

令和 4 年 6 月 17 日現在

機関番号：32660
研究種目：研究活動スタート支援
研究期間：2019～2021
課題番号：19K24368
研究課題名（和文）行列演算機構を有するアーキテクチャによる画像処理高速化フレームワークの構築

研究課題名（英文）Development of an acceleration framework for image processing using architectures with tensor processing

研究代表者
前田 慶博（Maeda, Yoshihiro）

東京理科大学・工学部電気工学科・助教

研究者番号：80843375
交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：本研究では、行列の積和演算を高速に実現可能な行列演算機構を、画像処理アルゴリズムにおいて有効に活用する方法について検討を行った。具体的には、画像処理における基本的な処理であるフィルタリングを対象に、行列演算機構に適した画像処理アルゴリズムの検討を行った。検討の結果、行列演算機構を用いることで、画像処理アルゴリズムの高速化を実現できることを示した。

研究成果の学術的意義や社会的意義

本研究は、行列の積和演算を高速に計算可能な行列演算機構を画像処理にも応用する方法を明らかにするものである。行列演算機構は、深層学習に特化した専用演算器であったが、本研究により、様々な処理においても行列演算機構が有効に活用できることを示した。これにより、行列演算機構の多種多様な用途での活用が期待される。

研究成果の概要（英文）：We investigated how to effectively utilize matrix arithmetic instructions, which can realize fast matrix sum-of-products operations, in image processing algorithms. In this study, we studied image filtering algorithms, which is a fundamental process in image processing, suitable for matrix arithmetic instructions. The results of this study showed that the matrix operation instructions could be accelerated image filtering algorithms.

研究分野：画像処理

キーワード：画像処理 行列演算機構 GPU FIRフィルタリング

1. 研究開始当初の背景

近年、4K・8K 画像の登場や高精細イメージセンサを搭載したスマートフォンによって画像の高解像度化が進んでいる。それに伴い、データ量及び処理量も増大しており、リアルタイム性を求められやすい画像処理分野においては、これまで以上の高速化が求められている。また、Deep Neural Network (DNN) の活躍に伴い、DNN の処理に特化した行列演算機構が登場している。これは行列の積和演算を実現する機構である。DNN を始めとする Convolutional Neural Network (CNN) の処理の多くは、行列演算として実現可能であるため、この演算機構は DNN の高速化を達成している。しかし、一般的な画像処理アルゴリズムにおいては、行列演算機構が必ずしも有効的に活用できるとは限らない。

2. 研究の目的

本研究課題では、画像処理において、行列演算機構を有する計算機アーキテクチャを有効的に活用した高速化手法を学術的に明らかにする。加えて、その知見を学術的に体系化することを目的とする。固定サイズの行列演算を行うことを前提とする行列演算機構の効率的な活用方法は未知である。加えて、行列演算機構は特殊な浮動小数点数型を用いるため、精度劣化の問題も存在する。これらを画像処理と計算機アーキテクチャの双方アプローチから明らかにする。

3. 研究の方法

本研究は、画像処理アルゴリズムと計算機アーキテクチャの観点から、行列演算機構を活用した画像処理として、これまでの画像処理研究の再考を行う。

(1) 行列演算機構に適した画像処理アルゴリズムの再考及び実装の検討

一般的な画像処理アルゴリズムは、必ずしも行列の積和演算で表現できるとは限らない。また、行列の積和演算で表現できたとしても、行列演算機構は性能を高めるために処理できる行列の大きさが固定であるため、それが効率的な実装となっているとは限らない。そのため、一般的な画像処理アルゴリズムであるフィルタリングを対象に、アルゴリズムを行列の積和演算で再定義を行う。そして、行列演算機構の特性に適した実装手法について検討を行う。

(2) 16bit 浮動小数点数による画像処理アルゴリズムの精度劣化の影響調査

一般的な画像処理アルゴリズムはデータ型として 32bit や 64bit 浮動小数点数を用いることを前提としている。しかし、行列演算機構では、16bit 浮動小数点数型が使用される。これは、高い高速化を実現するための犠牲であり、値の表現範囲が通常使用される 32bit 浮動小数点数よりも狭く、数値の分解能が低いことを表している。そのため、16bit 浮動小数点数型を用いた数値計算による誤差の影響について調査を行う。

4. 研究成果

(1) 行列演算機構に適した FIR フィルタリングアルゴリズム

FIR フィルタは、ガウシアンフィルタなど画像処理における基本的な処理を実現するアルゴリズムの一つである。FIR フィルタを行列の積和演算で表現できる形式に変換した場合、対象の行列が疎行列になり、行列演算機構を有効に活用できない。図 1 は、単純に重みと画素を行列表現した場合である。これでは、演算結果の行列の対角成分のみに有効な演算結果が格納される。そのため、行列演算機構を有効に活用できてない。

