Matematicko-fyzikální fakulta Univerzity Karlovy v Praze
Obor: Informatika
Specializace: Softwarové systémy – počítačová grafika

Směrový přístup v radiačních metodách

Diplomová práce

Radim Halíř

duben 1993

Vedoucí práce: RNDr. Josef Pelikán

Prohlašuji, že jsem tuto práci vypracoval samostatně a použil pouze uvedené literatury.
Souhlasím se zapůjčováním této práce.

V Praze dne 6. května 1993 Radim Halíř

Tato práce je věnována památce mého dědečka pana Karla Halíře

Obsah

Seznam obrázků

Kapitola 1

Úvod

V následující práci bude popsán nový přístup k výpočtu osvětlení scény pomocí tzv. radiační metody. Jedná se o zachycení směru osvětlení v rámci standardního radiačního výpočtu, kdy mluvíme o tzv. směrových radiačních metodách.

Po stručném seznámení s principy zobrazování umělých scén bude nejprve podána charakteristika základní radiační metody a poté uveden stručný přehled jejich nejdůležitějších modifikací. Dále se již budeme věnovat směrovému přístupu v radiačních metodách. Nejprve si uvedeme hlavní rysy dosud publikovaných metod, pak již bude následovat popis vlastní metody. Hlavní pozornost bude věnována reprezentaci směru propagované světelné energie v rámci vyzařovacích funkcí a způsobu, jak tyto funkce efektivně a dostatečně přesně realizovat. V práci však budou popsány i ostatní části směrové radiační metody počínaje reprezentací scény a světelným modelem přes použití ray-castingu při aktualizaci jednotlivých vyzařovacích funkcí až po kompletní algoritmus celého radiačního výpočtu. Metoda bude porovnána s ostatními dosud publikovanými směrovými metodami, v závěru pak budou uvedeny některé její modifikace a další možná vylepšení.

Ještě dříve, než se dostaneme k vlastnímu popisu navrhnuté směrové metody, je mou milou povinností poděkovat několika lidem, bez jejichž pomoci by tato práce stěží vznikla. V první řadě bych chtěl poděkovat RNDr. Josefu Pelikánovi za vedení této diplomové práce i za mnohé cenné podněty, rady a kritické připomínky. Za informace ohledně ray-tracingu a realistických metod zobrazování děkuji svým přátelům, studentům Josefu Vašicovi a Radomíru Měchovi, za mé matematické znalosti pak doc. RNDr. Janě Staré. Velký dík patří v neposlední řadě i mým rodičům za jejich podporu v průběhu celého studia.

Kapitola 2

Metody 3D renderingu, jejich princip, výhody a nevýhody

Při generování "reálných" obrázků uměle vytvořených scén (tzv. *image synthesis* respektive *3-d rendering*) je nutné vyjádřit a poté zobrazit vzájemné světelné poměry uvnitř dané scény. Je třeba popsat šíření světla v rámci pozorovaného prostředí, neboť právě toto určuje výsledný vjem, který získáme při pozorování skutečné scény. Popisem tohoto šíření se zabývají tzv. světelné modely (*illumination or shading models*), které se dle toho, jak danou scénu postihují, dělí na dvě základní kategorie:

- lokální světelné modely
- globální světelné modely

Lokální světelné modely postihují pouze šíření světla od primárních světelných zdrojů. Poté, co toto světlo dopadne na libovolné těleso uvnitř scény, není již dále propagováno. Oproti globálním modelům neposkytují příliš realistické obrázky, výhodou je však jejich rychlost, která umožňuje generaci obrazových sekvencí v reálném čase i na osobních počítačích. Základním představitelem této kategorie je Phongův světelný model [Pho75], dokonalejší (a také složitější) modely byly popsány např. v pracech [Bli77] [TS67] [Str90]).

Při potřebě generovat výstup v kvalitě fotografií je nutné použít některou z metod druhé kategorie. Při tomto přístupu se dané osvětlené plošky stávají sekundárními zářiči, ze kterých je světlo opět propagováno danou scénou. Tyto světelné modely, poskytují vynikající výsledky, generované obrázky jsou často nerozeznatelné od fotografií. Za kvalitu je ovšem nutné zaplatit, v tomto případě obrovskými časovými nároky. Složitost základních algoritmů má často exponenciální závislost vzhledem k počtu objektů ve scéně, je nutné používat nejrůznějších urychlovacích metod. Vlastní modely propagace světla vychází a nejnovějších fyzikálních poznatků (např. polarizace, útlum světla v prostředí, difrakce, interference či směrová difúze) a jsou často zdokonalením výše uvedených lokálních modelů. Mimo jiné umožňují i stochastické modelování povrchu těles (pomocí tzv. *textur*), zobrazení objektů rozmazaných pohybem (*motion blur*) a další "reálné" efekty.

Globální světelné modely nacházejí v dnešní době uplatnění nejenom v oblasti počítačového umění (*computer art*), ale i při výrobě animovaných sekvencí ve filmech či reklamách (3D studio firmy AUTODESK, Persistance of Vision Raytracer, Imagine atd.) nebo dokonce celých filmů (systém RENDERMAN), profesionálních leteckých a automobilových simulátorů, zkvalitnění výstupu z CAD produktů (např. program ARCHICAD pro architektonické aplikace) či v oblasti tzv. virtuální reality. I přes obrovské časové nároky se dnes již objevují systémy schopné generace obrázků nebo celých sekvencí v reálném nebo skoro reálném čase (jeden obrázek za několik minut až 20 obrázků za sekundu), některé dokonce s HW podporou těchto modelů.

Dle přístupu k výpočtu propagace světla uvnitř scény lze používané metody rozdělit do následujících tří kategorií:

- metody zpětného sledování paprsku (*ray-tracing*)
- radiační metody (někdy též termodynamický přístup, *radiosity*)
- kombinované metody

2.1 Ray-tracing

Předchůdcem této metody byl postup používaný optiky při návrhu složitých optických zařízení, kdy se navrhované zařízení simulovalo počítačovým modelem. V rámci této simulace byly skrze dané optické elementy vrhány světelné paprsky, jejichž průchod zařízením se sledoval. Jednoduchá modifikace umožnila pak tento postup použít pro generování velice věrných obrázků umělých scén (počátkem 80. let). Princip je následující: Pokud chceme vygenerovat pohled na nějakou scénu, potřebujeme vlastně určit barvu každého pixelu (tj. základního dále nedělitelného elementu) daného pohledu. Tuto barvu lze určit tak, že z místa, kde je umístěn pozorovatel, vyšleme skrze daný pixel v rovině pohledu paprsek a sledujeme, na jaký objekt ve scéně poprvé narazí. Pokud paprsek nenarazí na žádný objekt, přiřadíme danému pixelu barvu pozadí, jinak je barva určena pomocí použitého světelného modelu z charakteristiky zasaženého tělesa.

Budeme-li uvažovat, že se paprsek po zásahu tělesa dále nešíří, pixelu přiřadíme přímo barvu daného tělesa (mluvíme pak o tzv. vrhání paprsku *ray-casting*). Při tvorbě reálných obrázků je však nutné provádět tzv. sledování paprsku vyššího řádu (*ray-tracing*), kdy do úvahy bereme i sekundární paprsky odvozené z odrazivosti a průhlednosti zasažených těles. Z jednoho paprsku po nárazu na těleso vznikají obecně dva nové paprsky (odražený a lomený), jejichž šíření scénou dále sledujeme. Tak dochází k rekurzivnímu opakování základního postupu, dokud daný právě sledovaný paprsek nezasáhne světelný zdroj, neopustí scénu či dokud není překročena hranice, kolikrát se může původní primární paprsek dělit. Příspěvky od všech sekundárních paprsků (jejichž počet roste exponenciálně s řádem metody) pak ovlivňují barvu daného pixelu.

Výhodou uvedené metody je (při použití odpovídajících lokálních světelných modelů) dokonalá simulace lesklých a vyleštěných povrchů, zrcadel, dojem skutečných materiálů (pomocí textur, např. struktura dřeva, omítky či mramoru), dokonalé postižení tvaru těles (použití CSG reprezentace a spline-ploch), možnost zobrazení pohybem rozmazaných těles, dojem hloubky obrazu atd.

Základní nevýhodou je použitý na generovaném pohledu závislý (*wiew dependent*) přístup, kdy při jakékoliv změně pozice pozorovatele nebo směru pohledu je nutné vygenerovat celý obrázek znovu (existují ovšem modifikace, které se snaží využít při animaci úplné či částečné invariance dané scény). Tuto nevýhodu ještě umocňuje lineární závislost doby generace obrázku na jeho rozlišovací schopnosti. Metoda sama předpokládá, že scéna obsahuje pouze reflexní plochy. V řadě aplikací je generovaný kontrastní obraz vítaný, pro scény s opticky difúzními plochami osvětlenými plošnými zdroji však ray-tracing selhává.

