

## ディジタルエンターテイメントのための三次元映像メディア

齋藤 英雄<sup>†a)</sup> 植松 裕子<sup>†</sup> 内山 英昭<sup>†</sup> ピレ ジュリアン<sup>†</sup>

3D Visual Media for Digital Entertainment

Hideo SAITO<sup>†a)</sup>, Yuko UEMATSU<sup>†</sup>, Hideaki UCHIYAMA<sup>†</sup>, and Julien PILET<sup>†</sup>

あらまし 近年、三次元映像メディアに対する注目が高まっている。従来の映像メディアは二次元であり、三次元である実世界をカメラにより二次元映像に投影したものである。これに撮影されたものはそもそも三次元であり、その三次元性を考慮した新しい映像生成・表示法として、自由視点映像、三次元ビデオ、更には複合現実感・拡張現実感 (AR/MR) 表示が近年注目されるようになってきた。本論文では、実世界の三次元性を利用することによって実現される映像メディアを「三次元映像メディア」と位置づけ、これを実現するための基盤技術として AR/MR 表示と多視点画像からの自由視点映像生成に焦点を当て、これらをディジタルエンターテイメントに利用することを志向して筆者らがやっている研究を中心に紹介する。

キーワード 自由視点映像, 拡張・複合現実感, 多視点画像, カメラトラッキング, 画像検索

## 1. ま え が き

通常の映像メディアは二次元であり、三次元空間をカメラにより二次元に投影したものである。人間はこの映像メディアから多くの視覚的情報を認識できるため、映像メディアは情報伝達手段として広く使われている。また、これを用いたエンターテイメントも広く普及しており、映画やアニメーションのように一方的に映像情報を視聴者に見せるものから、ビデオゲームのようにインタラクションを伴った形で映像メディアがエンターテイメントに利用されているものまで、様々な形態で利用されている。

一方、カメラで撮影された二次元画像をコンピュータに認識理解させる方法論は、コンピュータビジョンと呼ばれ、1980 年ごろから盛んに研究されてきた。現在でも、コンピュータの性能向上の恩恵もあり、その進歩・進展は今後も継続していく勢いである。特に、カメラに撮影された対象の三次元情報を取得するための手法は 1990 年以降大きく進歩し、いろいろな応用が進んでいる。このようなコンピュータビジョン研究は、古くはロボットのための視覚機能の機械による実

現を目的としていたが、1990 年代中ごろから映像メディアを人間に見せることを目的にする研究 [1] が流行して以来、映像メディア生成や表示のための不可欠な技術として認識されるようになってきた。

本論文では、カメラに撮影された二次元映像・画像から推定された三次元情報を利用して生成・表示された映像メディアを「三次元映像メディア」と位置づけ、ディジタルエンターテイメントのための三次元映像メディアの生成と表示に関して、筆者らがやってきた研究について述べる。

これまでに筆者らは、三次元映像メディアの実現のための基盤技術として、自由視点映像生成技術と、拡張現実感・複合現実感 (AR/MR) 技術に着目してきた。自由視点映像生成技術は、観察者が両眼立体視ディスプレイなどの特殊な三次元表示ディスプレイを使ったり、観察視点を自由にコントロールしたりすることによって、映像の三次元性をできるだけリアルに体験できるようにする技術である。ここでは、生成した映像再生における視点の自由度や、再生映像の臨場感を高めることが重要となる。一方、AR/MR 技術は、ビデオシースルー、光学シースルーなどの表示装置を用いて、現場で撮影されたライブ映像上に CG やテキストなどの付加情報を表示することにより、ライブ映像に撮影された情報を拡張・強調表示したり、ライブ映像と CG 映像をシームレスに表示したりする技術で

<sup>†</sup> 慶應義塾大学大学院理工学研究科, 横浜市

Graduate School of Science and Technology, Keio University,  
3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi, 223-8522 Japan

a) E-mail: saito@hvrl.ics.keio.ac.jp

ある．ここでは、ライブ映像上に情報を付加するために、表示のリアルタイム性と表示情報とライブ映像の整合性を確保することが重要になる．

本論文では、まず、2. において、三次元映像メディアの基盤技術としての自由視点映像生成技術の動向について解説する．そして3. では、デジタルエンターテインメントのために自由視点映像生成技術を応用した事例として、野球映像においてキャッチャーと審判の真後ろに視点を設定し、キャッチャーや審判を消去することにより、ピッチャーの投球をキャッチャーの真後ろから観察可能な映像を生成するための手法について述べる．

続いて、三次元映像メディアのためのもう一つの基盤技術としての AR/MR 技術の動向について、特にカメラトラッキング技術に焦点を当てながら4. で解説する．そして5. では、カメラトラッキングをデジタルエンターテインメント向け AR/MR 表示に応用した事例として、スポーツ観戦システム、ビリヤード支援システム、ギター演奏支援システムについて述べる．更に、映像検索や特定物体認識技術を利用した新しい AR/MR 表示として、筆者らが最近取り組んでいる、地図や写真への CG 画像やアノテーション表示を行うシステムについて6. で述べる．そして、7. では今後の三次元映像メディアの展望についてまとめる．

## 2. 自由視点映像生成技術の動向

CMU で行われた Virtualized Reality Project [2] をきっかけにして、複数のカメラにより撮影された多視点映像から動的シーンの三次元形状を復元し、それに基づいて任意の視点からの映像を生成する研究が、1990 年代後半から国内外で盛んに行われるようになってきた．このような映像生成処理の基本的な流れは図1に示すようになる．まず、多視点カメラシステムを利用して同一の対象シーンを撮影し、撮影された多視点ビデオ映像から、対象シーンの三次元形状を復元する．復元された三次元形状を利用することにより、対象シーンに対して任意の視点の映像を再現することができる．この技術により、対象シーンを観察する視点を自由に選択することが可能になることから、近年では自由視点映像生成技術と呼ばれるようになっている．

