

1

音声・映像情報の 構造化と検索

越仲 孝文, 大網 亮磨, 細見 格, 今岡 仁
(NEC 情報・メディアプロセッシング研究所)

音声・映像アクセス技術の現状

●増大する音声・映像情報

昨今、音声や映像を媒体とする膨大なマルチメディア情報が流通し、人々はそれらを容易に入手、視聴できるようになった。その背景として、通信のブロードバンド化、記憶デバイスの大容量・小型化と低価格化、安価かつ手軽なカメラやマイクの普及などがあることは言を俟たない。

YouTube や Ustream に代表される動画共有サービスはインターネット先進国を中心に林立し、放送事業者らによる動画配信サービスも次々と始まっている。また、個人のブログや SNS (ソーシャルネットワークサービス) による情報発信においても、音声や映像を活用することは一般的になってきている。さらには、パブリックセーフティの観点で街頭や施設内に設置する監視カメラの普及が進んでいるが、この種の定点カメラ映像も、最近では長時間かつ高精細なデータが記録できるようになっている。いまや、音声・映像情報は世界中に氾濫しているといつてよい。

音声・映像情報は、視覚や聴覚に訴え、テキスト情報よりも圧倒的にリアルもしくは豊かな内容を受け手に与える。また送り手の立場からは、自身が意図した情報を誤解なく確実に受け手に送ることができるという利点もあろう。

一方で、音声・映像情報は、時系列データであるがゆえ、その内容を完全に把握するためには、再生して、基本的には始まりから終わりまでのすべてを視聴しなければならない。つまり、再生の実時間に等しい時間を割く必要がある。

そこで、世の中に溢れる膨大な量の音声・映像情報から、手軽かつ迅速に所望の情報を検索したいというニーズが生ずる。たとえば、テレビで特定のタレントが登場するシーンのみを視聴したい、サッカー中継でゴールシーンだけを視聴したい、ニュースや討論、講義などにおいて、ある話題について言及されているシーンだけを視聴したい、映像コンテンツ制作の場面などで、飛行機や高層ビルなど、特定の物体が映っているシーンを集めて再利用したい、等々である。

音声・映像情報は、テキスト情報と違って、そのままではテキストクエリ(いわゆるキーワード)による検索ができない。そのため、音声・映像情報には、その内容を記述するメタ情報(またはメタデータ)を付与して管理することが以前からよく行われている。制作者、制作日、制作場所、登場人物/物体、編集履歴、注釈などといったメタデータを付与しておくことによって、キーワードで音声・映像情報を検索することができる。

以前は、テレビ番組や映画、音楽など一部の音声・映像コンテンツに対して、人手で少量のメタデータ

が付与されるのみであった。しかし、多種多様な音声・映像情報が日ごと急速に増大している昨今、人手でのメタデータ付与には限界がある。このような背景から、音声・映像情報にメタデータを自動的に付与する自動インデクシング技術が望まれ、研究が進展している。

●情報にアクセスするための技術

自動インデクシング技術は、信号処理、パターン認識、言語処理、データマイニングなどの手法をベースとして発展し、これまでにテキストのみならず多岐にわたるメタデータ付与の試みがなされている。自動インデクシングにより生成可能なメタデータとしては、主に以下が挙げられよう。音声認識を用いて生成される発話内容の書き起こしテキスト、カット点検出により生成されるカット点（編集やカメラ切替えによって挿入される、映像の時間方向の切り替わり点）、音声認識と言語処理（トピック分割）によって得られるトピック境界、顔認識や話者認識によって生成される人物IDまたは特徴量、一般物体認識によって生成される物体（富士山、自動車、法廷、など）のIDまたは特徴量、カメラモーション検出によって生成される映像撮影時のカメラモーション（ズーム、パン、フィクスなど）。

自動インデクシング技術にかかわる研究分野の特徴として、整備された研究用データセットによるコンテスト（評価ワークショップ）が、各研究機関の研究活動を促進している点が挙げられる。

