

Ampere アーキテクチャの GPU による  
位相型計算機合成ホログラムの計算高速化

## Fast calculation of Phase-modulated Computer Generated Hologram with Ampere GPU

三谷永久\* 鈴木康平\* 高田直樹†

Towa Mitani Kohei Suzuki Naoki Takada

## 1 はじめに

電子化したホログラフィ（電子ホログラフィ）は「究極の立体テレビ」になると考えられている。しかし、計算機合成ホログラム（Computer-Generated Hologram, CGH）の計算は膨大であるため、未だ実用化されていない。GPU を用いた電子ホログラフィの研究は盛んに行われている。著者らが提案した GPU のための CGH 計算実装法 [1] を NVIDIA Ampere GA102 アーキテクチャの GPU に実装したところ、同アーキテクチャと Turing, Pascal GP104 アーキテクチャとの違いが原因で、理論性能に対して十分な計算性能を引き出すことができなかった。そこで本論文では、Ampere GA102 アーキテクチャの GPU においても、十分な計算性能を引き出すことのできる CGH 計算アルゴリズムを提案する。本稿では、1 枚の Ampere アーキテクチャの GPU (NVIDIA GeForce RTX 3080) を搭載した PC で、約 3 万点からなる三次元物体のリアルタイム動画再生に成功した。

## 2 CGH 計算

本論文では、三次元物体を点群で表したポイントクラウドモデルを用いる。物体を構成する点数を  $N_p$  とし、物体点を点光源と考える。位相型 CGH は、ホログラム面上の点  $(x_\alpha, y_\alpha)$  における三次元物体上の点光源の光の位相情報を、全ての点光源に対して求め重ね合わせる。その後、最終的な位相の変化を求め CGH を作製する。ホログラム面上の点  $(x_\alpha, y_\alpha)$  における位相変化  $\theta(x_\alpha, y_\alpha)$  は次式で求まる。

$$\theta(x_\alpha, y_\alpha) = \tan^{-1} \frac{\sum_{j=1}^{N_p} A_j \sin \left[ \frac{\pi}{\lambda z_j} \left\{ (x_\alpha - x_j)^2 + (y_\alpha - y_j)^2 \right\} \right]}{\sum_{j=1}^{N_p} A_j \cos \left[ \frac{\pi}{\lambda z_j} \left\{ (x_\alpha - x_j)^2 + (y_\alpha - y_j)^2 \right\} \right]} \quad (1)$$

ここで、三次元物体上の点  $j$  の座標を  $(x_j, y_j, z_j)$  とした。 $A_j$  は  $j$  番目の物体点の光の強度を示す。 $\lambda$  は再生光の波長である。

## 2.1 GPU アーキテクチャによる構造の違い

位相型 CGH の計算には式 (1) のように、 $\cos$  計算と  $\sin$  計算を多用する。よって、 $\cos$  計算と  $\sin$  計算は計算高速化のために、Streaming Multiprocessor (SM) 内の Processing Block に用意された、1 クロックで解を導出可能なコア（超越関数ユニット, SFU）に演算させている。一方、浮動小数点演算用の CUDA コアと比べて SFU は少ない。式 (1) を計算する場合、 $\cos$  計算と  $\sin$  計算以外は CUDA コアにより計算を行うが、CUDA コア数と SFU 数との比に合わないプログラムを実行した場合、SFU の演算待ちが生じてしまう。これまで著者らが用いていた CGH 計算実装法 [1] では、単精度浮動小数点演算回数と超越関数演算回数との比はほぼ 6:1 であった。Pascal GP104 アーキテクチャでは SM 内に CUDA コア (FP32) が 32 コア、SFU が 8 ユニット実装されている [2]。Turing アーキテクチャでは CUDA コアが 16 コア、SFU が 4 ユニット実装されている [3]。そのため、SM 内の CUDA コア数と SFU 数との比は 4:1 なので、これまで SFU の演算待ちが生じることはなかった。しかし、Ampere アーキテクチャでは、Turing アーキテクチャに実装されていた整数演算コア (INT32) 16 コアが CUDA コアとの選択式となり、整数演算が含まれない場合は実質的に CUDA コアが 32 コアに倍増していることになる [4]。しかし、SFU に関してはそのまま 4 ユニットであるため、CUDA コアが選択された場合、SM 内の CUDA コア数と SFU 数との比は 8:1 となり、従来の実装法 [1] の CUDA コアでの演算回数と SFU での演算回数との比より大きくなり、SFU の演算待ちが発生してしまう。

## 3 提案手法

Ampere アーキテクチャにおいて、CUDA コア数と SFU 数との比が変わったために従来の実装法 [1] では演算待ちが生じていた。そこで、CGH 計算における SFU 演算が占める割合を減らすことが出来れば、演算待ちを生じさせることなく計算が可能となる。前述のように、位相型 CGH の計算には式 (1) を利用する。この時、

\* 高知大学大学院総合人間自然科学研究科

† 高知大学教育研究部自然科学系理工学部門

式中の  $\cos$ ,  $\sin$  演算は三角関数の加法定理を用いて以下のように分解できる. ただし,  $\frac{\pi}{\lambda z_i}(x_\alpha - x_j)^2 = X$ ,  $\frac{\pi}{\lambda z_i}(y_\alpha - y_j)^2 = Y$  とする.

