Special issue

3. メディア情報

3-2. メディア工学

1節, 8節 吉田 俊之^{†1}, 2節 入江 豪^{†2, †3}, 2節 佐藤 隆^{†2}, 3節 芹沢 昌宏^{†4},
4節 多田 昌裕^{†5}, 5節 田川 憲男^{†6}, 6節 広明 敏彦^{†4}, 7節 青木 義満^{†7}

1. ま え が き

メディア工学がカバーする領域は、画像・映像を扱う分野に限っても非常に多岐にわたっており、また各分野の中心的研究対象も年々変化している。図1は、メディア工学研究会に申し込みのあった技術報告件数（過去4年間）を分野別に集計した結果である。図1(a)と図1(b)を比較すると、従来からの「画像/動画処理」や「画像符号化」に分類される件数が減少している一方、「3次元/多視点処理」や「感性情報処理」、「ITS」等の分野が伸びており、研究者・技術者の興味領域が変化している状況を明確に示している。

本稿では、従来からメディア工学がカバーする中心的な分野に加え、図1に見られる興味領域の変化、あるいは今後さらに活発化するであろう研究分野を取り上げ、その動向を概観する。

(吉田)

2. インターネット上のメディア流通

Flickr, YouTubeの登場から、5年ほどが経過した。インターネット上の画像・映像メディア流通は、その成長に陰りを見せるどころか、依然として急成長を続けている。2009年12月には、月間330億ビューを超えるインターネット映像の視聴がなされたという¹⁾。一人当たりの月間平均映像視聴時間は12時間を超える計算である。2010年3月の公式レポートによれば、YouTubeには、毎分およそ24時間分に相当する量の映像が投稿されている²⁾。また、Web技

(a) 2006～2007年度

(b) 2008～2009年度

図1 メディア工学研究会における技術報告の分野別内訳

†1 福井大学 工学部

†2 NTT サイバーソリューション研究所

†3 東京大学 大学院情報理工学研究所

†4 NEC 情報・メディアプロセッシング研究所

†5 株式会社国際電気通信基礎技術研究所 知能ロボティクス研究所

†6 首都大学東京 システムデザイン学部

†7 慶應義塾大学 理工学部

"Media Engineering" by Toshiyuki Yoshida (Faculty of Engineering, University of Fukui), Go Irie (NTT Cyber Solutions Laboratories, Kanagawa/Dept. of Information and Communication Engineering, University of Tokyo, Tokyo), Takashi Satou (NTT Cyber Solutions Laboratories, Kanagawa), Masahiro Serizawa, Toshihiko Hiroaki (Information and Media Processing Research Laboratories, NEC Corporation, Kanagawa), Masahiro Tada (ATR Intelligent Robotics and Communication Laboratories, Kyoto), Norio Tagawa (Faculty of System Design, Tokyo Metropolitan University, Tokyo) and Yoshimitsu Aoki (Faculty of Science and Engineering, Keio University, Kanagawa)

術に関する主要な国際会議、World Wide Web Conference (WWW)においても、2008年から、画像・映像メディアを対象としたRich Media Trackが設置され、最新のメディア処理・応用技術が報告されている。米国国立標準研究所(NIST)が主催する映像処理に関する国際コンペティションTRECVIDにおいても、2010年大会では、インターネット映像を対象としたタスクに大きくシフトすることが発表された³⁾。画像・映像メディアは、今まさにインターネットを流通するメディアの主役となりつつある。

インターネット上の画像・映像メディアは、文書(タグ、周辺テキスト)やログ(クエリログ、クリックログ)など、さまざまな情報と関連しあいながら流通している。2008年以降は、従来取組まれてきた一般物体認識、検索、コピー検出などに加え、これら関連する情報と協調したメディア処理手法に関する報告が目立った。本節では、特に発展の期待される三つの異なる情報との関連を扱ったメディア処理手法にフォーカスし、その動向に触れたいと思う。

2.1 位置情報とメディア処理

現在普及するカメラデバイスの多くはGPS機能を備えており、撮影した画像に対して「ジオタグ」と呼ばれるメタデータを付与することができる。ジオタグは、緯度経度からなる位置情報を表すEXIF準拠のタグであり、その画像がどこで撮影されたものであるかを特定する役割を果たす。2006年頃より、ジオタグによる位置情報との関連を扱った画像の検索、閲覧、あるいは、画像の撮影位置推定などの取組みが多く報告されるようになった。

- (1) 画像検索：ランドマークの写真について、代表的なものとそうでないものとを分類し、検索ランキングに反映する手法が提案されている⁴⁾。
- (2) 画像閲覧：ジオタグ付き画像を、Googleマップ⁵⁾のようなマップインタフェースにマッピングして閲覧できる手法⁶⁾が提案された。また、同一ランドマークを撮影した多数の画像群から、そのランドマークを3次元再構成し、実際の位置関係に則した閲覧を可能にする技術も開発された⁷⁾⁸⁾。この技術は、Googleマップ/ストリートビュー⁵⁾やPhotosynth⁹⁾に应用されている。
- (3) 撮影位置推定：ジオタグが付与されていない写真の撮影位置を、画像情報から推定する方法が提案されている¹⁰⁾。2009年には、撮影順序・時間差を考慮して位置推定精度を高める手法¹¹⁾も提案された。

このほか、写真共有サイトに蓄積されたジオタグ(位置情報)の時系列から、多数のユーザの旅行経路をマイニングし、旅行経路推薦を実現する手法が提案された¹²⁾。膨大なジオタグデータを、実世界での活動に有益な情報へ変えて活用するという、新しい視点を提案している。

