

# 1. SLIKE

Digitalizacija nam omogoča obdelavo informacij s pomočjo (digitalnih) računalnikov. Prav tako moramo znati digitalizirane vizualne informacije predstaviti uporabniku na razumljiv način.

## 1.1 Pridobivanje slik

Slika = posnetek svetlobe, ki gre skozi določeno točko v prostoru in pade na sliok ali ravnino senzorja v določenem časovnem intervalu. Da dobimo svetlobo v digitalni obliki, jo moramo digitalizirati (pretvoriti v matriko vrednosti, ki značuje, koliko svetlobe je padlo na določeno območje). Te informacije lahko pokažemo v matrični obliki ali pa kot digitalno sliko.

## 1.2 Človeško zaznavanje svetlobe

Svetloba je elektromagnetno valovanje (fotoni). Različne valovne dolžine predstavljajo različne barve (laser = svetloba z eno samo valovno dolžino). Vidimo valove v razponu od 400 do 700nm (infrardeča -> ultravijola).

Naše oči sestavljajo leča in mrežnica, ki je sestavljena iz dveh vrst celic, ki sta občutljivi na različne valovne dolžine svetlobe:

- **čepki:** večja jakost, občutljivost na barve
- **paličice:** nizka intenzivnost, barv ne zaznavajo dobro; v očesu imamo več paličic

Imenujemo se trikromatični, ker so čepki občutljivi na 3 različne valovne dolžine spektrov - rdeča, zelena, modra. Oko je najbolj občutljivo v sredini svetlobnega spektra, porazdelitev paličastih celic je 40:20:1 (RGB).

Krivulje občutljivosti (za občutljivost čepkov - odziv vsakega nevrona čepka) je enaka integralu produkta svetlobnega spektra in krivulje občutljivosti. Za doseganje enakega zaznavanja svetlobe moramo stimulirati čepke, da se odzivi ujemajo. Ta učinek se imenuje **metamerizem** - dva vira svetlobe z različnima spektroma zaznamo kot enako barvo.

Za simulacijo barve potrebujemo simulacijo receptorjev čepkov: ustvarimo svetlobo, ki ustreza njihovem intervalu visoke občutljivosti. To svetlobo imenujemo barvni primarij in se razlikuje v različnih standardih barvnih sistemov - različne kombinacije vrednosti v različnih standardih lahko zaznavamo kot isto barvo. Za natančno barvno reprodukcijo je potrebna dodatna kalibracija.

**Tristimulusni kolorimeter** je poskus za ugotavljanje razmerja med osnovnimi barvami in zaznano barvo. Oseba mora reproducirati referenčno barvo z nadzorovanjem intenzivnosti treh osnovnih barv.

Nekaterih barv ni mogoče reproducirati, zato je treba od referenčne svetlobe odšteti svetlobo, ki jo imenujemo negativna svetloba. Standardni opazovalec

določa, kakšen je kot opazovanja, tj. široko uporabljen standardni opazovalec CIE 1931 2<sup>ř</sup> določa kot opazovanja največ dve stopinji.

### 1.3 Digitalna kamera

Digitalna slika je pravilna dvodimenzionalna matrika vrednost (pikslov). Tehnično je zahtevno izmeriti tri barvne komponente na istem mestu, zato so bile izumljene številne postavitve senzorjev, ki to simulirajo z naknadno obdelavo. Najbolj znan je Bayerjev vzorec, ki ima dvakrat več pikslov zelene barve (barva + svetlost) kot rdeče in modre (kromatske komponente).

Barvni prostor CIE XYZ se ne uporablja (stroški, tehnične težave). Različni standardi se uporabljajo za različne namene. Glede na uporabljeni medij razvrščamo barvne modele:

- **Aditivni modeli:** začnejo se s črno barvo, barve se dodajajo z mešanjem primarnih barv; monitorji, TV, projektorji; RGB
- **Subtraktivni modeli:** začnejo se z belo barvo, dodajanje komponent - pigmenti; tiskanje, analogne slike, barvice; CMY ali CMYK (K pomeni črn pigment)

RGB je najbolj znan, ker se uporablja v strojni opremi za prikazovanje (temelj katodnih TV); tudi različni standardi RGB.

Pokritost barvnega prostora CIE XYZ se razlikuje med različnimi RGB in CMYK (pri pretvorbi lahko pride do izgub). Poznamo tudi druge barvne prostore:

- barvni prostor CIE Lab posnema zaznavanje barv (podobne barve so si v barvnem prostoru blizu)
- barvni prostor HSV ima bolj psihološko motivacijo - barvo ločuje na oddtenke, nasičenost in vrednost

### 1.4 Obdelava slik

Osnovni koncept pri obdelavi slik je digitalna slika (2D ali 3D matrika - prvi dve dimenziji sta višina in širina, treja (kanali) označuje posamezne barvne komponente). Drug pomemben vidik je ločljivost vzorčenja - št. barv, ki jih lahko predstavimo z eno vrednostjo; najpogostejše 8-bitna slika (vrednosti od 0 do 255).

Drugi pogled na sliko je pogled na obdelavo signalov: slika = diskretna funkcija iz 2D v 1 ali 3 dimenzionalen prostor.

Najpogostejši razred operacij je operacija na pikslu:

- **oddtenki sivine:** enokanalna slika, vrednost dobimo s povprečenjem treh kanalov -> svetlost, to velja za RGB, ne pa tudi za HVS (svetlost je tu že ločena)
- **inverzija:** vsako vrednost piksla odštejemo od največje možne vrednosti piksla - 255

- **pragiranje (thresholding)**: vsak piksel primerjamo z pragom
- **sprememba svetlosti**: svetlost = intenzivnost piksla glede na drug piksel; spreminjanje pomeni premikanje vrednosti vseh pikslov naenkrat
- **sprememba kontrasta**: kontrast = razlika med največjim in najmanjšim pikslom; spreminjanje pomeni pomikanje vseh vrednosti z danim faktorjem
- **nelinearno preslikavanje**: npr spreminjanje kontrasta slike

## 1.5 Histogram

Histogram je struktura, ki opisuje porazdelitev vrednosti - kolikokrat se določena vrednost pojavi na sliki; opišemo vsebino slike neglede na položaj posameznih slik.

Povedo nam o kakovosti slike - če je slika slaba, so vrednosti samo na nekem intervalu -> to prilagodimo z **raztezanjem histograma** (operacije na piksljih). Kompleksnejša tehnika za prilagajanje kontrasta je **izravnava histograma** - nelinearno želimo porazdeliti vrednosti v sliki, da bo histogram enakomeren. S histogrami lahko delamo le, če sliko prej pretvorimo v to primeren barvni prostor - pretvarjamo svetlost. Poznamo jih več:

- 3 histogrami: vsak kanal obravnavan ločeno, skupne informacije se ne ohranijo
- 3D histogrami: vsak kanal je dimenzija v 3D indeksu, ohranijo se skupne informacije

## 1.6 Filtriranje slik

Filtriranje je razred operacij, kjer je rezultat posameznega piksla odvisen od njegove okolice. Poznamo dve vrsti filtrov:

- linearni filtri: operacija nad piksli je linearna (obtežena vsota)
- nelinearni filtri: poljubna nelinearna operacija nad piksli (npr. min, max, mediana)

Linearne filtre interpretiramo kot obteženo vsoto. Jedro je operacija, ki določa utež v matriki. Ustrezni piksli se pomnožijo skupaj in seštejemo števila, da dobimo odziv filtra na piksel. Primeri jeder:

- enaka jedra: povprečenje, vse uteži so enake
- identitetno jedro: kopira samo izvorno središčno vrednost, slika se ohrani
- jedro za premik: sliko premakne v eno smer
- Gaussovo jedro: glajenje brez artefaktov. nizkoprepustni filter

Pomembna lastnost jeder je, da je vsota vseh njihovih elementov enaka 1, v nasprotnem primeru bo rezultat povečan oz. pomanjšan.

Kaj lahko storimo, ko pridemo do meje - obstajajo strategije:

- obrezovanje meje: nastala slika je manjša od izvirne

- konstantna vrednost: manjkajoče vrednosti nadomestimo s konstanto -> učinek vinjete
- vrednost robov: kopiranje vrednosti robov
- izkrivljanje: kopiranje vrednosti z druge strani slike
- odsev: zrcaljenje vrednosti slike

Druge operacije, ki jih lahko izvedemo z linearni filtri:

- zaznavanje robov
- izostritev: visoke frekvence lahko okrepimo najprej z odstranjevanjem nizkih frekvenc (zamegljena slika), nato pa združimo nizke in visoke frekvence (obtežena)

Najbolj znani nelinearni filtri:

- mediana: vzame srednjo vrednost v okolici
- max: vzame največjo vrednost v okolici
- min: vzame najmanjšo vrednost v okolici
- dvostranski filter: obtežena vsota, utež se izračuna za vsako soseso na podlagi podobnosti intenzivnosti s središčnim pikslom; ohrani robove <- vrednosti blizu skupaj se zgladijo

Preprost primer uporabe slikovnih filtrov je odstranjevanje šuma. Šum je naključni signal, ki se doda slikovnim informacijam. Dva najpogostejša profila šuma sta Gaussov šum (vrednosti na piksel se vzorčijo iz Gaussovega šuma z ničelno sredino, rezultat je dodajanje visokofrekvenčnih signalov) in šum salt-and-pepper (uporabimo mediana filter - velika odstopanja v okolici se izločijo). Gaussov filter je nizkoprepustni filter, ki odstrani visokofrekvenčne frekvence iz signala; lahko povzroči zamegljeno sliko -> v takem primeru uporabimo dvostranski filter, saj ohranja robove.

## 1.7 Transformacija slike

Filtriranje in operacije po pikslah so operacije intenzivnosti, geometrijske informacije spreminjajo samo geometrijo slike. So parametrične: transformirajo vse lokacije pikslov na enak način; lahko so tudi neparametrične: transformacije ne moremo opisati z param. trans. funkcijo, le kot preslikava med pikslji.

Parametrične transformacije so lahko linearne ali nelinearne. Osnovne linearne transformacije:

- translacija
- rotacija
- skaliranje
- evklidsko: translacija + rotacija
- podobnost: translacija + rotacija + enakomerno skaliranje
- afine: translacija, rotacija, skaliranje striženje (ohranja vzporedne črte)
- projektivna / homografija: projicira točke iz ene 3D površine na drugo

Linearne trans. so določene z matriko: 2D položaj je določen z 3D vektorjem, kjer je zadnja koordinata vedno 1. Izkrivljanje je postopek uporabe geometrijske transformacije na digitalni sliki.

Naivni postopek je preprost: za vsak piksel v izvorni sliki se izračuna lokacija v transformirani na podlagi transformacije in vanjo kopira vrednost. Težava nastane, da obiščemo vse piksele v izvorni sliki, ne pa tudi v transformacijski: ene vrednosti se ne prepišejo, druge pa so lahko večkrat prepisane. Postopek moramo obrniti, izračunamo inverzno transformacijo za vse piksele v ciljni sliki na ustrezne izvirne piksele.

Problem je tudi, da lokacija, s katere vzamemo barvo na izvorni sliki, ni koordinata samo enega piksla; to rešimo z uporabo najbližjega piksla.

Učinki interpolacije so najbolj vidni, ko spreminjamo velikost slike (jo povečujemo). Če sliko povečujemo, je potreba po interpolaciji jasna, saj želimo več podatkov, kot jih je na voljo. Boljše strategije vključujejo interpolacijo, kjer se barva piksla določi na podlagi sosednjih pikslov. Pri strategiji najbližjega sosedja se uporabi vrednost najbližjega sosedja. Uporabimo lahko linearno interpolacijo, ki prilega vrednost obema sosedoma. Kubična interpolacija deluje podobno, vendar uporablja štiri sosede, ki se prilegajo tretji stopnji polinoma.

Lanczosovo vzorčenje se uporablja pri spreminjanju velikosti ali vrtenju slike (gladka interpolacija, izognemo se aliasingu). Izvede se konvolucija signala z Lanczosovim jedrom. To vzorčenje je tudi najpočasnejše med omenjenimi.

Pri decimiranju (zmanjševanju slike) moramo paziti na podrobnosti, ki bi se ob uporabi naivnega pristopa (najbližji sosed) izgubile (prenizka frekvenca vzorčenja - nek teorem o vzorčenju). Najprej moramo odtisrati visoke frekvence in nato uporabiti vzorčenje po najbližjem sosedu.

