

Pro gradu -tutkielma

Valaistusyhtälö

Markus Kettunen

Toukokuu 2011

HELSINGIN YLIOPISTO
Matematiikan ja tilastotieteen laitos



Tiedekunta/Osasto – Fakultet/Sektion – Faculty Matemaattis-luonnontieteellinen		Laitos – Institution – Department Matematiikan ja tilastotieteen laitos	
Tekijä – Författare – Author Markus Kettunen			
Työn nimi – Arbetets titel – Title Valaistusyhtälö			
Oppiaine – Läroämne – Subject Soveltava matematiikka			
Työn laji – Arbetets art – Level Pro gradu -tutkielma		Aika – Datum – Month and year Toukokuu 2011	Sivumäärä – Sidoantal – Number of pages 59 + 8
Tiivistelmä – Referat – Abstract			
<p>Tässä tietokonegrafiikan alan tutkielmassa tutkitaan annetun kohteen tai maailman valaistuksen määrittämistä laskennallisesti. Ilmiöt kuvataan fysikaalisesti ja valaistusta mallinnetaan ilmiöitä kuvaavilla yhtälöillä. Yhtälöiden ratkaisu perustuu valonsäteiden etenemisen ja sironnan laskennalliseen seurantaan. Näin johdettua matemaattista mallia tutkitaan ja yhtälöiden ratkaisu muotoillaan tietokoneella laskettavaan muotoon. Lopuksi ohjelmoidaan esiteltyä teoriaa hyödyntävä numeerinen ratkaisija ja esitellään käytetyt menetelmät ja tulokset.</p> <p>Valaistuksen matemaattista mallintamista lähestytään fysikaaliselta pohjalta. Aluksi esitellään valon luonnetta ja yhteyttä sähkömagneettiseen säteilyyn ja ihmisen näköjärjestelmään. Tämän jälkeen paneudutaan valonsäteiden heijastumiseen ja sirontaan ja yleistetään klassiset ideaaliset sirontamallit huomioimaan pinnan hienorakenne sirontajakaumilla.</p> <p>Tutkittu ongelma muotoillaan matemaattisiksi yhtälöiksi jotka ratkaistaan analyttisesti. Ratkaisu osoittautuu intuitiiviseksi: valaistus saadaan laskemalla valonlähteistä emittoituvan valon kaikkien kertalukujen heijastukset.</p> <p>Analyttinen ratkaisu muotoillaan rekursiiviseksi ja ratkaistaan todennäköisyyslaskentaan perustuvalla Monte Carlo -integroinnilla, jonka suppenemista nopeutetaan tärkeysotannalla. Numeerinen ratkaisu osoitetaan odotusarvoisesti harhattomaksi ja ratkaisun virheen osoitetaan puolittuvan laskentapisteen määrän nelinkertaistuessa. Käytettävä todennäköisyyslaskenta esitellään pääpiirteittäin.</p> <p>Numeerinen ratkaisumenetelmä on stokastista säteenseurantaa yleistävä polunseuranta. Maailma määritellään kolmioverkko ja pintojen normaalit annetaan kolmioiden kärkipisteissä. Kolmioista muodostetaan kuvaus tasopinnalle, josta voidaan tarvittaessa lukea esimerkiksi pinnan tarkemmat normaalit, sirontaominaisuudet tai absorptiospektri.</p> <p>Numeerisen ratkaisun eniten aikaa vievä osuus on valonsäteen seuraavan osuapisteiden selvitys maailman pintojen välillä. Ratkaisua nopeutetaan tallentamalla maailman kolmiot tehokkaaseen tietorakenteeseen, kd-puuhun, joka mahdollistaa valonsäteen ja suurten kolmiojoukkojen nopeat leikkaustarkistukset. Kd-puun ajatus ja toteutus esitellään työssä lyhyesti.</p> <p>Lopuksi esitellään ratkaisun eri vaiheet ja teoria käytännössä ja nähdään konkreettisesti eri menetelmien merkitys numeerisen ratkaisijan tuottamaan kuvaan. Lisäksi esitellään tehokas prioriteettijonoon perustuva adaptiivinen menetelmä kuvaan jääneen kohinan pienentämiseksi tutkimalla näytteistykseen otoskeskihajonnan ja keskiarvon suhdetta kuvapisteittäin.</p>			
Avainsanat – Nyckelord – Keywords valaistus, simulaatio, säteenseuranta, sirontajakauma, tärkeysotanta, kd-puu			
Säilytyspaikka – Förvaringställe – Where deposited Helsingin yliopiston matematiikan ja tilastotieteen laitos			
Muita tietoja – Övriga uppgifter – Additional information			

Sisältö

1	Johdanto	1
2	Fysikaaliset prosessit	5
2.1	Valo ja väri	5
2.2	Sädeoptiikka	7
2.3	Radiometria	11
2.3.1	Säteilyvuo	12
2.3.2	Irradianssi	12
2.3.3	Intensiteetti	12
2.3.4	Radianssi	12
2.4	Sirontajakaumafunktiot	12
2.4.1	Sirontajakauma	13
2.4.2	Heijastus	13
2.4.3	Tahtuminen ja läpinäkyvyys	14
2.4.4	Emissio	14
3	Valaistusyhtälö	15
3.1	Määrittely	15
3.2	Analyyttinen ratkaisu	17
4	Plenoptisen funktion mittaaminen	19
4.1	Neulansilmäkamera	19
4.2	Linssisysteemit ja syväterävyys	20

5	Numeerinen ratkaisu	23
5.1	Maailman approksimointi	23
5.1.1	Barysentriset koordinaatit	24
5.1.2	UV-kartat	25
5.1.3	Sirontajakaumat ja teksturointi	26
5.1.4	Normaalikartta	27
5.1.5	Ympäristökartta	29
5.2	Todennäköisyyslaskentaa	30
5.2.1	Satunnaismuuttujat	30
5.2.2	Todennäköisyysjakaumat	30
5.2.3	Odotusarvo ja varianssi	31
5.2.4	Monte Carlo -menetelmät	32
5.2.5	Tärkeysotanta	33
5.2.6	Sirontajakaumaesimerkkejä	35
5.2.7	Monitärkeysotanta	37
5.3	Törmäystarkistukset	38
5.3.1	Kolmion leikkaustarkistus	38
5.3.2	Kd-puu	38
5.4	Klassinen säteenseuranta	41
5.5	Stokastinen säteenseuranta	42
5.6	Polunseuranta	43
5.6.1	Venäläinen ruletti	43
5.6.2	Menetelmä	44
5.6.3	Muotoilu polkuina	45
6	Toteutus	47
6.1	Kd-puu	48
6.2	Teksturointi	48
6.3	Heijastukset	49

<i>SISÄLTÖ</i>	vii
6.4 Antialiasointi	50
6.5 Normaalikartta	50
6.6 Emissiokartta	50
6.7 Sirontajakauma	51
6.8 Adaptio	52
6.9 Heikkouksia	54
6.10 HDR-kuvaus ja bloom	54
7 Yhteenveto	57

Luku 1

Johdanto

Tässä tietokonegrafiikkaa käsittelevässä soveltavan matematiikan tutkielmassa esitellään menetelmä valokuvamaisten kuvien muodostamiseen tietokoneella. Menetelmä perustuu valonsäteiden matemaattiseen seurantaan maailmassa, joka on määriteltä sopivalla tietokoneohjelmalla. Tällaista laskennallista kuvanmuodostusta käytetään esimerkiksi peleissä ja elokuvissa, mutta tärkeitä käyttökohteita ovat myös esimerkiksi arkkitehtuuri ja tuotesuunnittelu.

Käytettävä menetelmä perustuu valon simulointiin radiometrian keinoin. Malli on yleisempi kuin pelkkään sädeoptiikkaan perustuvat mallit mutta liian karkea selittämään joitakin kvanttimekaanisia ilmiöitä, kuten kaksoisrakokoetta ja valon polarisaatiota. Radiometriaan perustuva malli on kuitenkin huomattavasti kvanttimekaanisia menetelmiä soveltuvampi numeeriseen laskentaan.

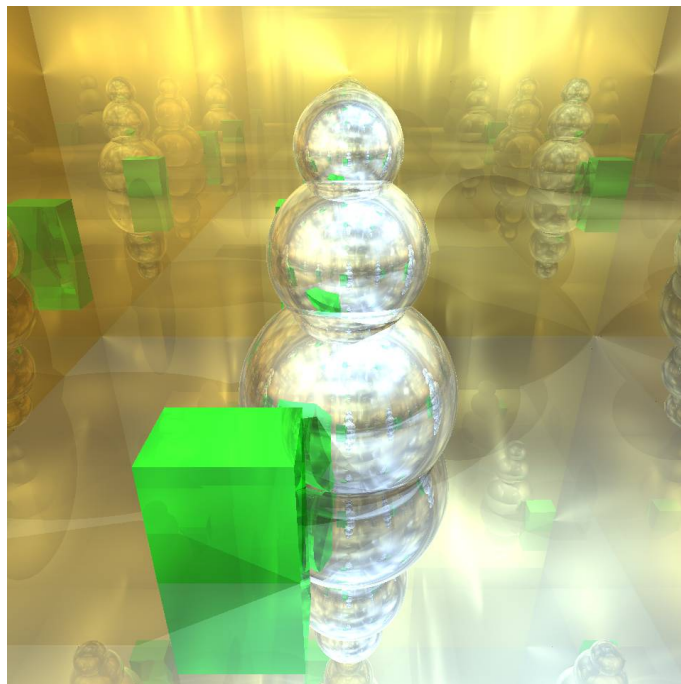
Radiometrian periaatteista johdettua valaistuksen määrittävää integraaliyhtälöä tutkitaan ratkaisemalla se analyttisesti operaattorisarjana. Ratkaisu auttaa ymmärtämään ongelman luonnetta ja kannustaa numeeriseen ratkaisuun stokastisella Monte Carlo -integroinnilla. Työssä esitellään todennäköisyyslaskentaa alkeista monitärkeisotantaan ja johdetaan näihin tietoihin perustuva ohjelmallinen ratkaisija.

Ratkaisu perustuu stokastisen säteenseurannan laajennokseen, polunseurantaan. Siinä missä klassinen säteenseuranta mallintaa lähinnä terävät heijastukset ja piste-mäiset valonlähteet (kuvat 1.1 – 1.2), sirontajakaumia käyttävä polunseuranta mallintaa myös valoa taittavien pintojen polttopintavalaistuksen (*caustic*), epäsuoran valaistuksen, valon hajoamisen prismassa sekä pehmeät varjot ja heijastukset (kuvat 1.3 – 1.4).

Vaikka malli ei sellaisenaan riitäkään täysin realistisen valaistuksen laskemiseen kaikissa olosuhteissa, se toimii erinomaisesti johdatuksena alan peruseriaatteiden ymmärtämiseen. Lisäksi tärkeimmät puuttuvat ilmiöt, kuten valon käyttäytyminen sumussa, savussa sekä ihon ja lumen kaltaisissa materiaaleissa, voidaan helposti liittää malliin.



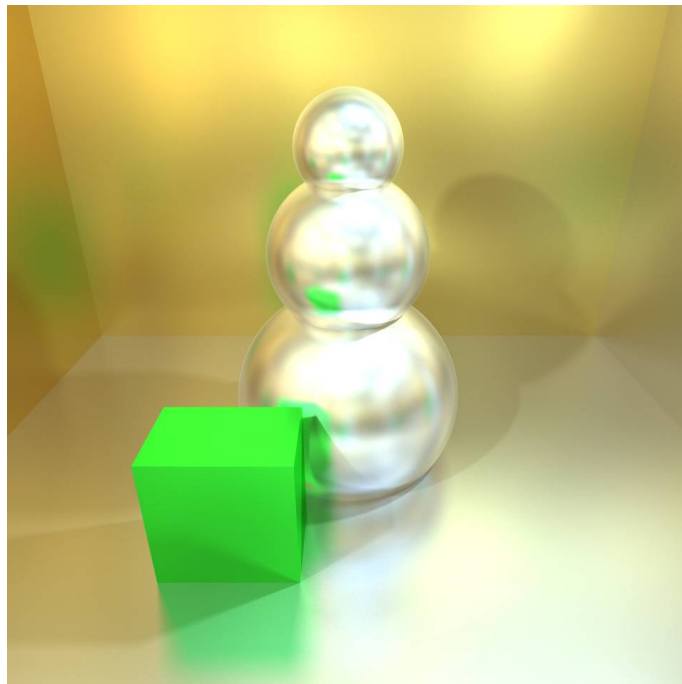
Kuva 1.1: Klassisella säteenseurannalla laskettu kuva kuvitteellisesta höyrypanssarivaunusta. Heijastukset ja varjot ovat teräviä. (Malli: Jussi Kekäläinen)



Kuva 1.2: Klassinen säteenseuranta ei kykene tuottamaan kovinkaan realistisia kuvia mutta toimii pohjana kehittyneemmille algoritmeille.



Kuva 1.3: Polunseuranta mahdollistaa laajat valonlähteet ja epäsuoran valaistuksen. (Vertaa kuvaan 1.1.)



Kuva 1.4: Pehmeät varjot ja sumeat heijastukset seuraavat työssä johdetusta mallista luonnollisesti.

Luku 2

Fysikaaliset prosessit

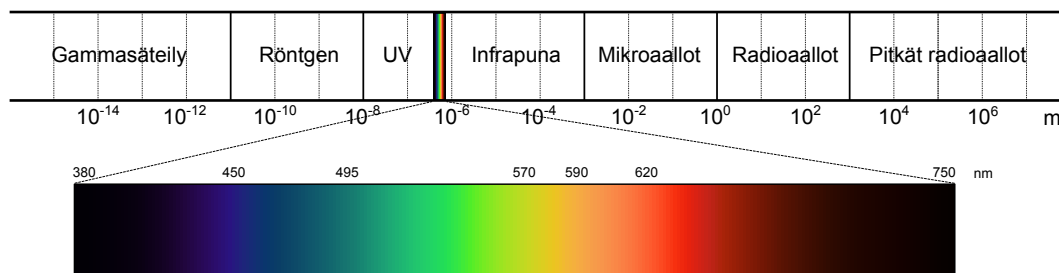
2.1 Valo ja väri

Sähkömagneettisen säteilyn, siten myös valon, välittäjähiukkanen on fotoni. Valo on hiukkasvirtaa ja sen aallonpituuksien jakaumaa kutsutaan spektriiksi. Klassinen valonsädemalli on nähtävä suuren mittakaavan approksimaationa.

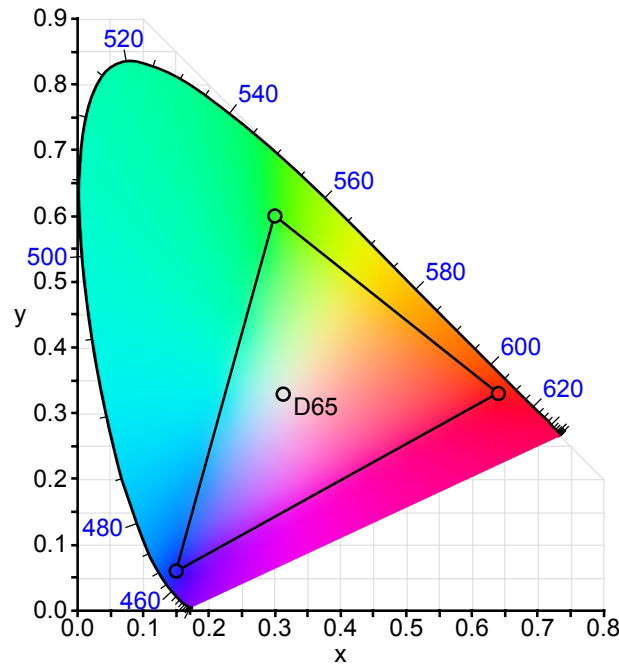
Ihmisen näköaisti on biokemiallinen järjestelmä, jonka ytimessä ovat silmän verkkokalvon sauva- ja tappisolut. Sauvasolut eivät havaitse lainkaan värejä, mutta ne mahdollistavat pimeänäön. Niitä on ihmisellä normaalisti noin 92 miljoonaa ja ne sijaitsevat pääasiassa verkkokalvon ulommissa osissa (Curcio et al. 1990).

Värinäkö perustuu silmän tappisoluihin. Niitä on kolmea eri tyyppiä, joista kukin reagoi eri suhteissa eri aallonpituuksiin. Tappisoluja on noin 4,6 miljoonaa ja ne ovat keskittyneet pääasiassa verkkokalvon keskiosaan, tarkan näön alueelle (Curcio et al. 1990). Yhdessä silmän tappisolut kykenevät havaitsemaan likimain aallonpituudet 390 nm – 750 nm, jota kutsutaan näkyvän valon alueeksi (kuva 2.1).

Usein valosta puhuttaessa jaetaan väri kolmeen komponenttiin: punaiseen, vihreään



Kuva 2.1: Sähkömagneettisen säteilyn luokittelu aallonpituuden funktiona ja näkyvän valon alue ihmissilmän näkemänä väreinä.

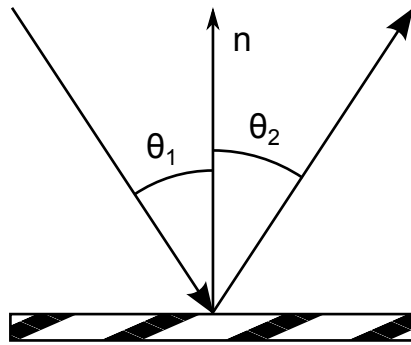


Kuva 2.2: Punainen, vihreä ja sininen sRGB-väriavaruudessa. Kolmion sisällä niiden avulla muodostettavat värit ja keskellä valkoinen piste. (Kuva: Wikimedia Commons)

ja siniseen. Jos kullekin näistä väreistä kiinnitetään jokin spektri, voidaan näitä värejä yhdistämällä stimuloida silmän tappisoluja tavalla, joka mahdollistaa suuren värijoukon näkemisen. Näin muodostettavien värien joukkoa kutsutaan väriavaruudeksi (kuva 2.2). Vaikka menetelmä on tehokas ja käytössä lähes kaikkialla, ei kolmella tai millään muullakaan määrällä päävärejä voida muodostaa kaikkia silmän näkemiä värejä.

Koska tietokoneiden näytöt näyttävät kuvan kolmella päävärillä, usein myös numeerisessa laskennassa käytetään valonsäteiden spektrinä tällaista kolmen komponentin jakoa. Tämä johtaa kuitenkin väärään lopputulokseen, sillä materiaalin väri perustuu eri aallonpituuksien absorbointiin eri voimakkuuksilla. Myös valon dispersio prismassa sateenkaaren väreiksi jää mallintamatta.