2.2 コミュニティ情報とメディア処理

動画共有サイトは、いわゆるコミュニティサイトの形態

をとるものがほとんどである。ここで活動するユーザのログ(履歴)を解析することで、有益な情報やコンテンツの発見に役立てようとする研究がなされてきている。

例えば、YouTubeのコメントデータを分析し、時々刻々と変換していくユーザの興味や話題を検出する手法が提案された¹³⁾。また、ニコニコ動画の映像に対して、その最大の特徴である映像時刻に対応したコメントを用いた研究も見られた。例えば、単位時間当たりのコメント量を映像要約に用いる試みが報告された¹⁴⁾。

2.3 メディアを利用したメディア処理

Web上の豊富なメディアコンテンツ自身を「道具」として使い、新しい効果を生み出そうとする試みがなされるようになった。

道具として用いられることの多いサイトは、Flickrである。例えば、Flickrを使って、二つのキーワードの意味的な距離を定義するFlickr Distanceが提案された¹⁵⁾。これは、異なるキーワードによる画像検索結果の差異(非類似度)を測ることで、キーワード間の距離を求める。また、Flickrの画像群を使って、モノクロ画像を自動カラー化する手法¹⁶⁾や、手書きスケッチを写真に合成してしまう手法¹⁷⁾も提案された。映像に対するアプローチとしては、シーンごとのユーザの興味のレベルを自動推定する手法が提案された¹⁸⁾。Flickr上に類似する画像が多くアップされているシーン(フレーム)ほど、ユーザが興味を持ちやすいという仮定のもと、画像のマッチングに基づいて興味レベルを推定するものである。

また、サイトをまたがるメディアコンテンツを統合して、新しい価値を提供しようとする試みもある。WikipediaのテキストとFlickrの画像とを関連付け、画像情報のための辞書を自動生成する取組み¹⁹⁾や、Flickrの画像に、関連するTwitterのテキストを重畳し、よりリッチで効率的な情報提示を実現する試みが報告された²⁰⁾。

2.4 その他の話題

上記のカテゴリには含まれないが、インターネットならではの話題に関する新しい提案を紹介したい。

- (1) 画像広告：Webにおける主要ビジネスの一つである画像広告をターゲットとした技術として、AdSenseの画像版ともいえるべき、ImageSense²¹⁾が提案された。ImageSenseは、Web上に流通する画像に対して、関連する広告画像(ロゴ)を最適な位置に自動挿入する。
- (2) 映像圧縮：膨大な映像データを効率的に蓄積する技術として、Near Lossless Video Summarization (NLVS)が提案された²²⁾。NLVSでは、映像を少数のキーフレームと、その間の動きを記述したXMLに変換する。これにより、画質を保持しながら劇的な容量削減を実現できる。
- (3) 映像検索：動画共有サイトの映像は、その作り手の

違い(スキルやツール)によって、編集のレベルが大きく異なる。これに着目し、映像の編集レベルを自動推定し、検索ランキングに利用する手法が報告された²³⁾。

- (4) 画像クエリ推薦: キーワードクエリ推薦と同様、画像検索に対して画像クエリを推薦する手法が提案された²⁴⁾。

2.5 まとめ

インターネットが、画像・映像メディアの主要な流通経路の一つとして確立したことは疑いようがない。クラウドサービスやIPTV、モバイルマルチメディアなど、更なる進化を遂げる環境も整備されている。今後もその成長から目が離せない。

(入江・佐藤)

3. 空間共有における臨場感技術

近年報告された臨場感に関する分析と、その空間共有に向けた技術を紹介する。人々が遠隔地で空間を共有する手段は、メッセンジャーのようなシンプルなものから、テレビ会議システムやVRシステムのようにマルチメディアを駆使した比較的大型なものまで多様化している。より高い臨場感を持って空間を共有することは、本物に近い状況や情報を交換することによって、ある程度達成できる。一方、本物に近づけなくても臨場感を感じることができる。これは、人の感じる臨場感が、与えられた情報をどのような意識で処理しているかに依存するためである²⁵⁾。すなわち、その人の意識に適合した情報が与えれば、ある程度臨場感のある情報の提供が可能となる。また一方で、本物に近くても意識とずれると逆に臨場感を失うこともある。

臨場感に関して、文献²⁶⁾では、空間要素(立体感、質感、包囲感)、時間要素(動感、同時感、リアルタイム感)、身体要素(自己存在感、インタラクティブ感、情感)の三構成要素を用いて説明している。これらの感覚要素が、「感覚器で得られる物理情報」に関する外的要因と「経験や学習記憶情報からの連想」に関する内的要因によって、どう組合せられるかで臨場感の程度や種類が決まる。臨場感を捉える時に、メディアを通じた事象であろうと、現実の事象であろうと、「心を揺さぶる事象」に人は臨場感を感じるという報告もある²⁷⁾。これは内的要因による。

提示する事象を本物に近づけることだけが、臨場感を必ずしも高めるわけではない。より正確に外的要因を与えられても、内的要因により、人によって異なる臨場感(場合によっては違和感)となる。リアリティを求める場合と逆のアプローチでありながら結果的にリアリティを強く想起させる意味で「リアルバーチャリティ」とも呼ばれるものがある²⁸⁾。やや大げさな表現、デフォルメによる変更などは臨場感を高め、意思疎通する上で重要な要因になる。

外的要因でできる限りリアル(本物)を実現する技術として、大画面や高精細、3D映像・音響等の実用化が進んでい

る。不自然な場合も未だあるが、互いの状況を混合して映像化する複合現実空間でもさまざまな臨場感の実現も試みられている。例えば、共有空間でインタラクティブに作業することによる高臨場感システムも実現されている^{29)~31)}。

一方、リアル感(本物っぽい)を実現する外部要因には、前述の明らかなデフォルメだけでなく、元映像と比較しないとわからない程度に改変された物理情報がある。例えば、ある事物の色を人が好ましいと考える記憶色で補正することにより、ある種の臨場感を感じる。「空は真っ青」で、「木々は深緑」の方が好まれ臨場感を感じる。また、ケータイ等の簡単な提示装置でもいくつかの要素を満たす要因が揃うことにより、臨場感をもって空間を共有することが可能となる。例え文字だけであっても、その文字を共有している人が特定の提示方法等によりある種の臨場感を共有することができる³²⁾。相手の状況を把握するための補助的な情報等により臨場感を高めることもできる。