Principiálním nedostatkem metody zpětného sledování paprsku je její základ v čistě geometrickém přístupu, který znemožňuje korektní postihnutí propagace světla ve scéně (rušivá je hlavně tvorba "ostrých" stínů). Existují modifikace, který tento nedostatek redukují na minimum, jedná se však často pouze o časově velice náročné metody používající "brutální sílu" (*brute force*): Většinou je místo jednoho paprsku používána celá skupina těchto paprsků, s nimiž se pracuje obdobným způsobem jako s původním paprskem (uveďme např. *distributed ray-tracing*, *Monte-Carlo ray-tracing* či *path ray-tracing*). Jedná se o značně umělá řešení, poskytující odpovídající výsledky za cenu řádového nárůstu výpočetní složitosti jednotlivých obrázků i celých animovaných sekvencí.

Vývoj základních algoritmů ray-tracingu je dnes již prakticky ukončen, současné práce se zaměřují na následující směry:

- postižení tvarově složitých těles (pomocí spline-ploch či NURBS)
- zvýšení kvality generovaných obrázků pomocí plošných nebo prostoro-

vých textur (simulace struktury povrchu reálných těles), antialiasingu, rozmazání pohybujících se těles či vytvoření dojmu hloubky obrazu

- použití dokonalejších světelných modelů
- urychlení výpočtu průsečíků paprsek – těleso lepší prostorovou organizací scény či reformulací algoritmů pro výpočet těchto průsečíků
- kombinace ray-tracingu s radiačními metodami

I přes výše uvedené nedostatky je metoda ray-tracingu a její modifikace nejpoužívanějším prostředkem pro generování fotorealistických obrázků v současnosti. O metodě existuje v dnešní době již nepřeberné množství článků a knih, uveďme si alespoň základní práci [Whi80] či výborný přehled [Gla91].

2.2 Radiační metody

Většina ploch v reálných scénách se chová jako difúzní zářiče, to znamená, že dopadající světelnou energii rozptylují bez ohledu na to, odkud daná energie přichází, do všech směrů v rámci polosféry nad danou plochou. Tuto vlastnost neumí výše uvedená metoda zpětného sledování paprsku korektně postihnout. Proto vznikly konkurenční přístupy k výpočtu globálního osvětlení zadané scény, společně nazývané jako radiační metody.

Tyto metody vychází, stejně jako ray-tracing, z postupu používaného fyziky. Tentokrát se jedná o výpočet šíření sálavého tepla v uzavřené soustavě, kdy je zadána skupina těles s uvedenými počátečními teplotami a úkolem je určit rozložení teploty ve scéně po tepelné stabilizaci celé soustavy [SH78] [SC78]. Modifikaci tohoto postupu, kdy teplo vyzařované a pohlcované jednotlivými tělesy bylo nahrazeno světlem, navrhla pro výpočet osvětlení scény v polovině 80. let Cindy M. Goralová [GTGB84]. V jistém smyslu se jedná o globální postihnutí lokálního Phongova modelu. V rámci tohoto přístupu je celá scéna chápána jako od okolí izolovaná soustava plošek, z nichž některé (světelné zdroje) na začátku vyzařují jistou světelnou energii a chovají se jako primární světelné zářiče. Energie je stejně jako v případě tepelného přenosu šíří soustavou a dopadá na jiné plošky, kde je částečně pohlcována a částečně odrážena zpět do scény (tyto plošky se chovají jako sekundární světelné zářiče). Přenos světelné energie scénou pokračuje tak dlouho, dokud nedojde ke stabilizaci vysílané a přijímané energie u každé z plošek. Barva každé plošky je pak určena intenzitou světelné energie, kterou tato ploška celkově vyzařuje do okolí.

Přenos energie mezi dvěma ploškami lze popsat integrální rovnicí, přenos energie v rámci celé scény pak soustavou těchto rovnic. Úkolem metody je tedy sestavit a vyřešit soustavu integrálních rovnic pro zadanou scénu. Jedná se ovšem o značně složitou úlohu, časově srovnatelnou s metodou ray-tracingu. Jejímu řešení a modifikacím bude věnována celá následující kapitola.

Uvedený postup má oproti ray-tracingu tři základní výhody:

- radiační metoda vychází z fyzikálně přesnějšího modelu než ray-tracing a umožňuje tak elegantním způsobem překonat jeho nedostatky (např. plošné světelné zdroje, "měkké" stíny, propagace barev při odrazu světla na tělesech, zbarvení stínů vlivem barvy odraženého světla od jiného tělesa (*color bleeding*), ...)
- výpočet světelných vztahů v celé scéně je zcela nezávislý na pozici pozorovatele a směru pohledu (*view independent*). Pro animované sekvence, kdy se pozorovatel "prochází" v čase neměnnou scénou, stačí jednou spočítat světelné poměry v celé scéně a pak již jen z těchto informací rychle kreslit jednotlivé pohledy
- výpočet propagace světla scénou je také zcela nezávislý na rozlišovací schopnosti generovaných obrázků

Samozřejmě však existují také nevýhody. Základní radiační metody uvažují pouze opticky dokonale difúzní plošky a nepostihují směr světla propagovaného v dané scéně, světlo je po dopadu na plošku nezávisle na jeho dosavadní historii rozptýleno rovnoměrně do všech směrů (mluvíme o tzv. ideálně difúzních nebo též kosinových zářičích). To s sebou přináší nemožnost ve scéně korektně reprezentovat zrcadla a lesklé plochy, kdy je naopak skoro všechno světlo odráženo ve směru daném zákonem odrazu. Také průhledná tělesa a lom světla nebyly dosud v radiačních metodách postiženy. Další značnou nevýhodou základní metody je nutnost povrchové reprezentace jednotlivých těles ve scéně pomocí rovinných plošek. To značně omezuje množinu těles, které lze použít; hladká a zaoblená tělesa (např. koule či válec) nebo různé plochy je nutné složitě a ne zcela přesně aproximovat. Za nevýhodu lze považovat i značné paměťové nároky metody související s tím, že (oproti ray-tracingu) popisuje globálně celou scénu. Také některé další "vymoženosti" ray-tracingu (rozostření těles pohybem, dojem hloubky obrazu, prostorové textury ...) nebyly dosud v radiačních metodách uspokojivě vyřešeny.

Radiační metody nejsou ještě v současné době příliš rozšířené, i když se objevují jejich první komerční aplikace (uveďme alespoň systém pro návrh svítidel VISULUX [AG91]). Vývoj není ještě zdaleka ukončen, objevují se nejenom nejrůznější modifikace, ale i zcela nové přístupy (např. postižení směru světla či popis šíření světla soustavou diferenciálních rovnic). Stejně jako u ray-tracingu existuje k radiačním metodám rozsáhlá bibliografie, její přehled viz [Hai91].

2.3 Kombinované metody

Cílem všech metod této kategorie je kombinací ray-tracingu s radiačními metodami překonat nedostatky obou těchto metod a spojit jejich přednosti (dokonalá simulace lesklých a průhledných těles v ray-tracingu a korektní propagace difúzně rozptýleného světla v radiačních metodách). Kombinované metody lze rozdělit na následující tři kategorie:

dvojkrokové metody - jedná se o velice jednoduché spojení obou přístupů. Nejprve je radiační metodou vygenerován na pozorovateli nezávislý popis propagace světla danou scénou. Pomocí ray-tracingu, jehož vstupem je právě tento radiační popis scény, je pak vygenerován daný pohled na scénu doplněný všemi efekty, které ray-tracing umožňuje (hlavně odlesk a odraz na vyleštěných površích a textury). Výhodou tohoto přístupu je možnost interaktivní volby pohledu, který chceme vygenerovat po ukončení první fáze. Nevýhodou je to, že v rámci radiační fáze není postižen úzce směrový odraz světla, vznikající na lesklých površích a zrcadlech. Pro dosažení reálných obrázků je proto nutné v rámci ray-tracingu provádět skoro tolik kroků jako bez použití předvýpočtu pomocí radiačních metod. Jako zástupce této kategorie uveďme alespoň předchůdce všech kombinovaných metod [WCG87].

vícekrokové metody - u těchto metod je pro dokonalejší postižení propagace světla scénou použito mnohonásobného opakování radiačních a ray-tracingových kroků, kdy typ a parametry metody, která se v daném kroku použije, je určován charakteristikami a vztahy mezi jednotlivými aktualizovanými ploškami. Metody této kategorie poskytují dnes nejvyšší kvalitu generovaných obrázků, nevýhodou je ovšem jejich extrémní časová náročnost. Snad nejlepší z uvedených metod [CRMT91] používá kombinaci radiačního přístupu, Monte Carlo ray-tracingu a path ray-tracingu, její rychlost je zhruba jeden obrázek za den.

syntetické metody - tyto metody stojí na pomezí radiačních metod a ray-tracingu. Jejich podstatou je často zahrnutí nějaké modifikace ray-tracingu přímo do použité radiační metody. Do této kategorie patří i dále popisovaný směrový přístup či [SAWG91].

