ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická
Katedra radioelektroniky

Oblasti zájmu v kompresi znakového mluvčího

Region of interest of sign language speaker compression

Diplomová práce

Studijní program: Komunikace, Multimedia a Elektronika
Studijní obor: Multimedialní technika

Vedoucí práce: Ing. Martin Bernas, CSc.

Bc. Lukáš Dvořák

Praha 2014

Vložit zadání

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských prací.

V Praze dne 12.května 2014 .....

podpis

Poděkování

Rád bych poděkoval Ing. Martinu Bernasovi, CSc., který mi po celou dobu psaní této práce byl velmi nápomocen, jak s častými konzultacemi, tak i s organizací testů. Také bych rád poděkoval Ing. Zdeňku Švachulovi za jeho rady a pomoc při programování. Dále děkuji všem, kteří se účastnili subjektivních testů pro tuto práci.

Nakonec bych rád poděkoval své přítelkyni a rodině za jejich psychickou podporu a zpříjemnění podmínek při psaní této práce.

Abstrakt

Diplomová práce se zabývá optimalizací kompresního standardu H.264 pro kompresi obrazu mluvčího českého znakového jazyka. Cílem práce je zajistit dobrou srozumitelnost znakových výrazů i dostatečnou subjektivní kvalitu obrazu mluvčího při minimálním bitovém toku.

Práce vychází ze zjištěných oblastí zájmu (ROI) českého znakového jazyka. Nejprve rozděluje tyto oblasti do tří úrovní a podrobně se zabývá jejich dynamickou detekcí ve videosignálu. Poté úpravami softwarového kodéru x264 zavádí tyto oblasti do standardu H.264. Výsledky této práce jsou ověřeny pomocí subjektivních testů s neslyšícími. Je zkoumána srozumitelnost znakových výrazů, obsahujících minimální páry pro různou úroveň komprese i různá rozlišení obrazu mluvčího znakového jazyka a je hodnocena i subjektivní kvalita výsledného obrazu pro dobrý dojem ze sledování.

Klíčová slova

komprese obrazu; H.264; MPEG 4 - Part 10; MPEG 4 - AVC; AVC; x264; Znakový jazyk; oblasti zájmu; ROI; televizní vysílání pro neslyšící

Abstract

This thesis deals with optimization of H.264 image compression of sign language speaker. The aim is to ensure good intelligibility of sing language expressions and sufficient subjective image quality of speaker at a minimum bit rate.

The work is based on the detected region of interest (ROI) of Czech sign language. Firstly, it divides these regions into three levels and it deals with the dynamic detection in video in detail. After editing software encoder x264 it introduces these regions to the H.264 standard. The results of this study verified using subjective tests with the deaf people. It examines the intelligibility of sign language expressions, containing minimal pairs for different levels of compression and various resolution of image with sign language speaker and evaluate the subjective quality of the final image for a good viewing experience.

Keywords

image compression, H.264; MPEG 4 - Part 10; MPEG 4 - AVC; AVC; x264; Sign language; region of interest; ROI; television broadcasts for the deaf

Obsah

Seznam zkratek

AVC	Advance Video Coding
MPEG	Moving Picture Experts Group
ROI	Region Of Interest
MB	Makroblok
HbbTV	Hybrid Broadcast Broadband TV
RGB	Red, Green, Blue
HSV	Hue, Saturation, Value
RLE	Run Length Encoding
DPCM	Differential pulse-code modulation
UHDTV	Ultra High Definition Television
HVS	Human Visual System
GOP	Group Of Pictures
DCT	Discrete Cosine Transform
SAD	Sum of Absolute Difference
QP	Quantization Parameter
DSIS	Double Stimulus Impairment Scale
ITU	International Telecommunication Union
ITU-R	RadioCommunication Standardization Sector of ITU
CAVLC	Context-based Adaptive Variable Length Coding
CABAC	Context-adaptive Binary Arithmetric Coding
FFUK	Filozofická fakulta Univerzity Karlovy

Úvod

Televizní vysílání představuje v současné době nejrozšířenější (sdělovací) médium, které slouží nejen jako zdroj informací, ale i zábavy. Neoddělitelnou součástí televizního vysílání je zvukový doprovod. Pro diváky se sluchovým postižením jsou velkým pomocníkem skryté titulky, které dnes doprovázejí většinu pořadů. Pro prelingválně neslyšící je však mateřským jazykem znakový jazyk a čtení českých titulků je pro ně značně obtížné.

Některé TV stanice mají ve svém vysílání pořady pro neslyšící doplněné o tlumočníka do znakového jazyka. Tlumočník je většinou umístěn v pravém spodním rohu a zakrývá poměrně velkou část obrazovky, což je rušivé pro běžného uživatele. To je také jeden z důvodů, proč je obecně takových pořadů pro neslyšící velmi málo.

Mezi jednu z nových technologií v televizním vysílání patří HbbTV (*Hybrid Broadcast Broadband TV*). HbbTV propojuje televizní vysílání s internetem a umožňuje přidávat k vysílání doplňky, které si televizor dokáže stáhnout z internetu a synchronizovat je s vysíláním. Jedním z těchto doplňků je i externí video, které lze vložit k přijímanému obrazu. Tato funkce se přímo nabízí k využití pro doplnění televizního vysílání o externí video s mluvčím ve znakovém jazyce. Externí video s mluvčím by si tak mohli zapnout pouze neslyšící, nijak by to neovlivnilo běžného televizního diváka.

Důležitým předpokladem pro správné porozumění mluvčího ve znakovém jazyce je kvalitní obraz, a to především těch oblastí znakového mluvčího, které se používají k tvorbě znakových výrazů (ruce, obličej). Ne všude je však dostatečně rychlé internetové připojení. Proto je zapotřebí minimalizovat datový tok videosignálu s mluvčím tak, aby kvalita signálu byla dostačující pro porozumění.

V této diplomové práci se zabývám kompresí videosignálu pro neslyšící a mým cílem je zavedení vhodných oblastí zájmu (*ROI - Region of interest*) do nejmodernějšího používaného kompresního standardu H.264/AVC. Zavedení oblastí zájmu umožní, aby i při nízkém bitovém toku byl obraz tlumočníka do znakového jazyka pro neslyšící srozumitelný.

1. Znakový jazyk

V České republice žije přibližně 10 000 sluchově postižených, jejichž základním dorozumívacím prostředkem je znakový jazyk. Znakový jazyk je jazykem nevokálním, nepracuje se zvukem. Je označován jako vizuálně motorický. Hlasový projev je v něm nahrazen pohyby rukou, pozicí těla a mimikou obličeje. Zjednodušeně se dá říci, že neslyšící "slyší" očima.

Ani tento jazyk není mezinárodní, i když stejně jako jiné vokální jazyky má s jinými jazyky určité podobnosti. To znamená, že Češi mají svůj český znakový jazyk, který se v komunitě neslyšících vyvíjel postupně. Stejně tak například v Německu, v Anglii, v USA a v jiných zemích mají samostatné znakové jazyky, a proto si většinou navzájem nerozumí. Hlavním rozdílem českého znakového jazyka oproti mluvené češtině je to, že se jedná o jazyk simultánní, zatímco čeština je jazykem lineárním. Lineární, tedy přímý jazyk, znamená postupné skládání jednotlivých základních složek v čase za sebou, ať se jedna o písmena, či slova. V určitý časový úsek dokážeme vyjádřit jen jednu danou informaci. Oproti tomu znakový jazyk je podstatně členitější, používá k vyjádření informace především tvarů, polohy a pozice rukou. Díky tomu je možné ve znakovém jazyce vyjádřit více informací současně.

Dalším rozdílem znakových jazyků oproti jazykům mluveným je rozdělení na manuální a nemanuální složku. Manuální složka je mechanická část znakování, tedy poloha, tvar a pozice rukou. Nemanuální složku tvoří především neverbální část. To znamená, že určité pohyby rukou jsou doplněny pohyby rtů, výrazem obličeje a pohyby hlavou. Nemanuální složka je nezbytnou částí českého znakového jazyka, v němž se některé výrazy znakují stejně. Správný význam zpřesní až nemanuální složka. Kromě vyjádření nálady či emoce hraje také důležitou roli v gramatice. Mimika a pohyby těla fungují na rovině lexikální jednotky jak u jednotlivých znaků, tak v celých proslovech. Nemanuální složka nahrazuje i intonaci, která také nese podstatnou část vyjádření.

Dobrý mluvčí ve znakovém jazyce, v našem případě českém, dovede pracovat s výrazem očí, s pohyby rtů, s jemnými pohyby hlavou. Rukama "hovoří", tedy popisuje, o čem je právě řeč, kdežto obličejem vyjadřuje citovou stránku. Nemá to však nic společného s jazykem mluveným, s hovorovou češtinou. Jestliže široce rozevře ústa, neznamená to artikulované písmeno O, ale jistý citový projev provázející pohyby rukou.

Podobně třeba pevně sevřené rty, povyplazený jazyk, vibrující rty, sešklebení, pousmání. Kromě toho může tato artikulace i doplnit nebo pozměnit význam znaků [4].

1.1 TV vysílání ve znakovém jazyce

Jsou pořady, které jsou na různých TV stanicích určeny přímo pro neslyšící. Těchto pořadů je málo. Jiné pořady jsou provázeny tlumočníkem do znakového jazyka, takže je neslyšící mohou sledovat. Když se však podíváme do televizního programu a hledáme písmenko Z, které označuje pořady ve znakovém jazyce, vidíme že i těchto pořadů je však velmi málo (krátké zprávy na ČT2).

Přitom televizní vysílání je médium, v němž by se mohli dozvědět nejen mnohé informace, ale mohli by sledovat naučné, kulturně zaměřené a zábavné pořady. S nástupem HbbTV, to bude možné. Neboť HbbTV dovolí přenést znakového tlumočníka po internetu a vložit ho do televizního obrazu. Nabízí se otázka, proč neslyšící nevyužijí funkce skrytých titulků, které jsou na teletextu dostupné pro většinu pořadů.

Odpověď je celkem jednoduchá. Pro člověka od narození neslyšícího je čeština (i jiné jazyky) jazykem nepřirozeným a téměř neznámým. Samozřejmě v mnoha zemích včetně České republiky jsou školy pro neslyšící. V nich se děti učí mluvit a psát, učí se i gramatice a běžným školním předmětům. Ale je v tom problém. Představme si, že jsme nikdy neslyšeli žádný zvuk, neumíme si představit, jak zní lidský hlas a teď máme něco přečíst. Číst umíme, jenže skryté titulky běží rychle, a přečíst je nestačíme. Dále bychom si mohli představit situaci, kdy slepec stojí před sochou. Nemůže ji vidět, ale může si ji osahat. Když to udělá, zjistí, že socha představuje sedící ženu. Trvá mu to však o dost déle, než lidem vidoucím, kterým k tomuto zjištění stačí jediný rychlý pohled. Nějak podobně je tomu se čtením u neslyšících. Musí si dát čtená slova do souvislostí, musí si je “ohmatat”. A skryté titulky jim k tomu nedají dostatek času.

1.2 Mluvčí ve znakovém jazyce

Pro komunikaci ve znakovém jazyce používá mluvčí tzv. znakovací prostor. Znakovací prostor je ohraničen dolní částí trupu, temenem hlavy ve vertikálním směru a roztaženými lokty v horizontálním směru. Většina znaků vyskytující se v běžném projevu ve znakovém jazyce se nachází právě v tomto prostoru. Při televizním vysílání

se snímá mluvčí na konstantním pozadí a záběr je o málo větší než znakovací prostor[4].

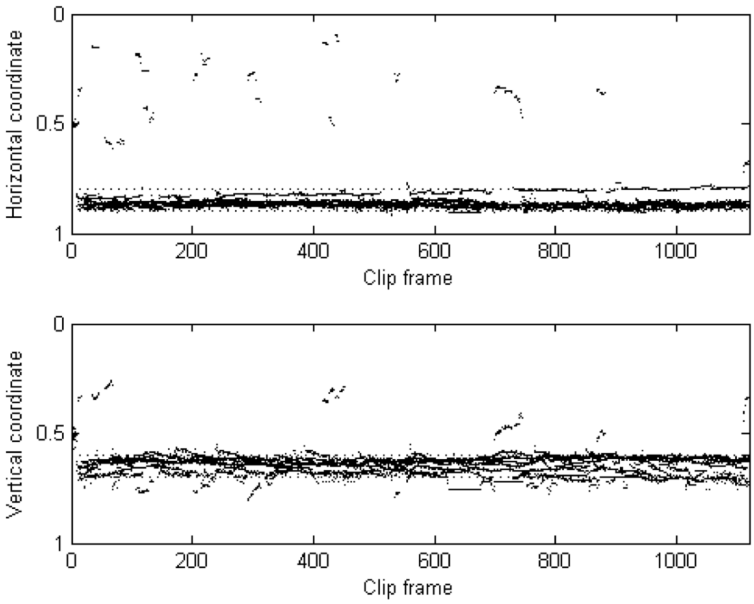
1.3 Oblasti zájmu - ROI

V experimentu [2] byla realizována měření, která se zabývala zjišťováním oblastí zájmů neslyšícího televizního diváka při sledování pořadu se znakovým mluvčím. V tomto měření byly provedeny testy na neslyšících. Účastníkům testu se promítaly připravené videosekvence, které obsahovaly klip se samotným tlumočníkem před neutrálním pozadím, a dále tři nahrávky z televizního vysílání s mluvčím znakového jazyka. Každý z těchto snímků se lišil umístěním a velikostí mluvčího. Jednalo se o zprávy v českém znakovém jazyce, kde je mluvčí před grafickým pozadím, dále pak o pořad Sama doma s tlumočníkem v rámečku v pravém dolním rohu a o televizní pohádku s dynamicky zobrazovaným mluvčím při promluvě herců, viz obr. 1.1.

Hodnotitelům se postupně promítaly jednotlivé videosekvence a s pomocí systému ViewPointEyeTracker se sledovalo, kam neslyšící divák při pozorování televize zaměřuje svou pozornost. Na obr. 1.2 je vidět výsledek měření pro video z pořadu Sama doma, kde se nachází mluvčí v pravém horním rohu. Z výsledků je patrné, že pozorovatel sledoval nejvíce část obrazu mezi tečkovanými čárami, což odpovídalo pozici obličeje.

