







## —— AR 研究の最先端は非マーカー型に移行しているようです。


大学はチャレンジすることを目的としているので、最先端の研究では位置合わせが難しい、マーカーを使わない方式が主流です。私自身も、現在はマーカーを使わない研究しかしていません。


ただ、マーカーも適材適所で使えばよいのではないのでしょうか。アプリケーションによって、マーカーがある方がよいものと、目障りになるのではない方がよいものがあると思います。特に携帯電話のARはヘッド・マウント・ディスプレイ(HMD)のARと異なり、ユーザーが意識的に端末を“かざして”使います。そうになるとQRコードのようなマーカーの存在が目印となってユーザーへの“分かりやすさ”や利便性を高めることもあると思います。また、GPSや電子コンパスが使えないシーンも多いですから、マーカー型とマーカーレス型を使い分ける、ということになるのではないのでしょうか。


## ——最先端の研究では英オックスフォード大学のジョージ・クライン氏が開発した「Parallel Tracking and Mapping for Small AR Workspaces」(PTAM)がマーカーレスのARとして注目を集めています。


PTAMは2007年に奈良で開催されたISMARで発表されましたが、私はそのときにプログラム委員長でした。論文をきちんと読んでみましたが、一番インパクトがあったのはやはりデモですね。


PTAMのすごいところは、トラッキング(位置の


マーカー=maker。標識。2次元バーコードのような白黒パターンや登録した画像などを使う。

ヘッド・マウント・ディスプレイ=頭に装着するディスプレイで、ゴーグルやヘルメットのような形状をしている。

QRコード=デンソー(現デンソーウェーブ)が開発した2次元コード。バーコードと異なり2方向に情報を持つので、表示面積が小さくても情報量を多くできる。ISO/IEC 18004、JIS-X-0510として規格化されている。URLが格納されたQRコードを携帯電話のカメラで読み込んで、Webサイトに誘導するなど、既に実用化が進んでいる。

GPS=Global Positioning System。全地球測位システム。

電子コンパス=地磁気から方位を検出するデバイス。地磁気センサーとも呼ばれる。

ISMAR=International Symposium on Mixed and Augmented Reality。ARや複合現実(Mixed Reality)に関する国際会議。

追跡)とマッピング(情報の重ね合わせ)を並列処理しながら、普通のパソコンで十分なフレームレートを確保できていることです。

優れた論文は数多くあっても、そのロジックをプログラムとして実装したときに性能が出ないというのは珍しくありません。PTAMの一つひとつの要素技術はロボティクスの研究などで既にあったものですが、それを非常に高い完成度でシステム化できた点が優れていました。

PTAMのソースコードの公開によって、他のメーカーのAR技術の評価軸が生まれた面もあります。ARを評価する客観的な手法は確立していないので、公平な評価が難しい。PTAMのように誰でも入手できるメーカーレス技術があれば、PTAMよりは良い、悪い、という議論が成り立ちます。

### —— ARの当面の課題は何でしょうか。例えば、コンピューティング・パワーは十分ですか？

パソコンであれば基本的に、コンピューティング・パワーを気にしなくて済む時代になっています。ARToolKit開発当時のマシンは、CPUがPentiumⅢの500MHz、解像度は640×480ドットのVGAという仕様でした。この頃はグラフィックス・カードの性能が十分ではなかったので、処理がかなり重かったです。

ARでは、人間の広い視野を実現するために広角のレンズのカメラを使います。広角では、撮影した画像に大きな歪みができます。これを補正するためにテクスチャマッピングで歪んだ画像を平面に戻すのですが、これをリアルタイムで行うのは当時のパソ


コンでは荷の重い処理でした。


一方、最近ではマーカーを使わないARでも120フレーム/秒で処理できます。120フレーム/秒くらいあれば、ARの対象物を激しく動かしても追従できます。Flashや携帯電話でも、処理速度の面では十分な水準に達しつつあるのではないのでしょうか。

### —では、アプリケーションが当面の課題ということになりますか。

そうですね。工場内のメンテナンスなど業務改善を目的としたAR技術を含めると、大学だけでなく日本の企業も相当な技術を抱え込んでいます。ただ実用化に向けたパスが少ないように思います。

