FORELÆSNINGSNOTER

TIL

INTRODUKTION TIL DIGITAL

BILLEDBEHANDLING

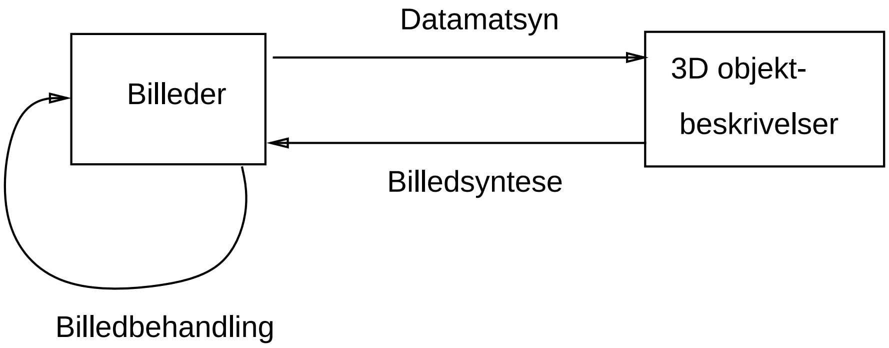
S. I. Olsen

11. oktober 1999

Dette skrift er 6. udkast til et notesæt i digital billedbehandling. Noterne omfatter emner som: filtrering af billeder, detektion af kanter og andre “features” i billeder, segmentering af billeder, kodning af billeder mv. Desuden introduceres i begrænset omfang til genkendelse af mønstre i billeder. Formålet med noterne er at give en praktisk såvel som teoretisk indføring i en række metoder til behandling af digitale billeder. Noterne henvender sig til 2. dels datalogistuderende med naturvidenskabeligt bifag. Det forudsættes at læseren har et modent (operationelt) forhold til matematik.

Notesættet ikke færdigudviklet, men er både ufuldstændigt, mangler mange figurer og billeder, savner referencer til meget vigtig litteratur etc. Desuden må det forventes at noterne indeholder er del fejl, såvel som afsnit der burde udvides eller afkortes. Visse dele af noterne er ikke skrevet endnu (og bliver det ikke i år). Især vil det nok undre at et notesæt i billedbehandling ikke indeholder nogen billeder. Årsagen er *ikke* et pædagogisk princip om at holde undervisningen på et rent teoretisk niveau. Tværdigmod er det forfatterens holdning, at uden en solid praktisk erfaring med anvendelse af metoderne, vil undervisningen være spildt. Derfor er der til hvert kapitel knyttet et antal opgaver, hvoraf mange omhandler eksperimenter med og sammenligning af de beskrevne metoder. Læseren anbefales indtrængende om ikke at stille sig tilfreds med en teoretisk forståelse af en given metode, men derimod eksperimentelt at afprøve metodens styrker og begrænsninger. Til slut kan det nævnes at der eksisterer en lang række lærerbøger i billedbehandling. Et eksempel på en lærerbog, der i noget omfang kan erstatte/supplere disse noter er: *Gonzales, Woods: Digital Image Processing, Addison-Wesley 1992.* Se iøvrigt afsnittet med litteraturreferencer sidst i noterne før indekset.