Yksinkertaisuutensa ja laskennallisen keveytensä vuoksi tällaista kolmen pääkomponentin jakoa käytetään kuitenkin yleisesti esimerkiksi reaaliaikaisessa kolmiulotteisessa grafiikassa, jossa realismia on usein huomattavasti kustannustehokkaampaa lisätä muilla menetelmillä kuin spektrin komponenttijakoa tihentämällä.



Kuva 2.3: Valonsäteen heijastuminen rajapinnasta. Ideaaliselle materiaalille $\theta_2 = \theta_1$.

2.2 Sädeoptiikka

Heijastuminen

Klassisesti valonsäteen törmätessä pintaan oletetaan monesti yksinkertaisissa laskuissa valonsäteen symmetrinen heijastuminen pinnan normaalin suhteen (kuva 2.3). Laadukkaille peileille tämä pitää hyvin paikkansa, mutta yleisesti heijastus noudattaa jotakin jakaumaa tarkan heijastuksen sijaan. Tämä jakauma on usein, joskaan ei aina, keskittynyt ideaalin heijastuksen ympärille ja aikaansaa myös epäsuoran valaistuksen.

Joskus yksinkertaiset tai reaaliaikaiset valaistuksen ratkaisijat olettavat kaikkien heijastusten olevan symmetrisiä heijastuksia normaalin suhteen ja approksimoivat epäsuoraa valoa jollakin muulla keinolla. Tämä on toisinaan kohtuullinen kompromissiratkaisu, sillä vaikka tulos onkin useimmiten virheellinen, symmetrisen heijastuksen suunta on monille materiaaleille kaikkein tärkein yksittäinen suunta.

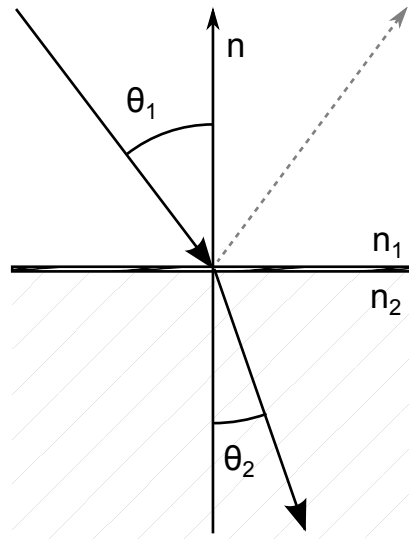
Taittuminen

Valonsäteen kohdatessa läpikuultavan pinnan osa valosta taittuu Snellin lain mukaisesti. Normaaliin nähden luetun tulokulman ollessa θ_1 saadaan taittumiskulma θ_2 yhtälöstä

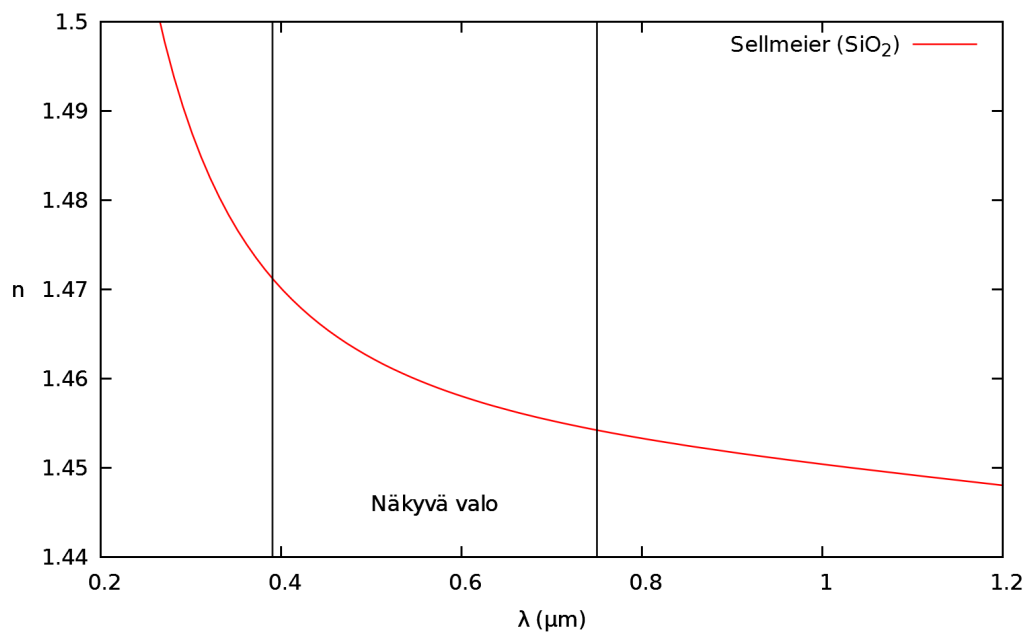
$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2} = \frac{n_2}{n_1}$$

kuvan 2.4 mukaisesti. Väliaineen i taitekerroin n_i määrittää valon keskimääräisen nopeuden kyseisessä väliaineessa: $v_i = c/n_i$.

Mikäli tällaista kulmaa θ_2 ei yhtälön mukaan ole olemassa, tapahtuu taittumisen sijaan *kokonaisheijastus*.



Kuva 2.4: Valonsäteen taittuminen rajapinnassa. Ideaaliselle materiaalille θ_2 saadaan Snellin laista.



Kuva 2.5: Valon aallonpituuden vaikutus kvartsilasin taitekertoimeen.

Valon matalampi keskimääräinen nopeus väliaineessa selittyy fotonien toistuvalla absorboitumisella ja emittoitumisella väliaineen hiukkasista. Tästä seurauksena on *dispersio* eli valon nopeuden ja siten taitekertoimen riippuminen valon taajuudesta (kuva 2.5). Tästä seuraa valon hajoaminen spektriä rajapinnassa eli *kromaattinen aberratio*, joka on monesti ongelma esimerkiksi valokuvauksessa vähemmän kehittyneillä linseillä.

Taitekertoimet voidaan selvittää kokeellisesti, mutta dispersion mallintamiseen tarvitaan tapa laskea taitekerroin erikseen kullekin aallonpituudelle. Tähän erinomaisen apuneuvon tarjoaa Sellmeierin kaava,

$$n^2(\lambda) = 1 + \sum_{j=1}^k \frac{A_j \lambda^2}{\lambda^2 - B_j^2},$$

jossa $k = 3$ on useimmiten riittävä (Tatian 1984).

Aine	n (589 nm)	A_1	B_1	A_2	B_2	A_3	B_3	A_4	B_4
Tyhjiö	1								
Ilma	1,0003								
Kvartsilasi (SiO ₂) ¹	1,4584	0,66347	0,06652	0,44065	0,11502	0,89901	9,9032		
Pii (Si) ¹	3,9314	10,6684	0,30152	0,00304	1,13475	1,54133	1104,00		
Safiiri (Al ₂ O ₃) ¹	1,7682	0,53148	0,02179	1,55060	0,10335	5,28958	17,9393		
Vesi (H ₂ O) ²	1,3332	0,5689	0,07149	0,17197	0,13510	0,02063	0,16199	0,11240	3,26727

Taulukko 2.1: Taitekertoimia ja Sellmeierin kaavan kertoimia. Lukujen B_j yksikkö on μm .

Monesti riittää käyttää epätarkempaa Cauchyn yhtälöä

$$n(\lambda) = A + \frac{B}{\lambda^2},$$

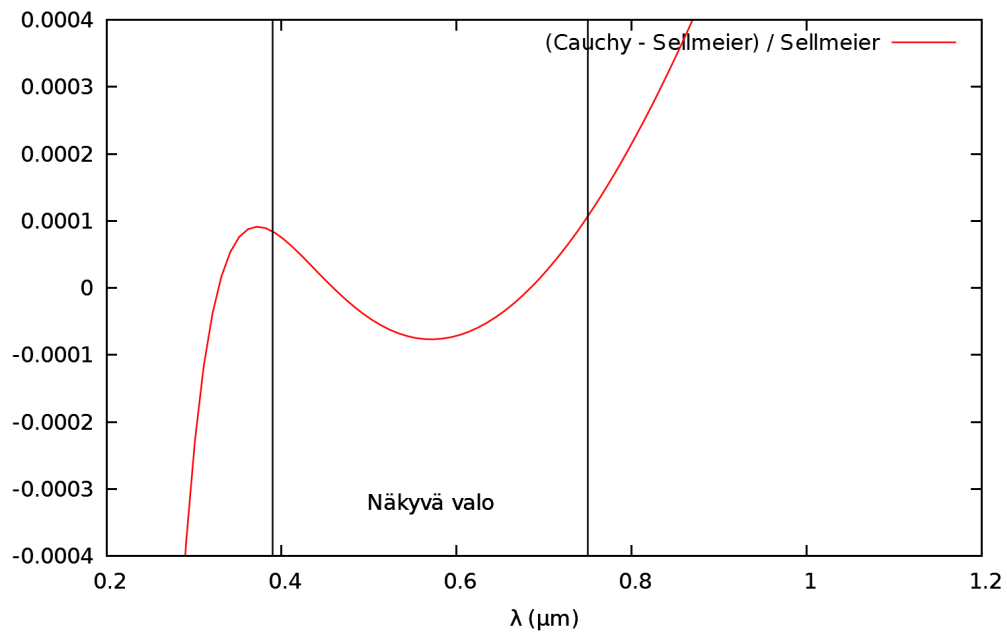
joka näkyvän valon alueelle sovitettuna antaa usein jo varsin hyvän approksimaation (kuva 2.6).

Valon kulkiessa tasaisen eri taitekertoimen levyn läpi valo taittuu mennessään sisään väliaineeseen. Poistuessaan väliaineesta saman suuntaisen rajapinnan läpi, valo taittuu takaisin alkuperäiseen suuntaan. Väliaineessa kuljettu matka aiheuttaa valonsäteiden siirtymisen sivusuunnassa eli *yhdensuuntaissiirtymän*. Koska taitekerroin riippuu valon taajuudesta, eri aallonpituudet siirtyvät eri verran (kuva 2.7). Tällä voi olla merkitystä joidenkin sovellusten kannalta.

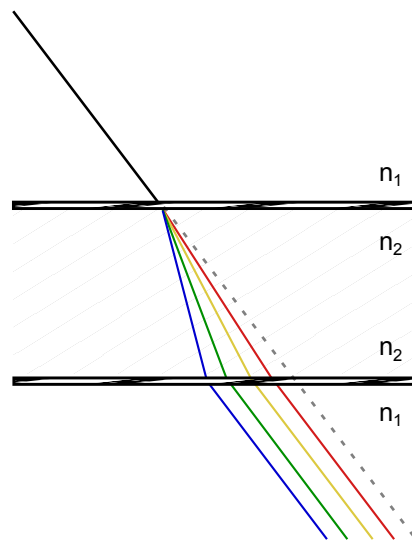
Myös valon taittumista rajapinnassa voidaan mallintaa taittumisjakaumalla. Näin saadaan paremmin huomioitua materiaalin hienorakenteen ja epätäydellisyyden aiheuttamat poikkeamat taittumiseen esimerkiksi sumean lasin tapauksessa.

¹Tatian (1984)

²Daimon and Masumura (2007)



Kuva 2.6: Cauchy-approksimaation $n = 1,4481 + 0,00354 \frac{(\mu\text{m})^2}{\lambda^2}$ suhteellinen virhe kvartsilasille.



Kuva 2.7: Havainnekuva eri aallonpituuksien yhdensuuntaissiirtymästä ja kromaattisesta aberraatiosta.

Fresnel-yhtälöt

Todellisuudessa myös läpikuultavissa pinnoissa tapahtuu sekä heijastumista että taittumista. Näiden välinen suhde kuvataan usein *heijastuskertoimella* F_r , joka saadaan Fresnel-yhtälöistä.

Polarisoimattoman valon tapauksessa heijastuvan valon osuudelle koko sironnasta saadaan arvio (Pharr and Humphreys 2004)

$$F_r = \frac{r_{\parallel}^2 + r_{\perp}^2}{2},$$

jossa heijastuskertoimet r_{\parallel}^2 ja r_{\perp}^2 riippuvat väliaineen johtavuudesta. Eristeissä heijastuskertoimille voidaan käyttää arvioita

$$\begin{aligned} r_{\parallel}^2 &= \left(\frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \right)^2 \\ r_{\perp}^2 &= \left(\frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \right)^2, \end{aligned}$$

jolloin taittuvan valon osuus sironnasta on $1 - F_r$.

Johteet eivät läpäise valoa mutta absorboivat sitä. Heijastuvan valon osuuden F_r laskemiseksi voidaan käyttää kaavoja

$$\begin{aligned} r_{\parallel}^2 &= \frac{(n_2^2 + k^2) \cos^2 \theta_1 - 2n_2 \cos \theta_1 + 1}{(n_2^2 + k^2) \cos^2 \theta_1 + 2n_2 \cos \theta_1 + 1} \\ r_{\perp}^2 &= \frac{(n_2^2 + k^2) - 2n_2 \cos \theta_1 + \cos^2 \theta_1}{(n_2^2 + k^2) + 2n_2 \cos \theta_1 + \cos^2 \theta_1}, \end{aligned}$$

jossa k on johteen *absorptiokerroin* (Pharr and Humphreys 2004).

Polarisaation huomiotta jättäminen voi joissakin tilanteissa johtaa virheelliseen lopputulokseen mutta yksinkertaistaa laskuja huomattavasti.

2.3 Radiometria

Yleistetään teoria sädeoptiikan yksittäisistä valonsäteistä säteilyenergian yleisempään käsittelyyn määrittelemällä aluksi radiometrian peruskäsitteet.

2.3.1 Säteilyvuo

Säteilyvuolla Φ tarkoitetaan jonkin pinnan läpi kulkevaa säteilyenergiaa aikayksikköä kohti. Säteilyvuo on siten säteilyn pinnalle aiheuttama kokonaisteho ja sen yksikkö on watti.

2.3.2 Irradianssi

Irradianssilla E tarkoitetaan sähkövuon pinta-alatiheyttä, $E = d\Phi/dA$. Pinnan säteilyvuo saadaan siten vaihtoehtoisesti integroimalla irradianssia pinnan yli. Irradianssin yksikkö on W/m^2 .

Jos pinta ei osoita kohti valonlähdetä, kerää pinta säteilyä valonlähteeseen nähdessä kohtisuoralta pinta-alalta $A^\perp = A \cos \theta$. Tällöin irradianssi $I = d\Phi/dA = (d\Phi/dA^\perp) \cos \theta$. Tätä yhtälöä kutsutaan *Lambertin laiksi*.

2.3.3 Intensiteetti

Intensiteetillä I tarkoitetaan sähkövuon avaruuskulmatiheyttä, $I = d\Phi/d\omega$. Kappaleen säteilyvuon voi siten laskea yhtäpitävästi integroimalla säteilylähteen intensiteettiä kappaleen peittämän avaruuskulman Ω yli,

$$\Phi_A = \int_{\Omega} I(\omega) d\omega.$$

2.3.4 Radianssi

Radianssilla L tarkoitetaan pienestä avaruuskulmasta pienelle kohtisuoralle pinnalle saapuvaa tai lähtevää irradianssia yksikköavaruuskulmaa kohti, $L = dE/d\Omega = d\Phi/(d\omega dA^\perp)$. Muut radiometriset suureet ovat johdettavissa radianssista.

Ellei toisin mainita, käsitellään jatkossa aallonpituutta kiinnitettynä vakiona. Lopullinen tulos saadaan yhdistämällä eri aallonpituuksien tulokset.

2.4 Sirontajakaumafunktiot

Valonsäteiden sirontaa epäideaaleista pinnoista on hyvä kuvailla sirontajakaumilla. Johdetaan seuraavaksi sirontajakaumia käyttäen sirontayhtälöt suurelle joukolle materiaaleja.

Merkitään avaruuden pisteeseen p suunnasta ω_i saapuvaa radianssia $L_i(p, \omega_i)$. Tämä radianssi aiheuttaa pisteessä p sijaitsevaan kohtisuoraan pintaan yksikköavaruuskulmaa kohti irradianssin $dE/d\Omega(p, \omega_i) = L_i(p, \omega_i)$. Lambertin lain mukaan kappaleeseen kohdistuva irradianssi yksikköavaruuskulmaa kohti on siten $L_i(p, \omega_i) \cos \theta$.

Pintaan kohdistuva irradianssi voidaan nyt laskea integroimalla nämä differentiaaliset irradianssit kaikkien suuntien yli, jolloin

$$E(p) = \int_{S^2} L_i(p, \omega_i) |\cos \theta| d\omega_i,$$

jossa θ on pinnan normaalin ja ω_i :n välinen kulma ja $\cos \theta = N_p \cdot \omega_i$.

2.4.1 Sirontajakauma

Pisteeseen p suunnasta ω_i saapuva differentiaalinen irradianssi jakautuu eri suuntiin jonkin jakauman $f(p, \omega_o, \omega_i)$ mukaisesti aiheuttaen kuhunkin suuntaan differentiaalisen radianssin. Tämä suuntaan ω_o lähtevä radianssi saadaan integroimalla differentiaalista radianssia kaikkien tulosuuntien yli

$$L_o(p, \omega_o) = \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos \theta| d\omega_i. \quad (2.1)$$

Tätä avaruuden pintojen pisteissä määriteltyä jakaumafunktiota f kutsutaan nimellä *kaksisuuntainen sirontajakaumafunktio*.

Määritelmä 1. Kaksisuuntainen sirontajakaumafunktio (*Bidirectional Scattering Distribution Function, BSDF*) on funktio $P \times S^2 \times S^2 \rightarrow \mathbb{R}_+$, jolle

- i) $f(p, \omega_o, \omega_i) = f(p, \omega_i, \omega_o)$
- ii) $\sup_{p, \omega} \int_{S^2} f(p, \omega, \omega_i) |\cos \theta| d\omega_i < 1$.

Ensimmäinen ehto tarkoittaa valonsäteiden käyttäytymisen suuntainvarianssia eli kaksisuuntaisuutta. Jälkimmäinen ehto on energian säilymlaki epäideaaleille pinnoille.

