

論文 / 著書情報
Article / Book Information

題目(和文)	ハードウェアアクセラレータを活用したクラウド基盤技術“Video Service Function Chaining”に関する研究
Title(English)	
著者(和文)	右近祐太
Author(English)	Yuta Ukon
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11937号, 授与年月日:2021年3月26日, 学位の種別:課程博士, 審査員:高橋 篤司,一色 剛,本村 真人,中原 啓貴,原 祐子
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11937号, Conferred date:2021/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信 情報通信	系 コース	申請学位 (専攻分野)： 博士 Academic Degree Requested Doctor of	(工学)
学生氏名： Student's Name	右近 祐太		指導教員 (主)： Academic Supervisor(main)	高橋 篤司 教授
			指導教員 (副)： Academic Supervisor(sub)	

要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

本論文は、「ハードウェアアクセラレータを活用したクラウド基盤技術“Video Service Function Chaining”に関する研究」と題し、和文 6 章から成る。

第 1 章「はじめに」では、研究背景について述べる。近年、コンピュータビジョンや人工知能 (AI) などの高度なクラウドサービスの需要が高まっている。特に、開発コストや開発時間を抑えつつユーザの個々の目的に合致した高度なサービスを高速かつ低消費電力で提供するために、本論文では柔軟なクラウド画像処理サービスを実現する Video Service Function Chaining (VSFC) を提案する。既存の Service Function Chaining (SFC) は、ネットワークサービスを柔軟に提供することを目的としているが、ユーザごとに異なった条件で高品質な画像処理サービスを提供するなどの高度な情報処理サービスに対する実現例はなく、その実現のためには低遅延かつ低消費電力のサーバノードの開発が必要である。そこで、本論文では 2 種類の技術によりサーバノード内の遅延を削減し、VSFC による高品質なクラウド画像処理サービスを実現する。

第 2 章「Video Service Function Chaining」では、ハードウェアアクセラレータ (HWA) を備えた汎用サーバを用いて画像処理機能を組み合わせる VSFC を提案する。まず、既存の SFC ネットワークの構成を紹介し、次に、SFC 技術を利用して複数のサーバノードに実装された画像処理機能を組み合わせる VSFC ネットワークのコンセプトを説明する。

第 3 章「パケット順序制御回路を備えた CPU-HWA サーバアーキテクチャの提案」では、サーバノード内で低遅延データ転送を実現する CPU-HWA サーバアーキテクチャを提案する。提案アーキテクチャでは、CPU が受け取った画像データを含むパケットを、ネットワークレベルで HWA に転送したのち、リングバッファを用いたリアルタイムパケット順序制御回路を用いてパケットを正しい順序に整列させ、適切なタイミングで処理回路に入力する。実験では、HOG 特徴量計算を、提案アーキテクチャで行うと、従来構成に比べてデータ転送および処理を高速かつ低消費電力で実現でき、パケット順序の入れ替えが大きい場合でも高速な処理が可能であることを示した。

第 4 章「Approximate Completion-Detection 方式の可変レイテンシ回路の提案」では、計算エラーを許容するアプリケーションにおいて、処理を高速化するための可変レイテンシ回路を提案する。提案の Approximate Completion-Detection (ACD) 方式では、高速なサブクロックを用いて演算の出力を監視する。出力が一定時間変化しない場合、演算は完了したとみなし、各演算のレイテンシを動的に変更することで、回路の高速動作を実現する。ACD 方式では、計算完了を誤検出した場合、計算エラーが発生する可能性があるが、エラー率は制御可能であり、低いエラー率で高速に動作できる。実験では、画像処理機能を提案の ACD 方式の可変レイテンシ回路で実現し、わずかな計算エラーを許容することで従来方式の回路と比べて処理性能が向上することを示した。また、設計した画像処理回路は、従来の回路構成と比べて面積遅延積が小さいことを示した。

第 5 章「VSFC 対応映像監視サービスの評価」では、2 種類の画像処理機能 (HOG 特徴量計算、フレーム間差分法による移動物体検出) からなる VSFC 対応映像監視サービスの性能を評価する。まず、提案アーキテクチャを FPGA 搭載サーバで実現し、2 種類の画像処理機能を組み合わせることで、50 台のネットワークカメラからの 60fps の VGA 映像をリアルタイムに監視できることを示した。次に、画像処理機能に ACD 方式の可変レイテンシ回路を用いることで、サービス品質をほとんど低下させることなく処理性能が向上し、低遅延な映像監視サービスを実現できることを示した。

第 6 章「おわりに」では、CPU-HWA サーバアーキテクチャによるサーバノード内のデータ転送遅延削減と ACD 方式の可変レイテンシ回路による画像処理遅延削減により、提案の VSFC が柔軟で高度なクラウド画像処理サービスを実現すると結果を総括するとともに、今後の展望について述べる。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信 情報通信	系 コース	申請学位 (専攻分野)： 博士 Academic Degree Requested Doctor of (工学)
学生氏名： Student's Name	右近 祐太		指導教員 (主)： 高橋 篤司 教授 Academic Supervisor(main)
			指導教員 (副)： Academic Supervisor(sub)

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

Image-processing services based on cloud computing are in great demand. These services should be customizable to achieve the purpose of each user. To flexibly provide image-processing services that match each user's purpose, we propose a technique called video service function chaining (VSFC) for combining image-processing functions in CPU-hardware accelerator (HWA) coupled servers. One of the most important requirements for providing a high-quality service in this proposal is that low latency in server nodes.

We reduced latency in server nodes by two kinds of techniques; one is a CPU-HWA server architecture with a real-time packet reordering circuit, and the other is a variable-latency circuit by an approximate completion-detection (ACD) method. The architecture achieves reduction of data transfer delay in the server node by forwarding packets containing image data to the HWA at the network level and performing image processing for the packets while controlling the packet order. Furthermore, the variable-latency circuit enables high-speed image processing effectively in which processing time for each task is changed by roughly detecting its completion.

To confirm the effectiveness of the proposed techniques, several experiments were conducted. The first experiment showed latency of an image-processing function implemented in the proposed architecture is shorter than that in conventional architectures. The second experiment showed that effective clock periods of image-processing circuits obtained by the ACD method with a small error are improved compared with circuits without error. Finally, we developed and evaluated a video monitoring service that consists of two image-processing functions: a HOG feature calculation circuit and an object detection circuit. The evaluation result showed that the service can be provided with low latency by combining image-processing functions using our architecture. We also evaluated the quality of the video monitoring service that is accelerated by an ACD-based HOG feature calculation circuit and showed that the throughput is improved with almost no decrease in object detection accuracy.

In summary, we have confirmed that the proposed techniques reduce latency in server nodes. In addition, we have confirmed that the server node using the proposed techniques can provide the low-latency video monitoring service based on VSFC.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note: Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1 copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).