この自由視点映像生成技術において重要な要素技術は、多カメラ（視点）画像からの三次元形状復元手法である．このために広く利用されている方法の一つに

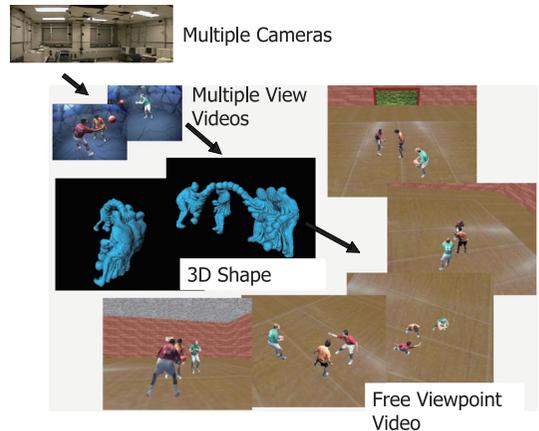


図1 自由視点映像生成処理の流れ  
Fig. 1 Flow of the process of generating free-viewpoint video.

視体積交差法 [3] がある．これは、各画像に撮影される対象物体のシルエットから決定される物体の存在可能範囲を複数カメラで交差させることにより決定される領域を対象物体形状とするものであるが、原理的に凹形状が復元できないなどの欠点があり、これを初期推定形状として形状推定精度を向上させる研究が広く行われてきた [4]．最近数年の間に、計算時間の短縮によるリアルタイム化 [5] や、形状復元精度を向上させるための研究 [6] のように、より高度な三次元復元を追求する研究が進んできた．

このような多視点カメラによる三次元復元を利用する映像処理技術では、多視点カメラが同一の対象シーン・物体を撮影しておく必要があり、更に、多視点カメラを幾何学的にキャリブレーションし、更に異なる視点でできるだけ被写体の明るさ・照明条件を同一にしておく必要があるなどのため、多視点カメラを備えた特別なスタジオ [7] が必要になったり、それを前提にしていたりすることが多い．

一方、このような多視点カメラ撮影をスポーツ映像の中継に利用しようとする試みは、CMU によるアメリカンフットボールの eyeVision 放映で利用されたことをきっかけにして盛んに行われるようになってきた [8], [9]．前述のように、多視点カメラの幾何学的なキャリブレーションが重要となるが、実際のスタジアム撮影では、それを行うための労力が大きく、実用上はキャリブレーションを行うことが困難なことが多い．

これに対して、筆者らは、スタジアムの観客席に並べて撮影しただけの多視点映像から、自由視点映像を

生成する手法を提案した [10] .ここでは, 対象シーンの三次元構造を平面の組合せとして考え, グラウンド, 遠方の観客席, そして, それぞれの選手領域, が平面構造になっていると近似することによりシーンの三次元構造復元を暗に行い, そこから得られる多視点映像間の画素の対応関係から任意視点における映像を合成している .

このときの仮想的な視点 (仮想カメラ) は, 実際に設置されたカメラ (実カメラ) の間を補間する位置に設定されるため, 実カメラとほぼ同様の視野での自由視点映像が生成される .したがって, このような自由視点映像生成法で生成できる映像は, 実際にそこにカメラを置きさえすれば, 実際に取得可能なものにすぎなかった .しかし, 自由視点映像生成の本来の意義は, 実際にはカメラを置くことが不可能な位置からの映像生成であり, 例えば, フィールド内にいる選手そのものの視点からの映像が合成できれば, そのインパクトは大きいものと考えられる .このような試みはいくつかなされている [11], [12] もの, 実際に複数視点のカメラを設置している場所と, 映像を合成する場所が大きく違いすぎるために, 生成される映像の画質が実際に耐え得るものにはなっていないのが実情である .

そこで近年我々が取り組んでいるのは, 実際にはカメラを置くことができない視点での映像を生成する試み [13] である .ここでは野球映像を対象とし, 従来の野球中継などでは提供されていない新しい視点を生成することを目的として, “キャッチャー視点”での映像を生成している .

ピッチャーが投球する瞬間は, 視聴者にとって非常に興味深いシーンであるが, 通常はピッチャーの背面であるバックスクリーン側からの映像がほとんどである .また, 実際にキャッチャーから見た映像を撮影するためには, キャッチャーにカメラを取り付けたり, キャッチャーの前方にカメラを配置したりすることが必要であるが, 試合の妨げになるので現実には不可能である .更に, キャッチャーの後方にカメラを設置するのでは, キャッチャー自身や審判が障害物となり, ピッチャーの投球を撮影することはできない .

そこで本研究では, 3 台のカメラをキャッチャーの後方に配置し, ピッチャー横に 1 台のカメラを設置した多視点カメラを利用して, キャッチャー真後ろの映像からキャッチャーと審判の領域を障害物として除去し, 他のカメラからの映像をもとにピッチャーの投球シーンをキャッチャーの後方カメラ視点に変換し, 合

成する .

このような映像生成は, 実際に撮影した視点とは異なる視点の映像を生成するという意味では, 自由視点映像生成技術を応用した映像生成である .また同時に, 現実世界に新たに仮想的な物体を追加表示する複合現実感 (Mixed Reality: MR) に対して, 現実中存在する物体を隠して消し去る技術とも位置付けることができる .この意味で, これを, 隠消現実感 (Diminished Reality: DR) [14], [15] と呼ぶ .次章では, この野球映像の隠消現実提示法の詳細について述べる .

