



# Palloharmoniset funktiot ja esilasketut siirtofunktiot valaistuslaskennassa

Timo Kellomäki  
Tampereen teknillinen yliopisto  
Ohjelmistotekniikan laitos

timo.kellomaki@tut.fi

## Tiivistelmä

Kolmiulotteisella, reaaliaikaisella tietokonegrafiikalla on nykyään keskeinen asema muun muassa simulaattoreissa ja peleissä sekä esineiden ja rakennusten suunnittelussa. Yksi alan suurimpia ongelmia on globaali valaistus, jossa pitäisi ottaa huomioon, miten kunkin pisteen koko ympäristö vaikuttaa sen valaistukseen. Suuri osa ympäristön vaikutuksesta johtuu kappaleesta itsestään, kuten esimerkiksi itsevarjostuksesta. Valaistuksen ja mallin muuttuessa ongelma on erityisen hankala. Eräs ratkaisu perustuu palloharmonisiin funktioihin ja esilaskettuun siirtofunktioon. Tämä artikkeli johdattelee aihepiiriin ja esittää tekniikan perusidean.

## 1 Johdanto

Kolmiulotteista tietokonegrafiikkaa käytetään nykyään hyvin monenlaisiin tarkoituksiin. Mallinnusohjelmilla suunnitellaan kaikkea pienistä käyttöesineistä rakennuksiin ja aina kokonaisiin kaupunkiympäristöihin asti. Simulaattoreilla koulutetaan esimerkiksi lentokoneiden, laivojen, metsäkoneiden ja erilaisten sotilasajoneuvojen kuljettajia. Jatkossa erilaisia virtuaaliympäristöjä voidaan käyttää yhä laajemmin vaikkapa kirurgien kouluttamiseen, sisustuselementteinä tai motivoimaan lonkkaleikkauksesta kuntoutuvan retkiä kävelymatolla. Jo nyt taloudellisesti merkittäviä ovat erilaiset viihdesovellukset, kuten pelit ja elokuvat.

Tämä artikkeli keskittyy interaktiiviseen grafiikkaan, jossa käyttäjä liikuttaa kameraa maailmassa reaaliaikaisesti. Ku-

va täytyy tällöin piirtää ohjelman ajan aikana uudesta näkökulmasta kymmeniä kertoja sekunnissa. Usein kameran lisäksi myös maailmassa sijaitsevat esineet voivat liikkua.

Useimmissa interaktiivisissa sovelluksissa on tarpeen toteuttaa kaksi ristiriitaista päämäärää: kuvan pitäisi näyttää mahdollisimman todenmukaiselta, mutta sitä pitäisi pystyä tuottamaan reaaliaikaisesti. Tietokonegrafiikan realismin kannalta nimenomaan valaistus on erittäin tärkeä osa-alue.

Realistiset valaistusmallit ovat suunnittelutyössä tärkeitä, jotta tuote näyttää suunnitteluvaiheessa mahdollisimman samalta kuin valmiina. Lentosimulaattoreissa voisi olla tärkeää mallintaa vaikkapa sokaisevia heijastuksia. Viihdesovelluksissa esteettisyys on erityisen tärkeää, ja

nykyään peleissä grafiikalta vaaditaan jommoikoista realismia.

Valaistuslaskennan tavoitteena on laskea kunkin kuvapisteen lopullinen kirkkaus ja väri. Luonnossa fotonit noudattavat fysiikan lakeja, mutta tämän tarkka mallintaminen ei grafiikassa onnistu fotonien valtavan määrän ja fysikaalisten prosessien monimutkaisuuden takia.

Kolmiulotteisen grafiikan piirto perustuu piirrettävien esineiden matemaattisiin malleihin kolmiulotteisessa avaruudessa. On kaksi päälähestymistapaa, joilla malli voidaan muuttaa kaksiulotteiseksi kuvaksi: *projektiopiirto* ja *säteenjäljitys* (ray tracing).

Säteenjäljitys pyrkii simuloimaan fotonien kulkua avaruudessa, mutta etenee kamerasta valonlähteisiin eikä päinvastoin. Tuloksena saadaan varsin realistinen kuva, mutta tekniikka on toistaiseksi liian hidas interaktiivisiin tilanteisiin. Reaaliajassa piirrettävässä grafiikassa tyydytään yleensä projektiopiirtoon, joka on kokonaan toisenlainen lähestymistapa. Projektiopiirto perustuu mallin muuttamiseen monikulmioiksi, jotka projisoidaan kuvaustasolle yksi kerrallaan.

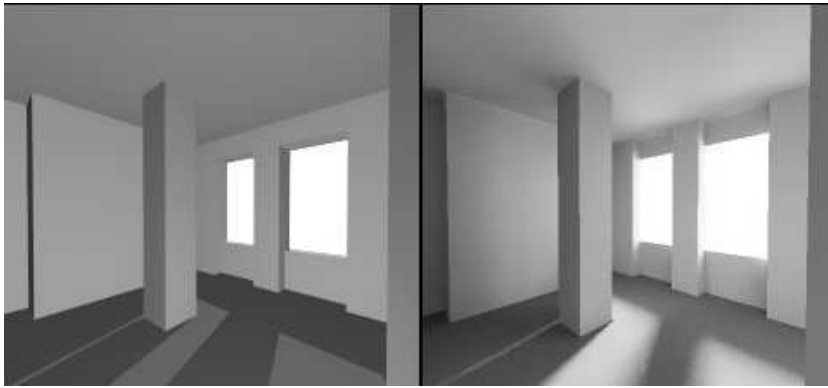
Perinteisesti reaaliaikainen tietokonegrafiikka on perustunut paikallisiin valaistusmalleihin. Valaistusta laskettaessa kutakin monikulmioiden kulmapistettä käsitellään erillisenä, aivan kuin muita esineitä ei olisi olemassa. Tämän vuoksi esimerkiksi kappaleet eivät langeta varjoja toisiinsa eivätkä myöskään varjosta itseään. Epäsuoraa valaistusta ei yleensä mallinneta lainkaan. Seurauksena syntyvä kuva ei ole realistinen. Realistisen kuvan aikaansaamiseksi valaistus täytyisi laskea globaalisti eli ottaa huomioon kussakin pisteessä koko muu malli, mutta tämä on aivan liian raskasta reaaliaikapiirtoon. Kuva 1 havainnollistaa paikallisen ja globaalisen valaistuksen eroa.

Reaaliaikaisen globaalisen valaistuksen ongelma on ollut yksi 2000-luvun tutkituimpia aiheita tietokonegrafiikassa ja siihen on kehitetty lukuisia erilaisia menetelmiä. *Palloharmonisten funktioiden* (spherical harmonics, SH) ja *esilaske-*  
*tun siirtofunktion* (precomputed radiance transfer, PRT) yhdistelmä on eräs näistä menetelmistä.

Reaaliaikaisessa piirroksessa voidaan onneksi käyttää hyödyksi monenlaista esilaskentaa. Tällöin mahdollisimman suuri osa piirtämiseen tarvittavista laskelmista tehdään etukäteen ja tulokset tallennetaan ajon aikana käytettäväksi. Näin ajon aikana tapahtuvaa laskentaa voidaan vähentää muistinkäytön kustannuksella. Mitä realistisempi tulos halutaan, sitä pidempään esilaskenta kestää ja sitä enemmän muistia se kuluttaa. Monet kehittyneet tekniikat perustuvat nokkeleihin tapoihin tallentaa raskaasti laskettavaa tietoa tiiviiseen mutta piirrettäessä helposti hyödynnettävään muotoon. SH/PRT-tekniikka on erinomainen esimerkki tällaisesta tilanteesta, sillä siinä pyritään laskemaan mahdollisimman paljon etukäteen ja tallentamaan tulokset tiiviisti palloharmonisten funktioiden avulla.

Yleensä ottaen esilaskenta toimii kohtuullisen hyvin, jos sekä valaistusolosuhteet että itse malli pysyvät vakioina, mutta jo jomman kumman muuttuminen tuottaa monesti ongelmia. Valaistus muuttuu vaikkapa vuorokaudenajan mukana, auton valojen liikkeessä tai hahmon kulkiessa taskulampun kanssa ympäri huoneistoa. Mallin kokonaisuus muuttuu, kun mikä tahansa esine liikkuu. Myös yksittäiset esineet voivat muuttaa muotoaan, mistä tyypillinen esimerkki on animoitu ihmishahmo.

PRT-menetelmä perustuu siihen, että valaistuksen laskeminen jaetaan valoista riippuvaan osaan ja kustakin kiinteästä



**Kuva 1:** Vasemmalla helposti reaaliajassa laskettava paikallinen valaistus ja erikseen lasketut terävät varjot. Oikealla esimerkki laskentaintensiivisestä globaalista valaistuksesta radiositeettitekniikalla. Kuvan lähde soveltaen Wikimedia Commons / Hugo Elias, GNU Free Documentation License.

kappaleesta itsestään riippuvaan osaan. Kappaleesta riippuva osa voidaan laskea etukäteen, jolloin ajon aikana tarvitsee laskea vain valaistuksesta riippuva osa ja yhdistää tulokset. Näin esilaskettuja tuloksia päästään hyödyntämään myös muuttuvissa valaistusolosuhteissa. Myös liikkuvista kappaleista selvittää, koska liikettä pystytään käsittelemään ajon aikana tehtävinä koordinaatistomuunnoksina.