2.4.2 Heijastus

Jos pinta on valoa läpäisemätön, on $L_i(p, \omega_i) = 0$, kun $\theta \geq 90^\circ$. Tällöin kaavassa (2.1) riittää kirjoittaa integraali pinnan puoleisen puolipallon $\mathfrak{H}^2 = \{\omega \in S^2 \mid \omega \cdot N_p > 0\}$ yli:

$$L_o(p, \omega_o) = \int_{\mathfrak{H}^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos \theta d\omega_i.$$

Tällöin funktiota f kutsutaan *kaksisuuntaiseksi heijastusjakaumafunktioksi* (*Bidirectional Reflectance Distribution Function, BRDF*).

2.4.3 Taittuminen ja läpinäkyvyys

Monet pinnat ovat läpikuultavia. Tällaisille pinnoille edellinen läpikuultamaton heijastusehto ei käy. Läpikuultavuutta kuvataan samankaltaisella konstruktiolla, jossa jakaumafunktiona on *kaksisuuntainen läpäisyjakaumafunktio* (*Bidirectional Transmittance Distribution Function, BTDF*). Kaksisuuntainen läpäisyjakaumafunktio kertoo, kuinka suuren differentiaalisen radianssin suuntaan ω_o aiheuttaa pintaan suunnasta ω_i saapuva differentiaalinen irradianssi. Toisin sanoen integroiden normaalia vasten olevan puolipallon yli

$$L_o(p, \omega_o) = \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos \theta| d\omega_i.$$

Monet pinnat ovat kuitenkin vain osittain läpinäkyviä. Tällöin taittumisen ja heijastumisen voi yhdistää sirontajakaumaksi.

2.4.4 Emissio

Jotkin pinnat eivät ainoastaan sirota valoa vaan myös lähettävät eli emittoivat sitä. Tällaisille pinnoille lähtevän radianssin yhtälöön lisätään emissiotermi emissiojakauman $L_e(p, \omega_o)$ mukaisesti, jolloin saadaan *sirontayhtälö* (*scattering equation*),

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos \theta| d\omega_i.$$

Tämä yhtälö mallintaa niin emittoivat, läpinäkymättömät kuin myös osittain läpikuultavat materiaalit. Kuitenkin monet tärkeät materiaalit, kuten lumi tai iho, tarvitsisivat monimutkaisempaa käsittelyä (Pharr and Humphreys 2004). Näissä materiaaleissa valonsäde siroaa materiaalin sisään, kimpoilee hetken ympäriinsä ja lopuksi siroaa jostakin muusta kohdasta ulos.

Tällaiset materiaalit voidaan mallintaa käyttämällä *kaksisuuntaista pintasirontajakaumafunktiota* (*bidirectional surface scattering reflectance distribution function, BSSRDF*). Tällöin integroitaisiin pisteen ympäristössä radianssia pintasirontajakaumaa vasten,

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_A \int_{S^2} S(p', \omega_i, p, \omega_o) |N_{p'} \cdot \omega_i| d\omega_i dA.$$

Tässä työssä jätetään kuitenkin tällaiset materiaalit vaille erityishuomiota. Lisätietoja pintasirontajakaumafunktioista löytyy esimerkiksi viitteestä Pharr and Humphreys (2004).

Luku 3

Valaistusyhtälö

Valonsäteiden mallintaminen kvanttitaso ulkopuolella on periaatteessa yksinkertaista: tyhjiössä valonsäde kulkee suoraan, kunnes kohtaa esteen. Este absorboi osan valosta ja loppu siroaa mitattujen tai laskettujen jakaumien mukaan eri suuntiin. Käytännön rajoitus äärellisestä fotonimäärästä voidaan unohtaa, sillä tarkoituksena on tutkia ideaalista, odotusarvoista tilannetta.

3.1 Määrittely

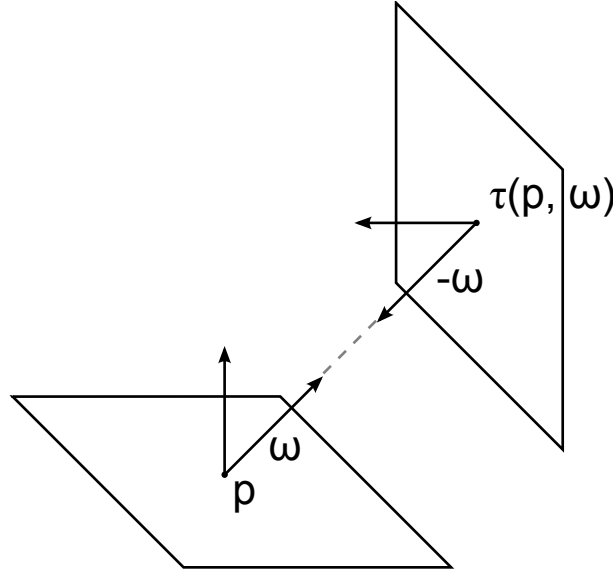
Johdetaan edellä esitettyjen tietojen perusteella yhtälö radianssille avaruuden eri pisteissä. Tätä valaistuksen määrittelevää ratkaisufunktiota L kutsutaan *plenoptiseksi funktioksi*.

Määritelmä 2. Plenoptinen funktio (*plenoptic function*) on funktio $L : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}_+$, jolle $L(p, \omega)$ on pisteeseen p suunnasta ω saapuva radianssi.

Maailman muodostaa mallinnettavien kappaleiden pisteistä ja äärettömässä sijaitsevasta pisteestä Ω_∞ koostuva pistejoukko $P \subset \overline{\mathbb{R}^3} = \mathbb{R}^3 \cup \{\Omega_\infty\}$. Jokaiselle joukon P pisteelle oletetaan lisäksi määritellyksi pinnan normaali N_p , sirontajakaumafunktio $f(p, \omega_o, \omega_i)$ ja emissiojakauma $L_e(p, \omega)$. Äärettömässä olevalle pisteelle sirontajakauma oletetaan nolaksi, mutta emissiojakaumalla voidaan mallintaa esimerkiksi kaukaisen ympäristön, kuten taivaan, valaistusta. Äärettömän pisteen normaalilla ei ole merkitystä.

Määritelmä 3. Maailma (*scene*) on nelikko (P, N, L_e, f) , jossa $P \subset \overline{\mathbb{R}^3}$ ja $\Omega_\infty \in P$, N on kuvaus $P \rightarrow S^2$, L_e on kuvaus $P \times S^2 \rightarrow \mathbb{R}_+$ ja f on joukon P kaksisuuntainen sirontajakaumafunktio.

Valonsäteet etenevät maailmassa suoraan, kunnes törmäävät johonkin. Törmäyspisteen selvittämistä varten tarvitaan *seurantafunktio*.



Kuva 3.1: Valon suora eteneminen avaruudessa seurantafunktion τ avulla.

Määritelmä 4. Seurantafunktio (*raycasting function*) on funktio $\tau : \mathbb{R}^3 \times S^2 \rightarrow \overline{\mathbb{R}^3}$, jolle $\tau(p, \omega) = p + \omega t$, jossa t on pienin positiivinen luku, jolla $p + \omega t \in P$. Mikäli tällaista lukua t ei ole, määritellään $\tau(p, \omega) = \Omega_\infty$. Määritellään lisäksi lyhennysmerkintä $p_\omega = \tau(p, \omega)$.

Seurantafunktion avulla voimme esittää matemaattisesti säteiden suoran etenemisen avaruudessa pintojen välillä (kuva 3.1):

$$L_i(p, \omega) = L_o(\tau(p, \omega), -\omega).$$

Tämä yhtälö yhdistettynä pintojen sirontayhtälöön

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos \theta| d\omega_i$$

muodostaa ratkaistavan yhtälösystemin. Plenoptinen funktio on L_i .

Valaistuksen määrittelevä integraaliyhtälö on siten väliaineettomassa ja pintasironnattomassa maailmassa

$$L_i(p, \omega) = L_e(p_\omega, -\omega) + \int_{S^2} f(p_\omega, -\omega, \omega_i) L_i(p_\omega, \omega_i) |N_{p_\omega} \cdot \omega_i| d\omega_i.$$

Tarkkaan ottaen tämä yhtälö ei vielä sellaisenaan määrittele koko valaistusta vaan ainoastaan säteilyenergian jakautumisen maailmassa. Valaistus saadaan tästä ratkaisemalla yhtälö jokaiselle aallonpituudelle λ_i erikseen — näin saadaan huomioitua myös valon aallonpituuden vaikutus valon jakautumiseen maailmassa, kunhan jokaiselle aallonpituudelle käytetään kyseisen aallonpituuden sirontajakaumafunktiota.

3.2 Analyttinen ratkaisu

Määritellään lineaarinen operaattori A , jolle

$$(AL_i)(p, \omega) = \int_{S^2} f(p_\omega, -\omega, \omega_i) |N_{p_\omega} \cdot \omega_i| L_i(p_\omega, \omega_i) d\omega_i.$$

Määritellään lisäksi $g(p, \omega) = L_e(p_\omega, -\omega)$. Tällöin yhtälö voidaan kirjoittaa muodossa $L_i(p, \omega) = g(p, \omega) + AL_i(p, \omega)$, tai yksinkertaisemmin

$$L_i = g + AL_i.$$

Tutkitaan seuraavaksi yhtälön ratkaisua analysoimalla operaattorin A normia. Lausekkeen $AL_i(p, \omega)$ integraalimuoto on positiivisten funktioiden tulon L_1 -normi, joten Hölderin epäyhtälöstä saadaan arvio

$$AL_i(p, \omega) \leq \int_{S^2} f(p_\omega, -\omega, \omega_i) |N_{p_\omega} \cdot \omega_i| d\omega_i \cdot |L_i(p_\omega)|_\infty \leq s |L_i(p_\omega)|_\infty,$$

jossa

$$s = \sup_{q, \omega} \int_{S^2} f(q, \omega, \omega_i) |N_q \cdot \omega_i| d\omega_i < 1.$$

Koska edellä oleva epäyhtälö pätee kaikille p ja ω , pätee myös $|AL_i|_\infty \leq s |L_i|_\infty$ ja siten operaattorin A normi $|A| \leq s < 1$.

Kirjoitetaan sitten yhtälö $L_i = g + AL_i$ muodossa $(I - A)L_i = g$, jossa I on identiteettioperaattori. Koska $|A| < 1$, tiedetään integraaliyhtälöiden teoriasta, että operaattorin $I - A$ käänteisoperaattori $(I - A)^{-1}$ saadaan Neumannin sarjalla $I + A + A^2 + A^3 + \dots$. Näin saadaan yhtälön ratkaisu:

$$L_i = (I + A + A^2 + \dots)g = g + Ag + A^2g + A^3g + \dots.$$

Ratkaisun tulkinta on varsin intuitiivinen: valaistus saadaan laskemalla yhteen suora valaistus ja sen kaikkien kertalukujen heijastukset.

Luku 4

Plenoptisen funktion mittaaminen

Valaistuksen laskemisessa on kaksi jonkin verran toisistaan poikkeavaa päätavoitetta. Ensimmäinen on plenoptisen funktion selvittäminen avaruuden kaikissa pisteissä. Toinen on pyrkimys kykyyn muodostaa laskennallisesti mahdollisimman realistisia valokuvamaisia kuvia.

Vaikka tavoitteet eivät ole täysin samat, ne eroavat vain vähän, sillä plenoptisen funktion tunteminen mahdollistaa valokuvatyylisten kuvien muodostamisen. Toisaalta valokuva toimii plenoptisen funktion approksimaationa.

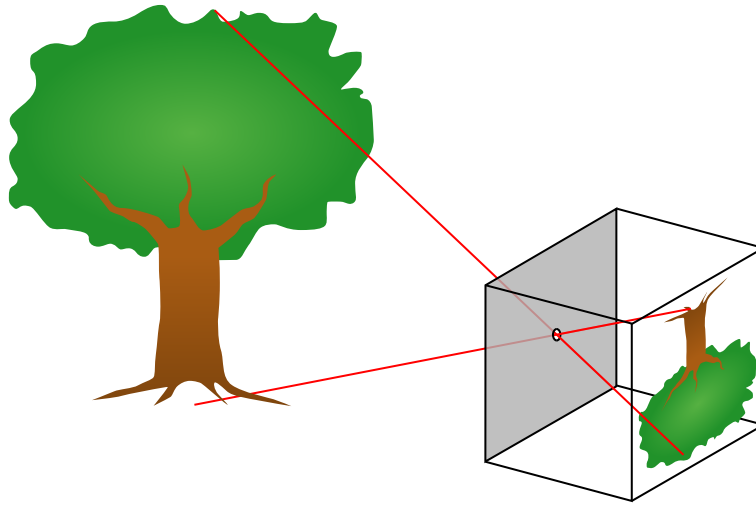
Mitatessa plenoptista funktiota L_i fyysisestä maailmasta, joudutaan valon hiukkasluonteen takia tyytymään approksimaatioon jonkinlaisella valonkeräimellä, kameralla. Yksinkertaisin tällainen kamera on neulansilmäkamera.

4.1 Neulansilmäkamera

Neulansilmäkamerassa maailmasta saapuva valo kulkee kameran päädyssä olevan pienen reiän läpi vastaanottavalle pinnalle kameran sisällä. Pinnalle muodostuu ylösalainen projektio maailmasta (kuva 4.1).

Ideaalisessa neulansilmäkamerassa kameran kuvapinnalle tai filmille kuhunkin pisteeseen pääsisi valoa vain yhdestä suunnasta. Tällöin plenoptisen funktion voisi konstruoida projisoimalla kuvapinnalle kerätyn valon takaisin neulansilmäpisteeseen.

Ideaalinen neulansilmäkamera on kuitenkin puhtaasti teoreettinen konstruktio ja kovin pienen reiän tapauksessa kvantti-ilmiöt ja reiän läpi kulkevan valon pieni määrä sotkisivat kuvauksen. Käytännössä reiän tulee olla suurempi, jolloin myös kuvasta tulee epäterävä, sillä jokaisen kuvan pisteen määrittää saapuvan radianssin integraali jonkin avaruuskulman yli.



Kuva 4.1: Neulansilmäkamera. Kuva muodostuu kameran takaosaan ylösalaisin. (Kuva: Wikimedia Commons)

Neulansilmäkamera on hyvä malli plenoptisen funktion teoreettiseen selvittämiseen mutta käytännön mittaamiseen lähes käyttökelvoton. Mallia käytetään kuvan muodostukseen plenoptisen funktion selvittämisen lisäksi sellaisissa numeerisissa ratkaisijoissa, joissa realismi ei ole laskennan tavoitteena tai joissa ratkaisu tarvitaan nopeasti ja suhteellisen vähällä laskennalla. Tärkeä tällainen sovellus on esimerkiksi pelien ja ohjelmien reaaliajassa laskettava kolmiulotteinen grafiikka.

4.2 Linssisysteemit ja syväterävyys

Käytännöllisempi tapa muodostaa kuva on kerätä valonsäteitä suuremmalta pinta-alalta ja kohdistaa ne linssillä tai kokoelmalla linsejä kuvapinnalle. Tämä menetelmä on nopea ja tehokas, sillä suuri linssi kerää paljon valoa. Menetelmässä on kuitenkin kaksi enemmän tai vähemmän suurta heikkoutta plenoptisen funktion määrittämisen kannalta.

Ensimmäinen, paljon pienempi haitta, on se, että lähellä olevan kappaleen jostakin pisteestä eri kohtiin linssiä saapuvat valonsäteet ovat sironneet eri kulmiin, joten ne ovat sironneet sirontajakauman eri kertoimilla. Kun linssi kokoaa nämä säteet takaisin yhteen pisteeseen, mitataankin keskiarvoistettua radianssia. Näin ollen plenoptisen funktion konstruoinnissa aiheutuu vähintään pientä epätarkkuutta. Tämä ongelma ei yleensä ole kovin merkittävä, sillä kappaleiden etäisyys kamerasta on tavanomaisesti suuri suhteessa linssin halkaisijaan. Tällöin kulmaero on pieni.

Suurempi heikkous projektion muodostamisessa linssikameralla on se, että linssi ei kykene tarkentamaan suurelta pinta-alalta kerättyjä säteitä pienelle alueelle kaikilta



Kuva 4.2: Syväterävyysalueen ulkopuoliset alueet näkyvät sumentuneina.

etäisyyksiltä samanaikaisesti (kuva 4.2). Tätä ilmiötä kutsutaan syväterävyydeksi (*depth of field*) ja se esiintyy myös ihmisen näköjärjestelmässä. Vaikka teknisesti kyse on projektion muodostamisen suhteen heikkoudesta, sen mallintaminen numeerisessa kuvanmuodostuksessa on tärkeää realistisen ja hyvältä näyttävän kuvan muodostamiseksi.

Kumpaankin ongelmaan auttaa kameran aukon pienentäminen, jolloin valoa pääsee sisään pienemmästä kulmasta. Haittapuolena on valon määrän väheneminen ja siten valotusajan kasvu.

Luku 5

Numeerinen ratkaisu

Yksinkertaistetunkin valaistusyhtälön numeerinen ratkaisu on haastavaa. Haasteena ei ole niinkään ratkaisun numeerinen epästabiilius, vaan vuorovaikutusten suuresta määrästä johtuva laskennan hitaus: lähtökohtaisesti jokainen maailman piste voi vaikuttaa jokaiseen muuhun maailman pisteeseen.

Valaistusyhtälön muotoisen integraaliyhtälön ratkaisu perustuu usein joko takaisinsijoitukseen tai operaattorisarjaratkaisuun. Takaisinsijoituksessa pinnasta lähtevään valoon lisätään jokaisella kierroksella seuraavan kertaluvun heijastukset. Sarjaratkaisussa lasketaan kerralla suora valaistus ja sen kaikkien kertalukujen heijastukset. Ensimmäinen johtaa *radiosityn* kaltaisiin menetelmiin, jälkimmäinen säteenseurantaan.

Tavanomaisissa maailmoissa nämä menetelmät suppenevat vaaditusta laadusta riippuen kohtuullisen nopeasti, sillä useimmat pinnat absorboivat suuren osan niihin saapuvasta valosta. Kuitenkin maailmat, jotka koostuvat esimerkiksi erittäin heijastavista peili- tai lasipinnoista, vaativat suuren määrän iteraatioita: valonsäteet kimpoilevat ympäriinsä pitkän aikaa ennen kuin niiden energia pienenee merkittäväksi.

