

# FPGAアクセラレーターによる ドメイン指向コンピューティング

## Domain Specific Computing Using FPGA Accelerator

● 渡部康弘      ● 藤澤久典      ● 小沢年弘

---

### あらまし

ドメイン指向コンピューティングとは、適用するアプリケーション領域を絞ることでサーバの桁違いな性能向上を図るアプローチである。FPGA(Field-Programmable Gate Array)は、アプリケーションに応じて演算器やメモリを適切に構成できるドメイン指向コンピューティングの実現に有用な汎用デバイスの一つである。今回、FPGAを活用した高速化技術により、画像検索処理に特化して高速化を実現した「メディア処理向けドメイン指向サーバ」を開発した。更に、ドメインに特化した高性能なアクセラレーターをFPGA上に効率的に設計するための技術として、モジュール間のデータ依存関係や性能ボトルネックを可視化し、アーキテクチャーレベルでの構造設計を容易にするFPGA設計支援環境を開発した。

本稿では、メディア処理向けドメイン指向サーバを実現するFPGAによる部分画像検索アクセラレーターと高速文書画像検索システムへの適用、および高性能なFPGAアクセラレーターの設計を可能にする設計支援環境について述べる。

### Abstract

Domain-specific computing is one approach to greatly improving server performance by specially designing server architecture for a particular application domain. A good way to implement such an approach is to use a field-programmable gate array (FPGA), a commodity device in which processing units and memory blocks can be configured depending on the application requirements. We have developed a domain-specific server for media processing that is specialized for high-speed image retrieval by using an FPGA accelerator. We also developed a design support environment that facilitates FPGA architecture design by enabling the data dependency between modules and a performance bottleneck to be visualized. This technology supports efficient design of high-performance domain-specific FPGA accelerators. In this paper, we describe a partial image retrieval accelerator that serves as a key component of a domain-specific server for media processing and its application to a high-speed image-based document search system. We also describe our design support environment that enables design of high-performance FPGA accelerators.

## まえがき

近年、社会で生成されるデータが爆発的に増加している。今後も、社会インフラのICT化・ネットワーク化や、スマートフォンなどの端末の進化、セキュリティカメラやIoT (Internet of Things) の発展などにより、画像やセンサーデータなどのデータがますます増加していくことは疑う余地がない。また、それらを賢く処理することで、データの中に埋もれている価値を引き出すことが、ビジネス的にも社会発展のためにも重要になっている。そのため、画像や音声などを処理するメディア処理や機械学習をはじめとする多くのアプリケーションにおいて、これまでとは桁違いの処理量と処理スピードが必要とされている。一方、コンピュータの性能向上はここ数年明白に鈍化しており、いわゆるムーアの法則に従った半導体の高性能化によるコンピュータの性能向上は限界に近づいている。

このように、大量データの高度・高速な処理というアプリケーションからの要請と、性能向上が鈍化しつつあるコンピュータの処理能力とのギャップは大きくなるばかりであり、コンピュータの性能向上に新たなアプローチが必要になってきている。このギャップを埋めるために筆者らは、適用するアプリケーション領域（ドメイン）を絞ることで桁違いの性能向上を図る「ドメイン指向コンピューティング」というアプローチを研究している。

本稿では、まずドメイン指向コンピューティングの考え方を説明した後、その具体的な適用例である、画像による画像の検索を高速化する「メディア処理向けドメイン指向サーバ」と、そのアクセラレーション手法、および応用について述べる。更に、様々な領域でタイムリーにドメイン指向サーバを開発する上で重要となるFPGA (Field-Programmable Gate Array) 設計支援環境について述べる。最後に、これまでの成果と今後の取り組みについて述べる。

## ドメイン指向コンピューティング

ドメイン指向コンピューティングとは、適用するアプリケーション領域を絞り、その領域の特性

に合わせてハードウェアとソフトウェアを最適化することで、サーバの性能および運用性を飛躍的に向上させるアプローチである。

例えば、汎用CPUだけでは十分な性能が得られない場合、FPGAやGPU (Graphics Processing Unit) などを採用する。また、I/Oの高速化や省電力化などが必要なドメインでは、フラッシュストレージや不揮発性メモリなどを活用する。このように、種々のデバイスをドメインの特性に合わせて最適に構成し、ドメインの主要機能を高性能に提供することで、そのドメインに属する様々なアプリケーションの性能や運用性の飛躍的向上を実現する。これがドメイン指向コンピューティングの考え方であり、筆者らはその考えを実現するサーバとして「ドメイン指向サーバ」の研究開発に取り組んでいる。

