Perspektiv och bildförvrängning av olika slag

Bakgrund

Tidiga avbildningar i mänsklighetens historia var platta i betydelsen att allt man ritade/målade av såg ut att vara på samma avstånd från betraktaren. Fram till och med medeltida kyrkomåleri saknade man i västerlandet ett systematiskt sätt att i målningar beskriva att föremål låg på olika avstånd. Att lika stora objekt ser mindre ut om de är avlägsna är förmodligen en stenåldersobservation. Men hur de ska minska i bildstorlek för att vi av en bild ska dra rätt slutsats om deras avstånd var länge okänt. Ett ännu värre problem var hur man skulle få avståndet mellan en rak rad av lyktstolpar utmed en väg som går rakt ut från betraktaren att se konstant ut (i ett tänkt medeltida konstverk med telefonstolpar).

Av uppenbara skäl var de första ansatserna till lösning på problemet gjorda av konstnärer, vilka vägde in de värderingar och känslor man ville förmedla i bilden. Ett stort antal konstnärsperspektiv (från olika åldrar) finns beskrivna:

- Fursteperspektivet där föremål eller främst personer avbildas större ju högre rang eller betydelse de har.
- Polycentriska perspektiv som är det man skulle få om man satte flera fotografier, tagna i olika vinkel bredvid varandra och på olika sätt gjorde övergången mellan dem kontinuerlig. Detta är absolut inte samma sak som om man toge ett foto med en vidvinkellins (om nu någon trodde det) vilket vi kommer att se.
- Grodperspektiv, kavaljersperspektiv och militärperspektiv är andra. Tekniskt används ofta isometriskt perspektiv som bevarar förhållanden mellan längder på ett sätt som är bra i exempelvis ritningar och plattformsspel.

Wikipedia (svenska) har en bra översiktartikel över dessa under rubriken perspektiv, men de är inte temat för denna lilla skrift.

För att ge helt korrekt avståndskänsla i en bild ska den naturligtvis egentligen vara det vi idag kallar 3D, dvs bilden ska ge korrekt stimuli till betraktarens öga med avseende på ackommodation, konvergens och stereoseende. Detta täcks emellertid av en annan föreläsning (3d-seende) och vi använder här en enögd person med oändligt skärpedjup som utgångspunkt för att slippa dessa problem.

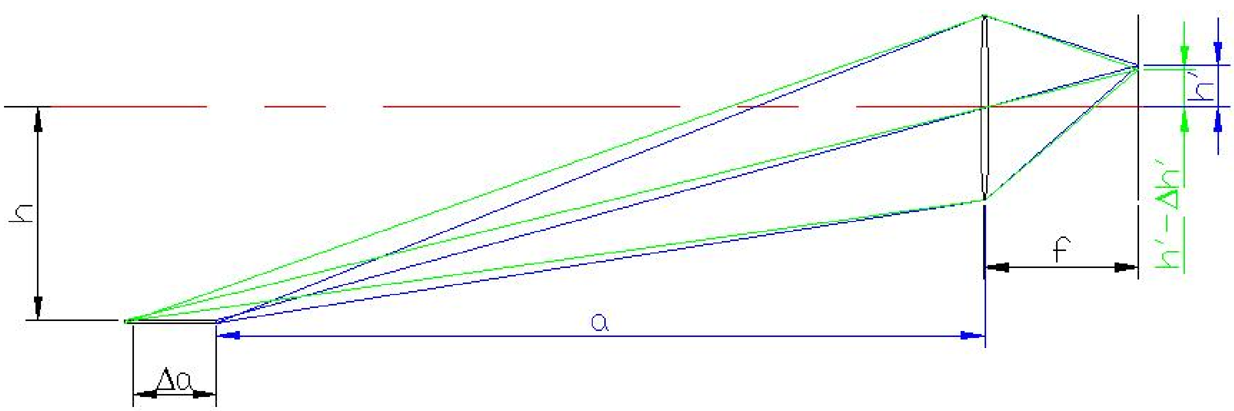
Fram till introduktionen av hålkameran, camera obscura, var fältet öppet för tyckande och känslomässiga tolkningar av begreppet perspektiv, men i och med dennas entré på scenen fick man ett facit på vad som var rätt. Man kalkerade av den bild som projicerades inuti hålkameran och nu hade ju naturen visat hur det skulle se ut. När man sedan kunde förse hålkameran med en lins och dessutom registrera bilden fotografiskt var ju saken klar. Ett kvarstående problem var att avståndet till objekten spelade roll på fler sätt än det uppenbara. En bild tagen med hålkamera på ca 10 m avstånd från en grupp objekt /(som ligger lite olika långt bort exvis 9, 10 och 11 m) är inte bara dubbelt så stor bilden av samma objektgrupp registrerad på 20 m avstånd, och det är bland annat dessa skillnader vi ska försöka reda ut.