遠隔地で空間を共有する典型的な実現例の一つとしてテレワークがある。近年、国による基盤や制度の整備が進みつつあり注目されている³³⁾。日本テレワーク協会では、テレワークを「情報通信技術(IT)を活用した場所や時間にとられない柔軟な働き方」と定義している³⁴⁾。

この働き方を実現するために、テレワークでは異なる場所の人々が、情報・状況を適切に間違いなく交換しながら、快適に働く環境を構築する技術が求められている。そのために、遠隔地にいるワーカーにとって、オフィスの状況を常時感じて、オフィスにいる人と同じ感覚で働ける環境を提供する超臨場感テレワークの実現が検討されている³⁵⁾。超臨場感テレワークの実現のためには、遠隔オフィスの雰囲気までも伝える映像・音響通信等に関する臨場感技術の開発が課題となっている³⁶⁾。臨場感のある共有空間を実現する基盤となる技術として、事象を正確に把握するために任意エリアの映像・音響情報にアクセスする自由視聴点技術^{37) 38)}、視聴者の視線や存在感を非視聴者に伝える情報提示技術³¹⁾、選択的に映像を強調/抑制する映像・音響技術³⁹⁾、特定者のみに情報を伝える提示技術等の取組み等が近年発表されている。

遠隔地で空間を共有する際に、円滑で快適にコミュニケーションを行うための臨場感技術がますます重要になる。今後は、外的要因としてリアル(本物)を求める大掛かりな技術がさらに進化の中で、広く人々が使用している携帯可能な端末であっても、リアル感(本物っぽい)を実現する臨場感技術の開発が望まれる。また、内的要因までもが人と人の間でアンビエントに交換され⁴⁰⁾、共通の意識を自然と有することによって、より円滑にコミュニケーションできる空間共有が実現されていくと考える。

(芹沢)

4. 感性情報処理とメディア

近年、産業界では「規格大量生産・大量消費」型のパラダ

イムに代わるものとして、「個人性・多様性」が注目されるようになってきた。このような流れの中で、人間と人間とをつなぐ役割を担うメディア技術に、個人性・多様性の概念をいかにして導入するかが重要な課題となりつつある。

人間の感覚器の仕組みは同一であるにもかかわらず、モノに対する主観的な解釈・評価結果は各人各様である。このような個人性・多様性は、感覚器を通して知覚したモノを、人間内部の情報処理系において、各人の経験や知識に基づき解釈・評価する過程で生ずると考えられる⁴¹⁾。感性の定義はさまざまになされているが、加藤⁴¹⁾や多田⁴²⁾は、個人性・多様性を生み出す「人間内部の情報処理系における解釈の違い」を感性と定義している。感性は、モノの主観的評価過程、知識の習得・体系化過程、身体行動を通して対象や環境に働きかける過程(モノや環境との相互作用)、など人間の知的活動のさまざまな場面で作用し⁴¹⁾、個人性・多様性を生み出す。

4.1 利用者の教示に基づく個人性・多様性のモデル化

個人性・多様性の概念をメディア情報システムに導入するためには、システムに利用者の多様な個性・特性を認識・理解させ、処理過程や結果に反映させるための技術が必要となる。

感性情報処理の分野では、モノに対する利用者の主観的印象を、「フレッシュな」「落ち着きのある」といった形容語(印象語)を用いて表現することが多い。その上で、そのような主観的印象を生じさせたモノの物理的特徴量と、印象語との関連性をコンピュータに学習させ、印象語を介した個人適応型の情報検索サービスとして供する試みがなされてきた。このようなモデル化手法は、利用者によって陽に事前教示されたモノの主観的印象・解釈情報を教師データとした、教師あり学習、あるいは半教師あり学習の枠組みとしてとらえることができる。

印象語を介した個人適応型情報検索サービスの初期の例としては、栗田らによる絵画検索システム⁴³⁾や、辻らによる楽曲検索システム⁴⁴⁾が挙げられる。近年は、都市景観デザイン支援に印象語を用いた例⁴⁵⁾、舞踊動作の分類を試みた例⁴⁶⁾など、さまざまな分野において感性情報処理技術が応用されている。

4.2 人の行動計測に基づく個人性・多様性の推定

従来の教師あり学習の枠組みに基づく感性情報処理システムでは、利用者自身が自らの主観的な印象や解釈をシステムに対して陽に教示することで、個人性・多様性の学習・適応を行っている。そのため利用者は、システム利用前に、個人性・多様性の学習データとして、しばしば膨大な量の事前データ入力(システムへの教示)をこなす必要があった。利用者に課せられる事前教示労力の低減は、感性情報処理技術の大きな課題である。

一方、人間の知的活動時の身体挙動をユビキタス・ウェアラブルセンシング技術により計測・解析することで、個

人性・多様性を陰に推定し、利用者に事前教示を課すことなく個人適応型のサービスを実現しようとする試みもはじまっている。

岡田ら⁴⁷⁾は、環境学習を題材に、学習者が実世界の中で環境との相互作用から知識を得ていく過程をウェアラブル技術により計測・解析することで、知識の習得過程の個人性・多様性を陰に推定し、学習者各々に適した学習支援を行うシステムを提案している。また、多田ら⁴⁸⁾は、安全運転講習を題材に、運転行動の個人性・多様性(運転者の癖)を装着型モーションセンサと行動認識技術によって計測・解析することで、個人の特性情報をシステムに陽に教示することなく、個人適応型の運転支援を行うシステムを提案している。

行動計測に基づく陰な個人性の推定技術は、まだ緒にのまないばかりであるが、個人の特性情報を陽に教示することなく、メディア情報システム上で個人性・多様性を取り扱うための方策として、今後ますますの発展が期待される。