Měřením bylo zjištěno, že pokud je při televizním vysílání přítomen mluvčí ve znakovém jazyce, je pohled pozorovatelů soustředěn hlavně na oblast obličeje, jenom v menší míře se přemístí na zbytek obrazu. Obličej je stanoven jako nejdůležitější oblast zájmu při sledování mluvčího znakového jazyka. Ruce tvoří další oblast zájmu, ale méně podstatnou [2].

Obr. 1.1: Ukázka videosignálů použitých při testech [2]



Obr. 1.2: Ukázka výsledku jednoho z měření [2]

2. Detektor objektů v obraze

Před uskutečněním komprese videosignálu s mluvčím znakového jazyka se, na základě experimentů popsaných v minulé kapitole, nejprve rozdělí obraz na části s oblastmi zájmu a na části bez oblastí zájmu. K nalezení oblastí zájmu se využije metody pro zjištění objektů v obrazu. Základní metody detekce objektů v obraze jsou popsány v této kapitole.

2.1 Detektor Viola-Jones

Nejvyužívanějším detektorem objektů v obraze je detektor popsaný P.Violou a M. Jonesem v roce 2001. Tento detektor se vyznačuje především svojí přesností detekce a rychlostí výpočtu. K detekci objektu v obraze jsou využity 4 základní principy, které budou v následujícím textu popsány. První z nich jsou klasifikátory pracující s Haarovými vlnkami, které slouží k samotné detekci. Pro zrychlení výpočtů detekce se využívá integrálního obrazu a kaskády klasifikátorů. Pro trénování detektoru se využívá AdaBoost algoritmus popsaný Y. Freundem a R.E. Schapirem z roku 1995 [5][6].

2.1.1 Haarovy vlnky

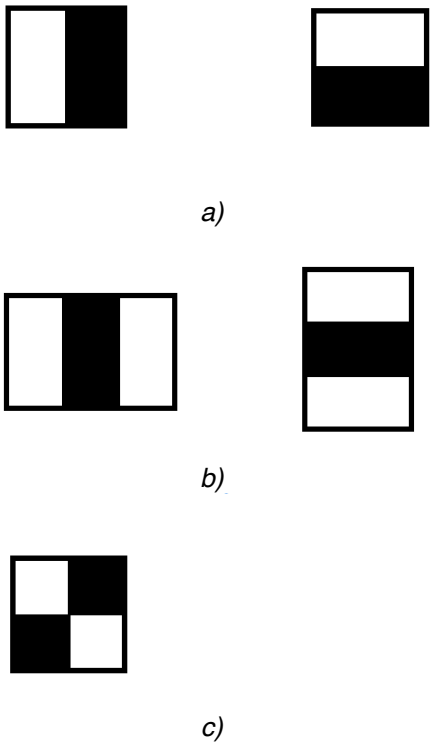
Snahou detektoru Viola-Jones je získat velkou řadu jednoduchých příznaků s minimálními výpočetními nároky. K nalezení jednotlivých příznaků se používají tzv. Haarovy vlnky neboli Haarovy příznaky (*Haar-like features*). K získání těchto příznaků se využívá několik obdélníků (obr. 2.1), které mohou být tvořeny dvěma (hranový příznak), třemi (čárový příznak) či čtyřmi (diagonální příznak) obdélníkovými oblastmi, v závislosti na typu informace, která má být detekována [6][7].

Hodnota příznaku **f(x)** ze vstupního obrazu **x** je vypočítána jako součet sumy bílé oblasti $\mathbf{r_0}$ a černé části $\mathbf{r_1}$, kde bílá část má váhu $\boldsymbol{\omega_0}$ = -1. Váha černé části je vypočtena jako podíl bílé a černé oblasti [6]

$$f(x) = \omega_0 r_0 + \omega_1 r_1 . \qquad (2.1)$$

Postup generování jednotlivých příznaků je následující: nejprve se nastaví základní okno, nejčastěji 24x24 pixelů, ve kterém jsou generovány příznaky. V tomto okně jsou jednotlivé příznaky posouvány vždy po jednom pixelu ve vertikálním, nebo horizontálním směru. Začne se s nejmenší velikostí příznaku a při každém projetí

celého okna se velikost příznaku zvětšuje, dokud není příznak větší než velikost okna. Při každém posuvu je generovaný příznak přidán do seznamu všech příznaků. Výsledný počet příznaku pak pro okno 24x24 pixelů je okolo 160 000 [5][6].



a)

b)

c)

Obr. 2.1: Haarovy vlnky a) Hranové příznaky b) Čárové příznaky c) Diagonální příznaky

2.1.2 Integrální obraz

Výpočet sumy hodnot pixelů pro definovanou oblast příznaků je časově velmi zdlouhavý, proto se využívá tzv. integrálního obrazu. Vstupní obraz je převeden na integrální, ve kterém hodnota každého bodu je suma všech předchozích pixelů ve vstupním obraze, viz obr. 2.2. K převodu vstupního obrazu na integrální se využívá rovnic

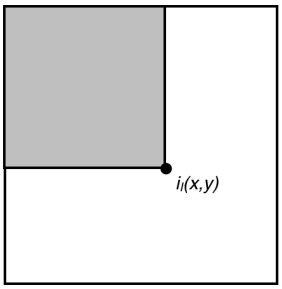
$$s(x,y) = s(x,y-1) + i(x,y) \tag{2.2}$$

$$i_I(x,y) = i_I(x,y-1) + s(x,y), \tag{2.3}$$

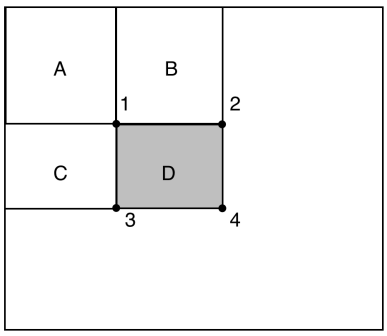
kde **s(x,y)** je kumulovaný součet pixelů v řádku pod podmínkou, že **s(x,-1) = 0** a $\mathbf{i_I(-1,y) = 0}$. **i(x,y)** je hodnota jednotlivých pixelů vstupního obrazu. Každý bod v integrálním obraze $\mathbf{i_I(x,y)}$ pak odpovídá hodnotě dle vzorce

$$i_I(x,y) = \sum_{x' \le x, y' < y} i(x',y'). \tag{2.4}$$

Hodnota obdélníku v obraze se v integrálním obraze vypočítá jednoduše viz obr. 2.3. Na obrázku jsou 4 oblasti (A,B,C,D) a čtyři body, kde každý bod udává sumu hodnot jednotlivých obdélníků. Pro bod 1 je to obdélník A, pro bod 2 jsou to A a B, pro bod 3 A a C a pro bod 4 A, B, C a D. Výsledná suma hledané oblasti se tedy vypočítá z bodů jako (1+4) - (2+3) [5][6].



Obr. 2.2: Hodnota bodu v integrálním obraze



Obr. 2.3: Výpočet sumy určitého obdélníku

Ve snímku je skoro vždy různá intenzita osvětlení, z toho důvodu je nutné Haarův příznak normalizovat pomocí vzorce

$$f(x') = \frac{f(x)}{wh\sqrt{\frac{i_{Sqr(w,h)} - \frac{i_I^2(w,h)}{wh}}{wh}}}, \qquad (2.5)$$

kde **f(x)** je Haarův příznak, **f(x)' je** normalizovaný příznak, **w** a **h** je šířka a výška detekčního okna, $\mathbf{i_I(w,h)}$ je suma hodnot pixelů vstupního snímku o velikosti **w**, **h** a $\mathbf{i_{ISqr(w,h)}}$ je suma kvadrátů hodnot pixelů v detekčním okně. Kvadrát hodnot pixelů integrálního obrazu $\mathbf{i_{ISqr}}$ je vypočten stejným způsobem jako $\mathbf{i_I}$ s tím rozdílem, že se s původními hodnotami pixelů **i(x,y)** počítá ve druhé mocnině $\mathbf{i^2(x,y)}$ [6].

2.1.3 Klasifikátory

Úkolem klasifikátoru je klasifikovat vstupní data do různých tříd s přesností větší než 50%. V detekci objektu klasifikátor určuje, zda data obsahují hledaný objekt či nikoliv, tedy klasifikace do dvou tříd: pozitivní a negativní. Lineární slabý klasifikátor obsahuje příznak **f** , polaritu **p** a prahové hodnoty **θ**. Klasifikátor nejprve z vstupního obrazu **x** vypočte odezvu příznaku, následně podle polarity **p** určí, jestli se pozitivní klasifikační třída nachází nad, nebo pod prahovou hodnotou. Poté je rozhodnuto, zda snímek **x** patří do pozitivní **h(x) = 1,** nebo do negativní klasifikační třídy. Rozdělení do jednotlivých klasifikačních tříd je vidět na následující rovnici [5][6]

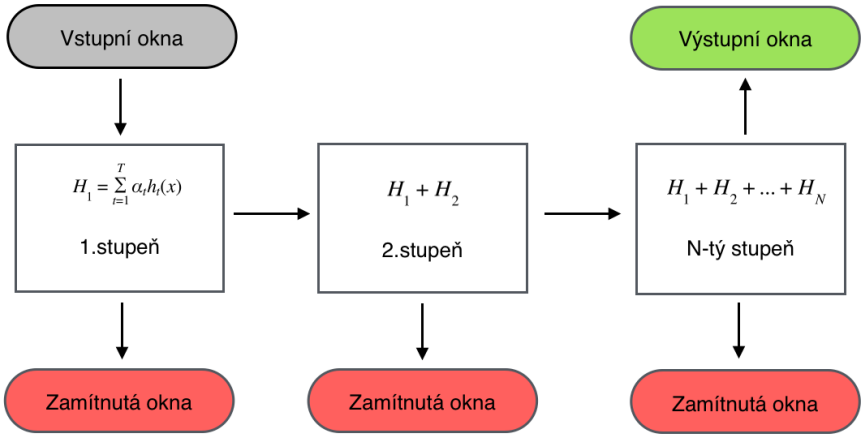
$$h(x,f,p,\Theta)=\begin{cases}1 & pf(x)<p\Theta\\ 0 & jinak.\end{cases} \tag{2.6}$$

Pro urychlení doby detekce hledaného objektu se využívá kaskády klasifikátorů. Kaskáda je sestavena z více stupňů, kde každý stupeň kaskády je složen z několika slabých klasifikátorů **h(x)**, tím je definován silný klasifikátor **H(x)**. Tento silný klasifikátor má svou prahovou hodnotu **P**, podle které rozhoduje, jestli dané podokno patří do pozitivní, či do negativní třídy. Funkce silného klasifikátoru je popsaná rovnicí

$$H(x)=\begin{cases}1 & \sum_{t=1}^{T}\alpha_t h_t(x)\geq P\\ 0 & jinak,\end{cases} \tag{2.7}$$

kde α_t je váha slabého klasifikátoru h_t.

Motivací kaskády je, aby v každém stupni zamítla co nejvíce negativních oken a do dalšího stupně poslala už pouze okna pozitivní, tím se ušetří spousta výpočtů. Na obr. 2.4 je vidět kaskáda klasifikátorů. První stupeň obsahuje několik klasifikátorů a měl by zamítnout co nejvíce negativních oken. Další stupně obsahují vždy všechny slabé klasifikátory z předchozího stupně. Každá kaskáda by měla být natrénována k vysoké úspěšnosti detekce a zároveň k co nejmenšímu počtu negativních detekcí [5] [6].



Obr. 2.4: Schéma kaskádového zapojení klasifikátorů

2.1.4 AdaBoost

AdaBoost (*Adaptive Boosting*) je klasifikační algoritmus vycházející z metody zvané boosting. Cílem je zlepšení klasifikační metody libovolného algoritmu. Základem je více slabých lineárních klasifikátorů **h(x)** z množiny klasifikátorů **H**, jejichž přesnost odhadu je o málo větší než 50%. Postupným přidáváním dalších klasifikátorů s podobnou mírou přesností je generován silný nelineární klasifikátor **H(x)**. Přesnost silného klasifikátoru je závislá na trénovací množině, čím větší trénovací množina, tím větší přesnost. Tento silný klasifikátor pak reprezentuje jeden stupeň kaskády klasifikátorů. S větším množstvím stupňů kaskády je zapotřebí více silných klasifikátorů. Vstupem algoritmu je trénovací množina **S**, která je složena z dvojice (x_i,y_i), kde x_i je získaná hodnota příznaku a y_i je třída odpovídající příznaku, negativní či pozitivní. V tomto případě jsou trénovací data obrázky obsahující hledaný objekt (pozitivní příznaky) a obrázky obsahující pozadí (negativní příznaky) [6][7].

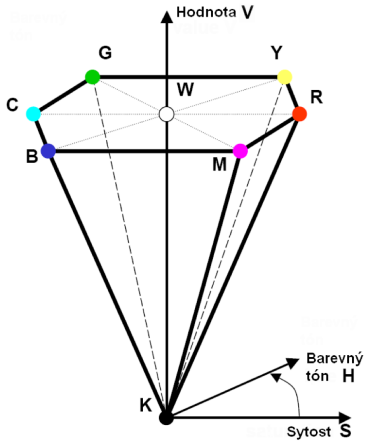
Obr. 2.5: Ukázka trénovací množiny tváří

2.2 Detekce objektů pomocí HSV barevného modelu

V obraze se můžou vyskytovat objekty, jejichž tvar nemusí být předem znám, nebo se v čase mění. Proto je potřeba užívat jiné metody detekce. U mnoha objektů však známe barvu, kterou lze využít k vyhledání objektu v obraze.

2.2.1 HSV barevný model

Třídimenzionální reprezentací HSV (*Hue, Saturation, Value*) barevného modelu je šestiboký jehlan. Centrální vertikální osa reprezentuje intenzitu jasu **V**. Určuje množství bílého světla v rozsahu 0 až 1. Vrchol jehlanu leží v počátku souřadnic, a určuje černou barvu. Barevný tón **H** je určován pomocí úhlu v rozsahu od 0 do 2π. Červené barvě odpovídá úhel 0, zelené 2π/3 a modré 4π/4. Sytost barvy S se mění od 0 do 1, kde 1 určuje tzv. čistotu barvy, neboli určuje množství šedi k poměru k odstínu. Čím dále je barva od středu, tím je sytější, čistší [9].



