

2014 年度 AVM 研究会最優秀賞受賞記念講演 「バイナリ特徴を用いた視点変化に頑健な三次元物体の姿勢推定手法」

小林 達也^{1,a)}

概要: 任意形状の三次元物体に対する姿勢推定はモバイル AR (拡張現実感) において重要な技術であるが、処理負荷と頑健性の両面で困難な課題が存在する。処理負荷の軽いバイナリ特徴に従来の特徴量選択手法を適用した場合、局所特徴量のマッチング性能が視点によって偏るため、姿勢推定精度の悪い視点が数多く存在していた。そこで提案手法は、オフライン時にマッチングが見込める特徴量数に基づいて各視点をスコア化し、スコアの低い視点でマッチングが見込める特徴量を優先的に選択し、データベースに登録することで、マッチング性能の偏りを解消する。従来手法との比較実験により、1 オブジェクトのみデータベースに登録する場合と、20 オブジェクトに登録する場合の両方において、提案手法が従来手法の姿勢推定の処理負荷を増やすことなく、頑健性を最大 38 %改善することを示す。さらに本稿では、提案手法を適用した具体的なサービスイメージを示すために作成した、スマートフォンで身の回りの様々な形状の市販製品を認識し、関連する 3DCG を AR 表示するデモアプリケーションについても紹介する。

1. まえがき

近年、拡張現実感 (AR: Augmented Reality) の分野において、予め登録された認識対象物に対し、3DCG 等のデジタルコンテンツを重量表示することで、広告販促や、作業支援等を実現する技術が注目されつつある。専用の AR マーカを利用する手法 [1], [2] に対し、自然特徴を利用して対象物の姿勢を推定するマーカレス AR は利便性が高い。特に、雑誌やポスター等の平面物体に加え、任意形状の様々な自然物体を AR の対象として扱うための技術として、三次元物体に対する姿勢推定技術が期待を集めている。しかしながら AR では、対象物の姿勢を様々な視点から安定的に推定するための頑健性と、リアルタイム処理を行うための効率性の両方が求められ、これらを両立する技術の実現が困難であった。

三次元物体を複数の視点から撮影した参照画像 (キーフレーム) と、入力画像との間で特徴点マッチングを行うことで、様々な視点から三次元物体の姿勢を推定する手法が提案されている [3], [4]。これらの手法は PC 上でリアルタイムに実行可能で、視点変化に対する頑健性にも優れるが、AR の主要なプラットフォームである携帯端末で実行するには、処理負荷やメモリ消費量の点で課題がある。一方で

頑健性には欠けるものの、処理負荷とメモリ消費量に優れるバイナリ特徴 [5], [6], [7] が提案されている。さらに、対象物を多数の視点から撮影した画像を自動生成し、生成画像から検出した多数の特徴量から、頑健なマッチングが見込める特徴量のみを選択することにより、特徴量の頑健性を効果的に補う手法が提案されている [8]。これらの併用により、三次元物体に対する頑健な姿勢推定を、携帯端末上で実現することが期待されるが、実際には姿勢推定困難な視点 (撮影角度や撮影距離) が数多く存在するため、更なる頑健性の向上が必須である。著者らは、三次元物体が撮影される視点の範囲が平面物体より広く、形状によって表面の見え方が異なるため、各視点でマッチング可能な特徴量数に偏りが生じることが問題の一因であると考え、各視点でマッチングが見込める特徴量数が均一に近づくようにバイナリ特徴を選択することで、姿勢推定困難な視点を削減し、頑健性を向上する手法を提案した [9]。提案手法はオフラインの特徴選択の際に、各視点をマッチングが見込める特徴量数でスコア化し、スコア最小の視点を補強するようにバイナリ特徴を選択することで、従来手法のマッチング数の変動を抑え、姿勢検出率を改善することができる。

本稿では、提案する姿勢推定手法の概要と、従来手法に対する優位性を示す評価実験結果について説明を行う。さらに、著者らが提案手法の具体的な適用イメージとして作成した、デモアプリケーションについても紹介を行う。作成したアプリケーションは、スマートフォンで身の回りの

¹ 株式会社 KDDI 研究所
埼玉県ふじみ野市大原 2-1-15
KDDI R&D Laboratories, Inc.
Ohara 2-1-15, Fujimino, Saitama, 356-8502 Japan
^{a)} ts-kobayashi@kddilabs.jp