Nelinearne transformacije so transformacije, pri katerih preslikava ni določena z linearno funkcijo, npr. odstranitev učinka popačenja objektiva, uporaba lokalne transformacije za izvedbo morfinga slike. Morfiranje je postopek postopnega spreminjanja ene slike v drugo (najlažje s uteženim povprečjem ustreznih pikslov s postopno spreminjajočimi se utežmi; boljši učinek dobimo, če postopoma geometrijsko preoblikujemo eno sliko v drugo).

Algoritem za morfiranje slike je naslednji:

- vhod: dve sliki, A in B, korespondenčni pari
- ponovitev za časovne korake  $t = 0 \dots 1$
- interpoliranje korespondence na položaj  $t$
- deformacija slike A od 0 do  $t$  in B od 1 do  $t$
- mešanje ukrivljene slike s faktorjem  $t$

## 1.8 Spreminjanje velikost slik glede na vsebino

V nekaterih primerih se želimo izogniti linearni transformaciji za spreminjanje velikosti slike, saj želimo ohraniti območja, ki so semantično pomembnejša. Spreminjanje velikosti glede na vsebino bo ohranilo pomembne strukture in zmanjšala nerealne artefakte. Ne moremo objektivno določiti, kaj je pomembno.

Nekateri hitri približki pomembnosti so robovi, hitre spremembe intenzivnosti slike, ... Z uporabo izpeljank prvega reda lahko ocenimo pomembnost vsakega piksla.

Da bi ohranili pomembnosti, moramo uporabiti algoritem - izrezovanje šivov. Temelji na dinamičnem programiranju za določitev poti pikslov iz ene strani slike na drugo. Pot se določi tako, da se za vsak končni piskel nenasitno izračuna najmanjša možna pot, nato pa od piksla z najmanjšo kumulativno energijo nazaj do vira.

## 1.9 Združevanje slik

Pogosto uporabimo operacijo združevanja za vse barvne kanale posebej. Najpogostejša oblika je združevanje z binarno masko. Imamo dve sliki in masko, ki so za vse enako velike. Maska vsebuje samo ničelne ali neničelne vrednosti (katera slika pripada kateremu območju). Velikokrat dobimo nerealne rezultate - težko določimo jasno pikselno mejo med objekti. Mešanje alfa je razširitev binarne maske, uporablja se maska alfa: vrednosti med 0 in 1 v maski alfa pomenijo stopnjo vrednosti pikslov iz prve in druge slike. Uporabno pri združevanju slik z enako vsebino vendar različnim kontrastom (panorama). Ustvari gladke prehode, vendar pregledki prehodi povzročijo učinek duhov.

Nizke frekvence združujemo z gladko masko, visoke pa z ostrejšimi prehodi. Tako deluje frekvenčno mešanje: sliko razgradi na več frekvenčnih pasov in vsakega od njih zmeša z lastno masko alfa, nato se pasovi združijo v končno sliko.

Za razgradnjo slike na pasove uporabljamo slikovne piramide.

Gaussova piramida se uporablja za razgradnjo slike v plasti, pri čemer vsaka višja plast vključuje samo nižje frekvence, kjer so določeni frekvenčni pasovi odstranjeni. To dosežemo z konvolvijo slik nižjih plasti s Gaussovim filtrom.

Druga piramida je Laplaceova piramida. To piramido dobimo iz Gaussove piramide: odštejemo naslednje plasti, razen zadnje plasti, ki je enaka kot v Gaussovi piramidi. To pomeni, da bo vsaka plast vsebovala le frekvence, ki so vključene v spodnjo plast in ne v plast nad njo. Iz Laplaceove piramide lahko rekonstruiramo izvirno sliko, če seštejemo vse plasti skupaj.

Algoritem mešanja, ki upošteva frekvenco:

- vhodni podatki: dve sliki, binarna maska
- za vse barvne kanale naredimo korake:
  - izdelamo Laplaceovo piramido za obe sliki in Gaussovo piramido za oba kanala
  - združimo ustrezne plasti slikovnih piramid s plastmi iz piramide maske
  - združimo nastalo združeno Laplaceovo piramido v združeni slikovni kanal

## 1.10 Segmentacija slike

Pri združevanju predpostavljamo, da je maska vnaprej znana. Včasih temu ni tako, zato se uporablja polavtomatski pristop, ki zahteva grob vnos in ustvari podrobno masko želenega predmeta - **GrabCut**.

Temelji na ugotovitvi, da so segmentacijske oznake zelo strukturirane (dva piksla, ki sta si podobna po barvi in / ali sta si blizu, si bosta verjetno delila tudi segmentacijske oznake). Problem segmentacije je formaliziran z uporabo Markovega naključnega polja. Uporablja Gaussov model mešanice za modeliranje posplošene barvne verjetnosti in Graph Cut za izboljšanje segmentacije v iteracijah. Začetno segmentacijo mora ugotoviti zunanji vir.

## 2. VIDEO

Digitalni video je zaporedje digitalnih slik, ki se prikažejo ena za drugo z zadostno hitrostjo, da se ustvari iluzija gibanja. Pogosto ga spremlja zvok, zato je video digitalni medij za snemanje, kopiranje, predvajanje, prenašanje in prikazovanje gibljivih vizualnih medijev.

### 2.1 Pridobivanje in reprodukcija

Človek lahko zazna v eni sekundi približno 10 do 20 slik kot posamezne slike, pri frekvencah nad 20fps pa so te slike združene v gibanje. Temu pravimo obstojnost vida.

Prve tehnologije video produkcije so bili filmski projektorji (osvetljevanje slik na filmskem kolutu, zapiranje zaklopa). Prvi so prikazovali približno 25 sličic na sekundo (problem, ker so bile hitre spremembe med svetlo in temno sliko -> svetlejšje kot so slike, krajši je interval za učinek vztrajnosti vida); povečali so frekvenco tako, da so večkrat prikazovali eno sliko.

Druga znana tehnologija je katodna televizija - elektronski snopi letijo v vakuumski cevi in udarjajo v fosforjeve delce na drugi strani, ki ob trku zažarijo. Pri barvnih televizijah so skupine treh vrst fosforja (rdeča, modra in zelena). Hirost osveževanja je 50Hz ali 60Hz, posamezniki tudi pri tej hitrosti zaznajo utripanje, zato so sistemi tudi z 75Hz ali 100Hz (računalniški monitorji). Razlog za utripanje: fosfor hitro začne izgubljati svojo svetlost, preden se ponovno zbudi.

Katodni zasloni so zastareli, nadomestili so jih tehnologije ploščatih zaslonov (predvsem LCD) - osvetlitev ozadja je blokirana ali pa prehaja skozi določene piksele iz tekočih kristalov, nato se filtrira skozi barvne filtre -> barvni piksli.

### 2.2 Prepletanje videoposnetkov

Prepletanje pri videu pomeni, da se lihe in sode vrstice v kadru snemajo z majhnim časovnim zamikom v različnih časovnih obdobjih. Pristop je bil uporabljen za podvojitve zaznave hitrosti sličic na račun pasovne širine. Slabost prepletanje

je, da v bistvu polovica podatkov manjka, kar povzroča artefakte (učinek česanja), npr. hitro ravno gibanje na posnetku. Nasprotna možnost je bila uporaba prograssivnega kodiranja, kjer je bila vsaka sličica zapisana v polni ločljivosti -> podvojitev velikosti videoposnetkov; počasi se opušča.

## 2.3 Barva v videu

V analognem videu kodiranje barvnih informacij na dva načina. Pri komponentnem videu se je vsak barvni kanal prenašal ločeno po linijah z ločenimi signali -> zmanjša navzkrižno govorjenje kanalov -> boljša kakovost slike, vendar tudi večja pasovna širina.

Pri kompozitnem videu je barva kodirana kot ločena komponenta, druga komponenta pa je še svetlost, vse komponente so pomešane v skupnem signalu. Tak pristop je primeren za manjše pasovne širine.

Pri digitalnem videu se ločevanje barvnih informacij uporablja tudi za ohranjanje pasovne širine ali prostora. Taki formati so bili poznani že pri analognem videu (YUV, YIQ), danes se pri digitalnem videu uporablja format YCrCb. Glavni razlog za zmanjšanje bitne hitrosti je možnost podvzorčenja barvnih kanalov brez opaznega poslabšanja kakovosti slike. To je posledica človeške narave - manj smo občutljivi na spremembe v barvi kot pri spremembah svetilnosti. Standardni zapis vrste podvzorčenja je J:A:B :

- J - horizontalna frekvenca vzorčenja, število vrednosti svetilnosti v vrstici
- A - št. vrednosti barv v lihih vrsticah
- B - št. vrednosti barv v sodih vrsticah

## 2.4 Televizijski formati

PAL v Evropi, NTSC v Severni Ameriki in SECAM v Rusiji, podobni so po ločljivosti in značilnostih. PAL je bil razvit za reševanje težav, ki bi jih imel NTSC v Evropi zaradi težje dostopnega terena in nepredvidljivega vremena. Danes večino držav uporablja digitalno televizijo: neposredna manipulacija, možnost vključitve videa v aplikacije, lažji dostop, boljše odpravljanje napak, brez poslabšanja kakovosti pri kopiranju; več standardov: DVB-t, ATSC, ...

## 2.5 3D videoposnetki

Korenine ima v stereoskopski fotografiji. Cilj je ustvariti iluzijo globine, tako da se prikažeta dve različni podobi istega prizora (sliki posneti v različnih položajih, oddaljeni med 50 in 75mm).

Različne naprave za prikazovanje 3D vide, ki jih je mogoče nositi: binokularni HDM (dva zaslona, nameščena na glavo). Drugi pristopi temeljijo na spremenjeni različici videoposnetka, ki vsebujejo obe sliki, in nosljivega filtra, ki ločuje slike in ju predstavi ustreznemu očesu. Najpreprostejša tehnologija je anaglif - slike so kodirane s komplementarnimi barvami, očala imajo leče, ki filtrirajo te barve.



Dražja možnost je uporaba polarizirane svetlobe in leč, ki prepuščajo le svetlobo navpične ali vodoravne polarizacije. Očala z aktivnim zaslonom so sinhronizirana s projektorjem, zaprejo posamezno oko.

Težave: tehnološke (ločljivost in kader hitrosti, zmanjšanje navzkrižnih zvokov), povečani stroški, zdravstvena vprašanja (slabost zaradi gibanja, glavobol, slabost, dezorinetiranost).

## 2.6 Večsmerni posnetek

Znan tudi pod imenom 360-stopinjski video. Sneman z več kamerami ali ogledali, posnete slike se projicira na skupno slikovno ravnino:

- ploska: 180 stopinjske projekcije
- ekvirektna: učinek “majhnega planeta”
- Cubemap: za preslikavo tekstur, boljši izkoristek slikovnih pik kot pri enakopravni projekciji

Lahkjo jih pretvorimo v običajne videoposnetke pod kotom kamere z uporabo projektivne geometrije. Omejitvi sta ločljivost in zorni kot kamere (ne sme biti preširok). Uporabno pri zabavi, športu in viralnem turizmu.

## 2.7 Video stabilizacija

Veliko posnetkov je posnetih z nezaželenim gibanjem kamere (tresenje) ali pa ne moremo gladko slediti objektu.

Mehanska stabilizacija: stabilizacija med zajemanjem videoposnetkov.

Optična stabilizacija: npr. kamera in objektiv, ki zazna nezaželene tresljaje in se premakne, da bi jih nevtralizirala.

Poznamo tudi posebno opremo (npr. Steadicam) ali pa stativi.

Digitalna stabilizacija: video stabiliziramo potem ko je že posnet. Poznamo dve vrsti: objekto ali lokalno stabilizaijo (kamera poskrbi, da ročno izbrani objekt ali točka v videu miruje) in globalno stabilizacijo (celotno gibanje kamere je gladko).

### 2.7.1 Stabilizacija s poravnavo

Najpogostejši pristop h globalni stabilizaciji je usklajevanje posameznih okvirjev in premikanje enega od njih, da se čimbolj ujema z drugim. Ne deluje, če se prizor preveč spreminja ali gibanje ni translacijsko.

### 2.7.2 Stabilizacija s funkcijami sledenja

Uporabimo določeno točko (npr. neka značilnost) in jo spremljamo na več posnetkih. Sprememba položaja točke se uporabi za izničenje translacije. Sledilnik značilnih točk uporablja korelacijo za iskanje najboljšega ujemanja za območje okoli značilne točke. Pristop je odporen na spremembe osvetlitve. Točke ne moremo postaviti na npr. vogale, značilnost mora biti vidna ves čas trajanja

videoposnetka; pomaga tudi lokacija gibanja, če je v novem kadru zaznanih več kandidatov. Lahko dodamo več točk -> kompleksnejše.