Varför känna till perspektiv

Att man behöver känna till grunderna i perspektivläran för att bli en bra (klassisk) konstnär är ganska uppenbart. Förlängningen därifrån till skapandet av animerad film eller scener i dataspel är också lätt att inse. I ett ingenjörsperspektiv är omvändningen också viktig: Hur kan man ur en given vanlig platt bild återskapa så mycket information som möjligt om den tredimensionella förlagan? Matematiken kommer att visa att den bild man får på ett ganska intrikat sätt beror på samband mellan fokallängden på det avbildande systemet och objektsavståndet. I många system kommer sedan aberrationer i optiken att komma in. En persons ögon har ju en given fokallängd på ca 20 – 25 mm och för att återskapa vad denna person ser måste det optiska systemet ien (tänkt) kamera avpassas efter detta.

En pinne på marken

Vi tänker oss situationen att vi betraktar en lång, rak väg och fixerar våra ögon eller mitten på kamerans synfält, mot vägens slut, långt bort. Vi måste då börja med att inse att den punkt vi tittar på är vårt avbildande systems optiska symmetriaxels skärningspunkt med vägslutet. De linjer som utgörs av trottoarkanter, diken och eventuella husfasaders över- och underkant vid sidan av vägen konvergerar mot denna punkt som brukar kallas centralpunkt, oändlighetspunkt eller perspektivpunkt. Linjerna brukar kallas perspektivlinjer. Någon gång på högstadiet brukar man i bildundervisningen upptäcka detta och många inser att de faktiskt kan rita ett hus så att det ser perspektivistiskt korrekt ut.



Men vad händer om vi lägger en pinne, med känd längd Δa på vägen (jfr fig där proportionerna mellan a och f inte är rimliga)? Pinnen börjar på avståndet a från betraktaren och är parallell med vägen. Till att börja med kan vi konstatera att pinnen (om den är rak) syns i bilden och ser ut som en rak pinne (detta är vetenskap på hög nivå). Hölle man däremot pinnen på betraktarens höjd äver marken dvs pinnen ligger utmed den optiska axeln (röd streckad), så syns den inte i bilden. Möjligen syns pinnens ändyta om det inte är en matematiskt oändligt tunn pinne. Frågan är nu hur lång bilden blir av pinnen på chipet inuti kameran om h inte är noll? Om vi använder kamera-approximationen och alltså säger att bildavståndet är f så kommer pinnens närmsta punkt att avbildas på avståndet

$$(1) \quad h' = M_T(a)h = \frac{f}{a}h$$

från bildens mittpunkt där MT är kameran transversella förstoring för objektsavståndet a. Pinnens bortersta punkt ligger också på avståndet h från symmetriaxeln och avbildas på avståndet

$$(2) \quad h' + \Delta h' = M_T(a + \Delta a) = \frac{f}{a + \Delta a}h$$

Subtraherar vi dessa från varandra får vi

$$(3) \quad \Delta h' = -hf\left(\frac{1}{a} - \frac{1}{a + \Delta a}\right) \approx \frac{-hf\Delta a}{a^2}$$

Den matematiskt väl bevandrade inser kanske att vi skulle kunnat differentiera (1) och direkt få (3). Grattis i så fall.