これまでは、ムーアの法則に従った汎用CPUの性能向上が継続していたため、FPGAやGPUを用いたアクセラレーターの優位性を長期にわたって維持することは困難であった。しかし、半導体の性能向上が鈍化しつつあるため、アクセラレーターの優位性が継続するようになっている。またFPGAやGPUは汎用化・低コスト化しているため、今後これらの汎用デバイスを用いながら、ドメインに特化して性能を差異化する技術の重要性は増していくと考えられる。

その中で、FPGAはアプリケーションが扱うデータや演算処理内容に応じて、演算器の構成やデータのビット幅、メモリなどを適切に構成することができ、ドメインに特化したサーバを実現するのに有用なデバイスと言える。しかし、当然のことながらFPGAは万能ではなく、あらゆる処理の性能が向上するわけではない。高性能を実現するためには、FPGAに適したアプリケーションの選択と処理特性に応じた適切なアーキテクチャー設計が必要である。また、アクセラレーター開発に当たっては、いかに容易に設計できるかという設計手法も重要である。

このような背景から、筆者らはFPGAを活用したアクセラレーション技術と、FPGAの回路設計を容易化する設計手法の研究開発を進めている。

## メディア処理向けドメイン指向サーバ

画像、音声などのメディアデータは、社会で生成されるデータの半分以上を占めると言われている。これらを有効に活用するためのメディア処理は、高速化が求められている重要なドメインの一つである。例えば、企業などの現場では、プレゼンテーション資料などの図や写真を含む文書が日々大量に生成されている。これらの文書の中から、欲しい画像を効率的に探し出して再利用できれば、資料作成などの業務の効率化が期待できる。

そこで今回、筆者らはドメイン指向サーバの第一弾として、メディア処理の一つである画像検索に特化した「メディア処理向けドメイン指向サーバ」の開発を行った。

### ● 部分画像検索

部分画像検索とは、画像そのものを検索のためのクエリとし、そのクエリ画像と一部でも一致する画像を網羅的に検索する機能であり、文書などで使用される図や写真を効率的に活用する技術として有効である。

図-1は、今回開発した部分画像検索の処理フローである。本手法では、まず検索キーとなるクエリ画像と蓄積されているデータベース画像のそれぞれに対して、画像の部分領域ごとの特徴点

におけるBRIEF (Binary Robust Independent Elementary Features)<sup>(1)</sup>と呼ばれる局所特徴量を算出する。BRIEFは、特徴点の周辺からランダムに決定された2点の画素値の大小関係の特徴量として用いる手法である。今回は、128対 (256画素) から抽出される128ビットを一つの特徴点に対する特徴量とする。

特徴量の算出後、クエリ画像とデータベース画像の各特徴点のBRIEF特徴量を比較することにより、画像間の部分領域の類似度を判定する。データベース画像の各特徴点に対するハミング距離が最小となるクエリ画像の特徴点を求めることで、画像間の各特徴点の対応ペアを決定する。これを全ての特徴点に対して総当たりで行うことにより、部分領域の類似も抽出できる。

本アルゴリズムでは、特徴点同士を比較するマッチング処理を粗マッチングと詳細マッチングの2段階で行う。粗マッチングで類似度がしきい値以下のものは、処理を打ち切ることによって演算量の削減を行っている。しかし、部分画像検索の処理は、画像間の各特徴点のBRIEF特徴量の比較を網羅的に行う必要がある。このため、依然として膨大な演算が必要であり、従来の汎用サーバでは非常に長い処理時間を要することが問題となっていた。

### ● FPGAによる高速化

今回、部分画像検索において、高負荷処理である特徴量算出とマッチング処理をFPGAにオフロードすることで、検索速度を飛躍的に向上させることに成功した。

FPGAによる高速化を実現するためには、アプリケーションの処理特性に応じて演算器やメモリの構成、および制御フローを適切に設計する必要がある。一般に、FPGA回路の動作周波数は数百MHz程度と、今日のCPUと比較すると1/10程度である。このため、FPGAによる高速化の実現には、①アプリケーションに特化した小型・高効率な専用演算器を、②FPGAリソース内で高並列に実装し、③高い稼働率で動作させる、ことが鍵となる。

今回の部分画像検索の実装においては、これら三つの要素を高いレベルで実現した。全体ブロック図を図-2に示す。特徴量算出部は32モジュール、マッチング処理部は6モジュールを実装している。マッチング処理部には、ハミング距離計算と最小