米国立標準技術研究所（NIST）が行うコンテスト TREC（Text REtrieval Contest）は、テキスト検索でよく知られたコンテストだが、Spoken Document Retrieval（SDR）と呼ばれるタスクが設定され、放送ニュースなどの音声情報の検索課題を取り扱っている。国内でも、本会の音声言語情報処理研究会（SLP）において、音声ドキュメント処理ワーキンググループが講演音声のデータ整備を推進している。前出の TREC から派生した TRECVID（<http://trecvid.nist.gov/>）は、映像情報検索に特化したコンテストである。放送ニュースなどの映像データを題材として、カット点検出、シーン検出、映像要約な

```
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" ...>
  <Description xsi:type="ContentEntityType">
    <MultimediaContent xsi:type="ImageType">
      <Image>
        :
      </Image>
    </MultimediaContent>
  </Description>
</Mpeg7>
```

図-1 MPEG-7 表記例
（本会 情報規格調査会 Web サイトより転載）

どのタスクが設定され、毎年各国の研究機関が参加している。

これらの研究開発と並行して、メタデータ記述の標準規格を定める活動も進められている。標準規格は、メタデータの活用、ひいてはインターネットでの音声・映像情報の流通を促進するための重要な役割を担う。MPEG-7 は、メタデータの表記方法に関する代表的な国際標準規格であり、我が国でも、本会の情報規格調査会や各研究機関が策定に関与し、普及に努めている（図-1）。

以上、音声・映像アクセスにかかわる技術、特に自動インデクシング技術について概観した。現状がある側面から見た場合、音声（聴覚）情報を用いたインデクシング、映像（視覚）情報を用いたインデクシングが研究され、それぞれが独立に発展してきたといえる。今後もこの方向での発展が期待されるが、それとともに、両者の融合領域、つまり聴覚情報と視覚情報を併用したインデクシング技術が重要になってくるのではなかろうか。

このような観点も考慮し、以降では、音声・映像情報アクセスにおいて筆者らが関与したこれまでの研究事例、実用化事例をいくつか述べ、最後に今後の展望について示したい。

音声認識による映像シーン検索

●音声ドキュメント検索

前章で紹介した TREC SDR をはじめとして、音声情報を手がかりとした検索、すなわち音声ドキュメント検索の研究は盛んに行われている。その基本的な考え方は、音声・映像情報に含まれる音声を音声認識により（タイムコード付きの）テキスト情報に変換することで、音声の検索をテキスト検索の問題

に帰着させるというものである。

したがって、音声認識をいかに精度よく実行できるかが、検索精度に大きく影響する。このような観点で音声認識の精度を向上させようとする研究は多くの研究機関で行われている。

たとえば、NISTと米国防高等研究計画局(DARPA)が1990年代後半に実施した放送ニュースプロジェクト(Broadcast News)では、テレビやラジオの放送ニュースを題材として音声認識システムの改良を継続的に行い、単語誤り率を10%前後にまで低減した。放送ニュースの場合、キャスターによる原稿読み上げが多く、そのようなシーンではとりわけ高い認識精度が得られた。

続いて実施されたRich Transcription(RT)プロジェクトでは、少人数によるミーティングの音声認識(Speech-to-text;STT)が題材とされた。ミーティングは放送ニュースよりも話し方の自由度が高く、したがって音声認識の難易度も高いが、最近の単語誤り率は20%程度に至っている。

RTで扱われているような自由な話し言葉の認識は、音声認識分野における近年の最重要テーマの1つである。我が国においても、2000年代に入って自由な話し言葉の研究用データベース「日本語話し言葉コーパス(Corpus of Spoken Japanese;CSJ)」が構築された¹⁾。CSJは主として学会講演音声とその詳細な書き起こしからなり、大学などの研究機関による日本語話し言葉の研究促進に広く貢献している。