### **2.7.3 Stabilizacija s ključnimi točkami**

Obstaja samodejno zaznavanje ključnih točk; več algoritmov za njihovo zaznavanje. Najbolj znani so SIFT (Scale Invariant Feature Transform). Pristop ne zagotavlja, da se bodo vse točke pravilno ujemale, zato uporabimo robusten algoritem za prilagajanje, kot je RANSAC.

### **2.7.4 Stabilizacija s optičnim tokom**

Algoritmi optičnega toka ocenjujejo gibanje gostih slikovnih pik med dvema zaporednima sličicama. Ne moremo ga oceniti za vse piksele, vendar je gostejše vektorsko polje kot ključne točke. Gostota je odvisna od uporabljenega algoritma (Lucas in Kanade algoritem: hiter, ne zagotovi globalno optimalne rešitve; Horn in Schunk algoritem: počasen zagotovi omejitve).

### **2.7.5 Stabilizacija v prostoru**

Prejšnji algoritmi stabilizacije se dejansko ne zavedajo položaja kamere in ga zato ne morejo neposredno izboljšati.

TODO

### **2.7.6 Manjkajoče regije in mozaičenje**

Nastanejo manjkajoče meje/regije. Če te manjkajoče regije niso prevelike, lahko posnetek obrežemo. Uporablja se tudi trajektorno glajenje (nizkoprepustni filter, omogoči počasno premikanje vendar zatire visokofrekvenčno drseče gibanje, kar je najbolj problematični del gibanja kamere). V nekaterih primerih ima glajenje prednost pred popolno stabilizacijo, ker je videti bolj realistično.

Najprednejša je možnost uporabe mozaičenja: zapolnimo manjkajoči prostor na podlagi informacij iz prejšnjih posnetkov. Deluje podobno kot seštevane panorame. Vsebina videoposnetka se ne sme spreminjati. Naprednejši pristop uporablja tudi optični tok.

### **2.7.7 Lokalna stabilizacija**

Ustvari se izrez enega predmeta na videoposnetku, pridobimo sled predmeta. Sledenje predmetu deluje na principu pomnjenja videza predmeta v začetnem posnetku. Če se predmet približa robu, imamo spet težave. Sled lahko tudi zgladimo.

## **2.8 Zaznavanje prehodov**

Video: več posnetkov, ki tvorijo več prizorov.

Vrste primitivnih prehodov:

- rez: jasen rez med dvema kadroma
- fade-out: postopno prelivanje v barvo
- prelivanje: postopen prehod iz barve
- razpuščanje: alfa prehod iz enega posnetka v drugega
- brisanje: postopna prostorska zamenjava s tekočo binarno masko

Zaznavanje prehodov je odvisno od vrste prehoda, ponavadi zadostuje opazovanje samo dveh sličic. Na splošno prehodi povzročijo hitro spremembo porazdelitve (barva, robovi, ...).

Preprostejša oblika zaznavanja je primerjava barvnih histogramov zaporednih posnetkov. Zaznavanje z prilagodljivim pragom pomeni, da opazujemo razlike med več posnetki pred in za trenutnim kadrom, prag nastavimo na višjo od obeh vrednosti. V primeru postopnih sprememb lahko uporabimo povprečenje značilnosti.

Druga možnost je pregledovanje delnih sprememb, tj. samo na delu slike. Sliko razdelimo na mrežo, v vsaki celici preverimo spremembe.

Poleg barve se uporabljajo tudi robovi - roboveprimerjamo prosotrsko, koliko robnih pikslov je skupno obema posnetkoma in koliko ne.

## 3. ZVOK

### 3.1 Pridobivanje zvoka

Je fizikalni pojav, ki ga povzroča valovanje tlaka v mediju. Valovi so vzdolžni valovi. Tipična medija za zvok sta zrak in voda, brez njiju zvoka ne bomo zaznali. Učinki, katerim je podvržen:

- odboj: zvok zadane ob površino in se odbije nazaj -> povzroči odmev
- lom: zvok vstopi v drug medij, kot se lahko spremeni
- difrakcija: zvok se upogne okrog ovire

Za snemanje zvoka moramo poznati lastnosti in način merjenja. Glavni lastnosti sta frekvenca (št. valov v časovni enoti; Hz) in amplituda (velikost spremembe tlaka; W/m<sup>2</sup>). Opazovani zvok ima tudi trajanje, smer in hitrost valovanja, ki temelji na podlagi medijea. Zvok skozi zrak potuje s hitrostjo 331 m/s.

### 3.2 Človeško zaznavanje zvoka

Bobnič = membrana, ki vibracije prenaša na kosti kostnice -> prenesejo se na tekočino v polžu. Ljudje zaznavamo vibracije med 20Hz in 20kHz. Za slišanje zvoka potrebujemo zadostno moč (amplitudo). Pri frekvencah proti koncu slišnega intervala potrebujemo več amplitude. Prag slišnosti: amplitude, pri katerih so osebkii zaznali vsaj polovico tonov. Mladi ljudje boljše slišijo visoke frekvence (nad 2kHz). Frekvenco zaznavamo kot višino tona, glasnost

je intenzivnost zvoka, barva tona je povezana z obliko valov. Dve ušesi -> zaznavamo prostorsko lokacijo.

### 3.3 Signal in šum

Zvok lahko ločimo od nezaželenega nihanja okolja, ki ga imenujemo šum. Zvok se v praksi meri z relativno glasnostjo v primerjavi s hrupom (dB).

### 3.4 Digitalni zvok

Prednost je lažja obdelava in shranjevanje s stiskanjem. Je analogni signal, ki ga lahko izmerimo kot spremenjen zračni tlak s časom, kar povzroči vibracije v snemalni membrani, ki se ga lahko prenesemo kot mehansko gibanje in nato v električni signal, ki se izmeri. Če ga želimo digitalizirati, ga moramo vzorčiti v določenih časovnih presledkih in dejanske vrednosti kvantizirati na določeno ločljivost. Vzorčimo ga v časovnih intervalih; frekvenca vzorčenja.

Nyquist-Shannonov teorem: frekvenca vzorčenja mora biti vsaj dvakrat višja od največje frekvence v signalu.

Da zagotovimo, da signal vsebuje le primerne frekvence, ga spustimo skozi analogno nizkoprepustni filter, preden ga vzorčimo.

Kvantizacijski šum se meri z razmerjem med signalom in kvantizacijskim šumom (SQNR), pri čemer je višja vrednost boljša, saj to pomeni, da je signal močnejši v primerjavi s šumom.

12 bitov je že dovolj za ustrezno reprodukcijo zvoka.

Če signalu dodamo majhno količino šuma, postanejo napake kvantizacije bolj statistično neodvisne in tako manj opazne.

Izraz pulzno kodna modulacija (PCM) je uradni izraz za kombinacijo vzorčenja in kvantizacije. Pristopi PCM se razlikujejo po vrsti uporabljene kvantizacije.

### 3.5 Frekvenčni spekter

Prikazuje, kako močno so določene frekvence prisotne v obliki zvočnega valovanja. Frekvence so predstavljenje s sinusnimi krivuljami, vsaki od njih pa je dodeljen koeficient, ki označuje njihovo moč.

Metoda, ki razgradi zvezne signale v niz osnovnih sinusoid, se imenuje Fourierjeva transformacija. Lahko ga transformiramo nazaj v časovno območje z inverzno Fourierovo transformacijo. V digitalnem svetu je Fourierjeva transformacija enakovredna diskretni Fourierjevi transformaciji (linearna transformacija  $n$ -točkovnega signala v časovnem območju v  $n$ -točkovni frekvenčni spekter).

### 3.6 Obdelava zvoka

Izvaja se z analognimi zvočnimi signali, ki jih predstavljajo zvezne funkcije v obliki spreminjanja električnega toka ali napetosti (lastnosti signala se spreminjajo z uporabo elektronskih komponent - kondenzatorji, upori, induktorji, ...).

Digitalna obdelava zvoka se izvaja z algoritmi, ki delujejo na računalnikih za splošne namene in obdelujejo digitalni signal. Digitalni filtri imajo običajno tudi boljše razmerje med signalom in šumom. Pri analognih filtrih je vsaka komponenta potencialni šum (večji filter -> večji šum).

Digitalne filtre lahko opišemo na več načinov, razredi filtrov imajo različne lastnosti in omejitve. Filter je linearen, če ga je mogoče zapisati kot linearno diferencialno enačbo, sicer je nelinearen. Filter, ki uporablja za izračun le pretekle vrednosti signala, je vzorčen, če pa uporablja tudi prihodnje vrednosti, je nevzorčen. Če filter vrne enak rezultat neglede na čas dovedenega vhoda, je časovno nespremenljiv, v nasprotnem primeru je časovno spremenljiv.

Filtri z najlepšimi lastnostmi so linearni časovno invariantni filtri (LTI), ki imajo lastnost aditivnosti in homogenosti (seštevanje in skaliranje). Frekvenčna analiza: preoblikovanje primitivnega sinusoidnega signala s filtrom glede na njegove parametre.

### 3.7 Linearni filtri

Najbolj splošni filtri se imenujejo podpasovni filtri.

Najpogostejši filtri tega razreda:

- nizkopasovni: prepuščajo samo frekvence pod pragom
- visokopasovni: prepuščajo samo frekvence nad pragom
- pasovna prepustnost: prepušča frekvence v intervalu
- band-stop: blokira frekvence v intervalu

Vse podpasovne filtre lahko izvedemo z enim samim prototipnim filtrom. Njihova izvedba je zahtevna, v praksi celo nemogoča. Obstaja pa veliko končnih približkov z različnimi prednostmi in slabostmi:

- Butterworth: frekvenčni odziv je v prepustnem pasu kar se da raven
- Chebyshev: čim manjša napaka med idealizirano in dejansko karakteristiko filtra v celotnem območju filtra, vendar z valovanji v prepustnem pasu v primeru filtrov tipa 1 ali v zaporednem pasu v primeru filtrov tipa 2
- eliptični: izenačeno valovanje v prepustnem in zapornem pasu

### 3.8 Izenačevanje

Klasičen primer uporabe podpasovnih filtrov je večpasovno izenačevanje. Signal je razdeljen na več pasov, ki se nato povečujejo z različnimi faktorji in združijo nazaj skupaj. Ta nam omogoča povečanje nizkih frekvenc ali visokega frekvenčnega razpona.

### 3.9 Filtri z glavniki

Filter z glavniki (grebenski fliter) ustvarja glavniku podoben vzorec, ko opazujemo amplitudni odziv pri frekvenčni analizi. Splošni filter kombinira signal s preteklim vhodnim vzorcem in preteklim izhodnim vzorcem, od katerih je vsak obtežen z ustreznim faktorjem.

$$y(t) = x(t) + \alpha x(t - a) + \beta y(t - b)$$

Dva primer, pri katerih sta vzorca preteklega vhoda ali preteklega izhoda obtežena z ničlo, se imenujeta povratni in napajalni grebenski filtri.

Eden lažje razumljivih aplikacijskih grebenskih filtrov je zakasnitev (širitev istega signala z določeno časovno razliko in ga sešteje z izvirnim signalom).

Odmevni filter je razširitev zakasnitvenega filtra z namenom simuliranja akustičnih lastnosti resničnih prostorov (odvijanje od več sten, majhen ali velik prostor).

Flanger je tudi izpeljan učinek zakasnitve. Razlika v času zakasnitve se spreminja z nizko frekvenco, kar povzroči učinek odbijanja. Filter ni časovno variabilen, je primer linearne časovno variabilnega filtra.

Chorus je tehnično podoben flangerju, le zakasnitve so daljše, višina podvojenega signala se lahko tudi nekoliko premakne; učinek igranja več inštrumentov hkrati, vsak z nekoliko drugačnim časom in/ali višino. Izhodni zvok je bogatejši kot vhodni.

### 3.10 Nelinearni filtri

So vsi filtri, ki jih ni mogoče formulirati kot linearno enačbo. Uporabljamo jih lahko v različnih primerih.

Prvi primer je odstranjevanje šuma z uporabo mediane. Drugi primer so filtri za stiskanje dinamičnega razpona, ki uporabljajo dinamično skaliranje amplitude signala glede na njegovo vrednost. Kompresija dinamičnega razpona se uporablja za učinkovito uporabo razpoložljivega razpona na magnetnih trakovih za snemanje.