### 3. 多視点カメラによる野球映像の隠消現実提示

本研究でのカメラ配置及び実験環境を図 2 に示す .キャッチャー後方に配置した 3 台のカメラのうち, 中央のカメラがキャッチャー視点映像の基準となる .図中に示すとおり, 中央カメラで撮影された映像には, キャッチャーと審判に隠れているので, ピッチャーの姿が写っていない .一方で, 左右のカメラによる映像中には, それぞれピッチャーの姿がとらえられている .そこで, この左右カメラ中のピッチャー領域を, 中央カメラ視点へと変換し合成する .ここでは, 図 3 に示すように, ピッチャー領域を平面とみなし, その平面射影変換により右カメラと左カメラの両方に撮影されたピッチャー領域を中央カメラの視点に変換する .こ

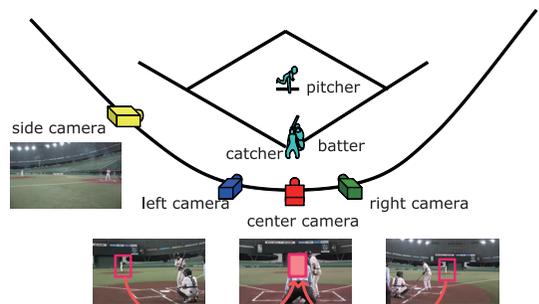


図 2 カメラ配置と実験環境  
Fig. 2 Camera arrangement and experimental environment.



図 3 ピッチャー領域に仮定する平面領域  
Fig. 3 Virtual plane at the pitcher area.

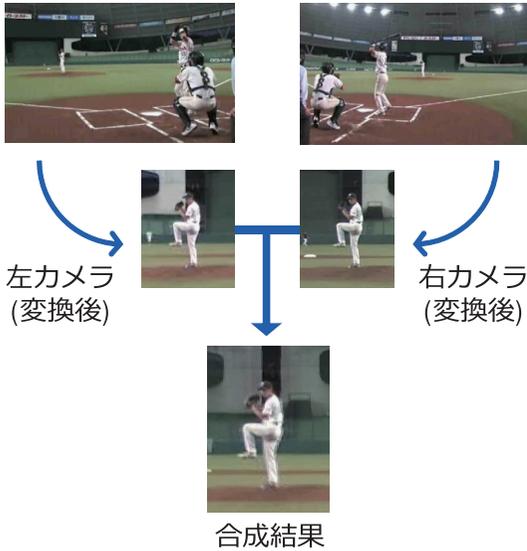


図 4 平面射影変換によるピッチャー領域の合成

Fig. 4 Rendering of pitcher area image via homography.

の平面射影変換に必要な Homography 行列は、ピッチャーマウンドのプレート上に事前に設置した平面を基準として、キャッチャー後方の 3 台のカメラについて事前に算出する。実際の合成例を図 4 に示す。

次に、中央カメラ映像中における障害物領域であるキャッチャー及び審判の領域を除去し、その上に合成されたピッチャー領域を重畳する。これにより、中央カメラでは審判とキャッチャーが除去され、その領域にピッチャーの投球の様子が表示された映像を生成することが可能になる。このとき、障害物領域の抽出には、選手や審判のユニフォームの色や、選手や審判が特定の範囲内に常に存在しているといった事前知識を利用しながら、Graph Cut による領域分割手法を適用した。

更に、ピッチャーと同じく、障害物に隠されてしまっているボールに関して、左右カメラから中央カメラへと視点変換を行い、重畳表示することで映像化する。ボールの視点変換のために、事前にキャッチャー後方の 3 台のカメラ間に成立する Trifocal Tensor を求めておく。そして、左右のカメラで撮影されているボール位置を追跡し、Trifocal Tensor によって中央カメラへと視点変換することで、図 5 のように、障害物に隠されている投球の様子が中央カメラにおいて映像化することが可能になる。

上記の手法の有効性を確認するために、西武ドーム

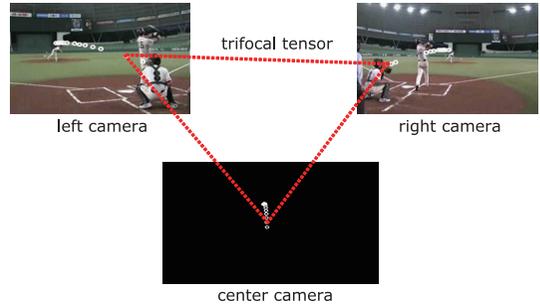


図 5 Trifocal Tensor によるボールの視点変換

Fig. 5 Viewpoint transfer of ball trajectory using Trifocal Tensor.

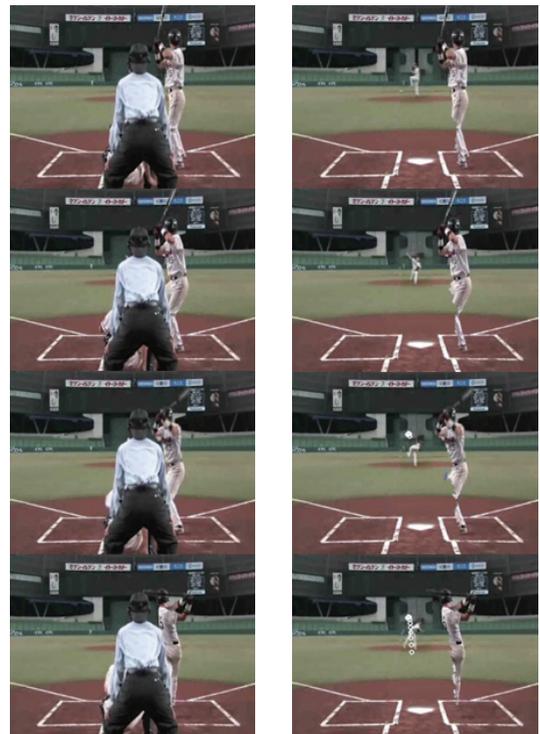


図 6 合成されたキャッチャー視点映像

Fig. 6 Synthesized catcher's viewpoint video.