●裁判員裁判での実用化事例

筆者らの研究チームでは、独自に話し言葉の認識技術の研究開発に取り組んでいる。その中で、2009年の裁判員制度開始に向けて、最高裁判所から音声認識システムの研究開発を請け負い、2008年度に開発を完了、全国の地方裁判所の裁判員裁判用法廷で実用化している。本節ではこの実用化事例について紹介する。

日本の裁判員制度では、一般の国民が裁判員として裁判に参加し、数日間で連日審理を行って、裁判官と裁判員が評議により判決を出すというスピーディな裁判手続きとなる。このため、評議において裁

判員や裁判官が、法廷での証言内容を確認する必要が生じた場合に、映像・音声によって迅速に証言内容を確認することができるツールが求められることとなった。

本システムはその一環として採用され、全国の裁判員裁判用法廷にて運用されている。本システムにより、法廷での証拠調べにおける裁判官・裁判員、検察官、弁護人の尋問や、証人・被告人の供述の内容がリアルタイムでテキスト化され、同時に供述者の映像が記録される(法廷AP、図-2)。これらのデータは、数日間の審理の後に開かれる評議の場に運ばれ、キーワードによる映像の検索、頭出しなど、裁判官と裁判員が証拠調べの内容を音声と映像で迅速に確認する用途に活用されている(評議室AP、図-3)。

法廷における音声認識技術の応用として、米国の裁判所では、尋問や供述の内容を裁判所の職員が聞いて、言い直すことにより音声入力する例はあった。



図-2 法廷 AP：公判での音声認識画面例

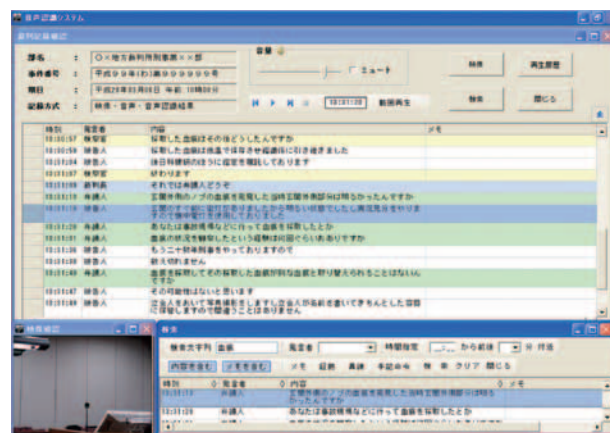


図-3 評議室 AP：評議での検索・再生画面例

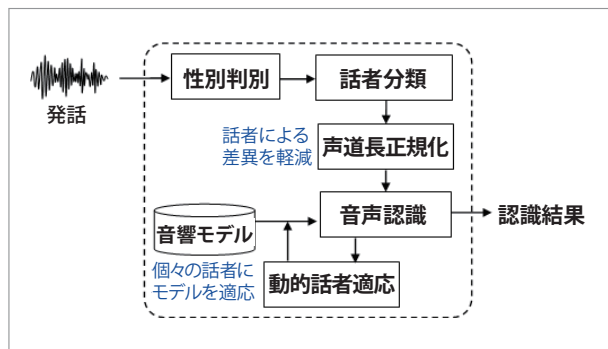


図-4 音声認識エンジンの概略構成：システムではエンジンが4プロセス起動し、各々裁判官・裁判員、検察官、弁護士、証人・被告人の音声进行处理する。

しかし、本システムのように尋問や供述の音声を直接認識する例は世界的に見ても前例がなく、初めての事例と考えられる。

以下、本システムの基幹部分である音声認識エンジンの主な技術的特長を2点示す。

多マイク環境下での音声検出：裁判員裁判の証拠調べでは大きく分けて、a)裁判官・裁判員、b)検察官、c)弁護士、d)証人・被告人という4種類の話者タイプが存在し、各々に与えられたマイクを使って発言する。システムはa)～d)のチャンネルごとに認識処理を行う。これにより、尋問者と供述者の音声に重複がある場合でも、各々の音声を認識できる。