Klasičen primer nelinearnih filtrov za obrezovanje pri obdelavi zvoka je učinek popačenja (preobremenitev vakuumskih elektronk). Digitalno je popačenje izvedeno kot nelinearn funkcija skaliranja, nizke amplitude ostanejo take kot so, visoke vrednosti pa so skalirane. Glede na uporabljeno funkcijo imenujemo izhomehko ali trdo obrezovanje.

## 4. KOMPRESIJA

Za shranjevanje in prenos podatkov. Senzorični podatki so zelo primerni za stiskanje, saj je del podatkov vedno odveč - stiskanje brez izgub (zmanjšana

velikost podatkov, vsebina ostane nedotaknjena).

Pri vizualnih informacijah govorimo o prostorski korelaciji, kjer se barva v sosednjih pikslah verjetno ne bo spremenila za večino pikslov, in kanalski korelaciji, kjer so informacije v treh barvnih kanalih prav tako kolerirane. Pri videzu poznamo tudi korelacijo gibanja (pri večini parov zaporednjih sličic se večina pikslov ne veliko spreminja).

Lahko si zamislimo stiskanje z izgubami, saj je človeški zaznavni sistem različno občutljiv na različne vrste informacij (bolj v intenzivnosti, kot v barvi).

## 4.1 Osnove stiskanja

Prva stopnja: preoblikovanje podatkov v obliko, ki je primernejša za stiskanje. Druga stopnja: preslikava, pri kateri se podatki preslikajo v simbole, ki so učinkovitejši za kodiranje (podatki se razdelijo na manjše dele). Druga znana tehnika preslikave je kodiranje po dolžini poteka.

Tranformacij in preslikava simbolov se štejeta za korake preobdelave, večina stiskanja se zgodi v končnem koraku kodiranja simbolov.

Glavni cilj stiskanja brez izgub je čim bolj zmanjšati število bitov, ne da bi pri tem izgubili informacije. Poznamo dva splošna pristopa:

- statistika simbolov se uporablja za kodiranje pogostejših simbolov s krajšimi besedami in manj pogostih z daljšimi besedami - **kodiranje s spremenljivo dolžino (VLC)**; Huffmanov algoritem je med najbolj znanimi
- grajenje slovarja, kjer je to mogoče, ta se ustvari dinamično brez uporabe statističnih podatkov; predstavnik Lempel-Ziv

Povratni postopek se imenuje dekompresija in se izvaja v obratnem vrstnem redu: dekodiranje, obratno preslikavanje in obratna transformacija.

Vidiki, ki določajo, kateri pristop stiskanja je najboljši:

- učinkovitost: razmerje med velikostjo surovih podatkov in stisnjenega sporočila
- zakasnitev: čas, potreben za stiskanje; tudi količina podatkov
- zapletenost izvedbe: kako zapletena je izvedba
- robustnost: ali je mogoče popraviti poškodovan stisnjen tok
- skalabilnost: podpora različnim profilom stiskanja

## 4.2 Metoda statističnega kodiranja

Huffmanov kod temelji na pojavu informacijske entropije, ki nam pove, koliko bitov je potrebnih za kodiranje enega znaka na podlagi verjetnosti njegovega pojavljanja v zaporedju simbolov.

Huffmanov algoritem je vrsta optimalnegapredponskega kodiranja, ki določa kodne besede, ki minimizirajo povprečno dolžino besed in izpolnjujejo dodatne zahteve:

- pravilnost: različnim simbolom so dodeljene različne besede
- edinstvenost: sporočilo je mogoče razumeti na en sam način
- takojšnje: vsako besedo je mogoče dekodirati takoj, ko je prebrana

Algoritem najprej razvrsti simbole po padajočih verjetnostih pojavljanja, nato iterativno združi dva simbola z najmanjšo verjetnostjo, jima dodeli en bit v končni besedi, da se razlikujeta, in to ponavljamo . . . .

Spodnja meja kodiranja je en bit za simbol, če imamo zelo pogost simbol, vsak simbol posebej zahteva neničelno število bitov. Huffmanov algoritem je optimalen, ko moramo kodirati vsak simbol posebej -> tej omejitvi se izognemo z aritmetičnim kodiranjem. Druga težava je, da bo zaradi odtisovanja statistične frekvence simbolov kodiranje postalo neoptimalno. Posodobitve kodnih besed so drage -> omejimo največjo možno velikost kodne besede.

### 4.3 Metode kodiranja slovarjev

Pri statističnem kodiranju je treba vnaprej poznati porazdelitev simbolov. Želimo, da se porazdelitev med zaporedjem tudi spremeni -> ustvarimo nov slovar za preslikavo. Taki algoritmi uporabljajo dinamično generiranje slovarja za zaporedja spremenljive dolžine simbolov. Kodne besede so fiksne dolžine in nam povedo, katero zaporedje iz kodiranega sporočila je treba uporabiti. Primerni so za sprotno stiskanje in dekompresijo, niso pa primerni za kratka zaporedja podatkov (lahko dobimo daljša kodna sporočila).

Temeljni algoritem je Lempel-Ziv oz LZ77. Ta uporablja drsno okno (nestisnjeni obdelani podatki), slovar je zato dinamičen (se spreminja s tokom podatkov) -> podatke lahko tudi dekompresiramo (nekatero kasnejše kodne besede se nanašajo na prejšnje porazdelitve simbolov).

Kodne besede v LZ77 so trojčki (odmik, dolžina, simbol (ki sledi trenutnemu segmentu)). Glavna težava algoritma je omejena velikost slovarja -> alternativni algoritem LZ78 -> slovar je izrecno zgrajen na podlagi celotnega zaporedja, kodne besede so preprosto indeksi v slovarju -> omogoča delno kompresijo (če imamo na voljo slovar). Še ena izboljšava je algoritem LZW, kjer je slovar predhodno inicializiran z vsemi možnimi simboli. Ko je najdeno neujemanje, se trenutni simbol predpostavi kot prvi simbol obstoječega niza v slovarju.

Oba pristopa (s slovarjem in statistični) je mogoče kombinirati - DEFLATE (najprej LZ77 za odpravo podvojenih podzaporedij simbolov, nato se dobljene kodne besede kodirajo z uporabo Huffmanovega algoritma). To kodiranje je skalabilno, nastavimo lahko čas, ki ga porabimo za iskanje podzaporedij. Uporablja se pri znanih formatih za stiskanje datotek - PKZIP, gzip.

### 4.4 Stiskanje slik

Brezizgubni PNG in izgubni JPEG format



## 4.5 Portable Network Graphics (PNG)

Uveden kot nadomestek za Graphic Interchange Format (GIF), ki je imel omejitve pri uporabi v svetovnem spletu. PNG zagotavlja popolno prosojnost kanala alfa in podpira način polne palete RGB in indeksirane palete. Za doseganje sprejemljivega tiskanja se uporablja dvostopenjsko stiskanje.

Prva stopnja: vrednosti pikslov se filtrirajo s prediktivnim filtriranjem - vrednost piksla se predvidi na podlagi prejšnjih vrednosti, ohrani se samo razlika.

Druga stopnja: razlike se stisnejo z uporabo DEFLATE.

Prediktivno filtriranje omogoča vsaj delno modeliranje zaporednih podatkov (pri slikah na vrstico). PNG uporablja več pristopov, ki se lahko preklapljajo na podlagi posamezne vrstice:

- nespremenjena vrednost: vsaka vrednost je kodirana brez napovedi
- uporaba vrednosti prejšnjega piksla v vrstici
- uporaba vrednosti prejšnjega piksla v stolpcu
- uporaba srednje vrednosti prejšnjih pikslov v vrstici in stolpcu (zaokroževanje navzdol)
- uporaba kombinacije več prejšnjih vrednosti

## 4.6 Kodiranje s krajšanjem blokov

Glavni plus pri stiskanju z izgunami je izkoriščanje prostorske korelacije (sosednji piksli so si običajno zelo podobni), to lahko izkoristimo z delitvijo slike na majhne regije in nato vsako regijo kodiramo tako, da nekatere informacije zavržemo. Tak primer kodiranja je BTC. Vsak blok pikslov je kodiran z dvema intenzitetama: srednja vrednost plus/minus standardna vrednost odklona, obtežena z razmerjem pikslov nad in pod standardno vrednostjo. Kodirna velikost bloka za 8-bitne piksele je 4 bajte (1 bajt za srednjo vrednost, 1 bajt za odstopanje, 2 bajta za 4x4-bitno polje); nestisnjena vrednost je 16 bajtov.

## 4.7 JPEG format

Joint Photographic Experts Group, najbolj priljubljen format za stiskanje z izgubami. Kodiranje in dekodiranje zasnovana tako, da nimata veliki računskih zahtev (uporaba v vgrajenih napravah). Podpira barvno globino med 8 in 12 bitov na kanal. Podpira zaporedno (slika je kodirana v eni ločljivosti) in progresivno kodiranje (slika kodirana najprej v nižji ločljivosti, nato pa se postopoma dodaja več podrobnosti). Razlika je pri nalaganju slike, ko vsi podatki niso še na voljo; progresivno - lahko prikažemo že v nižji ločljivosti v celoti, zaporedno - sliko lahko prikažemo le delno.

Datotečni format vsebuje podatke o sliki in nekatere metapodatke. Znan je tudi kot JPEG File Interchange Format (JFIF). Različica, kjer so še metapodatki o fotoaparatu je znana kot Exchangable Image File Format (EXIF).

Stiskanje brez izgub je podobno kot pri PNG, uporablja napovedno kodiranje,

ki mu sledi Huffmanovo kodiranje; ni uporabljen pogosto. Stiskanje brez izgub JPEG se uporablja pri kodiranju videoposnetkov (zavržejo se informacije, ki so za človeka manj pomembne).

Slika se razdeli na 8x8 bloke (dodajo se dodatni podatki, če ni deljiva z 8), vsak blok se pretvori v frekvenčno področje z DCT (podobna Fourierjevi transformaciji, večino signalov opišemo z manj osnovnimi funkcijami - spektralno zgoščevanje). Dct razgradi 64 bajtov v enosmerni izraz (current term) (to je srednja vrednost bloka), in 63 terminov izmeničnega toka (alternating current terms) (koeficienti različnih kosinusnih osnovnih funkcij).

Format podpira tudi barvno podvzorčenje kot še en način za izgubno zmanjšanje velikosti. Ker je človeško oko manj občutljivo na spremembe v barvi kot v svetilnost, uporablja barvni format, ločen na barve - YCrCb - ter zmanjša vzorce kanalov barve z določenim faktorjem.

## 4.8 Stiskanje videoposnetkov

Zahtevnejše kot stiskanje slik, saj se velikost pomnoži s časovno dolžino. Ena od možnosti je izkoriščanje same korelacije med sličicami, praktičen primer tega je kodek Motion JPEG, ki vsako sličico zakodira kot sliko JPEG -> kodiranje in dekodiranje se lahko izvajata v realnem času -> bitni tok ima nizko stopnjo stiskanja -> izkoriščamo časovno korelacijo.

Tipičen pristop k časovnem kodiranju je uporaba ujemanja blokov. Slika se razdeli na bloke, tem pa se poišče ujemanje v prejšnji sliki (popolnega ujemanja običajno ne najdemo -> iščemo najbolj podoben blok). Blok nato kodiramo z uporabo vektorjev prevajanja in razlik do referenčnega bloka.

