Precomputed radiance transfer

Martin Bulant

11. dubna 2011

Reprezentace funkce na sféře

Reálnou funkci na sféře $G(x)$ aproximujeme pomocí lineární kombinace lineárně nezávislých bázových funkcí $B_i(x)$:

$$G(x) = \sum_{i=1}^{n} c_i B_i(x)$$

To nám přináší určité výhody:

- potřebujeme si pamatovat jen váhy, neboli bázové koeficienty c_i
- pro danou aproximaci funkce z nekonečnědimenzionálního prostoru si stačí pamatovat jen konečně mnoho koeficientů

Volba bázových funkcí

Bázové funkce lze zvolit mnoha různými způsoby, volba báze tedy vždy závisi na použití. Když už máme bázi, je potřeba najít koeficienty c_i pro jednotlivé bázové funkce tak, abychom co nejlépe aproximovali požadovanou funkci $G(x)$. Chceme tedy najít koeficienty c_i, aby byla chyba aproximace minimální.

Chybu definujeme jako integrál z rozdílu požadované funkce a její aproximace:

$$E_{L_2} = \int_I [G(x) - \sum_i c_i B_i(x)]^2$$

Jelikož chceme tuto chybu minimalizovat, zderivujeme funkci chyby. Získáme soustavu lineárních rovnic. Tuto soustavu lze zapsat pomocí matice, ve které se vyskytují skalární součiny bázových funkcí, označíme jí B. Tato matice je nezávislá na aproximované funkci, ale je závislá na zvolené bázi. Vektor pravých stran už na aproximované funkci závislý je.

Postup hledání keficientů

Nejprve zvolíme jakou bázi budeme používat. Když máme již zvolenou bázi, spočítáme matici B a k ní inverzní matici B^{-1}. Pro danou funkci G potom spočítáme vektor pravých stran. Vektor koeficientů dostaneme jako součin matice B^{-1} a vektoru pravých stran:

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = B^{-1} \begin{bmatrix} \langle G|B_1\rangle \\ \langle G|B_2\rangle \\ \vdots \\ \langle G|B_n\rangle \end{bmatrix}$$

Takto spočítané koeficienty minimalizují chybu aproximace naší funkce při zvolené metrice.

Výhody ortonormální báze

Pokud máme ortonormální bázi, matice B se pak stane jednotkovou, což nám výrazně zjednoduššÍ hledání koeficientů. Skalární součin naší funkce s funkcemi z báze odpovídá projekci naší funkce do prostoru báze. Jednotlivé koeficienty pak udávají, jak moc se naše funkce G(x) podobá jednotlivým funkcím z báze (čím je koeficient větší, tím podobnější si obě funkce jsou).

Další nezanedbatelnou výhodou je, že při použití ortonormální báze lze integrál ze součinu funkcí $F(x) = \sum_{i=1}^{n} f_i B_i(x)$ a $G(x) = \sum_{i=1}^{n} g_i B_i(x)$ nahradit skalárním součinem koeficientů těchto funkcí vůči této bázi, jak popisuje tento vzorec:

$$\int_S F(x)G(x)dx = \sum_{i=1}^{n} f_i g_i$$

Toto je pro nás výhodné, jelikož zobrazovací rovnice má tvar integrálu součinu.

Sférické harmonické

Sférické harmonické jsou funkce definované na sféře, tedy na směrech v třírozměrném prostoru. Obecně jsou komplexní, my se však omezíme pouze na jejich reálnou část. Jejich argument, neboli směr je možné definovat v různých souřadných soustavách (např. polární, kartézská). My budeme používat polární souřadnou soustavu. Sférické harmonické jsou uzavřené vůči rotaci, respektive lineární podprostor tvořený bázovými funkcemi sférických harmonických do nějakého řadu je uzavřen na rotaci. Sférické harmonické lze zobrazovat dvěma způsoby:

- jako orbitaly
- jako barvu na povrchu koule

Je třeba si uvědomit že pro řád n sférické harmonické funkce je třeba si pamatovat n^2 koeficientů.

Indexování sférických harmocnických

Sférické harmonické se indexují dvěma indexy, nejčastěji l a m, kde l udává řád sférické harmocnické funkce a má rozsah $\langle 0, \infty)$. Oproti tomu index m udává typ harmocnické funkce v rámci řádu a má rozsah $\langle -l, l\rangle$ pro daný řád l.