しかしながら、法廷内は発言が隔々にまで通るように設計されており、ある話者の音声はすべてのマイクに回り込む。よって、どのマイクで発言がなされたかを逐次特定する技術が必要となる。我々は、話者とマイクの空間的な位置関係を利用して、マイクに近い話者の音声を（2人以上の発話が重なるケースも含めて）正確に検出する方式を開発した。

多様な話者への対応：法廷内で発言する話者は基本的に不特定であり、音声認識システムはさまざまな年齢、性別、職業等の話者を扱わねばならない。そこで、音声認識エンジンでは、1発話が発出されるたびに、それらを声質によって分類し、各話者への自動適応化（声道長正規化、動的な話者適応）を逐次実行する構成とした（図-4）。このような構成を採用することにより、話者ごとの個人差にきめ細かく対応し、多様な話者で高い認識性能を実現することができる。図-4のように要素技術を組み合わせたり

リアルタイム動作するようにした実用システムは、前例がなく、おそらく初めての例である。

以上、裁判員裁判での事例について述べた。本システムに搭載された技術は、裁判のみならず、人が集まって会話が発生する場面に広く応用できる可能性がある。たとえば、一般企業の会議などにも適用できるように技術改良を続けることにより、企業の生産性を向上するコミュニケーション・情報共有支援システムが実現できると考えられる。

自然な言葉による人物検索

● 定点カメラ記録映像の検索方法

冒頭でも述べたように、監視カメラなどの定点カメラは、最近では長時間かつ高精細な映像データが記録できるようになっている。施設内や街頭に設置されている定点カメラの用途として、不審者や行方不明者（迷子など）の発見がある。

不審者や迷子を捜す際、対象人物の写真などが手元にない場合が多く、記録された映像データの検索に類似画像検索のような技術は適用しにくい。主な手がかりとなるものは目撃者や捜索依頼者の証言であり、大抵は「グレーのジャンパーに野球帽」といった服装や付帯物の種類と色になる。そこで筆者らは、こうした言葉による手がかりから該当する人物の映像を検索するための技術を開発した。本技術は、自然言語による問合せ文（クエリ）から人物に関する画像特徴量を生成するクエリ変換処理、記録映像から人物の服に関する特徴量を抽出する映像解析処理、

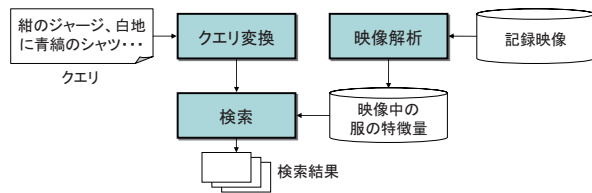


図-5 人物検索の機能構成

双方の特徴量を照合して合致度合いのスコアを算出する検索処理の3つからなる(図-5)。

●自然言語文からの画像特徴量生成

クエリ変換処理は、入力された自然言語文から、服や付帯物と色を表す語句、およびそれらの構成の判断に有効な助詞などの語句を抽出し、簡易な係り受け判定を行って人物の上半身および前後の色分布を推定する。たとえば、「紺のジャージ、白地に青縞のシャツと茶色っぽいジャケット」といった文が入力されると、服や色に関する辞書と服同士や着用時の身体位置との関係に関するオントロジーを用いて次のように解釈する。

- (1) 紺／ジャージ
→“ジャージ”が一般に身体を覆う領域から上半身と下半身が共に紺色
- (2) 白／に／青／シャツ(※「地」や「縞」は抽出せず)
→“に”が表す関係から上半身(シャツ)は大半(80%)が白で一部(20%)が青
- (3) 茶色／っぽい／ジャケット
→“っぽい”を反映し、上半身はHSV色空間上の茶色とその近傍色
- (4) (2)と(3)の服の種類(下着／上着)を比較
→上半身前面の一部(30%)のみ(2)の色、その他の上半身(前面の70%、背面の100%)は(3)の色
- (5) (1)と(4)を比較
→上半身は服装が明示的な(4)の色分布、下半身はほかに優先する情報がないため(1)の色分布

