

# Adaptivita pri kompresii a prenose obrazu a videa

## Motivácia

- prispôsobenie sa na špecifické vlastnosti prenosovej cesty
- prispôsobenie sa na schopnosti zobrazovacieho zariadenia
- preferenciami a požiadavkami konzumenta obsahu
- ...

## Cieľ

- spravidla maximalizácia koncovej kvality obsahu vzhľadom na možnosti prispôsobenia sa.
- Napr. z pohľadu
  - prevádzkovateľa prenosovej cesty - maximálna kvality prenášaného obsahu pri stanovenej záťaži siete, resp. jej jednotlivých častí.
    - Môže sa aj jednať o optimalizáciu (prispôsobenie sa) pri meniacich sa podmienok prenosovej cesty: mení sa chybovosť, mení sa latencia, a podobne.
    - Alebo sa mení samotná prenosová cesta (a jej poskytovateľ) v pozadí - prepnutie medzi mobilným pripojením, wifi a pevných pripojením.
  - správcu obsahu sa môže jednať o
    - minimalizáciu veľkosti úložiska.
- Z pohľadu užívateľa
  - maximalizácia kvality obsahu určeného jemu a prispôsobeného vzhľadom na jeho osobné požiadavky a zariadenie, ktoré používa.

## **Multimédia - vyhodnocovanie kvality – ako na to?**

- nie je jednoznačné – hoci čiastočné štandardy existujú ... (MOS, Video MOS, ...)
- je horšie video, kde je lepší obraz a horšie audio, alebo naopak?
- je kvalitnejšie do, čo ma mierne zníženú kvalitu dlhší čas, alebo veľmi zníženú kvalitu kratší čas?

## **Aké máme možnosti ovplyvniť kvalitu, napr. pri videu ?**

- Napr. pri zmenšení prenosovej rýchlosti kanála môžeme
  - zníženie snímkovej frekvencie, zachovanie ostrosti obrazu, zachovanie kvality audia
  - zachovanie snímkovej frekvencie, vznik artefaktov v obraze, zachovanie kvality audia
  - vypnutie audia a investovať ušetrené bity do obrazu
- Oficiálna kategorizácia:
  - výberom obsahu - vyberajú sa iba relevantné časti (vzhľadom na podmienky, preferencie, atď) multimediálneho obsahu
  - konverzia modality - využívana sa konverzia jednej zložky multimediálneho signálu na druhý (napr. text na reč)
  - škálovanie/transkódovanie obsahu - vykonáva sa napr. zmena formátu obsahu na podporovaný zariadením, obsah je doručovaný iba v potrebnom rozlíšení a podobne ...

Sprievodné závislosti od obsahu, iné správanie konzumenta pri:

- priamom prenose zo športového zápasu
- pri sledovaní akčného filmu
- pri sledovaní televízneho hlásateľa pri predpovedi počasia
- pri sledovaní koncertu

# Možnosti adaptácie v JPEG

Už ten špecifikoval 2 módy vhodné na využitie v procese adaptácie obsahu:

- progresívny mód ([JPEG] v dodatku G): informácie o obraze sú kódované postupne a to tak, že po prijatí časti informácií tieto predstavujú celý obraz, ale s menej detailami, po prijatí ďalších dávok informácií je obraz postupne vylepšovaný. Pritom platí, že najdôležitejšie informácie, t.j. tie, ktoré najviac zmenšujú chybu rekonštrukcie sú presúvané na začiatku a postupuje sa k menej dôležitým informáciám. Toto sa môže diať dvomi spôsobmi
  - volením koeficientov (spectral selection) - spektrálne koeficienty sú prenášané od dôležitejších k menej dôležitým
  - postupnou aproximáciou (succesive approximation) - hodnoty koeficientov sa prenášajú od najdôležitejších bitových rovín po najmenej dôležité.
- Hierarchický mód ([JPEG] v dodatku J) umožňuje lepšie prispôsobenie rozlíšeniu zobrazovacieho zariadenia. Obrázok je prenášaný tak, že najprv sú prenesené informácie potrebné pre rekonštrukciu zmenšenej verzie obrazu, potom postupný zvyšok informácií, ktoré umožnia obrázok kvalitne zobrazíť postupne pri väčšom a väčšom rozlíšení.