Palloharmonisia funktioita on käytetty jo pitkään esimerkiksi fysiikassa, sähkömagnetiikassa ja kemiassa. Tietokonegrafiikassa ne yleistyivät Ramamoorthin ym. [3] ja Sloanin ym. [5] artikkeleiden ilmestyttyä 2000-luvun alussa. Ramamoorthi ym. käyttivät palloharmonisia funktioita ympäristöstä tulevan valaistuksen tiiviiseen esittämiseen. Sloan ym. esitivät menetelmän, jossa palloharmoniset funktiot yhdistettiin esilaskettuun siirtofunktioon, mikä mahdollisti muuttuvan, globaaleja piirteitä sisältävän valaistuksen laskemisen reaaliajassa. Ideaa on tämän jälkeen kehitetty moniin eri suuntiin ja sovellettu käytäntöön etenkin peleissä.

Tämä artikkeli on jäsennetty seuraavasti: luvussa 2 esitellään yleisellä tasolla projektiopiirtämisen ja valaistuslaskennan periaatteita. Luku 3 käy läpi palloharmonisten funktioiden teoriaa ja luvussa 4 tutustutaan esilasketun siirtofunktion ideaan. Luku 5 käsittelee tekniikkaan liittyviä ongelmia ja rajoituksia. Luku 6 on yhteenveto.

## 2 Valaistuslaskennan perusteet

Kolmiulotteisessa mallinnuksessa kappaleita esitetään lukuisilla erilaisilla tavoilla. Tasaiset pinnat mallinnetaan usein joukkona monikulmioita. Maailma on täynnä myös kaarevia pintoja, joita on helpompi mallintaa esimerkiksi parametrisoiduilla pinnoilla. Tällöin pinnan  $x$ -,  $y$ - ja  $z$ -koordinaatit annetaan kahden parametrin funktiona. Pinta muodostuu, kun parametrit saavat kaikki arvot jossakin arvoalueessa. Myös implisiittisiä matemaattisia pintoja voidaan käyttää. Esimerkiksi pallo voitaisiin määrittellä pisteinä, jotka toteuttavat yhtälön  $x^2 + y^2 + z^2 = r^2$ .

Projektiopiirrossa ideana on, että matemaattisesti eri tavoin esitetyt kappaleet muunnetaan monikulmioverkoiksi esimerkiksi askeltamalla parametrisen pinnan parametreja sopivin välein. Kaarevat pinnat eivät ole tämän jälkeen aidosti kaarevia, vaan pyöreät muodot ruudulla ovat vain valaistuksen luoma illuusio.

Monikulmioina esitetyssä mallissa *kulmapisteet* ovat olennaisia. Kuhunkin kulmapisteeseen liittyy yleensä useita monikulmioita. Kullekin kulmapisteelle määritellään *pistenormaalivektori*, joka osoittaa ulospäin kappaleesta. Kaareville esineille normaali voidaan saada vaikkapa analyttisesti matemaattisesta mallista tai keskiarvona kulmapisteeseen rajautuvien monikulmioiden normaaleista. Esimerkiksi parametrisen pinnan kaksi tangenttia ovat osittaisderivaatat kahden parametrin suhteen. Kussakin pisteessä pistenormaali saadaan tangenttien ristitulona.

Monikulmiota piirrettäessä sen kulmapisteet projisoidaan kameran eteen sijoitetulle kuvaustasolle. Kuvaustasosta rajataan sopiva leikkausikkuna, joka jaetaan piirrettävän alueen resoluutiota vastaavaan määrään pikseleitä ja monikulmio rasteroidaan tähän resoluutioon. Rasteroinnissa tutkitaan, mitkä pikselit monikulmio peittää, minkä yhteydessä täytyy päättää kunkin pikselin väri jonkin valaistusmallin perusteella.

Kun valaistusmallia sovelletaan, on käytössä kaksi yleistä vaihtoehtoa. Nopeampi tapa on soveltaa valaistusmallia vain kussakin kulmapisteessä, jolloin saadaan selville kulmapisteen väri. Tämän jälkeen väri interpoloidaan monikulmion sisäpisteisiin (*Gouraud-sävytyt*). Tarkempi tulos saadaan interpoloimalla kulmapisteiden normaali (ja tarvittaessa muita olennaisia tietoja) monikulmion kattamille pikseleille ja soveltamalla tämän jälkeen valaistusmallia kullekin pikselille

(*Phong-sävytyt*). Koska valaistusmallin laskeminen on usein hidasta, tämä menetelmä on huomattavasti edellistä raskaampi.

Perinteisesti valaistus jaetaan ambienttiin, diffuusiin ja spekulaariin komponenttiin, joiden summana kokonaisvalaistus saadaan. Ambientti osuus on vain vakio, joka lisätään kaikkien pisteiden kirkkauteen. Koska fotonit voivat heijastua moneen kertaan, ennen kuin ne päätyvät silmään, toimivat kaikki pinnat tavallaan himmeinä valonlähteinä. Tämän vuoksi harvoin mikään piste on täysin varjossa, vaan kaikkialle pääsee valoa heijastusten kautta. Tällaisten ilmiöiden simuloinnin puute voidaan korvata ambientilla termillä, jotta pimeimmät kohdat eivät olisi liian pimeitä. Heijastelun vuoksi väri voi myös vuotaa pinnasta toiseen: punaisessa huoneessa valkea pöytä näyttää hieman punaiselta. Tätä ambientti termi ei tietysti mallinna mitenkään.

Diffuusi osuus olettaa, että fotonit siroavat pinnasta tasaisesti eri suuntiin, jolloin pinta näyttää samalta kaikista suunnista. Tällainen malli kuvaa hyvin mattapintaisia materiaaleja, kuten paperia. Jatkossa tulemme rajoittumaan usein pelkkään diffuusiin valaistukseen, koska se yksinkertaistaa esitystä huomattavasti. Lisäksi jo pelkkä hyvin toteutettu diffuusi valaistus luo paljon realismia.

Spekulaari komponentti taas perustuu ajatukseen pinnasta jonkinlaisena epätäydellisenä peilinä. Valon tulosuunnasta ja normaalista lasketaan fotonien heijastussuunta. Pinta on kirkkain kyseisestä suunnasta katsottuna, mutta himmenee nopeasti heijastussuunnan ja katselusuunnan välisen kulman  $\alpha$  kasvaessa. Tämä otetaan usein huomioon esimerkiksi Phongin lailla, joka on muotoa  $\cos^n \alpha$ , missä  $n$  on materiaalia kuvaava spekulaari eksponentti ( $1 \leq n < \infty$ ). Tällainen

malli tuottaa kirkkaita heijastustäpliä. Spekulaari valaistus on yleensä hankalampi toteuttaa kuin diffuusi, koska spekulaarista pinnasta lähtevä *radianssi* riippuu ratkaisevasti katsojan suunnasta. Radianssi kuvaa tiettyyn avaruuskulmaan ja pinta-alaan liittyvän valon määrää. Tämän artikkelin kannalta riittää ymmärtää, että se vastaa suunnilleen pinnan tai valonlähteen kirkkautta tiettyyn suuntaan.

Edellä kuvattu spekulaari-diffuusimalli on yksinkertainen ja täten nopea laskea, mutta se tuottaa muovisen näköisiä materiaaleja. Yleisemmin valaistumalleja kuvataan usein valaistussyhtälöllä, joka voidaan muotoilla esimerkiksi seuraavasti:

$$L_o(\omega_o) = \int_H L_i(\omega_i) f(\omega_i, \omega_o) V(\omega_i) \cos(\theta_i) d\omega_i, \quad (1)$$

missä  $L_o$  ("out") kertoo käsiteltävästä pisteestä avaruuskulman  $\omega_o$  suuntaan lähtevän valon radianssin. Tämä saadaan integroimalla kaikista pinnan etupuolella olevan puolipallon  $H$  eri suunnista  $\omega_i$  ("in") pisteeseen tuleva valo  $L_i$ . Tässä suunnat on ilmaistu pinnan normaalin suhteen, ja ne koostuvat oikeastaan elevaatiokulmasta ja atsimuuttikulmasta, esimerkiksi  $\omega_i = (\theta_i, \phi_i)$ . Kaikki funktiot ovat tietenkin myös käsiteltävän pisteen funktioita, mitä ei kuitenkaan ole kirjoitettu näkyville yhtälöön.

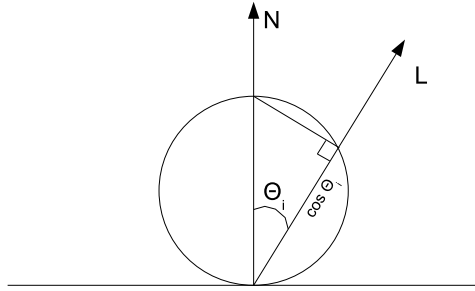
Saapuva valo täytyy painottaa kosinitermillä, jonka parametri  $\theta_i$  on pinnan normaalin ja valon tulosuunnan välinen kulma eli elevaatiokulma. Termi kuvaa maapallon pohjoisosissa asuville tuttua ilmiötä: mitä viistommin valo pinnalle saapuu, sitä vähemmän se vaikuttaa, koska valoteho jakautuu suuremmalle alalle. Tätä ilmiötä kutsutaan nimellä *Lambertin (kosini)laki*. Kuvassa 2 havainnollistetaan pinnan eteen asetetun pallon

avulla kerrointa, jonka tämä aiheuttaa eri suunnista tulevalle valolle. Samaa havainnollistuskeinoa hyödynnetään myöhemmin luvussa 4. Koska pelkän kosinin arvo olisi pinnan alapuolella negatiivinen, mutta alapuolelta valoa ei pisteeseen pääse, Lambertin termi muokataan usein muotoon  $U(\theta_i) = \max(0, \cos \theta_i)$ . Tämän funktion arvo pinnan alapuolella on 0, mutta yläpuolella haluttu elevaatiokulman kosini.