以上から、上半身前面は茶色とその近傍色が70%、白が24%(30%×80%)、青が6%(30%×20%)、上半身背面はすべてが茶色とその近傍色、下半身は前後面ともすべて紺色と推定する。個々の比率は事前に設定した固定値である。このようにして、身体

の部位ごとにHSV空間上の色の領域と各色領域の占有面積比率からなる画像特徴量を決定し、後述する映像解析処理で得た特徴量との照合に用いる。

対象の種類と色で画像を検索できる「Feelimage[®]」(<http://www.feelimage.jp/>)などの既存技術に比べ、本技術では複数の色や服の組合せを考慮した検索ができる点を特徴としている。警察が公開している不審者情報から抽出した101文と、他者の服装に対する印象を聞き取った25文を入力とし、出力された画像特徴量と実際の服装写真を目視で比較した際の正解率は、現在それぞれ90%と76%である。後者は曖昧な口語表現や省略が多い点が、精度の差に現れている。クエリ変換処理の詳細については文献2)を参照されたい。

●映像からの服の色分布解析

前節で述べた上半身による服の違いや、人の向き、重ね着等によって変化する人の見え方の違いを考慮して人物を検索できるようにするには、映像を記録する際に解析を行い、必要な情報をあらかじめ抽出しておく必要がある。本節では、この映像解析処理について述べる。

基本的には、映像から人物領域を抽出し、その服の色に関する特徴量(色のヒストグラム)を求める。人物領域の抽出では、まず、背景差分によって移動物体を抽出する。この際、背景の細かな動きに対して頑健にするため、時空間方向の相関を考慮した背景画素値の確率モデルを用い、事後確率最大化の枠組みで前景、背景を分離し、移動物体を抽出する³⁾。次に、画素値の勾配を反映した特徴量を抽出して一般化学習ベクトル識別器を用いることにより、得られた移動物体が人物かどうかの判定を行い、人物領域を抽出する⁴⁾。

服の色特徴の抽出においては、クエリ変換の処理結果が最大限活かせるように、上半身の服の自動分離と人物の向きの自動判定を行った上で色特徴を抽出する。以下ではそれぞれの技術について述べる。**上半身の服の自動分離**：上半身の分離可能性を判定した後、上半身の服の色の違いを検出し、分離する。分離可能性の判定では、画像上の人物の位

置による体の見え方の違いを事前知識として用いる。カメラから床面が見える位置に存在する人物は全身が映り、分離可能であるが、机などの障害物で下半身が隠れる位置にいる人物は上半身のみが映り、もともと分離できない。このため、画像上で人物領域の底部の位置を推定することで人物の存在位置を求め、事前知識を用いて分離可能性を判定する。そして分離可能と判定された場合は、分離境界を検出する。図-6に示すように、服装の画素値を鉛直方向の軸に射影した関数値が大きく変化する位置を検出することにより、分離境界を求める。なお、ワンピースなど上下半身が不可分な服の場合は、上下半身で服装の特徴が変化せず、境界は求まらないため、便宜的に上下方向の中央で分割する。

人物の向き判定：顔の向き、人物の動き、服の対称性の3つの情報を用いて人物の向きを判定する。顔の向きは顔の検出時に同時に推定している。人物の動きは、人物領域の動きベクトルから算出する。服の対称性は、図-6に示すように、服の画素値の水平方向の変化を各高さに対して調べ、得られる関数の対称性の度合いを算出する。正面や背面から人物を見た場合、服は水平方向に対称性を有する場合が多いため、この度合いを数値化して向きの判定に用いる。ただし、平坦な模様の服は向きによらず対称性を有するため、数値化の際は、平坦性も考慮する。これらの3つの情報を統合し、人物の向き（正面、背面、それ以外）を判定する。