Standard za stiskanje in kompresijo videa se imenuje video kodek. Izbera pravega kodeka je odvisna od več parametrov:

- stopnja stiskanja
- hitrost prenosa v primerjavi z izgubo
- zapletenost algoritma
- značilnost komunikacijskega kanala
- fiksna/spremenljiva bitna hitrost
- združljivost s standardi
- izguba informacij med stiskanjem

Danes najbolj znane video kodeke predlaga skupina Motion Picture Expert Group (MPEG). Nekatere pomembni standardi:

- video in zvok na digitalnih napravah (MPEG-1, MPEG-2)
- video, zvok, 3D grafika, splet, ... (MPEG-4)
- shranjevanje in priklic v videu (MPEG-7)
- interakcija z večpredstavnostnimi aplikacijami (MPEG-21)

## 4.9 MPEG-1

Vključuje video in avdio specifikacije stiskanja. Celoten standard ima 5 delov: vido, zvok, skladnost, referenčno izvajanje, sistem. Cilj: doseči shranjevanje in predvajanje za video in zvok kakovosti VHS ter omogočiti prenos v realnem času s pasovno širino okrog 1,5 Mbit/s (1:26 za video, 1:6 za zvok; brez pretirane izgube kakovosti). Predvidena uporaba je asimetrična, vsebina se stisne enkrat (kodirnik je lahko zapleten), počasno kodiranje; dekodirnik pa mora biti hiter in preprost ter delovati na različnih konfiguracijah strojne opreme. Video kodek mora izpolnjevati več zahtev:

- normalno predvajanje z naključnim dostopom
- podpora za urejanje videoposnetkov
- povratno in hitro predvajanje
- različna ločljivost, hitrost snemanja sličic
- poceni strojna izvedba dekodejev

Vsaka slička v MPEG-1 je predstavljena z barvni prostorom YCrCb, standard podpira vzorčenje 2:1:1 (blok štirih slikovnih pika ima deljeno barvno informacijo); to je prikazano v delitvi blokov - Y je razdeljen na bloke 16x16 slikovnih pik, kanala Cr in Cb pa na bloke 8x8 slikovnih pik. Bloki so združeni v skupine (makrobloki), ti so združeni v zaporedja (rezine). Vsaka rezina ima lahko različne parametre stiskanja, s katerimi se prilagodi velikost rezine glede na kakovost; vsaka rezina je kodirana ločeno (olajša odpravljanje napak).

Za upoštevanje časovne redundance so v video kodekih tri glavne vrste okvirjev: okvir I, okvir P in okvir B.

Okvirji I so medokvirji, vsak okvir je samostojen, v celoti ga lahko dekodiramo na podoben način kot sliko JPEG. Imajo kromatično podvzorčenje in so organizirani v korakih makroblokov. Imenujejo se tudi kj+ljučni okvirji, saj omogočajo naključen dostop do video vsebine (za kodiranje ne potrebujejo nobenih dodatnih informacij).

Okvirji P so predvideni okvirji, so podatkovno odvisni od prejšnjega okvirja P ali I in kodirajo le razlike do njega. Podatki za makroblok so kodirani kot vektor gibanja (izračuna se s pomočjo podobnosti kanala Y in razlike do ustreznih podatkov iz prejšnjega okvirja). Če so razlike prevelike, se makroblok dekodira tak kot je in ne kot razlika.

Okvirji B so dvosmerni predvideni okvirji, kodirani s pomočjo predhodnih in naslednjih okvirjev P ali I. Kot osnova se izračun razlike se uporabi povprečje zaplate iz prejšnjega in naslednjega referenčnega okvirja. Zahtevajo predhodno dekodiranje več okvirjev (tudi iz prihajajočega toka) -> postopek imenujemo zakasnjeno dekodiranje.

Podatkovni tok je organiziran hierarhično, zaporedje je sestavljeno iz več skupnih slik (GOP), vsaka slika iz več rezin, vsaka rezina iz več makroblokov, vsak makroblok iz 5 blokov, vsak blok pa kot koeficient DC in niz Huffmanovih kodnih besed.

## 4.10 MPEG-2

Višje ločljivosti posnetkov (HDTV) in večja pasovna širina prenosa (4 - 15 MB/s). Primeren je za digitalno televizijo in uvaža podporo za prepleteni video ter boljšo raobustnost (odpravljanje napak). Je razširljiv in združljiv z MPEG-1 (MPEG-2 dekodirji mnorajo dekodirati tudi MPEG-1).

Prepleteni video lahko dekodiramo na dva načina. V formatu okvirja so line in sode vrstice kodirane kot ena sama slika (kodirane z eno samo glavo - parametri); ta format se uporablja v progresivnem načinu. Po drugi strani pa lahko format ločeno kodira lihe in sode vrstice (vsako polje kodirano kot ločena slika z lastno glavo). Kodirnik lahko preklaplja med dvema načinoma in izbere primernejšega glede na velikost in kakovost slike.

Standardizira različne profile za različne aplikacije in ravni kakovosti. Vidki razširljivosti pomemben za uporabnike, ki zaradi pasovne širine ne morejo gledati vsebin v polni ločljivosti. Standard ne opredeljuje dvosmerne komunikacije -> video moramo ločiti na več slojev: osnovni sloj (samostojen, vsebuje le grobe informacije), sloj izboljšav (postopoma izboljšuje kakovost). Prenos plasti je ločen, odjemalec lahko dobi in dekodira le nekatere plasti in vseeno dobi dobro kakovost. Obstaja več vrst skalabilnost:

- prostorska skalabilnost (velikost slike): osnovna plast (posnetki z nižjo ločljivostjo), ločljivost boljša z plastmi izboljšav
- frekvenčna razširljivost: osnovna plast zagotavlja samo koeficiente enosmernega toka in vektorje gibanja, izboljšave vedno več koeficientov dvosmernega toka
- SNR skalabilnost (razmerje signal/šum): osnovna plast zagotavlja močno kvantizirano intenzivnost, kodirana v izvirni ločljivosti, plasti izboljšave prinašajo dodatne informacije
- časovna razširljivost: osnovna plast zagotavlja slike z nižjo hitrostjo sličic (npr samo I sličice) izboljšave prinesejo še vmesne sličice, z uporabo gibanja se rekonstruirajo

## 4.11 MPEG-4

Osrednja tema MPEG-4 je interaktivna večpredstavnostna vsebina. Vidoekodeki imajo še boljše stopnje stiskanja z večjo prepustnostjo. So robustni v okolju z napakami. Eden od vidikov standarda je tudi opredelitev medijskega objekta in način, kako je mogoče naključno dostopati do več medijskih objektov. Ti imajo lahko različne vire, lahko so prave ali sintetične slike, zvoki, vektorska grafika, video. Ključna zamisel: sodelovati s posameznimi objekti in z njimi manipulirati. Razdeljeni so na več kanalov in ločeno kodirani, prenašajo se v ločenih tokovi.

Objekte lahko hierarhično umestimo v prizor, ko so dekodirani na napravi za upodabljanje. Standard opredeljuje jezik za opis scene - Binary Format for Scenes (BIFS):

- zaporedje videoposnetkov - celoten vizualni prizor

- videoobjektiv - poljubna nepravokotna oblika
- plast video objektov - podpora za skalabilno kodiranje video objektov
- ravnina video objektov - posnetek videoobjekta v določeni časovni točki
- skupina ravnih video objektov

Stiskanje MPEG-4 opredeljujeta dva video kodeka: prvi kodek je MPEG 4 Part 2, bolj znan tudi kot H.263 (opredeljuje 21 video profilov; izboljšave: kompenzacija gibanja na četrtno pikslo, možnost opredelitve vektorjev gibanja na četrtno pikslo, globalna kompenzacija gibanja), drugi kodek: H.264 oz. napredno kodiranje videa. (uporablja se za prizemno televizijo HD, prevzet standard za video v HTML5; ponuja povečano razmerje stiskanja - do 1:50).

Glavne prednosti kodeka H.264 so:

- novi tehniki VLC: kontekstno prilagojeno kodiranje s spremenljivo dolžino in kontekstno prilagojeno binarno aritmetično kodiranje
- spremenljive velikosti blokov
- natančnejša kompenzacija gibanja
- napovedovanje znotraj okvirja
- uvaja singlano prilagodljive deblokirane filtre, ki zgladijo robove okoli blokov -> so manj opazni

## 4.12 Novejši standardi kodiranja videa

Naslednja generacija ni del standarda MPEG-4, temveč MPEG-H (H - heterogeno okolje), imenuje se tudi H.265 in velja za kodek naslednje generacije; njegova predvidena uporaba je pretakanje videoposnetkov - zmanjšati velikost posnetkov na spletnih straneh in povečati njihovo ločljivost. Ena glavnih izboljšav je spremenljiva velikost bloka (okvir je razdeljen na enote kodirnega drevesa).

H.266 je še naslednja generacija od H.265, ki uvaja nove vidike:

- ločljivost od 4k do 16k
- podpora za vsesmerne posnetke
- podpora visokemu dinamičnemu razponu
- pomožni kanali
- spremenljiva hitrost sličic

## 4.13 Vrednotenje stiskanja slik in videoposnetkov

Razvoj novih tehnik stiskanja je rezultat stiskanja brez izgub, ki se mora ujemati le toliko, da očitnost tega ne opazi. Najpogosteje uporabljeni merili vrednotenja sta PNSR (razmerje med najvišjim signalom in šumom) in SSIM (indeks strukturne podobnosti). Prva primerja slike od piskal do piskla, druga pa poskuša primerjati lastnosti na višji ravni in primerja vse piksele v oknu izbrane velikosti.

## 4.14 Stiskanje zvoka brez izgub

Zvočni kodeki za stiskanje brez izgub niso znani poslušalcu (namenjeni le za kasnejšo obdelavo ali pa ljudem, ki jih motijo zaznavni modeli z izgubo).

Eden od sodobnih zvočnih kodekov brez izgub je Free Lossless Audio Codec (FLAC), ponuja zmanjšanje za od 50% do 80%. Pri stiskanju se uporablja modelno napovedovanje signala na podlagi preteklih vrednosti, nato se kodirajo še ostanki. Podatki se kodirajo v več okvirjih., ki so razdeljeni na podokvirje (vsebujejo kratke dele zvočnih posnetkov). Podokvirji si delijo nekatere parametre kodiranja, medtem ko se lahko parametri popolnoma spremenijo med posameznimi okvirji. Podpira več načinov napovedovanja:

- ničelni: kodiranje digitalne tišine, napovedovanje s konstantno vrednostjo
- verbatim: napovedovalec ničelnega reda, ostanek je sam signal
- fiksni linearni: prileganje polinoma  $p$ -razreda na  $p$  točk
- FIR linear - linearna kombinacija prejšnjih vzorcev

Golombovo kodiranje je hibridna tehnika kodiranja, učinkovita, ko prevladujejo majhne vrednosti porazdelitve. Rižovo kodiranje je učinkovitejša različica Golbovega kodiranja, kar pomeni, da se delitev lahko učinkovito izvede na digitalnih računalnikih.

## 4.15 Stiskanje govora

Pri nelinearnem kvantiziranju signalov se upošteva, da imajo govorni signali neenakomerno porazdelitev amplitude (majhne amplitude signala so bolj verjetne kot večje). Podobno večkanalno kodiranje kodira različne frekvenčne pasove z različnimi ločljivostmi (več prostora se nameni frekvenčnemu spektru človeškega govora).

Kodirnik za govorni signal se imenuje tudi vokoder. Primeri vokoderjev, ki se uporabljajo v telefoniji:

- vokoder z linearnim prediktivnim kodiranjem (LPC) kodira okvire (dele signala, 30 do 50 posnetkov na sekundo) kot parametre časovno spremenljivega modela glasovnega trakta; priteneseni podatki so nabor parametrov, napovedovanje koeficientov z linearnim modelom prejšnjih koeficientov; LPC-10 uporablja 10 prejšnjih koeficientov; predpostavlja, da prenaša samo en človeški glas
- vokoder CELP (Code-excited Linear Prediction): glavni namen je še vedno prenos enogovornega govora, vendar deluje tudi na govoru več ljudi; signal kodira z uporabo kratkotrajne metode LPC ter prilagodljivega iskanja kodne knjige

## 4.16 Stiskanje zvoka MPEG-1

Opreddeljuje 3 zvočne kodeke, ki se imenujejo tudi plasti, saj ponujajo združljivost navzdol (tok s plasti 1 je mogoče dekodirati s dekoderjem plasti 3), vsaka plast

bolj zapleten kodirnik.

Glavni koncepti stiskanja so prilagojeni za izkoriščanje nepopolnosti v človeškem slušnem sistemu za zmanjšanje količine kodiranih informacij. Jedro mehanizma stiskanja se imenuje frekvenčno maskiranje. Ta psihoakustični pojav nastane, ko sta hkrati zabeleženi dve bližnji frekvenci, če je ena glasnejša, bo prekrila drugo.

Če zvok zavzame dva ali več pasov, ga zaznamo kot glasnejšega (človek ima približno 24 do 25 kritičnih pasov).

Drugi psihoakustični pojav je časovno maskiranje. Po glasnem zvoku traja nekaj časa, da slišimo tih zvok.

Najpomembnejši vidik stiskanja zvoka z izgubami je, koliko bitov je dodeljenih določenim pasovom, tako da to ne bo vplivalo na zaznavno kakovost signala.

MPEG-1 Layer 1: uporablja 32 filtrov za ločevanje signalov v enakomerno porazdeljene pasove, frekvenčni spekter se pridobi s FFT na 8ms blokih signala; namenjen uporabi za shranjen zvok.

