



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

**Fakulta elektrotechnická
Katedra radioelektroniky**

**Objektivní hodnocení kvality obrazu mluvčího
znakového jazyka**

**Objective quality evaluation of image of sign
language speaker**

Bakalářská práce

Studijní program: Komunikace, multimédia a elektronika

Studijní obor: Multimediální technika

Vedoucí práce: Ing. Martin Bernas, CSc.

Student: Marek Rousek

2016

České vysoké učení technické v Praze
Fakulta elektrotechnická

katedra radioelektroniky

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: **Marek Rousek**

Studijní program: Komunikace, multimédia a elektronika
Obor: Multimediální technika

Název tématu: **Objektivní hodnocení kvality obrazu mluvího znakového jazyka**

Pokyny pro vypracování:

Prostudujte specifické charakteristiky obrazu mluvího znakového jazyka a objektivní metody hodnocení kvality obrazu. Na jejich základě navrhnete metodu vhodnou pro objektivní hodnocení kvality obrazu mluvího znakového jazyka, metodu naprogramujete v prostředí Matlab a ověřte porovnáním se subjektivním hodnocením.

Seznam odborné literatury:

- [1] Wu,H.R.;Rao,K.R.: Digital Video Image Quality and Perceptual Coding. Taylor&Francis, 2006, ISBN-13: 978-0-824-7-2777-2
- [2] Kuchařová,L. Jazyk neslyšících. Výzkumná zpráva. Ústav českého jazyka a teorie komunikace, FFUK, Praha 2005.

Vedoucí: Ing. Martin Bernas, CSc.

Platnost zadání: do konce zimního semestru 2016/2017

L.S.

doc. Mgr. Petr Páta, Ph.D.
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 8. 1. 2016

Anotace

Bakalářská práce ve své teoretické části pojednává o modelu lidského vidění HVS, který lze využít pro objektivní hodnocení kvality obrazu v oblasti televizní techniky. Na základě přehledu jednotlivých objektivních metod a metrik je vybrána a podrobně zpracována metrika MSSIM, která je použita při návrhu metody hodnocení kvality obrazu mluvčího znakového jazyka. Závěr teoretické části se zabývá problematikou oblastí zájmu neslyšících při sledování pořadů s tlumočníkem do znakového jazyka.

V praktické části práce byla navržena, realizována a ověřena objektivní metoda hodnocení kvality videosekvence mluvčího znakového jazyka. Metoda je naprogramována v prostředí MATLAB, využívá metriku MSSIM a je rozšířena o respektování oblastí zájmu neslyšících. Metoda byla ověřena pomocí série videosekvencí obsahujících mluvčího znakového jazyka v různých bitových rychlostech a nastavených parametrech. Výsledky těchto testů byly porovnány s referenčními subjektivními testy, pomocí kterých byla navržená metoda kalibrována.

Klíčová slova: MSSIM, objektivní hodnocení kvality, MATLAB, ROI, znakový jazyk, videosekvence, obrazová komprese

Abstract

Bachelor's thesis in its theoretical part discusses the model of human vision HVS, which can be used for evaluation of image quality in television technology. There are analyzed objective evaluation methods and metrics of image quality. It is analyzed in detail MSSIM metric that is used in the design of methods for evaluation of image quality of sign language speaker. Conclusion of the theoretical part describes the problem areas of interest for the deaf when watching programs with an interpreter to the sign language.

In the practical part it was designed, implemented and verified an objective method for evaluating the quality of a movie of sign language speaker. The method is programmed into the environment MATLAB, it uses metrics MSSIM and it is extended to respect areas of interest for the deaf. The method was validated through a series of movies containing of sign language speaker in a variety of bit rates and parameter settings. Results of these tests were compared with reference subjective test by which the method was calibrated.

Keywords: MSSIM, objective quality assessment, MATLAB, ROI, sign language, movie, image compression

Poděkování

Rád bych poděkoval vedoucímu práce, panu Ing. Martinu Bernasovi, CSc., za velmi vstřícný, přátelský přístup a cenné připomínky a rady při zpracování mé bakalářské práce. Mé díky patří Ing. Karlovi Fliegelovi, Ph.D., za poskytnutí výkonné výpočetní techniky pro testování navržené metody, a také panu Ing. Petru Zatloukalovi, Ph.D., za cenné rady a poskytnutý testovaný materiál. V neposlední řadě chci také poděkovat své rodině, speciálně mé mamince, která mi byla oporou a pomohla se závěrečnou korekturou textu práce.

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 závěrečných prací.“

V Praze dne

.....
Podpis studenta

Obsah

Úvod	8
1 Teoretické základy pro lidský zrakový systém	10
1.1 Funkce oka.....	10
1.2 Psycho-fyzikální vlastnosti HVS.....	11
1.2.1 Foveální a periferní vidění.....	11
1.2.2 Adaptace na světlo.....	12
1.2.3 Citlivost na kontrast.....	12
1.2.4 Maskování.....	13
1.2.5 Diferenční metrika.....	14
1.3 Oční pohyby.....	14
1.3.1 Základy očních pohybů.....	14
1.3.2 Fixace.....	14
1.3.3 Sakády.....	15
2 Subjektivní metody hodnocení kvality obrazu	16
2.1 Metodiky subjektivního testování.....	16
2.2 Metody DS (Double Stimulus).....	16
2.2.1 DSIS.....	16
2.2.2 DSCQS.....	17
2.3 Metody SS (Single Stimulus).....	17
2.3.1 SSM.....	17
2.3.2 SSCQE.....	17
2.4 Výhody a nevýhody subjektivního testování.....	17
3 Objektivní metody hodnocení kvality obrazu	18
3.1 MSE a PSNR.....	18
3.2 Metody měření bez reference (NR).....	19
3.3 Metody měření s částečnou referencí (RR).....	19
3.4 Metody měření s plnou referencí (FR).....	20
3.4.1 Lubinův model HVS.....	20
3.4.2 VDP (Visible Differences Predictor)	21
3.4.3 Watsonův DCT model.....	22
3.4.4 Sarnoffův JND model.....	23
3.4.5 SSIM (Structural Similarity Index)	24

4 Charakteristika obrazu mluvěcího znakového jazyka	29
4.1 Znakový jazyk.....	29
4.2 Oblasti zájmu neslyšících diváků.....	30
4.3 Určení oblastí zájmu neslyšících diváků.....	31
4.4 Praktická implementace oblastí zájmu neslyšících diváků.	32
4.4.1 HbbTV a asistivní technologie.....	32
5. Testované videosekvence	33
5.1 Detekce ROI.....	33
5.2 Hpar.....	33
5.3 Kódování videosekvencí.....	34
6 Praktická realizace objektivní metody	35
6.1 Předpoklady pro návrh objektivní metody.....	35
6.2 Popis navržené metody.....	35
6.2.1 Načtení videosekvencí.....	36
6.2.2 Načtení masek.....	36
6.2.3 Výpočet algoritmu metriky.....	38
6.2.4 Zajištění aditivnosti metody.....	40
6.2.5 Konečné hodnota objektivní kvality obrazu.....	41
6.3 Uživatelský panel.....	44
7 Vyhodnocení objektivní kvality videosekvencí	46
7.1 Výsledky objektivní kvality obrazu u originální SSIM.....	46
7.2 Výsledky objektivní kvality obrazu u SSIM MATLABu...	50
7.3 Ohodnocení funkčnosti metody.....	54
Závěr	56
Seznam použitých obrázků	58
Seznam tabulek	59
Seznam použitých symbolů a zkratk	60
Seznam použité literatury	62
Obsah přiloženého DVD	65

Úvod

V současné době probíhá velký rozvoj multimediální techniky. Parametry multimediálních zařízení se každým rokem o stupeň zlepšují a nabízejí stále více funkcí a možností. S tím souvisí rozvoj v mnoha odvětvích, ve kterých lze tato zařízení využít díky jejich možnostem. Ve vyspělých zemích lze již nějakou dobu sledovat snahu vytvářet společensky rovné šance a příležitosti. Člověk se zdravotním handicapem, jakým je např. postižení sluchu nebo zraku, se za určitých okolností často dokáže plnohodnotně zapojit do pracovního života. Důležité ale je také zajistit handicapovaným lidem zlepšení v odvětví společenském.

Ve většině vědních oborů je věnována pozornost oblastem, které by přispěly ke zvýšení kvality života zdravotně postižených lidí. Pokud přeskočíme pomůcky na medicínské bázi, označují se dané postupy a principy termínem *asistivní technologie*. Tento termín zahrnuje veškeré hardwarové a softwarové prostředky, které nějakým způsobem pomáhají zdravotně postižené populaci usnadňovat jejich život. V rámci výzkumu týkající se neslyšící komunity takovou službu představuje přirozené tlumočení do znakového jazyka, které lze stále častěji spatřit ve vizuálních sdělovacích prostředcích, především v televizi.

Tato bakalářská práce je součástí rozsáhlého projektu, jenž si klade za úkol vytvořit automatizovaný animovaný model tlumočnicka do znakového jazyka, který by mohl být posléze použit tam, kde dochází ke kontaktu se sluchově postiženými a není k dispozici tlumočnick (pošta, úřady, banky, nádraží aj.). Za primární se považuje nasazení modelu tlumočnicka znakového jazyka do televizního vysílání. Tato práce vychází z již zpracovaných částí tohoto projektu. Cílem této práce je nalezení vhodné objektivní metody pro hodnocení obrazu mluvího znakového jazyka. Tato metoda by mohla poté nahradit subjektivní hodnocení kvality obrazu, které je v současné době jediným standardem pro určení kvality obrazu.

Zkoumání objektivní kvality obrazu je v současné době důležitým tématem díky prudkému rozvoji multimédií, internetu a mobilní komunikace. Důsledkem tohoto rozvoje je zavádění nových kompresních metod obrazu splňujících kritéria efektivního přenosu a dostatečné kvality obrazu. Díky tomu je tedy snaha nalézt vhodné efektivní, levné a časově nenáročné kritérium, podle kterého by bylo možné kvalitu obrazu hodnotit.

Teoretická část práce postupně rozebírá: lidský zrakový systém a jeho vlastnosti (kapitola 1), nejčastěji používané subjektivní metody hodnocení kvality obrazu (kapitola 2), nejznámější objektivní metody, které zásadně přispěly k vývoji ve svém odvětví (kapitola 3), oblasti zájmu neslyšících diváků (kapitola 4) a na jejich základě navržené kódování videosekvencí pro přenos mluvího znakového jazyka (kapitola 5).

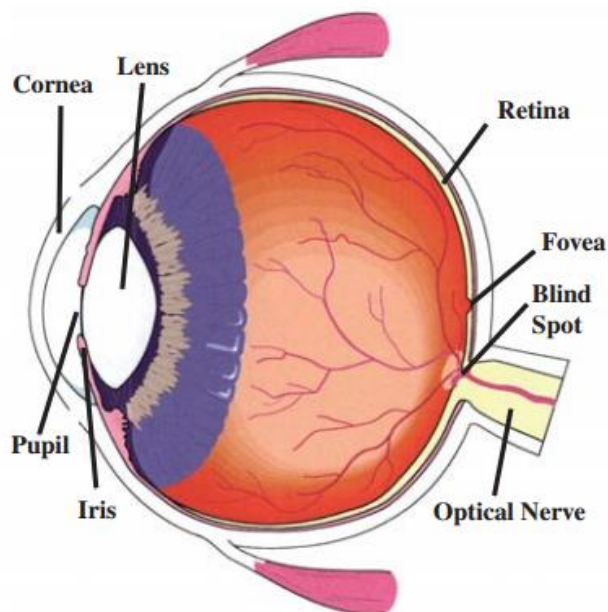
Praktická část se zabývá vlastním návrhem objektivní metody, která byla naprogramována v prostředí MATLAB (kapitola 6). V poslední části jsou zhodnocené výsledky zjištěné touto metodou porovnány se subjektivními testy a je ověřena funkčnost metody (kapitola 7).

Kapitola 1

Teoretické základy pro lidský zrakový systém

Pro vytvoření dobře fungující objektivní metody pro měření kvality obrazu je nutné správně extrahovat z obrazu vlastnosti, které co nejpřesněji odpovídají lidskému vnímání a lidskému zrakovému systému, označovanému též HVS (Human Visual System). K odvození takové metody proto potřebujeme znát základy zrakového systému a jeho vlastnosti, ovlivňující výsledný vjem toho, co vidíme.

1.1 Funkce oka



Obr. 1.1: Schématický řez lidským okem [2]

Obr. 1.1 znázorňuje příčný řez optickými komponenty lidského oka. Na začátku procesu vidění je dopadající světlo zaostřováno čočkou (lens) a prochází přes průhlednou výplň oka, tzv. sklivce (corpus vitreum), než dopadne na sítnici (retina). Sítnice je tenká vrstva tkáně obsahující dva druhy fotoreceptorů (tyčinky a čípky). Fotoreceptory obsahují fotopigmenty (Rodopsiny a Jodopsiny) a složitou kaskádu biochemických drah, které umožňují vznik nervového impulsu. Ten převádí příchozí světlo na elektrické impulsy, které jsou následně přenášeny optickým nervem (Optical nerve) do mozku.

Důležitou součástí oka, která má zásadní vliv na vnímaný obraz, jsou fotoreceptory: tyčinky a čípky. Tyto buňky snímají na sítnici dopadající světlo. Čípky jsou barevně citlivé receptory zajišťující vidění za normálních světelných podmínek, neboli zajišťují fotopické vidění. Existují 3 druhy čípků: L-čípky, M-čípky a S-čípky, odpovídající dlouhým (Long), středním (Medium) a krátkým (Short) vlnovým délkám, na něž mají nejvyšší citlivost. Čípky dělí obraz projektovaný na sítnici na tři vizuální proudy, které můžeme chápat jako tři různé barevné složky světla (červená, zelená a modrá). Tyčinky naproti tomu jsou citlivé na jas a tak se starají zejména o vidění za zhoršených světelných podmínek, neboli zajišťují skotopické vidění. Oba druhy fotoreceptorů nejsou na sítnici rozloženy rovnoměrně. Čípky mají nejvyšší hustotu ve foveě (jamce), která leží na vizuální ose vidění. S rostoucí vzdáleností od fovey se koncentrace čípků rapidně snižuje a naopak roste hustota tyčinek. Stejně tak distribuce gangliových buněk (tj. neuronů, které přenášejí elektrické impulsy z oka do mozku prostřednictvím optického nervu) je největší ve foveě a dále jejich počet klesá podobně jako u čípků.

Celkovým výsledkem je, že HVS není schopen vnímat celý vizuální podnět v rovnoměrném rozlišení [3]. Při modelování HVS se klade největší důraz právě na fotopické vidění. To znamená, že světelné podmínky budou normální. Tento popis je jen jednoduchým přiblížením celého procesu, který je ve své podstatě velmi složitý a detailněji je publikován např. v [1] a [2].

1.2 Psycho-fyzikální vlastnosti HVS

Pro samotné modelování objektivní metody respektující HVS musíme lépe pochopit jednotlivé jevy vidění, které mají vliv na výsledný zrakový vjem po jeho dopadu na sítnici. Ty nejzákladnější si přiblížíme v následující části.

1.2.1 Foveální a periferní vidění

Jak bylo řečeno výše, hustota čípků a gangliových buněk na sítnici není rovnoměrně rozdělena a je nejvyšší v oblasti jamky. S rostoucí vzdáleností od jamky hustota čípků a gangliových buněk výrazně klesá. Důsledkem toho je, že jak roste vzdálenost od pozorovaného bodu, klesá prostorové rozlišení obrazu. Toto vidění se sníženým rozlišením se nazývá periferní. Vidění s vysokým stupněm rozlišení, ke kterému dochází ve chvíli, kdy se pozorovatel na daný bod zaměří, se naopak nazývá foveální.

Většina modelů pro objektivní hodnocení kvality obrazu pracuje pouze v režimu foveálního vidění. Existují ale i některé modely, které počítají i s viděním periferním. Ty ale potřebují pro svou správnou funkčnost získat navíc informace o tom, kam se pozorovatel dívá. Tak je možné správně rozlišit, na kterou část obrazu může být aplikován daný druh vidění a tedy i příslušná metoda.

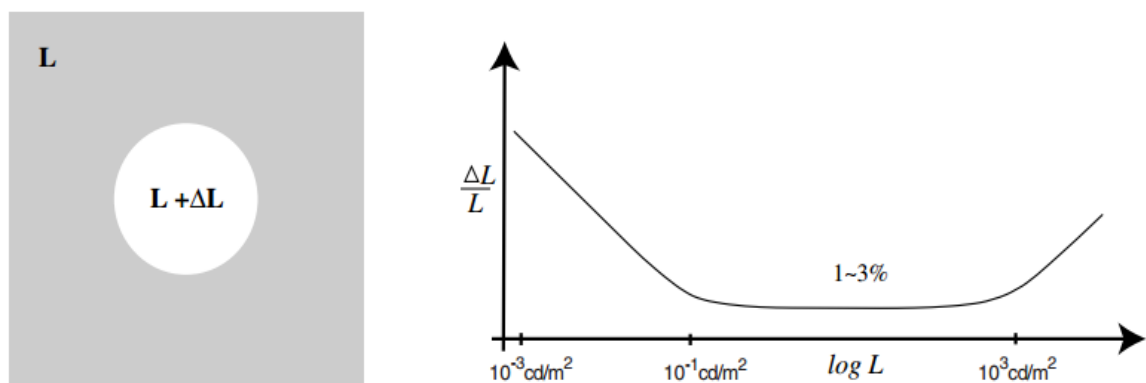
1.2.2 Adaptace na světlo

Lidské oko se dokáže přizpůsobit velmi širokému rozsahu hodnot intenzity světla. Od oblasti vidění při měsíčním svitu až po vidění při ostré sluneční záři. Tato schopnost funguje na základě regulace množství světla vstupujícího do oka pomocí zornice a také díky adaptačním mechanismům ve vlastních buňkách sítnice. Výsledkem je závislost, kterou popisuje Weber-Fechnerův zákon. Může být vyjádřen jako [4]:

$$\Delta S = k * \frac{\Delta L}{L}, \quad (1.0)$$

kde ΔS je změna intenzity subjektivního vjemu, k je Weberova konstanta, ΔL je nejmenší rozpoznatelný rozdíl jasu na daném pozadí a L je fyzikální intenzita podnětu na receptor. Vztah lze chápat tak, že subjektivně vnímaná malá změna počítku je přímo úměrná malé změně intenzity podnětu a nepřímo úměrná intenzitě počítku [34].

Tuto závislost lze měřit jednoduchým psychovizuálním experimentem popsaným na obr. 1.2. Měřicí obraz je složen z homogenního pozadí s jasem L a skvrny s jasem $L + \Delta L$. Roste-li jas pozadí od velmi tmavé po extrémně světlou, je naměřen právě rozpoznatelný rozdíl $\Delta L/L$ (JND). Ten je nutný pro zpozorování skvrny. Výsledek je na obr. 1.2. Konstantní část se nazývá Weber-Fechnerou částí a rozsah JND je v této části 1-3 %. To znamená, že jasový rozdíl 1-3 % je dostatečný pro vjem, je-li osvětlení pozadí v rozsahu 0,1 až 1000 cd/m² (oblast skotopického vidění). Z toho vyplývá důležitost lokálního jasu i z hlediska vnímání kvality obrazu [8].

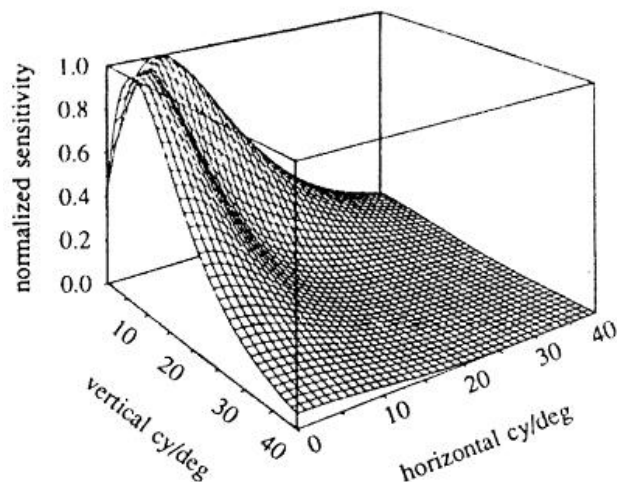


Obr. 1.2: Weber-Fechnerův zákon pro dvoubarevný experiment a jeho grafické vyjádření [2]

1.2.3 Citlivost na kontrast

Z výše uvedené adaptace na světlo vyplývá, jak je lidské oko citlivé na kontrast. Weber-Fechnerův zákon je stále velmi jednoduchý a nepostihuje kompletně celou jasovou adaptaci HVS. Ta nevychází jen z pozadí vizuálního podnětu, ale je

utvářena z celého vzoru obrazu. Proto je nutné definovat hodnoty kontrastu v příslušných frekvenčních pásmech. Tento popis představuje funkce kontrastní citlivosti CSF (Contrast Sensitivity Function) zobrazená na obr. 1.3. Tato funkce modeluje variace v citlivosti HVS na různé prostorové a časové frekvence, které jsou přítomny ve vizuálním podnětu. Ze závislosti Weber-Fechnerova zákona je vidět, že lidský zrak je nejcitlivější na střední prostorové frekvence. Pro vyšší prostorové frekvence citlivost rychle klesá. Pro nižší frekvence také klesá, ale pomaleji. To lze vysvětlit pomocí vlastností receptivních polí gangliových buněk nebo jako interní šumové vlastnosti neuronů HVS.