Obr. 2.6: HSV barevný model

2.2.2 HSV detektor

Pro detekci objektu pomocí HSV detektoru je důležitým předpokladem znalost barvy objektu. První možností je využít předem definovaných barevných odstínů a na jejich základě objekt detekovat. Další možností je nejprve v obraze najít oblast, o které víme, že obsahuje danou barvu. Z této oblasti je pak například průměrováním možné získat hledanou barvu, a tu v obraze hledat. K detekci barvy v obraze je nejvyužívanější vyhledávání v prostoru HSV. Barevný obraz, ve kterém chceme najít barevný objekt, je nejprve převeden z RGB prostoru do HSV pomocí:

$$R' = \frac{R}{255}, \quad G' = \frac{G}{255}, \quad B' = \frac{B}{255} \tag{2.8}$$

$$C_{\max} = \max(R', G', B') \quad C_{\min} = \min(R', G', B') \tag{2.9}$$

$$\Delta = C_{\max} - C_{\min} \tag{2.10}$$

$$H = \begin{cases} 60° \times (\frac{G' - B'}{\Delta} \bmod 6), & C_{\max} = R' \\ 60° \times (\frac{B' - R'}{\Delta} + 2), & C_{\max} = G' \\ 60° \times (\frac{R' - G'}{\Delta} + 4), & C_{\max} = B' \end{cases} \tag{2.11}$$

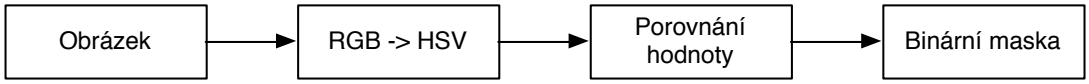
$$S = \begin{cases} 0, & \Delta = 0 \\ \frac{\Delta}{C_{\max}}, & \Delta \neq 0 \end{cases} \tag{2.12}$$

$$V = C_{\max}. \tag{2.13}$$

Následně je obraz pixel po pixelu porovnáván s referenčními hodnotami. Pokud platí, že barva leží mezi hranicemi

$$H \in \langle H_{\min}; H_{\max} \rangle, \quad S \in \langle S_{\min}; S_{\max} \rangle, \quad V \in \langle V_{\min}; V_{\max} \rangle, \tag{2.14}$$

je považován za hledaný objekt. Na základě detekce se vytváří binární maska. Z této masky můžeme říct, kde v obrazu je a není detekován hledaný objekt. Poté je na tuto masku použita morfologie. Jde o širokou množinu operací, které zpracovávají obraz na základě tvarů. Účelem morfologie je zlepšit účinnost detekce. Jde o přidání či odebrání pixelů na hraně objektu v závislosti na velikosti nebo tvaru stavebního prvku, který definuje sousední pixel. Nevýhodou detekce na bázi HSV je, že se v obraze může objevit mnoho předmětů s podobnou barvou, detektor se může často detekovat nepřesně. Proto je dobré detektor využívat v ideálních podmínkách, například se statickým pozadím, které neobsahuje stejné barvy [10][11].



Obr. 2.7: Blokové schéma HSV detekce

3. Komprese videosignálu

Videosignál je sekvence jednotlivých snímků, které tvoří pohyblivý obraz. Při, dnes typicky používaném, rozlišení 1920 x 1080 má jednotlivý obrázek přes 2 miliony obrazových bodů. Pokud uvažujeme studiový videosignál s 25 snímky za vteřinu, data, která musí být přenášena, dosahují objemu okolo 1.5 Gb/s , a jejich přenos je v reálném čase takřka nemožný. S rozvojem technologie zobrazovací techniky se dnes hodně mluví o UHDTV, kde jeden snímek má přes 33 milionů obrazových bodů. Přenos videosignálu se tedy nikdy neobejde bez komprese dat. V této kapitole se budu zabývat základními principy komprese videosignálu. Podrobněji dále popsán standard H.264/MPEG4 Part 10 AVC (*Advenced video coding*), který jako nejmodernější používaný standard bude v práci použit.

3.1 Komprese

Hlavním úkolem komprese je snížit datový tok odstraněním nadbytečných dat, ale přitom zachovat původní informaci. Pokud informace zůstává po dekompresi neměnná, jedná se o kompresi bezeztrátovou. Naopak pokud při kompresi dojde k neobnovitelné ztrátě informace, nazýváme tuto kompresi ztrátovou. Komprese videosignálů využívá jak bezeztrátové, tak ztrátové komprese.

3.2 Bezeztrátová komprese

Úkolem bezeztrátové komprese je redukovat objem dat pro přenos tak, aby výstupní data z dekodéru byla totožná se vstupními. Cílem bezeztrátové komprese je odstranit redundantní informaci. Toho dosahuje efektivnějším zakódováním přenášených dat. Využívá se tří základních metod: predikční metoda, slovníková metoda a pravděpodobnostní metoda.

Mezi predikční metody patří například RLE, kde dochází k nahrazení opakujících se dat. Další predikční metoda je DPCM. Ta místo celých dat přenáší pouze jejich rozdíl. DPCM je v kompresi obrazu využívána díky předpokladu, že mezi jednotlivými body v obraze je velká korelovanost.

Slovníkové metody, již podle názvu, obsahují slovníky, na jejichž základě jsou data nahrazena.

Mezi pravděpodobnostní metody patří např: Huffmanovo kódování. Principem je častěji se vyskytující datová informace, která je zakódována pomocí kratšího kódu. Zatímco informace s menší pravděpodobností výskytu je zakódována pomocí delšího kódu.

Bezeztrátové metody dosahují nízkých a především nezaručených kompresních poměrů, běžně tak 1:3. Pro dosažení větších kompresních poměrů se tedy využívá ztrátové komprese.

3.3 Ztrátová komprese

Při ztrátové kompresi dochází ke ztrátě informace. Výstupní data už nejsou identická se vstupními. Úkolem ztrátové komprese je z obrazu odstranit irelevantní informaci. Při kompresi videosignálu se využívá znalosti HVS (*Human Visual System*). Lidské oko je schopné vnímat jen určité rozmezí prostorových kmitočtů, proto je zbytečné přenášet informace o kmitočtech, které už oko není schopné vnímat. Z důvodu většího množství tyčinek v oku než čípků reaguje člověk více na změnu jasu než barvy. Díky tomu můžeme v obrazu některé informace o barvách zanedbat.

3.4 Komprese videa H.264/AVC

Videosignál se skládá ze snímků následujících v čase za sebou. Snímek je dvourozměrné pole pixelů. Každý pixel obsahuje 3 složky R, G a B, které dohromady definují úplnou barevnou informaci o pixelu. V kompresi obrazu je výhodnější převést obraz z RGB do YUV prostoru, nebo jemu podobných. Výhodou YUV je, že rozděluje jasovou složku od chrominačních. Chrominační složky U a V obsahují velké množství irelevantní informace. Ta může být například podvzorkováním odstraněna, aniž by to ovlivnilo jasovou složku.

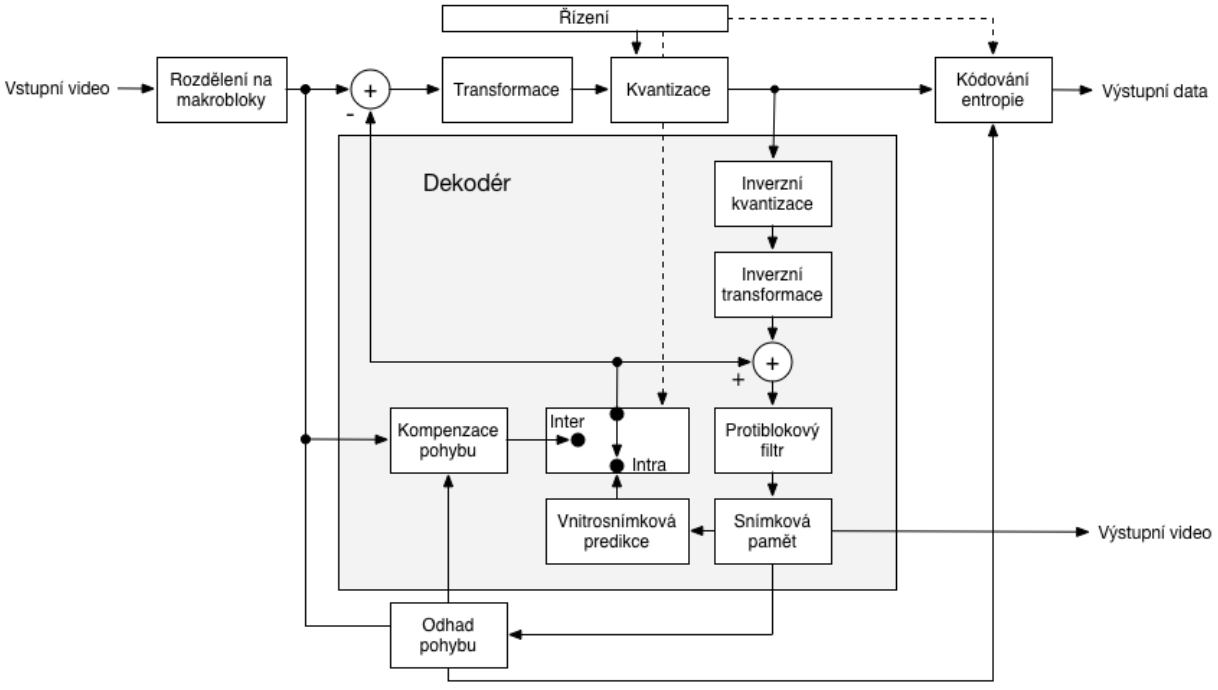
H.264 stejně jako další nejpoužívanější typy komprese je založená na transformaci obrazu. Tyto komprese rozdělují snímek na makrobloky (MB), se kterými se dále pracuje. Komprese probíhá ve třech základních technikách: predikce, transformace a kvantování, entropické kódování.

Predikce slouží k nalezení podobnosti MB tak, aby se nemusel kódovat celý MB, ale jen jejich rozdíl. V závislosti na tom, se kterými MB pracuji, se predikce rozděluje na intra-snímkovou a inter-snímkovou predikci. Intra-snímková predikce se využívá v I snímcích. Referenční MB se nachází ve stejném snímku a počítá se pomocí

matematických funkcí ze sousedních pixelů. Inter-snímková predikce se využívá u P a B snímků. Zde se referenční MB nachází v předchozích, nebo v budoucích snímcích. Referenční MB může také být vážená funkce z MB z několika snímků. I,P a B snímky dohromady tvoří GOP (*Group Of Pictures*)

Rozdíl mezi současným a referenčním MB nazýváme rozdílová data. Jednotlivé MB, nebo jejich rozdíly se pomocí modifikované diskrétní kosinové transformace (DCT) transformují z prostorové oblasti do frekvenční. HVS je více citlivý na nízké frekvence obrazu než na vysoké frekvence. Z toho důvodu se využívá kvantizace ve frekvenčním prostoru. Nízké frekvenci je přiřazena větší váha, zatímco vysoké frekvence jsou zanedbávány.

Třetí a poslední částí je entropické kódovaní. Kódování s proměnnou délkou přisuzuje pravděpodobnějším symbolům kratší kódová slova, a tím minimalizuje celkové množství potřebných bitů k přenosu obrazu. Na obr. 3.1 je vidět kodér a dekodér standardu H.264/AVC. Kodér obsahuje kromě kódující cesty také dekódující zpětnou část sloužící k zakódování snímků, které používají jako referenci již kódované části. V následujícím textu budou vysvětleny principy jednotlivých funkcí H.264/AVC[1].



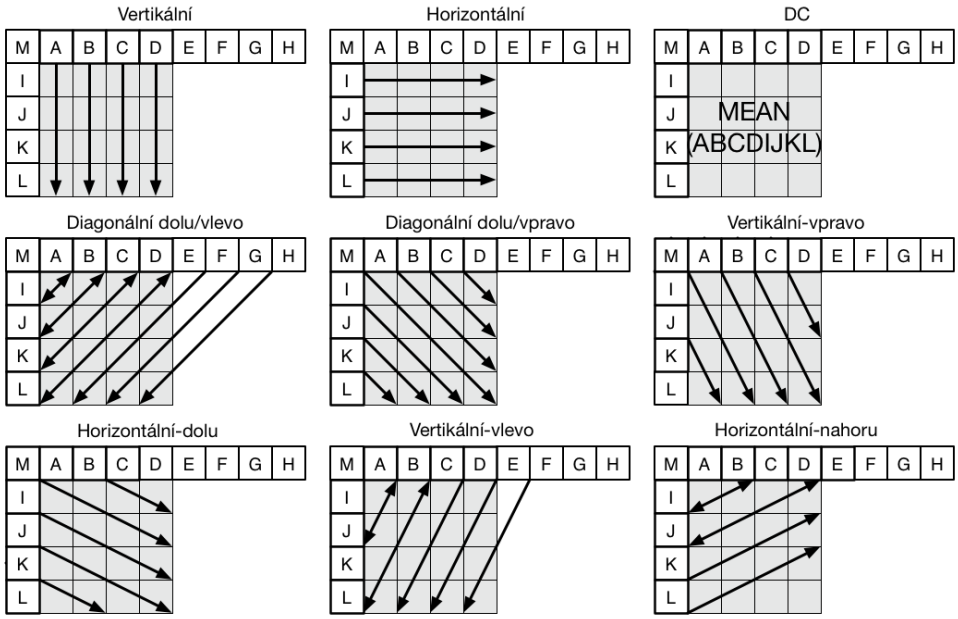
Obr. 3.1: Blokové schéma H.264/AVC

3.4.1 Predikce:

Predikce využívá prostorové nebo časové redundance videosignálu tak, aby mohl být kódován pouze rozdíl mezi originálním a predikovaným snímkem, namísto kódovaní originálních dat. Jak už bylo řečeno, existují dva druhy predikce: intra-snímková pro I snímky a inter-snímková predikce pro P a B snímky.

Intra-snímková predikce:

Obraz je dvouprostorové pole, ve kterém platí, že mezi blízkými body je vysoká korelovanost. Z toho vyplývá, že bod obrazu může být predikován ze sousedních pixelů již kódovaných a rekonstruovaných MB. Predikce se provádí pomocí sady matematických funkcí. V H.264/AVC formátu 4:2:0 je I snímek rozdělen do MB, kde jeden MB obsahuje jasový blok 16x16 a dva chrominační bloky 8x8. Každý blok je predikován samostatně. Blok jasových složek může být predikován celý 16x16, nebo rozdělen na 16 bloků 4x4. První varianta, predikce 16x16, je vhodná pro obrazové části, ve kterých není moc detailů. Využívá 4 módů predikce - vertikální, horizontální, na základě střední hodnoty, a prostorovou. Ve druhém případě, kde je predikováno po blocích 4x4, je predikce doplněna o další módy - diagonální dolů zleva, diagonální dolů zprava, vertikální doprava, vertikální doleva, horizontální dolů a horizontální nahoru. Chrominační bloky jsou predikovány po blocích 8x8 pomocí jednoho ze čtyř módů: vertikální, horizontální, střední hodnota, prostorový [1][14].