そこで、本研究では、行列演算機構を有効に活用する FIR フィルタアルゴリズムを提案した。この手法では、フィルタリング重みと対象画素を行列として表現し、それらの積和演算として FIR フィルタリング

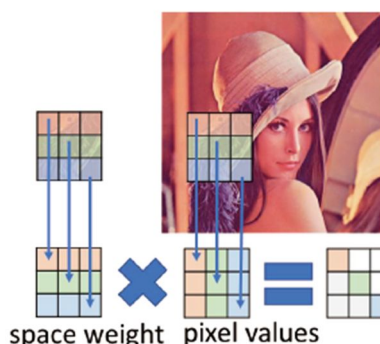


図 1 : 行列演算機構による単純な FIR フィルタ表現

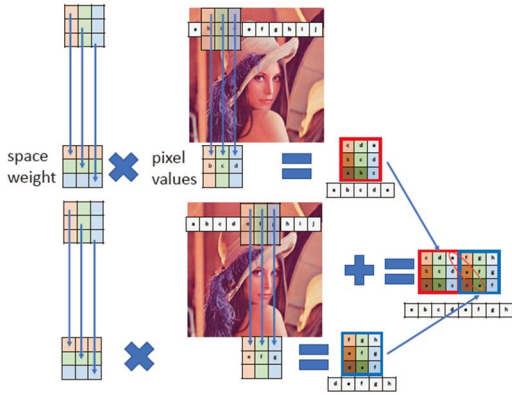


図 2 : 行列演算機構による効率的な FIR フィルタ表現提案手法

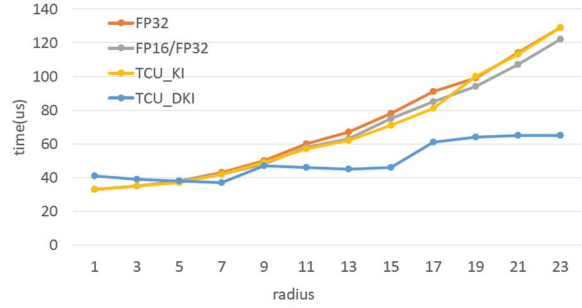


図 3 : 行列演算機構による FIR フィルタ表現の計算時間の比較

を実現する．提案手法（図 2）では，FIR フィルタが対象画素をループで走査した際に，隣の画素での使用されなかった計算結果を活用することで，より少ない行列積の積和演算数によって FIR フィルタを表現できる．図 3 に示す結果においても提案実装方式（TCU_DKI）は高速であることが確認された．

この成果は，行列演算機構を用いた FIR フィルタ表現を実現したものであり，行列演算機構は DNN 以外にも効果的に活用できることを示した．今後の展望として，更なる高速化を実現するために，タILING等の実装方式なども検討する必要がある．

(2) 行列演算機構に適した積分画像生成アルゴリズム

積分画像はガウシアンフィルタやボックスフィルタなど高速に実現する基盤技術である．本研究では，行列演算機構を用いた積分アルゴリズムを提案した．提案アルゴリズムでは，まず積分画像生成アルゴリズムを行列の積和表現によって表現した．画像を表現する行列を I としたときに，積分画像は全ての要素が 1 の下三角行列 L と上三角行列 U を用いて，行列の積 $L \cdot I \cdot U$ で表現することができる（図 4）．また，下三角行列 L と上三角行列 U の読み込み順序を変えることで，メモリ使用量の削減を行った（図 5）．表 1 の結果より，提案手法によって行列演算機構を使用して，従来手法と同程度の計算時間を達成した．また，16 ビット浮動小数点数型による誤差の影響についても調査を行った（表 2）．

この成果によって，積分画像生成アルゴリズムを行列演算機構によって表現できることを示した．今後の展望として，更なる高速化や 16bit 浮動小数点数型の誤差の影響を小さくするために，タILING等の実装方式なども検討する必要がある．

$$L \cdot I \cdot U = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & a_{1,1} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n,1} & \dots & a_{n,n} \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & a_{1,1} & \dots & \sum_{i=1}^n a_{1,i} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sum_{i=1}^1 \sum_{j=1}^n a_{j,i} & \dots & \sum_{i=1}^n \sum_{j=1}^n a_{j,i} \end{pmatrix}$$

