



開講にあたって



#E062A1



編集理事 甲 藤 二 郎

近年、拡張現実感 (Augmented Reality: AR) 技術の普及と発展が著しい。その背景としては、入出力デバイス、CPU、各種センサなどのハードウェアと画像処理アルゴリズムが相互に急速に発展したこと、また、現実世界に仮想世界を重ねることが、肉眼だけでは得られない新奇かつ有用な数々のアプリケーションを提示したことが挙げられる。

昨今のスマートフォンは、カメラに加え、デフォルトの機能として位置センサや姿勢センサを備え、従来のパソコン、ノートパソコンを優に凌ぐ高機能情報端末となっている。また、入出力デバイスそのものも、カメラに位置センサや奥行きセンサを統合したり、カメラとディスプレイを一体化したり、視覚・聴覚以外の五感を取り込んだりと、さまざまな進化を遂げている。また、画像処理アルゴリズムの進歩も著しく、実画像内のマーカを手掛かりに、カメラ姿勢や3次元空間を復元したり、実物体そのものを認識してマーカにするアプローチも報告されている (今月号《講座》第1回、ならびに、今月号《技術解説》も参考のこと)。

また、これらの基盤技術の発展を受け、拡張現実感の多

種多様な応用開発も進められている。放送、Web広告やゲームにおける特殊効果はもちろんのこと、スマートフォンをプラットフォームとする各種応用なども、本誌でもしばしば取り上げてきた (2008年1月号《私の研究開発ツール》、2011年5月号《知っておきたいキーワード》、など)。また、防災や医療など、われわれの日々の生活の安心・安全を高めるための試みも進められている。その一方で、提供される高臨場感が人間に与える影響を評価しておくことも重要である。

本講座では、以上の拡張現実感技術に関わるトピックを全12回にわたって取り上げる。最近、会誌へのアンケート結果として、特に若手の読者からの拡張現実感に関するリクエストが多く見られた。そこで本講座では、主に若手ならびに専門外の読者を対象に、拡張現実感に関する基盤技術から応用技術までを包含し、著名な著者陣による系統立った連載を試みる。

本講座の企画は、山下幸彦、太田 淳、大田原一成、近藤敏志、高取憲一各編集幹事と私 甲藤が担当いたしました。

回 題 目 (仮題)

- 1 総 論
- 2 視覚障害者のインターネット操作の支援
- 3 3次元距離計測カメラ
- 4 Inside-outカメラ
- 5 触覚・力覚ディスプレイ
- 6 嗅覚ディスプレイの仕組みと応用
- 7 臨場感と迫真性の心理分析
- 8 音響の高臨場感技術
- 9 Kinect
- 10 手のひらAR
- 11 防災AR
- 12 拡張現実感技術の将来展望

執 筆 者

- 加藤博一 (奈良先端大)
熊澤逸夫・小泉 亮・V-ris Jaijongrak (東工大)
池田 誠 (東 大)
藤吉弘亘 (中部大)
梶本裕之 (電通大)
中本高道 (東工大)
行場次朗 (東北大)
安藤彰男 (NHK)
西林 孝 (VOYAGE GROUP)
加藤晴久 (KDDI研)
細川直史 (消防庁)
暦本純一 (東 大)



1. ま え が き

近年、拡張現実感 (Augmented Reality: AR) に対する注目度が高まってきた。Googleトレンドでは、2009年にAugmented Realityの検索数がVirtual Realityの検索数を超えたという結果が表れている。また、スマートフォンの普及とともに、拡張現実感を取り入れたアプリケーションソフトウェアも増加し、IT系商用誌やウェブサイトでも、それらの解説記事が取り上げられるようになった。本稿では、今や一般用語となりつつある拡張現実感について、技術的な観点から解説する。

拡張現実感とは、コンピュータが生成した仮想物体や情報を現実世界の中に表示する機能として定義される。この機能を、現実世界を補強するもの、拡張するものと捉えることで、拡張現実感という言葉が使われている。ただし、重要なのは、この概念がコンピュータによる効果的な作業支援の手段として考案されたという点にある¹⁾。

現在、コンピュータは人間のあらゆる活動に利用され、さまざまな様式で人間の活動を支援している。通常のPCなどは、コンピュータ内で閉じた活動、例えば、電子メールによるコミュニケーションや文書作成などで効果的に使用され、それをより便利にかつ効率的にするために、GUIなどのユーザインタフェース技術が発達してきた。一方で、人間が現実世界で体を動かして行う活動も非常に多く存在する。そのような作業を支援するには、一般的なPCはあまりふさわしくない。通常、コンピュータ内に蓄積されている有益な支援情報はディスプレイ装置に提示される。このとき、人間がその情報を活用するには、ディスプレイ装置に提示された情報を現実世界の問題に照らし合わせて理解する必要がある。例えば、数多くのスイッチの中からONにするべきスイッチの情報がディスプレイ装置に表示



図1 拡張現実感技術を用いた作業支援の例

されていると、人間は、その情報を現実の操作パネル上のスイッチと照らし合わせて、実際にONにすべきスイッチを確認しなければならない。つまり、これは情報の提示されている場所と操作を行う場所が異なっていることを意味し、その対応付けの作業が常に発生する。この対応付け作業をわかりやすくするために、文字による表示ではなくグラフィックスを利用するといった工夫が行われるが、この対応付け作業のミスをなくすることは容易ではない。もちろんほとんどの場合、作業員、特に熟練者においてはこの対応付けを容易に行うことができると思われる。しかし、複雑な機械の組み立て、危険の伴う機械の操作、手術などの支援を考えたとき、この対応付けのミスをできる限り0に近づける努力は必要である。

この対応付けのミスをなくすための最も確実な方法は、対応付けの必要のない情報支援方法を採用することである。つまり、ディスプレイ装置に情報を表示するのではなく、現実世界の作業対象に直接支援情報を貼り付けることである。拡張現実感はこのようなアイデアを実現する技術として提案され、研究開発されてきた。

図1に、拡張現実感技術を用いた作業支援システムのプ

† 奈良先端科学技術大学院大学

"Technology Frontier of Augmented Reality (1): Introduction to Augmented Reality" by Hirokazu Kato (Nara Institute of Science and Technology, Nara)



ロトタイプにおけるユーザ視野の例を示す。このシステムは、ユーザがおもちゃのブロックを組み立てる作業を支援する。写真では、ユーザに対して次の作業がピンを手前の穴に刺すことであるということをビジュアルに示している。拡張現実感によって非常にわかりやすい作業指示が可能になることが容易にわかるであろう。

このような作業支援の他にも、拡張現実感技術の発展に伴い、さまざまな応用が検討されるようになった。例えば、ゲームのようなエンタテインメント分野においては、拡張現実感は革新的な映像表現技術として利用することができる。広告分野においては、デジタルサイネージへの応用や紙媒体の3次元コンテンツによる拡張などに利用できるであろう。特に、位置情報に基づく情報サービスとの連携は、拡張現実感の長所が活かせる応用である。

以下では、まず、拡張現実感を実現するシステムの構成について説明し、次に、そこで利用されるコア技術について説明する。さらに最近の技術動向と将来展望について述べる。

2. 拡張現実感システムの構成

情報をディスプレイではなく、現実世界の中(のありとあらゆるところ)に提示するという機能の実現は容易ではない。現状においては、HMD (Head Mounted Display) を利用する方式が主に検討されている。HMDは小型LCDのようなディスプレイ装置を頭部に搭載し、ユーザの前方に置かれた大型ディスプレイの映像と同じ効果を作り出す装置である。また、両目に視差のある映像を提示することで、ステレオ視の原理に基づき立体感のある映像をユーザに提示することもできる。HMDを用いても、情報はあくまでディスプレイ装置上に提示されているという点では従来の情報提示と変わりはないが、以下に説明する方式を用いることにより、ユーザにとっては情報が現実世界の中に提示されていると感じられる情報提示を実現できる。

HMDを用いる拡張現実感システムの構成を図2に示す。この図に示すように、二つのタイプが存在する。一つは、光学式シースルー方式というもので、小型のディスプレイに表示された映像と現実世界からの光をハーフミラーやプリズムなどの光学素子によって合成し、ユーザの目に提示する方式である。もう一つは、ビデオシースルー方式というもので、この方式ではHMDに小型のカメラを装着する。小型のカメラで撮影された映像は一旦コンピュータに入力され、そこでCG映像が合成される。そして、合成後の映像をHMDのディスプレイに提示する。光学式はハードウェア構成が簡単になるが、CG映像が透過するなどの問題がある。ビデオ式はハードウェア構成が複雑になり、実世界映像の画質にも問題があるが、合成処理の自由度が高いという長所もある。

また、HMDを使用せずに、標準的なディスプレイ装置

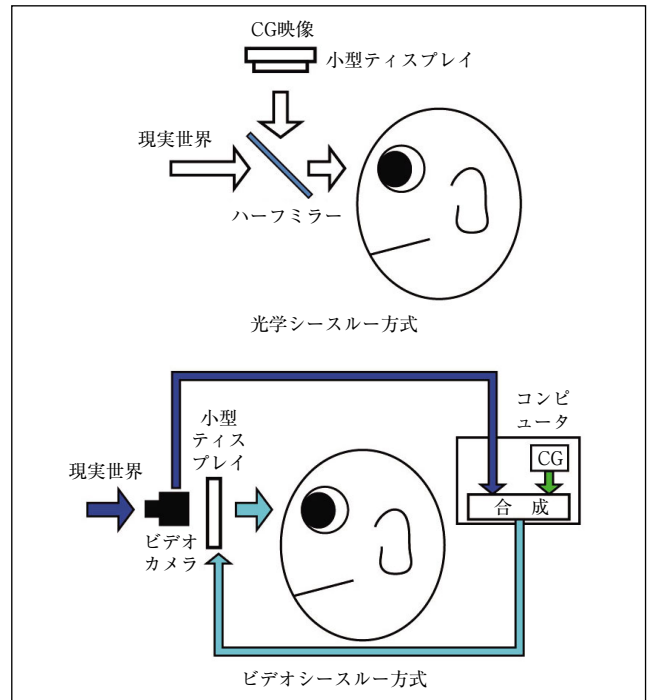


図2 システム構成

で代用する方式もあり、モニタ方式と呼ばれる。デスクトップディスプレイ装置や大型のディスプレイ装置にカメラを取り付け、ディスプレイを見るユーザ自身を撮影するようにする。映像の左右を反転させることで、ちょうど鏡を模倣し、その中にCG映像を合成する方式である。また、スマートフォンやタブレット端末などのカメラ付きモバイルディスプレイを使用する方式もある。カメラを現実世界に向けてかざし、撮影された映像中に、ビデオシースルー方式と同様に情報を合成し、それを端末のディスプレイに表示する方式である。このような方式では、HMDを用いる方式とは異なり、ユーザはディスプレイ装置を見ていることを意識するので、必ずしも本来の拡張現実感の定義である情報を現実世界の中に表示するという機能を実現しているとはいえない。そのため、この方式では、本来の拡張現実感の長所が活かされているのかという点で疑問が残る。ただし、最近のスマートフォンの爆発的な普及の中で、このタイプの実用アプリケーションやサービスが登場するようになり、このようなタイプこそが拡張現実感であるといった拡張現実感の新しい定義も普通に受け入れられているのが現状である。

3. コア技術

拡張現実感システムを実現するためには、いくつかの要素技術を組合せなければならない。現実世界とCGで構成される仮想世界を高品質に合成するために、次の三つの整合性を維持するためのレジストレーション技術が必要とされる。



図3 ARToolKitによるCG合成の例

- (1) 幾何学的整合性
- (2) 光学的整合性
- (3) 時間的整合性

幾何学的整合性とは、現実世界の中に仮想物体を3次元に適切な位置姿勢で表示することである。これは位置合わせ技術とも呼ばれている。これまで拡張現実感研究においては、最も多くの努力がこの幾何学的整合性の研究に費やされてきた。光学的整合性とは、仮想物体の陰影を現実世界のそれとあわせたり、仮想物体の影を適切に表示することである。作業支援という拡張現実感の本来の目的を考えると光学的整合性は必要とされない場合もあるが、エンタテインメント応用などでは非常に重要となる。時間的整合性は、CG合成における計算処理時間に起因する時間遅れをなくしたり、十分なフレームレートを確保することである。コンピュータの性能向上がこの問題をある程度は解決するがシステム全体の遅延時間を削減するための技術や、計算性能の弱いモバイル端末では依然として問題となっている。以下では、幾何学的整合性と光学的整合性について説明する。

3.1 幾何学的整合性

幾何学的整合性は、3次元CG映像描画における幾何学計算に用いられるパラメータを現実世界に定義された座標系や表示デバイスの幾何学的特性に基づき設定することで実現できる。

具体的には、CG描画における世界座標系からカメラ座標系への変換であるビューイング変換パラメータを現実世界の中での表示デバイスの位置と向きに応じて設定する必要がある。このパラメータは位置を表す3自由度のパラメータと、姿勢を表す3自由度のパラメータからなる。

例えば、GPSによって位置情報を取得し、電子コンパスとジャイロセンサによって姿勢をリアルタイムに計測できれば、ビューイング変換パラメータは設定できる。ほとんどのスマートフォンはGPSや慣性センサを内蔵しているので、それらのデバイスでは、これらの情報を容易に取得でき、システム開発が容易である。しかし、GPSは屋内では利用できなかったり、高層ビルが建ち並ぶ都心部では精度が悪くなるという問題を抱えている。慣性センサも方位角

においてドリフト誤差が発生しやすいという特性がある。

また、CG描画における仮想カメラの特性となる投影変換パラメータも表示デバイスなどの特性に応じて設定する必要がある。このパラメータは基本的にシステム運用時に逐次変化するものではなく事前に設定可能であるが、オフラインでの高精度なキャリブレーションが要求される。

ところで、ビデオシースルー方式やモニタ方式においては、ビデオカメラが使用されるので、それを用いて位置合わせを行うことができれば、他のセンサが不要となり、システム構築の面で非常に有効である。理論的には、カメラによる位置合わせとは、現実世界内で定義された世界座標系とカメラデバイス上に定義されたカメラ座標系との間の3次元関係を計測するカメラの位置姿勢推定に他ならない。伝統的には、現実環境にマーカを設置し、それをカメラで認識することでカメラの位置姿勢を推定する手法が使われてきた。近年では、マーカを使用せずに同様の機能を実現するマーカレス位置合わせ技術も実用レベルに到達してきた。

マーカを用いた位置合わせ技術の例としては、ARToolKitをあげることができる²⁾。ここでは、明確な輪郭線を持つ正方形をマーカとして利用している。正方形は画像内では一般的な四角形となり、その抽出は非常に容易である。さらに、四つの頂点位置を推定計算に利用することで、一つの正方形マーカの観測で位置姿勢推定が実現できる。ただし、正方形は90°毎の回転に対して対称性があるので、マーカの姿勢が一意に定まらない。そこで、正方形の対称性を解消する目的で正方形内部や周辺に2次元コード、文字、パターンなどを描くことが多い。これは、また、複数のマーカの識別にも利用される。図3に、正方形マーカを用いて3次元CGを合成表示した画像を示す。

マーカ方式では、環境中にマーカを配置する必要があるが、それが適さない応用も多い。そこで、マーカを使用せずに、環境中にもともとある特徴を利用して位置姿勢推定を行うマーカレス方式が必要となる。この技術は近年盛んに研究が行われ、モバイルデバイスでもリアルタイムに動作可能な手法も登場している。詳細に関しては、本誌「技術解説」に、その解説記事が掲載されているので、それを



参照いただきたい³⁾。

3.2 光学的整合性

光学的整合性に関する技術は、いくつかの種類に分類できる。最も基本になるのは、3次元CGの陰影特性を現実世界のそれと一致させることである。最近のリアルタイムCG技術は高品質な陰影を実現可能であり、問題は現実世界の照明環境をどのようにセンシングするかである。これをリアルタイムに実現するのは非常に難しいが、鏡面球を利用することで簡便に実現する手法も提案されている⁴⁾。また、ビデオスルー方式では、使用するカメラのシャッタースピードや焦点ずれによってボケが発生する場合があります。それをCG映像に付加する方法も提案されている⁵⁾。

最近では、現実世界にある物体によって隠されているものの可視化に関する研究も盛んに行われている。例えば、CTスキャナによって取得された人体内部の臓器をまるで人体が透けて見えているかのように表示したり、地中に埋設されている土管を透視したりする応用である。このような場合、例え幾何学的に正しくCGを描画しても、何も工夫をしなければ実物体が透けて見えるような印象を与えない。そこで、実写映像を加工し、透けて見えているかのような効果を与える技術も提案されている⁶⁾。さらには、現実世界の物体を完全に消し去る隠消現実感 (Diminished Reality; DR) と呼ばれる技術も登場するようになった⁷⁾。

4. 最近の動向と将来展望

前述のように携帯端末を用いた拡張現実感、すでに実用ソフトやサービスとして登場している。また、ウェブページに拡張現実感技術を組込むといった取り組みも行われている。このようなハードウェアの障壁の低い条件での応用が盛んに試みられるようになってきた。技術としてもマーカを使用した位置合わせからマーカレス方式への移行が始まっている。マーカレス方式においては、さまざまな手法がこれまでに提案されてきた一方で、どの手法がどのような特性を持つのかといったことが非常にわかりにくく、システム開発者にとっては、それぞれの問題設定においてどの手法に優位性があるのかを使用前に確認することができないという問題がある。そこで、このようなマーカレス方式の手法の特性比較を行うことを目的としたベンチマークテストの開発の動きが出てきている⁸⁾。

また、拡張現実感技術の普及のためには、コンテンツの流通が重要であり、そのためにはコンテンツの互換性を高める必要があり、システムやコンテンツの標準化が重要に

なると考えられている。そこで、拡張現実感技術に関連した標準化の議論がヨーロッパを中心に行われるようになった。今後、このような学術ではなくビジネス指向の議論はますます増えていくと思われる。

さて、拡張現実感の将来を考える際、前述の通り携帯端末などで実現されるモニタ方式の拡張現実感というのは、必ずしも拡張現実感の本質的な機能を実現しているわけではないということに注意しておく必要がある。真の拡張現実感の実現には高性能かつ社会的受用性の高いHMDの開発が重要であり、それが実現できることで拡張現実感の応用は目覚ましく広がると考えている。

一方で、拡張現実感を科学的な視点から捉えると、ほとんどの研究が人間の視覚を対象としているが、その他の感覚に対する情報の付加を考えることもできる。例えば、音の拡張現実感であったり、触覚の拡張現実感である。これらの課題に対する取り組みも盛んになってきており、将来的に人間の五感すべてに対しての究極の拡張現実感が実現されることも夢ではないかもしれない。(2011年11月28日受付)

〔文 献〕

- 1) D.W. Mizell: "Virtual Reality and Augmented Reality for Aircraft Design and Manufacturing", Proc. of Int. Conf. On Artificial Reality and Telexistence (ICAT95), pp.13-17 (2005)
- 2) 加藤博一: "ARToolKit", 映情学誌, 62, 1, pp.48-51 (2008)
- 3) 亀田能成: "マーカレスAR", 映情学誌, 66, 1, pp.45-51 (2012)
- 4) 神原誠之, 横矢直和: "光源環境の実時間推定による光学的整合性を考慮したビジョンベース拡張現実感", 信学技報, PRMU2002-190 (2003)
- 5) 奥村文洋, 神原誠之, 横矢直和: "焦点外れによるぼけとモーションブラーの推定に基づく拡張現実感における光学的整合性の実現", 信学誌D, J90-D, 8, pp.2126-2136 (2007)
- 6) S. Zollmann, D. Kalkofen, E. Mendez, G. ReitMayr: "Image-based Ghostings for Single Layer Occlusions in Augmented Reality", S&T Proc. of Int. Conf. On Mixed and Augmented Reality (ISMAR2010), pp.19-26 (2010)
- 7) 森尚平, 一刈良介, 柴田史久, 木村朝子, 田村秀行: "隠消現実感の技術的枠組と諸問題～現実世界に実在する物体を視覚的に隠蔽・消去・透視する技術について～", 日本バーチャルリアリティ学論誌, 16, 2, pp.239-250 (2011)
- 8) <http://trakmark.net/>



加藤 博一 ひろかず 1986年、大阪大学基礎工学部卒業。1988年、同大学院修士課程修了。1989年、同大学基礎工学部助手。1996年、講師。1999年、広島市立大学情報科学部助教授。2003年、大阪大学大学院基礎工学研究科助教授。2007年より、奈良先端科学技術大学院大学情報科学研究科教授。拡張現実感、ヒューマンインタフェースの研究に従事。博士 (工学)。



触覚・聴覚フィードバックを通じて、 視覚障害者の感覚を拡張する コンピュータ操作支援システム



小 泉◀◀ 亮†, V-ris Jaijongrak†, 正会員 熊 澤 逸 夫††

1. ま え が き

コンピュータの利用は私たちの日常生活の重要な一部分になり、コンピュータを所有している家庭はここ10年間で世界的に急増している。その潮流は視覚に障害を持った人々についても当てはまり、彼らのおよそ50%にも当たる人々がコンピュータを使用している。統計的に視覚障害者の数は、先進国における高齢化と発展途上国における糖尿病等の生活習慣病の進行とともに著しく増えている。社会の情報化は世界的に進展し、視覚障害者が本当の意味で自立した生活を送るためには、コンピュータを通じた情報の入手が不可欠となっている。肉体的活動を支援する介助者の数を単純に増やすだけでは対処できない問題がそこに存在するのである。

視覚障害者にとっては、利用するためには視覚に多くを依存する従来型の入出力デバイスは不都合である。視覚の不足を補って行動するため、他の感覚モダリティをユーザインタフェースに用いて、視覚障害者のコンピュータ操作を支援する研究が数多く存在するが、グラフィカルユーザインタフェース (GUI) が視覚に与える情報に比べ、獲得できる情報量が著しく不足している。

本稿の狙いは、触覚と聴覚に与えるフィードバック情報を工夫することにより、GUIが視覚モダリティに与える情報を、触覚と聴覚のモダリティを通じて把握できるように、視覚に障害を抱えた人々の感覚を拡張し、コンピュータの操作を効率的に手助けするようなユーザインタフェースのシステムを構築することである。

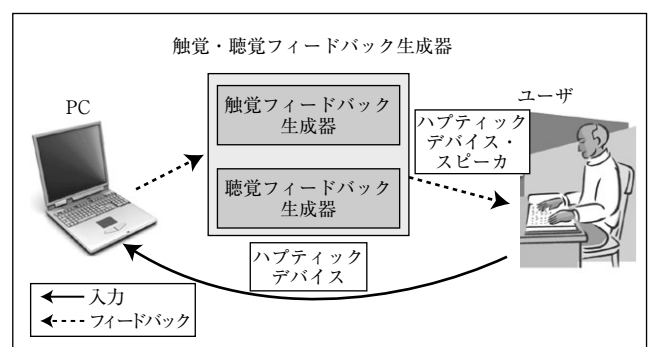


図1 提案する触覚・聴覚フィードバックシステムの構成図

2. システム構成

一般的に、コンピュータとユーザの仲立ちをするインタフェースは、アプリケーションの要求や、ユーザによってさまざまな様相を呈する。標準的なマシンインタフェースは、目の見える人々にとっての必要に答えているといえるが、目の見えない人の要求に答えているかどうかは疑わしい。筆者らは目の見えない人々を支援するために、ユーザに視覚以外のフィードバックを与え、彼らのコンピュータ操作を支援するシステムを提案する。その概要を図1に示す。

本稿の主な焦点は、触覚・聴覚フィードバックを生成する部分において、2種類の感覚に対するインタフェースのさまざまなタイプ、それらの良い組合せを探すことである。各タイプの細かい実装については次章で述べる。

3. 触覚・聴覚フィードバックの生成

コンピュータの入出力、そしてアプリケーション等の視覚的な状況を目の見えないユーザに感知させるためには、それらの情報がユーザにとってはっきりと理解しやすい形に翻訳される必要がある。以前の研究⁶⁾では、一つのホイールを持ち、カーソルがターゲットに向かう動きをそのホイールがトラッキングしつつ、抵抗を通じて触覚フィードバックとして返すマウスを作製した。この研究では、それ

† 東京工業大学 大学院物理情報システム専攻

†† 東京工業大学 像情報工学研究所

"Technology Frontier of Augmented Reality (2): Haptic and Auditory Assistive User Interface: Helping the Blinds on their Computer Operations" by Ryo Koizumi and V-ris Jaijongrak (Department of Information Processing, Tokyo Institute of Technology, Yokohama) and Itsuo Kumazawa (Imaging Science and Engineering Laboratory Tokyo Institute of Technology, Yokohama)

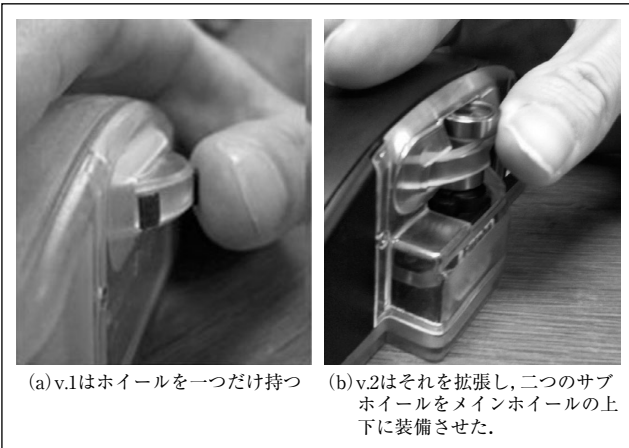


図2 ハプティックマウス

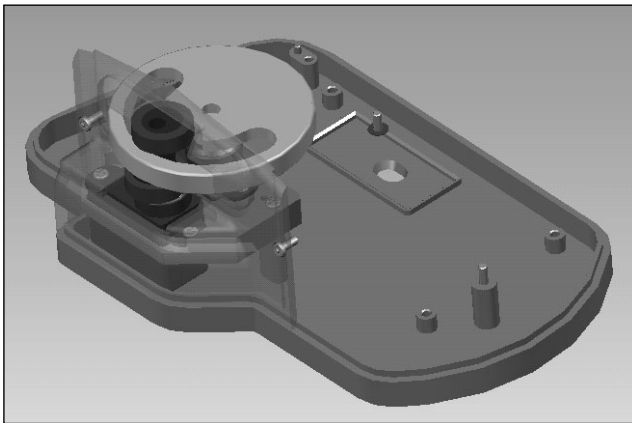


図3 ハプティックマウスv.2の内部構造

をさらに拡張したハプティックマウスと、聴覚フィードバックの情報を使い、目の見えないユーザがGUIを操作するために、マウスカーソルをターゲットへ移動させる課題の解決を試みる。

3.1節では、触覚フィードバックを生成する機構とその制御方式について簡単な説明を行う。3.2節では、聴覚フィードバックが本稿においてどのように使われるかを述べる。

3.1 ハプティックマウス

提案するシステムのための入力デバイスの概要を図2、図3に示す。これは、今回開発した触覚フィードバックを与えるために、モータによって回転するホイールを装備したハプティック(触覚)マウスと呼ぶものである。ハプティックマウスの以前のバージョン(図2(a))においては、マウスに取り付けた一つのホイールの回転を指で触り、その抵抗により、ユーザはポインタとターゲットとの1次元的な距離情報を感知できる。

この実装における一番の問題は、一つのホイールのフィードバックによって表現できる情報は1方向の軸のみであることである。この問題を打開するために、今回の実装で

は、より多くの情報を親指で感知できるようにするために、小さいホイールのペアをメインホイールの上下に取り付けた。ユーザはこれにより、メインホイールと合わせて2チャンネルの情報を親指から得ることができる。

サブホイールはメインホイールと違いベアリングを使用しており、回転ではなくベアリング自体が前後に動きフィードバックを与える構造になっている。メインホイールに比べて可動範囲が狭いものの、このアプローチは、ユーザがサブホイールのフィードバックをメインホイールから与えられるフィードバックと明確に区別するために必要である。

この触覚フィードバックは、マウスのセンサとアプリケーションの状態によって検知されるマウスカーソルの位置または動きに基づいて生成される。すなわち、画面上のターゲットをポイントするために、マウスカーソルがどのように移動しているかがフィードバックで感知できるようになっている。モータを駆動する信号は、PCを通じてマウスに内蔵されたコントローラによって生成される(図4)。触覚フィードバックを返すホイールの回転角は、現在のカーソルと目標との距離によって定義される。

触覚マウスのコントローラはマウスカーソルの現在の状態を受け取り、そしてX軸とY軸それぞれについての表現される回転角を再計算する。距離の変化を見て、マイナス方向であれば回転角を減少させ、プラス方向ならば増加させる。

ハプティックマウスv.2(図2(b)、図3)は三つのホイールを使い、2次元の軸を表現することを提案している。それらの触覚フィードバックの実装は、現実のアプリケーションに合わせて変化させる。

3.2 聴覚フィードバック

聴覚インタフェースによるコンピュータ操作の支援システムは、すでに広く使われている。例えば、スクリーンリーダ(アプリケーション)は、視覚障害者を補助するツールとしては一般的である。このようなスクリーンリーダの欠点は、ユーザはメッセージが読まれるまで操作を待たなくてはならないというリアルタイム性の不足にある。

本稿では、この操作が遅くなってしまうという問題を考えるため、聴覚フィードバックの形態を、情報量が多い「スピーチ(音声)」と即時性が高い「シグナル(信号)」という振る舞いが違う二つのタイプに分け、それぞれ独立に実験・評価を行い、結果について考察することにする。

本稿では、聴覚インタフェースはそれ単体で使われるか、触覚マウスの操作をサポートする役割で使われることになる。

(1)スピーチ(音声)による聴覚フィードバック(オーディオタイプ1)

このタイプの聴覚フィードバックは、アプリケーションがコマンドを受け取ったときに起動され、音声を読み上げられるもので、読み上げられる時間に起因する操作時間の遅延を引き起こす。音声指示の内容は、現在のカーソル位

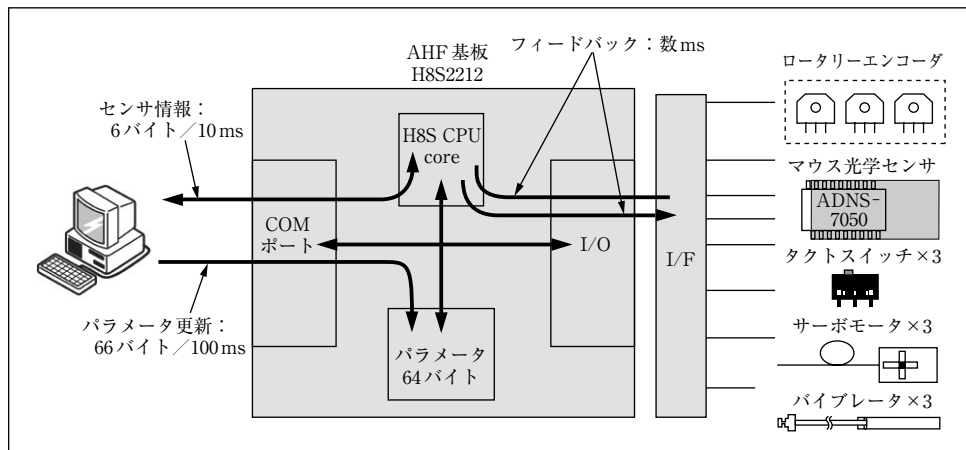


図4 ハプティックマウスとPCとの情報のやりとりの模式図

置からターゲットに向かう方向や、アプリケーションの現在の状況などである。音声の発音にはMicrosoft Speech Application Program Interface (MSAPI) を利用する。

(2) シグナル(信号)による聴覚フィードバック(オーディオタイプ2)

このタイプの聴覚フィードバックは、アプリケーションの現在の状態を、周波数や音量の違った音声シグナルの発音により伝える。即時性が高く、アプリケーションの現在の状態が逐次ユーザにアップデートされて伝えられる。音声シグナルの生成にはCreative Lab's Open Audio Library (OpenAL)が使われる。オーディオシグナルは一定間隔ごと、ここでは100ms毎に生成され、音量とピッチは現在のカーソル位置とターゲットとの距離で変化する。

4. 触覚・聴覚フィードバックの効果に関する実験とその結果

本章では、今回提案したシステムを利用して行った実験の結果を示す。実験の評価基準は、実際のターゲットとユーザがクリックしたポイントのずれ(距離)、およびそれを行うためにかった時間、そしてその操作を行うための理解のしやすさと快適さに関するユーザへのアンケートの平均で行った。評価基準についての詳細、データ集合および結果については以下に示す。

4.1 実験に用いたインタフェースの種類および組合せ

実験を行うに当たって、各インタフェースの種類および組合せの呼称を以下のように定義した。

- (1) タイプ1：ハプティックマウスのみ。ハプティックマウスv.2(図2(b))のみを利用する。
- (2) タイプ2：オーディオタイプ1のみ。ターゲット方向の指示(上下左右)と、カーソル位置がターゲットのどこにあるかの簡単な説明(ターゲットの外側、内側、近く)を、MS Samボイスで発声する。

- (3) タイプ1+2：触覚マウスとオーディオタイプ1の組合せ。上記のタイプ1とタイプ2を組合せて利用する。
- (4) タイプ3：オーディオタイプ2のみ。アプリケーションの現在の状態に対応した音声シグナルのみを利用する。2次元軸の座標をそれぞれピッチと音量にマッピングする。

基本シグナルは以下にセットされる。

- ・周波数：440Hz
- ・サンプルレート：22,000sample/秒
- ・ピッチ範囲：0.0～0.5
- ・ゲイン：0.0～1.0

ピッチ範囲がx座標、ゲインがy座標を個別に反映する。

- (5) タイプ1+3：触覚マウスとオーディオタイプ2の組合せ。上記タイプ1とタイプ3を組合せて利用する。

4.2 データ集合

実験データは、いずれも従来型のマウスの利用経験のある23歳から28歳の間の9人の被験者によって集められた。各被験者は、実験のタスクを行うために、どのようにそのデバイスを扱うかという簡単な説明を5分間受けた(ただし、オーディオタイプ2を用いたインタフェースを除く。このインタフェースは、どのように音量とピッチが位置座標にマッピングされているかを被験者が理解するまでトレーニングすることが許可されている)。

各被験者には視覚障害者をシミュレートするために、実験中はアイマスクを付けた。被験者が実験タスクをこなしている様子および実験に使用したアプリケーションのスクリーンショットを図5に示す。

クリックするターゲットは円形になっており、内側と外側の二重の円で構成されている。内側は実際にクリックを行う箇所、外側はカーソルがポイントするターゲットに近付いていることを示すための円である。半径は内側が60画素、外側は90画素にセットする。クリック箇所とタ

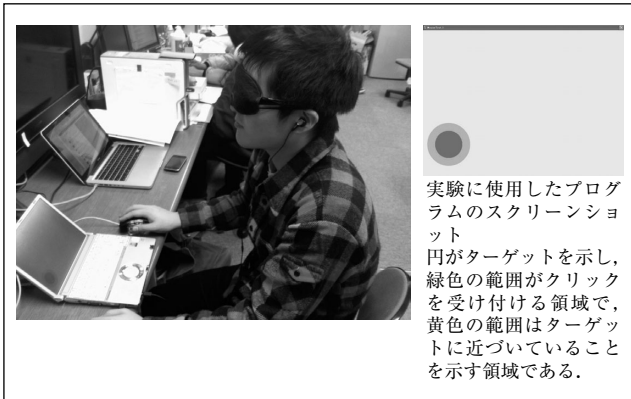


図5 アイマスクをした被験者が実験タスクを行ってる様子

ターゲットとの位置座標のずれの評価は、ターゲットの中心座標と、クリックされたポイントとの距離を測ることにより行う。

実験のルーチンは次のように繰り返される。被験者はマウスカーソルをターゲット円の真ん中に動かすよう指示される。ユーザがターゲット円を確認しマウスの決定ボタンをクリックすると、ターゲット円はランダムに違う場所に移動する。すべての被験者はこの手順を10回繰り返す。操作時間は、ユーザがターゲットをクリックするまでの1回ごとに記録される。

実験に用いたPCおよびその他の環境は表1の通りである。

4.3 結果

本節では、提案したインタフェースの評価結果を示す。それぞれのインタフェースタイプについて、すべての被験者のデータの平均と標準偏差をとった結果を表2に示す。また、実験終了後ただちに調査した各インタフェースについての操作の判りやすさと快適性に対するアンケートの評価値の平均値を表3に示す。ターゲットの中心とクリックした箇所の距離および時間の分布に関するヒストグラムとANOVA(分散分析)を図6と図7に示す。

4.4 分析

タイプ1+2の組合せは誤りが最小であり、また、ユーザの視点から見てもっとも快適性が高いことがわかる。操作時間に着目すると、タイプ1+3が最も早く、触覚と聴覚フィードバックの両方においてリアルタイム性が確保できていることがわかる。そのうえこれらの組合せは、ユーザの意見から充分な理解しやすさを持っていることがわかった。

実験結果の分布から見ると、視覚と聴覚を組合せたタイプは、単に一つのフィードバックインタフェースを使っただけのタイプより良く、特にクリック時のずれに関して大きな改善が見られることがわかる。逆に、タイプ2のフィードバックは操作時間の遅延が大きいことと、ユーザが正解率を上げるためにその情報を頼りにしないといけな

表1 実験に利用したPCおよびその他の環境

CPU	Intel Pentium-M ULV@1.06GHz
OS	Microsoft Windows XP® SP3
Screen	10.4インチ@1,024×768
Haptic Device	ハプティックマウスv.2(図2(b))
Audio Device	イヤホン

表2 インタフェース別の距離と操作時間の平均と標準偏差

	$d_{\mu}(\text{pixels})$	$d_{\sigma}(\text{pixels}^2)$	$t_{\mu}(\text{ms})$	$t_{\sigma}(\text{ms}^2)$
タイプ1	2.00	4.04	14089	7282
タイプ2	2.65	11.05	28131	26712
タイプ1+2	0.92	2.54	15747	7520
タイプ3	9.87	27.21	22458	20893
タイプ1+3	1.45	3.36	10307	8583

表3 理解しやすさと快適性についてのアンケート結果

各被験者がそれぞれについて1(低い)~5(高い)のポイントをつけたものの平均値。

理解しやすさの指標

	タイプ1	タイプ2	タイプ1+2	タイプ3	タイプ1+3
μ	3.56	3.22	4.33	2.78	4.56
σ	0.96	1.13	0.67	1.13	0.50

