

NII Today

National Institute of Informatics News

ISSN 1883-1966 国立情報学研究所ニュース

No. **57** Sep.
2012



国立情報学研究所
National Institute of Informatics

NII Interview | 「マルチメディアセンシング技術」で、
未来のコミュニケーションを切り開く

NII Special1 | 最新の音響信号処理技術で広がる
「音のセンシング」

NII Special2 | テレビ放送をセンサーとして分析する

That's Collabo.1 | 「見る」から「視られる」への転換が生む
新たな価値

That's Collabo.2 | 自由視点の3D映像を可能にする
イメージビジュアルコミュニケーション

特集

マルチメディア・センシング



「マルチメディアセンシング技術」で

カメラやマイクロフォンをはじめとしたマルチメディアセンサーは、この十年ほどの間に技術革新が進み、コストも大幅に下がってきた。こうしたセンサーが社会の隅々にまで普及していったとき、社会にどのような新しい価値がもたらされるのか。その最新動向と今後の応用分野について、チョン・ジーン准教授と東倉洋一編集長（特任教授）にお話を聞いた。

大量のマルチメディアデータを収集し、五感の特徴に合わせて加工

元村 「マルチメディアセンシング」とはどのようなものでしょうか？

ジーン カメラで撮った画像、マイクロフォンで録った音声など、センサーで受け取るデータはすべて信号です。これらを処理して新しい価値を生み出すことがマルチメディアセンシングと考えています。信号というのは生の状態、それが加工されて情報になっていくのです。

元村 もともと「情報」という言葉は「報」と書き、純粋に状態を知らせるものだったそうです。それが「情け」の情報に変わったのは、見た人の分析が加わったからだと言われています。つまり、状態を表すデータがいくらあっても、それを上手に処理しなくては意味を持たないということですね。

ジーン その通りです。マルチメディアセンサーから収集されるデータは「ビッグデータ」になっており、処理しなければただのノイズです。どうやって大量のデータを処理するかが課題なんです。



東倉 洋一

Yohichi Tokkura
国立情報学研究所 特任教授
NII Today編集長

元村 最近、そのビッグデータという言葉をよく聞きます。従来は処理し切れなかったような膨大なデータにブレークスルーを見出し、新たな価値を創り出すものと言われますが、マルチメディアセンシングで大量データを処理するためのブレークスルーはどのようなところにありますか？

ジーン マルチメディアデータは、カメラやマイクを入力源として、人間の目とか耳とか五感で受け止められるものなので、その特徴に合わせて処理すればいい。例えば、一つの画像でも視線があっちこっちへ行きます。だから集中して見ている部分は細かなところまで見えるけど、それ以外はぼやけていて気づかないのです。

元村 つまり、目を引く部分だけを丁寧に処理すればいいということですね。

ジーン そうです。光っている物や、動いている物などには人の目が行きやすい。そういった五感の持つ特徴に合わせて集中的に処理していくと、大量のデータもうまく処理できるというわけです。（P8-9参照）

元村 今、ジーン先生がいちばん興味を持っている研究テーマはどのようなものですか？

ジーン 私の専門は画像の圧縮と伝送です。今、ネットでも動画は見られるけど2次元なんです。それでは立体感や臨場感を感じることはできないので、3次元で再現するための画像圧縮技術を研究しています。（P10-11参照）

収集した大量データをいかに処理するかがポイント

元村 ほかに上手に処理する方法はありますか？



チョン ジーン

Cheung Gene
国立情報学研究所
コンテンツ科学研究系 准教授

ジーン 大量データを処理するにはスピードが重要です。アルゴリズムが人間にとって望ましい出力だけを着目し、無駄な演算をなくしたら、スピードも効率も上がります。防犯カメラも今はいろいろな所についている。非常に安価なカメラで撮れるが、解像度はあまり高くない。だけど、後から上手に加工することで解像度を上げることができるのです。それが新しい活用法に結びついていくと思います。

元村 素材を上手に処理することで、より鮮明さが際立った映像になるということですね？

ジーン 良くなります。1枚1枚撮っている画像を時間軸で関連づけると解像度が上がります。いろいろな角度から撮っている

、未来のコミュニケーションを切り開く



元村 有希子

Yukiko Motomura

毎日新聞社 科学環境部 副部長

九州大学教育学部卒。1989年毎日新聞社入社、2001年より科学環境部。日本の科学技術のあり方を理系人の現状を描きながら検証する連載「理系白書」を立ち上げる。2006年第1回科学ジャーナリスト大賞受賞。著書に「理系思考」など。

画像に、新しい角度からとっている画像も関連づけして、合成することができます。多様なデータでどういうふうな全体像を作り上げていくかがポイントです。

元村 先ほど人間の五感に着目することが大事とおっしゃっていましたが、今までの技術ではそういうメリハリの効いた処理はできなかったのでしょうか。

ジーン 例えば、コンピューターは撮影された画像を単なる数字の羅列としか受け止めません。でも、人間は違います。そこに机があって、人がいてという理解ができるんです。

東倉 人間は画像を見れば何が映っているかすぐわかります。だけど、機械にはなかなか理解できない。つまり、描かれている

物が何であるかを意味づけすることは非常に難しいのです。

身近なところから始まっている マルチメディアセンシング の応用

元村 今までご説明いただいた技術が応用されていくことで、私たちの社会はどう変わっていくのですか？

ジーン “行動認識”がわかりやすい例だと思います。例えば、防犯カメラに映っている人の顔を自動認識できることはもちろん、行動パターンも認識できます。だから、いつもの人と違う歩き方の人が映ると怪しいとわかります。

東倉 先日逮捕されたオウム事件の容疑者で言うと、防犯カメラに顔は映ってなくても、歩き方の特徴から95%くらいの確率で犯人だと認識できるのです。ですから、ある組織に属する人たちの歩き方や行動をパターンとして記録しておけば、それらと照合されない歩き方をする人が映ったときに部外者と判定できるわけで、いちいち身分証を確認したりする必要もなくなります。これはすでに実用レベルの技術です。