ゲームの「THE EYE OF JUDGMENT」や「電脳フィギュア ARis」、「セカイカメラ」など、注目を集めているAR製品は確かにありますが、後が続くかどうかやや不安です。私のところへもゲーム会社の方などがいらっしゃいますが、多くは「面白いですね」で終わってしまうのが現状です。世の中に出て行くには、まだ何か足りないのでしょうか。

ただ、 デジタルサイネージのようにそれ自体で収益を上げなくてもよいものであれば、可能性はありそうです。

 デジタルサイネージ=コンピュータ技術を用いる街頭広告の総称。

### —デバイスはどうでしょうか。最近では、手のひらサイズで持ち運べるモバイル・プロジェクタの製品化が相次いでいます。

モバイル・プロジェクタをディスプレイにしたARの研究は既にあります。ARはHMDを有力な

モバイル・ディスプレイの一つとして研究が進んできたのですが、そこで重ね合わせる情報はあくまでパーソナルなものです。モバイル・プロジェクトなら、モバイルでありながらARによる情報を複数人で共有できます。そこに何らかの可能性を見いだせるかもしれません。

有力なデバイスとして考えられているのは、今なら携帯電話、カーナビゲーション・システム、ゲーム機あたりでしょうか。いずれもカメラとディスプレイを兼ね備えていて、市場の規模も大きい。

ただ立体的に見せられるディスプレイでなければ、実用性に問題が出てくることがあります。3次元的に整合する位置に合成して表示することと、奥行き感は別物だからです。

例えばカーナビで、奥行き方向に複数の曲がり角があって、そこに情報をオーバーレイしたとします。すると「どの角だ?」と迷うシーンが出てくるわけです。

ユーザーに迷わせないという点では、情報のフィルタリングも課題です。ユーザーはほしい情報だけ見ればいい。例えば街にお店が30軒あって、そのすべてがお客さんを呼び込みたいとします。ですが、そのすべての情報を画面内に表示してしまうと、わけがわからなくなります。

さらに携帯電話や携帯ゲーム機は、バッテリー容量の少なさがネックになります。ARは映像の処理を多用するので、それこそ20分で電池が切れてしまう。おそらく最後に残る応用面での課題は、バッテリーになるのではないのでしょうか。

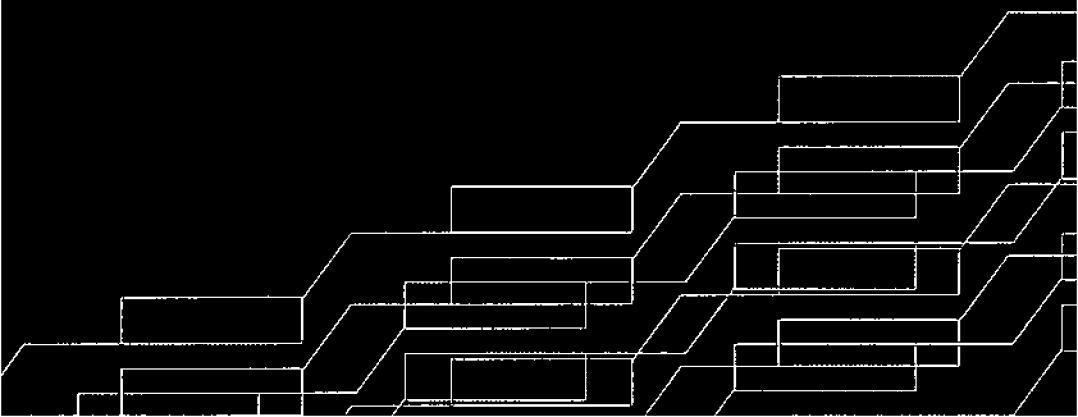
(このインタビューは2009年4月に行いました)

