解説

高速ビジョンのアーキテクチャと新展開

Architecture of High-speed Vision and Its New Evolution

石川 正 俊* *東京大学

Masatoshi Ishikawa* *The University of Tokyo

1. はじめに

2017年は、高速ビジョンの社会実装にとって、大きな飛躍の年になると考えられる。従来、実験室レベルでPOC (Proof of Concept) が提案され、先進的ないくつかの企業によりその実用化が進められていたが、ここにきて実用に向けた要素技術が具体的なハードウェアとして手に届く範囲で具体的な形が見えてきたからである。

本特集でも解説される1,000 [fps]の高速画像処理を実現する積層型のビジョンチップの実用化は、長年の夢の大きな第一歩であり、画像処理の今後の革新的進歩に寄与することは間違いない[1]。また、1,000 [fps]の高速プロジェクタも2016年に実用化され、高速画像処理と組み合わせることにより、ダイナミックプロジェクションマッピングや三次元形状計測に飛躍的進歩をもたらしている[2]。さらに、アクチュエータも高速化も進んでいることから、機械の目、それも高速の目を持った高速知能システムの実用化が現実味を帯びてきた。

本論文では、高速ビジョンや知能システムの基本構造を整理し、そのハードウェアの進歩を概観するとともに、応用システムの基本的なアーキテクチャならびに展開の可能性を議論する。

2. 高速ビジョンの基本構造

2.1 高速画像処理と高速ビジョン

筆者らは、従来から、高速画像処理と高速ビジョンを使い分けてきた。高速画像処理は、高速の撮像（できればグローバルシャッターを用いた撮像）とその画像の高速処理を意味し、1,000 [fps]を目標とする高フレームレートと画像の入力から処理結果の出力までの低遅延処理を実現することが目標となる。画像処理が低速（例えば30 [fps]）だった時代には、ほかのシステム要素は画像処理に比較して十分高速だったため、画像処理を高速化すれば、システムも

図1 高速ビジョンの構造

高速化が達成できていた。ところが、高速画像処理がほかの要素技術の帯域を超える1,000 [fps]程度を想定すると、高速画像処理だけを高速化してもシステム全体の高速化は達成できない。加えて、高速画像処理が十分な高速性を実現することにより、対象の動きやアクチュエータの動きを完全に把握、制御できる可能性が見えてきた。

これらのことから、高速画像処理を用いるシステムが生み出す新しい機能は、高速画像処理だけではなく、処理系や運動系との組み合わせによるシステムとしての機能を議論する必要が出てきた。このことから、高速画像処理を基本とし、処理系や運動系も含めたシステム全体の機能を論じるときには高速ビジョンとよんでいる。高速ビジョンは、システムのダイナミクスの把握と制御を高速画像処理を用いて実現する知能システムを意味している。この様子を図1に示す。

歴史的に見ても、1993年に提案されたSPE-4kは、高速画像処理の開発の第一歩であったが[3]、ほどなくして、アクティブビジョンという形で、高速ビジョンの開発へと移行してきた[4]。この流れは現在に引き継がれており、高速ビジョン、すなわち高速画像処理をベースに新しいシステム機能を創出することが大きな意味を持ち、社会の新しい価値を産み出す社会実装が大きな課題となっている。

2.2 高速画像処理は画像処理の高速化ではない

一般に、サンプリング定理が示していることは、対象やシステムのダイナミクスを把握・制御するには、対象の帯域をカバーするのに十分なサンプリングレートが必要であり、逆に十分なサンプリングレートを用いることにより、対象のダイナミクスを把握することが可能となる。不十分なサン

原稿受付 2017年8月29日

キーワード: CMOS Imager, Image Processing, Visual Feedback, Dynamic Compensation

*〒113-8656 文京区本郷 7-3-1

*Bunkyo-ku, Tokyo

プリングでは、ダイナミクスに対する情報の欠落を生じ、事前の知識やデータを用いたモデル、学習、推測・補間等を導入して、欠落している情報を補う必要があり、欠落している情報が補填されない限り完全な計測や制御は実現できない。

