

# VDTM を利用した AR/MR 用カメラトラッカ

松見 優一<sup>†1</sup> 酒井 章伸<sup>†2</sup> 森 尚平<sup>†2</sup> 柴田 史久<sup>†1</sup> 木村 朝子<sup>†1</sup> 田村 秀行<sup>†3</sup>

立命館大学 情報理工学部<sup>†1</sup> 同 大学院情報理工学研究所<sup>†2</sup> 同 総合科学技術研究機構<sup>†3</sup>

## 1. はじめに

拡張現実感 (AR) や複合現実感 (MR) において、マーカーレス・カメラトラッカの性能向上は最重要課題の 1 つである。我々は、長年の MR 研究の経験から、視点位置に依存して変化するシーンの見え方に対処できる新手法を開発した。事前登録したシーンの実画像テクスチャ付 3 次元メッシュを利用してカメラ位置姿勢を推定する手法で、実画像テクスチャ付 3 次元メッシュのレンダリング結果とカメラ画像から特徴点の 3D-2D 対応を取得する。こうしたカメラ位置姿勢推定の枠組みは Tracking by Synthesis [1] として知られるが、我々はこのレンダリング画像とカメラ画像との見た目の差異を吸収する目的で View Dependent Texture Mapping (VDTM) [2] を利用する方式を採用した。VDTM の導入により、特徴点の対応付けの高精度化が達成できた。

## 2. 関連研究

実画像テクスチャ付 3 次元メッシュを利用した特徴点ベースのカメラトラッカとして、Simon [1] は、カメラの焦点ボケをレンダリング画像上で再現し、レンダリング画像とカメラ画像間の差異を吸収することで、両画像間の特徴点マッチング精度を向上させた。我々は Simon の手法に倣いながらも、レンダリング画像とカメラ画像間の差異を吸収する枠組みとして、Image-Based Rendering の一手法である VDTM を導入する。

Zheng ら[3] は画素単位の高精度なカメラトラッカを実現するために 2 点の工夫を凝らした。まず、表示するコンテンツに応じてカメラ位置姿勢推定に利用する特徴点へ重み付けを行い、PnP 問題を解いた。続いて、カメラ画像とレンダリング画像間のオプティカルフローを計算し、画像間の位置合わせを行うことで見目上の誤差を最小化した。提案手法は、レンダリング画像とカメラ画像間でオプティカルフローを計算することで 3D-2D 対応を得るが、このオプティカルフロー計算時の精度を向上させる目的で VDTM を利用する。また、こうした精度向上の枠組みの導入に加え、破綻復帰処理を導入し、頑健性の向上を図る。

## 3. 提案手法

### 3.1. 概要

提案手法では、対象シーンを VDTM の枠組みに則って再構成した結果を利用する。より具体的には、3DCG 描画用 API のフレームワークを利用して VDTM によるシーンの再構成を行い、カラー画像及び奥行き画像 (Z バッファ) を取得して、現在のカメラ位置姿勢を推定する (図 1)。提案手法の処理手順を以下に示す。

#### 【事前処理】

(1) シーンの実画像テクスチャ付 3 次元メッシュの取得

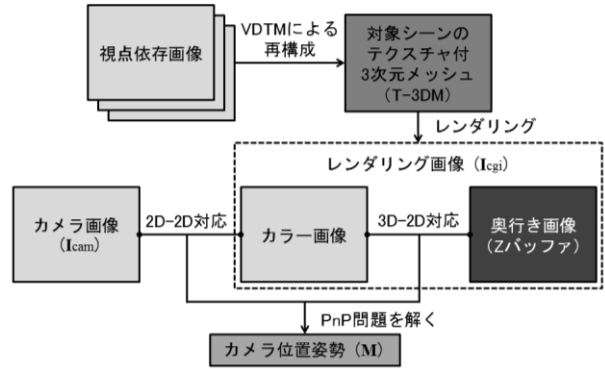


図 1 提案手法の流れ

(2) 任意視点での視点依存画像 (撮影時のカメラ位置姿勢と関連付けられた画像) の取得

#### 【実行時処理】

- (3) 初期位置姿勢推定
- (4) VDTM による対象シーンの再構成
- (5) カメラ位置姿勢推定 (カメラトラッキング)
- (6) 破綻復帰

以下では、事前処理で取得するシーンの実画像テクスチャ付 3 次元メッシュ及び視点依存画像を用いて VDTM により再構成されるシーンのテクスチャ付 3 次元メッシュを、「T-3DM」と呼ぶものとする。

Tracking by Synthesis の枠組みに VDTM を導入することで、Tracking by Synthesis において一般に設けられる以下の条件を緩和できることが期待される。