快適さの指標

	タイプ1	タイプ2	タイプ1+2	タイプ3	タイプ1+3
μ	3	3	4.11	1.78	3.33
σ	0.67	0.89	0.32	0.84	1.56

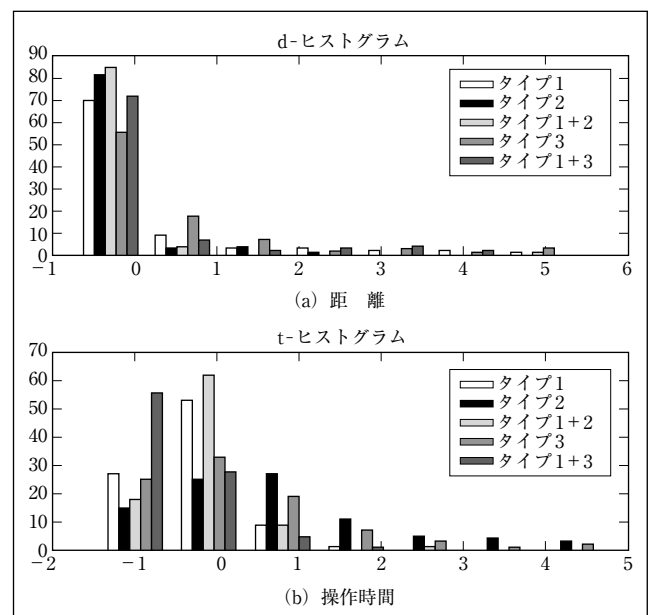


図6 インタフェースタイプ別のヒストグラム分布

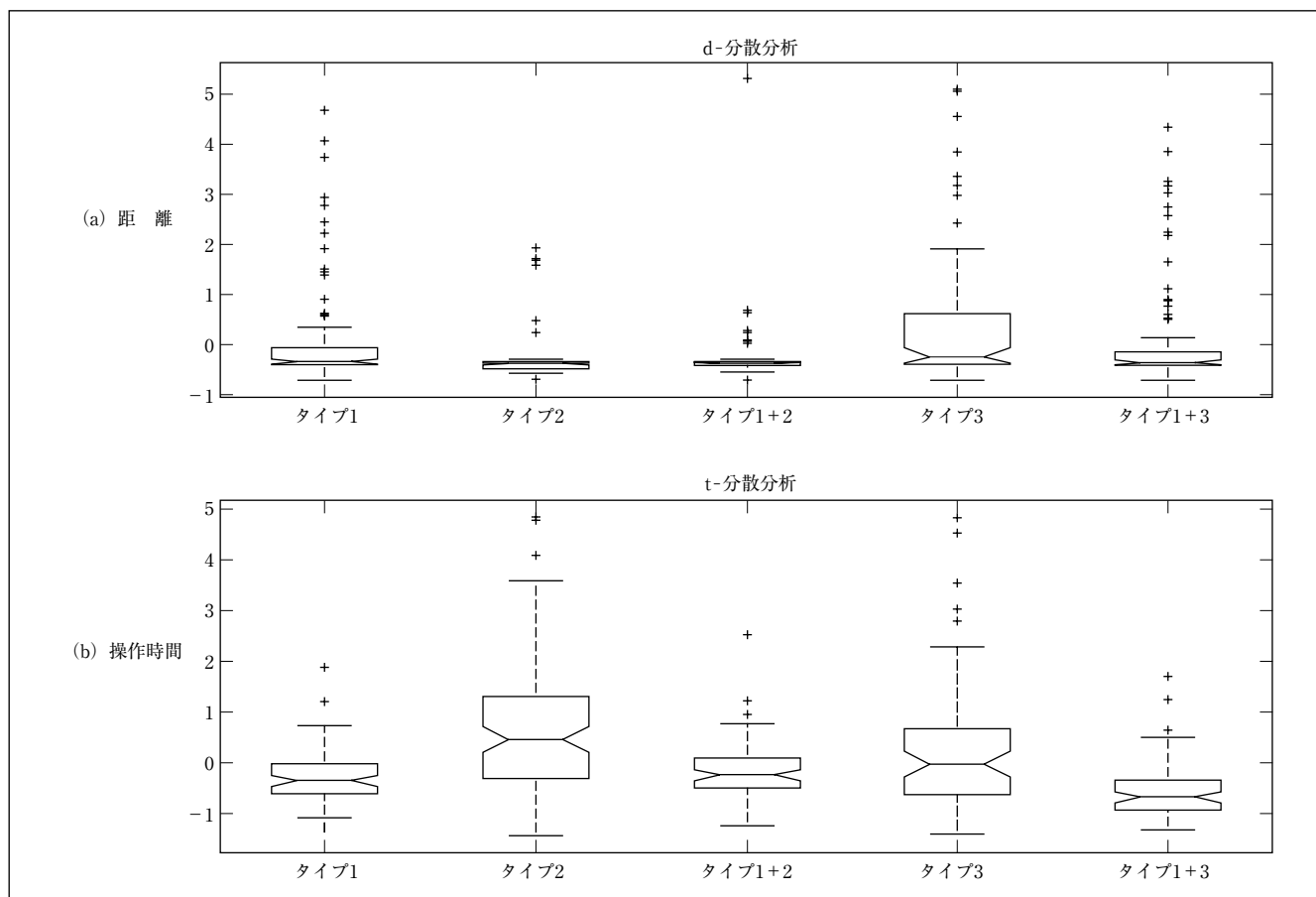


図7 インタフェースタイプ別のANOVA

とを考慮すると、タイプ1+2は、タイプ2のみを使った時と比較して遅くなる。その点ではタイプ1+3は、より応答性の高いフィードバックを使っており、ユーザはより素早く行動できる。

多くの被験者に目立った点として、タイプ2をベースとしたシステムは、クリック位置のずれにおいて大きな分散を示している。これは、発声が自然でないために、システムが指示した方向をユーザが区別できないためと考えられる。そして厳密な観察によって、タイプ3を用いたフィードバックにおいては、被験者が指示された方向をしばしば混乱してしまうことがわかった。これは、音量は負の値を持たないため、マッピングされたY軸方向の移動指示がわかりにくいためだと考えられる。

全体的に見て、提案した視覚と聴覚のマルチモーダルな入力デバイスシステムは、視覚がない状態において、ユーザが目標をクリックするというタスクにおいては、既存のシステムに比べて多くの優位点を持つことがわかった。しかし、触覚フィードバックの与え方として、ホイールをマウスに装備する現在のシステムは視覚障害者に対し有効であると考えられるが、三つのホイールによるフィードバックを親指のみに与えているため、単純な実装ではその情報の多さが問題になってしまうと考えられる。

次章では、筆者らが提案したシステムで、どのように視覚障害者を支援できるかを、ウェブブラウザアプリケーションの操作を例として示す。

5. 触覚・聴覚フィードバックによるウェブブラウジング

今日、コンピュータのアプリケーションの中で最もよく使われているものはウェブブラウザである。インターネット上の情報のほとんどは、ウェブブラウザを用いることによって容易に閲覧することができる。視覚障害者に向けたコンピュータ操作の支援を考えたとき、生活の情報源として不可欠になりつつあるウェブの閲覧を支援するシステムは、最も重要な研究課題となる。

ウェブブラウザ操作を支援するにあたって生じる問題は、本稿で行った実験と同じである。単にカーソルに対する移動の指示を発声するだけでは、ユーザに対するリアルタイム性に欠ける。また、単に触覚マウスを使うだけでは、逆にページ内の内容に関して何の意味も把握することができない。

しかし、この二つを組合せることにより、お互いを補完しあい、ユーザが情報に素早くアクセスしながら内容を把握できるようになる。ユーザは触覚マウスによってスクリ



図8 "Tokyo institute of Technology"という検索ワードで検索を行った時のウェブブラウザの動作例

ーン上のクリックすべきターゲットを知覚し、また、その間、コンピュータからの発声によってポイントされた箇所の前後の音声を聞いて、情報を効率的に探索することができる。

提案するウェブブラウザでは、ハプティックマウスv.2と、スクリーンリーダーアプリケーション(Amedia's Voice Screen-reader⁷⁾、日本語版)の組合せを用いる。クリックする位置は、触覚マウスからのフィードバックにより指示される。また、ポイントしている箇所はスクリーンリーダーによって音声で読み上げられる。また、カーソルキーが参照ポイントを移動するために使われる。

インターネットの検索エンジン(Google)を用いた情報検索のデモを行った。ここでは、ユーザは検索クエリを入力することができ、興味のあるコンテンツを自由に検索することができる。"Tokyo Institute of Technology"というクエリで検索を行ったときのウェブブラウザの動作例を図8に示す。

6. む す び

本稿では、視覚障害者がコンピュータを操作する上で有用な触覚と聴覚を用いた支援インタフェースを提案し、実際に視覚障害者をシミュレートして評価実験を行った。触覚と聴覚フィードバックを組合せて用いることで、それぞれを単独で用いるより良い支援が実現できることがわかった。

現状の視覚障害者支援システムで使われている発声による指示は、ターゲットをポイントする精度の向上のために良い働きを示し、音声シグナルによる指示は発声よりも情報を早く受け取ることができるが、いずれにしても情報を把握するために時間を必要とする。一方、触覚を組合せて

用いることにより、ポイント位置などの一部の情報に関してリアルタイムで把握できるようになる。また、聴覚を使った指示において、音色や音声は注意深く選択される必要がある。そうでなければユーザはそのシステムの操作を煩わしく感じてしまうだろう。

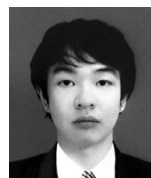
操作時間に関しては、触覚と聴覚を組合せによるタイプ1+3は触覚単独のタイプ1より40%早かった。また、組合せによるタイプ1+2の誤り率は、単独のタイプ1より50%低かった。タイプ2は、操作時間において他より明らかに劣っており、タイプ3は、タイプ1+2とタイプ1+3に比べて誤り率において劣っていた。(2011年11月30日受付)

〔文 献〕

- 1) M.D. Chinn and R.W. Fairlie: "ICT use in the developing world: An analysis of differences in computer and internet penetration", Review of International Economics, 18, 1, pp.153-167 (2010)
- 2) H. Kaye: "Computer and Internet use among people with disabilities. Disability Statistics Report, 13 (2000)
- 3) M. Sato: "Expanding welfare concept and assistive technology", in IIEE Annual Fall Conference (2000)
- 4) M.G. Parker and M. Thorslund: "Health trends in the elderly population: Getting better and getting worse", the Gerontological Society of America, 47, 2, pp.150-158 (2007)
- 5) J. Ivanson: "How to get to the cafeteria: Gesture and speech in blind and sighted children's spatial descriptions", Developmental Psychology, 35, 4, pp.1132-1142 (1999)
- 6) I. Kumazawa: "Haptic mouse with quick and flexible tactile feedback generated by double control loop", in Proceedings of the IEEE Internal Workshop on Robot and Human Interaction, 2010, pp.64-69
- 7) Voice surfing, speaking browser (in Japanese), <http://www.amedia.co.jp/product/vsurfing/>



小泉 亮 2011年、東京工業大学工学部情報工学科卒業。現在、同大学大学院物理情報システム専攻修士課程に在学中。入力インタフェースに関する研究に従事。



V-ris Jaijongrak 2007年、タマサート大学シリントーン国際工学部情報技術学科卒業。2010年、TAIST Tokyo Tech (タイ国科学技術開発庁タマサート大学シリントーン国際工学部および東京工業大学の連携大学院) 組込み情報システム (ICTES) プログラム修士課程修了。同年より、東京工業大学総合理工学研究科物理情報システム専攻博士課程所屬し、現在に至る。



熊澤 逸夫 1981年、東京工業大学工学部電気電子学科卒業。1986年、同大学大学院博士課程修了。現在、同大学情報工学研究施設教授。パターン認識・画像処理、人工ニューラルネットワーク、認知知覚モデルの研究に従事。正会員、博士(工学)。



～Smart Image Sensorを用いた高性能3次元形状測定システムの実現例～



3次元形状計測カメラ



正会員 池田 誠†

1. ま え が き

3次元計測には、主に三角測量法、光走行時間法(TOF法)、干渉法などがある。三角測量法は、その中で主に近距離を高精度に計測する手法として用いられており、その起源は古代エジプトにおける測地(おおよそ紀元前1400年ころ)にまでさかのぼることができ、受動法としてのステレオマッチング法、能動法としての光切断法、空間コード法(パターン光投映法)などがあげられる。ステレオマッチング法は、2章で述べるとおり、複数の汎用のイメージセンサ(カメラ)および汎用の画像処理プロセッサにより実現可能であることから、幅広く研究、実用化が行われている。光切断法は、投影された切断線の検出が容易であるため、パターンマッチングなどの後処理が不要であること、さらにそのパターンマッチングによる誤差が生じないことが利点としてあげられる。一方、一度に一切断面の形状情報しか取得できず、切断面をスキャンすることで全体の形状情報を取得することになるため、速度面で他の手法と比較し不利であること、一般的には機械的なスキャンが必要であることなどが欠点としてあげられる。これらの欠点に対して、3章では、スマートイメージセンサを用いることでの高速化、高精度化に関して紹介する。さらに、形状計測、距離測定に幅広く用いられる光走行時間法(ToF法)は、A. Fizeauによる光速測定(1849年)に起源を見出すことができ、4章で述べる。この他極近距離における高精度距離測定手法として、干渉法があり、こちらは、T. Youngによる干渉実験(1801年)にさかのぼることができるが、本稿では詳細は省略する。

図1に、高速3次元形状取得性能の現状を示す。ここで、Voxelは3次元体積素量を表しており、2次元画像における画素数(Pixel)に対応し、3次元空間の解像度を示す指標で

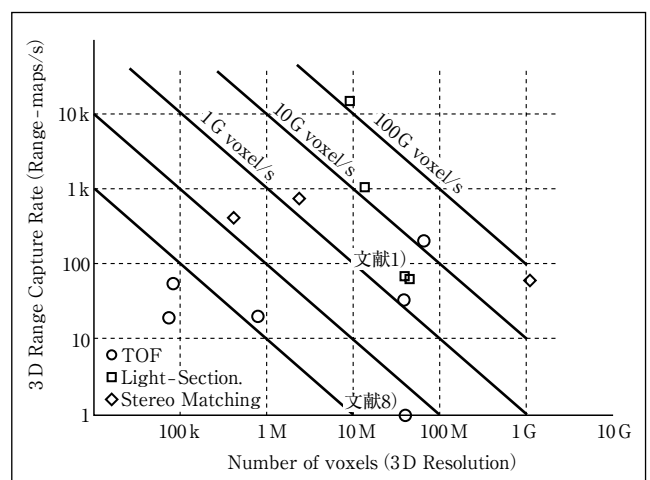


図1 高速3次元形状取得の現状

ある。なお、図1は各論文において示されているイメージセンサのピクセル数および3次元形状取得精度、3次元形状取得範囲から求めたものであり、論文に3次元形状取得精度の距離依存性が示されているものに関しては、その距離依存性を用いた空間解像度として示している。一方、3次元形状取得精度の距離依存性が示されていない場合には、(距離範囲)/(精度)により空間解像度を求めているため、実際より高めの空間解像度となっている可能性がある。

2. 受動法とステレオマッチング

ステレオマッチングは、視点の異なる2枚もしくはそれ以上の絵から、対応する部分を抽出し、その視差と撮像位置から奥行きを計算する手法であり、3次元形状と表面(テクスチャ)を同時に取得することが可能な方法である。例えば、図2の位置関係にある場合、距離 $Z = \frac{f \cdot d}{u \cdot (n_1 - n_2)}$ となり、ピクセルピッチが $5 \mu\text{m}$ のイメージセンサに焦点距離 20mm のレンズを使用し、カメラ間の基線距離が 10cm の設定で、100ピクセル分の視差 $(n_1 - n_2)$ が存在する場合、距離 $Z = 4\text{m}$ となる。また、この時、1ピクセルあたりの距離精度は、 $R = \frac{z^2}{f \cdot d} u = 4\text{cm}$ となる。このように三角測量に基づくステレオマッチング手法では、距離精度が測定対

† 東京大学 大規模集積システム設計教育研究センター

"Technology Frontier of Augmented Reality (3): 3D Range-finding Camera Systems" by Makoto Ikeda (VLSI Design and Education Center, the University of Tokyo, Tokyo)

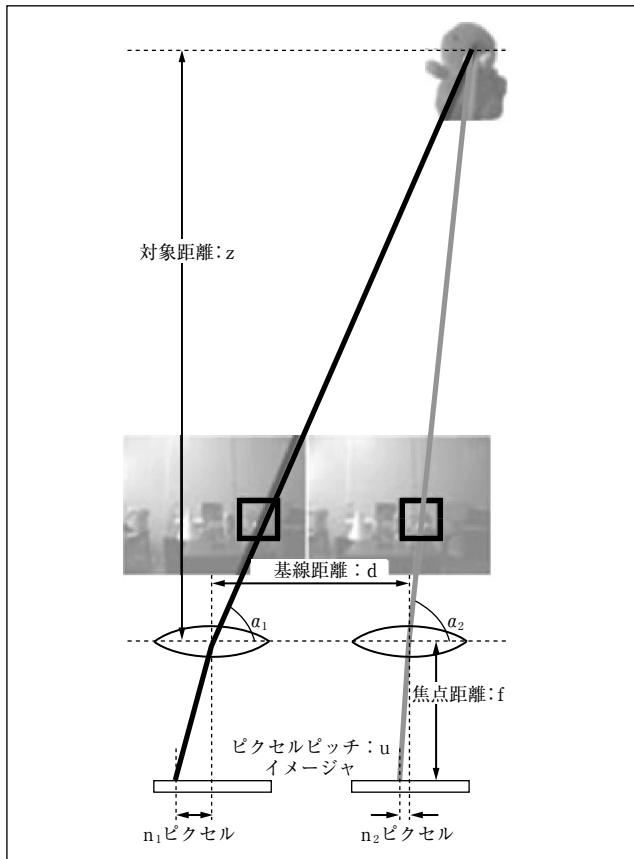


図2 ステレオマッチングの原理

象距離の2乗で劣化するため、近距離の距離測定に適することに加え、測定対象に対して適切な基線距離を設定する必要がある。

ステレオマッチングにおいては、2枚の画像から対応する点を抽出する必要があり、画素ブロックごとのSAD (Sum of Absolute Difference : 差分絶対値和) やSSD (Sum of Square Difference : 差分二乗和), NCC (Normalized Cross Correlation : 正規化相互相関) 等の手法を用いて探索するため、3次元の距離情報の抽出には多くの計算が必要である。近年では、FPGA (Field Programmable Gate Array) などを用いた専用の高速のハードウェアエンジンを用いることで、200 rangemap/sec程度の性能の実現例も報告されている。

一方で、模様のない表面等に対しては、距離を求めることができないといった欠点も存在する。

3. 光切断法とパターン光投影法

光切断法(図3)は、能動的な三角測量法の一種であり、パターン光の形状により0次元パターン光(ポイント光投影)、1次元パターン光(シート光投影: 光切断法)、2次元パターン光投影に分類することが可能である。0次元、1次元パターン光投影の場合、投影光から特徴点を一意に抽出可能であるため、前述のステレオマッチングのようにマッチングのために演算が不要であり、かつ、マッチングエ

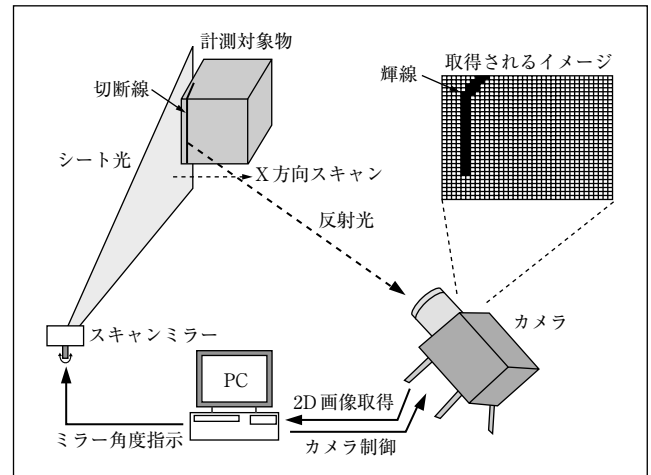


図3 光切断法の概念図

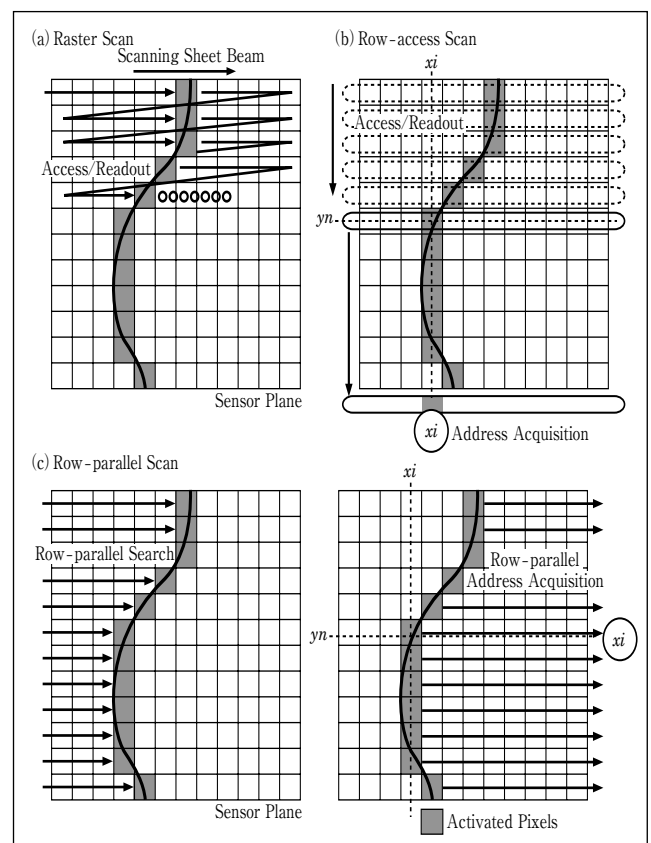
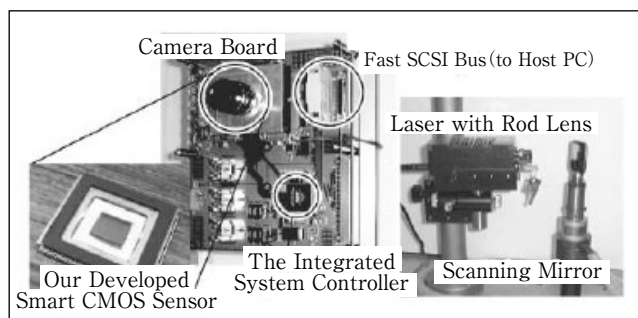
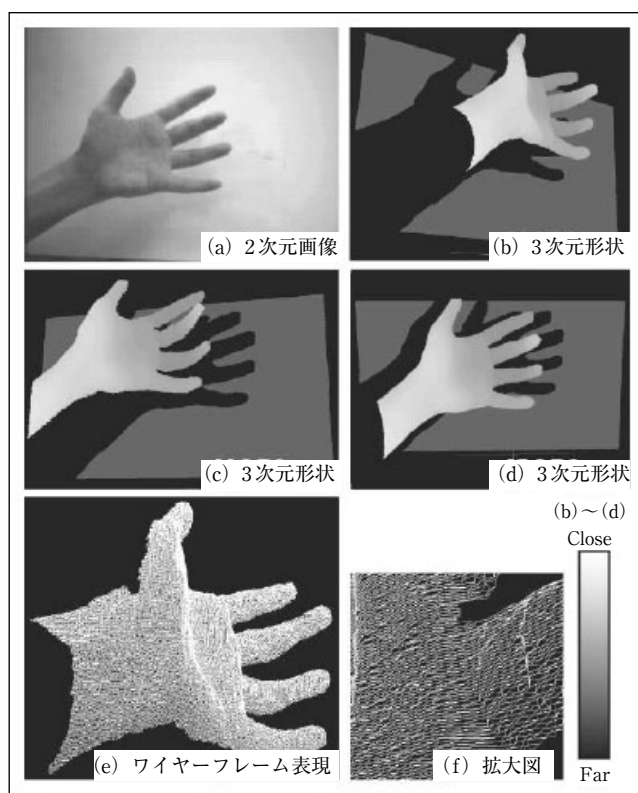


図4 光切断法における切断線の読出し

ラーによる誤差が生じないため、高精度な測距が可能である。ただし、0次元投影の場合、画面全面をスキャンする必要があるため、1次元投影の場合でも水平方向のスキャンが必要となるため、測距速度が投影光の抽出速度およびスキャン速度(多くの場合機械的速度)に律速されるほか、能動法の一般的な弱点として、背景光に対して充分強力な光投影が必要である点が問題点として挙げられる。

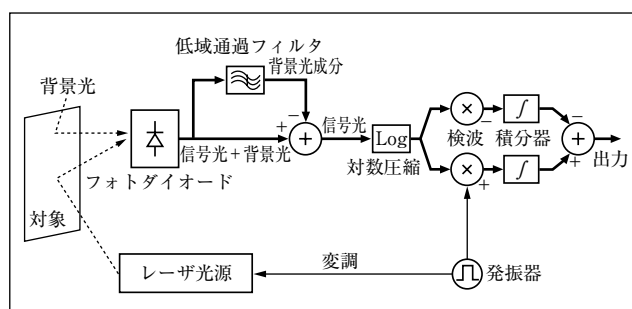
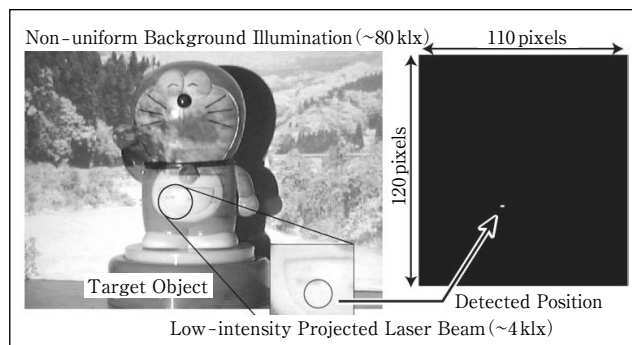
投影光の抽出速度に関しては、図4(a)に示すように、一般的なCIS (CMOS Image Sensor : CMOSイメージ

図5 SISを用い光切断法による3次元形状計測のセットアップ¹⁾図6 SISを用いて取得された画像¹⁾

センサ)やCCD(Charge Coupled Device：電荷結合型素子)のようにラスタスキャンにより全画素を読み出すことなく、センサ面上で輝点のみを検出し読出すことで高速化が可能である。図4(b)では、CISで一般的となっている行並列読出しに加えて、行並列で輝点検出を行うことで高速化を実現し、65.1rangemap/secを実現している¹⁾。同時に、1.2m先の対象に対して、最大誤差0.87mmの精度での3次元形状取得が可能であることを示している。図5には、Smart Image Sensor(用途に特化した機能を搭載したイメージセンサ、以下SIS)および制御用のFPGAを搭載したイメージセンサボード(カメラ)、シート光を生成するロッドレンズ付きレーザー光源とスキャンするガルバノミラーからなるシート光スキャナを示し、図6に、実際に得られた2次元画像、3次元形状を示す。表1は、SISの諸元および3次元形状測定結果のまとめである。

表1 行並列位置検出型3次元形状取得イメージセンサチップの概要と性能¹⁾

プロセス	0.6um CMOS
チップサイズ	8.9mm x 8.9mm
画素数	640×480 (VGA)
トランジスタ数	1.12M
画素寸法	12.0um x 12.0um
ピクセル形式	3Tr型
開口率	29.54%
電源電圧	5V
消費電力	305mW (@10MHz)
2Dフレームレート	13.0 fps
3D形状取得速度	65.1 rangemap/s
3D形状取得誤差	0.87mm (@1200mm)

図7 対数圧縮と検波を実現するピクセル回路²⁾図8 高照度下での変調光の検出結果²⁾

前述のように、能動法は一般的には背景光より強い参照光の照射が必要となるが、検波を用いた背景光除去手法を用いることで、この問題を解決することができる。図7に、検波および高照度状態での飽和を軽減するための対数圧縮を搭載したSISの模式図を示す²⁾。この方式をピクセル内に組み込むことで、背景光の1/100程度の変調光まで検出可能である。図8に、80klxの高照度下における4klx程度の変調光の検出例を示す。

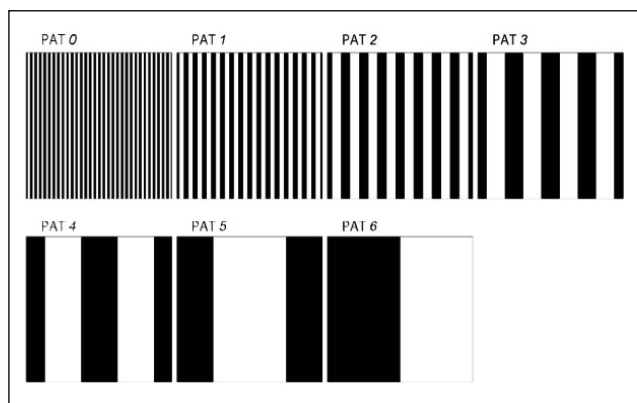
一方、0、1次元光を用いる場合に問題となる機械的スキャン速度律速からの脱却を目指したパターン光(2次元光)を用いた手法も数多く提案されている(図9)³⁾。

ここでは、当研究室で時間符号化手法(Time Multiplexing)

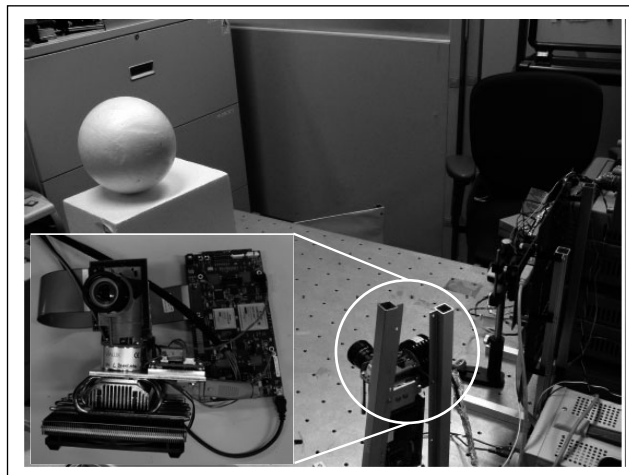
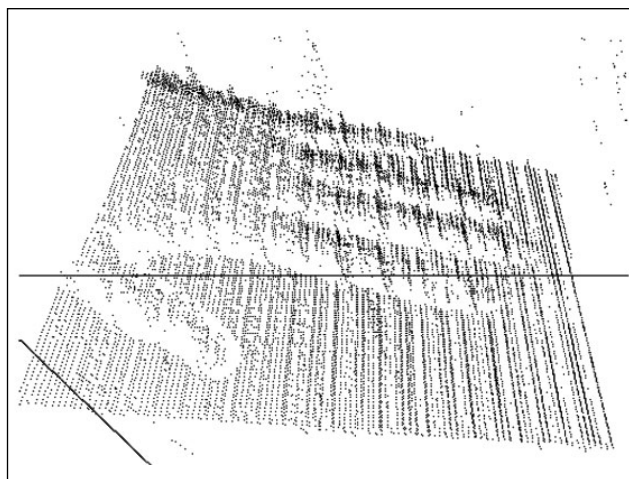


時間符号化 (Time-multiplexing)	二値符号 (Binary codes)		○	○	○	○
	N元符号 (N-ary codes)		○		○	○
	グレイ符号 (Gray code) + 移相 (Phase shifting)		○	○	○	○
	組み合わせ (Hybrid methods)		○	○	○	○
空間近傍符号化 (Spatial Neighborhood)	定式に基づかない符号化 (Non-formal codification)		○	○	○	○
	デブルーイン系列 (De Bruijn Sequences)		○	○	○	○
直接符号化 (Direct coding)	M配列 (M-arrays)		○	○	○	○
	グレイレベル (Grey levels)		○	○	○	○
	カラー (Color)		○	○	○	○

対象場面	静的 動的
階調	二値 白黒階調 カラー
符号化手法	周期的 非周期的

図9 パターン光投影手法³⁾図10 時間符号化投影パターン⁴⁾

に基づき、SISを用いた高速化を実現した3次元形状計測システムを紹介する⁴⁾。図10に、使用した時間符号化投影パターンを示す。これをチップ面上で効率的に認識可能なSISを用い、図11に示すように、DLP(Digital Light Processing:DMDデジタルミラーデバイスを用いた映像システム、テキサス・インスツルメンツの登録商標)を用いた高速投影によって、122rangemap /secの高速3次元

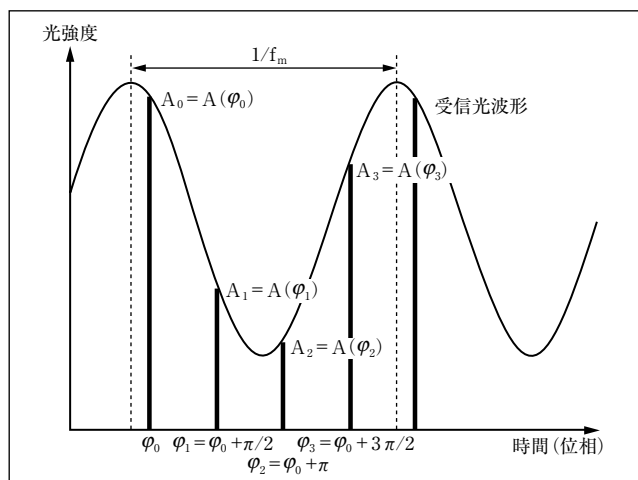
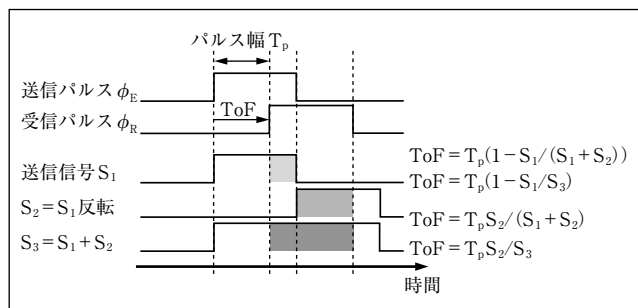
図11 時間符号化パターン光投影手法による3次元形状取得システム⁴⁾図12 取得した3次元形状例⁴⁾

形状取得(図12)を実現している。

4. 能動法とToF

MicrosoftからKinectが発表されるまで、3次元形状計測と言えば、ToF法が一般的であった⁶⁾。ToF法は、大まかに、連続波を用い光の位相を測定する手法と、単発パルス の走行時間を測定する手法に大別できる。

位相測定手法は、検波手法の一適用例であり、現在無線通信で広く用いられているIQ変調等と同様の手法により実現が可能である。ただし、この検波回路をピクセル内に納めるといった面積制約下では、図13に示すように、連続波の互いに等間隔な4サンプル値 A_0 、 A_1 、 A_2 、 A_3 から位相差 $\varphi = \tan^{-1} \frac{A_0 - A_2}{A_3 - A_1}$ を求める手法がLock-in Pixelという形で広く用いられている。本手法は、信号の差分を用いるため、背景光に強いほか、変調光(つまり反射光)の強度に依存せず位相差求めることが可能である⁵⁾。変調周波数はおおむね10MHz程度が上限となり、検波回数により誤差が軽減する特徴を有し、おおむね数mm～数cm程度の誤差となっている。なお、搬送波として光の代わりに電波を用い、

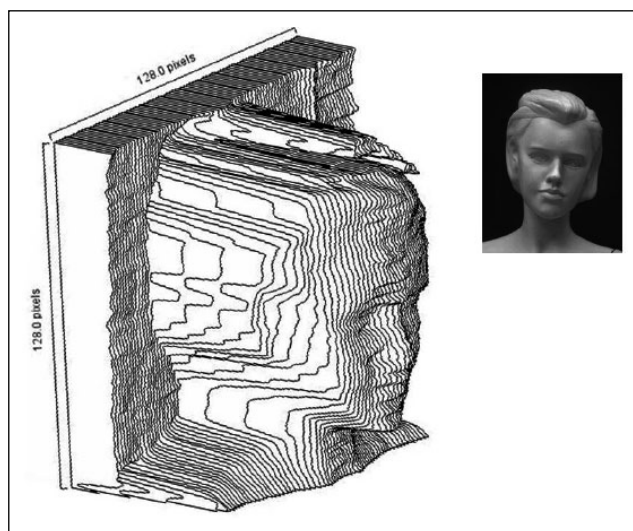
図13 連続波の位相測定⁵⁾図14 パルス方式による光走行時間測定の原理⁷⁾

高い変調周波数を用いる，もしくは音波を用いることで，変調波の波長を短くするとToF法の精度向上が可能となる．また，変調周波数の半分に相当する距離以上の位相は同定できないため測定が不可能となるが，複数の変調周波数および，そのビート周波数を用いることで解決することが可能である．原理的には測距精度は対象物体の距離に依存しないが，反射光強度の低下に伴い測距精度が劣化する．

パルスを用いる手法としては，図14に示すように，出射光，反射光の重なり部分から，距離 $Z = \frac{c}{2} T_p \left(1 - \frac{s_1}{s_1 + s_2} \right)$ 等として求める手法⁷⁾，および，SPAD (Single Photon Avalanche Diode：単一光子アバランシェダイオード)等光感度の非常に高い素子を用い，反射光を受光し光の到達時間そのものをTDC (Time-to-Digital Conversion：時間-デジタル変換)により測定する手法等が提案されている⁸⁾．図15にSPADを用いた3次元形状測定結果を示す．これらの手法では，測距精度はおおよそ数cm程度となる．

5. む す び

本稿では，数多く提案されている3次元形状測定手法・システムの中から，SISを用いることで高性能を実現している研究例，システム例について取り上げている．一方，1,000fps程度のCISが市販され，研究レベルでは，10Mfps程度の超高速イメージセンサが実現されており，

図15 SPADを用いたToF法による3次元形状測定例⁸⁾

また画像処理プロセッサも高性能CPU, GPU, FPGA，さらには専用LSI等を用いることで，SISを凌駕する性能が実現されつつある．システム的には，3次元形状にテクスチャを張り込んだ真の3次元映像を実時間で取得を可能とするシステム・イメージセンサの発表もなされるなど，今後ますますホットに研究・開発が進められるものと期待している．

(2011年12月12日受付)

〔文 献〕

- 1) Y. Oike, et al: "Design and Implementation of Real-Time 3-D Image Sensor with 640×480 Pixel Resolution", JSSC, 39, 4, pp.622-628 (2004)
- 2) Y. Oike, et al: "A 120x110 Position Sensor with the Capability of Sensitive and Selective Light Detection in Wide Dynamic Range for Robust Active Range Finding", JSSC, 39, 1, pp.246-251 (2004)
- 3) J. Salvi: "Pattern codification strategies in structured light systems", Pattern Recognition, 37, pp.827-849 (2004)
- 4) H. Yabe, et al: "CMOS Image Sensor for 3-D Range Map Acquisition Using Time Encoded 2-D Structured Pattern", 2011 Int. Image Sensor Workshop (IISW 2011)
- 5) R. Langer: "Solid-State Time-of-Flight Range Camera", IEEE J. QE, 37, 3, pp.390-397 (2001)
- 6) <http://www.csem.ch/site/card.asp?bBut=yes&pId=4288>
- 7) O. Elkhallil, et al: "3-D measurement applications", JSSC, 39, 7, pp.1208-1212 (2004)
- 8) C. Niclass, et al: "A 128x128 Single-Photon Imager with on-Chip Column-Level 10b Time-to-Digital Converter Array Capable of 97ps Resolution", ISSCC 2008, 2.1



池田 誠 1991年，東京大学工学部電子工学科卒業．1993年，同大学院工学系研究科電子工学専攻修士課程修了．1996年，同大学博士課程修了．同年，東京大学工学部電子工学科助手．同大学大規模集積システム設計教育研究センター助手を経て，現在，同大学大規模集積システム設計教育研究センター准教授．自己同期システムを用いた高信頼システム構築に関する研究．スマートタイムセンサによる高速実時間3次元形状計測手法などの研究に従事．博士(工学)．



First Person Visionにおける Inside-Outカメラ

関連動画
公開中!