様々な模様・形状の三次元物体を瞬時に認識し、関連情報や3DCGをリアルタイムに重量表示することができる。

2. 関連研究

三次元物体に対する姿勢推定手法には、姿勢変化の時間的な相関を利用し、前フレームの推定値を初期値として、フレーム間の小さな変化量を推定する姿勢追跡のアプローチと、毎フレーム独立に姿勢を推定する姿勢検出のアプローチが存在する。事前情報を利用する姿勢追跡の方が、高速かつ視点変化に頑健な姿勢推定を実現する上で有利であるが、追跡開始時や追跡失敗からの復帰時、追跡誤差の蓄積を解消する時等に姿勢の初期値を必要とする。そのため、従来手法の多くは姿勢検出と姿勢追跡を併用するハイブリッド型のアプローチを選択している [3], [4], [10], [11]。本稿が提案する姿勢推定手法は事前情報を必要としないため、姿勢検出に分類される。処理負荷と頑健性に優れる姿勢追跡と比較して、視点変化に頑健かつ、携帯端末でリアルタイム処理が可能な姿勢検出の実現は未だに困難であり、これを解決することが本稿の目的である。

SIFT [12] や SURF [13] といった局所特徴量を利用することで、視点変化に頑健なマッチングが可能であるが、処理負荷が高く、リアルタイム処理が困難である。そのため、SIFTGPU [14] を利用する手法 [3] や、オフライン学習を行うことで、オンラインでは高速に実行可能な Random Ferns [15] を利用する手法 [4] が提案されている。しかしながら PC より処理リソースの乏しい携帯端末上では、処理負荷や消費メモリ量の問題から、SIFTGPU や Random Ferns であっても、リアルタイムに実行することは難しい。

そこで提案手法は、より処理負荷と消費メモリ量の小さい、BRISK [5], ORB [6], FREAK [7] といったバイナリ特徴を利用する。バイナリ特徴は、アフィン歪みに対する頑健性に乏しい点が課題である。そこで ASIFT [16] のように、アフィン歪みの要因である様々な視点変化をシミュレートした多数の学習画像からバイナリ特徴を検出することで、頑健性を向上することが有効である。しかしながら、この方式は登録する特徴量数が増えるため、処理負荷が増大し、リアルタイム性が損なわれてしまう。これに対し、[8] で提案されている特徴選択手法は、学習画像から検出した特徴量から、頑健なマッチングが見込める特徴量のみを選択することで、データベースに登録する特徴量数を抑え、処理負荷の増大を防いでいる。[8] の手法は提案手法に近いが、各視点でマッチング可能な特徴数に偏りが生じることで、頑健に姿勢推定できない視点が多いことが課題である。一方提案手法は、各視点でマッチングが見込める特徴量数が均一に近づくように特徴量を選択することで、少ない特徴量数でも頑健な姿勢推定が可能である。

なお、提案手法と既存の姿勢追跡手法と組み合わせることで、より効率的なハイブリッド型のアプローチを取るこ

とが可能である。実際に著者らは提案手法が推定した姿勢を初期値とし、三次元物体の姿勢を頑健に追跡する姿勢追跡手法 [17] についても提案しており、本稿で紹介するデモアプリケーションはこれらを組み合わせたハイブリッド処理を実行している。ハイブリッド処理において提案手法の頑健性の高さは、追跡開始の速さや追跡失敗からの復帰の速さに直結し、AR アプリケーションのユーザビリティ向上に貢献している。

3. 提案手法

提案手法はオフラインの特徴量のデータベースへの登録処理と、オンラインの姿勢推定処理から構成される。提案手法の特徴は、オフラインの特徴量登録時に、姿勢推定が困難な視点を削減するように特徴量を選択する部分にあり、オンラインの姿勢推定処理は従来文献 [4], [6] と同様に、一般的な姿勢推定手法を用いることができる。提案手法は Randomized Trees [18] や Random Ferns といった事前学習型の手法と同様、高速かつ頑健なオンライン処理の実現と引き換えに、オフライン処理に時間を要する。ただし AR では認識対象物は予め定まっている場合が多いため、オフライン処理は PC で事前に行い、携帯端末は登録されたデータベースを用いてオンライン処理のみ実行することで、オフライン時の処理負荷の問題を解決可能である。