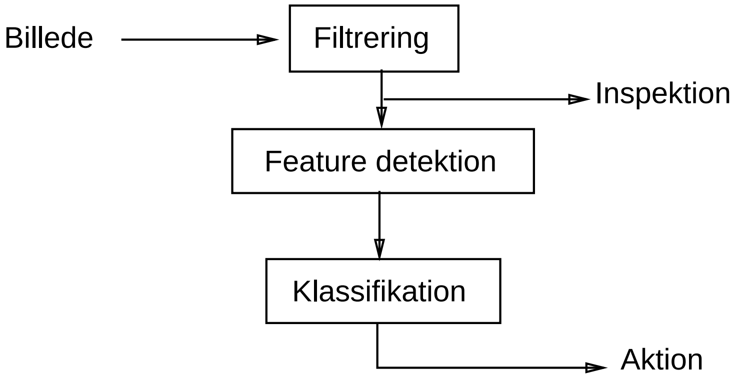
Digital billedbehandling kan defineres ved den process hvor et digitalt billede transformeres til et nyt digitalt billede. I praksis er definitionen for snæver, men rent forståelsesmæssigt velegnet til at skelne digital billedbehandling ud fra de nært beslægtede fagområder: *grafik*, *billedsyntese* og *datamatsyn* . I grafik og billedsyntese er formålet at *fremstille* et billede udfra en specifikation. I datamatsyn er formålet typisk at karakterisere den scene som et givet billede viser. Mange af de teknikker som anvendes inden for de nævnte felter er dog ens.



Digital billedbehandling kan ikke entydigt rubriceres under nogen af de traditionelle dicipliner (matematik, fysik, datalogi e.lign), men er multidiciplinær. Der ingår meget matematik og fysik i billedbehandling, men også fag som psykologi, neurofysiologi, geografi, fotogrammetri osv. Eksempelvis har perceptionspsykologien bibragt stor viden om hvorledes vi, som mennesker, behandler og fortolker de synsindtryk, der "dannes" på nethinderne når vi ser. Denne viden har bl.a. haft betydning for udvikling af metoder til detektion af kanter i billeder, for kodning af billeder etc. Ud fra en datalogisk synsvinkel er behandling og fortolkning af billeder interessant, fordi billeder er et nyt inddatamedie. Billeder indeholder overordentlige store datamængder. For blot 15-20 år siden var det svært at optage billeder fra en datamat. Endnu sværere var det at oplagre og behandle de store datamængder. Disse begrænsninger er i dag i stort omfang elimineret. Datalogens rolle i billedbehandlingen er derfor, dels at udvikle nye metoder, dels at kunne vælge og sammensætte de rigtige metoder til løsning af et givet problem.

Digital billedbehandling anvendes i dag inden for et meget stort antal dicipliner og fagområder. Et klassisk anvendelseseksempel er forbedring af røngtenbilleder. Et andet eksempel er fra astronomien, hvor man typisk ønsker at fjerne den forvrængning af billedet, der opstår når lyset passerer gennem atmosfærden. Et tredie eksempel er analyser af mikroorganismer og deres bevægelsesmønstre. Billedbehandling benyttes også i industriel sammenhæng. Kvalitetskontrol er en typisk anvendelse. Dette kan handle om at kontrollere en svejsesøm; om at måle kvaliteten af et stykke høvlet træ ved at tælle antal knaster i træet, således at kvalitetssorteringen kan foregå automatisk; at måle antallet af bakterier i en vandprøve, at kontrollere antal inklusioner af ædelmetal i en metallegering osv.

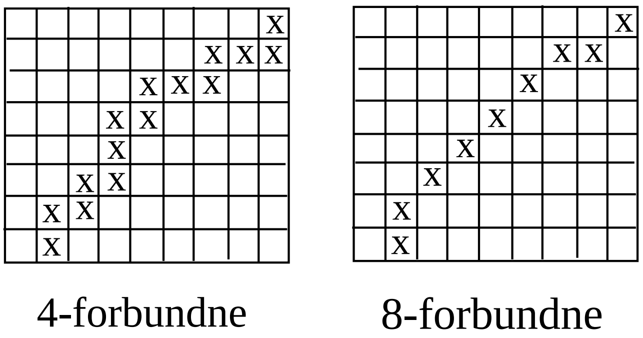
Digital behandling af et billede starter som regel med en *filtrering* af billedet. Formålet med filtreringen er af fjerne eller nedtone uinteressante detaljer samt at fremhæve væsentlige karakteristika. Resultatet af filtreringen kan ved inspektion af et menneske danne udgangspunkt for en beslutning; eller kan danne grundlag for en yderligere datamatisk analyse. Første skridt i en sådan analyse er ofte at foretage en datakompression. Denne kan foretages ved at detektere egenskaber (eng. features) i/ved billedet, samt at knytte mål til de observerede "features". Eksempler på "features" er kanter, herunder rette liniestykker, områder med svagt varierende intensitet, særlige intensitetsmønstre osv. De tilknyttede mål er ofte metriske: størrelse, orientering e.lign. Sidste trin er typisk at drage slutninger ud fra de observerede "features", samt ud fra prædefinerede prototyper af "features" for det, som analysen skal finde eller måle i billedet. Eksempler på sådanne *genkendelse af mønstre* kan være: Findes der en afbildet møtrik i billedet? Hvor i billedet er Sjælland afbildet? Hvor mange forekomster af en given art er synlige i billedet? Det skitserede forløb er naturligvis meget simplificeret, men kan, rent forståelsesmæssigt, anvendes/iagttages ved de fleste anvendelser.



