229

表 2 : 図 1,2 におけるカメラのパラメータ

	$x_w$	$y_w$	$z_w$	angle
視点	-774.1	-677.3	985.2	
参照視点	-156.2	-91.2	461.1	
回転角				19.3 度

#### 5 おわりに

提案した手法には、2つの問題点が挙げられるだろう。1つは、被写対象物の 3 次元的なモデルを事前に得なければならないという問題であり、もう 1つは、実映像から数多くの特徴点を抽出しなければならない

という問題である。前者の問題は謂わばグローバルなモデルである電子地図の利用が図れないか検討中である。後者の問題は極めて深刻であり、本手法の最大の問題点である。しかし、本手法の目的はあくまで実映像と 3 次元 CG 像の 3 次元的な合成を行なうことにある。例えば、謂わばグローバルなトラッキングシステムである GPS (Global Positioning System) や電子コンパスによってカメラの大まかな位置と向きが分かれば、それを基に被写対象物の CG 像を生成することが可能になる。その CG 像と実映像の比較を行なうことによって、漸時的にロバストな視点同定を行なうことが可能ではないかと考え、現在検討を行なっている。

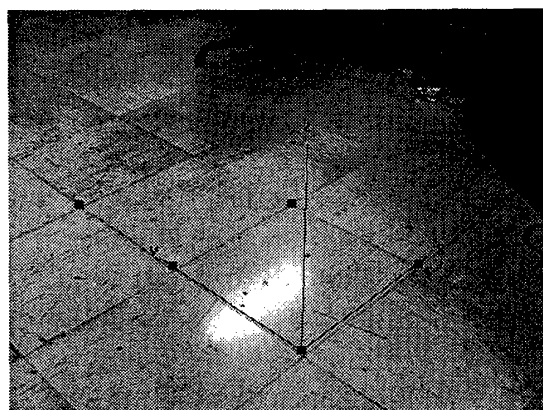


図 1: 原画像に特徴点と座標軸を 3 次元的に合成した例



図 2: 原画像に 2 つの標識を合成した例

#### 参考文献

- 1) M. Bajura, H. Fuchs and R. Ohbuchi: "Merging virtual objects with the real world", *SIGGRAPH'92*, pp.203-210(1992).
- 2) J. Rekimoto and K. Nagao: "The World through the Computer", *UIST'95*, pp.29-36(1995).
- 3) R. Y. Tsai: "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", *IEEE J. Robotics Automat.*, RA-3, No.4, pp.323-344(1987).