

پردازش موازی در واحد پردازش گرافیکی

تألیف:

دکتر پیمان کشاورزian

دانشیار گروه کامپیوتر دانشگاه آزاد اسلامی کرمان

مهندس حدیث احمدی موسوی

مهندس بهشید صاحب قلم

مهندس ماریه بازاری

انتشارات دانشگاه آزاد اسلامی واحد کرمان

سال ۱۳۹۷

عنوان و نام بنیادآور	: پردازش موازی در واحد پردازش گرافیکی / تألیف بهمان کشاورزبان ... [و دیگران].
مشخصات نشر	: کرمان؛ دانشگاه آزاد اسلامی، واحد کرمان، ۱۳۹۷.
مشخصات ظاهری	: ۲۸۴ ص؛ صورتی، نمودار (رنگی).
وضعیت فهرست نویسی	: فیا
یادداشت	: تألیف بهمان کشاورزبان، حدیثه احمدی موسوی، بهنید صاحب قلم، غایبه بازایی.
یادداشت	: واژه نامه.
موضوع	: پردازش موازی
موضوع	: Parallel processing (Electronic computers)
موضوع	: برنامه نویسی موازی
موضوع	: Parallel programming (Computer science)
ردیفی دیوپی	: ۰۰۴/۳۵۱
ردیفی کنکوره	: QA۷۶/۵۸/ب۳۶ ۱۳۹۷
شابک گزیده	: کشاورزبان، بهمان، ۱۳۹۰ -
شابک	: دانشگاه آزاد اسلامی، واحد کرمان، انتشارات
شابک	: ۴۵۷۳۵۱۲
شابک	: ۹۷۸-۹۶۴-۱۰-۴۵۴۳-۴
شابک	: ISBN :978-964-10-4543-4

چاپ این کتاب به اسناد مفاد بند ۲ صورتجلسه شماره ۳۸/م مورخ ۹۵/۰۵/۱۷ شورای
انتشارات دانشگاه آزاد اسلامی، واحد کرمان، به انجام رسیده است

عنوان:	پردازش موازی در واحد پردازش گرافیکی
مؤلفان:	دکتر بهمن کشاورزبان، مهندس حدیثه احمدی موسوی، مهندس بهنید صاحب قلم، مهندس غایبه بازایی
ویراستار ادبی:	مجید غلامی
طراح جلد:	نسرین شهاب پور
حروف نگاری و صفحه آرای:	رقیه کمالی
ناشر:	انتشارات دانشگاه آزاد اسلامی، واحد کرمان
سال و نوبت چاپ:	اول - ۱۳۹۷
ناظر چاپ:	فاطمه معین الدینی
شمارگان:	محدود
کد:	۹۷/۱۲
بها:	۷۰۰ ریال
شابک:	۹۷۸-۹۶۴-۱۰-۴۵۴۳-۴

(حق چاپ برای ناشر محفوظ است)

نشانی: کرمان - بلوار ولی عصر - دانشگاه آزاد اسلامی، واحد کرمان - ساختمان اداری شماره ۴، انتشارات

تلفن و نمابر انتشارات: ۳۱۴۲۱۷۶۵

کد پستی: ۷۶۲۵۱۳۱۱۶۷

به نام خدایی که توفیق از اوست

دل زنده را نور تحقیق از اوست

مرکز انتشارات دانشگاه آزاد اسلامی واحد کرمان می‌کوشد - به ضرورت معرفی توانمندی‌های علمی موجود در مجموعه و شناساندن روند خجسته‌ی تولید فکر و نظریه- پژوهش‌های علمی، تألیفات و ترجمه‌های گران سنگ اعضای هیأت علمی خود و دیگر پژوهشگران و اندیشمندان را به زیور طبع بیاراید. و به جهاد علمی مطرح در سطح کشور به ویژه مجامع هم پیوند با آموزش عالی، یاری رسانند سرعت بخشد. اهتمام به پذیرش آثار قلمی نویسندگان و مترجمان دانشگاهی و پذیرش مکتوبات آماده سازی، چاپ و انتشار هر اثر، آفرینندگان آثار را از پی‌گیری‌های ممتد و دراز دامنه در بیرون رها کند، بی‌نیاز نموده، زمینه‌های مناسبی برای ترغیب روز افزون اعضای هیأت علمی به ارائه آثار خوب، کاربردی و اثر گذار در موضوعات گوناگون فراهم آورده است.