MPEG-1 Layer 2: zasnovan za digitalno oddajanje zvoka.

MPEG-1 Layer 3: prvotno namenjen za prenos zvoka prek linij ISDN, danes se uporablja za shranjevanje glasbe (MP3 - ima skoraj univerzalno strojno in programsko podporo).

#### **4.17 Stiskanje zvoka MPEG-2**

Uvaja dve specifikaciji za kodiranje zvoka. Napredno kodiranje zvoka je namenjeno preglednimi zvočni reprodukciji v kinih. Uporablja se tudi na DVD-jih in sorodnih medijih.

### **5. ISKANJE INFORMACIJ**

#### **5.1 Iskanje besedila**

Cilj je vrniti ustrezne besedilne dokumente na podlagi iskalne poizvedbe. Poizvedba lahko temelji na preprostih izrazih, lahko pa je tudi bolj zapletena. Poleg postopka poizvedovanja po naboru podatkov imamo torej tudi postopek indeksiranja dokumentov (njihovo organiziranje). Osrednja sestavina iskanja informacij se imenuje poizvedovalni mehanizem. Motor (engine) je program, ki skrbi za indeksiranje dokumentov, razčlenjevanje uporabnikove poizvedbe, ujemanje dokumentov s specifikacijami poizvedbe, razvrščanje ujetih dokumentov po ustreznosti in vračanje rezultatov uporabniku.

#### **5.2 Logični izrazi**

Osnovni scenarij: uporabnik išče dokumente, ki vsebuje nekaj besed in ne vsebujejo drugih besed. Take poizvedbe se imenujejo logični izrazi. Iskanje po vseh dokumentih bi bilo zamudno tudi za preproste poizvedbe, zato hranimo

zapise, kateri dokumenti vsebujejo besede. Te informacije so shranjene v binarni matriki - matrika pojavnosti. Ppreprosto preverimo, kateri stolpci - dokumenti - ustrezajo vsem zahtevanim besedam - vrstice - in vrnemo seznam; ta pristop je neoptimalen. Bolje je hraniti seznime dokumentov, ki vsebujejo določeno besedo. Seznam identifikatorjev dokumentov za besedo se imenuje obrnjeni indeks za to besedo.

### 5.3 Dokumenti in slovarji

Zmogljivosti iskalnika so odvisne od načina izbire dokumentov in kako so indeksirani. Natančno iskanje bo imelo slabši odpoklic, saj lahko izgubimo pomembne informacije, ki jih poizvedba ne bo prikklicala. Pri grobozrnatem iskanju bo odpoklic dober, vendar rezultati ne bodo uporabnik, ker so dokumenti ogromni.

Med indeksiranjem, tj. gradnjo obrnjenih indeksov, se vsak dokument najpšrej razdeli na seznam žetonov, nato se z jezikovno obdelavo in normalizacijo žetonov zmanjša število različnih besed s preslikavo na skupni pomenoslovni koren. Nato je seznam normaliziranih žetonov združen z indeksi dokumentov, pari pa se razvrstijo po abecedi glede na žetone in združijo v skupine v seznime za posamezne žetone (v obrnjeni indeks), zapomnimo si tudi, v koliko dokumentih se izraz pojavlja.

Tokenizacija pretvori besedilo v posamezne atomske enote (različne kodirne sheme, smer besedila, besede z ločili, okrajšave, besede s pomišljaji, besede združene skupaj). Besede, ki se zelo pogosto pojavljajo v vseh dokumentih in nimajo posebnega pomena, se imenujejo stop besede. Te besede se odstranijo na samem začetku postopka. Seznami stop besed so za vsak jezik specifični. Dobri iskalniki zaznajo stop besede v besednih zvezah - te pustijo.

Enake besede se pojavljajo v različnih oblikah: male in velike črke, naglasna znamenja, jezikovne različice, oblikovanje števil in datumov. Naloga normalizacije je zmanjšati razlike z prepoznavanjem teh podobnosti in zamenjavo teh žetonov z enim samim.

Z normalizacijo sta povezana tudi procesa lematizacije (postopek normalizacije, ki temelji na jezikovnih pravilih) in izvornega zapisa (hevrističen pristop k normalizaciji, ki temelji na rezanju besed).

### 5.4 Besedne zveze

Obstajata dva pristopa za kodiranje besednih zvez:

- indeks dveh besed: pari žetonov se indeksirajo skupaj namesto posameznih žetonov; je preprost za izvajanje vendar ponuja le omejene količine besednih zvez v dokumentih
- položajno indeksiranje: vsak indeks dokumenta na seznamu zaizraz vsebuje tudi seznam položajev izraza, kjer se izraz pojavlja v dokumentih; je bolj zapleten, vendar podpira več in bolj zapletene besedne zveze



## 5.5 Dopustnost napak in nepopolnost

Nepopolne poizvedbe so običajno prepoznane po nadomesnem simbolu, ki pomeni, da ga lahko nadomesti karkoli, zato so izrazi kombinacija določenega zaporedja črk in delov, ki lahko vsebujejo poljubne črke. Običajni indeksu uporabljajo hash tabelo (ni primerna za poizvedbe z nadomestnimi znaki). V nekaterih primerih se lahko uporabi urejena posatkovna struktura; uporabljajo se tudi druge tehnike (Permuterm ali indeks K-gram), vendar je obdelava poizvedb z nadomestnimi znaki na splošno počasnejša od obdelave preprostih logičnih poizvedb:

- Permuterm: indeks vsebuje vse premike določenih izrazov s posebnimi simbolom, ki označuje konec besede
- indeks K-gram: slovar vsebuje vse podnize dolžine K, za označevanje začetka in konca se uporablja posebni znak

Popravek napak se lahko izvede na podlagi vsakega izraza posebej (za besede, ki jih ni v slovarju, se poiščejo najbolj podobne besede v slovarju). Za izračun podobnih izrazov se lahko uporabi Levenshteinova razdalja ali fonetična razdalja. Uporablja se tudi kontekst ali pogosto uporabljene besedne zveze.

## 5.6 Razvrščanje in primerjanje dokumentov

Bolj optimalno je, če najprej dokumente razvrstimo po njihovi ustreznosti glede na poizvedbo. Tisti dokumenti, ki večkrat vsebujejo poizvedovalni izraz, naj bi bili bolj pomembni - razvrščanje glede na pogostost izrazov. Zbirka vseh besed je predstavljena kot vektorski prostor - vsak dokument predstavljen kot točka glede na to, kolikokrat se je beseda pojavila. Dobimo oceno "pogostost terminov in obratna pogostost dokumentov". Dobimo utež: če je visoka, je izraz pogost le v majhnem številu dokumentov.

Primerjavo dokumentov lahko predstavimo kot razdaljo med točkama v vektorskem prostoru.

## 5.7 Povratna informacija o ustreznost

Majhna je verjetnost, da bo uporabnik poznal vse dokumente v naboru podatkov, zato ne bo vedel, kako določiti dovol specifično poizvedbo. Obstajata dva pristopa, kako narediti podatkovno zbirko bolj dostopno:

- globalni: razširitev poizvedbe na čim več možnosti s čim več možnimi izrazi s popravljanjem napak, sinonimi itd.
- lokalni: na podlagi interakcije med uporabnikom in sistemom uporabnik dobi rezultat in možnost, da s seznama izbere ustrezne in neustrezne dokumente ter na ta način izboljša poizvedbo - povratna informacija o relevantnosti

Naloga algoritma za izboljšanje je poiskati dokumente, ki so čim bolj podobni relevantnim rezultatom. To oceno je mogoče izboljšati v povratni zanki z več

iteracijami. Primer algoritma je Rocchio.

Povratno zanko relevantnosti je mogoče uporabiti tudi za izboljšanje poizvedb brez interakcije z uporabnikom - psevdo ali slepa povratna informacija o relevantnosti. V tem primeru uporabimo prevzeto metodo iskanja, predpostavimo, da je  $K$  najvišje uvrščenih dokumentov relevantnih. To je boljše kot če je uporabnik vključen v postopek.

## 5.8 Vrednotenje sistemov za iskanje

Pomaga razumeti, kateri pristopi delujejo in kateri ne. Za vsako poizvedbo pregledamo rezultate in preštejemo število dokumentov, ki so bili pričakovani (True Positive) in tistih, ki niso bili (False Positive). Pregledamo tudi TN in FN. Glede na to opredelimo dve glavni merili:

- natančnost: odstotek ustreznih dokumentov med pridobljenimi dokumenti, tj.  $TP / (TP + FP)$
- priklic: Odstotek vrnutih ustreznih dokumentov glede na vse ustrezne dokumente, tj.  $TP / (TP + TN)$ .

Ponavadi lahko izboljšamo le eno merilo. Da obe merili povzamemo v enem rezultatu, uporabimo  $F$  merilo, kar je harmonična sredina.

Za razširitev vrednotenja z neurejenih rezultatov poizvedb na rezultate poizvedb, razvrščene po pomembnosti, uvedemo prag podobnosti dokumentov (kako podoben mora biti dokument poizvedbi, da se šteje za relevantnega).

Drug način za pregled uspešnosti sistema je uporaba ROC krivulje glede na prag sprejemljivosti - prikazuje stopnjo resničnih pozitivnih rezultatov glede na stopnjo lažno pozitivnih. Najboljša točka delovanja sistema je točka na krivulji z najmanjšo razdaljo do točke  $(0, 1)$ .

## 5.9 Iskanje slik

Številni sistemi se zatekajo k besedilnim metapodatkom, ki spremljajo sliko ali uporabniškim opombam. Velika težava je tudi zaznavanje ustreznosti. Za vsako vneseno sliko v podatkovno zbirko izluščimo visokodimenzijski vektor njenih značilnosti, bodisi osnovnih bodisi bolj zapletenih. Tudi poizvedba je predstavljena kot slika, ki se skrči na vektorski opis. Nato se vse slike primerjajo z vektorjem poizvedbe in razvrstijo glede na podobnost.

## 5.10 Primitivni slikovni deskriptorji

Primitivni opisi vsebine slike so zelo blizu pikslom, ki jo sestavljajo, govorimo o barvi posameznih pikslov, lokalni teksturi, obliki. Cilj je te modalnosti nizke ravni prenesti v vektorski prostor, da bi lahko primerjali posamezne dokumente.

Ena najpreprostejših predstavitev na nizki ravni je porazdelitev barve. Barvo slike je mogoče opisati kot porazdelitev v barvnem prostoru, bolj splošno pa jo

opišemo z barvnim histogramom, neparametrično predstavitevijo porazdelitve, ki jo je mogoče preprosto interpretirati kot vektor z normalizacijo histograma (tako da ni odvisen od velikosti slike). Barvni histogrami so tudi robustni, saj so nespremenljivi glede na spremembo velikosti ali rotacijo ter delne okluzije. Slabost barvnih histogramov je, da ne vsebujejo nobenih prostorskih informacij, zaradi česar so neuporabni za kodiranje odnosov. Barvni histogrami so občutljivi tudi na spremembe osvetlitve. To je mogoče do neke mere odpraviti z uporabo barvnega prostora, v katerem je barva ločena od svetlosti, in primerjanjem samo kromatičnih komponente.

Tekstura je še ena lastnost, ki jo je mogoče enostavno pretvoriti v vektorski opis. Zaradi pomanjkanja natančne opredelitve teksture obstaja veliko načinov za njen opis. Vemo, da je tekstura ohlapno opis prostorske razporeditve barv ali intenzivnosti na sliki ali izbranem območju slike, kar pomeni, da mora deskriptor teksture na nek način odražati to razporeditev. Drugo vprašanje v zvezi s teksturo je raven podrobnosti. Če pogledamo prizorišče dovolj natančno, lahko ugotovimo, da je večina atributov, ki veljajo za teksturo, v resnici lastnosti oblike. Teksturo lahko opišemo s prostorskimi razmerji posameznih pikslov, s frekvenčno analizo in z uporabo zaznavnih lastnosti "visoke ravni", kot so periodičnost, grobost, prevladujoča usmerjenost in kompleksnost.

