Many light methods

Martin Kahoun

17. března 2011

Úvod

Pro připomenutí a úplnost uvedeme dva ekvivalentní zápisy zobrazovací rovnice. Představme si bod $\mathbf{x}$ na povrchu tělesa ve scéně a chceme vyjádřit celkovou radianci $L_o(\mathbf{x}, \omega_o)$ opouštějící tento bod ve směru ω_o. Bod $\mathbf{x}$ může být sám zdrojem světla a emitovat tak radianci $L_e(\mathbf{x}, \omega_o)$ v daném směru. Dále bude odrážet radianci

$$L_r(\mathbf{x}, \omega_o) = \int_\Omega f_r(\mathbf{x}, \omega_i \to \omega_o)\, L_i(\mathbf{x}, \omega_i)\, \cos\theta_i \,\mathrm{d}\omega_i$$

v daném směru ω_o, jíž spočteme integrováním přes celou polokouli (všechny směry ω_i, θ_i je pak úhel svíraný normálou povrchu v bodě $\mathbf{x}$ a směrem ω_i, $f_r(\mathbf{x}, \omega_i \to \omega_o)$ je BRDF). Dostáváme tak následující rovnici:

$$L_o(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + L_r(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + \int_\Omega f_r(\mathbf{x}, \omega_i \to \omega_o)\, L_i(\mathbf{x}, \omega_i)\, \cos\theta_i \,\mathrm{d}\omega_i \qquad (1)$$

Tuto rovnici lze též přeformulovat jako integrál přes povrch scény $\mathcal{M}$ místo integrace přes polokouli:

$$L_o(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + \int_\mathcal{M} f_r(\mathbf{y} \to \mathbf{x} \to \omega_o)\, L_i(\mathbf{y} \to \mathbf{x})\, V(\mathbf{y} \leftrightarrow \mathbf{x})\, G(\mathbf{y} \leftrightarrow \mathbf{x}) \,\mathrm{d}A_\mathbf{y} \qquad (2)$$

kde $V(\mathbf{y} \leftrightarrow \mathbf{x})$ značí viditelnost mezi body $\mathbf{y}$ a $\mathbf{x}$ (buď 1 nebo 0) a G je geometrický faktor definovaný:

$$G(\mathbf{y} \leftrightarrow \mathbf{x}) = \frac{\cos\theta_\mathbf{x} \cdot \cos\theta_\mathbf{y}}{\|\mathbf{y} - \mathbf{x}\|^2}$$

Platí $\mathrm{d}\omega_i = V(\mathbf{y} \leftrightarrow \mathbf{x}) \frac{\cos\theta_\mathbf{y}}{r^2} \mathrm{d}A$

Instantní radiozita

Ukážeme si nyní metodu publikovanou Kellerem v roce 1997, kterou nazval instatní (okamžitá) radiozita [2]. Celá třída metod odvozených z instantní radiozity je pak často nazývána *many-lights methods* nebo *virtual point light methods*. Metoda je v základu tzv. *biased* a byla uzpůsobena k výpočtu na GPU v roce 1997. Tehdy bylo k dispozici velmi málo prostředků jak pracovat s grafickými čipy a autoři se rozhodli využít akumulačního bufferu a mnohonásobného renderování scény s bodovým osvětlením, pomocí kterého aproximovali nepřímé osvětlení.

Metoda je dvoukroková. V prvním kroku se generují virtuální světla ve scéně. Jde o prakticky stejný přístup, jako když se generují fotonové mapy: dle hustoty pravděpodobnosti generujeme paprsky vysílané ze světel ve scéně. Na difúzních plochách vytvoříme virtuální bodový zdroj světla a dále sledujeme cestu paprsku. Tímto vznikne seznam virtuálních světel, kterým doplníme reálné osvětlení scény. Každý virtuální zdroj považujeme za čistě difúzní emitor a pro radianci přicházející od něj tedy platí:

$$L(\mathbf{y} \to \mathbf{x}) = \Phi \frac{\rho_d(\mathbf{y})}{\pi} \cos\theta_\mathbf{y},$$

kde Φ je "tok" nesený časticí, která vygenerovala virtuální světlo (výpočet Φ viz. photon tracing ve fotonových mapách). Dále $\rho_d(\mathbf{y})$ je difúzní odrazivost (albedo) v bodě $\mathbf{y}$.