正会員 藤吉 弘 亘[†], 木村 誠^{††}

1. ま え が き

近年、高性能カメラが小型化されたことにより、人の頭部や胸部等にカメラを装着して、装着者が見ているであろう視界に近い観測データから、装着者の行動を理解したり周囲の状況を認識しようとする研究が行われている。このような1人称視点を利用した画像解析、認識技術に基づく利用者への行動支援の枠組みをFirst Person Visionと呼ぶ。IEEEの国際会議Computer Vision and Pattern Recognition (CVPR) 2009にて、Workshop on Egocentric Vision¹⁾が開催され、First Person Visionの研究分野^{5) 14)}が注目を集めている*。

First Person Visionの研究では、人の頭部や胸部等にカメラを装着し、人の視点に近い位置から装着者が見ているであろう視界を観測する。この観測データを基にして、利用者の行動認識やライフログデータの収集、蓄積されたデータから解析や予測をして、利用者の行動の支援や生活の補助をすることが目的である。しかし、人の目は視界の一部に視線を向けて視対象を認識するため、視線情報を持たない観測データだけでは利用者の意図に沿う支援または補助をするには限界がある。この問題を解決するために、人の視線を計測する視線計測装置と、視界を観測するカメラを組合せた計測装置として、Inside-Outカメラが提案されている。

Inside-Outカメラは、図1に示すような、人の視点に近い位置から人の視界を撮影するSceneカメラと、人の眼球を撮影して視線計測等を行うためのEyeカメラにより構成されている。本稿では、First Person Visionに用いられるInside-OutカメラとInside-Outカメラの利用例について述べる。

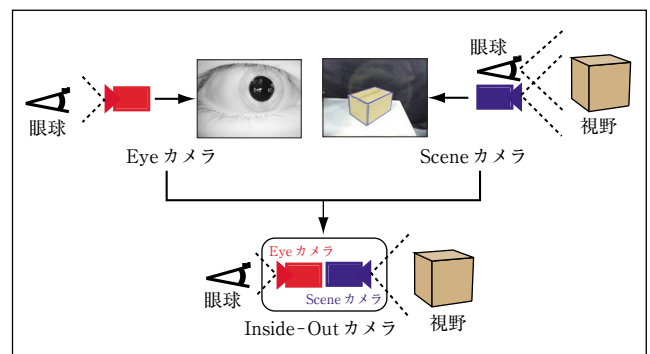


図1 カメラとSceneカメラ

て述べる。

2. さまざまなInside-Outカメラ

Inside-Outカメラは、EyeカメラとSceneカメラの構成によって使用目的が変わり、図2に示すように、大きく非ゴーグル型とゴーグル型に分類できる。非ゴーグル型は、主にメガネ等の装飾品にカメラを搭載したタイプで、軽量で装着者の負担は少ないが、カメラを小型にする必要があるため、パフォーマンスは低くなる。ゴーグル型は、頭部に装置を固定し、複数のカメラを搭載することのできるタイプである。

そのため、複数のカメラの観測データを基に、利用者の視界や視線の情報を密に計測可能である。本稿では、SceneカメラとEyeカメラをそれぞれ1台搭載したInside-Outカメラを単眼型と呼び、Eyeカメラの代わりに赤外センサを搭載したタイプをセンサ型と呼ぶ。また、複数のSceneカメラとEyeカメラにより構成されるInside-Outカメラを複眼型と呼ぶ。

本章では、各Inside-Outカメラにとその応用について紹介する。

2.1 単眼型

単眼型のInside-Outカメラは、EyeカメラとSceneカメラがそれぞれ1台搭載されているタイプのカメラである。

* 2012年に第2回目のWorkshop on Egocentric Visionが開催予定。

† 中部大学 工学部 情報工学科

†† 株式会社サムスン横浜研究所

"Technology Frontier of Augmented Reality (4): Inside-Out Camera in First Person Vision" by Hironobu Fujiyoshi (Department of Computer Science, Chubu University, Aichi) and Makoto Kimura (Samsung Yokohama Research Institute Co., Ltd., Yokohama)

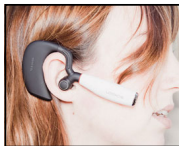
		← Inside-Outカメラ →				
		非ゴーグル型				ゴーグル型
デバイス						
カメラ数		なし 1台	なし 1台	1台 1台	1台 (センサ) 1台	2台 2台
視線推定		視線情報なし	頭部運動のみ	片目の視線 (粗) 粗	片目の視線 (密)	両目の視線 (密) 密
提案手法			Badinoら ²⁾	金出ら ⁵⁾ Landら ^{6, 7)} Pelzら ¹⁰⁾ Sailerら ¹¹⁾	石黒ら ⁴⁾	Shneiderら ¹²⁾ 清水ら ¹⁵⁾ 西山ら ¹⁶⁾ 折笠ら ¹⁷⁾
応用先				3次元マッピング	医療	
				ユーザインタフェース		
				ロボットアイ		
				Augmented Reality (AR)		
				3D表示		
				車載カメラ		

図2 First Person VisionとInside-Outカメラ



図3 単眼型のInside-Outカメラ

1999年にLandらは、図3に示す単眼型のInside-Outカメラを用いて人の視線を計測し、人がお茶を淹れる動作において、無意識的に共通の視線パターンが存在することを発見している⁷⁾。このような「視線の動きを調べる」研究は、1967年にYabusが「写真(の中のコンテンツ)を注視する」という被験者実験で視線の動きを調べたことが起源とされ、2001年にPelzらは、人の複雑な動作として洗面台で手を洗うという仕草の観測結果から、シーンの一部で視線

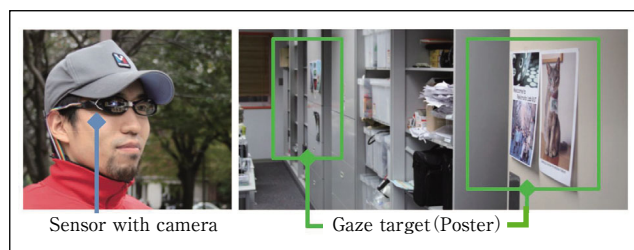


図4 石黒らのInside-Outカメラ(文献4)より抜粋

が先の動作場所へ動くという結果を報告している¹⁰⁾。

2005年には、Sailerらは知恵の輪のような簡単なタスクで知覚運動の学習をさせると、習熟に伴って視線の動きが変化することを報告している¹¹⁾。2006年にLandらは、ドライバが車を運転するシーケンスから、ハンドルの操作時に視線と体の動作に相関があることを報告している⁶⁾。

近年では、金出らが、生活の質の向上 (Quality of Life:QoL) を目的としたFirst Person Vision Wearable Deviceとして、メガネにEye, Sceneカメラを1台ずつ搭載させた単眼型のInside-Outカメラを提案し¹⁾、Inside-Outカメラを用いた視線解析以外の利用法について言及している。

2.2 センサ型

石黒らは、図4に示すように、単眼型のInside-OutカメラにEyeカメラではなく、光センサであるフォトトランジスタを搭載したセンサ型のInside-Outカメラを提案している⁴⁾。センサを用いることで、一般のカメラでは捉えるこ

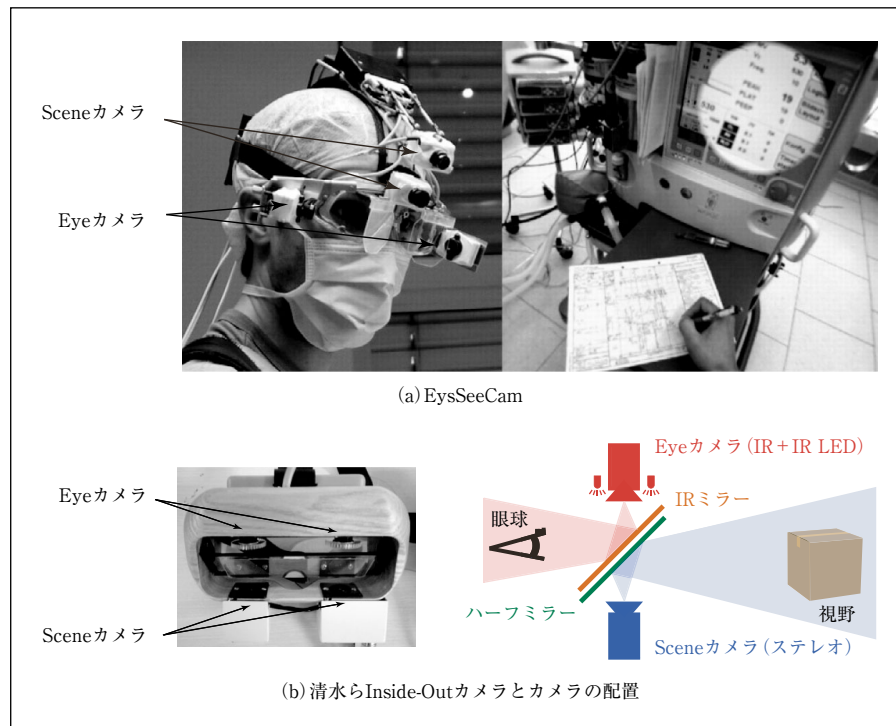


図5 複眼型のInside-Outカメラ(文献12)15)より抜粋)

とができない高速な眼球運動や瞬きの検出を実現している。センサ型は、カメラよりも小型化しやすく、低い計算コストで処理できる点が特徴であるが、環境光によるノイズの影響を受けやすい点を考慮する必要がある。文献4)において石黒らは、Inside-Outカメラを用いたライフログシステムを提案しており、人や物といった利用者に合った事前知識を登録しておくことで、注視領域内に存在する対象を高速に認識できるシステムとなっている。さらに、顔認識や文字認識を組合せることで、顔や文字などの認識対象の種類を拡張することも可能である。また、認識結果を蓄積することで、利用者の特徴に合わせた行動支援が可能となることから、データマイニングや物体認識への展開も考えられる。

2.3 複眼型

Schneiderらによって提案されたEyeSeeCam¹²⁾は、Eyeカメラ2台とSceneカメラ2台で構成されている。図5(a)に示すように、視界撮影用のカメラ1台は、人の視野全体を捉える広い画角を持ち、もう1台は、推定された注視点を追跡するPTZカメラとなっている。

EyeSeeCamは、両目の眼球映像と視界全体の映像に加えて、人の注視領域の高解像度な映像が撮影できるため、応用先は行動解析やライフログなど多岐にわたる。例として、熟練者の作業(医師の施術や看護師の機器確認)の注視点映像を撮影することで、ログデータを、経験の少ない新人や若手に対する教育目的に利用することができる。また、前庭動眼反射(VOR)実験では、頭部運動を打ち消すよう

な眼球運動を観測することで、健常者と異常者の診断にも利用されている。

著者の研究室では、図5(b)に示すような、ゴーグル型のフレーム上部にEyeカメラ2台と、下部に配置したSceneカメラ2台で構成されたInside-Outカメラを提案している¹⁵⁾。

このInside-Outカメラは、図5に示すように、眼球とEyeカメラの間で赤外光のみを反射するIRミラーと、シーンからの可視光をSceneカメラと眼球に分岐するハーフミラーが中央部に設置されているため、透明なカメラが存在するような光学的配置を実現している。著者らは、このInside-Outカメラを用いて、注視点の3次元位置を推定する手法を提案している¹⁵⁾。

3. Inside-Outカメラによる3次元注視点推定

本章では、著者らが提案しているInside-Outカメラにおける3次元注視点推定の手法について述べる。筆者らのInside-Outカメラは、他のInside-Outカメラとは異なり、Sceneカメラの位置が視線の位置とほぼ等しくなるよう配置されているため、シーン映像は人の見ている視界と同等に扱うことが可能である。そのため、カメラの位置ずれを考慮する必要がなく、眼球とEyeカメラ、Sceneカメラそれぞれの校正が容易に行える特徴がある。以下に、Inside-Outカメラによる3次元注視点推定の流れを示す。

(1) 視線ベクトル推定：瞳孔中心と角膜曲率中心から視線ベクトルを推定

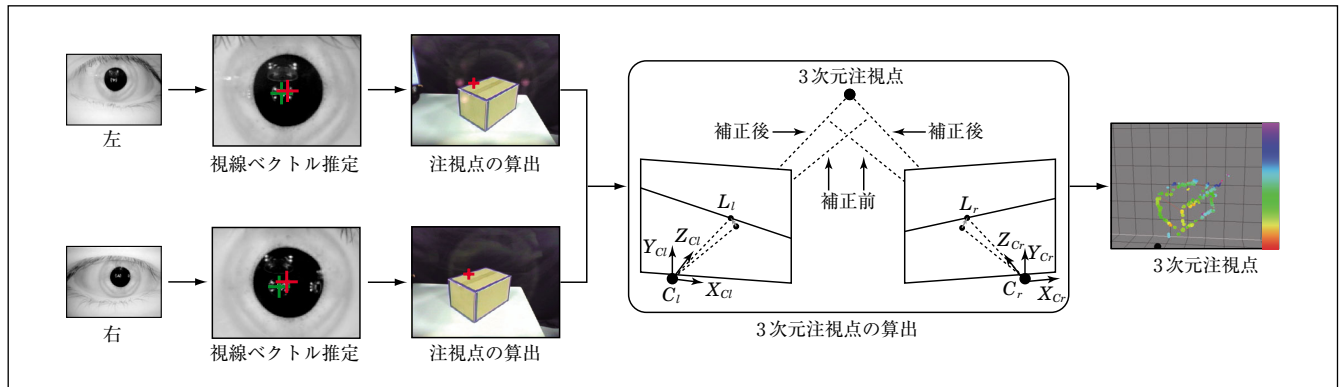


図6 注視点の算出の処理

(2) 注視点の算出：注視対象と視線ベクトルの相関関係から注視点算出

(3) 次元注視点の算出：Eyeカメラが2台ある場合、両眼の注視点を用いてステレオマッチングにより3次元の注視点算出可能

3.1 視線ベクトルの推定

視線検出手法により、眼球に赤外光が反射して映るブルキニエ像から角膜曲率中心を算出する。この角膜曲率中心と瞳孔中心を結ぶ直線を視線ベクトルとする。視線ベクトルは、角膜曲率中心とブルキニエ像から推定可能である。ブルキニエ像から角膜曲率中心を算出し、瞳孔中心を推定する。そして、角膜曲率中心と瞳孔中心から視線ベクトルを推定する。

3.2 変換式を用いた注視点の算出

3.1節により求めた視線ベクトルを用いて、画像上の注視点を推定する。

ここでは、変換式の変換パラメータを求める処理と、変換式を用いた注視点を推定する処理の、二つに分けて述べる。その処理の流れを図6に示す。

まず、視線ベクトルと注視点の関係を変換式で表現するパラメータを推定するため、視線ベクトルを入力としたときの注視点を算出する。次に、算出した変換パラメータを用いて、視線ベクトルから注視点を推定する。そして、算出した視線ベクトルを各成分に分割し、注視点を算出する。

3.3 3次元注視点推定

左右のEyeカメラから算出される注視点を、ステレオ視の問題に置き換えることで、3次元注視点を推定する。瞳孔中心と注視点を結ぶ線を視線とし、左右それぞれの視線の延長線上で交わる点が3次元の注視点となる。しかし、図6に示すように、左右の視線は正確に交わらないため、著者らは、最適補正により、二つの注視点と視線が対応するように位置の補正を行い、3次元注視点位置を推定している。図7に、著者らによる3次元注視点推定の結果を示す。3次元の注視点を推定することで、従来は2次元で計測されていたScan Path¹³⁾のような視覚機能を、3次元

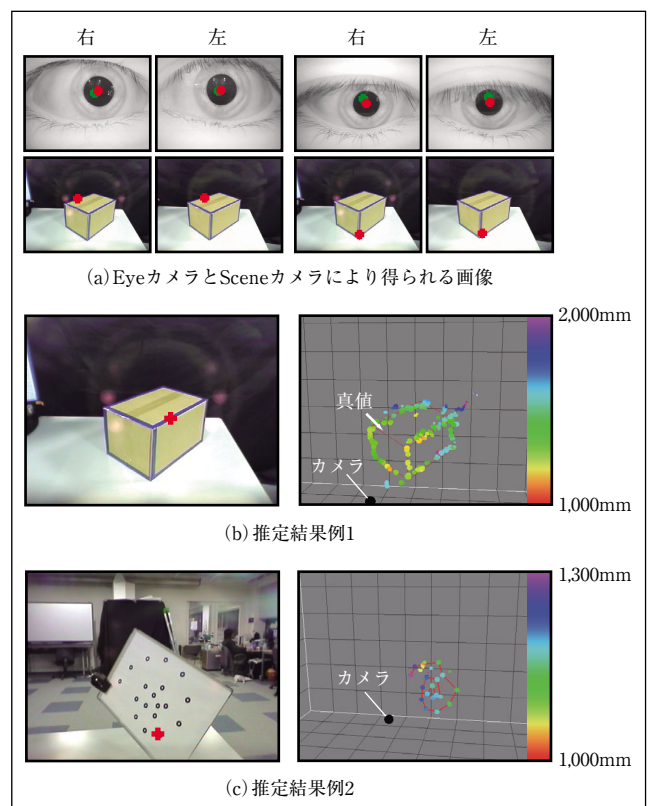


図7 3次元注視点の推定結果

的に計測・解析することができると考えられる。3次元の注視点の解析により、従来行われている視線計測や行動解析に適用することで、注視点に奥行き情報を含んだ解析結果が得ることができる。

4. 今後の展望

これまで述べたInside-Outカメラによる研究は、First Workshop on Egocentric Vision¹⁾で、金出らが提案しているように、Sceneカメラによるライフログ収集や、収集されたデータから建築物のマッピングなどの応用が考えられる。本章では、将来予測も含めた各分野における今後



の展望について述べる。

4.1 医療

2.3節で述べたEyeSeeCamは、元々医療研究用に開発されているため、さまざまな医療目的の応用が考えられている。例えば、多様な特性の色覚異常者を被験者として、自然環境中にさまざまな色のオブジェクトを配置し、どのような色に対してどのように視覚が反応するかを調べるような調査を行っている。このように、Inside-Outカメラでは、自然環境でのデータ取得が可能になるため、医療目的（診断、リハビリなど）の応用は今後も発展が期待できる。

4.2 3次元マッピング

Inside-Outカメラを装着した利用者が、歩行などの動作をすることで、多視点によるカメラ映像が取得できる。そのような無拘束の多視点映像から、特徴点抽出などの処理を経て3D形状を復元する研究として、Structure from Motion (SfM) が有名であるが、2011年に、人の運動モデルを考慮した3D形状復元手法をBadinoが提案した²⁾。

4.3 ユーザインタフェース

従来、視線計測システムに関する研究は数多く存在している。例えば、NAC社のEMR-AT Voxerのように、注視対象をディスプレイなどの平面に限定することで、非接触・無拘束な条件で、画面の中の注視点位置を特定するシステムがある。このような「ディスプレイ中の注視点の検出」というシステムは、研究や解析などを目的とした視点検出だけではなく、ユーザインタフェースとしての応用も期待されている。Inside-Outカメラを視線計測システムの拡張として考えると、Inside-Outカメラの応用によって、より一般的な条件下での視線ユーザインタフェースの実現が可能になると考えられる。

一方、ここ数年で多種多様な小型カメラ（例：メガネ型、耳掛け型、自転車のマウント用、ヘルメットへのマウント用など）が安価に流通するようになり、1人称視点からの映像の記録が容易になってきている。例えば、サーフィンやハングライダーなど、一般人が体験し得ない視点の映像が撮影できるため、YouTubeなどで1人称視点の動画が話題になることも多い。Inside-Outカメラを、このようなユーザ視点映像の拡張として考えると、2.3節で述べたEyeSeeCamの医療訓練のような熟練者の練習だけでなく、スポーツの練習などへの応用も考えられる。

4.4 ロボットアイ

EyeSeeCamの提案者であるSchneiderらは、人間の視線の動きを低レイテンシで再現できるロボットを作成し、Hannover Messe 2009において、EyeSeeCamでロボットの視線を遠隔制御するデモを行った。Schneiderらの研究では、このような視線制御を加えたロボットをアバターとして用いることで、遠隔コミュニケーションの質の向上が確認されている。また、人の視線の先をロボットが認識する「共同注視」という研究があり、多視点の監視システム

や、利用者の死角のアシストなどの応用分野が考えられている。ここにもInside-Outカメラは応用可能であると考えられる。

4.5 AR

従来、Virtual Reality (VR) やAugmented Reality (AR) において、ヘッドマウントディスプレイ (HMD) を用いたシステムは、研究、商用含めて数多く存在している。このようなHMD装着型のシステムにおいて、視線計測も含めたInside-Outカメラを融合することは、自然な流れであると予想される。

4.6 3D表示

ここ数年でメガネ型の3D-TVが多数発売されているが、それらの製品は、両眼視差のみで3Dを表現しているため、3D表示としてのさまざまな問題点が指摘されている。例えば、「運動視差がないため、頭部の移動に対して不自然になる」「実際の注視点の奥行き（表示面）と映像のフォーカスの不一致による眼球疲労」などが有名である。前者に対しては、Wiiリモコンを用いた頭部位置の観測や、iPhone、iPadの顔認識などを用いて、運動視差のみで3D的な表現を実現したシステムが提案されている。後者に対しては、注視点位置に合わせてフォーカスポケを調整して提示するシステムが提案されている¹⁷⁾。1人称システムを用いると、ディスプレイに対するユーザ視点位置の観測とディスプレイ内の注視点位置の推定を同時に行うことが可能になるため、3D表示の向上への応用が考えられる。

4.7 車載カメラ

近年、ドライブレコーダと呼ばれる車載カメラが普及している。これは、事故が起きた場合に、その直前の状況を確認することを主目的としたカメラである。一方、ドライバの居眠り検出や、眠気の予測¹⁶⁾などの研究は、過去に数多く行われている。これら「車の外の映像」、「ドライバの眼球の計測」を同時に行うカメラシステムは、ある種のInside-Outカメラと考えることができる。また、もしもInside-Outカメラの装着によって安全性を高めることが可能であれば、将来的に、ドライバ頭部へのInside-Outカメラの装着という展開も考えられるであろう。

5. む す び

本稿では、First Person Visionの研究に用いられるInside-Outカメラと各研究について紹介した。1人称カメラや視線計測の分野は、カメラやコンピュータの性能向上に伴い急速な発展を遂げており、実に多様な取り組みが行われるようになってきている。近年、視線計測の分野では、マイクロサッカードを計測することで人の潜在意識を解析する研究が進められている^{3) 8) 9)}。マイクロサッカードは、人が視線を動かす場合に、無意識に興味の対象へ小刻みに向く眼球運動である。そのため、二者択一のような実験から、人の内面、潜在意識を解析する手がかりとなる眼球運動と



して期待されている。マイクロサッカードは、極めて微小(0.1~1deg)で高速(500fps)な眼球運動であり、今後は、マイクロサッカードを観測するために、Inside-Outカメラの高解像度化かつ高速化が期待される。

First Person Visionの研究は、さまざまな分野や目的に適用することが可能であり、今後、私たちの生活の質の向上(Quality of Life: QoL)に欠かせない技術として、いっそうの研究開発が期待される。

(2012年2月6日受付)

〔文 献〕

- 1) First Workshop on Egocentric Vision, in conjunction with CVPR2009 (2009) <http://www.seattle.intel-research.net/egovision09/>
- 2) H. Badino and T. Kanade: "A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion", In IAPR Conference on Machine Vision Application (MVA) (2011)
- 3) K. Inagaki, Y. Hirata and S. Usui: "A model-based theory on the signal transformation for microsaccade generation", Neural Netw., 24, pp.990-997 (Nov. 2011)
- 4) Y. Ishiguro, A. Mujibiyi, T. Miyaki and J. Rekimoto: "Aided eyes: Eye activity sensing for daily life", In Proceedings of the 1st Augmented Human International Conference, p.25 (2010)
- 5) T. Kanade: "First-person, inside-out vision" (2009)
- 6) M. Land: "Eye movements and the control of actions in everyday life", Progress in Retinal and Eye Research, 25, 3, pp.296-324 (2006)
- 7) M. Land, N. Mennie, J. Rusted, et al: "The roles of vision and eye movements in the control of activities of daily living", PERCEPTION-LONDON, 28, 11, pp.1311-1328 (1999)
- 8) S. Martinez-Conde, S. Macknik, X. Troncoso and D. Hubel: "Microsaccades: A neurophysiological analysis", Trends in neurosciences, 32, 9, pp.463-475 (2009)
- 9) J. Otero-Millan, S. Macknik, A. Serra, R. Leigh and S. Martinez-Conde: "Triggering mechanisms in microsaccade and saccade generation: A novel proposal", Annals of the New York Academy of Sciences, 1233, 1, pp.107-116 (2011)
- 10) J. Pelz and R. Canosa: "Oculomotor behavior and perceptual strategies in complex tasks", Vision Research, 41, 25-26, pp.3587-3596 (2001)

- 11) U. Sailer, J. Flanagan and R. Johansson: "Eye-hand coordination during learning of a novel visuomotor task", The Journal of neuroscience, 25, 39, pp.8833-8842 (2005)
- 12) E. Schneider, T. Villgratner, J. Vockeroth, K. Bartl, S. Kohlbecher, S. Bardins, H. Ulbrich and T. Brandt: "Eyesecam: An eye movement-driven head camera for the examination of natural visual exploration", Annals of the New York Academy of Sciences, 1164, 1, pp.461-467 (2009)
- 13) A. Yarbus: "Eye movements and vision", Plenum press (1967)
- 14) 亀田能成, 藤吉弘亘, 羽下哲司, 山口晃一郎, 中村裕一: "First Person Visionへの挑戦 (First Person Visionのための認識・理解)", パターン認識・メディア理解研究会 (PRMU), 110, 28, pp.37-38 (2010)
- 15) 清水彰一, 西尾和見, 木村誠, 藤吉弘亘: "First Person VisionのためのInside-Outカメラの提案", 信学誌D, J94-D, 11, pp.1909-1918 (2011)
- 16) 西山潤平, 木下伸一, 平田豊: "前庭動眼反射による眠気予兆検出", 生体医工学, 48, 1, pp.1-10 (2010)
- 17) 折笠達郎, 鈴木孝明, 岡谷貴之, 出口光一郎: "注視反応ディスプレイによる両眼視差とボケの同時提示", コンピュータビジョンとイメージメディア (CVIM), 2011, 40, pp.1-8 (2011)



藤吉 弘亘 (ふじよし ひろのぶ) 1997年, 中部大学大学院博士後期課程修了。1997~2000年, 米カーネギーメロン大学ロボット工学研究所Postdoctoral Fellow。2000年, 中部大学講師, 2004年, 同大准教授を経て, 2010年より, 同大学教授。2006年, 米カーネギーメロン大学ロボット工学研究所客員研究員。計算機視覚, 動画像処理, パターン認識・理解の研究に従事。2005年度, ロボカップ研究賞。2009年度, 情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞。2009年度, 山下記念研究賞。2010年度, 画像センシングシンポジウム優秀学術賞。博士(工学)。正会員。



木村 誠 (きむら まこと) 2002年, 慶應義塾大学大学院理工学研究科博士課程修了。同年, 科学技術振興事業団CREST研究員。2006年, 産業技術総合研究所研究員を経て, 2010年より, (株)サムスン横浜研究所。コンピュータビジョン, 画像処理に関する研究開発に従事。2010年, ファナックFAロボット財団論文賞。2010年, 計測自動制御学会計測部門論文賞。博士(工学)。



触覚ディスプレイと拡張現実感

梶 本 裕 之[†]

1. ま え が き

われわれは感覚器という狭い窓によってのみこの世界を把握することができる。生物の進化とは一つにはこの窓の改良の歩みであり、例えば、ヒトに知覚可能な光の波長が太陽光エネルギー最大の領域と一致しているという事実は、進化による最適化の奥深さを感じさせる。しかし、いかにわれわれの感覚が最適化されていようと、むしろそのために、実世界には知覚不可能な物理現象の領域が多く存在する。これを知覚可能とすることは、最も直接的なヒトの能力拡張であるといえる。

Augmented Reality (以後AR) とは、こうした知覚能力の拡張技術の一種であり、現実感を損なわないままに、すなわち自分自身の能力として実現する点に特徴がある。本稿はこうしたARの、特に触覚に関する近年の研究の概説である。

近年、急速に話題となることが増えた視覚ARに対し、触覚AR (Haptic AR) は一般にはまだ耳慣れないかもしれない。しかし、そもそも1965年に開発されたHardiman¹⁾に端を発する外骨格型パワーアシスト装置自体、実世界の重さという触覚を変換するAR装置であったといえる。また、1982年に開発されたTELESAR²⁾などに代表されるTelexistence技術は、ロボットというインタフェースを介して遠隔地を眼前にある風景と等価にする技術であるから、それ自体がARであるともいえるし、触覚ARと見なせる技術要素が数多く含まれている。さらに触覚研究の分野では、ARという用語はすでに多用されており、いくつかの総説も書かれている³⁾⁴⁾。

本稿で扱う範囲は、特にユーザが自身の触覚的能力を拡張されたと感じられるものに限る。触覚ARの定義があいまいであった時期には、視覚ARに触覚手がかりを付与し

たものや、実世界で触覚呈示をするものを触覚ARと呼ぶ場合もあったが、本稿では範囲外とする。

本稿は、実現される「機能」に着目して触覚ARを分類して紹介する。第2章では、触覚ARの代表例として取り上げられることの多い、力覚による物性変化について述べる。力覚によるARは操作性向上を狙ったものも多いため、そうした操作性のARについて、第3章で述べる。第4章では、皮膚感覚に着目し、表面の材質感を変化させる試みについて述べる。第5章では、これまでのものと技術的共通点は多いものの変換元の情報が触覚以外であるものを、可触化のARとして紹介する。

2. 物性のAR

本章では、力覚によって接触対象の物性を主観的に変化させる試みについて述べる。ここで言う物性とは、粘弾性や質量など、対象の振る舞いを決める特性を指す。

最も代表的な物性のARとして、力覚呈示装置を用いて対象の粘弾性を変化させるものが挙げられる⁸⁾⁹⁾。システムは、力覚呈示装置と把持ツール先端に取り付けた力センサから構成される(図1)。ユーザがツールによって対象に接触すると、力センサによって反力が計測される。このとき反力に応じた適切な変位と速度を呈示すれば、対象とは異なる粘弾性を呈示することができる。同様の枠組みで、摩擦感と重量感に関しても呈示できる¹⁰⁾。

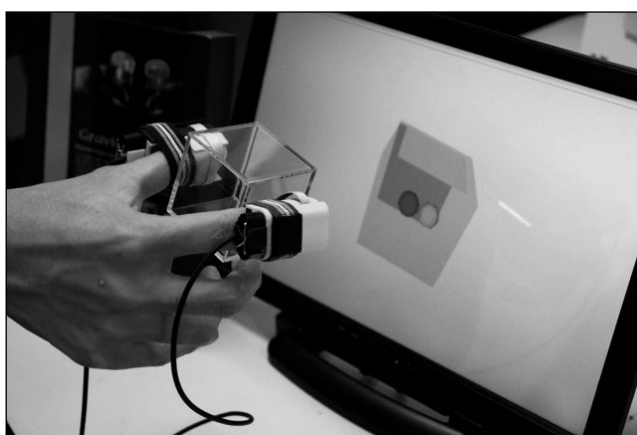
こうした物性のARの利点の一つは、作業に合わせて最適なインピーダンスを設定できる点である。例えば、柔らかいスポンジの表面に文字を書くことは通常困難であるが、硬いブロックのように振舞わせることで容易となる。

同様に、力覚呈示装置を用いた指先への触覚ARも提案されている¹¹⁾。触診トレーニングへの応用を想定したもので、平面状のシリコンゴムと指を吊り上げる力覚呈示装置によって構成される。シリコンゴムを触っている間に指を吊り上げることで、ゴムの固さを主観的に変化させ、「しこり」を表現することができる。

力覚呈示装置を指先に装着する別の試みでは、平面への

[†] 電気通信大学 総合情報学専攻

"Technology Frontier of Augmented Reality (5): Haptic Display and Augmented Reality" by Hiroyuki Kajimoto (Department of Informatics, The University of Electro-Communications, Tokyo)

図1 力覚呈示装置を用いた実物体の粘弾性感変化⁹⁾図2 皮膚感覚呈示による重量感呈示¹²⁾

手指の接触と掌内側からの力覚呈示の組合せによってボタンやスライダを押している感覚を再現する提案がある²⁶⁾。各指への力呈示を力覚呈示装置だけで実現するには大掛かりな機構が必要となるが、外力の大半を現実の反発力によって生じさせ、力覚呈示装置は補助的な力覚を返すという設計方針を採用している。

皮膚感覚の変化によって把持物体の重量感を変化させる試みも行われている^{12) 13)}。ベルトで指を締め付けることによって、実際に重量物を把持する際と同等の皮膚変形を生じさせる(図2)。軽い実物体(例えば図ではアクリル製の箱)を把持することで現実の力覚が生じ、皮膚変形による擬似的な重量感を重畳する。例えば、箱の中のボールの動きや液体の流動を呈示することができる。

3. 操作性のAR

前章で述べた物性のARの利点の一つは、行う作業に対して対象のインピーダンスを最適化することであった。この延長線上で、操作性向上という目的に限定したARを考えることができる。技術的には物性のARと違うところはないが、実現する触覚的表現は現実の物性を表していなくても良い。

操作性のARの代表的な例は、空間に「定規」を作り出すものである。例えば、糸による牽引で微小領域の描画作業をサポートするものや、脳外科手術において事前に取得したCTデータによって器具の進入不可領域を設定するものが提案されている^{14) 15)}。

最近の例では、彫刻刀を力覚ディスプレイに取り付けた木彫システムが提案されている¹⁶⁾。事前に用意したモデルデータによって、彫刻刀の進入不可領域を設定することによりハンドクラフト支援を行うというものである。これはまさに、夏目漱石が作品「夢十夜」で述べた「埋まっている仏像を掘り出す仏師」の喩えを具現化したものといえるかもしれない¹⁷⁾。

操作性のARは、対象が動的に変化する状況にも適用できる。例えば、心臓外科手術において、心拍を止めることなく施術するため、器具先端を心拍と同期して動かすことで、相対的に静止させる試みがある¹⁸⁾。

4. 材質のAR

これまでに紹介した触覚ARは、主に力覚呈示装置を用いたものであった。触覚のもう一つの要素である皮膚感覚に着目すると、対象の表面材質や細かな凸凹を変化させるAR技術を考えることができる。

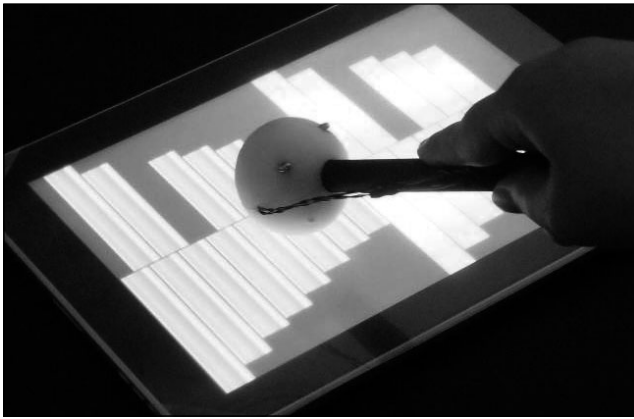
皮膚感覚ARの一例として、タッチパネル表面を振動させる試みが挙げられる^{19)~21)}。タッチパネル表面に触れた際にパネル全体を振動させることによって、クリック感に近い感覚を出すことができる。この機能は現在多くのタッチパネルインタフェースに採用されている。

振動するタッチパネルはパネル自体が情報呈示装置であり、さらに、振動付与の目的が材質感などの再現ではないため、ARの一種とみなされることは少ない。しかし、生じている現象に着目すると、タッチパネルという実物体の触感を振動によって変化させるということであるから、AR的状况であることは間違いない。

このような平面に対する触感の付与は、現在、クリック感にとどまらず数多く提案されている。例えば、平面と指先の間に生じる摩擦力を制御することにより、より精緻な表面材質感を呈示することができる^{22)~25)}。さらに、材質感の重要な要素である温度感覚についても、接触から数秒以内の温度変化を再現することによって正確な材質の表現が可能であることが示されている²⁷⁾。

平面の材質感を変化させる別の試みとして、床面を振動させて足に感覚呈示するものが提案されている²⁸⁾。床面パネルにスピーカ型アクチュエータとフィルム状力センサが取り付けられている。ユーザが床面を踏んだ瞬間にアクチュエータが振動することで、例えば、霜柱や薄氷を踏んだ感触がリアルに再現できる。

材質のARととらえられる別の試みとして、ペンを用いたものが数多く提案されている^{29)~32)}。ペンにアクチュエータが搭載されており、描画の瞬間に振動することにより書き味が変化する。例えば、滑らかな紙をザラザラした

図3 叩いた瞬間の振動重畳による材質感呈示³³⁾

紙やすりのように変化させることができる。

同様の単純なツールを介した材質感のARとしては、ペンによる「なぞり」ではなくスティックによる「叩き」に着目した提案もなされている³³⁾ (図3)。ヒトはスティックによってモノを叩いた時、ごく短時間の振動によって対象の材質を知覚できることが知られている³⁴⁾。この提案では、叩いた瞬間にスティックに内蔵された振動子で振動を生成することで、例えば、木琴か鉄琴かを識別させることができる。

以上の材質のARは、元々材質感のない接触対象に材質感を付与するものであった。これに対して、対象に接触した際の実際の振動を検出・増幅することで、対象の材質感をより明りように感じるという試みがいくつか存在する。

MicroTactusは、スティック先端でピックアップした振動を増幅して本体を振動させる³⁵⁾。Dayonは、同様の触覚増強機能を用いたコミュニケーション玩具である³⁶⁾。これらは共に、強調された振動によって現実よりも明りような材質感を呈示することができる。一方で、単純に記録した振動を再生すると触知覚可能な周波数領域を超えてしまうため、数十分の一の再生速度にすることで触覚のスローモーション再生を実現した例も挙げられる³⁷⁾。

以上の試みはすべて振動によって材質感を付与するものであったが、対象の微細な凹凸を機械的に増幅する試みも幾つか存在する。

乳がんの自己検診に用いられるTouch Enhancing Padは、薄いビニル製の袋に微量の潤滑油と空気が封入されている³⁸⁾ (図4)。腫瘍が存在すると思われる体部位上の皮膚にシートを乗せて触診を行うと、潤滑油の働きにより皮膚表面の摩擦がなくなり、結果として深部のしこり感が強調される。ARシステムと見るには単純に過ぎると思われるかもしれないが、人工的に触覚を増強していること、それを物理現象によって行っていることは特筆に値する。

同様に、物理現象を用いて対象の凹凸感を増幅する試みとして、触覚コンタクトレンズ(Touch Lens)が挙げられる³⁹⁾。触覚コンタクトレンズは、柔軟性のあるプラスチックで作成された微細な剣山構造をもつシートである(図5)。これを介して対象を撫でると、剣山底面が対象に沿って変

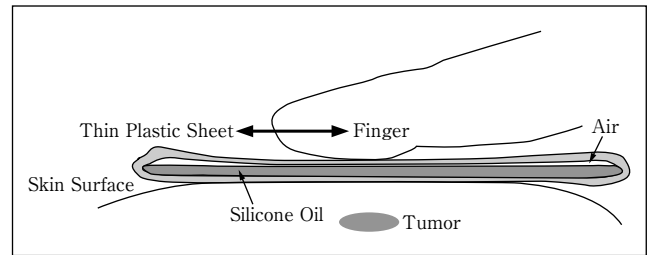
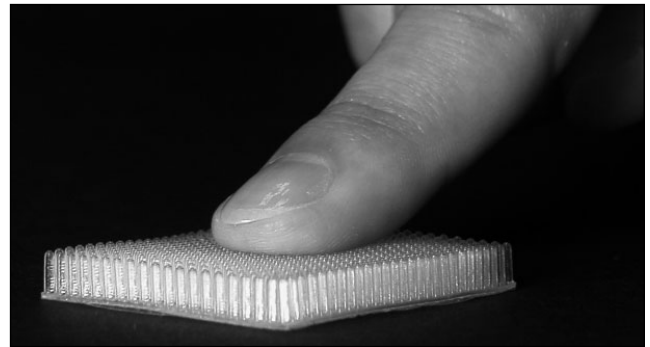


図4 摩擦の低減による腫瘍の探索(文献38)より再構成)

図5 微小な変位拡大機構による凹凸感の増幅³⁹⁾

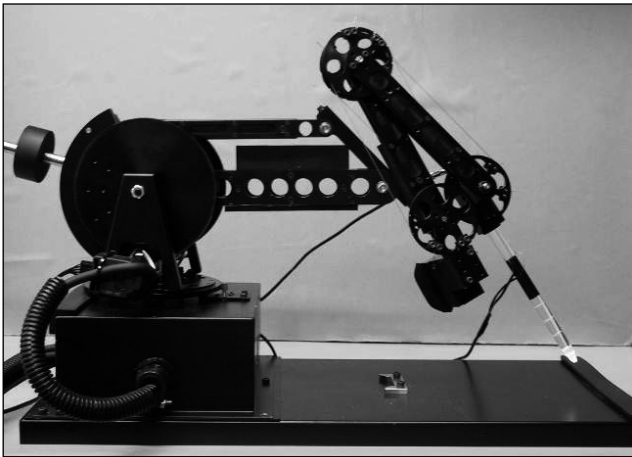
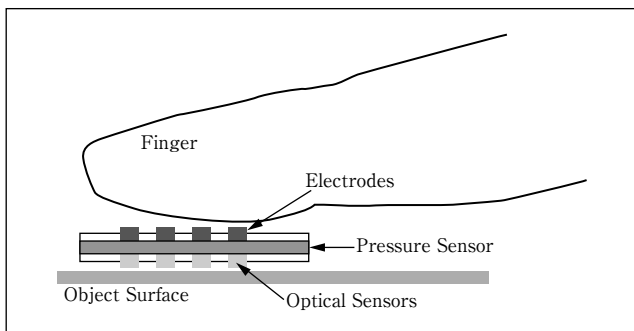
形し、その変形が剣山の微小なテコ機能によって増幅されて皮膚に伝えられる。例えば、5mm程度のゴムシートの下においた紙片を明りように知覚することができる。

5. 可触化のAR

これまでに紹介した触覚ARの多くは、対象の触覚的性質(固さ、材質)を知覚容易な領域に「移動させる」試みといえる。これに対して本来触覚としてはとらえられない実世界情報を触覚に変換する試みも数多く存在する。本章ではこれを可触化(Haptization)のARとして紹介する。

福祉分野における感覚代行技術は最も重要な可触化のARである。例えば視覚障害者に対して、カメラでとらえた画像情報を触覚的パターンとして呈示する試みが数多く提案されている^{40)~42)}。呈示部位は指先、背や腹、額などが用いられる(図6)。また聴覚障害者に対して、マイクでとらえた音声情報を触覚的パターンで呈示することも行

図6 額への画像情報の触覚呈示⁴²⁾

図7 センサ情報の可触化⁴⁷⁾図8 皮膚表面での光-触覚変換⁴⁹⁾

われている^{44)~46)}。呈示部位は腕や指先が用いられる。

福祉機器以外で実世界情報をリアルタイムに可触化した例として、SmartToolが挙げられる⁴⁷⁾。SmartToolは、ツール末端のセンサおよびツールに接続された力覚呈示装置から構成されている(図7)。センサを取り替えることによって、あらゆる情報を可触化することができる。例えば、センサとして色センサ、ツールとしてメスを用いることで、ゆで卵の黄身を傷つけることなく白身を切ることができる。また例えば、水と油の界面という本来触りえない情報を、ツール末端の導電センサによって触覚的な壁に変換できる。

