

プロジェクタ型複合現実感のためのスクリーン物体の追跡

永井 悠文[†] 向川 康博^{†, ††} 大田 友一^{†, †††}

[†]筑波大学大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1

^{††}先端学際領域研究センター, ^{†††}計算科学研究センター

E-mail: [†]{nagai, mukaigaw, ohta}@image.esys.tsukuba.ac.jp

あらまし プロジェクタ型複合現実感において, シーンの動的な変化への対応は, インタラクティブなコンテンツ制作のために欠かせない要素である. 本研究では, 動的なシーンに対応可能なプロジェクタ型複合現実感システムを実現するために, 性質の異なる2種類のスクリーン物体追跡手法を提案する. 1つ目は, ビジュアルマーカを物体底面に貼付し, カメラで作業空間下から観察を行うことで追跡を行う手法であり, 処理速度が速く, 精度がよいという利点がある. 2つ目は, 人の目には認識しづらいがカメラからは認識が容易なパターン情報をプロジェクタ投影画像に埋め込み, 3次元計測を行うことで追跡する手法であり, 空中を移動する物体の追跡も可能である. 本稿では, これらの手法について述べ, 実装・実験を行った結果を報告する.

キーワード プロジェクタ, 複合現実感, 物体追跡, 不可視情報

Screen Object Tracking for Projector-Based Mixed Reality

Hirobumi NAGAI[†] Yasuhiro MUKAIGAWA^{†, ††} Yuichi OHTA^{†, †††}

[†]Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573 Japan

^{††}Center for Tsukuba Advanced Research Alliance, ^{†††}Center for Computational Sciences

E-mail: [†]{nagai, mukaigaw, ohta}@image.esys.tsukuba.ac.jp

Abstract In Projector-Based Mixed Reality, the method to deal with dynamic scenes is considered as an important factor to achieve interactive systems. We propose two methods for screen object tracking. One uses a visual marker attached to the bottom of screen object and a camera to capture it under the working space. This method has advantages of fast processing and precise registration. The other method uses an image-embedded pattern which is well recognizable by the camera but is rather invisible to human eyes. This pattern can be extracted from captured images and used to achieve three-dimensional tracking of screen object.

Keywords projector, mixed reality, object tracking, invisible pattern

1. はじめに

近年, 仮想情報をプロジェクタから現実世界に直接投影することで, 仮想世界と現実世界の融合を図る“プロジェクタ型複合現実感”という新しいスタイルの複合現実感技術が提案されている. このプロジェクタ型複合現実感の「仮想情報を現実世界に直接投影する」という特徴は, ユーザが特別な機材を装着する必要がないことや, 仮想情報を与えられた物体に直接触れることができるなどの, Head-Mounted Display (HMD)にはない多くの利点をもたらしてくれる.

HMDを用いた複合現実空間での協調作業では, アイコンタクトなどの非言語情報の伝達が阻害され, ユーザ間の円滑なコミュニケーションが困難になる. しかし, 特別な機材の装着の必要がないプロジェクタ型複合現実感では, 互いの顔が直接見えるため, HMDを仮想的に消去する[1]などの特別な処理を必要とせず, 理想的なコミュニケーションが可能となる.

プロジェクタ型複合現実感のこれらの特徴は, 協調

作業システムへの応用における将来性を示している. しかし, 協調作業では, インタラクティブ性が要求されるため, なんらかのアクションを通してユーザの意思を複合現実空間に反映させる必要がある. そこで考えられるのは, スクリーンとなる物体を表示デバイスとして用いると同時に, 入力デバイスとしても用いることで, インタラクティブな複合現実感システムを実現する方法である.

従来のプロジェクタ型複合現実感では静的なシーンを扱ったものが多く, また, 動的なシーンを扱うためには大掛かりな装置を必要とした. そこで本研究では, 簡易な装置を用いたプロジェクタ型複合現実感のためのスクリーン物体の追跡手法を提案する.