Ve druhém kroku počítáme pomocí tohoto seznamu virtuálních světel přímé osvětlení scény a to následujícím způsobem:

$$L(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + \sum_{k=1}^{n} L(\mathbf{y}_k \to \mathbf{x})\, V(\mathbf{y}_k \leftrightarrow \mathbf{x})\, G(\mathbf{y}_k \leftrightarrow \mathbf{x})\, f_r(\mathbf{y}_k \to \mathbf{x} \to \omega_o) \tag{3}$$

Tato rovnice není ničím jiným než Monte Carlo estimátorem zobrazovací rovnice ve tvaru (2). Virtuální bodové zdroje pak představují vzorky radiance na plochách scény. Instantní radiozita je tedy podmnožinou obousměrného sledování cest (*bidirectional path tracing*), používáme ovšem jen první dopad paprsku z kamery a navíc vrcholy cest od světelných zdrojů (tj. virtuální bodové zdroje) se přepoužívají pro všechny pixely.

Pro výpočet viditelnosti mezi jedním bodovým zdrojem a všemi viditelnými body scnény lze s výhodou využít algoritmu *stínových map (shadow maps)*. Stínové mapy fungují ve dvou krocích: Nejprve vyrenderujeme scénu z pohledu světla a do bufferu ukládáme vzdálenost světelného zdroje a objektu scény, čímž vytvoříme stínovou mapu. Ve druhém průchodu renderujeme už normálně z pohledu kamery. V každém bodě scény se pak dotazujeme do stínové mapy a porovnáváme vzdálenost bodu od světla s hodnotou ve stínové mapě: pokud je vzdálenost bodu větší než hodnota ve stínové mapě, pak je bod zastíněn, jinak je osvětlen.

Stínové mapy trpí několika poměrně vážnými nedostatky. Buffer obsahuje po částech konstatní funkci a my musíme definovat od kterých hodnot je to stín a od kterých už není. Z toho plyne zubatost okrajů stínů či světlé fleky na zastíněných plochách. Rovněž si musíme pohlídat nelineární mapování hloubky. Tyto nedostatky se pak v praxi řeší manuálním nastavováním parametrů, které je nutné vyladit pro každou scnénu.

Instantní radiosita je vhodná pro difúzní scény, kde nám postačí zhruba 1000–2000 světel pro relativně dobré výsledky. Zároveň slouží jako základ pro *real-time* metody globálního osvětlení.

Problémy instatní radiozity

Podobně jako *bidirectional path tracing* trpí i instatní radiozita šumem. Zde je ovšem korelovaný mezi jednotlivými pixely a tvoří shluky s přibližně stejnou intenzitou jasu, což je způsobeno recyklací světelných cest pro sousední pixely. Tyto shluky se z valné části projevují jako světlá místa (*highlights*) a skutečně nepříjemně působí ve scénách s lesklými povrchy.

Eliminace problémů v difuzních scénách

Příčiny vzniku těchto artefaktů jsou vesměs dvě. Buď může být na vině BRDF, pokud lalok spekulární složky míří stejným směrem jako světlo. Druhou příčinou může být geometrický faktor G. Pokud je totiž vzdálenost mezi osvětlovaným bodem a virtuálním světlem příliš nízká, pak $G \to \infty$ a tato divergence má za následek přesvětlení bodu $\mathbf{x}$. Geometrický faktor dokonce vede ke vzniku singularit v rovnici (3). Lze je však poměrně snadno odstranit, pokud nahradíme součin $G \cdot f_r$ z rovnice (3) následujícím výrazem:

$$\min\{c, G(\mathbf{y} \leftrightarrow \mathbf{x}) f_r(\mathbf{y}_k \to \mathbf{x} \to \omega_o)\}$$

Tedy, slovně vyjádřeno, daný výraz shora omezíme (ořízneme - *clamping*) na nějakou vhodně zvolenou konstantu c. To má ovšem za následek ztrátu nestrannosti, jelikož tento zásah zavádí systematickou chybu (*bias*) a rovnice (3) přestává být nestranným Monte-Carlo estimátorem; důsledkem čehož jsou tmavá místa tam, kde by byl šum.