同様のセンサ情報の可触化としては、指先への振動を呈示するSmartFinger⁴⁸⁾や、指腹表面にパターンを呈示するSmartTouch⁴⁹⁾が提案されている(図8)。これらの試みでも光センサを用いている場合が多いため、福祉分野における感覚代行技術と技術的な共通点が多い。ただし、「実世界情報そのものを触っている」という主観的状况を作るために、センサと触覚呈示部の位置を極力近接させる工夫がなされている。

なお、可触化というキーワードからはTangible-Bitに関する数多くの研究が思い起こされる⁵⁰⁾。ただしTangible-Bitは、情報とヒトの関わりに関する概念であるため、RealityをAugmentするという今回の枠組みでは議論しなかった。

6. む す び

本稿では、現在盛んに研究が行われている触覚ARに関して分類を試みた。触覚ARが実現する「機能」に着目すると、「物性のAR」、「操作性のAR」、「材質のAR」、「可触化のAR」があるように思われた。ただし例えば、操作性と可触化には大きな重なりが存在し、物性と材質に関しても、力覚と皮膚感覚は通常同時に生じるものであるから、明確な切り分けはむづかしい。より明確な別の軸による分類が必要かもしれない。

触覚ARの別の分類方法の一つは、「目的」による分類であろう。第一の目的は、物性や材質感、異種感覚を知覚すること、すなわち「知覚能力のAugmentation」であり、第二の目的は、操作性のARでみられる「技能のAugmentation」である。この二つの目的を持つこと自体、触覚ARの重要な特徴であると考えられる。触覚ARは実際に対象に接するために、対象自体に変形などの作用を及ぼさざるを得ない。このため、対象を知覚することと、操作することの二つが同時に目的となる。

一方で、これまでに概観した触覚ARには第三の目的が散見された。それは、現在の非力な触覚ディスプレイの呈示能力を「補完」するために、現実の物理現象を利用するというものである。例えば、基本的な物性の呈示は現実の物理現象に任せ、触覚呈示装置には感覚の微小な修飾に専念させる手法が提案されてきた。こうした例ではARを、目的ではなく技術的手段としてとらえている。これも触覚ARの一つの特徴であるといえる。

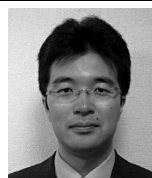
触覚呈示自体が広い分野であるが、その応用先は実は何らかの形でAR的な状況がほとんどであり、ARでない状況(コンピュータシミュレーションで閉じている状況)の方が少ないともいえる。よってこれまでに述べてきた分類はあまり意味がないかもしれないのだが、ARを研究分野の一部の名称としてではなく、研究分野全体を見る角度の一つと考えれば、こうした分類も何らかの気づき、新しい視点の提供につながるかもしれない。

(2012年2月13日受付)

〔文 献〕

- 1) Research and development prototype for machine augmentation of human strength and endurance - HARDIMAN I project. Specialty Materials Handling Products Operation, General Electric (1971)
- 2) S. Tachi, H. Arai, T. Maeda: "Development of an anthropomorphic tele-existence slave robot", Proceedings of the International Conference on Advanced Mechatronics, pp.385-390 (1989)
- 3) B. Bayart, A. Kheddar: "Haptic augmented reality taxonomy: Haptic enhancing and enhanced haptics", In Proceedings of EuroHaptics, pp.641-644 (2006)
- 4) S. Jeon, S. Choi: "Haptic augmented reality: taxonomy and an example of stiffness modulation", Presence, 18, pp.387-408 (2009)
- 5) 岩村: "タッチ", 医学書院 (2001)
- 6) 大山, 今井, 和気: "新編 感覚・知覚心理学ハンドブック", 誠信書房 (1994)
- 7) E.R. Kandel, J.H. Schwartz, T.M. Jessel: "Principles of Neural Science", McGraw-Hill (2000)
- 8) S. Jeon, S. Choi: "Modulating real object stiffness for haptic augment-

- ed reality", Proceedings of EuroHaptics 2008. LNCS, vol. 5024, pp.609-618. Springer, Heidelberg (2008)
- 9) Y. Kurita, A. Ikeda, T. Tamaki, K. Nagata, T. Ogasawara: "Haptic augmented reality interface Using the real force response of an object", Proceedings of 16th ACM Symposium on Virtual Reality Software and Technology, pp.83-86 (2009)
 - 10) S. Jeon, J.C. Metzger, S. Choi, M. Harders: "Extensions to haptic augmented reality: Modulating friction and weight", Proceedings of World Haptics Conference (WHC) 2011, pp.227-232 (2011)
 - 11) M. Solanki, V. Raja: "Haptic based augmented reality simulator for training clinical breast examination", Proceedings of Biomedical Engineering and Sciences (IECBES) 2010, pp.265-269 (2010)
 - 12) K. Minamizawa, H. Kajimoto, N. Kawakami, S. Tachi: "A wearable haptic display to present the gravity sensation - preliminary observations and device design", Proceedings of EuroHaptics 2007, pp.133-138 (2007)
 - 13) S. Scheggi, G. Salviettiand, D. Prattichizzo: "Shape and weight rendering for haptic augmented reality", In Proceedings of 19th IEEE International Symposium on Robot and Human Interactive Communication, pp.44-49 (2010)
 - 14) M. Furukawa, M. Ohta, S. Miyajima, M. Sugimoto, S. Hasegawa, M. Inami: "Support system for micro operation Using a haptic display device", Proceedings of 17th International Conference on Artificial Reality and Telexistence, pp.310-311 (2007)
 - 15) 野嶋, 稲見, 苗村, 川俣, 伊関, 館: "力覚のオーグメンテッドリアリティを用いた手術支援システムの研究, 日本VR医学論誌, 1, pp.78-82 (2002)
 - 16) A. Zoran, J. Paradiso: "FreeD" (2012), <https://www.youtube.com/watch?v=TxHiTjapTqg>
 - 17) 「夢十夜」 夏目漱石 (1908)
 - 18) S.G. Yuen, K.A. Dubec, R.D. Howe: "Haptic noise cancellation: restoring force perception in robotically-assisted beating heart surgery", proceedings of Haptics Symposium 2010, pp.387-392 (2010)
 - 19) M. Fukumoto, T. Sugimura: "Active click; tactile feedback for touch panels", Extended abstracts on Human factors in computing systems (CHI), pp.121-122 (2001)
 - 20) I. Poupyrev, S. Maruyama: "Tactile interfaces for small touch screens", Proceedings of Symposium on User Interface Software and Technology (UIST), pp.217-220 (2003)
 - 21) J. Lylykangas, V. Surakka, K. Salminen, J. Raisamo, P. Laitinen, K. Rönning, R. Raisamo: "Designing tactile feedback for piezo buttons", Proceedings of the 2011 annual conference on Human factors in computing systems (CHI), pp.3281-3284 (2011)
 - 22) T. Nara, M. Takasaki, T. Maeda, T. Higuchi, S. Ando, S. Tachi: "Surface acoustic wave tactile display", IEEE Computer Graphics and Applications Magazine, vol.21, pp.56-63 (2001)
 - 23) M. Takasaki, H. Kotani, T. Mizuno, T. Nara: "Transparent surface acoustic wave tactile display", Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp.3354-3359 (2005)
 - 24) L. Winfield, J. Glassmire, J.E. Colgate, M. Peshkin: "TPaD: tactile pattern display through variable friction reduction", Proceedings of World Haptics Conference (WHC) 2007, pp.421-426 (2007)
 - 25) B.O. Bau, I. Poupyrev, A. Israr, C. Harrison: "TeslaTouch: electrovibration for touch surfaces", Proceedings of Symposium on User Interface Software and Technology (UIST) (2010)
 - 26) C.W. Borst, R.A. Volz: "Evaluation of a haptic mixed reality system for interactions with a virtual control panel", Presence: Teleoperators and Virtual Environments, Vol.14, pp.677-696 (2005)
 - 27) A. Yamamoto, H. Yamamoto, B. Cros, H. Hashimoto, T. Higuchi: "Thermal Tactile Presentation Based on Prediction of Contact Temperature", Journal of Robotics and Mechatronics, 18, 3, pp.226-234 (2006)
 - 28) Y. Visell, A. Law, J.R. Cooperstock: "Touch is everywhere: floor surfaces as ambient haptic interfaces", IEEE Transactions on Haptics, 2, pp.148-159 (2009)
 - 29) I. Poupyrev, S. Maruyama: "Drawing with feeling: designing tactile display for pen", Proceedings of SIGGRAPH2002, Technical Sketch. Page 173 (2002)
 - 30) J.C. Lee, P.H. Dietz, D. Leigh, W.S. Yezazunis, S.E. Hudson: "Haptic pen: A tactile feedback stylus for touch screens", Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST), pp.291-294 (2004)
 - 31) K.U. Kyung, J.Y. Lee: Ubi-pen: "A haptic interface with texture and vibrotactile display", IEEE Computer Graphics and Applications, vol 29, pp.24-32 (2009)
 - 32) G. Wintergerst, R. Jagodzinski, F. Hemmert, A. Müller, G. Joost: "Reflective haptics: enhancing stylus-based interactions on touch screens", Proceedings of EuroHaptics 2010, pp.360-366 (2010)
 - 33) T. Hachisu, M. Sato, S. Fukushima, H. Kajimoto: "HaCHISick: simulating haptic sensation on tablet pc for musical instruments application", Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology (2011)
 - 34) A.M. Okamura, M.R. Cutkosky, J.T. Dennerlein: "Reality-based models for vibration feedback in virtual environments", IEEE/ASME Transactions on Mechatronics, vol 6, pp.245-252 (2001)
 - 35) H.Y. Yao, V. Hayward, R.E. Ellis: "A tactile magnification instrument for minimally invasive surgery", Lecture Notes on Computer Science (MICCAI), vol. 3217, pp.89-96 (2004)
 - 36) Dayon (2012), <http://www.otokinoko.com/dayon/index.html>
 - 37) Y. Hashimoto, H. Kajimoto: "Slow motion replay of tactile sensation", Proceedings of International Conference on Artificial Reality and Telexistence (ICAT) pp.51-56 (2010)
 - 38) D.A. Perry, H.E. Wright: "Touch enhancing pad", US Patent 4657021, 4793354 (1985)
 - 39) R. Kikuuwe, A. Sano, H. Mochiyama, N. Takesue, K. Tsunekawa, S. Suzuki, H. Fujimoto: "The tactile contact lens", Proceedings of the 3rd IEEE Conference on Sensors, pp.535-539 (2004)
 - 40) J.C. Bliss, M.H. Katcher, C.H. Rogers, R.P. Shepard: "Optical-to-tactile image conversion for the blind", IEEE Transactions on Man-Machine Systems, vol. MMS-11, pp.58-65 (1970)
 - 41) C.C. Collins: "Tactile television: mechanical electrical image projection", IEEE Transactions on Man-Machine Systems, vol. MMS-11, pp.65-71 (1970)
 - 42) H. Kajimoto, Y. Kanno, S. Tachi: "Forehead electro-tactile display for vision substitution", Proceedings of EuroHaptics (2006)
 - 43) S. Tachi, K. Tanie, K. Komiyama, M. Abe: "Electrocutaneous communication in a guide dog robot (MELDOG)", IEEE Transactions on Biomedical Engineering, BME-32, pp.461-469 (1985)
 - 44) G.V. Békésy: "Human skin perception of travelling waves similar to those on the cochlea", Journal of Acoustical Society of America, 27, pp.830-841 (1955)
 - 45) 伊福部, 湊, 吉本: "心理物理実験によるタクトイル・ボコーダの基礎的研究", 日本音響学誌, 31, pp.170-178 (1975)
 - 46) C.E. Sherrick: "Basic and applied research on tactile aids for deaf people: Progress and prospects", Journal of Acoustical Society of America, 75, pp.1325-1342 (1984)
 - 47) T. Nojima, D. Sekiguchi, M. Inami, S. Tachi: "The SmartTool: A system for augmented reality of haptics", Proceedings of IEEE Virtual Reality, pp.67-72 (2002)
 - 48) H. Ando, T. Miki, M. Inami, T. Maeda: "The nail-mounted tactile display for the behavior modeling", ACM SIGGRAPH 2002 Conference Abstracts and Applications, p.264 (2002)
 - 49) H. Kajimoto, M. Inami, N. Kawakami, S. Tachi: "SmartTouch: electric skin to touch the untouchable", IEEE Computer Graphics and Applications Magazine, 24, pp.36-43 (2004)
 - 50) H. Ishii, B. Ullmer: "Tangible Bits: Towards seamless interfaces between people, bits and atoms", Proceedings of Conference on Human Factors in Computing Systems (CHI), pp.234-241 (1997)



梶本 裕之 2003年, 東京大学大学院助手。2006年, 電気通信大学人間コミュニケーション学科助教授。2007年, 同大学准教授。2010年, 電気通信大学総合情報学専攻准教授となり, 現在に至る。触覚ディスプレイ, 触覚センサ, 人工現実感システムなどの研究に従事。博士(情報理工学)。



1. ま え が き

近年、視覚、聴覚以外の感覚情報を含めて人に提示するヒューマンインタフェースの研究が行われるようになってきた。今まで嗅覚に関してその情報提示技術が研究されることは少なかったが、近年コンピュータからの指令により目的とする香りを発生させて、人間に提示する嗅覚ディスプレイが人工現実感の分野で研究されるようになってきた¹⁾²⁾。映像情報メディア学会でも2006年に嗅覚ディスプレイの関する講演会が企画されている。視聴覚以外の感覚情報が加わることで、今までにない体験が可能になる。感覚情報の提示は視覚だけとは限らず、例えば、触覚情報の提示を行う装置は触覚ディスプレイである。よく展示会でディスプレイから匂いが出るのですかという質問を受けるが、視覚ディスプレイと一体になっているかどうかにかかわらず、嗅覚ディスプレイは香りを提示する装置を意味している。

絵画や音楽を楽しむ時に視聴覚が用いられるが、視覚や聴覚の情報の多くは論理的思考に用いられる。電車に乗ったり、歩いて目的地へ行くためにも視聴覚がなければ大変であり、目的を効率よく達成するために視聴覚は用いられる。

その一方で嗅覚は気分や情緒に大きな役割を果たし、感性情報ということができる。現代社会では基本的には論理的思考に基づき、なんらかの目的を達成するために最も効率的な方法を探索することが多くなされる。これまでのように、物質的な豊かさを求めている時代はそれでもよかったが、現代の多くの人は精神的な豊かさを追求するようになってきている。そのために視聴覚情報以外の感性情報提示技術は、これから必要とされる技術である。

人に嗅覚刺激を与える装置にはオルファクトメータがある³⁾。オルファクトメータは嗅覚誘発脳波を測定したり、嗅覚閾値を測定したりするために用いられてきた。この装

置を用いると濃度や提示時間等を正確に制御できるが、装置がおおがかりで価格も高い。また、市販のアロマディフューザは簡便であるが、通常は一つの香りの提示にのみ用いられ、香りを変えるにはカートリッジを交換しなければならない。そこで、コンピュータに接続して簡単に使用できるものとして嗅覚ディスプレイが注目され、人工現実感の分野で研究されるようになってきた。コンピュータ制御で匂いを切替えるディフューザも開発されたが、この装置は香りを調合する機能はない。コンピュータ制御のディフューザも有用な場面もあるが、多様な香りを出すのには香りの調合機能がほしい。

本稿では嗅覚ディスプレイの研究動向について述べ、多成分調合嗅覚ディスプレイの仕組みや遠隔匂い再現システム等の話題について説明する。

2. 嗅覚ディスプレイの研究動向

廣瀬によると嗅覚を含めた最初の五感ディスプレイはセンソラマであるという⁴⁾。その後、資生堂がKaori on demandを発表したり、トヨタのショールーム(アムラックスシアター)などで香りの提示が行われた。最近では辻らがフランステレコムおよびそのスピンオフした会社と香りwebという名前の装置を開発している⁵⁾。

また、柳田らは香りを人の鼻に供給するための技術が重要であるとして香りプロジェクトを開発した。香りプロジェクトは空気砲の原理を応用して特定の人のみに渦輪に香りをのせて届ける技術である。また、空気の圧力も感じてしまうので二つ空気砲を用いて人の鼻の前で渦輪を衝突させると圧力を感じなくなるという。さらに同氏のグループは、歩行者に対してその動きを予測して香りを届ける方法も研究している⁶⁾。

また、近年インクジェットを用いた嗅覚ディスプレイも研究されている。筆者らは、2004年頃にインクジェットを用いた香り提示法を提案したが⁷⁾、その後岡田らによって、短い時間の香り提示が可能な装置が開発された⁸⁾。この装置と呼吸との同期方法も研究されたが、呼吸のタイミ

† 東京工業大学 理工学研究科

"Technology Frontier of Augmented Reality (6): Olfactory display and its application" by Takamichi Nakamoto (Graduate School of Science and Engineering, Tokyo Institute of Technology, Tokyo)

ングを簡便な方法でセンシングすることは容易なことではないようである。

金らは、機能性高分子のゾルゲルの相転移を利用して香りの放出を制御するデバイスを開発した⁹⁾。ペルチェ素子による加熱冷却により、容器出口を覆っている機能性高分子膜の状態をゾルもしくはゲルの状態に可逆的に制御する。彼らはこのデバイスをアロマチップと呼んだ。また、熱による高分子の膨張/収縮を利用した素子を2次元状に配置して、指定した素子から香りを発生させるデバイスも提案されている¹⁰⁾。

また、香り提示の効果を客観的に検証しようという研究もある。伴野は、人間の視線を視線検出装置によりトレースすることによって何に注目しているかを調べた。適切な香りを付加すると視線は対象に誘導されやすく、記憶に残りやすいという¹¹⁾。

廣瀬らは、wearable嗅覚ディスプレイを製作し、それを持った人間がバーチャル空間上で匂い源を定位できるかどうかの研究を行った¹²⁾。匂い源からの距離に応じて香りの強さを変えるようにした結果、バーチャル空間上で空間的な匂い分布を体感することができるようになったという。石田らは、バーチャル空間上においた匂い源から放出される匂いの空間分布を数値流体シミュレーションで計算して、その計算結果で得られた濃度の香りを嗅覚ディスプレイで提示する研究を行った。匂いは風で運ばれるので、等方拡散モデルよりも現実的な匂いの空間分布が再現できる¹³⁾。

さらに、視覚等の他感覚との相互作用に関しても研究されている。河合らは、視覚に嗅覚の情報が影響されることがあり、その相互作用を利用することにより少ない香りの数でも多様な香りを表現できるとしている¹⁴⁾。廣瀬らも、同様な研究を行った後、味覚-嗅覚の相互作用に注目した¹⁵⁾。プレーンクッキーに拡張現実感の技術でチョコレート等の画像をクッキーに重ね合わせるとともに画像に合わせた香りをユーザに嗅いでもらう。そうすると、食べているのはプレーンクッキーなのに匂いに合わせたお菓子の味がするという。われわれが味わうおいしさは嗅覚に由来するものが多く、味と思っているものでも実は香りから来ているものが多い。このディスプレイは、このように嗅覚を用いてわれわれが思っている味を擬似的に実現したものである。

上に記したように、最近10年程度の間に嗅覚情報の提示技術が研究されるようになってきた。嗅覚ディスプレイに関連した技術は日本を中心に論文発表があり、欧米では初期的な製品は開発されても嗅覚ディスプレイに関する研究論文は非常に少ない。

3. 多成分調合嗅覚ディスプレイ

3.1 多成分調合方式

前節の嗅覚ディスプレイは、個々の香りを放出して香りを提示する技術である。XYの2次元状に素子を並べ、アド

レスを指定して対応する香りを放出する試みや、視覚等の相互作用を利用して少ない香りで多くの香りを擬似的に表現する試みもあるが、香りの種類は多様であるため、香りを調合して新しい香りを作り出す技術が有用である。実際に調香師が新しい香りを作り出すときは、この調合のプロセスを繰り返して行う。本節では、多成分調合の方法と要素臭の選択について述べる。

要素臭とは香りのもとであり、これを任意の比率で調合して嗅覚ディスプレイで提示する。図1(a)～図1(d)に、多成分調合を行ういくつかの方法を示す。図1(a)は、複数のマスフローコントローラを用いて調合を行う方法である¹⁶⁾。マスフローコントローラは電子的に流量を制御する装置であり、各マスフローコントローラの流量比によりさまざまな匂いのレシピを実現することができる。通常気体の流量は単位時間当たりの通過体積で表されるが、体積は圧力に依存するために、通常のプロトタイプ流量計を用いた調合では調合比が圧力によって変化してしまう。マスフローコントローラは、圧力に依存しない質量流量を用いて調合比を

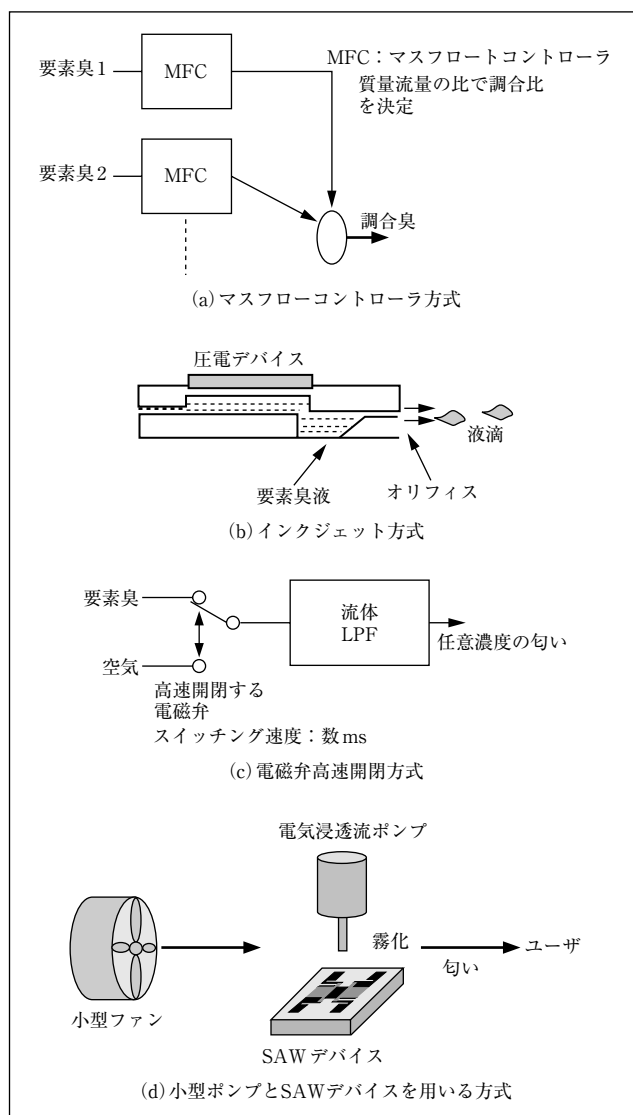


図1 多成分調合の方法



決定できる。マスフローコントローラは、流量を正確に制御できるが価格が高いという問題点がある。

図1(b)は、インクジェット方式を用いた匂い発生装置である⁷⁾。インクジェット素子は、ピコリットルオーダの微小な液滴を吐出することが可能なので、匂いの調合に利用することも可能である。複数の匂いの調合には、複数のインクジェット素子を用いて、吐出した液滴を気化させればよい。液滴の容量が十分に小さくない場合には、別途ヒータ等で気化させる必要がある。残り香が問題となる低揮発性香気成分の発生方法として有望であるが、扱いにノウハウを要する。

図1(c)は、電磁弁のみを用いて匂いの調合を行う方式である。電磁弁は流体のスイッチングデバイスであり、ON/OFFの状態のみを有してマスフローコントローラのような中間的な状態を持たない¹⁷⁾。この電磁弁のON/OFFは、デジタル回路の1と0と同様に考えることができる。そこで、アナログ信号を1ビットのデジタル信号に変換可能な $\Delta\Sigma$ 変調方式やPWM(Pulse Width Modulation)方式により電磁弁の開閉制御を行えば、任意の濃度の匂いを発生させることができるようになる。当初は、 $\Delta\Sigma$ 変調方式を用いていたが、筆者らはPWMを多成分用に拡張した方式に変えて、効率よく多成分調合できるようにした¹⁸⁾。この方式を用いるためには、濃度変化よりも高速に電磁弁を開閉制御することが必要であるが、数十ms程度のスイッチング速度で十分に滑らかな濃度変化が得られる。電子回路ではスイッチング雑音はLPF(Low Pass Filter)で除去する。この方式の場合は、配管の残留容量等が流体的なLPFとして働くためにスイッチング雑音は現れない。電磁弁は安価なので多数用いて多成分調合装置を構成するために適している。筆者らは、13-32成分を調合する装置をすでに実現している。最新版の嗅覚ディスプレイの写真を、図2に示す。この装置は13成分の調合が可能で、現在いくつかの大学や企業でもこの装置がすでに使われている。

図1(d)は、電気浸透流ポンプとSAW(Surface

Acoustic Wave) デバイスを用いた方式である¹⁹⁾。小型な電気浸透流ポンプで液滴をSAWデバイスの上に滴下する。そうすると、acoustic streaming現象により液滴が霧化され、瞬時に香りが提示される。複数の香りの調合には複数のポンプを使用する。上記の電磁弁高速開閉方式では、低揮発性香気成分を用いると配管に香りが残留する問題があった。本方式は、配管を用いないために瞬時に香りを提示することができ、残り香の問題も生じない。電気浸透流ポンプは小型で自己吸引可能で気泡に強いという特長がある。

また、リアルタイム性は低いが、液体クロマトグラフ用オートサンプラを使えば、大量の液体試料の調合も可能である²⁰⁾。

3.2 要素臭の選択方法

任意の比率で香りを調合する装置があったとしても、調合するもととなる要素臭はどのように用意すればよいのであろうか。Amooreは、匂い分子の形状を分類して7原臭説を唱えたが、実際はもっと多くの要素臭が必要と思われる²¹⁾。筆者らは、質量分析器で大量のデータを取得して、得られたマスペクトルから基底ベクトルを抽出し、その基底ベクトルに相当する要素臭を実際に存在する香りを調合して、近似的に作成する方法を提案している²²⁾。官能検査で大量のデータが取得できればそれが最良の方法であるが、実際は困難である。質量分析器は比較的短時間に大量のデータを安定に測定できるので、そのデータを解析することにした。さらにマスペクトルには、線形重ね合わせが成り立つ特徴がある。

図3に、要素臭を決める手順を示す。まず、数多くのサンプルのマスペクトルを測定し、そのデータをNMF(Nonnegative Matrix Factorization)法を用いて基底ベクトルを抽出する²³⁾。ここで要素臭の数を与える必要があり、要素臭の数を変えて求めた基底ベクトルで近似後のサンプルの残差平方和を算出して、最適な要素臭の数を決定する必要がある。NMF法は、マスペクトルのすべてのベクトル要素は非負であり、求める調合比も非負であると

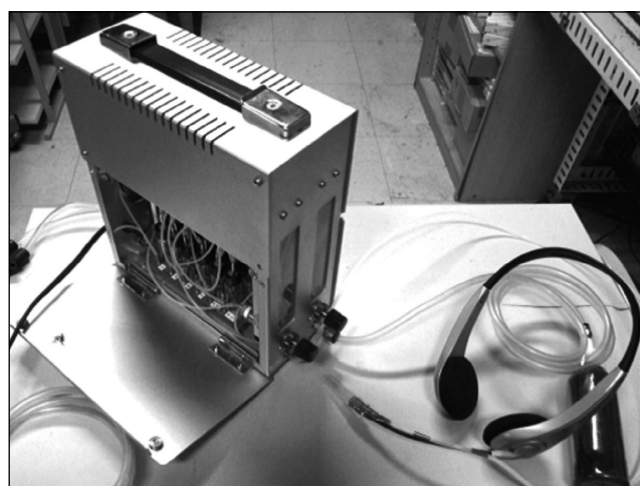


図2 電磁弁高速開閉方式嗅覚ディスプレイの外観（小野電機製）

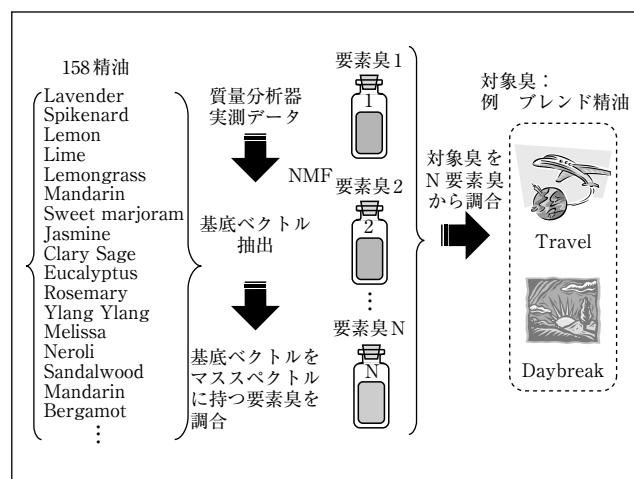


図3 質量分析器データを用いて要素臭を決める手順

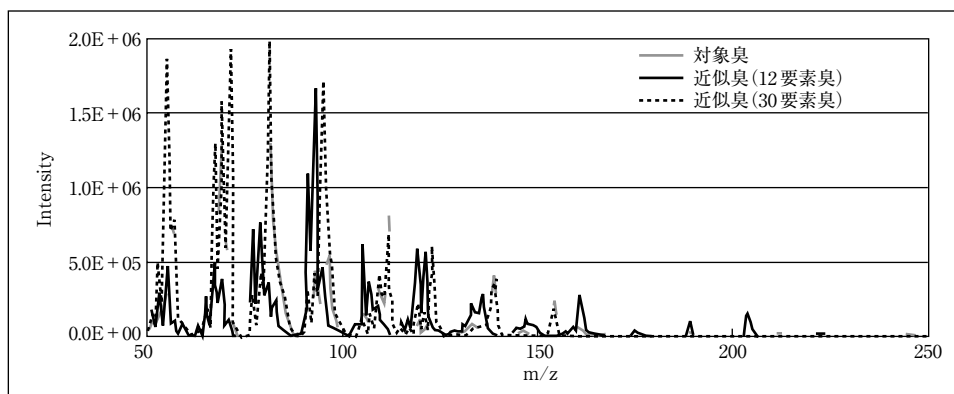


図4 オリジナルの香りと近似臭のマススペクトル (サンプル：ペパーミント)

いう拘束があるために用いた。多変量データ解析でよく用いられる主成分分析²⁴⁾は非負拘束がないので、ここでは使用できない。

ここで得られた基底ベクトルはデータから導かれたもので、実際に基底ベクトルと同じマススペクトルを持つサンプルが存在する保証はない。そこで、測定したサンプルのマススペクトルから非負拘束最小二乗法を用いて近似的に基底ベクトルを作成し、その時の調合比から要素臭を作成する。いったん、要素臭を作成しておけば調合比を変えることで多様な香りを作成することができる。そして、対象臭のマススペクトルを測定して再度非負拘束最小二乗法を適用すれば、要素臭の調合比を決定できる。

マススペクトルには一万から十数万種類の化合物に関するデータベースが存在し、当初はこのデータベースのデータを用いたシミュレーションで検討していた。しかし、実際の匂いは単一化合物ではなく、多数の化合物が複合されたものである。そこで、実際に158種類の精油²⁵⁾を測定して、そのデータから要素臭を作成することにした²⁶⁾。要素臭の数は12と30の場合を検討した。

対象臭と要素臭で近似したマススペクトルの例を図4に示す。図4の横軸は m/z (質量電荷比)、縦軸は検出器出力を表す。この時の対象臭はペパーミントで、要素臭の数は30と12の場合である。図4のように、12要素臭では対象臭と近似臭にずれが存在するが、30要素臭の場合には両者はほぼ重なっており、マススペクトルの近似精度が高いことがわかる。次に、オーガニックオレンジ、ブラックペッパー、ペパーミントの3種類の香りを選び出し、オリジナルの香りの場合と近似臭の場合に分類できるかどうかを、172人を対象に大学祭の機会を利用して調べた結果を表1に示す。表1(a)に示すように、オリジナルの香りではほとんどの人が間違えることなく分類することができている。表1(b)でも、9割前後の人が正しく分類することができおり、要素臭から近似的にこれらの香りを再現することができた。しかし、要素臭の数が12の場合はブラックペッパーとペパーミントを混同する人が多く、正しく分類できているとは言えなかった。マススペクトルの残差平

表1 3種類の香りの官能検査による分類結果

オリジナル精油			
	識別された精油		
検査対象となる精油	オーガニックオレンジ	ペパーミント	ブラックペッパー
オーガニックオレンジ	98.68%	0.66%	0.66%
ペパーミント	0.66%	96.05%	3.29%
ブラックペッパー	0.66%	3.29%	96.05%

(a) オリジナルの香り

30要素臭による近似精油			
	識別された精油		
検査対象となる精油	オーガニックオレンジ	ペパーミント	ブラックペッパー
オーガニックオレンジ	95.39%	3.29%	1.32%
ペパーミント	2.63%	87.50%	9.87%
ブラックペッパー	1.97%	9.21%	88.82%

(b) 30種類の要素臭で近似した場合

方を調べても、要素臭の数が12では残差が大きく30になると1桁程度小さくなることから、30程度の要素臭は精油を表すのに必要であると考えられる。

さらに、これらの精油を調合して調香師が作成したブレンド精油が13種類あったので、ブレンド精油に関して、オリジナルの香りと近似臭を人が区別できるかどうかを、三点識別法²⁷⁾により調べた。三点識別法とは、二つ同じで一つ異なる試料を用意し、三つの試料の中で異なるものはどれかを被験者に選んでもらう方法である。識別率が低いほど両者の類似性が高く、識別率が33%になると両者が同じであると言える方法である。官能検査の結果、12要素臭の場合は13ブレンド精油に関して44%から89%の識別率だったのに対し、30要素臭の場合は33%から56%の識別率になり、30要素臭の場合はかなり高い精度で香りを近似することができることがわかった。

なお、匂い調合に関しては、3.2の研究に関しては嗅覚ディスプレイを用いずに精油を直接調合して実験を行った。高速電磁弁開閉方式の嗅覚ディスプレイでは、30要素臭の調合は可能であるが、本研究では質量分析器で測定される時の形態に合わせることにした。



4. 遠隔地における香りと映像の同時再現

嗅覚ディスプレイの応用としては、映画やゲームがすぐに思いつくが、これらに関しては他誌でも解説しているため⁽²⁸⁾⁽²⁹⁾、本稿では、遠隔匂い再現システムとそれを応用したゲームについて述べることにする。

遠隔匂い再現システムでは、匂いセンサと嗅覚ディスプレイの両方を用いる。匂いセンサは、特性の異なる複数センサの出力パターンをパターン認識して、匂いの種類を判別したり、濃度を定量するものである。一方で嗅覚ディスプレイは香りを人に対して提示するシステムである。両者を組合せて、映像と匂いを同時にセンシングして記録し、それらを同期して再生させる技術をすでに開発していた⁽³⁰⁾。その技術をさらに発展させて、本システムでは情報をリアルタイムで遠隔地に伝送して、映像と香りを同時に再生させた⁽³¹⁾。このようなシステムは今までまだ実現されていなかった。

システム構成図を図5に示す。対象の匂いを匂いセンシングシステムで検出すると同時にウェブカメラで撮影する。匂いセンシングシステムでは、複数の水晶振動子ガスセンサ(20MHz, AT-CUT)の出力を計測し、コンピュータ内でニューラルネットワークによりパターン認識を行い、匂い種類を判別する。ここでは、LVQ (Learning Vector Quantization) 方式のニューラルネットワークを用いた⁽³²⁾。また、あらかじめ求めておいた較正曲線により匂いの強さも算出する。この匂いデータは、webカメラで取り込んだ映像とともにインターネットを介して遠隔地に送る。遠隔地では、嗅覚ディスプレイにより映像と同期させながら匂いを再生する。体験者は遠隔地にいながら映像と香りをリアルタイムで体験することが可能である。

ゲーム参加者からは、匂いを嗅いで探索動作を遠隔で行うことができる。図5のシステムでは、回転台を回転させてその上の場所を探索できるようになっている。

上記システムを用いて、2007年頃から実演公開を行ってきた。2011年には、大学祭にて遠隔地から匂い源を探知するゲームを実施した⁽³³⁾。この大学祭では、匂いセンサのパターン認識手法としてSVM (Support Vector Machine) を用いている。ゲームの参加者は、嗅覚ディスプレイで提示される匂いを嗅ぎながら遠隔地にある回転台を操作する。回転台の特定の箇所にはターゲットの匂いが浸み込ませてあり、遠隔地から回転台を動かして匂い源の場所を特定する。浸み込ませてあるものはターゲットの匂いだけでなく、ゲーム参加者を迷わせるためにダミーとなる匂いも埋め込まれている。これらの匂いは、匂いセンサで判別されて判別させた匂いが参加者側で再生される。正しく短い時間で探知に成功した参加者が勝者ということになる。

この大学祭では、ターゲットをオレンジとしてメロンとバナナの匂いをダミーとした。図6に、参加者側からみた遠隔地の回転台の様子を示す。果物の模型が映って見えるがこれは位置を確認するためのもので浸み込ませた匂いと

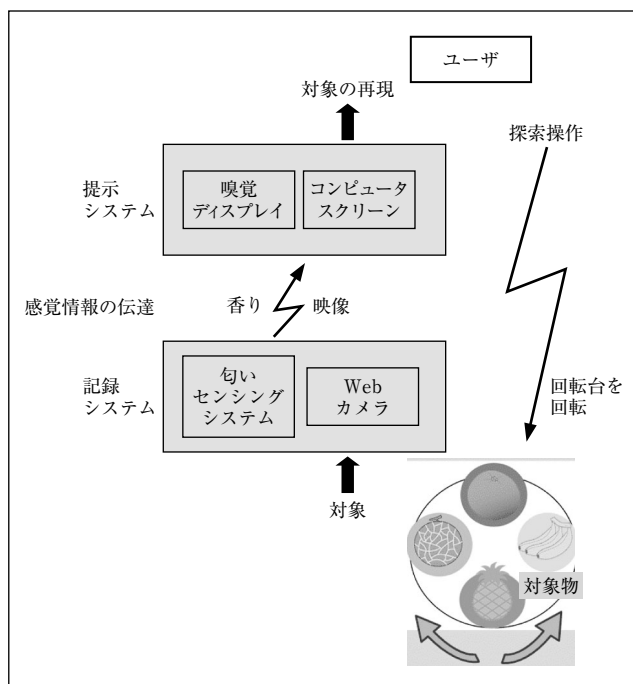


図5 遠隔匂い再現システムの構成

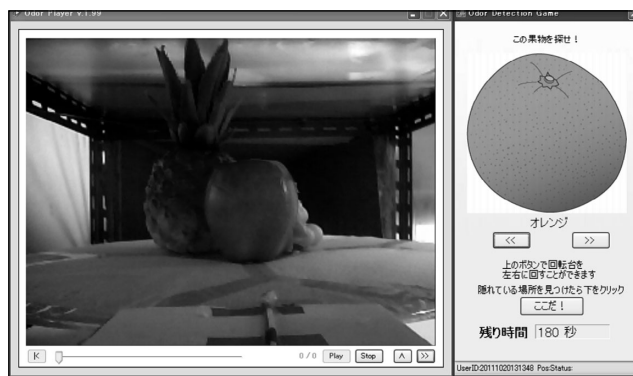


図6 遠隔匂い再現システムのユーザインタフェース

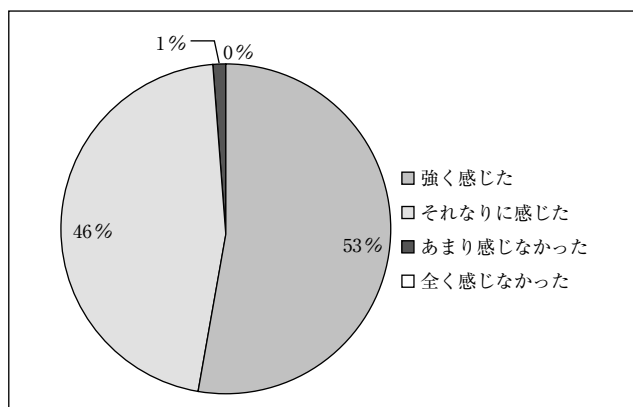


図7 遠隔匂い源探知ゲーム参加者に対する匂いの種類の違いに関するアンケート結果

は関係がない。参加者は右側のボタンをクリックすることで回転台の回転方向を遠隔制御できる。ターゲットを浸み込ませた場所から±18°を正解として、制限時間は3分以内で2回まで間違えてもよいこととしてゲームを行った。