علاوه بر آن مرکز در صدد است تا با همه بر تشریک مساعی عموم مسئولان، امکانات و تجهیزات خود را متناسب با چشم انداز توسعه‌ی واحد، توسعه داده و به وجه شایسته تقویت نماید؛ اما در عین حال استمرار موفقیت این مرکز را در ارتقاء "حلاف بیکران الهی و حسن توجه و حمایت پدیدآورندگان آثار ارزشمند علمی، ناممکن می‌نماید.

از این رو فروتنانه از پروردگار متعال مدد طلبیده و به جدی خود در ادامه‌ی مسیری که برگزیده‌ایم، پای می‌فشاریم. تا چه در نظر آید.

انتشارات دانشگاه آزاد اسلامی

واحد کرمان

فهرست مطالب

صفحه	عنوان
۱۳	پیشگفتار
فصل اول: مقدمه ای بر پردازش موازی	
۱۵	۱-۱- موازی سازی در پردازنده های مرکزی
۱۶	۱-۲- موازی سازی در سطح دستورات
۱۸	۱-۳- موازی سازی در سطح وظیفه
۲۰	۱-۴- موازی سازی داده
۲۱	۱-۵- تکنیک Turbo Boost
۲۲	۱-۶- چند پردازشی متقارن (SMP)
۲۳	۱-۷- پردازنده های مرکزی چند هسته ای
۲۶	۱-۸- چالش های چند هسته ای
۲۶	۱-۸-۱- توان و دما
۲۷	۱-۸-۲- انسجام حافظه نهان
۲۸	۱-۸-۲-۱- پروتکل جستجوگر
۲۸	۱-۸-۲-۲- پروتکل مبتنی بر directory
۲۸	۱-۹- چند نخه
۲۹	۱-۱۰- سیستم باس و شبکه های به هم متصل
۲۹	۱-۱۱- ابرنخی
۳۲	۱-۱۲- ساختار حافظه کامپیوترهای موازی
۳۲	۱-۱۲-۱- حافظه مشترک
۳۳	۱-۱۲-۲- حافظه توزیع شده
۳۴	۱-۱۲-۳- حافظه ترکیبی توزیع شده -مشترک
فصل دوم: آشنایی با واحد پردازش گرافیکی	
۳۵	۲-۱- مقدمه
۳۶	۲-۲- تاریخچه
۳۶	۲-۲-۱- دهه ۱۹۷۰

۳۶	۲-۲-۲ دهه ۱۹۸۰
۳۷	۲-۲-۳ دهه ۱۹۹۰
۳۷	۲-۲-۴ از سال ۲۰۰۰ تا به حال
۳۸	۲-۳ شتاب دهنده ی گرافیکی
۳۸	۲-۴ آشنایی با پردازنده گرافیکی
۳۹	۲-۵ پردازش موازی در پردازنده های گرافیکی
۴۱	۲-۶ گسترش پردازنده های گرافیکی
۴۲	۲-۷ ویژگی های پردازنده های گرافیکی پیشرفته
۴۳	۲-۸ آشنایی با GPU و کاربرد آن در پردازش تصویر
۴۴	۲-۹ ابزارهای برنامه نویسی GPGPU
۴۵	۲-۹-۱ API ها و زبان های برنامه نویسی گرافیکی
۴۵	۲-۹-۲ زبان های برنامه نویسی GPU و کتابخانه ها
۴۶	۲-۹-۳-۱ کتابخانه OpenVIDIA
۴۶	۲-۹-۳-۲ کتابخانه GPUCV
۴۷	۲-۹-۳ پلتفرم های تجاری
۴۷	۲-۱۰ کاربردهای غیر گرافیکی پردازنده گرافیک
۴۸	۲-۱۱ بهره گیری از قابلیت های پردازنده گرافیکی در الگوریتم های بینایی ماشین و پردازش تصویر
۴۸	۲-۱۱-۱ مشکلات متداول
۴۹	۲-۱۲ روند پیشرفت کارت های گرافیک
۵۰	۲-۱۳ ساختار یک پردازنده گرافیکی
۵۱	۲-۱۴ پردازش موازی
۵۳	۲-۱۵ کاربرد پردازنده گرافیکی در طرح FASTRA
۵۶	۲-۱۶ مقایسه ی پردازش در پردازنده مرکزی و پردازنده گرافیکی
۶۲	۲-۱۷ معماری VLIW
۶۴	۲-۱۸ SIMD
۶۵	۲-۱۹ واحد محاسبات
۷۱	۲-۲۰ ویژگی های معماری جدید GCN