Kapitola 3

Radiační metody, jejich vývoj a modifikace

V této kapitole se podrobněji seznámíme s principy používanými při výpočtu osvětlení scény některou z radiačních metod. Nejprve si odvodíme postup nutný při použití základní metody, ve druhé části této kapitoly se budeme věnovat různým modifikacím a vylepšením této metody.

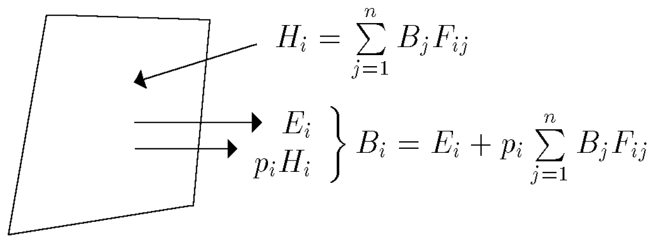
3.1 Základní princip, jeho odvození

Základní radiační metoda uvažuje pouze tzv. ideální difúzní zářiče, které dopadající světlo nezávisle na směru, odkud přichází, rozptýlí rovnoměrně do všech směrů viditelné polosféry. Pokud uvažujeme pouze tento typ zářičů, lze po dopadu paprsku na libovolné těleso zapomenout jeho dosavadní historii a pro každou plošku scény si pamatovat pouze celkové množství vyzařované světelné energie. Z důvodu zjednodušení také metoda předpokládá, že zobrazovaná scéna je energeticky izolovaná od okolí (jako by byla celá uzavřená do virtuální krychle, která veškerou vyzařovanou světelnou energii vrací zpět do scény). Metoda korektně postihuje termíny "difúzní" (*diffuse*) a "okolní" (*ambient*) světlo objevující se ve všech dokonalejších světelných modelech. Zde uvedené odvození této metody vychází z článků [Ser91] a [CG85].

Nechť se celá scéna skládá z N ploch; jednotlivé plochy označme indexy $i = 1 \dots N$. Každá z ploch scény vyzařuje světelnou energii B_i a dopadá na ni energie H_i. Vyzářená energie B_i (tzv. *radiace*) se skládá jednak z vlastní emise plochy E_i (pokud se jedná o primární zářič) a dále z energie, která vznikla odrazem světla H_i, vyslaného ostatními plochami směrem k dané ploše (tzv. sekundární záření). Platí tedy následující vztah, viz též obrázek 3.1:

$$B_i = E_i + p_i H_i \tag{3.1}$$

Číslo p_i s hodnotami mezi 0 a 1 se nazývá koeficient odrazivosti a udává, jaká část dopadající energie bude odražena zpět do scény.



Obr. 3.1: Energetické poměry na ploše

Energie dopadající na plošku H_i je součtem energií od všech okolních ploch, které i-tou plochu osvětlují. Podíl libovolné j-té plochy na ozáření i-té plochy bude záviset na velikosti obou ploch, na jejich vzájemné poloze a na energii, kterou j-tá plocha vyzařuje. Pro navzájem si nezaclánějící plochy můžeme napsat:

$$H_i A_i = \sum_{j=1}^{N} B_j A_j F_{ji} \tag{3.2}$$

kde A_i a A_j jsou velikosti příslušných ploch a F_{ji} je tzv. *konfigurační faktor*. Jedná se o číslo mezi 0 a 1 udávající, jaká část energie vyzářené z plochy j dopadne na plochu i. Z definice konfiguračních faktorů plyne:

$$A_i F_{ij} = A_j F_{ji} \tag{3.3}$$

Z rovnic (3.2) a (3.3) lze nejprve vyjádřit H_i:

$$H_i = \sum_{j=1}^{N} B_j F_{ij} \tag{3.4}$$

a po dosazení do rovnice (3.1) i tzv. *definiční rovnici pro radiaci i-té plochy*:

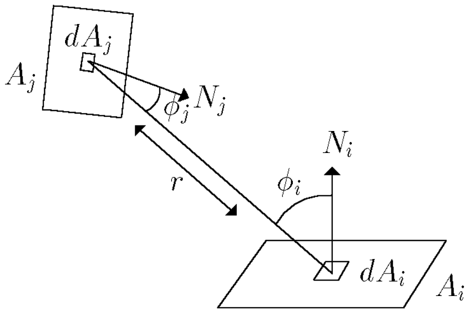
$$B_i = E_i + p_i \sum_{j=1}^{N} B_j F_{ij} \tag{3.5}$$

Poznamenejme ještě, že radiace se vyjadřuje jako energie na jednotkový čas a jednotkovou plochu a udává se v jednotkách Wm^{-2}.

Pokud aplikujeme vztah (3.5) na všechny plochy scény, dostáváme systém N lineárních rovnic:

$$\begin{bmatrix} 1-p_1F_{11} & -p_1F_{12} & \dots & -p_1F_{1N} \\ -p_2F_{21} & 1-p_2F_{22} & \dots & -p_2F_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -p_NF_{N1} & -p_NF_{N2} & \dots & 1-p_NF_{NN} \end{bmatrix} \cdot \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_N \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_N \end{bmatrix} \tag{3.6}$$

Řešením uvedené soustavy získáme hledané hodnoty B_i. Ještě před tím však musíme určit parametry E_i, p_i a F_{ij}. Emise jednotlivých ploch E_i jsou stanoveny uživatelem před zahájením výpočtu stejně jako koeficienty odrazivosti p_i. Zbývá tedy určit konfigurační faktory F_{ij}.



Obr. 3.2: Určení konfiguračního faktoru

Vyjděme ze situace znázorněné na obrázku 3.2. Na obou plochách A_i a A_j, jejichž vzájemný konfigurační faktor F_{ij} určujeme, zvolíme diferenciální plošky dA_i a dA_j. Pro konfigurační faktor mezi těmito elementárními ploškami $F_{dA_idA_j}$ bude platit (viz [SH78]):

$$F_{dA_idA_j} = \frac{\cos\phi_i \cdot \cos\phi_j}{\pi r^2} \cdot dA_j \tag{3.7}$$

Integrací přes celou plochu A_j dostaneme:

$$F_{dA_iA_j} = \int\limits_{A_j} \frac{\cos\phi_i \cdot \cos\phi_j}{\pi r^2}\, dA_j \tag{3.8}$$

Z tohoto vztahu již můžeme určit požadovaný konfigurační faktor F_{ij}, který získáme jako průměr po celé ploše A_i:

$$F_{ij} = F_{A_iA_j} = \frac{1}{A_i} \int\limits_{A_i} \int\limits_{A_j} \frac{\cos\phi_i \cdot \cos\phi_j}{\pi r^2}\, dA_j\, dA_i \tag{3.9}$$

Vztah (3.9) platí pouze v případě, že se plochy navzájem nezastiňují. V reálných scénách je proto nutné tento vztah upravit na:

$$F_{ij} = F_{A_iA_j} = \frac{1}{A_i} \int\limits_{A_i} \int\limits_{A_j} \frac{\cos\phi_i \cdot \cos\phi_j}{\pi r^2} \cdot HID(i,j)\; dA_j\, dA_i \tag{3.10}$$

kde dvouhodnotová funkce $HID(i, j)$ bude rovna jedné nebo nule podle toho, zda se plošky dA_i a dA_j navzájem vidí či nikoliv.

Pomocí rovnice (3.10) musíme určit konfigurační faktory mezi všemi plochami scény. Pokud se scéna skládá z N ploch, jedná se o N^2 výpočtů dvojného integrálu, tedy operaci značně složitou. Výhodou je, že jednou spočítané konfigurační faktory je nutné přepočítat pouze při změně geometrie scény, proti změně pozice pozorovatele nebo směru pozorování a proti změně optických vlastností jednotlivých ploch jsou tyto faktory invariantní. Po výpočtu všech konfiguračních faktorů je možné sestavit soustavu rovnic (3.6) a tuto soustavu vyřešit. Matice soustavy je značně řídká (většina ploch se navzájem nevidí) a též ostře diagonálně dominantní, lze ji proto s výhodou řešit iteračně. Často se používá Gauss-Seidelova metoda, která konverguje během několika málo kroků.

