8

ディジタル機器向け オブジェクト認識技術の動向

林　大輔
富士フイルム（株）

オブジェクト認識技術概観

近年，デジタルカメラや携帯電話を始めとするディジタル機器に次々とオブジェクト認識技術が搭載され始めてきた．一言でオブジェクト認識技術といっても定義は曖昧で範囲は非常に広い．デジタルカメラに搭載されている代表的なオブジェクト認識技術は顔検出技術である．また，動体追尾のように動いている物体を認識して追尾する技術もオブジェクト認識といってもよい．

ディジタル機器にとってはオブジェクト認識を搭載することが最終目的ではない．ディジタル機器がオブジェクトを認識することによって，ユーザにどのような価値を提供することができるようになるのか？　それが最も重要なのである．

本稿では，始めにオブジェクト認識がディジタル機器に搭載されることによってもたらす効果に触れ，オブジェクト認識がディジタル機器に搭載され始めた技術的背景，今後の展望について紹介する．

使いやすさを提供するオブジェクト認識

オブジェクト認識技術がディジタル機器に搭載されて，以前と比較して何が変化したのか？　デジタルカメラを例に，オブジェクト認識がもたらした効果について考えてみる．

まずオブジェクト認識技術の代表例である顔検出技術を例に挙げる．2006 年以降，各社が顔検出技術をデジタルカメラに搭載するまでは，デジタルカメラは単純にフイルムカメラに置き換わるものという考え方が一般的であった．数年前まではデジタルカメラを開発しているメーカは画質や画素数を競いあっていた．ところが，2006 年に顔検出技術がデジタルカメラに製品搭載されたことで状況が一変した．顔というオブジェクトを認識することによって，撮影者は誰でも自動で上手に人物を撮影することができるようになったのである．これはフイルムカメラではできなかった，まさにデジタルカメラだからこそできる新しい価値を生み出した瞬間である．

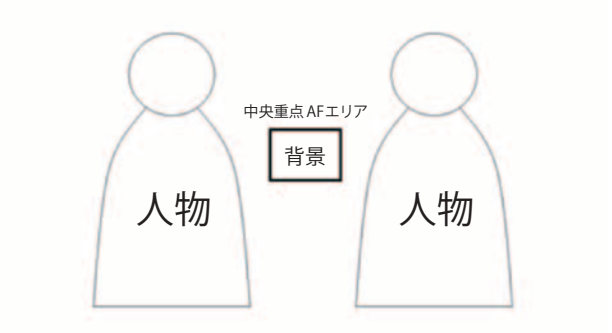

図 -1　中央重点 AF

1 つ，オートフォーカスを例に挙げて説明しよう．図 -1 に示すように 2 人の人物が並んでいたとする．この図の構図では撮影者は人物にピントを合わせたいと思っているはずだ．しかし，従来の中央重点のオートフォーカスだと背景にピントが合ってしまっていた．これは，デジタルカメラが主要被写体を認識できなかったためである．ここで，デジタルカメラが顔というオブジェクトを認識できるということは，主要被写体を理解できるということを意味する．デジタルカメラは従来から指定されたエリアにピ

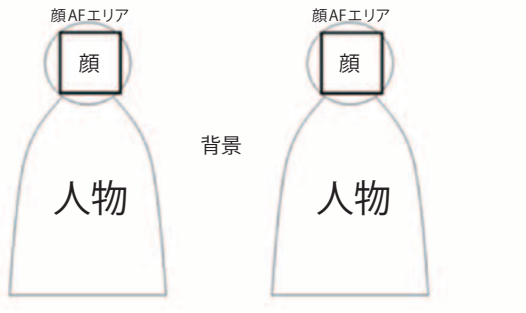

図 -2　顔重点 AF

ントを合わせにいく能力を持っていた（図 -2）ので，主要被写体の位置をそのエリアに設定すれば，主要被写体にピントを合わせることが可能となるという発想が自然と生まれる．顔という情報は人物撮影において非常に重要であり，オートフォーカスだけに利用されているものでない．露出やホワイトバランスも適切に設定してくれる．これはフイルムカメラの域を超えて，デジタルカメラが人間に近づいたことを意味する．人間は過去の記憶からたどって主要被写体を認識する能力を持っており，主要被写体に合わせて，焦点や明るさを調整することができる．この過去の経験を活かすことこそ，オブジェクト認識の「機械学習」にあたる能力である．人間は過去に出会った人物の顔と顔ではないオブジェクトを合わせて学習することによって，将来，目にするオブジェクトが人物の顔か顔ではないかを，100% に近い確率で判別することを可能としている．