Funktio  $V$  kuvaa valon näkyvyyttä pisteeseen. Se saa arvon 0, jos kyseinen piste on varjossa suunnasta  $\omega_i$  tulevan valon suhteen, ja muussa tapauksessa arvon 1. Yksinkertaisessa tapauksessa  $V$  huomioi vain kappaleen itseensä langettamat varjot, mutta myös muiden kappaleiden tuottamat varjot voidaan ottaa huomioon. Koska  $V$ :tä ei voida käytännössä laskea reaaliajassa, on pelkässä itsevarjostuksessa se hyvä puoli, että se pysyy samana ajon aikana, ellei kappale muuta muotoaan. On hyvä huomata, että kappaleen liike sen sijaan ei muuta itsevarjostusta, koska  $V$  on ilmaistu kappaleen omassa, paikallisessa koordinaatistossa. Tämä otetaan huomioon kiertämällä valonlähde kutakin kappaletta valaistaessa sen koordinaatistoon.

Esimerkiksi kasvoissa suurin osa varjoista johtuu yleensä kasvojen omasta varjostuksesta, joten itsevarjostuksen ottaminen huomioon riittää usein tekemään kuvasta paljon aiempaa realistisemmän. Kappale voi myös tarkoittaa suurtakin, monen esineen kokonaisuutta, kunhan se ei muuta muotoaan ajon aikana.

Vielä yksi valon määrää painottava funktio on  $f$ , jota kutsutaan nimellä BRDF (*bi-directional reflectivity distribution function*). BRDF kertoo pinnalta lähtevän radianssin suhteen saapuvaan valoon eli *irradianssiin* valon tulo- ja lähtösuuntien funktiona. Näin se kattaa



**Kuva 2:** Lambertin kosinilaki.  $N$  on pisteen normaali ja  $L$  on valon saapumissuunta. Pisteen eteen piirretty  $\frac{1}{2}$ -säteinen ympyrä kuvaa kertoimen  $\cos\theta_i$ , jolla kustakin suunnasta tuleva valo skaalataan. Ympyrällä voidaan esittää kosinitermin vaikutus suunnan funktioihin havainnollisesti. Tätä hyödynnetään myöhemmin artikkelissa.

kaiken, mitä materiaalin heijastusominaisuuksista on tarpeen tietää valaistuksen laskemiseksi. Toisin sanoen valaistusyhtälön perusteella voidaan laskea katselusuuntaan lähtevä radianssi, jos tunnetaan materiaalin ominaisuudet (eli BRDF) ja kaikista eri suunnista pisteeseen tuleva valo.

Tarkemmin ottaen BRDF ei tosin kuvaakaan kaikkia mahdollisia pinnan ilmiöitä. Esimerkiksi todellisissa materiaaleissa valon aallonpituus vaikuttaa heijastukseen. Lisäksi aallonpituus voi myös muuttua heijastuksen yhteydessä. Yleensä BRDF oletetaan joko kokonaan riippumattomaksi aallonpituudesta tai ainakin käsitellään punainen, vihreä ja sininen valo kukin omalla BRDF:llään riippumattomina toisistaan.

Edellä kuvattu diffuusi-spekulaarimalli voidaan ajatella myös eräänä BRDF:nä. Vastaavasti useat kehittyneemmät valaistusmallit tuottavat omanlaisensa BRDF:n, joten BRDF on valaistuslaskennan kannalta hyödyllinen abstraktio: voidaan unohtaa, millainen valaistusmalli on käytössä ja keskittyä valaistusyhtälön ratkaisemiseen joko

mielivaltaisille BRDF:ille tai sopivasti rajoitetuille BRDF:ille, jotka täyttävät jotkin annetut ehdot (esimerkiksi ovat riippumattomia katselusuunnasta).

Valaistusyhtälöstä (1) saadaan hieman yksinkertaisempi muoto, jos BRDF rajataan diffuusiksi, eli sen ajatellaan kuvaavan materiaalia, joka näyttää samalta kaikista suunnista. Koska BRDF:n täytyy fysiikan lakien mukaan olla symmetrinen tulo- ja lähtösuunnan suhteen eli  $f(\omega_i, \omega_o) = f(\omega_o, \omega_i)$ , myös riippuvuus valon tulokulmasta katoaa. Tällöin  $f$  on vain vakio  $\rho/\pi$ , missä  $\rho$  on pinnan *albedo*. Se kuvaa pinnan radianssin suhdetta pinnalle kaikista suunnista saapuvan valon määrään (irradianssiin). Kun tämä yksinkertaistus on otettu huomioon ja näkyvyysfunktio  $V(\omega_i)$  sekä Lambertin kosinitermiä kuvaava funktio  $U$  kerrottu yhdeksi funktioksi  $M(\omega_i)$ , saadaan yksinkertaistettu valaistusyhtälö diffuusille materiaalille:

$$L_o = (\rho/\pi) \int_S L_i(\omega_i) M(\omega_i) d\omega_i. \quad (2)$$

Huomataan, että integrointialue on vaihdettu puolipallosta ( $H$ ) kokonaiseen pisteittä ympäröivään palloon ( $S$ ). Funktio  $U$

rajoittaa integroitavan termin nolllaksi pinnan alapuolella, joten tulos ei muutu. Palloa on kuitenkin integrointialueena puolipalloa helpompi käsitellä myöhemmin palloharmonisten funktioiden yhteydessä. Tämä muoto havaitaan muutenkin hyväksi lähtökohdaksi palloharmonisia funktioita käytettäessä luvussa 3.

Edes tämän yksinkertaistetun valaistusyhtälön ratkaiseminen annetussa integraalimuodossa ei ole mahdollista reaaliajassa. Tämän vuoksi reaaliaikagrafiikassa käytetään tyypillisesti pistemäisiä valonlähteitä. Muista kuin valonlähteen suunnista tuleva radianssi oletetaan nolllaksi ja näkyvyysfunktio  $V$  ykköseksi. Tällöin integraali muuttuu summaksi muutaman erillisen pistevalonlähteen suunnan yli, joten  $L_0$  voidaan laskea hyvin nopeasti. Oletus ei valitettavasti ole kovin realistinen, sillä todelliset valonlähteet ovat harvoin lähelläkään pistemäisiä.

Lisäksi kuvattu perusmalli on täysin paikallinen. Se olettaa käsiteltävän kulmapisteen olevan äärettömällä tasolla, joka estää valoa tulemasta pinnan alta, mutta unohtaa koko muun mallin olemassaolon. Näin ollen menetelmä ei synnytä lainkaan varjoja eikä ota huomioon valon heijastelua pinnasta toiseen.

Varjojen laskemiseen on kehitetty omia, erikoistuneita algoritmeja, joista suosituin lienee *varjokartta*. Tässä artikkelissa ei perehdytä varjoalgoritmeihin sen syvällisemmin, vaan tyydytään toteamaan, että varjokartat perustuvat edelleen pistemäisiin valonlähteisiin. Tämän seurauksena syntyvät varjot ovat erittäin teräväreunaisia. Menetelmää on tutkittu ja käytetty erittäin paljon. Erilaisia varjoja pehmentäviä lisäyksiä on keksitty, mutta menetelmät ovat usein laskennallisesti raskaita tai niiden tulokset jättävät toivomisen varaa.

Todellisissa tilanteissa valonlähteet ovat usein laajoja alueita. Tyypillinen esimerkki tästä on pilvinen päivä ulkotiloissa, jolloin valoa tulee taivaalta joka suunnasta lähes yhtä paljon. Tällaisessa tilanteessa valaistusyhtälön integraalin laskemista ei voida välttää pistevalonlähteiden avulla. Taivaalta tuleva valaistus voidaan laskea esimerkiksi analyttisistä malleista tai eri suuntiin otettujen kuvien perusteella.

Valon heijastumista ympäristön muista pinnoista on vaikea mallintaa reaaliaikaisesti. Esilasketavaiheessa voidaan käyttää esimerkiksi erittäin raskasta *radiositeettimenetelmää* laskemaan kunkin pinnan kirkkaus, kun varjostukset ja heijastukset otetaan huomioon. Menetelmässä asetetaan valonlähteinä toimiville pinnoille alkukirkkaus, minkä jälkeen valoa levitetään iteratiivisesti kustakin pinnasta kaikkiin muihin näkyviin pintoihin. Koska iteraation kesto on näin ollen verrannollinen pintojen määrän neliöön, ei menetelmä sovellu reaaliajassa laskettavaksi mahdollisesti sadoista tuhansista pinnoista koostuvilla malleilla.

Radiositeettimenetelmän tulokset voidaan kuitenkin tallentaa kunkin pinnan kirkkaudeksi mallissa, jos oletetaan, että valonlähteet tai malli eivät muutu lainkaan. Tämä on erittäin vahva oletus, joten menetelmän käyttömahdollisuudet ovat rajalliset.