Obr. 1.3: Grafické znázornění CSF dle Dalyho [5]

V závislosti na těchto vlastnostech některé modely HVS implementují CSF jako filtrovací činnost, zatímco jiné implementují CSF jako zatěžovací faktory pro vlnová subpásma po frekvenční dekompozici. CSF také bere v úvahu vzdálenost od jamky, ale pro foveální vidění (v blízkosti jamky) se CSF modeluje jako prostorově invariantní funkce pásma propustnosti. Díky tomu metoda pro objektivní hodnocení kvality bere více v potaz změny v pozorovací vzdálenosti.

1.2.4 Maskování

Maskování (masking) je důležitou vlastností jak u sluchu, tak i u vidění. Ve své podstatě maskování představuje, že daný vjem původního signálu je potlačován jiným maskovacím signálem (maskovačem). Existuje celá řada různých maskovacích efektů, jako například: maskování kontrastu, maskování aktivity, maskování barvy, časové maskování atd. Maska obecně snižuje viditelnost původního signálu oproti skutečnosti, kdy není přítomna. V některých případech může ale detekci usnadňovat. Maskovací efekt je nejsilnější, když maska i testovaný signál mají stejný nebo podobný obsah a orientaci frekvencí.

Většina modelů hodnocení kvality obrazu využívá některý z modelů maskování. Jeho využití je důležité zejména v oblasti obrazové komprese, kdy jím lze maskovat nežádoucí artefakty. To se nejčastěji používá při maskování kontrastu

nebo barvy. Maskování kontrastu se využívá zejména v případech, kdy mají signál i maskovač podobné vlastnosti nebo vzhled a lze tak zakrýt nežádoucí rušení originálu. Maskování barvy má podobný charakter jako maskování kontrastu. Rozdíl je v tom, že poškození obrazu se nachází pouze v jednom barevném kanálu. Zde je pro HVS nejcitlivější zejména zkreslení v oblasti pleťové barvy.

1.2.5 Diferenční metrika

Existují i další jevy vidění, které mají vliv na výsledný zrakový vjem, jakými jsou např. prahování, vnímání barvy, Stiles-Crawfordův efekt atd., které jsou podrobněji popsány v [1] a [2]. Všechny tyto jevy mají rozdílný vliv na výsledný efekt vnímaného obrazu. Pro hodnocení kvality obrazu modelem HVS je proto nutné vytvořit diferenční metriku. Ta představuje finální vyhodnocení kvality na základě řady získaných parametrů, které vyplývají z jednotlivých jevů vidění. Výsledkem může být též mapa viditelných rozdílů (JND mapa). Praktickou implementaci v různých objektivních metodách uvádí kapitola 3.

1.3 Oční pohyby

V předchozí části byly popsány jevy, které mají vliv na obraz zpracovaný okem až po dopadu na sítnici. V kontextu této práce je ale také nutné se zaměřit na skutečnost, že vnímaný obraz zejména závisí na tom, kam je upírána pozornost pozorovatele.

1.3.1 Základy očních pohybů

Oči se pohybují neustále, a to kvůli získání komplexní informace z celého zorného pole (viz kap. 1.2.1.). Oční pohyby lze rozdělit do těchto základních skupin [28]: fixace, pohyby pro přesun pohledu, pohyby stabilizují pohled a adaptační pohyby.

Veškeré oční pohyby zajišťuje šestice svalů, které dostávají impulsy z centrální nervové soustavy. Tyto svaly se starají o šest základních pohybů: doprava, doleva, nahoru, dolů, a obě rotace podle optické osy oka. Tím získává HVS dostatečné množství informací, aby mohl docílit plné pozornosti v celém zorném úhlu. V kombinaci s pohyby hlavy dokáže HVS obdržet informace o celkovém charakteru scény. Oční pohyby ve většině případů nepřesahují výchylku 15°. Při vyšších výchylkách očí, například do stran, již zorné pole vzdálenějšího oka omezuje nos a vjem se stává monokulární [28].

1.3.2 Fixace

Ze všech čtyř základních skupin očních pohybů je právě fixace nejdůležitější, protože při ní dochází k získání informací o pozorovaném objektu zájmu (části scény). Ostatní pohyby slouží k přesunu očí k jinému místu fixace. Při fixaci se pohyb očí „zastaví“, dochází k zaostření pozorovaného objektu, obraz tím míří

do jamky a dochází k foveálnímu vidění (viz kap. 1.2.1). Slovo „zastaví“ je v uvo-
zovkách, protože i při fixaci dochází stále k miniaturním pohybům, které jsou
způsobeny podvědomými impulsy (nedokážeme je ovlivnit). Tyto miniaturní po-
hyby probíhají buďto z patologického hlediska nebo ze snahy mírně posouvat
obraz na jamce a tím udržet zrakový vjem.

1.3.3 Sakády

Mezi dvěma fixacemi probíhá velmi rychlý trhavý přesun označovaný jako
sakadický pohyb. Ten trvá v rozmezí 10 až 100 ms a v jeho průběhu se vyhledává
místo další fixace. Úkolem sakadického pohybu je zamezit přílišné adaptaci v zra-
kových drahách. K této adaptaci dochází zejména při dlouhodobém pozorování
stejného objektu [28]. Sakády mohou být jak vědomé (řízené vůlí), tak reflexivní.

Kapitola 2

Subjektivní metody hodnocení kvality obrazu

Při testování jakékoliv nové metody je potřeba srovnání s již některou ověřenou metodou, kterou lze nazývat jako referenční. Nejinak je to i zde. Protože ale v celé oblasti obrazové kvality zatím neexistuje žádná standardizovaná objektivní metoda pro měření kvality obrazu, je nutné použít jako referenci výsledky subjektivních testů kvality obrazu. Proto bude v následující části přiblíženo několik nejčastěji používaných subjektivních metod hodnocení kvality obrazu společně s jejich principy měření a vyhodnocování výsledků.

2.1 Metodika subjektivního testování

Jak z názvu vyplývá, subjektivní testy jsou založeny na pozorování daného obrazu skupinou pozorovatelů, kteří vyhodnocují vnímanou kvalitu. To popisuje doporučení [6] pro stanovení subjektivní kvality obrazu. Nejčastěji je pro měření vybrána skupina pozorovatelů, kterým se v laboratoři na přesném zobrazovači promítne série testovaných scén. Pozorovatelé pak podle definované stupnice subjektivně hodnotí kvalitu obrazu. Subjektivní metody lze rozdělit na dvě základní skupiny: metody DS (zde se hodnotí kvalita dvojice zobrazovaných scén) a metody SS (zde se hodnotí kvalita samostatného obrazu). Další subjektivní metody lze nalézt v [7].

2.2 Metody DS (Double Stimulus)

2.2.1 DSIS

Pozorovatelům se několikrát ukážou vždy dva obrazy pro různé scény. V každém páru je se stejnou scénou první obraz referenční a druhý obraz testovaný (se zhoršenou kvalitou). Hodnocení se provádí na stupnici od 5 (nepostřehnutelné zkreslení) do 1 (velmi nepříjemné zkreslení) podle zhoršení kvality obrazu. Nejčastěji je možné se s touto metodou setkat v momentě, kdy je měněno několik parametrů v obrazu (např. bitová rychlost, kompresní formát atd.). Zjišťuje se, kdy nastane zlom ve vnímání kvality. Metoda je vhodná pro hodnocení širokého rozsahu zkreslení.

2.2.1 DSCQS

Metoda DSCQS se liší od DSIS tím, že je měřena na kontinuální stupnici od 0 do 100. Místo referenčního a testovacího obrazu, jako v metodě DSIS, je zde obraz A a B, přičemž nemusí být jasně předem dáno, který obraz je z této dvojice referenční. Analýza je tak založena na hodnocení diferencí u každého páru, ale s ohledem na další scénu v páru. Je proto vhodná pro měření relativně nízkého stupně poškození.

2.3 Metody SS (Single Stimulus)

2.3.1 SSM

Pozorovatelům jsou postupně ukázány oddělené scény. Existují dva přístupy: SS (neopakuje se žádná zkušební scéna) a SSMR (zkušební scény jsou opakovány vícekrát). K vyhodnocení jsou užívány různé metodiky:

- 1) Nepřímá (adjectival) - zahrnuje 5-ti stupňovou stupnici s možností hodnocení po 0,5 stupně.
- 2) Nekategorická (non-categorical) - je plynulá stupnice bez čísel nebo s velkými rozsahy, například 0-100.

2.3.2 SSCQE

Pozorovatelé plynule hodnotí pouze testovací scény. K průběžnému hodnocení vnímané kvality obrazu mají pozorovatelé k dispozici posuvné ovládací prvky na spojitě stupnici (0 = Nejhorší kvalita, 100 = Nejlepší kvalita). Hodnocení obrazu probíhá po dlouhou dobu (10 až 20 min), kdy záznamové zařízení zapisuje výsledky každou polovinu sekundy. Tato metoda se dobře hodí k testování časově proměnné kvality současných digitálních systémů pro kódování videa, jakými jsou například MPEG-2, MPEG-4 atd.

2.4 Výhody a nevýhody subjektivního testování

Hlavní výhodou subjektivního testování je, že výsledky přesně korespondují s HVS, protože jsou přímo na něm měřeny. Z jejich výsledků lze vypočítat skalární hodnotu MOS (Mean Opinion Score), která je věrohodná pro široký záběr statických i pohybových obrazových aplikací. Nevýhodou je časová náročnost, finanční náročnost, dodržení předepsaných pozorovacích podmínek a vhodný výběr pozorovatelů. Výsledky jsou obvykle platné pouze pro dané pozorovací podmínky a daný typ scény.

Kapitola 3

Objektivní metody hodnocení kvality obrazu

Začátky modelování HVS pro stanovení objektivní kvality obrazu lze dohledat již s nástupem televizního vysílání. V současné době, hlavně díky masivnímu rozvoji digitální techniky v multimediální a telekomunikační oblasti, je k této problematice obrácena čím dál větší pozornost. Veškeré objektivní metody lze rozdělit dle způsobu zpracování a vstupních parametrů do tří základních skupin modelů: NR (no reference), RR (reduced reference) a FR (full reference) metriky.

3.1 MSE a PSNR

Přestože se tato práce zabývá HVS modely hodnocení kvality obrazu, je nutné v krátkosti zmínit i metriky nerespektující HVS. Mezi nejvýznamnější patří bezesporu pixelově orientované FR metriky MSE (Mean Squared Error) a PSNR (Peak Noise to Ratio), které jsou i v dnešní době kvůli nedokonalosti zpracování HVS v mnoha aplikacích stále nejpoužívanější, např. v [11].

Základem pixelově orientovaných metrik je střední kvadratická chyba MSE. MSE je definována jako střední hodnota druhých mocnin rozdílů dvou šedotónových hodnot pixelů obrazu. Pro třírozměrný obrazový signál je definována takto [9]:

$$MSE = \frac{1}{TXY} \sum_t \sum_x \sum_y [I(t, x, y) - \tilde{I}(t, x, y)]^2, \quad (3.0)$$

kde I a \tilde{I} představují jasové hodnoty dvou obrazů o rozměrech x , y a t . Pokud I bude bráno za původní nezkreslený signál a \tilde{I} za jeho zkreslenou verzi, potom lze MSE považovat za měřítko kvality obrazového signálu. Výraz $I - \tilde{I}$ pak udává hodnotu chybového signálu, který vznikne mezi hodnotami původního a zkresleného obrazového signálu. Tento přístup lze použít i pro barevné obrázky, kde se výše uvedený vztah použije pro výpočet každé barevné složky a upraví přidáním sumy sčítající chyby jednotlivých složek.

Z MSE je odvozená metrika PSNR. Hodnota PSNR v decibelech je dána vztahem:

$$PSNR_{dB} = 10 * \log_{10} \left(\frac{m^2}{MSE} \right), \quad (3.1)$$

kde m je maximální hodnota, kterou může pixel nabývat (např. 255 pro 8-bitový obraz). Typická hodnota PSNR pro kvalitní obraz by měla být kolem 50 dB.

Výhody těchto metrik jsou zejména v rychlosti výpočtu a snadné implementaci. Protože ale neberou v úvahu HVS, ve srovnání se subjektivními testy selhávají (korelace mezi PSNR, MSE (objektivní metody) a MOS (subjektivní hodnocení) je v rozsahu pouze $r = 0,4 - 0,7$). Je to způsobeno tím, že zpracovávají všechny numerické chyby se stejnou váhou, bez ohledu na jejich umístění ve scéně. Typickým příkladem těchto nepřesností je například šum ve členité oblasti obrazu (např. obraz se skalami), který není pozorovatelem téměř vnímán, PSNR jej však detekuje a objektivně určená kvalita obrazu bude proto nižší [9], [10].

3.2 Metody měření bez reference (NR)

Metody pro měření kvality obrazu bez reference nepotřebují žádnou původní informaci o testovaném obraze. Díky tomu není potřeba mít k dispozici zároveň referenční i testovanou část obrazu. Problémem tohoto přístupu je způsob, jak umět rozlišit, která část obrazu je rušením a která je naopak samotným obsahem. Kvůli tomu není prakticky možné navrhnout univerzální NR metriku. Místo toho se tyto metody specializují spíše na konkrétní typ rušení, který vyhodnocují. Proto je nutné předem vědět a rozhodnout, co je v dané chvíli nejdůležitější hodnotit (např. rozostření, blokové artefakty, šum atd.). I přes tato omezení jsou NR metriky rozšířené, protože nevyžadují na přijímací straně jakékoliv informace o originálním obraze. Je to také jediná možnost, jak zhodnotit obrazy, kde již originál není k dispozici. Protože výsledek v této práci není tímto způsobem hodnocen, podrobnější informace o těchto metodách a jejich metrikách může čtenář najít například v [12], [13], [14] a [15].

3.3 Metody měření s částečnou referencí (RR)

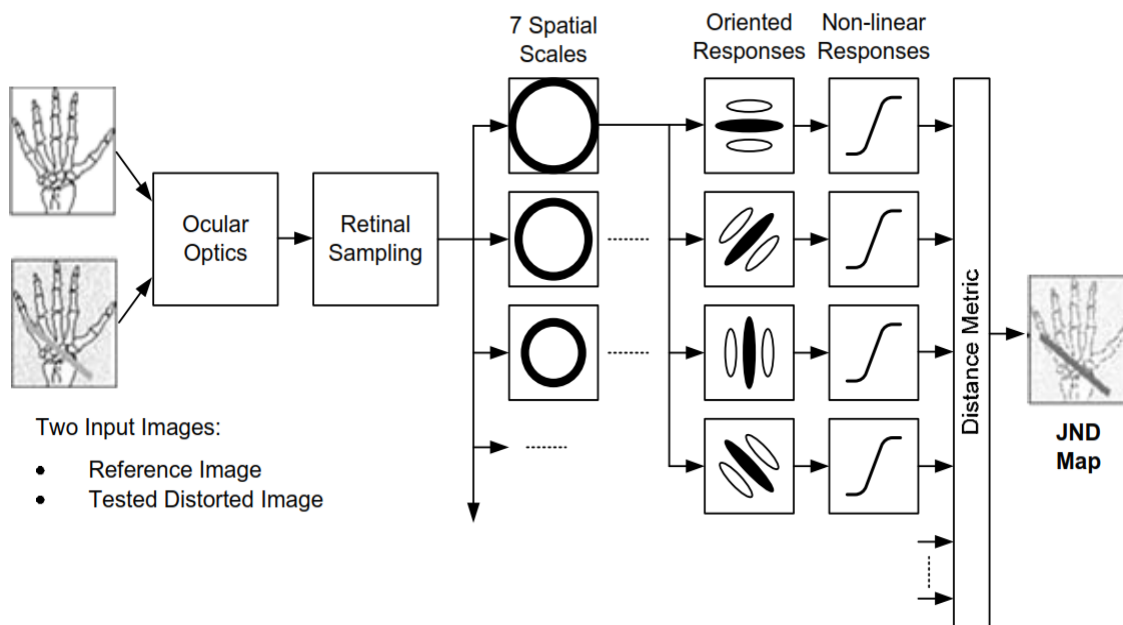
Metody pro měření kvality s částečnou referencí leží na půli cesty mezi FR a NR metodami. Oproti metodám FR se spokojí pouze s částečnou informací převzatou z původního obrazu, která je přenášena spolu s testovaným signálem a používá se pro měřítko kvality na konci přijímače. Příkladem použití může být například informace o hranách, kterou lze získat na základě hranových operátorů, informací o míře pohybu atd. z původního obrazu. Tyto informace se následně použijí k ohodnocení obrazu měřeného. Protože výsledek v této práci není tímto způsobem hodnocen, odkazy na tyto metody a jejich metriky může čtenář najít například v [13] a [14].

3.4 Metody měření s plnou referencí (FR)

Metody pro měření kvality s plnou referencí jsou z vyjmenovaných skupin nejrozšířenější a nejrozvinutější. Ke svému fungování potřebují mít k dispozici jak referenční obrazy, tak testované obrazy. Tyto metody tedy můžeme srovnat se subjektivními metodami DS, které většinou slouží pro porovnání zkoumané FR metody. Protože existuje mnoho stovek FR metod zaměřených na nejružnější aplikace, je následující přehled soustředěn pouze na ty klíčové, které zásadním způsobem přispěly k rozvoji celé problematiky a na jejichž základě je postavena metoda prezentována v této bakalářské práci.

3.4.1 Lubinův model HVS

První popsanou FR metodou je multikanálový model Lubina [16], [17]. Blokové schéma modelu zobrazuje obr. 3.1.



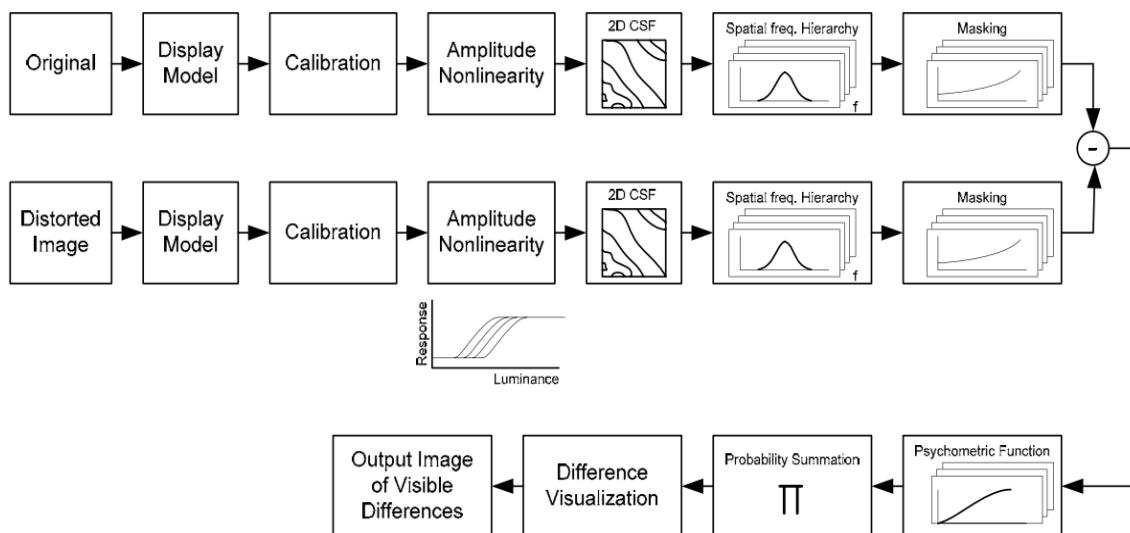
Obr. 3.1: Blokové schéma Lubinova modelu HVS [18]

První část modelu zahrnují optické a vzorkovací procesy. Ty je nejprve nutné nastavit na specifické pozorovací podmínky (např. rozteč obrazových pixelů, pozorovací vzdálenost, charakteristiky monitoru atd.). Při prvním zpracování signálu se aplikuje rozmazání. Následně jsou vstupní signály opětovně odebrány a díky tomu je napodoben způsob snímání obrazu pomocí fotoreceptorů na sítnici. V další části se využívá Laplaceova pyramida k dekompozici obrazů na sedm rozlišení. Každé rozlišení je vždy o polovinu nižší oproti vyššímu rozlišení na stupnici. Jasový signál je dále konvertován na lokální kontrast pomocí techniky podobné Peliho LBC algoritmu. Následuje sada osmi směrových filtrů, které zajišťují selektivitu ve čtyřech orientacích. Díky tomu se docílí ve výsledku 28 kanálů modelu. V každém jednotlivém kanálu se nejprve provádí maskování prostřednictvím sigmoidní nelinearity, která dokáže simulovat nalezení tzv. „dipper“

efektu pro shodné maskování a podněty [8]. Veškeré chyby jsou poté sdruženy do distorzní mapy za použití Minkowského sumace s exponentem 2.4, která je provedena přes jednotlivé kanály. Tyto chyby produkují výslednou JND mapu. Následující sumaci přes celou JND mapu se získá jediné číslo, které reprezentuje konečnou kvalitu obrazu a tím je možné toto číslo srovnat se subjektivními testy. Výsledky ukazují dobrou korelaci se subjektivními testy ($r = 0,94$) ve srovnání s metrikou MSE ($r = 0,81$), ačkoliv je nutné brát v úvahu, že v tomto experimentu byla použita velmi omezená sada obrazů (pouze 4 letecké snímky, 1 kodér, 8 pozorovatelů). Pozdější experimenty naznačovaly, že model díky složitosti moderních komprimovaných obrazů poskytuje větší toleranci k míře poškození. Tento model byl použit také pro stanovení kvality rentgenových snímků [18].