فصل سوم: پلتفرم های برنامه نویسی برای پردازش موازی

۷۳	۳-۱ مقدمه
۷۴	۳-۲ پلتفرم های همزمانی
۷۴	۳-۳ Cilk++
۷۶	۳-۳-۱ تجهیزات Cilk++ برای موازی سازی یک کد سریال
۷۷	۳-۴ OpenMp
۷۹	۳-۵ CUDA

۸۲	۳-۵-۱- تعریف نخ، بلاک، و شبکه در CUDA
۸۴	۳-۵-۲- انتساب توابع برای اجرا توسط یک هسته در CUDA
۸۵	۳-۵-۳- ارتباط بین قطعه CUDA و میزبان
۸۸	۳-۵-۴- همگام سازی و ارتباط نخ‌ها در CUDA
۸۸	۳-۵-۵- هسته ها و شبکه ها
۸۹	۳-۵-۶- بلاک ها
۸۹	۳-۵-۷- نخ ها

فصل چهارم: موازی سازی در متلب بدون استفاده از پردازنده گرافیکی

۹۱	۴-۱- مقدمه
۹۱	۴-۲- بردار سازی
۹۲	۴-۲-۱- عملگرهای عنصری
۹۴	۴-۲-۲- عملگر برداری ماتریس
۹۵	۴-۲-۳- ترفندهای مفید
۹۷	۴-۳- تخصیص اولیه
۹۹	۴-۴- حلقه ی For
۱۰۰	۴-۵- بررسی یک فرم ماتریس خلوت
۱۰۲	۴-۶- بهینه سازی کد
۱۰۲	۴-۶-۱- حداقل سازی خواندن/نوشتن فایل در حلقه
۱۰۳	۴-۶-۲- حداقل سازی تغییر دینامیکی مسیر و تغییر کلاس متغیر
۱۰۳	۴-۶-۳- حفظ تعادل بین خوانایی کد و بهینه سازی
۱۰۳	۴-۷- مثال ها

فصل پنجم: انجام تنظیمات برای برنامه نویسی موازی در متلب و CUDA

۱۱۳	۵-۱- مقدمه
۱۱۴	۵-۲- تنظیمات متلب برای برنامه ریزی c-mex
۱۱۴	۵-۲-۱- فهرست
۱۱۴	۵-۲-۱-۱- کامپایلرهای c/c++
۱۱۵	۵-۲-۱-۲- کامپایلر NVIDIA CUDA (nvcc)
۱۱۵	۵-۲-۲- انتخاب کامپایلر
۱۱۸	۵-۳- "hello-mex!" با استفاده از c-mex
۱۲۲	۵-۳-۱- خلاصه
۱۲۲	۵-۴- تنظیمات CUDA برای متلب
۱۲۲	۵-۴-۱- آماده سازی تنظیمات CUDA
۱۲۶	۵-۵- افزودن بردار ساده با استفاده از CUDA

۱۳۲	۵-۵-۱- خلاصه
۱۳۳	۵-۶- محاسبه‌ی کانولوشن تصویر
۱۳۵	۵-۶-۱- کانولوشن در نرم افزار متلب
۱۳۶	۵-۶-۲- کانولوشن در C-mex مرسوم
۱۳۹	۵-۶-۳- کانولوشن در C-mex رایج با CUDA
۱۴۴	۵-۶-۴- آگاهی از کارایی زمانی پروفایل
۱۴۶	۵-۷- خلاصه