Seuraavaksi esiteltävä palloharmonisten funktioiden ja esilasketun siirtofunktion yhdistelmä ratkaisee joitakin edellä esitetyistä ongelmista. Menetelmän avulla radiositeettimenetelmän tuloksia voidaan hyödyntää muuttuvissa valaistusolosuhteissa. Lisäksi menetelmällä saadaan mukaan laajat valonlähteet, joten sen avulla voidaan laskea pehmeitä varjoja. Itse asiassa nimenomaan terävät varjot tuottavat tekniikalle ongelmia.

### 3 Palloharmoniset funktiot

Palloharmoniset funktiot (SH, *spherical harmonics*) ovat funktioperhe, joka saadaan muun muassa eräissä fysiikan ongelmissa esiin tulevien differentiaaliyhtälöiden ratkaisueina. Kukin palloharmoninen funktio  $Y_\ell^m$  määritellään seuraavasti:

$$Y_\ell^m(\theta, \phi) = C_\ell^m P_\ell^m(\cos(\theta)) e^{im\phi},$$

missä  $\ell = 0, 1, \dots$  ja  $m = -\ell, \dots, \ell$ ,  $P_\ell^m$  ovat ns. Legendren liittofunktiot ja  $C_\ell^m$  ovat normalisointivakioita. Legendren liittofunktioiden määritelmää ei esitetä tässä, mutta niiden laskemiseen on olemassa joukko rekursiokaavoja [1].

Edellisten, kompleksiarvoisten palloharmonisten funktioiden avulla saadaan reaaliset palloharmoniset funktiot, jotka havaitaan varsin hyödyllisiksi tietokonegrafiikassa:

$$y_\ell^m(\theta, \phi) = \begin{cases} \sqrt{2} \operatorname{Re}(Y_\ell^m(\theta, \phi)), & m > 0 \\ \sqrt{2} \operatorname{Im}(Y_\ell^m(\theta, \phi)), & m < 0 \\ Y_\ell^0(\theta, \phi), & m = 0 \end{cases}$$

Koska tässä artikkelissa käytetään vain reaalisia palloharmonisia funktioita, viitataan lyhenteellä SH niihin. Kukin funktioista on määritelty pallon pinnalla eli ottaa parametrikseen suunnan  $s = (\theta, \phi)$  ja tuottaa reaaliarvon.

Funktion  $y_\ell^m$  alaindeksi  $\ell$  kertoo, mille *kaistalle* (band) funktio kuuluu. Lisäksi  $\ell$ :nnellä kaistalla on  $2\ell + 1$  funktiota, joita yksilöi yläindeksi  $m$ . Kunkin kaistan  $\ell$  funktiot vastaavat tiettyä paikkataajuutta pallon pinnalla. Matalilla taajuuksilla funktion arvo muuttuu hitaasti kulman mukana. Lisäksi  $\ell$ :nnen kaistan funktiot ovat  $\ell$ :nnen asteen polynomeja esitettynä karteesisten koordinaattien  $(x, y, z)$  avulla. Erityisesti kaista  $\ell = 0$  on siis vakiotermi. Indeksillä  $m$  erottelee samantaajuiset funktiot toisistaan. Itseisarvoltaan korkeampaa

$m$ :ää vastaavat funktiot ovat korkeampitaajuisia atsimuuttikulman  $\phi$  suhteen, mutta vastaavasti matalampitaajuisia  $\theta$ :n suhteen. Ääritapauksena funktiot, joille  $m = 0$ , ovat vakioita  $\phi$ :n suhteen. Funktioiden muodoista ja parametrien roolista saa parhaiten käsityksen tutkimalla niiden graafista esitystä kuvassa 3.

Usein SH:t halutaan esittää yhdellä yksikäsitteisellä indeksillä  $j$ , jolloin käytetään tyypillisesti muunnosta  $j = \ell(\ell + 1) + m$ .

Erityisen käyttökelpoisia SH:ista tekee se, että ne määrittelevät ortonormaalien kannan pallon pinnalla. Funktioiden ortonormaalius tarkoittaa, että funktioiden  $y_j$  ja  $y_k$  sisätulo on 0, jos  $j \neq k$ , ja 1, jos  $j = k$ . Tässä funktioiden sisätulo  $\langle a|b \rangle$  määritellään integraalina yli funktioiden määrittelyalueen  $S$ :

$$\langle a|b \rangle = \int_{s \in S} a(s)b(s) ds$$

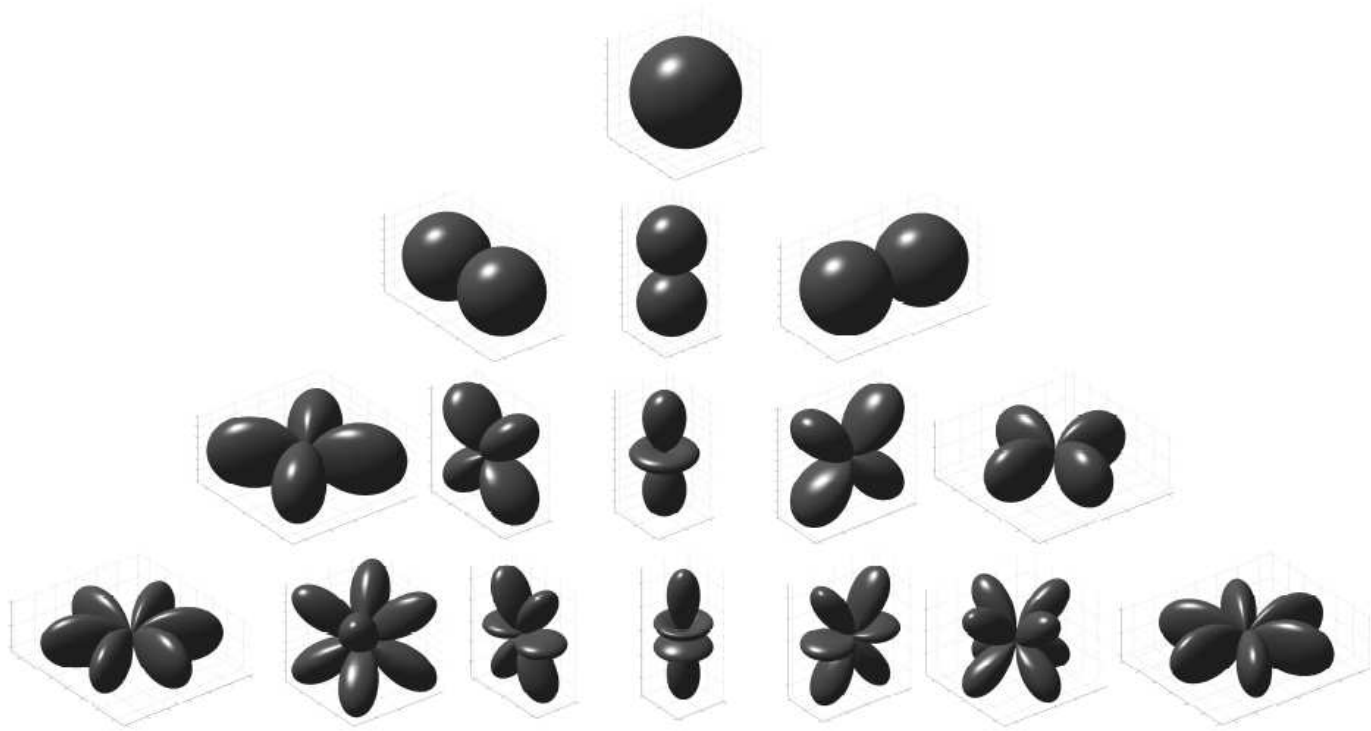
SH-kannan avulla mikä tahansa pallon pinnalla määritelty reaaliarvoinen funktio  $f(s)$  voidaan esittää äärettömänä summana painokertoimilla  $c_j$  skaalattuja kantafunktioita  $y_j$ :

$$f(s) = \sum_{j=0}^{\infty} c_j y_j(s)$$

Tilanne muistuttaa Fourier-muunnosta. Tunnetusti mikä tahansa signaali voidaan esittää äärettömänä summana eri taajuisia siniaaltoja. Kuten Fourier-muunnoksessa, myös SH-kehityksessä ensimmäiset kantafunktiot vastaavat pienitaajuisia funktioita pallon pinnalla, ja mitä suuremmaksi  $j$  kasvaa, sitä suurempia paikkataajuuksia saadaan mukaan.

Jos summaan otetaan mukaan vain  $n$ :ää ensimmäistä kaistaa vastaavat kantafunktiot, voidaan funktiota  $f$  approksimoida vastaavilla kertoimilla  $c_0, \dots, c_{n^2-1}$ . Tätä kutsutaan funktion projisoimiseksi SH-





**Kuva 3:** Ensimmäiset 16 palloharmonista funktiota. Kullakin rivillä on yksi kaista (ylinnä  $\ell = 0$ ), ja  $m$ :n arvo kasvaa vasemmalta  $-\ell$ :stä oikealle  $\ell$ :ään. Mitä suuremman itseisarvon funktio saa, sitä kauempana keskipisteestä pinta on kuvissa. Kuva lähteestä [10].

kantaan. Prosessissa funktiolle saadaan siis ikään kuin koordinaatit kantafunktioiden suhteen. Projisoitua funktiota merkitään tässä tekstissä funktion yläpuolisella aaltoviivalla ( $\tilde{f}$ ).