3.4.2 VDP (*Visible Differences Predictor*)

VDP je multikanálový model pro hodnocení kvality obrazu navržený Dalym [5], [20]. Jeho blokové schéma je na obr. 3.2.



Obr. 3.2: Blokové schéma Dalyho VDP multikanálového modelu [5]

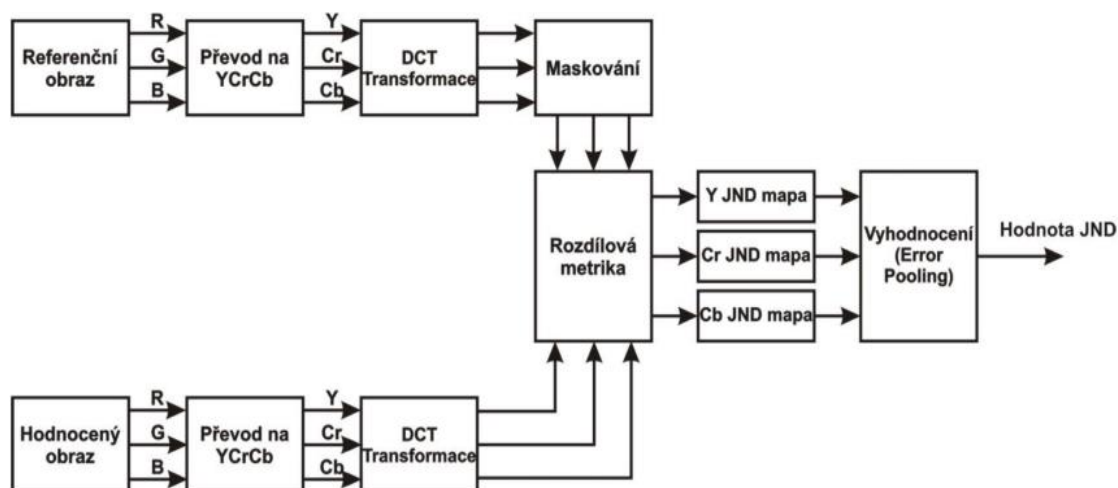
Tento model je nutné, stejně jako Lubinův, před začátkem měření zkalibrovat na dané specifické pozorovací podmínky. Referenční a testovaný snímek poté projde řadou procesů, mezi něž patří: jasová nelinearita (odpovídající Weberovu efektu), robustní CSF filtrování, selektivní kanálová dekompozice a orientace prostorových frekvencí, výpočet kontrastu a modelování maskovacího efektu. Kanálová dekompozice zahrnuje modifikovanou kortikální transformaci, která ve výsledku dává 31 nezávislých kanálů (5 prostorových úrovní krát 6 úrovní orientace + základní pásmo). Maskování využívá prahový zdvih (viz obr. 3.2), který pro každý kanál určuje z CSF práh viditelnosti na základě velikosti kontrastu v příslušném kanálu. V maskovacím procesu jsou použity jako maskovače referenční i testované obrazy. Tato implementace má význam, pokud je chyba obrazu maskována jak původním, tak testovacím snímkem (v závislosti na typu zkreslení). Po vyhodnocení veškerých chyb a jejich porovnáním s prahem viditelnosti, je pro

každý kanál vytvořena mapa pravděpodobnosti detekce chyb pomocí psychometrické funkce. Nakonec po sečtení všech kanálů vznikne celková JND mapa, kde každý bod značí, s jakou pravděpodobností bude člověk vnímat rozdíl mezi referenčním a testovaným obrazem.

Problém tohoto výstupního formátu je, že je obtížné ověřit přesnost JND mapy pro přirozené obrazy (toto ověření při testování modelu nebylo autorem provedeno) [8]. Daly následně doporučil, pro správnou schopnost detekovat maskovací prahy, použít špičkové hodnoty ve VDP mapě [21]. Díky tomuto přístupu Daly dokázal, že model je schopen správně predikovat data vycházející z širokého rozsahu psychofyzikálních experimentů [8].

3.4.3 Watsonův DCT model

Na rozdíl od výše zmíněných modelů je Watsonův model [22] založen na DCT transformaci. Blokové schéma principu modelu je na obr. 3.3.



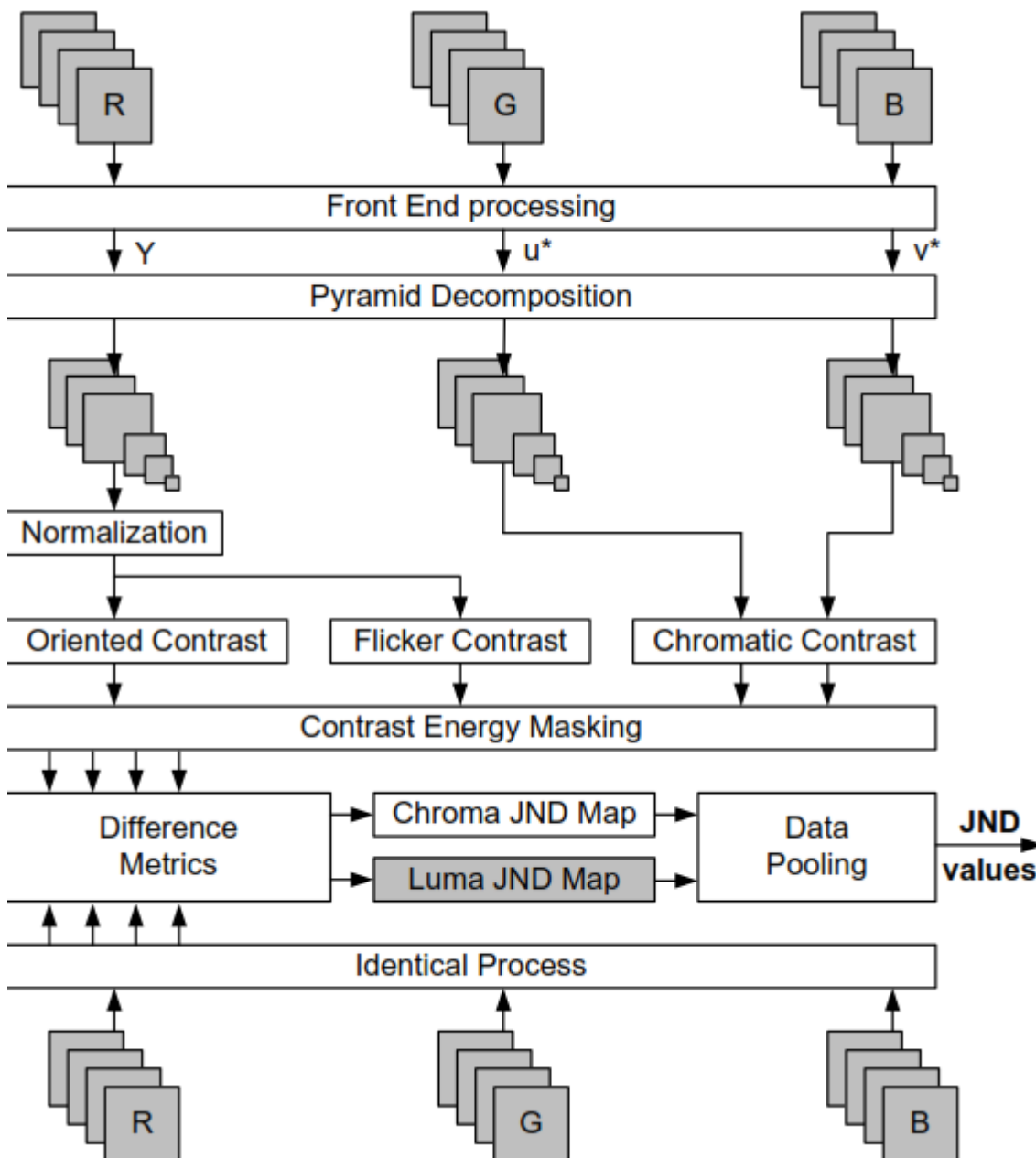
Obr. 3.3: Blokové schéma DCT modelu ve frekvenční oblasti [23]

Každá jednotlivá složka barevného obrazu ($Y C_r C_b$) je rozdělena do bloků o velikosti 8×8 pixelů. V každém z těchto bloků je provedena DCT transformace. Díky ní se získají frekvenční koeficienty z obou obrazů (referenční a testovaný). Koeficienty z referenčního obrazu jsou použity pro určení prahové úrovně, která je odvozena z kontrastní citlivosti, jasového maskování a kontrastního maskování. Jasové maskování (závislé jen na stejnosměrné složce každého bloku) a kontrastní maskování (funkce frekvenčních koeficientů a jasového maskovacího prahu) se provede pro každý frekvenční koeficient obrazu. To znamená, že pro jednotlivé koeficienty v každém bloku je nastaven jiný práh vnímání. Nyní lze již vypočítat míru poškození každého jednotlivého bloku, která je následně váhována celkovým citlivostním prahem. Výstupem je JND mapa vypočtená ve 2 krocích. V prvním kroku je vyhodnocena suma odpovídajících chyb frekvenčních koeficientů pro každý bod a všechny kanály. V druhém kroku je vzniklá matice chyb vyhodnocena

přes frekvence. Tímto výpočtem získáme celkovou chybu vnímání pro tři kanály (Y , C_r a C_b) a jejich sečtením vznikne celková chyba JND.

3.4.4 Sarnoffův JND model

Jedním z nejvýznamnějších multikanálových modelů pro hodnocení kvality barevných obrazů je Sarnoffův JND model vidění [24], [25]. Blokové schéma modelu je na obr. 3.4.



Obr. 3.4: Blokové schéma Sarnoffova multikanálového modelu [24]

Tento model byl komerčně vyvinut výzkumnou skupinou kolem Lubina v Sarnoffových laboratořích v Princetonu (USA) a firmou Tektronix. Model vychází z původních Lubinových modelů (viz obr. 3.1). Velký rozdíl ale přináší ve

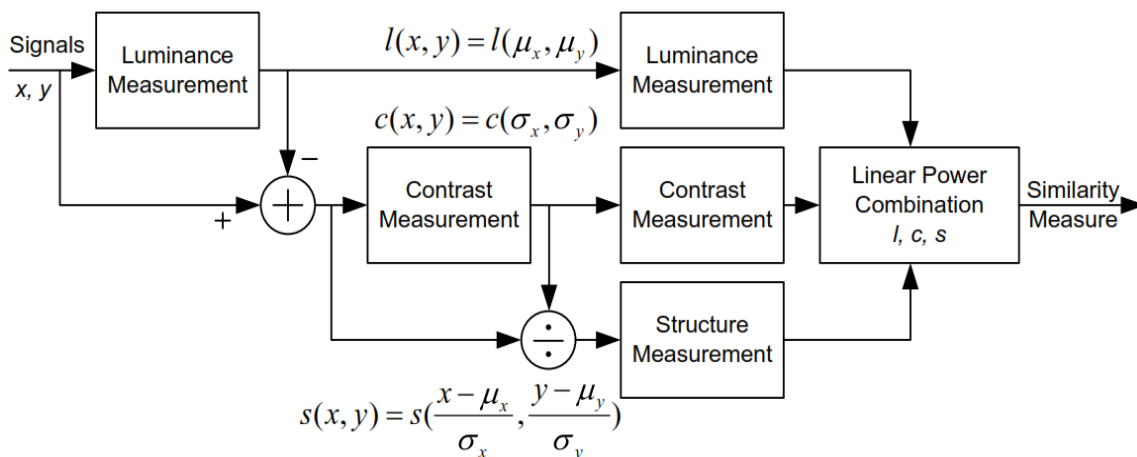
způsobu zpracování obrazu. Nejprve je každý obraz rozčleněn na 3 větve, které jsou reprezentovány jasovou a dvěma chrominančními složkami. V první kroku, nazvaném předzpracování, jsou složky Y' , C_r' , C_b' transformovány na signály R' , G' , B' . Následuje bodová nelinearita, která mimo jiné má za úkol oříznutí nízkých hodnot jasu v každé ploše zobrazovače. Hodnoty z barevného prostoru R , G , B jsou poté lineárně transformovány přes barevný prostor XYZ do konečného barevného prostoru CIE L^* , u^* , v^* (Wyszecki a Stiles, 1982). Větev L^* prochází jasovým zpracováním a větve u^* , v^* chrominančním zpracováním.

U jasového zpracování projde každá jasová hodnota nejprve kompresní nelinearitou. Poté následuje čtyř-úrovňová Gaussovská pyramidová dekompozice (Burt a Adelson, 1983), která rozděluje jasovou složku na různé prostorové frekvence. Po dekompozici se v každé úrovni vykonají jednoduché operace, mezi něž patří například prostorová a časová filtrace, výpočet kontrastu a maskování. Chrominanční zpracování je vůči jasovému zpracování paralelní v různých směrech. Meziobrazové diference chrominančních složek u^* , v^* prostoru CIE LUV jsou použity k definování detekčních prahů pro chrominanční model, analogicky k Michelsonovu kontrastu a Weber-Fechnerovu zákonu (viz kap 1.2.2), kde je definován detekční práh v jasovém modelu. V analogii s jasovým modelem jsou chrominanční kontrasty, definované pomocí diferencí u^* , v^* , také subjektem maskování. Oproti jasovému zpracování se u chrominančního zpracování provádí až sedmi-úrovňová pyramidová dekompozice, která zachycuje, že chrominanční kanály jsou citlivější na nižší prostorové frekvence než jasové kanály (Mullen, 1985). Následuje prostorová filtrace Laplaceovým jádrem, která generuje barevné rozdíly. Ty jsou propojeny s JND. Každá hodnota barevného rozdílu je následně váhována, absolutně ohodnocena a předána do části maskování kontrastu. Tato hodnota má stejnou funkci jako v jasovém modelu. Výstupem je, stejně jako u předcházejících modelů, JND mapa, složená z jasové a chrominanční části. Zároveň model produkuje jediné číslo označované Q_{norm} , které udává výslednou kvalitu obrazu.

3.4.5 SSIM (Structural Similarity Index)

Poslední FR model, o kterém se tato práce zmiňuje, prezentovala skupina kolem Wanga a Bovika [26], [27]. Oproti všem předcházejícím modelům, které ke stanovení kvality obrazu používají známé blokové modelování, používá SSIM odlišný přístup. SSIM počítá na základě statistických výpočtů strukturální podobnost referenčního a testovaného obrazu.

Jas povrchu objektu, který je pozorován, je výsledkem osvětlení povrchu a jeho odrazivosti. Struktura objektů je ale na velikosti osvětlení nezávislá. Proto je nutné vliv osvětlení nejprve omezit. Struktura obrazu je díky tomu definována naprosto nezávisle na průměrném jasu a kontrastu [27]. Výpočet kvality obrazu proto probíhá lokálním srovnáním korelace v jasu, kontrastu a struktuře obrazu. Blokové schéma modelu je na obr. 3.5.



Obr 3.5: Blokové schéma SSIM modelu [26]

Obrazové signály x a y představují bloky o stejných rozměrech, které byly vybrány ze dvou různých obrazů. Pokud první obraz je referenční a druhý je zkruslená verze referenčního, potom hodnota SSIM může být použita jako měřítko kvality.

V prvním kroku výpočtu algoritmu je provedeno srovnání signálů z hlediska hodnot jejich jasů. U diskrétních obrazových signálů je střední hodnota jasu μ_x (pro signál x) vypočtena jako průměr přes všechny hodnoty pixelů x_i z celkového počtu N v daném bloku:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i. \quad (3.2)$$

Funkce pro srovnání jasů $l(x, y)$ následně porovná střední hodnoty jasů μ_x a μ_y a je definována jako:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (3.3)$$

kde se konstanta C_1 zavádí kvůli stabilitě výpočtu v případě, že by se výsledek výrazu $\mu_x^2 + \mu_y^2$ blížil nule. Její hodnotu definuje rovnice:

$$C_1 = (K_1L)^2, \quad (3.4)$$

kde L značí maximální hodnotu, jakou může nabývat jeden pixel (255 pro 8 bitů na kanál) a $K_1 \ll 1$ je malá konstanta [27].

Pro výpočet kontrastu v obraze je využita definice rovnice směrodatné odchylky σ_x (druhá odmocnina hodnoty rozptylu), která má pro obrazový signál x podobu:

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}}. \quad (3.5)$$

Srovnání kontrastu dvou obrazů poté provede funkce $c(x, y)$ na základě porovnání hodnot jejich směrodatných odchylek σ_x a σ_y následujícím způsobem:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (3.6)$$

kde $C_2 = (K_2L)^2$ a $K_2 \ll 1$ mají podobný význam jako konstanty C_1 a K_1 . Důležitou vlastností této funkce je, že při stejném rozdílu kontrastů $\Delta\sigma = \sigma_y - \sigma_x$ je funkce více citlivá na změny nízkých hodnot kontrastů než na změny vysokých hodnot kontrastů. To odpovídá vlastnostem HVS, které jsou v modelu popsány pomocí funkce CSF a efektu maskování. Porovnání struktury obou obrazových signálů, jak bylo řečeno výše, lze provést až po odečtení střední hodnoty jasu od hodnot jejich vzorků a normování vůči hodnotě jejich směrodatné odchylky. Tuto definici lze zapsat pomocí výrazů $(x - \mu_x)/\sigma_x$ a $(y - \mu_y)/\sigma_y$. Korelace mezi těmito hodnotami odpovídá korelaci mezi vzorky signálů x , y a určuje míru strukturální podobnosti signálů. Samotná funkce pro porovnání struktury $s(x, y)$ je definována následovně [27]:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \quad (3.7)$$

Konstanta $C_3 = (K_3L)^2$ slouží, podobně jako konstanty v předchozích případech, k zajištění stability výpočtu. Symbol σ_{xy} reprezentuje rovnici:

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y). \quad (3.8)$$

Na rozdíl od předchozích funkcí může $s(x, y)$ nabývat i záporných hodnot. Konečná podoba matematického vyjádření výpočtu indexu SSIM je kombinací dříve uvedených funkcí z rovnic (3.3), (3.6) a (3.7). Její obecný tvar je:

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha * [c(x, y)]^\beta * [s(x, y)]^\gamma, \quad (3.9)$$

kde $\alpha > 0$, $\beta > 0$ a $\gamma > 0$ jsou parametry, které lze využít k zvýraznění důležitosti jednotlivých korelací. Pokud položíme $\alpha = \beta = \gamma = 1$ (nejčastěji používaný případ), a podobně konstantu $C_3 = C_2/2$, rovnice pak po drobných úpravách přechází ve finální tvar používaný při výpočtu indexu SSIM [27].

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (3.10)$$

Metoda SSIM je symetrická. To znamená, že nezáleží na tom, zda první načítaný obraz ze dvou je referenční nebo testovaný. Je rovněž omezená v rozsahu hodnot $-1 < \text{SSIM}(x, y) \leq 1$, kdy 1 reprezentuje, že oba porovnávané obrazy jsou stejné, naopak čím více se blížíme k mínus jedné, tím jsou obrazy rozdílnější.

