

社会の便利・安心に貢献する メディア処理技術の研究開発

Research and Development of Multimedia Processing Technology Dedicated to Safe and Convenient Society

長坂 晃朗

Nagasaka Akio

メディア処理技術とは

今日、IT（Information Technology：情報技術）は人々の暮らしや社会活動全般を支える基盤技術として欠かせないものとなっている。人間の行うさまざまな活動の中で、かつて人手で行っていた多くの面倒な仕事を代行したり手助けしたりするとともに、人間の持つ能力だけでは実現できなかった新しい価値を生み出し、社会の形を変えていく原動力にもなっている。

こうしてITが社会の多くの場面で活用されるようになった背景には、メディア処理技術の果たしてきた役割が小さくない。

メディア処理は、人間が活動するリアルな世界（実世界）と情報システムとをつなぐ界面を支える技術である（図1参照）。

実世界に存在、もしくはそこで発生する事物や事象の多くは、各種の計測デバイスによってデジタルデータ化でき、情報システムの中で扱えるようになる。これらのデータは、計測方法に応じて、音声や画像、動画といった各メディアの形式に分類できる。例えば、カメラを使って視覚的に計測すれば画像メディア、マイクによって聴覚的に計測すれば音声メディアとなる。そして、こうした多様なメディア形式のデータを通じて実世界を観測することで、情報シ

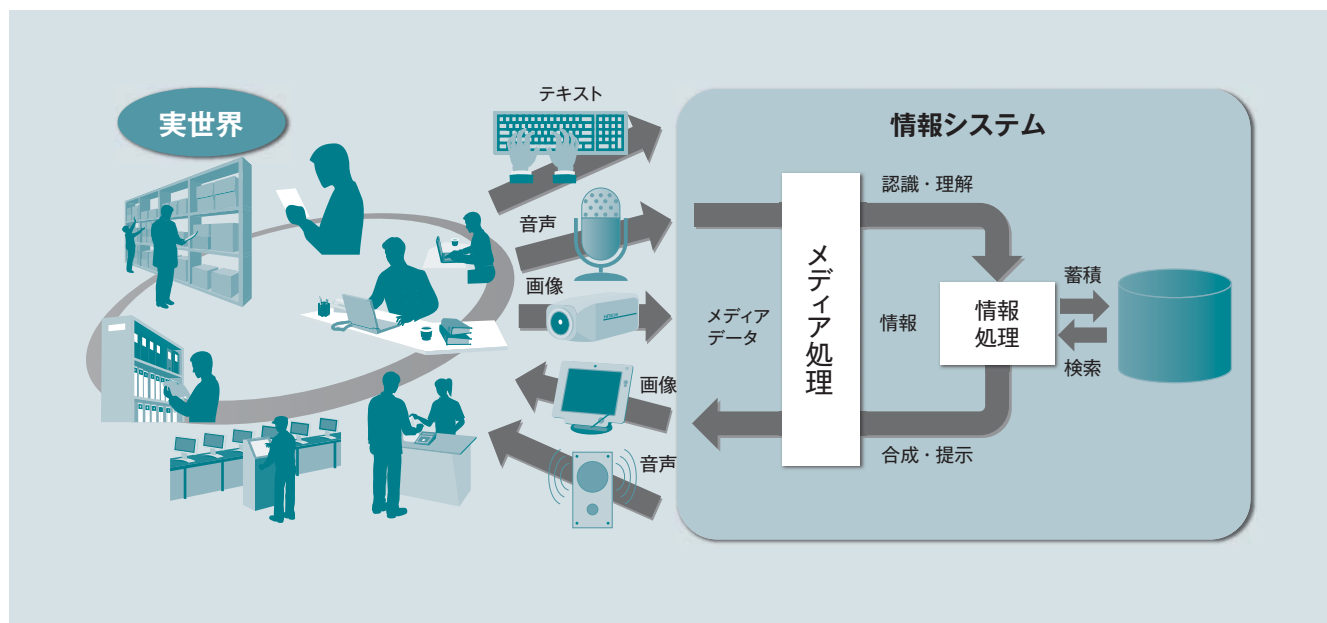


図1 | 情報システムの中におけるメディア処理の位置づけ

メディア処理は、実世界と情報システムとの界面を支える技術である。

システムは、実世界で生じる数多くの課題を余さず捉えられる。すなわち、ITの活躍の場を広げることができるということである。また一方で、情報システムの側から人間に対して、その置かれた環境や条件に最適なメディアを通して働きかけることができれば、人間にとってITはさらに使いやすくなる。