Approksimaatiossa syntyvän virheen neliösumma saadaan minimoitua [8], jos funktio projisoidaan integroimalla sen ja kunkin mukaan otetun kantafunktion tuloa:

$$c_j = \int_S f(s)y_j(s)ds.$$

Kuten arvata saattaa, tämän integraalin laskeminen on erittäin hidasta. Arvo voidaan laskea numeerisesti Monte Carlo-menetelmällä. Tämä vaihe on osaltaan vastuussa SH-menetelmiin liittyvistä pitkistä esikäsitelyajoista. SH-tekniikan sovelluksissa pidetäänkin huoli, että projektiot tehdään ajon aikana mahdollisimman harvoin.

On hyvä huomata, että ottamalla mukaan  $n$  ensimmäistä kaistaa, saadaan  $n^2$  kantafunktiota (ja vastaavasti kerrointa  $c_j$ ). Suuremmalla määrällä voidaan esittää funktiosta korkeampia taajuuksia, mutta sekä ajon aikana tehtävien laskutoimitusten viemä aika että kerrointen varastointiin käytetty tila kasvavat suhteessa neliöllisesti. Esilaskennan kesto kasvaa tätäkin nopeammin. Tämän kompromissin vuoksi yleensä otetaan mukaan sovelluksesta riippuen esimerkiksi kahdesta kahdeksaan ensimmäistä kaistaa.

Jo kahdella kaistalla saadaan esitettyä moniin sovelluksiin riittäviä matalataajuisia funktioita, joita voidaan käyttää valaistukseen. Tällaisen funktion esitys SH-kannassa vaatii siis vain neljä reaalilukua ( $c_0, \dots, c_3$ ).

Tiiviin esitysmuodon lisäksi SH-kannasta on muutakin hyötyä. Kannan ortonormaalius on erityisen kätevä ominaisuus. Oletetaan, että kaksi funktiota  $f$  ja  $g$  on approksimoitu SH-kannassa muutamien ensimmäisien

kertoimien avulla:  $\tilde{f}(s) = \sum c_j y_j(s)$  ja  $\tilde{g}(s) = \sum d_j y_j(s)$ . Funktioiden sisätulo eli tulon integraali  $\int \tilde{f}(s)\tilde{g}(s)ds$  redusoituu nyt kantafunktioiden ortonormaaliuden ansiosta pistetuloksi  $c^T d = \sum c_j d_j$ .

Jos tarkastellaan aiemmin esiteltyä yksinkertaistettua valaistusyhtälöä (2), havaitaan sen olevan valaistusfunktion  $L_i(\omega_i)$  ja funktion  $M(\omega_i)$  sisätulo vakiolla kerrottuna. Niinpä hankala valaistusyhtälö redusoituu nopeasti laskettavaksi pistetuloksi, jos  $L_i$  ja  $M$  esitetään SH-kannassa.  $M(\omega_i)$  voidaan laskea etukäteen kullekin kulmapisteelle ja  $L_i(\omega_i)$  kullekin valonlähteelle. Tämä on PRT-tekniikan idea, johon syvennyttään seuraavassa luvussa. Kuten todettua, integrointialueen muuttaminen koko pallon  $S$  alkuperäisen valaistusyhtälön puolipallosta  $H$  sopii hyvin pallon pinnalla määriteltyjä SH-funktioita käytettäessä.

Jos halutaan, että valonlähteet ja kappaleet voivat liikkua toistensa suhteen, täytyy SH-kannassa esitetty  $L_i$  pystyä kiertämään kappaleen paikalliseen koordinaatistoon. Tällöin erittäin hyödyllinen ominaisuus on SH-funktioiden kiertoinvarianttius. Kiertoinvarianttiudessa on taustalla se, että palloharmoniset funktiot on määritelty pallon pinnalla, joten ne ovat täysin sykliisiä. Tämän ansiosta kierrossa ei menetetä tarkkuutta, vaan tulos on yhtä hyvä ja luonnollinen kiertokulmasta riippumatta.

Oletetaan, että alkuperäistä funktiota  $f$  halutaan kiertää operaattorilla  $R$ . Tällöin voidaan rakentaa vastaava kiertooperaattori  $\hat{R}$ , jolla kierretään suoraan funktion projektiota  $\tilde{f}$  eli kerroinvektoria  $c$ . Kiertooperaattori  $\hat{R}$  paljastuu melko harvaksi ( $n^2 \times n^2$ )-matriisiksi, jonka rakentamista kuvaa esimerkiksi Green [1]. Ongelmana on se, että matriisin alkioiden määrä kasvaa neljanteen potenssiin

suhteessa mukaan otettujen kaistojen määrään  $n$  (tosin suurin osa alkioista on nollia). Kiertäminen tapahtuu kertomalla matriisi kerroinvektorin kanssa, joten kierron laskennallinen vaativuus kasvaa nopeasti  $n:n$  mukana.

#### 4 Esilaskettu radianssin siirtofunktio

Esilaskettu radianssin siirtofunktio (PRT, *precomputed radiance transfer*) on tekniikka, jota käytetään laskemaan kustakin kulmapisteestä eri suuntiin lähtevän valon radianssi. Oletetaan tässä vaiheessa, että materiaalit ovat diffuuseja, jolloin lähtevä radianssi ei riipu lähtösuunnasta. Tämä luku esittää PRT-tekniikan suunnilleen siinä muodossa, jossa Sloan ym. [5] sen alunperin esittivät.

##### 4.1 PRT:n perusajat

PRT:n ideana on jakaa kulmapisteen kirkkauteen vaikuttavat tekijät kahteen osaan: pisteeseen lankeavaan valoon  $L_i$  ja sen *siirtofunktioon* (radiance transfer function).  $L_i$  määritellään tilanteessa, jossa maailmassa ei ole mitään valon kulkua estäviä kappaleita. Se riippuu toisin sanoen valonlähteistä mutta ei mallista.

Siirtofunktio taas kuvaa, miten kulmapisteestä lähtevä radianssi voidaan laskea saapuvan valon  $L_i$  perusteella. Siirtofunktion  $M$  tehtäväksi jää ottaa huomioon mallin geometrian vaikutus: Lambertin kosinitermi, varjostus ja valon heijastuminen pinnasta toiseen.

Abstraktisti voidaan ajatella kulmapisteestä lähtevää radianssia kuvaavan funktion  $L'$  olevan muotoa  $L'(M(L_i))$ . Olennaista on siis se, ettei  $M$  vaikuta saapuvaan valoon  $L_i$ . Käytännössä, kun PRT-tekniikkaa käytetään SH:n kanssa, halutaan  $L'$ :n olevan valaistusyhtälön (2) mukainen integraali  $L_i$ :n ja  $M$ :n tulosta,

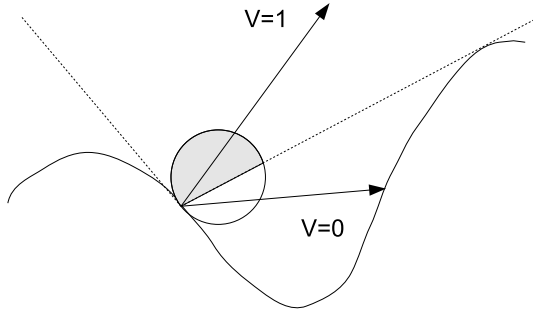
jotta päästään hyödyntämään SH-kannan ortonormaaliutta.

Kuten edellisessä luvussa todettiin, siirtofunktio  $M$  voidaan usein laskea etukäteen kussakin kulmapisteessä.  $M$  projisoidaan SH-kantaan esilaskentavaiheessa ja mukaan otetaan  $n^2$  ensimmäistä kerrointa. Funktion  $M$  projektio on siis  $\tilde{M} = \sum_{j=0}^{n^2-1} m_j y_j$ . Reaalisten kertoimien  $m_j$  muodostamaa  $n^2$ -ulotteista vektoria kutsutaan *siirtovektoriksi* (transfer vector), koska se sisältää kaiken tarvittavan informaation kyseisen kulmapisteen siirtofunktiosta.

Valaistus voidaan joko laskea kokonaan etukäteen tai määrittää ajon aikana kullekin kappaleelle. Näin saadaan funktio  $L_i$ , jonka kertoimet SH-kannassa muodostavat  $n^2$ -ulotteisen valaistusvektorin. Vaikka projektio on raskas operaatio, riittää projisoida valaistusfunktio vain kerran kappaletta kohti, ei jokaiselle kulmapisteelle. Tämän jälkeen valaistusvektori kierretään käsiteltävän kappaleen koordinaatistoon kertomalla se sopivalla  $\hat{R}$ -operaattorilla.

Lopulta kulmapisteen valaistus  $L'$  saadaan valaistusyhtälön (2) mukaisesta integraalista. SH-kannassa ilmaistuille funktioille se on sisätulo  $\langle \tilde{L}_i | \tilde{M} \rangle$ , joka redusoi- tuu valaistusvektorin ja siirtovektorin pistetuloksi SH-kannan ortonormaaliuden ansiosta. Kulmapistettä kohti tehdään siis vain yksi  $n^2$ -ulotteisten vektorien pistetulo.

Yksinkertaisimmillaan  $M$  koostuu vain kosinitermistä  $U$ , jolloin saadaan perinteinen paikallinen valaistusmalli. Kuvassa 2 nähty ympyrä kuvaa tällaisen siirtofunktion. Pinnan etupuolella sen arvo siis on suoraan elevaatiokulman  $\theta$  kosini ja pinnan alapuolella nolla.