元村 最後に、マルチメディアセンシングが開く可能性と未来について、東倉先生の展望をお聞かせください。

東倉 センサーが非常に発達して、いろいろな情報が取れるようになってきました。カメラやマイクだけでなく、ケータイやカーナビだって一つのセンサーです。例えば、最近のケータイやスマホは、音声・画像・映像情報はもちろん、GPSによる位置情報、ツイッターによるテキスト情報などもリアルタイムに発信できるマルチメディアセンサーと言えるわけで、これを1人が1台ずつ持ち歩いているのです。また、カーナビも、人間がクルマに置き換わったようなもので、センサーとして同様な特徴を持っています。東日本大震災の直後に、カーナビからの発信でクルマが走っている位置をつ

かんで、道路の復旧状況を知ることができたのは、カーナビのセンサーネットワークを効果的に役立てた好例と言えます。

元村 技術的課題があるとしたらどのようなものですか？

東倉 第一の課題は、多種多様なマルチメディアデータを、いかに横断的かつ時系列的に解析し、それを価値化して表現するかということ。言葉を換えれば、画像であれ、文章であれ、音であれ、その価値がいちばんよくわかる手段で、人間の五感に訴える方法を開発していくことです。ここで述べたように、「防犯」「減災」などは、社会に「安全・安心」を提供する大きな価値です。

第二の課題は、マルチメディアセンシングではビッグデータを相手にしないといけないということ。そのデータ量は10の15乗以上あると言われる。そこには計算量の壁とアルゴリズムの壁などがまだあります。第三の課題は、センサーから得たデータを「誰が、何のために使うのか」ということ。これは発信者や取得者の権利やデータのセキュリティに関わる大きな問題で、今後、社会のコンセンサスを得ていかなければならない課題です。そういう壁を乗り越えることで、マルチメディアセンシングは総合的な研究の出口として新たな応用分野を広げていくと考えます。

インタビューの一言

マルチメディアセンシングとはまだ一般に馴染みのない言葉ですが、いろいろな要素技術をどんなふうに組み合わせるかという概念であるという印象を受けました。一つ一つの端末の性能、それを連携するネットワーク、あとは応用法。プライバシー保護にも配慮が必要なので、いろいろな分野の人達が連携して成り立っていく総合的な技術であってほしいです。50年前は電話で遠くの人と話せるだけでうれしかったのが、それが映像になり、さらにリアルな3Dコミュニケーションになっていく。バラ色の未来の一端を見せてもらった気がします。

最新の音響信号処理技術で広がる

街中には防犯カメラが普及し、特定人物の追跡などさまざまな分野で応用されるようになった。このように、現在のセンシング技術は画像が主体となっているが、一方で、音情報も、視覚情報と同様、センシング技術において重要な役割を果たしている。ここでは、複数のマイクロフォンから特定の音響信号だけを取り出して、議事録作成やセキュリティシステムなどへの応用が可能な「マイクロフォンアレー」技術の研究を行う小野順貴准教授に、最新の「音のセンシング」について話を聞いた。

音を空間的に聞き分ける 機械の耳を目指して

ほとんどの動物は音を聞き分ける感覚器である耳を2つ備えており、両方の耳で取得した信号の時間と強さを比較して、音の到来方向を判断したり、ある方向から到来する音に注意を向けたりしている。聴覚は、野生動物にとって、視力の働かない暗闇で敵を察知したり、餌をつかまえたりする時に欠かせないもの。そして、こうした聴覚のようなセンシング機能を実現するのが「マイクロフォンアレー」技術を研究しているのが小野順貴准教授(情報学プリンシプル研究系)である。

「NIIで私が研究しているマイクロフォンアレーのアレーとは、並べたものという意味です。つまり、マイクを複数個並べることで、音を空間的に聞き分ける優れた音響センシング機能を作り出し、人工的なシステムやロボットに応用することを目標としています」と、小野准教授は語る。

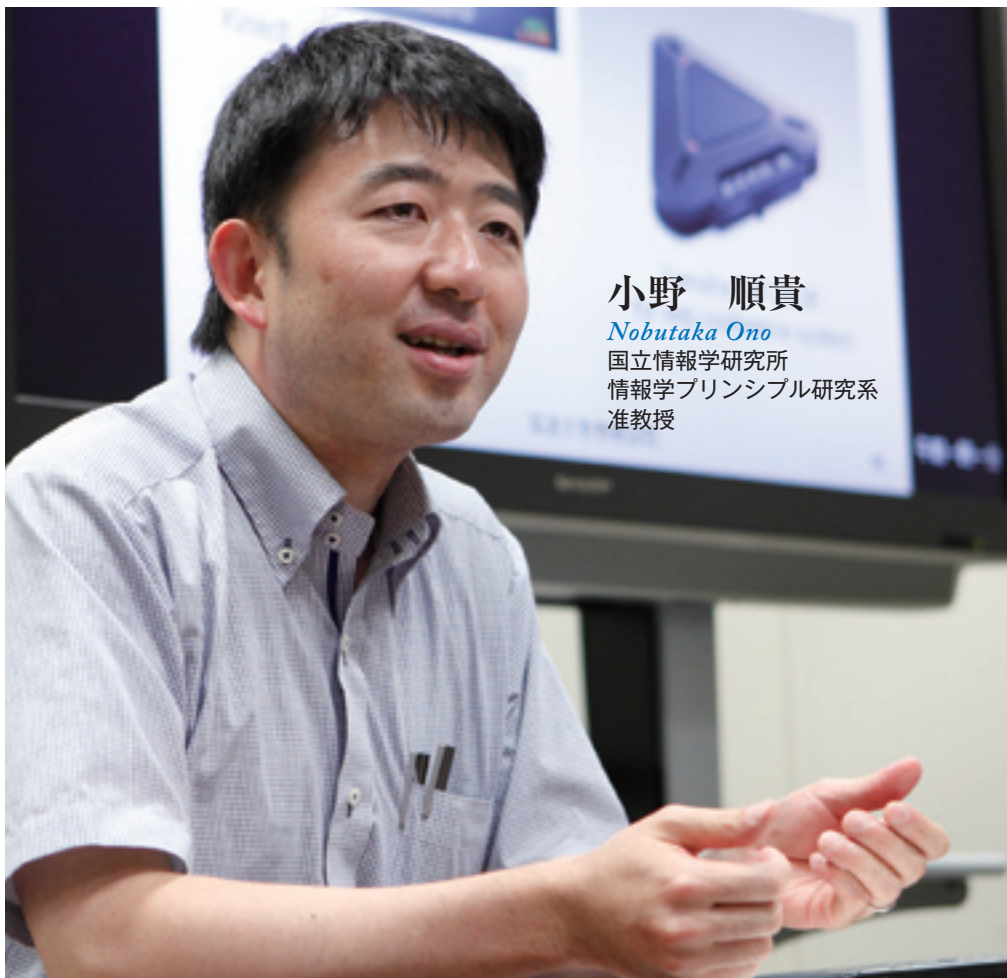
音声認識技術の開発は歴史が古く、現在も多くの研究者が研究を続けている。その中で小野准教授が注目するのが、信号処理によって特定の音声だけを強調する技術だ。例えば、人間の聴覚では簡単に聞き分けられる複数の人の声が、レコーダーの録音だと混ざり合って識別できないことはよくある。マイクロフォンアレーは、複数のマイクロフォンを並べて録音した音声を信号処理し、正確な音の位置特定や音声の抽出を可能にする技術なのだ。