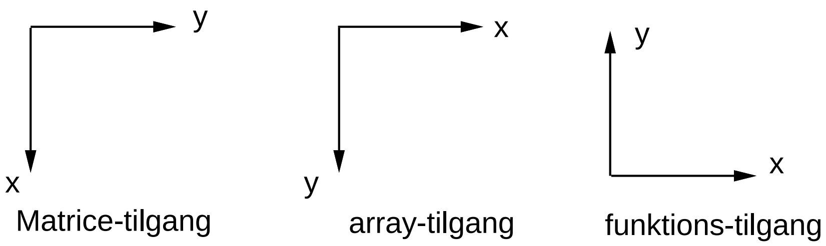
I takt med at prisen på "hardware" falder, og med at ydeevnen stiger bliver det mere attraktivt at benytte billedbehandling, dels i industriel sammenhæng f.eks. i kommunikations- og multimediebranchen, dels rent videnskabeligt ved at undersøge datamaters mulighed for fortolkning af billeder. Både nationalt og internationalt bliver der brugt mange midler til styrkning af forskningen inden for emnet. Både forskningsmæssigt og erhvervsmæssigt er der således god grund til at studere emneområdet.

Hvad er et digitalt billede

Et digitalt billede er en (normalt kvadratisk) opstilling af billedelementer . Hvert billedelement kaldes en *pixel* . Billedelementerne er som regel opstillet således at hver pixel har 4 side-naboer og 4 hjørne-naboer. I dette tilfælde svarer opstillingen til et sædvanligt todimensionalt array. I visse tilfælde opereres med en hexagonal opstilling hvor hver pixel har seks naboer. I disse noter skal vi betragte et N gange M billede som et to-dimensionalt array af dimension $N \times M$. Man siger at billedet har en *opløsning* på $N \times M$. Visse algoritmer (f.eks. til at tegne en ret linie) udnytter et bestemt naboskab. Der tales om *4-forbundne* hhv. *8-forbundne* billedelementer afhængigt af om kun de fire side-naboer eller om alle otte naboer betragtes.



Da det digitale billede kan betragtes som en matrice af tal (et array af tal) er det naturligt at indlægge et sædvanligt retvinklet *koordinatsystem* hvorved billedelementerne kan indiceres. Der er *ikke* nogen fast regel for navngivningen af koordinatakserne, eller for orienteringen af disse. Benyttes matrice-tilgangen er første koordinataksen rettet "nedad" og anden korrdinataksen orienteret "vandret mod højre". Benyttes array-tilgangen er første-aksen rettet "vandret mod højre" og anden-aksen rettet "nedad". Endelig benyttes funktions-tilgangen hvor første-aksen er rettet "vandret mod højre" og andenaksen rettet "opad". Eksempelvis benytter DIKU's **ITI-billedmaskine** matrice-tilgangen, billedvisningsprogrammet **imxv** benytter array-tilgangen og plot-programmet **gnuplot** benytter funktions-tilgangen. I disse noter vil jeg, så konsekvent som det er mig muligt, forsøge at benytte matrice-tilgangen. Første-koordinaten svarer således til rækkenummeret og andenkoordinaten til søjlenummeret.