**Kuva 4:** Itsevarjostus. Pisteeseen pääsee valoa vain suunnista, joihin lähtevä säde ei törmää kappaleeseen itseensä ( $V = 1$ ). Katkoviivat kuvaavat tämän alueen reunat. Harmaa alue kuvaa siirtofunktiota  $M = VU$ : paljonko kustakin suunnasta tulevaa valoa pitää skaalata, jotta saadaan pisteeseen pääsevä valo. Värittämättömällä alueella  $V = 0$  ja täten myös  $M = 0$ .

## 4.2 Itsevarjostus

PRT-menetelmä alkaa näyttää voimaansa, kun mukaan otetaan näkyvyysfunktio  $V$ , jolloin kulmapisteen siirtofunktioksi saadaan  $M(\omega_i) = V(\omega_i)U(\theta_i)$ . Kuten johdannossa todettiin, usein näkyvydessä rajoitutaan vain kappaleen itsevarjostukseen, jossa muut kappaleet unohdetaan kokonaan.

Itsevarjostuksen huomioon ottava  $M$ -funktio voidaan laskea kokonaan etukäteen, ja se pysyy vakiona myös kappaleen liikuessa, kunhan kappale ei muuta muotoaan. Kulmapisteen  $V$  selvitetään esilaskentavaiheessa ampumalla pisteestä säteitä lukuisiin eri suuntiin. Jos säde osuu kappaleeseen itseensä, ei ulkopuolelta tuleva valaistus pääse kyseisestä suunnasta perille pisteeseen. Näin ollen  $V$  ja täten myös  $M$  saa arvon 0. Muulloin  $V$ :n arvo on 1 ja  $M$ :n arvoksi tulee  $U(\theta_i)$ . Näin saatu  $M$  projisoidaan SH-kantaan, jolloin saadaan kullekin pisteelle itsevarjostuksen huomioon ottava siirtovektori. Tätä siirtovektoria voidaan käyttää aliluvussa 4.1 kuvatulla tavalla. Kussakin kulmapisteessä lasketaan vain siirtovektorin ja valaistusvektorin välinen

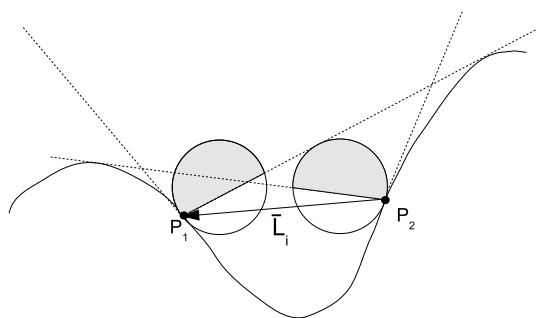
pistetulo. Varjojen lisääminen ei siis muuta ajon aikana tapahtuvaa prosessia mitenkään! Muutenkin esilaskettuun siirtofunktioon voidaan sisällyttää monimutkaisia ilmiöitä muuttamatta ajon aikana tapahtuvaa prosessia lainkaan.

Kuvassa 4 esitetään, kuinka  $V$ :n arvot rajaavat pelkästä Lambertin laista saatavan, ympyrän näköisen siirtofunktion vain niihin suuntiin, joista valo todella pääsee perille. Oikeanpuoleisen katkoviivan oikealta puolelta tuleva valo ei pääse pisteeseen, joten siirtofunktion arvo sitä vastaavissa suunnissa on  $\cos \theta$ :n sijaan 0. Pinnan alapuolelta tuleva valo karsiutuu jo funktion  $U$  vuoksi.

## 4.3 Valon heijastuminen

Vielä realistisempi siirtofunktio saadaan ottamalla huomioon myös valon heijastuminen kappaleen pisteestä toiseen. Tämä on juuri se ilmiö, joka usein korvataan kaikkien pisteiden valoisuuteen lisättävällä ambientilla vakioterminä. Heijastusvaikutusta voidaan mallintaa seuraavanlaisella valaistusyhtälöllä:

$$L'' = L' + (\rho/\pi) \int_S (1 - V) \bar{L}_i U d\omega_i, \quad (3)$$



**Kuva 5:** Itseheijastus. Pisteeseen  $P_1$  pääsee valoa suoran valaistuksen lisäksi heijastumalla esimerkiksi pisteestä  $P_2$ . Pisteestä  $P_2$  heijastuva valo  $\bar{L}_i$  perustuu puolestaan sinne saapuvaan suoraan valoon. Heijastuskierroksia voidaan laskea mielivaltaisen monta. Katso myös kuva 6.

missä  $L'$  on edellä laskettu valaistusfunktio (valaistusyhtälön mukainen integraali) ja  $\bar{L}_i$  kuvaa suunnassa  $\omega_i$  olevasta pinnasta heijastuneen valon radianssia.

Tässä lähestymistavassa  $L'$  voidaan laskea kussakin pisteessä ja sen jälkeen käyttää sitä eri suunnista tulevan toissijaisen valon radianssin arvioimiseen yhtälön mukaisesti. Iteroimalla tätä prosessia saadaan mukaan yhä useampia heijastumisia, jolloin tulos on yhä realistisempi ja realistisempi. Käytännössä ei tarvita kuin pari kierrosta, mutta toisaalta esilaskennan aikana tehdyt lisäkierrokset eivät hidasta ajon aikana tapahtuvaa laskentaa, joka on edelleen vain valaistusvektorin ja siirtovektorin pistetulo.

Käytännössä heijastuksia laskettaessa sovelletaan vielä yhtä yksinkertaistusta. Yhtälössä (3) valaistus ja siirt ominaisuudet eivät ole enää erotettavissa toisistaan, joten se ei ole enää valaistuksen ja siirtofunktion sisätulo, mikä taas olisi edellytyksenä SH/PRT-menetelmän käytölle. Niinpä menetelmää ei voida soveltaa siihen sellaisenaan. Sloan ym. käyttävät suoraviivaista approksimaatiota, jolla tilanne palautetaan haluttuun muotoon laskemalla siirtofunktio, joka

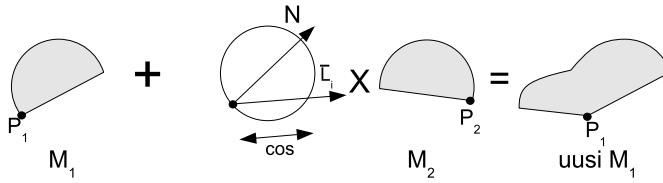
ottaa heijastelun huomioon. He käyttävät seuraavia kaavoja [5]:

$$m_k^0 = (\rho/\pi) \int_S V U y_k d\omega_i$$

$$m_k^j = (\rho/\pi) \int_S (1 - V) U m_k^{j-1} d\omega_i,$$

missä  $m_k^j$  sisältää  $j$ :n iterointia  $k$ :n SH-kertoimen ( $0 \leq k \leq n^2 - 1$ ). Tässä  $j$ :s iterointi mallintaa kaiken pisteeseen saapuvan valon, joka on kimmonnut kappaleesta  $j$  kertaa. Lopullinen siirtovektori kullekin kulmapisteelle saadaan yksinkertaisesti laskemalla kaikkien iterointien kertoimet yhteen:  $m_k = \sum m_k^j$ . Näin saadun siirtovektorin sisältämät kertoimet approksimoivat yhtälön (3) mukaista ilmiötä käytännössä tyydyttävästi, vaikkakaan eivät matemaattisen täsmällisesti. Menetelmässä käytetään oikeastaan edellisen kierroksen SH-kertoimia arvioimaan  $\bar{L}_i$ :tä.

Prosessin tuloksena saatu siirtovektori ottaa nyt siis huomioon myös valon heijastelun kappaleessa. Siirtovektorin soveltaminen ajon aikana tapahtuu täsmälleen kuten aiemminkin, mutta nyt tuloksena saatava valaistus kattaa itsevarjostuksen ja valon heijastumisen kappaleesta itsestään.



**Kuva 6:** Sloanin ym. [5] menetelmä heijastusten saamiseksi mukaan siirtofunktioon. Pisteeseen  $P_1$  saapuu valo sekä suoraan että eri suunnista heijastumalla. Kuvassa 4 saatuun siirtofunktioon lisätään tulokulman kosinilla kerrottu (ja albedosta riippuvalla vakioterminillä skaalattu) pisteen  $P_2$  siirtofunktio kuvasta 5. Vastaavasti pitäisi ottaa huomioon muistakin suunnista heijastunut valo. Lopputuloksena saadaan pisteelle  $P_1$  siirtofunktio, joka ottaa huomioon itseheijastuksen. Alkuperäisen siirtofunktion  $M_1$  mukaan lähinnä sivun oikeasta yläkulmasta tuleva valo pääsisi vaikuttamaan pisteen kirkkauteen, mutta uusi  $M_1$  kertoo, että myös vasemmalta yläkulmasta tuleva valo vaikuttaa heijastuksen kautta pisteeseen  $P_1$ , joskin vähemmän kuin oikealta tuleva.