Výpočet SSIM se prakticky provádí pro menší oblasti s pevně definovanou velikostí než pro celý obraz najednou. Důvodem je zejména proměnlivost statistických vlastností a zkreslení obrazu v závislosti na pozici v obraze. Kromě mnoha dalších příčin je jednou z nejdůležitějších správné postihnutí HVS, protože člověk nacházející se v určité vzdálenosti od obrazu je v daném časovém okamžiku schopen vnímat pouze určitou část obrazu s vysokou rozlišovací schopností [27]. Proto se prakticky používá výpočet SSIM pro více menších oblastí, nejčastěji pro oblast 11x11 pixelů. Uvnitř tohoto okna jsou následně hodnoty pixelů váhovány Gaussovou funkcí $w = \{w_i \mid i = 1, 2, \dots, N\}$ se směrodatnou odchylkou 1.5 vzorku [27]. Hodnoty vzorků jsou následně normovány tak, aby jejich výsledný součet byl roven jedné a nezměnil se tak celkový jas uvnitř okna. Výpočet hodnot μ_x , σ_x a σ_{xy} díky tomu probíhá podle těchto rovnic [27]:

$$\mu_x = \sum_{i=1}^N w_i x_i, \quad (3.11)$$

$$\sigma_x = \left(\sum_{i=1}^N w_i (x_i - \mu_x)^2 \right)^{\frac{1}{2}}, \quad (3.12)$$

$$\sigma_{xy} = \sum_{i=1}^N w_i (x_i - \mu_x)(y_i - \mu_y). \quad (3.13)$$

Použitím tohoto postupu dochází k výraznému potlačení blokových artefaktů v SSIM mapě. Pro stabilizaci výpočtu jsou zvoleny dle doporučení [27] hodnoty konstant $K1 = 0,01$ a $K2 = 0,03$. Základní podmínkou použitých konstant je, že jejich hodnoty musí být malá čísla. Jinak nemají žádný výrazný vliv na algoritmus výpočtu SSIM indexu.

Pro možnost porovnat výsledky této metriky s dalšími metrikami je nutné získat jedinou hodnotu indexu pro celý obraz. Proto se zavádí průměr všech hodnot SSIM vypočítaný pro jednotlivé oblasti nazvaný MSSIM (Mean SSIM), který je definován následovně [27]:

$$\text{MSSSIM}(X, Y) = \frac{1}{M} \sum_{j=1}^M \text{SSIM}(x_j, y_j), \quad (3.14)$$

kde X a Y představují referenční a testovaný obrázek, x_j a y_j jsou části obrázků vymezené j -tým posunem okna a M je celkový počet možných posunů okna po obrazu.

Metrika MSSIM byla použita v návrhu metody v této bakalářské práci. Její implementace je popsána v kapitole 6.2.3.

Kapitola 4

Charakteristika obrazu mluvího znakového jazyka

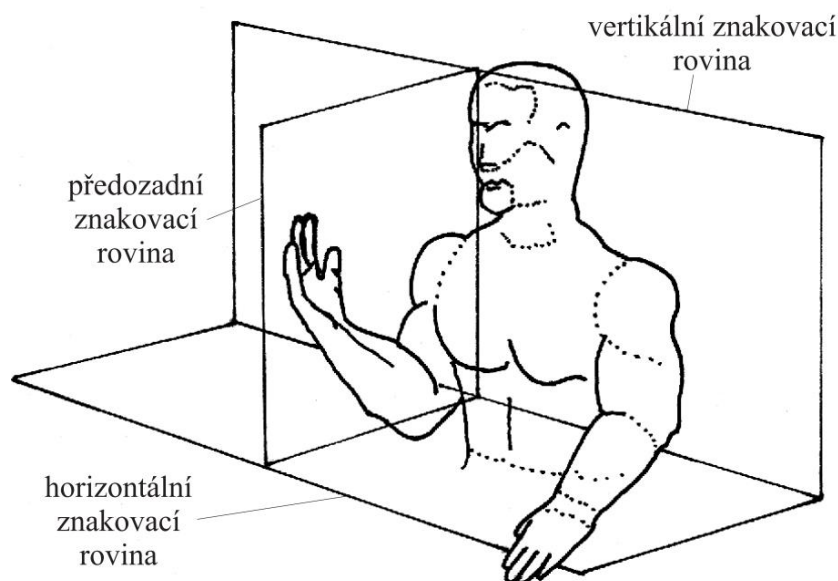
Pokud je potřeba navrhnout správnou metodu pro objektivní hodnocení kvality obrazu mluvího znakového jazyka, je nutné nejprve důkladně prozkoumat, které části obrazu jsou pro neslyšícího pozorovatele důležité.

4.1 Znakový jazyk

V České republice žije na základě průzkumu provedeném v roce 2007 Českým statistickým úřadem (ČSÚ) necelých 26 tisíc těžce nebo velmi těžce sluchově postižených lidí [28]. Nespokojenost s technikou kompenzačních pomůcek vyjádřilo v dotazníku 8 % neslyšících, což je nejvíce ze všech typů postižení (např. u nevidomých jsou to pouze 4 %). Ze zprávy navíc vyplývá, že jako největší omezení zúčastnění neslyšící respondenti cítí v příjmu informací a v komunikačních schopnostech. Tato omezení vycházejí z toho, jak se liší základní dorozumívací prostředky běžného slyšícího člověka a člověka neslyšícího.

Základním komunikačním prostředkem neslyšících je znakový jazyk (ZJ) [28]. Stejně jako mluvená řeč má svá „nářečí“ podle země (území), kde se používá, je i znakový jazyk odlišný podle toho, kde se s ním člověk setká. Rozlišujeme proto Britský znakový jazyk (BSL), Americký znakový jazyk (ASL), Český znakový jazyk (ČZJ) a mnoho dalších. Všechny druhy znakových jazyků se ale zásadně liší od mluveného slova.

ZJ se vytváří z největší části výhradně pomocí rukou, obličeje a horní poloviny těla. Tento prostor je znázorněn na obr. 4.1. ZJ je vnímán výhradně zrakem. Díky tomu se jedná o jazyk vizuálně-motorický a tím se velice liší od běžného audio-orálního (tj. mluveného a poslouchaného) jazyka. Grafické zachycení ZJ není elementární problém (prakticky neexistuje psaná forma ZJ), protože je nutné zachytit pohyb, tvar a orientaci. Snadný způsob zachycení ZJ je pouze jeho nasnímáním do videosekvence či série obrazů.



Obr. 4.1: Základní znakovací prostor [28]

4.2 Oblasti zájmu neslyšících diváků

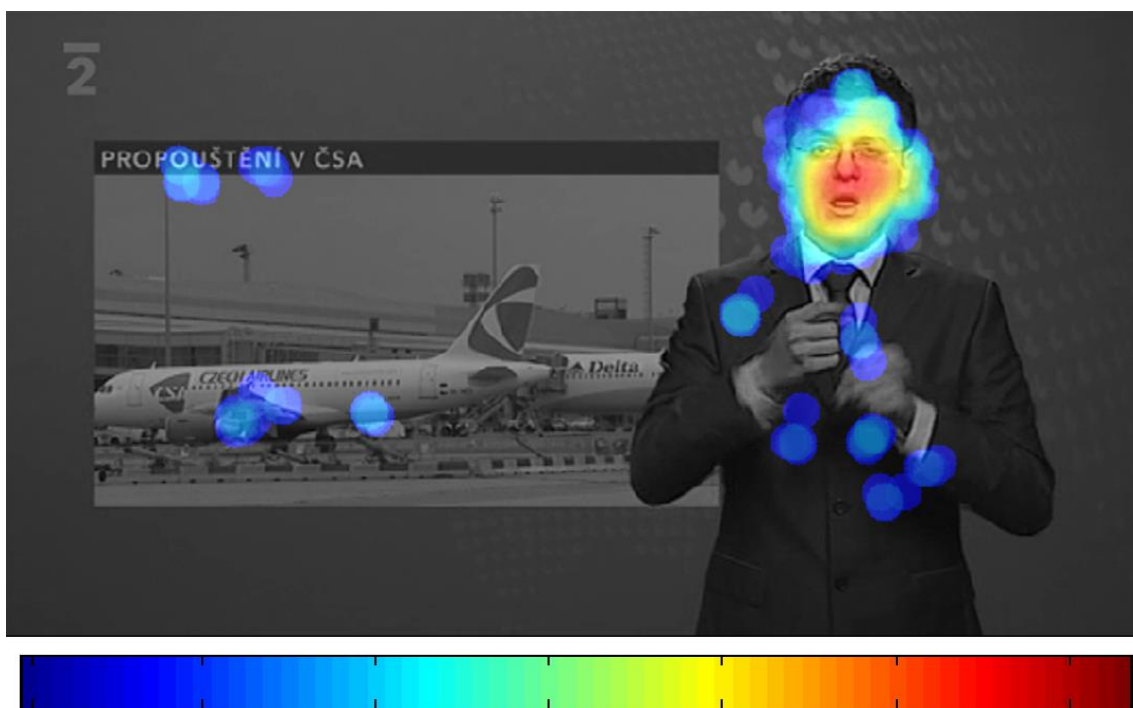
Důležitý faktor, který je nutný brát v potaz, vychází ze skutečnosti, že neslyšící vnímají vizuální scénu odlišným způsobem od slyšících. Slyšící divák používá při vnímání současně dva smysly: zrak a sluch. O oba tyto smysly se starají různá mozková centra. Mohou být proto vnímány paralelně, tedy pokud slyšící divák odvrátí zrak od vizuální scény, stále mu funguje sluch a může tak dále přijímat informaci o tom, co se děje. Naproti tomu neslyšící divák je plně závislý pouze na jednom smyslu - zraku. Tím se zásadně liší způsob, jak oba pozorovatelé (slyšící a neslyšící) vnímají tutéž scénu. Při komunikaci ve ZJ musí neslyšící správně vnímat všechny komponenty ZJ, které určují význam jednotlivých znaků a jsou důležité pro správné porozumění: polohu a tvar rukou, mimiku obličeje, tvar úst atd. Protože se všechny části znakové komunikace dějí současně, není možné je vnímat sériově (postupně za sebou). Lze tedy tvrdit, že pozornost neslyšícího (foveální vnímání) je zaměřena pouze na určité části obrazu a zbytek je vnímám periferně, aniž by docházelo k výraznému přesunu pohledu. Díky této skutečnosti lze rozčlenit obraz na oblasti zájmu (ROI). Každá ROI se vyznačuje různou dobou pozornosti, kterou neslyšící divák jednotlivým částem věnuje.

Praktická část zjištění ROI v obraze má kromě výzkumu pozornosti při sledování reklamy a vizuálního obsahu (webové stránky) především význam při vyvíjení kompresních standardů, které ROI zohledňují ve své kompresi. V této části bude v krátkosti přiblížena praktická implementace zjišťování oblastí zájmu neslyšících diváků. Výsledky vedly k vytvoření videosekvencí, které jsou použity k testování navrhované objektivní metody.

4.3 Určení oblastí zájmu neslyšících diváků

Cílem práce [28] bylo nalézt oblasti zájmu neslyšících ve vizuálních pořadech, jejichž součástí je tlumočení do ZJ. Z předchozích kapitol vyplývá, že pozornost neslyšícího je při libovolném pořadu soustředěna tak, aby mohl co nejlépe porozumět obsahu. Žádné další důvody při sledování nejsou brány v úvahu. Při zvážení těchto skutečností lze vyvodit, že většina pozornosti neslyšícího diváka bude soustředěna na mluvčího ZJ, a to zejména na jeho obličej.

Experiment byl prováděn na předem vybraných videosekvencích, které pokryly základní způsoby kompozice scén s tlumočníkem do ZJ. Jsou současně dostupné v televizním vysílání. Pro možnost sledování očních pohybů byl použit systém ViewPoint EyeTracker [28], který spadá do kategorie systémů měřících pozici oka vzhledem k hlavě. Výstupní data systému vytváří namapované body pohledu, které charakterizují, kam v dané části scény každý s účastníků zaměřoval svou pozornost. Tím lze vytvořit teplotní mapu pozornosti, která přímo na dané videosekvenci vytvoří barevnou paletu překrývající podnět a zároveň definuje četnost výskytu pohledu na danou část scény (viz obr. 4.2).



Obr. 4.2: Teplotní mapa videosekvence pro všechny subjekty a celou dobu klipu [28]

Modrá barva na obr. 4.2 značí oblasti s minimem pozornosti, zatímco tmavě červená barva značí místa s maximem pozornosti. Na šedé oblasti nebyla pozornost během promítání videosekvence soustředěna vůbec [28].

Teplotní mapy všech použitých videosekvencí potvrdily, že neslyšící divák věnuje většinu pozornosti oblasti kolem úst tlumočnicka, a to nezávisle na charakteru pozorované scény či velikosti tlumočnicka. V čase dochází ke krátkým přesunům

pozornosti na jinou zajímavou část scény, ale po chvíli se zrak opět vrací zpět na tlumočníka. Kompletní popis může čtenář nalézt v [28].

4.4 Praktická implementace oblastí zájmu neslyšících diváků

Hodnocení obrazu mluvího ZJ je zapotřebí zejména pro správné nastavení obrazové komprese tak, aby byla zajištěna nejen dostatečná srozumitelnost, ale i vizuální kvalita mluvího ZJ.

4.4.1 HbbTV a asistivní technologie

HbbTV představuje spojení klasického televizního vysílání společně s přenosem širokopásmového internetu. Cílem této služby je sjednocení výrobců a jejich interaktivních služeb tak, aby zákazník nemusel při nákupu řešit, která z nich mu bude fungovat a která ne. K příjmu platformy HbbTV je nutné mít set-top box nebo televizor s podporou HbbTV. Vzhledem k tomu, že zařízení, které dané služby podporují, jsou neustále dostupnější širšímu množství lidí, přináší tyto nové platformy možnost jejich využití i v asistivních technologiích.

V tomto případě je to možnost přenášet signál mluvího ZJ odděleně od datového toku televizního pořadu. Tato skutečnost by umožňovala neslyšícímu divákovi, aby si mohl libovolně zobrazit a sledovat mluvího ZJ společně s daným televizním vysíláním. Možnost volitelného zobrazení mluvího ZJ proto rozvíjí jeho použití v mnohem větším rozsahu než tomu je dnes, protože v současné době je obraz mluvího ZJ přenášen jako neoddelitelná součást televizního obrazu, což značně limituje počet televizních pořadů doprovázených s přenosem tlumočnicka ZJ. Zároveň by umístění a velikost mluvího ZJ mohla být přizpůsobena každému divákovi na základě jeho volby. V souvislosti s tím vzniká mnoho nových otázek a problémů.

Kapitola 5

Testované videosekvence

Z kapitoly 4 vyplynulo, že neslyšící divák při pozorování videosekvence, která obsahuje mluvčího ZJ, tráví většinu času sledováním hlavy mluvčího (foveální vidění) a ruce vnímá především periferně. Na těchto základech byly nasnímány videosekvence zobrazující pouze mluvčího ZJ před modrým pozadím [30]. Testované videosekvence jsou ve 2 rozlišeních: 544x544 pixelů a 640x640 pixelů. Tyto velikosti byly zvoleny na základě požadavků neslyšících diváků na velikost mluvčího ZJ v obraze. Kódování videosekvencí je založeno na ROI oblastech. Každé ROI oblasti je na základě její důležitosti přidělen určitý datový rozsah zabírající místo ve výstupním datovém toku.

5.1 Detekce ROI

Pro inteligentní vyhledávání ROI oblastí (obličej, pleťová barva, pozadí) byl použit Viola-Jonesův detektor, který dokáže odlišit části s pleťovou barvou a bez ní. Implementace detektoru byla použita tak, aby výsledná pleťová barva byla rozdělena na část zahrnující obličej a část zahrnující zbytek těla [30]. Výstupem algoritmu je mapa čísel, tzv. maska, kde každé číslo označuje makroblok o velikosti 16x16 pixelů a jeho hodnota určuje, do jaké skupiny ROI makroblok patří. Tyto masky jsou uloženy ve formě textových souborů s názvem charakterizujícím pořadí snímku, ke kterému maska náleží.

5.2 Hpar

Detekování ROI bylo použito při zakódování testovaného videa. ROI byly rozděleny na 3 oblasti: obličej, ruce a pozadí. Zakódování každé ROI oblasti závisí na kvantizačním parametru QP , který se stará o kompresi každé ROI oblasti ve snímku, podle následujících rovnic [30]:

$$QP_{face} = QP * \frac{1}{hpar}, \quad (5.1)$$

$$QP_{hands} = QP * \frac{\left(\frac{1}{hpar} + 1\right)}{2}, \quad (5.2)$$

$$QP_{signing_space} = \frac{N_b * QP - (a * QP_{face} + b * QP_{hands})}{c}, \quad (5.3)$$

kde $hpar$ je vstupní parametr definující rozdíl kvality každé oblasti ROI, N_b je celkový počet makrobloků ve snímku a a , b , c jsou počty makrobloků v oblasti obličeje, rukou a zbytku snímku. Z rovnic tedy vyplývá, že čím je hodnota $hpar$ vyšší, tím je vyšší kvalita obličeje a částečně i kvalita rukou vůči kvalitě pozadí.

5.3 Kódování videosekvencí

Referenční videosekvence jsou zakódovány v raw formátu UYVY se vzorkováním 4:2:2 a jsou zabaleny v AVI kontejneru. Tento formát je v podstatě stejný jako formát YUY2, ale s odlišným uspořádáním vzorků. Vzorky jasového a chrominančního signálu se v datovém formátu UYVY periodicky opakují v sekvenci U Y V Y po jednotlivých vzorcích v každé obrazové složce. Každý vzorek Y, V a U je vyjádřen osmi bity a dekadicky může nabývat hodnot od 0 do 255.

Testované videosekvence jsou zakódovány v otevřeném zdrojovém kódu x264 ve formátu YUV420 (vzorkování 4:2:0) a zabalených v MP4 kontejneru. Zdrojový kód x264 je odnož rodiny H.264, potažmo MPEG-4, který se běžně používá při kompresi videa v oblasti televizního vysílání. Vzorky jasového a chrominančního signálu jsou v tomto případě přenášeny ve formě, kdy na 8 vzorků jasového signálu připadají 2 vzorky U a 2 vzorky V chrominančních signálů. Každý vzorek Y, V a U je vyjádřen osmi bity a dekadicky může nabývat hodnot od 0 do 255. Kódování v x264 bylo použito kvůli kompatibilitě se standardem HbbTV, na který se plánuje budoucí použití tohoto systému (viz. kap. 4.4.1).

Referenční videosekvence obsahuje vždy jednu větu ve znakovém jazyce. Tyto věty patří do skupiny tzv. minimálních párů. Díky tomuto přístupu bylo možné kromě subjektivní kvality obrazu hodnotit také srozumitelnost nasnímané znakové řeči. Z referenčních videosekvencí byla vytvořena série testovaných videosekvencí, ve kterých se liší bitová rychlost (80, 120, 160 a 200 kbps) a parametr $hpar$ (1.0, 1.1, 1.2 a 1.3).

Více informací o způsobu kódování, průběhu testů a závěrečných výsledků může čtenář nalézt ve [30]. Výsledky subjektivních testů a seznam použitých videosekvencí je popsán v kapitole 7.

Kapitola 6

Praktická realizace objektivní metody

Praktická část této bakalářské práce zahrnuje návrh metody vhodné pro objektivní hodnocení kvality obrazu mluvícího ZJ a její naprogramování v programovém prostředí MATLAB.

6.1 Předpoklady pro návrh objektivní metody

Protože na poli objektivních metrik neexistuje žádná standardní metoda pro hodnocení obrazu, bylo nutné nejprve zjistit důležitá hlediska, která ovlivňují výslednou vnímanou scénu v případě mluvícího ZJ.

Za prvé bylo v této práci rozebráno, jak funguje zpracování obrazu v HVS. Na tomto základě bylo zjištěno, že vnímání obrazu není lineární ve všech směrech, ale závisí na několika faktorech (viz. kap. 1). Za druhé bylo v této práci rozebráno, čím se liší vnímání obrazu běžného slyšícího diváka a diváka neslyšícího. Z této části vyplynulo, že neslyšící divák věnuje většinu pozornosti pouze části obrazu (ROI), ve které se nachází mluvíčí znakového jazyka (viz. kap. 4). Třetím hlediskem byl způsob kódování obrazu mluvícího ZJ v hodnocených videosekvencích (viz. kap. 5). Posledním hlediskem byly výsledky subjektivních testů kvality obrazu hodnocených videosekvencí (viz. kap. 7), které zároveň sloužily jako referenční vzor pro výsledky objektivního testování kvality obrazu. Pro návrh metody byly také prozkoumány objektivní metody hodnotící kvalitu obrazu, které se používají (viz. kap 3).

6.2 Popis navržené metody

Realizace navržené metody je provedena v programu MATLAB ve verzi 2014a. MATLAB je programové prostředí a skriptovací programovací jazyk pro: vědecko-technické numerické výpočty, modelování, návrhy algoritmů, počítačové simulace, analýzu a prezentaci dat, měření a zpracování signálů. V této části bude postupně rozebráno, jak jdou jednotlivé části programu za sebou a jejich význam při výpočtu algoritmu metriky u navržené metody.