人間の聴覚機能の一つが、こうした音や声の認識だとすると、もう一方で自分の周りに危険な事象がないかを検知することもまた、聴覚の重要な役割の一つで

ある。ここでも、マイクロフォンアレーが活躍する。ユニークな実用例としては、アメリカのベンチャー企業が開発した「Shot Spotter」というシステムがある。これは、街中に多数のマイクを設置し、銃声などの異常音を検知した際に、ただちに発生位置を特定して警察や警備会社に通報するというシステムだ。いわば防犯カメラの音声版である。こうした音によるセキュリティも、これからは大きな効力を発揮していくと、小野准教授は考えている。

混ざった音の中から 望みの音だけを取り出す

小野准教授が現在研究しているのは「ブラインド信号分離」という技術だ。ブラインドとは、音の方向がわからないという意味。従来、放送局などで特定の人物の音声をクリアに録るためには、非常に指向性の強いマイクを人物に向けるという方法をとっていた。これに対してブラインド信号分離技術を用いれば、複数のマイクロ



小野 順貴

Nobutaka Ono

国立情報学研究所
情報学プリンシプル研究系
准教授

「音のセンシング」

フォンで録音した不特定の混ざり合った音の中から、望みの音だけを録音した後に分離して、自由に編集や加工することができる。学会会で娘のピアノ演奏を録音していたら、隣の人のくしゃみの音も入ってしまった。そんな時でも、後から欲しい音源だけを取り出すことができるという便利な技術と言える。

これまでも、こうした重複して録音された複数の音源を、コンピューターにより分離させる技術は研究されていたが、演算処理に時間がかかっていた。小野准教授は新たな高速アルゴリズムを開発することで、従来約50回必要だった演算を、10回程度にまで短縮することに成功した。これにより、スマートフォンなどのデバイスにも音源分離アプリケーションを実装することが可能になったという。

非同期マイクロフォンアレーが切り開く新しい音のセンシング

これまでのマイクロフォンアレーでは、複数のマイクロフォンで録った音の信号から、音の位置を推定したり、特定の音だけを抽出するには、録音信号を完全に同期させることが不可欠だった。例えば、マイクを2つ並べて、どの方向から音が来たかを調べるには、1秒間に340メートル進む音の時間差を、1万分の1秒単位で検出する必要がある。そのため、完璧な時間的同期がなされるよう、音をデジタル信号に変換する変換器が必要となるのだが、この部分に非常にコストがかかり、実用システムにおけるマイク（録音チャンネル）数の大きな制約となっていた。この点について、小野准教授は次のように語る。

「複数の携帯電話やパソコンについている、同期していないマイクロフォンで録音し、後で信号処理により同期させることで、音源位置の推定ができるようにならないかと研究を始めたわけです。音を強調したり、推定したりする性能は、マイクの数が多い

■iPhone用ブラインド信号分離アプリ



ステレオマイクを装着したiPhoneで、左右に位置する2人の声を同時に録音。アプリにある「セパレーション」ボタンを押すと、混ざり合った音の信号をすぐさま分離して、別個の独立した音声として再生できる。議事録作成やライブ録音など、さまざまな分野への応用が期待される。

ほど高くなります。幸い、ICレコーダーの低価格化やスマートフォンをはじめとしたモバイル端末の普及により、研究が非常に進めやすくなっています」。

スマートフォン時代に期待される様々な応用

非同期マイクロフォンアレーの研究により、スマートフォンを活用したさまざまな応用が開けてくる。その一つが議事録の聞き起こしだ。

「例えば、会議の出席者全員が自分のスマートフォンなどで会議音声を録音します。会議中、同時に話す人の音声为重なったり、雑音が混じったりするかもしれませんが、しかし、会議が終わった後で録音信号をインターネット上のサーバーにアップロードをすれば、サーバー側で自動的に時間の同期をとって、各人の発言を強調し、議事録を自動的に作成してくれるシステムを目指しています」と、小野准教授は解説する。