3.1 オフライン処理

図 1 に提案するオフライン処理のフローを示す。まず、認識対象物の 3D モデルを複数の視点（姿勢）で 3D レンダリングすることで、学習画像を生成する。この視点は [19] と同様に、対象物を中心としたジオデシック・ドームの頂点座標に配置する。提案手法は 162 頂点のジオデシック・ドームを、半径を変えて 5 層に配置し、計 810 点の視点をを用いる。162 頂点は認識対象物の全周囲にほぼ均等に配置されるため、様々な撮影角度をカバーできる。提案手法はこれを 5 層に配置することで、スケール変化に対する頑健性も向上する。スケールファクタは 1.2 とした。

次に、生成した 1 対象物あたり 810 枚の学習画像のそれぞれから、バイナリ特徴 $d_i \in D$ を 1 スケールで検出する。提案手法は、特徴量として ORB を使用した。各特徴量 $d_i \in D$ のピクセル座標 m_i を対象物の 3D モデル表面に逆投影することで、物体座標系の三次元座標 m'_i を計算する。逆投影には、Ray/Triangle 交差判定を用いる [20] の手法を利用した。

バイナリ特徴のセット D を検出後、従来手法 [8] と同様に D の視点別のサブセット間でマッチングテストを行い、各特徴量が別視点において正しくマッチング可能な特徴量であるか判定を行う。ここで、各視点 $v_j \in V$ から検出されたサブセット $D_j \subset D$ は、サブセット単位で他の全サブセットとマッチングされ、各マッチングにおいて、サブ

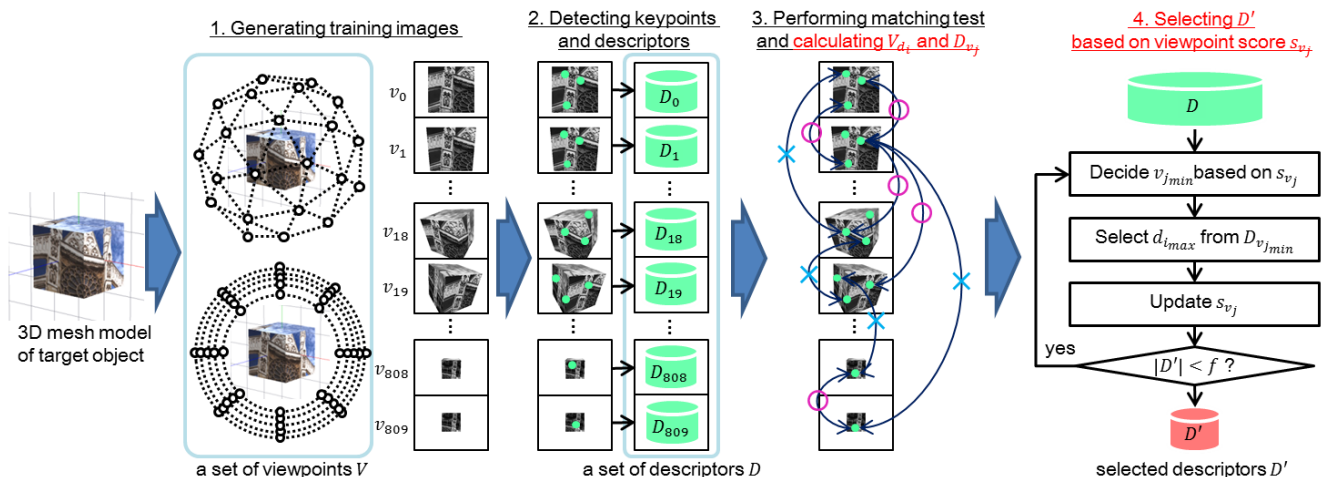


図 1 提案手法のオフライン処理のフロー

セット内の各特徴量は、別視点のサブセット内の特徴量のうち、最近傍と対応付けられる。マッチングの照合は物体座標系における距離計算で行われる。例えば、 D_{j_1} と D_{j_2} 間のマッチングの結果、 D_{j_1} の特徴量 d_{i_1} が、 D_{j_2} の d_{i_2} と対応付けられた場合、 $|m'_{i_1} - m'_{i_2}| < \tau$ であれば、 d_{i_1} は視点 v_{j_2} においてマッチング可能と判定する。提案手法は、このマッチングテストを通して、各特徴量 d_i がマッチング可能な視点のリスト（視点リスト） $V_{d_i} \subset V$ と、各視点 v_j においてマッチング可能な特徴量のリスト（特徴量リスト） $D_{v_j} \subset D$ を記録し、 V_{d_i} と D_{v_j} に基づいて特徴量の選択を行う。