فصل ششم: بهینه سازی برنامه از طریق Profiling

۱۴۷	۶-۱- مقدمه
۱۴۸	۶-۲- پروفایلینگ ۲: متلب برای یافتن تنگناها
۱۵۲	۶-۲-۱- پروفایلینگ دقیق تر: هسته های متعدد پردازنده ها
۱۵۴	۶-۳- پروفایلینگ کد C-mex با CUDA
۱۵۴	۶-۳-۱- پروفایلینگ CUDA با استفاده از Visual Studio
۱۵۸	۶-۳-۲- پروفایلینگ در CUDA با استفاده از NVIDIA Visual profiler
۱۶۹	۶-۴- تنظیمات محیط برای خطایابی C-mex

فصل هفتم: برنامه نویسی مداري در CUDA با C_mex

۱۷۹	۷-۱- مقدمه
۱۷۹	۷-۲- چینش حافظه در C_mex
۱۸۰	۷-۲-۱- ترتیب ستونی
۱۸۴	۷-۲-۲- ترتیب سطری
۱۸۵	۷-۲-۳- چیدمان حافظه برای اعداد مختلط در C-mex
۱۸۷	۷-۳- مدل برنامه نویسی منطقی
۱۹۰	۷-۳-۱- گروه بندی منطقی الف
۱۹۰	۷-۳-۲- گروه بندی منطقی ب
۱۹۱	۷-۳-۳- گروه بندی منطقی ج
۱۹۳	۷-۴- نکات ریز در GPU
۱۹۳	۷-۴-۱- موازی سازی داده
۱۹۴	۷-۴-۲- پردازنده جریانی
۱۹۴	۷-۴-۳- چندپردازنده جریانی
۱۹۴	۷-۴-۴- ریسمان
۱۹۶	۷-۴-۵- حافظه
۱۹۸	۷-۵- تحلیل روش ساده ی اولیه
۲۰۰	۷-۵-۱- بهینه سازی

۲۰۶	۷-۵-۲- بهینه سازی نوع ب
۲۰۹	۷-۵-۳- نتیجه

فصل هشتم: ابزارهای پردازش موازی در متلب

۲۱۱	۸-۱- مقدمه
۲۱۱	۸-۲- پردازش GPU برای توابع داخلی
۲۱۶	۸-۲-۱- چالش ها در پردازش GPU
۲۱۸	۸-۳- پردازش GPU برای توابع خارجی متلب
۲۲۰	۸-۴- پردازش وظایف موازی
۲۲۰	۸-۴-۱- ایجاد کد به متلب
۲۲۳	۸-۴-۲- دستور parfor
۲۲۵	۸-۵- پردازش موازی داده
۲۲۵	۸-۵-۱- spmd
۲۲۹	۸-۵-۲- آرایه های توزیع شده و هم توزیع
۲۳۴	۸-۵-۳- ایجادکننده ها با چند GPU
۲۳۵	۸-۶- استفاده ی مستقیم از فایل های CUDA و C-mex

فصل نهم: استفاده از کتابخانه های شتاب دهنده CUDA

۲۴۳	۹-۱- مقدمه
۲۴۳	۹-۲- CUBLAS
۲۴۵	۹-۲-۱- توابع cublas
۲۴۶	۹-۲-۲- ضرب ماتریس در ماتریس cublas
۲۴۸	۹-۲-۲-۱- مرحله اول
۲۴۸	۹-۲-۲-۲- مرحله دوم
۲۴۸	۹-۲-۲-۳- مرحله سوم
۲۴۹	۹-۲-۲-۴- مرحله چهارم
۲۵۰	۹-۲-۲-۵- مرحله پنجم
۲۵۲	۹-۲-۲-۶- مرحله ششم
۲۵۴	۹-۲-۲-۷- مرحله هفتم
۲۵۷	۹-۲-۲-۸- مرحله هشتم
۲۵۸	۹-۲-۲-۹- مرحله نهم
۲۵۸	۹-۲-۳- Cublas به همراه visual profiler
۲۶۱	۹-۲-۳-۱- خلاصه ای از cublas
۲۶۱	۹-۳- CUFFT
۲۶۳	۹-۳-۱- FFT دو بُعدی با CUFFT