また、音楽分野では楽器演奏の録音に威力を発揮する。演奏した楽曲を録音後に修正する必要がある場合、従来は音が混ざらないよう、ピアノやギターなど楽器ごとに音源を録音する必要があった。これが、非同期マイクロフォンアレーにより、スマート

フォンを数台並べて録音するだけで、混ざった音を後から分離して、まるでスタジオでプロが録音したように、各楽器の音の修正・加工ができるようになるかもしれない。

さらに、気象情報システムへの応用なども検討しているという。

「例えば、雷が発生したら、専用アプリを入れたスマホから、同時に音の情報を集めてサーバーで処理することで、雷の発生地点を推測する観測システムなどが考えられます。その予備実験として、花火大会に複数のスマホを持ち込み、異なる位置からの録音データを集めることで、どの場所で花火が鳴り、録音をした人たちがどの位置にいたのかを特定する実験を計画しています」と、小野准教授は語る。

「非同期マイクロフォンアレーは、同期信号技術に比べて信号分離に手間がかかるなど、まだまだ発展途上の研究ですが、マイクロフォンの数を増やすのが非常に容易で、配置も自由に行えるというメリットを生かして、さまざまな分野での実用化を目指していきたい」（小野准教授）。

スマートフォンやマイクロフォン付きデバイスを1人が1台ずつ持つ時代が訪れつつある今、ネットワークを活用した音のセンシング技術の可能性はますます広がっていくに違いない。

（取材・構成 浅川 仁）

テレビ放送をセンサーとして分析する

テレビ放送は視聴者にとって重要な情報源だが、同時に、社会や文化、経済などの動向を観測するセンサーとしての機能も持ち合わせている。例えば、ニュース番組は世の中で起きている出来事のセンサーとなるし、コマーシャルをウォッチすることにより、景気動向や企業活動を分析することもできる。そこで、テレビ放送ならではのリッチなコンテンツを解析し、センサーとしての利用を目指すNIIの研究最前線をレポートする。



佐藤 真一

Shin'ichi Satoh

国立情報学研究所
コンテンツ科学研究系 教授

人びとに役立つ情報を 膨大なテレビ放送記録から 取り出したい

NIIで視覚情報の研究に携わる佐藤真一教授(コンテンツ科学研究系)が、テレビ放送を研究対象として注目し始めたのは2000年のこと。きっかけは、記録した数多くの番組から必要な情報を検索して抽出できるようになれば、新たな価値を紡ぎ出すことができるようになるはずと考えたからだ。

そのシステムが本格稼働したのは2001年8月。立ち上げ当初は、NHKのニュース番組「ニュース7」をクローズドキャプション(文字字幕放送)付きでアーカ

イブしていた。その後、記録メディアの低価格化に伴い、2009年8月からはシステムを拡充し、在京7放送局の全番組を24時間録画できる体制を整えた。

「テレビ放送をセンサーにするという着想は、2011年度に文部科学省のプロジェクト『サイバー・フィジカル・システムのためのフィージビリティ・スタディー(費用対効果調査)』に関わる機会があって生まれました。それは、防犯カメラ映像や全国の電力網使用状況データなど、サイバー空間にあるセンサーを用いて、現実世界で起きている事象を解析しようという試みです。プロジェクトに携わる中で、このシステムをテレビ放送にも応用できないかと思いついたのです」と、佐藤教授は振り返る。

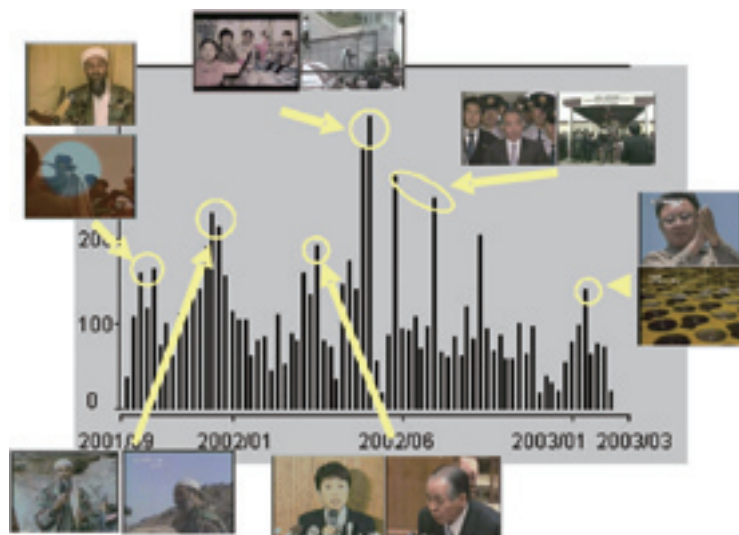
テレビ放送の解析結果から 新たな価値を発見する

映像の検出にはいくつかのアプローチがあるが、その一つに映像の繰り返しパターンがあると佐藤教授は語る。

「2002年の日韓ワールドカップで日本がチュニジア戦に勝利した際に、宮本恒靖選手のフェイスガード姿が連日のように放映されました。従来、画像解析の世界では、映像の中に何が映っているかを機械的に分析することは困難な作業でしたが、テレビ放送の繰り返し映像では非常に簡易な処理で抽出できます。映像が繰り返される頻度、期間や発生パターンを分析すると、どのような話題が社会に提供されたのかがわかるというわけです」。

また、佐藤教授の研究では、放映されるテレビコマーシャルをすべて記録する実験にも着手している。録画された大量の映像の中で、繰り返し現れる15秒と30秒のまとまりをコマーシャルであると定義。超高速のアルゴリズムを開発して、民放5チャンネル3年分のCMをわずか1週間で解析できるようにした。こうした分析は広告代理店でも行っているが、手作業なので放送から1カ月を経ないと解析が終わらない。NIIのシステムはこれを高速化した結果、例えば、ある自動車メーカーのCMが1週間の放送時間帯中、土曜深夜に集中してオンエアが多いことがわかれば、その時間にテレビを見る確率の高いビジネス向けであると見て取れる。あるいは、2011年以降、日中の時間帯にお菓子のCMが激減していることがわかり、菓子業界に与える景気の影響が原因ではないかと推論できる。このように長期間にわたるCMの全記録を残すことは、経済や景気