Hvert billedelement kan rumme et eller flere tal. Hvis tallet tilhører mængden $\{0, 1\}$ kaldes billedet binært . Typisk skal vi i disse noter beskrive operationer på *gråtonebilleder* . De billeddannende sensorer, der oftest benyttes kvantificerer de enkelte billedværdier til 8 pixel, dvs. til værdier mellem 0 og 255 (begge incl.). Kvantificering til 12 bit ses også. For farvebilleder rummer hver pixel tre 8 bit størrelser angivende mængen af rødt, af grønt og af blåt. Multispektrale billeder indeholder mange (fra 5 op til flere hundrede) farvebånd , ofte fra den infrarøde del af det elektromagnetiske spektrum og ikke kun rødt, grønt og blåt. Multispektrale billeder optages ofte af skannere monteret i fly eller satellitter. Fortolkningen af sådanne billeder afhænger naturligvis af et grundigt kendskab til sensoren, der har produceret billedet. Endelig benyttes billeder indeholdende flydende (pseudo-reelle) tal, eller flydende komplekse tal.

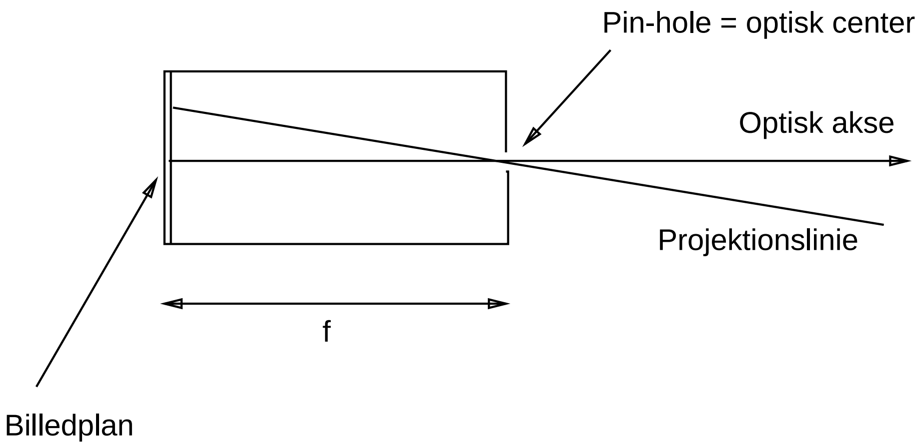
Digitale billeder kan naturligvis have forskellige størrelser. Det mest typiske er at billederne er kvadratiske med en kantlængde der er en potens af 2. De fleste billeddannende sensorer producerer billeder i formatet 512×512 pixel (en anden "standard" er 640×480 pixel). Der findes også sensorer med meget høj opløsning, f.eks. 2048×2048 pixel, men disse er dyre.

I mange situationer er det tilstrækkeligt (og tidsbesparende) at "nedfotografere" billedstørrelsen til 256×256 pixel eller helt ned til 128×128 pixel. Dette afhænger naturligvis af formålet. Til mange eksperimentelle formål er det tilstrækkeligt at benytte små billeder. Et gråtonebillede på 512×512 pixel fylder $\frac{1}{4}$ MB. I dag er dette ikke meget, men for blot 10 år siden var det en stor brøkdel af mængden af indre lager på de fleste PC'er og arbejdsstationer. I en produktionsmæssig sammenhæng, hvor der opereres med tusinder af billeder, er en effektiv metode til kompression af billeder nødvendig. Tilsvarende gør sig gældende for transmission af billeder. Taler vi om sand-tids billedanalyse forstår vi en proces hvor mindst 25 billeder analyseres hvert sekund. Dette svarer til en analysehastighed af 50 megabit per sekund. Dette tal kan sammenlignes med hastigheden på et sædvanligt ether-net, der er i samme størrelsesorden.