## 加藤 博一 氏

1963年福岡県生まれ。1986年大阪大学基礎工学部制御工学科卒業。大阪大学講師、米国ワシントン大学客員研究員、広島市立大学情報科学部助教授などを経て、2007年から現職。90年代後半から画像計測技術を応用したバーチャルリアリティやARの研究を開始。様々な入出力インタフェースを使うヒューマン・コンピュータ・インタラクションに興味を持っている。2009年3月、ARToolKit開発の功績によりIEEE vgTCの「The 2009 Virtual Reality Technical Achievement Award」を受賞。同年4月、「日経BP技術賞 情報通信部門」を受賞。博士(工学)。

# 第 **3** 部

ARを支える技術



# ARの仕組みを理解する

author: 橋本 直

工学ナビ主宰  
独立行政法人科学技術振興機構 ERATO  
五十嵐デザインインタフェースプロジェクト 研究員 博士(工学)



最初に、私が考える拡張現実という言葉の定義について説明したいと思います。私はARを「現実の環境から受ける知覚情報に、コンピュータによって生成された情報を融合させることで現実を拡張、強化する技術」と定義しています。

「ARは映像技術である」と説明されることが多いようですが、それは少し違います。必ずしも映像情報に限るものではなく、実際には研究段階ではありますが、聴覚、触覚、平衡感覚、味覚といったあらゆる感覚情報がARのターゲットになります。

聴覚のARでは、立体音響を出力するヘッドフォンや指向性スピーカーなどを使って、仮想の物体の存在感を音によって提示する研究が行われています。

触覚のARとは、現実の物体に対して新たな触感、または異なる触感を付加することを意味します。ある研究では、爪の上に振動を与えることによって、指先の皮膚感覚を提示するという試みが行われています。この技術を応用すれば、タッチディスプレイ

のツルツルの表面に対して仮想的に凹凸を付加することが可能になります。また別の研究では、ARの技法によって、物体表面に新たなテクスチャ画像を重畳表示すると、その視覚効果によって、物体を触ったときの質感が変化するという面白い現象が報告されています。

平衡感覚に関する研究では、耳の辺りに電気刺激を与えて平衡感覚を“乗っ取る”ことにより、人の歩く方向をコントロールするといった実験が行われています。これは加速度を重畳するARと言えます。この技術は視覚障害者の安全な歩行を支援するシステムへの応用が期待されています。

味覚のARとは何かと言われれば、例えば薄味の料理があったときに、本物の調味料を入れずに感覚器に対して何らかの刺激を与えることによって味を濃く感じさせたり、コンピュータ上でデザインされた食べ物を食べているような感覚を実現するといったことが考えられます。現状では、味覚に関するARの研究はほとんど見られませんが、可能性としてはあり得るでしょう。

## ■ センシング技術がカギを握る

ARの実現のためにはどのような基礎技術、要素技術が必要でしょうか。大きく分けるとエレクトロニクス技術、グラフィックス技術、ネットワーク技術の三つです。

エレクトロニクス技術では④センシングが一番重要です。具体的には、カメラや⑤GPS、⑥ジャイロセンサーなどの各種センサーが必要になります(図9.1)。

表示のための映像提示装置も必要です。AR用に

④ センシング=sensing。  
検知、検出の意味。

⑤ GPS = Global Positioning System。全地球測位システム。米軍が打ち上げた24個の人工衛星と地上の制御局、利用者の移動局から構成されるシステム。移動局と三つ以上の衛星との距離を計測することで、移動局自身の平面上の位置を知ることができる。四つ以上の衛星を利用すれば3次元的な位置を測定できる。

⑥ ジャイロセンサー=物体の角速度を検出することで、物体の向いている方向を取得できる。ジャイロスコープとも呼ばれる。



決まった形というのはありませんが、現実空間に映像情報を重ね合わせるためのスタンダードなデバイスとして、④ヘッド・マウント・ディスプレイ(HMD)が挙げられます。いわゆるメガネ型のディスプレイです。次によく用いられる方法が、カメラとディスプレイを搭載したモバイル端末(いわゆる携帯)を使う方法です。虫眼鏡を使うときのように、デバイスを対象にかざして観察します。最近では、プロジェクタを使って、対象に直接映像を投影する方法も数多く提案されています。