従来手法 [8] は、マッチングテスト後、各特徴量 d_i をマッチング可能な視点の数 $|V_{d_i}|$ でスコア化し、スコア上位の特徴量から一定数 $f = |D'|$ を順に選択することで、データベースに登録する特徴量のサブセット $D' \subset D$ を選択する。一方で提案手法は、選択された特徴量 D' のうち、各視点においてマッチング可能な特徴量数で当該視点をスコア化し、より低スコアの視点を補うように特徴量を選択していく。ここで、各視点 v_j のスコア s_{v_j} は、視点リストに当該視点を含む特徴量の数 $|\{d_k | v_j \in V_{d_k}, d_k \in D'\}|$ で表される。 s_{v_j} は D' の選択状況によって変動するため、提案手法は選択した特徴量が一定数 f に達するまで、以下のステップを反復する。

- スコア最小の視点 $v_{j_{\min}} = \operatorname{argmin}_{v_j}(s_{v_j})$ を選択する。ここで、各視点スコアの初期値は 0 とし、同スコアの視点が存在する場合はランダムに一視点を選択する。
- $v_{j_{\min}}$ の特徴量リスト $D_{v_{j_{\min}}}$ のうち、マッチング可能な視点の数が最も多い特徴量 $d_{i_{\max}} = \operatorname{argmax}_{d_i \in D_{v_{j_{\min}}}} |V_{d_i}|$ を選択し、データベースに登録する。
- 登録した特徴量 $d_{i_{\max}}$ の視点リスト $V_{d_{i_{\max}}}$ に含まれる視点のスコアを 1 加算することで、各視点のスコアを更新する。

姿勢推定の失敗の主な原因は十分な対応点数（インライア数）が得られないことであるが、従来の選択手法ではマッチング可能な特徴量数が特定の視点に集中し、姿勢推定困難な視点が生じることを防ぐことができない。一方で、提案するフローで特徴量を選択することにより、マッチング可能な特徴量数が少ない視点を削減することができ、視点変化に対する頑健性を向上することが可能である。

3.2 オンライン姿勢推定

オンラインの姿勢推定処理は、まず、連続的に入力されるフレームから ORB 特徴量を検出し、データベースの特徴量 D' とマッチングすることで、2D-3D 対応点を取得する。マッチング方法は、登録されている対象物が 1 体のみの場合や、フレーム中の対象物が既知の場合（姿勢推定のみ行う場合）と、フレーム中の対象物を複数登録された対象物から認識する必要がある場合（物体認識と姿勢推定を行う場合）で異なる。姿勢推定のみ行う場合は、検出した特徴量を、対応するデータベースの特徴量と総当たりでマッチングを行う。物体認識と姿勢推定を行う場合は、データベースの全特徴量と近似最近傍探索でマッチングし、マッチングされた三次元座標が最も多い対象物を決定し、それ以外の対象物に関する特徴量を対応点から除外する。提案手法は、FLANN [21] (Fast Library for Approximate Nearest Neighbors) ライブラリの LSH [22] (Locality Sensitive Hashing) を使用する。マッチング後、2D-3D 対応点を対応点間のハミング距離でソートし、PROSAC [23] と EPnP [24] を使用して姿勢を推定する。

4. 評価実験

評価実験では、まず、従来の特徴量選択手法 [8] と、提案する特徴量選択手法を実装し、実験用の三次元物体に対する姿勢推定性能の比較評価を行うことで、提案手法の有効性を示す。次に、身の回りの三次元物体を実際に対象物