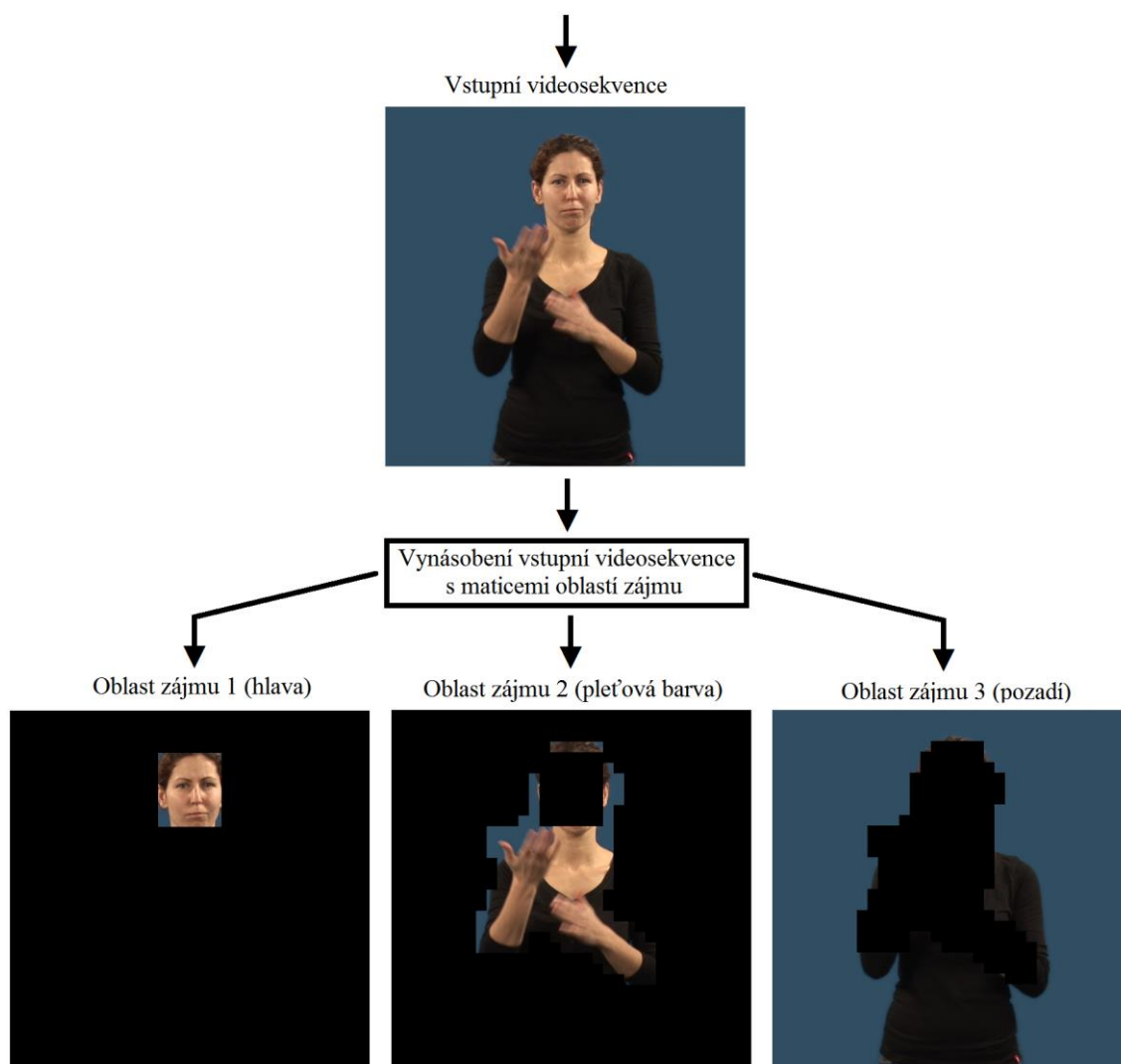
6.2.1 Načtení videosekvencí

V první části programu je zpracováno načtení videosekvencí, které byly popsány v kapitole 5. Načtení obou druhů videosekvencí (referenčních a testovaných) je realizováno v MATLABu pomocí funkce `videoreader`. Tato funkce jednotlivé videosekvence načte do 4D matice, kde první dva rozměry charakterizují výšku a šířku snímku. Třetí dimenze charakterizuje barevnou informaci snímku a čtvrtá dimenze udává počet snímků ve videosekvenci. Výškou snímku je myšlena hodnota označující počet aktivních řádků v jednom snímku. Šířkou snímku je vyjádřen počet aktivních vzorků na jednom řádku snímku. Barevná informace je uložena u obou videosekvencí ve formátu RGB24. Tato informace je uložena MATLABem v buňkách `uint8` (8-bitový integer), které odpovídají velikosti kvantování jednotlivých barevných vzorků ve videosekvencích.

6.2.2 Načtení masek

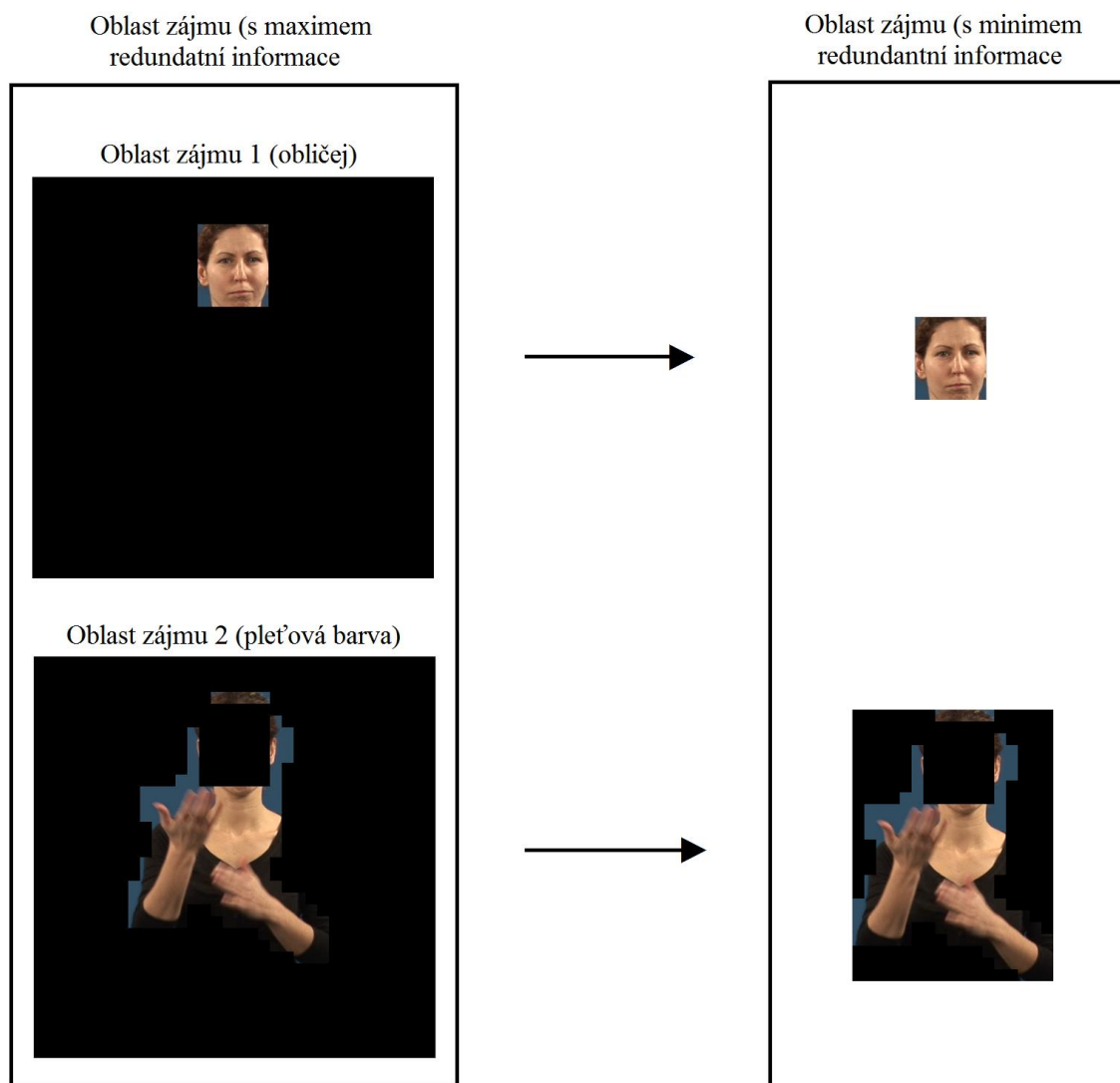
Na základě předpokladů pro návrh použité metody (viz. kap. 6.1) navržená metoda bere v úvahu oblast foveálního a periferního vidění. Díky výsledkům zjištěných při zkoumání oblastí zájmu neslyšících byly k dispozici údaje o tom, kam bude neslyšící divák zaměřovat většinu své pozornosti (viz kap. 4). Na základě těchto skutečností byl každý snímek videosekvence rozdělen na tři oblasti zájmu: obličej, pleťová barva (kromě obličeje) a pozadí. Každá ROI v celkovém výsledku metody představuje jiný parametr, který reprezentuje důležitost oblasti v jednom snímku.

Protože detektor ROI, který byl popsán v kapitole 5, nebyl k dispozici v programovém prostředí MATLAB, byla k detekci ROI v každém snímku použita data z výstupních textových souborů tohoto detektoru. O jejich načtení se stará skript `kompletmaska.m`. Každý textový soubor obsahuje řadu čísel (1156 pro rozlišení 544x544 a 1600 pro rozlišení 640x640), kde každé číslo reprezentuje oblast zájmu pro makroblok 16x16 pixelů. Po jednoduchém roznásobení (počet čísel krát velikost makrobloků) vyjde hodnota, která reprezentuje všechny pixely v jednom snímku (295936 pixelů pro rozlišení 544x544 a 408600 pixelů pro rozlišení 640x640). Textové soubory jsou nejprve postupně načteny, aby byly správně přiřazeny k příslušným snímkům, podle jejich pořadového čísla. Následují matematické operace, při kterých je z dat každého textového souboru vytvořena 4D matice o totožných rozměrech všech dimenzí jako mají načtené videosekvence. Posledním krokem při vytvoření ROI je vzájemné vynásobení matice textových dat ve čtyřech `for` cyklech (každý pro jednu dimenzi matice) spolu s maticemi referenčních a testovaných videosekvencí. Výstupem z každé videosekvence vznikne soustava tří 4D matic charakterizujících tři oblasti zájmu. Grafické znázornění tohoto postupu je na obr. 6.1.



Obr. 6.1: Schéma postupu vytvoření oblastí zájmu v programu

Tímto pronásobením matic se získají z každého snímku tři nové snímky o stejné velikosti, jakou měl původní snímek. Tyto tři nové snímky (snímky oblastí zájmu) ale dohromady nesou stejné množství informace, jako snímek původní. Pro zbavení se maximálního množství redundantní (nadbytečné) informace (černého pozadí) je použit `for` cyklus, který vyhledává minima a maxima v prvních dvou dimenzích u snímků zahrnujících první a druhou oblast zájmu. Ve snímcích zahrnujících třetí oblast zájmu již tento proces neprobíhá, protože ze snímku se tímto způsobem žádná nadbytečná informace neodstraní. U snímků zahrnujících první a druhou oblast zájmu použitý `for` cyklus zmenší rozlišení na nejmenší možnou množinu pixelů, která obsahuje veškerou informaci daného snímku. Grafické znázornění tohoto procesu ukazuje obr. 6.2.



Obr. 6.2: Odstranění redundantní informace z vytvořených oblastí zájmu

Jak je vidět z obr. 6.2, v oblasti zájmu 1 (obličej) bylo možné se zbavit veškeré redundantní informace. V oblastí zájmu 2 a 3 ovšem stále redundantní informace zůstala. Tento problém řeší zajištění aditivnosti algoritmu, která je popsána v kapitole 6.2.4. V současné chvíli je tedy na výstupu programu šest 4D matic reprezentujících oblasti zájmu vstupních videosekvencí (3 matice pro referenční videosekvenci a 3 pro testovanou videosekvenci). Každá matice reprezentuje jednu oblast zájmu ROI a obsahuje zároveň maximálně dosažitelné minimum redundantní informace, které bylo popsáním způsobem možné docílit. Nyní následuje výpočet hodnotícího algoritmu metriky.

6.2.3 Výpočet algoritmu metriky

Jak bylo řečeno v kapitole 3, existuje velké množství objektivních metod pro hodnocení kvality obrazu. Každá je specifická a použitelná pro určitou kompozici, členitost či jiné vlastnosti obrazu. Z této velké množiny bylo nutné vybrat takovou metriku, která by dokázala zachytit potřeby neslyšícího diváka. Jak

vyplývalo z kapitoly 4, pro neslyšícího diváka je nejdůležitější porozumění předkládanému vizuálnímu podnětu, přičemž ostatní části obrazu vnímá pouze periferním viděním. Po zvážení těchto skutečností a dalších diskutovaných předpokladů v kapitole 6.1 byla použita metoda SSIM. Tato metoda používá odlišný způsob výpočtu kvality oproti jiným metodám (viz. 3.4.5). Díky tomu dosahuje metoda SSIM dobré výpočetní rychlosti. Na výstupu poskytuje jediné číslo reprezentující výslednou kvalitu. A v neposlední řadě z mnoha experimentů, které byly provedeny [31], poskytla výsledky dobře postihující HVS.

U SSIM metody existuje mnoho modifikací, které byly později provedeny jinými autory. Proto v programu byla použita originální implementace SSIM metody [31] a implementace dostupná přímo v programu MATLAB od verze 2014a [32].

Základem výpočtu u originální implementace [31] je funkce `ssimO.m`. Vstupem jsou vždy dva porovnávané snímky, výstupem je pak výsledná hodnota MSSIM metriky a mapa SSIM indexů. Algoritmus výpočtu SSIM indexů je popsán v kapitole 3.4.5. Velikost procházejícího okna w o velikosti 11x11 pixelů je nastavena využitím předdefinované funkce MATLABu:

```
w = fspecial('gaussian', 11, 1.5).
```

Výpočet parametrů $\mu_x(\mu_y)$, $\sigma_x(\sigma_y)$ a σ_{xy} vychází z rovnic (3.11), (3.12) a (3.13). Okno w si lze představit jako konvoluční jádro, které je aplikováno na testovaný snímek, tj. snímek, u kterého zjišťujeme parametry. Tato konvoluce pro dva rozměry je v MATLABu realizována funkcí `filter2`. Příkaz v programu pro zjištění např. hodnot $\mu_x(\mu_y)$ může vypadat následovně:

```
ux = filter2(w, img_x, 'valid'),
```

kde `img_x` představuje testovaný snímek. Stejným způsobem se zjistí hodnoty u parametrů $\sigma_x(\sigma_y)$ a σ_{xy} . Mapa indexů se následně spočítá z rovnice (3.10) a výsledná metrika MSSIM se spočítá jako průměr indexů SSIM podle (3.14).

U výpočtu implementace dostupné přímo v MATLABu [32] je postup obdobný jako u originálního SSIM. Základem je funkce `ssimM.m`. Rozdílem oproti originální metodě SSIM je, že konvoluční jádro je zde aplikováno zároveň na tři snímky obsahující vždy jednu barevnou složku (R , G , a B). Proto je u originální implementace SSIM nutné nejprve každý snímek videosekvence převést do šedotónové stupnice podle vzorce [19], který má tvar:

$$Y = 0.2126 R + 0.7152 G + 0.0722 B \quad (6.1)$$

U výpočtu implementace dostupné přímo v MATLABu je konvoluce pro tři barevné složky realizována funkcí `imfilter`. Příkaz v programu pro zjištění hodnot potom vypadá takto:

```
Ux = imfilter(img_x, w, 'conv', 'replicate');
```

kde `img_x` představuje testovaný snímek. Hodnota MSSIM a mapa SSIM indexů se poté počítá stejným způsobem jako v originální implementaci SSIM indexu. Pro zajištění výpočtu algoritmu přes všechny snímky v testované videosekvenci slouží `parfor` cyklus. Jeho použití vypadá následovně:

```
parfor i=1:frame
[ssimval(i)] = ssim(ref{i}(:, :, :), test{i}(:, :, :));
end
```

kde `frame` označuje počet snímků ve videosekvenci, `ssimval` je výsledná hodnota MSSIM, `ref` je referenční snímek a `test` je testovaný snímek. `Parfor` cyklus je oproti klasickému `for` cyklu rozdílný v tom, že dovoluje použití paralelního výpočtu algoritmu (využitelné na vícejádrovém procesoru výpočetního stroje), což výpočet znatelně urychluje. `Parfor` cykly jsou celkově použity tři, protože porovnáváme tři dvojice snímků (pro každou ROI jeden). Výstupy všech tří `parfor` cyklů jsou zapisovány do 1D matice, kde v každém řádku je jedna hodnota MSSIM reprezentující i -tý snímek ve videosekvenci.

Díky tomuto postupu jsou na výstupu tři matice s hodnotami všech MSSIM indexů jednotlivých snímků ve videosekvenci (každá matice pro jednu ROI).

6.2.4 Zajištění aditivnosti metody

Jak bylo popsáno výše, při vytváření oblastí zájmu se v programu vytvořily z každého snímku tři nové snímky (viz. kap. 6.2.2). Každý nový snímek obsahuje pokaždé pouze jednu část ROI z původního snímku. Tímto postupem se ale do nových snímků přidala redundantní informace (černé pozadí), což je vidět na obrázku 6.1. Zároveň se po dalších korekcích, znázorněných na obrázku 6.2, změnila velikost jednotlivých snímků. Pro zajištění aditivnosti metody proto nelze hodnoty metriky MSSIM jednotlivých oblastí zájmu jednoduše sečíst, protože by se nerovnal výsledku, který by poskytla metrika při hodnocení původního snímku bez rozdělení na ROI. Pro kontrolu aditivnosti je proto nutné každý snímek reprezentující část ROI sečíst se správnou váhou. Toho je v MATLABu docíleno pomocí následujícího `for` cyklu:

```
for i=1:framenorm
MSSIMA(i) = (1-ssimval1(i))*percent1(i);
MSSIMB(i) = (1-ssimval2(i))*percent2(i);
MSSIMC(i) = (1-ssimval3(i));
end
```


kde $ssimval1(i)$, $ssimval2(i)$ a $ssimval3(i)$ označuje vypočtené MSSIM hodnoty jednotlivých ROI z původního i -tého snímku videosekvence. $Percent1(i)$ a $percent2(i)$ označují procentuální zastoupení velikosti ROI oproti původnímu i -tému snímku (u třetí ROI není potřeba počítat její procentuální zastoupení, protože pokaždé tvoří velikost stejnou jako každý původní i -tý snímek) a $framenorm$ označuje počet snímků ve videosekvenci. $MSSIMA(i)$, $MSSIMB(i)$ a $MSSIMC(i)$ tedy vyjadřují normovanou hodnotu MSSIM každého i -tého snímku jednotlivých částí ROI.

Vypočtené hodnoty MSSIM jsou vždy v rozsahu $0 < MSSIM(x,y) \leq 1$, kde se zvyšující se hodnotou MSSIM klesá množství rozdílů mezi porovnávanými obrazy, tj. roste kvalita testovaného snímku. V tomto případě je ale nutné osu MSSIM invertovat kvůli sjednocení obou stupnic hodnocení kvality obrazu (objektivní a subjektivní). Proto jsou ve výpočtu hodnoty $ssimval$ odečítány od jedné.

Aditivnost metody SSIM je potom potvrzena, pokud se hodnota součtu $MSSIMA(i)$, $MSSIMB(i)$ a $MSSIMC(i)$ vyjádřená vzorcem:

$$MSSIMZ(i) = 1 - ((MSSIMA(i) + MSSIMB(i) + MSSIMC(i))) \quad (6.2)$$

každého i -tého snímku rovná hodnotě MSSIM spočtené z původního i -tého snímku, tj. bez rozdělení na ROI. Tímto způsobem je vyřešen problém veškeré redundantní informace, která byla přidána do vytvořených snímků ROI a která tím mohla zkreslovat výsledné hodnocení. Odstranění této informace, popsané v kapitole 6.2.2 na obr. 6.2 ve snímcích zahrnujících oblasti zájmu 1 a 2, je pouze z důvodu zrychlení výpočtu celého algoritmu u metody SSIM MATLABu. Pokud by tato redundantní informace nebyla odstraněna způsobem, který je znázorněn na obr. 6.2, stačilo by k získání hodnoty $MSSIMZ(i)$ výsledky MSSIM jednotlivých částí ROI pouze jednoduše sečíst bez počítání procentuálního zastoupení, protože by všechny tři části zabíraly 100% velikost původního snímku. Tento postup je využit při výpočtu originální metody SSIM, kde by odstraňování redundantní informace zpomalovalo celý výpočet algoritmu metriky MSSIM.