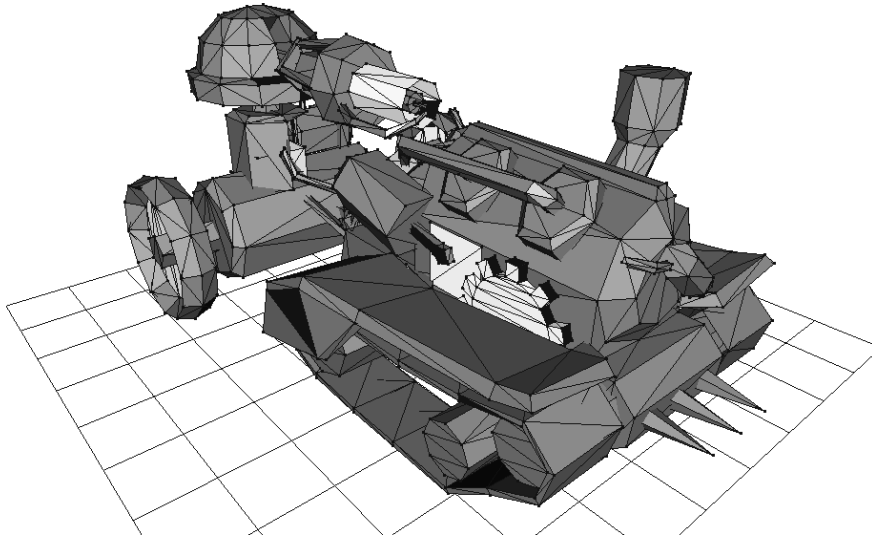
Myös integroitavan funktion approksimointi tehokkaammilla integrointimenetelmillä on haasteellista, sillä seurantafunktio τ tekee integrandista vahvasti epäjatkuvan. Varsinkin sarjaratkaisussa plenoptisen funktion moniparametrisuus ja korkeamman kertaluvun sirontaintegraalien korkeat ulottuvuudet aiheuttavat perinteisesti tehokkaampien integrointimenetelmien jäämisen Monte Carlo -menetelmien jalkoihin.

5.1 Maailman approksimointi

Äärettömän pistejoukon tallentaminen tietokoneen muistiin ei tietenkään ole mahdollista. Jäljelle jäävät pääasiassa seuraavat kaksi vaihtoehtoa: koko maailman dis-

kreointi ja tallennus muistinkäytön kannalta tehokkaaseen puurakenteeseen (esim. octree tai kd-puu), tai pintojen approksimointi paloittain parametrisilla pinnoilla tai kolmioilla. Keskitytään näistä yleisimpään ratkaisuun eli kolmiointiin (kuva 5.1).

Pintojen pistejoukon P lisäksi tarvitsee jokaiselle pisteelle määrittää normaali N_p . Kolmioiden tapauksessa normaalit määritellään tavanomaisesti kolmioiden kärjissä ja interpoloidaan pinnan pisteisiin.



Kuva 5.1: Höyrypanssarivaunun matalapolygoniversiön kolmioverkkoa. Malli: Jussi Kekäläinen.

5.1.1 Barysentriset koordinaatit

Interpolointi kolmion kärkipisteistä sisäpisteisiin tapahtuu usein käyttäen *barysentrisiä koordinaatteja*. Mikäli kolmion piste p on kolmion (p_1, p_2, p_3) pinnalla, on helppoa löytää α_1 , α_2 ja α_3 , joille

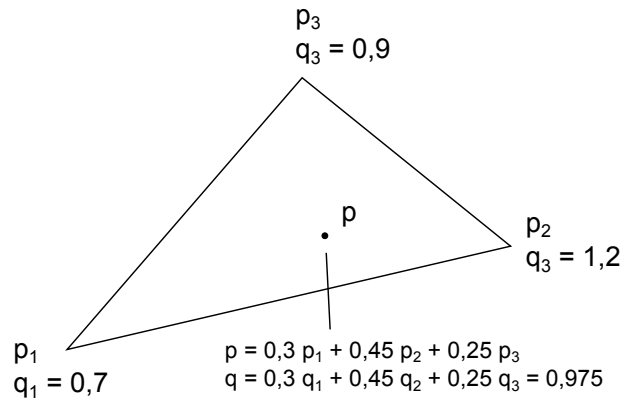
$$p = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3.$$

Lisäksi on helppoa nähdä, että tällöin $\alpha_1, \alpha_2, \alpha_3 \geq 0$ ja $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Kertoimet $(\alpha_1, \alpha_2, \alpha_3)$ ovat pisteen p barysentriset koordinaatit.

Barysentristen koordinaattien avulla mikä tahansa kolmion kärkien ominaisuus q voidaan interpoloida lineaarisesti kolmion pisteisiin seuraavasti (kuva 5.2):

$$q(p) = \alpha_1 q(p_1) + \alpha_2 q(p_2) + \alpha_3 q(p_3).$$

Varsinkin normaalien tapauksessa tulee huolehtia, että pinnan kolmioverkko on riittävän tiheä, jotta kunkin kolmion kärkipisteissä määritellyt normaalit osoittavat



Kuva 5.2: Kolmion pisteen p barysentriset koordinaatit ja kärkien ominaisuus q interpoloituina pisteeseen p .

suunnilleen samaan suuntaan. Tällöin lineaarisestakin approksimaatiosta tulee tarkka, eivätkä normaalit voi summautua nolllaksi. Lopuksi interpoloidut normaalit normitetaan.

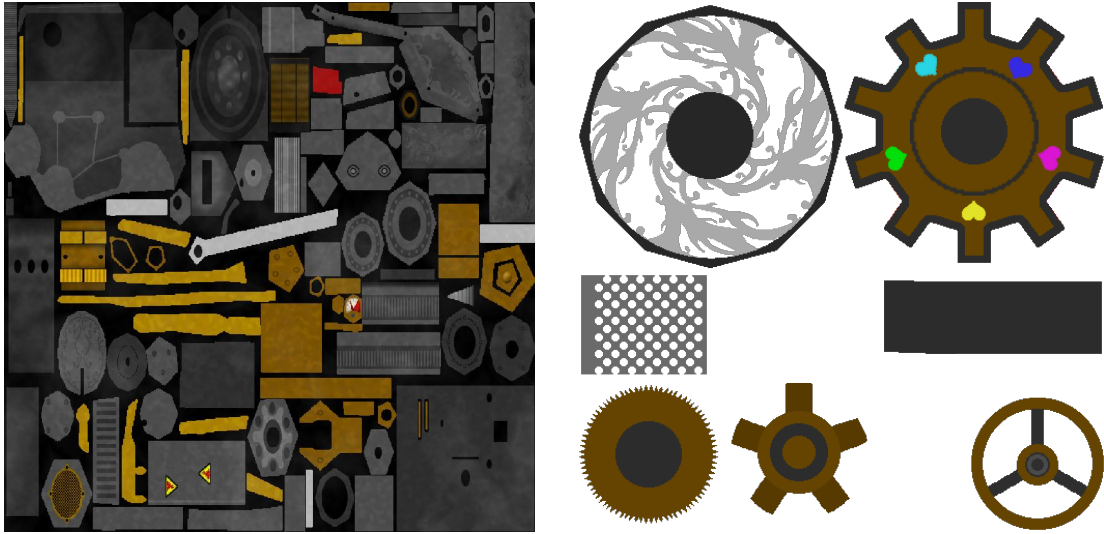
5.1.2 UV-kartat

Periaatteessa kolmiojakoa tihentämällä saadaan mielivaltaisen tarkkoja tuloksia. Usein on kuitenkin laskennallisesti merkittävästi tehokkaampaa ja datan antamisen kannalta helpompaa liittää ominaisuuksia suoraan kolmion pisteisiin ulkoisella kartalla. Periaatteessa olisi mahdollista määrittää ominaisuudet suoraan avaruuden \mathbb{R}^3 pisteisiin, mutta on käytännöllisempää ja muistinkäytöllisesti merkittävästi halvempaa ensin kuvata kolmioiden pisteet tasolle ja liittää ominaisuus näihin kuvapisteesiin.

Kuvataan ensin kukin avaruuden kolmio affiinisti johonkin tason \mathbb{R}^2 osajoukkoon. Tämän kuvauksen määrittävät täydellisesti pisteiden p_1 , p_2 ja p_3 kuvat (u_1, v_1) , (u_2, v_2) ja (u_3, v_3) . Näitä pisteitä on tapana ajatella kolmion kärkipisteen ominaisuuksina, jolloin kuhunkin kärkeen on tämän jälkeen liitetty pisteen $p = (x, y, z)$ lisäksi normaali $N_p = (n_x, n_y, n_z)$ ja pisteen kuva tasokartalla, $uv(p) = (u, v)$. Näin saadaan jokaiselle pisteelle p uv -koordinaatit:

$$uv(p) = \alpha_1 uv(p_1) + \alpha_2 uv(p_2) + \alpha_3 uv(p_3).$$

Tämä kuvaus mahdollistaa yhdessä barysentristen koordinaattien kanssa mielivaltaisten ominaisuuksien liittämisen kolmioiden pisteisiin: Olkoon f mielivaltainen kartta jostakin tason \mathbb{R}^2 osajoukosta haluttuun ominaisuusjoukkoon. Tällöin pisteen p ominaisuus voidaan lukea kartasta f kohdasta $uv(p)$.



Kuva 5.3: Hörypanssarivaunun pintavärit tekstuurikarttoina.

Käytännössä kartat annetaan useimmiten taulukoina tai esimerkiksi kuvatiedostoina, joista pisteiden p ominaisuudet $f(uv(p))$ luetaan sopivasti interpoloituna tai suodatettuna.

5.1.3 Sirontajakaumat ja teksturointi

Useimmiten kappaleiden kolmioihin liitetään *tekstuurikartta*, joka määrittää kappaleen värin kolmella päävärillä (kuva 5.3). Todellisuudessa valo koostuu suuresta määrästä aallonpituuksia, joiden jakauma riippuu valonlähteestä ja siitä, mitä valolle on matkalla tapahtunut. Kolmen aallonpituuden approksimaatio vastaa koko näkyvän spektrin jakamista kolmeen osaan ja johtaa tässä vaiheessa laskentaa joissakin tapauksissa merkittävästi vääriin lopputulokseen.

Kolmen päävärin etuna on kuitenkin datan keräämisen helppous, sillä pinnan väritys on mahdollista taltioida esimerkiksi kameralla tai tehdä tietokoneella kuvankäsittelyohjelmalla. Näissä tarkoituksissa kolmen värikanavan approksimaatio on erinomainen, sillä datan käyttötarkoitus on katselu ihmissilmällä. Valaistuksen laskemiseen sen sijaan tarvitaan koko spektri, ja jako kolmeen pääväriin on mahdollista tehdä vasta monimutkaisten spektrilaskujen jälkeen. Koko kappaleen spektrin mittaaminen joka pisteessä kuitenkin vaatii kehittyneempiä työkaluja, joita ei useinkaan ole saatavilla. Tämän lisäksi kappaleen heijastama spektri riippuu sekä valon tulo- että lähtösunnista; tehtäväksi jää joko mitata tai arvioida koko sirontajakauma jokaisessa pisteessä kaikille aallonpituuksille.

Koska sirontajakauman sisältämän informaation määrä on valtava, käytännössä joudutaan useimmiten tyytymään karkeisiin approksimaatioihin. Usein käytetty kei-

no on sirontajakaumien approksimointi muutamilla parametrisoiduilla sirontajakaumaperheillä, joiden parametrit valitaan vastaamaan mahdollisimman tarkasti todellisuutta.

Usein maailman kolmiot jaetaan aluksi eri materiaaleista koostuviin osiin, joista kullekin tallennetaan materiaalin käyttämä sirontajakaumaperhe ja sen parametrit. Näitä sirontajakaumia voidaan moduloida esimerkiksi kertomalla ne tekstuuri- tai spektrikartasta saadulla kertoimella. On myös mahdollista varioida kappaleen materiaalin sirontaominaisuuksia kuvaamalla sirontajakauman parametrit kappaleelle esimerkiksi UV-kartoituksen avulla.

Läpikuultavien pintojen tapauksessa huomioidaan sirontajakaumassa lisäksi valon taittuminen. Jos materiaali on tunnettu, voidaan taitekerroin kullekin aallonpituudelle laskea Sellmeierin tai Cauchyn kaavalla. Näin saadaan mallinnettua dispersio.

5.1.4 Normaalikartta

Normaalien antaminen kolmioiden kärkipisteissä riittää periaatteessa täysin kuvaamaan kappaleen muodon, mikäli kolmioverkko on riittävän tiheä. Monesti on kuitenkin käytännöllisempää käyttää harvempaa kolmioverkkoa ja antaa tarkemmat normaalit kappaleiden pintoihin UV-kartoituksen avulla.

Tangenttiavaruus

Normaalikartoituksessa kullekin pinnan pisteelle p määritellään pinnan normaali suoraan UV-kartassa. UV-kartan kantavektorit \hat{e}_u ja \hat{e}_v vastaavat standardikantaan kuvattuna joitakin pinnan pisteen p tangenttivektoreita \hat{t}_u ja \hat{t}_v . Määritellään näiden vektoreiden jatkoksi kolmas kantavektori \hat{e}_w , joka vastaa standardikannassa pinnan normaalia $\hat{t}_w = \hat{n}$. Muodostetaan koordinaatisto, jonka origo standardikannassa ilmaistuna sijaitsee pisteessä p ja jonka kantavektorit standardikannassa ovat \hat{t}_u , \hat{t}_v ja \hat{t}_w . Tätä koordinaatistoa kutsutaan *tangenttiavaruudeksi* ja UV-kartan normaalivektorit annetaan tässä avaruudessa.

Seuraavaksi tulisi normaalikartasta luettu normaalivektori $\tilde{n}_{tan} = \hat{e}_u \tilde{n}_u + \hat{e}_v \tilde{n}_v + \hat{e}_w \tilde{n}_w$ kuvata standardikantaan. Mikäli tiedetään edelliset tangenttivektorit, ongelma on jo ratkaistu:

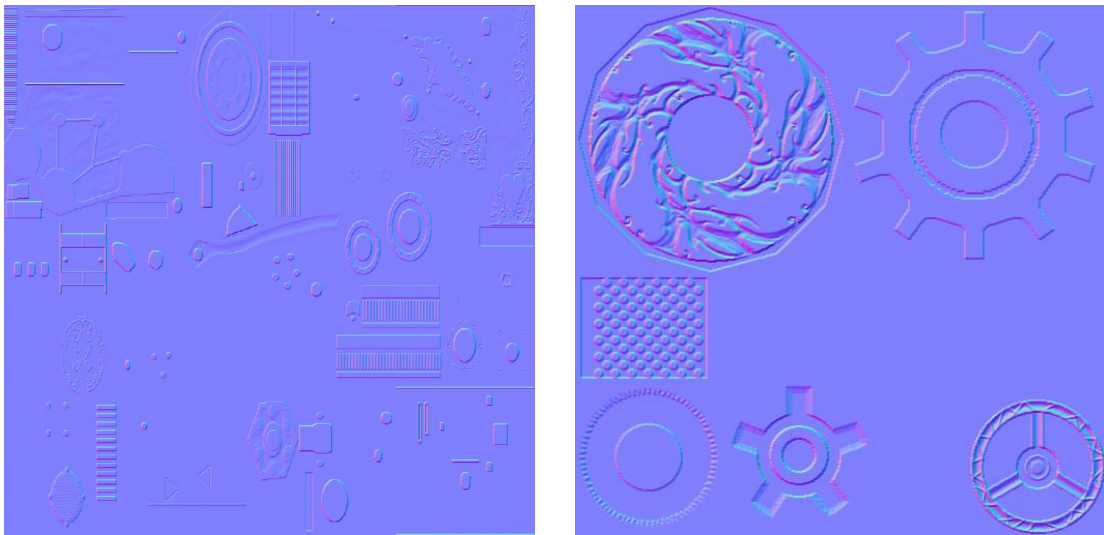
$$\vec{n} = \hat{t}_u \tilde{n}_u + \hat{t}_v \tilde{n}_v + \hat{t}_w \tilde{n}_w.$$

Tämän jälkeen vektori \vec{n} on vielä syytä normittaa, sillä barysentristen koordinaatien avulla kolmion kärjistä interpoloidut kantavektorit eivät välttämättä ole täysin ortonormaaleja.

Normaalikartat

Normaalikarttoja voi tehdä esimerkiksi kuvankäsittelyohjelmissa korkeuskarttojen gradientteina. Toinen vaihtoehto on tehdä tarkempi malli 3D-mallinnusohjelmassa ja pyytää ohjelmaa tekemään karkeamman mallin pintaan liimattava normaalikartta, joka takaa lähes yhtä hyvän tuloksen huomattavasti vähemmällä laskennalla. Tämä menetelmä on käytössä erityisesti reaaliaikaisessa 3D-grafiikassa.

Normaalikartat tallennetaan usein hyvin pakkautuviin kuvaformaatteihin. Tällöin punainen kanava tulkitaan useimmiten x-komponenttina, vihreä y-komponenttina ja sininen z-komponenttina. Koska kuvien väriarvot ovat useimmiten väliltä $[0, 1]$, kuvataan värikomponentit välille $[-1, 1]$ affiinisti. Koska suoraan ulospäin osoittavan normaalin komponenttiesitys on $(0, 0, 1)$, ovat tangenttiavaruuden normaalikartat useimmiten enimmäkseen vaaleansinisiä (kuva 5.4).



Kuva 5.4: Höyrypanssarivaunun normaalikartat kuvina. Punainen komponentti kertoo normaalin x-komponentin, vihreä y-komponentin ja sininen z-komponentin.

Tangenttien laskeminen

Mikäli kolmioiden kärkien tangenttivektoreita ei ole annettu valmiiksi, ne voidaan myös laskea kolmioverkon perusteella. Tällöin etsitään tangenttiavaruuden kantamatriisi T , joka kuvaa kolmion kylkivektorien \vec{p}_{12} ja \vec{p}_{13} kuvat \vec{u}_{12} ja \vec{u}_{13} UV-kartasta takaisin 3D-avaruuteen ja UV-kartan vektorin $(0, 0, 1)$ kolmion normaali-vektoriksi $\vec{p}_{12} \times \vec{p}_{13}$. Normittamalla matriisiin T pystyvektorit saadaan tangenttiavaruuden kantavektorit kolmiolle.