人物の顔というオブジェクトは応用範囲が非常に広い．顔全体を検出するだけではなく，顔の中のパーツ（目，鼻，口など）を認識することで，表情を認識したり，目つぶりを認識したり，個人の識別，年齢，性別などの認識も行うことができる．また，赤目を認識して補正する画像補正技術もある．デジタルカメラではこれらの認識結果をさまざまなアプリケーションに応用している．

次に動体検出について紹介する．動いているオブジェクトを認識することもカメラの使いやすさを向上させる．デジタルカメラやデジタルビデオカメラにおける動体検出の主な目的は動体にフォーカスを合わせ続けることである．撮影者が撮影したい主要被写体が動体であったとしよう．もし，カメラが動体を認識できないとすると，近づいてきたり，遠ざかったりする被写体に撮影者が即座にピントを合わせるのは困難である．ところが，カメラが動体を検出し続けることで，撮影者は何もしなくても，自動でフォーカスをあわせ続けることが可能となるのである．

顔検出と動体検出だけを見ても分かる通り，オブジェクト認識技術はデジタルカメラの使いやすさを大幅に向上させた．ただし，1 つ注意しなければいけないことは，顔や動体といったオブジェクト認識は写真の撮影に詳しい人にとっては必ずしも必要な機能ではないということである．図 -1 の被写体もフォーカスロックを知っていれば，人物にフォーカスを合わせることが可能であるし，露出やホワイトバランスも同様である．現在，デジタルカメラに搭載されているオブジェクト認識技術は，誰でも手軽に上手な写真が撮れるという価値を提供しているといってもよいだろう．もし，オブジェクトを認識することによって，撮影者のテクニックや努力では実現不可能なことが実現できるようになるアプリケーションがあったとすれば，それはこれまで以上の価値を提供するアプリケーションであると言える．次世代のオブジェクト認識にはそのような価値が求められているのかもしれない．

たとえば，富士フイルムでは 2006 年に顔検出専用の LSI を開発し，これまでになかったハードウェアによる高速高精度顔検出をデジタルカメラで実現した．それ以降，顔検出 AF（自動フォーカス調整），AE（自動露出調整），AWB（自動ホワイトバランス調整），自動赤目補正，個人認識など顔というオブジェクト認識を通じて，簡単に人物撮影できる機能を提供してきた．また，ペット（犬，猫）検出など新しいオブジェクト認識への広がりも見せており，ペットを認識した瞬間にカメラが自動でシャッターを切ることによって，瞬間のシャッターチャンスを逃さない価値を提供するなどのユニークなアプリケーションも搭載されている（図 -3）．