以上、映像処理解析について述べたが、監視映像は通常24時間録画され続けるため、リアルタイムで処理できることが重要である。上述の処理は、NTSC解像度の画像に対し、Pentium 4 3GHzクラスのPCで1秒間に10フレーム以上の処理を行うことが可能である。また、処理精度については、照明条件や人物の混雑状況等によって変化するが、人物抽出が可能な状況においては、精度は8割程度である。今後の実用化に向け、さらに精度向上を図っている。

●特徴量の照合による検索

以上のようにして自然言語文から生成された画像特徴量を記録映像から抽出したそれと照合すること

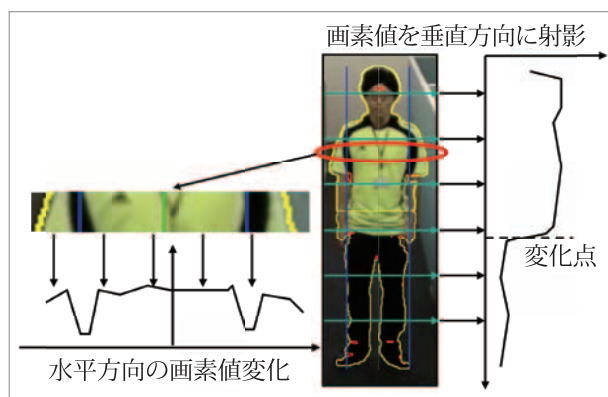


図-6 服の上下半身の自動分離と対象性判定

により、検索を実行する。その際、各特徴量の抽出状態を考慮した照合を行う。基本的には、上下半身ごと、向きごとに服の特徴量を照合する。たとえば、正面方向で上下半身の服の特徴量が共に得られた場合には、上下半身正面のクエリ側特徴量と照合を行い、検索スコアを算出する。一方、上半身の特徴量しか得られていない場合には、上半身の特徴量のみで照合する。また、特徴量抽出時に向きを判別できなかった場合には、正面・非正面の両方のクエリ側特徴量と照合を行い、スコアの高い方を選択する。上下半身が分離できない場合には、上下半身の特徴量を合成して1つの特徴量にしてから照合を行う。これにより、特徴量の抽出状態に応じた柔軟かつ精度の良い人物検索を可能にしている。

検索例を図-7に示す。クエリとして、「白シャツと紺のデニム、らくだ色のパーカー」という自然言語文を与えた際の検索結果を示している。先にも述べたように、パーカーの方がシャツより上に着るため、上半身のクエリの色特徴量は、正面は、らくだ色70%、白30%、背面は、らくだ色100%として生成され、検索が実行される。上述の通り、向きを考慮して色特徴量を照合するため、正面と背面の両方のシーンが検索されている。

本技術によって、文字によるメタデータが付与されていない映像からでも、自然言語文から特定の服装をした人物を検索できる。クエリが自然言語で書けるため、さらに音声認識技術を組み合わせることにより、口頭での目撃証言から直接検索を行うことも可能になる。