Lopuksi kolmioiden normaalit ja tangentit tulisi siirtää kolmioiden kärjille interpolointia varten. Usein käytetty tapa on laskea jokaiselle kärkipisteelle keskiarvo

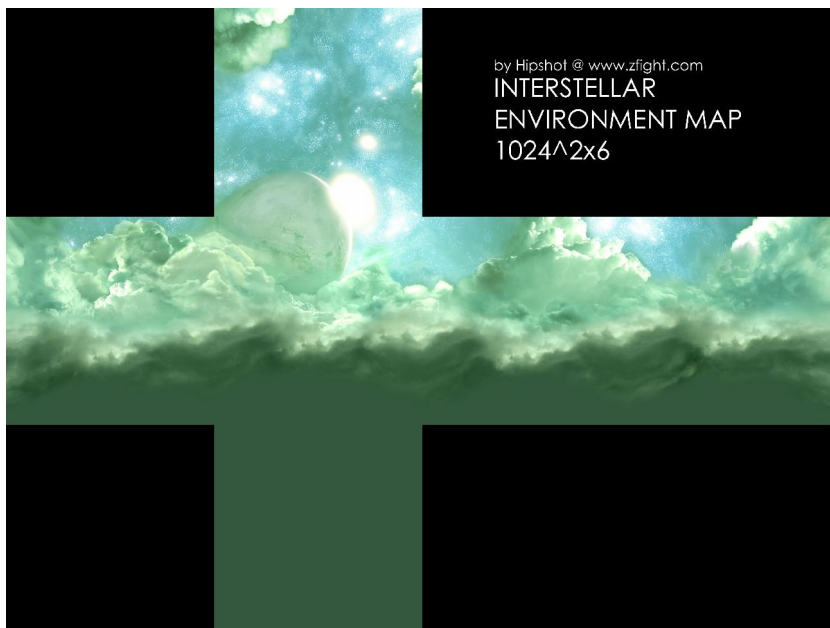
kärkeä koskevien kolmioiden normaaleista ja tangenteista lopuksi normittaan. Toinen tapa on painottaa arvoja kolmioiden pinta-aloilla ennen yhteenlaskua. Tällöin suurempien kolmioiden merkitys korostuu.

5.1.5 Ympäristökartta

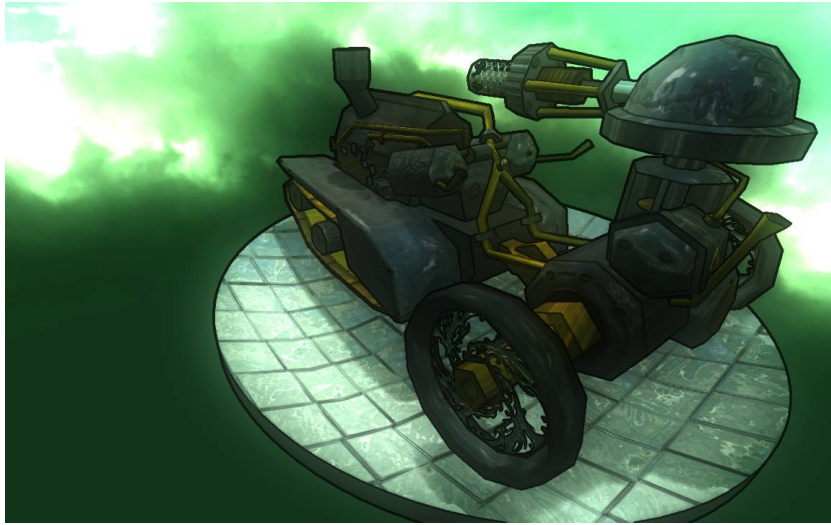
Useimmiten käsiteltävä maailma on kooltaan pieni suhteessa ympäristöön. Ympäristönä voi olla esimerkiksi tähtitaivas, korkealla olevia pilviä tai kaukaista vuoristoa. Tällaisia ei useinkaan ole syytä lähteä mallintamaan erikseen, vaan niiden vaikutuksen voi huomioida ympäristökartalla (*environment map*). Ympäristökartan voi luoda esimerkiksi valokuvien pohjalta tai ohjelmallisesti.

Mikäli seurattavalle säteelle ei mallinnetun maailman alueella löydetä osumakohtaa, voidaan katsoa säteen karanteen äärettömyyteen. Tästä suunnasta valoa tulee ainoastaan kaukaisesta ympäristöstä. Koska ympäristö on kaukana, ainoastaan säteen suunta ω_i on merkitsevä. Tällöin kappaleeseen säteen suunnasta saapuva radianssi voidaan lukea suoraan ympäristökartasta.

Eräs tapa toteuttaa ympäristökartta on kuutiokartta (*cube map*). Kuutiokartassa ympäristökartta (kuvat 5.5 – 5.6) levitetään kuutioksi tarkasteltavan pisteen ympärille ja luetaan saapuva radianssi suuntaan ω_i ammutun säteen osumapisteestä. Muita tapoja ympäristökartan toteuttamiseen ovat esimerkiksi pallokartta (*sphere map*) ja paraboliset ympäristökartat (*parabolic environment map*).



Kuva 5.5: Ympäristökartta kuutiokarttana. (Kuva: Hipshot / zfight.com)



Kuva 5.6: Ympäristökartta reaaliaikaisessa kolmiorasteroinnissa. (Malli: Jussi Kekäläinen)

5.2 Todennäköisyyslaskentaa

Koska valaistussyhtälön numeerisessa ratkaisussa tullaan käyttämään runsaasti todennäköisyyslaskentaa, kerrataan lyhyesti todennäköisyyslaskennan peruseriaatteita intuitiotasolla.

5.2.1 Satunnaismuuttujat

Todennäköisyyslaskennan ydin ovat satunnaismuuttujat. Intuitiivisesti satunnaismuuttuja tarkoittaa nimensä mukaisesti muuttujaa, jonka arvo arvotaan kullakin kerralla satunnaisesti jonkin jakauman mukaan. Tärkein jakauma lienee välin $[0, 1]$ tasajakauma, josta kukin välin $[0, 1]$ luku arvotaan yhtä todennäköisesti.

5.2.2 Todennäköisyysjakaumat

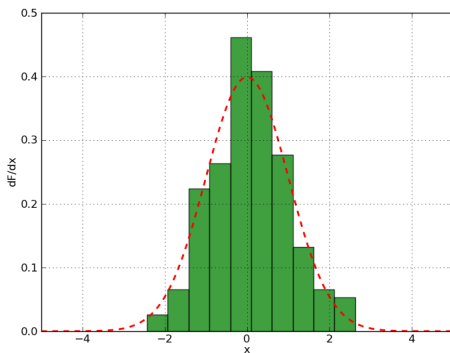
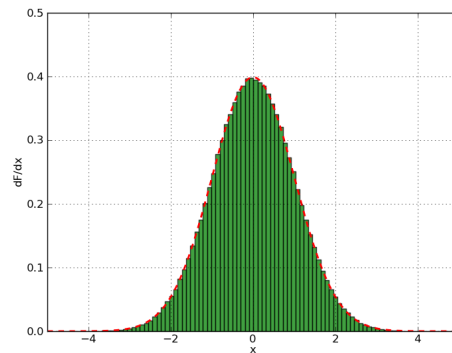
Arvotaan satunnaismuuttujalle X toistuvasti N arvoa peräkkäin ja annetaan jokaiselle ilmentymälle sama pinta-ala $1/N$. Jaetaan avaruus palkkeihin kuvan 5.7 mukaisesti ja kerätään kullekin palkille sen sisältämien satunnaismuuttujan X ilmentymien pinta-alat.

Lukumäärän N kasvaessa rajatta ja palkkien leveyden pienetessä kohti nollaa päästään raja-arvona Riemannin integraalin kaltaiseen pinta-alakonstruktion. Toden-

näköisyysmassan kertymäfunktio on

$$F(x) = P(X < x) = \int_{-\infty}^x f(t) dt,$$

ja kertymäfunktion derivaattana saadaan jakauman tiheysfunktio $f(x)$, joka on piirretty kuvaan 5.7 punaisella katkoviivalla.

(a) $N = 150$ (b) $N = 800000$

Kuva 5.7: Satunnaismuuttujan näytteiden jakauma lähestyy satunnaismuuttujan jakaumaa näytteiden lukumäärän kasvaessa.

5.2.3 Odotusarvo ja varianssi

Useille käytännössä esiintyville todennäköisyysjakaumille satunnaismuuttujan X näytteiden keskiarvo lähestyy jotakin lukua näytteiden määrän N kasvaessa. Tätä raja-arvoa kutsutaan satunnaismuuttujan X odotusarvoksi ja merkitään $E(X)$.

Jatkuvalle jakaumalle, jonka tiheysfunktio on f , odotusarvo määritellään kaavalla

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

ja se toteuttaa yhtälön

$$E(aX + bY) = aE(X) + bE(y).$$

Satunnaismuuttujan X näytteiden keskittymistä odotusarvon ympärille kuvaa satunnaismuuttujan X arvojen keskineliöpoikkeama odotusarvosta,

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx.$$

Varianssin lausekkeen voi kirjoittaa muotoon $\text{var}(X) = E(X^2) - E(X)^2$ ja se toteuttaa riippumattomien satunnaismuuttujien tapauksessa yhtälön

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y).$$

Toinen usein käytetty tunnusluku on keskimääräisen neliöpoikkeaman neliöjuuri, keskihajonta $\sigma(X) = \sqrt{\text{var}(X)}$.

Mikäli satunnaismuuttujan X jakaumaa ei tunneta, mutta siitä kyetään laskemaan näytteitä, voidaan varianssia $\text{var}(X)$ arvioida näytteiden x_1, x_2, \dots, x_N otosvarianssilla

$$\text{var}(X) \approx \frac{\sum_{i=1}^N (x_i - \mu)^2}{N - 1},$$

jossa μ on otoksen keskiarvo.

5.2.4 Monte Carlo -menetelmät

Olkoon tavoitteena laskea numeerisesti likiarvo lausekkeelle $\int_A h(x) dx$ jonkin $|A|$ -pinta-alaisen joukon A yli. Lähestytään integrointia arpomalla joukosta A pisteet x_1, x_2, \dots, x_N tasaisella jakaumalla. Tällöin kukin vaihtoehto on yhtä todennäköinen ja satunnaismuuttujan X tiheysfunktio $f(x) = 1/|A|$.

Edellisen perusteella keskiarvo

$$\frac{\sum_{i=1}^N h(x_i)}{N}$$

suppenee todennäköisyydellä 1 arvoon

$$E(h(X)) = \int_A h(x) \frac{1}{|A|} dx = \frac{1}{|A|} \cdot \int_A h(x) dx.$$

Tästä voidaan päätellä lausekkeen

$$|A| \cdot \frac{\sum_{i=1}^N h(x_i)}{N}$$

suppenevan haluttuun integraaliin $\int_A h(x) dx$.

Teoreettisen suppenemisen lisäksi tulee kyätä arvioimaan menetelmän suppenemisnopeutta. Olkoot satunnaismuuttujat X_i riippumattomia ja samoin jakautuneita kuin X . Tutkitaan keskiarvosatunnaismuuttujan

$$Y = \frac{|A|}{N} \cdot \sum_{i=1}^N h(X_i)$$

poikkeamaa tarkasta arvosta $E(Y) = \int_A h(x) dx$:

$$\begin{aligned} \text{var}(Y) &= \text{var} \left(\frac{|A|}{N} \cdot \sum_{i=1}^N h(X_i) \right) \\ &= \frac{|A|^2}{N^2} \cdot \sum_{i=1}^N \text{var}(h(X_i)) \\ &= |A|^2 \cdot \frac{\text{var} h(X)}{N}. \end{aligned}$$

Tällöin keskiarvon virhe puolittuu näytteiden lukumäärän nelinkertaistuessa:

$$\sigma(Y) = \sqrt{\text{var}(Y)} = |A| \cdot \frac{\sigma(h(X))}{\sqrt{N}}.$$

5.2.5 Tärkeysotanta

Edellä esitelty Monte Carlo -integrointi on tehokas menetelmä useampiulotteisia integraaleja laskettaessa. Koska kuitenkin virheen puolittaminen vaatii näytämäärän nelinkertaistamisen, tulisi virhettä yrittää pienentää myös muilla keinoin mikäli mahdollista.

Ensiapua tähän ongelmaan tuo *tärkeysotanta* (*importance sampling*), joka hyödyntää tietoa integroitavan lausekkeen $h(x)$ muodosta siirtämällä laskentapistettä vähämerkityksisiltä alueilta alueille, joilla funktion h arvo on suuri. Vaikka funktion h muotoa ei tunnetaisi täydellisesti, jo suuntaa-antavasta jakaumasta on paljon hyötyä.

Oletetaan tunnetuksi ei-negatiivinen funktio g , joka approksimoi integroitavaa funktiota h jossakin mielessä. Muodostetaan funktiosta g tiheysfunktio f normittamalla:

$$f(x) = \frac{g(x)}{\int_A g(x) dx}. \quad (5.1)$$

Tällöin arvioitavalle integraalille saadaan yhtälö

$$\int_A h(x) dx = \int_A \frac{h(x)}{f(x)} f(x) dx = E \left[\frac{h(X)}{f(X)} \right],$$

jossa satunnaismuuttujan X tiheysfunktio on f .

Arvotaan lopuksi tämän jakauman X mukaisesti pisteet x_1, x_2, \dots, x_N ja approksimoidaan odotusarvoa otoskeskiarvolla. Tällöin saadaan approksimaatio alkuperäiselle integraalille:

$$E \left[\frac{h(x)}{f(x)} \right] \approx \frac{\sum_{i=1}^N \frac{h(x_i)}{f(x_i)}}{N}.$$

Perustellaan seuraavaksi täsmällisemmin, miksi tärkeysotanta tuo parannusta suppenemiseen. Oletetaan aluksi, että h tunnetaisiin täysin¹. Tällöin voitaisiin muodostaa jakauma tiheysfunktiolla

$$f(x) = \frac{h(x)}{\int_A h(y) dy}.$$

Tällöin olisi

$$\begin{aligned} \text{var}(X) &= E(X^2) - E(X)^2 \\ &= \int_A \frac{h(x)^2}{f(x)^2} f(x) dx - \left[\int_A \frac{h(x)}{f(x)} f(x) dx \right]^2 \\ &= \left[\int_A h(x) dx \right]^2 - \left[\int_A h(x) dx \right]^2 \\ &= 0. \end{aligned}$$

Jo yksikin näyte integroisi funktion täysin oikein:

$$\frac{h(x_i)}{f(x_i)} = \int_A h(x) dx.$$

Aivan näin hyvään lopputulokseen ei tietenkään käytännössä päästä.

Jos merkitään $h(x)/g(x) = e(x)$, nähdään, että varianssin lauseke voidaan kirjoittaa muotoon

$$\begin{aligned} \text{var}(X) &= \int_A \frac{h(x)^2}{f(x)^2} f(x) dx - \left[\int_A \frac{h(x)}{f(x)} f(x) dx \right]^2 \\ &= \int_A \frac{h(x)}{e(x)} dx \int_A h(x)e(x) dx - \int_A h(x) dx \int_A h(x) dx, \end{aligned}$$

sillä yhtälön (5.1) ja funktion $e(x)$ määritelmän mukaan

$$\frac{h(x)^2}{f(x)} = h(x) \frac{h(x) g(x)}{g(x) f(x)} = h(x)e(x) \int_A \frac{h(x)}{e(x)} dx.$$

Tästä voidaan päätellä, että mitä lähempänä funktioiden h ja g suhde on vakiota, sitä pienempi on näytteistykseen kohina.

Konkretisoidaan tätä vielä seuraavasti: Olkoon

$$\frac{|e(x) - c|}{c} \leq k.$$

¹Tämä on tietenkin teoreettinen oletus, sillä mikäli $\int_A h(y) dy$ kyetään laskemaan, ei Monte Carlo -integrointia enää tarvita.

Tällöin

$$\begin{aligned} \text{var}(X) &\leq \int_A \frac{h(x)}{1-k} dx \int_A h(x)(1+k) dx - \int_A h(x) dx \int_A h(x) dx \\ &= \frac{2k}{1-k} \cdot \left[\int_A h(x) dx \right]^2, \end{aligned}$$

ja nähdään varianssin pienenevän nopeasti kohti nollaa jakauman painottuessa kohti parempia otantapisteitä.

5.2.6 Sirontajakaumaesimerkkejä

Kuvantamisessa tärkeysotantaa käytetään sirontayhtälössä

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos \theta| d\omega_i$$

esiintyvän integraalin laskemiseen Monte Carlo -integroinnilla.

Sovellettaessa tärkeysotantaa olisi periaatteessa hyvä tuntea integrandi

$$f(p, \omega_o, \omega_i) L_i(p, \omega_i) |N_p \cdot \omega_i|$$

mahdollisimman tarkasti. Koska tekijää $L_i(p, \omega_i)$ ei käytännössä tunneta etukäteen, joudutaan tyytymään karkeampaan painotukseen. Huomattavasti tyhjää parempaan arvioon päästään pudottamalla tuntematon tekijä lausekkeesta pois ja painottamalla otantaa funktion $f(p, \omega_o, \omega_i) |N_p \cdot \omega_i|$ mukaan.

Nyt kun tiheysjakauma osataan selvittää, on jäljellä kysymys tiheysjakauman mukaisesta vektorien arpomisesta. Havainnollistetaan tätä muutamien esimerkkien kautta.

Diffuusi jakauma

Diffuusi, kaikkiin suuntiin tasaisesti sirottava pinta on yleinen approksimaatio rosoisille pinnoille. Olkoon koordinaatisto sellainen, jossa normaali N_p on vektorin k suuntainen. Tällöin sirontajakauma on $f(p, \omega_o, \omega_i) = k_d$ normaalin puoleisella puolipallolla ja nolla toisella puolella.

Huomioidaan lisäksi tekijä $|N_p \cdot \omega_i| = \cos \theta$ ja etsitään jakauman tiheysfunktio pal-

lokoordinaateissa:

$$\begin{aligned}
 1 &= \int_0^{2\pi} \int_0^{\frac{\pi}{2}} C \cdot k_d \cos \theta \sin \theta \, d\theta \, d\phi \\
 &= C \cdot 2\pi k_d \int_0^{\frac{\pi}{2}} \cos \theta \sin \theta \, d\theta \\
 &= C \cdot \pi k_d [\sin^2 \theta]_0^{\frac{\pi}{2}} \\
 &= C \cdot \pi k_d.
 \end{aligned}$$

Pallokoordinaatiston tiheysfunktio on siten

$$f(\theta, \phi) = \frac{1}{\pi} \cos \theta \sin \theta.$$

Tästä saadaan laskettua $f(\theta)$ marginalisoimalla:

$$f(\theta) = \int_0^{2\pi} \frac{1}{\pi} \cos \theta \sin \theta \, d\phi = 2 \cos \theta \sin \theta.$$

Integroimalla tiheysfunktioita saadaan kertymäfunktio:

$$F(\theta) = \int_0^\theta 2 \cos t \sin t \, dt = \sin^2 \theta.$$

Kertymäfunktion käänteisfunktiona saadaan kuvaus välin $[0, 1]$ tasajakaumalta marginalisoidulle tiheysfunktiolle: $\theta = \arcsin \sqrt{x}$, $x \in [0, 1]$.