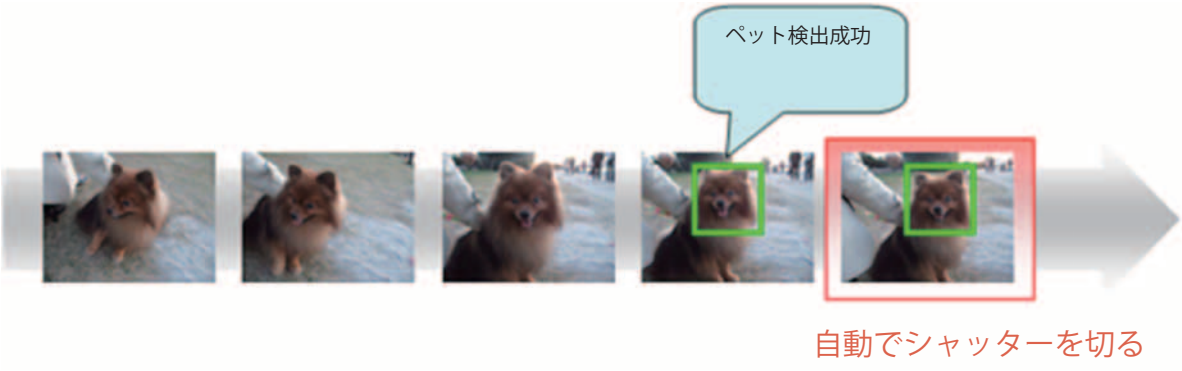

図 -3　ペット検出応用オートシャッター

オブジェクト認識が近年ディジタル機器に搭載され始めた技術的背景

デジタルカメラやデジタルビデオカメラをはじめとするディジタル機器にオブジェクト認識が搭載され始めた技術的背景について，以下 4 つの要因を紹介する．

① アルゴリズムの高速化，シュリンク
② ディジタル機器の能力向上（CPU 性能向上，並列処理化，メモリの大容量化／低価格化，高速撮像）
③ ワークステーションなどローカルマシンの能力向上
④ ネットワークの高速化などの外部インフラ整備

♦ アルゴリズムの高速化，シュリンク

デジタルカメラなどの機器にオブジェクト認識を搭載するためには，アルゴリズムの高速化とアルゴリズムのシュリンクが必須となる．シュリンクとは処理の簡略化やデータのビット精度の削減である．たとえば，画像処理で 15 × 15 のフィルタをかける処理を 9 × 9 のフィルタに変更したり，32 ビット整数精度のデータを 16 ビット整数精度のデータに落としたりすることである．

顔検出を例に挙げる．2004 年に Viola と Jones は Haar-Like 特徴量（図 -4）の利用と複数の弱識別器を複数連結（Cascade）することで，高速に顔を検出する手法を提案した [1]．この手法では，顔を図 -4 に示したような明暗の集合体であると考える．入力画像に対して図 -4 の明暗パッチを順次適用していき，それぞれのパッチごとに顔らしさの評価値を算出する．従来はそれらの評価値をすべて連結加算して最終評価値を算出し，閾値を設けて顔か非顔の判定を行っていた．しかし，Viola と Jones は評価値を加算していく過程でそれぞれの閾値を設け，途中段階で評価値の低いものは非顔と判定する処理の高速化手法を提案した（図 -5）．この手法は画期的で，デジタルカメラなどのモバイル機器に実装できる可能性を感じさせた．このように実現性のある画期的な処理高速化手法が提案されると，各メーカが急速に研究，開発を進め，数年後には製品化される可能性が非常に高くなる．

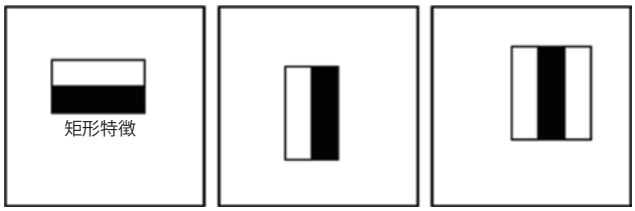

図 -4　Haar-Like 特徴量

また，アルゴリズムのシュリンクもオブジェクト認識のディジタル機器搭載には欠かせない．これは，処理高速化のためだけではなく，メモリ節約（= コスト削減）の問題も含む．具体的には，表 -1 の施策がとられるのが一般的であるが，それぞれメリット，デメリットがあることを念頭において，開発が進められる．

当然，アルゴリズムのシュリンクにはオブジェクト認識の精度低下を伴う．しかし幸いにも，現在のデジタルカメラやデジタルビデオカメラの撮像素子で CMOS が一般化しはじめ，高速に連続画像を取