(多田)

5. CVおよびCG技術とメディア処理

コンピュータビジョン(CV)とコンピュータグラフィックス(CG)は互いに逆計算の関係にあり、一般にCGが順問題、CVは逆問題となる。あるCVの手法が安定に逆問題を解くために、その解法の中にCGとしての順問題計算を含むこともあり、逆に効率的にCGの順問題を解くために、CVとしての逆問題計算を利用することがある。最近の研究の傾向として、後者に属する広い意味でのCG手法に関する研究が多いように見受けられる。一方で、撮影対象の3次元形状や表面反射率を主要な出力とするCV手法の研究が少なくなっているのが印象的である。

上述した後者の枠組み、すなわちCVを利用したCG技術の代表は、モデルベースの任意視点画像・動画生成である。そこでは、CGのために必要な3次元モデルを多視点カメラで撮像した画像群を用いてCVにより計算し、画像生成に利用する。完全な3次元モデル、すなわち対象の3次元形状と表面反射率が獲得されれば、任意視点からの任意光源下の画像生成が陰影計算で可能となる。しかし当然ながら、CV計算の誤差により画質の低下が問題となる。そこで文献49)では、レンジセンサとフラッシュ付カメラを用いたセンサフュージョンの立場で解決を図っている。

任意視点画像生成の研究に着目すると、上記の他にはイメージベースの方法、および3次元ビデオ生成と呼ばれる方式が存在する。前者は、光線空間法などの詳細な研究が数年前まで大量に推進されたが、校正済みカメラの稠密な配置が必要であることから、こここのところはやや落ち着きをみせているように感じる。なお、文献50)では、フリーハンドで収録された画像群(イメージクラスタ)に基づくイメージベースの手法を提案しており、撮像にかかる手間の

削減を図っている。3次元ビデオ生成法は、ちょうどモデルベースとイメージベースの中間に相当する方式である。通常のイメージベースでは、対象の3次元形状を考慮しないか、簡単な構造(例えば平面)で近似するのに対し、3次元ビデオ生成では、モデルベースと同様にCVによって3次元形状を獲得したのち、これに基づいて多視点撮画像の輝度情報を利用しながら任意視点画像を生成するものである。すなわち、形状はCVにより獲得するが、反射率分布は求めず、イメージベースな方法でレンダリングを行う。この方式でも、形状の復元誤差が問題になるが、無理にCVの精度向上を図るのではなく、複数視点から撮影された画像をブレンドして任意視点画像を作る際の、形状誤差に起因する画像不一致性を考慮した画像補正(幾何補正)により、画質向上を実現することができている⁵¹⁾。

任意視点画像生成にもそのまま利用可能であるが、CGモデルとして広く活用可能な動的3次元対象のデータ構造という立場から、Time-Varying Mesh (TVM) に関する研究も、精力的に進められている⁵²⁾。TVMは実世界の物体を動的に変化する3次元メッシュデータ(形および色の情報を保持)であり、物体の3次元的な動きの扱いを工夫し、レンダリング負荷の低減等が検討されている。

画像生成研究の別の方向として、一般論としてではなく、VR (Virtual Reality) やMR (Mixed Reality), AR (Augmented Reality) のための工夫や提案も盛んに行われている。文献53)では、映像の広視野提示による臨場感提示のために、3次元空間を意識したCG技術が使用されている。文献54)では、少ないカメラで可能な効果的なフローモーション(被写体の周囲をぐるりと回転するようなカット)を、幾何光学に基づくCG技術を使いながら、イメージベースで生成している。文献55)ではAR技術を取り上げ、実画像と仮想物体の整合性を向上させる高速な陰影計算法を提案している。このように、CVやCG技術を駆使したタスク指向型の画像生成技術の研究が盛んになってきており、顔画像分野もそのようなタスクの一つに数えることができる。顔表情のCG合成、さまざまに表情を変える顔ロボット、似顔絵生成など、多くの課題が扱われている⁵⁶⁾。

以上のように、全般的な方向性として、広い意味での画像生成に関する研究が非常に活発であり、さまざまな応用における高度な画像生成技術の必要性の高まりに応じて、自然物や自然環境の表現手法といったCGの基礎技術に関しても興味深い報告がみられる。今後、さらなる写実性の追及が期待される。一方で、CVの基礎研究が少なくなっているのは残念である。これまでの流れに沿ってのCV研究が限界に達している可能性があり、従来の枠組みを越えるブレークスルーが期待される。

(田川)

6. 画像認識技術の動向

6.1 はじめに

画像認識技術の最近のトレンド(主として製品化段階のもの)を概観してみる。画像処理を行う計算資源の小型高性能化や低価格・低消費電力化が進んだ結果、従来では導入が難しかったさまざまな用途へ画像認識技術が利用されるようになり、コモディティ化が急速に進んでいる。特徴的な動向としては、デジタルカメラや携帯電話(スマートフォン)、家庭用ゲーム機などのコンシューマ向け製品へ画像認識技術を積極的に導入する動きがでてきたこと、さらには、画像をネットワーク経由でサーバへ送り、サーバ側で画像認識処理をするサービスが普及してきたことなどが挙げられる。その他、ビジネス用途では、デジタルサイネージ(電子看板)などにカメラを付けて、周辺の人物の特性や行動を取得し活用するものなどが登場してきた。

6.2 顔画像照合

高精度な個人識別を目的とした顔照合技術については、屋内などの照明を制御した環境下において、正面から撮影した顔画像に対する認識性能が実用的水準(明確な基準や定義等はないが、一般的には他人許容率0.1%で本人の識別率が98%程度以上は必要とみられる)になってきており、出入国管理などの用途への導入例が増えてきた。すでに技術開発の関心は、屋外監視などを想定した、低解像度・高圧縮率な映像中における、正面を向いていない人物の顔検出や照合へと移ってきており、米国国立標準研究所(NIST)がこれら技術課題を組込んだ公開ベンチマークテストMBGC/MBEを実施している(2010年夏頃に結果が公表される予定)^{57) 58)}。