にて撮影した映像シーケンスに対して本手法を適用した。使用した機材は以下のとおりである。

- カメラ：Canon XHG1 (1440 × 1080)
- PC：Windows XP Professional

Intel Core 2 Duo 2.80 GHz 3.2 GB

生成されたキャッチャー視点映像を図 6 に示す。キャッチャー及び審判の領域を削除し、隠されていたピッチャーの姿が現れている様子が分かる。ピッチャーの

投球動作が中央カメラ視点において良好に合成できていることが分かる。

また、ボールの軌跡も同様に映像化されていることが分かる。本実験では 30 fps のカメラで撮影したため、ボールの詳細な軌跡を取得することは不可能であったが、今後は高速カメラを利用することによって、より正確なボール軌跡を追跡できるものと考えられる。

本実験では、障害物領域の検出処理以外は 16 fps で動作している一方で、Graph Cut による領域検出処理においては、1 フレーム当り 1 秒以上かかっている。本研究は、将来的にはテレビの野球中継において、投球シーンのリプレイなどに用いることを想定しているため、準リアルタイムの処理が要求される。したがって、HD サイズの画像に対してリアルタイムに領域検出を行うことができる手法を確立することが今後の課題である。

#### 4. 拡張現実感・複合現実感の動向

撮影した映像にリアルタイムで CG 映像やアニメーション情報を重ね合わせることによる映像情報提示技術である拡張現実感・複合現実感 (AR/MR) においては、処理のリアルタイム性はいうまでもないが、撮影されたライブ映像と、その上に重ね合わせて表示する CG 映像やアニメーション情報との間の整合性を確保することが不可欠である。

その整合性として第一に挙げられるのが、ライブ映像により撮影されたシーンと重畳表示する CG 映像やアニメーションの間の幾何学的整合性である。これを実現するための重要な技術がカメラトラッキングである。カメラトラッキングとは、移動するカメラの位置姿勢を推定することであり [16]、移動ロボットや自動車の自己位置推定 [17] や、映画製作におけるポストプロダクションにおける CG 映像と撮影映像との重ね合わせ [18] など、いろいろな目的に利用されている。

AR/MR のためには、リアルタイム性が重要となるので、映画のポストプロダクションのように後処理として使う場合は異なり、カメラトラッキング処理のリアルタイム性を確保することが必須となる。そこでカメラに取り付けた位置センサや加速度センサなどを利用してリアルタイム性を確保することも考えられるが、このようなセンサだけでは十分な精度を確保することが困難なため、撮影されたライブ映像からのカメラトラッキングを併用したり、センサを使わずにカメラで撮影したライブ映像のみを利用したりすることが

ほとんどである。

カメラで撮影した映像のみを利用した AR/MR のためのカメラトラッキング法としては、撮影対象範囲内に、形状が既知な白黒の正方形マーカを配置する方法が広く利用されている [19]。マーカが撮影範囲内に存在する限り、リアルタイムで安定にカメラトラッキングを行うことができるので、AR/MR のためのカメラトラッキング手法として広く利用されている。

しかしながら、マーカが視覚的に美しくないことや、そのデザインにも制約があるなどといったことから、特殊なマーカを使わずに AR を実現するために、一般的な物体を対象として、その自然特徴を用いるアプローチも盛んに研究されてきた [20]。今日では、端末として携帯電話や UMPC (Ultra-Mobile PC) などが用いられており、限られた計算機環境においても AR を実現するための処理の高速化がなされている [21]。

上記はいずれの場合も、カメラトラッキングで利用する対象物体の形状が既知 (大抵は平面) であることを前提とした方法であった。これに対して、最近では対象物体形状が未知な環境におけるカメラトラッキングの研究も行われている。PTAM (Parallel Tracking And Mapping) [22] などに代表される SLAM (Simultaneous Localization And Mapping) に基づくアプローチでは、カメラの位置姿勢を持続的に推定するとともに、未知な空間の三次元復元を同時に行うことで、形状が未知の広い範囲における AR/MR が可能となった。特に PTAM では、カメラの位置姿勢推定と空間の自然特徴点の三次元復元を別スレッドで行うことで、ノート PC 上で SLAM ベースのリアルタイムな AR を実現した。

次章では、このような AR/MR 表示を利用したエンターテインメント応用事例として筆者らが提案した、スポーツ観戦システム、ピリヤード支援システム、ギター演奏支援システムについて述べる。更に、映像検索技術も併用した AR/MR 表示の新しい応用事例として、複数の地図や写真上への CG 画像・アニメーション情報の提示システムについて述べる。

### 5. AR/MR 表示を利用したデジタルエンターテインメントシステム

#### 5.1 スポーツ映像の観戦システム

テーブル上に配置したサッカースタジアム模型の上に、多視点カメラ映像から生成したサッカー選手の自由視点映像を重畳表示し、目の前で試合が行われてい

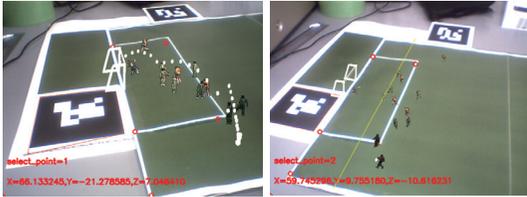


図 7 テーブル上でサッカーの試合を AR/MR 観戦するシステム

Fig. 7 AR/MR observing system of soccer match.