Tilannetta on havainnollistettu kuvissa 5 ja 6. Pisteeseen  $P_1$  siirtofunktioon lisätään heijastunut valo kaikista niistä suunnista, joille  $V = 0$ . Eräs tällainen suunta on pisteen  $P_2$  suunta. Pisteeseen  $P_2$  siirtofunktio kerrotaan vakiolla  $\rho/\pi$ , jolloin tulos vastaa pisteeseen  $P_1$  pisteen  $P_2$  suunnasta saapuvaa valoa  $\bar{L}_i$ . Tämä täytyy vielä skaalata saapumissuunnan ja  $P_1$ :n normaalin välisellä kosinilla, minkä jälkeen se voidaan lisätä pisteen  $P_1$  alkuperäiseen siirtofunktioon. Uusi siirtofunktio ottaa nyt huomioon myös heijastuneen valon.

Kuva 7 havainnollistaa tässä luvussa käsiteltyjen erilaisten siirtofunktioiden vaikutusta yksinkertaiseen malliin.

#### 4.4 Muita laajennuksia

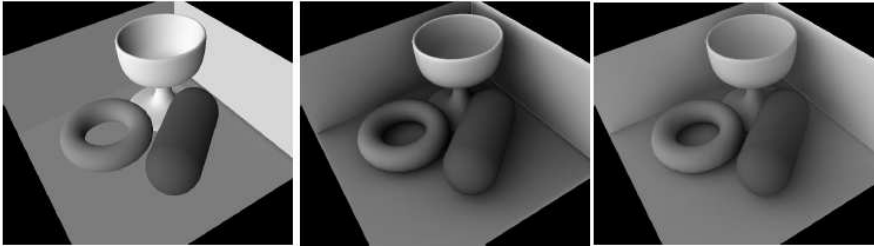
Käsitelty menetelmä ei ole vielä aidosti globaali, koska valaistukseen vaikuttaa vain yhden, muotoaan muuttamattoman kappaleen geometria. Tietenkin täysin staattisissa malleissa, joissa esineet eivät liiku toisiinsa nähden, ”kappale” voisi olla koko malli. Yleensä ottaen kappaleen käsite tarkoittaa tässä yhteydessä mitä tahansa joukkoa pintoja, jotka eivät liiku

toistensa suhteen ajon aikana, jolloin siirtofunktio pysyy vakiona.

Tähän asti on yksinkertaisuuden vuoksi oletettu, että BRDF on diffuusi. SH/PRT-menetelmä soveltuu silti ainakin teoriassa myös yleisille BRDF:ille [2]. Hieman yksinkertaistaen ajatuksena on, että tällöin kutakin pistettä varten tallennetaan siirtovektorin sijaan taulukollinen siirtovektoreita. Kukin taulukon alkio vastaa yhtä katselusuuntaa, joten käytettävä siirtovektori haetaan ajon aikana sen hetkisen katselusuunnan perusteella. Toisaalta mikään ei estä laskemasta kirkkaita, spekulareja heijastuksia jollakin muulla tekniikalla ja yhdistämästä tuloksia SH/PRT-tekniikalla saatuun diffuusiin valaistukseen.

## 5 Ongelmia ja ratkaisuja

Vaikka SH/PRT-tekniikka on osoittautunut käyttökelpoiseksi tietyissä tilanteissa, siihen liittyy myös monia ongelmia ja rajoituksia. Koska vaadittavien SH-kerrointen määrä kasvaa suhteessa mukaan otettujen kaistojen neliöön ja esimerkiksi kiertojen laskennallinen vaativuus vielä tätäkin nopeammin, ei tekniikal-



**Kuva 7:** Erilaisten siirtofunktioiden avulla piirrettyjä kuvia. Vasemmalla pelkkä paikallinen, diffuusi valaistus. Keskellä mukana itsevarjostus ja oikealla lisäksi valon heijastuminen. Heijastumisen ansiosta pimeimpiin paikkoihin pääsee aiempaa enemmän valoa. Kuvat lähteestä [1].

la käytännössä voida esittää kovinkaan korkeita paikkataajuuksia. Tämä koskee kaikkia funktioita, jotka projisoidaan SH-kantaan, eli valonlähteitä ja erilaisia siirtofunktioon otettavia termejä.

Valonlähteiden osalta taajuusrajoitus tarkoittaa, ettei valonlähteissä voi olla teräviä reunoja, vaan niiden tulee hiipua vähitellen. Parhaiten tähän sopivat ulkotilat, joissa valaistuksena toimiva taivas voidaan usein esittää melko tarkasti pienilläkin taajuuksilla. Monissa muissaakaan tilanteissa valonlähteiden epätarkka mallinnus ei ole visuaalisesti suuri ongelma verrattuna tekniikan mahdollistamien pehmeiden varjojen hyötyihin.

Mataliin taajuuksiin rajoittumisen ongelma koskee myös itsevarjostusta. Tyyppillisesti varjostusta laskettaessa tarvitaan enemmän kaistoja kuin pelkissä valaistuslaskelmissa, mutta tarkka määrä riippuu toivotusta visuaalisesta laadusta. Esimerkiksi kuvassa 7 kaksi oikeanpuoleisinta pikkukuvaa on tuotettu viidellä kaistalla, jolloin tulos on jo paljon vasemmanpuolimmaisesta kuvasta paikallista valaistusta luonnollisempi.

Jos projisoitavissa funktioissa on paljon informaatiota sellaisilla taajuuksilla, joita valitulla kaistamäärällä ei voida

esittää, voi syntyä hyvin silmiinpistäviä laskostumisartefakteja. Artefakteilla tarkoitetaan grafiikassa käytetyn tekniikan rajoituksista tai approksimaatioista johtuvia, selvästi epärealistisia visuaalisia tuloksia. Tämä voi ilmetä esimerkiksi valonlähteiden tapauksessa aivan väärälle puolelle ilmestyvinä kirkkaina valoina. Ongelma voidaan onneksi melko tehokkaasti kiertää, kuten signaalinkäsittelyssä yleensäkin, ajamalla muunnettavat funktiot sopivan alipäästösuotimen läpi [8].

Eräs ongelma on se, että valaistusolosuhteet ovat todellisuudessa erilaiset eri puolilla kappaletta varsinkin tilanteissa, joissa valonlähteet ovat lähellä käsiteltävää kappaletta. SH/PRT-tekniikassa oletetaan, että samaa valaistusta kuvaavan funktion projektiota  $\tilde{L}_i$  voidaan käyttää kaikille mallin kulmapisteille.  $L_i$ :n projisoiminen erikseen kullekin kulmapisteelle olisi liian raskasta. Ongelmaa voi lieventää selvittämällä valaistus monessa pisteessä eri puolilla kappaletta ja interpoloimalla näin saatuja tuloksia kulmapisteen sijainnin perusteella. Jälleen kerran ulkotiloissa tämä ongelma ei ole kovin paha, koska taivas on hyvin kaukana, joten sieltä tuleva valaistus on kaikissa mallin pisteissä käytännössä samanlaista.

Edellä keskityttiin lähinnä diffuuseihin BRDF:iin, mikä ei ole vain sattumaa. Spekulaarien ja varsinkin kehittyneempien BRDF:ien tapauksessa esilaskettua dataa tarvitaan moninkertainen määrä verrattuna diffuusiin tapaukseen. Ongelmana on myös se, että yleisissä BRDF:issä on usein korkeita taajuuksia, jolloin kaistoja täytyy ottaa mukaan diffuusialaistusta enemmän. Niinpä SH/PRT-menetelmää sovelletaankin eniten diffuusin valaistuksen laskemiseen. Samalla voidaan laskea spekulaari valaistus ynnä muut erikoisemmat BRDF:t toisilla, niihin paremmin sopivilla tekniikoilla. Jos yleisiä BRDF:iä kuitenkin käytetään, muodostuu siirtomatriisien muistinkulutus ongelmaksi. Niinpä yksi tutkimuskohde on ollut siirtofunktioiden informaation esittäminen mahdollisimman tiiviisti esimerkiksi pääkomponenttianalyysin avulla [6].

Vaikka sopivasti toteutettuna ja kaikkien helpottavien oletusten täyttyessä SH/PRT-tekniikka ei vaadi ajon aikana laskettaessa paljon tehoa, esilaskenta saattaa silti kestää hyvinkin kauan. Yksi tutkimuskohde onkin esilaskennan nopeuttaminen. Nopeammasta esilaskennasta olisi hyötyä erityisesti sen vuoksi, että tekniikka olisi tällöin helpompi integroida osaksi vaikkapa pelien tekemisen liukuhihnaa taiteilijoilta valmiiksi peliksi. Jos mallintaja joutuu odottamaan seuraavaan päivään nähdäkseen jokaisen mallin muutoksen sellaisena kuin se lopulta tuotteessa esiintyy, tekniikka on monissa tilanteissa käyttökelvoton.

Kuten edellä on jo käynyt ilmi, PRT-tekniikka ei sellaisenaan toimi kappaleen muuttaessa muotoaan esimerkiksi animaatioissa. Tällaisessa tilanteessa siirtofunktio jokaisessa mallin kulmapisteessä tulisi laskea ja projisoida uudestaan, mitä ei voida tehdä reaaliajassa. Samasta syystä tekniikkaa ei yleensä edes yritetä käyttää

koko mallille, vaan tyydytään yksittäisen esineen itsevarjostukseen. Koko mallin tapauksessahan varjostustilanne muuttuu aina minkä tahansa kappaleen liikkeessa.