高速画像処理は、このような帯域を確保することを前提として、高い時間密度の中で十分な情報を得て画像処理を行うものであるから、不十分な帯域で長い処理時間をかけて画像処理を行おうとする画像処理とは、基本的な考え方、アルゴリズム、ハードウェア、システムとの関係等の観点で、まったく違うものと考えるべきである。別の言い方をすれば、通常の画像処理をそのまま高速化しても、高速性を利用したことにはならないし、高速性を維持することもできず、ハードウェアでの実装に大きな問題を残すことになる。

2.3 ダイナミクス整合

筆者らは、帯域をカバーする高速画像処理の実現のための設計思想として、システムのすべての要素を対象のダイナミクスをカバーするように設計することをダイナミクス整合とよんでいる [5]。対象あるいはシステムのダイナミクスを把握し、そのうえで必要な帯域を設定し、センサ、アクチュエータ、処理系等、すべての要素の設計仕様をその帯域に設定することにより、高速ビジョン特有のシステムが実現できる。このことは、従来のビデオレートを基本としたシステムが人間の目の実現にあったことに対して、機械の眼を用いて機械の限界でシステムを動かすことであり、人間をはるかに超えるスピードの知能システムの実現を目指すものである。

2.4 低遅延処理

一般的に画像処理は、画像の撮像、処理、出力の間に、各モジュールでの入力および出力バッファ等も含むパイプライン処理となる。そのため、ビジュアルフィードバックでは、画像入力から出力までの遅延は、各モジュールでの遅延の積算となり、処理だけを高速化しても他の要素の遅延も削減しない限り、低遅延のビジュアルフィードバックは実現できない。高速画像処理、各モジュールでのバッファリング・同期構造等を低遅延のもので実現する必要がある。

2.5 Dynamic Compensation

プレイバック制御をはじめとして、従来のロボットは、繰り返し精度に関しては十分な性能が実現されていたが、ギアのバックラッシュ等に起因して絶対位置精度は高くなかった。そのため、対象とロボットの双方を絶対位置で計測・制御して、対象への操作を高精度で実現することは困難であった。そこで、従来のプレイバック制御を含む絶対位置制御によるロボットシステムの手先部分に、新たに高速のビジュアルフィードバックモジュールを導入し、画像内相対座標を用いた視覚フィードバック制御を高精度で実現する方法が提案されている [6]。この方法では、対象とロボットの相対位置を制御することになるため、ロボットの位置

誤差の補償や動的応答の不足はもちろんのこと、剛体の仮定を含むロボットや環境の不確定性を補償することが可能となる。

これにより、従来のロボットの性能の向上、特に高速性と高精度の両立 (ms でサブ mm オーダ) が実現でき、従来必要であった撮像のためにロボットを停止されること、精緻なカメラ校正、長時間の教示時間等は必要なくなり、環境や対象の変化への対応、柔軟対象物への対応、ロボット側の剛体性の不足への対応、ダイナミクスのリアルタイム同定等において、大きな改善が望めることとなる。教示やライン設計も含むコストパフォーマンスの向上、教示方法の革新が、高速ビジョンにより、実現できる。

また、このことは、人間ロボット協調作業における人間の作業精度の低さをも補償することが可能となる。

3. 高速画像処理のアーキテクチャ

3.1 ハードウェアの構造

高速画像処理のハードウェアは、古くから人間の目の構造を模した知能化が目標とされてきた。1980 年代後半には、アメリカを中心に初期視覚とよばれる単純な演算機能をアナログ回路で作成する研究が進んだが、実用化されることはなかった [7]。その理由の一つが汎用性・プログラマビリティの欠如であって、筆者らはこの時代からデジタル処理を導入した汎用・プログラマブルビジョンチップの実現を提唱し、実際にスケールアップモデルとよんでいる SPE-4k を製作している [3]。その後の IT の進展を見れば、デジタル化ならびにプログラマビリティの確保は、必須の課題である。