2. スクリーン物体の追跡

2.1. 関連研究

プロジェクタ型複合現実感では, 画像投影の対象と

なるスクリーン物体は剛体であると仮定し、光学的な見え方のみが変化するような応用が考えられてきた。例えば、向川ら[2]は、プロジェクタ型複合現実感におけるシーンの光学的な見え方の変化を取り扱い、光源位置や反射特性を任意に変化させることのできる仮想光学環境を提案し、その技術をデジタルアーカイブされた文化財の提示法として応用している[3]。また、我々は、写実性の向上を目的として、現実世界の照明環境を記録し、それを忠実に再現する SpaceRelighter と呼ぶシステムを提案した[4]。これは、イメージベースドな手法を用い、CGではモデル化が困難な複雑な照明環境の記録と再現が容易に行える。しかし、これらのシステムでは、投影対象のシーンが静的でなければならないという制約があった。

シーンの動的な変化に対応するためには、シーンの変化を何らかの方法で検知する必要がある。例えば、Bandyopadhyay ら[5]は、赤外線センサや磁気センサをスクリーン物体に取り付ける Dynamic Shader Lamps を提案している。また、白井ら[6]は、可視光と不可視光である赤外線光とを同時に投影するプロジェクタを“秘映プロジェクタ”と呼び、秘匿情報を公共の場で共有する手段としてその有用性を確認している。このシステムは、画像投影のためのプロジェクタの他に、赤外線フィルタを取り付けたプロジェクタを別に用意し、可視光による投影画像に不可視光によるパターンを重ねて投影したシーンを赤外線カメラで観察するという手法を用いている。

しかし、赤外線センサや磁気センサを用いる手法や、不可視光情報を用いる手法は、いずれも高価あるいは特殊な機材を必要とし、システムとしても大規模なものになってしまうという問題がある。

2.2. キーアイデア

本稿では、簡易なシステムでスクリーン物体を追跡できる手法を提案する。簡易なシステムとは、通常のプロジェクタ1台とカメラ1台という構成を指す。

プロジェクタ型複合現実感において、カメラを用いた物体追跡を困難にしている要因は、プロジェクタから画像が投影されていることである。一様でない画像がシーンに投影されている場合、スクリーン物体を動かしたことにより、シーンの見え方がどの様変わったかをカメラ画像から推定するのは困難である。そこで、スクリーン物体の追跡手法として次の2つのアイデアを考えた。

- (a) プロジェクタ光が投影されないスクリーン物体の裏面に情報を付加する。
- (b) 投影画像に人間が気づきにくい情報を埋め込む。
 - (a)の、スクリーン物体の裏面に情報を付加・観察す

る方法は、ビジュアルマーカを物体底面に貼付することで具体化を試みた。ビジュアルマーカは高い検出精度と引き換えに、マーカが環境の外観を損ねるという欠点がある。しかし、スクリーン物体の底面にマーカを貼付すれば、物体が接地しているときはユーザにはマーカが見えず、外観を損なうことはない。また、物体の底面にはプロジェクタ光が照射されないためマーカの認識にも影響を与えない。

(b)の、人間が気づきにくい情報を投影画像に埋め込む方法は、人の目には認識しづらいが、カメラからは認識が容易なパターン情報を投影画像に埋め込み物体追跡の手掛かりとすることで具体化を試みた。人間の感覚特性とカメラの感度特性の違いを考慮して投影画像にパターン情報を埋め込むことで、赤外線による不可視情報を用いるようなシステムと同等の効果を得ることが期待できる。

以下、3章ではビジュアルマーカを用いた物体追跡について、4章では投影画像へのパターン情報埋め込みを用いた物体追跡について、それぞれ詳しく述べる。

3. ビジュアルマーカを用いた物体追跡

3.1. システムの構成

ビジュアルマーカを用いた物体追跡システム (Object Tracker with Visual Marker: OTVM) では、図1に示すように、平面上に複数のスクリーン物体を置くようなシーンにおける協調作業を取り扱う。スクリーン物体は白色の完全拡散反射面を持つものとし、形状はレンジファインダであらかじめ計測しておく。平面下の USB カメラで物体の位置・回転角を検出し、それに応じて、上方のプロジェクタから画像を投影する。ユーザは物体を接地させながら自由に動かすことが可能である。