Nicméně míra námi zavedené chyby se dá pohodlně vyčíslit a kompenzovat, protože jediné co děláme, je ignorování části energie jdoucí od světel ke kameře. Pokud tedy integrujeme přes plochu scény $\mathcal{M}$

$$\int_{\mathcal{M}} L_i(\mathbf{y} \to \mathbf{x})\, V(\mathbf{y} \leftrightarrow \mathbf{x})\, \min\{c, G(\mathbf{y} \leftrightarrow \mathbf{x})\, f_r(\mathbf{y} \to \mathbf{x} \to \omega_o)\}\, \mathrm{d}A_{\mathbf{y}} \tag{4}$$

můžeme ořezávací funkci vyjádřit též jako hodnotu, jíž vyvážíme dotyčný integrál:

$$w_1 = \min\left\{1, \frac{c}{G(\mathbf{y} \leftrightarrow \mathbf{x}) \cdot f_r(\mathbf{y}_k \to \mathbf{x} \to \omega_o)}\right\}$$

Po dosazení do (4) dostaneme výraz pro výpočet osvětlení s opravou singularit (pro větší čitelnost vynecháváme parametry funkcí):

$$L_w = \int_{\mathcal{M}} L_i \cdot V \cdot G \cdot f_r \cdot w_1 \cdot \mathrm{d}A_{\mathbf{y}}$$

Energii, kterou ztratíme pak budeme kompenzovat pomocí integrace celé polokoule nad bodem $\mathbf{x}$:

$$L_{w'} = \int_{\Omega} L_i \cdot f_r \cdot \cos\theta \cdot (1 - w_1) \cdot \mathrm{d}\omega$$

Když tedy oba výrazy dáme dohromady dostaneme rovnici instantní radiozity, která zahrnuje opravu na šum a zároveň kompenzaci tím zanesené chyby:

$$L_o = L_w + L_{w'} = \int_{\mathcal{M}} L_i \cdot V \cdot G \cdot f_r \cdot w_1 \cdot \mathrm{d}A_{\mathbf{y}} + \int_{\Omega} L_i \cdot f_r \cdot \cos\theta \cdot (1 - w_1) \cdot \mathrm{d}\omega$$

Shrňme si tedy výše popsaný přístup (podrobně pospaný v [3, str. 245-257]). Rovnice (3) je Monte-Carlo estimátor s vysokou variancí. Tu snížíme zavedením členu w_1, díky čemuž získáme L_w, zde se nám ovšem ztratila nějaká energie. Tuto ztrátu kompenzujeme pomocí $L_{w'}$, což nastává typicky v rozích. Při implementaci, bychom se podívali na hodnotu w_1 a v případě, že se nerovná 1, bychom zavolali *path tracer*, aby dopočítal ztracenou energii. Tento lze navíc omezit, aby paprsky nevrhal dále než nějaká vzdálenost od bodu $\mathbf{x}$, která se dá spočíst ze znalosti konstanty c. Tato metoda funguje rychle pro difúzní a mírně lesklé scény.

Eliminace problémů v lesklých scénách

Chceme-li eliminovat vznik *highlights* ve scénách s lesklými povrchy, záhy zjistíme, že výše popsaná metoda ořezávání a kompenzace, kterou použil Keller ve své implementaci, nefunguje (viz [3, str. 245-257])[1].

Je tedy nutné poohlédnout se po jiném řešení. Tím je přechod od bodových zdrojů světla k plošným - konkrétně sférickým, odtud plyne název *virtual spherical lights* [1]. Představme si scénu, ve které chceme vyhodnotit radianci v bodě $\mathbf{x}$ přicházející od virtuálního světla na nějakém dalším povrchu nacházejícím se v bodě $\mathbf{p}$. Předpokládejme, že toto světlo rozprostírá svoji energii na všechny povrchy scény uvnitř koule o poloměru r okolo pozice světla $\mathbf{p}$. Tato energie je následně transportována do bodu $\mathbf{x}$ (ve své podstatě jde o analogii fotonových map). Pak můžeme radianci v bodě $\mathbf{x}$ způsobenou virtuálním sférickým světlem v bodě $\mathbf{p}$ vyjádřit následovně:

$$L_o(\mathbf{x}, \omega_o) = \frac{\Phi}{\pi r^2} \int_{\Omega} f_r(\mathbf{x}, \mathbf{l} \to \omega_o) \cos\theta_{\mathbf{x}}\, f_r(\mathbf{y}, \omega_i \to -\mathbf{l})\, (\|\mathbf{p} - \mathbf{y}\| < r)\, \mathrm{d}\mathbf{l} \tag{5}$$