図7に、168名のゲーム参加者に対して行ったアンケート結果の一部を示す。図7のように、9割以上に人が匂いの種類の違いを、「はっきりとわかった」、もしくは、「それなりにわかった」と答えている。なお、大学祭では参加者は説明を受けながらゲームを行ったので、探知時間についてはあまり意味をなさないと判断して割愛する。また、アンケート記述からゲーム全体は好評であったが、対戦形式にした方がよい、ゲームのストーリーが欲しいという意見もあり、今後の参考にしたい。

5. む す び

本稿では、嗅覚ディスプレイとその応用について、最近の研究成果をまじえて解説した。ハードウェアとしては、残り香にない多成分嗅覚ディスプレイを実現することがまず必要である。筆者らは、電気浸透流ポンプとSAWデバイスを用いた方法を発展させようとしている。要素臭の探索に関しては、精油からさらに多くの匂いに範囲を拡大する必要がある。また遠隔匂い再現システムでは、現在はまだ分類した匂いを遠隔地で再生するだけである。将来的には、要素臭のレシピを得て遠隔地でレシピに基づき再現させる仕組みが、システムの汎用性を増すために必要である。

嗅覚ディスプレイは、最近その名前が知られてきたとはいえまだ研究者人口が少ない。情緒まで含めた感性情報の表現に嗅覚は重要な役割を果たす。映像が専門の方々も、今後嗅覚ディスプレイを用いた研究にも進出していただければ幸いである。

最後に、本稿の内容の一部は日本学術振興会科学研究費基盤研究B (No. 22300073) の成果を含み、謝意を表する。

(2012年3月1日受付)

〔文 献〕

- 1) “匂いとVR”, 日本バーチャルリアリティ学誌, 9, pp.6-37 (2004)
- 2) 中本編著: “嗅覚ディスプレイ”, フレグランスジャーナル社 (2008)
- 3) 栗岡, 外池編: “匂いの応用工学”, 朝倉書店, p.87 (1994)
- 4) 廣瀬: “五感情報通信と嗅覚メディア”, アロマリサーチ, 3, 60-65 (2002)
- 5) 辻: “日本料理の神髄と香りの心”, フレグランスジャーナル社 (2009)
- 6) K. Murai, T. Serizawa and Y. Yanagida: "Localized Scent Presentation to a Walking Person by Using Scent Projectors", IEEE International Symposium on Virtual Reality Innovation, pp.67-70 (2011)
- 7) T. Nakamoto, H. Takigawa, T. Yamanaka: "Fundamental study of odor recorder Using inkjet devices for low-volatile scents", Trans. On IEICE, 2004, E87-C, 2081-2086 (2004)
- 8) J. Sato, K. Ohtsu, Y. Bannai, K. Okada: "Effective Presentation Technique of Scent Using Small Ejection Quantities of Odor", IEEE VR2009, pp.151-158 (2009)
- 9) 金, 三浦, 李, 柳, 西本, 川上, 国藤: “機能性高分子を用いた嗅覚ディスプレイの開発およびビデオへの応用”, 情報処理学論誌, 49, pp.160-175 (2008)
- 10) H. Kim, K. Noh, C.J. Gardner, S.D. Kong, J. Kim and S. Jin: "An X-Y addressable matrix odor-releasing system Using on-off switchable device", Angew. Chem. Int. Ed. 50, pp.1-6 (2011)
- 11) 伴野, 神田, 大竹: “映像への香り付加が誘目性と記憶に与える影響”, 電気学論誌, 128-E, pp.478-486 (2008)

- 12) T. Yamada, T. Tanikawa, K. Hirota and M. Hirose: "Wearable olfactory display: Using odor in outdoor environment", Proc. IEEE Virtual Reality, pp.199-206 (2006)
- 13) 石田, 吉田, 中本: “数値流体シミュレーションを導入した嗅覚ディスプレイ”, 電気学論誌, 128-E, pp.472-477 (2008)
- 14) 河合, 高津: “視覚情報を随伴した芳香提示システムの開発と評価”, アロマリサーチ, 3-4, pp.26-30 (2002)
- 15) T. Narumi, T. Kajinami, S. Nishizaka, T. Tanikawa and M. Hirose: "Pseudo-Gustatory Display System Based on Cross-Modal Integration of Vision, Olfaction and Gustation, IEEE Virtual Reality, 127-130 (2011)
- 16) T. Nakamoto, S. Utsumi, N. Yamashita, T. Moriizumi and Y. Sonoda: "Active Gas Sensing System Using Automatically Controlled Gas Blender and Numerical Optimization Technique, Sensors and Actuators B, 20, pp.131-137 (1994)
- 17) T. Yamanaka, R. Matsumoto, T. Nakamoto: "Study of odor blender Using solenoid valves controlled by delta-sigma modulation method", Sensors and Actuators B, 87, p.457 (2002)
- 18) T. Nakamoto, Pham Hai Dinh Minh: "Improvement of olfactory display Using solenoid valves", Proc. IEEE Virtual Reality, pp.171-178 (2007)
- 19) アーリヤクン, 中本: “超小型液体ポンプとSAW霧化器を用いた嗅覚ディスプレイの基礎的研究”, 日本VR学会, 15, pp.589-594 (2010)
- 20) P. Somboon, B. Wyszynski and T. Nakamoto: "Novel odor recorder for extending range of recordable odor", Sensors and Actuators B, 121, pp.583-589 (2007)
- 21) J.E. Amoore: "Molecular basis of odor, Charles C Thomas Publisher (1970)
- 22) T. Nakamoto and K. Murakami: "Selection method of odor components for olfactory display Using mass spectrum database, IEEE Virtual Reality, pp.159-162 (2009)
- 23) A. Cichocki, R. Zdunek, A.H. Phan and S. Amari: "Nonnegative matrix and tensor factorization", Wiley (2009)
- 24) W.R. Dillon and M. Goldstein: "Multivariate Analysis", Wiley, 1984, 23-52
- 25) 三上: “エッセンシャルオイル総覧”, フレグランスジャーナル社 (2007)
- 26) T. Nakamoto, M. Ohno and Y. Nihei: "Odor Approximation Using Mass Spectrometry", IEEE Sensors Journal, 掲載予定
- 27) 日科技連官能検査委員会編: “官能検査ハンドブック”, pp.252-253 (1973)
- 28) 中本: “多成分嗅覚ディスプレイとその応用”, 応用物理, 80, pp.231-234 (2011)
- 29) 中本: “香りはどうすれば記録・再生できるか, ベルシステム24総合研究所編, 交感する科学, pp.140-161 (2007)
- 30) T. Yamanaka, N. Nimsuk and T. Nakamoto: "Presence, Concurrent Recording and Regeneration of Visual and Olfactory Information Using Odor Sensor, 16, pp.307-317 (2007)
- 31) T. Nakamoto, N. Nimsuk, B. Wyszynski, H. Takushima, M. Kinoshita, N. Cho: "Experiment on Teleolfaction Using Odor Sensing System and Olfactory Display Synchronous with Visual Information, ICAT (International Conference on Artificial Reality and Telexistence), pp.85-92 (2008)
- 32) T. Kohonen, Self-Organization and Associative Memory, Springer-Verlag, Berlin (1988)
- 33) 米澤, B. Wyszynski, 中本: “遠隔匂い再現システムを用いた匂い源遠隔探知ゲーム”, インタラクション, IEXB-04 (2012)



中本 高道 1982年, 東京工業大学電気電子工学科卒業。1984年, 同大学院電気電子工学専攻修士課程修了。同年, 日立製作所(株)入社。1987年, 東京工業大学助手。1993年, 同大助教授となり, 現在, 同大電子物理工学専攻准教授。1996年～1997年, 米国バンフィックノースウェスト研究所客員研究員。匂いセンサ, 嗅覚ディスプレイの研究に従事。工学博士。



臨場感と迫真性

行 場 次 朗[†], 寺 本 渉^{††}

1. ま え が き

ICTやVR研究・開発の進展にともなって、今日、臨場感や高臨場感、さらに超臨場感や拡張現実感に関する議論や検討が盛んである¹⁾。それらの概念や新技術についての啓発的解説や研究成果については、たくさんの論文や報告がすでにあるので、本稿では、筆者の専門である感性・認知心理学的立場から、臨場感とはどのような心的特性に基づくのかについて検討する。そして、臨場感だけではカバーすることのできない別の感性があることを説き、それを「迫真性」と呼んで基本的概念を紹介する。さらに臨場感と迫真性の違いをうきぼりにする実験を紹介し、今後のICTやVR研究・開発への示唆を述べる。

2. 臨場感をめぐる問題点

これまで臨場感 (Sense of Presence) については、「現在いる場所とは異なる場所にいるような感覚」²⁾や、「その対象・環境を提供するメディアに対する気づきを失っている状態」³⁾などの概念的あるいは操作的定義が行われてきた。こうした考え方は重要な示唆を与え、それを実現するシステムの開発がなされ続けている。臨場感を生起させる要因は、外部要因(メディアやコンテンツ側要因)と内部要因(ユーザ側要因)に大きく分けられている⁴⁾。外部要因には、感覚情報の提示範囲、インタラクティブ性、ユーザによる環境の変更可能性⁵⁾やコンテンツの性質⁶⁾、他者とのコミュニケーション可能性³⁾などが含まれ、内的要因には、ユーザの知覚・認知能力、運動能力、パーソナリティ、年齢、性別、観察時のムードなどが含まれている。

ところが、実は、研究者間でも「臨場感とは一体どのよ

うな感覚であるのか」という根本的な問題に対して統一的な回答は得られていない³⁾。これまでの研究では、個々の研究者が事前に測定・抽出するよう決定した臨場感の特定の要素や要因が主な研究対象となっていたと考えられる。誰もが自然で快適に使用できるコミュニケーションシステムを実現するためには、多くの人が日常生活で抱えている臨場感を反映させたシステムの構築が必要である。そのためには、研究者側が定義した「臨場感」だけでなく、一般の人がイメージする「臨場感」を調査・検討することは、臨場感の定量化や高臨場感創出要因の特定のためには極めて重要な課題であると考えられる。そこで筆者ら⁷⁾は、一般の人が抱く臨場感のイメージについて、感性心理学的手法を用いて分析を行うこととした。

3. 臨場感の素朴な理解

一般の人が日常生活で「臨場感」に対してどのようなイメージを持っているのかを明らかにするために、108名の文系大学生を対象に、臨場感を感じる事象と、それらの事象の印象について意味微分法 (Semantic Differential Method, 以後SD法) を実施した⁷⁾。その結果、「好きな」、「良い」、「楽しい」などの極性をもつ形容詞対に代表される因子、「迫力のある」、「強い」、「緊張した」などの極性であらわされる因子、「動的な」、「騒がしい」などの極性を示す因子、および「人工的な」、「冷たい」、「複雑な」などの極性をもつ因子の四つが抽出され、それぞれ評価性因子、迫力因子、活動性因子、メカニック性因子と命名した。次に、感覚モダリティ微分法 (Modality Differential Method, 以後MD法)⁸⁾を用いて、同じ対象者に各自が臨場感を感じるためにはどのような感覚が重要であるかを、「まったく重要ではない」から「非常に重要である」の5段階尺度で測定した。MD法で用いた感覚モダリティは8種類であったが、聴覚、視覚、平衡感覚や身体運動感覚は有意に臨場感と関連性が高いと評定され、いわゆる近感覚群(味覚、嗅覚、痛覚、触覚)はかなり低い関連性を示した。

これらのSD法とMD法の分析結果をあわせて考えると、

[†] 東北大学

^{††} 室蘭工業大学

"Technology Frontier of Augmented Reality (7): Senses of Presence and Verisimilitude" by Jiro Gyoba (Department of Psychology, Graduate School of Arts & Letters, Tohoku University, Sendai) and Wataru Teramoto (Department of Computer Science and Systems Engineering, Muroran Institute of Technology, Hokkaido)

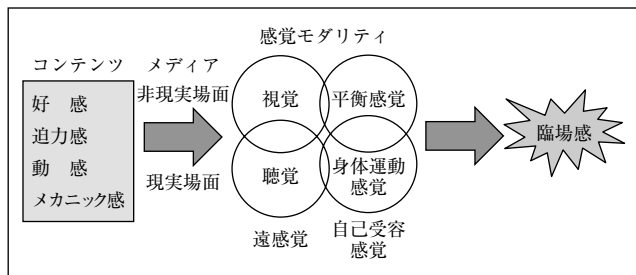


図1 臨場感にかかわる感性因子と感覚モダリティ選択性

臨場感を創出するには、「その場で実際に体験しているような感じ」という辞書的定義どおりだけでなく、また、必ずしも現実的な場面にとらわれる必要もなく、「好感」、「迫力感」、「動感」を持ち合わせた事象であることが大切であり、そしてそれらのコンテンツが、視覚や聴覚などの遠感覚、ならびに、前庭感覚や身体運動感覚などの自己受容感覚を主に介して感受されることが重要であることがわかった(図1)。したがって、今後、臨場感の定量化および高臨場感創出要因の特定するためには、このような臨場感に内包される複数の感性因子と感覚モダリティ選択性に配慮する必要があることが示唆された。

4. 古くて新しい感性概念「迫真性」の重要性

今日、ICTやVR研究・開発の目覚ましい進展にともなって、図2に示すように、伝送・提示する情報量を拡大させ、現実世界のもつ情報量に近づけて高臨場感を生み出す試みが盛んである。さらに、超臨場感や拡張現実感とも呼ばれるように、現実世界のレベルを通り超えて、日常では体験できないことを可能にする研究フォーラムも推進されている¹⁾。3章で述べた臨場感の感性心理学的分析でも明らかになったように、現実世界にとらわれない「好感」、「迫力感」、「動感」、「メカニック感」も大切であるので、超臨場感や拡張現実感のアプローチは重要である。しかし、感性心理学を専攻してきた筆者らは、あえてこのような動向と方向の異なるアプローチも重要であることを主張したい(ただし、超臨場感アプローチのうち、「超越臨場感(メタリアリティ)」とよばれる考え方は、以下に述べる迫真性の概念と一部、共通性がある)。

例えば、古くから培われてきた日本文化・様式を少し振り返ってみたい。枯山水や水墨画、侘び茶や能舞などに代表される日本美の特徴は、刺激情報を可能な限り還元し、美の中心となる感性エッセンスを抽出し、それを的確に提示することにあると考える(図2参照)。また、落語家はカゼとよばれる扇子と、マンガラとよばれる手ぬぐいなどの実に単純な道具で、蕎麦をすする情景や煙草をふかす情景などを、実世界よりも真実味をおびて見事に表現する。このような還元された情報により創出される本物らしさに対応する感性を、「迫真性(Sense of Verisimilitude)」と定義することにする。また、日本美は豊富な多感覚性ないし共感覚性を有していることも大きな特徴である。最近、刺激強度が小さいほど、

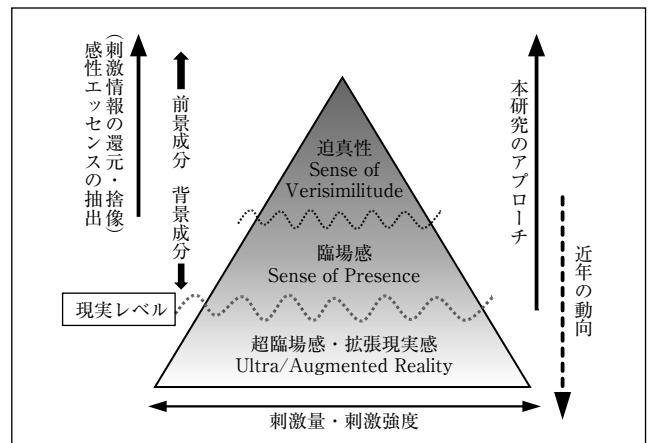


図2 迫真性および臨場感と、超臨場感の研究動向をあらわす概念図

種々の感覚間の相互作用が大きくなることが知られており、逆理的效果(Inverse Effectiveness)とよばれている⁹⁾。現実性を適切に捨象することにより刺激量を減らせば、多感覚間の相互作用がかえって強まって創発効果(Emergent Effect)を生起させ、また鑑賞者自身による補完処理(Completion Process)を励起することにより、より迫真性のあるコンテンツとして感受される可能性がある。このように迫真性は、侘び寂び、粋、洒脱、軽み、移ろい、はかなさ、未完の美などの特性とも深い関連があるといえる。クールジャパンとも総称される日本発信のコンテンツやプロダクトの一部にも共通した特性があり、映像や音響をひたすら強烈に、しかも3次元的に提供する傾向が目立つハリウッド的コンテンツと対比をなしている。

5. 異なった感性特性を示す臨場感と迫真性

臨場感にとっては、背景的な場の時空間的な広がりであるアンビエントな情報も重要である。一方、迫真性にとっては、焦点的注意が向いている対象や事象などの前景的要素の情報の方がより大切と考えられる。知覚心理学の知見では、背景的情報は「地」、前景的情報は「図」と呼ばれ、それぞれ別々の知覚システムが関与する可能性が指摘されてきた¹⁰⁾。地は媒質としての性質をもって自動的に概略処理される一方、そこから分離された図は事物としての性質をもって意識的で詳細な分析を受けるとされる。

そこで筆者らは、臨場感と迫真性では異なった感性特性が示されるはずと予想して、日本庭園における「鹿威し」と、交響楽団における「シンバル演奏」の視聴覚コンテンツを用いて、背景と前景となる対象の画角や音量を操作し、参加者に臨場感と迫真性をどれくらい感じるかを測定する感性評価実験を実施した^{11) 12)}。このときに参加者には、臨場感については「自分があたかもその場にいる感じがする」、迫真性については「本物らしい趣がある」などの教示を加えて、一対比較法や7段階尺度でそれらの強弱の判断を行ってもらった。その結果、臨場感は画角や音量が大きくなるにつれて線形に増加するのに対して、迫真性は、中程度の

画角や音量に最大値を持つ逆U字型非線形的特性を持つことがわかった。これらの中程度の画角と音量条件では、背景情報（日本庭園や楽団の包囲的な情報）よりも前景情報（鹿威しやシンバル奏者）が背景情報とより適切な比をもってクローズアップされていたと考えられる。

さらに著者らは、鹿威しとシンバル演奏に伴う視覚情報と聴覚情報のSOA (Stimulus Onset Asynchrony) を操作して同様の実験を行った^{11) 13)}。その結果、臨場感については、SOAを大きくずらしても画角や音量さえ大きければ、比較的滑らかな増減を示すのに対し、迫真性は中程度の画角や音量で、SOAに対して時間窓がより狭く、チューニングが厳しい特性を示すことがわかった。これは、視聴覚的に調和のとれた刺激が提示されている条件で、前景に適切に焦点が向けられ、時間ずれの許容範囲がより狭くなったと考えられる。

6. むすびと今後の課題

視聴覚コンテンツの臨場感は表示画角サイズが大きくなるほど、また実際に収録した音圧以上の音圧で提示することによって高まることは先行研究の報告と一致している^{14) 15)}。したがって、臨場感は、背景(地)の情報量により大きく依存し、総量としての感覚経験(クオリア)によって感受されると考えることができる。一方、迫真性は、相対的に刺激量が少ないときに、前景(図)の情報が適度にクローズアップされ、事物に対する気付き(アウェアネス)が高まった時により強く喚起されると推定される。

臨場感と迫真性の両方の感性を鑑賞者が求めていることは、以下のような行動でも知ることができる。例えば、野球やサッカーなどの競技場あるいは大画面・サラウンド音響のライブ中継会場に小型テレビや携帯電話などを持ち込んで観戦するケースで、臨場感を競技場や大画面で実際に感じながらも、観客個々人が望むここぞという決定的シーンを携帯端末でズームアップやスローで再生して、それらのコンテンツの迫真性をじっくり何度も味わおうとする。現在普及している携帯情報端末でも、通信容量や画面サイズは限られているものの、後者のような目的には適合していると考えられるので、配信コンテンツを作成する際にも臨場感と迫真性を別に考えて補完しあう工夫をする必要がある。また、クールジャパンと呼ばれるアニメの一部には、3D大画面・サラウンド音響装置などよりは、迫真性の原理を取り入れた小型・少容量携帯情報端末のほうがよりコストパフォーマンス良く、的確に表現できる期待も大きい。

今後、臨場感と迫真性の感性概念をより明確化するとともに、両者の相補性や交互作用を支える表示技術、表現方法、効果量の測定方法などをさらに整備していく必要がある。

最後に、ここで紹介した臨場感や迫真性にかかわる筆者らの研究は、すべて科学研究費補助金特別推進研究(19001004)の助成を受けた。共同研究者の吉田和博、神田敬幸、本多明生、柴田寛、日高聡太、浅井暢子、坂本

修一、岩谷幸雄、鈴木陽一の各氏に感謝したい。

(2012年4月4日受付)

〔文 献〕

- 1) 榎並和雅, 岸野文郎: “今後の超臨場感にかかわる研究はどこを目指すべきか”, 信学誌, **93**, pp.363-367 (2010)
- 2) B.G. Witmer and M.J. Singer: "Measuring presence in virtual environments: A presence questionnaire", Presence: Teleoperators & Virtual Environments, **7**, pp.225-240 (1998)
- 3) M. Lombard and T. Ditton: "At the heart of it all: the concept of presence", Journal of Computer Mediated Communication, **3**, pp.1-43 (1997)
- 4) M. Slater, A. Steed, J. McCarthy and F. Maringelli: "The influence of body movement on presence in virtual environments", Human Factors, **40**, pp.469-477 (1998)
- 5) T.B. Sheridan: "Musings on telepresence and virtual presence; Presence: Teleoperators & Virtual Environments, **1**, pp.120-126 (1992)
- 6) J. Lessiter, J. Freeman, E. Keogh and J. Davidoff: "A cross-media presence questionnaire: the ITC-Sense of Presence Inventory", Presence: Teleoperators & Virtual Environments, **10**, pp.282-298 (2001)
- 7) 寺本渉, 吉田和博, 浅井暢子, 日高聡太, 行場次朗, 鈴木陽一: “臨場感の素朴な理解”, 日本バーチャルリアリティ学論誌, **15**, pp.7-16 (2010)
- 8) 鈴木美穂, 行場次朗, 川畑秀明, 山口浩, 小松紘: “モダリティ・ディファレンシャル法による形容詞対の感覚関連性の分析”, 心理学研究, **77**, pp.464-470 (2006)
- 9) D. Senkowski, D. Saint-Amour, M. Hofle, J.J. Foxe: "Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness", NeuroImage, **56**, pp.2200-2208 (2011)
- 10) 行場次朗: “図と地の知覚－視覚の心理－”, 信学誌, **74**, pp.315-320 (1991)
- 11) 寺本渉, 吉田和博, 日高聡太, 浅井暢子, 行場次朗, 坂本修一, 岩谷幸雄, 鈴木陽一: “「迫真性」を規定する時空間情報”, 日本バーチャルリアリティ学論誌, **15**, pp.483-486 (2010)
- 12) 神田敬幸, 本多明生, 柴田寛, 浅井暢子, 寺本渉, 坂本修一, 岩谷幸雄, 行場次朗, 鈴木陽一: “視聴覚コンテンツの臨場感と迫真性に関する実験的研究－シンバル演奏場面における視野制限と音圧レベルの効果－”, 信学技報111, HIP2011-44, pp.7-10 (2011)
- 13) T. Kanda, A. Honda, H. Shibata, Y. Asai, W. Teramono, S. Sakamoto, Y. Iwaya, J. Gyoba and Y. Suzuki: "The sense of verisimilitude has different spatial-temporal characteristics from those producing the sense of presence in the evaluation process of audiovisual contents", i-Perception, **2**, pp.807 (2011)
- 14) 江本正喜, 正岡顕一郎, 菅原正幸, 野尻祐司: “広視野静止画像による臨場感の提示視覚依存性と評価指標間の関係”, 映情学誌, **60**, pp.1288-1295 (2006)
- 15) 小澤賢司: “聴覚臨場感の基礎特性－コンテンツ臨場感とシステム臨場感－”, 信学技報, **108-333**, pp.83-88 (2008)



行場 次朗 1981年、東北大学大学院文学研究科博士課程心理学専攻修了。信州大学人文学部助教授、九州大学文学部助教授を経て、2002年より、東北大学大学院文学研究科心理学講座教授。専門は感覚知覚心理学・認知心理学・感性心理学。博士(文学)。



寺本 渉 2004年、神戸大学大学院文化学研究科博士課程修了。日本学術振興会特別研究員(PD)、東北大学電気通信研究所・同大学院文学研究科研究員を経て、2011年より、室蘭工業大学情報電子工学系学科助教。知覚心理学、実験心理学、特にマルチモーダル感覚情報処理に関する研究に従事。博士(学術)。



音響の高臨場感技術

正会員 安藤 彰 男[†]

1. ま え が き

本稿では、高臨場感音響を実現するための技術の中で、拡張現実感にも利用可能な技術、すなわち、コンピュータが生成した音を現実の空間の中に表示できる技術¹⁾を概観する。近年は、録音から再生に至るあらゆるプロセスでコンピュータ技術が活用されている。したがって、「コンピュータが生成した音」には、電子音だけでなく、自然音や楽音なども含むこととする。

音を現実空間の中に表示するためには、音の空間性に配慮する必要がある。音の空間性を表す要素として、音源 (Sound Source) と環境 (Environment) が挙げられる²⁾。音源は特定の位置に知覚可能なものである。その位置を知覚することを一般に音源あるいは音像定位 (Sound Localization) と呼ぶ。環境は、音像としてその位置を知覚しうる音ではなく、空間的な広がりとして知覚されるもので、このような知覚は音による包み込まれ感 (Sound Envelopment) と呼ばれる。また、包み込まれ感がある方向に限定される場合には、単に音の広がり感と呼ぶことにする。これらの知覚を適切にもたらす技術、言い換えれば、所望の位置に音像を定位させたり、所望の包み込まれ感を実現したりする技術のことを、本稿では、空間音響技術と呼ぶ。また、音源や環境も含めた空間的な音の場を、音場 (Sound Field) と呼ぶ。

空間音響技術は大雑把に分類すると、心理音響に基づく技術と、物理音響に基づく技術に分けることができる。前者は、所望の音の知覚を実現する技術である。2chステレオの拡張に基づくマルチチャンネル音響が、前者に含まれる。一方後者は、所望の音場の物理量なるべく損なわずに実現するための技術である。

2. 心理音響に基づく高臨場感技術³⁾

2.1 2chステレオ

2chステレオでは、前方に置いた二つのスピーカを用いて、音源定位や音の広がり感を実現する。音像定位や広がり感は録音技術によって実現可能であるが、さらにコンテンツを作り込む場合には、音像定位は振幅パンニング (Amplitude Panning) で制御され、音の広がり感は残響付加などで実現される。2chステレオにおける振幅パンニングとは、目的音の信号を、強度差を付けて左右のチャンネルに分配する方法である。一般のマルチチャンネル音響でも、この方法で音像定位を制御する。このように定位された音像を、“仮想音像”と呼ぶ。以下、特に振幅パンニングについて概説する。

一般に、人間の音の方向定位の知覚は、左右の耳における音の強度差、位相差 (時間差)、音色の違いの三つの要素によって決まる。第3の要素は、耳介での反射や頭の周りの回折などによって生じるものである。2chステレオの場合、例えば、Lchの信号は右耳にも到達するため、振幅パンニングによるチャンネル間の強度差が、そのまま両耳の強度差となる訳ではない。振幅パンニングに基づく2ch間の強度差は、両耳の位置において、二つのスピーカからの音波の加算によって得られる位相差⁴⁾と、頭の周りの音の回折に起因する強度差を生み出す。このメカニズムを図1に示す。

図1に示すように、ある信号 $s(t)$ が、重み係数 w_L と w_R によって、左右のスピーカに分配される場合を考える。左スピーカから左耳に音波が到達する時間と、同スピーカから右耳に音波が到達する時間の差を 2τ とすると、右スピーカからの音波についても、両耳の位置で同様な時間差を仮定することができる。左(右)スピーカから右耳(左耳)に伝搬する場合には、頭による遮蔽効果によって音波の減衰が起こる。その減衰率を α とする。この減衰は、低域では音の回折によって補われるため無視できる。左スピーカへの入力信号と、左耳での音圧 (音による大気圧の変動分) の比例係数を G とすると、周波数 ω の正弦波信号を左スピーカに

[†] NHK放送技術研究所

"Technology Frontier of Augmented Reality (8): Sound Technology Providing Strong Sense of Presence" by Akio Ando (NHK Science and Technology Research Laboratories, Tokyo)

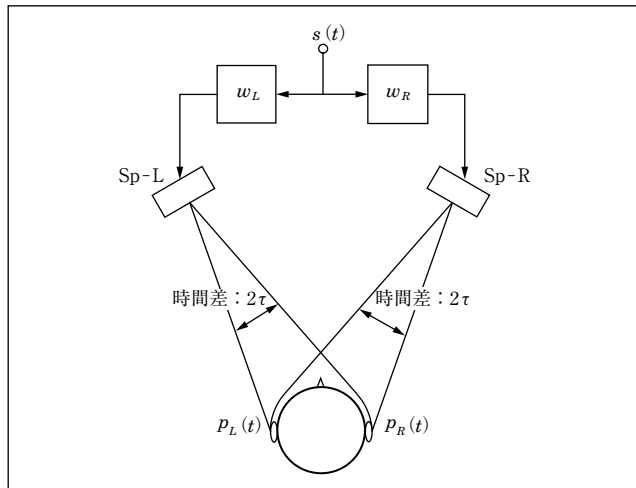


図1 振幅パンニングにおける音の伝搬

入力したときの左右の耳における音圧 $p_L(t)$ と $p_R(t)$ は、

$$\begin{aligned}
 p_L(t) &= Gw_L \sin \{ \omega(t + \tau) \} + G\alpha w_R \sin \{ \omega(t - \tau) \} \\
 &= G \sqrt{w_L^2 + \alpha^2 w_R^2 + 2\alpha w_L w_R \cos 2\omega\tau} \\
 &\quad \sin \left\{ \omega t + \tan^{-1} \left(\frac{w_L - \alpha w_R}{w_L + \alpha w_R} \tan \omega\tau \right) \right\} \\
 p_R(t) &= G\alpha w_L \sin \{ \omega(t - \tau) \} + Gw_R \sin \{ \omega(t + \tau) \} \\
 &= G \sqrt{\alpha^2 w_L^2 + w_R^2 + 2\alpha w_L w_R \cos 2\omega\tau} \\
 &\quad \sin \left\{ \omega t - \tan^{-1} \left(\frac{\alpha w_L - w_R}{\alpha w_L + w_R} \tan \omega\tau \right) \right\}
 \end{aligned} \quad (1)$$

と表される。頭部の遮蔽効果による音の減衰が無視できる中低域の周波数では、 α はほぼ1となるため、左右の耳における音の強度差は無視できる。一方、位相差は、

$$\phi \approx 2 \tan^{-1} \left(\frac{w_L - w_R}{w_L + w_R} \tan \omega\tau \right) \quad (2)$$

で与えられる。これに対して、高域では、音の位相差は知覚困難となる一方、頭部の遮蔽効果による音の減衰が無視できなくなるため、音像定位は、音のレベル差

$$\begin{aligned}
 &10 \times \log (w_L^2 + \alpha^2 w_R^2 + 2\alpha w_L w_R \cos 2\omega\tau) \\
 &- 10 \times \log (\alpha^2 w_L^2 + w_R^2 + 2\alpha w_L w_R \cos 2\omega\tau)
 \end{aligned} \quad (3)$$

によって知覚される。

方向定位知覚においては、少なくとも700Hzより低い周波数帯域では位相差が支配的となり、高い周波数になるほど強度差が支配的となることが知られている²⁾。人間は、この位相差や強度差により、仮想音像の定位を知覚するが、前述の3要素を同時に満足する実際の音源と比べ、この定位知覚は錯覚に近いものと考えられている。

2chステレオにおける代表的な振幅パンニング法として、サイン則 (Sinus Law) とタンゼント則 (Tangent Law) が知られている。振幅パンニングのブロック図を図2に示す。図2において、 ϕ_0 は正面から各スピーカまでの角度であり、 ϕ は正面から仮想音像までの角度である。サイン則では、角度 ϕ に音を定位させるために、左右のスピーカに分配する信

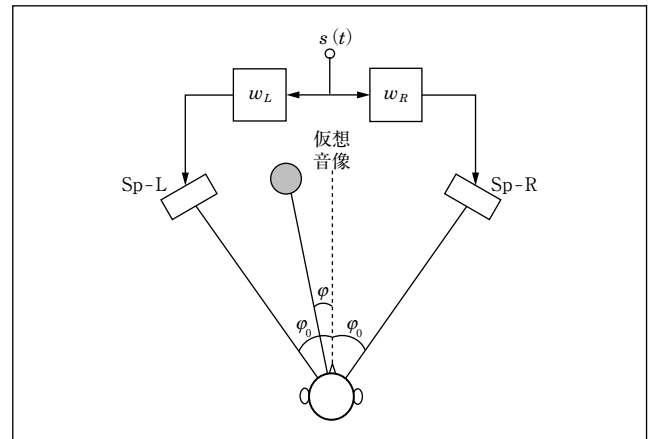


図2 振幅パンニングによる仮想音像の生成

号の利得を、

$$\frac{\sin \phi}{\sin \phi_0} = \frac{w_L - w_R}{w_L + w_R} \quad (4)$$

にしたがって定める。一方、タンゼント則では、利得を

$$\frac{\tan \phi}{\tan \phi_0} = \frac{w_L - w_R}{w_L + w_R} \quad (5)$$

にしたがって定める。サイン則は、頭を固定した場合の音像定位モデルであり、タンゼント則は、仮想音像の方向に頭を向けたときの音像定位モデルと考えられている⁵⁾。

2.2 3次元マルチチャネル音響

近年、2chステレオの再生能力を3次元方向に拡大したマルチチャネル音響の開発が進んでいる。このような3次元マルチチャネル音響では、音像定位を制御するため、3次元音響パンニング法が利用される。以下、代表的な3次元音響パンニング法である、VBAP (Vector Based Amplitude Panning)⁶⁾について解説する。

VBAP法では、再生空間を、3個のスピーカからなる三角領域で分割する。そして、各三角領域において、音源信号を、重み係数 g_1, g_2, g_3 によって各スピーカに分配することにより、振幅パンニングを行う。図3に、VBAP法の概念図を示す。受音点を原点とし、仮想音像方向に向いた単位ベクトルを $\mathbf{p} = (p_1 \ p_2 \ p_3)^T$ で表す。ここに、 T はベクトルまたは行列の転置を表す。また、原点から三つのスピーカ方向への単位ベクトルを、それぞれ、 $\mathbf{l}_1 = (l_{11} \ l_{12} \ l_{13})^T$, $\mathbf{l}_2 = (l_{21} \ l_{22} \ l_{23})^T$, $\mathbf{l}_3 = (l_{31} \ l_{32} \ l_{33})^T$ とする。このとき、ベクトル $\mathbf{p}$ を、スピーカ方向ベクトルを基底として

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2 + g_3 \mathbf{l}_3 \quad (6)$$

あるいは、係数ベクトル $\mathbf{g} = (g_1 \ g_2 \ g_3)^T$ によって、

$$\mathbf{p} = \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \mathbf{g} \quad (7)$$

と表す。スピーカが受音点から等距離に配置された場合、この距離を単位距離とすると、ベクトル $\mathbf{p}, \mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3$ は受

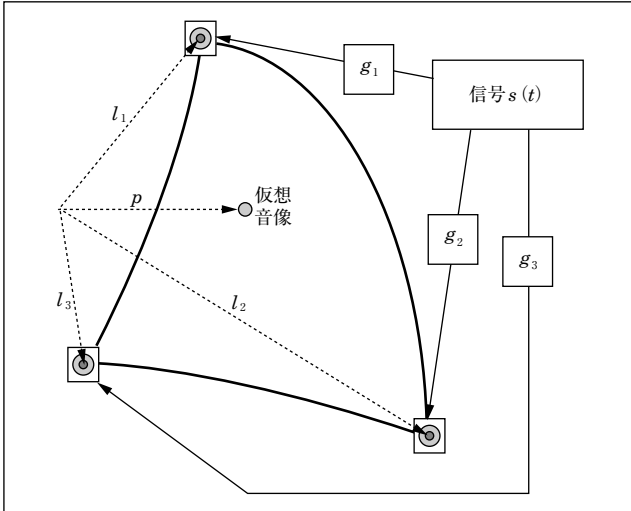


図3 VBA3次元パンニング

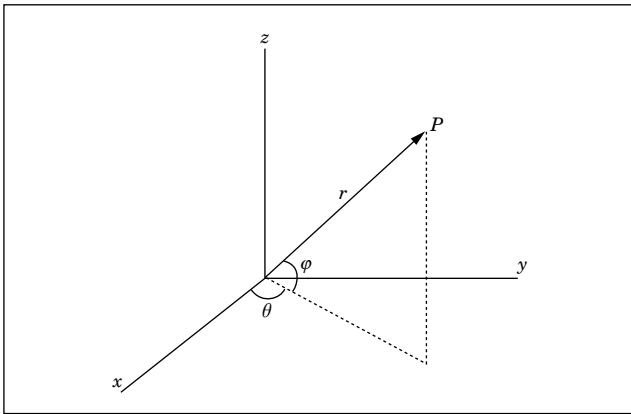


図4 極座標系

音点を原点とする極座標系(図4)により,

$$l_1 = (\tilde{\theta}_1, \tilde{\varphi}_1), l_2 = (\tilde{\theta}_2, \tilde{\varphi}_2), l_3 = (\tilde{\theta}_3, \tilde{\varphi}_3), \\ p = (\theta, \varphi)$$

と表される。このとき、式(7)は,

$$\begin{pmatrix} \cos \tilde{\theta}_1 \cos \tilde{\varphi}_1 \cos \tilde{\theta}_2 \cos \tilde{\varphi}_2 \cos \tilde{\theta}_3 \cos \tilde{\varphi}_3 \\ \sin \tilde{\theta}_1 \cos \tilde{\varphi}_1 \sin \tilde{\theta}_2 \cos \tilde{\varphi}_2 \sin \tilde{\theta}_3 \cos \tilde{\varphi}_3 \\ \sin \tilde{\varphi}_1 \sin \tilde{\varphi}_2 \sin \tilde{\varphi}_3 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} \\ = \begin{pmatrix} \cos \theta \cos \varphi \\ \sin \theta \cos \varphi \\ \sin \varphi \end{pmatrix} \quad (8)$$

と変形される。式(8)の解は,

$$g_1 = \frac{D_1}{D}, g_2 = \frac{D_2}{D}, g_3 = \frac{D_3}{D} \quad (9)$$

で表される。ここに,

$$D = \sin(\tilde{\theta}_2 - \tilde{\theta}_1) \cos \tilde{\varphi}_2 \cos \tilde{\varphi}_1 \sin \tilde{\varphi}_3 \\ + \sin(\tilde{\theta}_3 - \tilde{\theta}_2) \cos \tilde{\varphi}_3 \cos \tilde{\varphi}_2 \sin \tilde{\varphi}_1 \\ + \sin(\tilde{\theta}_1 - \tilde{\theta}_3) \cos \tilde{\varphi}_1 \cos \tilde{\varphi}_3 \sin \tilde{\varphi}_2 \\ D_1 = \sin(\tilde{\theta}_3 - \tilde{\theta}_2) \cos \tilde{\varphi}_3 \cos \tilde{\varphi}_2 \sin \tilde{\varphi} \\ + \{\sin(\tilde{\theta} - \tilde{\theta}_3) \cos \tilde{\varphi}_3 \cos \tilde{\varphi}_2 \\ - \sin(\tilde{\theta} - \tilde{\theta}_2) \cos \tilde{\varphi}_2 \sin \tilde{\varphi}_3\} \cos \varphi \\ D_2 = \sin(\tilde{\theta}_1 - \tilde{\theta}_3) \cos \tilde{\varphi}_1 \cos \tilde{\varphi}_3 \sin \tilde{\varphi} \\ + \{\sin(\tilde{\theta} - \tilde{\theta}_1) \cos \tilde{\varphi}_1 \cos \tilde{\varphi}_3 \\ - \sin(\tilde{\theta} - \tilde{\theta}_3) \cos \tilde{\varphi}_3 \sin \tilde{\varphi}_1\} \cos \varphi \\ D_3 = \sin(\tilde{\theta}_2 - \tilde{\theta}_1) \cos \tilde{\varphi}_2 \cos \tilde{\varphi}_1 \sin \tilde{\varphi} \\ + \{\sin(\tilde{\theta} - \tilde{\theta}_2) \cos \tilde{\varphi}_2 \cos \tilde{\varphi}_1 \\ - \sin(\tilde{\theta} - \tilde{\theta}_1) \cos \tilde{\varphi}_1 \sin \tilde{\varphi}_2\} \cos \varphi$$