スクリーン物体の位置・回転角の推定には、スクリーン物体の底面に貼付したビジュアルマーカを用いる。下方に取り付けた USB カメラからビジュアルマーカを観察する。各スクリーン物体にユニークなマーカを貼付し、物体が接地していると仮定することで、1台のカメラからの観察だけで、どの物体が、どのような位置・回転角であるかを認識することができる。なお、スクリーン物体を置く平面を透明なアクリル板にすると、プロジェクタの投影光が透過してしまうため、カメラでマーカが観察できて、且つプロジェクタの投影光を拡散反射させるように、アクリル板の上にトレーシングペーパーを敷いた。

3.2. キャリブレーション

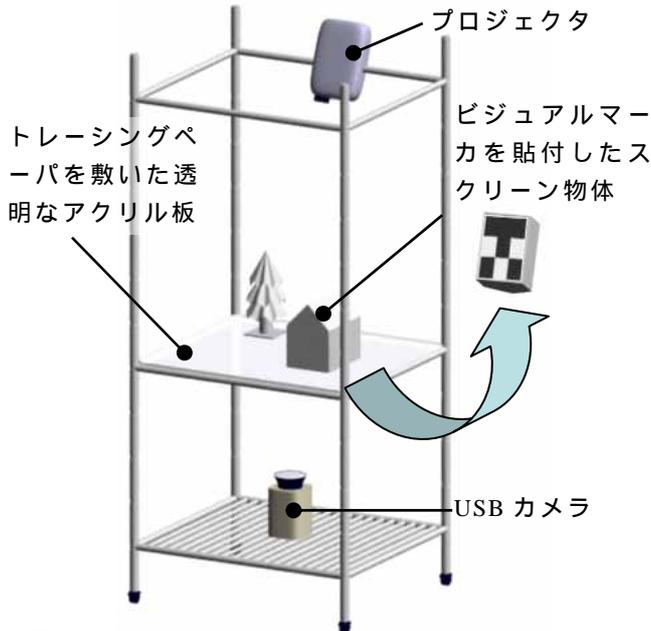


図 1 : ビジュアルマーカを用いた物体追跡が行えるシステム (OTVM) の構成

OTVM で取り扱う座標系は、カメラ座標系 x_c 、プロジェクタ座標系 x_p 、ワールド座標系 x_w の 3 つであり、それぞれを次のように定義する。

$$\mathbf{x}_c = [x_c \ y_c \ 1]^T \quad (1)$$

$$\mathbf{x}_p = [x_p \ y_p \ 1]^T \quad (2)$$

$$\mathbf{x}_w = [x_w \ y_w \ z_w \ 1]^T \quad (3)$$

ここで、物体を置く平面のワールド座標系における 2 次元座標を \mathbf{x}_w^{2D} とする。 \mathbf{x}_w^{2D} は次式のように、 $z_w = 0$ となる平面である。

$$\mathbf{x}_w^{2D} = [x_w \ y_w \ 1]^T \quad (4)$$

また、射影行列 \mathbf{H} 、 \mathbf{P} を次のように定義する。ここで s 、 t はスカラーである。

$$s\mathbf{x}_w^{2D} = \mathbf{H}\mathbf{x}_c \quad (5)$$

$$t\mathbf{x}_p = \mathbf{P}\mathbf{x}_w \quad (6)$$

まず、これらの座標系間のキャリブレーションが必要である。キャリブレーションは、アクリル板の上にグリッド線を印刷した紙を敷き、その上に形状が既知である直方体を重ねていくという手法を取った。射影行列 \mathbf{P} は、3 次元点と、その 2 次元像の座標の対応から求める[7]。