Často se však indexování zjednodušuje pouze na jeden index i, který odpovídá pořadí sférické harmonické funkce při procházení indexu l od 0 a pro každý řád l indexu m od $-l$ do l. Pro toto indexování platí:

$$i = l(l+1) + m$$

Aproximace funkce pomocí sférických harmonických

Funkce na sféře, může být aproximována pomocí sférických harmonických. Z praktických důvodů se používá dvojitá suma pro indexy l a m, místo jednoduché sumy pro index i. Pro aproximovanou funkci při použítí polárních souřadnic platí:

$$G(\theta, \varphi) = \sum_{l=0}^{n-1} \sum_{m=-l}^{m=l} c_{l,m} Y_{l,m}(\theta, \varphi)$$

Bohužel pro složité funkce je nedostačující i řád 10, který odpovídá 100 koeficientů sférických harmonických. Jak už jsme uvedli, sférické harmonické tvoří ortonormální bázi.

Výpočet koeficientů $c_{l,m}$ se provádí pomocí integrace součinu aproximované funkce s funkcemi bázickými přes celou sféru. Platí pro ně vzorec:

$$c_{l,m} = \int_0^{2\pi} \int_0^{\pi} G(\theta, \varphi) Y_{l,m}(\theta, \varphi) \sin\theta d\theta d\varphi$$

Většinou se tyto koeficienty počítají pomocí numerické integrace, např. pomocí Monte Carlo integrace.

Renderování v reálném čase - real-time rendering

Chceme v reálném čase zobrazovat scény s realistickým osvětlením, stíny a globálním osvětlením. Přičemž reálným časem je myšleno, spočítat výsledný obrázek za zlomek sekundy. Bohužel pro současné algoritmy je to nemožné, je potřeba provést příliš velké množství výpočtů během velmi krátkého časového úseku. To se řeší dvěma různými přístupy.

- Prvním je omezení požadavků, například osvětlení přicházející z nekonečna, nezabývání se viditelností.

- Druhým je rozdělení výpočtu na 2 části, předvýpočet a samotný výpočet obrázku.

Mapování prostředí - Environment mapping

V tomto přístupu se použije mnoho předpokladů, například, že světlo přichází z nekonečna, neovlivňuje ho viditelnost, máme jen zrcadlové povrchy atd. Za použití těchto předpokladů je pak rendering triviální.

Pokud však naše předpoklady omezíme tím, že povolíme i difúzní a lesklé povrchy, začnou se objevovat komplikace. Pro Phongův světelný model se do výsledného směru ω_o neodráží jen světlo ze směru R, tedy zrcadlový odraz, ale i nějaké světlo z okolí dle laloku BRDF. Při vyhledávání hodnoty v mapě prostředí se jako index použije vektor R.

To lze vyřešit tak, že se použitá mapa už předem rozmaže pomocí filtru vycházejícího z dané BRDF. I toto řešení má bohužel vady, pro každý exponent je třeba mapu rozmazat jinak. Dalším závažnějším problém ale je, že pro každý pixel rozmazané mapy musím nějak váženě zkombinovat polovinu pixelů té mapy. Celá polovina je to proto, že BRDF je nenulová na jedné polovině sféry. To ovšem znamená, že rozmazávání mapy je velmi náročné.

Pokud máme difúzní povrch, rozmazává se jádrem, které odpovídá Phonogovu osvětlovacímu modelu s exponentem rovným jedné. V tomto případě se narozdíl od phonga pro vyhladávaní v mapě prostředí použije jako index normálový vektor N (protože difúzní osvětlení je nezávislé na úhlu pohledu).

Odbočka o zvláštním využití difúzní koule

Pokud je potřeba do filmu umístit nějakou umělou postavu tak, aby měla správné osvětlení, tak se při natáčení na místo, kde bude ve filmu umělá postava umístí difúzní koule. Po nafilmování se pak nastavují světla v animačním programu pro difúzní kouli tak, aby byl výsledek stejný jako ta nafilmovaná koule. Takto nastavenými světly se pak osvítí vkládaný model.

Mapování prostředí pomocí sférických harmonických

Mapu prostředí budeme reprezentovat pomocí sférických harmonických. Hodnota iradiance je funkcí normály, lze ji tedy reprezentovat vůči bázi sférických harmonických. Abych dostal koeficienty iradiance, stačí koeficienty mapy prostředí přenásobit konstantami A_l.

Konstanty A_l mají malou hodnotu pro velká l. To znamená, že po vynásobení touto konstantou uřežeme vysoké frekvence, což nám ale nevadí, jelikož chceme zpracovat difuzní povrch. Když vezmeme toto v potaz, tak nám stačí použít pouze 9 koeficientů sférických harmonických pro reprezentaci mapy prostředí. Dá se ukázat, že pro difuzní povrch je chyba výsledné radiance při 9 koeficientech menší než 3 procenta. To je díky tomu, že difúzní povrch funguje jako filtr, v tomto případě nízkofrekvenční propusť.