しかしながら、こうして実世界を計測して得られる各メディア形式のデータは、そのままでは記号の羅列に過ぎず、不必要な情報も多く含まれている。それらのデータを各メディアの特性に合わせて適切に解釈し、情報システムの中で有効に使えるように変換する仕組みが不可欠である。また、情報システムが処理した結果を、実世界の表現形態に戻す逆変換の仕組みも欠かせない。メディア処理は、こうした実世界と情報システムとのやり取りを支える役目を持つ。

ここでは、まずメディア処理技術の発展の流れについて俯瞰（ふかん）し、次いで日立グループにおける研究開発の取り組みを紹介しながらメディア処理技術の現状と今後について解説する。

メディア処理技術の潮流

メディア処理の研究には、その先駆けと言えるものまで含め、長い歴史がある。20世紀初頭には周波数スペクトルに着目した音声処理の手法がベル研究所のDudleyによって確立され、1939年のニューヨーク万国博覧会で、それを応用した音声合成のデモンストレーションが行われている¹⁾。また、画像処理に関する先駆けとしては、1920年代終盤、文書を読み取って各文字を認識する、現在の光学式文字読み取り装置OCR^(a)につながるコンセプトの特許をTauschekが取得している。ここで示された方式は、機械式の実装ながら、今でも画像処理の代表的な技法として知られるテンプレートマッチングの原型であった²⁾。

その後、1950年代から1960年代にかけ

てコンピュータが商用利用できるようになると、コンピュータを用いた本格的なメディア処理の研究開発が一気に進んだ。前述の文字認識の流れでは、伝票などに記載された金額の自動読み取りによる売上管理の効率化、郵便番号識別による郵便仕分け作業の支援といった具体的な産業応用が早々に始まった。コンピュータにカメラをつなぎ、撮影された空間内の物体などを認識するコンピュータビジョンの研究もこのころ始まっている。音声認識においても、この当時、数字のみに限定されるものの、人間の発話音声を認識する技術が世に示された。

1970年代に入ると、メディア処理、中でも画像認識技術の産業応用の動きが一段と活発になっていった。コンピュータビジョンの応用先として、工場での生産工程の自動化、すなわちFA (Factory Automation) が注目されるようになった。日立グループでも、電子部品製造現場向けに、プリント基板中の傷を自動抽出する装置を開発している³⁾。この装置は、基板を1枚ずつカメラで撮影し、複雑パターン中に埋もれた微小な傷を、1画面当たり $\frac{1}{60}$ 秒で自動的に発見することができた。さらには、検査だけでなくトランジスタを全自動で組み立てる装置の開発にも成功している。この装置は、トランジスタの電極位置を自動的に認識し、従来、人手で行っていた電極と外部リード端子との配線作業を自動化した。

1980年代になり、マイクロプロセッサの普及など、コンピュータの小型化と高性能化が進むと、メディア処理はオフィスへと進出していった。設計図や地図などの図面をコンピュータ上でデータベース化して管理したいという要求から、図面認識の研究が盛んになった。さらには、図面に限らず業務で使われるさまざまな定形・不定形の文書を電子化したいというニーズに向け、レイアウト認識や構造認識などを含む文書画像理解の研究も活発になった。

1990年前後からは、コンピュータが動画などデータ量の多いメディアも扱えるようになり、またそうしたメディアが大量に

(a) OCR

Optical Character Readerの略。光学式文字読み取り装置。手書き文字や活字をイメージスキャナなどで画像として光学的に読み取り、あらかじめ記憶させたパターンとの照合により文字を特定し、電子的な文字データに変換する装置をさす。

生産・蓄積されるようになって、認識の対象は文書や図面からマルチメディアコンテンツへと拡大していった。動画や音声を解析し、その内容に基づく検索や構造化を目標としたマルチメディアデータベースの研究も進んだ。ここで生まれたコンセプトは、現在のインターネット上の画像検索エンジンや写真共有サービス、特に写真などを自動的に解析して整理してくれる機能を持った高付加価値サービスの源流となった。