6.3 コンシューマ向け機器(デジタルカメラ、ゲーム機)

デジタルカメラへの顔認識技術の導入が一般化し、現在では、国内で発売されるほとんどの機種に顔検出・認識技術が搭載されている。従来は顔検出のみだったが、笑顔を判定して自動的にシャッターを切るものや、個人識別を行う製品なども登場した。普及が加速した背景としては、画素数や画質による製品の差異化が難しくなり、画像認識への取組みが必須になってきたことや、組み込み型ハードウェア上でも複雑な認識処理が可能になり、かつ、Boostingなどの計算量が少なくても済む有力なアルゴリズムが登場したことなどが挙げられる。最近では、ペット(犬や猫)などの人以外の識別に対応した製品⁵⁹⁾も開発されている。これら認識技術は、撮影モードの自動切り替えや、ピント・色調の自動調整に加え、撮影した画像の自動分類などにも用いられており、人物認識からオブジェクト認識へと関心が広がってきている。

家庭用ゲーム機においても、画像認識技術の導入が進みつつある。例えば、任天堂はカメラで顔の動きを検出し利用するゲーム⁶⁰⁾を、ソニーはトレーディングカード(紙)

をカメラにかざしゲーム内のキャラクタを操作する⁶¹⁾といったアイデアを製品化している。また、Microsoftは、カメラ映像からユーザの体の動きを解析してゲームの操作を行うコントローラ「Project Natal」を開発中であることを発表した⁶²⁾。

6.4 デジタルサイネージ⁶³⁾

顔認識技術を販促支援に活用するデジタルサイネージ(電子看板)システムが登場してきた^{64) 65)}。具体的には、店頭や街頭に設置した広告用の大型ディスプレイにカメラを付設し、広告周辺の人物が広告を見ているか(広告効果の測定)、さらには、性別や年齢等の推定値に基づき、表示する広告の内容を切替え制御するもの(ターゲティング広告)などが実現されているほか、ユーザの興味を引くために電子看板にCGを重畳表示した自分の姿が映るなど、エンタテインメント性を高めるさまざまな工夫が試みられている⁶⁶⁾。

6.5 画像認識技術のSaaS化

携帯電話やデジタルカメラで撮影した画像を、端末(PCを含む)側では処理せずに、サーバ側へ送り認識処理するSaaS(Software as a Service)が次々に登場している。それらを大別すると、①画像の解析や処理結果自体がサービスとなるもの、②画像の管理や検索を効率化するもの(主としてインデキシング)、に分類できる。前者の例には、有名人と似ている程度を自動評価するものや⁶⁷⁾、撮影した場所や位置などを知らせるものなどがある^{10) 68)}。また後者には、撮影した画像中の文字などを認識して自動的にインデキシングし、その画像を後からキーワードで検索できるものが挙げられ、Evernote⁶⁹⁾がその代表例といえる。さらには、画像自体を検索クエリにするサービス(例えばGoogle Goggles⁷⁰⁾)も提案されている。

画像認識技術を活用し、複数のユーザを対象としたネットワークサービスの出現も活発になってきている。Googleは無償の画像管理ソフト「Picasa」に顔画像認識技術を組み込み、写真を解析して名前でタグ付けを行い、名前でも自動分類する機能を提供しているが、これはPC内に保存した画像だけでなく、Web上で他人とアルバムを共有する機能にも対応している⁷¹⁾。また、米国では数億人規模の会員を有する大手SNS(ソーシャルネットワークサービス)であるFacebookが、会員の顔画像を認識する機能を提供して話題となっている⁷²⁾。

6.6 学会の動向、今後の技術チャレンジ

この領域の代表的な国際学会としてはCVPR, ICCV, ICIPなどがあり、活発な議論がなされている。例えば、CVPR2009が公表しているテーマ別投稿件数の統計⁷³⁾をみると、オブジェクト検出・認識や統計的手法・学習、モーショントラッキングなどの投稿が多く、この傾向は他の国際学会などでも同様とみられる。特に、一般オブジェクト認識に関する議論は、国内でも積極化しており、技術の進

展が最も期待される領域の一つといえる。

また、画像認識技術として取組むべき今後の課題について、鷲見らが10項目のグランドチャレンジとしてまとめている⁷⁴⁾。数年後の実現が見込めるものから究極とも言える課題までが列記されており、画像認識技術の開発ロードマップを俯瞰的に検討する上でも大いに参考になるであろう。

(広明)

7. ITSにおける画像認識技術

7.1 はじめに

高度道路交通システム(Intelligent Transport Systems)分野における研究開発の方向性は大きく、安全運転支援と快適性の向上の二つがあり、事故防止のための安全運転支援、ナビゲーションシステムの高度化、道路交通管理の効率化、最適化などが代表的な取組みである。その中で、自動車の高知能化を目指す先進安全自動車(ASV:Advanced Safety Vehicle)の実現へ向けて、車載センサ群と情報処理技術の高度化に関する取組みが継続的に進められている。特に、車載画像処理システムは、カメラの高機能化、画像認識技術の進歩と、リアルタイムでの画像処理を可能とする画像認識専用プロセッサの登場により、前方監視をはじめとするさまざまな機能により運転者の危険予知、視覚補助の実現がなされている。近年では特に、多様な状況が想定される実環境における有用性を高めるため、各処理やシステムのロバスト化を実現する工夫がハード、ソフト両面からなされている。ここでは、ITSの安全運転支援分野におけるビジョン技術応用の話題として、最近の車両周辺監視技術、運転者の状態推定技術、車載向け画像認識ハードウェア技術について紹介する。