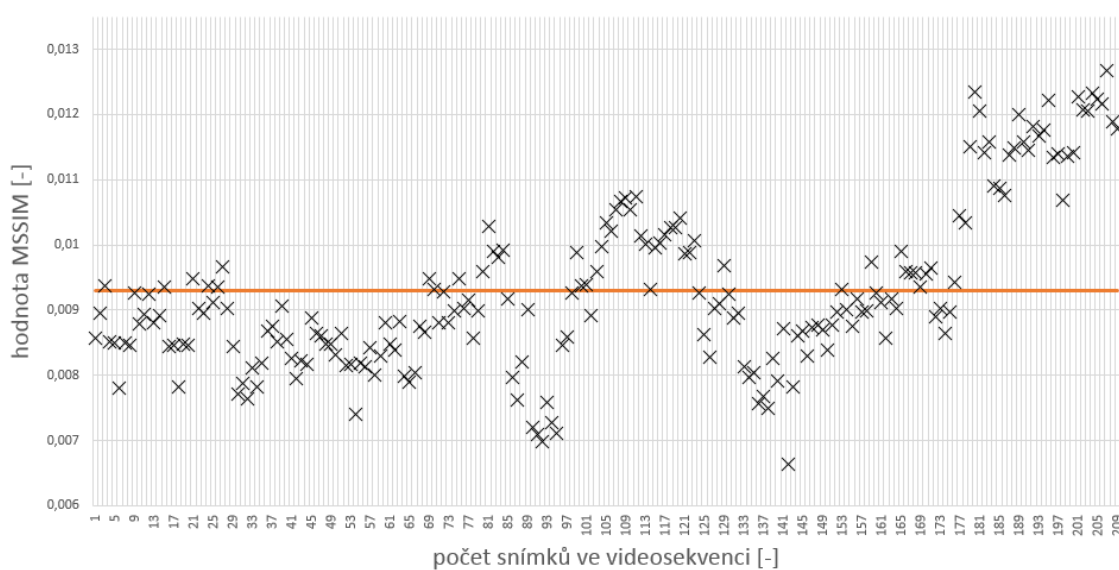
Nyní jsou na výstupu tři matice s normovanými hodnotami všech MSSIM indexů jednotlivých snímků ve videosekvenci (každá matice pro jednu oblast zájmu). Poslední částí programu je výpočet konečné hodnoty určující objektivní kvalitu obrazu.

6.2.5 Konečná hodnota objektivní kvality obrazu

Jelikož výstupem objektivní metody má být (stejně jako u subjektivních testů) jedno číslo reprezentující výslednou kvalitu testované videosekvence, je nutné k tomu uzpůsobit výstup programu. Jak bylo řečeno na konci kapitoly 6.2.4,

v současné chvíli jsou na výstupu tři matice zachycující normovanou hodnotu MSSIM pro každý snímek videosekvence a také pro každou část ROI.

Prvním krokem pro získání objektivní kvality obrazu je určení jednotné normované hodnoty MSSIM pro každou část ROI ve všech snímcích. To je provedeno jednoduchým spočtením průměru všech normovaných MSSIM hodnot ze všech snímků pro danou část ROI. Tím jsou získány pouze tři normované hodnoty MSSIM (jedna pro každou oblast ROI), se kterými se dále pracuje. Grafické znázornění výsledku tohoto postupu je vidět na obr. 6.3. Body v obrázku nastiňují normované hodnoty MSSIM každého i -tého snímku jedné části ROI ve videosekvenci. Oranžová čára znázorňuje průměrnou hodnotu MSSIM v dané části ROI.



Obr. 6.3: Zjištění průměrné normované hodnoty MSSIM ze všech snímků jedné části ROI videosekvence

Druhým krokem je váhování vypočtených hodnot MSSIM podle důležitosti oblastí ROI. Smyslem váhování je určení výsledných hodnot FQR tak, aby konečné hodnocení co nejlépe odpovídalo subjektivním testům u daných videosekvencí. Váhování je implementováno v programu pomocí následující části:

$$FQR = k * ((A * MSSIM1)^2 + B * MSSIM2 + C * MSSIM3) ;$$

kde $MSSIM1$, $MSSIM2$ a $MSSIM3$ označují vypočtené celkové normované MSSIM hodnoty jednotlivých částí ROI, FQR obsahuje výslednou objektivní kvalitu videosekvence, k je konstanta a A , B , C označují parametry, kterými je každá část ROI ve výsledku váhována.

Váhování je v příkazu složeno ze dvou částí. První část zahrnuje vynásobení každé ROI jedním parametrem, který vyjadřuje její důležitost v celkovém hodnocení. Pro určení správných hodnot parametrů A , B a C musíme vyřešit soustavu lineárních rovnic podle vzorce:

$$S = (A * MSSIM1)^2 + B * MSSIM2 + C * MSSIM3 \quad (6.3)$$

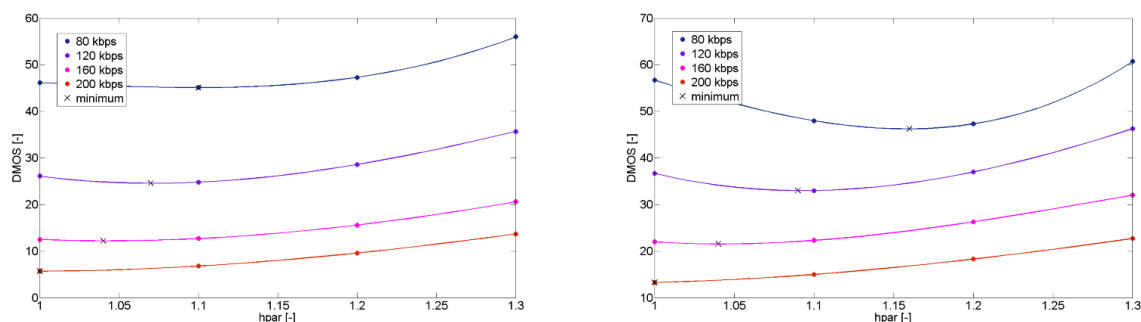
kde S značí výsledek subjektivního hodnocení pro danou videosekvenci, $MSSIM1$, $MSSIM2$, $MSSIM3$ vyjadřuje vypočtené normované hodnoty MSSIM jednotlivých ROI a A , B , C jsou hledané parametry. Počet rovnic se rovná součtu testovaných videosekvencí u jedné referenční videosekvence. Tento součet je pak vynásoben počtem hodnocených referenčních videosekvencí. Například pokud se v testovaných videosekvencích liší bitová rychlost (80, 120, 160 a 200 kbps), parametr $hpar$ (1.0, 1.1, 1.2 a 1.3) a testují se celkem 4 referenční videosekvence, potom soustavu lineárních rovnic tvoří 64 členů. Pro účely vyřešení takové soustavy lineárních rovnic byla použita metoda nejmenších čtverců. Podrobný popis této metody může čtenář nalézt např. v [33].

Implementace metody nejmenších čtverců je v programu MATLAB provedeno ve skriptu `ctverce.m`. Vstupem jsou zde hodnoty subjektivních testů ve formě vektoru b a vypočtené normované hodnoty MSSIM jednotlivých ROI částí ve formě matice A , kde počet sloupců odpovídá počtu hledaných parametrů. Počet řádků v matici A (a také počet prvků ve vektoru b) vyjadřuje celkový počet řešených lineárních rovnic. V tomto případě jich bylo celkem 108. Zjištění parametrů A , B , C je pak v MATLABu zajištěno jednoduchým příkazem:

```
par=A\b;
```

kde proměnná `par` obsahuje tři prvky určující hodnoty parametrů A , B a C . Posledním krokem je vydělení všech parametrů konstantou k , která se rovná nejmenší kladné hodnotě vypočtených parametrů. Díky tomu získáme hodnoty parametrů odpovídající jejich důležitosti (parametr s nejmenší vahou má tak hodnotu 1). Při porovnání se subjektivními testy jsou parametry znovu konstantou k vynásobeny, aby nebyla porušena rovnost rovnice 6.3.

Druhá část váhování je založena na povaze výsledků subjektivních testů. Na obr. 6.4 je vidět grafické zpracování výsledků subjektivních testů pro obě testovaná rozlišení [30].



Obr. 6.4: Závislost hodnot subjektivních testů DMOS na parametru $hpar$ pro videosekvence o velikosti 544x544 pixelů (vlevo) a 640x640 pixelů (vpravo) [30]

Z charakteru grafů lze dobře rozpoznat, že se snižující se bitovou rychlostí roste hodnota $hpar$ pro nejlépe hodnocenou subjektivní hodnotu. Zatímco u rychlosti 200 kbps je nejlépe hodnocen výsledek s $hpar = 1.0$, u rychlosti 80 kbps je ve videosekvencích s rozlišením 544x544 pixelů nejlépe hodnocen výsledek s $hpar = 1.1$ a u videosekvencí s rozlišením 640x640 pixelů dokonce výsledek s $hpar = 1.15$. Z těchto výsledků vyplývá, že se snižující se hodnotu kbps roste důležitost kvality obličeje vůči kvalitě pozadí.

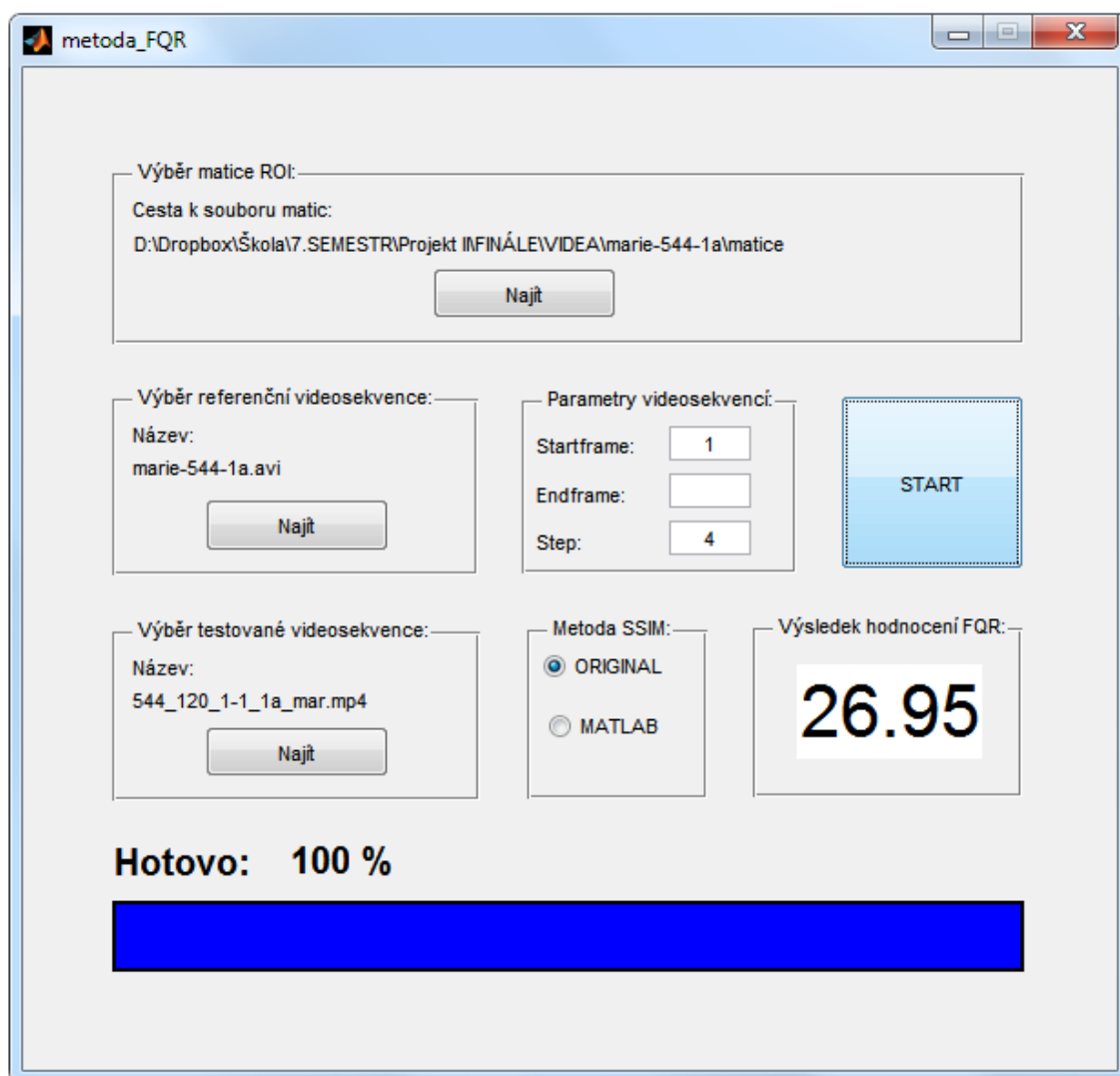
Tyto výsledky jsou v navržené metodě implementovány pomocí kvadrátu první normované hodnoty MSSIM (oblast zájmu 1 - obličej). Tento způsob při nižších bitových rychlostech zvýrazní viditelnost chyb pro objektivní hodnocení kvality obrazu v hlavní oblasti zájmu (obličej), což lépe odpovídá popsaným výsledkům subjektivních testů.

Nyní je na výstupu jediné číslo, které charakterizuje výslednou objektivní kvalitu (FQR) testované videosekvence. Výsledky normovaných hodnot MSSIM pro jednotlivé oblasti zájmu, zjištěných parametrů a porovnání se subjektivními testy je popsáno v kapitole 7.

6.3 Uživatelský panel

Výše popsaný program se spouští z uživatelského panelu sloužícímu k jednoduchému a intuitivnímu ovládní programu, který realizuje navrženou objektivní metodu. Obsahuje základní ovládací prvky k nastavení parametrů videosekvencí a přehledně zobrazuje dosažené výsledky. Uživatelský panel je rozdělen na několik bloků. Tři bloky slouží k načtení videosekvencí a matic určující ROI. Kliknutím na tlačítko „Najít“ v těchto blocích se otevře nové okno, ve kterém uživatel najde cestu k hledanému souboru. Po nalezení cesty se u bloků výběru videosekvencí zobrazí název souboru a u bloku výběru matic ROI se zobrazí kompletní cesta ke složce obsahující matice pro dané videosekvence. Blok „Parametry videosekvencí:“ poskytuje nastavení několika parametrů. Parametr „Startframe:“ charakterizuje, od jakého snímku v pořadí videosekvence má začít hodnocení objektivní kvality obrazu. Parametr „Endframe:“ naopak určuje, do jakého snímku v pořadí videosekvence má hodnocení probíhat. Posledním parametrem je „Step:“, který slouží k vybrání kroku hodnocených snímků. Díky tomu lze hodnotit pouze každý druhý, třetí, čtvrtý... až n -tý snímek ve videosekvenci. Blok „Metoda SSIM:“ slouží ke zvolení metody, která bude pro výpočet metriky použita. Tlačítko START slouží ke spuštění výpočtu. Po jeho zmáčknutí se spustí algoritmus metody a postup měření je zobrazován v dolní části panelu. Po dosažení stavu „Hotovo: 100 %“ se v okně „Výsledek hodnocení FQR:“ zobrazí hodnota odpovídající naměřené kvalitě FQR testované videosekvence.

Na obr. 6.5 je uživatelský panel zobrazen po dokončení hodnocení kvality jedné videosekvence.



Obr. 6.5: Vzhled uživatelského prostředí

Hodnoty parametrů A , B , C a konstanty k jsou uloženy uvnitř finálního programu a uživatel je v uživatelském panelu nemůže měnit. Hodnoty těchto parametrů a konstanty k jsou nastaveny na základě výsledků popsaných v kapitole 7.

Výpočty byly prováděny na pracovní stanici HP Z420 s 12ti-jádrovým procesorem Intel® Xeon® CPU E5-1650 s taktovací rychlostí 3,2 GHz a pamětí RAM 24 GB. Doba výpočtu díky této stanici trvá podle nastavení vstupních parametrů u jedné videosekvence od několika sekund do cca deseti minut. Velký rozdíl doby výpočtu vzniká zejména na základě zvolení druhu použité metody. U originální SSIM trvá výpočet až několikrát kratší dobu než u SSIM MATLABu. Je to způsobeno tím, že originální SSIM má podstatně jednodušší zpracování, kdy je konvoluce prováděna pouze přes jasovou složku atd.

Kapitola 7

Vyhodnocení objektivní kvality videosekvencí

Funkčnost navržené metody byla vyzkoušena na souboru videosekvencí, pro které byly k dispozici hodnoty subjektivních testů. Odhady kvality snímků videosekvence (FQR) byly porovnávány s průměrnými hodnotami subjektivních hodnocení DMOS (Pro subjektivní testování vizuální kvality těchto videosekvencí byla zvolena metoda DSCQS [30]).

Pro nastavení parametrů popsaných v kapitole 6.2.5 bylo použito celkem 8 referenčních videosekvencí (4 v rozlišení 544x544 pixelů a 4 v rozlišení 640x640 pixelů). U tří referenčních videosekvencí byly testované videosekvence k dispozici ve všech možnostech nastavení parametrů (bitová rychlost: 80, 120, 160 a 200 kbps, *hpar*: 1.0, 1.1, 1.2 a 1.3), ve zbylých chybělo nastavení *hpar* 1.0. Výsledky hodnocení MSSIM testovaných videosekvencí jsou společně s výsledky subjektivních testů uvedeny v tabulkách u jednotlivých vyhodnocení.

Výsledné porovnání subjektivních testů a zjištěných objektivních hodnot po váhování je zobrazeno do grafu. V grafu jsou vždy na horizontální ose vyneseny hodnoty FQR a na vertikální ose průměrné hodnoty DMOS pro danou videosekvenci. Obě osy jsou cejchovány v procentech, kde 100% znamená nejhorší kvalitu videosekvence a 0% znamená nejlepší kvalitu videosekvence. Diagonální modrou přímkou jsou vyznačeny ideální hodnoty, kterým odpovídá situace, kdy jsou hodnoty FQR naprosto totožné jako průměrné hodnoty subjektivních hodnocení DMOS.

7.1 Výsledky objektivní kvality obrazu u originální SSIM

Výstupy normovaných hodnocení MSSIM pro jednotlivé části ROI (viz kap. 6.2.4) testovaných videosekvencí u jedné referenční videosekvence jsou zobrazeny vždy v jedné tabulce, kde BR je bitová rychlost [kbps] testované videosekvence, *hpar* značí hodnotu *hpar* nastavenou u testované videosekvence a MSSIM1, MSSIM2 a MSSIM3 značí hodnoty výstupu normovaných hodnocení MSSIM pro jednotlivé části ROI. U výstupů subjektivních testů jsou v tabulce místo hodnot MSSIM hodnoty DMOS.