るかのように見せるサッカーの AR システムを開発した [23]。ここでは、サッカースタジアム模型に描かれた直線を利用してカメラトラッキングを行ったが、安定したトラッキングができなかった。そこで、正方形マーカと、サッカースタジアム模型の直線を併用することにより、カメラトラッキングの安定性を向上することができた [24]。図 7 に、このシステムを使って実際のサッカーの試合をテーブル上で観戦している様子を示す。

サッカーだけではなく、野球の観戦にも上記のシステムと同様な AR/MR 表示技術を使ったシステムを構築した。ここでは、実際に撮影された映像を AR/MR 表示するのではなく、野球のスコアブックにより試合の流れが状態遷移で表せることを利用して、既に行われた野球の試合を CG キャラクタによって再現するシステムとなっている [25]。CG キャラクタはテーブルに配置された野球盤モデル上に重畳表示され、ユーザは自分の好きな視点へ動きながら試合を観戦することができる。本システムでは、野球盤モデル上に複数個正方形マーカを配置し、野球盤モデルの一部の領域しか撮影していない状況においても、最低一つのマーカが撮影される可能性を向上させることにより、野球盤モデルに対して広範囲な位置と方向からの観戦を可能にしている。また、複数マーカの配置があらかじめ未知であっても、それらの相互関係を自動推定する方法を新たに提案し、それを実装している。

上記のシステムは、AR/MR 表示技術を実世界のテーブル上で映像鑑賞に利用したものであるが、AR/MR 表示技術を使えば、実世界とのインタラクションを行うことも可能である。この実世界とのインタラクションを利用した例として、テーブル上に配置したボウリングレーンモデル上で、現実物体であるボールを転がすと、その動きに応じて仮想物体であるボウリングピンが倒れる AR ボウリングシステムも提

案した [26]。ここでは、ユーザ自身が動かした現実のボールによって、CG 映像であるピンが倒れる仕組みになっている。

これらの応用事例では、スタジアム模型、野球盤モデル、及びマーカという、平面形状であることが既知な物体を利用することにより、カメラトラッキングをリアルタイムで実現できたことが重要な要素技術である。

## 5.2 ビリヤード

上記と同様に平面物体に基づくカメラトラッキングを応用した AR/MR 表示の応用例として、ビリヤード支援システム [27] の開発を行った。これは、ユーザがビリヤード台を撮影することで、その球の配置に対する打ち方などの支援情報を AR によって提示するシステムである。

本システムで AR/MR 表示するのは、球の落とし方に関する支援情報である。はじめに、ユーザが撮影したビリヤード台の画像から、台上の球の配置を算出する。次に、コンピュータシミュレーションに基づく球の配置の解析を行うことで、支援情報を算出する。最後に、ユーザが再びビリヤード台の撮影を行い、支援情報を重畳表示する。

更に本システムでは、ボール位置の推定、更に支援情報をビリヤード台上に AR 表示するために、ビリヤード台を基準としたカメラトラッキングを行う。ここでは、ビリヤード台が特有の色である緑色をした長方形の平面物体であることに着目し、このビリヤード台の四つのコーナを追跡することによりカメラトラッキングを実現している。その流れは下記のとおりである。

前提として、ビリヤード台を世界座標系として扱うために、台と球の大きさは既知とする。また、ビリヤード台とすべての球を撮影し、画像からそれぞれの RGB の値を計測しておく。AR/MR 表示の実行時は、まず、カメラに撮影された入力画像中の全画素に対し、あらかじめ計測したビリヤード台の RGB と比較することで、ビリヤード台領域を検出する。次に、検出した領域の輪郭線に対し、ハフ変換を用いた直線群の検出を行うことで、ビリヤード台の四つの辺を検出する。最後に、ビリヤード台のコーナを隣り合う辺から算出することで、四つのコーナの世界座標と画像座標の関係からカメラの位置姿勢を算出する。

図 8 に本カメラトラッキング手法を利用して、ビリヤード台上で支援情報を AR/MR 表示している様子



図 8 ビリヤード支援情報の AR/MR 提示  
Fig. 8 AR/MR display of supporting instruction.

を示す。カメラトラッキングをリアルタイムで行うことができるので、カメラを動かしても、ビリヤード台上に表示される支援情報はビリヤード台に固定されたまま表示されるようになっており、このように、実際のビリヤード台上に支援情報が提示されることで、直感的な理解が可能となる。

### 5.3 ギター演奏支援システム

これまで紹介したシステムは、固定されているテーブルやビリヤード台に対するカメラの位置姿勢を推定するものであった。一方、固定されたカメラに対して移動する物体の位置姿勢の推定も同様な手法で行うことが可能であり、ここでは、このような状況におけるカメラトラッキングを応用したシステムとして、ギター演奏支援システム [28] を紹介する。

図 9 に示すように、カメラとディスプレイが固定され、これの前でギターを演奏することを考える。このとき、カメラに撮影される画像を左右反転してディスプレイに表示すると、あたかも鏡に自分の演奏する姿が映ったような状況をつくりだすことができる。このとき、ギターの位置姿勢を推定し、これに合うように、CG 映像で演奏すべき手の形状を表示すると、演奏者はその手の CG 映像に合致するように自分の手を動かすだけで正しい演奏を行うことができるようになる。

このシステムを実現するために、まず、先に紹介したような正方形のマーカをギターに取り付け、ギターの位置姿勢を推定し、それに合わせて CG 映像を表示するシステムを構築した。このとき、カメラトラッキングの安定性を向上するために、ギターのエッジ検出を行い、このエッジの画像内の位置もカメラトラッキ

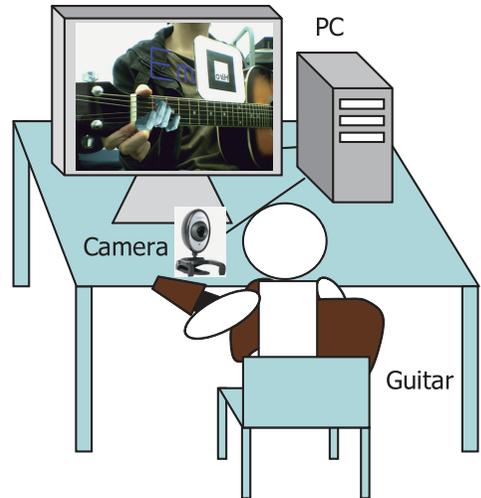


図 9 AR/MR 提示によるギター演奏支援システム  
Fig. 9 Guitar play supporting system via AR/MR display.