■ マルチメディアアーカイブから新たな価値を発見する



延べ1万時間にわたるニュース映像の中から、繰り返し現れる重要ショットのパターンを解析。出現頻度や時間帯などから、通常のテレビ視聴からは見えてこない、情報の潜在的価値を導き出すことができる。

の動向を探るセンサーとして非常に有効であると言える。

複数メディアとの比較、 他分野との共同研究も推進

ニュース番組を解析すれば、テレビ局がどのような情報を発信しようとしているかをうかがい知ることができる。例えば、東日本大震災発生後に論議となったのが、テレビ局は果たして情報を正しく伝えているかということだった。そこで、佐藤教授はニュース番組を定点観測して、Web上の有力な情報源であるtwitterの膨大なつぶやきと比較することで、テレビが提供するニュースに偏りはないか、Webでは盛り上がるのにテレビで放送されないのはなぜかなど、センサーから得られるデータを通してメディアの関連性も研究してきた。

また、佐藤教授のプロジェクトでは、Web上の情報解析についても研究を進めている。それが、文部科学省主催の「多メディアWeb解析基盤の構築及び社会分析

ソフトウェアの開発」である。東京大学の喜連川優教授・豊田正史准教授、早稲田大学の山名早人教授と協力しながら、Web情報を10年間にわたって収集し、研究結果の連携をはかっている。

「例えば、野田内閣をテーマにした研究では、Webアーカイブとテレビ放送映像のそれぞれから画像、テキストを抽出し、報道される首相の言動がメディア間でどのくらい温度差があるかを解析したりしています。このようにWeb情報と放送映像を組み合わせることで解析することによって、社会で何が起きているかをより深く分析する可能性を探っています」と、佐藤教授は付け加える。

テレビ放送の価値向上の 可能性にも挑戦していく

テレビをセンサーにするこうした研究は、技術的にどのような段階にあるのだろうか。佐藤教授はこう語る。

「視覚情報に何が映っているかを解析する研究は40年ほど前から始まっています

が、人間に近いレベルの理解はコンピューターにとっては難しい作業です。今のところ、顔情報の認識については、一部突出して技術が進んでいますが、未知の画像を与えられて、これは何々ですと自信をもって答えられるコンピュータープログラムはいまだ存在しません。さまざまな表現があり得る画像を、無数に存在する概念へと認識する手法について、多くの研究者が日夜試行錯誤しているのが現状です」。

多様な視覚表現・無数の概念が混在するテレビ放送を大規模に解析することにより、画像の意味解析技術のブレークスルーが得られるのではないかと佐藤教授は考えている。画像の意味解析の研究が特段の進展を得れば、一般のテレビ視聴者にも、必要なときに必要な情報が提供できるようになるだろう。

「将来的には、受け身だったテレビ視聴が、Web ページと比較したり、他チャンネルの情報を同時に取り込んだりできることで、より俯瞰的な情報を視聴者が利用できるようにしたい。そうならば、放送局はより質の高いコンテンツを提供するようになり、ひいては、視聴者がさらにレベルの高い情報を享受する世の中になるのではないかと考えます」(佐藤教授)。

教授自身、テレビ放送を見るのが好きで、テレビには優良なコンテンツを提供するポテンシャルがあると確信している。視聴率だけを優先した使い捨てのような番組ではなく、長期的な視野に立った映像価値を創り出し、テレビ復権の一助となる研究をしたいと言う。

佐藤教授が最終的に目指すもの。それは、人間がテレビ放送から読み解ける内容のすべてを、コンピューターも自動的に理解できるようにすることだ。それが現実となったとき、「テレビコンシェルジュ」ともいうべき究極の映像情報アシスタントが誕生するのかもしれない。その日が来るのを待ち望みたい。

(取材・構成 浅川 仁)

「見る」から「視られる」への転換が生む

視線を向けるだけでディスプレイ上の情報が操作できる。隠れた危険やトラブルの種に自然に目が行き届く公共施設が実現する。これからの情報デザインに創造的な影響を与えるのが、NIIと東京大学が共同研究を進める注視推定・注視誘導の技術だ。膨大なメディアセンサーから得られる情報をもとに、人の視覚的注意の仕組みを解明し、人の行動を紐解く研究開発の現状と展望を二人の研究者に伺った。

人の注視をセンシングする、 世界初の研究開発

まず、お二人が共同研究を始められた経緯について伺えますか？

杉本 十数年前、私がある研究プロジェクトに参加していた時に、佐藤先生と出会いました。そして2002年、私が京都大学からNIIに移った時、特定領域情報学という別のプロジェクトで佐藤先生との共同研究を始めたのです。このプロジェクトで行った研究は、環境に埋め込まれたカメラやセンサーからの情報をもとに、人の身振り手振りからジェスチャを認識したり、顔の動きを高速に計測するというもので、注視推定の研究につながるテーマでした。2009年になって科学技術振興機構のCRESTプロジェクトとして採択され、「日常の生活空間における注視の推定と誘導」を情報インフラの構築に応用するための共同研究をスタートしたのです。

注視推定とは聞き慣れない言葉ですが、どのような研究をされているのでしょうか？

佐藤 人が次に何に注目するかを予測す

ることで、人の内面、次の行動や意識を知る研究です。そのためには、人の注視をセンシングすることが必要です。当初は、人の頭に装着するカメラや環境埋め込み型のカメラを使って、顔の向きや目の動きをセンシングしていました。しかし、この方法は研究室など一定の環境のもとでは有効ですが、街の中など日常の生活空間で人の注視を推定することには不向きです。また、高精度な計測技術と、煩雑な調整作業が必要になります。そこでより使いやすい注視推定技術として着目したのが、“視覚的顕著性”の応用でした。