Obr. 3.2: Druhy intra-snímkové predikce

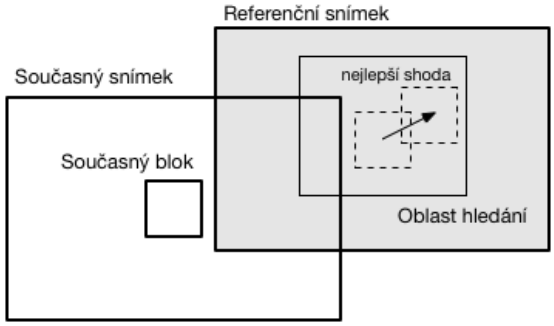
Inter-snímková predikce:

Videosignál má v Evropě ve většině případů 25 nebo 50 snímků za vteřinu. Z toho vyplývá, že po sobě následující snímky si, ve většině případů, budou podobné. Cílem inter-snímkové predikce je využít časové redundance pro snížení dat potřebných pro kódování. Pokud máme statický obraz, ve kterém je například jedoucí auto, je zbytečné kódovat celý obraz, když většina obrazu bude stejné. Stačí zakódovat rozdíl mezi současným a referenčním snímkem.

Ve většině video kompresních standardů se dále využívá odhadu pohybu bloků. Pokud dochází v obraze k pohybu nebo ke švenku, bloky už nezůstávají na stejném místě jako v referenčním snímku, ale je pravděpodobné, že se blok bude nacházet někde v okolí. Pro každý blok v aktuálním snímku se hledá nejpodobnější blok z referenčního snímku. Pro P snímky mohou být použity předchozí, nebo budoucí snímky jako snímky referenční. V případě B snímku se mohou předchozí i budoucí snímky kombinovat. Informace o posunu bloků se pak přenáší pomocí pohybových vektorů. K nalezení této podobnosti se využívá funkce SAD (součet absolutních rozdílů)

$$SAD(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} | p(x+i,y+j) - p'(x+i+dx,y+i+dy) | \qquad (3.1)$$

kde **p** je současný snímek a **p'** je referenční snímek[1][14].



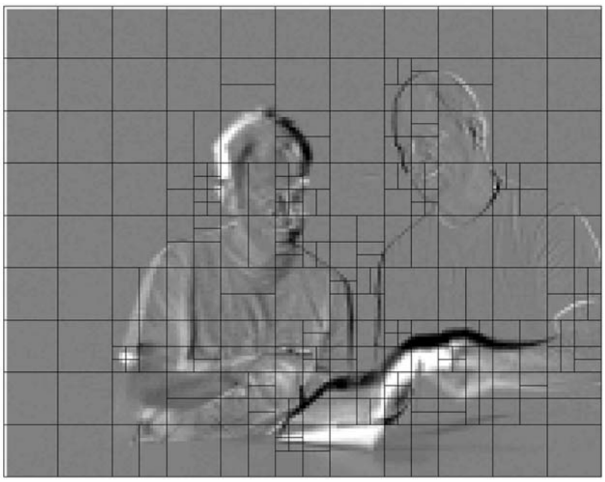
Obr. 3.3: Vyhledávání bloku na základě odhadu pohybu

Standard H.264/AVC přišel oproti původním kompresním standardům s novými technikami odhadu pohybu - variabilní velikost bloků, vícenásobné referenční snímky.

Variabilní velikost bloků:

V předchozích kompresních standardech byla velikost bloku pevná, např. 8x8 nebo 16x16. Odhad pohybu byl stejný jak u statických částí obrazu, tak u pohybujících se objektů. To způsobuje nižší účinnost kódování. V H.264 může být každý 16x16 makroblok rozdělen do dalších dílčích sub-makrobloků, dvou 16x8, dvou 8x16 nebo čtyř 8x8. Pokud se jedná o poslední variantu rozdělení na bloky

po 8x8, je možné tyto sub-makrobloky znovu rozdělit podobně jako v předchozím případě.



Obr. 3.4: Rozdělení na makrobloky [1]

Variabilní velikost bloků používá rozdělení na menší bloky pro pohybující se objekty, zatímco na statické části jsou použity velké bloky. Tím dochází ke zvýšení kvality obrazu a efektivity kódování [1][14].

Vícenásobné referenční snímky:

H.264/AVC narozdíl od předchozích standardů, které pro odhad pohybu využívaly jen jeden referenční snímek, využívá více referenčních snímků, a to 5 pro P snímky a 10 pro B snímky. To má za následek menší predikční chybu, a tedy menší datový tok. Naopak dochází ke zvýšení výpočetní náročnosti a nutnosti větší paměti pro přenos [14].

3.4.2 Protiblokový filtr

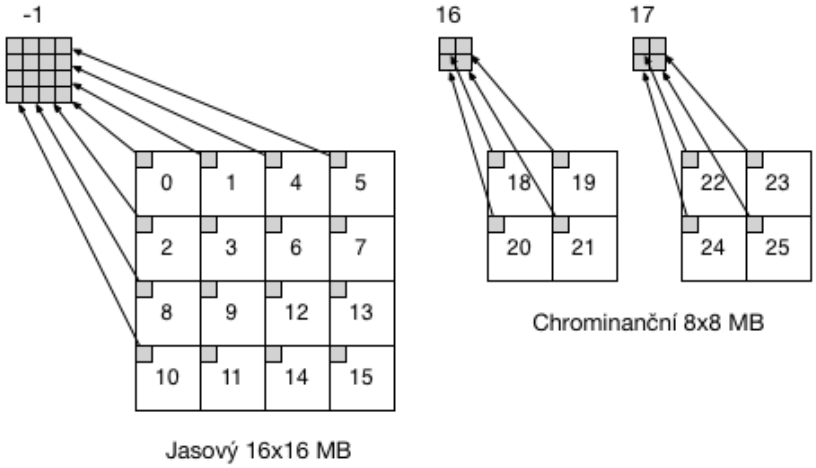
Nevýhodou komprese videa, založená na blokové struktuře, je to, že při rekonstrukci obrazu dochází k viditelným artefaktům způsobených právě blokovou strukturou. K odstranění těchto artefaktů je využit protiblokový filtr, který vyhlazuje přechody mezi jednotlivými bloky. Protiblokový filtr je umístěn jak ve zpětné dekódující části kodéru, tak v dekodéru. Filtrovaný obraz je věrohodnější, než obraz s blokovou strukturou, a proto se využívá pro predikci budoucích snímků, což zlepšuje vlastnosti komprese. K filtraci dochází vždy na hranici 4x4 bloku, a to jak v horizontálním, tak vertikálním směru. To znamená, že dochází ke dvěma horizontálním a dvěma

vertikálním filtracím v chrominačním bloku a čtyřem filtracím v každém směru v jasovém bloku [14].

3.4.3 Transformace a kvantizace

Transformace se u komprese videa využívá k převodu obrazové informace z prostorové oblasti do oblasti prostorových kmitočtů. V této oblasti dochází ke kvantizaci koeficientů a kódování dat pro přenos.

H.264/AVC využívá tři druhy transformace - transformaci založenou na diskrétní kosinové transformaci pro transformaci rozdílových dat (obr. 3.5 bloky 0-15 a 18-25), 4x4 Hadamardovu transformaci pro jasové DC koeficienty v 16x16 módu (blok -1), 2x2 Hadamardovu transformaci pro chrominační DC koeficienty (bloky 16 a 17). Bloky jsou odesílány v pořadí od -1 do 25 [1][14].



Obr. 3.5: Pořadí makrobloků pro přenos

3.4.4 Transformace rozdílových dat

H.264/AVC využívá modifikované DCT pro transformaci jednotlivých 4x4 bloků videosignálu resp. rozdílových dat.

Diskrétní kosinová transformace DCT vychází z Fourierovy transformace, ale využívá pouze reálné kosinové složky. Výhodou je kromě jednodušších výpočtů také lepší kompresní poměr. V obrazové kompresi se využívá DCT k převodu bloku **X**

s NxN vzorky v prostorové oblasti do **Y** s NxN koeficienty ve frekvenční oblasti. Dopředná DCT je definována vztahem:

$$Y = AXA^T \tag{3.2}$$

a inverzní DCT:

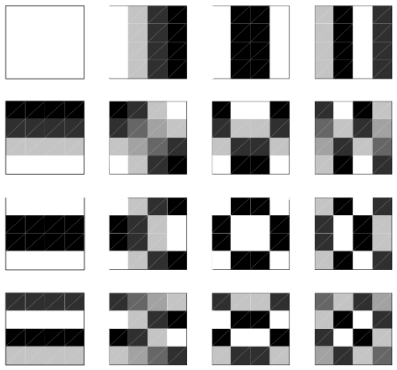
$$X = A^T YA, \tag{3.3}$$

kde **X** pole vzorků, **Y** pole koeficientů a **A** je transformační matice o velikosti NxN, kde jednotlivé bázové funkce jsou:

$$A_{ij} = C_i \cos\frac{(2j+1)i\pi}{2N} \tag{3.4}$$

$$C_i = \begin{cases} \sqrt{\frac{1}{N}}, & i = 0 \\ \sqrt{\frac{2}{N}}, & i > 0 \quad . \end{cases} \tag{3.5}$$

Výstupem dopředné DCT je pole NxN koeficientů, kde každému koeficientu náleží jeden ze základních vzorů viz obr.3.6. Velikost koeficientů nám určuje, kolikrát je v původním bloku každý vzor obsažen. Samotná DCT je bezeztrátový proces. Ke ztrátám dochází až v následující části, kvantizaci.



Obr. 3.6: Šablona 4x4 DCT matice

Základní změny modifikované DCT v H.264 jsou:

- Využívá celočíselné transformace (pouze celočíselná aritmetika)
- Transformace je provedena pouze použitím sčítání a posunů (rychlejší výpočet)
- Změna měřítka (násobení) je součástí kvantizéru - snížení počtu násobení

Při použití 4x4 DCT, která je dána vztahem

$$Y = AXA^T = \begin{pmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{pmatrix} (X) \begin{pmatrix} a & b & a & c \\ a & c & -a & -b \\ a & -c & -a & b \\ a & -b & -a & -c \end{pmatrix}, \tag{3.6}$$

kde

$$a = \frac{1}{2}, \quad b = \sqrt{\frac{1}{2}} \cos\left(\frac{\pi}{8}\right), \quad c = \sqrt{\frac{1}{2}} \cos\left(\frac{3\pi}{8}\right) \tag{3.7}$$

je matice faktorizována do následující rovnice

$$Y = (CXC^T) \otimes E = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & d & -d & -1 \\ 1 & -1 & -1 & 1 \\ d & -1 & 1 & -d \end{bmatrix} [X] \begin{bmatrix} 1 & 1 & 1 & d \\ 1 & d & -1 & -1 \\ 1 & -d & -1 & 1 \\ 1 & -1 & 1 & -d \end{bmatrix} \right) \otimes \begin{bmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \end{bmatrix}, \tag{3.8}$$

kde

$$a = \frac{1}{2}, \quad b = \sqrt{\frac{2}{5}}, \quad d = \frac{1}{2} \tag{3.9}$$

a **CXC^T** je jádro 2D transformace. **E** je matice koeficientů změny měřítka. Každý koeficient **E** odpovídá elementu matice **CXC^T** na stejné pozici. Následně dochází k úpravě matice **CXC^T**[1]

$$Y = (C_f X C_f^T) \otimes E = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} [X] \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix} \right) \otimes \begin{bmatrix} a^2 & \frac{ab}{2} & a^2 & \frac{ab}{2} \\ \frac{ab}{2} & \frac{b^2}{4} & \frac{ab}{2} & \frac{b^2}{4} \\ a^2 & \frac{ab}{2} & a^2 & \frac{ab}{2} \\ \frac{ab}{2} & \frac{b^2}{4} & \frac{ab}{2} & \frac{b^2}{4} \end{bmatrix}. \tag{3.10}$$

3.4.5 Transformace jasových DC koeficientů

Pokud je zvolen mód 16x16, je každý rozdílový blok 4x4 nejprve transformován pomocí modifikované DCT. Předpokládá se, že jednotlivé DC koeficienty jsou

korelované, proto se využívá Hadamardovy transformace k dekorelaci stejnosměrných složek

$$Y_D = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} [W_D] \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \right) / 2, \qquad (3.11)$$

kde W_D je 4x4 blok DC koeficientů a Y_D je blok po transformaci[1].

3.4.6 Transformace chrominačních DC koeficientů

Stejně jako u jasové transformace se předpokládá velká korelace stejnosměrné složky u chrominačních 4x4 bloků. DC koeficienty jsou přivedeny do matice 2x2, která je transformována pomocí Hadamardovy transformace[1]

$$Y_D = \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} [W_D] \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right). \qquad (3.12)$$

3.4.7 Kvantizace

Hlavním úkolem kvantizace je snížit velikost koeficientů, a tím zjednodušit náročnost kódování. V reálných obrazech je největší množství informací obsažených v nízkých prostorových frekvencích. Ty jsou v DCT spektru soustředěny v levé horní části. Protože HVS je méně citlivý na vyšší frekvence, využívá větší kvantizace na koeficienty vyšších prostorových kmitočtů. Tím může dojít až k jejich odstranění. Kvantizací DCT spektra dochází v nevratným ztrátám informace.