SH/PRT-tekniikan tuottamat varjot esittävät siis lähinnä kappaleen itsevarjostusta. Nämäkin varjot ovat todenmukaiset vain hitaasti paikan mukaan muuttuville varjoille, koska SH/PRT-tekniikalla valaistus lasketaan vain kulmapisteissä ja interpoloidaan sisäpisteille Gouraud-varjostuksella. Suuret, laakeat monikulmiot voidaan pilkkoa pienemmiksi, jotta kulmapisteitä olisi tiheämmässä. Realismiin pyrkivässä piirrosta tämäkään ei riitä, vaan myös kappaleiden toisiinsa heittämät varjot täytyisi laskea ja toisinaan monikulmion keskelle syntyy teräviä varjoja, joita SH/PRT-tekniikka ei tuota.

Muodonmuutosten käsittelyä varten on kehitetty joitakin tekniikoita [7, 9]. Yleispätevää ratkaisua ei silti ole olemassa, vaan esimerkiksi yhtäaikainen kappaleiden muodonmuutos ja liikkuminen toistensa suhteen aiheuttaa yhä ongelmia [4]. Jos mallin muodonmuutos on pieni, kuten vaikkapa ihmisen kasvojen ilmeet, muutos ei välttämättä vaikuta visuaalisesti merkittävästi. Joskus malli muuttuu vain muutaman etukäteen tunnetun version välillä, kuten vaikkapa ihmiskasvojen mallin muutamat suun asennot. Tällöin voi olla mahdollista laskea siirtovektorit kullekin versiolle etukäteen ja saada tyydyttävä tulos interpoloimalla näiden välillä.

Kaikista tässä luetelluista ongelmista ja rajoituksista huolimatta SH/PRT-tekniikka on yleisesti käytössä varsinkin peleissä. Useimmat puutteet lähinnä rajoittavat tekniikan käyttöaluetta ja tutkimus on keskittynyt käyttötilanteiden laajentamiseen. Oikeanlaisessa tilanteessa käytettynä tekniikka toimii erinomaisesti.



## 6 Yhteenveto

Tässä artikkelissa esitettiin palloharmoonisten funktioiden (SH) ja esilasketun siirtofunktion (PRT) nimillä tunnetut tekniikat, joita käytetään usein yhdessä. SH/PRT-tekniikka mahdollistaa joidenkin globaalien valaistuksen piirteiden ottamisen mukaan reaaliaikaiseen grafiikkaan ja soveltuu myös muuttuviin valaistusolosuhteisiin. Mikäli globaaliusominaisuuksissa rajoitetaan vain jäykkien kappaleiden itselleen aiheuttamiin valaistusilmiöihin, tekniikkaa voidaan käyttää myös muuttuvalle mallille.

Vaikka SH/PRT-tekniikka saadaankin toimimaan yleisille materiaaleille, se on parhaimmillaan diffuusin valaistuksen laskemisessa. Se ei myöskään tuota realistisia teräviä varjoja eikä aidosti globaalia valaistusta muuttuville malleille. Niinpä SH/PRT on yleensä vain yksi osa valaistuslaskennan kokonaisuutta, jossa sitä käytetään rinnan muiden tekniikoiden kanssa.

Palloharmooniset funktiot muodostavat ortonormaalin kannan. Niitä käytetään valaistuslaskennassa esitettäessä suunnasta riippuvia funktioita. Kun funktio projisoidaan SH-kantaan eli sille etsitään koordinaatit  $n^2$  ensimmäisen kantafunktion suhteen, saadaan sille erittäin tiivis esitys. Nämä  $n^2$  koordinaattia tallennetaan kerroinvektoriksi. Mukaan tulevat vain  $n$  alinta paikkataajuutta, koska ensimmäiset SH-funktiot vastaavat alimpia paikkataajuuksia pallon pinnalla ja kunkin taajuuden esittäminen vaatii sitä enemmän kantafunktioita, mitä korkeammasta taajuudesta on kyse.

Tyypillisiä SH-kannan avulla esitettäviä funktioita ovat esimerkiksi pisteeseen kustakin suunnasta saapuvan valon radianssi ja erilaiset esilasketut siirtofunktiot, jotka kuvaavat, miten pisteen valaistusympäristön perusteella saadaan

pisteestä lähtevä valo. Siirtofunktio lasketaan erikseen kullekin mallin kulmapisteelle. Se sisältää yleensä tietoa kappaletta ympäröivien pintojen vaikutuksesta valon käyttäytymiseen. Tällaisia vaikutuksia ovat kappaleen itsevarjostus ja valon heijastelu kappaleen pinnasta toiseen. SH/PRT-tekniikka perustuu valaistusyhtälön jakamiseen näihin kahteen osaan. Tällöin valaistus voidaan laskea kappalekohtaisesti ajon aikana ja esilaskettu siirtofunktio tallentaa kunkin kulmapisteen yhteyteen.

Erityisen kätevän SH-kannasta tekee sen ortonormaalius, jonka takia hankala, integraalimuotoinen valaistusyhtälö voidaan ratkaista kahdelle SH-kannassa esitetylle funktiolle yksinkertaisesti SH-kerroinvektorien pistetulona. Ajon aikana valaistus lasketaan ja projisoidaan SH-kantaan kullekin kappaleelle aina kuvaa piirrettäessä. Tämän jälkeen valaistus kierretään kappaleen koordinaatistoon hyödyntäen SH-kannan kiertoinvarianttiutta. Kulmapistettä kohti ei jää muuta tehtävää kuin valaistusyhtälön ratkaiseminen laskemalla juuri saadun valaistusvektorin pistetulo kyseisen pisteen esilasketun siirtovektorin kanssa, mikä on erittäin nopeaa.

Yhtenä tekniikan ongelmana on se, että funktioiden muuntaminen SH-kantaan on hidasta. Tämän vuoksi siirtofunktio ei saa muuttua, sillä sen laskeminen uudelleen jokaiselle kulmapisteelle on mahdotonta reaaliajassa. Siirtofunktio puolestaan muuttuu aina kun geometria, jolle se on laskettu, muuttuu. Niinpä on rajoitettava kokonaan staattisiin malleihin tai esittämään vain jäykän kappaleen itsevarjostus. Tästäkin huolimatta SH/PRT-tekniikkaan liittyvä esilaskenta voi kestää edelleen varsin pitkään, mikä osaltaan rajoittaa tekniikan käyttökelpoisuutta. Lisäksi rajoittuminen vain mataliin paikkataajuuks-

siin vähentää tekniikan käyttökelpoisuutta monissa tilanteissa.

Tässä artikkelissa on esitetty palloharmonisten funktioiden ja esilasketun siirtofunktion käyttämiseen tarvittavia perusideoita. Palloharmonisia funktioita käytetään grafiikassa muissakin yhteyksissä, kun suunnasta riippuvia funktioita täytyy esittää ja käsitellä tiiviissä muodossa.

Palloharmoniset funktiot ovat jo nykyään käytössä pelimoottoreissa. Vähitellen tämän kaltaiset peruslähestymistapaa raskaammat ja vaikeammat tekniikat tulevat yleistymään myös muissa reaaliaikapiirtoa käyttävissä sovelluksissa, kuten simulaattoreissa ja arkkitehtuurin havainnollistamisessa. Tätä varten tekniikoista täytyy kuitenkin tehdä yleiskäyttöisiä ja helposti sovellettavia.

## Viitteet

1. Green, R.: *Spherical harmonic lighting: the gritty details*. Game Developers Conference, 2003. Saatavilla <http://www.research.scea.com/gdc2003/spherical-harmonic-lighting.html> (maaliskuu 2011).
2. Kautz, J., Sloan, P., Snyder, J.: Fast, arbitrary BRDF shading for low-frequency lighting using spherical harmonics. *Proceedings of the 13th Eurographics Workshop on Rendering*, ss. 291–296. Eurographics Association Aire-la-Ville, Sveitsi, 2002.
3. Ramamoorthi, S., Hanrahan, P.: An efficient representation for irradiance environment maps. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 01)*, ss. 117–128. ACM New York, USA, 2001.
4. Ramamoorthi, S.: Precomputation-based rendering. *Foundations and Trends in Computer Graphics and Vision*, 3(4), ss. 281–369. Now Publishers Inc, USA, 2007.
5. Sloan, P., Kautz, J., Snyder, J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics – Proceedings of ACM SIGGRAPH 2002*, 21(3), 527–536. ACM New York, USA, 2002.
6. Sloan, P., Hall, J., Hart, J., Snyder, J.: Clustered principal components for precomputed radiance transfer, *ACM Transactions on Graphics – Proceedings of ACM SIGGRAPH 2003*, 22(3), ss. 382–391, 2003.
7. Sloan, P., Luna, B., Snyder, J.: Local, deformable precomputed radiance transfer. *ACM Transactions on Graphics – Proceedings of ACM SIGGRAPH 2005*, 24(3), ss. 1216–1224. ACM New York, USA, 2005.
8. Sloan, P.: Stupid spherical harmonics (SH) tricks. *Game Developers Conference 2008*. Saatavilla <http://www.ppsloan.org/publications/StupidSH36.pdf> (maaliskuu 2011).
9. Zhou, K., Hu, Y., Lin, S., Guo, B., Shum, H.: Precomputed shadow fields for dynamic scenes. *ACM Transactions on Graphics – Proceedings of ACM SIGGRAPH 2005*, 24(3), ss. 1196–1201. ACM New York, USA, 2005.
10. Öhrström, M.: *Spherical harmonics, precomputed radiance transfer and realtime radiosity in computer games*. Diplomityö, Lundin yliopisto, 2003.