- ・対象シーンの幾何形状は精確に得られている
- ・対象シーンは完全拡散反射面で構成されている
- ・対象シーンのテクスチャは高解像度である

また、提案手法では、事前処理 (1), (2) の方法は限定しない。よって次節では実行時処理に関して示す。

### 3.2. 実行時処理

(3) 初期位置姿勢推定

初期位置姿勢  $M_0$  の推定には、カメラ画像  $I_{cam}$  と T-3DM のレンダリング画像  $I_{cgi}$  を用いる。任意のカメラ位置姿勢を入力し T-3DM をレンダリングすると  $I_{cgi}$  が得られる。 $I_{cam}$  と  $I_{cgi}$  で特徴点検出を実行し、対応付けることで  $I_{cam}$  と  $I_{cgi}$  間の対応関係 (2D-2D 対応) が得られる。更に、レンダリング結果の奥行き情報から T-3DM と  $I_{cgi}$  間の対応関係 (3D-2D 対応) が得られるため、T-3DM と  $I_{cam}$  間の対応関係 (3D-2D 対応) が得られる。この結果から、PnP 問題を解き  $M_0$  を推定する。

(4) VDTM による対象シーンの再構成

VDTM の処理は、(2) で取得した視点依存画像を用いて、T-3DM の頂点毎に実行される (図 2)。頂点毎に視点依存画像をブレンディングしてテクスチャを生成するため、視点依存画像の重み  $w$  を決定する必要がある。重み  $w$  は Buehler ら[4]の手法を基に計算する。Buehler らは Angular, Resolution, Field of View (FoV) の 3 つのペナルティ  $p$  を設定し、重み  $w$  を計算している。

“AR/MR camera tracker using VDTM”

<sup>†1</sup> College of Information Sci. and Eng., Ritsumeikan University

<sup>†2</sup> Graduate School of Information Sci. and Eng., Ritsumeikan University

<sup>†3</sup> Research Organization of Sci. and Tech., Ritsumeikan University

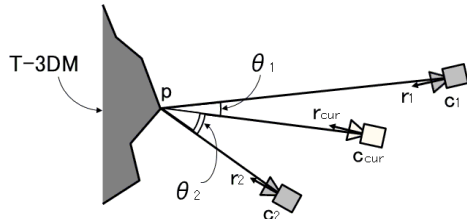


図2 プレンディングの重み計算に用いるパラメータ Angular ペナルティには[2]と同様、視点依存画像の撮影位置  $c_i$ 、T-3DM の頂点  $p$ 、現在のカメラ位置  $c_{cur}$  の 3 点により得られる角度  $\theta_i$  を基に設定する (式 (i)) .

$$p_{ang}(i) = \theta_i \quad (i)$$

Resolution ペナルティには「 $c_i$  と  $p$  間の距離」と「 $c_{cur}$  と  $p$  間の距離」との差を基に設定する (式 (ii)) .

$$p_{res}(i) = \max(0, \|p - c_i\| - \|p - c_{cur}\|) \quad (ii)$$

FoV ペナルティには  $c_i$  と  $c_{cur}$  におけるカメラの光軸ベクトル  $r_i$ ,  $r_{cur}$  の内積を基に設定する (式 (iii)) .

$$p_{fov}(i) = -r_i \cdot r_{cur} \quad (iii)$$

3 つのペナルティと比率から T-3DM の頂点毎に視点依存画像のプレディング時の重みを決定する (式 (iv)) .

$$w(i) = 1 - (\alpha p_{ang}(i) + \beta p_{res}(i)) \gamma p_{fov}(i) \quad (iv)$$

(5) カメラ位置姿勢推定 (カメラトラッキング)

$M_0$  の推定後は、1 フレーム前のカメラ位置姿勢  $M_{i-1}$  を用いて現フレームのカメラ位置姿勢  $M_i$  を推定する。まず、 $M_{i-1}$  で T-3DM をレンダリングして  $I_{cgi}$  を得る。この時、 $M_i$  と  $M_{i-1}$  は 1 フレームの違いであるため、変化量が小さいと仮定できる。したがって  $I_{cgi}$  で特徴点検出を実行し、 $I_{cam}$  とのオプティカルフローを計算することで、 $I_{cam}$  と  $I_{cgi}$  間の対応関係 (2D-2D 対応) が得られる。この後は、(3) と同様に T-3DM と  $I_{cgi}$  間の対応関係 (3D-2D 対応) から、T-3DM と  $I_{cam}$  間の対応関係 (3D-2D 対応) が得られるため、PnP 問題を解くことで  $M_i$  を推定する。以降は、毎フレーム同様の処理を繰り返し、カメラトラッキングを実現する。