3.3. 物体位置の推定

アクリル板の上にはトレーシングペーパーを敷いているので、ビジュアルマーカが認識できるのは物体が床に接地しているときのみである。つまり、物体の位置・回転角は、ビジュアルマーカの x_w, y_w 平面座標と、 z_w まわりの回転さえ分かれば一意に決まる。図 2 は、トレーシングペーパーを敷いたアクリル板越しに USB カメラから観察したマーカの様子である。

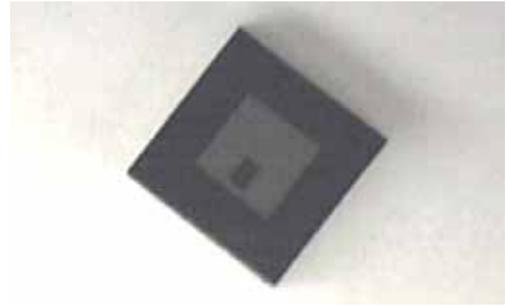


図 2 : USB カメラから見た物体底面

ビジュアルマーカの位置と ID の認識には ARTToolKit[8]を用いた。正方形マーカの内部にビジュアルパターンを設定し、テンプレートマッチングによってマーカの識別を行う。正方形マーカの 4 隅のカメラ座標 x_c と、それがどの物体であるかを示すマーカ内部のビジュアルパターンの認識から、マーカの位置と ID を推定する。そして、マーカのカメラ座標 x_c を射影行列 \mathbf{H} を用いてワールド座標 \mathbf{x}_w^{2D} に変換する。物体は接地していると仮定しているので、物体の位置姿勢が一意に求まる。ワールド座標系で定義された座標 \mathbf{x}_w を、射影行列 \mathbf{P} を用いてプロジェクタ座標 x_p としてレンダリングする。これは、プロジェクタをカメラと見なしたときに、ワールド座標系のシーンを描画することに相当する。

3.4. 実験

プロジェクタは PLUS U4-136 を、USB カメラは Logicool QV-4000 を、PC は PentiumIII 600MHz を用いて実験を行った。USB カメラで物体の底面画像をキャプチャし、物体の位置・回転角を推定した後に、画像を生成・投影する一連の処理に要する時間を表 1 に示す。ビデオレートで処理が行えており、動的なシーンへの対応が実現できた。図 3 は、白色立方体に木目調のパターンを投影した例である。

表 1 : 各処理に要する時間

処理内容	時間[msec]
物体位置・回転角推定	5.1
レンダリング	4.6

OTVM では、白色物体の底面にビジュアルマーカを貼付するだけで容易にスクリーン物体を増やすことができる。また、位置・回転角の推定には USB カメラ 1 台のみを用いているので、安価にシステムを構築できる。しかし、床面への画像投影も考慮して、アクリル板の上にトレーシングペーパーが敷かれているため、物体が接地していないと物体の位置姿勢が分からない。つまり、物体の 3 次元的な追跡ができないという問題

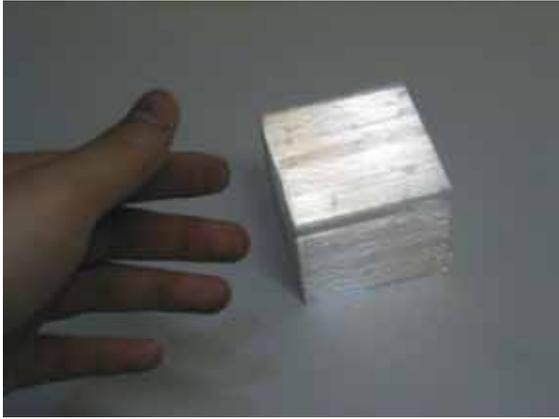


図 3：白色立方体に木目調のパターンを投影した例

がある。床面への投影は行わず、スクリーン物体のみへの画像投影を行うようなアプリケーションならば、平面マーカを用いた物体の3次元的位置追跡は可能である[8]。しかし、多くのアプリケーションを考えた場合、床面への画像投影は重要であると考えられる。