である。式(9)は、タンゼント則⁵⁾の3次元への拡張となっている。

3. 物理音響に基づく高臨場感音響

3.1 回折理論に基づく音場再現

3.1.1 キルヒホッフ-ヘルムホルツの積分定理に基づく音場再現

波の伝搬や回折を説明する原理として、ホイヘンス-フレネル(Huygens-Fresnel)の原理が知られている。この原理は、図5に示すように、波源(1次波源)から伝搬する波の波面上の2次波源から2次波面が生成され、2次波面の包絡によって次の波面が生成されるという原理である。この原理を、数学的に表現したのが、キルヒホッフ-ヘルムホルツ(Kirchhoff-Helmholtz)の積分定理である⁷⁾⁸⁾。

音波の到来している場(音場)を音圧 $p(\mathbf{r}, t)$ で表す。ここに、 $\mathbf{r}$ は音場内の位置、 t は時間である。また、音圧 $p(\mathbf{r}, t)$ のフーリエ変換を、 $p(\mathbf{r}, \omega)$ で表す。以下、フーリエ変換した後の周波数領域を用いて音場を記述する。この音場

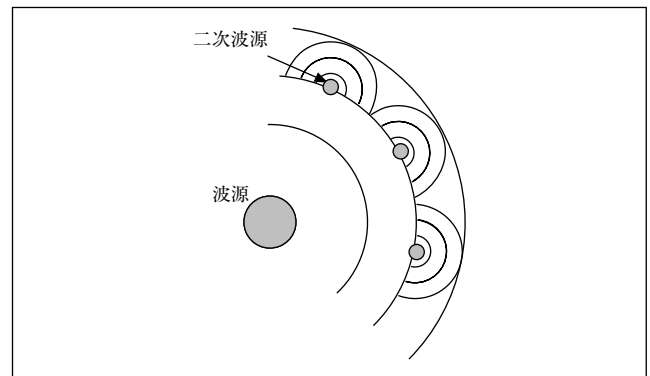


図5 ホイヘンス-フレネルの原理

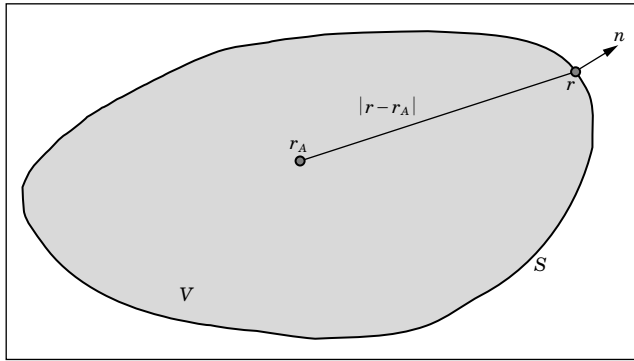


図6 音場内の領域V

内に、図6のような領域Vを考える。図6において、 S はVを囲む閉曲面、 $\mathbf{r}_A$ はV内の任意の点、 $\mathbf{n}$ は S 上の点 $\mathbf{r}$ における外向き法線方向の単位ベクトルである。このとき、キルヒホッフヘルムホルツの積分定理は、

$$p(\mathbf{r}_A, \omega) = \frac{1}{4\pi} \int_S \left(\frac{e^{-ik|\mathbf{r}-\mathbf{r}_A|}}{|\mathbf{r}-\mathbf{r}_A|} \frac{\partial p(\mathbf{r}, \omega)}{\partial n} - p(\mathbf{r}, \omega) \frac{\partial}{\partial n} \left(\frac{e^{-ik|\mathbf{r}-\mathbf{r}_A|}}{|\mathbf{r}-\mathbf{r}_A|} \right) \right) dS \quad (10)$$

で表される。ここに、 i は虚数単位、 $k=c/\omega$ は波長定数、 c は音速、 ω は角周波数、 $\partial/\partial n$ は法線方向($\mathbf{n}$ 方向)の偏微分を表す。式(10)において、関数

$$\frac{e^{-ik|\mathbf{r}-\mathbf{r}_A|}}{|\mathbf{r}-\mathbf{r}_A|} \quad (11)$$

は、モノポール関数と呼ばれ、音源の大きさが波長に比べて充分小さい場合に、点 $\mathbf{r}_A$ におかれた全指向性音源による音波をよく近似する関数である。また、

$$\frac{\partial}{\partial n} \left(\frac{e^{-ik|\mathbf{r}-\mathbf{r}_A|}}{|\mathbf{r}-\mathbf{r}_A|} \right) \quad (12)$$

はダイポール関数と呼ばれ、音源の大きさが波長に比べて充分小さい場合に、点 $\mathbf{r}_A$ におかれた両指向性音源(主軸は $\mathbf{n}$ 方向)による音波を表す関数である。

式(10)の被積分項は、境界上の点 $\mathbf{r}$ に置かれた二つの2次音源、すなわち音圧勾配 $\partial p(\mathbf{r}, \omega)/\partial n$ に等しい強さを持つ全指向性音源と、音圧 $-p(\mathbf{r}, \omega)$ に等しい強さを持つ両指向性音源による $\mathbf{r}_A$ 点における音圧を表している。この音圧を境界面 S 上のすべての点 $\mathbf{r}$ で積分することにより、点 $\mathbf{r}_A$ における音圧を求めることができる。

キルヒホッフヘルムホルツの積分定理は、コンサートホールなどの原空間に設定された受信領域境界で音圧と音圧勾配を観測し、これらをスタジオなど別空間の境界面上に再現することによって、原音場をそのまま再現できることを示している。

3.1.2 レイリー積分に基づく音場再現

図6の境界面を平面 S_1 と部分球面 S_2 とに分割した後(図7

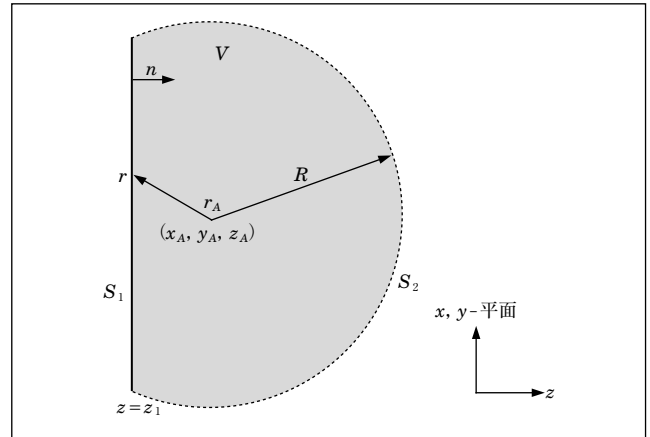


図7 図6の境界面の変形

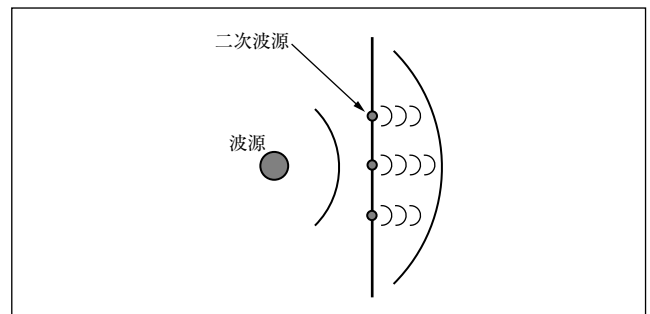


図8 レイリー積分に基づく波面合成

参照)、 S_2 の半径 R を無限大にすることによって、キルヒホッフヘルムホルツ積分は、レイリー(Rayleigh)積分に変形される⁸⁾。第1種レイリー積分は、

$$p(\mathbf{r}_A, \omega) = -\frac{1}{2\pi} \int_{S_1} \frac{\partial}{\partial z} (p(\mathbf{r}, \omega)) \frac{e^{-ik|\mathbf{r}-\mathbf{r}_A|}}{|\mathbf{r}-\mathbf{r}_A|} dS \quad (13)$$

で表され、第2種レイリー積分は、

$$p(\mathbf{r}_A, \omega) = \frac{1}{2\pi} \int_{S_1} p(\mathbf{r}, \omega) \frac{\partial}{\partial z} \left(\frac{e^{-ik|\mathbf{r}-\mathbf{r}_A|}}{|\mathbf{r}-\mathbf{r}_A|} \right) dS \quad (14)$$

で表される。ここで、法線方向 $\mathbf{n}$ と z 軸方向が等しい方向であるため、 ∂n を ∂z で置き換えている。図8にレイリー積分を用いた波面合成の例を示す⁹⁾。

第1種レイリー積分は、平面上に配置した両指向性マイクロホンによって音圧勾配 $\partial p(\mathbf{r}, \omega)/\partial z$ を観測し、この観測信号を用いて、モノポール関数で記述される2次波源を駆動することで原音場を再現する。一方、第2種レイリー積分は、平面上に配置した全指向性マイクロホンによって音圧 $p(\mathbf{r}, \omega)$ を観測し、この観測信号を用いて $-z$ 軸に主軸を持つダイポール関数で記述される2次波源を駆動して、原音場を再現する。

3.1.3 Wave Field Synthesis

レイリー積分では、原音場の平面 S_1 上で音圧あるいは音

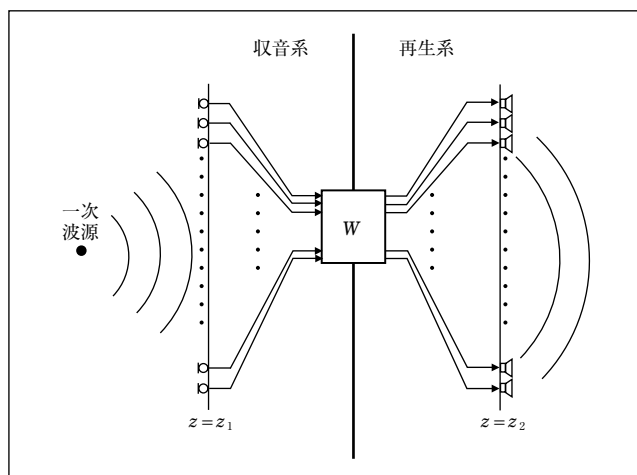


図9 WFSによる音場合成

圧勾配を観測し、観測した信号で別の空間における同じ平面 S_1 に配置したダイポール2次波源あるいはモノポール2次波源を駆動して、原音場と同じ音場を再現するものであった。ここで、原音場と再生空間との関係を切り離したらどうなるだろう。例えば、 S_1 とは異なる平面 S_2 上に2次波源を配置して音場を構成する場合や、実在する原音場の代わりに仮想空間を考えて、仮想空間上に1次波源を配置した場合がこれに相当する。このような考えに基づいて波面合成を行う方法を、WFS (Wave Field Synthesis) と呼ぶ。図9にWFSによる音場再生の例を示す。

図9に示すように、WFSでは、原音場における境界面 S_1 と再生音場における境界面 S_2 を時間的にも空間的にも異なる位置に設置可能とするため、信号処理ユニット W が必要である。また、図9の1次波源を、マルチチャンネル音響のスピーカ位置に設定し、その信号を1次波源に入力することにより、マルチチャンネル音響による再生波面を合成することができる。

WFSの研究開発はヨーロッパで盛んであり、フランスの IRCAM (国立音楽音響研究所) が平板パネルの後ろに数個の振動素子をつけたWFS再生システムを開発している。また、ベルリン工科大学の円状スピーカアレイを用いたWFSサークル、ドイツのフラウンホーファーIDMT (Institute for Digital Media Technology) 研究所のWFSシアターなどが良く知られている⁹⁾。

3.2 波動方程式の極座標表現に基づく音場再現

3.2.1 音の到来方向のフーリエ表現

Cooperと志賀は、2次元音場の受音点で観測される音の到来方向をフーリエ表現する手法を提案している¹⁰⁾。ある受音点に入射する音の方向パターンを方位角 θ の関数 $S(\theta)$ で表すと、 $S(\theta)$ は図10のように、方向に依存しない定数成分と1次の $\cos$ 成分、 $\sin$ 成分、2次の $\cos$ 成分、 $\sin$ 成分…という形に分解できる。定数成分に対応する指向性を持つ収音デバイスは全指向性マイクロホンであり、1次の成

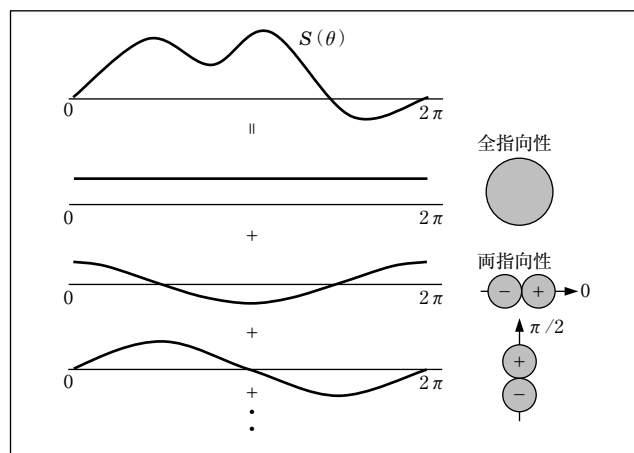


図10 受音点に入射する音の方向性パターンのフーリエ表現

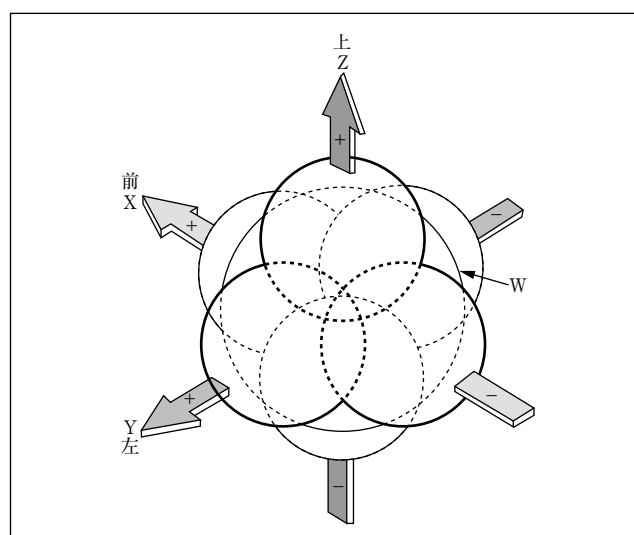


図11 1次アンビソニックス収音方式

分に対応する収音デバイスは両指向性マイクロホンである。したがって、全指向性マイクロホンを利用して定数成分が、両指向性マイクロホンを利用して1次の成分が観測できる。Gerzonは、このような音の方向性パターンの考え方を3次元音場に拡張して、モノフォニックから2chステレオ、水平面内のサラウンド、3次元音響までを階層的に記述するアンビソニックス方式を提案した¹¹⁾。アンビソニックスでは、受音点における音圧を、立体角に関する直交関数である球面調和関数で展開し、一定次数以上の展開係数を無視することで音の方向性パターンを近似的に表す。図11に、1次までの展開係数を利用した場合のアンビソニックス空間収音方式を示す。

3.2.2 波動方程式の極座標表現

音圧の波動方程式は直交座標系によって、

$$\nabla^2 p(x, y, z, t) - \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2} = 0 \quad (15)$$



で表される。ここに、

$$\nabla^2 p = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

はラプラスの演算子(ラプラシアン)である。式(15)の変数を極座標 (r, θ, ϕ) に変換し、変数分離法で解くと、

$$p(r, \theta, \phi, t) = \sum_{n=0}^{\infty} \sum_{m=0}^n (R_1 j_n(kr) + R_2 y_n(kr)) Y_n^m(\theta, \phi) k^{-i\omega t} \quad (16)$$

を得る¹²⁾。ここに、 r は原点からの距離、 θ は方位角、 ϕ は仰角を表す。 j_n と y_n は第1種および第2種の球ベッセル(Bessel)関数であり、 r 方向の音圧の変動を表す。また、

$$Y_n^m(\theta, \phi) \triangleq \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\sin \phi) e^{im\theta} \quad (17)$$

は球面調和関数¹³⁾であり、角度による音圧の変動を表す。球面調和関数における $P_n^m(x)$ は第1種ルジャンドル(Legendre)陪関数であり、

$$P_n^m(x) = \frac{(2n)!}{(n+m)!} \frac{(1-x^2)^{m/2}}{2^n n!} \left\{ x^{n-m} - \frac{(n-m)(n-m-1)}{2(2n-1)} x^{n-m-2} + \frac{(n-m)(n-m-1)(n-m-2)(n-m-3)}{2 \cdot 4 \cdot (2n-1)(2n-3)} x^{n-m-4} - \dots \right\} \quad (18)$$

で表される¹³⁾。ただし、式(18)の $\{ \}$ 内は、 x のべき乗の係数が0になる項で打ち切られる。球面調和関数において、方位角 θ に関しては、Cooperらの方法と同じく $\cos$ 関数と $\sin$ 関数で表されることに注意されたい。

いま、任意の方向 (Ψ, ϕ) から平面波が到来する場合を考える。ここに、 Ψ は到来方向の方位角、 ϕ は到来方向の仰角である。また、図4に示す点 P の極座標を (r, θ, ϕ) とする。このとき、波動方程式の極座標表現を用いると、点 P での音圧は、

$$p(r, \theta, \phi, k) = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) \sum_{m=-n}^n Y_n^m(\theta, \phi) \overline{Y_n^m(\Psi, \phi)} \quad (19)$$

で表される¹²⁾。ここに、 Q は音源の出力であり、 $\bar{Y}$ は Y の複素共役である。また、便宜上、角周波数 ω の代わりに波長定数 $k=\omega/c$ を変数として利用している(c は音速)。

3.2.3 アンビソニックスによる音場再現

受音点 P を取り囲む半径 α の球面上に N 個配置されたスピーカによって、方向 (Ψ, ϕ) から到来した平面波を再現する問題を考える(図12)。いま、 l 番目のスピーカ方向を (θ_l, ϕ_l) 、そのスピーカ入力 $\alpha_l(k)$ とし、スピーカからの音波も平面波で近似すると、 N 個のスピーカによる合成音圧は、

$$\tilde{p}(r, \theta, \phi, k) = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) \sum_{m=-n}^n \sum_{l=0}^N \alpha_l(k) Y_n^m(\theta, \phi) \overline{Y_n^m(\theta_l, \phi_l)} \quad (20)$$

で表される。式(19)と式(20)が一致した場合に、 (Ψ, ϕ) から到来した音波がスピーカによって再現される。この条件を、球面調和関数の正規直交性を用いて解き、展開次数を次数 M までに制限することで、アンビソニックスの基本式

$$QP_n^m(\sin \phi) e^{-im\Psi} = \sum_{l=0}^N \alpha_l(k) P_n^m(\sin \phi_l) e^{-im\theta_l}, \quad (21) \\ n=0, \dots, M, -n \leq m \leq n$$

を得る。式(21)の左辺は (Ψ, ϕ) 方向から到来した音波の直交関数展開係数を表し、右辺は N 個のスピーカで再生した音波の直交関数展開を表す。式(21)を連立させて解くことにより、所望の平面波を再現するためのスピーカ駆動信号 $\alpha_l(k)$ 、($l=1, \dots, N$)が求められる。

アンビソニックスの收音については、球状のマイクロホンアレイが良く研究されている。また、32個の小型マイクロホンカプセルを持ち、4次までのアンビソニックス係数を観測できるマイクロホンが商品化されている¹⁵⁾。

4. バイノーラル方式

以上紹介した技術は、再生時は、基本的にスピーカを用いることを想定したものであった。これに対して、ヘッドホン再生のための技術として、バイノーラル方式が知られている¹⁶⁾。3次元空間におけるあらゆる音場も、人間は二つの耳で観測するため、両耳の入り口で音の物理量が再現できれば、3次元音場が知覚されるはずである。バイノーラル方式は、このような考えに基づくものであり、頭部伝達関数(HRTF: Head Related Transfer Function)、あるいは頭部インパルス応答(HRIR: Head Related Impulse Response)を用いる。HRTFとHRIRはフーリエ変換/逆変換によって相互に変換される。HRTFは、さまざまな方向から両耳に至る伝達関数であり、2.1節で述べた音の方向定位知覚に必要な三つの要素を統一的に表現したものである。最も簡単なバイノーラル録音方式として、ダミーヘッド録音が知られている。これは、人間の上半身を模擬した人形の耳の入り口にマイクロホンを設置したもので、これによって、この人形のHRIRが畳み込まれた音が收音される。

マルチチャネル音響をバイノーラル再生するための方式として、各チャネル信号に対して対応する方向のHRIRを畳み込んでそのチャネルに対する左右耳での信号を計算し、これらを全チャネルで足し合わせる方式が考えられる。また、キルヒホッフヘルムホルツの積分定理によって得られた音場をバイノーラル再生する方法も検討されている¹⁷⁾。

5. む す び

現状では、高品質でコンテンツとしての作り込みができるという意味で、心理音響モデルに基づくマルチチャンネル音響の優位性は疑いがない。物理音響モデルに基づく方式の研究開発は、コンサートホールなどの原音場を別の空間に再現することから始まったが、WFSのようにマルチチャンネル音響信号の再生に利用可能な方式が開発されるに至った。最近では、マルチチャンネル音響とアンビソニックスの間の信号変換も研究テーマとして考えられている。さらに、さまざまな方式の信号をバイノーラル再生するための技術も開発されている。本稿で紹介した音響技術が、今後、仮想現実感の実現に活用されていくことを期待している。

(2012年5月17日受付)

〔文 献〕

- 1) 加藤博一：“講座拡張現実感技術の最前線 第1回：総論”，映情学誌 66, 1, pp.53-56 (2012)
- 2) F. Rumsey: "Spatial Audio", Focal Press (2001)
- 3) 安藤彰男：“高臨場感音響技術とその理論”，IEICE Fundamental Review, 3, 4 (2009)
- 4) H.A.M. Clark, G.F. Dutton and P.B. Vanderlyn: "The 'stereophonic' recording and reproduction system", J. Audio Eng. Soc., 6, 2, pp.102-117 (1958)
- 5) B. Bernfeld: "Attempts for better understanding of the directional stereophonic listening mechanism", Proc. 44th AES Convention (Feb.1973)

- 6) V. Plukki: "Virtual sound source positioning Using vector base amplitude panning", J. Audio Eng. Soc., 45, 6, pp.456-466 (1997)
- 7) B.B. Baker and E.T. Copson: "The Mathematical Theory of Huygens' Principle, 3rd ed.", Chelsea Publishing Company (1987)
- 8) M. Born and E. Wolf: "Principles of Optics, 7th ed.", Cambridge University Press (1999)
- 9) D. de Vries: "Wave Field Synthesis", AES monograph, Audio Engineering Soc. (2009)
- 10) D.H. Cooper and T. Shiga: "Discrete-matrix Multichannel Stereo", J. Audio Eng. Soc., 20, 5, pp.346-360 (1972)
- 11) M.A. Gerzon: "Periphony: With-height Sound Reproduction", J. Audio Eng. Soc., 21, 1, pp.2-10 (1973)
- 12) E.G. ウィリアムス：“フーリエ音響学，音の放射と近距離場音響ホログラフィの基礎”，シュプリンガーフェアラーク東京 (2005)
- 13) 寺沢寛一：“自然科学者のための数学概論”，岩波書店 (1983)
- 14) B. Rafaely: "Analysis and Design of Spherical Microphone Arrays", IEEE Trans. Speech and Audio Proc., 13, 1, pp.135-143 (2005)
- 15) http://www.mhacoustics.com/mh_acoustics/Eigenmike_microphone_array.html
- 16) 平原達也，大谷真，戸嶋巖樹：“頭部伝達関数の計測とバイノーラル再生にかかわる諸問題”，IEICE Fundamental Review, 2, 4, pp.68-85 (2008)
- 17) 矢入聡，岩谷幸雄，鈴木陽一：“聴覚ディスプレイ”，映情学誌, 61, 10, pp.1427-1430 (2007)



安藤 彰男 あんどう あきお 1980年，九州芸術工科大修士課程修了。同年，NHK入局。1983年から，放送技術研究所にて，音声認識，収音システム，空間音響再生，オーディオ符号化などの研究に従事。2010年より，東京工業大学連携教授を兼務。主な受賞として，本会業績賞，信学会論文賞など。博士(工学)。正会員。



Kinectの仕組みと ナチュラルユーザインタフェース



正会員 西 林 孝†

1. ま え が き

Kinectは、Xbox 360用のゲームコントローラとして2010年末に発売されたデバイスである。Kinectを使うとプレーヤーは手に何も持たずに、身振り手振りや声を使ってゲームを操作することができる。本来はゲームコントローラなのであるが、USBインタフェースを備えており、KinectをWindowsやLinuxから利用できるドライバがいくつか公開されている。USBケーブルを使って手元のコンピュータに接続するだけで、Kinectから得られたデータを使ったり、モーションキャプチャデバイスとして利用したアプリケーションを手軽に作ることができる。そのため、ゲーム開発とは関係のない所でも注目され、インタラクションの研究などにも使われている。Kinectには、距離画像センサ、ビデオカメラ、アレイマイクといった人間とインタラクションするためのセンサが搭載されており、加えて、いくつかのKinectドライバには、高度な姿勢認識を行う機能が実装されている。本稿では、距離画像センサの技術や姿勢認識機能、ジェスチャ認識を使ったシステムの可能性について、技術的な側面から解説する。

2. Kinectの概要

Kinectセンサは、人間とインタラクションを行うための複数のセンサとプロセッサ、USBインタフェースから構成されている(図1)。KinectのUSBケーブルをPCやMacに差すと、カメラ、チルトモータ、アレイマイクの三つのデバイスとして認識される。細かく挙げると、ビデオカメラ、距離画像センサ、赤外線センサ、アレイマイク、チルトモータ、加速度センサがPCから使える。ただし、これらのセンサだけでゲームをジェスチャで操作するような仕組みを実現している訳ではない。Kinectセンサは、これらのセンサを統合して扱えるドライバソフトウェアでの利用を想

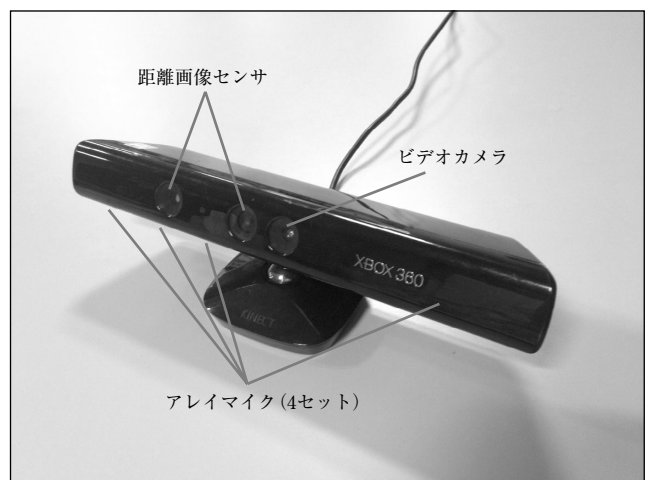


図1 Kinectセンサの外観

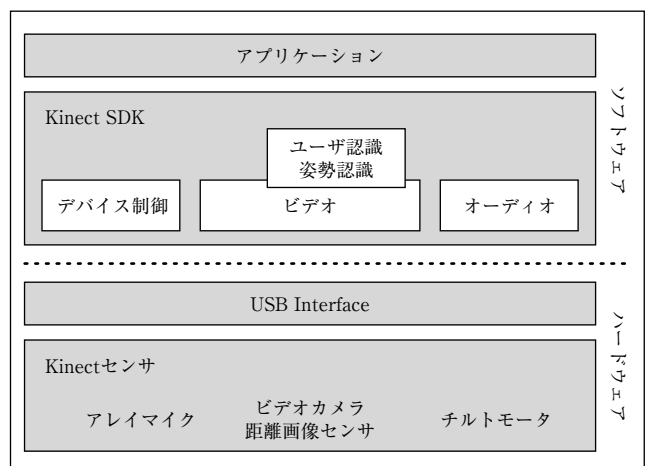


図2 Kinectのアーキテクチャ

定して作られており、ドライバソフトウェアも重要な働きをしている。図2に、KinectセンサとKinect用のドライバソフトウェアである、Kinect for Windows SDK (以降Kinect SDKと表記する)の関係を示す¹⁾。例えば、モーションキャプチャ(ボートラッキング)機能は、NUI API

† 株式会社ジェネシックス

"Technology Frontier of Augmented Reality (9): Kinect and Natural User Interface" by Takashi Nishibayashi (genesix, Inc, Tokyo)



表1 Kinectセンサのスペック項目スペック

ビデオカメラ最大解像度	1280×1024ピクセル
距離画像センサ最大解像度	640×480ピクセル
フレームレート	30fps
距離取得可能範囲	0.5～10m
姿勢認識可能範囲	0.8～4m
垂直視野	43°
水平視野	57°
チルトレンジ	-30°～+30°
アレイマイク	本体下部に4機
音声フォーマット	16-kHz, 16ビットモノラルPCM

(図中「ユーザ認識, 姿勢認識」), 音声方向認識はAudio API(図中「オーディオ」)に含まれている。Microsoft社は、人間の認識, 姿勢の認識を使ったユーザインタフェースを, NUI (Natural User Interface) と名付けている。

ハードウェアのおおまかなスペックは、表1の通りであるが、実際には利用するドライバによって、センサから得られる値の範囲は大きく異なってくる。最初に発売されたKinect for Xboxと、2012年2月に発売されたKinect for Windowsでも、若干異なる。Kinect for Windowsは、Nearモードという新しい動作モードが追加されており、40cmから距離を取ることができる。

Kinectの最も特徴的な機能である、人間の動き・姿勢をリアルタイムに認識する機能は、距離画像センサと姿勢認識という技術の組み合わせによって実現されている。次章でそれぞれについて解説する。

3. Kinectの距離画像センサ

距離画像センサとは、各画素値が距離の情報を持つ画像(深度画像)が得られるセンサのことを言う。深度センサと呼ばれることもある。筆者を映した深度画像をグレースケールにしたものが図3である。光学カメラと比較して物体の形がはっきりと捉えられるのがわかる。そのため、入退室のセキュリティチェックや荷物のサイズチェックに用いられる。Kinect for Xbox + Kinect SDKでは、各ピクセルが80～4,000cmの値のデータが得られる。

3.1 他の距離画像センサとの比較

Kinectの登場以前、距離画像センサといえば、価格帯が数十万から数百万円の業務用・研究用の機器であった。例えば、研究用途でよく使われるMesa Imaging社のSR4000は、安いネット通販でも4,200USドルである²⁾。それに比べればKinectは、12,000円程度という安価で²⁾、家電量販店に行けばいくらかでも手に入る。しかし、Kinectはゲームコントローラなどだけに屋外では利用できない。屋内でも直射日光が当たる部分の距離データが欠落する。これは、屋内利用を想定した強度の赤外線レーザを照射して距離を測定しているためである。逆にこの現象を



図3 深度画像

利用して、距離情報が不要な部分に赤外線を照射して、測定範囲を限定することもできる。

3.2 パターン照射による距離の測定

距離画像センサが距離を測定する方法として最もよく使うのは、TOF (Time of Flight) 方式である。TOF方式は、赤外線レーザを照射し、受信した反射光の時間差を利用して距離測定を行う。一方、Kinectではそれとは異なり、特定のパターンを赤外線レーザで照射し、反射光のパターンを解析して距離を測定している。反射光のパターンは対象までの距離に応じて歪みが生じるため、各パターンドットの移動量から三角測量でセンサからの距離が計算できる。

Kinectセンサには、イスラエルPrimeSense社の距離画像センサチップのPS1080が使われている。PrimeSense社は、距離測定用の赤外線照射パターンの特許を取得³⁾しており、同社のWebサイトには、LIGHT CODING INFRARED PATTERNと記載されている。図4は、LIGHT CODINGの照射パターンである⁴⁾。

LIGHT CODINGの照射パターンは、ランダムドットではなく、周期性を持たせたパターンで構成されており、単一の点の位置を特定できるという特徴がある。この方式により、周辺光の影響を減らし、高速な3Dマップの生成が可能になった。

さて、物体の形がよくわかる深度画像が距離画像センサで得られることを説明したが、どうして距離画像センサが使われているのであろうか。Kinectの目的は、手に何も持たずにゲームが操作できることである。つまり、人間がこういった動きをしているか、を把握する必要があったためである。次の章で、深度画像を使った姿勢認識について解説する。

4. 深度画像を使った3次元姿勢認識

深度画像センサに続いてもう一つ重要な技術は、人物の姿勢認識 (Human Pose RecognitionまたはHuman

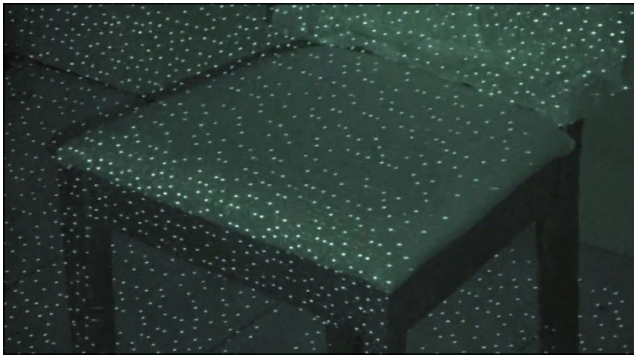


図4 LIGHT CODINGの照射パターン

Pose Estimation)である。姿勢認識は、人間がどのような姿勢を取っているかを推定する技術で、Kinect SDKを使って、認識結果から骨格を描画したものが図5になる。Kinectで採用された方法は、一枚の深度画像から、画像の各ピクセルが人体のどのパーツであるかを、右上腕や頭部といったレベルで識別するものである⁵⁾。各パーツを識別した後、関節の位置を導出する。ここで深度画像を使うメリットは、光学画像に比べて、照明条件・服の色の影響を受けず、人の形がくっきりと得られることである。この方法の特徴を二つ挙げる。

- (1) 前後(動き)の時系列の情報を一切使わずに、一枚の深度画像のみを基に体のパーツ推定を行った：GPUでの並列処理が可能になった。さらに初期ポーズが不要で、ロバストな姿勢認識ができる。

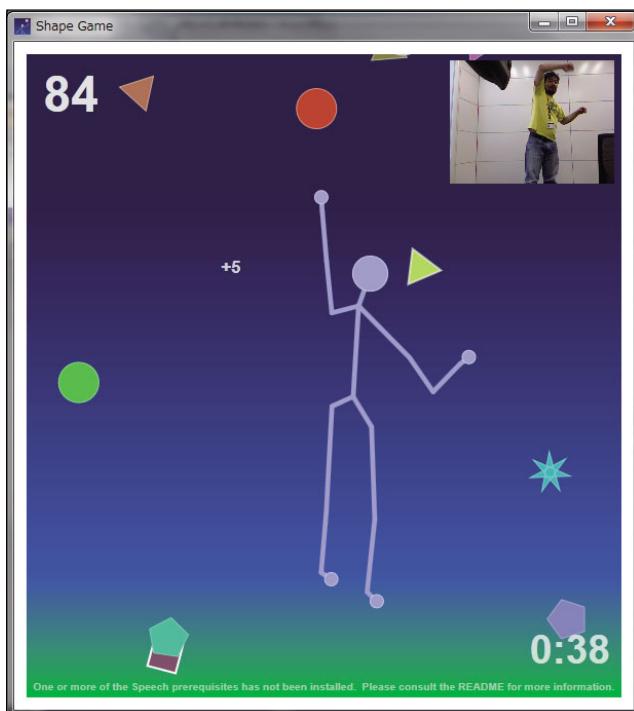


図5 姿勢認識結果を使った骨格表示

- (2) 学習データを大量に用意して機械学習で結果を出した：Xbox 360では、すべての人間の動きを認識する必要はなかったため、ゲームに必要なボールを蹴るポーズやダンスのポーズに限定して学習データが作成された。この時の学習データは、良い結果を得るために9百万セット程度必要だった。

従来のリアルタイム姿勢認識は、認識を開始する前に初期ポーズを取る必要があり、利用時の煩わしさがあった。それに比べると、初期ポーズ不要なこの方式は、圧倒的に手間が少なく快適であることから、今後のリアルタイム姿勢認識のスタンダードになるであろう。ここまでで、Kinectが人の体勢を認識する方法を解説した、次章は、それをどのように使えるかを解説する。

5. 人とのインタラクションと課題

例として、Xbox 360のメニューを挙げる(図6)。Xbox 360のメニューをKinectで操作するモードは、プレーヤの手の動きに合わせてカーソルが動き、カーソルをボタンの上で一定時間静止させるとボタンを選択できる。また、音声コマンドも可能で、「エクスポックス」と呼びかけた後にBing検索などができる。これらの操作の特徴や課題を次に述べる。

5.1 ジェスチャインタラクション

手の動きにあわせて、マウスカーソルを動かすことは、姿勢認識(ポインタラッキング)の結果を使えば容易にできる。ポインタラッキングの結果、手の位置の世界座標が取れるので、表示座標系に変換して適度にフィルタをかければよい。さて、Xbox 360では、メニューの選択に静止ジェスチャが使われているが、あまり直感的ではない静止ジェスチャが使われているのは、やや違和感があると思われるであろう。手を前後にプッシュしたり、握ったりする方が「選択」というアクションに近い感じがする。筆者の経験では、ジェスチャ動作の認識しやすさは、そのまま誤認識のしやすさにも繋がってしまう。人が無意識に行う動きを

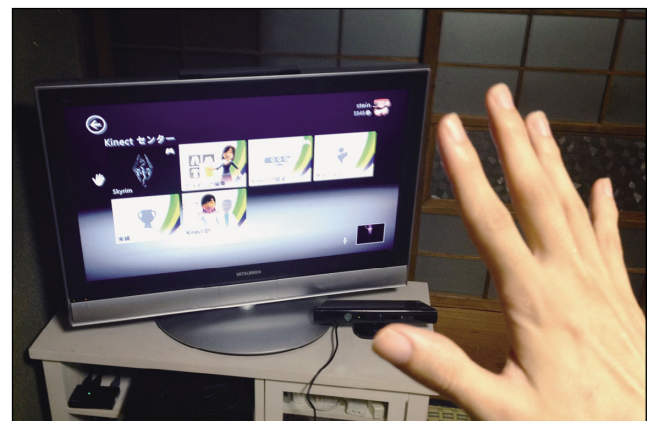


図6 Xbox 360のメニュー



誤って検知してしまうようでは使い物にならないため、ジェスチャ認識は特徴的な動きを利用する必要がある。一方、コンピュータにとって認識しやすくても、人間に厳しい複雑な動作も利用できない。その兼ね合いから、静止ジェスチャが採用されたのだと考えられる。筆者は一度、手を握る・解放する動きを認識してドラッグアンドドロップを実装してみたが、誤検知が多く、とても実用に耐えられる物にはならなかった。

もう一つ課題を挙げるとすると、腕を肘より高い位置に上げた状態を維持するのはとても体力に負担がかかることがある。手を上空に掲げてコンピュータを操作する、まるで映画の世界だが、カッコ良さの代償はそれなりにある。しかし、コンピュータに触れることが制限された外科手術現場での使用例があり⁶⁾、そういった特殊な空間では十分にメリットがある。

5.2 音声認識

ここ1年の間、音声認識は特に携帯電話端末に一気に普及した。Kinect SDK自体には音声認識の機能はないが、Microsoft Speech Platformと組合せることで実装ができる。音声認識もジェスチャ認識と同様で、いかに誤認識を減らすかが重要になってくる。先に挙げたXbox 360のメニューでは、「エクスポックスプレイ」、「エクスポックスホーム」と、音声コマンドの先頭に「エクスポックス」を付けることで、日常会話中の単語を拾ってしまうのを回避している。

5.3 無意識の動きを捉える

人間の無意識の動作にはさまざまな情報が含まれている。例えば、椅子に座っている時の姿勢の良し悪し(猫背)や画面に向かって集中しているかどうか等。これらの特徴を認識してフィードバックを送ったり、記録したりするシステムの可能性は充分にあるだろう。

5.4 ARシステムでの利用

人の動きに合わせて、まるで服を着ているかのように見せる試着システムをはじめ、ARシステムにKinectが利用されることもある。ビデオカメラの画像と姿勢認識結果を組み合わせることで、人の姿勢自体をマーカとして扱うことができる。初期ポーズが不要な姿勢認識エンジンが使えるようになったことで、このようなシステムの運用は非常に楽になった。

6. Kinect以外の家庭用モーションキャプチャデバイス

人間の動作を認識するセンサデバイスは、Kinectだけではない。ゲーム用途以外にも、リビングに設置して家電をコントロールするといった用途の製品も発売されている。これらのセンサは、Kinectと同様に姿勢認識などの機能を持ったドライバソフトウェアとプログラミング環境が付属しており、独自のシステムを開発するのに利用できる。