グラフィックス技術は3次元コンピュータ・グラフィックス(CG)などの描画技術も重要ですが、2次元や3次元データの作成技術がポイントになります。例えば、商品情報を現実空間に重ね合わせようとしたときに、商品のグラフィックス情報をどうやって用意するのが問題です。

既に⑤CADデータがあるような商品であればそれを流用できますが、職人が作ったような工芸品

④ヘッド・マウント・ディスプレイ＝頭に装着するディスプレイで、ゴーグルやヘルメットのような形状をしている。

⑤CAD＝Computer Aided Design。コンピュータを使い、設計や製図を行うこと。

## ■ 図 9.1 ARの実現に必要な要素技術

### エレクトロニクス技術

- センシング技術：カメラ、GPS、ジャイロセンサー、RFID、光学式モーション・キャプチャなど
- ディスプレイ技術：一般的な液晶テレビ、モバイル端末用小型ディスプレイ、ヘッド・マウント・ディスプレイ、携帯型プロジェクタなど

### グラフィックス技術

- 2次元の注釈情報、3次元CGモデルによるコンテンツの作成(実際の商品のCADデータを再利用/3Dスキャナで作成)
- それを画面に表示するための高速な描画エンジンや物理エンジン

### ネットワーク技術

- HSDPA、無線LAN、WIMAXなど、高速なデータ通信技術

④3次元スキャナ=レーザーなどを用いて多方向から物体をスキャンすることで3次元データを生成する機器。ローランド ディー・ジーなどが販売している。

などCADデータが存在しないものについては、  
④3次元スキャナを使ってデータを取り込むといった工夫が必要になります。ビジネスとして展開する際には、3次元CGのデータをいかに効率よく用意するかが鍵になると思います。

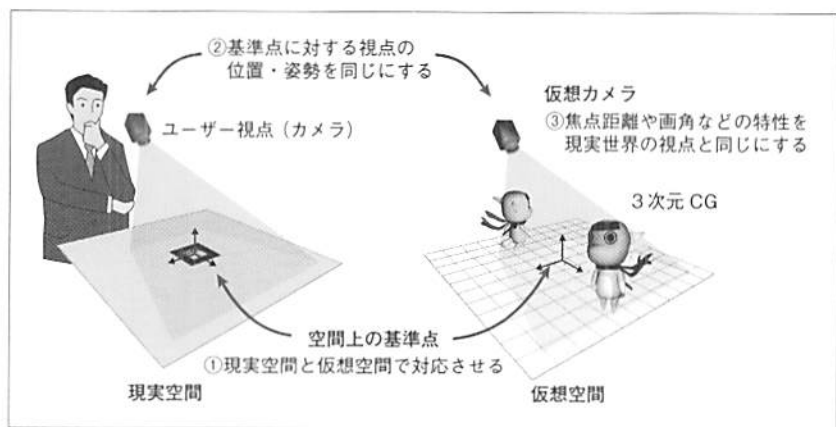
ネットワーク技術では、コンテンツをユーザーへ配信するための高速回線が必要になります。モバイル端末であれば、無線LANや携帯電話のデータ通信サービスなどを使うことになります。

## ■ 現実空間と仮想空間の位置を合わせる

現実の世界に仮想の世界を重ね合わせるというと、言葉としては簡単に聞こえますが、これを実現するためには技術的に三つのポイントがあります(図9.2)。

①現実空間と仮想空間の基準点を対応させる作業。現実空間の絶対座標に対して、「この座標は仮

■ 図 9.2 現実空間と仮想空間の融合に必要なこと



想空間ではここでは」と対応を取ります。簡単な方法の一つは、現実空間の中に④マーカーを置いて現実空間と仮想空間を結び付ける手法です。

②現実空間と仮想空間において、基準点に対するカメラの位置・姿勢を同じにする作業。こうしないと現実空間と仮想空間をぴったりと合わせることができません。

③カメラの特性を設定する作業。カメラにはレンズ歪みや焦点距離、⑤画角といった特性値があります。これらのパラメータは3次元物体の見え方に影響を及ぼします。このため、仮想空間を撮影しているカメラ(仮想カメラ)と現実世界を撮影しているカメラの特性を同じにする必要があります。