杉本 “視覚的顕著性”とは、簡単にいえば人は目立つところに視線がいく、ということです。色彩や明るさのコントラスト、特定の動きなど、人が注視する対象はある程度モデル化されます。そこに網膜の特徴や人が歩く向きなどを考慮すれば、“次に何を視るか”を予測することができます。

佐藤 これまでは、環境からの働きかけをもとに注視を推定するという発想はありませんでした。いわば見る側ではなく視られる側に着目するという発想の転換に基づくもので、視覚的顕著性を手掛かりとした注視推定は世界初の研究テーマです。



視覚的顕著性を応用し、 注視をコントロールする

“見る”という行為を、さまざまな側面から研究されているのですね。

杉本 人の注視には、大きく二つのモデルがあります。一つはボトムアップモデル。これは、ピカッと光ったところに目がいくといった反射的で自然な注視です。もう一つはトップダウンモデル。これは、人が経験の中で培った注視行動で、対象物の認識と密接に関係しています。人の注視がどちらに

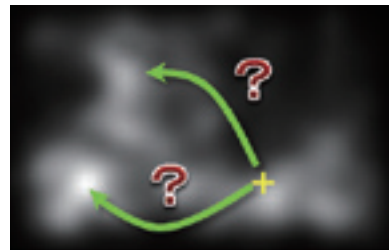
■視覚的顕著性モデルを応用した注視推定の流れ



A. 通行中の男性(+印)を捉えている視線が、次にどこを視るかを予測します。



B. 注視する中心では輝度や色彩、周辺では明るさなどに刺激を受けやすくなります。こうした視野による違い(視野特性)を考慮します。



C. 人の顔や目、特定の動きなど視覚的顕著性の高いモデルと、視野特性を考慮した上で、注視の順序を予測します。予測精度の向上が課題です。

新たな価値

杉本 晃宏

Akibiro Sugimoto

国立情報学研究所
コンテンツ科学研究系 教授



佐藤 洋一

Yoichi Sato

東京大学
生産技術研究所 教授

当たるかによって、視覚的顕著性のモデルも異なってきます。例えば人の顔や目、自動車といったものに自然と目が向くというのは、トップダウン型の視覚的顕著性によるものです。色彩や明るさなどはボトムアップ型です。これらのモデルを積み上げることで、注視推定はさらに精度を増していきます。

もう一つの研究テーマ、注視誘導とはどのような研究ですか？

佐藤 注視誘導は、人の注視をコントロールする考え方です。これを自然に行うためには視覚的顕著性の活用が不可欠となります。例えば、人に注視させたい対象の色や明るさを視覚的顕著性を踏まえて適切に変えてあげることにより、さりげなく人の視線を引き付けることが可能となります。また、トップダウン型の視覚的顕著性モデルに人の顔や目がありますが、まずロボットの顔に注視を引き寄せてから、共同注視によって注視を誘導する研究も行われています。さまざまな環境に応じた視覚的顕著性の高いモデルを活用することで、人は自然にそれを注視します。こうした注視誘導の技術は、さまざまな応用が可能になるのです。

例えば、事故が頻発する交差点で何が起きているか？

注視推定と注視誘導の技術は、社会の中でどのような可能性を持つのでしょうか？

杉本 自動車事故が頻繁に起きる交差点があると思います。さまざまな原因が考えられますが、視覚的顕著性の観点から現場を検証してみると、人の注視をひきつける看板やデザインが設置されているため事故が多いのかも知れません。こうした事故の原因となる注視をなくすことも、注視誘導の一つです。もちろん、事故への注意をうながす視覚的顕著性の高い情報看板を安全な場所に設置してもいい。注視推定・注視誘導は、公共分野の情報デザインの可能性を大きく広げると思います。

佐藤 研究を始めたばかりの頃は、学校や保育所などでの応用を考えていました。授業中に黒板に目を向けさせる注視誘導や、目が届きにくいところに自然に注視を向ける環境づくりです。こうした情報デザインはもちろん可能ですが、広告宣伝などビジネスの分野でも応用の可能性は高いと思

います。例えば注視誘導の考え方で広告看板をデザインし、環境埋め込み型カメラで“注視率”を測定、さらにそれを手掛かりにデザインへとフィードバックするというのが考えられます。

「注視」を軸に、新しい社会や暮らし方をデザインする

共同研究のこれからの展望についてお聞かせください。

杉本 この研究にとって、ゴールは人です。注視を推定することも、誘導することも、人の行動をひも解き、よりよいものにするための研究です。社会や公共分野の情報デザインも大きなテーマですが、人の日常生活そのものに関わる応用研究を進めたいと考えています。

佐藤 注視を誘導するために適当な位置に情報を配置して、ユーザーの視線の向きだけで情報を操作するという注視反応モデルを利用したディスプレイも開発されています。実用化にはまだ至っていませんが、“注視”を軸とした情報デザインは、これからの生活を変えていくと思います。

杉本 NIIのチョン・ジーン准教授が取り組んでいるマルチメディアセンシングや、イマーシブ（臨場感）ビジュアルコミュニケーションとの連携も考えられます。人や環境の情報をセンシングし、注視推定や注視誘導につながる情報をアウトプットするのはマルチメディアセンシングへの応用の一つです。また、相手の視線を予測して3D画像のズレをなくし、目の前にいるのと同じ臨場感を生む技術に、注視推定の研究が役立つでしょう。私と佐藤先生、ジーン准教授、それぞれの研究成果を活かして、注視推定・誘導の技術をいかに活用していくかを考えたいと思います。