Hvordan dannes et digitalt billede

De fleste billeddannende sensorer er følsomme inden for et vist bånd af det elektromagnetiske spektrum. Således er et almindeligt CCD-baseret kamera følsomt over for synligt lys. Andre sensorer , f.eks. radar- og røngtenapparater er mere følsomme over for andre bånd i det elektromagnetiske spektrum. Endelig findes der sensorer, der ikke benytte elektromagnetiske bølger som medium. Eksempler på sådanne sensorer er sonar-skannere, ultralyds-skannere, elektronmikroskoper, og Magnetisk-Resonans-skannere. Det er meget vigtigt at lære det billeddannende instrument godt at kende inden metoder til billedanalyser udvikles. Dette omfatter både sensorens følsomhed og dynamik, samt typen og omfanget af støj i billederne. Endelig kræver det ofte en del oplæring at fortolke billederne korrekt.

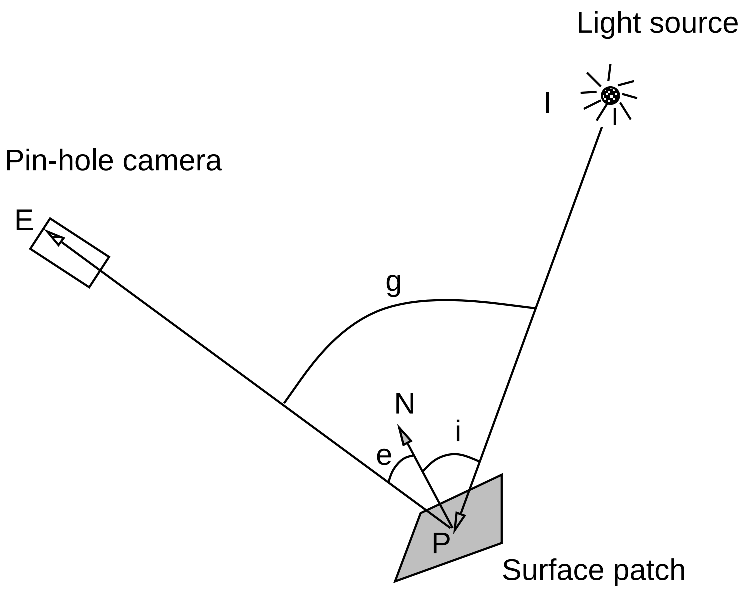
I det følgene skal vi koncentrere os om sensorer, der benytter synligt lys som medie. Som grundlæggende eksempel skal vi betragte et simplificeret kamera, et *pin-hole*-kamera. Dette består af en kasse hvor den ene side udgør billedplanet og der i den modstående side er gjort et punktformigt hul hvorigennem lyset udefra kan trænge ind i kassen.



Gennem hullet (det optiske center) bliver lyset projiceret, og afbildet på billedplanen. Den del af scenen der afbildes er bestemt af brændvidden f . Hvis f er lille vil projektionen eksempelvis være vidvinklet. Da alle projektionslinier går igennem det optiske center er projektionen perspektivisk . Det grundlæggende problem ved et pin-hole kamera er at der kun kan trænge få fotoner ind gennen det lille hul. Gøres hullet større bliver billedet uskarpt, fordi lyset som reflekteres fra et scenepunkt mod hullet “rammer” et helt område af billedplanen. Det kan bemærkes at gamle fotografiapparater faktisk var opbygget efter pin-hole modellen. Problemet med at samle lys nok blev løst dels ved at benytte magnesiumbomber (giver et stærkt lys) og ved at benytte lang eksponeringstid. I dag benyttes et system af linser i stedet. I de datamatiske analyser hvor en kameramodel er påkrævet er pin-hole modellen dog ofte en god approksimation.