Tab. 7.1: Normované hodnoty MSSIM pro videosekvenci marie-544-1a

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
MSSIM1	0,003801	0,003241	0,002846	0,002480	0,002690	0,002277	0,002046	0,001829
MSSIM2	0,013063	0,012866	0,013171	0,013066	0,009518	0,009335	0,009418	0,009528
MSSIM3	0,010078	0,011311	0,012763	0,013684	0,008350	0,009131	0,010240	0,011098
BR [kbps]	160				200			
MSSIM1	0,002109	0,001851	0,001631	0,001523	0,001788	0,001574	0,001427	0,001312
MSSIM2	0,007691	0,007598	0,007516	0,007788	0,006631	0,006501	0,006541	0,006612
MSSIM3	0,007453	0,008029	0,008798	0,009823	0,006839	0,007297	0,007852	0,008640

Tab. 7.2: Normované hodnoty MSSIM pro videosekvenci marie-544-2a

BR [kb/s]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
MSSIM1	0,003746	0,003245	0,002870	0,002497	0,002696	0,002281	0,002013	0,001782
MSSIM2	0,012065	0,012106	0,012443	0,012600	0,009020	0,008853	0,008927	0,009011
MSSIM3	0,010356	0,011807	0,013279	0,014894	0,008681	0,009541	0,010675	0,011632
BR [kb/s]	160				200			
MSSIM1	0,002071	0,001777	0,001584	0,001441	0,001709	0,001500	0,001361	0,001224
MSSIM2	0,007146	0,007026	0,007158	0,007230	0,006113	0,006053	0,006127	0,006078
MSSIM3	0,007585	0,008141	0,009008	0,009875	0,006899	0,007462	0,008084	0,008754

Tab. 7.3: Normované hodnoty MSSIM pro videosekvenci marie-640-1a

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
MSSIM1	0,004783	0,004020	0,003476	0,003065	0,003385	0,002894	0,002494	0,002240
MSSIM2	0,013910	0,013698	0,013854	0,014030	0,010290	0,010052	0,010087	0,010194
MSSIM3	0,012002	0,013320	0,015308	0,016725	0,009839	0,011030	0,012115	0,013250
BR [kbps]	160				200			
MSSIM1	0,002701	0,002304	0,002052	0,001858	0,002259	0,001950	0,001759	0,001616
MSSIM2	0,008368	0,008285	0,008276	0,008365	0,007213	0,007125	0,007115	0,007250
MSSIM3	0,008775	0,009565	0,010550	0,011782	0,008094	0,008733	0,009331	0,010195

Tab. 7.4: Normované hodnoty MSSIM pro videosekvenci marie-544-1b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,003146389	0,002553953	0,002488789	0,002242175	0,001987378	0,001751624
MSSIM2	0,012500029	0,011602798	0,013092846	0,009058694	0,009061385	0,009069769
MSSIM3	0,011207506	0,012122513	0,014156376	0,009094390	0,009985960	0,011120916
BR [kbps]	160			200		
MSSIM1	0,001751384	0,001540056	0,001401484	0,001486512	0,001316144	0,001209906
MSSIM2	0,007153633	0,007154530	0,007200395	0,006224878	0,006160136	0,006215061
MSSIM3	0,007882271	0,008521802	0,009298135	0,007178602	0,007692777	0,008363256

Tab. 7.5: Normované hodnoty MSSIM pro videosekvenci marie-544-2b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,003280362	0,002942937	0,002458800	0,002165363	0,001898244	0,001663497
MSSIM2	0,013566453	0,014138904	0,013768110	0,009237568	0,009391766	0,009273095
MSSIM3	0,012453058	0,014827595	0,015969432	0,009557763	0,010752321	0,011749751
BR [kbps]	160			200		
MSSIM1	0,001690626	0,001490682	0,001356640	0,001398449	0,001290248	0,001191439
MSSIM2	0,007434215	0,007372198	0,007471571	0,006263836	0,006341452	0,006476839
MSSIM3	0,008210424	0,008890439	0,009985657	0,007462482	0,008124790	0,008845819

Tab. 7.6: Normované hodnoty MSSIM pro videosekvenci marie-640-1b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,003799352	0,003327170	0,002891414	0,002709727	0,002356776	0,002090152
MSSIM2	0,013488898	0,013551390	0,013485575	0,009790317	0,009759796	0,009653946
MSSIM3	0,013144553	0,014748258	0,016662689	0,010673578	0,011780081	0,012929478
BR [kbps]	160			200		
MSSIM1	0,002139997	0,001922902	0,001723710	0,001820507	0,001651588	0,001506748
MSSIM2	0,007859272	0,007925913	0,007964172	0,006823342	0,006861607	0,006908389
MSSIM3	0,009240755	0,009939840	0,011164306	0,008361503	0,009077269	0,009828883

Tab. 7.7: Normované hodnoty MSSIM pro videosekvenci marie-640-2a

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,004337691	0,003656024	0,003144130	0,002919998	0,002526998	0,002237387
MSSIM2	0,013761618	0,013385482	0,013468037	0,009349428	0,009394961	0,009342417
MSSIM3	0,014250339	0,016114627	0,017974325	0,011082647	0,012253340	0,013421515
BR [kbps]	160			200		
MSSIM1	0,002253881	0,002001968	0,001837829	0,001927543	0,001707883	0,002018901
MSSIM2	0,007447906	0,007569130	0,007750617	0,006508911	0,006438270	0,009216723
MSSIM3	0,009619386	0,010520409	0,011871436	0,008722774	0,009434935	0,013834642

Tab. 7.8: Normované hodnoty MSSIM pro videosekvenci marie-640-2b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,003878423	0,003108270	0,002894601	0,002629966	0,002224669	0,002119927
MSSIM2	0,013200220	0,012190558	0,012838046	0,009111782	0,008815293	0,009373172
MSSIM3	0,014137324	0,015484200	0,017695644	0,010805719	0,012118125	0,013964398
BR [kbps]	160			200		
MSSIM1	0,002190851	0,001834991	0,001682435	0,001832975	0,001637886	0,001501864
MSSIM2	0,007669870	0,007287234	0,007364762	0,006489784	0,006443709	0,006443709
MSSIM3	0,009606773	0,010529054	0,011740931	0,008668078	0,009292212	0,012724092

Tab. 7.9: Hodnoty DMOS pro videosekvence o rozlišení 544x544 pixelů

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
DMOS	46,2	45,1	47,3	56,0	26,1	24,8	28,6	35,7
BR [kbps]	160				200			
DMOS	12,5	12,7	15,6	20,6	5,7	6,8	9,6	13,7

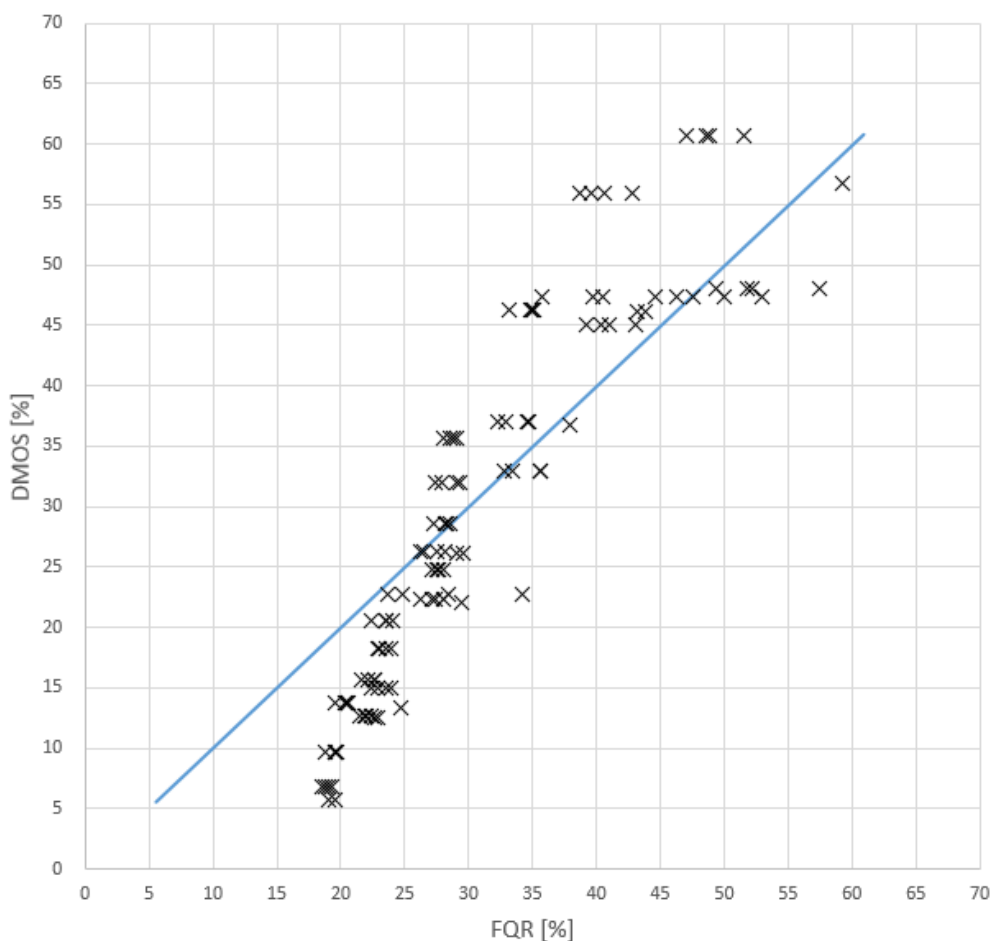
Tab. 7.10: Hodnoty DMOS pro videosekvence o rozlišení 640x640 pixelů

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
DMOS	56,7	48,0	47,3	60,7	36,7	33,0	37,0	46,3
BR [kbps]	160				200			
DMOS	22,0	22,3	26,3	32,0	13,3	15,0	18,3	22,7

Na základě výše uvedených výstupů hodnot MSSIM a DMOS byly pomocí metody nejmenších čtverců určeny následující hodnoty parametrů A , B , C a konstanty k (viz. kap. 6.2.5):

Tab. 7.11: Výsledky parametrů A , B , C a konstanty k

Parametry	A	B	C	k
Hodnoty	53,81065026	1	3,663807248	477,0119536



Obr. 7.1: Srovnání výsledků DMOS a FQR hodnocení kvality obrazu originální SSIM

Na obr. 7.1. je zobrazen graf s výsledky objektivních a subjektivních testů hodnocení kvality obrazu při použití originální metody SSIM, které jsou porovnány ve stejném měřítku. Modrá čára v grafu znázorňuje ideální stav, kdy se subjektivní testy přesně rovnají testům objektivním. Výsledky objektivních testů jsou vůči testům subjektivním v drobných odchylkách, protože ne vždy lze subjektivní hodnocení kvality obrazu přesně postihnout. Z grafu lze pozorovat, že v oblasti kolem hodnoty 25% směrem dolů objektivní metoda více podhodnocuje výsledky hodnocení kvality obrazu. To lze vysvětlit tím, že metoda i v těchto bitových rychlostech stále detekuje určité chyby v obraze, které ale pro neslyšícího diváka jsou důležité (rušící) a tudíž je nevnímá.

Z tabulky 7.11., která obsahuje vypočtené parametry oblastí zájmu lze pozorovat, že nejvyšší zastoupení má podle předpokladů část ROI 1, tedy oblast obličeje. Je hodnocena jako 50x důležitější než oblast rukou, což potvrzuje předpoklady uvedené v kapitole 4. Hodnota parametru u oblasti zájmu 2 (ruce), je algoritmem hodnocena cca 3,6x méně než hodnota parametru u oblasti zájmu 3 (pozadí). Na tento výsledek mají velký vliv subjektivní testy, ze kterých vyplývá, že pozadí má v subjektivní kvalitě nezanedbatelnou váhu. S velkým snižováním kvality pozadí (vyšší hodnoty *hpar*) se také rapidně snižuje subjektivní hodnocení, jak vyplývá z tabulek 7.9 a 7.10. Proto je i v algoritmu výpočtu navrhované metody parametr zdůrazňující pozadí trochu vyšší než parametr zdůrazňující část rukou. Ovšem stále je parametr zdůrazňující pozadí několikrát menší než parametr zdůrazňující obličej.

Z grafu na obr. 7.1 lze pozorovat, že výsledky objektivního hodnocení kvality obrazu nemají naprosto lineární závislost vzhledem k výsledkům subjektivních měření. Důvodem může být nedostatek zhodnocených videosekvencí s parametrem *hpar* 1.0, které nebyly k dispozici. Díky tomu byly při výpočtu parametrů metodou nejmenších čtverců převáženy výsledky parametrů *hpar* 1.0, které byly k dispozici celkově pouze u tří referenčních videosekvencí, výsledky ostatních parametrů *hpar* (1.1, 1.2, 1.3), které byly k dispozici u všech osmi referenčních videosekvencí.

7. 2 Výsledky objektivní kvality obrazu u SSIM MATLABu

Stejně jako v kapitole 7.1 jsou i v této kapitole výstupy normovaných hodnocení MSSIM pro jednotlivé části ROI (viz kap. 6.2.4) testovaných videosekvencí u jedné referenční videosekvence zobrazeny vždy v jedné tabulce, kde BR je bitová rychlost [kbps] testované videosekvence, *hpar* značí hodnotu *hpar* nastavenou u testované videosekvence a MSSIM1, MSSIM2 a MSSIM3 značí hodnoty výstupu normovaných hodnocení MSSIM pro jednotlivé části ROI. U výstupů subjektivních testů jsou v tabulce místo hodnot MSSIM hodnoty DMOS.

Tab. 7.12: Normované hodnoty MSSIM pro videosekvenci marie-544-1a

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
MSSIM1	0,001341	0,001140	0,001015	0,000881	0,000934	0,000797	0,000710	0,000638
MSSIM2	0,006756	0,006750	0,006973	0,006952	0,005002	0,004948	0,005060	0,005118
MSSIM3	0,011204	0,013126	0,016034	0,018273	0,008946	0,010066	0,011853	0,013033
BR [kbps]	160				200			
MSSIM1	0,000736	0,000640	0,000567	0,000524	0,000615	0,000538	0,000486	0,000448
MSSIM2	0,004079	0,004079	0,004059	0,004202	0,003578	0,003531	0,003551	0,003597
MSSIM3	0,007715	0,008473	0,009681	0,011375	0,006967	0,007550	0,008365	0,009376

Tab. 7.13: Normované hodnoty MSSIM pro videosekvenci marie-544-2a

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
MSSIM1	0,001272	0,001126	0,000994	0,000880	0,000914	0,000784	0,000696	0,000621
MSSIM2	0,006111	0,006239	0,006390	0,006526	0,004609	0,004554	0,004585	0,004685
MSSIM3	0,011441	0,013424	0,015647	0,019150	0,008995	0,010357	0,011585	0,013508
BR [kbps]	160				200			
MSSIM1	0,000707	0,000610	0,000550	0,000495	0,000585	0,000509	0,000465	0,000414
MSSIM2	0,003723	0,003659	0,003744	0,003788	0,003206	0,003189	0,003212	0,003190
MSSIM3	0,007613	0,008461	0,009696	0,010927	0,006809	0,007512	0,008389	0,009423

Tab. 7.14: Normované hodnoty MSSIM pro videosekvenci marie-640-1a

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
MSSIM1	0,001666	0,001391	0,001202	0,001068	0,001138	0,000977	0,000853	0,000771
MSSIM2	0,006553	0,006520	0,006589	0,006861	0,004894	0,004810	0,004854	0,004967
MSSIM3	0,014083	0,016396	0,018491	0,023838	0,010363	0,012225	0,014011	0,016106
BR [kbps]	160				200			
MSSIM1	0,000908	0,000779	0,000690	0,000624	0,000748	0,000652	0,000583	0,000536
MSSIM2	0,003954	0,003945	0,003969	0,004039	0,003429	0,003407	0,003408	0,003499
MSSIM3	0,008839	0,010006	0,011566	0,013046	0,007975	0,008857	0,009977	0,011321

Tab. 7.15: Normované hodnoty MSSIM pro videosekvenci marie-544-1b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,001096341	0,000884362	0,000864946	0,000772911	0,000684288	0,000600122
MSSIM2	0,006524153	0,006154987	0,006866008	0,004770903	0,004834109	0,004815633
MSSIM3	0,012836748	0,014003983	0,017046382	0,009892330	0,011240556	0,012533040
BR [kbps]	160			200		
MSSIM1	0,000591037	0,000526631	0,000473119	0,000499460	0,000445276	0,000408075
MSSIM2	0,003811676	0,003838405	0,003863695	0,003317619	0,003293734	0,003340149
MSSIM3	0,008267641	0,009203213	0,010283293	0,007382576	0,008045704	0,008920438

Tab. 7.16: Normované hodnoty MSSIM pro videosekvenci marie-544-2b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,001116218	0,001035492	0,000862077	0,000744251	0,000648698	0,000572619
MSSIM2	0,006738290	0,007216470	0,007070576	0,004730601	0,004824350	0,004734176
MSSIM3	0,014608240	0,018799595	0,020912881	0,010377541	0,011947610	0,013609698
BR [kbps]	160			200		
MSSIM1	0,000572554	0,000512100	0,000464308	0,000476033	0,000435438	0,000401853
MSSIM2	0,003849636	0,003805619	0,003880113	0,003267044	0,003331306	0,003390723
MSSIM3	0,008533220	0,009584463	0,010860705	0,007475308	0,008484573	0,009721371

Tab. 7.17: Normované hodnoty MSSIM pro videosekvenci marie-640-1b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,001324586	0,001147994	0,001028465	0,000945717	0,000811199	0,000715417
MSSIM2	0,006432533	0,006478426	0,006451778	0,004677886	0,004670592	0,004641143
MSSIM3	0,015322296	0,017680852	0,021540158	0,011584505	0,013267379	0,015159110
BR [kbps]	160			200		
MSSIM1	0,000726932	0,000649994	0,000586402	0,000612242	0,000558118	0,000506476
MSSIM2	0,003712993	0,003739569	0,003792159	0,003232553	0,003241426	0,003291193
MSSIM3	0,009460569	0,010760246	0,012259675	0,008322733	0,009331277	0,010577656

Tab. 7.18: Normované hodnoty MSSIM pro videosekvenci marie-640-2a

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,001535230	0,001300387	0,001120475	0,001016848	0,000870732	0,000775048
MSSIM2	0,006344205	0,006324023	0,006395645	0,004380130	0,004412691	0,004434520
MSSIM3	0,016576019	0,020386342	0,023199160	0,011851012	0,013739230	0,015580541
BR [kbps]	160			200		
MSSIM1	0,000772895	0,000696854	0,000632618	0,000664131	0,000584540	0,000501288
MSSIM2	0,003502455	0,003534541	0,003662122	0,003044025	0,003031004	0,003032630
MSSIM3	0,009900480	0,011395913	0,013199878	0,008619596	0,009771447	0,019624656

Tab. 7.19: Normované hodnoty MSSIM pro videosekvenci marie-640-2b

BR [kbps]	80			120		
hpar	1.1	1.2	1.3	1.1	1.2	1.3
MSSIM1	0,001298140	0,001028333	0,000971561	0,000865581	0,000731408	0,000706230
MSSIM2	0,006156811	0,005804859	0,006212192	0,004258557	0,004131929	0,004488365
MSSIM3	0,016708479	0,019744628	0,023823436	0,011775674	0,013782340	0,016464518
BR [kbps]	160			200		
MSSIM1	0,000716540	0,000595692	0,000554202	0,000592829	0,000525591	0,000482532
MSSIM2	0,003564091	0,003410154	0,003485625	0,003042901	0,003015974	0,003016020
MSSIM3	0,010077154	0,011109558	0,013157550	0,008681656	0,009692950	0,011216444

Tab. 7.20: Hodnoty DMOS pro videosekvence o rozlišení 544x544 pixelů

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
DMOS	46,2	45,1	47,3	56,0	26,1	24,8	28,6	35,7
BR [kbps]	160				200			
DMOS	12,5	12,7	15,6	20,6	5,7	6,8	9,6	13,7

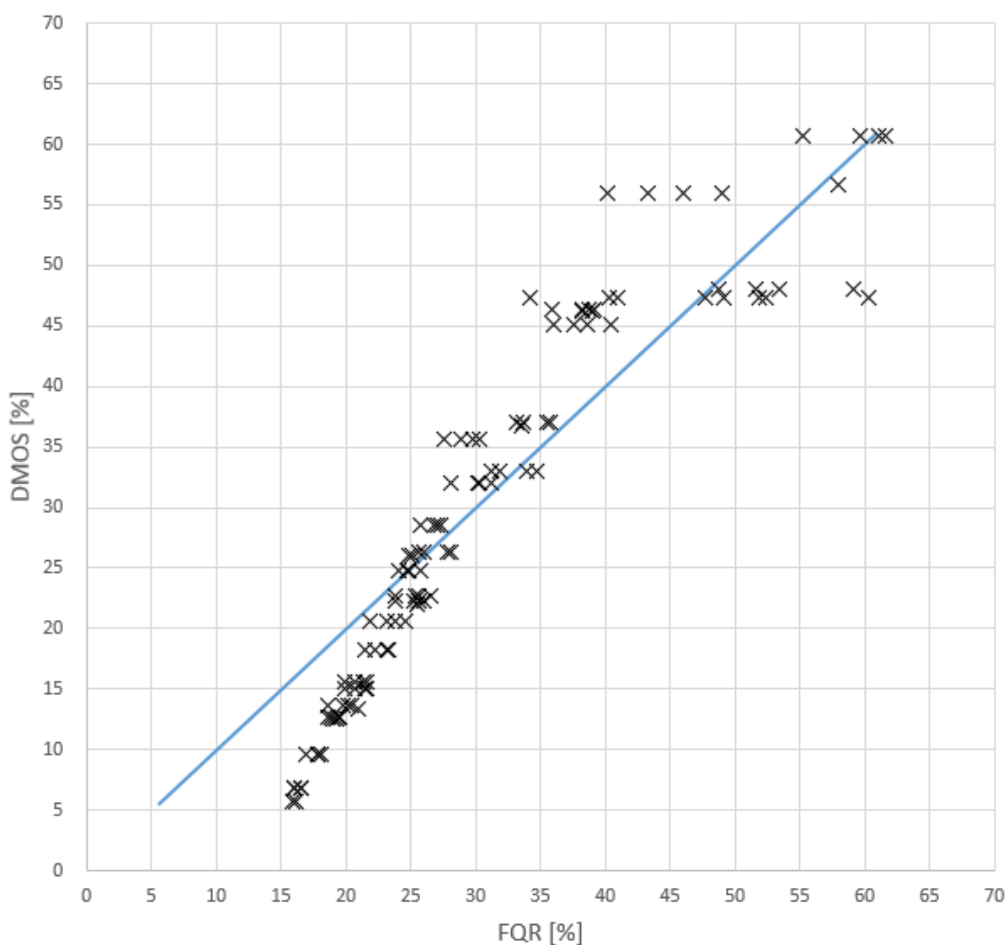
Tab. 7.21: Hodnoty DMOS pro videosekvence o rozlišení 640x640 pixelů

BR [kbps]	80				120			
hpar	1.0	1.1	1.2	1.3	1.0	1.1	1.2	1.3
DMOS	56,7	48,0	47,3	60,7	36,7	33,0	37,0	46,3
BR [kbps]	160				200			
DMOS	22,0	22,3	26,3	32,0	13,3	15,0	18,3	22,7

Na základě výše uvedených výstupů hodnot MSSIM a DMOS byly pomocí metody nejmenších čtverců určeny následující hodnoty parametrů A , B , C a konstanty k (viz. kap. 6.2.5):

Tab. 7.22: Výsledky parametrů A , B , C a konstanty k

Parametry	A	B	C	k
Hodnoty	71,00174906	-0,594768075	1	2396,773847



Obr. 7.2: Srovnání výsledků DMOS a FQR hodnocení kvality obrazu SSIM MATLABu

Na obr. 7.2. je zobrazen graf s výsledky objektivních a subjektivních testů hodnocení kvality obrazu při použití metody SSIM MATLABu, které jsou porovnány ve stejném měřítku. Modrá čára v grafu znázorňuje ideální stav, kdy se subjektivní testy přesně rovnají testům objektivním. Výsledky objektivních testů jsou i zde v drobných odchylkách vůči testům subjektivním, protože ne vždy lze subjektivní hodnocení kvality obrazu přesně postihnout.