# JPEG2000

- zaviedol paketizáciu
- zaviedol koncept vstiev kvality ("quality layers")

## Audio

- adaptivita obsahu napr. v telekomunikačných systémoch. Napr. AMR (adaptive multi rate) kodek kóduje 20ms úseky rečového signálu, reprezentujúce 160 vzoriek signálu a v závislosti od aktuálnych podmienok v mobilnej sieti sa volí verzia údajov s primerane silným kódovaním.
  - pri dobrých prenosových podmienkach sa volí verzia, ktorá umožní preniesť maximum užitočných informácií avšak slabochránených proti chybám, pri zlých podmienkach sa zvolí verzia prenášajúca málo užitočných informácií, avšak oveľa odolnejšia voči prípadným chybám.

## MPEG-2

- Zaviedol pojem škálovateľnosti pri videu
- obrazový stream je kódovaný v dvoch vrstvách
  - v prvej, základnej vrstve je prenášaný základ, ktorý môže byť ďalej vylepšený
  - doplnkovou vrstvou, ktorá obsahuje dodatkové informácie k dosiahnutiu plnej kvality videa.
- Základné metódy škálovania
  - škálovanie na základe SNR (Signal to Noise Ratio): doplnková vrstva posiela informácie, ktoré vylepšujú kvalitu videa pri danom rozlíšení

- priestorové škálovanie: základná vrstva kóduje video pri nižšom (polovičnom) rozlíšení a doplnková vrstva kóduje rozdielové informácie potrebné na zobrazenie v plnom rozlíšení.
- Štandard MPEG-2 nepodporuje škálovateľnosť audia.

## MPEG-4

- prináša škálovateľnosť audia vo forme MPEG-4 SLS (Scalable Lossless Coding)
  - umožňuje kódovať zvyškovú informáciu ako ďalšiu vrstvu a dosiahnuť až kvalitu bezstratového kódovania.
- rozširuje škálovateľnosť kódovania obrazovej informácie MPEG-2 tak, že umožňuje využiť viac ako 1 doplnkovú vrstvu.

## MPEG-4 časť 10 /AVC (Advanced video coding)/H.264

- navyše okrem priestorového a SNR škálovania zavádza aj časové škálovanie
- Časové škálovanie je, že v čase existuje viacero sekvencií snímkov a teda v prípade potreby je možné doplnkové sekvencie zahadzovať.
  - Toto má za následok redukciu snímkových frekvencií
- V dodatku G štandardu MPEG-4 AVC po názvom SVC (Scalable Video Coding) sú uvedené doplnkové odporúčania ako škálovateľnosť efektívne využívať.

## MPEG-H časť 2 / HEVC (High efficiency video coding)/ H.265

- prináša škálovateľnosť pre videá až s rozlíšením 8K, teda 8192x4320 bodov
- o desiatky percent zlepšuje kompresiu oproti MPEG-4 AVC.
- Časová škálovateľnosť bola prítomná už v prvej verzii štandardu
- priestorová, SNR škálovateľnosť a škálovateľnosť na základe bitovej hĺbky a gamutu a ich kombinácie boli pridané až v druhej verzii štandardu vo forme SHVC (rozšírenia Škálovateľnosti HEVC) v októbri 2014.
- Pridané nástroje na efektívny prenos
  - 3D (stereoskopické, s mapou hĺbky) videa
  - multiview (MV) videa.

## MPEG-21

- finalizovaný v roku 2003
- definuje otvorený framework pre multimediálne aplikácie
- venuje sa otázke (časť 7) adaptácie obsahu vo forme Digital Item Adaptation (DIA)
- Pojem Digital item zahŕňa nejaký multimediálny obsah spolu s jeho všetkými potrebnými metadatami.
- DIA špecifikuje množinu popisných nástrojov na optimalizáciu adaptácie multimediálneho obsahu.
- Špecifikuje ako by malo byť popisované prostredie pri adaptácii a to popisom charakteristík siete, zariadenia, užívateľských preferencií a okolia užívateľa:
  - schopnosti zariadenia: dostupné kodeky, vstupno-výstupné možnosti (displej, audio, mikrofón, ...)