一方、この時期から人々の生活を支えるさまざまな機器の電子化・デジタル化が進み、機器の中に普通にコンピュータが組み込まれるようになると、人と機械のよりよいインタラクション（HMI：Human Machine Interaction）のためにメディア処理技術を活用しようとする動きが広まっていった。例えば、カーナビゲーションシステムでは、ドライバーが運転中に画面を見なくても安全に使えるように、車両が交差点など重要地点に到達するのに合わせて音声合成で案内をしたり、音声認識で操作を指示したりできる機能が採用された。こうした流れは現在さらに広がっており、顔認識を利用して被写体の顔に自動で焦点を合わせてくれるデジタルカメラや、人間の体の動きをカメラで認識してゲームが楽しめる入力デバイスなどが生まれている。最近ではスマートフォンやタブレット端末が、それら単品の機器を統合する勢いで高機能化しており、**拡張現実感^(b)**など高度なメディア

処理が時と場所を選ばず提供できるようになった。

メディア処理への日立の取り組み

このように歴史を俯瞰してみると、メディア処理技術は、コンピュータや通信技術の進化、それに伴う社会構造や生活スタイルの変化と歩調を合わせた技術革新によって、その適用範囲を拡大してきたことが分かる。

その黎明（れい）期には、コンピュータが認識しやすいよう整えられた環境や条件の下で、限定された作業の一部を自動化できるだけであったが、長年の研究開発により、そうした制約が次第に取り払われていった。

こうして制約が外れていくと、メディア処理は、ある時点で、限られた職人だけが使いこなせる仕事のツールから、一般の人が便利に使える日常的なツールへと位置づけが変化し、社会生活に不可欠なものになっていく。扱える対象の規模が大きくなるにつれ、ツールの枠を超え、社会全体を見守る基盤へと、その期待される役割も大きくなっていく。そして、こうしたイノベーションの先に、人々が便利で安心して暮らせる豊かな社会が形づくられる。

日立グループにおいても、このような社会の実現に向け、メディア処理の技術革新に古くから取り組み続けている。それらの取り組みは、データ授受の流れのどこに技術的力点を置くかによって、次の三つのカテゴリに分類することができる（図2参照）。

- (1) データの出力（見せる・示す技術）
- (2) データの入力（見分ける・理解する技術）
- (3) データの入出力（蓄える・探す技術）

最初の(1)は、情報システム側から実世界へのデータの流れを主に支える技術から成るカテゴリである。情報を利用者にとっていかに分かりやすく伝えるか、美しく感じさせるかが重要になる。これらの技術を必要とする用途では、もともと情報に不足や欠落があったり、出力に利用できる機器の性能に制限があったりすることが多

(b) 拡張現実感

Augmented Reality (AR)。現実の世界から得られる知覚情報に、コンピュータが作り出した情報を重ね合わせることで、情報を補足したり、情報を付加して強調したりする技術。例えば、カメラが捉えた目の前の画像や映像に、コンピュータグラフィックスの文字やオブジェクトなどを重ね合わせて表示することで、肉眼では見えない部分の映像を補ったり、関連情報を提供したりする。

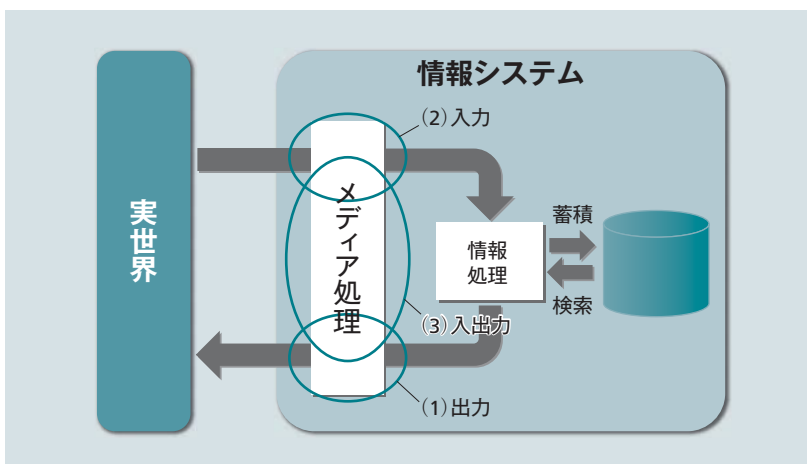


図2 | メディア処理技術の分類

メディア処理は、データ授受の流れのどこに技術的力点を置くかによって、三つのカテゴリに分類できる。

い。そうした各種の制限をどう両立させ、受け手の印象を高めていくかが技術の特長になる。

日立グループでは、映像や画像を高画質化する技術や、高品質な音声を合成して出力する技術などの研究開発を進めている。

映像高画質化の分野では、ぼやけ感やノイズなどによって平板な印象を受ける低品質映像に対して、「複素信号処理」の利用などによって高精細なテクスチャを再現する超解像技術を開発し、デジタルテレビ向けに製品化している。また、映像データをコンパクトに蓄積したり伝送したりするために行われる符号化圧縮の過程で生じるノイズの除去や、人間の視覚特性を利用したコントラスト補正といった技術開発も進めている。