Visse billeddannende sensorer skanner billedfeltet. Dette opnås ofte v.h.a. et sæt af roterende spejle. Metoden anvendes ofte i fly- og satellit bårne instrumenter. Billedet opbygges række for række ved at sensoren bevæger sig hen over scenen. Der er derfor kun tale om rækkevis perspektivisk projektion. Man kan også bemærke at der er et vist tidsrum mellem optagelsen af første og sidste pixel. I det følgende skal vi antage pin-hole modellen.

Det lys, der afbildes stammer oprindeligt fra en lysgiver. Sædvanligvis er vi omgivet af mange lysgivere, hvoraf solen ofte er den kraftigste. I mange modeller antages at lyskilden er uendelig fjern og punktformig. Flere og større lyskilder kan derefter fås ved “integration”. Antag derfor at en enkelt uendelig fjern og punktformig lyskilde er givet og betragt et punkt på billedplanet. Til dette svarer en projektionslinie, der skærer scenen i et punkt P. Antag endvidere at vi kan modellere scenefladen i en lille omegn om P ved en plan flade med normal $\mathbf{N}$.



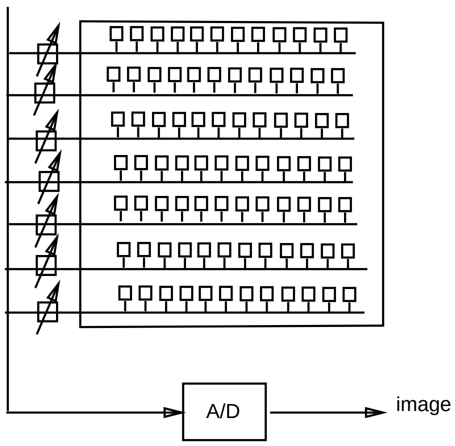
Når lyset rammer den lille fladeflis vil en vis del af energien i lyset blive omsat til varme. Den restende del vil blive reflekteret. Brøkdelen ρ af mængden af reflekteret lys er materialeafhængig og kaldes *albedoen* . Måden hvorpå lyset reflekteres er også materialespecifik.

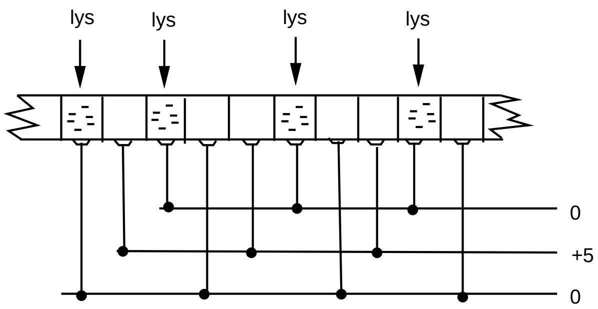
Den bidirektionale *reflektionsfordelingsfunktion* R angiver hvor meget lys der reflekteres i en given retning, når fladeflisen belyses fra en anden given retning. For de fleste materialer (de isotrope) er R en funktion af indfaldsvinklen i (vinklen mellem fladenormalen og en vektor rettet fra P mod lyskilden), af udfaldsvinklen e (defineret tilsvarende) og vinklen g mellem det indfaldende og udfaldende lys.

$$E = I\rho R_M(i,e,g)$$

hvor E er mængden af observeret lys, og I er mængden af lys udsendt af lysgiveren i retning mod P. Funktionen R er naturligvis materialespecifik. For et spejl gælder at $R = 1$ hvis $i = e$ og $g = i + e$. Ellers er $R = 0$. Et andet grænsetilfælde er et perfekt diffust materiale (hvor lyset reflekteres ligeligt i alle retninger). Her er $R = cos(i)$.