- Matrika sovpadanja je matrika, ki kodira, kolikokrat se piksel z vrednostjo  $V_1$  pojavi poleg piksla z vrednostjo  $V_2$  v določenem vnaprej določenem razmerju. To matriko je mogoče izračunati za območja različnih velikosti, kar pomeni, da je slika pravzaprav zbirka matrik cooccurence. Te matrike so lahko precej velike, zato se slike običajno najprej kvantizirajo na manjše število vrednosti na piko. Za vsako matriko se lahko izračunajo različne lastnosti, npr. energija, entropija, kontrast, homogenost, korelacija. Te količine, vzete iz več matrik, nato sestavljajo opisni vektor.
- Lokalni binarni vzorci so še en deskriptor teksture na nizki ravni. V tem primeru vsaka piksla dobi vrednost LBP na podlagi tega, katera od njenih osmih sosednjih pikslov ima višjo ali nižjo vrednost od nje same (osem sosedov pomeni, da je binarna vrednost, dobljena s primerjavo, spet 8-bitna). Te vrednosti LBP se nato povzamejo podobno kot pri barvi s histogramom.
- Avtokorelacija je merilo samopodobnosti slikovne zaplate, pri čemer premikanje pikslov in njihova primerjava z izvirno zaplato z uporabo normaliziranega skalarnega produkta daje odziv, ki lahko jasno pokaže, ali zaplata vsebuje ponavljajočo se strukturo, saj samopodobnost premaknjenih zaplat povzroči več močnih vrhov.
- Fourierova transformacija kodira sliko kot niz sinusnih osnovnih funkcij. Pri deskriptorjih teksture nas zanima predvsem energija frekvenčnega spektra. Značilnosti, ki jih pridobimo iz spektra, so običajno vsote območij v frekvenčnem prostoru, tj. značilnosti so občutljive na določene dele spektra.

Oblika predmetov je tesno povezana s konceptom teksture. Oblike so sprva

robovi ali binarne maske, ki so kodirane s številčno predstavitvijo, kot so momenti ali diferencialne kode. V nekaterih primerih lahko deskriptorje oblike primerjamo z uporabo običajne vektorske podobnosti, druga možnost pa je primerjava z uporabo transformacijske razdalje, npr. količine transformacije, ki je potrebna za pretvorbo oblike poizvedbe v dano obliko.

Ker predstavitev na nizki ravni, običajno kodirane kot histogram, ne posredujejo dovolj dobro prostorskih informacij, se slika razdeli na več regij, za vsako regijo posebej se izračuna lokalni histogram, histogrami pa se združijo z uporabo konkatenacije. Naprednejša oblika te tehnike se imenuje prostorska piramida; celoten opis histograma slike se združi z več ravnmi podrazdeljenih opisov histogramov slike. To je kompromis, ki kodira tako prostorske kot tudi globalne lokacijsko invariantne informacije.

### 5.11 K semantičnemu opisu

Velik korak k bolj redkemu in semantičnemu opisu slik je bil narejen z uvedbo redkih lokalnih značilnosti, ki so združene v grozde, da združijo svojo varianco videza s fiksno velikostjo slovarja. Pristop se na splošno imenuje vreča besed, lokalne značilnosti se zaznajo in dodelijo ustreznim besedam. Nato nam pogostost teh besed na sliki da deskriptorski vektor, ki ga lahko uporabimo podobno kot pri iskanju besedila. Prednost tega pristopa pred primitivnimi značilnostmi nizke ravni je, da nam proces pridobivanja (učenja) vizualnih besed daje osnovne semantične informacije, pojavijo se nekateri osnovni pogosti deli (npr. kolesa na avtomobilu), ker se pogosto pojavljajo na učnih slikah.

Eden od zgodnjih pristopov z vrečko besed je uporabljal odkrivanje stabilnih območij, npr. vogalov in kapljic, ter opisoval njihovo normalizirano (na rotacijo in merilo invariantno) lokalno okolico z deskriptorjem SIFT (Scale invariant feature transform). Ta deskriptor razdeli regijo na 16 podregij in izračuna histograme kvantiziranih histogramov robov v vsaki od njih za 8 orientacij; skupni histogram ima standardno velikost 128 binov ( $16 \times 8$ ) in je normaliziran. Za pridobitev slovarja besed se uporablja nenadzorovan pristop, deskriptorji se pridobijo iz niza učnih slik in se grupirajo z uporabo grupiranja K-means. Tako nastane K prototipov besed, ko je slika predstavljena sistemu, se značilnosti pridobijo in ujemajo z najbližjo besedo. Nato se besede preštejejo v obliki K-dimenzionalnega histograma.

Drug pristop k oblikovanju semantičnih značilnosti je hierarhično združevanje značilnosti robov, zaznanih z Gaborjevimi filtri. Ti elementi se naučijo na podlagi njihove sočasnosti v naboru podatkov za učenje. Na ta način se vzpostavi večplastna hierarhija, ki lahko (na zgornji plasti) opiše zelo zapletene dele. Ker je število delov spet končno, se lahko histogram teh lastnosti izračuna na zgornji ali več ravneh.

Zadnji velik konceptualni preskok pri semantičnem iskanju slik je prišel v obliki učenja od konca do konca, kar pomeni, da se tako značilnost kot klasifikator hkrati usposobita v enem samem okviru, ki neposredno obdeluje slikovne pike

in izpiše odločitev o sliki na visoki ravni. Takšno učenje je zahtevalo konceptualne izboljšave, velike količine podatkov in dovolj zmogljivo strojno opremo. Konceptualne izboljšave, ki so bile potrebne za ta preskok, so prišle v obliki naprednih in učinkovitih umetnih nevronske mreže, ki so bile dobro znana metoda že od šestdesetih let prejšnjega stoletja in imajo korenine v bioloških sistemih. Pred kratkim so bile te metode nadgrajene s konvolucijskim pristopom k porazdelitvi uteži nevronov, ki je zaradi prostorske invariance in zmanjšanja števila parametrov izboljšal učinkovitost pri nalogah razvrščanja slik. Konvolucijska nevronska mreža je sestavljena iz niza slojev, od katerih vsak obdeluje odzive prejšnjega sloja, zgornji sloj pa je podvržen stroškovni funkciji, ki primerja njegove odzive z želenimi vrednostmi, ki običajno predstavljajo koncept na visoki ravni, npr. kategorijo. Usposabljanje takšne mreže je omejeno na povratno razširjanje napak stroškovne funkcije navzdol po plasteh, pri čemer se popravijo uteži, tako da bi te ustvarile odziv, ki bi bil bližje želenemu. Da bi bil proces učenja učinkovit, obstaja več dodatnih slojev, ki obdelujejo odzive na druge načine, npr. max-pooling, dropout, zaradi kratkosti tega opisa pa o njih ne bomo govorili. Konfiguracija in vrstni red teh plasti se imenujeta arhitektura omrežja. Ko takšno omrežje dovolj dolgo treniramo, začnejo biti uteži podobne lokalnim semantičnim delom.

Prvi primer uporabe konvolucijskih nevronske mreže je bilo kategoriziranje in zaznavanje predmetov, vendar se lahko odzivi funkcij na visoki ravni uporabljajo tudi kot vektorji v sistemih za iskanje. Poleg tega se lahko omrežje za odkrivanje več predmetov uporabi za indeksiranje slik z uporabo semantičnih besed na podlagi vizualnih kategorij, npr. z odkrivanjem predmetov na sliki in njihovo uporabo v scenariju Booleovega iskanja za pospešitev iskanja, indeks kategorij zagotavlja začetno podmnožico kandidatnih slik, ki se nato razvrstijo na podlagi vektorske podobnosti. V zadnjem času se pristopi globokega učenja uporabljajo tudi v še ambicioznejših scenarijih iskanja, npr. pri opisovanju prostorskih odnosov predmetov in opisovanju prizorov kot celote.

## 5.12 Segmentacija pri iskanju

Eden od pomembnih vidikov semantičnega iskanja slik je pravilna dekompozicija slik. Če iščemo slike z določeno semantično kategorijo, je treba celotno sliko opisati z z enim samim deskriptorjem to nalogo otežuje šum v deskriptorju (ozadje, druge predmeti). Opišemo lahko samo dele slike, vendar ne vemo vnaprej, koliko ali oblike regij. Kot smo omenili v prejšnjem razdelku, lahko uporabimo objekt detektorje za zaznavanje predmetov določenih kategorij, vendar lahko to uporabimo le za vnaprej usposobljene kategorije.

Segmentacija razgradi slike v niz regij. Delitev regij lahko temelji na na podlagi preprostih lastnosti na nizki ravni ali pa imajo nekatere semantične lastnosti. Klasični pristop k segmentaciji uporablja tehnike nenadzorovanega učenja za določitev delitve, slika pa je obravnavana kot množica elementov z različnimi lastnostmi, ki so združeni v skupine. Ključna ideja je opisati piksele ali zaplate z vektorji atributov, ki jih je nato mogoče uporabiti pri grozdenju. Ti lastnosti

so lahko zelo preproste, npr. barva piksla, ali bolj zapleteni deskriptorji lokalne teksture.

- barva - vsaka piksla je predstavljena kot točka v barvnem prostoru, barve se nato združijo v grozde.
- kookurenca - matrika kookurence je bila opisana že v prejšnjem razdelku, lokalna se lahko uporabijo tudi vektorji lastnosti matrike
- tekstoni - tekstoni so deskriptorji, ki se jih naučimo iz podatkov, vsaka piksla je opisana z odzivi na banko filtrov (npr. 24 filtrov). Nato se vektorji odzivov združijo v grozde, pri čemer se grozdi nato predstavljajo posamezne tekstone (podobno kot besede v ideji vreče besed). Lokalno teksturo lahko opišemo s histogramom tekstonov - ti histogrami so opisi, ki se združijo v grozde za segmentacijo.

Pri delu z informacijami o videzu je malo verjetno, da bodo grozdi prostorsko koherentni, če je to je treba informacije o lokaciji kodirati kot del opisa. To je pri pretirani segmentaciji, tj. pri superpikslih, kjer gre za razdelitev slike na majhne lokalne enote več pikslov, ki sledijo strukturi slike in se lahko nato uporabijo za obdelavo (npr. segmentacijo) namesto surovih pikslov.

Za grupiranje značilnosti za segmentacijo se lahko uporabi več metod. Na splošno vprašanje, ali število grozdov poznamo vnaprej ali pa se mora metoda odločiti o njihovem številu. na podlagi nekega parametra praga podobnosti. Nekatere pomembne metode so: Središča grozdov so naključno izbrana in se jih ne da razporediti. inicializirani, vektorji funkcij so dodeljeni najbližjemu grozdu. Nato se središča ponovno izračunajo kot povprečje vseh dodeljenih vektorjev značilnosti. Ta pristop konvergira, če je merilo razdalje evklidsko.

- sredinski premik: grozd sestavljajo vsi vektorji značilnosti, ki konvergirajo k istemu modusu. z uporabo iskanja načina srednjega premika. Število grozdov se določi samodejno, vendar je jedro je treba navesti pasovno širino in vrsto. Jedro določa polje privlačnosti, tj. območje v prostoru, kjer imajo vse točke enak modus. Žal algoritem ne meri dobro na veliko število dimenzij.
- širjenje sorodnosti: ključna značilnost algoritma je, da lahko deluje s poljubnim merilom podobnosti (sorodnost) med značilnostmi. Ni nujno, da je sorodnost v skladu s trikotnikom neenakosti in niti ni nujno, da je opredeljena za vse pare lastnosti. Število grozdov se določi samodejno, grozdi so opredeljeni z vzorčnimi značilnostmi in ne z s povprečnimi vektorji lastnosti. Algoritem deluje iterativno s sporočili med vozlišči v grafu.

Doslej opisana segmentacija bi teoretično lahko segmentirala slike v različne predmete, če bi bili njihovi videz je bil različen, vendar so bili segmenti še vedno le skupine podobnih pikslov. V spletnem mestu semantični segmentaciji so skupinam že dodeljene semantične kategorije. Najpogostejše pogled na semantično segmentacijo je kategorizacija po pikslih, tj. vsakemu pikslu je dodeljena kategorijo. Klasičen primer tega je uporaba besedilnih značilnosti skupaj s klasifikatorjem, ki je usposobljen za razvrščanje histogramov tekstonov (vreča

tekstonov) v semantične kategorije.

Tako kot pri klasični kategorizaciji predmetov se je tudi pri semantični segmentaciji bistveno izboljšalo uporabo konvolucijskih nevronske mreže. Namesto usposabljanja za eno samo odločitev na najvišji ravni, so te mreže zasnovane za klasifikacijo na piksel. Ključna zamisel je, da se ne uporablja nobena popolnoma povezanih slojev ali obravnavati značilnosti slike kot vektorje, temveč ohraniti prostorske informacije, dokler se ne do vrha omrežja. Zaradi tega se segmentacijske mreže imenujejo tudi popolnoma konvolucijske.