ングのために利用した。

更に、マーカなしでもトラッキングを可能とするために、ギターのフレットの寸法などのギターの幾何学的形状に関する事前知識を活用する手法も提案した [29]。この方法により、前述のマーカを不要としたギターのトラッキングを実現することができた。

## 6. 映像検索技術に基づく AR/MR 表示

AR/MR 表示技術における新たな研究の動向として、検索や認識技術の利用が挙げられる。AR/MR 表示というと、4. に述べたようなカメラトラッキングが重要であり、盛んに研究も行われているが、これはあくまでも撮影しているライブ映像と表示コンテンツである CG 画像やアニメーションとの幾何学的整合性を確保するための技術にすぎない。AR/MR 表示においては、その状況に適した必要な表示コンテンツを検索し表示することも重要な要素技術であり、リアルタイム性が必須である AR/MR においては、この表示コンテンツ検索を効率的に行うことが AR/MR 表示技術においては極めて重要となる。

そこで筆者らは、このような表示コンテンツの検索技術に基づく AR/MR 表示に関する二つの研究を行ってきた。

一つ目は、地図画像検索技術に基づく地理データの AR/MR 提示 [30] である。これは、紙に印刷された地図をカメラで撮影することで、その地図がどの領域

の地図であるかを識別し、その領域の三次元地理データを紙地図上に重畳表示するものである。この研究では、地図検索手法として、LLAH (Locally Likely Arrangement Hashing)[31]を用いた交差点マッチングを提案した。単純なマーカトラッキングや特徴点トラッキング、緑領域トラッキングでは、幾何学的な位置姿勢の推定は可能であるが、どの領域の地図か、ということを確認することは困難である。そこで、LLAHを導入し、地図の検索を容易に行えるようにした。

二つ目は、特定物体認識技術を用いたコンテンツのAR/MR提示[32]である。これは、データベースに登録されている物体をカメラで撮影することで、その物体に関連付けられた様々なコンテンツを物体上に重畳表示するものである。物体の登録、検索及びトラッキングの三つの技術の一つのフレームワークに統合することで、新たなAR/MRに対するアプローチの提案を行った。以下に、それらの詳細について述べる。

### 6.1 地理データのAR/MR提示

はじめに、地図を撮影した画像から交差点の抽出を行い、地理データベースに登録された交差点とのマッチングを行うことで、地図の識別を行う。次に、対応づけられた交差点を用いて地図に対するカメラの位置姿勢を求めることで、地図に関連づけられた地理データを地図上に重畳表示する。

地図の識別を行うことで、複数の地図に対する地理データの提示が可能となる。そこで、地図の識別を行う特徴量として、交差点の局所的な位置関係を利用し、特徴量の記述はLLAHを用いて行う。なお、地理データベースから作成した交差点をプロットされた地図を利用する。

前処理では、地理データベースから交差点の座標群を抽出し、各交差点に対してLLAHの特徴量を算出しておく。オンライン処理では、画像中から抽出した交差点に対して前処理と同様にLLAHの特徴量を算出することで、地理データベースに登録されている交差点とのマッチングを行うことができる。最も多くの交差点が対応づけられた地図をユーザが撮影している地図として出力する。

交差点を特徴点として扱うことで、地図に対するカメラの位置姿勢を算出する。地図の識別において対応付けされた交差点群の中には、誤対応が含まれるため、カメラの位置姿勢を算出する際に、RANSACを利用する。位置姿勢が算出できた場合には、図10のように地理データベースに登録されている地理データを重

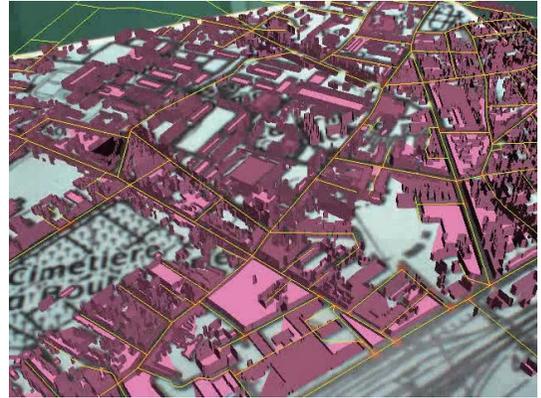


図 10 建物の三次元モデルの重畳表示  
Fig. 10 AR/MR display of 3D model of city buildings onto map.

表 1 地図に対する識別成功率  
Table 1 Successful rate of recognition of map.

交差点の数	地図の数	検索に成功した数
21 ~ 40	11	8
41 ~ 60	11	11
61 ~ 80	9	9
81 ~ 100	6	6
101 ~	11	11
合計	48	45

畳表示することができる。このように、交差点マッチングを用いることで、紙地図と地理データの関連付けを行うことが可能となった。

交差点のマッチングに基づく地図画像検索の有効性を示す実験として、交差点数と識別の成功率の関係を検討した。本実験では、フランスのナント市の地理データベースを用いて48枚の地図を作成し、地図を上部から撮影した。表1に示したように、48枚中45枚の地図について、識別が成功した。特に、1枚の地図に含まれる交差点数が40を超えている場合は100%の識別成功率になっている。これは、各交差点の局所的な位置関係が異なることによって、特徴量として用いることが可能であったことを示している。