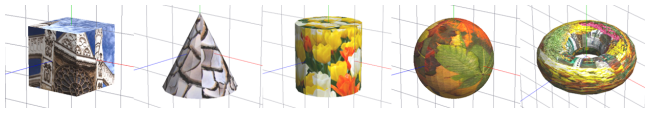


図 2 評価実験でテスト用の対象物として使用した、形状と模様の異なる 5 種類の 3D モデル。

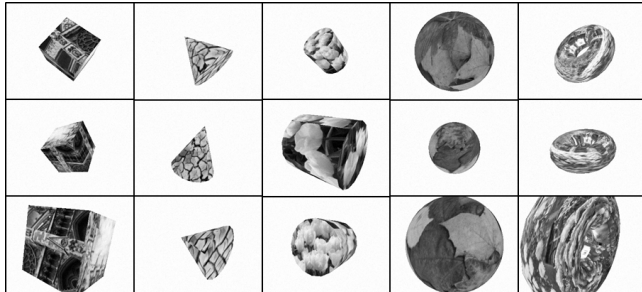


図 3 5 種類の 3D モデルから生成したテスト画像の例。

として登録し、関連する 3DCG を重畳するモバイル AR アプリケーションを作成することで、提案手法を適用した具体的なサービスイメージを示す。

4.1 提案手法の性能評価

図 2 に、実験に使用した 5 種類のテスト用の対象物を示す。各対象物について、全方位の回転行列を含むランダムなカメラパラメータで 3D レンダリングを行うことで、QVGA (320 × 240) サイズのテスト画像を 1,000 枚ずつ (計 5,000 枚) 作成した。図 3 に、テスト画像の一部を示す。対象物はフレーム中央に写ることを仮定し、レンダリングの距離は 200 mm から 400 mm の範囲で一様に変動させた。データベースには、図 2 の 5 物体に加え、テクスチャの異なる 15 物体 (計 20 物体) を登録した。提案手法は 3.1 節のフローで、各対象物について 810 枚の学習画像を生成し、各学習画像から 100 点の ORB 特徴量を検出し、最終的に 2,000 特徴量 (64 KB) を選択した。一方、従来手法は 362 頂点のジオデシック・ドーム (1 層) を使用し、362 枚の学習画像から、5 スケール (スケールファクタ 1.2) で 200 点の ORB 特徴量を検出し、提案手法と同数の 2,000 特徴量を選択した。オンライン処理は両方式とも、3.2 節の手法を用いた。LSH のパラメータはハッシュテーブル数 20、ハッシュキー次元数 16 に設定した。実装は C/C++ で行い、オフライン処理のみ Intel Core i5-3380 CPU (2.9 GHz) を搭載したノート PC で行い、それ以外の実験は Qualcomm Snapdragon 600 CPU (1.9 GHz) を搭載した Android スマートフォン (Samsung Galaxy S4) で行った。

表 1 に、5,000 枚の全テスト画像に対する認識率と姿勢推定率、マッチングに成功した 2D-3D 対応点 (インライア) の数の標準偏差 σ_{n_l} の結果を示す。上 2 段は、テスト画像の対象物の種類を既知とし、姿勢推定のみ行った場合 (テスト画像から検出された特徴量を、対応する 2,000 特

表 1 全テスト画像に対する平均の認識率、姿勢推定率、インライア数の標準偏差。(Conv.: 従来手法 [8], Pro.: 提案手法)

	認識率 [%]	姿勢推定率 [%]	σ_{n_l}
Conv. (1 object)	-	81.48	14.90
Pro. (1 object)	-	87.16	12.77
Conv. (20 object)	94.42	76.70	11.71
Pro. (20 object)	98.88	85.54	10.04

徴量のみとマッチングした場合) の結果を示している。本評価実験では、推定した姿勢を用いて認識対象物の 3D モデルの各頂点をテスト画像に投影することで、各頂点の再投影誤差 (真値に対するズレ) を計算した。全頂点中で最大の再投影誤差が 20 ピクセル以内の場合に、姿勢推定に成功したと判定した。提案手法により、姿勢推定率が 5.68 ポイント改善し、姿勢推定の失敗を約 31 % (18.52 % から 12.84 %) 削減することを確認した。 σ_{n_l} を比較すると、提案手法の方が、全テスト画像に対するインライア数のばらつきが少ない。このマッチング性能の安定性向上が、姿勢推定の頑健性向上に寄与したことが分かる。