Až dosud jsme uvažovali pouze černobílé scény, při generování barevných scén musíme uvedený opakovaně provést pro červenou, zelenou a modrou složku. Poznamenejme, že konfigurační faktory jsou pro všechny složky stejné, mění se pouze emise a koeficienty odrazivosti jednotlivých ploch. Pro každou složku tedy stačí pouze sestavit a vyřešit odpovídající soustavu lineárních rovnic.

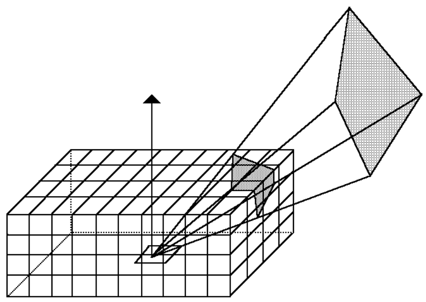
Po výpočtu radiace všech ploch (jedná se vždy o třísložkové vektory udávající radiaci dané plochy v červené, zelené a modré oblasti) může dojít k postupnému zobrazování jednotlivých pohledů na danou scénu. Pro každý pohled jsou nejprve dle pozice pozorovatele a směru pohledu určeny viditelné oblasti jednotlivých ploch (pro tento krok zobrazovacího procesu se používají standardní algoritmy pro řešení plošné viditelnosti, např. řádkový rozklad či Z-buffer). Všechny viditelné oblasti jsou pak vykresleny pomocí Gouradovy metody spojitého stínování [Gou71], kdy intenzita a barva každého pixelu oblasti je určena bilineární interpolací z hodnot ve vrcholech plochy, do které daná oblast patří. Samotné hodnoty radiací ve vrcholech se určí jako aritmetické průměry z hodnot radiací okolních ploch.

3.2 Přehled vývoje a modifikací

Po tom, co jsme si popsali princip základní radiační metody, věnujme se ještě chvíli vývoji těchto metod. V následujících bodech bude uveden pouze přehled nejdůležitějších přístupů a modifikací, podrobnější informace lze získat v citované literatuře. Jednotlivé metody jsou řazeny chronologicky v pořadí, v jakém byly publikovány:

Explicitní numerické řešení [Gor85] [GTGB84]
Jedná se o historicky první radiační metodu, použitou Cindy M. Goralovou jako konkurence ray-tracingu. Metoda funguje přesně tak, jak bylo uvedeno v předchozím odstavci, dvojné integrály při výpočtu konfiguračních faktorů jsou vyčíslovány explicitně. Jelikož analytické řešení těchto integrálů je obecně nedostupné, bylo nutné se omezit pouze na jisté konfigurace ploch, pro které byly odpovídající konfigurační faktory tabelovány (viz např. [How82]). Hlavní nevýhodou metody byla nemožnost zpracovávat rozsáhlejší scény, neboť nebyl znám žádný postup, jak v těchto scénách vůbec určit vzájemné konfigurace jednotlivých ploch.

Užití polokrychlí pro výpočet konfiguračních faktorů [CG85]
Výše uvedený nedostatek základní metody se podařilo elegantně odstranit aplikací *Nusseltovy geometrické analogie* s použitím tzv. polokrychlí (*hemicubes*).



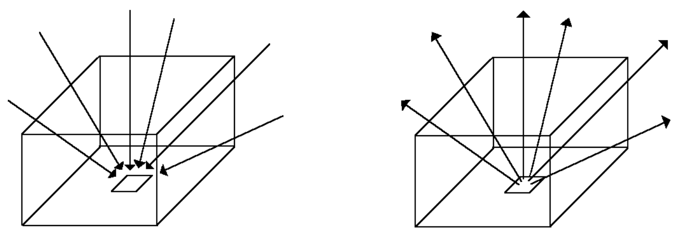
Obr. 3.3: Polokrychle

Při výpočtu konfiguračních faktorů mezi konkrétní aktualizovanou plochou a všemi ostatními plochami je nad touto plochou umístěna virtuální polokrychle skládající se z elementárních plošek s explicitně určenými konfiguračními faktory. Jednotlivé plochy scény jsou pak na tuto polokrychli promítnuty s vyřešením jejich viditelnosti z aktualizované plochy (polokrychle se vlastně v tomto okamžiku chová jako prostorový Z-buffer), viz obr. 3.3. Konfigurační faktor každé z ploch vůči aktualizované ploše je pak určen jako součet konfiguračních faktorů všech elementárních plošek polokrychle, na něž je daná plocha promítnuta.

Dělení ploch na elementy [CGIB86]
Z popisu základní radiační metody je zřejmé, že přesnost aproximace jednotlivých konfiguračních faktorů a tím i kvalita generovaných obrázků značně závisí na velikosti jednotlivých ploch scény (a případně ještě na počtu elementárních plošek, z nichž se skládá polokrychle). Se zmenšováním rozměrů jednotlivých plošek však roste jejich počet a vzhledem k tomu, že časová i paměťová náročnost radiačního algoritmu je kvadraticky závislá na počtu použitých ploch, není tento postup v mnoha případech výhodný. Citovaný přístup proto dělí základní plochy na tzv. elementy (*patch*), ovšem pouze v rámci výpočtu konfiguračních faktorů, při nezměněném počtu základních ploch (a tím pádem i rovnic, které je nutné řešit).

Adaptivní dělení ploch [CGIB86]
Toto vylepšením předchozí metody vychází z pozorování, že dělení ploch na elementy je potřebné hlavně v oblastech velké změny gradientu osvětlení (např. v oblastech přechodu mezi osvětlenou a zastíněnou částí), zatímco rovnoměrně osvětlené plochy není třeba dělit vůbec. Dělení ploch proto probíhá adaptivně, rozděleny jsou jen ty plochy, na nichž rozdíl intenzit osvětlení přesáhne jistou hranici. S nově vzniklými plochami se pracuje úplně stejně jako s původními, postupně se tedy zvyšuje počet rovnic i paměťové nároky metody, přesto je však výpočet stejně kvalitního obrázku mnohokrát rychlejší než u předchozích metod.

"Střílení" (*shooting*) [CCWG88]



Standardní postup (gathering) Střílení (shooting)

Obr. 3.4: Střílení

Autorům této metody se podařilo přeformulovat základní radiační přístup tak, že došlo ke snížení paměťových i časových nároků z $O(n^2)$ na $O(n)$ (n je počet ploch ve scéně). Idea je velice jednoduchá: Místo toho, abychom pro jednotlivé plošky počítali, kolik na ně dopadne energie z okolních ploch (*gathering*), budeme naopak distribuovat dosud nevyzářenou energii dané plošky do jejího okolí (*shooting*), jak je znázorněno na obrázku 3.4. V rámci jednoho kroku je tak vždy aktualizováno mnoho ploch současně, navíc si můžeme vybrat, kterou z ploch scény budeme aktualizovat. Pokud si vždy vybereme plochu s největším množstvím dosud nevystřelené energie (zpočátku to budou světelné zdroje, dále plochy blízko těchto zdrojů atd.), bude metoda rychle konvergovat k přijatelnému řešení. Po provedení každého "střílecího" kroku je možné scénu nakreslit a při dostatečné kvalitě obrázku výpočet ukončit (většinou postačuje několik kroků). V rámci výpočtu není nutné vytvářet matici soustavy (3.6), není tedy ani nutné předem počítat a uchovávat všechny konfigurační faktory. Potřebné faktory se vždy spočtou až v rámci dané aktualizace.

Užití ray-castingu pro výpočet konfiguračních faktorů [MBG86]

Značně omezujícím faktorem se u všech výše popsaných metod ukázalo použití polokrychlí pro výpočet konfiguračních faktorů. Tyto polokrychle totiž tvoří uniformní dělení prostoru, které spolu s pravidelným dělením ploch na menší plošky způsobují značný *aliasing* (zkreslení nedostatečným vzorkováním) generovaných obrázků. Aliasing se v jistých případech nedá odstranit ani zvýšením počtu elementů polokrychlí. K výpočtu konfiguračních faktorů byl proto použit alternativní přístup, kdy se na daných plochách zvolí síť vzorkovacích bodů (*sampling points*) a pomocí ray-castingu zjistí jejich vzájemná viditelnost. Odpovídající konfigurační faktor je pak určen z počtu navzájem viditelných bodů a z polohy daných ploch.

Aktualizace vrcholů místo plošek [WEH89]

Při generaci obrázků dané scény je nejprve nutné určit radiace v jednotlivých vrcholech kreslených ploch. Tyto radiace se standardně určují zprůměrováním z hodnot radiací okolních ploch. V souvislosti s použitím ray-tracingu při výpočtu konfiguračních faktorů se však ukázala výhodným aktualizace radiace přímo v těchto vrcholech. Toho lze dosáhnout velice jednoduchou modifikací ray-tracingového přístupu: vrcholy se jednoduše zvolí jako vzorkovací body dané plochy.

