

Posudek oponenta bakalářské práce

Autor/autorka práce: Václav Löffelmann

Název práce: Paralelizace geometrických výpočtů pro analýzu biomolekul

Obsah práce:

Bakalářská práce Václava Löffelmanna se věnuje paralelnímu výpočtu molekulárního povrchu s využitím architektury OpenCL. Je to aktuální problém v oblasti studia molekul proteinů. Práce je přehledně dělena do 10 kapitol. V první kapitole je obecný úvod do problematiky a stanovení cílů práce. V následujících kapitolách jsou popsány aditivně vážené Voronoi diagramy, které jsou stěžejní pro analýzu biomolekuly, paralelní výpočty na současném hardwaru, architektura OpenCL a současný stav výpočtu molekulárního povrchu. Úroveň podrobností je přiměřená. Nicméně struktura práce a výklad jsou místy nelogické. Příkladem může být hned úvodní kapitola, kdy se na jejím konci dozvídáme shrnutí dosažených výsledků, což by mělo být předmětem závěru. Dále úplně nedává smysl použití pojmu bottleneck již u popisu algoritmu trasování hran, jelikož pojem bottleneck se vztahuje k problematice výpočtu molekulárního povrchu, ale z hlediska aditivně váženého Voronoi diagramu, pro jehož konstrukci algoritmus trasování hran slouží, není obecně znám, a mělo by zde tudíž být uvedeno, že dané tři sféry definují hranu aditivně váženého Voronoi diagramu, nikoliv bottleneck. Navíc autor objasňuje čtenáři význam tohoto pojmu až o několik odstavců později, než je jeho první výskyt v textu. V práci taktéž není sjednoceno názvosloví a autor volně přechází mezi používáním anglických pojmů a jejich českých ekvivalentů, aniž by na tuto skutečnost nějak upozornil. V teoretické části práce se taktéž vyskytuje několik faktických chyb. Mezi tyto chyby patří:

- V definici Voronoi regionu (1) chybí, že uvedená nerovnost musí být platná pro všechny generátory, tj. správně by měla vypadat nějak takto

$$R_i = \{\vec{x} | \forall \vec{p}_j, j \neq i: d(\vec{p}_i, \vec{x}) \leq d(\vec{p}_j, \vec{x})\},$$

to samé platí i pro definici oblasti v aditivně váženém Voronoi diagramu (2).

- V práci se vyskytují nepřesná označení objektů, např. na str. 3 autor popisuje speciální příklad Voronoi hrany, kterým je eliptická hrana, ale hned v další větě o ní mluví jako o kruhu. Na téže straně autor také popisuje, co je znázorněno na Obrázku 2, kdy o kruzích mluví jako o sférách, což je ale v případě 2D prostoru zobecněný pojem pro kružnici.
- V popisu algoritmu trasování hran (str. 5) autor uvádí, že na zásobník ukládá nově nalezené vrcholy, ale hned v další části věty z něj vybírá hrany (ty se na něj také správně ukládají).

Od šesté kapitoly je popsána autorova vlastní práce, přičemž kapitola 6 se zabývá vhodností paralelizace jednotlivých částí algoritmu pro výpočet molekulárního povrchu. Kapitola 7 popisuje detaily implementace a v kapitole 8 jsou zdokumentovány dosažené výsledky. I tato část se neobešla zcela bez chyb, jelikož v kapitole 7.1.2 autor uvádí, že sféra definující sférický trojúhelník je řešením soustavy rovnic (4), nicméně tato soustava hledá tečnou sféru ke čtyřem atomům, zatímco pro sférický trojúhelník je potřeba nalézt sféru tečnou ke třem atomům, čehož docílíme řešením soustavy rovnic (3).

Rozsah práce čítá 31 stran čistého textu a pohybuje se tak u minimální požadované délky. K práci je přiloženo doprovodné CD se zdrojovými kódy, příklady vstupů a ilustračním videem.