高速画像処理のアーキテクチャとして、必須の要素は、撮像素子、A/D 変換、画像処理であり、これらの要素間の画素構造との関係でいくつかのタイプに分類することができる。この分類と特徴を表 1 に示す。

第 1 のアーキテクチャは完全並列タイプであり、一つの画素の中に撮像素子、A/D 変換、プロセッシングエレメント (PE) を実現するものである [3]。この場合は画素数と同数の PE による並列処理を実現することになる。高速性

表 1 高速画像処理アーキテクチャ

分類	アーキテクチャ	特徴
完全並列処理	一つの画素内に、撮像素子、A/D 変換、PE を配置。	高速性、撮像・処理の同期性に優れるが、A/D 変換の精度、集積化に難点。
列並列 (列並列 PE)	一行に対して、走査回路、A/D 変換、PE を配置。	集積化、A/D 変換の精度に優れるが、同期性に難点。
列並列 (完全並列 PE)	一行に対して、走査回路と A/D 変換を配置。転送後に画素数分の PE を配置。	A/D 変換の精度、処理の同期性に優れるが、集積化に難点。
部分並列	列並列の転送線数を標準インターフェイス等の活用で削減。転送後に並列 PE、GPU、CPU 等で処理。	A/D 変換の精度に優れるが、高速性に難点。GPU、CPU での高速性は、アルゴリズム依存性が強い。

PE: Processing Element

や処理の同期性等の利点があるが、A/D 変換の精度を確保することが難しい。

第2のアーキテクチャは列並列タイプであり、行方向にはスキャンするが、列方向は並列化するものである。A/D 変換は列の数だけ必要であるが、PE の数を列の数とする場合は行方向のスキャンと同期して転送・処理を実行することとなり [1], PE を画素数分揃えた並列処理を用いる場合は転送のみ列並列となり、処理は画像全体で同期処理となる [8]。両者とも A/D 変換の精度は確保しやすく、高速性は少し落ちるものの実用的には十分な速度を実現できるが、処理の同期性に関しては注意が必要となる。

第3のアーキテクチャは部分並列タイプであり、列の数の並列転送ではなく、マルチプレキシングして少数の転送線の高速転送を導入するタイプで、標準の高速インタフェイス、例えばカメラリンクのようなものを活用するものである。標準の高速インタフェイスの速度が不十分な場合もあって、実現する画素数やフレームレートによっては限界が存在する。

一般に画像処理のアーキテクチャはアルゴリズム依存性が高いので、アルゴリズムによっては汎用画像処理チップや GPU の活用が可能な場合もあり、ときには CPU のみでも高速の画像処理が可能な場合も存在する。このことは、多くの場合、処理を必要とする対象が全画面でなく、一部の領域に限られる場合に有効となる。

3.2 ハードウェアの新展開

2017 年に、高感度・高速の CMOS イメージャーと列並列型の A/D 変換器と PE のチップを積層型のチップとして製作したものが開発された [1]。具体的にこのチップには、1,304 個の 4 [bit] PE と同数の 1 [bit] PE、さらには特徴量抽出プロセッサが搭載されていて、140GOPS の性能を 0.4 [W] を切る消費電力で実現している。今まで、ボードレベルでしか実現されていなかった 1,000 [fps] の高速画像処理が、低消費電力の積層化チップで実現されている。

高速ビジョンの高速化という観点からは、高速画像処理を速くしただけでは不十分な場合も多い。例えば、構造照明を用いた三次元計測やダイナミックプロジェクションマッピング等では、ディスプレイも高速化する必要があり、その際にも画素構造や転送方式は全体のパフォーマンスに大きく影響する。

また、ビジョンネットワークを実現する場合にも、アルゴリズムによっては、ビジョンモジュール間のデータ転送のアーキテクチャ、特に画素構造を考慮したうえでのデータの同期性の確保が重要となる。