一方、音声に関しては、日本語の細かな韻律特徴の考慮や話者の話し方の癖なども再現できる、統計的機械学習を用いた韻律モデリング技術、ならびに音質や韻律の違いによって**波形重畳と波形接続^(c)**の二つの方式を動的に切り替える波形合成技術を開発している。これらによって、自然で肉声感のある音声合成を実現し、製品化している（図3参照）。

次の(2)は、上記(1)とは逆方向、すなわち実世界側から情報システムへのデータの流れを主に支える技術カテゴリである。実世界の事物・事象を人間に代わって認識し、情報を取り出すための方法論が中心課題であり、特に実世界を計測した玉石混交のデータ中から、価値ある情報をどのようにして見分け選び取るかが重要になる。このとき、実世界中のあらゆる対象を一律のやり方で認識しようとしても実用的な精度の達成は難しいため、認識対象を特化し、その対象に固有の特徴を使って識別を行うことが多い。そのため、認識の対象ごとに技術が分化しており、それぞれに特色ある技術が生まれている。最近では、大規模な事例データの収集や蓄積・活用が容易になり、大量の事例データを機械に自動学習させることで、より汎用性の高い認識をめざした研究も進んでいる。

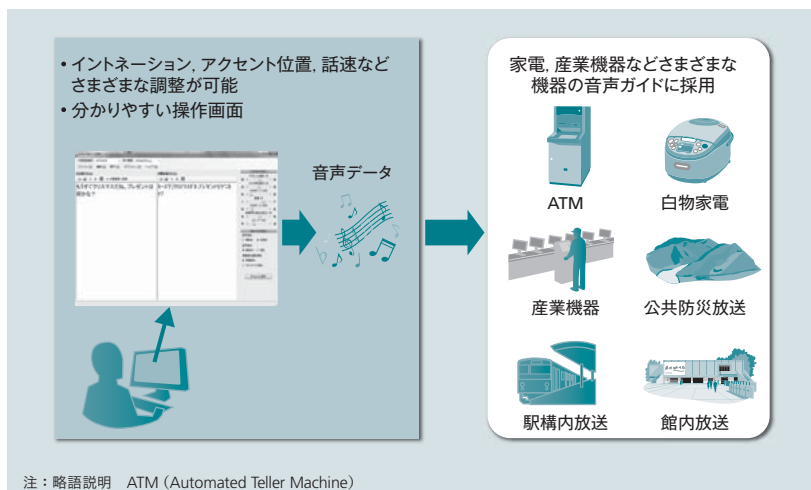


図3 | 音声合成システムの応用分野

家電製品のガイダンスや公共空間のアナウンスなど、さまざまな分野に活用されている。

日立では、FA応用、セキュリティ応用、ITS (Intelligent Transport System) 応用、オフィス応用などの分野で、ハードウェアからアプリケーションまで幅広い技術ラインアップを有している。

例えば、FA分野では、配線設計データと検査画像を照合することで微細加工に起因する配線不良を検出する半導体検査装置、セキュリティ分野では人間のさまざまな動きや多様な顔をロバストに抽出し、侵入者などの異常を検知する監視システムや、人間の指の内部にある静脈パターンを画像計測し、そこから抽出した特徴を用いて個人を識別する指静脈認証装置などの製品がある（図4参照）。

ITS分野では、交通事故防止を目的として、さまざまな天候や昼夜による照明変動、路面標示のかすれといった悪環境においても誤動作なく車両、歩行者、レーンを検知する車載カメラシステムを製品化している。また、文字パターンの局所的・大局

(c) 波形重畳と波形接続

合成音声を作成する技術の代表的な方式。波形重畳は、基本単位(1ピッチ)の音声波形を、目標の音の高さとなる間隔でずらしながら重ね合わせていく方法。自由な高さの合成音声を作成できるが、重ね合わせ処理によって音がこもる傾向がある。波形接続は、あらかじめ録音した音声を音素や音節などの単位でデータベース化しておき、目標の高さ・長さに近いものを選んで単純に接続することで合成音声を作成する方法。波形重畳のような信号処理を行わないため、よりクリアな音声を得られるが、高さ・長さが一致する音声がない場合に抑揚が不自然になりやすい。