表 1 の下 2 段は、20 物体をデータベースに登録し、物体認識と姿勢推定を行った場合 (テスト画像から検出された特徴量を、40,000 特徴量とマッチングした場合) の結果を示している。認識率は登録された 20 物体中、正しい対象物を認識できた割合を表し、姿勢推定率はさらに姿勢推定にも成功した割合を表す。従来手法、提案手法共に 9 割以上の精度で物体認識することができたが、提案する特徴点選択により、認識の失敗を約 8 割削減 (5.58 % から 1.12 %) できることを確認した。また、最終的な姿勢推定率は 8.84 ポイント改善し、提案手法により、姿勢推定の失敗を約 38 % (23.3 % から 14.46 %) 削減できることを確認した。提案手法の方が認識率に優れるため、最終的な姿勢推定率の差分は、姿勢推定のみを行った場合より大きくなっている。また、姿勢推定のみ行った場合と比較して、従来手法、提案手法共に σ_{n_l} が小さくなっている。これは誤対応の増加により、全体的にインライア数が低下したことが関係している。ただし、提案手法の方が、全テスト画像に対するインライア数のばらつきが少ない傾向は変わっていない。

図 4 に、QVGA サイズのテスト画像 1 枚に対する平均の処理時間を示す。データベースに登録する特徴量数が等しく、オンライン処理も等しいため、提案手法の処理時間は従来手法とほぼ等しい。つまり、提案手法が従来手法の処理負荷を高めることなく、頑健性の向上を実現できていることが分かる。また、バイナリ特徴と近似最近傍探索を用いたことにより、1 物体が登録された場合に 1 フレームあたり約 80 msec、20 物体が登録された場合でも 1 フレームあたり約 90 msec と、登録する対象物数が増えた場合でも、携帯端末上でリアルタイムに近い処理速度 (10 FPS 以上) で姿勢推定を実行できることが示された。

なお、オフライン処理はノート PC で実行したが、従来

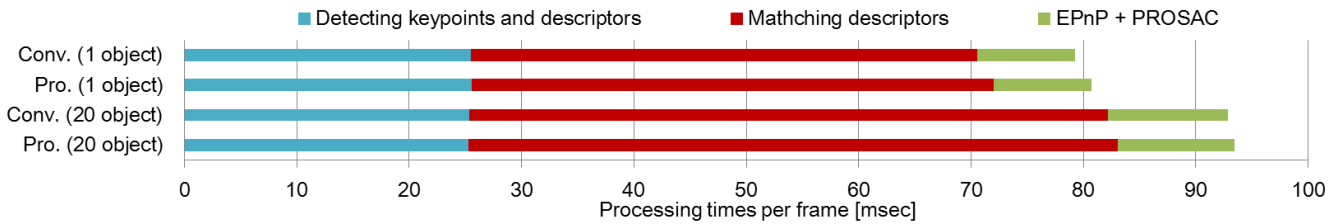


図4 スマートフォン上での1フレームあたりの平均の処理速度。(Conv.:従来手法 [8], Pro.: 提案手法)

手法と提案手法の両方式とも1物体あたり3~5分の処理時間がかかった。この処理時間の大部分は、特徴量検出後のマッチングテストに要した。予め認識対象物が定まっている場合は、携帯端末上でオフライン処理を行う必要はないが、携帯端末上で実行するためには、マッチングテストの処理負荷を大幅に削減する必要がある。また、本実験に使用した20物体(40,000特徴量)からさらに登録する対象物数を増やす場合、LSHの消費メモリが増加し、現状のLSHのパラメータではモバイル端末上での実行が困難になることが課題となっている。今後、より大規模な認識を実現するため、LSHのパラメータや登録する対象物数、1対象物あたりの登録する特徴量数と、推定精度、処理速度、消費メモリ量との関係性を詳細に評価する予定である。

4.2 提案手法の適用イメージ

提案手法の具体的な適用イメージとして、身の回りの6種類の三次元物体を認識して、3DCGを重畳するモバイルARアプリケーションを作成した。アプリケーションを実行している様子を図5に示す。主に市販製品を対象物として想定しており、ユーザが購入検討の際に詳細情報や関連情報をAR表示で確認する用途や、電化製品の故障時にマニュアル等を確認することで作業を支援する用途において、提案手法を適用したモバイルARアプリケーションが利用できると考えている。