Koska jakauma on kulman ϕ suunnassa kiertosymmetrinen, kulma ϕ saadaan helposti tasajakaumasta: $\phi = 2\pi y$, $y \in [0, 1]$.

Tästä saadaan algoritmi diffuusin jakauman otosvektoreille: Olkoon s_1 tasaisesti jakautunut satunnaisluku väliltä $[0, 1]$ ja s_2 tasaisesti jakautunut satunnaisluku väliltä $[0, 2\pi]$. Tällöin vektori

$$\left(\sqrt{s_1} \cos s_2, \sqrt{s_1} \sin s_2, \sqrt{1 - s_1^2} \right)$$

noudattaa puolipallon pinnalla kulman kosinilla painotettua tasajakaumaa.

Kosinieksponttijakauma

Phong-sirontajakaumassa ja sen varianteissa esiintyy usein termi $(\omega_i \cdot \omega_r)^n$, jossa ω_r on vektorin $-\omega_o$ ideaalin heijastuksen suunta. Muodostetaan tämän jakauman mukainen satunnaisvektori koordinaatistossa, jossa vektori k osoittaa suuntaan ω_r .

Tiheysfunktio johdetaan vastaavasti kuin edellä ja tulokseksi saadaan

$$f(\omega) = \frac{n+1}{2\pi} \cos^n \theta.$$

Satunnaisvektori muodostetaan seuraavasti: olkoon s_1 tasaisesti jakautunut satunnaisluku väliltä $[0, 1]$ ja s_2 tasaisesti jakautunut satunnaisluku väliltä $[0, 1]$. Tällöin

$$\begin{aligned}\theta &= \arccos\left(s_1^{\frac{1}{1+n}}\right) \text{ ja} \\ \phi &= 2\pi s_2.\end{aligned}$$

Haluttu satunnaisvektori on

$$(\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta)).$$

Mikäli edellinen vektori muunnettuna alkuperäiseen koordinaatistoon osoittaa väärälle puolelle pintaa, ei sitä voida pitää fysikaalisesti mahdollisena ja se hylätään. Tätä ongelmaa ei olisi, jos edellinen kaava olisi huomionnut tekijän $N_p \cdot \omega_i$. Valitettavasti tekijän huomioiminen monimutkaistaa laskuja merkittävästi.

Tekijän puuttumisella on myös muita seurauksia. Mikäli tärkeysotannalla valittu otosvektori osuu epätodennäköiseen suuntaan, jossa integrandin arvo on suuri, on näytteistykseen arvo $h(x_i)/f(x_i)$ erittäin suuri. Tällöin kuvaan ilmestyy pieni kirkas piste. Tätä harmillista ilmiötä voidaan torjua sekoittamalla tärkeysotannan tiheysfunktioon esimerkiksi tasajakaumaa. Tällöin nimittäjä ei koskaan pääse liian lähelle nollaa ja pahimmilta artifakteilta vältytään.

5.2.7 Monitärkeysotanta

Tärkeysotantaa parantamaan käytetään monesti *monitärkeysotantaa* (*multiple importance sampling*), joka yhdistää tunnetuista painotusjakaumista jakaumia, joiden tiheysfunktiot ovat annettujen tiheysfunktioiden lineaarikombinaatioita. Menetelmää käytetään myös monen jakauman summana määriteltyjen sirontajakaumien integroimiseen.

Olkoot funktiot p_i tunnettujen satunnaismuuttujien X_i tiheysfunktioita ja $c_i \geq 0$ annettuja ykköseksi summautuvia kertoimia. Tarkoituksena on löytää menetelmä tiheysfunktion

$$p(x) = \sum_{i=1}^n c_i p_i(x)$$

määräämän jakauman vektoreiden arpomiseksi.

Menetelmä on varsin yksinkertainen. Asetetaan annetuille jakaumille todennäköisyydet c_i ja arvotaan näistä jokin jakauma. Näin valitusta jakaumasta arvotaan tärkeysotannalla vektori x ja palautetaan se.

Koska kukin jakauma olisi voinut tulla valituksi todennäköisyydellä c_i ja koska kullekin jakaumalle todennäköisyystiheys vektorin x palauttamiselle on $p_i(x)$, on satunnaisvektorin x tiheysfunktio haluttu $p(x) = \sum_i c_i p_i(x)$.

Erinomainen kuvaus monitärkeysotannasta laajemmassa kontekstissa löytyy lähteestä Veach (1998).

5.3 Törmäystarkistukset

Kun muut osat valaistusyhtälöstä osataan ratkaista, jää jäljelle vielä seurantafunktion τ toteutus. Tarvitsisi selvittää annetun säteen lähin osumapiste maailman muodostavien kolmioiden joukosta. Naiivi toteutus tarkistaisi jokaiselle säteelle maailman kolmiot yksitellen.

5.3.1 Kolmion leikkaustarkistus

Kolmion (p_1, p_2, p_3) ja säteen $s + vt$ leikkauspiste lasketaan usein projisoimalla säde kolmion pisteiden muodostamalle tasolle. Mikäli $t > 0$, lasketaan tämän jälkeen leikkauspisteelle

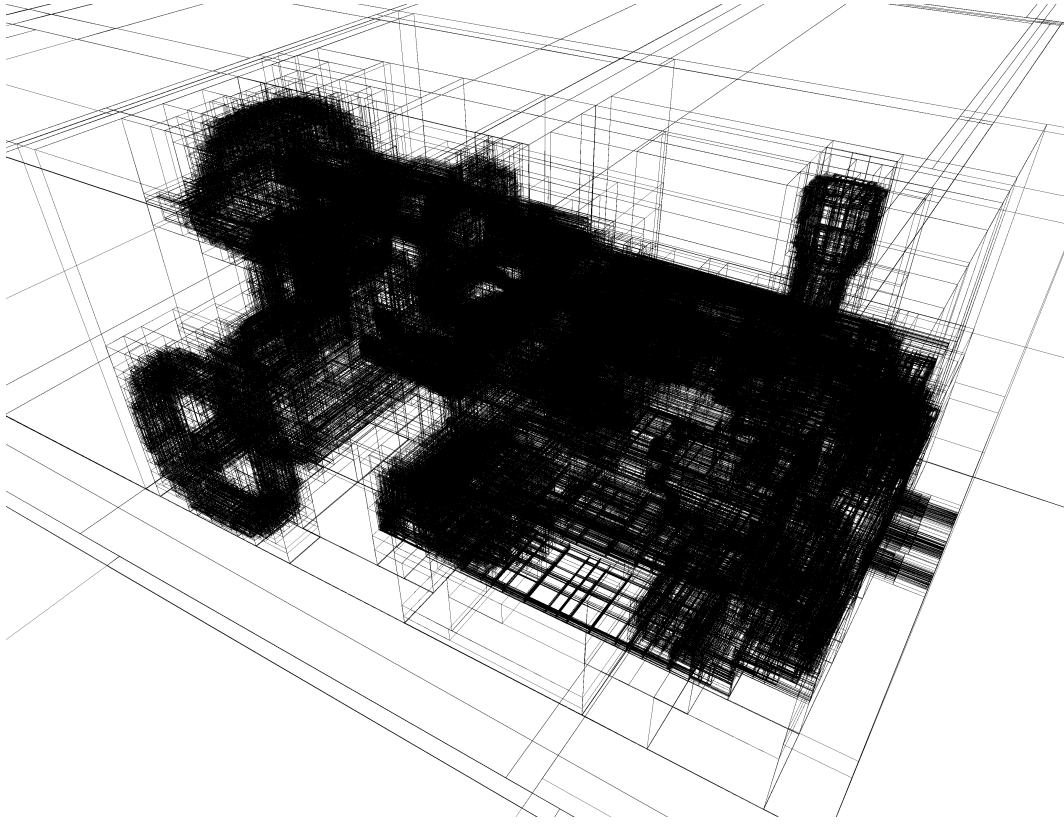
$$p_h = s + \frac{(p_1 - s) \cdot n}{v \cdot n} v$$

barysentriset koordinaatit $(\alpha_1, \alpha_2, \alpha_3)$ ja hylätään piste mikäli $\sum \alpha_i > 1$ tai $\alpha_i < 0$ jollekin i .

5.3.2 Kd-puu

Kolmioiden määrän kasvaessa edellinen menetelmä jää nopeasti käyttökelttomaksi, sillä osumat kaukanakin oleviin kolmioihin tarkistetaan samoilla raskailla laskuilla. Tarvitaan tietorakenne, jolla saadaan nopeasti pudotettua pois mahdollisimman suuri osa säteen kanssa leikkaamattomista kolmioista.

Eräs tehokkaimmista tällaisista kiihdytysrakenteista on kd-puu. Kyseessä on binääripuu, jonka solmut ovat rekursiivisesti akselien suuntaisilla tasoilla kahtia jaettuja tilavuusalkioita. Jako tapahtuu työn määrää minimoivalla heuristiikalla, ja lopullisessa puussa maailman kolmiot on jaettu lehtisolmujen kesken (kuva 5.8).



Kuva 5.8: Höyrypanssarivaunulle muodostetun kd-puun solmut visualisoituna.

Kd-puun läpikäynti on nopeaa, sillä tilavuusalkiot, joiden kautta säde ei kulje, voidaan kaikkine lapsisolmuineen sulkea heti pois. Jäljelle jää pieni joukko lehtisolmuja, joiden kolmiot voidaan käydä läpi likimääräisessä etäisyysjärjestyksessä.

Pinta-alaheuristiikka ja rakennus

Kd-puun useimpia muita puutyyppejä parempi tehokkuus perustuu pitkälti säteenseurantaan hyvin soveltuvaan pinta-alaheuristiikkaan (*Surface Area Heuristic, SAH*).

Lasketaan ensin arvio jakamattoman solmun käsittelyn hinnalle kaavalla

$$W(s) = n_{\Delta} w_{\Delta},$$

jossa $W(s)$ on solmun tarkistuksen hinta, n_{Δ} on kolmioiden lukumäärä solmussa ja w_{Δ} on yhden kolmion tarkistuksen hinta.

Tämän jälkeen etsitään koordinaattiakselien suuntaisilla tasoilla kolmioiden kärkipisteistä jakamalla se taso, joka minimoi lausekkeen

$$W(s) = w_a + P(B_1|B_1 \cup B_2)(n_{\Delta_1} w_{\Delta}) + P(B_2|B_1 \cup B_2)(n_{\Delta_2} w_{\Delta}),$$

jossa $W(s)$ on jaon yhteishinta, w_a on kd-puussa yhden laskeutumisaskeleen hinta, B_1 ja B_2 ovat tasolla jakamisen seurauksena muodostuneet suorakulmaiset särmiöt ja $B_1 \cup B_2$ on alkuperäinen suorakulmainen särmiö. Lauseke $P(B_i|B_1 \cup B_2)$ on satunnaisen säteen ehdollinen todennäköisyys leikata laatikkoa B_i , mikäli säde leikkaa myös laatikkoa $B_1 \cup B_2$. Termi n_{Δ_i} on laatikon B_i leikkaamien kolmioiden lukumäärä.

Koska joukot B_1 , B_2 ja $B_1 \cup B_2$ ovat suorakulmaisina särmiöinä konvekseja, voidaan edelliset todennäköisyydet laskea pinta-alojen suhteina (Santaló 1976):

$$P(B_i|B_1 \cup B_2) = \frac{A(B_i)}{A(B_1 \cup B_2)}.$$

Mikäli jollakin tasoista jakaminen pienentää hinta-arviota jakamattomaan nähden, jaetaan solmu kyseisen tason kohdalta kahteen lapsisolmuun. Pelkästään laatikkoa B_i leikkaavat kolmiot siirretään solmuun B_i ja molempia laatikoita leikkaavat kolmiot lisätään molempiin lapsisolmuihin.

Erinomainen kuvaus kd-puun tehokkaasta rakentamisesta löytyy lähteestä Wald and Havran (2006). Pharr and Humphreys (2004) suosittelee kokemukseen perustuen arvoja $w_a = 1$ ja $w_{\Delta} = 80$, jotka toimivat erinomaisesti myös myöhempanä esitellyssä kokeellisessa toteutuksessa.

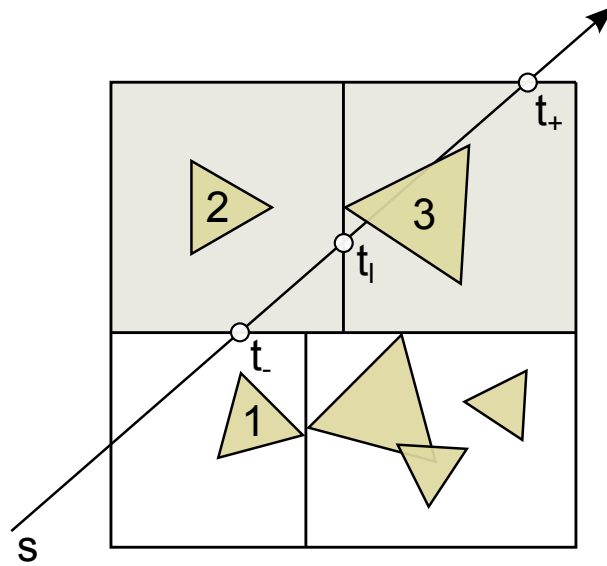
Kd-puun läpikäynti

Kd-puun läpikäynti on periaatteessa yksinkertaista. Tarkistetaan aluksi, että säde $s+vt$ leikkaa koko maailman ympärille asetettua laatikkoa. Mikäli näin on, lasketaan osumapisteiden t-arvot t_- ja t_+ , $t_- < t_+$.

Koska mahdolliset osumat alkavat kuitenkin vasta säteen alkupisteen s positiiviselta puolelta, asetetaan $t_- = \max(t_-, \epsilon)$, jossa ϵ on liukulukuepäätarkkuuden vuoksi jonkin verran nolaa suurempi vakio.

Rekursiovaiheessa lasketaan ensin säteen sijainti sen edettyä nykyisen solmun reunalle: $p = s + vt_-$. Mikäli säteen leikkauspiste solmua jakavan tason kanssa, $s + vt_l$, on solmun sisällä, eli $t_- \leq t_l \leq t_+$ (kuva 5.9), asetetaan vastakkainen lapsisolmu pinon odottamaan käsittelyä t-rajoilla (t_l, t_+) ja jatketaan lähemmästä lapsisolmusta t-rajoilla (t_-, t_l) . Muussa tapauksessa jatketaan lähemmästä lapsisolmusta t-rajoilla (t_-, t_l) , tai mikäli lapsisolmua ei ole, otetaan pinosta seuraava solmu käsittelyyn.

Solmujen kolmiot käsitellään solmujen löytämisjärjestyksessä. Läpikäynti voidaan lopettaa, kun kd-puun t_- on suurempi kuin lähimmän löydetyn leikkauspisteen etäisyys t , sillä tällöin lähin leikkauspiste on jo löydetty.



Kuva 5.9: Kd-puun läpikäyntiä. Mikäli maalatussa solmussa t_l on arvojen t_- ja t_+ välissä, edetään myös oikeanpuoleiseen solmuun. Tarkistettavat kolmiot on numeroitu.

5.4 Klassinen säteenseuranta

Klassisessa säteenseurannassa tarkoitus ei ole varsinaisesti selvittää plenoptista funktiota vaan laskea ideaalin neulansilmäkameran kuva. Kaikkien heijastusten ja taittumisten ajatellaan olevan ideaaleja, valonlähteiden olevan pistemäisiä ja varjojen teräväreunaisia (kuva 5.10).

Laskentaa ei useinkaan tehdä spektreillä vaan kolmella värikanavalla. Maailma ei edes pyri olemaan realistinen. Laskut tehdään suoraan välin $[0, 1]$ luvuilla eikä suinkaan watteina tai lumeneina. Näin ollen kuvan esittämisessä ei käytetä sävykartoitusta, joten järkevän kuvan muodostamiseksi pistemäisen valonlähteen intensiteetin $1/r^2$ -riippuvuus on tapana jättää pois tai muuttaa esimerkiksi $1/r$ -riippuvuudeksi, kuten ehdotetaan lähteessä Whitted (1980).

Klassinen säteenseuranta ratkaisee valaistusyhtälön ainoastaan singulaaristen δ -jakaumien tapauksessa. Valaistusyhtälön täyteen ratkaisemiseen tarvitaan Monte Carlo -integrointia, jolloin saadaan stokastinen säteenseuranta.



Kuva 5.10: Tyypillinen klassisen säteenseurannan esittelykuva, palloja ja peilipintoja.

5.5 Stokastinen säteenseuranta

Stokastisessa säteenseurannassa (*stochastic ray tracing, distribution ray tracing*) valaistussyhtälön ratkaisu

$$L_i = g + Ag + A^2g + A^3g + \dots$$

kirjoitetaan muotoon $L_i = g + AL_i$, joka voidaan kirjoittaa rekursioyhtälönä

$$L_n(p, \omega) = L_e(p_\omega, -\omega) + \int_{S^2} f(p_\omega, -\omega, \omega_i) L_{n+1}(p_\omega, \omega_i) |N_{p_\omega} \cdot \omega_i| d\omega_i, \quad (5.2)$$

kun merkitään $p_\omega = \tau(p, \omega)$.

Tästä saadaan stokastinen säteenseuranta, kun integraalit korvataan Monte Carlo-integroinnilla ja rekursio katkaistaan joltakin syvyydeltä m :

$$L_m = 0,$$

$$L_n(p, \omega) = L_e(p_\omega, -\omega) + \frac{1}{N_n} \sum_{k=1}^{N_n} \frac{f(p_\omega, -\omega, X_k) L_{n+1}(p_\omega, X_k) |N_{p_\omega} \cdot X_k|}{p(X_k)},$$

jossa luvut N_n ovat haarautumisten lukumäärät eri syvyyksillä, p on tärkeysotannan tiheysfunktio ja yksikköpallon pinnan satunnaisvektorit X_i arvotaan jokaiselle funktion evaluointikerralle erikseen.