7.2 車両周辺監視技術

車両の周辺を監視する技術としては、運転者の死角となる部分を画像生成技術により再構成し、画面表示する視覚補助機能技術と、走行レーンや他車両、歩行者を認識する物体認識技術がある。前者については、複数のカメラ映像を車両上方からの仮想視点映像として実時間生成し、車両周辺の状況を容易に把握可能なシステムが実用化されている⁷⁵⁾。安全運転を支援するための物体認識技術としては、ステレオカメラによる奥行き情報を手がかりに道路上の障害物を検知する技術が主流であるが、Low-levelな画像特徴と識別器の組合せによる、アピランスベースの画像認識手法の導入が進められている。例えば、複数のHistograms of Oriented Gradients(HOG)特徴を組合せたJoint HOG特徴と、AdaBoostによる学習によって獲得した識別器を利用し、ロバストに車両や歩行者を検出するものなどがある⁷⁶⁾。また、視覚補助機能においては全方向のモニタリングが可能となっているが、視覚補助機能に加え、画像認識機能の搭載による周辺監視技術の開発が活発に行われている。

7.3 運転者の状態推定技術

交通事故の要因として、車外の外的要因に加え、ドライバー自身の不注意や意識低下がある。近年の顔認識技術の進展により、運転者の時系列の顔表情変化を捉えて状態を推定する研究が進められている。昼夜を問わず、顔の検出と状態認識を行う必要があるため、撮像上の工夫が必須であるが、異なる2波長のマルチバンド光源による近赤外LEDのアクティブな点灯により、照射画像の差分処理のみで肌を検出し、暗い環境下でも高精度に顔領域を検出可能とする技術がある⁷⁷⁾。また、顔の表情をロバストに認識する技術として、パーティクルフィルタと勾配法を組合せた変動テンプレートにより頭部姿勢と表情を同時推定するものが注目されている⁷⁸⁾。眼の開眼度の時系列推移を指標とした運転者の意識低下状態の検出に関しては、すでに実用化事例もあるが、今後、さらに詳細な運転者挙動を捉える画像認識技術の導入が進むものと考えられる。

7.4 車載向け画像認識ハードウェア技術

車載カメラで外界環境の認識をリアルタイムで実現するためには、高度な画像認識ソフトウェアと、高速な処理性能を持ったハードウェアの両方が必要となる。車載向け画像認識用プロセッサ技術は、研究フェーズで生まれた高度な画像認識技術を車載システムとして実現するためのキー技術であり、車両周辺監視のためにステレオ演算で奥行き情報を得て実時間センシングするといったケースなどで必要不可欠なものとなっている。車載向け画像認識用プロセッサとして、さまざまなものが開発、実用化されているが⁷⁹⁾⁸⁰⁾、特に近年、注目されているものとして、画像認識処理アクセラレータを搭載したカーナビゲーション装置向けSoC (System on a Chip)がある⁸¹⁾。カーナビゲーション装置は、すでに車両周辺の映像を液晶ディスプレイ上に表示する機能を持ち、車載システムの中で最も高性能なプロセッサと大容量メモリーを搭載している。カーナビゲーションシステムに画像認識機能を付加することができれば、ナビゲーション機能と画像認識機能を同時に並列処理することが可能となり、走行制御やカーナビゲーション機能との画像認識機能の連携が容易となるため、ビジョン技術の車載化を大きく推進させる可能性を持っている。

7.5 まとめ

この2年間に於いてITSに関連するビジョン研究は数多くの成果が得られ、実用化が進んでいる。これには、車載用カメラの性能向上、画像処理アルゴリズムのロバスト化とそれを高速に演算可能とする車載画像処理専用LSIの普及が大きく寄与している。電気系学会における研究会や、ビジョン系の大規模なシンポジウムにおいてもITS関連のセッションは盛況であり、産学官の研究連携体制も充実化がはかられている。これまでの予防安全実現へ向けての取組みに加え、近年注目されているエコドライブ実現へ向けてのビジョン技術の適用も想定される。今後さらにITS分野

への画像センシング技術活用と実利用化に拍車がかかるものと期待される。

(青木)

8. む す び

以上、メディア工学を取り巻く状況について、最近2年間の動向を概観した。画像処理を一つの源泉として進展してきた同分野は、今後、従来からの画像映像処理技術に加え、3次元処理や画像認識技術などの基礎研究成果を基に、高度な空間共有技術やコンテンツ処理、ネットワーク応用、あるいはITSなどの各分野に発展していくと考えられる。また、認識・認証処理や感性情報処理等の研究分野はここ2年間で従来以上に活発化しており、今後のメディア工学の大きな柱に成り得る領域と期待される。さらに、ここ最近の3次元ブームに見られるように、従来の静止画像・動画画像を対象とした処理から、3次元や4次元データ等へのシフトも加速することになろう。次回12年度の年報では、どのような項目が並んでいるか楽しみである。

(吉田)

(2010年5月13日受付)