各対象物の3Dモデルの作成方法については、対象物の形状に応じて複数の方法を使い分けた。プリミティブ形状である図5(a)~(e)の対象物については、市販の3Dモデリングソフトを使用してモデルを作成した。対象物のテクスチャには、対象物を撮影した写真や、対象物のラベルのスキャンデータを用いた。図5(f)の卓上プリンタについては、KinectFusion [25] を使用して3Dモデルを作成した。また、重畳する3DCGについても、3Dモデリングソフトを使用して手動で作成した。

アプリケーションはAndroid/iOSの両方に対応し、カメラから取得したプレビュー画像から対象物の姿勢を推定し、プレビュー画像と対応する3DCGをOpenGLでディスプレイにリアルタイムで描画することで、AR表示を実現している。提案する姿勢推定手法を、描画を実行するメインスレッドとは別スレッドで実行することで、3DCGの



図5 提案手法を適用したモバイルARアプリケーションが身の回りの三次元物体にCGを重畳している様子。(a) コーヒー豆の缶にコーヒー豆の詳細情報を重畳。(b) 菓子箱にキャンペーン情報を重畳。(c) ワインボトルに産地情報を重畳。(d) 動物の菓子箱に動物の3Dモデルを重畳。(e) 地球儀に各国の国旗を重畳。(f) 卓上スキャナにマニュアル情報を重畳。

レンダリングを常にリアルタイム(30FPS)で実行している。また、提案する姿勢推定手法が対象物を認識し、姿勢を推定した後は、より高速に実行可能な姿勢追跡処理がメインスレッドで実行され、見失うまで対象物の姿勢を更新し続ける。この姿勢追跡処理には、著者らが過去に提案したモデルベースの姿勢追跡手法 [17] を使用している。提案する姿勢推定と姿勢追跡のハイブリッド処理により、対象物の姿勢に応じて3DCGの表示がリアルタイムに更新され、3DCGが対象物にピッタリと張り付いている様なAR表示が実現できた。

身の回りの人工物はプリミティブな形状であることが多いため、3Dモデリングソフトによってモデルを生成することが可能である。また、より複雑な形状の三次元物体で

あっても、KinectFusion を利用することでモデルを生成することが可能である。しかしながら、高精度なモデリングに熟練を要する点と、モデリング可能なサイズに制約がある点が現状の課題である。そのため、より簡易にモデリングが可能な 3D スキャナの利用や、対象物を撮影した画像群から SfM (Structure from Motion) でモデルを生成する技術等についても今後検討する予定である。

5. まとめ

本稿では、任意形状の三次元物体に対する、視点変化に頑健な姿勢推定手法を提案した。提案手法は、オフライン時に、各視点でマッチングが見込める特徴量数が均一に近づくようにバイナリ特徴を選択することで、姿勢推定の頑健性を向上する。複数の形状と模様の 3D モデルを用いた評価実験により、提案手法が、従来手法の処理負荷を増やすことなく、姿勢推定の失敗を約 38 %削減することを示した。また、データベースに 20 物体が登録された場合においても、携帯端末上で 10 FPS 以上の処理速度で姿勢推定が実行できることを確認し、頑健性と効率性の両立を実現可能であることを示した。さらに、身の回りの 6 種類の市販製品を認識して、対応する 3DCG をリアルタイムで重畳するモバイル AR アプリケーションを作成し、提案手法を活用した具体的なサービスイメージを示した。