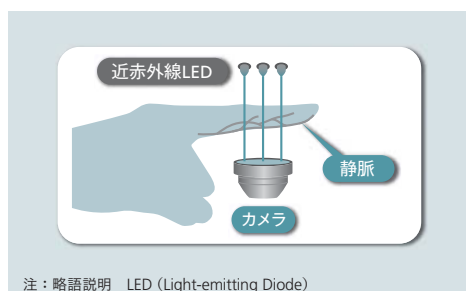


図4 | 指静脈認証技術

近赤外線を指に透過させて得られる指の静脈パターンの画像によって個人認証を行う。

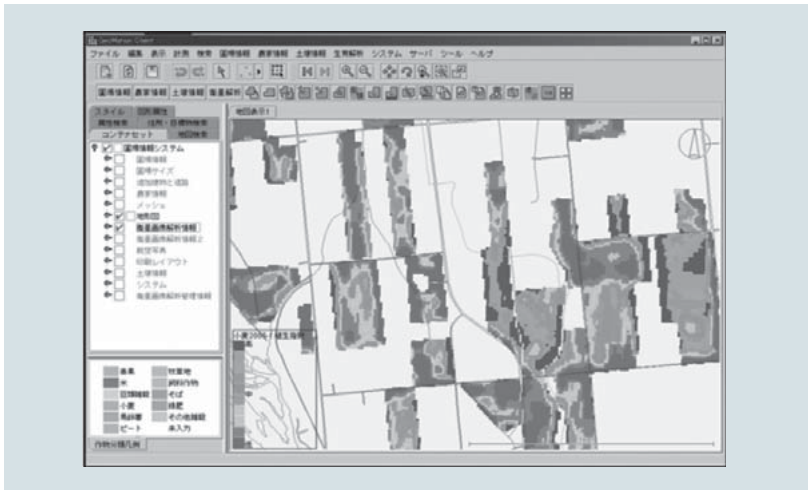


図5 | 衛星画像による小麦の生育状態の解析画面例
小麦や稲の最適な収穫時期を視覚的に表示する。

的特徴を抽出するパターン特徴抽出器の組み合わせを大量データによる統計学習で最適化することで文字の誤読を低減したOCR技術や、PDF (Portable Document Format) などの電子文書をデータベースへ変換する構造化技術なども手がけている。さらには、農業への応用として、衛星画像の解析によって、小麦や稲の最適な収穫時期を視覚的に表示するシステムも製品化している(図5参照)。

最後の(3)は、実世界からの情報を情報システムの中に取り込んで蓄え、利用者からのリクエストに応じてその中から必要な情報を選び出して実世界側に提示する、いわば(1)と(2)の両方にまたがるカテゴリである。(1)や(2)と違って、認識の

正確さや出力の美しさよりも、利用者が求めている情報が何であるかをいかに的確に情報システム側に伝え、大量の蓄積情報の中から、どれだけ分かりやすくタイムリーに提供できるかが重要な技術課題となる。

日立グループでは、大量に蓄積された画像や音声のデータに対して高速に検索が行える技術の研究を進めている。

まず画像データについては、利用者が指定した画像に類似した画像を、対象となる画像集合の中から探し出す類似画像検索技術を開発している。この技術では、各画像から特徴量を抽出し、それらの更新や格納方法に工夫を凝らすことで高速な検索を実現した。これにより、1万時間の放送映像の検索、100台のカメラを接続した映像監視、Webから収集した1億件の画像を対象とした検索を実用的な時間内で可能としている(図6参照)。

次に音声に関しては、テキストで入力されたキーワードと同じ言葉が話されている音声データを探し出す技術や、話者を識別する技術の開発を進めている。この技術では、音声データを単語列として認識する代わりに、音素や音節などのサブワードの列として認識し、検索実行時には、入力キーワードもサブワード列として表現することで、あらかじめ辞書に登録されていない未知語に対しても検索が可能となっている。これによって、人名や製品名などの固有名詞での検索ニーズが強いコールセンターなどへも適用が可能になった。

さらなるイノベーションをめざして

昨今の計測デバイスの目覚ましい進歩や普及により、部分的には人間の五感をも超える感覚器が容易に情報システムに備えられるようになった。また、通信ネットワークの高速化・モバイル化、スマートフォンやクラウドなどによるコンピューティング基盤の遍在化・高性能化は、「ビッグデータ」のキーワードに代表されるように、実世界から得られる大量のデータを瞬時に処理していけるだけの能力を着実に情報システム