Näin lasketun arvion L_0 odotusarvo on

$$E(L_0(p, \omega)) = \left[\sum_{k=0}^{m-1} A^k g \right] (p, \omega),$$

joten näin katkaisemalla ei saada keskimäärin oikeaa tulosta. Sen sijaan valitsemalla riittävän suuri heijastussyvyys m saadaan virhe mielivaltaisen pieneksi, sillä

$$\lim_{m \rightarrow \infty} \sum_{k=m}^{\infty} A_k g = 0.$$

Jälleen kerran yhtälöt on muotoiltu ilman aallonpituuksia. Edellinen yhtälö pätee jokaiselle aallonpituudelle erikseen, ja eri aallonpituuksien radianssit voidaan laskea joko approksimoimalla ratkaisua jokaiselle aallonpituudelle erikseen tai laskien komponenteittain spektrivektoreilla.

5.6 Polunseuranta

Edellä esitetyn rekursiivisen säteenseuranta-algoritmin heikkoutena on laskennan määrän eksponentiaalinen kasvu rekursion myötä. Heikkoudeksi tämän tekee se, että laskennasta suurin osa tapahtuu syvällä rekursiossa, jossa poikkeamien merkitys lopputulokseen on vähäisin. Menetelmää voidaan parantaa, jos löydetään tapa kohdistaa laskutoimituksia merkittävämpiin termeihin, matalampiin rekursion asteisiin.

Luonnollinen parannusaskel menetelmään onkin *polunseuranta* (*path tracing*). Polunseurannassa lasketaan eksponentiaalisesti haarautuvien säteiden sijaan suuri määrä yksittäisten valonsäteiden reittejä.

5.6.1 Venäläinen ruletti

Oleellinen parannus aiempaan on odotusarvon virheellisyyden korjaaminen ns. *venäläisellä ruletilla* (*Russian roulette*). Tällöin luovutaan kokonaan maksimisyvyyden käsitteestä ja tehdään rekursion pysähtymisestä stokastista. Jos jatkamistodennäköisyydeksi asetetaan q ja alkuperäisen lausekkeen odotusarvo on $E(Y)$, on venäläisellä ruletilla katkaistun lausekkeen odotusarvo

$$E(q \cdot Y + (1 - q) \cdot 0) = qE(Y).$$

Tämä voidaan korjata jakamalla lauseke Y arvolla q , sillä tällöin saadaan odotusarvoksi

$$E(q \cdot Y/q) = E(Y).$$

Arvon q valinnalla voidaan siirtää laskennan painopistettä. Mikäli jatkamistodennäköisyys q on kovin suuri, iteraatio suppenee hitaammin, sillä korkeiden kertalukujen k termejä

$$\frac{A^k g(p_k, \omega_k)}{q^k}$$

lasketaan usein, vaikka niiden vaikutus on pieni.

Toisaalta, jos jatkamistodennäköisyys on kovin pieni, saadaan aikaan häiritsevää kohinaa, sillä silloin harvoin, kun korkeiden kertalukujen k termejä lasketaan, ovat termit

$$\frac{A^k g(p_k, \omega_k)}{q^k}$$

valtavan suuria ja voivat näkyä kuvassa kirkkaina pisteinä. Kumpaakin ongelmaa voi kompensoida myös säätämällä laskettavien näytteiden lukumäärää.

5.6.2 Menetelmä

Polunseurannassa rekursio (5.2) kirjoitetaan näin ollen muotoon

$$L_n(p, \omega) = L_e(p_\omega, -\omega) + [Q] \frac{1}{q} \frac{f(p_\omega, -\omega, X) L_{n+1}(p_\omega, X) |N_{p_\omega} \cdot X|}{p(X)},$$

jossa $[Q]$ saa todennäköisyydellä q arvon 1 ja todennäköisyydellä $1 - q$ arvon 0, p on tärkeysotannan tiheysfunktio ja X on yksikköpallon pinnan satunnaisvektori, joka arvotaan jokaiselle funktion evaluointikerralle uudestaan.

Näytteiden $L_0(p, \omega)$ lukumäärää kasvattamalla voidaan parantaa laskennan tarkkuutta. Tällöin lopullinen arvio on

$$L(p, \omega) = \frac{\sum_{k=1}^N L_0(p, \omega)_k}{N},$$

jossa termit $L_0(p, \omega)_k$ ovat edellä esitellyllä kaavalla laskettuja näytteitä. Koska keskiarvo suppenee odotusarvoon ja yksittäisen termin odotusarvo on oikea, on myös keskiarvo harhaton.

5.6.3 Muotoilu polkuina

Palataan edellä esiteltyyn rekursiokaavaan 5.2,

$$L_n(p, \omega) = L_e(p_\omega, -\omega) + \int_{S^2} f(p_\omega, -\omega, \omega_i) L_{n+1}(p_\omega, \omega_i) |N_{p_\omega} \cdot \omega_i| d\omega_i.$$

Tässä kaavassa tekijä $f(p_\omega, -\omega, \omega_i) |N_{p_\omega} \cdot \omega_i|$ kertoo, kuinka suuri osuus pisteeseen p_ω suunnasta ω_i saapuvasta differentiaalisesta radianssista siroaa pisteen p suuntaan.

Yritetään kirjoittaa yhtälö intuitiivisempaan muotoon määrittelemällä tätä osuutta kuvaava funktio:

$$G(r \rightarrow q \rightarrow p) = f(q, \omega_{q \rightarrow p}, \omega_{q \rightarrow r}) |N_q \cdot \omega_{q \rightarrow r}|.$$

Tässä $\omega_{p \rightarrow q}$ tarkoittaa pisteestä p pisteeseen q osoittavaa yksikkövektoria. Näillä merkinnöillä rekursiokaava saa muodon

$$L_n(p, \omega) = L_e(p_\omega, -\omega) + \int_{S^2} G(p_{\omega, \omega_i} \rightarrow p_\omega \rightarrow p) L_{n+1}(p_\omega, \omega_i) d\omega_i.$$

Määritellään seuraavaksi

$$G(p_{n+2} \rightarrow p_{n+1} \rightarrow \cdots \rightarrow p_1) = G(p_{n+2} \rightarrow p_{n+1} \rightarrow p_n) \cdot G(p_{n+1} \rightarrow p_n \rightarrow \cdots \rightarrow p_1).$$

Määritelmän motivaation voi ymmärtää seuraavalla ajatusmallilla: Jos piste p_{n+1} sirottaa pisteestä p_{n+2} saapuvasta valosta pisteeseen p_n osuuden α ja piste p_n sirottaa pisteestä p_{n+1} saapuvasta valosta pisteeseen p_1 osuuden β , on osuus pisteestä p_{n+2} pisteeseen p_1 siroavasta valosta $\alpha\beta$.

Edellisellä määritelmällä rekursio L_0 voidaan kirjoittaa auki muodossa

$$L_0(p, \omega) = \sum_{k=0}^{\infty} \int_{(S^2)^k} G(p_{\omega \dots \omega_k} \rightarrow \cdots \rightarrow p_\omega \rightarrow p) L_e(p_{\omega \dots \omega_k}, -\omega_k) d\omega_k \cdots d\omega_1,$$

jos ymmärretään tapauksen $k = 0$ tarkoittavan lauseketta $L_e(p_\omega, -\omega)$.

Tästä juontaa polunseurannan perusajatus: integroidaan kaikkien mahdollisten pisteeseen p suunnasta ω saapuvien valonsäteiden reittien yli. Tämä voidaan kirjoittaa myös kompaktimmassa muodossa,

$$L_0(p, \omega) = \int_{R(p, \omega)} L_e(p_r, \omega_r) d\mu(r).$$

Tässä $R(p, \omega)$ on pisteeseen p suunnasta ω päättyvien valonsäteiden joukko, r on valonsäde, p_r on valonsäteiden alkupiste, ω_r on valonsäteiden alkusuunta ja $d\mu(r)$ on valonsäteiden energiasta pisteeseen p saapuva osuus.

Luku 6

Toteutus

Havainnollistetaan mallia toteuttamalla ratkaisija tärkeimmillä edellä mainituista ominaisuuksista. Ajatellaan aluksi kameran olevan ideaalinen aineeton neulansilmä-kamera, jossa kuva muodostuu väärinpäin kameran sisään ja valonsäteitä saapuu kuvapinnalle ainoastaan pistemäisen aukon läpi. Muokataan konstruktiota siirtämällä kuvapinta symmetrisesti aukon toiselle puolelle, jolloin kuva muodostuu oikeinpäin.

Muotoillaan tämän jälkeen pseudokoodina naiivi säteenseuraja:

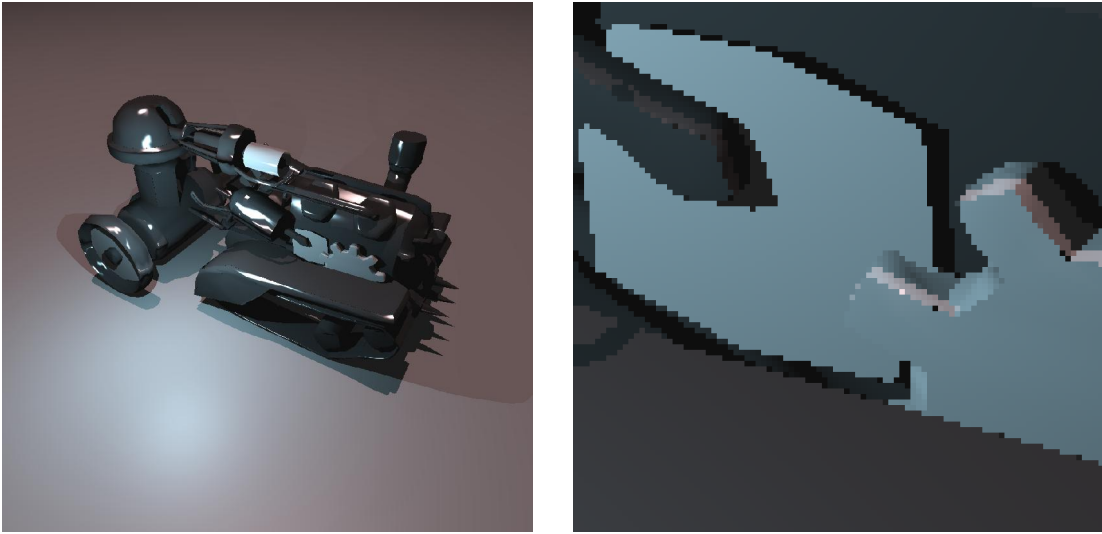
Jokaiselle kuvapisteele:

```
säde = säde kameran aukosta kuvapisteen sijaintiin avaruudessa
kuvapisteen suunnasta siroava valo = seuraa(säde)
```

Funktio seuraa(säde):

```
alusta tulosuuntaan siroava valo
jokaiselle pistemäiselle valolle:
    jos valo näkyvissä,
        lisää valon vaikutus tulosuuntaan siroavaan valoon
jos syvyys < maksimisyvyys:
    säde ulos = säteen ideaali heijastus
    saapuva valo = seuraa(säde ulos)
    lisää saapuvan valon vaikutus takaisin siroavaan valoon
palauta tulosuuntaan siroava valo
```

Osumapiste selvitetään etsimällä lähin säteen $s + vt$, $t > 0$ ja maailman kolmioiden leikkauspisteistä. Numeerisesta epätarkkuudesta johtuva väärä törmäys lähtöpisteseen eliminoidaan vaatimalla $t > \epsilon$ jollekin ennalta määrätyle turvarajalle.



Kuva 6.1: Yksinkertaisella säteenseurajalla renderöity kuva ennen teksturointia ja antialiasointia.

6.1 Kd-puu

Tehokas tietorakenne on helposti säteenseurajan tärkein yksittäinen komponentti. Naiivi kaikkien kolmioiden läpikäynti säteen leikkauspisteitä laskiessa on hitautensa vuoksi lähes käyttökelvoton ja suoriutui testikokoonpanolla ¹ 16000-kolmioisen höyrypanssarivaunun laskennasta noin 23 minuutissa.

Kd-puun rakennus höyrypanssarivaunulle valmistuu noin kahdessa sekunnissa. Tämän jälkeinen renderöinti kestää noin kolme sekuntia, josta suurin kuluu 72 miljoonan kd-puun solmun läpikäymiseen ja 14 miljoonaan säde-kolmio-tarkistukseen.

6.2 Teksturointi

Annetaan kappaleille tarkemmat värit liimaamalla kappaleen pintaan tekstuurikartta, joka kertoo kullekin spektrin aallonpituudelle osuuden, joka siroaa materiaalista. Annetaan lisäksi säteen osumalle kolmioon lisäehto, jonka mukaan osuma kolmioon hyväksytään vain, jos pisteen tekstuurikartan läpinäkyvyysarvo on annettua rajaa korkeampi. Ehdon myötä panssarivaunun kyljen sininen laatta muuttuu rattaaksi (kuva 6.2).

¹Intel Core 2 Duo P8700, renderöintiresoluutio 960×960



Kuva 6.2: Säteenseuraja tekstuurikartalla. Laskenta-aika n. 5,4 sekuntia.



Kuva 6.3: Säteenseuraja lisättyä heijastuksilla, heijastussyvyys 8. Laskenta-aika noin 11 sekuntia.

6.3 Heijastukset

Koska kd-puu mahdollistaa nopean renderöinnin, voidaan algoritmin heijastusten maksimisyvyttä nostaa (kuva 6.3). Kuvassa on kuitenkin edelleen rumia artefakteja, joista voidaan huolehtia myöhemmin.

6.4 Antialiasointi

Edellisen kaltaiset artifaktit johtuvat siitä, että laskenta suoritetaan liian harvoissa pisteissä. Asetetaan ohjelma laskemaan jokaisen kuvapisteen tulos useampaan kertaan satunnaisesti eri kohdista kuvapistettä. Pahimmat virheet ja sahalaidat katoavat (kuva 6.4). Parannus ei kuitenkaan tule ilmaiseksi, sillä 2,8 miljardin kd-puun solmun ja 700 miljoonan säde-kolmio-leikkauksen tarkistuksessa kestää jo yli kaksi minuuttia.



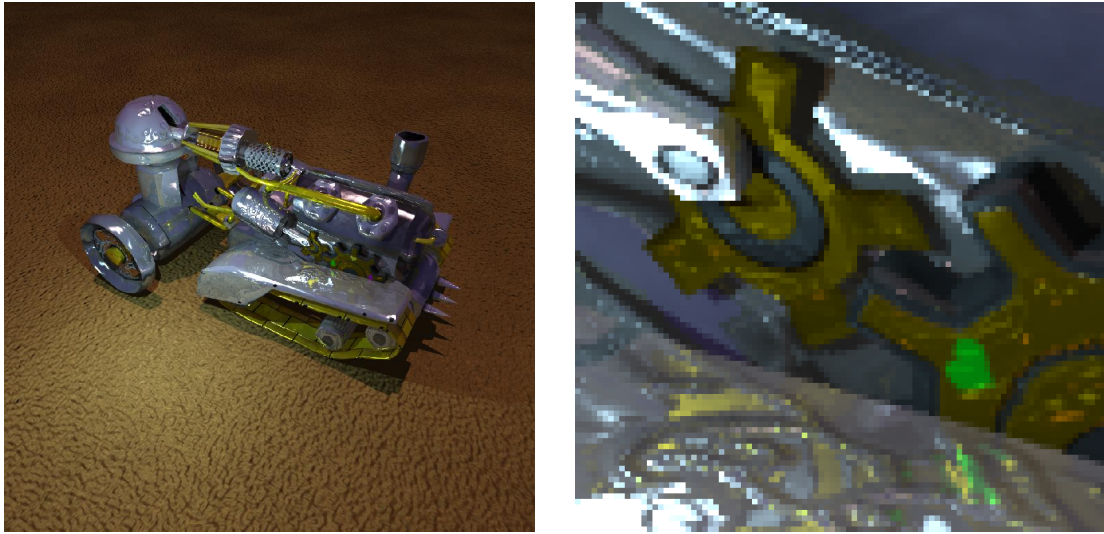
Kuva 6.4: Vasemmalla säteenseuraja ilman antialiasointia. Oikealla antialiasoinnilla, 16 näytettä kuvapistettä kohti. Laskenta-aika noin 2 minuuttia ja 13 sekuntia.

6.5 Normaalikartta

Aiemmin esiteltyllä normaalikartalla malliin saa huomattavasti lisää yksityiskohtaisuutta lisäämättä ainoatakaan kolmiota (kuva 6.5).

6.6 Emissiokartta

Emissiokartalla voidaan kappaleiden emissiospektrit esittää kätevästi. Tässä on käytetty diffuusia emissiota, joka on kaikkiin suuntiin yhtä voimakasta (kuva 6.6).



Kuva 6.5: Säteenseuraja normaalikartalla. Laskenta-aika noin 2 minuuttia ja 28 sekuntia.



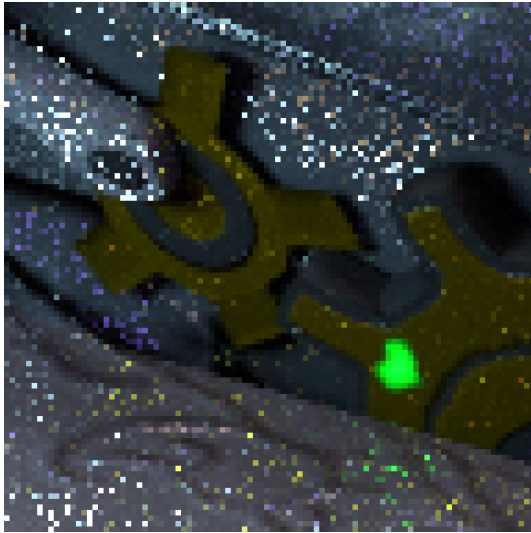
Kuva 6.6: Säteenseuraja emissiokartalla. Laskenta-aika 2 minuuttia ja 31 sekuntia.