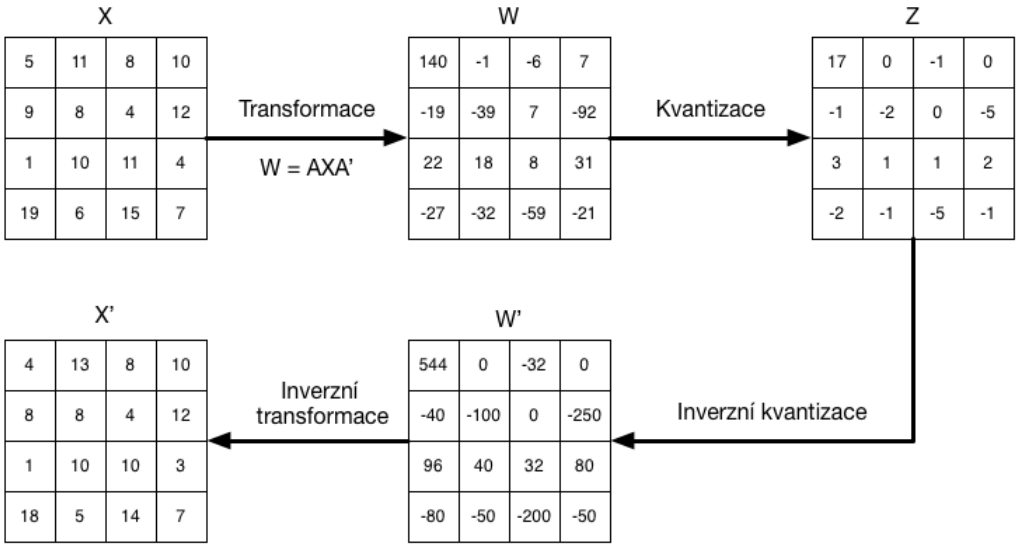
V H.264/AVC se využívá skalární kvantizace a je vyjádřena následujícím vztahem

$$Z_{ij} = round\left(W_{ij} \cdot \frac{PF}{Q_{step}} \right), \qquad (3.13)$$

kde $\mathbf{Z_{ij}}$ je parametr po kvantizaci, $\mathbf{W_{ij}}$ je transformační koeficient (matice $\mathbf{CXC^T}$), **PF** je $\mathbf{a^2}$, **ab/2** nebo $\mathbf{b^2/4}$ v závislosti na pozici (i,j) a $\mathbf{Q}_{step}$ je kvantizační krok. Celkové množství 52 kvantizačních kroků je indexováno pomocí kvantizačního parametru (QP) od 0 do 51 viz tab. 3.1. Na obr. 3.7 je znázorněna transformace a kvantizace vstupního bloku a následně rekonstrukce bloku [1].

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	…
Q_{step}	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	
QP	…	18	…	24	…	30	…	36	…	42	…	48	…	51
Q_{step}		5		10		20		40		80		160		224

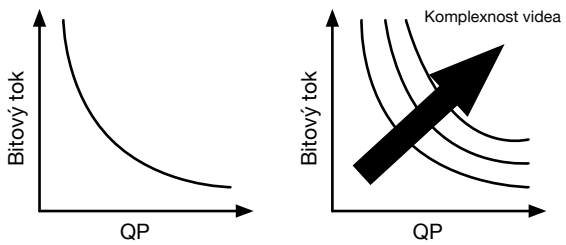
Tab. 3.1: Velikost kvantizačního kroku



Obr. 3.7: Postup transformace a rekonstrukce dat

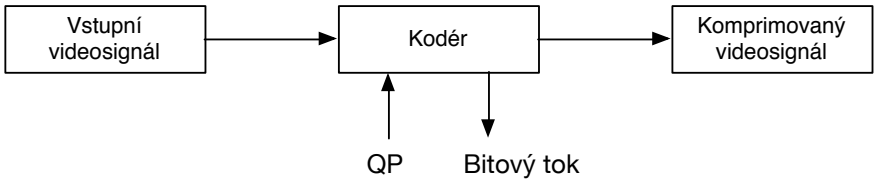
3.4.8 Rate-control

Velikost kvantizačního parametru (QP) určuje, jak velké množství prostorových detailů obrazu bude zachováno. Čím menší hodnota QP, tím více prostorových frekvencí bude obraz obsahovat. S větším množstvím detailů ve videu ovšem stoupá jeho spotřeba dat na přenos. Se zvýšením hodnoty QP tedy klesá bitový tok videa, ale s tím také kvalita obrazu. Velikost bitového toku je společně s QP také závislá na komplexnosti videa viz obr. 3.8.

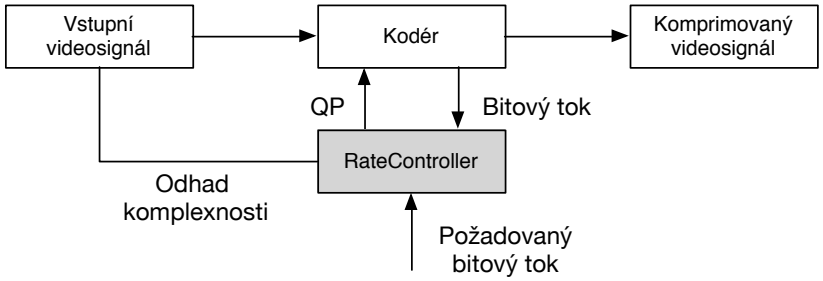


Obr. 3.8: Závislost bitového toku na QP

Jednoduchý princip kodéru je vidět na obr. 3.9. Do kodéru vstupuje nekomprimované video a QP, který nastavuje kvalitu videa. Výstupem je komprimované video s konstantní kvalitou, ale bitová rychlost se může velmi výrazně měnit. Při konstantním QP se v závislosti na komplexnosti videosignálu mění bitový tok. V reálném použití je však přenos závislý na velikosti vyrovnávací paměti dekodéru a šířce pásma přenosového kanálu. Proto je potřeba kódovat video s konstantním bitovým tokem. Toho je dosaženo pomocí bloku nazvaného Rate-controller. Úkolem Rate-controlleru je udržovat nastavenou konstantní bitovou rychlost pomocí dynamické změny QP. Základem algoritmu je princip znázorněný na obr. 3.8. vztah mezi QP bitovým tokem a komplexností videa. Rate-controller však mění bitový tok jen pro rozdílová data, protože QP nastavuje kvantizaci jen pro transformovaná rozdílová data. QP nemá žádný přímý vliv na bitový tok spojený s predikčními daty nebo s pohybovými vektory [1][15].



Obr. 3.9: Jednoduchý princip kodéru



Obr. 3.10: Kodér doplněný o Rate-controller

3.4.9 Entropické kódování

Entropický kodér je zodpovědný za převedení všech elementů kodéru (kvantizované koeficienty, pohybové vektory, metody predikce atd.) na bitový tok tak, aby se všechna data dostala do dekodéru. Entropické kódování je založeno na faktu, že každý signál nese unikátní informaci, a délka kódu je svázaná s entropií zdroje informací. Kódová slova pro každý přenášený symbol mohou mít buď pevnou, nebo proměnou délku. Při kódování s proměnou délkou se přidělují kratší kódová slova

slovům s menší entropií, neboli těm, které mají větší pravděpodobnost výskytu. Základními typy entropických kódů jsou například Huffmanovou a aritmetické kódování. Standard H.264/AVC definuje dvě metody entropického kódování CAVLC (*Context Adaptive Variable Length Coding*) a CABAC (*Context Adaptive Binary Arithmetic Coding*) [1].

3.5 x264

V této práci byl využit kodek x264. Jedná se o open source kodek H.264/MPEG-4AVC v jazyce C vyvíjený Laurent Aimar, Loren Merritt, Eric Petit (OS X), Min Chen (vfw/asm), Justin Clay (vfw), Måns Rullgård, Radek Czyz, Christian Heine (asm), Alex Izvorski, Alex Wright and Jason Garrett-Glaser. Kodek x264 je dostupný na [12].

4. Kodér x264 s ROI

Úkolem této diplomové práce je realizovat kodér H.264/AVC s využitím oblastí zájmů - ROI. Z již dříve provedených testů bylo zjištěno, že neslyšící se při TV vysílání doplněném o mluvčího v českém znakovém jazyce, soustřeďují především na obličej mluvčího a zbytek obrazu sledují spíše periferně. Tento fakt byl zohledněn při úpravě kodéru určeného pro kompresi videa s mluvčím ve znakovém jazyce. Prvním úkolem bylo realizovat detektor jednotlivých oblastí zájmu ve videosignálu. Dalším úkolem pak bylo jednotlivé části obrazu komprimovat v různých kvalitách na základě informací z detektoru.

Kompresí oblastí zájmu pro znakový jazyk se jako jeden z prvních zabýval, ve své doktorské práci [16] vydané v roce 2008, M. Preda. Především se zabýval animací virtuálního mluvčího ve znakovém jazyce. Ve čtvrté kapitole této práce, ve které se zabýval kompresí, se kromě komprese animovaného mluvčího také zaměřil na kompresi reálného mluvčího znakového jazyka. Ke kompresi využil objektového přístupu navrženého ve standardu MPEG-4 Part 2. Cílem bylo zkomprimovat video na maximální bitový tok 64 kb/s pro přenos po telefonní lince tak, aby bylo srozumitelné. Video s reálným mluvčím rozděluje na objekty pomocí klíčování videa. Nejprve testuje obraz jako jeden objekt (NSV). Následně z videa odstraňuje statické pozadí a komprimuje jen samotného mluvčího (NSV2O). Ve třetí variantě se pak věnuje jen kompresi důležitých částí pro znakový jazyk, obličej a ruce (NSV3O). Segmentovaný obraz kóduje pomocí MPEG-4 kodéru a zkoumá bitový tok pro různé úrovně kvantování, velikosti rozlišení a počet snímků za vteřinu.

Z testů vyplývá, že při překročení kvantizačního kroku přes hodnotu 12 video dosahuje tak nízké kvality, že už není srozumitelné. Další experimenty provádí pro různý počet snímků za vteřinu. Nejmenšího datového toku dosahuje u třetího typu videa, kde se zabývá pouze přenosem obličeje a rukou. Při rozlišení 352x280 dostává pod 64 kb/s až při 10 snímcích za vteřinu. S klesajícím rozlišením také klesá bitový tok, ale s tím klesá i kvalita obrazu důležitá pro porozumění znakovému jazyku. Více hodnocení testů naleznete v publikaci[16].

Obr. 4.1: Ukázka komprese NSV3O z [16]

Vzhledem rozšířenosti standardu H.264/AVC, který dnes podporují všechny nové televizory, jsem se rozhodl upravovat právě tento standard. Ve své práci jsem modifikoval nejznámější H.264/AVC open source kodér x264 [12].

O doplnění x264 o oblasti zájmu se již pokusili v roce 2010 na univerzitě v Kapském Městě v JAR, v rámci dizertační práce - Open mobile video communication for the deaf [13]. Zabývali se zde videokomunikací sluchově postižených pomocí mobilních telefonů. Kvalita, v té době, nebyla pro porozumění znakové řeči dostačující, a tak vyvinuli kodér, který zohledňoval oblasti zájmů. Více v dizertační práci[13].

Po nějaké době testování tohoto kodéru jsem ho však nedokázal zprovoznit do funkčního stavu. Po detailním prostudování kódu jsem ale uznal za vhodné využít některých jejich již napsaných funkcí a struktur pro práci s oblastmi zájmu.

4.1 Detektor oblastí zájmů

Pro kompresi oblastí zájmu je nejprve nutné tyto oblasti definovat a najít. Podle již provedených testů [2] je známo, na které části v obrazu se zaměřit. Úkolem je jednotlivé části v obrazu detekovat v reálném čase a předat informaci o nich do kodéru videa. Detektor oblastí zájmu byl naprogramován v jazyce C++ s knihovnou OpenCV[18]. OpenCV je open source knihovna sloužící k počítačovému vidění a ke zpracování obrazu v reálném čase. Kód byl napsán v operačním systému OS X 10.9.2 a k editaci byl použit program Xcode 5 od firmy Apple [19]. Po rozchození detektoru byl poté program upraven i pro operační systém Linux, na kterém také funguje. Program je připraven pro jednoduché ovládání v terminálu, a to jak na OS X, tak na Linuxu.

Detektor rozděluje obraz na základní čtyři části: obličej, ruce, znakovací prostor a zbytek obrazu, který tvoří nedůležitou část pro porozumění mluvčího ve znakovém

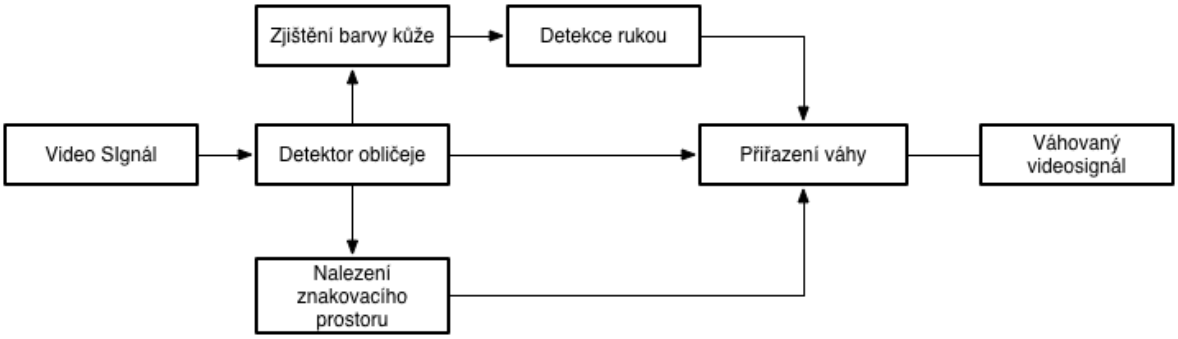
jazyce, viz obr. 4.3. Obraz s těmito částmi je následně rozdělen do bloků 16x16 px. Každému bloku je přiřazena následující váha podle oblasti, ve které blok leží: 0 - pokud jde o část bez oblastí zájmu,1 - znakovací prostor, 2 - ruce, 3 - obličej. Takto naváhovaný obraz je poslán do kodéru x264, který s ním následně pracuje.

První fází celého detektoru je nalezení obličeje. K tomu je využito principu detektoru Viola-Jones, který využívá již natrénované databáze obličejů. Více o problematice vyhledání objektů v obraze v kapitole 3. Detektor postupně projde celý snímek a hledá v něm pozice a velikost obličejů v obraze. Pro přenos mluvčího ve znakovém jazyce předpokládáme videosignál pouze s jedním mluvčím, proto je program nastaven tak, aby vyznačil jen obličej s největším nalezeným obsahem. Pokud obličej není nalezen, program označí snímek jako snímek bez obličeje a celému obrazu nastaví stejnou váhu. Program je také opatřen zabezpečením proti náhodným chybám při hledání obličeje. Pokud nastane chyba v detekci obličeje způsobená například zakrytím obličeje rukou při znakování, nebo špatným nalezením obličeje v jiné části obrazu, program využije nalezený obličej z minulého snímku.