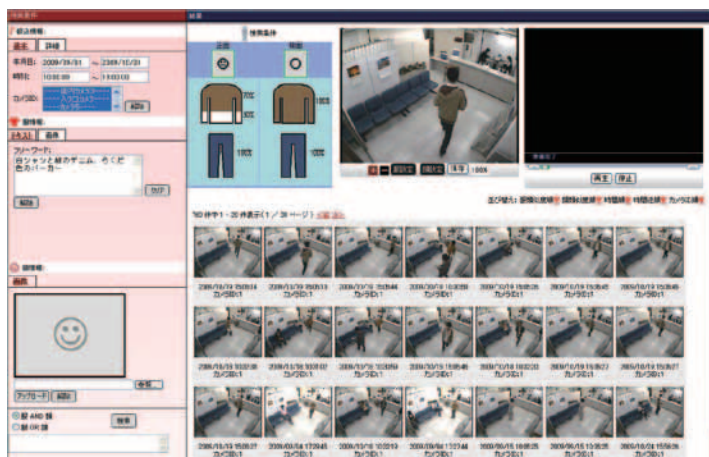


図-7 人物検索結果の例

顔認識による人物検索

●監視映像に対する人物検索

自然な言葉による人物検索で述べた服特徴量は、人物を検索する上で重要な手がかりとなる一方、それだけでは、人物を一意に特定することが困難な場合が多い。特に似た服を着た人が数多く映っている映像の場合には、検索結果から探したい人のシーンのみを抽出することが困難になり、個人が特定可能な特徴を用いた検索が必要となる。

これを可能とするのが顔認識技術である。人物特定が可能なバイオメトリクス情報のうち、顔は監視映像等に直接映っている情報であるため、監視映像からの人物検索に利用できる。このため、服と顔を合わせて人物を検索するシステムが、これまでにいくつか提案されている（たとえば文献5）。これらはいずれも、検索したい人物の写った写真をクエリとしてシステムに入力すると、顔の特徴量を抽出し、あらかじめ解析しておいた映像中の人物の顔特徴量と比較することで、人物を検索するものである。実際、図-7で示した人物検索システムでも、この機能が搭載されている。

しかしながら、顔照合技術は、正面に近い向きの顔に対しては高い照合率を出す一方、正面から外れた斜めの顔になると、照合率が大きく低下する場合が多い。また、逆光や影の映りこみなど、照明条件が厳しい状況下では、照合率が低くなる。このように、顔の向きや照明の変化への対応が、実用化する上で重要なファクタになる。特に監視映像の場合に

は、撮影対象人物が撮影に協力的なホームビデオ等とは異なり、撮影対象人物は通常カメラの方を向かず、顔の向きが正面でない場合が多い。また、照明条件が厳しい場合も多々ある。

このような状況下であっても人物検索を可能にするには、正面でない顔の向きや照明条件の変化に対して頑健に照合可能な技術が必要になる。次節では、顔照合技術の強化の動向について述べる。

●検索精度向上に向けた顔照合技術の進展

NISTでは、顔認証技術の開発の加速と実用性能評価を目的として、定期的に顔認証技術のベンチマークテストを開催している。最近では、2008年から2010年にかけて、MBGC (Multiple Biometric Grand Challenge, <http://face.nist.gov/mbgc/>) と MBE (Multiple Biometrics Evaluation, <http://www.nist.gov/itl/iad/ig/mbe.cfm>) の2つのテストが実施された。前者は薄暗い照明や直射日光などのさまざまな照明条件下での照合性能を評価するものであり、後者は百万名規模の大量の顔画像を用いた顔画像検索の実用性を検証することを目的としている。

筆者らの研究チームもこれらのテストに向けて顔認識技術の強化を行ってきており、顔検出、顔照合の両方の技術を改良している。顔検出では、顔の特徴点ごとに構築した一般化学習ベクトル量子化識別器により、特徴点候補位置を決定し、最小メジアン推定により、高精度に位置を検出する方式を開発した。また、顔照合では、顔の3次元形状をベースに向きや照明の異なる顔画像を生成する摂動空間法に

加え、変動に対して安定かつ個人識別に有効な空間に特徴を射影する多元特徴識別法を併用し、識別精度を向上させている。この結果、MBGCのテストでは、2～4%の誤照合率(他人許容率が0.1%のとき)、MBEのテストでは、犯罪記録から抽出した160万人の顔画像の検索で92%、ビザ申請時に使われた180万人の顔画像の検索では95%という高い精度を達成している⁶⁾。