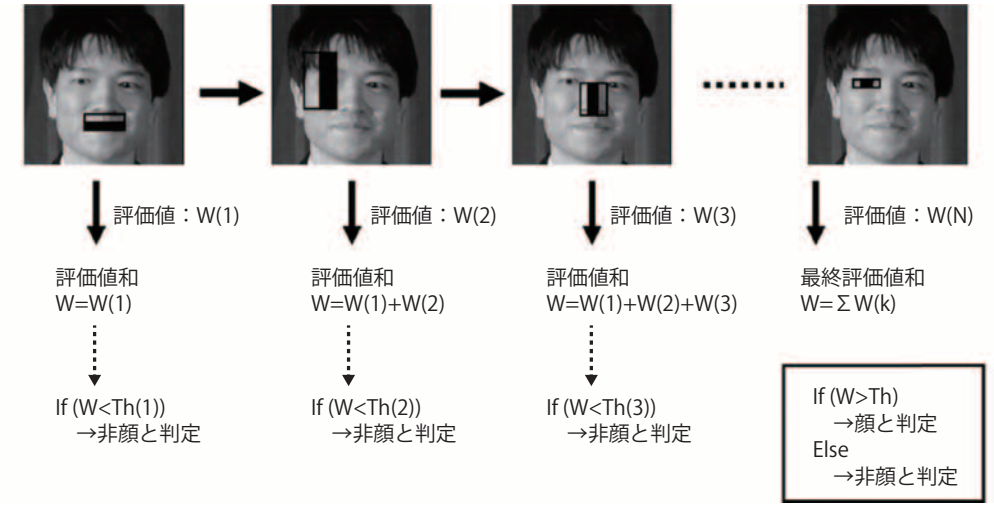

図 -5　Haar-Like 特徴量を Cascade 接続した顔検出例

得できるようになってきた．その連続撮像による複数枚の時系列画像でオブジェクト認識することによって，全体的な精度低下を抑制する手法が用いられることもある．図 -6 では連続フレームの顔検出の例を示す．第 1 フレーム，第 2 フレームで顔検出に成功し，第 3 フレームで顔検出に失敗したとする．デジタルカメラではノイズや画角の微小変動の影響で，時間間隔の短い連続フレームであったとしても顔検出に失敗することが多々ある．この場合，人物の動きや画角変動が微小であると仮定し，第 1 フレームや第 2 フレームの顔検出結果を利用して，顔検出に失敗した第 3 フレームを補間する処理が用いられる．これは動画像に対して，デジタル的にフレームレートを上げるためのフレーム補間処理や動画の超解像技術に近い考え方である．

♦ ディジタル機器の能力向上

ディジタル機器の能力向上も要因として挙げられる．以下 3 点について紹介する．

① CPU の性能向上
② RAM，ROM の大容量化
③ 撮像技術，画像処理技術の進化

近年，組込み機器向けの CPU の能力も着実に向上してきている．それと同時に組込み機器の世界でもデュアル CPU による並列処理が行われるようになってきた．プロセスも 45 ナノが普及し始め，32 ナノプロセスも登場しようとしている．

また，RAM，ROM ともに大容量化，低価格化してきており，大容量の参照データを必要とするオブジェクト認識には追い風となっている．

撮像技術，画像処理技術の進化もオブジェクト認識には必要である．前述した CMOS などの高速撮像による時系列データの利用や，画像処理技術の進歩による低ノイズ，広ダイナミックレンジの画像がオブジェクト認識の入力画像として利用できるようになったことは，オブジェクト認識技術の性能を向上させた．

施策	メリット	デメリット
①整数演算化	演算高速化	演算精度低下
②固定長化	演算高速化	メモリの増大
③ビット精度削減	メモリ削減（＝コスト削減）	演算精度低下

表 -1　ディジタル機器の能力向上施策のメリット／デメリット

♦ ワークステーションなどオフライン学習マシンの能力向上

オブジェクト認識技術の実用化のためには，搭載されるディジタル機器の能力が向上するだけでは不十分である．現在，ディジタル機器に搭載されているオブジェクト認識技術の大多数はオフラインで学習した学習データを機器の不揮発性メモリに格納しており，それを参照することで認識を行ってい