6.2 特定物体認識技術のAR/MR表示への応用  
カメラで撮影した物体に応じて提示するコンテンツの切換を実現するために、BOF (Bag-of-Features) の考え方に基づく特定物体認識を利用したAR/MR表示システムを構築した。本システムでは、あらかじめ色々な物体を登録しておき、カメラで撮影したライブ映像から登録した物体を検出して認識し、認識した物体に応じたコンテンツをAR/MR表示するもので

ある。

物体の登録の際には、認識対象物体をカメラにより撮影し、この映像中から検出されたキーポイントを検出・追跡する。そして、追跡された同一キーポイントの様々なパターンを、BOF (Bag-of-Features) の考え方を利用して Visual Word 表現することにより、カメラで撮影される条件によって見え方が様々に変化する同一キーポイントのパターン変化を一つの表現形態に置き換える。このキーポイントを表す Visual Word ごとに、それが属する物体 ID を登録した転置インデックスを作成しておく。同時に、各 Visual Word に対し、物体上の画像座標も保持する。更に、物体ごとに AR/MR 表示を行うコンテンツとの関連付けも行っておく。

一方、物体の検索時には、カメラにより撮影されるライブ映像から検出されたキーポイントに対応する Visual Word を求め、転置インデックスからそれに対応する物体の ID 群を取得することで、画像中に写っている物体の候補を得ることができる。次に、各物体候補に対して、複数個検出された Visual Word を用いることで画像とデータベース中の物体の特徴点マッチングを行う。最後に、このマッチングの結果に対して、ホモグラフィ(平面な物体)や基本行列(非平面な物体)を用いた検証を行うことで、画像中に写っている物体を決定する。

このマッチングの結果からカメラの位置姿勢を算出することで、図 11 のようにコンテンツを物体上に重畳表示する。一度検索された物体については、フレーム間の特徴点トラッキングを行うことにより、カメラ

が動いた場合にも継続的にコンテンツを重畳表示することが可能になっている。

## 7. む す び

本論文では、デジタルエンターテインメントのために、カメラに撮影された二次元映像・画像から推定された三次元情報を利用して生成・表示された映像メディアである「三次元映像メディア」の基盤技術として自由視点映像技術と AR/MR 表示技術に着目し、これらの動向について解説した。そして、三次元映像メディアを実際に利用したデジタルエンターテインメントのために筆者らが開発したシステムを紹介した。

本論文では、「三次元映像メディア」の「三次元性」の再現・表示のための処理に主眼を置いて説明した。この「三次元性」というのは、我々が存在している「実世界」のもつ基本的性質そのものであり、その意味では、「実世界志向映像メディア」という方向に発展していくものである。つまり、単に「三次元性」の再現・表示だけではなく、三次元映像メディアの実時間性やインタラクティブ性などを更に高め、実世界志向を実現するために必要な技術を更に発展させていくことが重要と考えている。

謝辞 本論文で述べた研究の一部は、(独)情報通信研究機構(NICT)の高度通信・放送研究開発委託研究「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発」、(独)科学技術振興機構戦略的創造研究推進事業(CREST)研究領域「デジタルメディア作品の製作を支援する基盤技術」、文部科学省グローバル COE プログラム「アクセス空間支援基盤技術の高度国際連携」、それぞれの一環としてなされたものである。

## 文 献

- [1] 大田友一, “映像メディアのためのパターン認識・理解” 信学技報, PRU 95-36, May 1995.
- [2] T. Kanade, P.W. Rander, S. Vedula, and H. Saito, “Virtualized reality: Digitizing a 3D time-varying event as is and in real time,” International Symposium on Mixed Reality (ISMR99), pp.41-57, Yokohama, Japan, 1999.
- [3] Lavakusha, A.K. Pujari, and P.G. Reddy, “Linear oc-trees by volume intersection,” Comput. Vis., Graph. Image Process., vol.45, pp.371-379, 1989.
- [4] 富山仁博, 片山美和, 岩館祐一, 今泉浩幸, “視体積交差法とステレオマッチング法を用いた多視点画像からの 3 次元動オブジェクト生成手法” 映像学誌, vol.58, no.6, pp.797-806, June 2004.
- [5] T. Matsuyama, X. Wu, T. Takai, and T. Wada,



図 11 関連付けされたコンテンツの重畳表示

Fig. 11 AR display of contents related to pictures.