①～③を適切に実行すると、現実の世界と仮想の世界の映像をぴったりと違和感なく合わせることが出来ます。

## ■ センサーやカメラを使う

①や②を実現するために多数の方法が提案されていますが、(A)各種センサーを使ってセンシングする方法と、(B)カメラによるビジョン・センシングの方法に大別できます(次ページの図9.3)。カメラもセンサーの一つですが、ARにおいてはユーザー提示用の画像を取得する目的でも使うので、ここでは便宜上区別します。実際には、(A)と(B)を組み合わせることで精度を高めます。

(A)は磁気式3次元位置姿勢センサーやジャイロセンサー、⑥電子コンパス、GPSなどが挙げられます。磁気式3次元位置姿勢センサーはバーチャル・リアリティの研究でよく使われていたもので、部屋の隅などに磁気を発する装置を置いて、その磁

④マーカー=marker。標識。2次元バーコードのような白黒パターンや登録した画像などを使う。

⑤画角=カメラで撮影するときの被写体の写る範囲を、カメラのレンズを中心にした角度で示したもの。

⑥電子コンパス=地磁気から方位を検出するデバイス。地磁気センサーとも呼ばれる。

■ 図 9.3 位置合わせの手法

専用のセンサーを用いる方法	カメラによるビジョン・センシング
<ul style="list-style-type: none"><li>■ 磁気式 3 次元位置姿勢センサー</li><li>■ ジャイロセンサー</li><li>■ 電子コンパス</li><li>■ GPS</li><li>■ RFID</li><li>■ 光学式モーションキャプチャ</li><li>■ 無線 LAN 測位</li></ul>	<ul style="list-style-type: none"><li>■ マーカーを壁・床・対象などに貼付し、それをカメラで認識・計測する手法</li><li>■ 対象物体のテクスチャや形状特徴などのモデル情報を事前に作成しておき、画像とのマッチング &amp; トラッキングを行う手法</li><li>■ Visual SLAM (Simultaneous Localization and Mapping) 画像中の特徴点を抽出し、フレーム間の動きから各点の相対的な 3 次元位置を計測し、連続的にマップを構築する手法</li></ul>

どれか一つということではなく、組み合わせて使うことにより精度を上げる

☞ PlaceEngine = クラウドが運営する無線 LAN アクセス・ポイントを使う現在位置取得サービス。

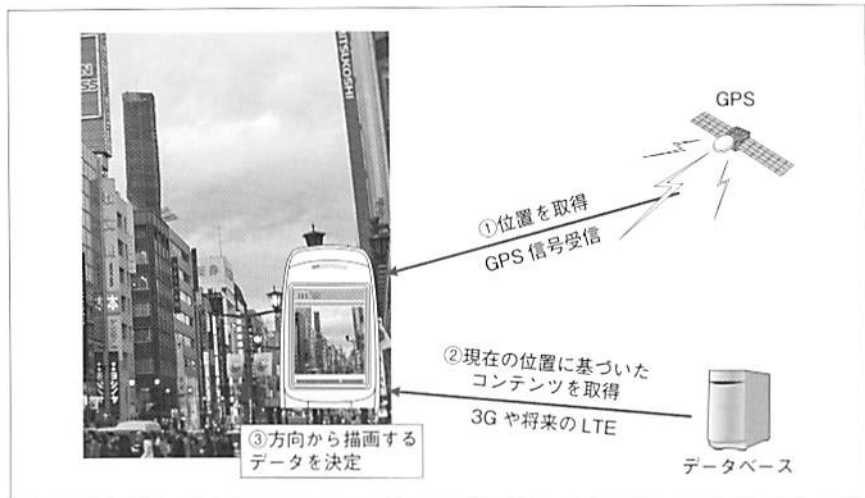
☞ RFID = 無線を使って、タグに埋め込まれた ID 番号を認識する技術。タグには CPU 機能やメモリーを持ったものもある。