る．このオフライン学習はワークステーションなどのマシンで実施するわけだが，現在の最新スペックのマシンでさえ，学習に数十時間から数日を要することも少なくない．これが 10 年前のマシンなら数週間の学習時間を要しても不思議ではないだろう．実は，この学習時間が直接ディジタル機器の商品化時期に大きな影響を与えるのである．オフライン学習では学習アルゴリズムの開発だけではなく，適切な教師データを入力することが重要なポイントとなってくる．これはトライアンドエラーの要素を多く含む．学習結果データを分析して，教師データを調整して再学習することを繰り返す必要があるからである（図 -7）．

学習の失敗例を 1 つ挙げる．人物を認識する学習をしたとする．人物の学習用画像と人物ではない学習用画像を収集し，教師データとして与えて学習を実施した．同様に収集した検証用の人物画像で検証すると TP[☆1] は 100% 近い数値を示した．しかし，この結果を検証してみると，学習によって生成された参照データに，偏った特徴があることが分かった．それは人物の足元から離れた位置の特徴量を必ずとっているという特徴である．そこで改めて教師用の画像を見てみると，学習に用いた人物の教師データには必ず人物の影が写っていたことに気づく．この学習は，「影のついた」人物に特化した学習結果データが得られたという失敗例である．もちろん，この参照データでは影のない人物の認識はほとんどできない（図 -8）．

このように，学習はトライアンドエラーの繰り返しの側面も大きいため，学習時間の短縮は製品開発において，非常に重要なファクタとなることがお分かりいただけるだろう．

近年，CPU 単体のクロック向上競争は停滞し，マルチ CPU 化よる並列処理が進んでいる．そのため，当然ながら学習プログラムも並列処理を考慮してプログラミングされるのが望ましい．また，以前はグラフィックボードに搭載されている GPU は画像処理専門の演算に用いられていたが，近年，GPU を一般数値演算用に利用する GPGPU（General Purpose GPU）が大きな潮流となっている．GPU の性能は著しい向上を続けているが，GPU は単純なデータ演算を大量に処理することを得意としており，条件分岐などが苦手なため，完全に CPU に取って代わるものになるとは言えない．しかし，機械学習のように，同じ計算を繰り返し行うアプリケーションにおいては大きな力を発揮する．DirectX9.0 の登場以降，NVIDIA は統合開発環境「CUDA」を，AMD は「ATI Stream」を発表