4. 投影画像へのパターン埋め込みによる物体追跡

4.1. システムの構成

投影画像へのパターン埋め込みを用いた物体追跡を行うシステム（Object Tracker with Pattern Embedding: OTPE）では、図4に示すように通常のプロジェクタと通常のカメラを作業空間の上方に設置し、3次元的な追跡を可能にする。

図5にOTPEの処理のフローチャートを示す。スクリーン物体が動いていないときは同じ画像を投影し続ける。フレーム間差分をもとにシーンの変化を検出すると、投影画像に3次元形状復元の手掛かりとなるようなパターンを、人の目には気づかれにくく、カメラでは認識が可能のように埋め込んだ画像を投影する。次に、カメラでシーンをキャプチャして、埋め込んだパターンのみを抽出する。そのようにして得られたパターンを元に、疎な形状復元を行う。そして、あらかじめレンジファインダで獲得しておいたスクリーン物体形状を手掛かりに、スクリーン物体の位置姿勢を推定し、新しいシーンに対応した画像の生成・投影を繰り返す。

ここで問題となるのは以下の2点である。

- ・ 人の目には気づかれにくく、カメラでは認識が可能なパターンの埋め込み方とは。
- ・ 動的なシーンを解析するので、少数のフレームの情報からスクリーン物体の追跡を行わなければならない。

これら2点について以降の節で詳しく述べる。

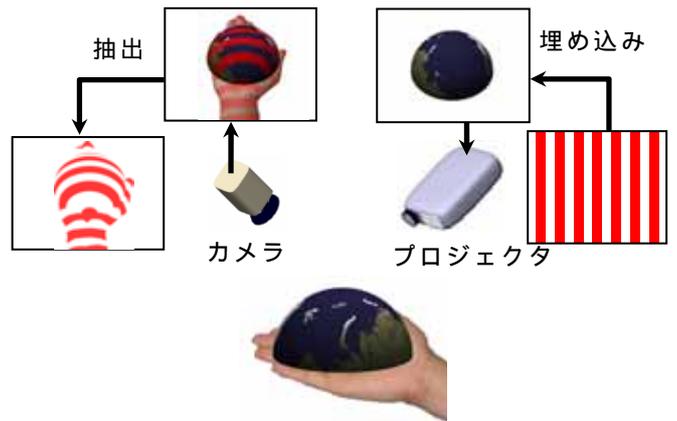


図 4：投影画像へのパターン埋め込みによる物体追跡の原理

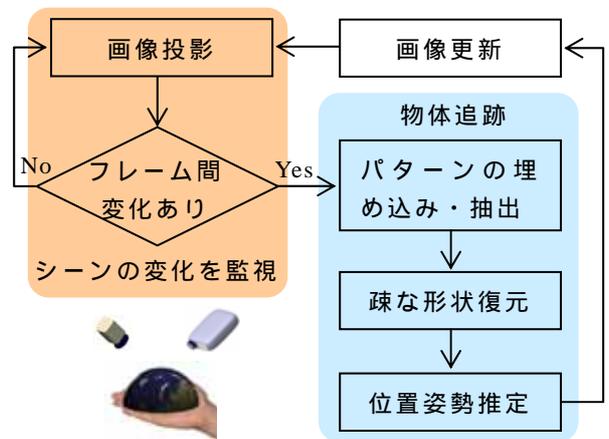


図 5：OTPE のフローチャート

4.2. 投影画像へのパターン埋め込み

投影画像の中に人の目には気づかれにくく、カメラでは認識が可能であるようなパターンの埋め込み方を検討する。一つの方法として、プロジェクタが再現できる色空間において、人間の感覚特性ではほぼ同じ色に見えるが、カメラの感度特性では異なる色として検出できるような光を用いて、3次元形状復元の手掛かりとなるパターンを画像に埋め込むことが考えられる。人間の心理的な色空間上では、白と黄色の距離は短い。そのため、白地に配した黄色のパターンは人の目には認識しづらいことをステガノグラフィとして利用した研究[9]にも、カメラと人間の目の感度特性の違いが利用されている。本研究でも、色差の中にパターンの埋め込みを行った。