上述のNISTによるベンチマークテストは、実利用を想定したデータベースで行っており、ここで高精度な検索が実現できていることは、顔による検索が監視映像にも適用可能になりつつあることを示している。今後、顔の向きや照明変動への頑健性に加え、低解像度・高圧縮画像に対する頑健性を向上させていくことにより、監視映像検索で顔認識が利用可能になることが期待される。

今後の展望

本稿では、映像、音声、テキストのさまざまな解析結果を密接に組み合わせることにより、時間という縛りがある音声・映像情報へのアクセスを改善する事例について述べてきた。冒頭で述べた音声・映像情報の増加は、今後さらに拍車がかかり、増加の一途を辿ると予想される。このため、複数の解析手段を有機的に組み合わせることで解析し、音声・映像情報へのアクセスを容易にしたいという要望は、さまざまな場面でますます大きくなるであろう。そして、アクセスが容易になることにより、従来死蔵されていた音声・映像データが有効に活用される事例が増えてくるのが期待できる。本稿で紹介した自然な言葉による人物検索は、複数の解析手段の組合せの一例といえよう。ほかにも、たとえば顔認識と音声認識(話者認識)を組み合わせることで相補的に動作させることができれば、照明や雑音の条件が厳しい環境でも、いずれかの解析手段が働いて頑健な人物検索が実現できると考えられる。また、一般物体の検索などでも、たとえばヘリコプターをその形状だけから認識するよりも、プロペラの旋回音も併せて認識し

た方がよいことは容易に想像がつく。

今後、このような流れの中で、さまざまな分野で解析技術を組み合わせることを想定すると、複数の技術をいかに容易に組み合わせられるかが重要になってくる。本稿で述べた各システムは、それぞれ個別に作り上げているため、あるシステムで得られた結果を他のシステムに利用することが、まだそれほど容易ではないのが現状である。

この促進に向け、さまざまな解析技術を収容し、それらを組み合わせることで活用することが容易になるプラットフォームの構築が重要になる。そして、さまざまな解析モジュールが、共通の仕様で扱うことができるようになれば、従来はコスト高で困難だった分野でも技術が活用できる可能性が高まる。NECでも、この基盤の整備に取り組み、本稿で述べた技術が適用可能な分野を広げていく予定である。

参考文献

- 1) 国立国語研究所 日本語話し言葉コーパス, <http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/>
- 2) 細見：自然文から視覚情報を復元する一手法と人物検索への応用, 情報処理学会研究報告, Vol.2010-DD-74, No.4 (Jan. 2010).
- 3) 池田, 石寺：尤度ベースの背景モデルを用いた物体検出手法, FIT2008, H-051 (2008).
- 4) 細井, 石寺：動き領域の見えに基づく物体認識, FIT2006, I-030 (2006).
- 5) 大網, 細見, 中島, 原田：車両・人物向けメタデータ解析技術とその応用, NEC 技報, Vol.63, No.3 (Sep. 2009).
- 6) 今岡, 早坂, 森下, 佐藤, 広明：顔認証技術とその応用, NEC 技報, Vol.63, No.3 (Sep. 2010).

(平成 22 年 11 月 11 日受付)

■越仲孝文 koshinak@ap.jp.nec.com

1993年京都大学大学院修士課程修了。日本電気(株)情報・メディアプロセッシング研究所主任研究員。パターン認識の研究に従事。

■大網亮磨 r-oami@az.jp.nec.com

1995年東京大学大学院修士課程修了。日本電気(株)情報・メディアプロセッシング研究所主任研究員。画像認識・検索の研究に従事。

■細見 格 (正会員) i-hosomi@ay.jp.nec.com

1993年神戸大学大学院修士課程修了。日本電気(株)情報・メディアプロセッシング研究所主任研究員。自然言語処理・セマンティック技術の研究に従事。

■今岡 仁 (正会員) h-imaoka@cb.jp.nec.com

1997年大阪大学大学院博士課程修了(工博)。日本電気(株)情報・メディアプロセッシング研究所主任研究員。脳視覚情報処理・顔認証技術の研究に従事。