図7 Xtion Pro live

Kinectのような処理をやりたいがWindowsマシンを持っていない、といった場合には、以下に挙げるデバイスの導入を検討すると良いだろう。

6.1 ASUS Xtion PRO LIVE

距離画像センサ、ビデオカメラ、アレイマイクが2機という構成で、距離画像センサはKinectと同じPrimeSense社のPS1020である(図7)⁷⁾。チルトモータがない分、消費電力は低く5VのUSB給電のみで動作し、大きさはKinectの半分程度で取り回し良いのが特徴である。公式ドライバソフトウェアはOpenNI。動作プラットフォームはWindows, Linux, MacOS, Android (ARM) がサポートされており、開発はC, C++, Java, C#で行えるため、開発環境、動作環境が限定されたKinectに比べて、広いシーンで利用可能である。価格は24,000円程度、amazon等のネット通販で購入することができる。ジャケットのポケットに入れて持ち運べる大きさなので、外出の多い筆者は愛用している。

6.2 OpenNI

OpenNI⁸⁾は、Kinectのようなジェスチャ認識、姿勢認識、音声認識を行うセンサデバイスの共通インタフェースを、アプリケーションに提供するフレームワークである。汎用性を持たせるため、OpenNI本体、ハードウェアドライバ、ミドルウェアという構成になっており、クロスプラットフォームなAPIが用意されている。Kinect用のハードウェアドライバを使うと、OpenNIを使ってKinectを操作することもできる。Kinect for Windows SDKよりも早く2010年末には公開されていたため、今でも根強い人気がある。

6.3 SoftKinetic DepthSense 311

距離画像センサ、ビデオカメラ、アレイマイク2機という構成のセンサである。ベルギーのSoftKinetic社が開発した⁹⁾。Kinectとは異なり、ToF方式を使った距離画像センサを採用している。ToF方式ながら499USドルという価格設定には驚きである。ドライバソフトウェアは、ジェ



スチャ認識、姿勢認識ができるiisu SDK¹⁰⁾が用意されている。

6.4 LeapMotion LEAP

本稿執筆時点では注文受付中で、販売は開始されていない。詳細は明らかにされていないが、小型の距離画像センサで指先の動きを認識する。既存の3Dセンサの100倍超の解像能があり、マウスの代りになるデバイスになると発表されている¹¹⁾。利用するのに広いスペースが必要なKinectに比べて、机に上向きに設置するLEAPは設置場所に困ることはないだろう。開発者向けSDKの公開も予定されており、発売が楽しいデバイスの一つである。

7. む す び

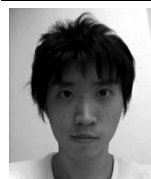
Kinectは、ただのゲームコントローラではなく、人とコミュニケーションするのに特化したデバイスであることを解説した。音声でのやりとりも可能なことから、モーションキャプチャデバイスという単語では表現しきれない能力がある。3Dセンシング技術の発達、一般家庭のリビングに恩恵をもたらす段階にまで来た。コンピュータと人間が自然に対話できる時代になったと言える。これからは人の動きを正確に捉えるだけでなく、人の意志をいかにして認識するか、という技術がますます必要になるであろう。例えば、発せられた状況によってその意図が変わる言葉をいかに認識するかという取り組みもなされている¹²⁾。デバイスが安価になったことにより、さまざまな分野の人がこの

課題に取り組むことが可能になった。今後の発展が非常に楽しみである。

(2012年6月29日受付)

〔文 献〕

- 1) Kinect for Windows SDK, <http://www.microsoft.com/en-us/kinectforwindows/>
- 2) <http://www.acroname.com/robotics/parts/R329-SR4000-10M-USB.html>
- 3) US Pat., US 2010/0118123 A1: "Depth Mapping Using Projected Patterns" (May 13, 2010)
- 4) J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake: "Real-Time Human Pose Recognition in Parts from Single Depth Images", CVPR (2011).
- 5) Kinect filmed with a night vision camera - HD, <http://www.youtube.com/watch?v=6CFoOFZ6ifc>
- 6) Xbox useful to surgeons, <http://sunnybrook.ca/media/item.asp?c=1&i=616>
- 7) ASUS Xtion PRO LIVE, http://www.asus.com/Multimedia/Motion_Sensor/Xtion_PRO_LIVE/#specifications
- 8) OpenNI, <http://openni.org>
- 9) Softkinetic DepthSense 311, <http://www.softkinetic.com/Solutions/DepthSensecameras.aspx>
- 10) iisu SDK, <http://www.softkinetic.com/Solutions/iisuSDK.aspx>
- 11) Leap, <http://live.leapmotion.com>
- 12) MSDNマガジン: Kinectを使った状況に応じた会話, <http://msdn.microsoft.com/ja-jp/magazine/hh882450.aspx>



にしばやし たかし
西林 孝 2003年、姫路工業大学理学部生命科学科卒業。ソフトウェアエンジニアとして、Web系の会社を転々としつつ、Kinect等の面白デバイスを使ったハッカソンや勉強会を東京で開催している。現在、VOYAGE GROUP (株) にて、スマートフォンアプリケーションの開発に従事。 http://d.hatena.ne.jp/hagino_3000



1. ま え が き

スマートフォンは、カメラを含めさまざまなセンサとディスプレイをコンパクトに格納していることから、Augmented Reality (AR) システムを実現する上で好適なプラットフォームとして利用され始めている¹⁾。現在のモバイルARは、実現手法によって大きく2つに分類できる(図1)。一つはGPSや電子コンパス等を用いた測位によるARであり、一つはカメラを用いた画像認識によるARである。位置情報ARに関しては、レストランやホテル、ランドマーク等の情報をユーザの緯度経度方位に紐付けて提示するアプリケーションが公開されている²⁾³⁾。しかし、位置情報ARは未だ座標の精度が不十分という問題がある。一方、画像認識ARに関しては、サーバで実行するシステムとモバイル端末で処理するシステムに分類できる。前者は、豊富な計算資源を利用できるものの送受信に時間がかかるという問題がある。後者に関しては、人工的な白黒マーカを用いる手法⁴⁾⁵⁾が、比較的少ない計算量で正確な認識を可能としている。しかし、不自然な白黒マーカを予め配置しておく必要があるため、利用できるシーンが限られるという問題がある。一方で、白黒マーカに限定されず任意平面を認識するマーカレスARシステム^{6)~9)}が商用化されている。雑誌やパンフレット等を認識することにより、詳細情報の提示や、購入サイトへ直接導くことで、紙面とネットとを繋ぐサービスを展開している。ただ、何らかの紙面を別途用意する必要があることには変わりはなく、使い勝手には改善の余地がある。また、平面には多数の特徴点が存在することを前提としているため、テクスチャに乏しい平面の認識には対応できていない。さらに、ユーザの立場からすると、ARシステムが眼前の紙面を認識対象としているのか否かを事前に判断する手段がないという課題がある。どの紙面にカメラをかざすべきなのかを

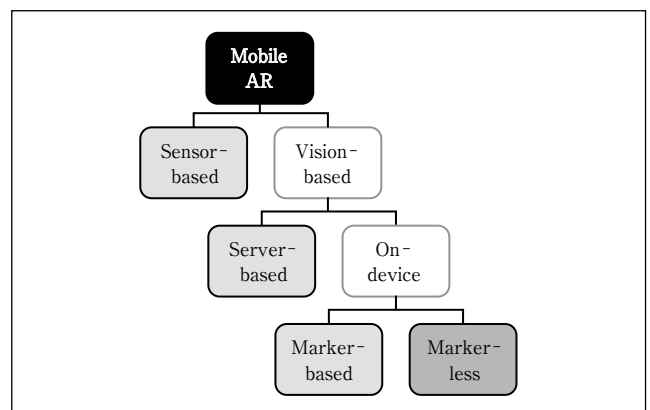


図1 モバイルARの分類

示すためのマーカを付与するとなると、白黒マーカ方式との差異を打ち出しにくくなりかねない。複雑なテクスチャを持つ紙面の利用は、提示情報と姿勢情報とを同時に取得する点で有効であるが、提示情報を別途指定する場合には、姿勢推定にのみに注力することが、レスポンス向上の観点からも望ましい。

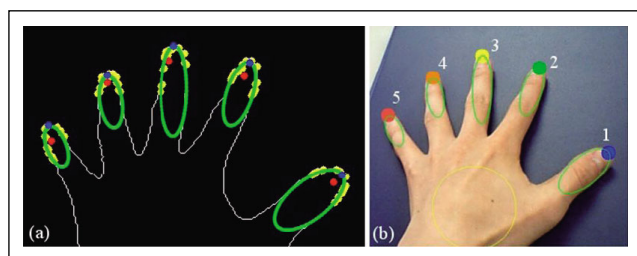
本稿では、上記課題の解決策の一つとして、てのひらを認識対象とする手法について解説する。てのひらには十全な特徴点が存在せず、個人差も大きいことから、前述のARシステムではまったく対応できていない。認識対象をてのひらに限定することで、比較的非力なスマートフォンでもリアルタイム動作を可能とするとともに、てのひらの位置と姿勢を利用した新たな利用方法を提示する。次章では、関連研究の概要と問題点を紹介し、第3章でアルゴリズムの詳細を解説する。第4章では幾つかの応用事例を示し、第5章でまとめを述べる。

2. 関連研究とスマートフォンでの課題

てのひらをマーカの代用として利用する研究は幾つか報告されている。Handy AR¹⁰⁾は、色情報を用いて肌色領域を抽出した上で、肌色領域の境界線において図2のよう

† 株式会社 KDDI 研究所

"Technology Frontier of Augmented Reality (10): Palm-top AR" by Haruhisa Kato (KDDI R&D Laboratories Inc., Saitama)

図2 HandyAR¹⁰⁾による指先検出

に、曲率が大きな箇所を指先として検出する。複数の指先の配置からてのひらの姿勢を推定し、情報を重畳する。しかし、処理負荷が大きい楕円フィッティングを用いて指先を検出するので、5本の指先を検出するには、処理時間がかかりすぎるという課題がある。特に、スマートフォンで実現するには相応の高速化が必要となる。文献¹¹⁾では、同境界線に対する凸包多角形の頂点として指先を認識する。楕円フィッティングの省略により、HandyARと比較して7倍以上の高速化を実現しており、スマートフォンでのリアルタイム動作を可能とした。本手法¹¹⁾は、Androidアプリとして実装され、無償で公開されている¹²⁾。ただし、両手法とも、てのひらの輪郭抽出に肌色検出¹³⁾を利用しているため、認識精度は肌色検出精度に依存する。HandyARは、RGB色空間においてGMMによる肌色モデルを構築して、過去数フレームの情報から肌色モデルを更新しているが、むしろ不安定になっているという報告¹⁴⁾もある。一般に、色情報に依存した検出法は、背景色や光源変化に対する耐性が低い。しかも、両手法は指を開き続けておく必要があることから、指の間から見える背景の影響を受けやすい。さらに、スマートフォンは、機種によってオートホワイトバランスや自動露出調整の性能にばらつきがあるため、色情報に強く依存する既存手法は認識精度が低下しやすいという問題がある。

また、スマートフォンのカメラは画角が十分に広くないため、指先すべてを画面内に収めにくいという問題がある。HandyARに関しては、指先が楕円として見えるように撮影するため、カメラからの距離を短くしなければならない。一方で、画角が狭いスマートフォンのカメラにおいて、5本の指先すべてをフレームに収めるには、カメラから十分に距離を取る必要がある。実際、両手法は5つの指先の配置から姿勢を推定するので、指先が一本でもカメラフレームの外に出ると認識に失敗する。相反する制約によって、認識可能な距離は狭い領域に制限されるため、ユーザは適切な距離を常に保つ必要があり、インタフェースとしては使い勝手に難がある。実際のユースケースでは、てのひらに提示される情報をよく見ようと手を近づける状況において、指先がカメラフレームから外れて認識に失敗する事例が多い。

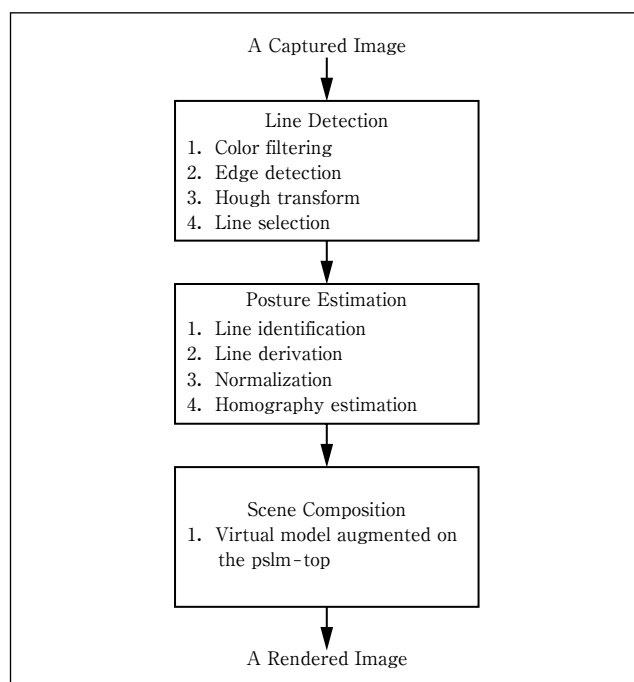


図3 てのひらARの処理手順

3. てのひらAR

3.1 概要

てのひらAR¹⁵⁾では、手の線分を利用することで、新たに指を揃えた手に対応した。図3に、てのひらAR全体の処理手順を示す。まず、入力画像(図4)にエッジ検出(図5)とHough変換を適用し、線分の長さ、色、平行性および近傍色の均一性を予め設定した閾値と比較することで、4本の指の境界線および端点を抽出する(図6)。次に、端点を接続した線分を含めて正規化し、最小二乗法によりHomography行列を算出する。最後に、3Dモデル情報にHomography行列を適用することで、てのひらの上に存在するように重畳する(図7)。次節以降では、処理の詳細について解説する。

3.2 てのひらの線分検出

てのひらの線分検出では、入力画像から指と指の境界線を抽出する。まず、Canny等のエッジ検出を行い、ノイズ除去およびエッジの拡張等の整形をする。肌色検出を適用することで、エッジ検出する領域を限定することも可能である。この場合、既存手法とは異なり、肌色検出は高速化のための処理であるため、検出精度は比較的重要ではない。

次に、Hough変換によってエッジ成分の座標を極座標系で表現し、複数の線分を検出する。

続いて、線分選択では、下記の条件を選別基準として、Hough変換で検出された線分の中から、指と指の境界線に対応する線分を抽出する。

- (1) 線分の長さが閾値より長いこと
- (2) 線分の輝度が閾値より小さい(暗い)こと



図4 入力画像



図5 エッジ検出結果

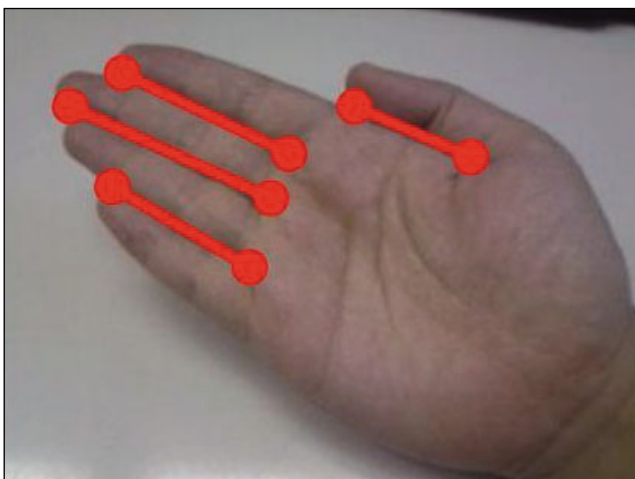


図6 線分検出結果

- (3) 近傍に平行な線分が存在すること
- (4) 線分の近傍領域の色変化が閾値より小さいこと
- (5) 複数の線分の傾きが閾値内に収まること

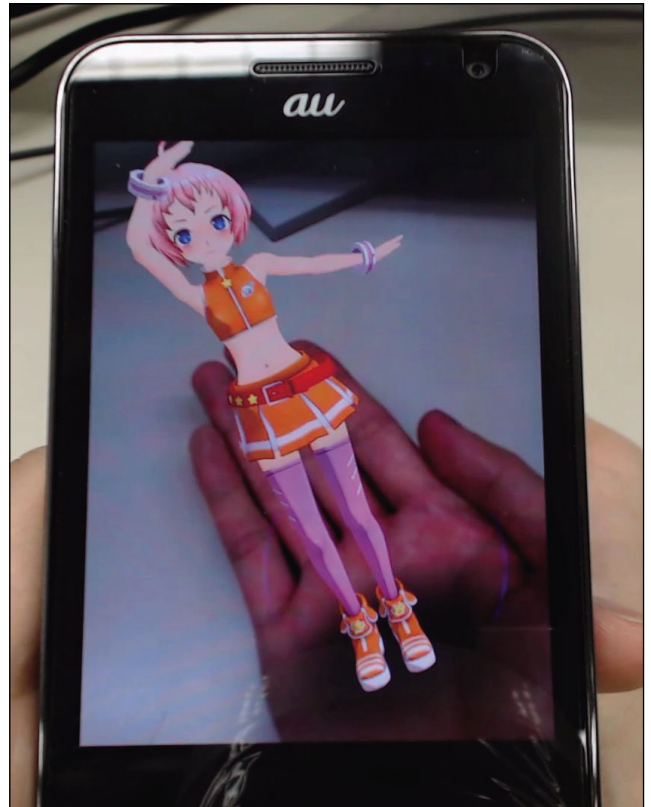


図7 3D CGキャラクターの重畳

最後に、選別された線分と指同士の境界線との対応関係を同定する。複数の線分毎に中心座標 c_i を求め、当該座標を除いた平均との差が最大となる座標を、親指とてのひらが構成する線分として判断する。

$$\max \left\| c_i - \frac{1}{n-1} \sum_{j \neq i} c_j \right\| \quad (1)$$

複数の線分における相対的な位置関係は保持されるので、当該線分と垂直な線分が交差する順に、各指同士が構成する線分に割当てて。

3.3 てのひらの姿勢推定

てのひらの姿勢推定では、既知の姿勢における指同士の境界線を基準線とし、前節で検出された線分と比較する。

選別された線分 l' および対応する基準線 l は、直線の係数としてそれぞれ次式で表現する。

$$l = (a, b, 1)^T \quad (2)$$

$$l' = (a', b', 1)^T \quad (3)$$

ここで、 T は転置を表す。座標 $x = (x, y, 1)^T$ との関係は次式で表現できる。

$$l^T x = 0 \quad (4)$$

$$l'^T x' = 0 \quad (5)$$

一般に、Homography 行列 H を使って、任意の座標 x

は座標 x' に写像される。ただし、スケーリングには依存しないため、係数 s を介して式 (6) が成り立つ。

$$sx' = Hx \quad (6)$$

ここで、 H は下記の 3×3 行列で表される。係数は 9 個であるが、スケーリングの不定性のため自由度は 8 である。

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (7)$$

式 (6) を式 (5) に代入すると、式 (4) との関係から、

$$sl^T = l'^T H \quad (8)$$

$$\therefore sl = H^T l' \quad (9)$$

が得られる。ベクトルおよび行列の要素を書き表すと、

$$s \begin{pmatrix} a \\ b \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} a' \\ b' \\ 1 \end{pmatrix} \quad (10)$$

となるので、各行で展開すると、

$$sa = h_{11}a' + h_{12}b' + h_{13} \quad (11)$$

$$sb = h_{21}a' + h_{22}b' + h_{23} \quad (12)$$

$$s = h_{31}a' + h_{32}b' + h_{33} \quad (13)$$

となる。スケーリング係数 s を除いて、

$$a(h_{31}a' + h_{32}b' + h_{33}) = h_{11}a' + h_{12}b' + h_{13} \quad (14)$$

$$b(h_{31}a' + h_{32}b' + h_{33}) = h_{21}a' + h_{22}b' + h_{23} \quad (15)$$

とした上で、改めて行列表現に戻すと、

$$Ah = \mathbf{0} \quad (16)$$

$$A = \begin{pmatrix} -a' & 0 & aa' & -b' & 0 & ab' & -1 & 0 & a \\ 0 & -a' & ab' & 0 & -b' & b'b' & 0 & -1 & b \end{pmatrix} \quad (17)$$

ここで、 h は行列 H をベクトル化した 9×1 のベクトル、 $\mathbf{0}$ はすべて 0 で構成された 2×1 のベクトルである。式 (16) は一組の線分の対応関係を表したものであるから、複数組の線分を同時に表す場合は、行列 A およびベクトル $\mathbf{0}$ を拡張すれば良い。すなわち、 n 組の線分の対応関係を並べたとき、 A は $2n \times 9$ の行列、 $\mathbf{0}$ は $2n \times 1$ のベクトルで表される。このとき、式 (16) 左辺のノルムを最小化するベクトル h は、 $A^T A$ の最小の固有値に対する固有ベクトルとして求められる。

3.4 姿勢推定に用いる線分の生成

3.2 節で 4 本の線分が得られているため、行列 A から Homography 行列 H を求めることができる。しかし、4 本の線分はほぼ平行に近いので、姿勢推定に直接用いると誤差の影響が大きくなりやすい。したがって、姿勢推定には直線の端点を接続した新たな線分を用いる。



図8 検出された線分の端点を接続して生成された線分

端点の接続はさまざまな組合せを取り得る。しかし、3 本以上の線分が一点で交差する共点が存在する場合は情報が欠損するため、共点を持たない組合せが望ましい。例えば、図 8 のような組合せを利用することで、姿勢推定精度を向上させることができる。情報量的には、3 点の特徴点（親指、人差し指、小指の根元）と 1 本の線分（中指と薬指の境界線）に相当するが、すべてを線分として扱うことで、後述する正規化処理を統一的に扱うことができる。また、線分への統一であれば指先方向の端点座標が不要であるため、指先が画面外に見切れても認識を継続することができるメリットがある。

3.5 姿勢推定に用いる線分の正規化

線分が原点付近を通過する場合も、誤差に左右されやすくなるため、予め正規化しておくことで、姿勢推定精度をさらに向上させることができる。正規化は、複数の直線の中心を原点に移動するとともに、原点からの距離の平均をスケーリングする。

線分 l および l' をそれぞれ行列 N , N' で正規化した線分 L , L' を求めておき、改めて線分 l および l' として扱う。

$$L = Nl \quad (18)$$

$$L' = N'l' \quad (19)$$

正規化後の直線 L および L' に対応する行列 H_L を算出した後、正規化前の直線 l および l' に対応する行列 H_l を次式で算出する。

$$H_l = N'^{-1} H_L N \quad (20)$$

4. てのひら AR の応用

本章では、てのひらの位置と姿勢を認識することによる応用例を提示する。

4.1 アプリランチャーとしての UI

図 9 は、てのひらの上にアプリケーションのアイコンを



図9 アプリランチャーのUI



図11 3D CGキャラクタの等身大表示



図10 指輪の仮想試着



図12 平面推定

立体的に表示し、手のジェスチャによって直感的な操作を可能とする。具体的には、手を傾けた方向へ球状に配置されたアイコンを回転させるとともに、傾けた角度によって回転速度を制御できる。てのひらを閉じることで、球の正面に位置するアイコンが選択され、アプリケーションが起動するランチャーのUIとして利用できる。タッチパネルでの入力との相違は、スマートフォンの小さな画面を太い指が隠蔽しないというメリットがある。また、将来的にカメラを搭載したHMDが登場した際には、ハンズフリーのUIとしての利用が考えられる。

4.2 仮想試着

図10は、指輪や腕時計、ネイルアート等の3D CGを自らの手に重畳することで、仮想的な試着を可能とする。さまざまなデザインや色を簡単に差し替えることができ、気に入ったものをECサイトで購入するシステムの構築も考えられる。てのひらの動きに応じて、見え方の変化やCGのエフェクトを連動させることで、商品の魅力を十二分に伝えることができるため、広告としての利用に付加価値を与えるこ

とができる。顔の認識による眼鏡の仮想試着^{16) 17)}はすでに商業利用されており、てのひらに限らず、今後も身体を認識する用途は、ECサイト等にて導入されていくものと考えられる。

4.3 等身大表示

重畳する仮想情報は手の上に載せるだけでなく、拡大表示することで、より親近感が湧くような体験を提供できる。図11では、実空間と連動した親しみのあるエージェントを携帯端末上に実現し、身近に感じられる演出が新たな体験を生み、キャラクタービジネスにも新たな価値基準を生むことが期待される。

4.4 平面推定

図12では、手の甲の姿勢を認識している。手が設置している平面の姿勢と手の甲の姿勢とが同一であると仮定して、手の甲の姿勢から平面および周辺環境の相対的な位置関係を推定することで、より広い領域を含む実写映像に仮想物体を重畳できる。

同様の機能は、Visual SLAM (Simultaneous



Localisation and Mapping) 等で研究が報告¹⁸⁾されている。しかし、十分な数の特徴点が存在していることを前提としているため、図 12 のように周囲に特徴点が少ない場合、認識精度が低下しやすい。また、スマートフォンでのリアルタイム動作させるには、Visual SLAM は処理負荷が大きすぎるという問題がある。一方、手の甲を画面に収めることで手の姿勢を手がかりとすることができ、どこでも簡単な平面姿勢推定が可能となる。さらに、てのひらに隣接する領域を含む範囲を抽出することで、仮想情報を重畳させる空間とすれば、周辺環境との整合性を保ちリアリティの向上が期待できる。

5. む す び

本稿では、てのひらとカメラとの相対的姿勢および位置を推定する手法について紹介した。指先の座標を利用する既存手法と比較して、指の線分を利用する手法は下記の点で優位であると言える。

- (1) 指の境界線を基準とするため、画面外に指先が出て
も認識が可能
- (2) 指を揃えているため、背景の影響は軽微
- (3) 色情報への依存が小さいため、光源変化に頑健

また、応用例として、てのひらの姿勢を変化させるだけで、提示情報を直観的な操作で制御できる事例を示した。スマートフォンのハードウェア性能は急速に発展しているため、今後更なる研究開発が進められるものと考えられる。

(2012 年 6 月 26 日受付)

〔文 献〕

- 1) 小林亜令: “モバイル AR プラットフォーム”, 映情学誌, 65, pp.660-662 (2011)

- 2) Sekai Camera, <http://support.sekaicamera.com/>
- 3) Augmented Reality Browser: Layer, <http://www.layar.com>
- 4) H. Kato and M. Billinghurst: "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System", In Proc. of the 2nd Int'l. Workshop on Augmented Reality, pp.85-94 (1999)
- 5) ARToolkit, <http://www.hitl.washington.edu/artoolkit/>
- 6) Vuuforia, <http://www.qualcomm.com/solutions/augmented-reality>
- 7) metaio Mobile SDK, <http://www.metaio.com/software/mobile-sdk/>
- 8) Aurasma, <http://www.aurasma.com/>
- 9) String, <http://www.poweredbystring.com/>
- 10) T. Lee and T. Hollerer: "Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking", In Proc. of the 11th IEEE Int'l. Symp. On Wearable Computers, pp.83-90 (2007)
- 11) H. Kato and T. Kato: "A Marker-Less Augmented Reality Based on Fast Fingertip Detection for Smart Phones", In IEEE International Conference on Consumer Electronics (ICCE), pp.127-128 (2011)
- 12) てのりん, <http://tenor.in>
- 13) M.J. Jones and J.R. Rehg: "Statistical Color Models with Application to Skin Detection", In Int'l. Journal of Computer Vision, 46, pp.1-23 (2002)
- 14) 田中隆治, 吉岡理文: “拡張現実感のための手認識”, 電気学会論文誌 C, 131, 11, pp.1895-1900 (2011)
- 15) H. Kato and A. Yoneyama: "A Line-based Palm-top Detector for Mobile Augmented Reality", In International Working Conference on Advanced Visual Interfaces, pp.204-207 (2012)
- 16) D'fusion Studio, <http://www.t-immersion.com/>
- 17) Satch, <http://au-ar.jp/>
- 18) G. ReitMayr, T. Langlotz, D. Wagner, A. Mulloni, G. Schall, D. Schmalstieg and Q. Pan: "Simultaneous Localization and Mapping for Augmented Reality", In International Symposium on Ubiquitous Virtual Reality, pp.5-8 (2010)



加藤 晴久 1997 年, 神戸大学工学部電気電子工学科卒業。1999 年, 同大学大学院修士課程修了。同年, KDDI (株) 入社。現在, (株) KDDI 研究所ソフトウェアインテグレーショングループ研究主査。動画像の高効率符号化方式, 動画像認識に関する研究・開発に従事。博士 (工学)。正会員。



消防防災分野における 拡張現実の活用



#E062A1



細川直史[†]

1. ま え が き

GPSとジャイロセンサが搭載されたスマートフォンが普及し、それとともに、拡張現実（AR: Augmented Reality）を利用した観光案内やエンタテインメントを目的としたアプリケーションが数多く登場している。拡張現実感の原理や機能については、すでに本連載で詳細に解説されていることから、本稿では、防災分野での拡張現実の活用事例について解説する。

本解説では、防災マップの避難所や過去の災害履歴などの防災情報を、拡張現実を利用してフィールドで情報提示可能とすることで、地域における災害リスクの理解を支援するツールについて解説する。

2. 避難とリスクコミュニケーション

2011年3月11日に発生した東北地方太平洋沖地震は、20,000人を越える死者・行方不明者、40万棟を越える建物の全半壊などの大きな被害を引き起こした。これらの被害の大部分は津波によるものであり、歴史的にみても数十年という周期で地震と津波に襲われている地域において、多くの防災対策がなされていたにもかかわらず、多数の人が被災したことは、沿岸地域のまちづくりや津波からの避難のあり方について、さまざまな問題を提起した。

地域住民が、地震を感じたり、役場からの避難指示を携帯メールや防災無線の放送で受け取った場合、本当に逃げなければならないと判断するかどうかは、その情報を受け取る人の経験や災害リスクに対する理解に大きく依存する¹⁾。住民が受け取った避難に関する情報が有効活用されて、適切な避難が行われるためには、行政から出される避難指示内容と地域の災害リスクとの関係が、地域住民に予め理解されている必要があり、いわゆる日頃からのリスクコミュニケ



図1 DIGにおいて地域の防災マップを作成している様子

ーションが、逃げ遅れを発生させない重要な要素となる。このような観点から、地域の災害リスクに関する情報を共有する手法として、DIG (Disaster Imaging Game) に代表される図上訓練を取り入れた手法が防災ワークショップなどで活用されている²⁾。DIGは、訓練対象地域の白地図の上に透明シートをかけ、10名程度の住民が、油性ペンなどを用いて、その地域の川や狭い道路、崖崩れ箇所など危険が予測される地帯または事態をシートの上に書込んでいく訓練手法のことである（図1）。住民自身が、地形や町並みの状況を基に、その地域の災害リスクについて自ら考え、どういう行動すれば良いかということに気づいて貰うことが、DIGの重要な目的である。

一般的なDIGは地図を囲む机上の訓練であるが、地域の災害リスクや災害対応への理解を深めるために、さらに、DIGの結果の再確認や足りない情報を探すための現地確認として「街歩き」が行われる場合がある。筆者らは、「街歩き」で使用するリスクコミュニケーション支援ツール（図2）として、拡張現実を用いて、現実風景を表示した携帯電話のディスプレイに防災情報を付加提示することが可能な可視化ツールを開発し、訓練などにおいて評価実験を行っている。

[†] 総務省 消防庁 消防研究センター

"Technology Frontier of Augmented Reality (11): An Application of the Augmented Reality for Fire and Disaster Prevention" by Masafumi Hosokawa (National Research Institute of Fire and Disaster, Tokyo)

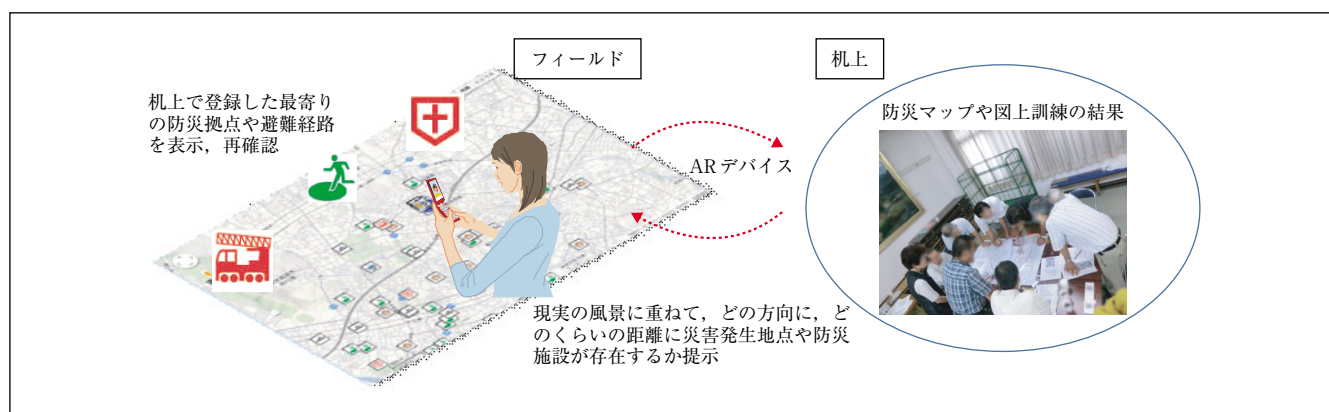


図2 拡張現実を利用したリスクコミュニケーション支援ツール

3. 拡張現実とリスクコミュニケーション

GPS、ジャイロ、電子コンパスなどのセンサを備えたスマートフォンの普及により、セカイカメラなどの拡張現実アプリケーションが誰でも利用できる環境になっており、観光案内や人の所在地やその人に好みにあわせて広告を提示する位置情報サービスなど、さまざまな分野への応用が期待されている³⁾。実際の風景が映し出されたカメラ映像に3次元のグラフィックスを重ねて表示したり、サーバ上に蓄えられた最寄りのエアタグ (AirTag) を、GPS (Global Positioning System) 情報に基づいて検索し、同じくカメラ映像に重ねてスマートフォンのディスプレイに表示する機能は、単に、ゲームや観光の分野にとどまらず、以下の点において、防災分野においても有効であると考えられる。

- (1) 図形や文字などを実際のカメラ映像に重ねて表示できる機能は、防災拠点の種別や名称、津波の浸水深の被害想定結果や避難経路をその場で表示することに利用できる。
- (2) 位置や方位の測位結果に基づいて付近のエアタグ検索・表示できる機能は、自分の立っている場所を中心にして、どの方向に、どのくらいの距離に防災施設が存在するか、もしくは災害が起こっているかを提示してくれる。
- (3) マーカレスのエアタグの活用により、表示看板はもとより、バーコードやRFIDなどを現地に設置する必要がなくなり、安価にシステム構築できる。
- (4) エアタグにURLのリンク情報を加えることで、Webページ上の詳細な説明や映像を提示可能とする。
- (5) スマートフォンやタブレットPCなどで利用できることは、学生など若年層への防災意識啓発に役立つ。

これら拡張現実の機能は、DIGの結果の再確認や新たなリスクへの気付きを目的とした“街歩き”に必要な要素を備えており、地域におけるリスクコミュニケーションの支援ツールになることが期待できる。

4. システム構成

本稿で紹介するリスクコミュニケーション支援ツールの構成を図3に示す。エアタグを格納するサーバアプリケーションとスマートフォン端末で動作する拡張現実アプリケーションからなる。拡張現実アプリケーションは、iPhone3GS以降およびすべてのAndroid端末に対応したLayar (Layar Reality Browser)⁴⁾を利用している。サーバアプリケーションは、Google App Engine (GAE) に基づくもので、所在場所の緯度・経度、アイコン、詳細説明、画像、ホームページアドレスなどの属性を含んだ以下のエアタグを登録可能である。

- (1) 防災マップの情報 (避難所、防災倉庫、病院、警察・消防署などの防災拠点、AEDの設置場所)。
- (2) 過去の災害履歴 (写真、災害についての解説文、関連情報のホームページアドレス)
- (3) 要救助者、火災の発生などの警報情報

サーバアプリケーションは、防災拠点や気づいた危険地点を、ウェブブラウザに表示されたGoogleマップ上で位置を確認しながら1件ずつ入力できる機能と、エクセルなどで編集されたエアタグデータのリストを一括登録できる入力用のインタフェースを備えている (図4)。

また、登録された防災施設を街歩きにおいてLayarで確認するだけでなく、フィールドで新たに発見した防災上重要な施設や危険箇所を、専用のアンドロイドアプリによって、移動ログや説明文などを付加してエアタグとして登録可能である (図5)。

5. 拡張現実による防災情報の提示例

5.1 防災マップ⁵⁾

拡張現実を利用して防災マップを表示した場合の有効性を検証するために、調布市の防災マップ⁶⁾に記載されている、避難所、消防署、警察署、病院などの情報を、エアタグとしてサーバに登録した。図6はエアタグとして登録された防災マップ情報をサーバアプリケーションで地図表示

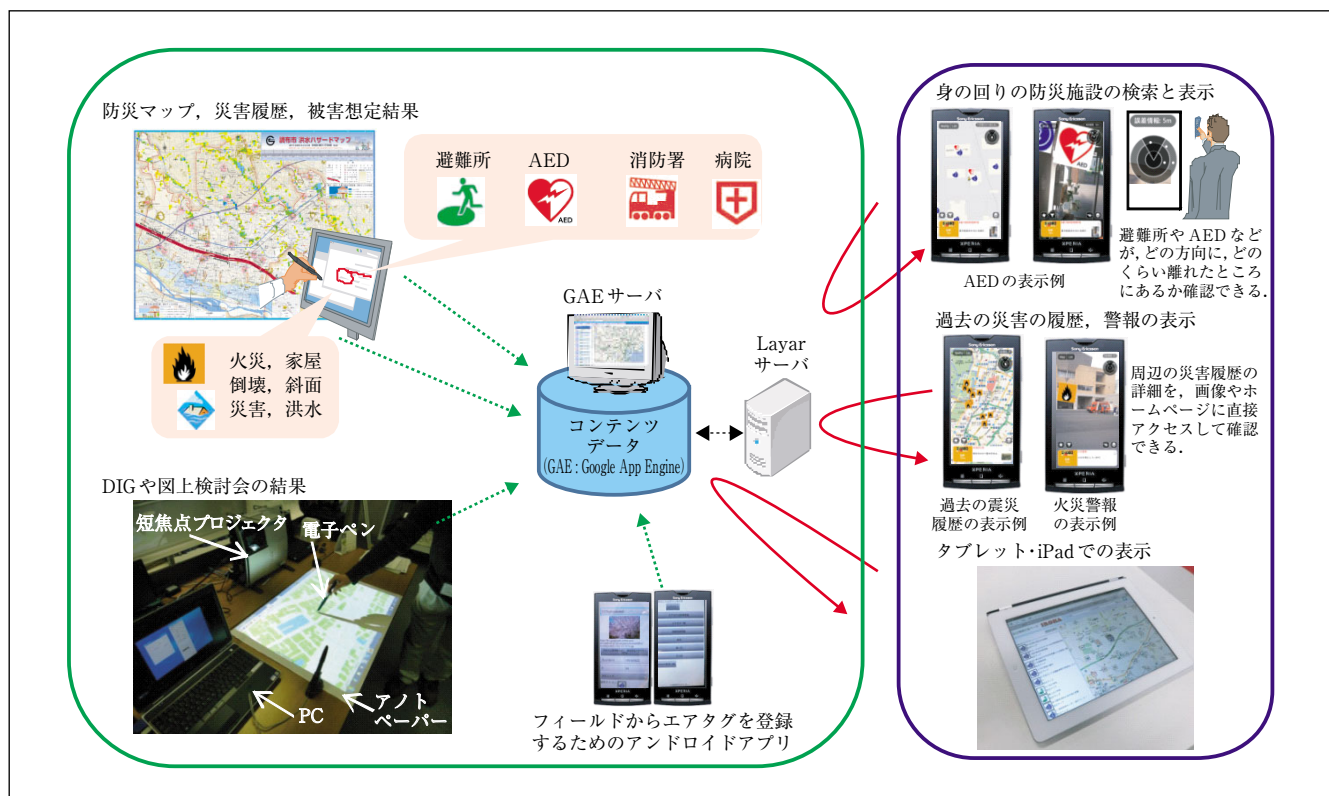


図3 システムの構成



図4 エアタグを登録するためのサーバ画面

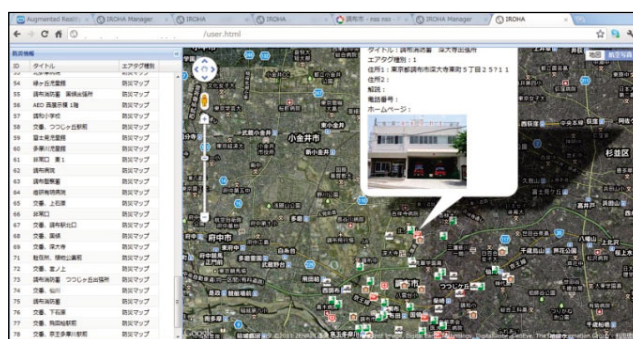


図6 エアタグ登録された防災マップ情報のサーバアプリケーションでの地図表示例



図5 エアタグ登録用のアンドロイドアプリとGAEサーバに登録された移動ログ

した例である。施設の所在を示すアイコンをクリックすることで、画像や詳細説明を表示可能である。

登録されたエアタグ情報とLayarアプリケーションを用いてフィールドで検索・表示することで、エアタグの表示精度ならびにシステムの有効性を確認・検証した。図7はスマートフォンを使った拡張現実の表示例で、左の写真は、携帯電話を風景にかざすことにより撮影された現実の映像の上に情報が重ねて表示されている様子である。スマートフォンは、Sony Ericsson Xperia (OS : Android 2.2, 画面サイズ: 約4.0インチ, 解像度: 480ドット×854ドット)を用いている。図7の中央の画像は、AEDのエア