Minustecknet betyder att pinnens bortersta punkt ligger närmre bildens mittpunkt än den närmsta. Vidare ser vi att bildstorleken är proportionell mot fokallängden precis som vid vanliga objekt. En avlägsen pinne kan alltså fås att se större ut om man använder ett teleobjektiv. Det som möjligen är lite överraskande är att pinnens längd minskar med kvadraten på 1/avståndet. En pinne på dubbla avståndet ger alltså en bild som är en fjärdedel så stor. Hade pinnen legat tvärs över vägen hade bildstorleken minskat med 1/avståndet. Detta innebär att förhållandet mellan längd och bredd inte bevaras. Lägger man av outgrundlig anledning ut kvadratiska plattor på vägen i rad bort ifrån betraktaren kommer visserligen bredden på dem att minska när de ligger långt bort med längden kommer att minska snabbare. Avlägsna plattor kommer alltså att se ut som brädor.

Vidare ser vi ur formeln, som väntat, att om h =0 så syns inte pinnen. Praktiskt innebär detta till exempel att om man fotograferar en husrad i stadsmiljö och står intill husen syns avlägnsa hus dåligt. Går man över gatan och fotograferar längs gatan syns avlägsna hus bättre. h har ju blivit större.

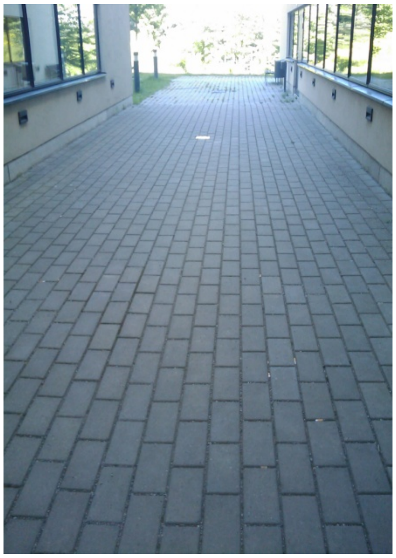
Vad händer i detta avseende om man jämför inzoomning med att gå närmre objektet? Tänk att vi jämför fotografering på 60 m avstånd med fokallängd 300 mm (motsvarar inzoomat läge i ett ganska bra zoomobjektiv) med 10 m avstånd och 50 mm fokallängd (ganska utzoomat). Den vanliga kameraförstoringen blir ju då lika i bägge fallen dvs -1/200. Bilden av upprättstående objekt blir alltså lika stor. Men om vi tittar på den ständigt återkommande pinnen i ovanstående exempel, får vi att förhållandet mellan bildstorlekarna blir med hjälp av ekv (3)

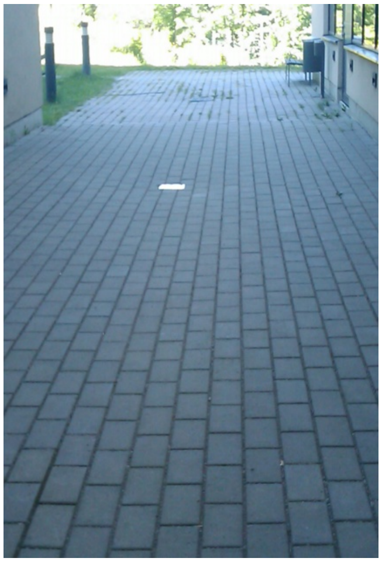
$$(4) \quad \frac{\Delta h'_{inzoomat}}{\Delta h'_{utzoomat}} = \frac{f_{in} a_{ut}^2}{f_{ut} a_{in}^2} = \frac{a_{ut}}{a_{in}} = \frac{1}{6}$$

Den inzoomade bilden har alltså en pinne som bara är en sjättedel så lång. Bortsett från vårt perverterade intresse för pinnar på marken, innebär ju detta att alla sträckor i djupled i bilden blir kortare. Den inzoomade bilden upplevs alltså som platt. Ett vanligt trick vid animeringar är att överdriva djupleds-sträckor eftersom detta av betraktaren tolkas som att man är mycket nära.