(6) 破綻復帰

破綻復帰処理は、(3) と同様の処理で行う。 $I_{cgi}$  にはトラッキング中に注視していた領域のものを利用する。トラッキング時のカメラの光軸ベクトルと T-3DM の交点を、カメラが注視する 3 次元点  $pos_{cur}$  とし、毎フレーム観測する。現フレームまでの注視点の平均  $pos_{avg}$  と標準偏差  $\sigma$  を計算し、 $pos_{cur}$  が条件 (式 (v)) を満たす場合は、その時の  $pos_{avg}$  をトラッキング破綻時の T-3DM の描画パラメータとして登録する。

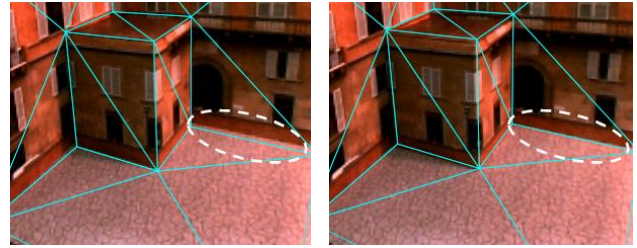
$$\|pos_{avg} - pos_{cur}\| > k\sigma \quad (k > 0) \quad (v)$$

破綻復帰時に、登録された注視点情報から  $I_{cgi}$  を作成することで、トラッキング中にカメラが注視していた領域の  $I_{cgi}$  を得ることができる。この  $I_{cgi}$  を  $I_{cam}$  と対応付ける。

## 4. 実験

### 4.1. 実験内容と実行環境

VDTM の利用によって、カメラ画像とレンダリング画像との差異を吸収し、特徴点の対応付けを高精度に行うことができることを確認するため実験を行った。複数平面で構成され、一部にスペキュラ反射が発生するシーンをトラッキング対象とし、VDTM を利用した場合と利用しない場合とで、対象シーンのワイヤーフレームの



(a) VDTM 利用なし (b) VDTM 利用あり

図3 対象シーンのワイヤーフレームの重畳結果

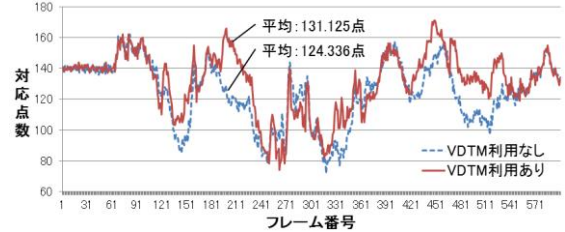


図4 特徴点の対応点数

重畳結果を目視で比較した。VDTM に利用する視点依存画像は 6 枚とした。提案手法の実装には、描画用 API に OpenGL (GLSL 3.3 以上での動作を確認) を、カメラには PointGrey 社製 Flea 3 を用いた。画像サイズは  $640 \times 512$  [画素]、600 フレームの連番画像を入力とした。

### 4.2. 実験結果と考察

対象シーンのワイヤーフレームの重畳表示結果を図 3 に示す。VDTM を利用した場合には、正しい位置に重畳されていることが分かる。また、特徴点の対応点数を図 4 のグラフに示す。VDTM を利用することで、特徴点の対応点数が増加していることが分かる。以上の結果から、VDTM の利用によってレンダリング画像とカメラ画像の差異を吸収し、特徴点の対応付けを高精度に実現できていることを確認した。

## 5. むすび

本論文では、視点位置の移動により、対象シーン中の特徴点が大きく変化する場合に適用可能な、新しいカメラトラッカを提案した。本手法では、シーン中にスペキュラ反射が発生する実物体が存在する場合でも、安定した幾何位置合わせが達成できる。実験的には、VDTM を利用した場合と利用しない場合を比較し、VDTM の利用によってカメラ画像とレンダリング画像との差異が吸収され、特徴点の対応点数が増加することを確認した。

本手法は、様々な AR/MR 用途に利用できる汎用性を有しているが、より高精度なトラッキングを必要とする隠消現実感 (DR) で威力を発揮することが期待できるので、今後は DR トラッカとしての性能向上を図っていく。

本研究の一部は、科研費・基礎研究 (S)「複合現実型情報空間の表現力基盤強化と体系化」による。

### 参考文献

[1] G. Simon: "Tracking-by-synthesis using point features and pyramidal blurring," Proc. ISMAR, pp. 85 - 92, 2011.  
 [2] P. Debevec, et al.: "Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach," Proc. SIGGRAPH, pp. 11 - 20, 1996.  
 [3] F. Zheng, et al.: "Pixel-wise closed-loop registration in video-based augmented reality," Proc. ISMAR, pp. 135 - 143, 2014.  
 [4] C. Buehler, et al.: "Unstructured lumigraph rendering," Proc. SIGGRAPH, pp. 425 - 432, 2001.