気をレーザースキャナで受信し、磁気の計測値から 3 次元空間での位置と姿勢を推定するものです。ジャイロセンサーは 3 次元の角度情報を得るセンサー、電子コンパスは地磁気から方角を取得するセンサーです。これらは現在位置を取得できる GPS や、「☞ PlaceEngine」のような無線 LAN 測位の手法と組み合わせてもよく使われます。

☞ RFID を使う方法もあります。RFID を細かく配置し、そこから現在位置を計測します。またはユーザーに RFID を取り付け、部屋の入り口などに RFID リーダーを置いておけば、ユーザーがどのエリアにいるのかを判別できます。その後、他の技術で細かくセンシングしていきます。

現在話題となっているモバイル端末の AR の多くは、(A)の方法を使っています。最初に GPS や無線 LAN 測位で位置情報を取得します。その後、無線ネットワーク経由でデータベースから、現在の位置情報に基づいたコンテンツをダウンロードしま

■ 図 9.4 モバイル端末を使う AR の一般的な手法



す。次に、電子コンパスで取得したユーザーが向いている方向に合ったコンテンツを携帯電話のライブビュー映像の中に重畳表示する、といった流れになります(図 9.4)。

例えば、「Wikitude」という T-Mobile G1 で動く AR アプリケーションは、現在地の近くにある「Wikipedia」の緯度経度情報を持つ記事を表示します。このような外部のサービスとのマッシュアップが既に可能になっています(次ページの図 9.5)。

## ■ マーカーとマーカーレス

(B) のカメラによるビジョン・センシングは、(I) ビジュアル・マーカーを使う、(II) 対象物体の形状情報を用いる、(III) Visual SLAM と呼ばれる技術を使うという三つのタイプに大別できます

☞ Wikitude = T-Mobile G1 向けに開発された AR アプリケーション。Wikipedia の記事を AR 的に検索できる。開発はオーストリアのモビリジ社。

☞ T-Mobile G1 = 台湾 HTC が開発した初代 Android 搭載の携帯電話端末。6軸の加速度センサー / 電子コンパスを内蔵している。

☞ Wikipedia = オンラインのユーザー記述型百科事典。

☞ マッシュアップ = 外部にある他のサービスなどを組み込んで新しいサービスやアプリケーションを作る手法。

## 図 9.5 実稼働している携帯電話の AR アプリケーション

- [Enkin]
- [Wikitude] (Wikipedia 記事検索)
- [ING Wegwijizer] (オランダの ING 銀行の ATM 検索)
- [セカイカメラ]

コンテンツはネット  
にある既存のものも  
利用可能

緯度・経度情報を持つ  
コンテンツは多数ある。  
(例) Wikipedia, Flickr

☞ ARToolKit = AR アプリケーションを開発するためのオープンソースの C 言語ライブラリ。奈良先端科学技術大学院大学の加藤博一教授が開発した。

☞ QRコード = デンソー (現デンソーウェーブ) が開発した2次元コード。バーコードと異なり2方向に情報を持つので、表示面積が小さくても情報量を多くできる。ISO/IEC 18004、JIS-X-0510として規格化されている。URLが格納されたQRコードを携帯電話のカメラで読み込んで、Webサイトに誘導するなど、既に実用化が進んでいる。