Resonemanget om zoomning kan naturligtvis översättas till skillnaden mellan att titta på ett sceneri med kikare jämfört med att minska avståndet en faktor lika med kikarförstoringen.

Jämför gärna bilderna till höger. Den ena är inzoomad 2x den andra inte. Förhållandet mellan längd och bredd på stenplattorna är i verkligheten lika stort överallt





Stora femåringar

En annan skillnad som beror på betraktningsavstånd är relativa tranversella storlekar. Tänk er scenen med en pappa med höjden H och ett barn med höjden h, där barnet befinner sig en sträcka Δa= 3 m framför pappan. Det hela fotograferas mer eller mindre rakt framifrån (inte så att barnet skymmer pappan men nästan) på avståndet a = 3m, från barnet.

$$(5) \quad h'_{barn} = h\frac{f}{a} \quad h'_{pappa} = H\frac{f}{a+\Delta a} \Rightarrow \frac{h'_{barn}}{h'_{pappa}} = \frac{h(a+\Delta a)}{Ha}$$

Om pappan är 1,80m och femåringen 1,20m blir kvoten 1,33, dvs barnet blir 33% större än pappan. Detta upplevs möjligen som självklart (att föremål i förgrunden får överdriven storlek), men även i detta fall finns en skillnad mellan zoomade och icke-zoomade bilder. Om man i stället fotograferar på 18 m avstånd och zoomar så att bilden blir lika stor blir ovanstående kvot 0,77, dvs barnet ser 23% mindre ut än pappan. Detta närmar sig det ”riktiga” värdet på 0,67, men frågan är vilken bild vi upplever som riktigast? Svaret är nog att vår förväntade storlek på bilden av ett objekt är väldigt nära knuten till vår avståndsbedömning så att om Δa/a inte är alltför stor behöver vi ofta mäta med linjal i bilden för att inse att bilden av femåringen är större.

Om det då av övriga detaljer i bilden framgår att det är ett avstånd mellan far och dotter kommer hjärnan att försöka få ihop de motstridiga synintrycken, med resultatet att den inzoomade bilden även i detta avseende upplevs som platt. Detta trots att fokallängden inte ingår i slututtrycket i ekvation 5.

Vänster bild inzoomad och höger bild normal. Bilderna är tagna från samma höjd. Mät med linjal storleken på pallen relativt bordsbenen. Vilket ser mest naturligt ut, vilket är ”rätt” och vilket ger mest djup i bilden?