Vprašanje, ki ga je treba obravnavati v zvezi s semantično segmentacijo s konvolucijskimi nevronskimi mrežami je, kako doseči segmentacijo polne ločljivosti, če so sodobne arhitekture mrež močno zanašajo na plasti združevanja, kjer je celotno odzivno polje podvzorčeno (zmanjšano v prostorski ločljivost) in se ohrani le največji odziv v lokalnem območju. Ti sloji zmanjšujejo število parametrov in povečajo prostorsko robustnost. Privzeta maska za segmentacijo je zato zelo groba. Preprosta rešitev za to je uporaba interpolacije, da jo povečamo in morebitno uporabo naknadne obdelave, kot je Markovo naključno polje, za izboljšanje lokalne natančnosti. Novost rešitve vključujejo tudi razširjene konvolucijske plasti, ki ne zmanjšujejo ločljivosti, hkrati pa še vedno in dekonvolucijske plasti, ki dejansko povečajo prostorsko ločljivost, a hkrati povečajo prostorsko robustnost, ter plasti dekonvolucije, ki dejansko povečajo prostorsko ločljivost in povečajo prostorsko robustnost. Plasti dekonvolucije se uporabljajo v arhitekturah kodirnik-dekoder, kjer se konvolucijske plasti najprej kodirajo vizualne informacije v vektorje z visoko razsežnostjo, vendar nizko ločljivostjo; dekonvolucijski sloji nato dekodirajo te informacije iz visokodimenzionalnega prostora funkcij v kategorizacijo na piksel.

### 5.13 Iskanje v videu

Konceptualno je lahko iskanje v videu zelo podobno iskanju v posameznih slikah. Vsaka slička je dokument, medtem ko je celoten videoposnetek korpus. Poleg tega videoposnetek ne vsebuje le fotografij predmete in kategorije predmetov, ampak tudi kratkoročna dejanja in dolgoročna časovna razmerja. med entitetami. Trenutno so tovrstni semantični podatki še vedno predmet raziskav in niso v veliki meri uporabljali pri iskanju.

Ker je videoposnetek sestavljen iz številnih precej podobnih kadrov, je veliko bolj učinkovito, če samo indeksiramo podmnožico reprezentativnih kadrov, ki jih imenujemo ključni kadri. Ključne kadre je mogoče odkriti na podlagi vnaprej določenih meja posnetkov, tako da vzamemo njihov prvi, zadnji ali srednji okvir. Lahko pa tudi pridobljeni iz "surovega" videoposnetka z uporabo grozdenja ali podvzorčenja. Da bi zagotovili časovno skladnost, značilnosti, zaznane v ključnih posnetkih, lahko spremljamo v več posnetkih, da zagotovimo njihovo skladnost. Prostorsko doslednost je mogoče uporabiti tudi v primerih, ko je določen predmet iščejo (elementi morajo biti prikazani blizu drug drugega).

## 5.14 MPEG-7

V nasprotju z večino drugih standardov MPEG standard MPEG-7 ne opisuje video in avdio kodekov, temveč standard za večpredstavnostne vsebine. Dopolnjuje standard MPEG-4 in je namenjen omogočanju učinkovitega dostopa do večpredstavnostnih vsebin in manipulacije z njimi s standardizacijo elementov iskanja objektov brez besedila. Zato standardizira naslednje komponente:

- deskriptorje objektov - kaj in kako se lahko opiše
- opisne sheme - struktura in semantika odnosov med sestavnimi deli
- opisno poizvedbo - kako iskati vsebino

Video in avdio vsebino je mogoče opisati na ravni objektov, ki so v določenem semantičnem razmerju. Glavna težava te zamisli je, da standard ne obravnava problema semantične vrzeli, velikega razkoraka med vizualnimi značilnostmi na nizki ravni in smiselnimi koncepti na visoki ravni, s katerimi operirajo ljudje.

## 6. INTERAKTIVNOST

### 6.1 Povečana resničnost

Svet zaznavamo s čutili: vidom, sluhom, vonjem, otipom in ravnotežjem. Ker vplivanje na čutne informacije spreminja naše razumevanje sveta, lahko rečemo, da je resničnost subjektivna. Cilj razširjene resničnosti je razširiti čutne informacije tako, da združene informacije koristijo našemu razumevanju sveta (ali da nudijo kakšno drugo vrednost, npr. zabavo).

Čeprav imamo ljudje več čutil, se v prvi vrsti zanašamo na vid, morda tudi zato je razširjena resničnost najbolj razvita v tej smeri. Osnovni sistem vizualne razširjene resničnosti mora v realnem času določiti položaj kamere in zaslona v prostoru, da ohrani iluzijo, da so razširjene informacije vstavljene v ta prostor. Obstaja veliko možnosti, kako to doseči samo s kamero, od računsko manj intenzivnih in robustnejših do bolj zapletenih, a tudi bolj splošnih. Poleg tega lahko sistemom za določanje položaja s kamero pomagajo tudi drugi senzorji, kot so GPS, WiFi za določanje položaja, IMU.

Za določanje položaja kamere v prostoru moramo določiti njene zunanje parametre: rotacijo in translacijo. To lahko dosežemo z reševanjem sistema linearnih enačb, ki vključuje referenčne podatke o točkah v svetovnem koordinatnem sistemu in njihovih projekcijah v slikovnem prostoru, te pare imenujemo korespondence. Pridobivanje zanesljivih korespondenc je na splošno lahko težavna naloga, zato obstajajo različni pristopi, ki bodisi povečajo zanesljivost bodisi znajo obravnavati potencialno nezanesljive korespondence.

### 6.2 Uporaba markerjev

Najpreprostejša oblika vizualnega določanja položaja, ki jo bomo obravnavali, temelji na umetno ustvarjenih črno-belih oznakah, ki jih je mogoče prepoznati



z osnovno obdelavo slik. Najpogosteje so ti označevalci kvadratni s črnim robom in identifikacijsko notranjostjo, ki omogoča njihovo prepoznavanje in tudi razločevanje rotacije. En tak označevalec zadostuje za oceno položaja kamere, vendar jih je mogoče uporabiti več za izboljšanje odpornosti na okluzijo in smer pogleda.

Težava pri uporabi označevalcev (poleg dejstva, da jih je treba vstaviti v prizorišče) je, da algoritmi, ki se uporabljajo za njihovo zaznavanje, niso odporni na okluzijo.

### 6.3 Uporaba predlog

Namesto umetnih označevalcev lahko za določanje položaja kamere uporabite tudi katero koli teksturirano površino. Koncept prav tako temelji na zaznavanju ujemanj, vendar je možnost, da so v nizu napačna ujemanja, veliko večja. Ustreznosti se zaznajo z ujemanjem značilnih točk (njihov opis lokalnega območja) z referenčne predloge slike s posnetkom kamere za poizvedbo. Za zanesljivo oceno zunanjih parametrov kamere iz takega nabora se uporabi algoritem RANSAC.

### 6.4 Grajenje zemljevida scene

Najnaprednejši pristop za določanje položaja kamere, ki ne zahteva vnaprej določenih označevalcev, je gradnja 3D reference iz dejanskega prizora, ki mu sledimo. To se imenuje tudi VSLAM (Visual simultaneous localization and mapping) in se lahko poleg razširjene resničnosti uporablja v številnih scenarijih (mobilna robotika, 3D rekonstrukcija). Načeloma je tovrstni pristop računsko drag, saj je treba posodabljati tako lokacijo kamere kot zemljevid. Pristop je prilagojen scenarijem razširjene resničnosti v realnem času v algoritmu PTAM (Parallel tracking and mapping), ki uporablja dve niti, eno za odzivno lokalizacijo kamere v realnem času in drugo za počasnejše posodabljanje zemljevida.

### 6.5 Interaktivne površine

Od prihoda prvega telefona iPhone je tehnologija dotika postala dejanski način interakcije z večpredstavnostnimi sistemi. Ogledali si bomo tehnologije, ki omogočajo delovanje vmesnikov na dotik, in načine uporabe različnih nastavitev v različnih scenarijih uporabe. Interaktivno površino sestavljata dve komponenti: zaslon in senzor. Ti dve komponenti sta lahko integrirani skupaj (npr. pametni telefoni) ali nameščeni na ustreznih mestih (npr. projektor in globinska kamera).

Oblikovanje namiznega vmesnika na dotik predstavlja več izzivov, od ergonomskih vprašanj (velikost in položaj površine, lahko povzroči obremenitev vratnih mišic ali težave s hrbtom) do vprašanj uporabnosti (vidnost in dosegljivost v scenarijih z več uporabniki, primeri uporabe, dodana vrednost).

## 6.6 Uporni senzorji

Uporni senzorji so izdelani iz dveh prozornih plasti, prevlečenih s prozornimi prevodnimi snovmi, in izolacijske plasti med obema plastema. Sila, ki deluje na zunanjo plast, stisne izolacijo in jo približa nasprotni plasti, kar je mogoče zaznati kot spremenjeno upornost sistema. Krmilnik izmenično preklaplja plasti, pri čemer na eni poganja električni tok in meri tok na drugi plasti. Na ta način se zaznavanje vodoravnega in navpičnega položaja združi v celovito informacijo. Uporni senzorji so poceni in imajo majhno porabo energije, vendar lahko zmanjšajo kakovost prikaza prekritega zaslona.

## 6.7 Kapacitivni senzorji

Obstajata dva načina uporabe kapacitivnih senzorjev. Pri površinskih kapacitivnih senzorjih se na stekleno ploščo nanese enakomerna prozorna prevodna prevleka, elektrode pa so nameščene v vsakem kotu plošče. Elektrodi ustvarjata enakomerno električno polje na prevodni plasti. Ko se prst (ali drug prevodni predmet) dotakne površine, električni tok teče iz štirih vogalov skozi prst. Meri se razmerje električnega toka, ki teče iz štirih vogalov, da se zazna točka dotika. Izmerjena vrednost električnega toka bo obratno sorazmerna z razdaljo med točko dotika in štirimi vogali. Ta pristop ima visoko položajno natančnost, vendar je težko zaznati več dotikov.

Alternativni pristop se imenuje projicirani kapacitivni senzor. V tem primeru je površina razdeljena na senzorsko mrežo elektrod, pri čemer vsaka elektroda pokriva le majhen del površine. Ta pristop omogoča natančno zaznavanje večkratnih dotikov in ima visoko položajno natančnost, vendar ni primeren za velike plošče zaradi počasnejšega prenosa električnega toka na velike razdalje. Vendar je to prevladujoča tehnologija v sodobnih mobilnih telefonih in prenosnih računalnikih z možnostjo dotika.

## 6.8 IR kamere

Ker senzorji, ki temeljijo neposredno na električnih pojavih, niso zelo primerni, so raziskovalci uporabili druge načine za zaznavanje dotika, predvsem z uporabo optike. Običajen pristop je uporaba infrardeče svetlobe, saj jo je mogoče zaznati z (namensko) kamero in ne moti našega vidnega zaznavanja. Pristop razpršene osvetlitve je neposreden primer tega, polprozorna projekcijska površina je od zadaj osvetljena z viri IR-svetlobe. Če na površino položimo prst (ali kateri koli drug predmet), se IR-svetloba od njega odbije in jo zazna IR-kamera.

FTIR (Frustrated Total Internal Reflection) izboljša zanesljivost dotika, saj odpravlja napačne zaznave dotika, kadar je prst le blizu površine. IR-svetloba se uporablja s strani akrilne plošče in se zaradi TIR (totalnega notranjega odboja) ujame v materialu. Ta učinek se prekine, ko se predmet dotakne plošče s strani, zaradi česar se svetloba od dotikajočega se predmeta odbije v stran in jo kamera IR ponovno zazna.

## 6.9 Globinske kamere

Še en pristop, ki uporablja optiko in se vse pogosteje uporablja za HCI, so globinske kamere. Večina kamer na nek način uporablja IR-svetlobo za oceno globine s senzorja, bodisi iz projicirane vzorca ali z uporabo časa preleta. Druga možnost je uporaba stereokamere. Zaznavanje dotika z uporabo informacij o globini je računsko zahtevnejša. Algoritem mora zaznati površine, roke in prste. Vendar pa ta pristop ponuja tudi dodatne možnosti HCI, ki presegajo prste zaznavanje dotika prstov, npr. prepoznavanje gest v 3D prostoru nad površino interakcije in zaznavanje predmetov.