Charakter výsledků ve srovnání s obr. 7.1 je podobný. Znovu největší váhu ve výsledku zastupuje oblast zájmu 1 (obličej). Stejně i druhá nejvyšší hodnota parametru zdůrazňující ROI je u oblasti zájmu 3 (pozadí). Zajímavá je ale hodnota parametru B (oblast zájmu 2), která je záporná. Znaménko u této hodnoty lze jedinečně vysvětlit tím, že metoda nejmenších čtverců se snaží nalézt nejlepší možné matematické řešení rovnic vycházejících z výsledků normovaných měření MSSIM (viz kap. 6.2.5). To v tomto případě znamená, že algoritmus při výpočtu nejmenší chyby mezi výsledky testů místo aditivního sčítání všech oblastí zájmu s různou vahou oblast zájmu 2 odečítá.

I v tomto případě chybělo více videosekvencí s parametrem $hpar$ 1.0, které mohli ovlivnit výsledky výpočtu parametrů zdůrazňujících části ROI.

7.3 Ohodnocení funkčnosti metody

K ohodnocení funkčnosti metody byl vypočítán Pearsonův koeficient korelace, který udává vztah mezi subjektivními hodnoceními a hodnoceními objektivní metody. Pearsonův koeficient korelace nabývá hodnot od 0 do 1, kde hodnota 1 značí zcela přímou závislost obou hodnocení a hodnota 0 naopak značí nulovou závislost mezi oběma hodnoceními. Pearsonův koeficient korelace lze popsat následujícím matematickým vztahem [35]:

$$\rho(A, B) = \frac{1}{N - 1} \sum_{i=1}^N \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right), \quad (5.3)$$

kde A představuje výsledky subjektivních testů, B značí výsledky objektivních testů, μ_A a σ_A jsou hodnoty standardní odchylky A , μ_B a σ_B jsou hodnoty standardní odchylky B , N značí počet výsledků a ρ udává hodnotu výsledného Pearsonova koeficientu korelace. Pearsonův koeficient korelace je spočten v MATLABu ve skriptu `ctverce.m`. Následující tabulka uvádí výsledné hodnoty u obou způsobů objektivního měření využívající originální metodu SSIM a metodu SSIM MATLABu.

Tab. 7.23: Porovnání Pearsonova koeficientu korelace u objektivních měření

Metoda	Original SSIM	SSIM MATLABu
Hodnoty	0,913756057	0,935041832

Z tabulky 7.23 vyplývá, že Pearsonův koeficient korelace vyšel pro obě měření vysoký (přesahující hodnotu $\rho = 0,91$). Dokonce vyšel lépe pro metodu využívající metodu SSIM MATLABu (i přes zvláštní přístup výpočtu parametrů zdůrazňujících kvalitu jednotlivých částí ROI) než metodu využívající originální SSIM.

V tuto chvíli se už musí rozhodnout samotný uživatel, zda chce využít metodu využívající metodu SSIM MATLABu (pomalejší výpočet, $\rho = 0,935$) nebo metodu využívající originální metodu SSIM (rychlejší výpočet, $\rho = 0,914$).

Závěr

Tato bakalářská práce pojednává o návrhu metody objektivního hodnocení obrazu mluvčího ZJ. Při návrhu metody byla pozornost opřena o výsledky oblastí zájmu neslyšících diváků a využití již vytvořené metody SSIM ve dvou verzích. V práci je nejprve popsán lidský zrakový systém HVS, na jehož základě je objektivní metoda navržena. Dále jsou rozebrány metody objektivního hodnocení kvality obrazu a podrobně je rozebrána metoda SSIM včetně kompletní popisu výpočtu její MSSIM metriky. Dále jsou převzaty výsledky zkoumání oblastí zájmu neslyšících diváků, které jsou aplikovány při výpočtu konečné hodnoty objektivního měření kvality obrazu. Daná problematika byla řešena a naprogramována v prostředí MATLAB, které je uzpůsobeno k práci s různými druhy signálů a obsahuje vhodné nástroje pro zpracování videosekvencí.

Hlavním cílem při návrhu metody objektivního hodnocení obrazu bylo kladení důrazu na oblasti zájmu neslyšících a vlastnosti HVS. Hlavním kamenem celé metody je použití metriky MSSIM, která splňuje požadavky pro dobré postihnutí HVS. Metrika byla použita ve dvou variantách: originální SSIM, která pochází přímo od autora metody a SSIM MATLABu, která je součástí samotného programu MATLAB. Oblasti zájmu neslyšícího diváka byly zakomponovány do algoritmu metody ve formě rozčlenění obrazu na jednotlivé části ROI a jejich závěrečné váhování podle jejich důležitosti. Jejich správné extrahování a zakomponování do metody představuje základ, aby metoda správně fungovala.

Metoda byla vyzkoušena na souboru videosekvencí zahrnujících mluvčího ZJ před modrým pozadím. Každá testovaná videosekvence byla kódována ve zdrojovém kódu x264 s různým nastavením rychlosti toku informací a parametrem *hpar*, který určuje kvalitu jednotlivých ROI částí vůči sobě navzájem. Výsledky testování byly použity k samotnému kalibrování výsledné navržené metody. Kalibrace byla prováděna na základě výsledků subjektivních testů testovaných videosekvencí. Pro kalibraci byla použita metoda nejmenších čtverců.

Funkčnost navržené objektivní metody byla nakonec ověřena vypočítáním Pearsonova koeficientu korelace, který udává vztah mezi subjektivními hodnoceními a objektivními výsledky metody. U obou způsobů výpočtu metody vyšel Pearsonův koeficient nad hodnotou 0,91, při využití SSIM MATLABu dokonce nad 0,935, což značí vysokou korelaci se subjektivními testy a tedy i funkčnost navržené metody při hodnocení obrazu mluvčího ZJ.

Dalším krokem kam by se práce mohla ubírat, je rozšíření počtu metrik kromě MSSIM i o další a jejich vzájemné porovnání. I přes dobrou korelaci navržené metody je stále dostatečný prostor pro zlepšení. Dalším cílem by také mohlo být aplikování metody na animovaného mluvčího ZJ, jehož vytvoření je cílem projektu. Součástí i tato bakalářská práce.

Seznam použitých obrázků

Obr. 1.1: Schématický řez lidským okem [2]	10
Obr. 1.2: Weber-Fechnerův zákon pro dvoubarevný experiment a jeho grafické vyjádření [2]	12
Obr. 1.3: Grafické znázornění CSF dle Dalyho [5]	13
Obr. 3.1: Blokové schéma Lubinova modelu HVS [18]	20
Obr. 3.2: Blokové schéma Dalyho VDP multikanálového modelu [5]	21
Obr. 3.3: Blokové schéma DCT modelu ve frekvenční oblasti [23]	22
Obr. 3.4: Blokové schéma Sarnoffova multikanálového modelu [24]	23
Obr. 3.5: Blokové schéma SSIM modelu [26]	25
Obr. 4.1: Základní znakovací prostor [28]	30
Obr. 4.2: Teplotní mapa videosekvence pro všechny subjekty a celou dobu klípu [28]	31
Obr. 6.1: Schéma postupu vytvoření oblastí zájmu v programu	37
Obr. 6.2: Odstranění redundantní informace z vytvořených oblastí zájmu	38
Obr. 6.3: Zjištění průměrné normované hodnoty MSSIM ze všech snímků jedné ROI části videosekvence	42
Obr. 6.4: Závislost hodnot subjektivních testů DMOS na parametru <i>hpar</i> pro videosekvence o velikosti 544x544 pixelů a 640x640 pixelů [30]	43
Obr. 6.5: Vzhled uživatelského prostředí	45
Obr. 7.1: Srovnání výsledků DMOS a FQR hodnocení kvality obrazu originální SSIM	49
Obr. 7.2: Srovnání výsledků DMOS a FQR hodnocení kvality obrazu SSIM MATLABu	53

Seznam použitých tabulek

Tab. 7.1: Normované hodnoty MSSIM pro videosekvenci marie-544-1a	47
Tab. 7.2: Normované hodnoty MSSIM pro videosekvenci marie-544-2a	47
Tab. 7.3: Normované hodnoty MSSIM pro videosekvenci marie-640-1a	47
Tab. 7.4: Normované hodnoty MSSIM pro videosekvenci marie-544-1b	47
Tab. 7.5: Normované hodnoty MSSIM pro videosekvenci marie-544-2b	48
Tab. 7.6: Normované hodnoty MSSIM pro videosekvenci marie-640-1b	48
Tab. 7.7: Normované hodnoty MSSIM pro videosekvenci marie-640-2a	48
Tab. 7.8: Normované hodnoty MSSIM pro videosekvenci marie-640-2b	48
Tab. 7.9: Hodnoty DMOS pro videosekvence o rozlišení 544x544 pixelů	49
Tab. 7.10: Hodnoty DMOS pro videosekvence o rozlišení 640x640 pixelů	49
Tab. 7.11: Výsledky parametrů A , B , C a konstanty k	49
Tab. 7.12: Normované hodnoty MSSIM pro videosekvenci marie-544-1a	51
Tab. 7.13: Normované hodnoty MSSIM pro videosekvenci marie-544-2a	51
Tab. 7.14: Normované hodnoty MSSIM pro videosekvenci marie-640-1a	51
Tab. 7.15: Normované hodnoty MSSIM pro videosekvenci marie-544-1b	51
Tab. 7.16: Normované hodnoty MSSIM pro videosekvenci marie-544-2b	52
Tab. 7.17: Normované hodnoty MSSIM pro videosekvenci marie-640-1b	52
Tab. 7.18: Normované hodnoty MSSIM pro videosekvenci marie-640-2a	52
Tab. 7.19: Normované hodnoty MSSIM pro videosekvenci marie-640-2b	52
Tab. 7.20: Hodnoty DMOS pro videosekvence o rozlišení 544x544 pixelů	53
Tab. 7.21: Hodnoty DMOS pro videosekvence o rozlišení 640x640 pixelů	53
Tab. 7.22: Výsledky parametrů A , B , C a konstanty k	53
Tab. 7.23: Porovnání Pearsonova koeficientu korelace u objektivních měření	54

Seznam použitých symbolů a zkratek

1D - jednodimenzionální
4D - čtyřdimenzionální
ASL - americký znakový jazyk
AVI - Audio Video Interleave
B - Blue (modrá)
BSL - Britský znakový jazyk
 C_r - chrominanční složka barevného prostoru Y $C_r C_b$
 C_b - chrominanční složka barevného prostoru Y $C_r C_b$
CIE L^*u^*v , CIE Luv - CIE barevný prostor stejných barevných rozdílů
CQE - Continuous Quality Evaluation
CSF - Contrast Sensitivity Function (Funkce kontrastní citlivosti)
ČSÚ - Český statistický úřad
ČZJ - Český znakový jazyk
DCT - Discrete Cosine Transform (diskrétní kosínová transformace)
DMOS – Differential Mean Opinion Score
DS - Double Stimulus
DSCQS - Double Stimulus Continuous Quality Scale
DSIS - Double Stimulus Impairment Scale
FQR – Final Quality Rating
FR - Full Reference
G - Green
H.264, x264 - High 264 (formát multimedialního kodeku)
Hbb TV – Hybrid broadcast broadband TV
HVS - Human Visual System (lidský vizuální systém)
JND - Just Noticeable Difference(s) (právě pozorovatelný rozdíl(y))
kbps – kilobits per second (kilobit za sekundu)
L - jasový kanál barevného prostoru CIE L^*u^*v
LBC - Local Band-limited Contrast (pásmově omezený kontrast)
MATLAB – MATrix LABoratory
MOS - Mean Opinion Score
MP4 – Motion Picture 4
MPEG - Motion Picture Experts Group
MSE - Mean Squared Error (střední kvadratická odchylka)
MSSIM - Mean Structural Similarity Index (celkový index strukturální podobnosti)
NR - No Reference
PSNR - Peak Signal to Noise Ratio (odstup signál šum)
R - Red
RGB - barevný prostor základních barev R, G, B
ROI - Region of Interest (oblast zájmu)
RR - Reduced Reference
SS - Single Stimulus
SSCQE - Single Stimulus Continuous Quality Evaluation
SSIM - Structural Similarity Index (index strukturální podobnosti)
SSM - Single Stimulus Method
SSMR - Single Stimulus with Multiple Repetitions
u - červeno-zelený kanál barevného prostoru CIE L^*u^*v
v - žluto-modrý kanál barevného prostoru CIE L^*u^*v
VDP - Visible Differences Predictor (predictor viditelných rozdílů)

XYZ - barevný prostor nereálných základních barev X, Y, Z
Y - Luminance (jas)
ZJ - znakový jazyk

Seznam použité literatury

- [1] WANG, Zhou, BOVIK, Alan C.: *Modern image quality assessment*. [s.l.] : Morgan & Claypool publishers, 2006, s. 18-26
- [2] NADENAU, Marcus.: *Integration of Human Vision Models into High Quality Image Compression*, These no. 2269, Ecole polytechnique federale de Lausanne, 2000, str. 9-29.
- [3] WANG, Zhou, SHEIKH, Hamid R., BOVIK, Alan C.: *Objective video quality assessment. The Handbook of Video Databases: Design and Applications*, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas, USA, 2003, kapitola 41, s. 1041-1078.
- [4] HECHT, Selig.: *The visual discrimination of intensity and the Weber-Fecher law*, Department of Physical Chemistry in the Laboratories of Physiology, Harvard medical School, Boston, 1924.
- [5] DALY, S.: *The visible difference predictor: An algorithm for the assessment of image fidelity*, Digital Images and Human Vision, ed. A.B. Watson, MIT press, Cambridge, MA, USA, 1993, s.179-206.
- [6] ITU-R Recommendation BT.500-11.: *Methodology for the Subjective Assesment of the Quality of the Television Pictures*, 2002, s. 48.
- [7] ITU-T Recommendation P.910.: *Subjective video quality assessment methods for multimedia applications*, 1999.
- [8] DUŠEK, Jaroslav.: *Objektivní hodnocení subjektivní kvality obrazu na základě modelu*, Disertační práce, Ústav radioelektroniky, ČVUT FEL, Praha, 2008.
- [9] WINKLER, Stefan.: *Digital video quality, Vision Models and Metrics*, London, Wiley, 2005, ISBN 0-47002404-6, s. 54.
- [10] ZACH, Ondřej.: *Měření kvality obrazu při příjmu analogového a digitálního televizního vysílání*, Bakalářská práce, Ústav radioelektroniky, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Brno, 2011, s. 57.
- [11] CIARAMELLO, Francis M., HEMAMI, Sheila S.: *'Can you see me now?' An objective metric for predicting intelligibility of compressed american sign language video*, Visual Communication Laboratory, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 2007.

- [12] WU, H.R., YUEN, M.: *A generalized block-edge impairment metric for video coding*. In: *IEEE Signal processing Letters*, 4(11):317-320, Nov. 1977.
- [13] WINKLER, Stefan.: *Perceptual Video Quality Metrics - A review*. In: *Digital video image quality and perceptual coding*, Boca Raton: Taylor, 2006, 640 s ISBN 0 82472777-0.
- [14] SLANINA, Martin.: *Methods and tools for image and video quality assesment*, Disertační práce, Ústav radioelektroniky, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Brno, 2008.
- [15] ZACH, Ondřej.: *Nástroje pro měření kvality videosekvencí bez reference*, Diplomová práce, Ústav radioelektroniky, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Brno, 2013.
- [16] LUBIN, Jeffrey.: *The use of psychological data and models in the analysis of display system performance*, In *Digital images and Human Vision*, ed. By A.B. Watson, MIT press, Cambridge, MA, USA, 1993, s. 163-178.
- [17] LUBIN, Jeffrey.: *A visual discrimination model for imaging system design and evaluation*, In *Vision models for target detection and recognition*, ed. By E.Peli, World Scientific, River Edge, NJ, USA, 1995, s. 245-283.
- [18] JACKSON, W. B., BEEBEE, P., JARED, D. A., BIEGELSEN, D.K., TRIMER, J. O., LUBIN, Jeffrey, GILLE, J.L.: *X-ray Image System Design Using a Human Visual Model*, Proc. SPIE, 2708 Medical Imaging, vol. 2706, 1995, s. 29-40.
- [19] ITU-R Recommendation BT.709.: *Parameter values for the HDTV standards for production and international programme exchange*, 2015.
- [20] DALY, S.: *Method and apparatus for determining visually perceptible differences between images*, USA, 1995.
- [21] DALY, S.: *A visual model for optimising the design of image processing algorithm*, In ICIP 94, Austin, Texas, USA, 1994.
- [22] WATSON, A. B.: *Perceptual optimization of DCT color quantization matrices*, in Proc. IEEE Int. Conf. Image Processing, Austin, TX, 1994.

- [23] ŠEVČÍK, Martin.: *Modelování vlastností modelu HVS v MATLABu*, Diplomová práce, Ústav radioelektroniky, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Brno, 2008, s. 20.
- [24] TEKTRONIX.: *A Guide to Picture Quality Measurement for Modern Television Systems*, Tektronix, 4/97 FL5372 25W-11419-0, 1997, s. 20.
- [25] LUBIN, Jeffrey.: *Sarnoff JND Vision Model*, Sarnoff Corporation,, Princeton, NJ, USA, 1998.
- [26] WANG, Zhou, BOVIK, Alan C., SHEIKH, Hamid R., SIMONCELLI, E. P.: *The SSIM Index for Image Quality Assessment*, New York Univ. & Univ. Of Texas, Austin, 2003.
- [27] WANG, Zhou, BOVIK, Alan C., SHEIKH, Hamid R., SIMONCELLI, E. P.: *Image Quality Assessment: From Error Measurement to Structural Similarity*, IEEE Trans. On Image Proc., Vol. 13, No. 4, 2004, s. 600-612.
- [28] ZATLOUKAL, Petr.: *Oblasti zájmu ve videosekvencích znakového jazyka*, Diplomová práce, Ústav radioelektroniky, ČVUT FEL, Praha, 2011.
- [29] UYVY Video Picture Encoding [online], dostupné z: <http://www.digitalpreservation.gov/formats/fdd/fdd000365.shtml>
- [30] ZATLOUKAL, Petr, BERNAS, Martin.: *Optimized H.264 compression of sign language video*, (při zpracování práce nebylo zatím publikováno)
- [31] The SSIM Index for Image Quality Assessment [online], dostupné z: <http://www.cns.nyu.edu/lcv/ssim/>
- [32] SSIM MATLAB [online], dostupné z: <http://www.mathworks.com/help/images/ref/ssim.html>
- [33] MONSPORTOVÁ, Jana.: *Aproximace funkcí – metoda nejmenších čtverců*, Bakalářská práce, Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno, 2013.
- [34] Weber-Fechnerův zákon [online], dostupné z: http://www.wikiskripta.eu/index.php/Weber-Fechner%C5%AFv_z%C3%A1kon
- [35] Pearsonův korelační koeficient [online], dostupné z: <http://www.mathworks.com/help/matlab/ref/corrcoef.html>

Seznam obsahu přiloženého DVD

Program	skripty, funkce a příslušné soubory použité při vytváření výsledků navržené metody
Testovaná videa	soubor všech použitých referenčních a testovaných videí
Text	text práce ve formátu .pdf