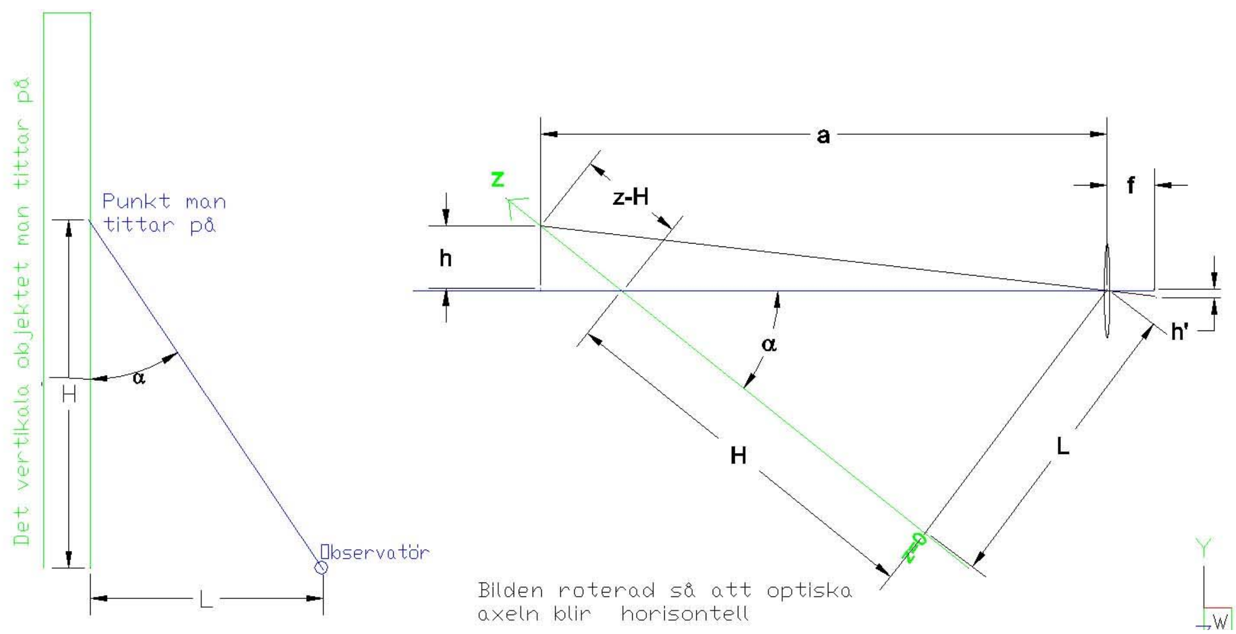
Observera att vid inzoomning händer två saker:

Längsgående sträckor ser kortare ut vid inzoomning och relativa höjdskillnader minskar. Bägge påverkar emellertid synitrycket på samma sätt. Bilden ser plattare ut vid inzoomning.

Grodperspektiv och von Oben-perspektiv

Grodperspektiv innebär att man från marknivå tittar upp mot någonting stort. Von Oben-perspektiv innebär det omvända, dvs att man tittar ner på sitt objekt. Bägge har identisk matematisk beskrivning. Vi väljer att titta på grodperspektivet.

Låt avståndet utefter marken mot det (vertikala) objektet vara L och den punkt mot vilken man tittar vara belägen H över marken. Inför vidare en vertikal z-axel (grön) utefter objektet med nollpunkt vid marken.



Betraktar vi figuren och använder den vanliga förstoringsformlen h'=hf/a för en punkt som inte ligger i bildmitten dvs inte på höjden H över marken, så får vi med

$$(6) \quad a = \frac{L}{\sin\alpha} + (z - H)\cos\alpha$$

$$(7) \quad h = (z - H)\sin\alpha$$

ett uttryck för var bilden i kameran hamnar i förhållande till mittpunkten

$$(8) \quad h' = f\,\frac{(z-H)\sin^2\alpha}{L + (z-H)\sin\alpha\cos\alpha} = \frac{(z-H)L^2}{L\left(H^2 + L^2\right) + (z-H)HL}$$

Det sista ledet fick vi genom att använda

$$\sin\alpha = \frac{L}{\sqrt{L^2 + H^2}} \quad och \quad \cos\alpha = \frac{H}{\sqrt{L^2 + H^2}}$$

Av dessa kan vi nu dra ett antal halvroliga slutsatser:

Perspektivpunktens läge i bildplanet (i eller utanför bilden) fås ur (8) genom att låta z gå mot oändligheten. Detta ger

$$(9)\quad h'_{persppunkt} = \lim_{z\to\infty} \frac{(1-H/z)\sin^2\alpha}{L/z+(1-H/z)\sin\alpha\cos\alpha} f = f\tan\alpha = \frac{fL}{H}$$

Observera alltså att perspektivpunkten inte ligger på symmetriaxeln längre.