- charakteristiky siete: schopnosti siete (max. kapacita, minimálna garantovaná šírka pásma, ...), podmienky - oneskorenie, jitter, chybovosť, dostupná šírka pásma
- charakteristiky užívateľa: užívateľské preferencie, história používania, všeobecné informácie o užívateľovi, preferencie na prezentovanie jednotlivých modalít - napr. pri audio preferovaná hlasitosť a nastavenie ekvalizátora, pri obraze rozlíšenie, teploty farieb, saturácia, kontrast.
- prístupnosť: charakteristiky pre ľudí s postihnutím napr. vizuálneho systému (napr. stupeň a typ poruchy farebného videnia) sluchového systému (obmedzenie sluchu pri rôznych frekvenciách)
- fyzické okolie užívateľa: miesto a čas používania, aududiovizuálne vlastnosti prostredia (napr. úroveň a charakteristika hluku), úroveň osvetlenia
- Štandard sa venuje aj škálovateľnému obsahu ako MPEG-4 SVC a poskytuje generický nástroj na popis syntaxe bitového streamu a použitie tohoto popisu na vytvorenie adaptovenej verzie obsahu.
- Okrem využitia škálovateľnosti popisuje aj možnosti **konverzie modality obsahu (modality translation)**, napr. audio na text a naopak.

# MPEG-DASH

- MPEG-DASH (Dynamic Adaptive Streaming over HTTP),
- štandardizovaný v roku 2012
- popisuje interakciu klienta zo serverom pri adaptívnom doručovaní obsahu cez HTTP protokol
- MPEG-DASH je agnostický od kodeku - môže sa použiť s ľubovoľným moderným kódovacím postupom na kódovanie videa, ako napr. H.264, H.265, VP9 atď.
- Princíp
  - na strane servera je obsah rozdelený na segmenty a každý segment je zakódovaný pri rôznych bitových náročnostiach, resp. rôznym spôsobom optimalizovaný obsah.
  - Informácia o segmentoch a ich verziách je dostupná prostredníctvom t.z.v. MPD (Media Presentation Descriptor) súboru.
  - Klientské zariadenie si najprv stiahne MPD súbor, na základe toho začne sťahovať video, pričom v závislosti od aktuálnych podmienok na prenosovom kanáli môže optimalizovať akú verziu stiahne pre ďalšie segmenty videa.
  - Samozrejme kritérium je aby stihol včas stiahnuť čo najkvalitnejšiu resp. najvhodnejšiu verziu pre konzumenta.
- MPEG-DASH používajú napr. Netflix a Youtube pri distribúcii obsahu aj pri streamovaní priamych prenosov.
- Existujú systémy, ktoré vedia efektívne využiť škálovaťnosť MPEG-4 SVC pri použití MPEG-DASH
- V súvislosti s MPEG-DASH vyvstáva otázka, prečo klient, keď sú dobré podmienky nestiahne celý stream k sebe, aby keď sa zhoršia podmienky prenosu, nemusel zhoršovať kvalitu? Odpovede sú viaceré, napr.:
  - obmedzenie dostupných prostriedkov klienta (pamäť)

- načo sťahovať dáta, ktoré si užívateľ možno ani nepozrie
  - prečo zbytočne zaťažovať prostriedky siete v čase, keď ich možno urgentnejšie potrebujú iní
- MPEG-DASH teda umožňuje riadiť doručovanie obsahu pri rôznych podmienkach prenosového kanálu a optimalizovať kvalitu a prenesený objem dát resp. iné aj kritéria (napr. nepriamo výdrž zariadenia na batériu) z pohľadu jednotlivých užívateľov.
- Optimalizácii rozhodovania v klientovi (ktorú štandard MPEG-DASH) nepredpisuje je v súčasnosti venovaného veľa úsilia, napr. [DASH-LOGIKA]

# DASH-VR

- V MPEG-DASH v súčasnosti prebieha výskum a štandardizačné úsilie ako rozšíriť adaptivitu pre doručovanie VR obsahu (DASH-VR). Predpokladajú sa 3 úrovne:
  - základná úroveň - doručované celé 360°video
  - SRD (spatial relationship description) - poskytuje techniku delenia na oblasti (tiles) aby užívateľ dostával len ten obsah na ktorý sa aktuálne pozerá (predstavuje druhý dodatok MPEG-DASH)
  - SRD kombinované so SVHC - pri kombinácii SRD so škálovateľným rozšírením HEVC sa pri delení videa na priestorové oblasti dá efektívne škálovať kvalita pre jednotlivé oblasti.