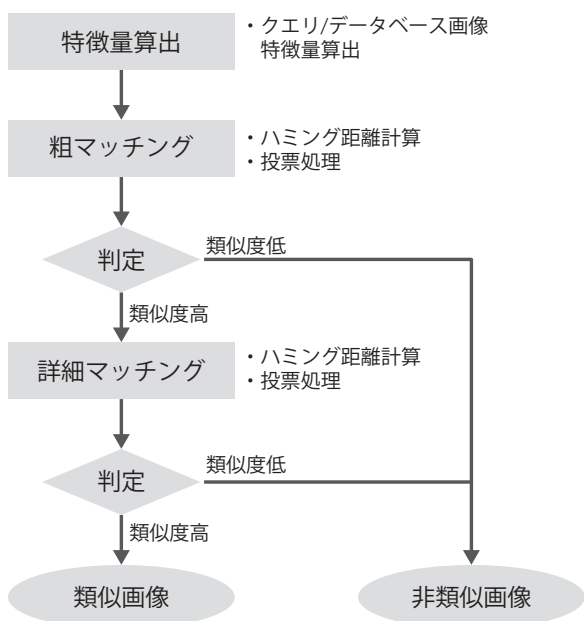


図-1 部分画像検索処理フロー

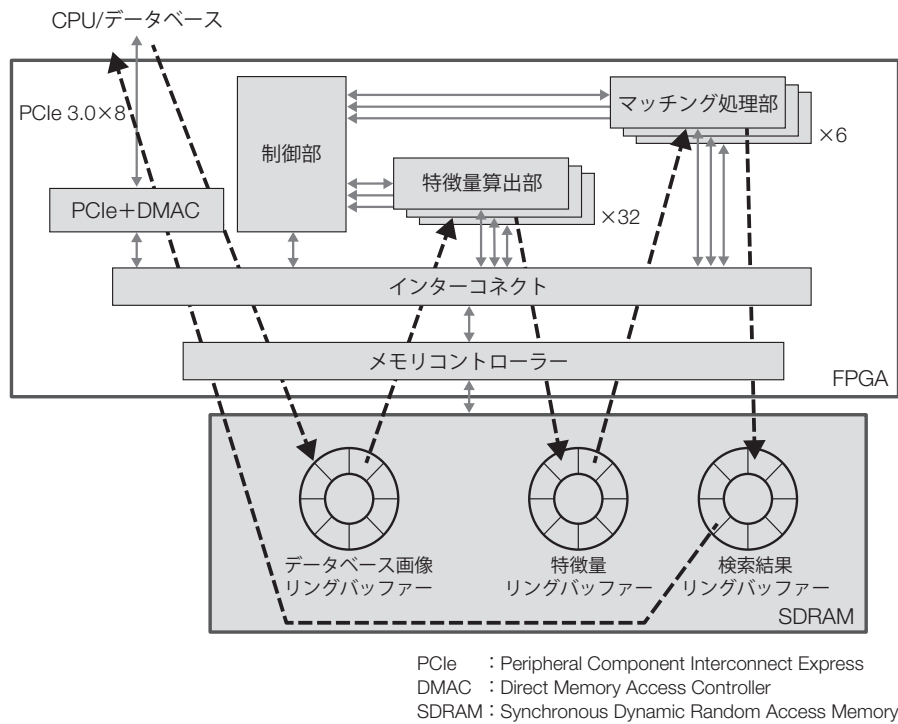


図-2 部分画像検索アクセラレータ全体ブロック図

値検出処理を実行する専用演算器を並列に実装し、1クロックあたり64特徴点ペアの対応点検出を実現している。これを1モジュールとして、6モジュール構成で実装することにより、毎クロック384個の特徴点ペアの対応点検出が可能である。

また、高い演算器稼働率を実現するために、これらの高並列な演算器に適切にデータを流し込むことが可能なデータ供給回路と、データフローに着目して読み出されたデータを無駄なく活用するスケジューリング技術を開発した。本アルゴリズムでは、粗マッチングの判定結果によって詳細マッチングを実行するかどうかが決まるため、通常の処理フローではスムーズなパイプライン実行にはならない。しかし、判定結果によって次に行う処理対象が変わった場合であっても、先読みしたデータが無駄にならないように処理順序を入れ替えることにより、スムーズなパイプライン実行を可能な構成としている。これらの技術により、高並列演算器を高い稼働率で動作させることを可能にし、高い検索性能を実現した。