Směrový přístup [ICG86] [RT90]

Základní nevýhodou základní radiační metody je to, že považuje všechny plochy scény za ideální difúzní zářiče. Aby bylo možné postihnout úzce směrový odraz, vznikající na lesklých površích a další ray-tracingové "vymoženosti", bylo v radiačních metodách nutné postihnout možnost, že intenzita radiace plochy bude záviset na směru, ve kterém tuto radiaci pozorujeme (tj. zahrnout i obecné difúzní zářiče). Základem všech metod této kategorie je reformulace propagace světla scénou spojená s uchováváním a modifikací celých tzv. *vyzařovacích funkcí*, které popisují radiaci jednotlivých ploch v závislosti na směru pozorování (místo dosud uchovávané jediné hodnoty, která určovala radiaci plochy ve všech směrech). Podrobněji budou tyto metody popsány v následující kapitole.

Výpočet konfig. faktorů pomocí polorovin nebo jehlanů [Kir92]
Cílem obou modifikací je urychlení a zvýšení přesnosti při výpočtu konfiguračních faktorů. Standardní polokrychle je nahrazena buď polorovinou (kdy se zanedbají boční stěny polokrychle) nebo čtyřbokým jehlanem (kdy se naopak ignoruje horní stěna a boční stěny se změní v trojúhelníky).

Popis vlastností povrchů pomocí distribučních funkcí [SAWG91]
Jedná se o rozšíření směrového přístupu v radiačních metodách, kdy je pro každý povrch ve scéně uložena jeho *distribuční funkce*, tedy popis, jak daný povrch rozptyluje dopadající světelnou energii. Pokud hodnoty distribuční funkce závisí i na směru, odkud světelná energie přichází, lze dosáhnout skoro přesné simulace skutečných materiálů (např. hrubý či leštěný hliník nebo soustružená ocel), nevýhodou jsou však značné paměťové nároky pro uložení popisu těchto funkcí. Častěji se proto používají tzv. dvousměrové distribuční funkce (*BRDF - Bidirectional Reflectance Distribution Function*), kde směr přicházejícího paprsku není postižen. Hodnoty distribučních funkcí pro jednotlivé materiály se určují explicitně přímo z fyzikálních měření.

Urychlení konvergence pomocí přestřelování (*overshooting*) [Pur92]
Zajímavou metodou značně urychlující konvergenci výpočtů je modifikace "vystřelovacího" přístupu, kdy je z dané plochy vždy vystřeleno o něco více energie než plocha v daném okamžiku skutečně vyzařuje. Vychází se přitom z předpokladu, že tato "nadvystřelená" (*overshooting*) energie na plochu dopadne (a bude ji tedy nutné vystřelit zpět do scény) v některém z dalších "střílecích" kroků. Pokud k tomu nedojde, je nutné provést korekci vystřelením energie záporné velikosti. Koeficient, kolikrát více energie se má vystřelit, se určuje z množství dosud nevystřelené energie obsažené v celé scéně.

Kapitola 4

Směrový přístup v radiačních metodách

V této kapitole se budeme věnovat směrovému přístupu v radiačních metodách, tedy prostředku, který umožňuje reprezentovat směr světla propagovaného ve scéně. Postižení tohoto směru má velký význam pro kvalitu generovaných obrázků, umožňuje totiž simulovat odraz světelných paprsků na lesklých površích a také jejich lom a průchod průhlednými a průsvitnými tělesy. Společným cílem všech postupů této kategorie je přiblížit kvalitu obrázků realitě alespoň na úroveň obrázků generovaných ray-tracingem při zachování všech výhod radiačních metod, způsoby, kterými toho jednotlivé metody dosahují, jsou však značně různorodé. V následujících odstavcích budou nejprve stručně shrnuty vlastnosti dosud publikovaných řešení, dále bude následovat podrobný popis vlastního přístupu.

4.1 Dosud publikované metody

Pokud pomineme přístupy patřící do kategorie dvoukrokových kombinovaných metod, u nichž jde pouze o jednoduché spojení klasických radiačních metod s ray-tracingem, lze směrové radiační metody v zásadě rozdělit do následujících skupin:

- Důmyslnější kombinace radiačních metod a ray-tracingu
- Rozšíření radiačních metod jejími vlastními prostředky
- Nové přístupy na pomezí radiačních metod a ray-tracingu

4.1.1 Důmyslnější kombinace radiačních metod a ray-tracingu

Typickým zástupcem (a jednou z nejlepších metod po stránce kvality generovaných obrázků vůbec) je [CRMT91], kde střídavě probíhají následující tři kroky:

1. **radiační krok** – poskytující hrubý popis propagace světla celou scénou. Je použita standardní radiační metoda, kdy se všechny plošky považují za ideální difúzní zářiče. Směr světla tedy není vůbec postižen a musí být dodatečně "přidán" v následujících krocích.
2. **vysokofrekvenční krok** – kde se pomocí ray-tracingu do generovaného obrázku přidává vysokofrekvenční informace, tedy i vliv směru jednotlivých světelných paprsků. Jedná se o simulaci různých obrazů vznikajících v zrcadlech, ostrých stínů, odlesků, průhledných těles atd.
3. **nízkofrekvenční krok** – cílem tohoto kroku je zlepšit nízkofrekvenční charakteristiku daného obrázku (tedy zkvalitnit postupné přechody mezi barevnými odstíny, vylepšit generované polostíny a podobně). Ve své podstatě je jedná o simulaci směrového radiačního přístupu pomocí distribuovaného ray-tracingu, kdy se pro dosažení kvalitních obrázků používá několik desítek až stovek paprsků na každý pixel.

Metoda poskytuje vynikající výsledky, generované obrázky jsou často nerozlišitelné od fotografií. Po volbě pozice pozorovatele a směru pozorování lze uvedené kroky výpočtu libovolně kombinovat (fixace pozorovatele je nutná z důvodu použití ray-tracingu ve vysokofrekvenčním a nízkofrekvenčním kroku). Ve své podstatě se tedy jedná o směrovou metodu, propagace směru světla však není postižena globálně, ale pouze v rámci právě generovaného obrázku. Místo o směrovém přístupu v radiačních metodách bychom tedy měli hovořit spíše o použití radiačních metod v ray-tracingu.

4.1.2 Rozšíření radiačních metod jejími vlastními prostředky

Jako zástupce této skupiny lze jmenovat postup popsaný v práci [ICG86], kde je použit "střílecí" algoritmus a polokrychle, směr světla je postižen následující modifikací: Pro každou plochu není uchováváno jediné číslo, určující radiaci plochy ve všech směrech, ale množina čísel, udávající radiaci plochy v daných pevně zvolených směrech. Vzhledem k tomu, že se používá aktualizace pomocí polokrychlí, odpovídají jednotlivé směry směrům, v nichž leží středy jednotlivých elementárních ploch, z nichž se daná polokrychle skládá. Při aktualizaci jednotlivých plošek v rámci "vystřelovacího" kroku pak každá ploška obdrží jen tolik energie, kolik odpovídá počtu a pozici elementů polokrychle, na něž byla promítnuta.

Výhodou metody je, že směr světelné energie je propagován už v rámci jednotlivých "vystřelovacích" kroků, tedy nikoliv až na konec jako v případě dvoukrokové kombinace základní radiační metody s ray-tracingem. To umožňuje dosáhnout podstatně věrnějšího popisu šíření světla scénou, lze například postihnout propagaci vícenásobně odraženého světla či zbarvení stínů vlivem barvy odraženého světla.

Největší slabina metody spočívá v jejím nedokonalém postižení směrového vyzařování jednotlivých plošek, hodnota radiace plošek je totiž uchovávána pouze v několika předem určených směrech. Radiaci plošky v libovolném požadovaném směru je nutné získávat aproximací těchto hodnot, což způsobuje drobné nepřesnosti a ve svém důsledku aliasing generovaného řešení. Aliasing nelze zcela potlačit ani při značném zvýšení počtu elementů polokrychlí (při kterém navíc rychle rostou požadavky na paměť pro uchování jednotlivých směrových koeficientů).

4.1.3 Nové přístupy na pomezí radiačních metod a ray-tracingu

Ani jedna z dosud uvedených skupin směrových metod nepostihuje směr světla ve scéně zcela korektně. V rámci první skupiny je směr světla popsán dostatečně přesně, často ale pouze v rámci jediného obrázku, nikoliv celé scény. Metody druhé skupiny zase korektně popisují propagaci směrově šířeného světla celou scénou (a to dokonce v rámci vlastních radiačních metod, bez použití ray-tracingu), chybí však odpovídající postižení tohoto směru v rámci radiace jednotlivých plošek.