$$\theta(x_\alpha, y_\alpha) = \tan^{-1} \frac{\sum_j A_j \sin \phi_j}{\sum_j A_j \cos \phi_j}$$

$$\phi_j = \frac{\pi}{\lambda z_i} \left\{ (x_\alpha - x_j)^2 + (y_\alpha - y_j)^2 \right\} = X + Y$$

$$\cos(X + Y) = \cos X \cos Y - \sin X \sin Y$$

$$\sin(X + Y) = \cos X \sin Y + \sin X \cos Y$$

これにより  $\cos$ ,  $\sin$  計算は, CGH 上の座標  $(x, y)$  の各方向のみに依存する計算になった. よって, 予め  $x$  方向,  $y$  方向の  $\cos$ ,  $\sin$  計算の結果を保持しておき, その後に各画素で加法定理により最終的な値を計算することが可能となる. この時, 加法定理による演算が単純な加算・乗算となるため, SFU を使用する必要がなくなる. 加えて, これまで  $\cos$ ,  $\sin$  計算は CGH 上の全画素で行っていたため, 三角関数部分の演算回数は CGH の縦画素数を  $H$ , 横画素数を  $W$  とおいた時,  $2(H \times W)$  であった. しかし, 提案手法では, 各方向を予め計算してその結果を保持するという性質上, 演算回数は  $2(H + W)$  であるため, 従来の実装法 [1] と比べ演算回数が低減されている.

## 4 結果

本論文で使用した GPU の仕様を表 4.1 に示す. また, CGH 計算を行う PC は, CPU に Intel Core i7 4700 (クロック周波数 3.4GHz, 4 コア 8 スレッド), メモリに DDR3-1600 のものを 8GB, OS に Linux (CentOS 7.9 x86\_64) を使用している. ソフトウェア環境としては, CUDA 11.2 を利用している. CGH の解像度は  $1920 \times 1024$  とした. なお, 表 4.1 において, NVIDIA GeForce GTX 1080 Ti, 同 RTX 2080 SUPER, 同 RTX 3080 をそれぞれ 1080 Ti, 2080 SUPER, 3080 と略して表記した. 表 4.1 に示した 3 枚の GPU, 1080 Ti, 2080 SUPER, 3080 のそれぞれの場合において, 従来の実装法 [1] と提案するアルゴリズムでの計算を比較した結果を図 4.1 に示す. 3080 では前述の SFU 演算待ちが解消されたため, 大きく高速化された. 1080 Ti および 2080 SUPER においては, SFU の演算待ちは生じていなかったが, 三角関数の演算回数が低減されたため, こちらも従来の実装法 [1] と比べ高速化された.

## 5 まとめ

三角関数の演算回数を低減させた CGH 計算アルゴリズムを提案した. 本アルゴリズムを Ampere アーキテク

表 4.1: NVIDIA GeForce GPU の仕様

GPU	1080 Ti	2080 SUPER	3080
コアクロック	1582 MHz	1830 MHz	1740 MHz
CUDA コア	3584	3072	8704
メモリサイズ	11 GB	8 GB	10 GB
メモリバンド幅	484 GB/s	496 GB/s	760 GB/s
理論性能	11.34 TFLOPS	11.24 TFLOPS	30.29 TFLOPS

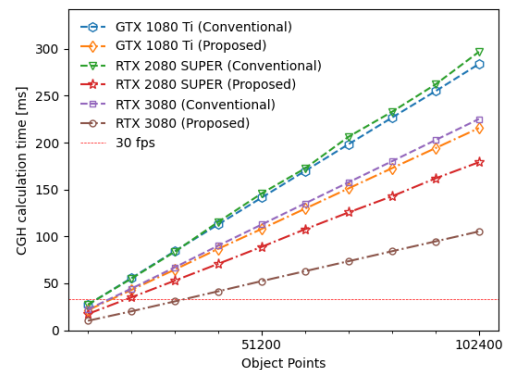


図 4.1: CGH 計算時間の比較

チャの GPU (NVIDIA GeForce RTX 3080) に実装したところ, GPU 1 枚で 3 万点までの三次元物体のリアルタイム動画再生に成功した.

## 謝辞

本論文の一部は, 日本学術振興会の科研費・基盤研究 (C) (課題番号 21K11996), I-O DATA 財団第 3 回研究開発助成によって行われた.

## 参考文献

- [1] Naoki Takada et al. *Appl. Opt.*, 51(30):7303–7307, Oct 2012.
- [2] NVIDIA. Whitepaper NVIDIA GeForce GTX 1080. [http://international.download.nvidia.com/geforce-com/international/pdfs/GeForce\\_GTX\\_1080\\_Whitepaper\\_FINAL.pdf](http://international.download.nvidia.com/geforce-com/international/pdfs/GeForce_GTX_1080_Whitepaper_FINAL.pdf).
- [3] NVIDIA. NVIDIA TURING GPU ARCHITECTURE. <https://images.nvidia.com/aem-dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf>.
- [4] NVIDIA. NVIDIA AMPERE GA102 GPU ARCHITECTURE. <https://www.nvidia.com/content/PDF/nvidia-ampere-ga-102-gpu-architecture-whitepaper-v2.pdf>.