Ω zde odpovídá kuželu s vrcholem v bodě $\mathbf{x}$ a základnou obepínající světelnou kouli o poloměru r se středem $\mathbf{p}$. Φ je zářivý tok světla, $f_r(\mathbf{x}, \mathbf{l} \to \omega_o)$ je BRDF v bodě $\mathbf{x}$ ze směru $\mathbf{l}$ (všechny směry omezené kuželem přes které integrujeme) do směru k pozorovateli. Rovněž $f_r(\mathbf{y}, \omega_i \to -\mathbf{l})$ je BRDF v bodě $\mathbf{y} = ray(\mathbf{x}, \mathbf{l})$, který se nachází na povrchu uvnitř sféry a světlo na něj dopadá ze směru ω_i (tj. ze směru incidence částice, která vygenerovala virtuální světlo) a je odráženo ve směru $-\mathbf{l}$. $\theta_{\mathbf{x}}$ odpovídá úhlu sevřenému normálou bodu $\mathbf{x}$ a směrovým vektorem $\mathbf{l}$ a konečně $(\|\mathbf{p} - \mathbf{y}\| < r)$ je výraz nabývající 1 nebo 0, pokud se paprsek strefí na povrch uvnitř koule nebo nikoliv.

Implementace rovnice (5) by však byla výpočetně náročná (vrhání paprsků pro nalezení bodu $\mathbf{y}$) a proto zavedeme několik předpokladů pro její zjednodušení:

[1] Vzniká nám zde další *bias* způsobený tím, že virtuální bodové zdroje jsou považováný za difúzní emitory. Správně bychom do radiance emitované virtuálním bodovým zdrojem měli zahrnout spekularní složku BRDF v bodě $\mathbf{y}_k$. Tu instatní radiosita ignoruje, aby zabránila vzniku artefaktů v obrázku.

- Viditelnost v kuželu Ω definujeme jako viditelnost na spojnici $\mathbf{x}$ a $\mathbf{p}$.
- Normálu a hodnotu BRDF pro všechny povrchové body uvnitř sféry aproximujeme normálou a BRDF v bodě $\mathbf{p}$.
- Funkci ($\|\mathbf{p} - \mathbf{y}\| < r$) nahradíme faktorem $\cos\theta_{\mathbf{p}}$, který ošetří případy kdy nám paprsek protne sféru, ale dopadne na povrch mimo ni.

Díky těmto předpokladům můžeme zkonstruovat mnohem jednodušší rovnici pro výpočet radiance z povrchového vzorku při použití *virtual spherical light*:

$$L_o(\mathbf{x}, w_o) = V(\mathbf{x} \leftrightarrow \mathbf{p}) \frac{\Phi}{\pi r^2} \int_{\Omega} f_r(\mathbf{x}, \mathbf{l} \to \omega_o) \cos\theta_{\mathbf{x}} \, f_r(\mathbf{p}, \omega_i \to -\mathbf{l}) \cos\theta_{\mathbf{p}} \, \mathrm{d}\mathbf{l} \tag{6}$$

Použití rovnice (6) ve scéně s lesklými povrchy odstraňuje nepříjemné odlesky, kterými trpí instantní radiozita. Jistou daní je ovšem rozmazání zrcadlových obrazů. Metoda je nicméně konzistentní k počtu světel a zvýšíme-li tedy jejich počet, zlepší se i vizuální kvalita zrcadlových obrazů.

Reference

[1] Hašan M., Křivánek J., Walter B., Bala K., *Virtual Spherical Lights for Many-Light Rendering of Glossy Scenes*, Proc SIGGRAPH Asia '09.

[2] Keller Alexander, *Instant radiosity.* Proc. SIGGRAPH '97.

[3] Kollig Thomas, Keller Alexander, *Illumination in the Presence of Weak Singularities*, Monte Carlo And Quasi-monte Carlo Methods, 2004.