もとの色情報の中にパターンを埋め込む場合、もとの色と埋め込んだ後の色のカメラの色感度特性上の距離を ΔC 、人間の色感度特性上の距離を ΔH とすると、 $\Delta C/\Delta H$ を最大化するような埋め込み方ができればよ

い。L*a*b*などの均等色空間は、人間が等しく感じる色差が等距離になるように設計された色空間であるので、色差による埋め込みを検討するには適している。一方、カメラの色感度特性上の距離 ΔC の検討には、プロジェクタの分光分布、カメラの分光感度特性が必要になり、個々の機器に依存する問題となる。

最適な埋め込み法は今後の検討課題とし、本稿では、人間の視覚特性を表す色空間として YCbCr 色空間を用いて色差成分への埋め込みを試みた。人間の視覚では、色度における空間周波数特性は明度に比べて劣ることから、色差成分である Cb,Cr 空間上でパターンの埋め込みを行えば、人間には気づかれにくいと考えられる。

4.3. 物体位置の推定

プロジェクタからシーンに情報を投影し、カメラで撮影することによりシーンを解析する、いわゆるアクティブビジョンは数多くの研究がなされており、中でも空間コード化光投影法[10]は基本的かつ有効な3次元復元方法である。OTPE では、空間コード化光投影法を応用した疎な3次元復元によって物体追跡を行う。空間コード化光投影法は、カメラ座標とプロジェクタ座標(1次元)から物体形状を推定する手法である。プロジェクタ座標をカメラ座標と対応付けるために、プロジェクタ座標の分解能のビット数と同じフレーム数のパターンを投影する必要がある。

しかし、4.1 節で述べたように、動的なシーンを解析するためには、できるだけ少ないフレーム数の情報からプロジェクタ座標が取得できることが望ましい。そこで OTPE では、図 6 のように等間隔に配置された帯状のパターンの中にその帯の ID を示す 0/1 のビットパターンを付加する。これにより、帯の両端におけるプロジェクタ座標が取得可能になる。我々は、このパターンを ID スリットと呼ぶ。ID スリットをプロジェクタから投影し、カメラ画像を 0 が埋め込まれた帯、1 が埋め込まれた帯、帯以外に 3 値化することで、離散的にカメラ座標とプロジェクタ座標の対応を表した離散的コード画像を取得する。

しかし、わずかな色差情報から 1 フレームで 3 値化を安定に行うことは難しい。そこで本稿では、ポジ画像とネガ画像を交互に表示させ、それらの差分から 3 値化する。元画像のある点における RGB 値を YCbCr 値に変換したものを Y, Cb, Cr, とする。Cb 成分に埋め込む色差を Δb とし、Cr 成分値に埋め込む色差を Δr とする。ポジ画像の 0 とネガ画像の 1 は Y, Cb+ Δb , Cr+ Δr とし、ポジ画像の 1 とネガ画像の 0 は Y, Cb- Δb , Cr- Δr とする。

キャプチャ画像を 3 値化するときには、ある点におけるポジ画像とネガ画像の YCbCr 空間における距離を

ΔE とし、閾値を T とすると、 $\Delta E \geq T$ のときポジ画像の Cb (または Cr) 値がネガ画像の Cb (または Cr) 値より大きければ 0, 小さければ 1 とする。一方、 $\Delta E < T$ はビット情報のない部分とする。

また、床面のコード画像をあらかじめ獲得しておき背景コード画像とした。取得した離散的コード画像において背景コード画像のコード値周辺は、床面とみなし削除しておくことで、図 7 右上のようなスクリーン物体のみのコード画像を得ることができる。このように獲得したコード画像と、あらかじめキャリブレーションで求めておいたプロジェクタとカメラの射影行列を用いることで、図 7 下のようにスクリーン物体のみの疎な 3 次元形状復元を行う[10]。