図6 | 類似画像検索の応用例
類似画像検索の応用例(類似顔画像検索システム)を示す。

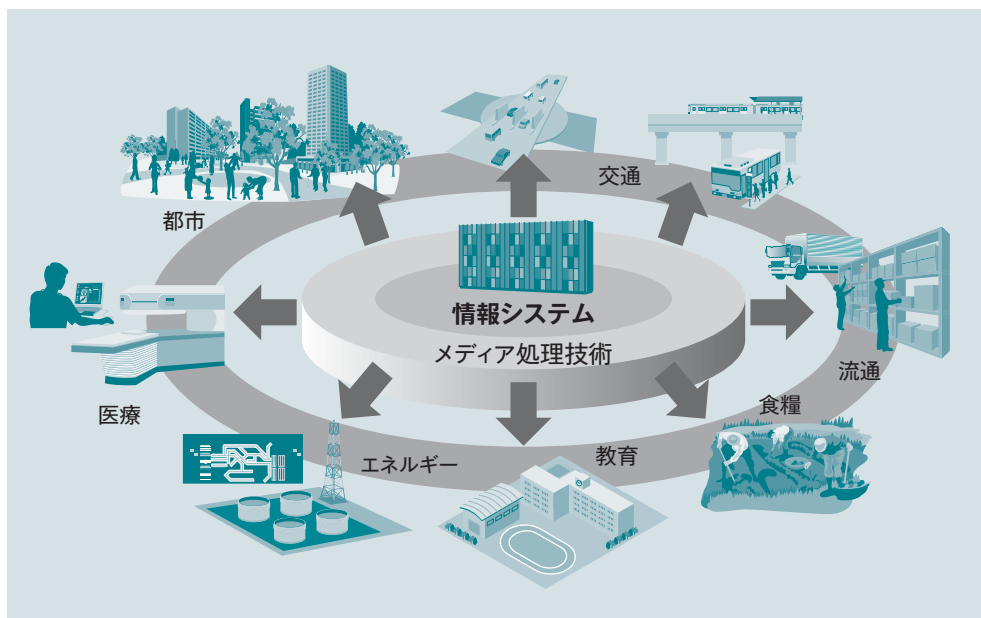


図7 | メディア処理の広がり

メディア処理は、われわれの社会生活の中に一層溶け込み、身近なものとなっていく。

に授けつつある。メディア処理は、これまで以上に複雑で大規模な課題にも対応できるようになる。

現在、メディア処理のフロントエンドとなり得る機器は実世界の至る所に広がり、ネットワークで相互に接続されて地域社会・地球社会を包み込む巨大システムへと成長しつつある。メディア処理は、われわれの社会生活の中に一層溶け込み、身近なものとなっていく（図7参照）。今後は、さらに社会との関係性を強め、環境、エネルギー、教育、医療といったより大きな課題の解決も視野に入れ、持続可能な社会づくりに貢献していくことが求められるようになる。

そのためには、解決していくべき課題に関わる社会の仕組みやそれを支えるインフラまで理解を深め、その中でメディア処理をうまく活用していく必要がある。メディア処理は必ずしも万能なわけではなく、認識精度や出力品質など制約がある。課題に対し、その何を解決したいのか、どのような条件下でどこまでの性能を達成すれば価値につながられるのかなど、大局的な視点で最適解を導いていくことが重要である。日立グループは、これまで蓄積したメディア処理に関する技術やノウハウに加え、長きにわたって社会インフラを支えてきた知見を生かした新たなイノベーションを創出していく。

参考文献

- 1) S. Mori, et al. : Historical Review of OCR Research and Development, Proc. IEEE, Vol. 80, No. 7, pp. 1029-1058 (1992.7)
- 2) B. H. Juang, et al. : Automatic Speech Recognition-A Brief History of the Technology, Elsevier Encyclopedia of Language and Linguistics, Second Edition, Elsevier (2005)
- 3) 江尻：視覚を中心とした新ロボット技術の展開，日立評論，57，10，819～824（1975.10）

執筆者紹介



長坂 晃朗

1991年日立製作所入社，中央研究所 情報システム研究センタ 所属
現在，生体認証技術の研究開発に従事
博士（工学）
電子情報通信学会会員