参考文献

- [1] Kato, H. and Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system, *Proc. of IWAR*, (online), DOI: 10.1109/IWAR.1999.803809 (1999).
- [2] Wagner, D. and Schmalstieg, D.: Artoolkitplus for pose tracking on mobile devices, *Proc. of CVWW* (2007).
- [3] Kim, K., Lepetit, V. and Woo, W.: Keyframe-based modeling and tracking of multiple 3D objects, *Proc. of ISMAR*, (online), DOI: 10.1109/ISMAR.2010.5643569 (2010).
- [4] Park, Y., Lepetit, V. and Woo, W.: Extended Keyframe Detection with Stable Tracking for Multiple 3D Object Tracking, *IEEE Trans. on VCG*, (online), DOI: 10.1109/TVCG.2010.262 (2011).
- [5] Leutenegger, S., Chli, M. and Siegwart, R.: BRISK: Binary Robust invariant scalable keypoints, *Proc. of ICCV*, pp. 2548–2555 (online), DOI: 10.1109/ICCV.2011.6126542 (2011).
- [6] Rublee, E., Rabaud, V., Konolige, K. and Bradski, G.: ORB: An efficient alternative to SIFT or SURF, *Proc. of ICCV*, (online), DOI: 10.1109/ICCV.2011.6126544 (2011).
- [7] Alahi, A., Ortiz, R. and Vandergheynst, P.: FREAK: Fast Retina Keypoint, *Proc. of CVPR*, (online), DOI: 10.1109/CVPR.2012.6247715 (2012).
- [8] Kurz, D., Olszowski, T. and Benhimane, S.: Representative feature descriptor sets for robust handheld camera localization, *Proc. of ISMAR*, (online), DOI: 10.1109/ISMAR.2012.6402540 (2012).
- [9] 小林達也, 加藤晴久, 柳原広昌: バイナリ特徴を用いた視点変化に頑健な三次元物体の姿勢推定手法, 情報処理学会研究報告オーディオビジュアル複合情報処理 (AVM), No. 5 (2014).
- [10] Choi, C. and Christensen, H.: Real-time 3D model-based tracking using edge and keypoint features for robotic manipulation, *Proc. of ICRA*, (online), DOI: 10.1109/ROBOT.2010.5509171 (2010).
- [11] Park, H., Mitsumine, H., Fujii, M. and Park, J.-I.: Analytic fusion of visual cues in model-based camera tracking, *Proc. of VRCAI*, New York, NY, USA, pp. 215–220 (online), DOI: 10.1145/1670252.1670298 (2009).
- [12] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, *IJCV*, Vol. 60, pp. 91–110 (online), DOI: 10.1023/B:VISI.0000029664.99615.94 (2004).
- [13] Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V.: SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110, pp. 346–359 (2008).
- [14] Wu, C.: SiftGPU: A GPU Implementation of Scale Invariant Feature Transform (SIFT) (<http://cs.unc.edu/ccwu/siftgpu>, 2007).
- [15] Ozuysal, M., Calonder, M., Lepetit, V. and Fua, P.: Fast Keypoint Recognition Using Random Ferns, *IEEE Trans. on PAMI*, (online), DOI: 10.1109/TPAMI.2009.23 (2010).
- [16] Yu, G. and Morel, J. M.: A fully affine invariant image comparison method, *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1597–1600 (online), DOI: 10.1109/ICASSP.2009.4959904 (2009).
- [17] 小林達也, 加藤晴久, 柳原広昌: モバイル AR のための三次元物体のリアルタイム姿勢追跡, 映像情報メディア学会冬季大会講演予稿集, pp. 10–5–1–10–5–2 (2014).
- [18] Lepetit, V. and Fua, P.: Keypoint recognition using randomized trees, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 9, pp. 1465–1479 (online), DOI: 10.1109/TPAMI.2006.188 (2006).
- [19] Hinterstoisser, S., Lepetit, V., Benhimane, S., Fua, P. and Navab, N.: Learning Real-Time Perspective Patch Rectification, *IJCV*, Vol. 91, No. 1, pp. 107–130 (online), DOI: 10.1007/s11263-010-0379-x (2011).
- [20] Vacchetti, L., Lepetit, V. and Fua, P.: Fusing online and offline information for stable 3D tracking in real-time, *Proc. of CVPR*, Vol. 2, pp. II – 241–8 vol.2 (online), DOI: 10.1109/CVPR.2003.1211476 (2003).
- [21] Muja, M. and Lowe, D. G.: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration, *Proc. of VISSAPP* (2009).
- [22] Lv, Q., Josephson, W., Wang, Z., Charikar, M. and Li, K.: Multi-probe LSH: Efficient Indexing for High-dimensional Similarity Search, *Proc. of VLDB*, (online), available from <http://dl.acm.org/citation.cfm?id=1325851.1325958> (2007).
- [23] Chum, O. and Matas, J.: Matching with PROSAC - progressive sample consensus, *Proc. of CVPR*, (online), DOI: 10.1109/CVPR.2005.221 (2005).
- [24] Lepetit, V., Moreno-Noguer, F. and Fua, P.: EPnP: An Accurate O(n) Solution to the PnP Problem, *IJCV*, (online), DOI: 10.1007/s11263-008-0152-6 (2009).
- [25] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S. and Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking, *Proc. of ISMAR*, pp. 127–136 (online), DOI: 10.1109/ISMAR.2011.6092378 (2011).