スクリーン物体の位置姿勢は、得られた 3 次元形状と、あらかじめ登録しておいたスクリーン物体の 3 次元形状を手掛かりに求める。しかし、得られた 3 次元形状は、コード画像生成時の 3 値化誤りや ID スリットの誤読などの誤差によって、大きな外れ値を含んで

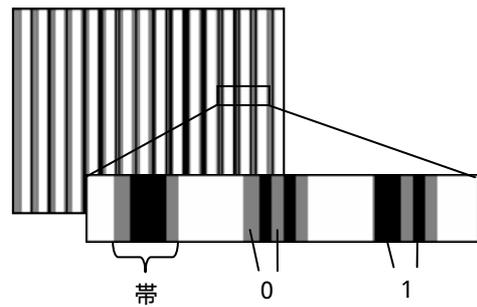


図 6 : ID スリット

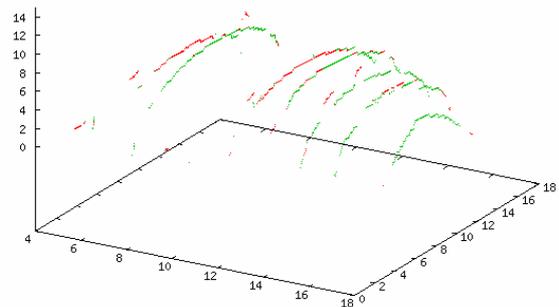
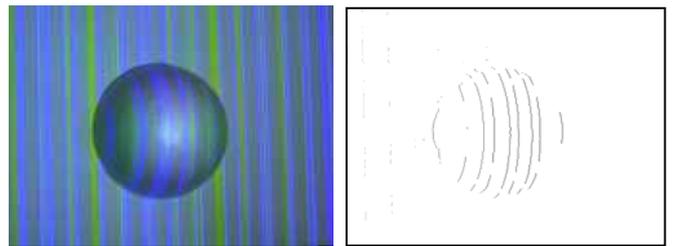


図 7 : 投影された ID スリットのポジ画像 (左上), 得られた離散的コード画像 (右上), 得られた疎な 3 次元形状 (下)

いる可能性があるため、ロバスト推定法の一つである RANSAC によって位置・姿勢を推定する。なお、本稿では問題の簡単化のために、スクリーン物体として半径が既知である半球を用いることで姿勢の計算を省き、位置推定問題のみを取り扱っている。

以下に具体的な処理を述べる。\$k\$ 回目のサンプリングにおいて、\$n\$ 個の点群データから、ランダムに 3 点を選択し、スクリーン球の半径 \$R\$ が既知であることを利用して、その 3 点を通るような球の中心 \$\mathbf{c}_k\$ を求める。復元された 3 次元形状の \$i\$ 番目の点座標を \$\mathbf{p}_i\$ とすると、\$\mathbf{c}_k\$ から \$\mathbf{p}_i\$ までの距離は、\$r_{ki} = \|\mathbf{c}_k - \mathbf{p}_i\|\$ と表すことができる。このとき、半径の許容誤差を \$m\$、評価関数 \$e(r_{ki})\$ を、

$$e(r_{ki}) = \begin{cases} 1 & : |R - m| \leq r_{ki} \\ 0 & : \text{otherwise} \end{cases} \quad (7)$$