# HEIF

- HEIF = High Efficiency Image File
- V súčasnosti najefektívnejší štandard na kompresiu obrazu štandard
- Je to vlastne časť 12 odporúčania MPEG-H.
- Využívajú sa tam algoritmy použité v HEVC pre vnútrostránkovú predikciu (de facto je to vyňatá časť HEVC vhodná na kódovanie statického obrazu).
- Tento štandard by mal byť efektívnym nástupcom klasického JPEG algoritmu, na rozdiel od rozpačite akceptovaného JPEG2000 získal HEIF širokú podporu, napr. natívnou súčasťou OS High Sierra.

# Siete 5G a adaptivita

- Pri komunikačných sieťach 5. generácie je dôraz nielen na rastúcu kapacitu siete, ale aj na to, rastúcou inteligenciou siete túto kapacitu ďalej zvyšovať
- boli navrhnuté viaceré rozšírenia siete, ktoré zavádzajú oproti MPEG-DASH aj inteligenciu na strane siete, napr. použitím vyrovnávacej pamäte (cache), a to čo najbližšie k užívateľom
  - ETSI tento komponent nazýva MEC (mobile edge computing) - typicky sa nachádza na okraji mobilnej siete, t.j. napr. v NODE-B.
  - Takto sa dá zmenšiť záťaž zbytku siete, keďže viac záťažená bude iba časť prístupovej siete medzi cache a koncovými užívateľmi.
- Na realizáciu adaptácie na strane siete sa dá s výhodou použiť architektúra SDN/NFV sieťach (Software-defined Networking / network function virtualization) a v moduloch VNF (virtualised network functions), ktoré môžu byť v pozícii EC (edge computing)
- Poskytovaním výpočtového výkonu na hranici siete sa dá efektívne budovať adaptivita na strane siete v sieťach 5 generácie.

# Videoanalýza – algoritmy na odčítavanie pozadia

- Čo to je? Sú to metódy na detekciu pohyblivých objektov pomocou statických kamier
- Pohybujúci sa objekt = objekt ktorý sa pohybuje voči „pozadiu“
- Čo je pozadie a čo nie? Potrebujeme spraviť odhad pozadia.
- Potom stačí pozadie odčítať od aktuálneho obrazu a máme obraz pohybujúcich sa objektov
- Jednotlivé snímky videa budeme označovať ako  $obr_n(x, y)$ , kde  $n=1...N$  je pozícia snímky.

## Metóda jednoduchkej diferencie [IJCSA2014][PIC2008]

- Založené na jasovej zložke
- aktuálny obrázok je odčítaný od predchádzajúceho
- pre každý pixel je odhadnuté či patrí pozadiu. Pixel  $obr_n(x, y)$  pozadiu nepatrí ak
$$|obr_n(x, y) - obr_{n-1}(x, y)| > T_h$$
- $T_h$  je stanovená prahová hodnota

| Pozitíva                                                                                                                                                                                                                                  | Negatíva                                                                                                                                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>• nízke výpočtové nároky</li><li>• nízke pamäťové nároky</li><li>• extrémne rýchlo sa adaptuje na pozadie</li></ul> (však je aj model pozadia mimoriadne povrchný, je to iba predchádzajúca snímka) | <ul style="list-style-type: none"><li>• citlivý na šum</li><li>• vnútro väčších objektov je často interpretované ako pozadie</li><li>• veľmi citlivé na threshold</li></ul> |

# Metóda s detekciou tieňa (r. 1999) [IJCSA2014]

- Vhodné pre RGB video
- Sledujeme 2 druhy rozdielov
  - jasový
  - chromatický
- metóda odlišujem tieň od pohybujúcich sa objektov od samotného objektu
  - pohybujúci objekt má aj jasovú aj chromatickú zložku význačne rozdielnu od pozadia
  - tieň má podobnú chromatickú zložku ako pozadie, ale jasovú má menšiu

## Poznámka:

- ako získať lepší model pozadia?
- Pozadie ale nesmie byť statické, musí zohľadňovať (ak je to potrebné)
  - zmeny osvetlenia
    - postupné (deň/noc)
    - náhle (mraky, zapnutie/vypnutie osvetlenia)
  - pohyby
    - chvenie kamery
    - pohyb konárov, trávy, vlny, ...
  - zmeny geometrie pozadia
    - zaparkované pozadia na parkoviskách

# Pozadie ako priemer, resp. medián [PIC2008]

## Bez selektivity

- pozadie je tvorené ako priemer resp. medián predchádzajúcich K obrázkov
- veľká spotreba pamäti (K obrázkov v pamäti)
- určuje sa pre každý pixel zvlášť (žiadna priestorová korelácia medzi susediacimi pixelmi)