Ur (6) får vi att den transversella förstoringen som funktion av z blir

$$(10)\quad M_T = \frac{f\sin\alpha}{L+(z-H)\sin\alpha\cos\alpha}$$

Förstoringen av sträckor utmed det vertikala objektet fås genom att differentiera h' med avseende på z

$$(11)\quad M_{vert} = \frac{dh'}{dz} = (tungt\ arbete) = \frac{fL\sin^2\alpha}{(L+(z-H)\sin\alpha\cos\alpha)^2}$$

Det intressanta för perspektivet är nu kvoten mellan transversell förstoring och det vi kallar vertikal förstoring

$$(12)\quad K = \frac{M_T}{M_{vert}} = \frac{L+(z-H)\sin\alpha\cos\alpha}{L\sin\alpha} = \frac{L^2+zH}{L\sqrt{L^2+H^2}}$$

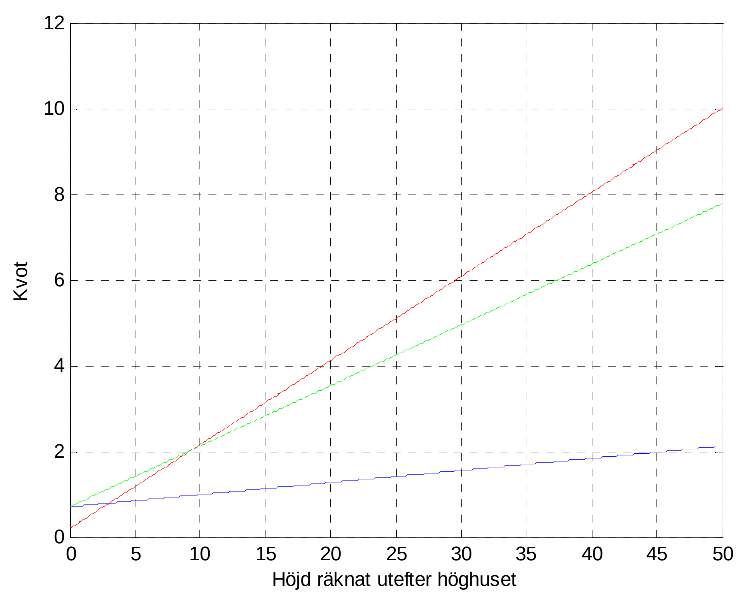
För given kameraplats (ges av L) och given kamerariktning (ges av H) ökar alltså kvoten mellan bredd och höjd med stigande z. Om objektet vore ett höghus, skulle fönstren högt upp se förhållandevis mindre höga ut. Notera att både bredd och höjd minskar men höjd minskar snabbare.

Vi gör ett numeriskt exempel med ett höghus som är 50 m högt.

Blått: L = 25 m , H = 25 m

Rött: L = 5 m , H = 25 m

Grönt: L = 5 m, H = 5 m



Några bilder

Den högra bilden är tagen med en vinkel på ca 45° mot husväggen (motsvarar blå kurva på förra sidan). I den vänstra bilden har jag krupit närmre och har en vinkel på ca 20°. Man ser bl a att förhållandet längd/ bredd ändrar sig mycket mer dramatiskt i den vänstra. Alla fönster är lika stora.

Jag har i bilderna också lagt in det man brukar kalla perspektivlinjer som konvergerar mot oändlighetspunkten, dvs den som ges av ekvation (9).

I det vänstra fallet ska den ligga $f \tan 20°$ från bildmitten och i det högra fallet $f \tan 45°$ vilket ju i alla fall ser rimligt ut.

Att perspektivlinjerna inte följer fönsterkanterna exakt beror inte på byggslarv utan på distorsion i kameran. Jag har använt en enkel mobilkamera. Detta är en aberration.