Po detekci obličeje přichází na řadu vyhledání rukou. Kvůli častému pohybu a komplikovanosti tvarů rukou je jejich vyhledávání v obraze obtížnější a není realizovatelné pomocí trénovací množiny. Ruce jsou v obraze nalezeny pomocí detektoru barvy. Lidská barva je po celém těle podobná, a tak detektor využije již nalezeného obličeje a vypočte z něj barvu kůže. Obličej je nejprve převeden z RGB do HSV prostoru. Z výřezu obličeje jsou vypočtené průměrné hodnoty H, S a V. Následně jsou jednotlivé hodnoty HSV všech pixelů v celém obraze porovnány s průměrnými hodnotami. Pro testování detektoru jsem měl k dispozici videosekvence se třemi různými mluvčími. Každý z nich se trochu lišil odstínem pleti viz obr. 4.3, 4.4 a 4.5. Spolehlivé detekce jsem u všech třech mluvčích dosáhl při nastavení prahu ± 60 od všech průměrných hodnot H, S i V. Z toho důvodu jsem se tento práh rozhodl v kódu použít. Porovnáním celého obrazu s prahem HSV je vytvořena maska pro celý obraz obr. 4.3. 4.4 a 4.5. Důležitým předpokladem pro správnou funkčnost detektoru je nahrávání mluvčího v prostředí neobsahující barvu podobnou barvě lidské kůže. Nejvýhodnější je natáčení mluvčího před homogenním pozadím s neutrální barvou.

Další částí vyhledanou detektorem je znakovací prostor. Detektor ho nalezne na základě fyziologických předpokladů lidského těla. Podle nalezeného obličeje v kódu předpokládá, že znakovací prostor je v rozmezí dvou velikostí obličejů na obě strany od nalezeného obličeje v horizontálním směru, půl velikosti obličeje nad původním

vyhledaným obličejem a čtyři velikosti směrem dolů ve vertikálním směru viz obr. 4.3. Znakovací prostor je samostatně váhován z toho důvodu, aby nedocházelo k velkým skokům kvality mezi hlavními oblastmi zájmu a nedůležitou částí obrazu.



Obr. 4.2: Blokové schéma detektoru oblastí zájmu

Obr. 4.3: Ukázka detekce obličeje, znakovacího prostoru a barvy kůže pro mluvčího 1

Obr. 4.4: Ukázka detekce barvy kůže pro mluvčího 2

Obr. 4.5: Ukázka detekce barvy kůže pro mluvčího 3

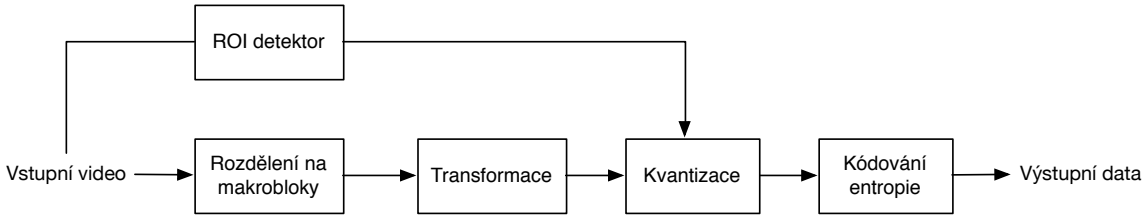
4.2 Kodér

Úkolem kodéru je komprimovat videosignál s využitím informací z detektoru a na základě nich upravit kódování videa tak, aby v oblastech zájmu byl obraz kvalitnější než mimo oblasti zájmu. Zlepšením kvality v oblastech zájmu ale také logicky dochází ke zvýšení bitového toku. Cílem komprese videosignálu pro neslyšící však není jen zlepšit kvalitu v oblastech zájmu, ale i udržet požadovaný výstupní bitový tok. Z toho důvodu musí také dojít ke snížení kvality mimo oblasti zájmu.

Ke ztrátě kvality dochází v kompresi obrazu při kvantizaci. Na základě kvantizačního parametru QP se rozhoduje jaká bude kvalita obrazu. Obecně platí, že čím větší QP, tím horší kvalita videa. V H.264/AVC standardu je QP po celém snímku videa stejný a dosahuje hodnot od 1 do 51. K odhadu QP dochází pomocí algoritmů v bloku Rate-controller, na základě parametrů, kterými jsou: požadovaný bitový tok, požadovaná kvalita, velikost obrazu apod. Pokud má být kvalita v oblastech zájmu zlepšena, je důležité pro bloky s ROI zmenšit jejich hodnotu QP, a naopak ji zvětšit mimo oblasti zájmu.

Kodér nejprve rozděluje jednotlivé snímky videosignálu do makrobloků 16x16 px. Každý makroblok je po transformaci kvantizovaný pomocí odhadnutého QP. Detektor oblastí zájmu rozděluje snímek do stejných makrobloků jako kodér videa a jednotlivým blokům přiděluje váhu, zdali jde o obličej, ruce, znakovací prostor, nebo nedůležitou část obrazu. Kodér tak při práci s jednotlivými makrobloky videosignálu zároveň čte

jejich přidělenou váhu detektorem a na základě toho přiřazuje blokům nově přepočítaný QP.



Obr. 4.6: Blokové schéma kodéru s přidaným detektorem oblastí zájmu

Pro výpočet QP určitých oblastí zájmu se nejprve pomocí vstupního parametru **hpar** určí, o kolik má být QP v oblasti obličeje menší, než původně odhadnutý QP. Následně se vypočtou QP pro hlavu, ruce, znakový prostor, a zbytek obrazu pomocí následujících vztahů

$$QP_{hlava} = QP \cdot \frac{1}{hpar} \tag{4.1}$$

$$QP_{ruce} = QP \cdot \frac{\left(\frac{1}{hpar} + 1\right)}{2} \tag{4.2}$$

$$QP_{zn.\ prostor} = QP \tag{4.3}$$

$$QP_{zbytek} = \frac{(pb \cdot QP) - (a \cdot QP_{hlava} + b \cdot QP_{ruce} + c \cdot QP_{zn.\ prostor})}{d}, \tag{4.4}$$

kde **hpar** je vstupní parametr,který určuje o kolik má být QP obličeje lepší než původní QP, **pb** je celkový počet bloků, **a** je počet bloků hlavy, **b** je počet bloků pro ruce, **c** je počet bloků znakovacího prostoru a **d** je zbývající počet bloků. Pokud nastane situace, že znakovací prostor je přes celý snímek videosignálu, nebo počet bloků nedůležité části obrazu není dostatečný na vyvážení bitového toku, je $QP_{zn.\ prostoru}$ přepočítán podle následujícího vzorce

$$QP_{zn.\ prostor} = \frac{(pb \cdot QP) - (a \cdot QP_{hlava} + b \cdot QP_{ruce})}{d}. \tag{4.5}$$

Na obr. 4.7 je vidět srovnání videa kódovaného bez ROI a s ROI. Video je v rozlišení 576x576 a bylo kódované s předem zadaným výstupním datovým tokem 100 kb/s, parametr hpar je 1.5, to znamená 2/3 původní odhadnuté hodnoty QP. Z obrázku je možné vidět, že při kompresi bez ROI, jsou obličej a ruce ve špatné kvalitě, na to aby neslyšící rozeznal, co mluvčí chce vyjádřit. Oproti tomu při kompresi s ROI jsou obličej a ruce čisté a mluvčímu je dobře rozumět. Nevýhodou je zhoršení kvality pozadí snímku a viditelné artefakty. Tyto artefakty mohou působit velmi rušivě

a odvádět pozornost od mluvčího. Na obr. 4.8 je vidět, že při kompresi s větším bitovým tokem (200 kb/s) je v obraze artefaktů daleko méně. Při detailním pozorování si můžeme všimnout, že se artefakty objevují v místech s častým pohybem, například na hranách mezi pozadím a u límečku košile v horní části rukou, dále na temeni hlavy.

Obr. 4.7: Srovnání videa kódovaného bez ROI a s ROI (100kb/s, rozlišení 576x576px)

Jako původní verze kodéru standardu H.264, posloužil open sorce kodér x264. Tento kodér bylo nutné upravit tak, aby byl schopný pracovat s nalezenými oblastmi zájmu. Kodér x264 je určený pro platformy Linux, OS X a Windows. Funkčnost kodéru s úpravami pro ROI je vyzkoušena pouze na OS X a Linuxu. Kodér se ovládá pomocí příkazu v terminálu. K originálnímu kodéru byli přidány 2 ovládací vstupní parametry. První je **--roi,** který zapne podporu ROI v kódování, a druhý parametr je **-- hpar <float>.** Výčet nejdůležitějších příkazů pro kódování videa s ROI je v tab. 4.1. Více o ovládání x264 v [20].

Obr. 4.8: Srovnání videa kódovaného s ROI (200kb/s, rozlišení 576x576px)

./x264	Volání kodéru x264
--help	Nápověda ovládání
<input>	Vstupní video
-o <output>	Výstupní video
--bitrate<int>	Nastavení požadovaného výstupního toku
--vf resize: width=<int>,height=<int>,method=<string>	Změna velikosti videa, mezi metody patří: fastbilinear, bilinear, bicubic, experimental, point, area, bicublin, gauss, sinc, lanczos, spline
--roi	Zapnutí kodéru s ROI
--hpar <float>	Vstupní parametr, určuje zlepšení QP hlavy

Tab. 4.1: Nejdůležitější ovládací parametry pro X264

5. Subjektivní testy

V praxi má být kodér použit pro doplňkové vysílání mluvčího ve znakovém jazyce přes systém HbbTV. Obraz tlumočníka ve znakovém jazyce bude vkládán většinou do pravého dolního rohu televizní obrazovky. Je tedy zapotřebí zjistit nejen dosažitelný kompresní poměr, ale i vhodné rozlišení obrazu, které v předpokládaném televizním formátu 1920x1080/50p bude odpovídat i relativní velikosti vkládaného obrazu. Pro tyto účely jsem realizoval dva testy.

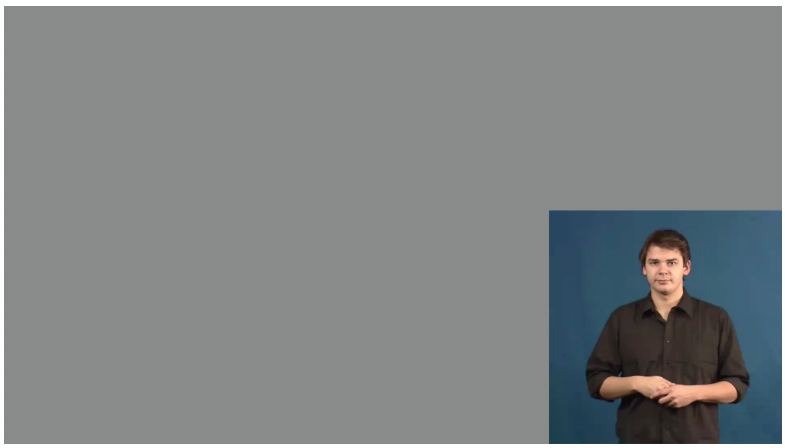
V prvním testuji srozumitelnost v závislosti na kompresním poměru a rozlišení. Protože se ukázalo, že zavedením dynamických oblastí zájmu ROI do komprese obrazu mluvčího znakového jazyka mohou být znakové výrazy srozumitelné i při celkově nízké kvalitě obrazu, zkoumal jsem ve druhém testu subjektivní kvalitu obrazu pro zajištění dostatečného dojmu ze sledování obrazu mluvčího.

Před samotným měřením bylo nejprve nutné rozhodnout, jaké použít rozlišení pro testy. Na konzultaci s Lucií Petráňovou z FFUK jsem se dozvěděl, že neslyšící preferují vyšší rozlišení obrazu (větší relativní obraz) mluvčího znakového jazyka. Proto jsem při volbě rozlišení vyšel ze současného vysílání České televize pro neslyšící a zvolil pro testy jedno rozlišení vyšší (576x576 px), jedno stejné(512x512 px) a jedno nižší (448x448 px). Tyto hodnoty platí pro HD vysílání České televize. Ukázky relativní velikosti doplňkového obrazu pro HD. jsou vidět na obr 5.1.

K testům jsem využil videosekvence mluvčího ve znakovém jazyce, které byly již dříve natočeny na katedře radioelektroniky. Videosekvence byly nahrány ve formátu 16:9 v rozlišení 1920x1080 px. Před samotným měřením jsem tedy videosekvence v programu Final Cut Pro X ořízl na formát 1:1 a exportoval podle uvedených formátů tak, aby nedošlo ke ztrátě kvality.

Dále bylo zapotřebí určit vstupní parametr hpar, který určuje poměr kvality mezi obličejovou částí a okolím obrazu. Tento parametr jsem zvolil 1.5 (2/3 původní hodnoty QP).

Takto upravené videosekvence jsem pomocí programu Motion 5 vložil na šedé pozadí s rozlišením 1920x1080 px.



a)

b)

c)

Obr. 5.1: Ukázky relativních velikostí mluvčího a) 576x576 px, b) 512x512 px ,c) 448x448 px

5.1 Subjektivní testy srozumitelnosti

V tomto testu jsem ověřoval srozumitelnost videosekvencí mluvčího znakového jazyka komprimovaných pomocí kodéru x264 s podporou oblastí zájmu. Úkolem bylo

zjistit, při jakém bitovém toku je pro dané rozlišení obraz znakového mluvčího pro neslyšící diváky již hůře srozumitelný.

Pro testy jsem měl k dispozici 12, již předem nahraných vět ve znakovém jazyce, viz tab. 5.1. Věty byly vybrány tak, aby obsahovaly tzv. minimální páry. Tedy znakové páry, které se liší pouze jedním parametrem. Vybrané znakové minimální páry použité v testu jsou uvedeny v tab. 5.2.

Tím, že v testu byla zkoumána srozumitelnost jednotlivých vět, každý hodnotitel mohl vidět každou větu maximálně jednou. Rozhodl jsem se pro použití následujících bitových toků : 100, 200, 300 a 400 kb/s.