Kvalita řešení a dosažených výsledků:

Paralelní výpočet molekulárního povrchu byl implementován pomocí standardu OpenCL, přičemž hostitelská část programu je implementována v jazyce Java. Doprovodný program je funkční, nicméně pro větší poloměr sondy (cca od 9 Å) výpočet nefunguje nebo program přestane fungovat z důvodu nedostatečného množství zdrojů a to i pro malé molekuly. Navíc pro některé molekuly, např. pro molekulu s PDB ID 4MFD, je stanovený molekulární povrch zcela nekorektní. Autorem navržené řešení taktéž velice plýtvá s paměťovými prostředky, přičemž tato skutečnost by šla dle mého názoru poměrně jednoduše odstranit, aniž by se nějak významně zvýšila časová náročnost výpočtu.

Přiložené zdrojové kódy jsou příliš stručně komentované, což občas činí kód obtížně čitelným. Metody nemají dostatečně popsané vstupní parametry a návratové hodnoty, což platí i pro kernely OpenCL standardu, kdy v komentářích není vůbec uveden význam jednotlivých parametrů. Kernel pro výpočet ořezávacích rovin má dokonce ve zdrojovém kódu méně parametrů, než kolik jich je vyjmenováno v dokumentaci, jelikož došlo ke spojení několika parametrů do jednoho pole, aniž by tato skutečnost byla někde uvedena. Dále se ve zdrojových kódech vyskytují konstanty pojmenované `MAX_AVG_CROP_PLAMES_PER_TRIANGLE` a `MAX_TOROIDAL_PATCHES_PER_ATOM`, přičemž nikde v práci není vysvětleno, proč jim byla přiřazena uvedená hodnota, a jak k této hodnotě autor došel. Také stanovení hodnoty `BOXES_COUNT = [spTrianglesCount * 1.3 / 512] * 512` pro prostorové hashování se jeví jako velice experimentální volba a nikde se nenachází odůvodnění této hodnoty.

U porovnání implementace z hlediska časové náročnosti si nejsem jistá korektností výsledků, jelikož autor uvádí, že pro původní implementaci ze softwaru CAVER Analyst bylo měření provedeno vždy pětkrát a výsledná hodnota je průměrem naměřených časů, zatímco pro autorovu paralelní implementaci bylo měření provedeno třikrát a jako výsledná hodnota byla použita střední hodnota naměřených časů. Není mi vůbec jasné, proč byla měření provedena odlišným způsobem, a dle mého názoru tento rozdíl může značně zkreslit výsledky. Poslední má poznámka k výsledkům experimentů je, že podíly Δ_1 a Δ_2 v Tabulce 3 by bylo vhodnější stanovit v opačném poměru, aby jejich hodnota udávala urychlení navrženého řešení vůči původní implementaci.

V práci taktéž postrádám vizualizační nástroj, pomocí kterého by bylo lehce ověřitelné, zda získaný textový výstup korektně popisuje molekulární povrch.

Formální úroveň:

Práce obsahuje větší množství pravopisných chyb a překlepů. Navíc stavba většího množství vět není zcela korektní, ve větě například chybí sloveso, je v ní použito špatné skloňování či věta dokonce nedává smysl a čtenář si musí domyslet, co chtěl autor popsat. Uvedená skutečnost se samozřejmě projevuje na srozumitelnosti a text je těžko čitelný. Celkově práce působí značně uspěchaným dojmem.

Práce s literaturou:

Použitá literatura je relevantní a dostatečná pro řešení bakalářské práce.

Splnění zadání:

Zadání práce bylo splněno bez výhrad.

Dotazy k práci:

- Zkoušel jste porovnávat shodu výsledků ze softwaru CAVER Analyst a z vaší implementace jinak než vizuálně (tj. jinak než metodou „kouknu a vidím“)?
- Proč byla měření časové náročnosti implementace ze softwaru CAVER Analyst a vaší implementace provedena různým způsobem (implementace v CAVER Analyst – zprůměrování časů pěti měření, vaše implementace – střední hodnota ze tří měření)? Nejsou poté výsledky porovnání zkresleny?
- Na základě čeho byly stanoveny hodnoty konstant BOXES_COUNT, MAX_AVG_CROP_PLAMES_PER_TRIANGLE a MAX_TOROIDAL_PATCHES_PER_ATOM?

Navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni 20.7.2016

Ing. Zuzana Majdišová

Majdišová