図 4 : 行列積による積分画像生成

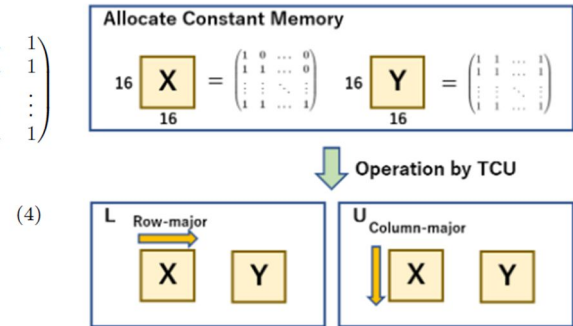


図 5 : 三角行列の性質を考慮した省メモリ化

表 1 : 積分画像生成アルゴリズムの計算時間

| 使用型 | 入力画像サイズ | 提案手法 [ms] | 比較手法 [ms] |
|-----------|-------------|-----------|-----------|
| FP16/FP32 | 128 × 128 | 0.4928 | 0.3340 |
| FP16/FP32 | 256 × 256 | 1.116 | 0.8924 |
| TF32/FP32 | 512 × 512 | 2.489 | 2.062 |
| TF32/FP32 | 1024 × 1024 | 10.22 | 5.694 |
| TF32/FP32 | 2048 × 2048 | 55.61 | 17.78 |
| TF32/FP32 | 4096 × 4096 | 345.0 | 66.18 |
| TF32/FP32 | 8192 × 8192 | 2422 | 256.9 |

表 2 : 積分画像生成アルゴリズムの演算誤差

| 使用型 | 入力画像サイズ | 提案手法 | 比較手法 |
|-----------|-------------|---------|---------|
| FP16/FP32 | 128 × 128 | 110.163 | 0.00000 |
| FP16/FP32 | 256 × 256 | 107.227 | 0.00000 |
| TF32/FP32 | 512 × 512 | 100.430 | 0.00000 |
| TF32/FP32 | 1024 × 1024 | 97.0139 | 0.00000 |
| TF32/FP32 | 2048 × 2048 | 95.1709 | 0.00000 |
| TF32/FP32 | 4096 × 4096 | 94.8667 | 0.00000 |
| TF32/FP32 | 8192 × 8192 | 93.1102 | 0.00000 |

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計11件（うち招待講演 0件 / うち国際学会 6件）

| |
|---|
| 1. 発表者名 角谷勇仁, 大塚友貴, 前田慶博, 福島慶繁 |
| 2. 発表標題 フーリエ級数展開によるローカルブラシアンフィルタの高速化 |
| 3. 学会等名 電気・電子・情報関係学会東海支部連合大会 |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 角谷勇仁, 大塚友貴, 前田慶博, 福島慶繁 |
| 2. 発表標題 定数時間アルゴリズムによるローカルブラシアンフィルタの高速化 |
| 3. 学会等名 画像符号化シンポジウム(PCSJ)/映像メディア処理シンポジウム(IMPS) |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 T. Otsuka, N. Fukushima, Y. Maeda, K. Sugimoto, S. Kamata |
| 2. 発表標題 Optimization of Sliding-DCT based Gaussian Filtering for Hardware Accelerator |
| 3. 学会等名 International Conference on Visual Communications and Image Processing (VCIP) (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 N. Fukushima, Y. Kawasaki, and Y. Maeda |
| 2. 発表標題 Accelerating Redundant DCT Filtering For Deblurring and Denoising |
| 3. 学会等名 International Conference on Image Processing (ICIP) (国際学会) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 N. Fukushima, T. Tsubokawa, and Y. Maeda |
| 2. 発表標題 Vector Addressing For Non-Sequential Sampling In FIR Image Filtering |
| 3. 学会等名 International Conference on Image Processing (ICIP) (国際学会) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 福嶋慶繁, 前田慶博 |
| 2. 発表標題 バイラテラルフィルタの数値計算精度 |
| 3. 学会等名 画像符号化シンポジウム(PCSJ)/映像メディア処理シンポジウム(IMPS) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 坪川哲平, 前田慶博, 福嶋慶繁 |
| 2. 発表標題 乱択フィルタにおけるサンプリングカーネルの最適化 |
| 3. 学会等名 画像符号化シンポジウム(PCSJ)/映像メディア処理シンポジウム(IMPS) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Y. Kameda, Y. Maeda, and T. Hamamoto |
| 2. 発表標題 Operator overloading for cv::UMat converted to equivalent function calls at compile time |
| 3. 学会等名 International Workshop on Advanced Image Technology (IWAIT) (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 H. Tajima, T. Tsubokawa, Y. Maeda, and N. Fukushima |
| 2. 発表標題 Fast Local LUT Upsampling |
| 3. 学会等名 International Conference on Computer Vision Theory and Applications (VISAPP) (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 T. Sasaki, N. Fukushima, Y. Maeda, K. Sugimoto, and S. Kamata |
| 2. 発表標題 Constant-Time Gaussian Filtering for Acceleration of Structure Similarity |
| 3. 学会等名 International Conference on Image Processing and Robot (ICIPRoB) (国際学会) |
| 4. 発表年 2020年 |

| |
|-------------------------------------|
| 1. 発表者名 堀内慧, 前田慶博, 杉村大輔, 浜本隆之 |
| 2. 発表標題 周辺人物の動きの一貫性を考慮した歩行者経路の予測 |
| 3. 学会等名 電子情報通信学会技術研究報告 |
| 4. 発表年 2020年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

| 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|---------------------------|-----------------------|----|
|---------------------------|-----------------------|----|

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
|---------|---------|