*ジーン准教授の研究成果は、P2およびP10参照

（取材・構成 高橋義和）

自由視点の3D映像を可能にする

地球の反対側にいる相手と、つい握手をしたくなる。そんな遠距離でも臨場感あふれるコミュニケーションを可能にする技術が「イマーシブ(臨場感)ビジュアルコミュニケーション」だ。多数のカメラで撮影された画像データを処理・伝送・合成し、ユーザーはどんな角度からでも3Dシーンを捉えられる。マルチメディアセンシングが開く進化したコミュニケーションが、ここにある。

目の前で起きているのと同じ臨場感を、ネットワークで再現する

テレビ会議の相手と、アイコンタクトをしながら話し合える。入院している患者が、ベッドに寝たままで好きな角度からスポーツ観戦を楽しめる。遠く離れていても、まるで目の前で起きているような3Dの映像環境を実現するのがイマーシブビジュアルコミュニケーションだ。NIIのコンテンツ科学研究系チョン・ジーン准教授は、南カリフォルニア大学のアントニオ・オルテガ教授と、この技術の共同研究を進めている。ジーン准教授はこう語る。

「現在、3Dと呼ばれている画像技術は、私たちから見ればまだ2.5Dです。私とオルテガ教授が目指しているのは自由視点でどこから見ても3Dに見える、つまり、側面や反対側からも見ることができる、真にリアルな3Dです。多数のカメラであらゆる角度から対象物を撮影し、必要な画像を選んで合成することで、イマーシブな3Dビジュアルが再現できます」。

ジーン准教授は、大学院生だった約15年前にオルテガ教授と知り合い、5年前企業ラボにいた時に共同研究をスタートした。現在はオルテガ教授もNIIの客員教授として年2回は日本を訪れている。普段はテレビ会議を通してお互いの研究成果を共有し、グローバルなコラボレーションを実践しているのだ。

自由視点3Dを実現する、奥行き画像の圧縮技術

イマーシブビジュアルコミュニケーションを実現する技術上のポイントは三つある。まず多くのカメラで撮影した大容量の画像



アントニオ オルテガ

Antonio Ortega

南カリフォルニア大学 教授

チョン ジーン

Cheung Gene

国立情報学研究所
コンテンツ科学研究系 准教授

データを符号化し、圧縮すること。次にネットワークでデータを伝送する際のバケットロスの映像への影響を最低限に抑えること。さらに人間の視線と頭の動きを予測して、必要なデータを改めて送り手側に要求することだ。相手の頭の動きなどによって生まれる視線のズレをリアルタイムで調整することで、遠方の相手と常に目を合わせながら対話することが可能になる。この中でも、特に重要となるのが画像データの圧縮技術だ。オルテガ教授はこう語る。

「3Dビジュアルを実現するカギとなるのが、奥行き画像です。通常の画像撮影では、対象物の表面は捉えられますが、奥行きまで捉えることはできません。私たちの研究では、特殊なカメラを使って奥行き画

像までも撮影し、合成することで、対象物の実際の形を好きな角度から自由視点で見ることができるのです。そのためには、撮影した画像データをネットワークで伝送できるよう圧縮しなければなりません。しかし、通常の画像と奥行き画像は、効果的な圧縮方法が異なるのです。そこで、最終的に目にする3D画像に歪みが生じないように、最適な圧縮方法を研究しています」。

また、データの冗長性を最適化することで、伝送中のデータが消失してしまうバケットロスを防ぐシステム設計や、人の視線を予測して常に相手の動きと同期した画像を撮影する技術の研究(P8-9参照)など、イマーシブビジュアルコミュニケーションは着実に実現へと向かっている。

イマーシブビジュアルコミュニケーション

ニーズの数だけ、新たなビジュアルコミュニケーションの世界が広がる

イマーシブビジュアルコミュニケーションは、今後どのように実用化されるのだろうか？

「例えば、インターネットの動画サイトでコンサートなどを見る場合、今は決まった角度からしか楽しむことができませんが、私たちの技術が実用化されれば、実際に会場にいて、好きな席に座って見るような自由な楽しみ方ができます。しかも、コンサートの途中で移動までできるのです」(ジーン准教授)。

また、手術などの様子を3Dで実況し、多くの医学生が実際に手術に立ち会っているのと同じシーンを見るというような、教育・シミュレーションツールとしても有望だという。

さらに、自動車に360度撮影できるカメラを載せ、東日本大震災の被災地の様子を撮影している東北大学の先生から、その画像を3D化し、自由視点で被災地の現状を見せることができないかという提案も来ており、実現に向けて話し合っているとジーン准教授は語る。

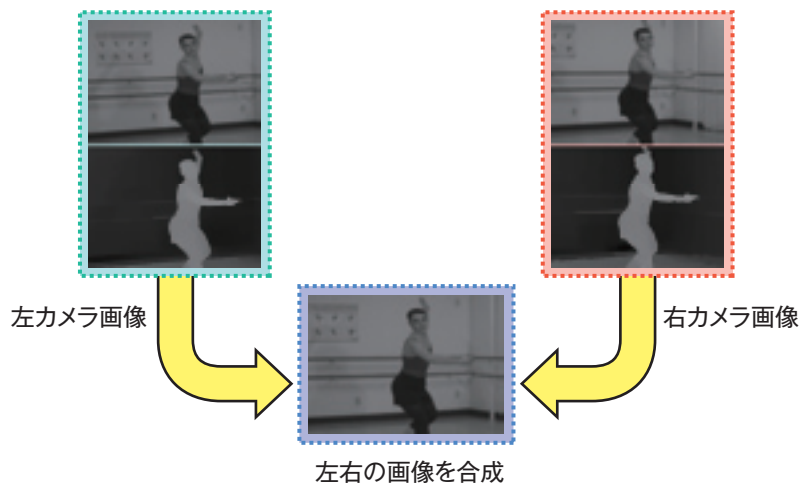
イマーシブビジュアルコミュニケーションは、次世代コミュニケーションの基盤となるツールとして、さまざまなアプリケーションに対応できる。オルテガ教授はこう語る。