1	Na den dětí škola připravila pro své žáky různé hry.
2	Film, který dávali včera večer v televizi, trval téměř 100 minut.
3	Můj bratr velmi nerad myje nádobí.
4	Na procházce v lese mi přítel povídá: „Vidíš ten hrad? To je nebezpečná oblast.“
5	Ve škole jsme včera psali náročný test.
6	Tato nová socha je vyrobena z papíru.
7	Na tom domě byl velký nápis „Obchod”.
8	Dnes jsem se zeptal babičky na její oblíbené jídlo. Odpověděla mi, že má nejraději jablka.
9	Na bílém tričku je hnědá skvrna.
10	Čtu knihu o botách.
11	Pojeďte na dovolenou do Španělska, je tam úžasný život.
12	V pátek večer jdeme na ples.

Tab. 5.1: Věty použité pro testy srozumitelnosti

PŘÍTEL	VLASTNÍ
JABLKO	OMÁČKA
BOTY	FRANCIE
VČERA	PŘEDEVČÍREM
100	200

Tab. 5.2: Vybrané znakové minimální páry

5.1.1 Testování

Pro testy srozumitelnosti bylo důležité provést testy na prelingválně neslyšících, pro které je český znakový jazyk hlavním komunikačním nástrojem. Proto jsem provedl testy v České unii neslyšících (ČUN) v Praze, kde se neslyšící pravidelně scházejí. K dispozici jsem měl tlumočnici znakového jazyka, která mi pomáhala jak s komunikací s neslyšícími, tak s vyhodnocením testu. Testy jsem prováděl na full HD LCD televizoru SONY BRAVIA KDL 40EX500 s úhlopříčkou 40 palců (102 cm). K dispozici jsem měl místo v kanceláři, ve které se prováděly nejen testy, ale v místnosti stále pracovali i jiní pracovníci unie, podmínky tedy byly omezené. Testy byly provedeny za běžného osvětlení. Hodnotící seděli ve vzdálenosti 1.5 m od televizoru.

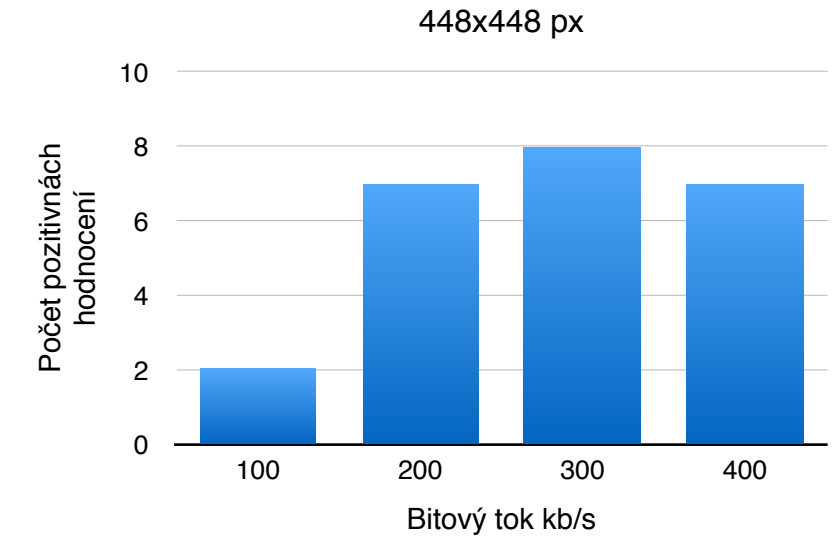
Testy jsem provedl na 10 dobrovolnících ve věku od 29 do 84 let. Většina účastníků však byla starší 60 let. Každému hodnotiteli samostatně jsem promítl 12 vybraných vět, každou v jiném rozlišení s různou kvalitou videosignálu.
Úkolem hodnotitele bylo každou větu zopakovat. Pokud se obsah jejich opakování shodoval s předem zadanou větou a minimální pár byl pochopen správně, vyhodnotil jsem výsledek jako pozitivní. Pokud hodnotitel rovnou nerozuměl, nebo se jeho věta lišila minimálním párem, nebo celým obsahem, vyhodnotil jsem výsledek jako negativní.

5.1.2 Výsledky

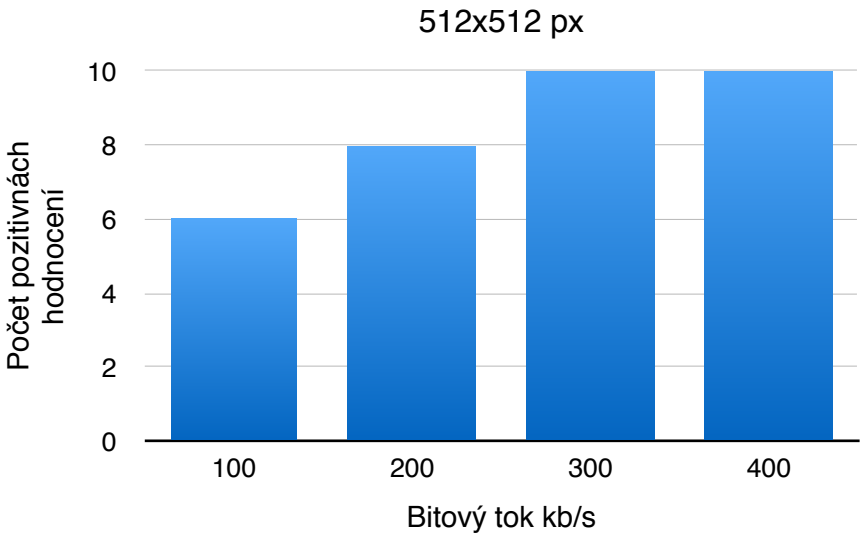
Z výsledku testů je patrné, že více hodnotitelů větám spíše rozumělo. To potvrzuje význam využití kodéru s oblastmi zájmu. I přes nízký bitový tok tak neslyšící dokáže mluvčímu ve znakovém jazyce porozumět. Na výsledcích je podle očekávání také vidět, že s rostoucím bitovým tokem stoupá porozumění obrazu. Pro komprese s bitovým tokem od 200 kb/s výše rozumělo 70%, či více hodnotitelů u všech třech rozlišení, viz obr. 5.2, 5.3, 5.4.

Nejlepších výsledků dosahují videosignály s rozlišením 512x512 px. Pro vyšší obrázek s rozlišením 576x576 px je nižší srozumitelnost nejspíše způsobena tím, že pro větší rozlišení je potřeba více obrazových bodů. Tím stoupá potřebný bitový tok pro přenos a je tím zhoršena kvalita obrazu. Nejnižší rozlišení 448x448 px dosahuje nejnižších výsledků. To potvrzuje i samostatné hodnocení jednotlivých hodnotitelů, kteří mi při testech potvrdili, že při menším rozlišení mají větší problém s vnímáním orálních komponentů, a tím pádem se musí více soustředit na tlumočníka, než na samotný obraz.

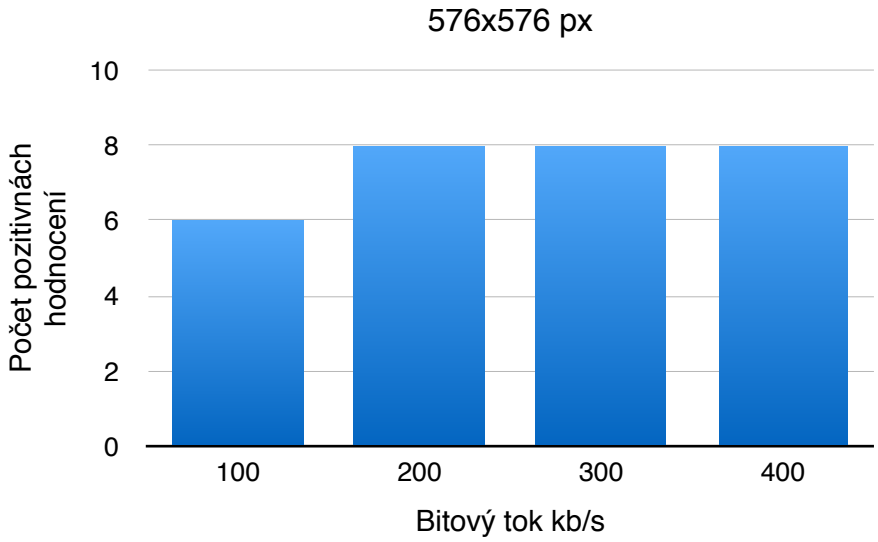
Dalším poznatkem, na kterém se shodli všichni účastníci testu je, že obraz působí velice rušivě při kompresi s nižším bitovým tokem. Artefakty na pozadí, nebo v nedůležitých částech těla odvádí pozornost diváka od samotného mluvčího. Pro porozumění se tedy musejí více soustřeďovat a sledování televizoru pro ně není příjemné.



Obr. 5.2: Výsledky testu srozumitelnosti pro rozlišení 448x448 px



Obr. 5.3: Výsledky testu srozumitelnosti pro rozlišení 512x512 px



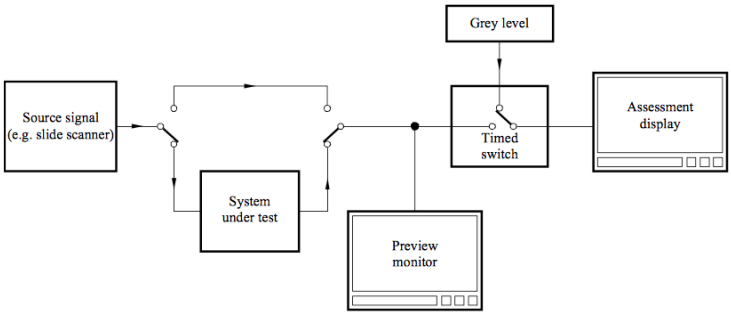
Obr. 5.4: Výsledky testu srozumitelnosti pro rozlišení 576x576 px

5.2 Subjektivní testy kvality

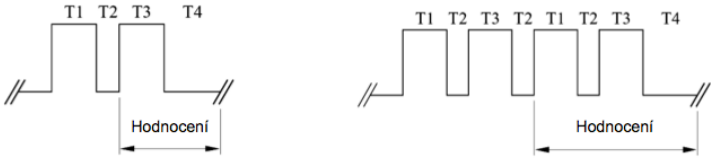
Pro tyto testy byla zvolena metoda DSIS, podle doporučení standardu ITU R BT. 500-11, která srovnáváním s referenčním obrazem poskytuje poměrně dobré výsledky.

5.2.1 Metoda the Double-Stimulus Impairment Scale (DSIS)

Metoda DSIS je založena na srovnání komprimovaného a referenčního snímku. Využil jsem první variantu průběhu testovací sekvence, viz obr. 5.6, ve které respondenti srovnávali referenční a hodnocený obraz bez opakování. Mezi jednotlivými snímky je třívteřinová část se střední úrovní šedé, aby si oči mezi jednotlivými videi odpočinuly. Za každým testovacím snímkem následuje 10 vteřin pro zhodnocení videa. Stupnice hodnocení je vidět v tabulce 5.3. Více o metodě DSIS v [17].



Obr. 5.5: Princip metody DSIS [17]



Obr. 5.6: Varianty testovacích sekvencí DSIS [17]

Hodnota	Popis
1	nepostřehnutelný rozdíl
2	postřehnutelný rozdíl, ale neruší
3	mírně ruší
4	rušivé
5	velmi rušivé

Tab. 5.3: Tabulka hodnocení DSIS [17]

5.2.2 Testování

Subjektivní testy jsem provedl v audiovizuálním studiu na katedře radioelektroniky, ČVUT FEL. Testovací sekvence byly promítány na stejném televizoru SONY BRAVIA KDL 40EX500 jako testy srozumitelnosti. Testování proběhlo v zatemněné místnosti. Pozadí za televizorem bylo ozářeno pomocí běžné žárovky tak, aby nevznikal velký kontrast mezi pozadím a televizorem. Testující seděli ve vzdálenosti 2 m od televizoru. Postupně jsem provedl tři testy, každý pro jedno ze tří rozlišení. Pro každý test bylo vybráno 9 videosekvencí. Jednalo se o šest různých komprimovaných videí, jednu referenci, a dvě opakovaná, již použitá, komprimovaná videa pro ověření správného hodnocení testujících. Bitový tok pro kompresi jsem zvolil tak, aby zahrnoval kvalitu stěží rozlišitelnou od originálu (400 kb/s) až po kvalitu obsahující výrazné kompresní artefakty (50 kb/s) dle tab. 5.4. Pořadí jednotlivých testů je vidět v tab. 5.5.

Testů se zúčastnilo 35 respondentů. Maximální počet pro jeden test bylo 5 lidí. Předem jsem hodnotitele testu informoval o problematice a průběhu testu. Nejprve byla promítnuta ukázka jedné referenční videosekvence a následně s nejvyšší kompresí.

576x576 px	400 kb/s	300 kb/s	250 kb/s	200 kb/s	150 kb/s	100 kb/s
512x512 px	400 kb/s	300 kb/s	250 kb/s	200 kb/s	150 kb/s	100 kb/s
448x448 px	300 kb/s	250 kb/s	200 kb/s	150 kb/s	100 kb/s	50 kb/s

Tab. 5.4: Použité bitové toky

576x576 px	věta1 **300 kb/s**	věta8 **400 kb/s**	věta1 **200 kb/s**	věta6 **REF.**	věta1 **300 kb/s**	věta8 **150 kb/s**	věta6 **100 kb/s**	věta8 **400 kb/s**	věta6 **250 kb/s**
512x512 px	věta6 **250 kb/s**	věta1 **200 kb/s**	věta8 **150 kb/s**	věta1 **300 kb/s**	věta1 **200 kb/s**	věta6 **100 kb/s**	věta8 **400 kb/s**	věta6 **REF.**	věta6 **250 kb/s**
448x448 px	věta8 **50 kb/s**	věta1 **300 kb/s**	věta6 **250 kb/s**	věta1 **200 kb/s**	věta6 **REF.**	věta8 **150 kb/s**	věta6 **100 kb/s**	věta1 **300 kb/s**	věta8 **50 kb/s**

Tab. 5.5: Varianty testovacích sekvencí

5.2.3 Vyhodnocení

Prvním krokem vyhodnocení bylo vyřazení výsledků, které se výrazně lišily od průměru [17]. Po vyřazení nekorektních údajů se ze zbylých hodnot vypočetla střední hodnota, pomocí vztahu

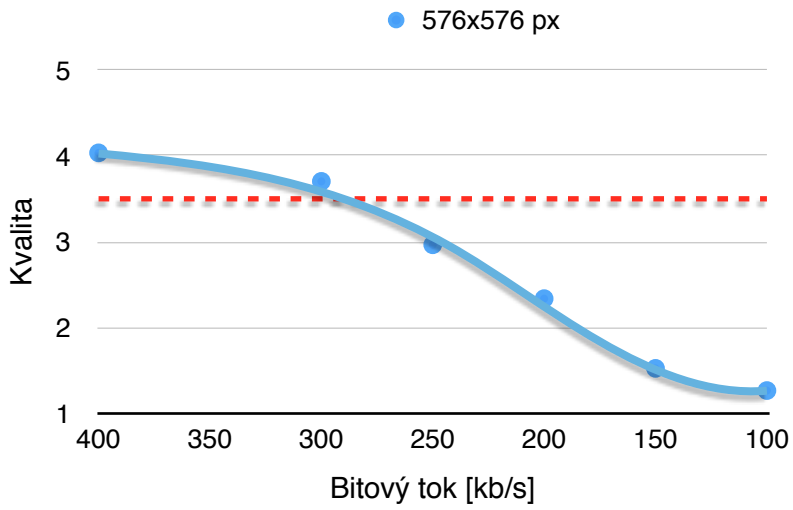
$$\bar{u}_{jkr} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr}, \qquad (5.1)$$

kde $\bar{u}_{jkr}$ je obecný průměr, u_{ijkr} jsou jednotlivé hodnoty z testu, **N** je počet účastníků testu. Index **i** označuje účastníky, **j** podmínky testování, **k** různé typy videosekvencí a **r** označuje pořadí opakování stejného páru sekvencí.