● **メディア処理向けドメイン指向サーバの性能と応用**

これらの技術を適用した部分画像検索アクセラ

表-1 FPGAへの実装結果

ロジック数 (使用率)	313,816 (87%)
レジスタ数 (使用率)	434,916 (30%)
DSPブロック数 (使用率)	72 (20%)
ブロックメモリビット数 (使用率)	43,901,724 (81%)
クロック周波数	200 MHz

DSP : Digital Signal Processor

ラレータをFPGAに実装することにより、汎用サーバと比較して50倍以上に高速化する高速部分画像検索サーバを開発した<sup>(2)</sup>。FPGAボードは米BittWare社のS5-PCIe-HQ (Altera Stratix V GX AB FPGA搭載) を、サーバはFUJITSU Server PRIMERGY CX2570 M1を使用した。FPGA回路のリソース使用量などの実装結果を表-1に示す。今回開発した高速部分画像検索サーバ1台で、1万枚以上の画像データベースからクエリ画像の任意部分に一致する画像を検索する処理を、約1秒で完了できる。

また、本サーバを文書検索システムに適用することにより、従来は分単位の時間がかかっていた文書中の部分画像の網羅的な検索処理を秒単位の短縮でき、画像によるインタラクティブな検索が

可能となる。これにより、企業内に蓄積されている大量の文書を効率的に再利用できる。

**FPGA設計支援環境**

回路の再構成が可能なFPGAは、ドメイン指向コンピューティングにとって重要なデバイスであり、近年の高位合成技術の進歩により効率的な設計が可能になってきている<sup>(3)</sup>。しかし、高性能なFPGA回路の設計には、前述の部分画像検索のように、限られたリソースの中に演算器を高並列に実装し、かつ高い稼働率で動作させる必要があるため、高度な設計スキルが要求される。特に重要となるのが、大きな処理単位での最適な並列性や、スムーズなパイプライン処理を考慮したアーキテクチャーレベルでの構造設計である。

アーキテクチャーレベルの構造設計で重要な検討項目の一つは、パイプライン処理の各モジュールの性能および構造を決めることである。パイプライン処理では、最も性能が低いモジュールの処理が全体の性能を決定するため、ボトルネックとなるモジュールの性能を向上させることが重要である。一方、性能向上には並列化が有効であるが、その分ハードウェアリソースが必要になることから、これを考慮した構造の決定が大切である。

筆者らは、このような各モジュールの処理性能が最適となるような構造設計を支援するために、性能ボトルネックなど、アーキテクチャー設計に有用な情報を設計者が直感的に分かるように、統合的に可視化する設計支援環境を構築した。本設計支援環境は、C言語で記述されたコードを入力と

して、ボトルネックとなるモジュールなどの解析結果を可視化する。設計者は、可視化された解析結果を基に該当モジュールの改良を検討し、目標性能を決定する。検討結果は、設計支援環境上において該当モジュールの数値として直接入力することで、可視化情報に反映される。この作業を繰り返すことにより、ボトルネックとなるモジュールの性能を改良しながら全体性能の最適化を実施できる(図-3)。

本設計支援環境では、各モジュールの処理性能を把握するために、横軸を時間にとった帯状のグラフでデータ処理の開始間隔を表示する(図-4)。ここでデータ処理の開始間隔とは、一つのデータセットの処理を開始してから、次のデータセットの処理が開始できるようになるまでの時間である。この開始間隔が最も長いモジュールが全体のスループットの性能を決定するボトルネックであり、強調して画面上に表示している。また、階層構造を持っているモジュールに対しては、リストから表示したいモジュールを選択することで、その下位階層のモジュールの開始間隔が表示され、下位階層中のボトルネックとなっている箇所が提示される。更に、データの流れを設計者が把握できるように、モジュール間のデータ依存関係を示し、データ受け渡しの前段のモジュールの終端とデータを受け取るモジュールの始端が同じになるように表示している。