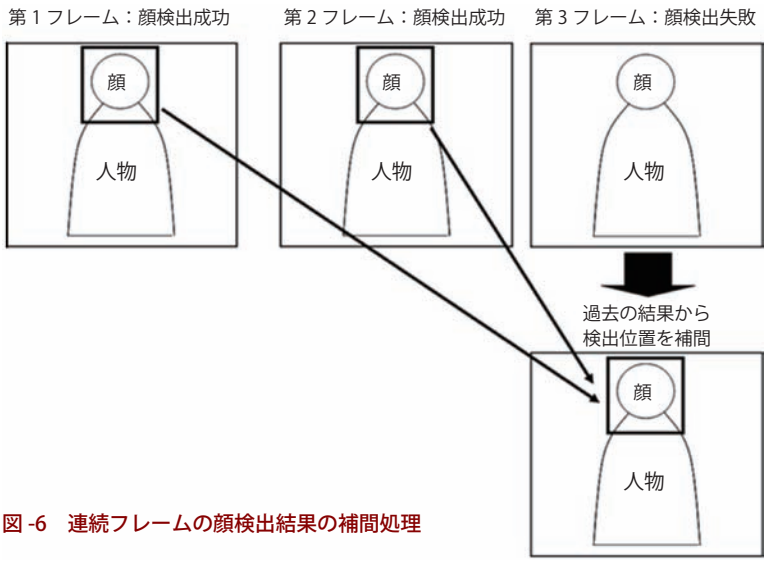

図 -6　連続フレームの顔検出結果の補間処理

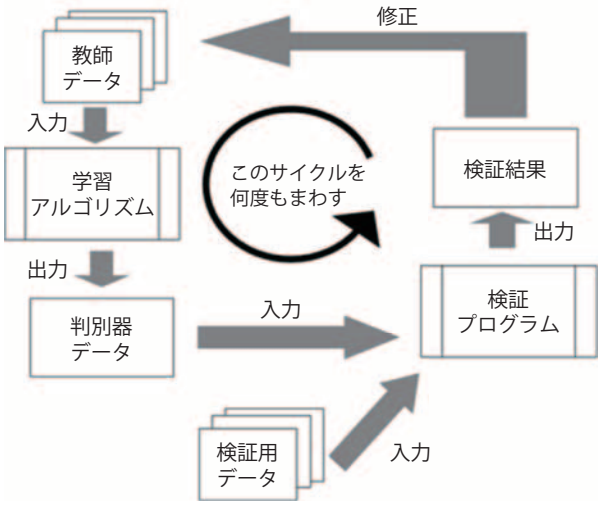

図 -7　機械学習のサイクル

☆1　TP：True Positive. 正解を正解として判定した割合（≒検出成功率）．

図 -8　人物認識失敗例

し，それに対応するグラフィックチップも次々と発売されている．今後，学習時間短縮の鍵を握るのは，CPU，GPU での並列処理をいかに実現していくかである．

♦ ネットワークの高速化などの外部インフラ整備

近年，光回線の整備など，ネットワーク回線の高速化により，世界中のコンピュータの間でデータのやりとりが短時間で可能となった．それに伴い，クラウドコンピューティングサービスが次々と生まれてきている．これまでオフライン学習はローカルの単体，または数台のコンピュータで実施されることが多かったが，今後はクラウドコンピューティングの登場によって，ネットワーク上の数百，数千のコンピュータで学習を行うことができるようになる可能性が高い．そして，LTE（Long Term Evolution）などの高速無線ネットワークの構築により，クラウド上のコンピュータとオブジェクト認識を搭載するディジタル機器とが当然のように接続される時代がくるはずだ．それによって，新たな価値を生み出すことになるだろう．この点に関しては将来のインフラ整備に期待したいところである．

オブジェクト認識技術の今後の展開

本稿ではデジタルカメラやデジタルビデオカメラなど，小型のモバイルディジタル機器に搭載されているオブジェクト認識について説明してきた．それ以外にも，大型小型問わず，さまざまなディジタル機器にオブジェクト認識は応用されている．たとえば，不審人物を認識して防犯に役立てるために監視カメラに応用されていたり，ナンバプレートや標識を認識して，ドライバの運転をサポートするために車載カメラのシステムとして応用されていたり，人物の年齢層や性別を認識してマーケティングに役立てるために，ショッピングセンタや自動販売機に応用されていたりする．知らない間に我々の身の回りには多くのオブジェクト認識技術が広まってきているのである．

今後，オブジェクト認識技術はますますディジタル機器に搭載されていくであろう．これまではオフライン学習によって作成された参照データを用いたオブジェクト認識が主流であったが，統計的学習手法として代表的な Boosting（従来，Boosting はオフライン学習であった）をオンラインでの物体追跡に拡張した Online Boosting などのオンライン学習手法も次々と登場してきている．オンライン学習はディジタル機器の楽しみ方を変えるだろう．従来，ディジタル機器を提供するメーカが設計した機能，性能の範囲内でユーザは楽しむしかなかった．しかし，ディジタル機器にオンライン学習が搭載されることで，ユーザが機能をカスタマイズできるようになる．そして，高速無線ネットワークによって，ユーザ同士で共有することも可能となる．それらの技術が発展していくための条件として，高速無線ネットワークのインフラ整備には注目しておきたい．

参考文献

1）Viola, P. and Jones, M. : Robust Real-Time Face Detection, International Journal of Computer Vision, Vol.57, No.2, pp.137-154 (2004).

（平成 22 年 9 月 29 日受付）

林　大輔　daisuke_hayashi@fujifilm.co.jp
2004 年京都大学大学院工学研究科機械物理工学専攻修士課程修了．同年，富士写真フイルム（株）入社．現在，富士フィルム（株）電子映像商品開発センター所属．画像認識を始めとするデジタルカメラの開発・設計に従事．