4. 知能システムの基本構造

4.1 知能の3要素

知能システムがシステムとして意味のある動作を実現す

るためには、三つの主要な要素が存在する。計算理論、情報表現とアルゴリズム、ハードウェアの三つである。計算理論は、知能を実現する計算理論であり、ときに並列処理を前提とした計算理論である。情報表現とアルゴリズムは、計算理論を実際の処理として記述するための内部モデルやその情報表現、それを用いたアルゴリズムである。ハードウェアは、それらを実現する実際のハードウェアであり、処理系としてのプロセッシングユニットはもちろんのこと、感覚系としてのセンサや運動系としてのアクチュエータやその処理構造が実現される。

一般的に、知能システムは階層的並列分散構造を取ることが多く、この構造を用いて感覚運動統合を実現することになる。その際、高速性を実現するためのタスク分解として、従来の直列分解から並列分解の導入が重要であり、感覚系と運動系の相互作用、例えばアクティブビジョンの実現には並列分解の構造の上に高速のビジョンと高速アクチュエータを実現する必要がある。

4.2 高速化アルゴリズムと知能の新しい形

時間密度の高い撮像が実現されれば、像面における対象の変化がわずかであることが仮定できる。もし仮定できないのであれば、フレームレートを上げることで、この仮定が実現するように設定することができる。像面における変化がわずかであれば、変化を認識する処理は簡素化されるので、高速化はアルゴリズムを簡素化することとなり、その結果、簡素化されたアルゴリズムで処理の高速性が引き出されることになる。

このことは、システムの時間密度（時間稠密性）と空間密度（空間稠密性）が、要求仕様や人間の感覚系、できれば支配する物理現象を完全にカバーし、システムの知能としての認識行動のすべてが実世界、実時間、遅延なしで動く知能システムの実現を目指すことになる。Brooks が言う「象はチェスをしない」という例えが意味する二つの知能のうち、象の知能を高速に実現することを目指すものである [9]。

5. 応用システムのアーキテクチャの新展開

5.1 応用システムの基本構造

応用システムの観点から高速ビジョンを考えると、高速ビジョンと対象との関係からいくつかの構造に分類できる。

高速対象を光軸を固定したカメラで捕らえる場合、高速ビジョンのフレームレートは設定したレンズ系で結像される対象の像面での動きに対して決定することになる。これに対して、アクティブビジョンとして、対象のトラッキングを実現する場合は、像面における対象の動き、特に位置の変動に対してトラッキング装置のアクチュエータの帯域とカメラのフレームレートを決定することになる。そのうえで、トラッキングした対象に関する処理、特にトラッキン

図2 アクティブビジョンの構造

グがロックオンした後の処理では、対象の位置の変動がキャンセルされているため、対象自体に対する処理のフレームレートは、対象の変形や回転等から決定されることになる。

このことは、構造照明やプロジェクションマッピングといった投影系を用いる場合にも同様の設計手順となる。つまり、固定視点のプロジェクタを用いる場合には対象に対する投影像の関係からフレームレートを決定することになり、トラッキングを使う場合には、ロックオンした後の同軸の光学系を用いれば、対象の変形に応じたフレームレートを決定することとなる。対象が剛体で回転がなければ、低速のプロジェクタを利用できるが、対象の変化、回転、振動等並進運動以外の動きがある場合は、それらの帯域をカバーするプロジェクタを用いる必要がある。この様子を図2に示す。

5.2 応用システムの基本アーキテクチャ

応用システムにおける高速ビジョンの構造は、光学系によって生成される像面の速度と帯域に対して、固定光軸カメラによるトラッキング撮像に依存して、そのフレームレートが決定される。固定光軸でも、並進運動領域を十分にカバーできる場合には、高速画像処理の導入で高速低遅延の知能システムが構築できる。例えば、固定視野内の対象またはロボットの運動の制御、着席でのジェスチャー認識、並進運動要素が小さいダイナミックプロジェクションマッピング、固定視野の高速検査等においては高速画像処理の導入効果が大きい。