Nejlepší by bylo výhody obou skupin zkombinovat a zahrnout směrový přístup přímo do dané radiační metody s tím, že popis směrového vyzařování plošek by měl být stejně kvalitní jako v případě ray-tracingu. Právě o to se snaží nové metody, které často kombinují radiační a ray-tracingový přístup přímo ve svém základním principu.

Z této kategorie si uveďme alespoň metodu popsanou v článku [SAWG91]. Závislost radiace plošky na směru, ve kterém ji pozorujeme, je v této metodě vyjádřena jako spojitá funkce, tedy nikoliv pouze hodnotami této funkce v několika zvolených směrech, jak tomu bylo v předchozí metodě. To umožňuje určení hodnoty radiace v libovolném požadovaném směru značně přesně, bez nutnosti aproximace. Funkce sama je popsána pomocí jednotlivých funkcí z nějakého předem zvoleného systému (konkrétně je použit systém tzv. normovaných sférických harmonických funkcí – *normalized spherical harmonics*). Vzhledem k tomu, že danou aproximací nelze radiaci plošek postihnout dostatečně přesně, je v rámci výpočtu používán ještě ray-tracing. Oproti předchozím metodám je však příspěvek ray-tracingu zahrnut do radiací jednotlivých plošek a dále propagován scénou společně s příspěvkem určeným pomocí polokrychlí.

4.2 Cíl vlastního návrhu, jeho základní rysy

Postižení směru propagované energie v rámci radiačních metod není žádná novinka, první pokusy pochází již z roku 1986 ([ICG86]). Přesto zatím nedošlo k výraznějšímu rozšíření a popularizaci této metody, současné trendy v radiačních metodách směřují spíše k distribuovaným a paralelním algoritmům a ke kombinaci standardního (tj. nesměrového) přístupu s ray-tracingem. Přitom vývoj nesměrových radiačních metod i ray-tracingu je dnes již prakticky ukončen, obě metody jsou na vrcholu svých možností a budoucnost patří (alespoň dle mého názoru) právě směrovým metodám. Důvodů malé popularity směrového přístupu je hned několik: jednak jsou směrové metody jak časově, tak hlavně paměťově velice náročné při nepříliš kvalitních výsledcích, jednak je lze do jisté míry nahradit kombinací osvědčeného ray-tracingu se standardními radiačními metodami. Hlavním důvodem je však nedostatečné teoretické zázemí směrových metod, v době, kdy se o ray-tracingu publikuje ročně několik stovek článků, bylo o těchto metodách za celou jejich 7-mi letou historii napsáno zhruba 20 – 30 prací. Právě tento nedostatek byl největší motivací k napsání této diplomové práce.

Práce vychází z metod popsaných v předchozích kapitolách (hlavně z metody popsané v článku [SAWG91]) a jejím hlavním cílem bylo formulovat postup, který by umožnil efektivně reprezentovat směr světelné energie propagované scénou v rámci radiačního výpočtu. Postižení směru v rámci tzv. vyzařovacích funkcí jednotlivých elementů bude věnována hlavní část celé práce, celkově však bude charakterizována celá nově navrhnutá směrová metoda počínaje reprezentací scény a světelným modelem až po popis propagace světelné energie scénou a formulaci směrového radiačního algoritmu. Důraz bude kladen hlavně na teoretické základy popisované metody, podrobněji také budou popsány některé nestandardní postupy jako použití vystřelovacího algoritmu s aktualizací typu vrchol – vrchol či aplikace ray-castingu. Nyní se však již věnujme vlastní metodě.

4.3 Reprezentace scény

Dříve než se dostaneme k popisu vlastní směrové metody, věnujme se způsobu, jakým lze popsat generovanou scénu. Je zřejmé, že při použití radiačního přístupu nestačí pouze popsat jednotlivé objekty, z nichž se scéna skládá a postihnout jejich vzájemné vztahy, s každým objektem musí být svázány i informace o jeho radiaci a další charakteristiky (např. optické vlastnosti). V této kapitole se zaměříme pouze na možnosti, jak reprezentovat prostorové vztahy mezi objekty, tj. geometrii dané scény. Popis použité reprezentace bude rozdělen do následujících bodů:

- porovnání objemové a plošné reprezentace
- modifikace plošné reprezentace
- užití trojúhelníků (trojúhelníkových plošek) pro aproximaci povrchu
- zobecnění na trojúhelníkové pláty

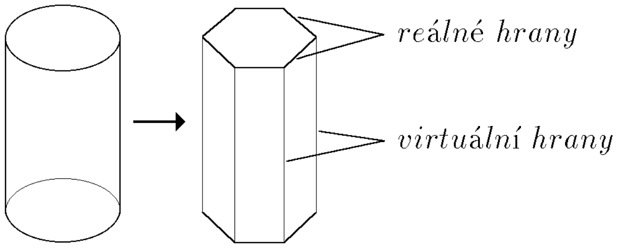
4.3.1 Porovnání objemové a plošné reprezentace

V zásadě existují dvě možnosti, jak danou scénu reprezentovat:

- popisem objemu jednotlivých těles pomocí tzv. objemové reprezentace (*solid representation*)
- popisem povrchu jednotlivých objektů pomocí tzv. povrchové reprezentace (*boundary representation*)

Použití objemové reprezentace je z uživatelského hlediska jednoznačně výhodnější. Scény lze velice jednoduše vytvářet pomocí *množinových operací* nad množinou daných primitivních těles (sjednocení dvou těles, jejich průnik a odečtení jednoho tělesa od druhého), přirozený je i popis scény pomocí tzv. CSG stromů (*Computer Solid Geometry*) či hierarchických grafů. Animované sekvence lze vytvářet přímo popisem trajektorie jednotlivých těles a jejich změn (např. jazyk systému RENDERMAN), ve všech případech se těleso chápe jako skutečný třídimenzionální objekt. Nevýhodou je snad jen složitá reprezentace těles, jejichž povrch tvoří obecné plochy. Při zobrazování takto reprezentovaných scén však narážíme na řadu problémů, kromě ray-tracingu či ray-castingu není možné tyto scény zobrazit přímo bez jejich převodu do jiné reprezentace.

Oproti objemové reprezentaci pracuje plošná reprezentace pouze s dvojrozměrnými útvary – plochami. Každé těleso je v této reprezentaci charakterizováno soustavou ploch, které tvoří jeho povrch. Tento postup není z uživatelského hlediska příliš výhodný, uplatňuje se pouze tam, kde je nutné modelovat složité povrchy vyhovující jistým kritériím (hlavně letecký a automobilový průmysl). Popis scény při použití povrchové reprezentace je značně nepřehledný (jedná se vlastně pouze o seznam jednotlivých ploch), nad plochami se také složitě provádí jakékoliv animační operace. Hlavní výhodou této reprezentace je možnost danou scénu poměrně jednoduše zobrazovat a provádět nad ní nejrůznější "grafické" operace (ořezávání, řešení vzájemné viditelnosti, odstraňování neviditelných hran a ploch, stínování atd.). Pro popis povrchu se používají nejrůznější aproximační plochy (uveďme alespoň Bèzierovy a β-spline), často je povrch aproximován přímo pomocí trojúhelníků, obdélníků či konvexních polygonů. Hrany mezi použitými aproximačními plochami lze rozdělit do dvou kategorií – na tzv. reálné a virtuální. Reálné hrany odpovídají skutečným hranám reprezentovaných těles, naopak virtuální hrany ve skutečnosti neexistují a používají se pouze pro rozdělení povrchu tělesa na menší části, které se pak nahrazují jednotlivými aproximačními povrchy. Jinak řečeno, reálné hrany jsou ty, které vidíme na generovaných obrázcích, virtuální hrany jsou pouze pomocné a na obrázcích neviditelné (povrch tělesa je v daném místě hladký). Příklad použití reálných a virtuálních hran pro reprezentaci válce je zobrazen na obrázku 4.1



Obr. 4.1: Reálné a virtuální hrany

Aby se využilo výhod obou přístupů, často se používají obě reprezentace současně. Objemová reprezentace se používá jako vstupní, uživatelská, povrchová reprezentace pak jako výstupní, systémová. Propojení mezi oběma reprezentacemi je následující: Po zadání či modifikaci scény v objemové reprezentaci se tato reprezentace převede na povrchovou (např. pomocí *Eulerových operací*), která se pak použije pro kresbu dané scény. Vzhledem k tomu, že povrchová reprezentace je ve skutečnosti pouze aproximace, jsou při kresbě používány i informace uložené v reprezentaci objemové (např. pro správné určení normál jednotlivých povrchů).