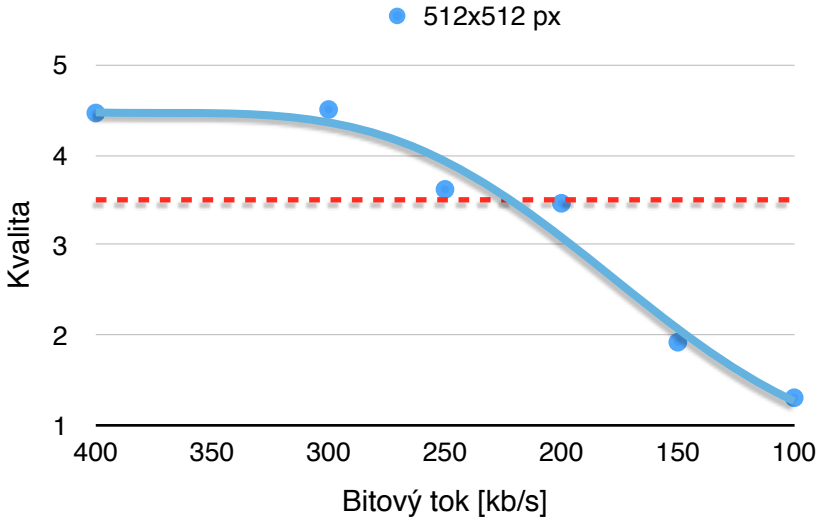
5.2.4 Výsledky

Jednotlivé střední hodnoty jsem vynesl do grafů obr. 5.7, 5.8 ,5.9 , ze kterých jsem dále zkoumal, jaké hodnoty bitového toku lze použít jako minimální při kompresi s oblastmi zájmu. Přijatelné hodnoty kvality pro vysílání jsem se po zvážení s vedoucím práce rozhodl zvolit hodnocení vyšší než 3.5.

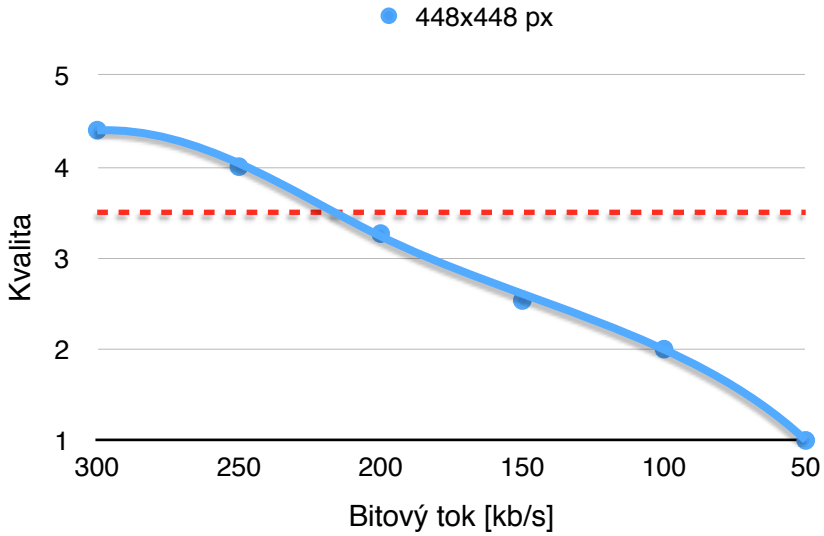
Z jednotlivých grafů je vidět, že pro videosignál s větším rozlišením je zapotřebí většího bitového toku pro kvalitnější obraz, což jsem předem očekával.
Z testů je však vidět, že rozdíly mezi bitovým tokem pro různé velikosti nejsou příliš vysoké, a to především mezi velikostmi 512x512 px a 448x448 px. Pro rozlišení 576x576 px se dá výsledek považovat za kvalitní okolo bitového toku 300 kb/s. Videosignály s rozlišením 512x512 px a 448x448 px se podle výsledků zdají být kvalitativně srovnatelné při stejném bitovém toku. Za kvalitní se dají považovat mezi hodnotami 200 - 250 kb/s. Stejných výsledků nejspíše dosahují z důvodu, že subjektivně na člověka obraz s mluvčím ve znakovém jazyce v nižším rozlišení (tedy menší velikostí) působí kvalitativně hůře, což posunuje hodnocení uživatele směrem dolů.



Obr. 5.7: Výsledky subjektivního testu kvality pro rozlišení 576x576 px



Obr. 5.8: Výsledky subjektivního testu kvality pro rozlišení 512x512 px



Obr. 5.9: Výsledky subjektivního testu kvality pro rozlišení 448x448 px

Závěr

Obsahem této diplomové práce je komprese videosignálu s mluvčím ve znakovém jazyce. Komprese je založena na úpravě v současné době nejrozšířenějšího kodéru H.264/AVC. Úprava spočívá v zavedení dynamických oblastí ROI, které výrazným způsobem zvyšují srozumitelnost komprimovaného obrazu s mluvčím ve znakovém jazyce.

V první kapitole se zabývám českým znakovým jazykem a popisem testu, který posloužil k nalezení oblastí zájmu ROI. Ve druhé kapitole popisuji teoretický princip detekce oblastí ROI pomocí detektoru Viola-Jones a detektoru založeném na HSV barevném prostoru. Ve třetí kapitole je popsán základní princip komprese videosignálu, a to především standard H.264/AVC. V následující kapitole se věnuji praktické realizaci detektoru a zavedení oblastí zájmu do softwarového kodéru x264. Tato úprava umožňuje řízeně měnit poměr kvality jednotlivých částí obrazu. V poslední kapitole jsou popsány dva testy. První je test srozumitelnosti, který byl proveden s neslyšícími v České unii neslyšících . Výsledky testu potvrdily, že komprese videa s oblastmi zájmu ROI pomáhá neslyšícím k úspěšnému porozumění mluvčímu ve znakovém jazyce, komprimovaného s větším kompresním poměrem. Nevýhodou je však zhoršení kvality mimo oblasti zájmu, která sice nezpůsobuje snížení srozumitelnosti, ale zhoršuje celkový subjektivní dojem ze sledování obrazu. Z tohoto důvodu je potřeba volit parametr, který určuje poměr kvality jednotlivých částí obrazu, velmi opatrně. Proto jsem provedl i testy kvality. Cílem testů bylo nalézt, při kterém bitovém toku vnímá divák obraz jako dostatečně kvalitní při různých velikostech obrazu (rozlišeních).

Z výsledků je zřejmé, že je zbytečné uvažovat o vysílání v rozlišení 448x448 px, protože požadovanou kvalitu získáme překvapivě při stejném bitovém toku, jako při rozlišení 512x512 px. Pokud divák bude vyžadovat menší obraz, nižší rozlišení tlumočníka do znakového jazyka, bude si moci obraz zmenšit. Otázkou je, zdali se vyplatí uvažovat o vysílání v rozlišení 512x512 px, případně rozlišení 576x576 px. Videosignál s rozlišením 576x576 px se dá podle testů považovat za kvalitní při kompresi s bitovým tokem 300 kb/s. To je zhruba o 75 kb/s více, než při rozlišení 512x512 px. Na základě diskuze s neslyšícími, kteří preferovali větší pole pro mluvčího i za cenu vyššího překrytí televizního obrazu, doporučuji využít rozlišení 576x576 px.

Rozdíl 75kb/s v současné době nemusí být při přenosu po internetu podstatný, zmenšit si obraz může divák vždy.

Za uvážení do budoucnosti stojí, zda při vysílání využít adaptivního kodéru, který by při zmenšení obrazu v televizoru, nebo při snížení dostupného bitového toku, automaticky přepnul rozlišení na vysílací straně.

Použitá literatura

[1] RICHARDSON, Iain E. H.264 and MPEG-4 video compression: video coding for next-generation multimedia. Chichester: Wiley, 2003, 281 s. ISBN 04-708-4837-5.

[2] ŠVACHULA, Zdeněk, Martin BERNAS, Petr ZATLOUKAL a Lucie KLABANOVÁ. Investigation of Regions of Interest for Deaf in Czech Sign Language Content. International Journal of Software and Web Sciences (IJSWS). roč. 2013, č. 3, s. 44-49. Dostupné z: http://iasir.net/IJSWSpapers/IJSWS12-344.pdf

[3] ZATLOUKAL, Petr, Zdeněk ŠVACHULA, Martin BERNAS. Relationship between Intelligibility and Visual Quality in Czech Sign Language as a Supplement to Television Broadcasting. International Journal of Software and Web Sciences (IJSWS). roč. 2013, č. 5, s. 1-6. Dostupné z: http://iasir.net/IJSWSpapers/IJSWS13-205.pdf

[4] MACUROVÁ, Alena. Poznáváme český znakový jazyk. Speciální pedagogika. roč. 2001, č. 2, s. 69-75. Dostupné z: http://ruce.cz/clanky/8-poznavame-cesky-znakovy-jazyk-i

[5] VIOLA, PAUL a MICHAEL J. JONES. Robust Real-Time Face Detection. International Journal of Computer Vision. roč. 2004, č. 57, 137–154. Dostupné z: http://www.vision.caltech.edu/html-files/EE148-2005-Spring/pprs/viola04ijcv.pdf

[6] MAŠEK, Jan. Detekce objektů v obraze s pomocí Haarových příznaků. Brno, 2012. Diplomová práce. VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ.

[7] PŘINOSIL, Jiří a Martin KROLIKOWSKI. Využití detektoru Viola-Jones pro lokalizaci obličeje a očí v barevných obrazech. Elektrorevue. roč. 2008, č. 31 [cit. 2014-04-27]. Dostupné z: http://www.elektrorevue.cz/cz/download/vyuziti-detektoru-viola-jones-pro-lokalizaci-obliceje-a-oci-v-barevnych-obrazech/.

[8] VIOLA, Paul a Michael JONES. Rapid Object Detection using a Boosted Cascade of Simple Features.[online] 2001. Dostupné z: https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf

[9] SARAVANAKUMAR, S., A. VADIVEL, C. G. Saneem AHMED a A. VADIVEL. Multiple object tracking using HSV color space. Proceedings of the 2011 International Conference on Communication, Computing. New York, New York, USA: ACM Press, 2011, s. 247-. DOI: 10.1145/1947940.1947993. Dostupné z: http://portal.acm.org/citation.cfm?doid=1947940.1947993

[10] CERLINCA, Tudor Ioan, Stefan Gheorghe PENTIUC, Radu Daniel VATAVU a Marius Cristian CERLINCA. Hand posture recognition for human-robot interaction. Proceedings of the 2007 workshop on Multimodal interfaces in semantic interaction - WMISI '07. New York, New York, USA: ACM Press, 2007, s. 47-50. DOI: 10.1145/1330572.1330580. Dostupné z: http://portal.acm.org/citation.cfm?doid=1330572.1330580

[11] BHAT, Vandana S. a Jagadeesh D. PUJARI. Face detection system using HSV color model and morphing operations. Proceedings of National Conference on 'Women in Science & Engineering' (NCWSE 2013), SDMCET Dharwad. 2013, s. 200-204. Dostupné z: http://inpressco.com/wp-content/uploads/2013/09/Paper39200-204.pdf

[12] *X264: A free h264/avc encoder* [online]. [cit. 27.04.2014]. Dostupné z: http://www.videolan.org/developers/x264.html

[13] LAIDLER, Chris. Open mobile video communication for the deaf. JAR, 2010. PhD These. University of Cape Town. Dostupné z: http://people.cs.uct.ac.za/~claidler/MoVidCom/files/Thesis_LDLCHR004.pdf.

[14] CHEN, Jian-Wen, Chao-Yang KAO a Youn-Long LIN. Introduction to h.264 advanced video coding. Asia and South Pacific Conference on Design Automation, 2006. IEEE, 2006, s. 736-741. DOI: 10.1109/ASPDAC.2006.1594774. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1594774

[15] Rate Control and H.264. In: [online]. [cit. 27.04.2014]. Dostupné z: http://www.pixeltools.com/rate_control_paper.html#mad

[16] MARIUS, PREDA. Advance virtual character animation within the MPEG - 4 framework. Paříž, 2008. PhD These. UNIVERSITE RENE DESCARTES - PARIS V.

[17] Rec. ITU-R BT.500-11. *Methodology for the subjective assessment of the quality of television pictures*. Geneva: ITU, 2012. Dostupné z: http://www.cecs.uci.edu/~papers/aspdac06/pdf/p736_7D-1.pdf

[18] *OpenCV: Open Source Computer Vision Library* [online]. [cit. 27.04.2014]. Dostupné z: http://opencv.org

[19] Xcode 5: [software].[přístup: 6.5.2014]. Dostupné z: https://developer.apple.com

[20] Project357: X264_Settings. [online]. [cit. 27.04.2014]. Dostupné z: http://mewiki.project357.com/wiki/X264_Settings

Seznam obrázků

Seznam tabulek