〔文 献〕

- 1) ComScore, http://www.comscore.com/Press_Events/Press_Releases/2010/2/U.S._Online_Video_Market_Continues_Ascend_as_Americans_Watch_33_Billion_Videos_in_December
- 2) YouTubeBlog, <http://youtube-global.blogspot.com/2010/03/oops-pow-surprise24-hours-of-video-all.html>
- 3) NIST TRECVID, <http://www.nllpr.nist.gov/projects/trecvid/>
- 4) L. Kennedy and M. Naaman: "Generating Diverse and Representative Image Search Results for Landmarks", Proc. WWW, pp.297-306 (2008)
- 5) Googleマップ, <http://maps.google.co.jp/>
- 6) D. Crandall, L. Backstrom, D. Huttenlocher and Jon Kleinberg: "Mapping the World's Photos", Proc. WWW, pp.761-770 (2009), <http://www.cs.cornell.edu/~crandall/photomap/>
- 7) N. Snavely, S.M. Seitz and R. Szeliski: "Modeling the World from Internet Photo Collections", Int. J. of Computer Vision (IJCV), 8, 2, pp.189-210 (2008)
- 8) S. Agarwal, N. Snavely, I. Simon, S.M. Seitz and R. Szeliski: "Building Rome in a Day", Proc. ICCV, pp.72-79 (2009), <http://grail.cs.washington.edu/rome/>
- 9) Photosynth, <http://photosynth.net/>
- 10) J. Hays and A.A. Efros: "IM2GPS: Estimating Geographic Information from a Single Image", Proc. CVPR, pp.1-8 (2008) <http://graphics.cs.cmu.edu/projects/im2gps/>
- 11) E. Kalogerakis, O. Vesselova, J. Hays, A.A. Efros and A. Hertzmann: "Image Sequence Geolocation with Human Travel Priors", Proc. ICCV, pp.253-260 (2009)
- 12) 倉島健, 岩田具治, 入江豪, 藤村考: "写真共有サイトにおけるジオタグ情報を利用したトラベルルート推薦", 信学技報, 109, 450, LOIS2009-88, pp.55-60 (2010)
- 13) M.D. Choudhury, H. Sundaram, A. John and D.D. Seligmann: "What Makes Conversations Interesting? Themes, Participants and Consequences of Conversations in Online Social Media", Proc. WWW, pp.331-340 (2009)
- 14) 谷直紀, 山崎俊彦, 相澤清晴: "コメント数の動的な変化に基づくCGM動画要約", 第5回デジタルコンテンツシンポジウム (2009)
- 15) L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma and S. Li: "Flickr Distance", Proc. ACM Multimedia, pp.31-40 (2008)
- 16) 森本悠嗣, 田口裕一, 苗村健: "Web上の画像群を用いたモノクロ画像の自動Colorizationに関する基礎検討", 第5回デジタルコンテンツシンポジウム (2009)