4.3.2 Modifikace plošné reprezentace

V rámci radiačních metod dochází k výpočtu propagace světelné energie mezi jednotlivými plochami scény, je tedy nutné použít plošnou reprezentaci zobrazované scény. Toto je jedna z hlavních nevýhod radiačních metod vůči ray-tracingu, která značně omezuje jejich použití při generaci složitějších animací (kdy dochází k deformacím jednotlivých těles či kdy se výrazně mění topologie celé scény).

Nutnost použít plošnou reprezentaci se nepodařilo odstranit ani v této práci, došlo však k jejímu přiblížení reprezentaci objemové. Oproti dosavadním radiačním metodám, kdy je scéna chápána jako množina plošek stejných vlastností, je v popisované metodě použit přístup, který tyto plošky navzájem diferencuje. Scéna je zde chápána jako množina těles, každé těleso je pak reprezentováno skupinou plošek, které tvoří jeho povrch. Tento přístup, kdy můžeme pro každou plošku určit, do jakého tělesa patří, umožňuje v radiačních metodách reprezentovat nejenom odraz světelné energie od povrchu těles, ale i její průchod danými tělesy (postižení lomu světelných paprsků). Po modifikaci světelného modelu (viz kapitola 4.4) lze pak zcela přirozeně používat nejenom neprůhledná, ale i průsvitná a průhledná tělesa. V popisované metodě jsou jednotlivé plošky scény ještě navíc rozděleny na následující dvě skupiny:

- *plochy* – jedná se o původní plošky, jimiž byla popsána daná scéna, v průběhu radiačního výpočtu mohou být děleny na stále menší a menší plošky
- *plošné elementy* – vznikají postupným dělením jednotlivých ploch; ty z nich, které již nejsou dále rozděleny, se v daném okamžiku výpočtu považují za aproximaci diferenciálních plošek použitých ve světelném modelu (viz kapitola 4.4)

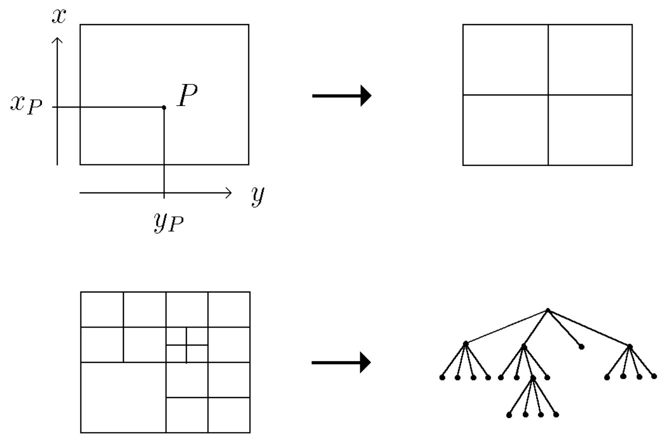
4.3.3 Užití trojúhelníků pro aproximaci povrchu

U dosud publikovaných metod se pro aproximaci povrchu jednotlivých těles scény používají nejčastěji obdélníky. Důvodů je hned několik:

- jednoduché zavedení souřadnic na jednotlivých obdélnících (kartézská soustava souřadnic)
- triviální dělení obdélníku na elementy
- výhodný popis výsledného dělení dané plochy pomocí tzv. *kvadrantových stromů*

- rychlé určení průsečíku obdélník – úsečka
- jednoduchá interpolace hodnot nějaké funkce zadané ve vrcholech – používá se bilineární interpolace, její popis viz např. [Gou71]
- jednoduchá lokalizace bodu – tj. jsou zadány souřadnice bodu uvnitř obdélníku, chceme určit, do jakého elementu patří; problém lze řešit průchodem odpovídajícího kvadrantového stromu s testem bodu vůči dělícím přímkám

Jednotlivé body jsou shrnuty na obrázku 4.2.



Obr. 4.2: Použití obdélníků

U popisované metody bylo použito řešení jiné, a to reprezentace povrchu pomocí trojúhelníků. Hlavní motivací bylo umožnit přirozenou aproximaci složitějších těles a také zjednodušit operace spojené s výpočtem propagace světla scénou. První cíl byl vyřešen triangulací daného povrchu; pro triangulaci lze použít mnoha standardních algoritmů, některé triangulace (např. Delaunayova triangulace) mají navíc oproti použití obdélníků velice výhodné vlastnosti (nemohou vznikat "protáhlé" trojúhelníky). Jak se podařilo splnit druhý cíl vyplyne z následujících odstavců.

Nejprve se věnujme souřadnicím jednotlivých bodů daného trojúhelníku. Je možné použít pravoúhlý souřadný systém (kartézské souřadnice) jako u obdélníků, daleko výhodnější je však použití tzv. *barycentrických souřadnic.* Tyto souřadnice se zavádějí následovně (viz [RJ92]): Uvažujme trojúhelník s vrcholy A, B a C a libovolný bod P ležící v rovině tvořené těmito body. Pak je možné bod P vyjádřit jako lineární kombinaci daných vrcholů:

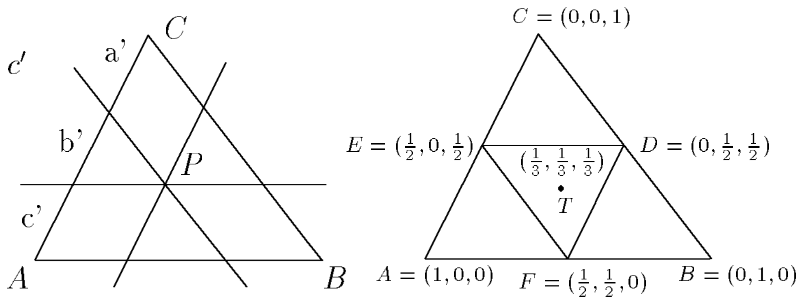
$$P = a \cdot A + b \cdot B + c \cdot C \quad \text{za podmínky} \quad a + b + c = 1 \tag{4.1}$$

Koeficienty (a, b, c) se nazývají barycentrickými souřadnicemi bodu P vůči trojúhelníku ABC. Těmito souřadnicemi je možné specifikovat všechny body roviny, v níž daný trojúhelník leží, pro body uvnitř trojúhelníku musí navíc platit:

$$a, b, c \in \langle 0, 1 \rangle \tag{4.2}$$

Celá situace je znázorněna na obrázku 4.3, kde jsou uvedeny i souřadnice některých význačných bodů (D, E a F dělí jednotlivé strany $\triangle ABC$ na poloviny, bod T je těžiště tohoto trojúhelníku).

$$\begin{aligned} d &= a' + b' + c' \\ a &= a'/d \\ b &= b'/d \\ c &= c'/d \\ P_{\triangle ABC} &= (a, b, c) \end{aligned}$$



Obr. 4.3: Barycentrické souřadnice

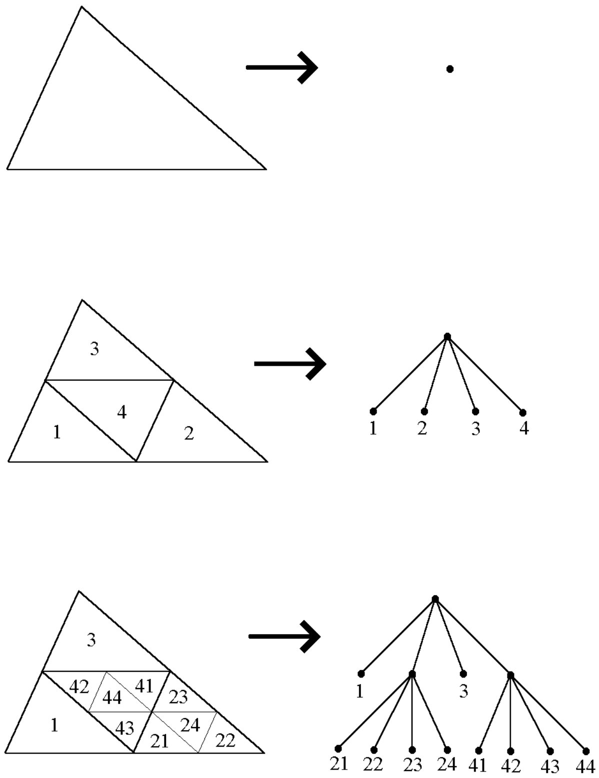
Barycentrické souřadnice umožňují velice jednoduchou interpolaci hodnot nějaké funkce zadané ve vrcholech A, B a C pro libovolný bod P uvnitř trojúhelníku ABC: Nechť $f(A)$, $f(B)$ a $f(C)$ jsou hodnoty dané funkce ve vrcholech $\triangle ABC$ a bod $P \in \triangle ABC$ má barycentrické souřadnice (a, b, c). Pak

$$f(P) = a \cdot f(A) + b \cdot f(B) + c \cdot f(C) \tag{4.3}$$

Tato možnost bude v popisované metodě s výhodou použita pro určování hodnot tzv. vyzařovací funkce v rámci kresby scény nebo při adaptivním dělení jednotlivých plošek (blíže viz kapitoly 4.13 a 4.12).