また、FPGAではハードウェアリソースの使用量を考慮しながら設計する必要があるため、FPGAの論理素子であるロジックエレメント、FPGAの内部

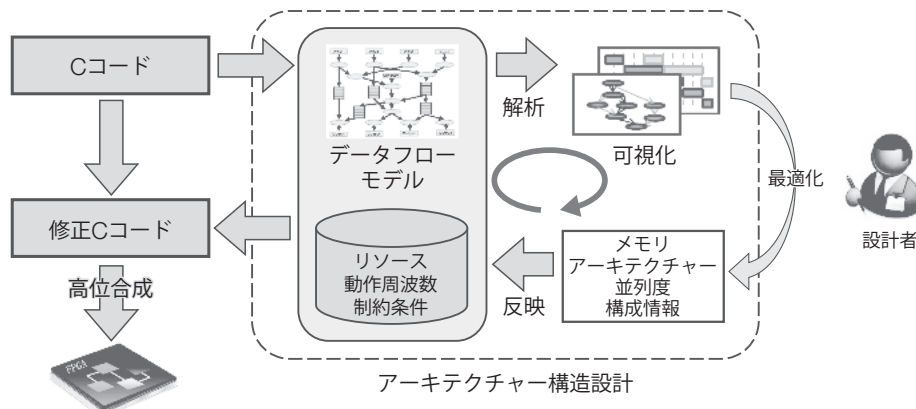


図-3 アーキテクチャー設計フロー

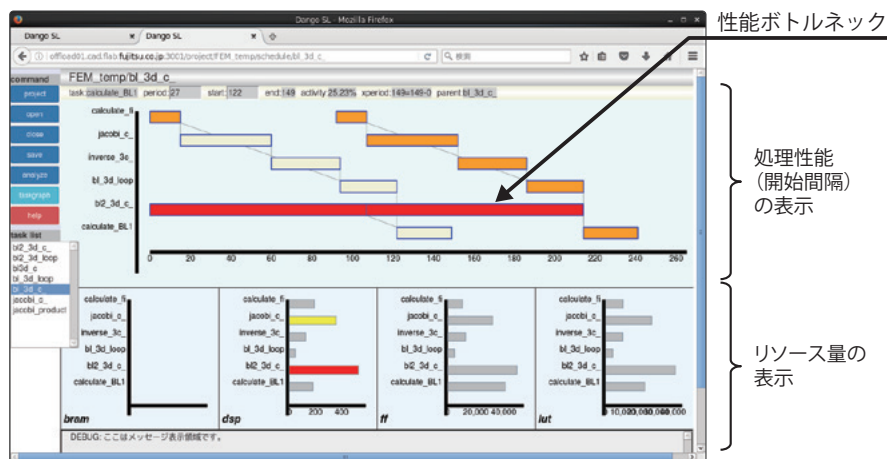


図-4 FPGA設計支援環境の表示画面

RAM, および算術演算に特化した専用回路 (DSP: Digital Signal Processor) の各リソースについて、モジュールごとの使用量を表示している。一定量以上のリソースを使用している場合には、グラフに色を付けて強調する。更に、回路全体のデータ入力の開始間隔に対する各モジュールのデータ入力の開始間隔の比を、稼働率として表示している。

これらの可視化情報を活用することで、設計者はアーキテクチャーレベルの最適化を行うことができる。例えば、リソースが不足している場合には、稼働率に余裕のあるモジュールの性能を落とすことでリソースを削減できる場合がある。例えば、稼働率が50%のモジュールは、半分の時間しか動作していない。このようなモジュールは、並列化されていれば並列数を減らしたり、類似処理のモジュールの共通処理を統合したりすることによるリソース量削減の検討が可能となる。

む す び

本稿では、ドメイン指向コンピューティングのコンセプトを紹介し、それを実現するために有用なデバイスであるFPGAを活用したメディア処理向けドメイン指向サーバと、その開発を効率化する技術としてFPGA設計支援環境について述べた。

今後、様々な分野で処理量が増大しており、汎用サーバでは性能要求を満たさないケースがますます増えると予想される。富士通研究所では、特定分野で飛躍的な性能向上を実現するドメイン指向サーバと、その開発を支えるFPGA設計支援環境

の実用化を進めていく。

参考文献

- (1) M. Calonder et al. : BRIEF : Binary Robust Independent Elementary Feature. European Conference on Computer Vision, p.778-797 (2010).
- (2) H. Matsumura et al. : An FPGA-accelerated partial duplicate image retrieval engine for a document search system. IEEE Winter Conference on Applications of Computer Vision (WACV 2016).
- (3) R. Nane et al. : A Survey and Evaluation of FPGA High-Level Synthesis Tools. IEEE Trans. on CAD of Integrated Circuits and Systems, Vol.35, No.10, Oct., p.1591-1604 (2016).

著者紹介



渡部康弘 (わたなべ やすひろ)

コンピュータシステム研究所  
次世代コンピュータシステムプロジェクト  
ドメイン指向コンピューティングの研究開発に従事。



**藤澤久典** (ふじさわ ひさのり)

コンピュータシステム研究所  
次世代コンピュータシステムプロジェクト  
ドメイン指向コンピューティングの研究開発に従事。



**小沢年弘** (おざわ としひろ)

ソフトウェア研究所  
ドメイン指向コンピューティングの研究開発に従事。