☞ 再帰性反射材 = 入射した光を光源の方向に反射する特殊な素材。交通標識などで使われている。

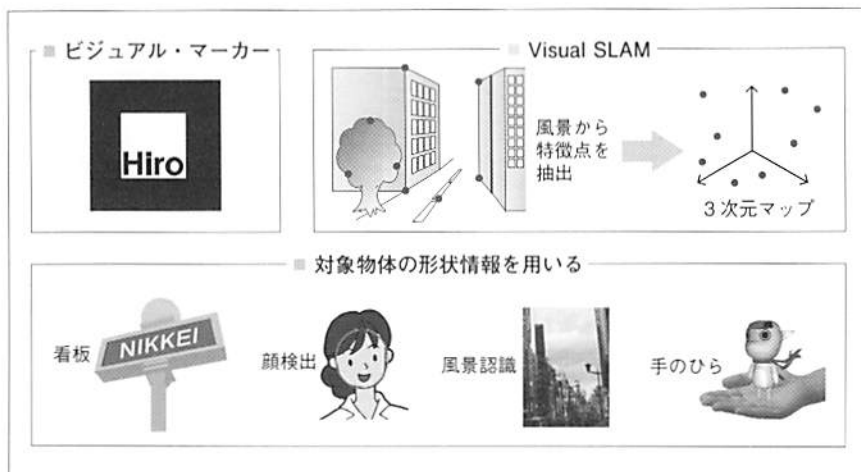
(図 9.6)。

(I) ビジュアル・マーカを使う。これは、最近人気のある「☞ ARToolKit」が採用している手法です。ビジュアル・マーカ (以下、マーカ) には幾何学的な特徴 (例えば、正方形という制約) があり、これを利用して、マーカの3次元位置・姿勢を計測します。また、マーカには☞ QRコードのような識別用のパターンが描かれており、これにより、マーカと提示したい情報を対応付けることができます。

例えば、マーカを壁に貼り付けておけば、それを撮影することによってナビゲーション情報などを空間に重畳表示することができるようになります。マーカを使う方法は手軽さで長けていますが、空間中のあちこちにマーカを貼り付けると、目障りになってしまうという問題もあります。

そこで最近の研究では、マーカを透明にして見えないようにする工夫が提案されています。例えば、透明な☞再帰性反射材という特殊な素材でマーカを作り、壁に貼り付けると、人間の目からは透明なシールが貼ってあるようにしか見えないの

■ 図 9.6 カメラによるビジョン・センシングの種類



ですが、赤外線カメラで撮影すると、カメラからは認識できるようになります。

(Ⅱ) 対象物体の形状情報を用いる。例えばTシャツの模様や看板などのテクスチャ情報や、物体の3次元形状特徴を使います。対象物体に関するそれらのデータを事前取得しておき、カメラ画像に対してモデルマッチングを行うことにより、対象物体の位置・姿勢を計測します。この方法では、情報を提示する対象の物体に対してマーカーのようなセンシング用の特別なしかけをほどこす必要がないというメリットがあります。

この手法を用いたある研究では、プリンタをカメラで撮影すると、プリンタ上に使い方の説明が重畳表示されるということが行われていました。例えば、「ここが電源ボタンですよ」、「メンテナンスのときはここを開けてください」といった説明です。家電のように製品形状があらかじめわかっているも

のに対しては、このような手法が有効でしょう。

形状情報として、手のひらを用いる研究もあります。Taehee Lee、Tobias Höllerer 氏らが提案した「Handy AR」では、手を開いたときの指の位置関係を利用して、手のひらの3次元姿勢を計測します。手をカメラで撮影すると、手のひらの上に3次元CGを表示することができます。さらに、人間の顔を用いる方法も考えられます。最近では、特定の風景から特徴点を抽出して、風景全体をモデル化する研究も進んでいます。

(Ⅲ) Visual SLAM (Simultaneous Localization and Mapping) と呼ばれる技術を使う。Visual SLAM はロボット研究では古くから取り組まれている方法で、画像中の特徴点を利用します。例えば物体の角やエッジなどに基づいた特徴点を画像中から抽出し、得られた複数の特徴点をフレーム間でトラッキングします。そしてフレーム間での特徴点の位置情報から3次元構造を計算し、連続的に特徴点の3次元マップを構成していきます。作成された3次元マップによって、空間中に3次元CGや注釈情報を配置することができるようになります。

## ■ ブレークスルーだった PTAM

Visual SLAM に関しては、最近画期的なブレークスルーがありました。2007年にISMAR という世界的に権威のある拡張現実の国際会議で最優秀論文賞に選ばれた「Parallel Tracking and Mapping for Small AR Workspaces」という研究です。Visual SLAM の一種で、略して PTAM と呼ばれています。

PTAM の優れたところは、特徴点の抽出とマッピングの作業をスレッド化することで非常に高

☞ ISMAR = International Symposium on Mixed and Augmented Reality.