## So selektivitou

- každý pixel aktuálneho obrázku je klasifikovaný ako v popredí, alebo ako súčasť pozadia.
- Iba pixely, ktoré sú súčasťou pozadia sa použijú na aktualizáciu modelu pozadia
  - model pozadia je tvorený ako priemer resp. medián pixelov pozadia z predchádzajúcich K obrázkov
  - veľká spotreba pamäti (K obrázkov v pamäti)
  - určuje sa pre každý pixel zvlášť (žiadna priestorová korelácia medzi susediacimi pixelmi)

# Pozadie pomocou IIR filtra [PIC2008]

Bez selektivity:

- $B_n$  - je model pozadia pri snímke  $n$
- Aktualizácia modelu pozadia
  - $B_n(x, y) = \alpha obr_n(x, y) + (1 - \alpha)B_{n-1}(x, y)$
- $\alpha$  určuje rýchlosť učenia, typicky 0.05

So selektivitou

- $B_n(x, y) = \begin{cases} \alpha obr_n(x, y) + (1 - \alpha)B_{n-1}(x, y) & \text{ak } obr_n(x, y) \text{ patrí pozadiu} \\ B_{n-1}(x, y) & \text{ak } obr_n(x, y) \text{ nepatrí pozadiu} \end{cases}$

## Gaussovské pozadie [PIC2008]

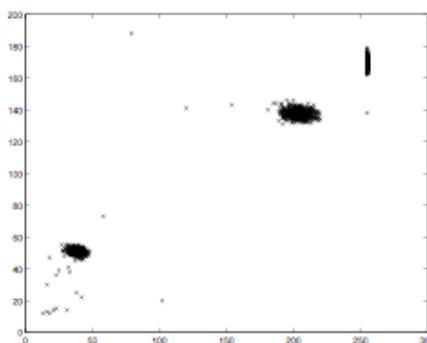
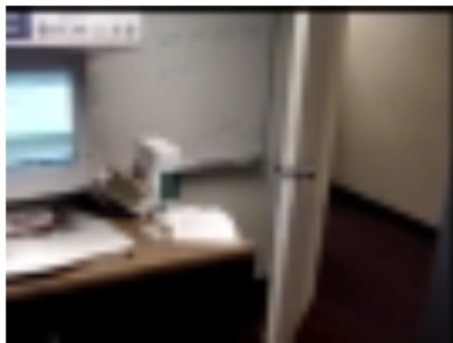
- umožňuje automaticky voliť prah
- okrem priemeru pozadia  $\mu$  sa modeluje aj rozptyl pozadia  $\sigma$ 
  - funguje iba pre unimodálne rozdelenie jasu pozadia v histograme (jeden význačný vrchol)
- $\mu_n(x, y) = \alpha obr_n(x, y) + (1 - \alpha)\mu_{n-1}(x, y)$
- $\sigma_n^2(x, y) = \alpha [obr_n(x, y) - \mu_n(x, y)]^2 + (1 - \alpha)\sigma_{n-1}^2(x, y)$
- pri teste sa testuje, či  $|obr_n(x, y) - \mu_{n-1}(x, y)| > T_h$ , pričom  $T_h = k\sigma$

# Pozadie ako gaussovský mix [PIC2008] [GM] [GM\_FG]

- ak na pozadí sa mení pomaly iba osvetlenie, potom stačí 1 mód aby sa eliminoval šum prítomný z procesu snímania
- čo ak sú ale prítomné, napr. pravidelné odlesky (napr. od vodnej hladiny), vypnuté/zapnuté osvetlenie, blikanie monitora
- multimodálne rozdelenie (3-5 módov)
- mód  $i$  je chrakterizovaný tripletom  $(\mu_i, \sigma_i, \omega_i)$ , kde  $\omega_i$  je váha
- $\mu_i, \sigma_i$  sú aktualizované ako pri 1 gaussovskom pozadí
- $\omega_i$  sú aktualizované pri každom novom obrázku videa
- či pixel patrí nejakému módu je zisťované pomocou  $T_h = 2.5\sigma_i$

## Poznámky:

- Podobne ako pozadie, môže byť aj popredie modelované ďalšími módm



Príklda: Bimodalita prítomná v jase jedného pixelu v priebehu času. Spôsobené blikaním monitora.