- 17) T. Chen, M.M. Cheng, P. Tan, A. Shamir and S.-M. Hu: "Sketch2Photo: Internet Image Montage", Proc. SIGGRAPH Asia (2009), <http://cg.cs.tsinghua.edu.cn/montage/main.htm>
- 18) F. Liu, Y. Niu and M. Gleicher: "Using Web Photos for Measuring Video Frame Interestingness", Proc. IJCAI, pp.2058-2063 (2009)
- 19) R. Hong, J. Tang, Z.-J. Zha, Z. Luo and T.-S. Chua: "Mediapedia: Mining Web Knowledge to Construct Multimedia Encyclopedia", Proc. MMM, pp.556-566 (2010)
- 20) 丸山悠斗, 山崎俊彦, 相澤清晴: "マイクロブログと画像の集約によるライフログとその要約", 第5回デジタルコンテンツシンポジウム (2009)
- 21) T. Mei, X.-S. Hua and S. Li: "Contextual In-Image Advertising", Proc. ACM Multimedia, pp.439-448 (2008)
- 22) L.-X. Tang, T. Mei and X.-S. Hua: "Near-Lossless Video Summarization", Proc. ACM Multimedia, pp.351-360 (2009)
- 23) G. Irie, K. Hidaka, T. Satou, T. Yamasaki and K. Aizawa: "A Degree-of-Edit Ranking for Consumer Generated Video Retrieval", Proc. ICME, pp.1242-1245 (2009)
- 24) Z.-J. Zha, L. Yang, T. Mei, M. Wang and Z. Wang: "Visual Query Suggestion", Proc. ACM Multimedia, pp.15-24 (2009)
- 25) 板東武彦: "高臨場感が与える脳や生体へのインパクト", 映情学技報, 32, 49, pp.7-12 (2008)
- 26) 井ノ上直己: "超臨場感コミュニケーションにおける人の感じる臨場感評価", 信学技報, CQ2008-47, ICM2008-44 (2008)
- 27) 寺本ほか: "臨場感の素朴な理解", TVRSJ, 1, 1, pp.7-16 (2010)
- 28) 栗田ほか: "リアルバーチャリティ映像の中のうそー", 映情学技報, 33, 27, pp.33-38 (2009)
- 29) 南谷ほか: "変型ビルボードを用いた人物像の提示による複合現実感卓上作業の遠隔共有", 日本バーチャルリアリティ学会論文誌, 13, 3, pp.363-373 (2008)
- 30) 高見ほか: "道具型デバイスを活用した複合現実空間での木材加工", 信学技報, 109, 376, pp.91-96 (2010)
- 31) 能登ほか: "高臨場感対面遠隔コミュニケーションシステム:Tele-Face", 信学技報, MVE, 109, 215, pp.17-22 (2009)
- 32) 稲垣ほか: "文字提示における時間構造の表現効果", 信学技報, NLC, 108, 283, pp.13-18 (2008)
- 33) http://www.soumu.go.jp/main_sosiki/joho_tsusin/telework/
- 34) <http://www.japan-telework.or.jp/about/index.html>
- 35) 超臨場感コミュニケーション産学官フォーラム, <http://www.scator.jp/urcf/>
- 36) 徳満ほか: "超臨場感テレワークシステム", Okiテクニカルレビュー, 76, 2, pp.58-61 (2009)
- 37) 國田ほか: "3次元イメージファクシミリの実用:遠隔地間の3次元物体多人数共有システム", 映情学技報, 33, 47, pp.1-6 (2009)
- 38) 石川ほか: "ウォークスルーを実現する自由視点映像合成方式", 情学技報, CVIM, 2008, 27, pp.201-206 (2008)
- 39) 野村ほか: "超臨場感テレワークシステムにおける能動的メディア処理の実現に向けた音源分離定位品質の基本評価", 信学技報, 109, 472, EA2009-119, pp.25-30 (2010)
- 40) 岸野文郎: "人に優しい究極のヒューマンインタフェースを目指して", 映情学技報, 33, 44, pp.17-20 (2009)
- 41) 加藤俊一: "感性ロボティクス", 人工知能学会誌, 21, 2, pp.183-188 (2006)
- 42) 多田昌裕, 加藤俊一: "類似する画像領域の特徴解析と視覚感性のモデル化", 信学論, J87-D-II, 10, pp.1983-199 (2004)
- 43) 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ: "印象語による絵画データベースの検索", 情処学論, 33, 11, pp.1373-1383 (1992)
- 44) 辻康博, 星守, 大森匡: "曲の局所パターン特徴量を用いた類似曲検索・感性語による検索", 信学技法, 96, 565, pp.17-24 (1997)
- 45) 内山直子, 柴田滝也: "都市景観と建築物との調和分析とモデル化に関する研究", 日本建築学会計画系論文集, 73, 623, pp.241-248 (2008)
- 46) 沼口直紀, 中澤篤志, 竹村治雄: "印象語による舞踊動作データの分類法", 情報処理学会CVIM167予稿集 (2009)
- 47) 岡田昌也, 鳥山朋二, 多田昌裕, 角康之, 間瀬健二, 小暮潔, 萩田紀博: "実世界重要体験の抽出・再現に基づく事後学習支援手法の提案", 信学論, J91-D, 1, pp.65-77 (2008)
- 48) 多田昌裕, 瀬川誠, 岡田昌也, 蓮花一己, 小暮潔: "装着型センサを用いた運転技能自動評価システムの開発と講習現場への導入の試み", 信学技法, 108, 263, pp.1-6 (2008)
- 49) 肥後智昭, 宮崎大輔, 池内克史: "センサフュージョンによる効率的な3次元モデルの推定と表現", 映情学誌, 64, 1, pp.112-119 (2010)
- 50) 巴図斯仁, 土橋宜典, 山本強: "イメージクラスタからの任意視点画像生成", 映情学誌, 63, 10, pp.1453-1458 (2009)
- 51) 高井勇志, 松山隆司: "Harmonized Texture Mapping", 映情学誌, 63, 4, pp.488-499 (2009)
- 52) 山崎俊彦, 前田崇, 中川卓, 相澤清晴: "スケルトンによるTime-Varying Meshの動き抽出と合成", 映情学誌, 63, 10, pp.1459-1466 (2009)
- 53) 本多健二, 橋本直己, 佐藤誠: "奥行きモデルを用いた擬似広視野映像の実時間生成", 映情学誌, 62, 4, pp.550-557 (2008)
- 54) 和田稔, 伊藤浩, 杉本和夫, 山岸秀一, 小林翼: "外部パラメータが未知なカメラの多視点映像による移動体のフロアモーション生成法", 映情学誌, 62, 4, pp.533-544 (2008)
- 55) 角田哲也, 大石岳史, 池内克史: "影付け平面を用いた複合現実感における高速陰影表現手法", 映情学誌, 62, 5, pp.788-795 (2008)
- 56) 原島ほか: "小特集 いま"顔"が面白い〜顔の画像処理とその応用〜", 映情学誌, 62, 12, pp.1909-1945 (2008)
- 57) <http://face.nist.gov/mbgc/>
- 58) <http://face.nist.gov/mbe/>
- 59) http://www.fujifilm.co.jp/corporate/news/articleffnr_0351.html
- 60) <http://www.oki.com/jp/press/2007/08/z07055.html>
- 61) http://www.sonycs.co.jp/research_gallery/cybercode.html
- 62) <http://www.xbox.com/en-US/live/projectnatal/>
- 63) <http://www.digital-signage.jp/>
- 64) <http://www.nec.co.jp/press/ja/0908/2002.html>
- 65) 須崎昌彦, 井上清司: "画像による人の年齢・性別推定技術の応用. 広告効果測定ミドルウェア「Signage Eye TM」の解説", 画像ラボ, 20, 12, pp.80-84, 日本工業出版 (2009)
- 66) <http://tokyo.interactive.ad.awards.jp/Results09/info/82.html>
- 67) <http://www.mobilefactory.jp/service/kaocheiki/>
- 68) Y.-T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua and H. Neven: "Tour the World: Building a Web-Scale Landmark Recognition Engine", Proc. CVPR (June 2009), <http://www.comp.nus.edu.sg/~g0501938/papers/cvpr09-landmark-yantao.pdf>
- 69) <http://www.evernote.com/about/home.php>
- 70) <http://www.google.com/mobile/goggles/>
- 71) <http://picasa.google.com/features.html>
- 72) <http://jp.techcrunch.com/archives/20090721its-facebook-photo-tagging-time-with-facecom-photo-tagger/>
- 73) <http://www.cvpr2009.org/cvpr09-paper-registration-and-submission>
- 74) 驚見ほか: "パターン認識・メディア理解の10大チャレンジテーマ", 信学誌, 92, 8, pp.665-675 (2009)
- 75) 金岡晃廣, 高野照久, 菅原大輔: "アラウンドビューモニタの開発", 日産技報, 63, pp.37-41 (2008)
- 76) 三井相和, 山内悠嗣, 藤吉弘亘: "Joint特徴量を用いた2段階Boostingによる物体検出", 信学論, J92-D, 9, pp.1591-1601 (2009)
- 77) 鈴木康弘, 山本和彦, 加藤邦人, 安藤道則, 小島真一: "近赤外マルチバンドによる肌検出手法の提案", 電学論C, 127, 4, pp.583-590 (2007)
- 78) 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一: "パーティクルフィルタとこう配法の組合せによる頭部姿勢と表情の同時推定", 信学論, J92-D, 8, pp.1349-1362 (2009)
- 79) 岡崎一郎: "高並列プロセッサIMAPCARと車載画像認識への適用", 信学技報, PRMU, 107, 57, pp.53-58 (2007)
- 80) 佐々木一人, 河本新二: "安全運転を支援する車載画像処理システム", 東芝レビュー, 59, 4, pp.28-31 (2004)
- 81) 村松彰二, 門司竜彦, 松本芳幸: "画像認識技術の自動車応用とプラットフォーム開発", 日立評論, 89, 08, pp.638-639 (2007)