I et “rigtigt” kamera er billedplanen en film bestående af et tyndt meget finkornet lag af meget lysfølsomme partikler. Funktionsmåden af et moderne CCD-kamera (CCD = Charge Coupled Device) er helt anderledes. En CCD-chip består primært af et array af små felter, der hver fungerer som ladningsreservoire. Ladningerne kan skiftes ud til en forstærker ved skiftevis at sætte høj spænding på en af tre styreledninger. Den resulterende strøm af ladninger transporteres således (række for række) ud af chippen, forstærkes og omsættes til et digitalt signal v.h.a. en analog-til-digital konvertering .





Det skal bemærkes at der allerede på CCD-chippen er tale om en diskret sampling af fordelingen af lys på chippen. Mægden af lys, dvs. den ladning der opstår når lyset rammet et reservoire er analog (så meget som elektronens ladning er analog). I analog-til-digital konverteringen foretages en egentlig digitalisering af værdierne. Desuden foretages en ny diskretisering. Der er altså **ikke** overensstemmelse mellem reservoire-enheder og pixels. Et eksempel er en CCD-chip med 800 horisontale søjer og 582 rækker af elementer. Disse omsættes, på en for brugeren sædvanlig ukendt måde, til et billede med 512 rækker og 512 søjler. Hvorledes ekstraktionen sker, og hvor den optiske akse skærer billedplanet opgives sædvanligvis ikke af producenten.

At se på billeder

Et væsenligt element inden konstruktionen af et billedbehandlingssystem er at vurdere kvaliteten af og problemerne i de billeder, der skal danne udgangspunkt for analysen. Hvis det er muligt at identificere centrale problemer ved billederne på et tidligt tidspunkt er det lettere at vælge analysemetode og mange problemer, der ellers ville opstå på et senere tidspunkt, kan undgås. Det er håbet at kursusdeltagerne lærer at betragte billeder på en ny kritisk måde, således at de forelagt et billede hurtigt kan vurdere om en given analyse er muligt, samt hvilke metoder, der vil være hensigtsmæssige at anvende.

Når man betragter et digitalt billede er det vigtigt at huske den måde hvorpå billedet er vist. Sædvanligvis er billedet vist på en monitor eller er trykt på papir af en laserprinter. Disse medier er meget forskellige. Ofte er tætheden (pixels/tomme) forskellig, antallet af mulige gråtoner/farver er forskellige og baggrunden er forskellig (sort for en monitor, hvidt for papir). Dette kan give meget forskellige synsoplevelser. Eksempelvis kan mange laserprintere ikke gengive sort korrekt, og antallet af forskellige gråtoner er ofte lille (fx. 32). Mange billigere farveprintere kan kun meget dårligt gengive de farver der kan vises på en farvemonitor (dette skyldes bl.a. at der ofte benyttes “dithering”). Mennesker er fantastisk gode til at skelne farvenuancer, hvorimod vi er dårligere til at skelne gråtonevariationer. Eksempelvis kan gråtonebilleder ofte gengives godt med 64 eller færre forskellige gråtoner (selv om de fleste monitorer samtidigt kan vise 256 forskellige gråtoner). Et “rigtigt” farvebillede indeholder ofter mange millioner forskellige farvenuancer. De fleste farvemonitorer kan imidlertid kun vise 256 samtidige forskellige farvenuancer. Denne forskel er ofte klart synlig. Ved *fuld-farve* (eng. true-color) repræsenteteres hver pixel ved 24 bit. True-color monitorer er dyrere, men kan gengive billeder så godt, at mennesker ikke kan iagttage nogen fejl i farvekvantificeringen. Farver kan også være nyttige ved analyse af gråtonebilleder. Eksempelvis kan identificerede genstande i et gråtonebillede fremhæves ved en klar farve. Ved segmentering af et billede opdeles dette i områder. En gengivning af segmentbilledet ved falske farver (eng. false color) gør det ofte lettere visuelt at vurdere resultatet.

Indhold