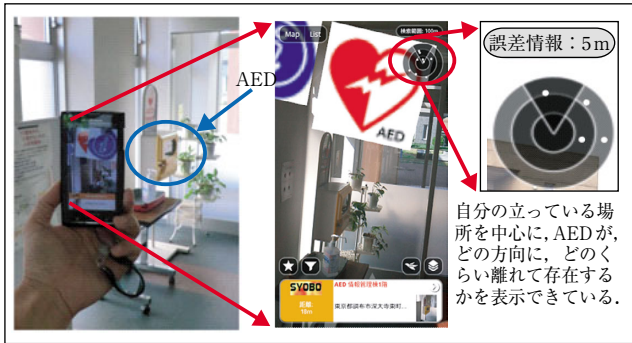


図7 スマートフォンを使った拡張現実の表示例



図8 “お助け隊”が到着した時の様子

タグが、スマートフォンのディスプレイへ表示された画面例である。自分の立っている場所を中心に、AEDのエアタグが、どの方向に、どのくらい離れて存在するかを表示可能であった。

情報提示実験の結果としては、

- (1) 屋外(かつ近くに高いビルが存在しない場所)であれば、携帯電話端末の測位誤差は約5m以内であった。
- (2) 自分が立っている位置を中心として、どの方向に、対象物があるかを確認することが可能であった。
- (3) 対象物までの距離に応じてエアタグアイコンが距離と一緒にスケール変換されて携帯電話のディスプレイに表示されることから、おおよその距離を把握することが可能であった。
- (4) エアタグに登録された対象物の画像やホームページへのリンク機能が、エアタグと現実空間との対応付けに有効であることが確認できた。

これらの結果から、携帯電話端末のGPSや電子コンパスが有効に機能し、自分が立っている位置を中心として、避難所や消防署などの防災拠点がどの方向に、どのくらい離れたところにあるかを把握することが可能であることが確認できた。

5.2 要救助者情報の提示実験

拡張現実を用いることによって、警報の発信場所や要救助者の位置を確認し、それを基にその場所まで駆け付けることが可能かどうかの実証実験を行った。実験は、平成22年12月19日に山梨県中央市における豪雨災害を想定した防災訓練において、要援護者の逃げ遅れが発生し、救助班「お助け隊」が救助に向かうという以下の想定で行われた。

- (1) 山梨大学が開発・整備した地域防災SNS⁷⁾によって、要援護者の逃げ遅れが発生したことによる救助要請の通報が入る。
- (2) 要救助者の自宅位置の座標を、エアタグとしてサーバに入力する。
- (3) スマートフォンのLayarアプリケーションをつかって、要救助者の位置を検索し、スマートフォンのカメラで撮影される映像上に、要救助者の所在を示す

アイコンとその場所までの距離や方向を重ねて表示する。

- (4) 表示された情報だけで、要救助者の発見、所在場所まで駆け付けられるかを確認する。

図8は、お助け隊が到着した時の様子で、状況を確認後、ドアに連絡票を貼り付け、要救助者とともに避難所に向かった。スマートフォンに表示された要救助者の位置や方向の精度については、GPSの精度が比較的保証されている屋外の実験だったので、的確に情報提示がなされ、速やかに現場到着することが可能であった。

図9は、拡張現実を用いた要救助者探索実験について、スマートフォンの画面表示とともに、要救助者の位置と端末の位置および移動経路を地図上に表示したものである。この実験で明らかになった課題としては、

- (1) Layarのカメラモードの方向だけの表示(図9の左側の表示画面)で要救助者の位置に向かう場合は、川やブロック塀などの遮蔽物があった場合にはその都度迂回する必要があった。結果、図9の右側の地図表示モードの併用も必要であった。
- (2) 端末の実際の位置と表示位置にずれが生じる場合があり、測位と再表示の時間的間隔を短くする必要がある。
- (3) 画面に集中することで足下がおろそかになることや、長時間スマートフォンを支えているのは困難であることから、HMD(Head Mounted Display)などUI(User Interface)の工夫が必要である。

この訓練によって、拡張現実を用いることで、要救助者が、自分のいる位置から、どの方向に、どのくらい離れた場所に存在するかを容易に把握できた。特に、要救助者の位置特定が困難と考えられる都市部での使用や、地域の地理を把握していない人が救助に向かう場合において、この仕組みは有効であろうとのコメントが訓練参加者から得られた。

5.3 水害履歴の提示⁸⁾

富士川・釜無川・笛吹川における過去の水害履歴情報を、拡張現実によって、現場付近において表示・確認できるよう



図9 拡張現実を用いた要救助者探索実験について



図10 エアタグ化された富士川水系の水系履歴マップ

図11 エアタグ化された富士川水系の水系履歴マップ
(詳細情報を表示した状態)



図12 スマートフォンに表示された水害履歴
(Layarのカメラモードの表示例)

に、エアタグデータ化した。ここでは、文献9)に記載されている、昭和34年9月(台風15号伊勢湾台風)、昭和52年8月熱帯低気圧、昭和57年8月台風10号による水害の記録から現場写真と解説文を、橋の名前・地名や地元の人証言などから場所を推定してエアタグとしてサーバに登録した。図10は、エアタグ化された富士川水系の水害履歴をサーバアプリケーションのGoogle マップ上に表示したもので、アイコンをクリックすることで、図11に示すように、当時の現場写真とともに説明文が表示される。

図12は、笛吹市役所の前の堤防の上から、過去の水害時の被害写真を拡張現実で表示した例である。この地域に土地勘があるシステム評価の被験者からは、過去の水害での被害状況を具体的にその場で見ることで、水害の危険度を身近に感じることができたとの回答を得た。拡張現実を使うことにより、水害時の水位の高さや水量の多さ、堤防の整備や地形の変化などをその場で比較することが可能であり、フィールドでの水害リスクの理解に有効であることがわかった。

6. 今後の動向・課題

従来の紙の防災マップを活用するためには、地図を読取る能力が求められるが、スマートフォンのGPSやジャイロセンサの機能を活用する拡張現実実装は、エアタグとして登録された防災情報と自分の位置との関係をよりわかりやすく示すことができる仕組みである。ハザードマップの想定震度や予測水位などのハザード情報を、図形やアニメーションなどによって、フィールドでよりわかりやすく提示す

る仕組みとして大きな可能性を秘めている。一方、フィールド実験では、スマートフォンの電池の持ちが悪い、画面が小さく見づらいなどの率直な声も聞かれた。新しい技術や普及途上のスマートフォンを用いた仕組みであり、表示精度や災害現場での運用の可能性について、さらに検討を行う必要がある。現在、実用化が期待されているメガネ搭載型のヘッドマウントディスプレイが利用できるようになった場合は、さらに、拡張現実の防災利用がすすむものと思われる。

7. むすび

拡張現実を防災分野へ活用した事例として、リスクコミュニケーション支援ツールについて解説した。このツールは、まだ試作段階であり、実際の防災訓練での実証実験などをとおして、さらに改良して行く予定である。

最後に、水害履歴データの作成に当たっては、国交省関東地方整備局甲府河川国道事務所の皆様の協力を頂いた。関係者の皆様に対し感謝の意をここに記す。

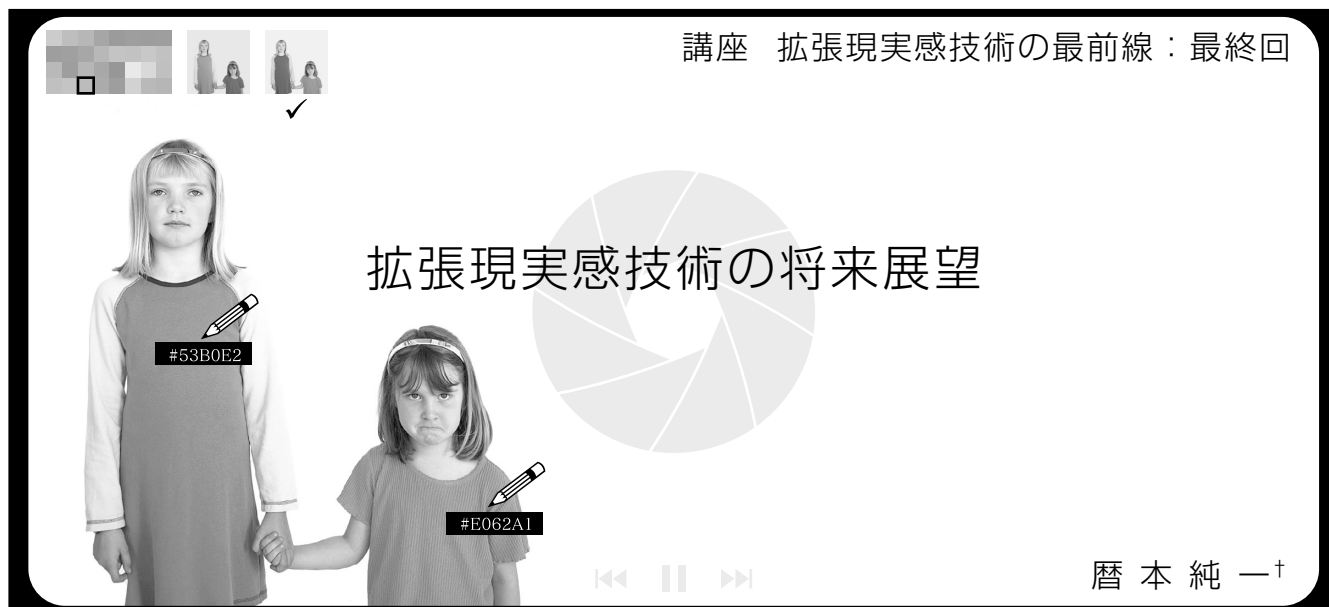
(2012年7月30日受付)

〔文 献〕

- 1) 広瀬弘忠：“人はなぜ逃げ遅れるのか—災害の心理学”，集英社新書(2004)
- 2) 小村隆史，平野昌：“図上訓練DIG (Disaster Imagination Game) について”，地域安全学論報，p136-139 (1997)
- 3) “ARのすべて—ケータイとネットを変える拡張現実”，日経コミュニケーション編 (June 2009)
- 4) <http://www.layar.jp/layar>
- 5) 柴山明寛，武原靖，細川直史：“拡張現実 (AR) を用いた現場における警報と防災情報可視化の試み”，第27回 (平成22年度) 地域安全学会研究発表会梗概集，27，pp.79-80 (Nov. 2010)
- 6) 調布市洪水ハザードマップ，<http://www.city.chofu.tokyo.jp/www/contents/1183592386922/>
- 7) 鈴木猛康，秦康範，佐々木邦明，大山勲：“住民・行政協働による減災活動を支援する情報共有システムの開発と適用”，災害情報，9，pp.46-59 (2011)
- 8) 細川直史：“拡張現実 (AR) を用いたリスクコミュニケーションツールの開発”，信学総大，シンポジウムセッション安全・安心な生活のための情報通信システム，AS-5-6 (Mar. 2012)
- 9) “富士川・釜無川・笛吹川，洪水概要 平成20年度版”，国土交通省関東地方整備局，甲府河川国道事務所



ほそかわ 直史 1987年，徳島大学工学部情報工学科卒業。1989年，同大学院修士課程修了。1991年，自治省消防庁消防研究所入所。現在，総務省消防庁消防研究センター地震等災害研究室長。リモートセンシングデータ画像解析の研究，災害時の情報システムに関する研究に従事，博士(工学)。



1. ま え が き

ここ数年、画像・映像を使った情報サービスの新しい可能性として、AR(拡張現実感: Augmented Reality)が注目を集めている。また、ARを利用したエンタテインメントや、スマートフォンを利用した具体的なサービスの展開が始まっている。本稿では、ARの技術を概観し、現状および将来の可能性について議論する。

2. ARの定義

ARという用語にはいろいろな解釈がありえるが、ここでは「現実世界を何らかの手段で増強、拡張する」ための技術の総称をARと呼ぶことにする。この意味では、ARと、マーク・ワイザーが1990年代に提唱したユビキタスコンピューティングの概念とは近いものがある。

ただし、ARで特徴的な点は情報が遍在するだけでなく、現実世界の視界と電子的な情報を重畳(オーバーレイ)して表示することにある。仮想現実感(ヴァーチャルリアリティ: VR)が、視覚を完全にコンピュータの世界で覆ってしまうのに対して、ARは現実の視界とコンピュータの情報が共存している。すなわち、視覚オーバーレイによって、現実情報が添付され強化されたような感覚を利用者がもつことになる。このような現実に対する電子的なオーバーレイは、SF映画やアニメーションの世界ではすでに頻繁に取り上げられており、「ターミネーター」、「ロボコップ」あるいはコミックの「ドラゴンボール」のスカウター、アニメーションの「電腦コイル」のように、AR的コンセプトのインタフェースをもつものは多い。また、このような実映像と仮想情報の合成は複合現実(Mixed Reality: MR)と呼ばれることもある。

3. 黎明期のAR

ARの歴史は、1965年にユタ大学のアイバン・サザランドが開発した「Ultimate Display」に遡る¹⁾。これは頭部搭載型ディスプレイ(Head-Mounted Display, HMD)であり、VRの元祖なのだが、同時に世界初のシースルー型ARシステムでもあった。

サザランドは、GUIの元祖でもあるSketchpadの研究でも有名であるが、グラフィックスで世界と対話するという概念を2次元から3次元に拡張し、世界初のHMD, Ultimate Displayを開発した(図1)。Ultimate Displayは、両眼にそれぞれディスプレイを配したステレオ画像システムだが、HMDにはロッド(棒)が接続されており、ロッドの関節角を測ることで頭部の位置と方向をメカカルに算出する。HMDはハーフミラーを採用しており、ディスプレイの映像と視界とが光学的に合成されるようになっている。つまり、頭を動かすと、ハーフミラー越しにシースルーで見える現実世界といっしょに、CGの仮想世界も追従して変化する。この機構により、現実世界にあたかもCG映像が浮いているような効果を得ることができた。さらに、手にもった別のロッドを操作すると現実世界と仮想世界のインタラクションができるようになっていた。例えば、CGとして表示されたメニューをロッドで選択するこ



図1 サザランドのUltimate Display

† 東京大学 大学院情報学環/ソニーコンピュータサイエンス研究所
"Technology Frontier of Augmented Reality (Final Study): Augmented Reality: Current and prospects for the future" by Jun Rekimoto (Interfaculty Initiative in Information Studies, The University of Tokyo/Sony Computer Science Laboratories, Inc., Tokyo)



図2 KARMA

とができた。現実世界に電子情報を重畳するというARの基本機能は、1960年代のUltimate Displayですでに試みられていたことになる。

Ultimate Displayは1960年代の画期的な成果であったが、他のARシステムがこれに続くにはしばらく時を要した。そして90年代に入ると、より明確にARやMRを意識した研究が多くなってくる。

「KARMA」(図2)は、コロンビア大学で開発されたARシステムである²⁾。KARMAは、機器のメインテナンスを想定したARシステムで、プリンタを修理するときに、コンピュータが「この部分を引き出せばいいですよ」と指示するCG情報をHMD越しに見ることができる。情報を正確に提示するためには、利用者の視野と対象物(この場合はプリンタやプリンタのペーパートレイ)との位置関係を正確に計測する必要がある。KARMAでは、頭部位置の計測には超音波センサを使い、プリンタと頭の位置や方向を測定し、現実の視界と仮想の映像をハーフミラーで合成している。KARMAの方式は、具体的なアプリケーションをもつARとして画期的であった一方で、対象物のすべてにも超音波センサを取り付ける必要があり、機器が増えるにつれてシステムの複雑さが増大してしまう問題を顕著にした。

1990年代初頭にノースカロライナ大学で開発された医療用ARシステムでは、患者の体内を超音波診断装置によってスキャンし、その情報を医師の視野にオーバーレイする³⁾。ここでも、超音波診断装置、利用者(医師)の頭部位置などを正確に計測する必要がある。

戦闘機のヘッドアップディスプレイもARの先駆的な例といえる。パイロットの現実の視野に、ロックオンされた敵機の情報などを、ハーフミラーで合成してオーバーレイ表示する。このようなハーフミラー合成は、自動車の情報提示手段としても用いられるようになってきている。

4. コンテキストウェアネス

上記の例で示されているように、ARで重要なことは現実世界に応じた情報の提示であり、そのためには対象物と利用者との位置関係をセンシングする必要がある。これには超音波センサ、磁気センサ、あるいは画像認識などの技術が利用できる。

より一般的には、ARのインタフェースで重要なものは、「コンテキストウェアコンピューティング」(状況連動型

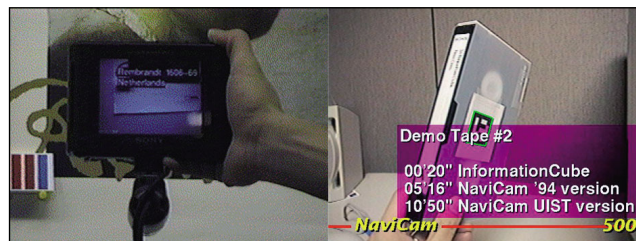


図3 NaviCam:ハンドヘルドARシステム

コンピューティング)という発想である。利用者が今、何をしようとしているのか、どんな場所にいるのか、近くにどんな人がいるのかなど、利用者の周りの状況(コンテキスト)をコンピュータが察知(センシング)し、コンピュータの力で、例えば、データベースから状況に適した情報を瞬時に引き出してくるようなインタラクションである。

コンテキストウェアなコンピューティングに通ずるインタラクションの例として、筆者自身が1980年代に経験しインスパイアされたのは、歌舞伎座のイヤホンガイドだった。歌舞伎座へ行くと、イヤホンガイドをレンタルで利用することができる。これは、歌舞伎を見ながらイヤホンを付けて聞くシステムなのだが、単純な「同時通訳装置」ではない。実際に芝居で演者が台詞をしゃべっているときはほとんど何も言わず、演者が間を空けたときに、そのときの状況(コンテキスト)に応じた説明を音声で流す。例えばこの掛け軸にはこういった歴史的な意義があるとか、時代背景はこうである、といった情報を補足的に説明する。

イヤホンガイドで利用されている技術は単純ではあるが、「芝居の進行に合わせて」というところがコンテキストウェアネスに相当する。この情報を芝居の進行と無関係に聞いてしまうと、ほとんどガイド情報としての意味をなさないのであろう。このようなガイドが、芝居ではなく、現実世界の利用者のコンテキストを理解するように適用できれば、コンテキストウェアのインタフェースになる。

5. モバイルマーカ型AR : NaviCam

コンテキストウェアという発想の下で、筆者自身がソニーコンピュータサイエンス研究所において1993年に開発したのが、世界初のハンドヘルド型ARシステム「NaviCam」(図3)である⁴⁾。ビデオ映像に映っている現実に対して、コンテキストに合った情報の提示を実現する、ビデオシースルー型のARシステムである。当時の他のARシステムと異なり、頭部搭載型ではなく、ハンドヘルド型の機器構成をとった。ハンドヘルド機器の画面越しに現実をみると、そこに情報がオーバーレイされるという方針をとった。これにより、複雑な頭部搭載型の装置を用いることなく、より簡易にARインタフェースを利用することができるようになる。今日、モバイルARとして、スマートフォン上で実現されている数多くのハンドヘルド型ARの世界最初の構成例である。



図4 TransVision:マルチユーザAR

情報を表示させたい対象物にはIDとしてバーコードを貼り付け、それを認識させる方法を採用した。当時の非力な計算処理能力では、直接画像を解析して対象物を理解することが現実的ではないという判断だったが、IDを添付してあることが、そこに情報が存在することを利用者に示すという利便性もあった。NaviCamで採用されたID認識の手法は、のちに広く利用される「マーカ型AR」の源流となった。

その後、NaviCamにジャイロセンサを付加した。バーコードから得た絶対位置とジャイロセンサから得た相対位置によって、空間に矢印などの情報を映すナビゲーションを実現した。また、空間や特定のモノに情報を貼り付けるインタフェースも実現した。例えば、誰かが本棚に対してメールを送っておくと、その本棚の前を通りかかった人はNaviCamによってその情報を取得できるといったことが可能になった。現在のモバイルARで「空中にタグを貼り付ける」操作に相当する。

6. マルチユーザAR: TransVSION

NaviCamの発展形として、複数の人が同じものを共有する仮想空間、仮想作業環境をARシステムとして構築したものが1996年の「TransVision」(図4)である⁵⁾。これはマルチユーザ型ARシステムの最初期の開発例である。

TransVisionは、磁気センサを利用して複数のハンドヘルドARシステムの位置を計測している。これにより、マルチユーザ間で共有できる仮想空間を、現実の風景上にオーバーレイして表示する。このような共有空間をARとして構成する利点は、各利用者の身体行為が自然に統合される点にある。VRでは利用者は仮想世界のみを見るため、他の利用者の身体や作業に必要な行為(例えば何かを指し示したり、特定の方向を向いたり)を共有することが難しい。TransVisionでは利用者の姿もそのまま見えるので、仮想物体を現実世界で共有して作業しているような感覚を得られる。また、描画するCGが単純でも、背景が現実世界で、CGの大きさに空間的な整合性を持たせるため迫真度が増している。CGの「影」を現実のテーブルの位置に合わせて表示すると、テーブルのサイズとの対比でCGがどれくらいの大きさが直感的に把握できるようになる。

TransVisionのようなマルチユーザ型ARシステムは、近年ではプレイステーションポータブル(PSP)のInvisimalsのようなARゲームタイトルに応用されている(図5)。

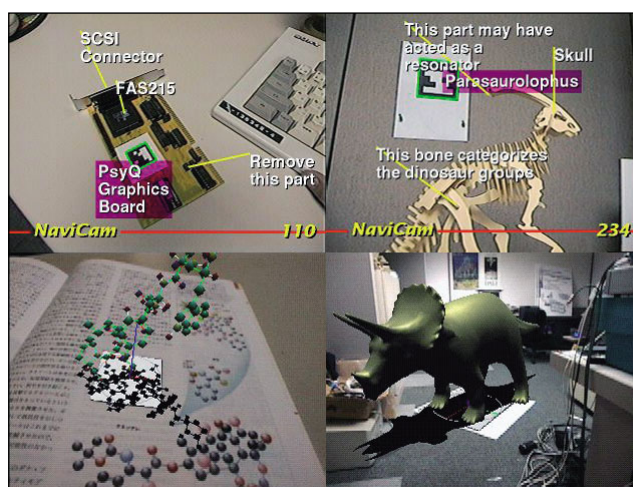
図5 PSPのInvisimals
Sony Computer Entertainment Inc.

図6 CyberCodeによるマーカAR

7. CyberCode:ビジュアルマーカによるAR

NaviCamのID認識によるARの考えを発展させて、カメラ位置に対して正確な情報の重畳を可能にしたのが「CyberCode」(1996年)である(図6)。これがのちに多くのARシステムで利用されている「マーカ型AR」の最初の実現例である⁶⁾。

CyberCodeは、2次元バーコードとして情報をエンコードするが、中に描かれているバーコードからどのような物体かを読取る(ID情報)と同時に、全体の四角形の変形具合から、空間の中でどの位置や角度からカメラが見ているかを逆算する。逆算によって得たカメラパラメータでCGを描画することで、現実世界の位置関係に即したオーバーレイが可能になる。このように、ビジュアルマーカを位置計測とID認識の両面で使うことがCyberCodeの特徴である。

マーカ方式は、ARを実現するための非常に簡易な方法といえる。現実世界にマーカを貼れば、カメラによる画像認識によって3次元空間の中で整合されたARが非常に簡単に実現できる。この発明により、磁気センサのような高価なハードウェアを要せずにARシステムが構築できるようになった。またマーカは、「どの場所に情報が添付され

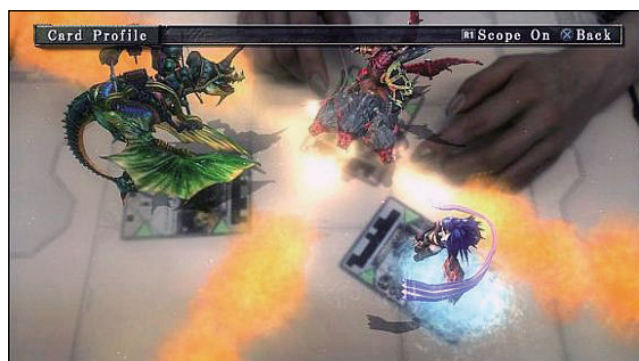


図7 The EYE of JUDGMENT
Sony Computer Entertainment Inc.

ているか」を示す手がかりとしても有効であり、利用者がその場所を見つけるためのガイドとなる。

マーカによってカメラのパターンが得られれば、その空間上にCGを置くことは容易である。例えば絵本の中からCGが飛び出すような演出や、マーカを床においてCGのソファを実物大で部屋の中に配置する家具のレイアウト・シミュレーションといった応用が考えられる。

CyberCodeではじまったマーカ型ARは多くの派生型を生み出し、現在、世界でもっとも広く用いられているARの実現手法である。エンタテインメント、書籍や雑誌のCGによる拡張、広告、レイアウト、シミュレーションなど各種の応用が活発に試みられている。

1998年にソニーは「VAIO C1」というカメラ内蔵の横長型のノートパソコンを製品化したが、そこに搭載されたCyberCodeが世界最初の商用ARシステムであり、CyberCodeカードをかざすと3次元CGが表示されるといった機能が提供されていた。

その後、2007年にPlayStation3で「THE EYE of JUDGMENT」(図7)が、世界初のARエンタテインメントコンテンツで、CyberCodeのテクノロジーをもちいて製品化された。プレーヤが手にとって遊べるトレーディングカードの上にCGがオーバレイ表示される。ビジュアルマーカを使い、ネット対戦ができ、ARを複数のユーザで共有しようという、いろいろなARのエッセンスがとりこまれている。また、カード以外にも指を認識し、指でバーチャルなキャラクタを動かすという技術も入っており、近年のナチュラルユーザインタフェース(NUI)の先駆的な事例にもなっている。

8. ARと可視光通信

ビジュアルマーカ以外にもID情報が発信できれば、ARシステムを構築する手段として使うことができる。例えば、「ID Camera」では、可視光通信を用いている⁸⁾。

可視光通信は、照明光の高速な点滅や変調によって情報を重畳する技術である。照明がLEDになると、LEDの高速点滅にIDの情報をエンコードして乗せることができる。



図8 モバイルARの例
左: Layar (layar), 右: セカイカメラ (tonchidot)

この光を高速撮影できるカメラによって撮影すれば、自然の映像と同時に、LEDの点滅にエンコードされている情報を取出すことができる。LEDのピクセルごとに異なる情報を乗せることも可能である。これをマーカとして使うことで、ARに応用することができる。

可視光通信とARの組合せにより、例えば、未来のネオンサインは、人間に見える視覚パターンを発光するだけではなく、人間の眼にはわからないほど高速点滅することで情報を発信する源にもなるだろう。また、人間に取り付けられる簡単なLEDバッジにより、人間の位置を正確に把握するARシステムが構築できる。

9. モバイルAR

スマートフォンの普及により、携帯電話をモバイルARのプラットフォームとして利用する試みが盛んになってきている。例えば、LayAR⁹⁾やセカイカメラ¹⁰⁾は、利用者が作り出す情報をタグとして空中に置く操作をサポートし、新しいコミュニケーションプラットフォームを提案している(図8)。

モバイルARで使われている技術は、GPSやWiFiセンシングによる位置認識と、電子コンパスによる方角情報である。これらの認識技術の精度限界によって、必ずしも精密な重畳情報の位置合わせが達成できているわけではない。今後は以下で述べるマーカレス方式などとの併用によって、より正確な情報提示が可能になるだろう。

10. マーカレスAR

2000年代に入ると、それまでのマーカ型のARに代わって、現実の情景そのものから位置や対象物を認識する、「マーカレスAR」の研究が盛んになってきている。

この背景には、高精度かつ頑強な画像認識アルゴリズムの発明がある。プリティッシュコロンビア大学のLoweによって開発されたSIFT (Scale Invariant Feature Transform, スケール不変な特徴変換) と呼ばれるアルゴリズムは、対象映像のサイズや傾きが変化しても、安定的に特徴点を追跡することができる¹¹⁾。さらに、特徴点付近の画像特徴量をベクトル化することで、どの特徴点とどの特徴点に対応するかを正確に決定することができる。この特性を利用すると、画像データベース中に蓄積された特

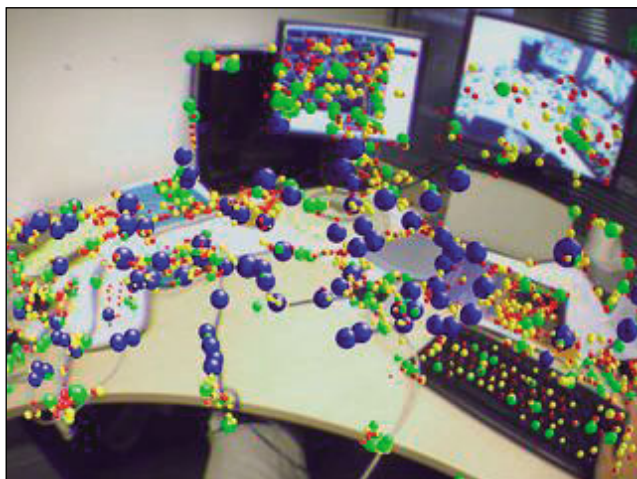
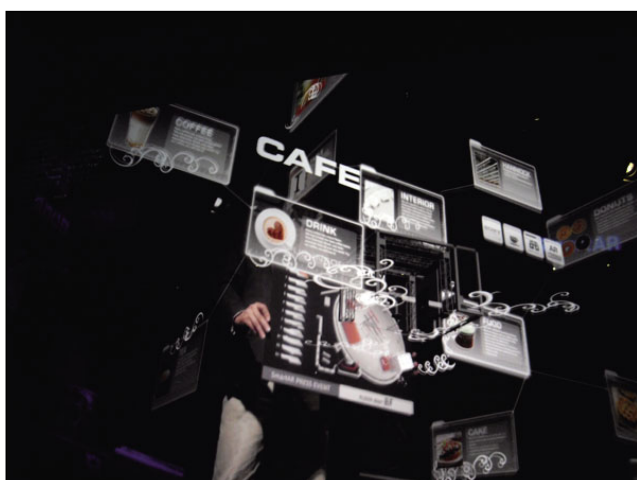


図9 PTAMによるマーカレスAR

図10 SmartAR
Sony

微量と、AR機器によって取得された現実の映像から得られる特徴量を照合し、場所やモノを特定することができる。SIFTは、当初は実時間で利用するには計算負荷が高すぎたが、ここ10年間のCPU性能の発展により、現在では計算量の問題はほぼ解決されている。また、SIFTはその後多くのバリエーションを生み出し、現在ではオープンソースのコンピュータビジョンライブラリーとして広く利用されているOpenCVにも、その発展形であるSURFが搭載されている。これらの環境整備によって、誰でもが容易にマーカレスARの実験を行えるようになった。

ケンブリッジ大学で開発されたPTAM (parallel tracking and mapping)¹²⁾は、マーカレスARの一種であるが、機器の相対的な移動によるオーバレイ情報の変化処理を、計算コストの少ない特徴点追跡部と、計算コストの高い3次元マッピングの二つのプロセスに分割することで、見かけ上の処理速度を向上させている(図9)。

近年では、さらに商用利用を念頭にいたマーカレスARがいくつか登場している。Qualcomm社のAR

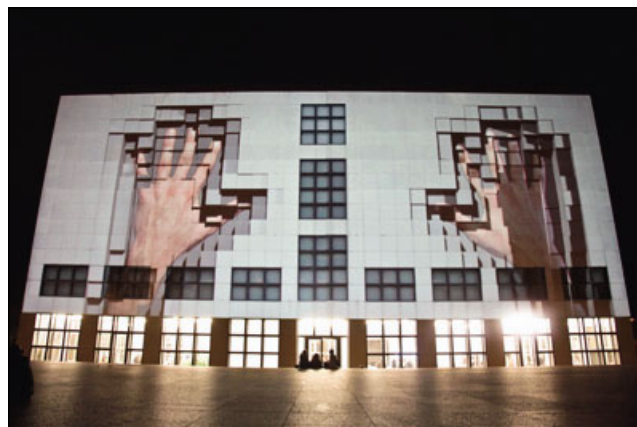
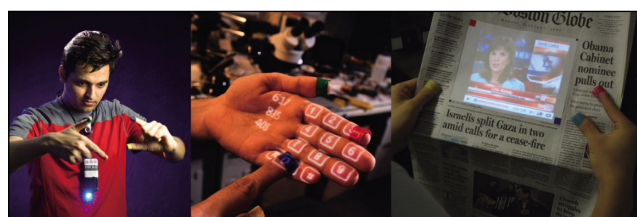
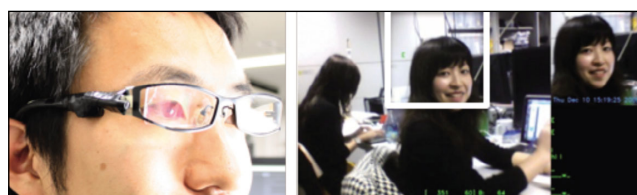


図11 プロジェクション式AR

図12 sixthsense: 身体や周辺環境の情報の拡張
(提供: Pranav Mistry氏)図13 視線認識機能を装備した眼鏡型コンピュータAidedEyes¹⁶⁾

SDK¹³⁾や、ソニーのSmartAR(図10)¹⁴⁾など、スマートフォンの計算資源でもリアルタイムに実行可能なマーカレスARが提案されている。

ただし、世の中にあるすべての風景の特徴量をあらかじめ蓄積することは不可能である。位置認識テクノロジーとの組合せが考えられる。例えば、位置情報によって渋谷のハチ公前にいるということがわかれば、渋谷のハチ公前付近の画像の特徴量だけをデータベースで検索し、マッチングを行えば良い。このように位置情報と画像マッチングを組合せることで、マッチングの計算量を適度に抑えたARシステムが構築できる。

11. プロジェクション型AR

以上の事例は、利用者のもつHMDやハンドヘルドディスプレイを介して現実には情報を重畳するものだったが、プロジェクションマッピング、あるいはプロジェクション型ARは、プロジェクタによって直接対象物に情報を提示する方式である。例えば、ビルの壁面や立体模型の表明に情



報を提示する。高い光度をもつ強力なプロジェクタが利用可能になったことで、近年注目を集めている。

図11は、壁面に別の壁面を表示するインスタレーションで、とつぜん窓の位置が変化したり、壁がくずれたり、建築自体が自在に変化するような視覚表現が可能になっている¹⁶⁾。投影対象物は平面に限らず、自動車模型の表面に投影すれば、任意のカラーリングなどを動的に試すことができる。

12. ARの今後

ARの基盤技術や要素技術はさまざまなものが研究されており、過去10年以上にわたってISMAR (International Symposium on Mixed and Augmented Reality)などの国際会議で盛んに研究成果が発表されている。マーカレス方式は、現在勢力的に研究が行われており、計算量や安定性、精度など、いろいろな点で急速に発展しつつある状況である。

近い将来、家の改装のARでシミュレーション、渋谷で携帯電話を取出すと、周りのビルを全部認識して、AR的に情報を付加する、などは実現可能性が高いサービスである。また、衣服の試着のように人体をトラッキングしてARでオーバーレイすることも多く行われるようになるだろう。

13. ARからAugmented Humanへ

一方、人間は「物を見る」、「音を聴く」、「ものに触る」など、現実空間と常にインタラクション(相互作用)を行っている。そこにコンピュータが介在し、現実世界と人間が相互作用する能力や、知覚能力を強化することがより広義のARといえる。この観点では、視覚のARだけでなく、聴覚のAR、触覚のAR、あるいは人間の身体能力や記憶能力を強化することもARに含めることができるかも知れない。筆者らは、とくに人間の能力の強化・拡張という観点からの研究を進めており、Augmented RealityからAugmented Human (AH)という概念に発展すると考えている¹⁸⁾。

例えば、Googleのような情報検索企業が眼鏡型コンピュータの開発に取り組んでいる(Project Glass)。キーワードを入力して情報を検索するのではなく、利用者の置かれた状況にしたがって情報が自動的に検索され、眼前に提示される。コンピュータやインターネット上の膨大な情報と、利用者の個人的なコンテキストが結合した場合の可能性を示している。デバイスの小型軽量化が進むと、このような眼鏡型のコンピュータを日常生活で実際に利用する社会が到来するかもしれない。

図12は、MIT Media Laboratoryで開発されたsixth-senseとよぶウェアラブルコンピュータのプロトタイプである¹⁵⁾。小型のプロジェクタと認識用のカメラを装着し、利用者の身体や周辺環境をインタラクティブな空間として

利用可能にする。例えば、手のひらはタッチディスプレイのように利用することができ、周りの物体の表面に情報を表示し、操作することができる。また、商品のパッケージを手にとると、その商品に関する情報がパッケージ上に投影される。このようなインターフェースは、利用者の能力を拡張するとともに、周辺の空間を拡張しているとも言えるだろう。

図13は、筆者らが開発したAidedEyesという視線トラッキング機能つきの眼鏡型コンピュータである¹⁶⁾。利用者が何に注目しているかに依存して情報を提示したり、利用者が興味をもった事物をライフログの中でタグ付けしておくことができる。このように利用者とコンピュータによる計算能力がより緊密に連携し、利用者の能力を自然に拡張する方向に、拡張現実感技術が発展していくと考えている。

(2012年9月18日受付)

〔文 献〕

- 1) I.E. Sutherland: "The Ultimate Display", Proceedings of IFIP Congress, pp.506-508 (1965)
- 2) S. Feiner, B. MacIntyre and D. Seligmann: "Knowledge-based augmented reality", Communications of the ACM, **36**, 7, pp.52-62 (1993)
- 3) Bajura, Michael, H. Fuchs and R. Ohbuchi: "Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient", Proc. SIGGRAPH '92, pp.203-210 (1992)
- 4) J. Rekimoto and K. Nagao: "The World through the Computer: Computer Augmented Interaction with Real World Environments", Proceedings of UIST'95, pp.29-36 (1995)
- 5) J. Rekimoto: "TransVision: A Hand-held Augmented Reality System for Collaborative Design", Virtual Systems and Multi-Media (VSMM) '96 (1996)
- 6) 暦本純一: "2次元マトリックスコードを利用した拡張現実感システムの構成手法", インタラクティブシステムとソフトウェアIV, 近代科学社 (1996)
- 7) N. Matsushita, D. Hihara, T. Ushiro, S. Yoshimura, J. Rekimoto and Y. Yamamoto: "ID CAM: A Smart Camera for Scene Capturing and ID Recognition", in Proc. International Symposium on Mixed and Augmented Reality (ISMAR 2003) (2003)
- 8) <http://www.layar.com>
- 9) <http://sekaicamera.com>
- 10) G. Klein and D. Murray: "Parallel Tracking and Mapping for Small AR Workspaces", in Proc. International Symposium on Mixed and Augmented Reality (ISMAR'07) (2007)
- 11) <http://developer.qualcomm.com/dev/augmented-reality>
- 12) <http://www.sony.co.jp/SonyInfo/News/Press/201105/11-058/>
- 13) <http://www.urbanscreen.com/>
- 14) The 2nd International Conference on Augmented Human, <http://www.augmented-human.com>
- 15) <http://www.pranavmistry.com/projects/sixthsense/>
- 16) Y. Ishiguro, A. Mujibiy, T. Miyaki and J. Rekimoto: "Aided Eyes: Eye Activity Sensing for Daily Life", Proc. Augmented Human (2010)



暦本 純一 1986年、東京工業大学理学部情報科学科修士過程修了。1994年より、(株)ソニーコンピュータサイエンス研究所に勤務。2007年より、東京大学大学院情報学環教授兼ソニーコンピュータサイエンス研究所副所長。ヒューマンコンピュータインタラクション、拡張現実感、テクノロジーによる人間の拡張に興味を持つ。1990年、情報処理学会30周年記念論文賞、2003年、日本文化デザイン賞、2007年、ACM SIGCHI Academy等を受賞。理学博士。