- “Real-time dynamic 3D object shape reconstruction and high-fidelity texture mapping for 3D video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.14, no.3, pp.357–369, 2004.
- [6] J. Starck and A. Hilton, “Spherical matching for temporal correspondence of non-rigid surfaces,” *IEEE International Conference on Computer Vision (ICCV2005)*, vol.2, pp.1387–1394, Beijing, China, 2005.
- [7] J. Starck, S. Nobuhara, A. Maki, A. Hilton, and T. Matsuyama, “The multi-camera 3D production studio,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.19, no.6, pp.856–869, 2009.
- [8] J.Y. Guillemaut, J. Kilner, and A. Hilton, “Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes,” *IEEE International Conference on Computer Vision (ICCV2009)*, Sept. 2009.
- [9] 北原 格, 橋本浩一郎, 亀田能成, 大田友一, “サッカーの自由視点映像提示における気の利いた視点選択手法,” *日本バーチャルリアリティ学会論文誌*, vol.12, no.2, pp.171–180, 2007.
- [10] 稲本奈穂, 斎藤英雄, “多視点スポーツ映像からの自由視点映像合成と提示,” *信学論 (D-II)*, vol.J88-D-II, no.8, pp.1693–1701, Aug. 2005.
- [11] K. Hayashi and H. Saito, “Synthesizing free-viewpoint images from multiple view videos in soccer stadium,” *IEEE International Conference on Computer Graphics, Imaging and Visualisation (CGIV2006)*, pp.220–225, July 2006.
- [12] N. Kasuya, I. Kitahara, Y. Kameda, and Y. Ohta, “Automatic player’s view generation of real soccer scenes based on trajectory tracking,” *3DTV-Conf. 2009*.
- [13] 橋本昂宗, 植松裕子, 斎藤英雄, “多視点カメラ撮影による野球のシースルー映像生成,” *信学技報*, CQ2009-67, PRMU2009-166, SP2009-107, MVE2009-89, Jan. 2010.
- [14] 榎本暁人, 斎藤英雄, “複数のハンディカメラの協調利用による遮蔽物体除去映像のオンライン生成,” *映情学誌*, vol.62, no.6, pp.901–908, June 2008.
- [15] S. Jarusirisawad, T. Hosokawa, and H. Saito, “Diminished reality using plane-sweep algorithm with weakly-calibrated cameras,” *Progress in Informatics*, no.7, pp.11–20, March 2010.
- [16] V. Lepetit and P. Fua, “Monocular model-based 3D tracking of rigid objects: A survey,” *Foundations and Trends in Computer Graphics and Vision*, vol.1, no.1, pp.1–89, Oct. 2005.
- [17] W. Burger and B. Bhanu, “Estimating 3D egomotion from perspective image sequence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.12 no.11, pp.1040–1058, Nov. 1990.
- [18] H.S. Sawhney, Y. Guo, J. Asmuth, and R. Kumar, “Multi-view 3D estimation and applications to match move,” *Proc. IEEE Workshop on Multi-View Modeling & Analysis of Visual Scenes*, pp.21–28, June 1999.
- [19] H. Kato and M. Billinghurst, “Marker tracking and HMD calibration for a video-based augmented reality conferencing system,” *Proc. IWAR*, pp.85–94, 1999.
- [20] U. Neumann and S. You, “Natural feature tracking for augmented reality,” *TMM*, vol.1, pp.53–64, 1999.
- [21] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg, “Pose tracking from natural features on mobile phones,” *Proc. ISMAR2008*, pp.125–134, 2008.
- [22] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” *Proc. ISMAR2007*, pp.225–234, 2007.
- [23] N. Inamoto and H. Saito, “Immersive observation of virtualized soccer match at real stadium model,” *Second International Symposium on Mixed and Augmented Reality (ISMAR03)*, pp.188–197, Oct. 2003.
- [24] A. Enomoto and H. Saito, “AR display for observing sports events based on camera tracking using pattern of ground,” *Third International Conference on Virtual and Mixed Reality (VMR2009)*, *Lect. Notes Comput. Sci.*, vol.5622/2009, pp.421–430, July 2009.
- [25] Y. Uematsu and H. Saito, “Visual enhancement for sports entertainment by vision-based augmented reality,” *Advances in Human-Computer Interaction*, vol.2008, Article ID 145363, 2008.
- [26] Y. Uematsu and H. Saito, “Interactive AR bowling system by vision-based tracking,” *Proc. International Conference on Advances in Computer Entertainment Technology (ACE ’07)*, pp.236–237, June 2007.
- [27] 内山英昭, 斎藤英雄, “ハンディカメラ入力からのシーン解析に基づくピリヤード戦略発想支援 AR 表示システム,” *日本バーチャルリアリティ学会論文誌*, vol.12, no.2, pp.159–170, 2007.
- [28] 元川洋一, 斎藤英雄, “拡張現実表示技術を用いたギターの演奏支援システム,” *映情学誌*, vol.61, no.6, pp.789–796, 2007.
- [29] 元川洋一, 斎藤英雄, “ギター演奏支援のための構造特徴追跡を利用したマーカレス AR 表示,” *日本バーチャルリアリティ学会論文誌*, vol.13, no.2, pp.267–278, 2008.
- [30] H. Uchiyama, H. Saito, M. Servieres, and G. Moreau, “AR city representation system based on map recognition using topological information,” *Third International Conference on Virtual and Mixed Reality (VMR2009)*, *Lect. Notes Comput. Sci.*, vol.5622/2009, pp.128–135, July 2009.
- [31] 中居友弘, 黄瀬浩一, 岩村雅一, “特徴点の局所的配置に基づくデジタルカメラを用いた高速文書画像検索,” *信学論 (D)*, vol.J89-D, no.9, pp.2045–2054, Sept. 2006.
- [32] J. Pilet and H. Saito, “Virtually augmenting hundreds of real pictures: An approach based on learning, retrieval, and tracking,” *IEEE Virtual Reality (IEEE-VR2010)*, March 2010.

(平成 22 年 7 月 2 日受付)



齋藤 英雄 (正員)

1992 慶應義塾大学大学院理工学研究科博士課程了。博士(工学)。1992より、同大理工学部助手、専任講師、助教授を経て、2006より、同大理工学部情報工学科教授。1997から1999まで、CMU ロボティクス研究所訪問研究員。コンピュータビジョン、仮想現実感、三次元映像などに関する研究に従事。映像情報メディア学会、日本バーチャルリアリティ学会、画像電子学会等各会員。IEEE シニア会員。



植松 裕子 (正員)

2004 慶大・理工・情報工学卒。2009 同大大学院理工学研究科後期博士課程了。博士(工学)。2009より、同大理工学部情報工学科助教。コンピュータビジョン、複合現実感に関する研究に従事。



内山 英昭 (学生員)

2006 慶大・理工・情報工学卒。2010年9月同大大学院博士課程了。博士(工学)。2010年9月より、フランス INRIA Rennes 訪問研究員。画像処理に関する研究に従事。



ピレ ジュリアン

2008 EPFL (スイス連邦工科大学ローザンヌ校) 博士課程了。Ph.D. 2008年12月慶應義塾大学大学院理工学研究科特別研究助教。2010年5月よりスイスグーグル社に勤務。コンピュータビジョンを用いた AR/MR 表示に関する研究に従事。

CVPR2005 Best Paper Award.