۲۶۴	 مرحله اول ۹-۳-۱-۱
۲۶۴	 مرحله دوم ۹-۳-۱-۲
۲۶۴	 مرحله سوم ۹-۳-۱-۳
۲۶۵	 مرحله چهارم ۹-۳-۱-۴
۲۶۵	 مرحله پنجم ۹-۳-۱-۵
۲۶۸	 مرحله ششم ۹-۳-۱-۶
۲۷۰	 مرحله هفتم ۹-۳-۱-۷
۲۷۲	 مرحله هشتم ۹-۳-۱-۸
۲۷۳	 مرحله نهم ۹-۳-۱-۹
۲۷۴	 استفاده از CUPET به همراه Visual profiler ۹-۳-۲
۲۷۵	 خلاصه CUPET ۹-۳-۲-۱
۲۷۷	 Thrust ۹-۴
۲۷۸	 مرتب سازی با Thrust ۹-۴-۱
۲۷۸	 مرحله اول ۹-۴-۱-۱
۲۷۸	 مرحله دوم ۹-۴-۱-۲
۲۷۸	 مرحله سوم ۹-۴-۱-۳
۲۷۹	 مرحله چهارم ۹-۴-۱-۴
۲۸۰	 مرحله پنجم ۹-۴-۱-۵
۲۸۰	 مرحله ششم ۹-۴-۱-۶
۲۸۰	 مرحله هفتم ۹-۴-۱-۷
۲۸۰	 thrust همراه با visual profiler ۹-۴-۲
۲۸۱	 خلاصه ای از Thrust ۹-۴-۲-۱
۲۸۳	 پیوست ۱
۲۸۳	 واژه نامه

پیشگفتار

امروزه، پردازنده‌های گرافیکی که بر روی کارت‌های گرافیک گران‌قیمت نصب می‌شوند، توان پردازشی خارق‌العاده‌ای نسبت به پردازنده‌های مرکزی دارند. این موضوع، موجب گسترش کاربردهای این پردازنده‌ها در حوزه‌هایی فراتر از بازی‌های کامپیوتری شده است. پردازنده‌های گرافیکی مدرن با معماری موازی خود، پردازنده‌های بسیار سریعی به شمار آمده و در عین حال با قیمت و توان مصرفی کمتری، عرضه می‌شوند. جهت پیاده‌سازی الگوریتم‌ها و برنامه‌های بینایی ماشین و پردازش تصویر که بار پردازشی زیادی را جهت رسیدن به نرخ فریم بیشتر طلب می‌کنند، استفاده از پردازنده‌های گرافیکی، یک راه‌کار اقتصادی و کارآمد به شمار می‌رود. به گونه‌ای که برنامه نویس بدون نیاز به فرآیند واسطه‌های برنامه نویسی گرافیکی به کمک کتابخانه‌های موجود برای این کار می‌تواند بار پردازشی، نامۀ خود را به سادگی از پردازنده مرکزی به پردازنده گرافیکی منتقل کند.

اشتیاق فراوان طرفداران بازی‌های کامپیوتری به بازی‌های زیبا و طبیعی‌تر، موجب به‌وجود آمدن تحولات متعددی در کارت‌های گرافیک و فناوری‌های وابسته به آن شده است. به طوری که کارت‌های گرافیکی مدرن می‌توانند به درمان سرطان کمک کنند، زمین لرزه بزرگی در شهر سانفرانسیسکو را پیش‌بینی کنند و با فراهم آوردن توان پردازشی خارق‌العاده بسیار پیچیده محاسباتی و ریاضی را، به سادگی حل کنند. آیا تا کنون به این موضوع فکر کرده‌اید که کارت‌های گرافیکی چندصد هزار تومانی به غیر از بازی‌های کامپیوتری، چه کاربردهای دیگری دارند؟

سال‌ها است که افزایش روز افزون عملکرد کارت‌های گرافیکی، محققین را به فکر بهره‌گیری از توان پردازشی آنها در کاربردهای غیرگرافیکی، انداخته است. در همین راستا، شاخه جدیدی در علوم کامپیوتر به نام GPGPU ایجاد گشته که هدف نهایی فعالان این عرصه، بهره‌گیری از کارت‌گرافیک به عنوان یک کمک پردازنده محاسباتی در برنامه‌های غیرگرافیکی و برنامه‌های عمومی است.

مؤلفان

پاییز ۱۳۹۵