Algoritmus

Algoritmus se skládá ze dvou částí, první z nich je předvýpočet. Ten se provádí při každé změně osvětlení, v našem případě při změně mapy prostředí. Nejprve si najdeme projekci mapy prostředí do prvních devíti bází sférických harmonických a poté se tyto koeficienty přenásobí konstantou A_l, čímž dostaneme koeficienty iradiance jako funkce normály. Druhou částí je samotný výpočet, osvětlení. Pro kazdý pixel potřebujeme vyhodnotit iradianci pro danou normálu. Díky tomu, že sférické harmonické do koeficientu 9 jsou maximálně kvadratické funkce, lze jí spočítat jen pomocí maticového násobení a skalárního součinu.

Další možné vylepšení

Chceme použít libovolnou BRDF, máme mapu prostředí a stále nás nezajímá viditelnost vzhledem k této mapě. BRDF i mapu prostředí reprezentujeme pomocí sférických harmonických. Toto má ale drobnou vadu, kterou je to, že BRDF není funkce definovaná na sféře, protože má 2 parametry. Zafixujeme si tedy odchozí směr, čímž se nám z BRDF stane funkce definovaná na sféře. Odchozí směr si diskretizujeme a spočítáme si koeficienty BRDF pro daný směr vůči sférickým harmonickým a tento vektor si uložíme do textury. Pro každý odchozí směr tedy máme jeden vektor koeficientů.

Výpočet hodnoty pixelu

Integrál ze součinu mapy prostředí a BRDF lze nahradit skalárním součinem vektorů koeficientů vůči bázi sférických harmonických. Je třeba si ovšem dávat pozor, vůči kterému souřadnému systému máme definice. Je třeba souřadné systémy zarovnat, to se provede pomocí rotace sférických harmonických. Pro provedení této operace se využije toho, že sférické harmonické jsou uzavřené na rotaci.

Precomputed radiance transfer

Je to druhá cesta, jak rychle počítat osvětlení. V tomto případě nezmírňujeme požadavky, chceme tedy komplexní osvětlení včetně stínů a globálního osvětlení. Ale i tak musíme použít určité předpoklady, je nutné zafixovat geometrii, materiál a pro některé techniky také směr pohledu.

Příkladem je Daytime relighting, kdy chceme mít scénu osvětlenou v různých denních dobách. To se řeší tak, že se vypočítá osvětlení scény v předem určených časových okamžicich a v ostatních případech se ošetlení intepoluje z těchto přepočítaných obrázků.

Transport světla

Na EM se lze dívat jako na vektor pixelů. Obdobně i na výsledek, například obrázek, se lze dívat jako na vektor pixelů (v real-time grafice barev vrcholů). Transport světla je lineární a lze ho tedy ho zapsat maticí. Označme ji M, je to předpočítaná matice přenosu osvětlení ve scéně. Lze do ní zakódovat stíny, globální osvětlení a další.

Výpočet matice transportu světla M

Matici lze spočítat dvěma způsoby:

- po sloupcích — to znamená, že celý obrázek je osvětlený jen jedním pixelem EM
- po řádcích — to znamená, že jeden bod obrázku je osvětlený celou EM

Tento přístup má ale problém s příliš velkou pracností, takováto matice by byla obrovská a ani maticové násobení s ní by nebylo upočitatelné, natož její výpočet. Řeší se různými způsoby, jením z nich je SH-based PRT.

SH-based PRT

Je to předpočítaný přenos radiance, který využívá reprezentaci funkce na sféře pomocí sférických harmonických. Rozlišují se zde 2 pojmy:

- Source radiance, což je radiance samotné environment mapy.
- Transfered Incident radiance, což je radiance přicházející do bodu, pokud započítáváme stíny a odrazy. Pokud bychom stíny a odrazy zanedbali, byla by stejná jako Source radiance.

Algortimus funguje tak, že si předpočítám vzhled objektu při osvětlení jednotlivými bázemi SH. Pro skutečnou mapu prostředí si zjistím její koeficienty vůči zvolené bázi SH a výsledný obrázek spočítám váženou lineární kombinaci předpočítaných vzhledů pro jednotlivé báze.

Zvolením vhodné báze pro reprezantaci osvětlení jsme řádově zmenšili potřebnou dimenzi. Tento postup lze použít pro nasvětlování exteriéru při různých světelných podmínkách (změny počasí, změny denní doby, atd.)

Light mapa je extrémní případ PRT. Když se chci dívat z různých úhlů, budu si ukládat více hodonot, pro každou bázi SH jednu.

Reference

[1] Křivánek, J.: Function Approximation & Spherical Harmonics. 2011. URL <http://cgg.mff.cuni.cz/~jaroslav/teaching/npgr031/07-stgi-sh.pdf>

[2] Křivánek, J.: Pre-computed Radiance Transfer. 2011. URL <http://cgg.mff.cuni.cz/~jaroslav/teaching/npgr031/07-stgi-prt.pdf>