Na obrázku 4.3 je také znázorněn způsob, jakým se daný trojúhelník ABC v případě potřeby dělí na čtyři menší. Jedná se o trojúhelníky $\triangle AFE$, $\triangle BDF$, $\triangle CED$ a $\triangle DEF$, které jsou mezi sebou navzájem odděleny třemi nově vytvořenými virtuálními hranami. Drobné problémy jsou s novými vrcholy D, E a F, tyto vrcholy mohou totiž někdy existovat (pokud již byl rozdělen odpovídající sousední trojúhelník), jindy je nutné je vytvořit.

Trojúhelníky se dělí vždy na čtyři části, pro popis výsledného dělení nějakého trojúhelníku lze tedy použít kvadrantových stromů stejně jako v případě obdélníků. Kvadrantový strom (někdy též čtyřstrom – *quadtree*) je strom, kde každý uzel je buď list nebo má právě čtyři následníky. Listy reprezentují elementární,



Obr. 4.4: Kvadrantové stromy

dále nerozdělené trojúhelníky, uzly naopak ty trojúhelníky, které byly rozděleny na menší. Pořadí následníků je pro vnitřní uzly pevné a je zřejmé z obrázku 4.4.

Z obrázku 4.4 vyplývá ještě jedna důležitá vlastnost výše popsaného dělení trojúhelníků: Pokud zaručíme, aby počáteční triangulace povrchů těles ve scéně byla taková, že každý vrchol této triangulace bude ležet v nejvýše 6 trojúhelnících, bude tato vlastnost vrcholů zachována i po libovolně složitém dělení jednotlivých trojúhelníků.

Zbývá vyřešit poslední dva problémy spojené s použitím trojúhelníků: určení průsečíku trojúhelník – úsečka a lokalizace bodu. Problém určení průsečíku lze jednoduše převést na problém lokalizace bodu, stačí určit případný průsečík úsečky s rovinou, do níž trojúhelník patří a pokud tento průsečík existuje, pomocí lokalizace určit, zda leží uvnitř nebo vně daného trojúhelníku.

Problém lokalizace bodu lze formulovat následovně: Je dán trojúhelník ABC a bod P ležící ve stejné rovině. $\triangle ABC$ je výše popsaným způsobem rozdělen na elementární trojúhelníky $\triangle T_i$, odpovídající kvadrantový strom je znám. Úkolem je určit, zda bod P leží uvnitř trojúhelníku ABC a pokud ano, určit ještě ve kterém elementárním $\triangle T_i$ leží a jaké jsou jeho barycentrické souřadnice vůči tomuto elementárnímu trojúhelníku. Tento problém lze rozdělit na dvě části: jednak určit souřadnice bodu P vůči $\triangle ABC$ a dále pokud bod leží uvnitř tohoto trojúhelníku, postupně jej lokalizovat.

Chceme-li určit souřadnice bodu P vůči $\triangle ABC$, je nutné vyřešit následující soustavu rovnic pro neznámé a, b a c:

$$\begin{aligned} P_x &= a \cdot A_x + b \cdot B_x + c \cdot C_x \\ P_y &= a \cdot A_y + b \cdot B_y + c \cdot C_y \\ P_z &= a \cdot A_z + b \cdot B_z + c \cdot C_z \\ 1 &= a + b + c \end{aligned} \tag{4.4}$$

Po dosazení $c = 1 - a - b$ do prvních tří rovnic a jejich úpravě dostáváme:

$$\begin{aligned} P_x - C_x &= a \cdot (A_x - C_x) + b \cdot (B_x - C_x) \\ P_y - C_y &= a \cdot (A_y - C_y) + b \cdot (B_y - C_y) \\ P_z - C_z &= a \cdot (A_z - C_z) + b \cdot (B_z - C_z) \end{aligned} \tag{4.5}$$

Potřebujeme určit pouze dvě neznámé, třetí rovnici proto můžeme ignorovat. První dvě přepíšeme do maticového tvaru:

$$\begin{bmatrix} A_x - C_x & B_x - C_x \\ A_y - C_y & B_y - C_y \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} P_x - C_x \\ P_y - C_y \end{bmatrix} \tag{4.6}$$

Vyřešením soustavy (4.6) získáme následující vztahy:

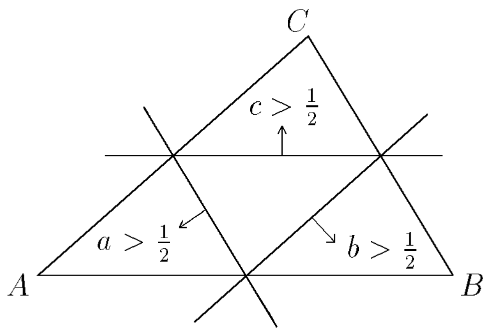
$$\begin{aligned} A' &= A - C \\ B' &= B - C \\ P' &= P - C \\ d &= A'_x \cdot B'_y - A'_y \cdot B'_x \\ a &= (P'_x \cdot B'_y - P'_y \cdot B'_x)/d \\ b &= (A'_x \cdot P'_y - A'_y \cdot P'_x)/d \\ c &= 1 - a - b \end{aligned} \tag{4.7}$$

Pokud body A, B a C nejsou kolineární (tj. neleží na jedné přímce, tento předpoklad je v případě, kdy dané body tvoří vrcholy trojúhelníku automaticky splněn), má soustava (4.7) právě jedno řešení. Vektor (a, b, c) pak určuje barycentrické souřadnice bodu P vůči trojúhelníku $\triangle ABC$.

Není-li splněna podmínka (4.2) (tj. alespoň jedna ze souřadnic neleží v intervalu $\langle 0, 1 \rangle$), leží bod P mimo $\triangle ABC$ a nemusíme se o něj dále zajímat.

V opačném případě je nutné bod P lokalizovat do nějakého z trojúhelníků T_i, z nichž se $\triangle ABC$ skládá. Vzhledem k tomu, že jednotlivé $\triangle T_i$ jsou hierarchicky uspořádány v kvadrantovém stromu trojúhelníku ABC, bude nejlepší lokalizaci provádět postupným průchodem daného stromu. Začneme v kořeni (víme, že $P \in \triangle ABC$) a určíme, do kterého ze čtyř trojúhelníků, na které byl původní trojúhelník rozdělen, bod P náleží. Ve stromu sestoupíme do odpovídajícího uzlu a pokud se nejedná o list, postup opakujeme. Bod P náleží do elementárního $\triangle T_i$, který odpovídá listu, do kterého jsme se při průchodu stromem dostali. Zbývá ještě popsat, jakým způsobem přecházíme od daného trojúhelníku k jedné z jeho čtyř částí. Označme $P_{\triangle ABC} = (a, b, c)$ barycentrické souřadnice bodu P vůči $\triangle ABC$. Postup je pak následující (viz obrázek 4.5):

$$\begin{aligned} a > 1/2 \;\Rightarrow\; P \in \triangle AFE \;&\Rightarrow\; P_{\triangle AFE} = 2 \cdot (a, b, c) - (1, 0, 0) \\ b > 1/2 \;\Rightarrow\; P \in \triangle FBD \;&\Rightarrow\; P_{\triangle FBD} = 2 \cdot (a, b, c) - (0, 1, 0) \\ c > 1/2 \;\Rightarrow\; P \in \triangle EDC \;&\Rightarrow\; P_{\triangle EDC} = 2 \cdot (a, b, c) - (0, 0, 1) \\ \text{jinak} \;\Rightarrow\; P \in \triangle DEF \;&\Rightarrow\; P_{\triangle DEF} = (1, 1, 1) - 2 \cdot (a, b, c) \end{aligned} \tag{4.8}$$



Obr. 4.5: Lokalizace bodu

4.3.4 Zobecnění na trojúhelníkové pláty

V předchozích odstavcích jsme se zabývali použitím trojúhelníků pro aproximaci povrchu jednotlivých těles scény. Oproti standardně používaným obdélníkům mají trojúhelníky některé výhody (např. možnost triangulace obecných povrchů či jednodušší lokalizace bodu), stále se však jedná o rovinné plošky. Použitím tzv. virtuálních hran lze zobrazit i zakřivené povrchy, aproximace je ale značně složitá a nepřesná. Zkusme proto pro popis povrchu použít obecnějších nerovinných ploch.