*Poznámka: jasne vidíme a môžeme kvantifikovať „neviditeľné veci v obraze“*

# Priestorová korelácia

- Doteraz sme hovorili iba o modeli v čase, každý pixel mal vlastný model, nezávislý od modelov ostatných pixelov
- Toto dáva veľký priestor náhodným chybám
- Náhodné chyby sa dajú do veľkej miery eliminovať, keď blízke body zkorelujeme - spravíme ich závislými, napr. DP filtráciou (toto sme robli aj pri matlab deme)
  - stratíme na citlivosti
  - získame na robustnosti

# Počítačové videnie (computer vision)

- Ciele, napr.
  - Detekcia objektov v obraze (**objekt**=oblasť v obraze, ktorá sa niečím vyznačuje, špecifickými príznakmi)
  - Popis/Identifikácia objektov v obraze
- Jedným zo základov je analýza príznakov v obraze (**predspracovanie, extrakcia, výber/redukcia, klasifikácia**)
  - Príznačky kvantifikujú nejaké vlastnosti obrazu/objektu
    - Nízkoúrovňové príznaky (základné vlastnosti objektov – tvar, farba, textúra, hrany, rohy, ...)
      - Bodové (opisujú jednotlivé body pozíciou, intenzitou, farbou, ...)
      - Globálne (opisujú celý obraz, alebo časti (objekty) vzniknuté segmentáciou (rozdelenie obrazu na segmenty, opisujú napr. pomocou farby, textúry, tvaru ...)
      - Lokálne
    - Strednoúrovňové príznaky (spájajú nízkoúrovňové príznaky a idú do sémantických kategórií – napr. voda, sneh, skala, tvár, postava, ...)
    - Vysokoúrovňové príznaky (sémantika /popis celej scény - napríklad osoba sediaca na pláži)
  - príznaky sa spájajú do vektorov -> **vektor príznakov**
  - objekt sa vyznačuje nejakým vektorom príznakov (**deskriptor objektu**)
    - dva rovnaké objekty musia mať rovnaké deskriptory a naopak, dva rovnaké deskriptory musia opisovať rovnaké objekty,
    - podobné objekty musia byť klasifikovateľné do jednej triedy, ak majú podobné (blízke) deskriptory,
    - príznaky by mali byť invariantné. Najmä voči geometrickým transformáciám ak máme rozpoznať objekty snímané z rôznych zorných uhlov,
    - deskriptor by mal byť kompaktný. Mal by obsahovať najmenšiu množinu informácií, ktorá je potrebná na identifikáciu daného objektu a jeho odlíšenie od objektov iných tried.

# Príklad – príznaky tvaru

- **obvod** - napr. ak máme objekt popísaný reťazovým kódom („osemsmerák“), ľahko vypočítame obvod objektu
- **priemerná energia zakrivenia** – množstvo energie potrebné na transformáciu uzavretej krovky na kruh s rovnakým obvodom
  - definovaná pomocou krivosti v jednotlivých bodoch obvodu
  - krivosť krivky = prevrátená hodnota polomeru vpísanej kružnice
- **obsah** – zrejme ...
- **priemer** – max. vzdialenosť medzi bodmi objektu
- **hlavná os** – smer v ktorom je stanovený priemer
- **vedľajšia os** – kolmá na hlavnú os
- **signatúra** - 1D signál opisujúci 2D krivku, napr. vzdialenosť bodov obvodu od ťažiska
- ...



A takto by sme mohli pokračovať celým panteónom priznakov ....

## SW základ pre počítačové videnie

- Matlab – súčasť výučby (najlepšie prepojenie), proof of concept (PoC)
  - computer vision toolbox
  - matlab vie volať openCV funkcie pomocou interfejsových knižníc
- openCV (java, python, C/C++ API) – reálne, životaschopné projekty

# Expertná videoanalýza

- nesnažiť sa napodobiť ľudské oči, ale im trochu „pomôcť“
- využitie obrovského množstva údajov vo viditeľnom spektre
- chceme vidieť „neviditeľné“ zachytenie a štatistické spracovanie jemných detailov (a ich prípadné následné zvýraznenie v pôvodnom videu (ideálne v reálnom čase)) napr.
  - pulz krvi v žilách
  - trasenie rúk (diagnostika napr. parkinsonovej choroby)
  - detektor lži (nie posudzovanie, len zvýrazňovanie príznakov)
  - ...
- klasické
  - meranie vzdialeností