とすると、推定された球の中心の評価関数 \$S(k)\$ は、

$$S(k) = \sum_{i=1}^n e(r_{ki}) \quad (8)$$

となり、\$S(k)\$ が最大となる \$\mathbf{c}_k\$ をスクリーン球の推定中心位置とする。

4.4. 実験結果

プロジェクタは EPSON EMP-74 を、カメラは SONY DFW-VL500 を、PC は Pentium III 600MHz を用いて実験を行った。埋め込む色差は \$C_b\$ 成分のみとし、\$\Delta b = 15, \Delta r = 0\$ として行った結果を図 8 に示す。図 8 左のように、スクリーン物体が動かされてテクスチャの位置がずれても、人の目には見えにくいよう ID スリットが埋め込まれた画像を投影し、抽出された ID スリット情報から 3 次元形状を復元し、球の位置を推定することで、図 8 右のようにスクリーン物体の位置を推定し適切な位置に描画を行うことができた。ただし、現時点では、スクリーン物体を動かしてから投影画像が更新されるまでに約 2 秒かかっているため、更なる高速化が必要である。

5. まとめ

本稿では、プロジェクタ型複合現実感のためのスクリーン物体の追跡手法について検討し、ビジュアルマーカを用いる手法 (OTVM) と、投影画像にパターンを埋め込む手法 (OTPE) を提案した。

それぞれのシステムには長短所があり、ビジュアルマーカを用いる手法は計算コストが低いので追跡が速く、追跡精度もよいが、アプリケーションによっては 3 次元的な追跡ができない。一方、投影画像にパターンを埋め込む手法は、3 次元的な追跡が可能だが、計算コストが高い。しかし、それぞれの手法の短所は、

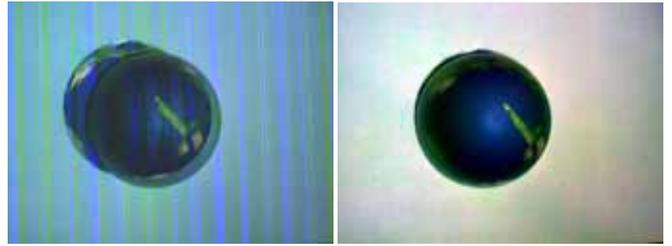


図 8 : 動かされた半球に投影されたポジ画像(左)、半球推定位置にテクスチャを投影した結果(右)

互いの手法を組み合わせることによって解消できると考えられる。例えば、接地時にはマーカ情報を用い、非接地時には接地時の情報を元に探索空間を狭めることで計算コストを抑えることも可能である。

今後の課題として、色差情報をどのような色空間へ埋め込めば人間が気づきにくいかという議論が不十分であるため、その点を更に検討していく必要がある。

文 献

- [1] M.Takemura and Y.Ohta, "Diminishing Head-Mounted Display for Shared Mixed Reality", IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002), pp.149-156, 2002.
- [2] 向川康博, 西山正志, 尺長健, "スクリーン物体への光学パターン投影による仮想光学環境の実現", 信学論 D-II, Vol. J84-D-II, No.7, pp.1448-1455, 2001.
- [3] 向川康博, 西山正志, 尺長健, "仮想光学環境における実物体反射特性の学習", 文化財のデジタル保存 自動化手法開発プロジェクト平成 13 年度成果報告会, pp.75-85, Mar.2002.
- [4] 向川康博, 永井悠文, 大田友一, "SpaceRelighter: 現実空間における照明環境の記録と再現", 信学技報 PRMU 2003-202, pp.25-30, 2004.
- [5] D.Bandyopadhyay, R.Raskar, H.Fuchs, "Dynamic Shader Lamps:Painting on Real Objects", The Second IEEE and ACM International Symposium on Augmented Reality (ISAR2001), pp.207-216, 2001.
- [6] 白井良成, 松下光範, 大黒毅, "秘映プロジェクト: 不可視情報による実環境の拡張", The 11th Workshop on Interactive Systems and Software (WISS2003), pp.115-122, 2003.
- [7] 徐剛, "写真から作る 3 次元 CG", 近代科学社, 2001.
- [8] H.Kato and M.Billinghurst, "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System", International Workshop on Augmented Reality (IWAR1999), pp.85-94, 1999.
- [9] <http://pr.fujitsu.com/jp/news/2004/06/30.html>
- [10] 井口征士, 佐藤宏介, "三次元画像計測", 昭晃堂, 1990.