6.7 Sirontajakauma

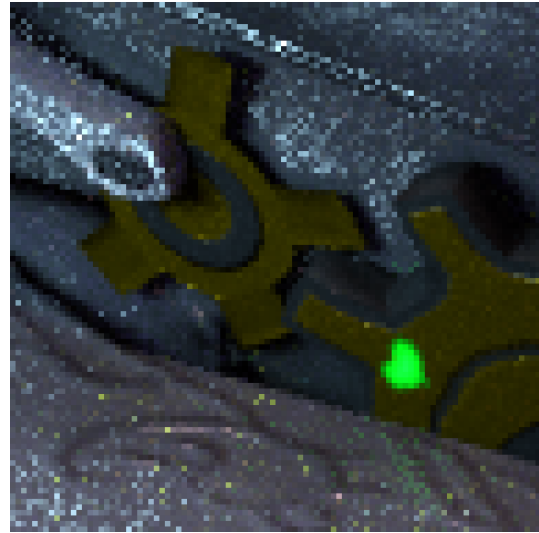
Kun tärkeimmät perusominaisuudet on ohjelmoitu, voidaan ideaalit heijastukset vaihtaa satunnaisiksi painottaen tuloksia sirontajakaumalla. Samalla otetaan käyttöön venäläinen ruletti ja levitetään valonlähteet suuremmiksi (kuva 6.7(a)).

Kun säteet valitaan tärkeysotannalla painottaen tärkeimpiä suuntia ja valonlähteitä, saadaan huomattavasti parempia tuloksia (kuva 6.7(b)). Modifioidulle Phong-

jakaumalle erinomaiseksi painotusjakaumaksi osoittautui diffuusin valaistuksen \cos -painotuksen, spekulaarin valaistuksen \cos^n -painotuksen ja tasajakauman yhdistelmä.



(a) Tasajakauma painottaa liikaa vähämerkityksisiä suuntia ja vaatii runsaasti näytteitä. Laskenta-aika noin 2 minuuttia ja 43 sekuntia.



(b) Sopivasti painotettu tärkeysotanta on huomattavasti tehokkaampi. Laskenta-aika noin 2 minuuttia ja 43 sekuntia.

Kuva 6.7: Polunseurantaa sirontajakaumilla, 16 näytettä kuvapistettä kohti. Tehokkaampi menetelmä pienentää näytteistykseen varianssia eli vähentää kohinaa.

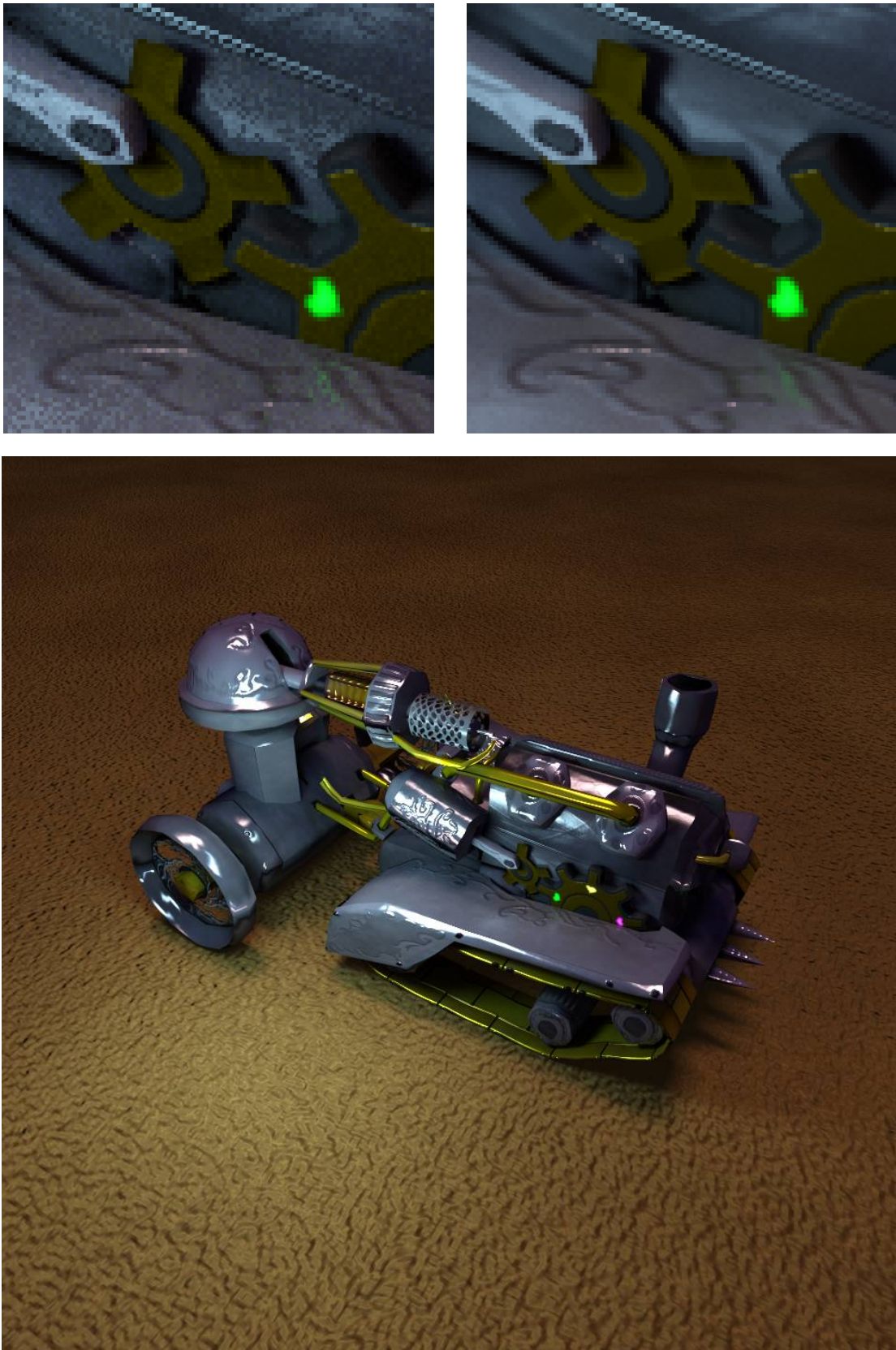
6.8 Adaptio

Toimivaksi menetelmäksi kohinan pienentämiseen osoittautui adaptio, jossa alun näytteistykseen jälkeen näytteistetään lisää niitä kuvapisteitä, joille suhteellinen otos-hajonta $\sigma[S]/E[S]$ on suurin. Toteutus tapahtui prioriteettijonoa käyttäen laskemalla kullekin kuvapisteelle näytteiden summat ja neliösummat, sillä tällöin

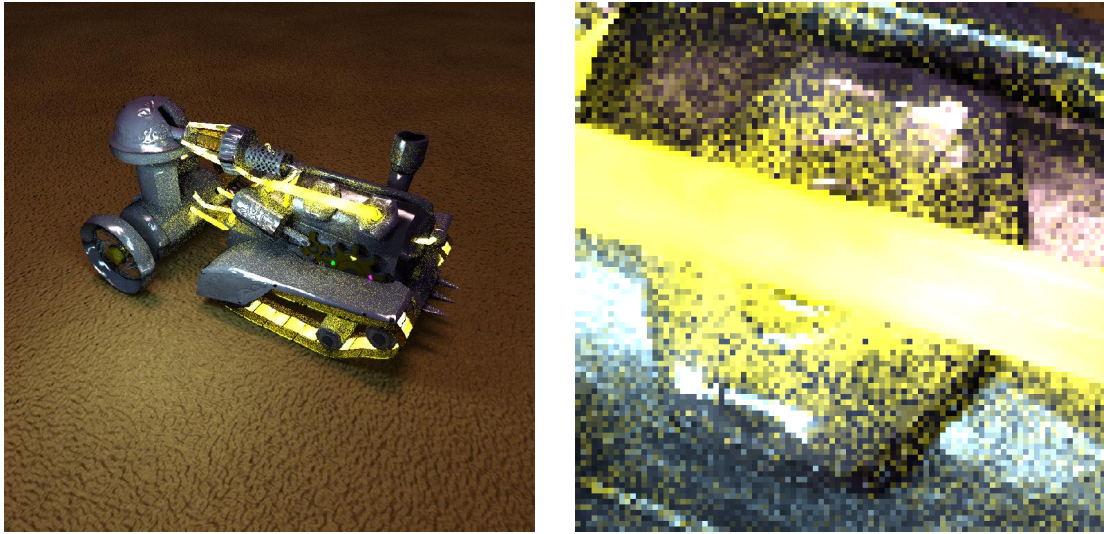
$$\frac{\sigma[S]}{E[S]} = \sqrt{\frac{N^2}{N-1} \frac{\sum(s_i^2)}{(\sum s_i)^2} - \frac{N}{N-1}} \approx \sqrt{N \frac{\sum(s_i^2)}{(\sum s_i)^2} - 1}.$$

Prioriteettifunktioksi voidaan näin ollen valita $N \cdot \sum(s_i^2)/(\sum s_i)^2$.

Pahin kohina poistuu muutamassa minuutissa ja adaptiota voi jatkaa pidempäänkin (kuva 6.8).



Kuva 6.8: Adaptiolla parannetut renderöinnit. Laskenta-ajat 10 minuuttia (vasemmalla) ja kolme tuntia (oikealla ja alla).



Kuva 6.9: Polunseuraaaja 16 näytteellä ja voimakkaasti emittoivia pintoja. Laskenta-aika 2 minuuttia ja 45 sekuntia.

6.9 Heikkouksia

Polunseuranta ratkaisee esitetyn valaistusmallin täydellisesti, mutta menetelmässä on silti runsaasti parantamisen varaa. Vaikka menetelmä on yleensä suhteellisen nopea, joissakin tilanteissa esiintyy kohinaa, josta on vaikea päästä eroon.

Esimerkki hankalasta tilanteesta on runsaasti valoa emittoivat pinnat, joita on hankalampi huomioida tärkeysotannan painojakaumassa. Seurauksena on runsasta kohinaa, joka vaatii erittäin suuren määrän näytteitä varianssin pienentämiseksi siedettävälle tasolle (kuvat 6.9 – 6.10).

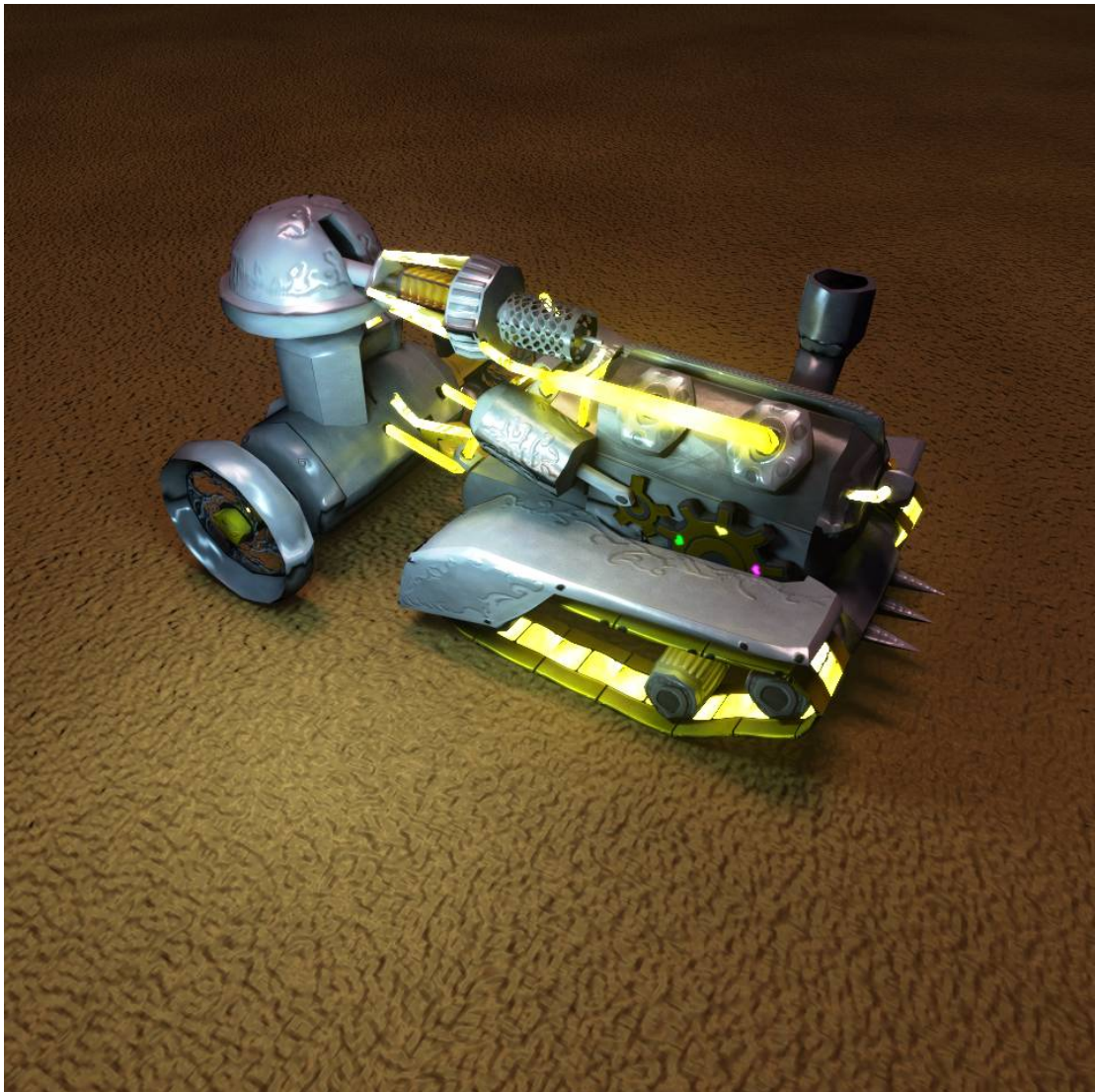
6.10 HDR-kuvaus ja bloom

Koska näyttöpäätteen kirkkausdynamiikka on rajattu, ei kirkkaimpia väriarvoja voi näyttää sellaisenaan. Väriarvot kuvataan näytön käyttämälle dynamiikka-alueelle sävytyskuvauksella (*tone map*). Edellä lasketuissa kuvissa on sävytyskuvauksena käytetty luonnollista logaritmia.

Koska dynamiikkaa joudutaan rajoittamaan, parhaallakaan sävytyskuvauksella kuva ei voi näyttää silmälle täysin samanlaiselta kuin sen olisi tarkoitus. Osittain tätä puutetta voi korvata jälkikäsitteilyllä. Eräs yleisimmistä jälkikäsitteilyefekteistä on *bloom*, joka levittää kirkkaita alueita ja luo illuusion kirkkaudesta. Tyypillinen bloom-toteutus lisää kuvan väriarvoihin gaussisesti pehmennetyn kopion kuvasta ennen sävytyskuvausta (kuva 6.11).



Kuva 6.10: Sama maailma 1024 näytteellä ja adaptiolla. Laskenta-aika neljä tuntia, jonka aikana käsiteltiin 273 miljardia kd-puun solmua ja 67 miljardia kolmiotarkistusta.



Kuva 6.11: Jälkikäsittelyllä lisätty bloom-efekti luo illuusion kirkkaudesta.

Luku 7

Yhteenveto

Edellä on johdettu todellisuuteen perustuva radiometrinen malli maailman valaistuksen laskemiseksi. Malli on kompromissi suorituskyvyn ja realismin väliltä. Jotkin fysikaaliset ilmiöt jäävät selittämättä, mutta toisaalta erinomaisia tuloksia saadaan jo minuuteissa. Malli on kolmiulotteisessa grafiikassa de facto -standardimalli, jota pyritään approksimoimaan erilaisin keinoin myös realismiin pyrkivässä reaaliaikagrafiikassa.

Ratkaisun määrittävän integraaliyhtälön analyttinen ratkaisu vahvistaa sen intuitiivisen tosiasian, että valaistus määräytyy suoraan valonlähteistä ja valon kaikkien kertalukujen heijastuksista. Tämä mahdollistaa erilaisia menetelmiä valaistuksen laskemiseen, joista on kuvattu tarkimmin stokastinen säteenseuranta ja polunseuranta.

Valaistuksen määrittävän yhtälön ratkaisu on ohjelmoitu lähtien yksinkertaisesta säteenseurajasta kohti kehittyneitä polunseurajaa. Ratkaisuun tarvittavat menetelmät on esitelty sekä teoreettisesti että esimerkein. Tärkeimmät optimointimenetelmät, kuten kd-puu ja monitärkeysotanta, on myös kuvattu yksityiskohtaisesti.

Vaikka malli on kattava, se ei kuitenkaan ole täydellinen. Suurimmat mallin puutteet ovat luultavasti pintasironajakaumien (BSSRDF) ja väliaineen mallinnuksen puuttuminen. Näistä kiinnostuneen kannattanee turvata lähteeseen Pharr and Humphreys (2004).

Myös ratkaisussa olisi paljon kehittämisen varaa. Suurin puute liittyynee monitärkeysotantaan, jossa tulisi huomioida tarkemmin voimakkaasti emittoivat pinnat. Toinen tärkeä parannus olisi lähteessä Veach (1998) esitelty Metropolis Light Transport, jonka tulisi merkittävästi nopeuttaa menetelmän suppenemista.

Kirjallisuutta

Christine A. Curcio, Kenneth R. Sloan, Robert E. Kalina, and Anita E. Hendrickson. Human photoreceptor topography. *The Journal of Comparative Neurology*, 292(4):497–523, 1990. ISSN 1096-9861. doi: 10.1002/cne.902920402. URL <http://dx.doi.org/10.1002/cne.902920402>.

Masahiko Daimon and Akira Masumura. Measurement of the refractive index of distilled water from the near-infrared region to the ultraviolet region. *Appl. Opt.*, 46(18):3811–3820, Jun 2007. doi: 10.1364/AO.46.003811. URL <http://ao.osa.org/abstract.cfm?URI=ao-46-18-3811>.

Matt Pharr and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004. ISBN 012553180X.

Luis A. Santaló. *Integral geometry and geometric probability /*. Addison-Wesley, Reading, MA, 1976. ISBN 0-201-13500-0.

B. B. Tatian. Fitting refractive-index data with the Sellmeier dispersion formula. *Applied Optics*, 23(24):4477–4477, 1984. URL <http://search.proquest.com/docview/734244800?accountid=11365>.

Eric Veach. *Robust monte carlo methods for light transport simulation*. PhD thesis, Stanford, CA, USA, 1998. AAI9837162.

Ingo Wald and Vlastimil Havran. On building fast kd-trees for ray tracing, and on doing that in $O(N \log N)$. In *Proceedings of the 2006 IEEE Symposium on Interactive Ray Tracing*, pages 61–69, 2006.

Turner Whitted. An improved illumination model for shaded display. *Commun. ACM*, 23:343–349, June 1980. ISSN 0001-0782. doi: <http://doi.acm.org/10.1145/358876.358882>. URL <http://doi.acm.org/10.1145/358876.358882>.