「私たちの技術は、ユーザーニーズとの組み合わせで、考えてもいなかったような応用が可能になります。東日本大震災の画像の3D化がいい例でしょう。ユーザーの知恵とのシナジーで、まったく新しいビジュアルコミュニケーションを創造できる。想像できなかった未来の映像世界の可能性が広がっていきます」。

次世代のスタンダードへ、グローバル企業との共同研究もスタート

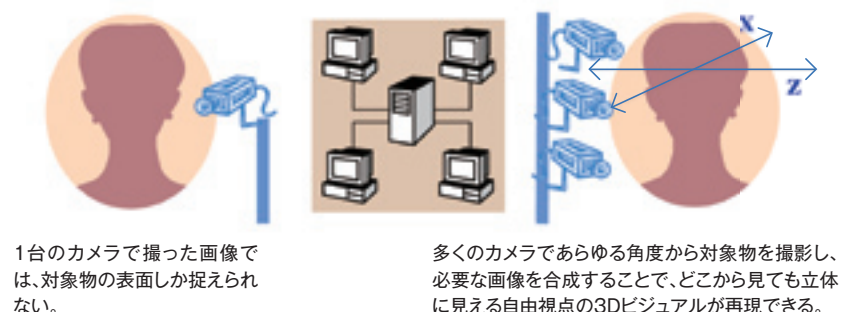
よりイマーシブな3Dビジュアルコミュニケーションに向け、共同研究はこれからど

■通常の画像と奥行き画像の合成



上が通常の画像、下が奥行き画像。左右の画像を合成することで、自由視点で奥行きまで捉えることができる3Dビジュアルが再現される。

■3Dビジュアルを再現する仕組み



のように進展していくのだろう。ジーン准教授はこう語る。

「データの圧縮技術は重要な研究テーマですが、より高度な解決策としては“コンパクトな表現”という考え方があります。これは、撮影した多くの画像データの中から、適切な変換により必要な要素だけを符号化することです。その仕組みやデータ選別の基準を含めて研究中です」。

イマーシブビジュアルコミュニケーションと同様の研究は世界中で進んでいる。その中でジーン准教授とオルテガ教授のチームが先行しているのは、3Dジオメトリのコンパクトな表現だ。これは、3次元空間の物体の座標をディスプレイ上の座

標に適切に変換する技術であり、奥行き画像の圧縮もその手法の一つである。この研究にテレビ会議やエンターテインメントなどを事業とするグローバル企業も注目し、共同研究に参画している。こうした技術が実用化されれば、またたく間にグローバル市場が構築され、次世代のデータ伝送や3Dコミュニケーションの世界標準となる。そんな将来が開かれているのだ。

リビングルームにいながら、3Dビジュアルが生み出す“もう一つの現実”を楽しめる。そんなコミュニケーションの未来が、もう目の前にある。

(取材・構成 高橋義和)

「十人十色」―世界共通でない虹の色―

佐藤 いまり

国立情報学研究所
コンテンツ科学研究系 准教授

色って何だろう？私達が知覚する色の正体

は「光」。このことに気がついたのは、かの有名な科学者ニュートンです。ニュートンは、白色光をガラスでできた三角柱のプリズムを通して観察すると虹色のスペクトルに分かれるという実験を通して、色の正体を明らかにしました。その一方で、ニュートンは光自体には色がないことについても言及しています“For the Rays, to speak properly, are not colored. In them there is nothing else than a certain power and disposition to stir up a sensation of this or that Color”（著書*Optics*より）。色の正体は光とっておいて、でも光自体には色がついていないとは、どういうことでしょうか？

私達にとって赤く見える光があったとしても、その光自体に色がついている訳ではありません。たとえば、目の前に赤いリングが見えたとしても、「赤いリング」が実在する訳ではなく、人間の視覚を通して観察することで初めてリングに色がつく。すなわち色とは、光という刺激を通して引き起こされる感覚であることを

ニュートンは知っていました。

人間の視覚系は、可視域の光を「赤」「青」「緑」の三刺激値に変換して色を知覚することが知られています。もちろん、人間のカラーセンサにも個人差があり、自分が見ている色と他人が見ている色が同じである保証はありません。例えば、虹の色ひとつをとっても、世界共通の概念は存在しません。虹色は七色という考えが日本では一般的ですが、イギリスやアメリカでは一般に六色、ドイツでは五色、四色、三色と答える国まであるとのこと。

興味深いことに、異なる生物の光センサを通して観察してみると実世界はさらに異なった装いを見せます。例えば、私達には同じく黄色に見えるモンシロチョウの雄と雌も昆虫の目（昆虫は紫外線を感じ）を通して観察すると見え方が異なり、容易に区別できるといいます。



さらに、鳥の視覚は人間より優れていると言われており、人間が知覚する「赤」「緑」「青」に加えて「紫外線」の4色の組み合わせに基づき色彩豊かな世界を見ることができるといいます。ちなみに、赤いマントでおなじみの闘牛、実は、牛は赤色を識別することができません。赤色は、観客を興奮させるために用いられているとのこと。興奮させられているのが私達とは、色の世界は奥深く、その話題はつきません。

情報から知を紡ぎだす。

NII

国立情報学研究所ニュース[NII Today]第57号 平成24年9月

発行：大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
http://www.nii.ac.jp/
〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター
編集長：東倉洋一 表紙画：小森 誠 写真撮影：熊谷美由希
制作：株式会社廣済堂

本誌についてのお問い合わせ：総務部企画課 広報チーム
TEL：03-4212-2164 FAX：03-4212-2150 e-mail：kouhou@nii.ac.jp

表紙イラスト

マルチメディアセンシングの進化により、五感に訴えるリアルなコミュニケーションが可能になりつつある。例えば、テレビ会議も2Dから3Dへ。遠く離れていても、目の前に映し出された3D画像を自由な角度から見ることができるようになる。映画で見た未来はもう手の届く所まで来ようとしている。