☞ スレッド = ソフトウェアで並列的な処理を行うための仕組み。



速な処理を実現した点です。PTAMはオックスフォード大学の研究者が特許を申請していますが、ソースコードが2008年に公開されたので、誰でも利用できるようになっています。☞YouTubeなどの動画共有サイトには、PTAMのソースコードを使って作られたさまざまなアプリケーションのデモ動画が投稿されています。

☞YouTube=米グーグルが運営する動画投稿サイト。

## ■ 技術的には成熟しているマーカー型

ここからはマーカー型ARに関して説明します。マーカー型ARは、研究レベルではほぼ成熟した技術で、既に実用化もされています。

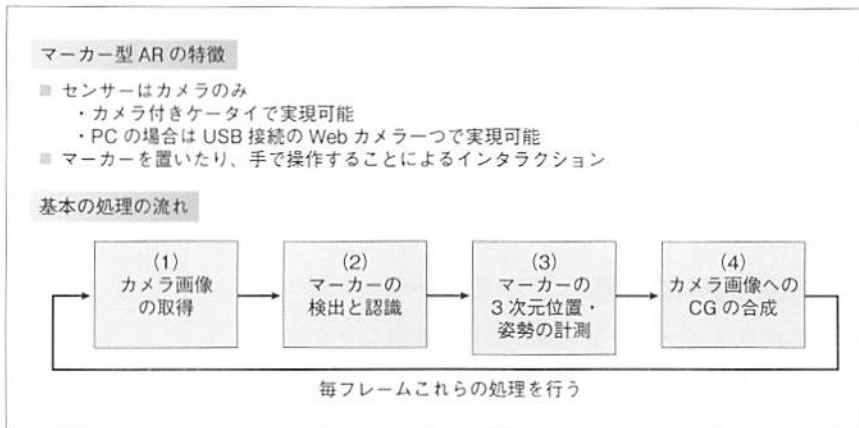
マーカー型の良い点は、紙のマーカーとカメラだけで実現できることです。携帯電話内蔵のカメラでも、パソコン用の安価なUSBカメラでも、ARアプリケーションを動かせる手軽さがあります。

基本的な処理の流れは、(1)最初にカメラ画像を取得、(2)画像処理によってマーカーを検出し、マーカーに描かれているパターンを認識、(3)画像中のマーカーの位置情報から射影幾何学的な計算を行い、3次元空間でのマーカーの位置と姿勢を計測、(4)計測した位置と姿勢情報をもとに3次元CGなどを合成、となります(次ページの図9.7)。

この一連の処理を1フレーム中に1回行います。30フレーム/秒の実行速度を出すためには、33ミリ秒以内に(1)から(4)を完了する必要があります。以前は画像認識の計算コストが非常に大きかったのですが、最近ではコンピュータの進歩によって、家庭用パソコンでもリアルタイムで動くレベルにまで到達しています。

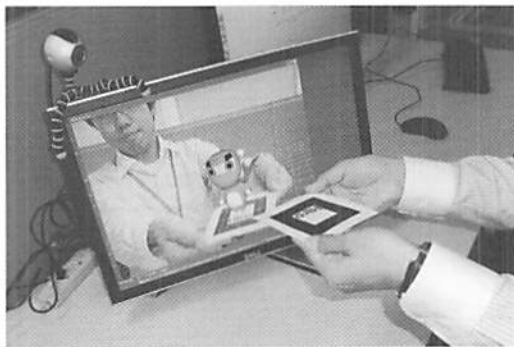
(2)の画像認識では次のようなことを行います。

## 図 9.7 マーカー型 AR の仕組み



## 写真 9.1 ARToolKit のデモ

マーカーを認識して、CG を合成している。



画像認識の検出アルゴリズムは、最初に画面全体をモノクロ画像に変換し、黒い太枠の存在する領域を探します。そして、ある程度の大きさの領域を抽出し、その後で輪郭線を抽出します。ここで角が4個ある領域があれば、四角形の領域だと判断します。そして、中に描かれているパターン領域を一定の間