これに対して、並進運動領域が大きく、必要とする画角に対して、対象物が小さくなってしまいう場合には、対象の分解能が不足するため、ビジュアルフィードバックを導入したトラッキング撮像を導入することにより、並進運動をキャンセルし、視野を狭くすることにより対象の分解能を高くできる。例えば、移動量の大きい対象の運動の制御、並進運動要素が大きい運動解析やダイナミックプロジェクションマッピング、対象生物の移動量が大きいマイクロビジュアルフィードバック、手先の高速ビジュアルフィードバックを用いたダイナミック補償、高速道路検査等で導入効果が大きい。

6. おわりに

高速画像処理と高速ビジョンの今後の展望を主としてアーキテクチャやハードウェアの側面から展望した。高速性、低遅延、低消費電力、高いコストパフォーマンスの点で、高速画像処理チップの実用化は、従来の方法に比べて、格段の性能の向上が実現できる。この性能を活かすことにより、新しい応用システムのアーキテクチャが生まれ、知能システムとしての高速性が実現できる新しい世界が目前に迫っており、ロボット制御は大幅な進展・展開が期待できる。

参考文献

- [1] T. Yamazaki, H. Katayama, S. Uehara, A. Nose, M. Kobayashi, S. Shida, M. Odahara, K. Takamiya, Y. Hisamatsu, S. Matsumoto, L. Miyashita, Y. Watanabe, T. Izawa, Y. Muramatsu and M. Ishikawa: "A 1 ms High-Speed Vision Chip with 3D-Stacked 140GOPS Column Parallel PEs for Spatio-Temporal Image Processing," Proc. Int. Solid-State Circuits Conf., pp.82-83, 2017.
- [2] Y. Watanabe, G. Narita, S. Tatsuno, T. Yuasa, K. Sumino and M. Ishikawa: "High-speed 8-bit Image Projector at 1,000 fps with 3ms Delay," Proc. Int. Display Workshops, pp.1064-1065, 2015.
- [3] M. Ishikawa, A. Morita and N. Takayanagi: "High Speed Vision System Using Massively Parallel Processing," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.373-377, 1992.
- [4] I. Ishii, Y. Nakabo and M. Ishikawa: "Target Tracking Algorithm for 1 ms Visual Feedback System Using Massively Parallel Processing," Proc. IEEE Int. Conf. Robotics and Automation, pp.2309-2314, 1996.
- [5] M. Ishikawa, T. Komuro, A. Namiki and I. Ishii: "1 ms Sensory-Motor Fusion System," Robotics Research (J.M.Hollerbach and D.E. Koditschek eds.), pp.359-364, Springer, 2000.
- [6] S. Huang, Y. Yamakawa, T. Senoo and M. Ishikawa: "Dynamic Compensation by Fusing a High-speed Actuator and High-speed Visual Feedback with Its Application to Fast Peg-and-hole Alignment," Advanced Robotics, vol.28, no.9, pp.613-624, 2014.
- [7] Carver Mead: Analog VLSI and Neural Systems. Addison-Wesley, 1989.
- [8] Y. Nakabo, M. Ishikawa, H. Toyoda and S. Mizuno: "1 ms Column Parallel Vision System and Its Application of High Speed Target Tracking," Proc. IEEE Int. Conf. Robotics and Automation, pp.650-655, 2000.
- [9] R.A. Brooks: "Elephants Don't Play Chess," Robotics and Autonomous Systems, vol.6, pp.3-15, 1990.

石川正俊 (Masatoshi Ishikawa)

1979年東京大学工学系研究科計数工学修士課程修了。同年通商産業省工業技術院製品科学研究所入所。平成元年東京大学工学部計数工学科助教授、1999年同大学工学系研究科計数工学専攻教授、現在同大学情報理工学系研究科研究科長および創造情報学専攻教授。この間平成2002年度東京大学総長特任補佐、2004年東京大学副学長、2005年東京大学理事・副学長（2006年まで）を歴任。センサフュージョン、超高速ロボット、超並列・超高速ビジョン、ダイナミックイメージコントロール、メタパーセプション、情報処理等の研究に従事。工学博士。2011年紫綬褒章受章。

（日本ロボット学会正会員・フェロー）